Physics

Teachers' guide Unit 8
Electromagnetic waves

Nuffield Advanced Science

copy 2

Physics Teachers' guide **Unit 8**

Electromagnetic waves

14 082718 8
Nuffield Advanced Science

Nuffield Advanced Physics team

Joint Organizers

Dr P. J. Black, Reader in Crystal Physics, University of Birmingham

Jon Ogborn, Worcester College of Education; formerly of Roan School, London SE3

Team members

W. Bolton, formerly of High Wycombe College of Technology and Art

R. W. Fairbrother, Centre for Science Education, Chelsea College, formerly of Hinckley Grammar School

G. E. Foxcroft, Rugby School

Martin Harrap, formerly of Whitgift School, Croydon

Dr John Harris, Centre for Science Education, Chelsea College, formerly of Harvard Project Physics

Dr A. L. Mansell, Centre for Science Education, Chelsea College, formerly of Hatfield College of Technology

A. W. Trotter, North London Science Centre; formerly of Elliott School, Putney

Evaluation

P. R. Lawton, Garnett College, London

Acknowledgements

The Organizers wish to express their gratitude to the many people who gave freely of their time in criticizing draft proposals for this Unit, and in stimulating new ideas, through a series of meetings, conversations, and correspondence.

Professor J. T. Allanson, W. K. Mace, Esq., Dr E. S. Shire, Dr R. W. Whitworth, formed a working party which met many times to discuss early drafts.

Professor R. G. Chambers and Dr I. Smith joined the working party at a later stage, when the drafts of the final materials were being considered.

Physics Teachers' guide Unit 8

Electromagnetic waves

Nuffield Advanced Science
Published for the Nuffield Foundation by Penguin Books

Penguin Books Ltd, Harmondsworth, Middlesex, England
Penguin Books Inc., 7110 Ambassador Road,
Baltimore, Md 21207, U.S.A.
Penguin Books Ltd, Ringwood, Victoria, Australia

Design and art direction by Ivan and Robin Dodd
Illustrations designed and produced by Penguin Education

Filmset in 'Monophoto' Univers
by Keyspools Ltd, Golborne, Lancs.
and made and printed in Great Britain
by C. Tinling & Co. Ltd, London and Prescot

Contents

Consultative Committee

Professor Sir Nevill Mott, F.R.S. (Chairman)
Professor Sir Ronald Nyholm, F.R.S. (Vice-Chairman)
Professor J. T. Allanson
Dr P. J. Black
N. Booth, H.M.I.
Dr C. C. Butler, F.R.S.
Professor E. H. Coulson
D. C. Firth
Dr J. R. Garrood
Dr A. D. C. Grassie
Professor H. F. Halliwell
Miss S. J. Hill
Miss D. M. Kett
Professor K. W. Keohane
Professor J. Lewis
J. L. Lewis
A. J. Mee
Professor D. J. Millen
J. M. Ogborn
E. Shire
Dr J. E. Spice
Dr P. Sykes
E. W. Tapper
C. L. Williams, H.M.I.

Foreword

It is almost a decade since the Trustees of the Nuffield Foundation decided to sponsor curriculum development programmes in science. Over the past few years a succession of materials and aids appropriate to teaching and learning over a wide variety of age and ability ranges has been published. We hope that they may have made a small contribution to the renewal of the science curriculum which is currently so evident in the schools.

The strength of the development has unquestionably lain in the most valuable part that has been played in the work by practising teachers and the guidance and help that have been received from the consultative committees to each Project.

The stage has now been reached for the publication of materials suitable for Advanced courses in the sciences. In many ways the task has been a more difficult one to accomplish. The sixth form has received more than its fair share of study in recent years and there is now an increasing acceptance that an attempt should be made to preserve breadth in studies in the 16–19 year age range. This is no easy task in a system which by virtue of its pattern of tertiary education requires standards for the sixth form which in many other countries might well be found in first year university courses.

Advanced courses are therefore at once both a difficult and an interesting venture. They have been designed to be of value to teacher and student, be they in sixth forms or other forms of education in a similar age range. Furthermore, it is expected that teachers in universities, polytechnics, and colleges of education may find some of the ideas of value in their own work.

If the Advanced Physics course meets with the success and appreciation I believe it deserves, it will be in no small measure due to a very large number of people, in the team so ably led by Jon Ogborn and Dr Paul Black, in the consultative committee, and in the schools in which trials have been held. The programme could not have been brought to a successful conclusion without their help and that of the examination boards, local authorities, the universities, and the professional associations of science teachers.

Finally, the Project materials could not have reached successful publication without the expert assistance that has been received from William Anderson and his editorial staff in the Nuffield Science Publications Unit and from the editorial and production teams of Penguin Education.

K. W. Keohane
Co-ordinator of the Nuffield Foundation Science Teaching Project

The Teachers' guide

This volume is intended to contain whatever information and ideas are required for the day to day teaching of the Unit. Not every teacher will need all of it all of the time; sometimes the summary and the list of experiments will come nearer to meeting the need.

The main text contains, on the righthand pages, a detailed suggested teaching sequence, which teachers can adopt or adapt. The facing lefthand pages carry practical details, suggested questions, references, and background information for teachers in the form of a commentary on the text. This commentary also indicates aims of the teaching, and points out links with other parts of the course.

At the end, there are some appendices containing material needed on occasion only, and lists of apparatus and teaching aids for the Unit. These include details of books and articles referred to in this *Guide*.

Introduction

This Unit is concerned with light as a wave motion. It represents the culmination, or end-point, of one line of thought in the course as a whole, for in it, earlier work on waves, on electric fields, and on magnetic fields come together in a (simplified) description of what an electromagnetic wave might be thought to be like. The velocity of light, hitherto merely an empirical constant, is related to two other constants, ε_0 and μ_0, taken from electricity. Ideas of fields and action at a distance form a theme which runs through the whole course, and in this Unit such ideas are seen to have a power far greater than might have appeared when they were first introduced. The 'calculating devices' B and E seem almost to acquire a life of their own, when they are seen to be able to propagate away into empty space, leaving behind the charges which produced them.

Too much should not be claimed, however, for the treatment of electromagnetic waves in Part Three of this Unit. Despite its importance as a high point in classical physics, the theory of electromagnetic waves receives here a very simplified treatment, intended to be introductory only. The ideas are very difficult, and we aim only to indicate to students that such a theory exists, with an outline of what it has to say about the nature of electromagnetic waves. One relatively easy example is chosen for discussion: waves travelling in the space between two long, parallel, flat conductors. This example is chosen, partly because it links well with earlier work on capacitors (Unit 3) and solenoids (Unit 7), but also because it avoids having to introduce Maxwell's 'displacement current'. The example is offered to students as an example of what can be done, and not as a complete or rigorous argument. It is simply asserted that a more general discussion leads to a similar description of an electromagnetic wave seen in terms of changes in a pair of mutually perpendicular E- and B-fields propagating at right angles to the directions of both of these fields.

There is a further, optional discussion of the behaviour of travelling fields which is intended to show that there could be reasons for thinking that the waves might propagate in empty space, as well as within a space between conductors on which there are charges and currents to guide them. But this discussion is liable to lead into quite deep matters (see Appendix B) and is not an essential part of the course.

We are not entirely satisfied with our solution to the problem of giving a good physical understanding of what an electromagnetic wave is like, and why people think it is like it is. We have found the subject especially hard to understand, and despite our efforts to pare away mathematical and conceptual difficulties so as to leave a simple, intelligible core of ideas, it is likely that the work will present difficulties to teachers as well as to students. For this reason, there are a number of notes on the commentary pages about difficult background theory, together with a long appendix (Appendix B) which tries to describe how the ideas offered in this Unit relate to those which are found in conventional University level texts. It is our hope that at least, these ideas may stimulate others to do better.

In this Unit the subjects 'physical optics' and 'electromagnetic waves' are combined. The link is strengthened by a discussion of polarization which, it is suggested, can come after the theoretical description of an electromagnetic wave which indicates that the waves are of a transverse variety. Polarization effects show how this feature is reflected in the actual behaviour of light and of radio waves. It was to make this unified treatment (with light and radio waves being treated almost casually as all of one sort) seem sensible, that Part One of Unit 4 spent time discussing the empirical evidence for the unity of the electromagnetic spectrum. In Unit 4 also, there was developed a theoretical account of the speed of propagation of a mechanical compression wave in a solid. One reason for this was to make it seem a sensible enterprise to try to explain the speed of the much more tenuous waves considered in Unit 8, again in terms only of fundamental principles. As it happens, there is a close analogy between the masses-and-springs wave model used in Unit 4 and the inductor–capacitor transmission line looked at in Unit 8, so the results of that earlier work have a direct as well as a general application.

The discussion of electromagnetic waves represents less than half the whole Unit. The first half contains, in Parts One and Two, a discussion of diffraction at an aperture and at a grating. Only Fraunhofer diffraction is discussed. So far as the course is concerned, besides developing ideas given more simply in Unit 4, this is part of the preparation for Unit 10. The essential idea is that, to calculate the light intensity at a place on a screen, one must add the amplitudes of waves coming from different places, taking account of their phases, and square the resultant. The treatment of the single slit is offered to students as an example of this general process, rather than as something one must know about for its own sake. Nevertheless, the matter is a useful one to understand, and it is used here to relate the problem of locating and resolving distant sources, to the discoveries of radio astronomy. The latter forms the subject of an article in the *Students' book*.

The arguments about diffraction suggested in the text use a minimum of mathematics, so that they may be accessible to the widest number of students within a modest time. It is possible to go further with students who would benefit from a more detailed and careful discussion. This can not only assist the understanding of those who have the necessary mathematical fluency (and it is fluency rather than knowledge which counts here) but can be of service to them in later learning. For this purpose, Appendix A contains a treatment based on rotating vectors or 'phasors', which may be of use to some teachers with some of their students. The phasor treatment will not, however, be regarded as necessary for the purpose of answering examination questions.

Gratings and spectra are dealt with briefly, mainly so as to summarize work from O-level and from Units 1 and 4, in readiness for the discussion of line spectra and of electron diffraction, in Unit 10. It is, of course, the existence of line spectra and their link with photons, which represents a kind of time-bomb ticking away under the point of view implicit in the rest of this Unit: that light is nothing else but an electromagnetic wave motion.

The study of the spectra of sources of electromagnetic waves helped finally to undermine the wave theory itself. But the theory also contained within itself the seeds of another radical revision of physical ideas: the theory of relativity. Relativity is too interesting to students to be passed over completely, but too difficult to be discussed in satisfying depth in the time available. We have proposed, in Part Four, an extremely limited and cursory discussion of one or two ideas (mainly time dilation), so as to suggest to students what kind of problem is involved. We hope that they will emerge from this knowing that the constancy of the velocity of light has led to a basic rethinking of fundamental concepts, and that they will be interested to learn more. We do not expect them to understand the matter at all well; indeed the subject may prove a good opportunity for the teacher to demonstrate his or her own lack of omniscience. We suggest this last point not only in joking fashion. It seems to us unfortunate that some students grow up with the belief that adult scientists really do understand everything in science. The fact is that not only does every individual not understand some things, but also that there are some things in science which nobody yet understands. As an 'end-point' in a school physics course ought, this Unit looks forward to further problems quite as much as it looks back over earlier work.

The Unit is based on the assumption that students have done some geometrical optics equivalent to that in Nuffield O-level Physics (Year III). If this is not so, the deficiency ought to be made good, with an appropriate sacrifice of material from Parts Three and Four.

Summary of Unit 8

Time: about 4 weeks; no more than 5.
(Numbers in brackets refer to suggested experiments, listed on page 7.)

Part One
Looking through holes

Time: less than 2 weeks.

This Part is concerned with diffraction at an aperture, as an example of the use of the general principle that the light reaching a screen is calculated by combining amplitudes. This principle is needed again in Unit 10. One example – the single slit – is discussed. The matter of diffraction is presented as a practical question, concerned with what the eye can see, and with what a radio telescope can detect. The work also provides a place for students to gain firsthand experience of diffraction, and to learn to distinguish it from other defects of images.

Suggested sequence

The limitations on what can be seen through a small aperture, including practical experience of difficulties of resolving sources (8.1). Spreading of waves going through a hole (8.2). Practical experience of the effect of diffracting screens on an image, including image defects not due to diffraction (8.3). Explaining diffraction at a slit. Diffraction of microwaves (8.4) and light (8.5) at a slit, examined more quantitatively. Radio astronomy. Overcoming limitations in direction finding; interferometers (8.6). Wave amplitude and intensity (8.7).

Part Two
Spectra

Time: two or three periods.

This Part revises ideas about diffraction gratings, and tries to set them in a general perspective, showing how the analysis of spectra can give useful information. In this context, 'spectroscopy' is seen not just as the analysis of visible light, but as the recording of the emission or absorption of radiation by matter at almost any frequency, from gamma rays to radio waves. In addition, there is an opportunity to indicate why the grating is a good instrument with a high resolving power, though this could be omitted in favour of revision of simpler ideas where necessary.

Suggested sequence

Information to be obtained from the study of the emission or absorption of radiation. The diffraction grating (8.8). $n\lambda = d \sin \theta$. The sharpness of the spectra. Looking at line spectra.

Part Three
Electric waves

Time: less than 2 weeks.

The aim of this Part is to describe electromagnetic waves, and to make that description seem reasonable by giving a simple example of how a theoretical argument about their propagation in one instance might be developed. So far as possible, the discussion is based on actual experiments. These should bring out the link between the propagation of electromagnetic waves, and the simpler steady field situations discussed in Units 3 and 7.

Suggested sequence

Radio sets: waves and aerials. A Hertz-type spark transmitter (8.9). The speed of electrical signals around a circuit (8.10). Reasons for a finite speed of propagation of a signal along a row of inductors and capacitors; a slow electrical wave (8.11). Analogy with masses-and-springs wave model in Unit 4. Waves between a long pair of flat, parallel, wide, conducting plates, as an example of a theoretical argument about electromagnetic waves. E-field and B-field (8.12) in such a situation. The velocity $1/\sqrt{\varepsilon_0\mu_0}$. Simple arguments about why waves near such conductors might be able to get out. Optional further arguments about moving fields (8.13), suggesting that electromagnetic waves might travel in empty space at the velocity $1/\sqrt{\varepsilon_0\mu_0}$. Polarization (8.14, 8.15).

Part Four
Relativity

Time: two or three periods.

A very brief introduction to the character of the problems and ideas of relativity, using time dilation as an example.

Suggested sequence

Material which may be drawn upon, as time allows, is as follows. The constant speed of light; reasons why it might be constant and evidence that it is (8.16). The principle of relativity, that the 'physics' does not depend on steady motion. A practical consequence of relativity: the over-long life time of π-mesons from an accelerator. Time dilation, from a consideration of radar methods of distance, velocity, and time measurement (Bondi's k-calculus). Doppler shift (8.17, 8.18). Clocks, forces and magnetism: a hint of how magnetic effects might be related to electric effects between moving charges.

Choosing one's own path

We hope and expect that teachers will find their own ways of using the material in this Unit. The detailed teaching programme laid out in the following pages represents as good a way of handling the material as we have been able to find in the light of experience in the trials, but should not be thought of as more than a possible, fairly well tested way of achieving the aims we decided upon. No doubt others can and will do better.

But teachers will know that it is the detail that counts in successful teaching, and so the *Guide* is full of particular teaching suggestions and practical details. We hope that these will help those who are uncertain how to handle either new material, or old material taught in a new way for unfamiliar aims.

The summary and list of experiments will, it is hoped, assist those who have taught the course a few times and no longer need to refer to all of the detailed teaching suggestions, as well as those who feel confident that they can make up their own teaching programme out of their previous experience. We also hope that the summary will provide an overall view of the work suggested. Such a view is necessary for keeping a sense of perspective and direction, both when one is immersed in particular detailed teaching suggestions and comments, and when students lead the teaching off in an unpredictable direction by contributing their own ideas.

It seems fair to add that the summary, taken on its own, could mislead. It cannot easily indicate the aims of pieces of work in any precise way, or find words to express the relative seriousness or lightness of particular episodes. Nor should a phrase one might find in a current examination syllabus always be taken here to imply the same work as it would imply there.

Experiments suggested for Unit 8

Part One

Looking through holes

Time: less than two weeks.

Introduction to diffraction at an aperture

Individual experiments in which diffraction effects are examined with some care (8.3) could be used as the starting point. We suggest starting from a rather general point of view, and with some simple experiments for the class as a whole, before getting down to detail, partly so that the point of the later work may be clearer. The later experiments use lenses for practical reasons, and might seem, wrongly, to be about lenses rather than about light going through holes. In any case, the general point of view is a valuable one, and an initial emphasis on sight may prove more interesting than something which seems, initially, to be a technical matter of the apertures in an optical system.

Suggestions from students that what they see has to do with 'waves' or 'interference' are welcome, but not yet essential. The main point is to use the experiments to suggest things that may be worth exploring by experiment and by theoretical argument.

Demonstrations and experiments

8.1 Looking through holes

1067/1 J	holder for two halves of a razor blade to be used as a single slit
1067/3 R	set of three colour filters (red, blue, green)
1053	aluminium foil
1053	cardboard 35 mm slide mounts
1054	copper wire, 36 s.w.g., bare or enamelled
94 A	lamp, holder, and stand
27	transformer
1063	multiple light source (one for the whole class)
1053	card
116	plane mirror
1053	transparent ruler with millimetre graduations

8.1a Looking at a lamp through a slit and through a pin-hole

One or more lamps with straight vertical filaments can be set up around the room. They should be viewed from a distance of about three metres.

The razor blades are adjusted by the student so that the edges form as narrow a parallel-sided slit as possible, as judged by looking at the slit against an illuminated background such as a window. This takes time, but is good practice for later work. The tungsten lamp is then viewed with the slit held close to the eye. If necessary, the slit should be narrowed until the light from the filament is seen to spread widely. The filters (red, green, and blue) should be placed in turn in front of the lamp or slit so that students may see and compare the diffraction patterns. After the slits are adjusted, the room needs to be darkened.

The copper wire is used to prick a small hole in the aluminium foil, which may be mounted in slide mounts. First the wire is stretched until it breaks and then, with one of the pieces held between the fingers a few millimetres from the break, the point is pressed slowly but firmly through the foil. With the hole held close to the eye, the lamp filament should be viewed from a distance of 3 m or more.

Detecting wave energy through apertures

Why does a person's eye have an aperture of variable size? One obvious answer is concerned with controlling the light energy entering the eye. Why are radio telescopes often very large? It seems likely that they need to be, so as to collect enough energy from weak or distant sources.

This Part is about a different aspect of the detection of wave energy, the *diffraction* effects which inevitably accompany the passage of a wave through a restricted aperture. Everything we see is seen through a hole; the few millimetres' width of the eye pupil. Should the pupil be regarded as a 'large' or as a 'small' hole? The answer must be another question: large or small compared with what? As far as diffraction effects are concerned, the standard of comparison is the wavelength of light. Some of the implications of having to look through a hole can be gained by looking through a narrower hole. Doing this should suggest other, more careful experiments to try.

Demonstrations and experiments

8.1 Looking through holes

The following experiments are about some aspects of the difference it makes when one looks at the world through a small hole.

8.1a Looking at a lamp through a slit and through a pin-hole

The class can be invited to look at a distant lamp filament through slits which they can make narrow and parallel-sided, and through a small hole pricked in metal foil. Filters can be used to vary the colour of the light seen.

In discussion students should be able to report that when they look through a small hole they see a pattern consisting of light and dark regions (or colours, when no filters are used); that the shape of the pattern depends on the shape of the hole, that the smaller the hole the wider the pattern is; that long wavelength (red) light gives a broader pattern than short wavelength (blue) light.

They should all see the effect of a narrow, parallel-sided slit, and of a small hole, in red and in blue light. Some may like to try more than one hole, or may try the effect of a V-shaped slit, inadvertently or on purpose.

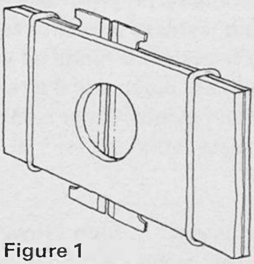


Figure 1
Holder for two halves of a razor blade to be used as a single slit (item 1067/1 J).

8.1b Distinguishing lamps as separate

The multiple light source (figure 2) is set up with a green filter in front of it and in such a position that students can all be more or less in front of it.

Viewing the lamps through a narrow parallel slit, each student should adjust first the slit width and then his viewing distance until the lamps can just (but only just) be seen not to be a continuous strip of light. When everyone has found this position, the green filter should be replaced by the blue filter, and when all have seen that the lamps can be distinguished, the blue filter should be replaced by the red filter. The lamps cannot then be resolved.

Resolving power

Using several lamps gives a sharper change from resolution to non-resolution than using just two, so that the change of resolving power with wavelength can be seen more clearly. But the presence of several lamps means that Rayleigh's criterion is not strictly applicable.

8.1c Resolving detail with the eye

Two black lines are drawn on card, parallel to one another and 2 mm apart. Students are to measure the greatest distance at which these two lines can be seen as separate. The paper on which the lines are drawn should be well illuminated.

The diameter of the eye pupil is worth measuring. This can be done by holding a ruler to the eye and looking at it in a mirror, placed beside the illuminant to make the pupil close.

Students' book

Questions 1 to 7 are general questions about waves and diffraction.

Two-source superposition effects

It may be necessary, if students are in danger of confusing the diffraction effects seen in experiment 8.1 with interference effects seen previously in Unit 4, to add a further experiment with two closely spaced holes (or slits) so that they may see the extra interference fringes 'on top of' the broadening of the source by one slit or hole.

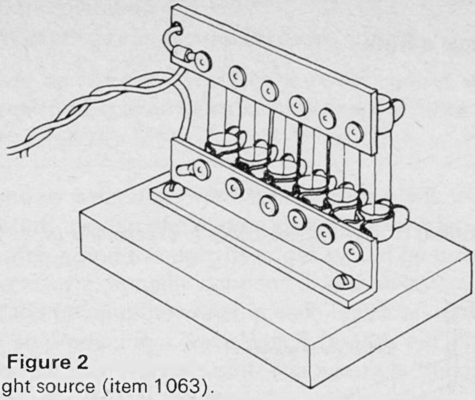


Figure 2
Multiple light source (item 1063).

8.1b Distinguishing lamps as separate

This experiment shows that what we see depends on the size of the hole we look through, and on the colour of the light used.

Let students try looking at the multiple light source (covered by a green filter) through narrow slits. When they have found that through a narrow slit, or at a large distance, the five lamps cannot be seen separately, the effect of the colour of the light can be demonstrated. If the slit width and viewing distance are made such that the lamps viewed through a green filter can just, but only just, be distinguished from a continuous strip of light, then, viewed through a blue filter they can be clearly resolved, but not if seen through a red one. It should seem possible to the class that this is an effect of a change of wavelength.

8.1c Resolving detail with the eye

What would the world look like if one could only ever look at it through a narrow hole or slit? (Dim and blurred.) But do we ever see anything except through a small hole? (No, since the pupils in our eyes are small holes a few millimetres across.) It is at least possible that the size of this hole is important in deciding how much detail an eye can distinguish.

Two clear black lines on a card, about 2 mm apart, make a suitable object to look at. Students can find the greatest distance at which they can just distinguish the lines, which may be about 5 metres. If the lines are closer, the distance is smaller. It should be clear that it is the angle the lines subtend at the eye which matters: the smallest angle for detail to be resolved by the eye thus seems to be of the order $2 \times 10^{-3}/5$, or some 4×10^{-4} radian. Later experiments and arguments will try to explain the size of this angle.

It may be worth noting, even at this stage, that the ratio of the wavelength of visible light to the pupil diameter is also a few times 10^{-4}.

Demonstration

8.2 Water waves going through a hole

90 ripple tank kit

47 illuminant

1033 cell holder with two U2 cells

541/1 rheostat (10 to 15 Ω)

For details of the use of the ripple tank, see Nuffield O-level Physics, *Guide to experiments III*, experiment 4 (particularly 4r and 4s).

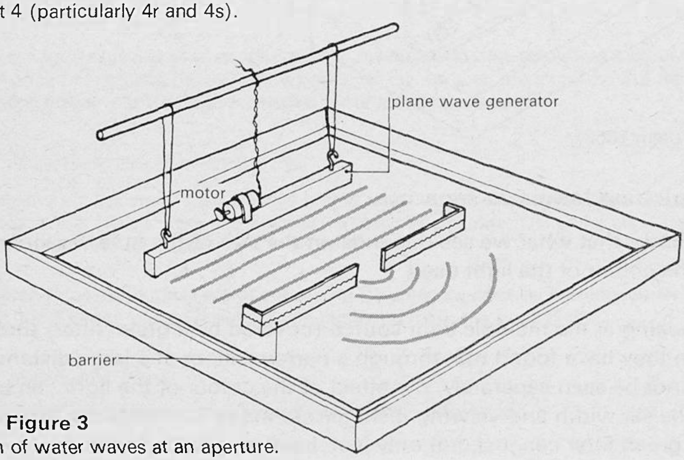


Figure 3
Diffraction of water waves at an aperture.

A group of experiments on diffraction and lens defects

We suggest that the experiments 8.3a to h are best shared out among the class, so that each student investigates one or two aspects of the group of problems, being allowed to take the initiative as much as he or she wishes. Little that can be done with the apparatus is without value.

Of the experiments suggested, a, b, and c are not essential later, and nor is h. The first three are about defects of lenses, while the last is about the refractive index of air. All four could be shown as quick demonstrations in a small class with too few students for all the experiments. Experiments d, e, and f are the vital ones for students to do themselves.

After a double period of experimenting, students can report what they saw and did. There will be opportunities later to repair any omissions. The important thing here is to provide an opportunity for firsthand experience. A second session, in which students set up an experiment they did not do before and show it to the rest, may be useful.

Each experiment uses a lamp, a lens, and a screen (or eyepiece) arranged in much the same way. The group of experiments can be started off with a demonstration, which at least serves to let students see where, and how far apart, the parts are put. So far as possible, the kit of parts is arranged so that all the components come naturally at the same height above the bench. Even so, a word about alignment will be needed.

Demonstration

8.2 Water waves going through a hole

Water waves can be seen, while light waves cannot, and the reasons why water waves are diffracted at an aperture may seem simpler to students than reasons why light is diffracted.

Figure 4 shows what does *not* happen when plane waves pass through a gap. Before going through the gap, the middle strip of wave had water at its own height on either side of it; after going through, its edges would be unsupported if what happens in figure 4 were what actually happens. But as everyone knows, an isolated hill of water will not stand up, and the water in each crest can now collapse outwards and sideways, so producing a wave which spreads out. This is not the whole story, but it is an important practical part of it. The narrower the gap and the longer the wavelength, the greater the spreading of the wave.

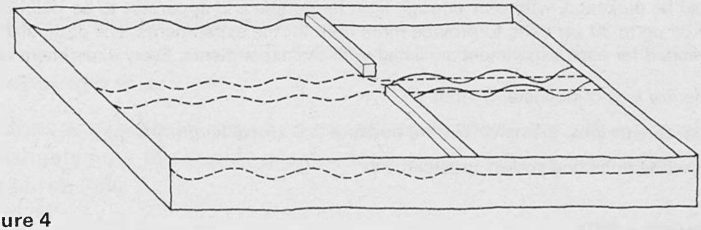


Figure 4
Imaginary water waves not diffracted at a gap.

Diffraction effects and other effects

It may seem plausible to suppose that the effects seen in experiment 8.1 are wave effects, but what is seen is not simple, and wave effects may only be a part of the story. The eye is a complex system, with defects such as lack of the proper curvature of the retina, which might explain its lack of resolution. Another explanation of the same thing might have to do with the finite number of light receptors in the retina, and it is possible that the brain could enable one to 'see' detail not clearly present in one retinal image, by storing information from successive images, and comparing the different sets of information.

Not every blurred defect in what one sees is bound to be a diffraction effect. The eye contains a lens, and no lens produces a perfectly sharp image. The next group of experiments offers a chance to disentangle some of the many things that spoil the perfection of an optical system, and, having distinguished the wave-like diffraction effects from others, to experiment with them in greater detail.

The various effects can be shown, at least in part, in the introductory demonstration. For example, experiment d could be introduced somewhat in this style.

> 'I have put a sharp image of the lamp filament on the card. Now if I mask the lens with this slit and narrow it, you all see the image gets dimmer, and do you see what happens just before it disappears altogether? (It widens.) You may be able to measure something, or explain it, or see a difference for different colours. You may want to use a narrower source, like this slit propped against the lamp, and a translucent screen, or an eyepiece like this when it gets very dim.'

Group of experiments

8.3 Effects of optical systems on light waves

1067 physical optics kit

94 A lamp, holder, and stand

27 transformer

The room should be darkened, with only enough light for people and apparatus to be visible. The lamps can be overrun by up to 30 per cent, to provide more light for the experiments. The particular parts from the optics kit needed for each experiment are listed with the experiments. Every experiment uses:

1067/1 B holder for lens of diameter 37 mm

1067/1 C plano-convex lens, diameter 37 mm, power +2 D (focal length 0.5 m)

1067/3 O matt white reflecting screen (postcard)

8.3a Stopping a lens

Extra items from the physical optics kit:

1067/1 E set of stops for lens in holder

1067/1 H small translucent screen

The lamp and screen should be about 3 m apart, with the lens positioned to produce a magnified image on the screen. The screen should be moved to find the best image for each stop in turn. The translucent screen is easier to use than the opaque screen.

Difficulties which arise may be due to:

1. The plane of the lens not being perpendicular to the lamp-screen axis.
2. The stops not being placed so as to be coaxial with the lens aperture.
3. The screen position not being adjusted to find the best image.

8.3b Change of image distance with zone of lens used

Extra items from the physical optics kit:

1067/1 D plano-convex lens, diameter 37 mm, power +7 D (focal length 0.15 m)

1067/1 F set of masks with holes at different zone radii

1067/1 H small translucent screen

The effects are more easily observed with the stronger lens.

With the stronger lens, lamp and screen should be about 2 m apart with the lens positioned to produce a magnified image of the filament. The mask with the largest circle of holes is then placed against the lens and the best image position found. The other masks produce clear images in different places. The translucent screen is easier to use than the opaque one.

Group of experiments
8.3 Effects of optical systems on light waves

In each experiment, students inspect the real image of a lamp filament or of a slit in front of the lamp, formed on a screen by a lens. Stops, slits, or screens of various kinds are placed against the lens. The screen is sometimes replaced by an eyepiece in the same place.

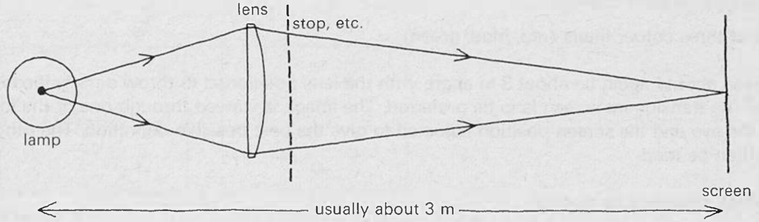


Figure 5
The basic optical system used in experiment 8.3.

8.3a Stopping a lens

A series of stops with holes of varying sizes, all too big to show diffraction effects, are used to investigate how the quality of the image varies with the size of the unmasked central area of the lens.

Points which may emerge are:

1. The image is dim with small stops.
2. The depth of focus is greater with small stops.
3. The image is clearer with small stops.

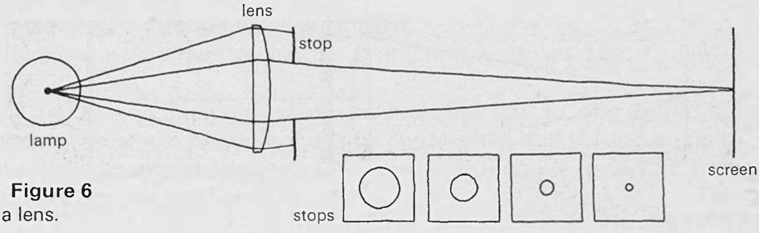


Figure 6
Stopping a lens.

8.3b Change of image distance with zone of lens used

Lens masks are used in which circles of holes have been cut. The total open area of all the masks is about the same, but the exposed area lies at different distances from the centre of the mask. The sharpest image is formed at a distance from the lens which depends on the zone radius.

Difficulties which arise may be due to:

1 The plane of the lens not being perpendicular to the lamp-screen axis.
2 The circle of holes not being coaxial with the lens aperture.

8.3c **Change of image distance with colour**

Extra items from the physical optics kit:

1067/1 H small translucent screen

1067/3 R set of three colour filters (red, blue, green)

Lamp and screen should again be about 3 m apart, with the lens positioned to throw a magnified image on the screen. The translucent screen is to be preferred. The image is viewed through one of the filters held close to the eye and the screen position adjusted to give the best possible definition. The other filters should then be tried.

Difficulties which arise may be due to:

1 The plane of the lens not being perpendicular to the lamp-screen axis.
2 Dirty, fingerprinted colour filters.

8.3d **Masking the lens with an adjustable slit**

Extra items from the physical optics kit:

1067/1 A sheet with slit and holes

1067/1 G big stop to stand on bench

1067/1 H small translucent screen

1067/1 I holder for eyepiece or adjustable slit 2

1067/2 L eyepiece

1067/2 M slit of variable width

1067/3 R set of three colour filters (red, blue, green)

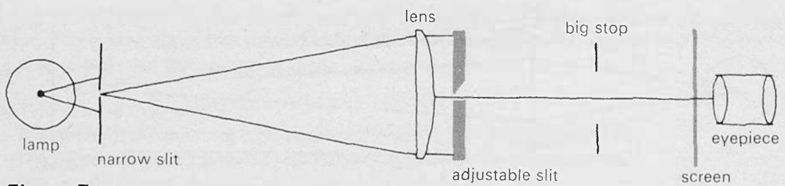


Figure 7
Diffraction by a single slit.

The lamp and screen should be just over 2 m apart with the lens midway, forming an image of the filament on the screen. The adjustable slit in its holder should be placed close to the lens with its length parallel to the filament. As the slit is narrowed, the image dims and then widens. The mounted eyepiece can now be substituted for the screen. The light must fall centrally on the eyepiece, and it will usually be necessary to adjust the height of the lamp to achieve this.

The big stop is provided as a means of cutting out light which has not gone through the slit. Its best place must be found by trial and error. It will be somewhere between the slit and the eyepiece.

A narrow slit propped against the lamp and parallel to the adjustable slit will improve the visibility of the pattern. Fringes other than the central maximum should be visible, especially in coloured light. The screen or eyepiece will need to be moved when the extra slit is put over the lamp, so that the image of the slit is in focus, rather than the filament.

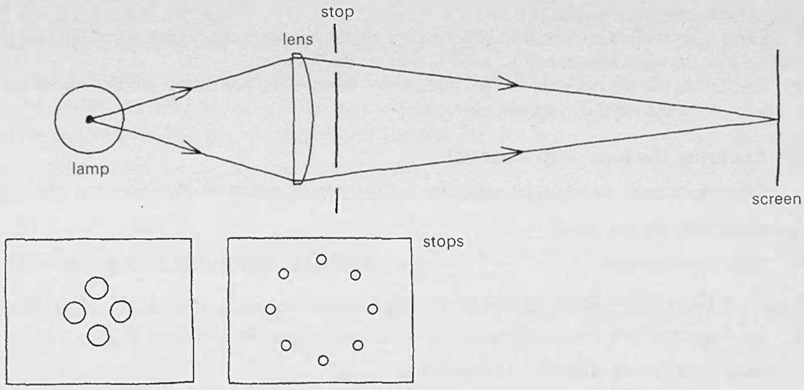


Figure 8
Change of image distance with zone radius.

Points which may emerge are:

1 *The image is nearer to the lens when the edges of the lens are used.*

2 A plano-convex lens gives greater variation of image distance when its curved face is towards the lamp.

8.3c Change of image distance with colour

The magnified image of a white hot filament tends to have coloured edges, and the experiment helps to show why, and to distinguish the effect from other coloured fringes due to diffraction, seen elsewhere.

Points which may emerge are:

1 Images seen through filters are clearer.

2 Red, green, and blue images are at different distances from the lens.

In principle it makes no difference where the filters are put, so long as the light goes through them somewhere between the lamp and the eye. But the experiment is easier if the filter is put between the screen and the eye. Students will probably follow the lead if they see the teacher do it.

8.3d Masking the lens with an adjustable slit

The lens forms an image of the filament; object and image are roughly equidistant from the lens. An easily adjustable slit is placed with its length parallel to the length of the lamp filament, close to the lens. As the slit is slowly closed, the image becomes dimmer, but before it quite vanishes it becomes wider again. This experiment is important, and to be sure that all students see the problem during the initial review of the experiments the lamp may be over-run. The students who investigate should be given another narrow slit to put in front of the lamp, so as to make a narrower source, and an eyepiece to examine the effect in more detail. They may also have colour filters.

Difficulties which arise may be due to:

1 Light from the source slit not falling on the adjustable slit – the source slit should be positioned so that the light from it can be seen falling on the holders.

2 Light from the slit not entering the eyepiece – remove the adjustable slit and check by holding a piece of card over the eyepiece aperture.

8.3e Covering the lens with material

Extra items from the physical optics kit:

1067/1A	sheet with slit and holes
1067/1H	translucent screen
1067/3R	set of three colour filters (red, blue, green)
1067/3S	small piece of fine black chiffon
	other finely woven fabrics (optional)
	Nuffield O-level Chemistry diffraction grids (optional)
	diffraction grid from Unit 10 (see below, optional)

Lamp and screen should be just over 2 m apart with the lens about midway between them producing an image of the filament on the screen. The materials available are placed close to the lens so that the light passes through them. Prop the sheet with slit and holes against the lamp so that light coming through a hole reaches the screen, adjusting the position of the hole for the best illumination. The patterns should also be viewed through the colour filters, using the translucent screen.

A diffraction grid which will also be useful in Unit 10 can be made by photographing figure 25 from that Unit. Its use now could be helpful later on when Unit 10 is taught.

Fourier transforms

It should become clear that the pattern of brightness and darkness on the screen is related to, but is not the same as, the pattern of openings and opacities in the diffracting grid. It may interest some students to know that there is a general functional relationship between the two, valid for every type of diffracting screen. One pattern is the Fourier transform of the other (when the light going through the diffracting screen is parallel). Some of the properties of the transform can be studied empirically by using a variety of screens, and by deforming some of them.

8.3f Coarse line gratings

Extra items from the physical optics kit:

1067/1 A	sheet with slit and holes
1067/1 G	big stop to stand on bench
1067/1 H	small translucent screen
1067/1 I	holder for eyepiece
1067/2 L	eyepiece
1067/2 N	set of coarse gratings
1067/3 R	set of three colour filters (red, blue, green)

Lamp and screen need to be just over 2 m apart with the lens about midway forming an image of the filament on the screen. The gratings should be placed close to the lens. A hole or slit in front of the lamp makes the effects clearer and the translucent screen may then be used. The effects should also be observed with the eyepiece. Care is necessary to ensure that the lamp height is such that the diffracted light enters the eyepiece. Any extraneous light should be reduced as much as possible by positioning the large stop between the grating and the eyepiece.

Points which may emerge are:

1 The narrower the slit, the wider the light spreads.
2 The narrower the slit, the dimmer the illumination is.
3 The broad spread of light has narrower bands of light on either side.
4 The pattern is smaller for blue than for red light.

Measurements are difficult to make and should not be expected.

8.3e Covering the lens with material

The material forms, in effect, a two-dimensional diffraction grating, and the screen shows a considerable number of bright spots or bands.

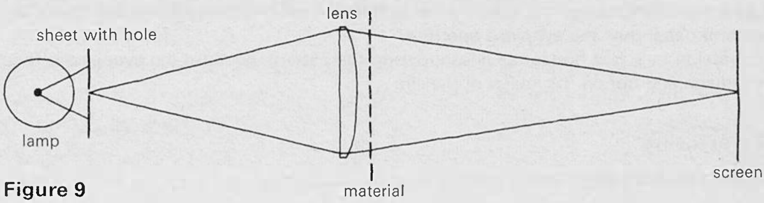


Figure 9
Covering the lens with material.

Students can also hold the material close to their eyes and look at a filament lamp through it. Each student can be given a few square inches of material to take away so that he can look through it at street lamps, traffic lights, car lights, etc., out of school.

Points which may emerge are:

1 The pattern on the screen is *not* identical with the pattern of holes in the material.

2 A more closely woven or knitted material produces broader spaced diffraction patterns.

3 Rotating the material about the lens's axis rotates the diffraction pattern.

4 If the light passes obliquely through the material it increases the spacing of the diffraction pattern in one direction.

5 Stretching the material in one direction narrows the diffraction pattern in that direction.

8.3f Coarse line gratings

Gratings coarse enough for the lines to be visible, but fine enough to give spectra on a small scale, are provided as part of the optics kit.

The lens forms an image of the filament; object and image are roughly equidistant from the lens. Gratings of various spacings are put close and parallel to the lens. A hole or slit in front of the lamp may make the effects clearer. A ground glass screen and colour filter may be used.

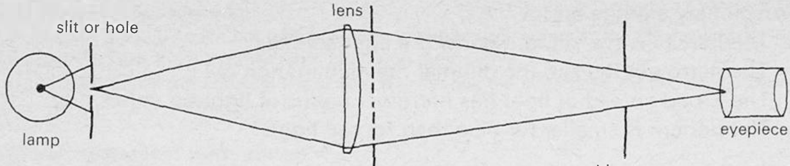


Figure 10
Use of coarse gratings.

The patterns should be viewed with the coloured filters as well as with white light.

Difficulties which arise may be due to:

1 Light from the source slit or hole not falling on the lens and grating – the source slit should be wide enough for this light to be visible where it falls on the holders.

2 Light from the gratings not entering the eyepiece – remove the gratings and check by holding a piece of paper over the eyepiece aperture.

3 Confusion over fine horizontal lines crossing the pattern, because the eyepiece is focused on the filament's image and not on the image of the slit.

8.3g Several slits

Extra items from the physical optics kit:

1067/1 A	sheet with slit and holes
1067/1 G	big stop to stand on bench
1067/1 I	holder for eyepiece
1067/1 K	support for a set of slits
1067/2 L	eyepiece
1067/2 O	set of 1, 2, 3, 4, 5, and 6 parallel slits of pitch about 0.25 mm
1067/3 R	set of three colour filters (red, blue, green)

With lamp and screen separated by just over 2 m, the lens is positioned about midway to form an image of the filament on the screen. The slit support is placed adjacent to and on the screen side of the lens so that it is coaxial with the lens, and the mounted eyepiece replaces the screen. The slits go in the support, parallel to the lamp filament. Care is necessary to ensure that the light enters the eyepiece through the centre of the lens and some adjustment to the height of the lamp may be necessary.

With, say, the double slits in place, fringes should be observed in the eyepiece. Any extraneous light should be reduced as much as possible by careful positioning of the large stop between the slits and the eyepiece. A slit in front of the lamp will improve the fringes, but it must be parallel to the slits near the lens. The pattern from each set of slits should be examined in white light, and with the coloured filters.

Difficulties which arise may be due to:

1 Light from the source slit not falling on the lens and multiple slits – the source slit should be wide enough for this light to be visible on the holders.

2 The source slit and diffracting slits not being approximately parallel.

3 Light from the source slit not entering the eyepiece – remove the diffracting slits and check by holding a piece of paper over the eyepiece aperture.

4 The eyepiece not being near the image of the source slit.

5 Confusion over horizontal lines crossing the image, because the eyepiece is seeing the filament image, not the slit image.

Points which may emerge are:

1 Finer gratings give broader patterns.

2 The light is spread out perpendicularly to the lines of the grating.

3 Fairly pure spectra are produced.

4 The lens's original image does not move when the grating is rotated, though the diffraction pattern rotates.

Optionally, two gratings can be used together, with their rulings crossed.

8.3g **Several slits**

The set of 1, 2, 3, 4, 5, and 6 slits bridges the gap experimentally between the diffraction pattern of a single slit, and the spectra produced by a grating.

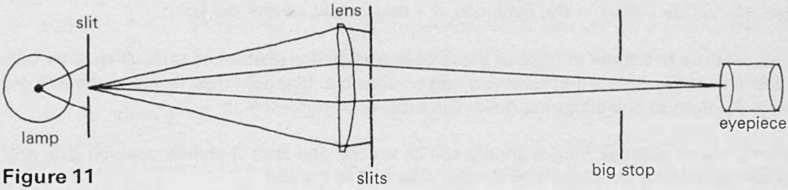


Figure 11
Diffraction by several slits.

As in 8.3f, the lens forms an image of the filament, the light passing through slits close to the lens. An eyepiece is needed to see the faint pattern produced by only a few slits. A narrow slit in front of the lamp helps a good deal to improve the clarity of the fringes.

Points which may emerge are:

1 The single slit gives a wide spread of light.

2 The pair of slits gives Young's fringes within the wide spread.

3 Greater numbers of slits give brighter patterns.

4 The greater the number of slits, the narrower the bright fringes, but their spacing is unaltered, and the pattern approaches that of a coarse grating.

5 With a few slits, the pattern is essentially the two-slit one, with some extra dark or bright lines within it.

Discussion of the experiment, which may also recall earlier work on wave superposition in Unit 4, can begin to clarify the relation between interference and diffraction.

8.3h **Removing air from the path of one of two interfering beams**

Extra items from the physical optics kit:

1067/1 A sheet with slit and holes

1067/1 G big stop to stand on bench

1067/1 I holder for eyepiece

1067/2 L eyepiece

1067/2 P air pressure cell for interference

1067/3 R set of three colour filters

1067/3 T card with slits for air pressure cell (see below)

Figure 13 shows the arrangement of the apparatus. The screen is put 3 m from the lamp, with the lens forming a *diminished* image of the filament on it, and the eyepiece is then put in place of the screen so that light enters it centrally. The double slit is then propped against the lens, on the lamp side. Fine fringes should be visible in the eyepiece, if a narrow slit covers the lamp.

Now the air pressure cell is put in front of the double slits, and is positioned until its shadow passes between the slits and is as fine and narrow a ring as possible. It should then lie along the axis of the system, with the light to one slit going down the tube, as in figure 14.

If the tube is properly aligned, fringes should still be visible, and should shift in position if air is drawn out of the cell by sucking on a long rubber tube attached to the cell.

If fringes cannot be seen with the tube in place, the slit separation may be too small. The slit spacing must be a compromise between allowing light to go through and alongside the cell tube, and giving visible fringes.

Construction of suitable double slits

Figure 12 shows a construction which has proved successful. A 6 mm × 4 mm rectangle is cut out of the centre of a card with a razor blade. Across the middle of this hole is stuck a 1.5 mm wide strip of sticky paper label, cut with clean parallel edges. Two more straight edged strips of label are stuck on either side of the first, leaving slits about 0.2 mm wide. It is best to do this on a dark surface, judging the equality of the slits by eye.

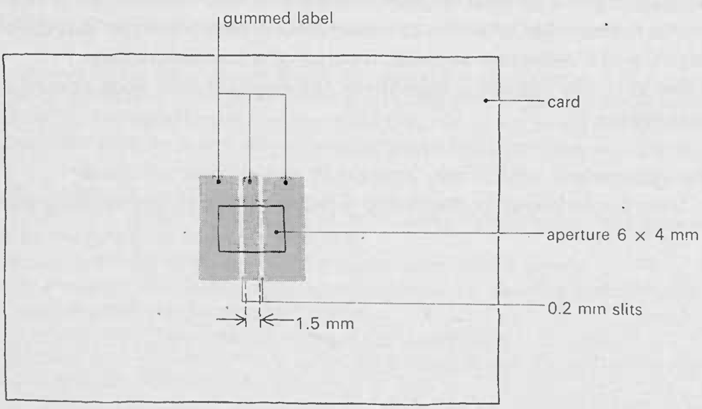


Figure 12
Construction of double slits.

8.3h Removing air from the path of one of two interfering beams

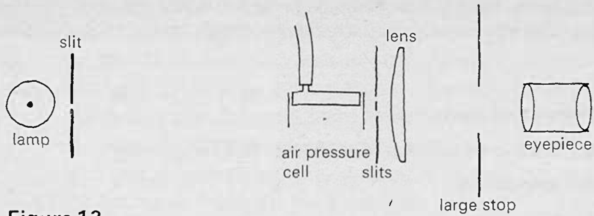


Figure 13
Air pressure cell used to remove air from the path of one of two interfering beams.

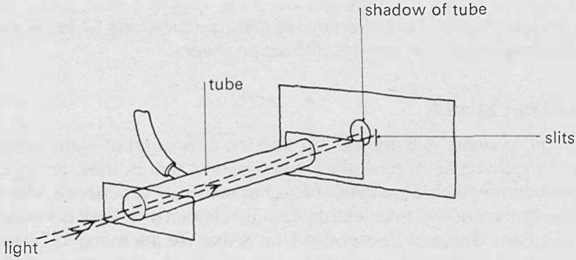


Figure 14
Alignment of air pressure cell and slits.

The experiment shows that the interference pattern shifts when air is withdrawn from one of the paths of two interfering beams of light. It has interest as showing the importance of refractive index, in that if one beam has traversed less air than another, they arrive out of step, having a small time difference, despite having travelled the same length of path. It would seem that a view which suggests that light goes at a different speed through materials than it does through a vacuum would be reasonable. Incidentally, the experiment points up the great sensitivity of interference methods. (A shift of one fringe represents a time difference of one cycle, about 10^{-15} s.)

It is worth mentioning the quantity known as the refractive index, which may be understood as the ratio of the wave velocity in a vacuum to that in a material, and to refer back to the bending of light by glass lenses in earlier experiments.

Explaining the single-slit diffraction pattern

The outcome of experiments 8.1 and 8.3 can best be summarized by developing an argument for the shape of the single-slit diffraction pattern.

When a plane wave (wave fronts straight and parallel) passes through a slit it spreads out. The explanation of the bright and dark bands depends on this effect, which can be seen clearly with water waves (experiment 8.2).

Demonstrations of diffraction

It may help the theoretical discussion to have a demonstration of what is being talked about, to add to previous experience. Experiment 8.3d may serve, or better, experiment 8.5. Best of all is a demonstration with a laser.

Photographs of diffraction patterns

Photographs of diffraction patterns are to be found in many books. See particularly:

PSSC *Physics* (second edition), Chapter 18.

PSSC *College Physics*, Chapter 9.

Students' book

Questions 8 to 13 are mainly for revision, concerning interference effects. Questions 14 to 24 are about diffraction. Question 18 covers the simpler parts of one-slit diffraction theory.

Why do waves spread out at all?

Everyone knows that an isolated 'hill' of water will not stay up, and the diffraction of water waves is an obvious consequence. Electric and magnetic fields have the same property, that is, they do not stay where they are put without external contrivance. However, this is not everyday experience, and cannot be pictured. Moreover, at this stage in the course, the electromagnetic character of light is hearsay. So the teaching standpoint throughout Parts One and Two of this Unit is that we are trying to explain the diffraction of light as if light were a wave motion, thinking first of water waves when we try to think what wave behaviour is. In passing we also look at the diffraction of any other waves which seem interesting, or contribute to the appearance of unity of the electromagnetic spectrum.

Phasor treatment of diffraction

Appendix A outlines a more mathematical treatment of diffraction than that suggested in the text. This treatment, based on phasors, has advantages for students who would enjoy a more rigorous and complete discussion, and would profit from it. But it is not a necessary part of the course.

Judging a theoretical method by results

As well as explaining an observed result, this piece of work can be used to introduce some discussion of the way theoretical procedures in physics are often justified because they seem to work rather than because they are known beforehand to be right.

Some students may feel that they cannot justify the division of the slit into strips or the simple addition of the contributions from the strips. They may feel that light from two strips could combine in some much more complicated way for all they know about the real nature of light. It is probably unwise for teachers to put such doubts into the minds of students who do not have them spontaneously; however, such doubts are an important part of the best students' process of understanding. The reason for using this artificial process is not because we know at first that it will be correct, but because it gives predictions that can be tested against measurements we can make of the light intensities in various directions. The process we use gives the right results, and this is our reason for thinking it may have sense.

Plane waves in a ripple tank spread out and the wave fronts become nearly circular after passing through a narrow hole. As was suggested previously (figure 4), the reason why water waves spread out is not hard to understand. With light, the effect is much the same, though it is harder to explain. Although intensities of visible light are too small for the light to be seen after passing through a really narrow slit, a very considerable spreading can be seen from a slightly wider one. Careful measurements show that there is less intensity at big angles, but in an A-level course it is best to admit to students that considering this reduction would make the problem too complicated for the time that can be afforded.

Why, if light waves spread out like this, are there bright and dark bands on the screen? Why isn't there just a bright area on the screen, wider than it would be if the waves passed through without spreading?

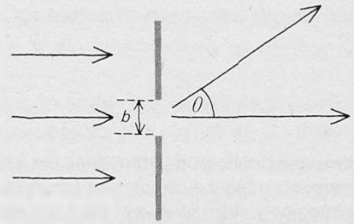


Figure 15

Suppose a parallel beam of light, with plane wave fronts, comes to a narrow slit of width b in an opaque screen which is parallel to the wave fronts (figure 15). The light which goes through the slit does not all continue in its original direction, but some of it goes in other directions. How much goes at angle θ to the original direction? This is the problem to be argued about in what follows.

Angles at which there is no light

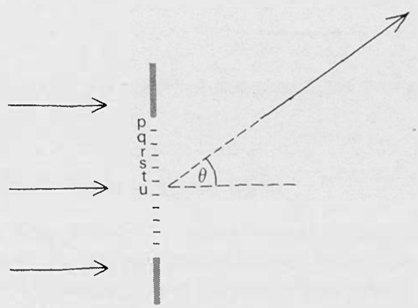
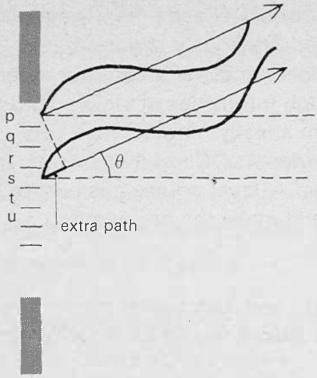
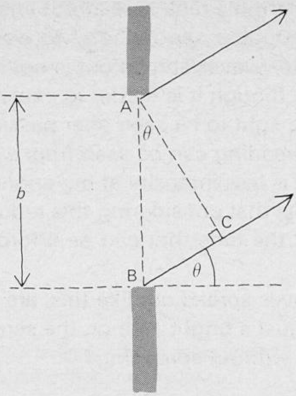


Figure 16

From a wider slit (figure 16) the light arriving at a distant place at an angle θ can be thought of as the total of the contributions from many very narrow strips or slits, p, q, r, s, etc. The width of the whole slit is the sum of the widths of all the slits





Figures 17 and 18

An incomplete argument

The argument is approximate and incomplete and conceals the difference between intensity and amplitude. But complete mathematical treatments merely evaluate more accurately the total amplitude from the strips of slit. Dividing the slit into imaginary strips is always the first step. The discussion is also limited to plane wave fronts; that is, only Fraunhofer diffraction is being considered. The general case, Fresnel diffraction, is similar, but harder still.

The argument is also two-dimensional, not three-dimensional, as it ought to be if a real wave front were under discussion.

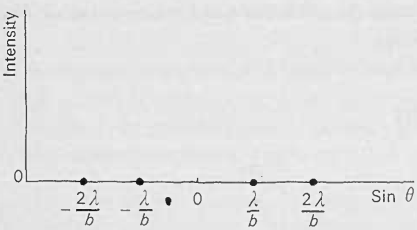


Figure 19

p, q, r, s, etc. These very narrow slits are imaginary in the sense that nothing intervenes between them. Clearly, then, it does not matter how large a number of them we imagine or how little light comes from each, provided that each is narrow enough to emit light nearly equally in all directions. If we imagine the slit to be divided into twice as many slits, each will emit half as much light and the total quantity of light will be unchanged.

Why doesn't the whole slit emit light equally in all directions in the same way that we imagine the individual slits doing so? Because for a particular angle θ, the light from such a slit as s will have to travel further than light from p, before it reaches the screen, and so the two will not be in phase when they reach the screen (figure 17).

What is the difference in path length for light from the two slits at the extreme edges of the slit, that is, at a distance b apart? A line AC (figure 18) can be drawn so that the two paths from A and C to a distant point, are the same. The path difference is then $BC \approx b \sin \theta$.

What happens if $b \sin \theta$ is exactly one wavelength λ? Then the light emitted by the slit, say t, exactly half-way across the slit, at angle θ would arrive $\lambda/2$ behind the light from p. So light from slit t would cancel out light from slit p. But similarly light from slit u would cancel out light from slit q, and light from every slit would be cancelled out by light from some other slit a distance $b/2$ from it. So there is no light at angle θ where $\sin \theta = \lambda/b$.

The number (or size) of the small imaginary slits does not appear in this result. (If it did one would be quite sceptical of the result. The result says that for a particular slit width, b, there is, for each wavelength λ, a direction given by $\sin \theta = \lambda/b$ (corresponding to two points or lines on the screen, one on each side of the original direction of the light beam) in which no light will appear.

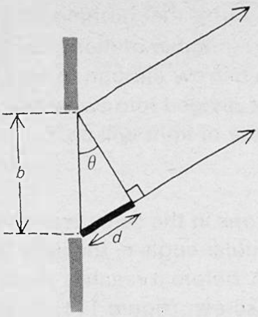
Similar arguments can be used to show that there will be no light in directions given by $\sin \theta = \dfrac{2\lambda}{b}, \dfrac{3\lambda}{b}, \ldots$

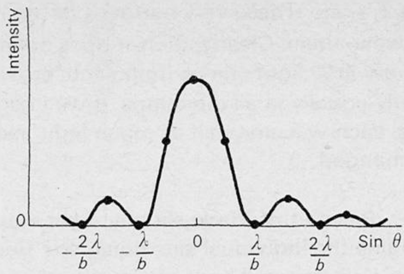
A start can be made on plotting a graph of the predicted amount of light against $\sin \theta$ (figure 19).

Angles at which there is some light

What happens to the energy of the light from individual strips emitted at angle θ if eventually no light goes in that direction? One possibility is that more energy than we expect is going in other directions. It will be possible to check on this later (see experiment 8.7).

How much light goes at angle $\theta = 0$? If $\theta = 0$, paths from all strips to the final point on the screen (at infinity or in the focal plane of the lens) are the same length and





Figures 20 and 21

Use of a laser

If a laser is being used to demonstrate diffraction effects along with the argument, it is simple to show that the central maximum is twice as wide as those on either side.

Connection of diffraction theory with later work

In Unit 10, *Waves, particles, and atoms*, we shall want to be able to say that the number of photons arriving at a place on a screen is to be calculated as if a wave had gone through the system, and that the net wave amplitude – the sum, taking account of phase, of all the pieces of wave that arrive at a place – gives (squared) the chance of arrival of a photon. For that purpose, it is important that students see the diffraction theory here not as an isolated bit of work, but as an example of a general problem and of a general way of solving such problems.

Demonstration

8.4 Microwaves going through a hole

184/1/2	3 cm wave transmitter and receiver
1045	diode probe for microwave experiments
181	general purpose amplifier
183	loudspeaker (if not part of item 181)
1014	wax lens *2*
1053	metal screen about 0.3 m square *2*
501	metre rule
1000	leads

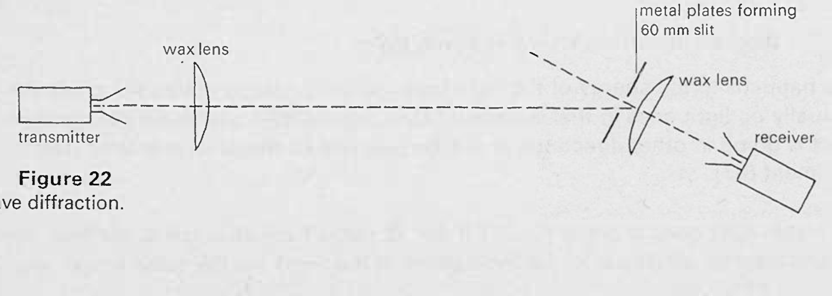


Figure 22
Microwave diffraction.

light from all strips is in phase. So the greatest intensity of light is in this direction: in other directions the intensity must be smaller.

What about some intermediate angles?

How much light goes at the angle whose sine is $\lambda/2b$? In this case the greatest path difference d (heavy line in figure 20) is given by

$$d/b = \sin\theta = \lambda/2b$$
$$d = \lambda/2.$$

Light from the last strip cancels out light from the first, but there is no other pair of strips whose contributions differ by exactly $\lambda/2$. So all the strips, except a pair at each end, contribute something.

It is difficult to say how much light there will be in the direction given by $\sin\theta = \lambda/2b$. But there will obviously be less than in the straight-through position ($\theta = 0$), for in that case there is no phase difference between contributions from individual slits, whereas here there are phase differences to be considered.

Similar arguments, which need not be gone into in detail, lead one to suppose that the graph of intensity would look like figure 21, with a strong bright central band, flanked by less intense bands. One striking prediction emerges from the simplest arguments: the central band should be twice as wide as those on either side of it. Those who did experiment 8.3d may be able to confirm this prediction, or it can be examined later (experiment 8.5).

Single-slit diffraction as an example of a general method

Single-slit diffraction has its own interest, but its main purpose here is as an example of a general method. If one wants to know how light will be distributed in an image, one must add up wave contributions, taking account of their phase. The light at one place is the sum of a lot of waves from different parts of the incident wave front. In other problems, such as X-ray diffraction or the analysis of the performance of a diffraction grating being used to measure wavelengths accurately, the process is the same, though usually it is more complicated.

Demonstration
8.4 Microwaves going through a hole

The ideas about diffraction at a single slit can be given some measure of experimental test (8.4 and 8.5).

The first experiment starts from the observation that for 'light' with a wavelength of 3 cm, that is, for microwaves, the slit can be quite big and still give substantial angles of diffraction. If the slit is two wavelengths wide, for the first 'zero',

$$\sin\theta = \lambda/b = 1/2$$
$$\theta = 30^\circ.$$

The transmitter is set up at the focus of one lens (about 0.5 m from the lens), to send a more or less parallel beam across the room. This lens may not be necessary if the transmitter is well away from the receiver.

Two metal plates, with a 60 mm gap between them, cover all but the middle of a second lens, which has the receiver at its focus, as in figure 22. It must be possible to turn the slit, lens, and receiver as a whole, or to place the slit over the transmitter and swing the receiver in an arc in front of it.

When the receiver points at the transmitter, the signal is large, but falls off as the receiver is turned. The angle, which may be in the vicinity of 30°, at which the signal first falls to a low value, should be found. It is necessary to use the modulation facility, and to listen to the output of the receiver via an audio amplifier.

In the experiment shown in figure 23, the transmitter and receiver can be about 0.5 m apart, with the receiver offset some 0.1 m, pointing at the transmitter. Slide one plate between them until it cuts off the radiation. Then add the second plate, having found beforehand by trial and error what space to leave between the plates for the greatest effect. From 30 to 60 millimetres is about right.

Demonstration

8.5 Measurement of the diffraction pattern from a single slit

1067/1 J	holder for two halves of a razor blade to be used as a single slit
	or
1067/2 O	single slit from set of 1, 2, 3, 4, 5, and 6 parallel slits
1067/2 M	slit of variable width
1067/1 H	small translucent screen
	or
46/1	translucent screen
1053	transparent ruler with mm graduations
501	metre rule
	slide projector (1000 W),
	or the following:
94 A	lamp, holder, and stand
94 B	housing shield
59	l.t. variable voltage supply
1067/1 A	sheet with slit and holes
1067/1 B	holder for lens of diameter 37 mm
1067/1 C	plano-convex lens, diameter 37 mm, +7D (focal length 0.15 m)

If a laser is available, the experiment is easy. If not, one of the arrangements in figure 24 will serve. The source is either a 1000 W slide projector with a single slit (width 0.1 mm or less) in its slide carriage, or a 12 V lamp overrun by 30 per cent, with a slit propped against it and a +7 D lens in front to cast an image of the slit on a screen.

Focus the slit's image sharply on a translucent screen with lines 10 mm apart ruled on it, check that the slit image is narrow compared with the rulings, and adjust the screen so that the image lies in the middle of the rulings, and parallel to them. Then add the adjustable slit in front of the lens, and, having blocked stray light from reaching the screen with suitable shields, narrow the second slit until the diffraction maximum just fills the 10 mm space. Record the distance from slit to screen.

A beam of microwaves can be made more or less parallel by a wax lens and then be passed through a 6 cm slit, being detected by a probe (or horn receiver) at the focus of a second lens, as in figure 22. The signal detected falls off on either side of the straight through direction, and reaches a low value at an angle of about 30°, though the precision of the test is poor, because the lenses are themselves diffracting apertures not too many wavelengths across.

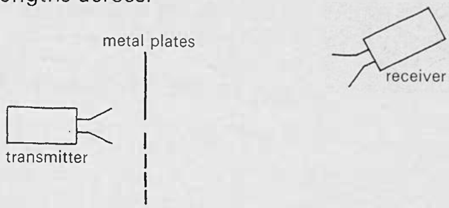


Figure 23
A 'less means more' experiment.

Another simple and striking experiment, illustrated in figure 23, is worth showing quickly. One metal plate is placed so as just to cut off the signal from the transmitter to the receiver placed at an angle a little to one side of the main beam. Then a second plate is added, as shown by the broken line, so that even *less* wave energy passes the plates. If a gap is left, the receiver output *increases* – less means more. The experiment may make it seem more plausible that waves from parts of an extended wave front can interfere destructively, since removing some of them adds to the net effect.

Demonstration
8.5 Measurement of the diffraction pattern from a single slit

In this experiment, the width of the central maximum of a diffraction pattern is measured to see if it conforms to the value expected from the slit width and the wavelength of the light.

The image of a slit, diffracted by passing the light through a second, adjustable, slit is cast on a translucent screen carrying a pair of pencil lines 10 mm apart. The width of the adjustable slit is then reduced until it is judged that the dark fringes on either side of the central maximum fall on the pencil lines, so that the central maximum is known to be 10 mm wide.

Having found the width of the adjustable slit for this setting, by projecting its image directly on a screen, the wavelength of the radiation which would produce such a diffraction maximum can be calculated. In one experiment, the screen was 3.3 m from the slit, so that the angle from the centre of the maximum to its edge was $5 \times 10^{-3}/3.3 = 1.5 \times 10^{-3}$ radian. The slit was 0.33 mm across, so setting the angle equal to $\lambda/$*slit width* gives $\lambda = 5 \times 10^{-7}$ m. The wavelength agrees with what might be expected for the wavelength of light.

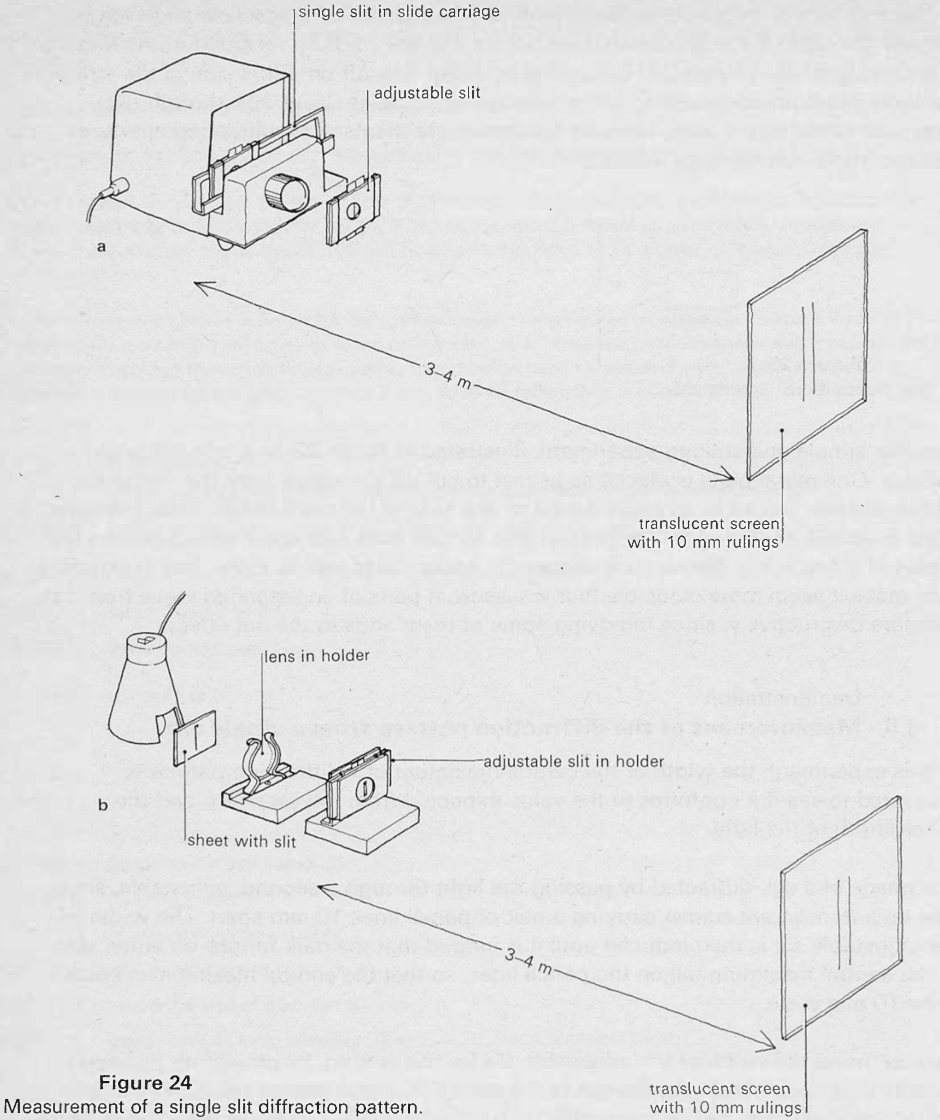


Figure 24
Measurement of a single slit diffraction pattern.

To measure the width of the adjustable slit, use it as an illuminated object, and project its image on the screen. Mark the edges of the image, and then use the same projection system to cast an image of a transparent ruler onto the markings.

A large slit to screen distance is important, but efficient screening from stray light is essential. It is not possible to adjust the slit and watch the pattern on the screen; an assistant must watch the pattern and direct the opening or closing of the slit until the pattern is the right width. It may be found that the slit of variable width (1067/2M) is easier to alter than the adjustable slit shown in figure 24.

Seeing things distinctly

The wavelength used in experiment 8.5 is roughly 10^5 times smaller than that in experiment 8.4 with microwaves, and the difference shows. With the larger wavelength, diffraction at a slit is a large-scale effect; with the smaller wavelength even a small slit produces only a modest spreading. The comparison, and the theory discussed previously, explain why the diffraction of light is not a thing we observe easily everyday. For $b = 1$ mm, which might in everyday speech be called a 'narrow slit', $\sin\theta$ for the first minimum is $\lambda/b = \dfrac{0.5\times10^{-6}}{1\times10^{-3}} = 0.5\times10^{-3}$.
So $\theta \approx 0.5\times10^{-3}$ radian (or 0.025 degree) as $\sin\theta \approx \theta$ if θ is small. This amount of spreading is hard to notice.

Resolution

In experiment 8.1b, the class saw that whether or not they could distinguish lamps in a row from one another, 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 ability of an optical system to *resolve* things.

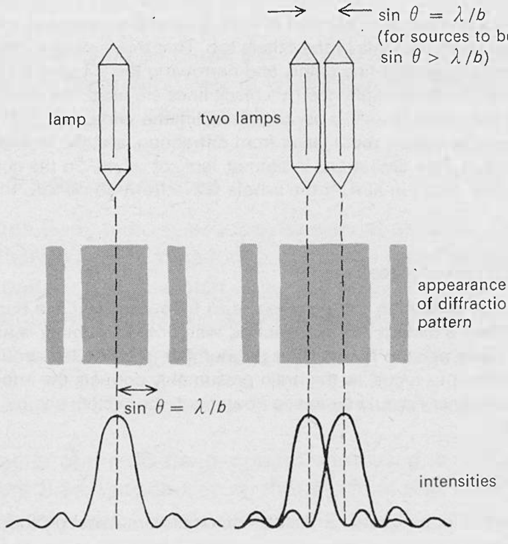


Figure 25
Overlapping diffraction patterns of two sources.

We cannot resolve, or see separately, two small objects if their angular separation at the eye is too small compared with the width of the central maximum resulting from diffraction by whatever restricts the light entering the eye. 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 not be distinguished as separate (figure 25). Where a slit of width b is

Students' book

Questions 19 to 24 are examples of the use of the idea of the limitation of resolving power by an aperture.

Resolving power

We suggest that the discussion of resolution be limited to the fundamental, simple point that the limit on resolution is at angles *of the order of* λ/b, without going into greater detail.

The Rayleigh criterion is restricted to a pair of incoherent point sources, and ought not to be presented as a sure guide in all situations: it is not strictly applicable to the row of five lamps, for example, nor will it serve directly for the resolution of a microscope.

For a circular aperture, the result $\sin \theta = \lambda/b$ becomes $\sin \theta = 1.22\lambda/b$, the difference arising because if the circle is divided into strips, the shorter ones at the edges contribute less wave amplitude than those near the middle.

There are of course other factors that could be involved in determining the resolving power of the eye. If two objects subtend an angle of 4×10^{-4} radian at the eye, so will their two images on the retina. If the pupil-retina distance is 20 mm, then the images will be about 8 microns apart. To resolve these would require light sensors closer together than this. In fact in the central region of the retina, the receptors are about 2 microns apart. Another factor that may be important is the quality of the lens and cornea.

In fact it is probably true that these different components are well matched: an improvement in any one of them alone would be worthless without improvements in the others too. That this is so is suggested by looking through a wide slit at any object containing fine detail, and narrowing the slit until it first becomes noticeable that detail is being lost. If, for example, the two black lines are used, the slits should be parallel to the lines. Detail starts to be lost when the slit is about one millimetre wide, less than the width of the pupil. The loss of detail for smaller widths must result from diffraction, and the lack of change in detail for greater widths must result from limitations in cornea, lens, or retina. So the cornea, lens, and retina are good enough to process detail in almost the whole wave front going into the eye out of doors on a sunny day.

An experiment on the 'resolving power of the brain'

The limitations of the 'camera' view of visual resolution can be revealed as follows. Stand at a bus stop with a friend who knows the bus route. When a distant bus approaches, wait until its number is just not distinguishable, but can almost be read. Then ask the friend what the number is; as he tells you, the 'unresolvable' number usually seems to snap into focus, as the brain presumably corrects the imperfect information it is getting. An interesting experiment would be to see how much the system can be deceived by being told the wrong answer.

Students' book

The article 'Radio astronomy' in the *Students' book* contains, besides much else, material on the resolution of various sorts of radio telescope.

Further reading about radio astronomy

Burbidge and Hoyle, 'The problem of the quasi-stellar objects', *Scientific American* Offprint No. 305.

Smith, F. Graham, *Radio astronomy*.

responsible for the diffraction patterns, the angle between the main maximum and the first zero is λ/b (taking $\theta \approx \sin\theta$). So λ/b radians is a good rough guide to the limit of angular resolution of a system with an aperture of width b.

What is the smallest angular separation between two objects that one could expect the eye to resolve? Taking 2 mm for the diameter of the eye pupil and 5×10^{-7} m for the wavelength of light, gives $\lambda/b = 2.5\times10^{-4}$ radian (taking $\theta \approx \sin\theta$). Students can estimate b for each other, or for themselves by holding a ruler to the eye and looking in a mirror. In experiment 8.1b they probably found that they could resolve two lines 2 mm apart at a distance of about 5 m so that $\theta \approx \frac{2\times10^{-3}}{5} = 4\times10^{-4}$ radian. The agreement is quite good for this kind of rough measurement, especially as the pupil is a circle, not a slit.

It is important also to remember that the eye is an extension of the brain, not a camera, and what a person 'sees' is not an image on his retina. Indeed, he may – *perhaps usually does – 'see' things which are not there, as in various optical illusions.* It is, for example, normal for the 'seen' image to have some of its 'defects' corrected: a rectangle with a bit missing will be seen complete. No doubt our impression of detail depends a good deal on this sort of process.

The subject of vision and how the eye and brain work together to give the sensation of seeing, is a fascinating area in which biology, physics, chemistry, psychology, and information theory all play a part.

Radio astronomy

The Jodrell Bank radio telescope consists of an 80 m diameter paraboloidal mirror with a receiving aerial at its focus. Although it can be steered to point to any part of the sky, the precision with which it can locate a radio source – or its ability to separate two neighbouring sources – is limited by diffraction. At a wavelength of 0.2 m, $\lambda/b = \frac{0.2}{80} = \frac{1}{400} = 0.0025$ radian (0.15°). The human eye can do several times better; about 0.0004 radian, as was seen in experiment 8.1c.

The diameter of the 36 s.w.g. copper wire, used to make a hole to look through in experiment 8.1a, was chosen so that the hole was about 400 light wavelengths across, so the view through such a hole is analogous to Jodrell Bank's view of the radio sky. It may be good to repeat the experience of looking through such a hole, at this stage, so that students may appreciate more clearly the limitations of one of the best radio telescopes.

When a student looks through this hole at a tungsten lamp, the lamp is as sharply defined to his eye as a radio source can be to the Jodrell Bank telescope. But he still sees the lamp 'better' than a radio astronomer 'sees' a radio source, for he has not just one sensitive cell on his retina to sweep across the diffraction pattern, but enough cells to show him a whole picture instantaneously.

Demonstration

8.6 The principle of an interferometer type of radio telescope

184/1 3 cm wave transmitter

184/2 3 cm wave receiver

1014 wax lens *2*

1053 metal screen about 0.3 m square *2*

181 general purpose amplifier

183 loudspeaker (if not part of item 181)

1053 bent metal plate (0.2 m square, bent to form two 0.2 m × 0.1 m surfaces)

table on wheels (see below)

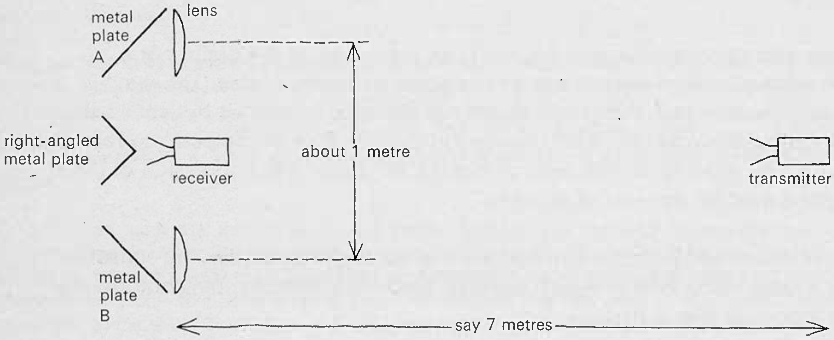


Figure 26
3 cm analogue of radio interferometer.

The transmitter is placed at one end of the room, whilst the items on the lefthand side in figure 26 are set up on a table near the other end, so that they may be turned through small angles as a unit. (If a suitable table is not available, the bits can be mounted on two trolley runways on a bench. The runways should be upside down, laid side by side and securely tied together at several points through the metal angle.)

First the lenses are placed a metre apart and in such a way that the perpendicular bisector of the line joining them points towards the transmitter. The receiver, on a support block and connected to the amplifier, is then also placed on this bisecting line with the right-angled reflector near to and pointing towards the horn. The reflector must be placed symmetrically. Reflector A is now positioned so as to give maximum receiver response, this position being carefully marked. A is then removed and B is similarly adjusted. If the receiver response is not the same in the two cases, the right-angled reflector should be readjusted until approximate equality is achieved.

The table or trolley boards should now be carefully rotated (with only one reflector in position) to show that the receiver response goes through a main maximum with maxima and minima at an appreciable fraction of a radian on either side of it. Reflector A is now replaced in its marked position and the table or runways slowly rotated again. This time there should be several maxima and minima within the broader diffraction pattern of the single-beam arrangement.

If this is done, and if there is time, students can try looking at two close sources through one hole (two holes in a mask over a lamp) and, as an analogue of the interferometer principle in the next experiment (8.6), through two small holes about 1 mm apart, at single sources of light.

Demonstration

8.6 The principle of an interferometer type of radio telescope

It would be impracticable for mechanical reasons to make a movable mirror several times as wide as the one at Jodrell Bank, so as to approach the sharpness of definition of a naked human eye. But a radio telescope can be made with two smaller mirrors some distance apart and can in this way achieve greater sharpness of definition, by using interference between the signals from the aerials. The electrical output from the aerials is combined before reaching any rectifier or amplifier, so that if the two aerials give equal amplitudes in antiphase, the total signal is zero.

A radio interferometer at Cambridge uses two sets of aerials up to a mile apart, and uses the Earth's rotation to change the direction in which they point. As the interferometer rotates past a radio source, the combined output is recorded and the superposition pattern obtained.

Another analogue to the radio interferometer, this time using 3 cm waves, may give a clearer idea of the principle.

The transmitter is placed on one side of the room and all the apparatus on the lefthand side of figure 26 is mounted on a board or table so that it can be turned through a small angle. When one flat metal plate is removed, radiation is received only through one lens. As the receiving apparatus is turned the output passes through a maximum, but the angle between the minima on either side is large, perhaps $\frac{1}{4}$ radian, so that the direction of the source is only crudely located. If the metal plate is replaced and properly adjusted, radiation is received through both lenses and is mixed before entering the receiver. Now, as the equipment is turned, the receiver's output passes through several maxima within the same angle ($\frac{1}{4}$ radian). If the receiving array is symmetrical, the narrower new central intensity maximum locates the direction of the source much more accurately. Similarly, when one looks through two small holes at a lamp, the position of the central fringe fixes the direction of the lamp more precisely than the diffraction blur of a single hole.

Accurate location of radio sources

As it happens, the accurate location of radio sources in the sky has turned out to be of the greatest importance for astronomy. Many of the sources – even the strongest – turn out to be very distant objects, which are not easy to identify with objects on a photograph. Only a few correspond to relatively nearby objects, such as the Crab nebula, whose appearance is striking and unusual and makes identification easy, given an approximate position of the radio source.

Aims

Radio measurements lead to deductions about the Universe which are of general interest, and one reason for including a short discussion of how these measurements are made is that students will be better able to understand the articles which appear from time to time in newspapers and magazines. Of the parts of physics in which important new developments are taking place, radio astronomy is one where the nature of the experimental data is comparatively easy to understand. But another reason for its inclusion is that it emphasizes the similarity of electromagnetic radiation of different wavelengths.

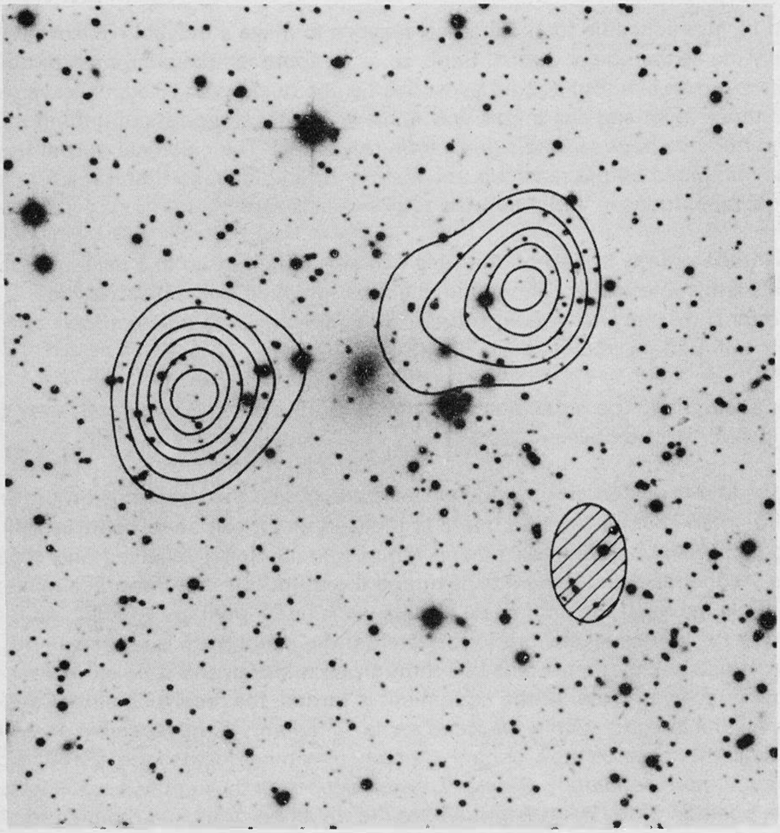


Figure 27
Location of an astronomical radio source. The striped ellipse indicates the radio resolution.
Photograph, Moffet, A. T. (1966) Annual review of astronomy and astrophysics, **4**, *149*.

Intensity and the square of the amplitude

There is a need, in Unit 10, for the idea that the intensity of a wave is given by the square of its amplitude, so that we may argue that the rate of arrival of photons (or the chance of arrival of one photon) might be related to the square of a wave amplitude.

It is not easy to give a general and rigorous argument for this point, especially for electromagnetic waves, until ideas about field energy can be discussed, which we cannot attempt in this course. So we suggest only a simple argument concerning sound waves, based on experiment, which may make the point seem a reasonable one, without pretending to establish it securely.

In passing, it may be noted that the Crab nebula is especially interesting. It seems to be the debris of an enormous stellar explosion, and most of its radio emission is from the expanding, swirling, ionized gas clouds still flying away from the explosion hundreds of years after it happened. But recently, it has been found that at the centre of the object, where there is a faint star, there is one of the newly-discovered, rapidly-pulsating radio sources, or 'pulsars'. Present feeling is that a pulsar is likely to be a spinning star made mostly of matter compressed nearly to the density of an atomic nucleus, such as is expected on theoretical grounds to be the remnant of the largest kind of stellar explosion (a supernova).

Some of the radio sources seem to be distant galaxies, and there are so many of them at great distances that accurate positions are essential for making any identification. Another recent puzzle depends even more on accurate location. Some radio sources seemed at first to have no obvious visual object of any interest at the place the radio waves came from. Accurate positions drew attention to what looked like rather ordinary stars at such positions, but it was soon found that these may well be very extraordinary indeed. There is some evidence that they are very distant, and that their radio emission is enormous, not less than the total radiation emitted from a whole galaxy, yet the evidence is that they are not very enormous in size. Other recent evidence suggests that they may not be so very distant, however. If they are far away, they must be emitting galaxies' worth of energy from a space much smaller than a galaxy (so that they look star-like, and have been called quasi-stellar objects, or quasars for short), and this interpretation of the evidence presents theoretical astronomers with one of their toughest problems. Various theoretical models have been proposed, but none as yet commands general assent.

Amplitude and energy in a diffraction pattern

Radio waves, light waves, and sound waves carry energy, and in a diffraction or interference pattern the energy is not evenly distributed over the screen, but a lot goes to places where the pattern has a maximum, and only a little to places where the pattern has a minimum. The positions of maxima and minima are found by working out the net amplitude of the wave arriving at each place, adding (allowing for phase differences) the wave contributions from different parts of the diffracting slit, grating, or object. But the rate of arrival of energy is not exactly the same as the amplitude.

For a harmonic oscillator, the energy turned out (in Unit 4) to be proportional to the square of the amplitude of oscillation. A similar relationship gave the rate of delivery of energy by an alternating current, in Unit 6.

A guess that the *intensity*, or rate of arrival of energy, of a wave is given by the square of the amplitude turns out to be correct. The next experiment gives an idea of how this comes about for sound waves.

Demonstration

8.7 Wave amplitude and energy when waves are superposed

1009	signal generator	
1035	pre-amplifier	
64	oscilloscope	
183	loudspeaker	2
157	microphone	
541/1	rheostat (10 to 15 Ω)	
501	metre rule	
503–6	retort stand base, rod, boss, and clamp	3
1000	leads	

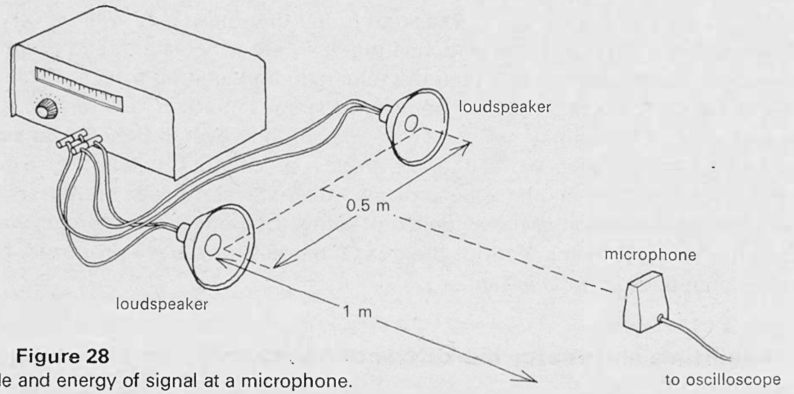


Figure 28
Amplitude and energy of signal at a microphone.

The two loudspeakers are driven, in parallel, by the oscillator. They are placed, as in figure 28, about 0.5 m apart, with the microphone about 1 m from them. Room reflections are a nuisance in this experiment. It may be best to work at a low sound level. To make this possible, the microphone is connected to the oscilloscope, which is switched to its most sensitive range, via the pre-amplifier. The microphone is to be moved through the interference pattern and it should initially be placed in the zero path difference position. The oscillator should be set at about 4 kHz and the 'gain' advanced until the oscilloscope trace nearly fills the screen. If the response is small for this microphone position and larger on either side of it, then the loudspeakers are in antiphase and the connections to one of them should be reversed.

The microphone should now be moved to the first 'zero' in the pattern. It is not likely that this will be exactly zero because the microphone is nearer to one of the speakers than the other, because there are reflections, or because the speakers may have slightly different characteristics. Should there be an appreciable trace height in the 'zero' position, then the rheostat can be connected in series with the stronger loudspeaker and adjusted to make the wave amplitudes more nearly equal. At the first maximum, each loudspeaker can be covered in turn by a coat to show that the amplitude is half of what it is for both loudspeakers together. In doing this experiment, care must be taken not to have reflectors in the vicinity of the apparatus. There is something to be gained by having the loudspeakers near the edge of one table and the microphone near the edge of another, with a gap of about a metre between the tables.

Demonstration

8.7 Wave amplitude and energy when waves are superposed

A pair of loudspeakers are connected to an oscillator and directed at a microphone. The microphone is connected to an oscilloscope and moved to the central maximum. Each loudspeaker is covered up in turn so as to emit no sound: the output voltage from the microphone drops to half what it was for both loudspeakers together.

This should not seem surprising. What does it show? Probably that the movement of the microphone's diaphragm is twice as great when both loudspeakers are emitting, and that the movement of air in front of the microphone is twice as great.

If the microphone were connected to a resistor instead of to the oscilloscope, how much more energy would be dissipated in the resistor when both loudspeakers are emitting? Four times as much, because at each instant the voltage across the resistor is doubled and therefore the current through it is doubled as well. At every moment during each cycle, four times as much power is given out by the microphone.

So the energy in the movement of the air in front of the microphone must be four times as great. But each loudspeaker emits the same power, and the total power emitted by loudspeakers can only be twice as much when both of them are uncovered. Where has the extra power come from? If the class cannot answer the question, repeat the demonstration with the microphone placed at a zero of sound intensity. Here the two lots of power emitted by the two loudspeakers add up to give zero. Where has the power gone to? The class ought to suggest that it is going to the loud places. The suggestion is correct: energy is not destroyed when waves superpose, it is just distributed differently.

A microphone and oscilloscope measure amplitude. The eye and photographic processes respond to the intensity (rate of arrival of energy). Intensity varies as $(amplitude)^2$ and is the appropriate quantity to consider in graphs like figure 21.

Consider the amplitude of sound (or any other wave motion) at a point as the wave goes by. The amplitude is alternately positive and negative, and the 'average' value of amplitude is zero. Yet clearly some energy is passing. $(Amplitude)^2$ is always positive and its 'average' value is not zero.

Part Two

Spectra

Time: this Part should not occupy much time. With students who have taken Nuffield O-level Physics, in which the diffraction grating appears in Year 5, three periods may perhaps be enough. The work here is not very new, and this section of the Unit serves as a place to collect together and revise ideas developed earlier.

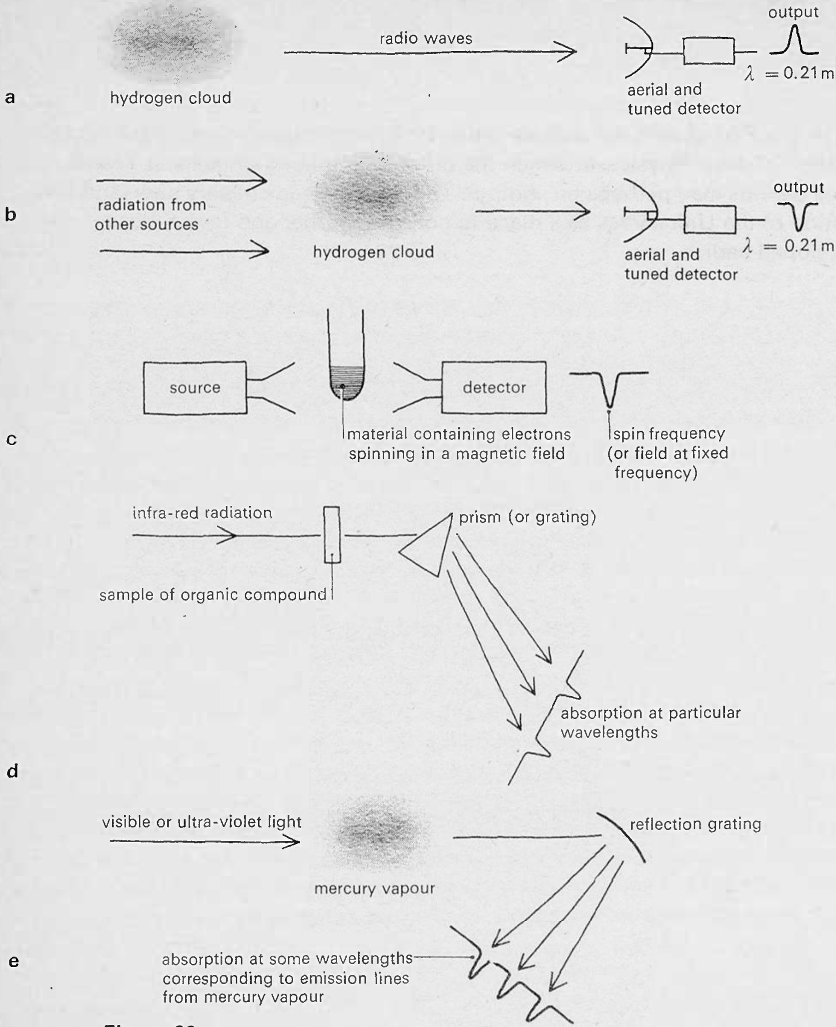


Figure 29
Different sorts of spectroscopy.
a Radio astronomy, wavelength 0.21 m. Emission from a hydrogen cloud in our galaxy, due to small rearrangements of electrons in hydrogen atoms (spin axis).
b Radio astronomy, wavelength 0.21 m. Absorption of radiation by hydrogen. The presence of some other molecules has been detected in the same way.
c Electron spin resonance, wavelength ≈ 1 m. Radiation absorbed by electrons in the presence of a magnetic field. Information obtained about the environment of electrons in the sample.
d Infra-red spectroscopy, wavelength $\approx 10^{-6}$ m. Radiation absorbed by oscillations of atoms bonded together in a molecule. Information obtained about the springiness of bonds: also used to identify bonds in a new substance or to identify substances.
e Spectra and energy levels, wavelength $\approx 10^{-7}$ m. Emission, or, as here, absorption of radiation by electrons changing energy levels. Information obtained about energy levels of electrons, usually outer electrons. (Continued.)

Looking at spectra and measuring wavelengths

The wavelengths of the radiation emitted or absorbed by a thing can tell one something about that thing. From the mixture of notes emitted by an organ pipe, one could tell something about the length, and perhaps the shape of the pipe, without looking at it. In Unit 4, the frequency with which the ions in a crystal of salt might be expected to oscillate was used to predict that it would absorb radiation in the infra-red region. The argument could be stood on its head, and one could deduce something about the forces between the ions, from the wavelength at which the crystal absorbs radiation. Chemists do this sort of thing often, and can sometimes decide whether a molecule contains say C—C, or C=C bonds, from the infra-red wavelengths at which molecules containing these bonds absorb radiation as the bonds oscillate, each with its own special frequency not much affected by what else is attached to them.

In astronomy, it is of interest that hydrogen atoms broadcast on a wavelength of 0.21 m, much the same wavelength as u.h.f. television signals. Such signals reaching the Earth indicate the presence of a cloud of hydrogen, and the 0.21 m hydrogen radiation has been used to map out our galaxy, confirming that it has a roughly spiral shape, with ourselves in one arm towards the outer rim of the whole flat spiral. Another way to detect hydrogen clouds is to look for absorption of other radiation coming through them, at just this frequency and wavelength. In both methods, the 0.21 m radiation is picked out for detection by tuning a receiver to it, just as one tunes a television set to the required channel. Figure 29 illustrates this principle.

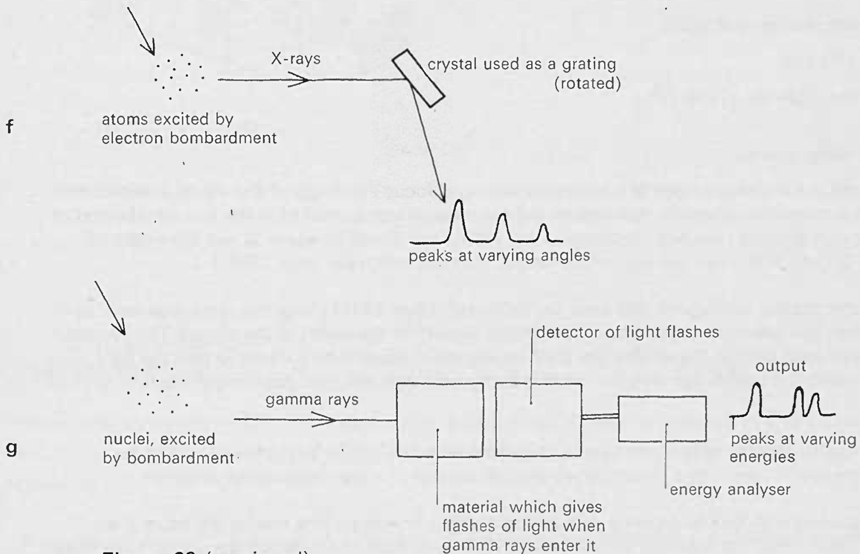


Figure 29 (continued)
f X-ray spectroscopy, wavelength $\approx 10^{-10}$ m. Radiation from electrons deep within atoms, giving evidence of deep energy levels.
g Gamma ray spectroscopy, wavelength $\approx 10^{-12}$ m. Radiation from nuclei excited by bombardment. Information obtained about nuclear energy levels.

Students' book

Questions 25 to 31 are about spectra and gratings. Question 25 develops $d \sin \theta = n\lambda$. Question 31 deals with the resolution of a grating. Questions 32 to 34 are about the uses of spectroscopy, while questions 35 to 37 cover other aspects of the wave theory of light.

Demonstration

8.8 The diffraction grating

191/1	coarse grating
1067/1 J	holder for two halves of a razor blade
1067/2 M	slit of variable width
1067/2 N	set of coarse gratings
1067/2 O	set of 1, 2, 3, 4, 5, and 6 parallel slits
1067/3 R	set of three colour filters
1067/3 S	small piece of fine black chiffon
1073	concave reflection grating
1053	aluminium foil
1053	cardboard 35 mm slide mount
23	microscope
46/1	translucent screen
59	l.t. variable voltage supply
94 A	lamp, holder, and stand
501	metre rule
	slide projector (1000 W)

Grating spectra

Put a single slit in the slide carriage of a slide projector, and focus the image of the slit on a translucent screen about a metre from the lens. It is best to use the variable slit, so that its width can be adjusted to give the best compromise between illumination and purity, but it may be easier to use the single slit, item 1067/2 O (which is really too narrow) or better, the razor blade slit, item 1067/1 J.

Hold the coarse grating with about 100 lines per millimetre (item 191/1) over the projection lens, as in figure 30, when the spectrum with three or four orders should fill the width of the screen. The spectra are best viewed from behind the screen, the ideal arrangement being with a mirror to turn the light through 90°, so that students can also see what is being held over the lens. Add colour filters in front of the grating.

Replace the coarse grating with a very coarse one (from item 1067/2 N) with about 10 lines per millimetre. The spectra are now a few millimetres apart instead of a few centimetres as before.

Finally try a grating with four to six lines, comparing it with a coarse grating having the same line spacing but many lines. It is essential to have a fully darkened room and to shield the screen from stray light if anything is to be seen with the grating that has only a few lines. Figure 31 shows an alternative arrangement if there is too little light. A laser is, of course, even better.

There is no receiver which can be tuned through a range of wavelengths for radiation in the infra-red or shorter wave regions of the spectrum, but the radiation can be spread out by a prism or its equivalent into a spectrum, and radiation of a particular wavelength selected by placing a detector in the right direction. Both this and the use of a tuned detector are examples of spectroscopy: gathering information about materials from the wavelengths at which they absorb or emit. Figure 29 illustrates a number of types of spectroscopic systems.

Demonstration
8.8 The diffraction grating

Students have already seen something of what a grating will do, but a demonstration may help to refresh their memories.

A slide projector with a slit in the slide holder and a coarse diffraction grating in front of the projection lens are used to project spectra onto a distant wall or screen.

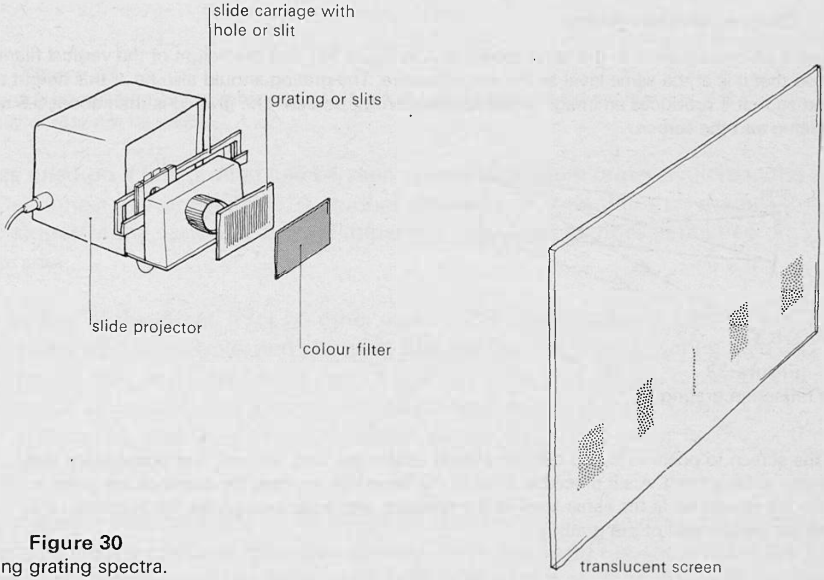


Figure 30
Projecting grating spectra.

After showing the spectra in full colour, place a filter (red or green) in the beam to show what happens to light of a narrower band of wavelengths going through the grating.

A much coarser grating, with perhaps 10 lines to a millimetre instead of nearer 100, gives spectra which are much more closely spaced. Indeed, it is not too easy to identify them as spectra at all. In a way, they look like bright Young's fringes, and so

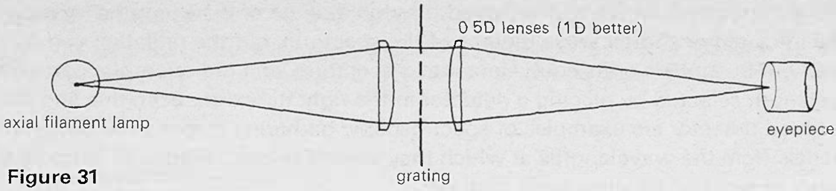


Figure 31
Alternative arrangement for use with small numbers of slits.

Two-dimensional gratings

Show that as the 100 line per mm grating is rotated, the spectra rotate with it, having replaced the slit with a slide mount covered with aluminium foil in which a 1 mm pin prick has been pierced. Then hold two gratings over the projection lens, with their rulings crossed.

Fine materials of various sorts give pretty effects, though the scale of the spectra is rather small unless the materials are very finely woven. Umbrella material is as good as anything. See Unit 10, Part Two, for a later use in the course of two-dimensional gratings.

Concave reflection grating

The screen is placed adjacent to the lamp (position A in figure 32) and the height of the vertical filament adjusted so that it is at the same level 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.

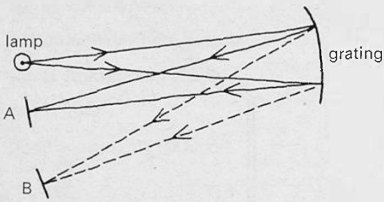


Figure 32
Concave reflection grating.

Moving the screen to position B and beyond should enable the first, second, and possibly the third order spectra to be shown. A slit placed in front of the lamp will improve the purity of the spectra. Of course, the slit should be at the same level as the filament, and wide enough for the light from it to illuminate the greater part of the grating.

they are, after a fashion. Two slits give fringes too faint to see with this apparatus, but a screen with four, five, or six slits gives fringes which have the same spacing as those from a very coarse grating, but which are broader where they are bright, as well as being fainter. The difference is not easy to see, but it is worth a try. If fringes from a pair of slits as closely spaced as the lines on the not-so-coarse grating could be seen (they would be too faint) the fringes would lie where the bright spectra lie, but would not be so sharp. Figure 33 illustrates the effects.

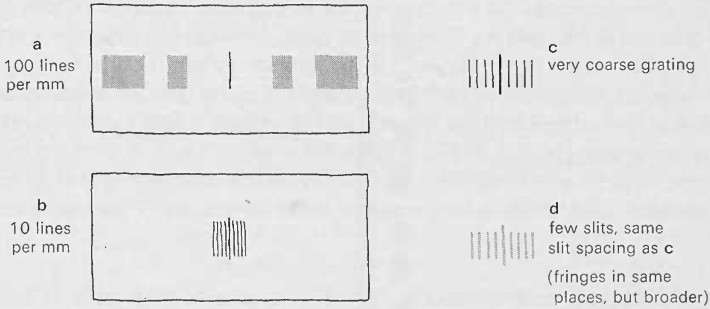


Figure 33
Effects of grating spacing and number of slits. (*Note.* **d** will include subsidiary maxima, not shown, which may or may not be visible.)

The lines ruled on a fine grating can be seen without difficulty under a microscope, using the highest power available. A student should have seen that they are not the opaque stripes of the textbook, but are imperfect, even kinked, lines scratched on a hard surface.

For the beauty of the effect, if for no other reason, it is worth replacing the slit in the slide projector by a small hole, and showing first the spectra from a grating as it is turned through 90°, and then from a pair of gratings with their rulings crossed. The 'extra' spectra along diagonal lines should be pointed out. A piece of fabric is a crossed grating too, and the diffraction pattern produced by fine fabric held over the lens is worth seeing.

Finally, a gesture may be made in the direction of more modern devices, by showing the spectrum from a concave reflection grating. Such a grating has the advantage that light need not go through it, so it can cope with ultra-violet or deep infra-red radiation which would be absorbed by most grating materials.

One obvious disadvantage of the gratings seen so far is that precious light is sent into many spectra, when it only needs to go into one spectrum to be analysed. Modern spectroscopists have developed a range of grating-like devices which send the light more nearly where it is needed, and also spread the differing wavelengths out over greater angles, so that spectroscopy is one of the fields where extremely accurate work is possible, and, indeed, is almost routine.

$n\lambda = d \sin\theta$ in Nuffield O-level Physics

Year V of Nuffield O-level Physics gives a simple derivation of the grating equation, and the argument here is intended mainly for students who have seen the simpler argument earlier. If they have not, it may be best to use the O-level argument first, and maybe even to omit the further discussion if the matter is likely to drag.

The main difference is that the argument opposite pays attention to the sharpness of the spectra as well as to their position; this is what makes the grating a useful tool.

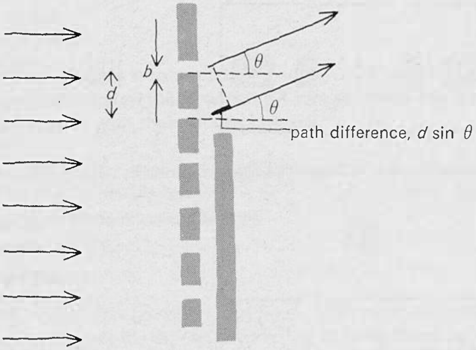


Figure 34

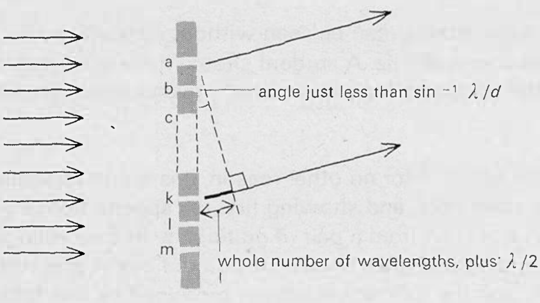


Figure 35

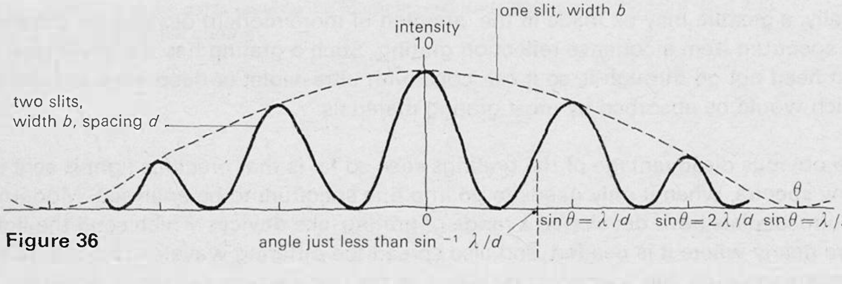


Figure 36

Many slits: sharply defined spectra

The thing which makes a grating useful is not that it sends light of one wavelength off at an angle, for two slits can do that, but that there is hardly any light except at certain angles. The following argument indicates how a grating with many slits achieves this result.

Real gratings are not so simple, but it will be enough to think about one made of narrow parallel slits of width b, recurring at an interval d in an opaque screen. Suppose that at first all but two slits are covered, as in figure 34. Then the two open slits will give a Young's fringe pattern like that in figure 36. One slit would give a pattern like the broken line in figure 36, and since b must be less than d, the one-slit pattern will be broader than the two-slit pattern within it. The one-slit pattern has its first minimum where θ is given by $\sin\theta = \lambda/b$. The two-slit pattern has its first maximum away from the centre ($\theta = 0$) where the path difference $d \sin\theta$ is equal to λ, so that waves from the two slits are in phase. There will be other maxima where $n\lambda = d \sin\theta$ for whole numbers n.

If a large number of slits, such as a, b, c, . . . k, l . . . in figure 35, are exposed, the light from all the slits will still be in phase at angles θ where $n\lambda = d \sin\theta$, so the maxima will be at the same angles as before (figure 37). There will be much more light energy at these angles, because the extra slits let through additional light. (If figure 37 had its intensity to the same scale as figure 36, the peaks would go off the page.)

The important difference is in the sharpness of the peaks. At an angle just less than $\sin^{-1}\lambda/d$, with two slits there is an appreciable intensity, because the light from the two slits is only slightly not in phase. Such an angle is marked on both of figures 36 and 37. At the same angle, with many slits as in figure 35, slits b and a will contribute light with the same phase difference as that from the two slits just discussed. Slits c and a will contribute light with a larger phase difference: somewhere there will be a slit k such that light from it travels half a wavelength more than light from a, and arrives in antiphase with it, so producing zero effect. If so, slits b and l will also produce zero effect, as will slits c and m, and similarly all the way across the grating.

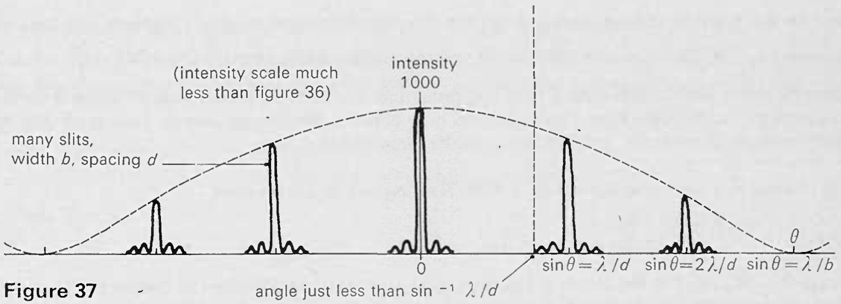


Figure 37

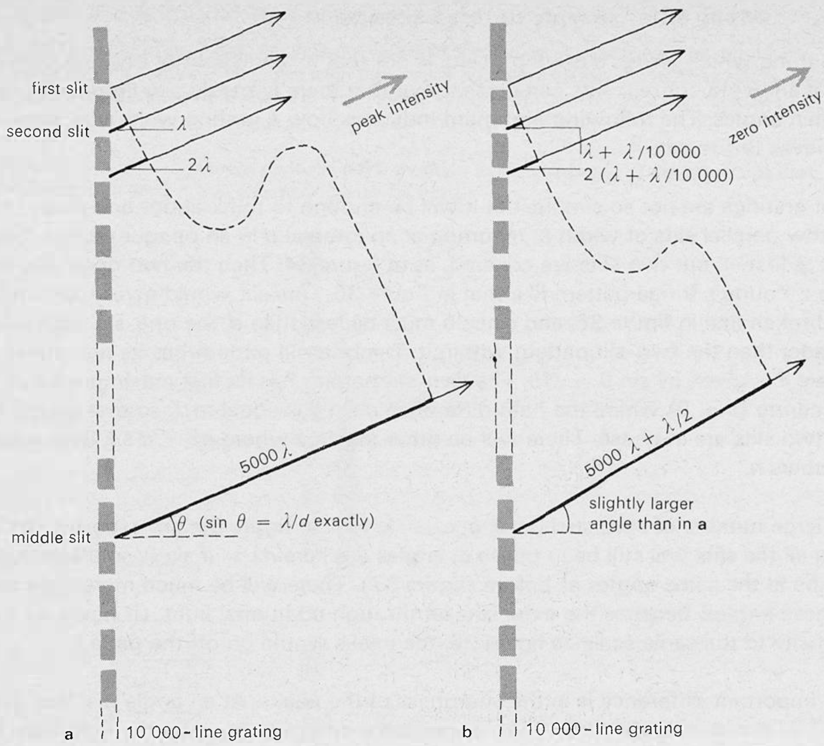


Figure 38
Illustration of resolution of a 10-000 line grating.

Algebraic version of the argument for the sharpness of the peaks

An algebraic version of the argument may seem easier for some students. At the exact angle for the first peak, the difference in path between successive slits is λ. If the grating has N slits the path difference between the first and the middle slit is $\frac{N}{2}\lambda$.

Now let the angle be increased slightly so that the path difference between adjacent slits happens to become $\lambda + \lambda/N$. That between the first slit and the middle slit becomes $\frac{N}{2}(\lambda + \lambda/N) = \frac{N}{2}\lambda + \frac{\lambda}{2}$. It is larger by half a wavelength than it was, and previously the path difference was an integral number of wavelengths, so the light from these two slits now cancels. Similar arguments, pairing off slits $N/2$ apart, indicate that the net amplitude sent at the larger angle is zero.

The change in $d \sin \theta$ was $\delta(d \sin \theta) = \lambda/N$. Treating this as a derivative,

$$\cos\theta\,\delta\theta \approx \lambda/Nd$$
$$\delta\theta \approx \lambda/Nd\cos\theta = \lambda/D$$

where $D = Nd\cos\theta$ is the width of the grating as seen from the direction of the peak. Clearly, the width of a spectral line from a grating is fixed by the fact that the grating is an aperture of effective width D. There is no escape from 'looking through holes'.

Except in the special directions $\sin^{-1} n\lambda/d$, the grating sends very little light at all, as illustrated in figure 37 by the sharp narrow peaks. Suppose k is the fortieth slit from a. If there are 10 000 lines in the grating, every fortieth slit can pair off with an earlier one, and there will be very few left over without partners.

How far away from an angle $\sin^{-1} \lambda/d$ will the intensity fall to zero, that is, how sharp are the peaks? A numerical example can make the principle of such a calculation clear, though an algebraic version is not very difficult.

A grating might have lines 10^{-6} m apart, with 10 000 such lines in a total width of 10 mm. Using light of wavelength 5×10^{-7} m, there will be a maximum first at $\theta = 30°$, where $\sin\theta = \lambda/d = 0.5$. The path difference between the first slit and the second is one wavelength; between the first and the third is two wavelengths, and so on. The path difference between the first and the five thousand and first slit, at the middle, is 5000λ. This is shown in figure 38 *a*.

Now think about light coming at an angle a shade more than 30°, but one so little larger that the path difference $d \sin\theta$ between the first and second slits is only $\frac{1}{10\,000}$ of a wavelength greater than the one whole wavelength it was at exactly 30°. Between the first and the third slit, the path difference will exceed two wavelengths by only $\frac{2}{10\,000}$ of a wavelength. But the first and the five thousand and first slits will have an extra half wavelength added onto their previous whole-number path difference ($\frac{5000}{10\,000}$ of λ onto 5000λ). See figure 38 *b*. The light from these two slits will now combine to give zero. So will that from the second and the five thousand and second, and so on, across the whole grating. The grating as a whole gives no light in this direction.

How much bigger than 30° is this angle? Its sine is not 0.5 but is 0.500 05, since $d \sin\theta$ must now be bigger than its previous value by $\frac{1}{10\,000}$ of that value, and d is unchanged. If there are N slits, the intensity will be zero at an angle where $\sin\theta$ is larger or smaller than the value at a peak by almost $1/N$ of that value. Roughly, the angle itself will be different by about the same fraction of the angle. No real grating will perform quite so well, because the lines cannot be ruled accurately enough.

Because the peaks for one wavelength are so sharp, it is easy for the grating to produce close but distinct peaks if light containing several wavelengths which differ only a little from one another is shone on it. This is, of course, the grating's main job.

The previous argument showed that $\sin\theta$ might change by $1/N$ of itself across the width of a peak from light of one unique wavelength λ, where, for $n = 1$,

$$d \sin\theta = \lambda.$$

The resolution of the grating can also be written down. Light of wavelength λ from one slit and from another one N slits further along the grating will, if it goes in a direction at which there is a maximum, have a path difference of $N\lambda$ in the first order spectrum, and $nN\lambda$ in the nth order.

If the wavelength is $\lambda+\delta\lambda$, the corresponding path difference is $nN(\lambda+\delta\lambda)$.

If the wavelength is λ, but the direction is that for the minimum intensity closest to the maximum, the path difference is $nN\lambda+\lambda$.

Light of wavelength $\lambda+\delta\lambda$ might be resolved from light of wavelength λ if the maximum of the first falls in the same direction as the minimum of the second. If so, then equating the two path differences (the direction being the same),

$$nN(\lambda+\delta\lambda) = nN\lambda+\lambda$$

so that

$$nN\delta\lambda = \lambda$$

or

$$\delta\lambda/\lambda = 1/nN.$$

Simple hand spectroscopes

Simple, cheap, hand spectroscopes can be bought, or can be made from a cardboard tube with a slit at one end, and a piece of transmission grating glued over an aperture in a cap which fits onto the other end.

Light of slightly longer wavelength $\lambda + \lambda/N$ (0.01 per cent greater if N is 10 000) will have a peak where $\sin\theta$ is bigger by the same fraction, that is, by $(\sin\theta)/N$. Such a peak, falling where that from light of smaller wavelength λ has just reached zero, might possibly just be distinguished from the peak at wavelength λ.

Thus a grating with 10 000 lines, if it were perfectly ruled (and they never are), could distinguish wavelengths in the first order spectrum which differ by only 0.01 per cent. The higher order spectra, with $n = 2$ or more, are spread out still more, and the resolution is even better.

Seeing spectra

A good way to finish this Part is to let students use gratings to look at spectra, and one of the best ways of doing this is to give each student a cheap piece of transmission grating mounted in a 35 mm slide holder, and let him look at sodium and mercury vapour street lamps, and at neon signs, on the way home. Obviously, discharge tubes could also be set up in the laboratory (and would have to be, in schools in an isolated rural area lacking the doubtful advantage of advertising signs to look at).

The reason for being interested in line spectra is that they give evidence for the spacing of atomic energy levels. See figure 39. This link was made in Unit 5 (experiment 5.18) and will be used again in Unit 10, *Waves, particles, and atoms*. It might be well to recall, 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, each of energy hf.

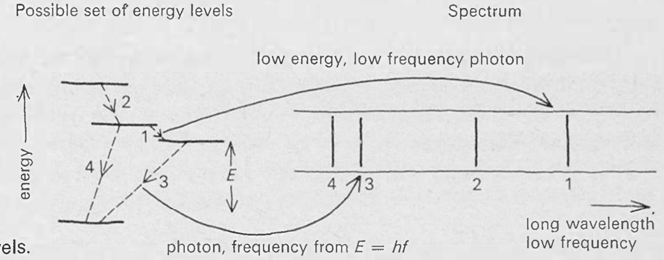


Figure 39
Spectra and energy levels.

Part Three

Electric waves

Time: less than two weeks.

This Part contains some sketches, none of them rigorously pursued, of electromagnetic wave theory. If these pieces of theory were examined in detail, with discussion of their inadequacies, the Part could take a very long time, and be extraordinarily difficult. It would be better to do nothing, than to get into such a situation, at this level of education. The aim is to give students a description of what people imagine an electromagnetic wave to be like, supported by some arguments which might make that description less implausible.

'The velocity of light through space is about 190 000 miles in a second; the velocity of electricity is, by the experiments of Wheatstone, shown to be as great as this, if not greater. . . . The view which I am so bold as to put forth considers, therefore, radiation as a high species of vibration in the lines of force which are known to connect particles and also masses of matter together.'

(M. Faraday, in a letter to R. Phillips Esq., April 1846. From his Experimental researches in electricity, *Volume III, pages 447–52. Dover.)*

'The velocity of transverse undulations in our hypothetical medium, calculated from the electromagnetic experiments of MM. Kohlrausch and Weber, agrees so exactly with the velocity of light calculated from the optical experiments of M. Fizeau, that we can scarcely avoid the inference that *light consists in the transverse undulations of the same medium which is the cause of electric and magnetic phenomena.*'

(J. C. Maxwell. From The scientific papers of James Clerk Maxwell, *1890, ed. W. D. Niven; Volume I, page 500. Dover.)*

Aims and purpose of teaching about electromagnetic waves

Much of science has consisted of patient research into how and why familiar things happen. More dramatic and fascinating are the unhoped-for developments which have been brought about by theoretical speculations or predictions.

The greatest example is Newton's work on gravitation, which is part of the Nuffield O-level Physics course. In the middle of the nineteenth century, physicists, still wondering what light consisted of, thought it might be something electrical and noticed that the speed of light is very close to a combination of electrical constants ($1\sqrt{\varepsilon_0 \mu_0}$ in modern terms). Maxwell gave a mathematical explanation of light as electrical waves which implied unlimited possibilities of other radiation, different only in wavelength and frequency. Twenty-five years later the first artificial electrical waves were made. Today they are made everywhere. Electricity, magnetism, and optics have become one subject.

Because of this and because they are made everywhere electromagnetic waves have a place in this course. The purpose of this Part is to show, in a very simplified form, how an electromagnetic wave is described and to link this information with some of the ways which are commonly used to set them going and pick them up. The use of such waves is now so common a part of ordinary people's lives that a good deal of interest in their nature may be expected, and should be satisfied. And those who go on may profit more from a complete and rigorous treatment of a hard topic if they have thought about some simple cases first.

We hope that students will emerge knowing that an electromagnetic wave is thought of as a pair of interlinked fields travelling at a unique velocity, and that the existence of the waves and the value of the velocity are related to the laws which govern electric and magnetic fields, discussed in earlier work. The theoretical arguments are intended to show just that such arguments do exist, not to be complete or rigorous. The description suggests that such waves could be polarized, which leads to a discussion of the polarization of light.

Teachers who wish to follow the arguments through more completely, for their own satisfaction, or to help the occasional very gifted student, may find some guidance in Appendix B.

Links with Unit 4

Unit 4 contained a preliminary, empirical survey of the electromagnetic wave family, discussing the similarities and differences which exist between radiations across the spectrum. It would seem to be a good idea to recall that discussion briefly, before beginning a more theoretical study.

The electromagnetic wave family

'Electromagnetic' waves have been referred to so far in this course without an attempt to justify the term or to elucidate their nature.

We find that a particular device emits radiation with a wide range of wavelengths, varying in quality only as wavelength varies. For example a tungsten filament lamp emits white light, which with a diffraction grating can be seen to be a mixture of colours each having a wavelength between about 4 and about 7×10^{-7} metre. The human eye is a detecting device only within this range, but a heat sensitive detector (e.g. a thermopile) moved across the spectrum and beyond the red end still responds to 20×10^{-7} metre. Although a thermopile shows hardly any response when moved to the violet end of the spectrum, photographic paper or a photo-electric cell shows a response down to 2.5×10^{-7} metre or less if no absorbing material intervenes between the source and the detector.

It seems to be a reasonable inference, though not a logically unavoidable deduction, that the radiation from the lamp differs only in wavelength, and not in a more fundamental way, over the range from 2.5 to 20×10^{-7} metre. However, there are other sources, which, with a grating to produce a spectrum, will cause a photo-electric cell to respond from, say, 1 to 4×10^{-7} metre, so that we have no reason to suspect any change in the basic nature of the radiation between 10^{-7} and 20×10^{-7} metre.

The same style of argument applies to radio sets. We can tune a receiver to any wavelength between 3 metres and 4 metres by turning a knob (and, if we want to, we can measure the wavelength by an interference method).

And we could have a transmitter capable of being tuned to any frequency between 3.5 and 5 metres. Using these two together we could assume that a single type of signal-carrying radiation exists between 3 m and 5 m wavelengths. And from the great variety of radio devices transmitting and receiving on overlapping ranges, we have no reason to expect a basic change in the quality of the radiation between 3 cm and 3000 m wavelengths.

Between 20×10^{-7} metre and 3 cm the range is not covered by well known techniques. But experiments 4.3 and 4.4 may have helped to show that the speeds at which radiation of visible and radio types travels are the same. Complicated astronomical measurements lead to the same conclusion about the propagation of light and radio in space.

Details of evidence for wavelengths much shorter than that of visible light would be out of place here, but it is evidence of the same type. A GM tube can be used over a big range of wavelength including γ-radiation and X-rays, and a film shows a measurement of the speed of γ-radiation which is the same as that of light. (Critical students should, however, be able to object that a GM tube also detects β-radiation which is in a different class. An apparent similarity in the method of detecting radiation is not sure evidence that the radiation is similar.)

Film: 'The velocity of gamma rays'

This film could be shown now, if it has not been shown earlier, in Units 4 or 5. It shows the direct determination of the speed of gamma rays, which were manufactured by letting the beam of charged particles from a cyclotron fall on a metal plate. The film was made in association with the Advanced Physics project. Details are given in the list of films, page 156.

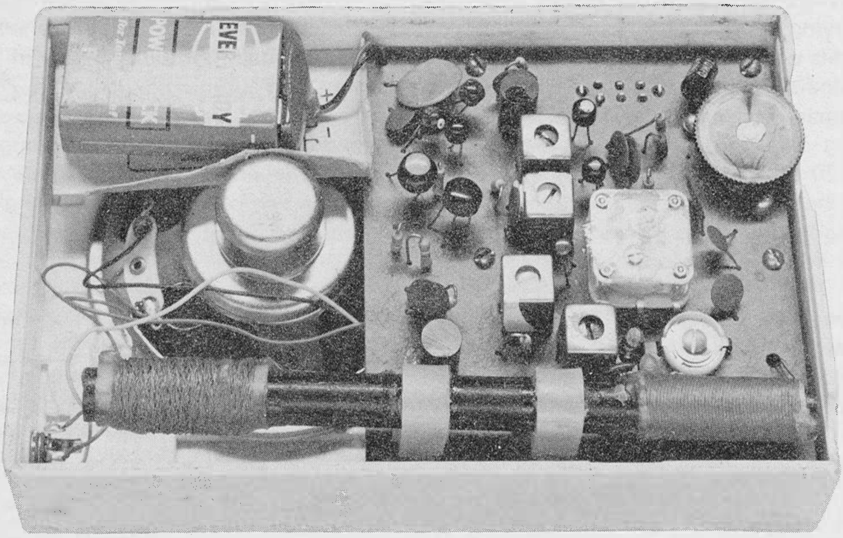


Figure 40
Ferrite rod aerial for a portable radio set.
Photograph, Michael Plomer.

Experiments with radio or television sets

A medium-wave v.h.f. portable radio is not a piece of recommended apparatus but if possible one should be opened in class. It would be best if a student could provide the set and show the aerials.

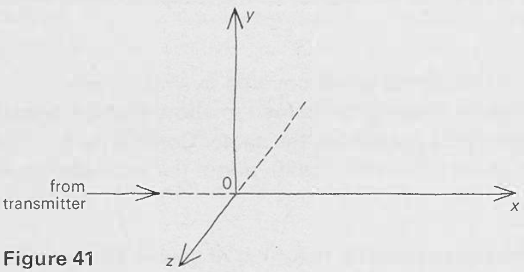


Figure 41

Ideally, in their respective cases, the electric and magnetic field changes can be shown to be in particular directions at right angles to the direction in which the signal is coming from the transmitter. If the signal is travelling in the 0*x* direction the v.h.f. rod aerial is most effective along some direction in

At various points in this Unit, and in Unit 4, experiments with light and with longer electrical waves have been associated with one another (for example, experiments 8.4 and 8.5). A reason for this was to associate the two radiations in students' minds and present the similarities, in order that the importance of electromagnetic wave theory should be clear. But a further lesson may be learned, even though teaching it too openly may confuse. How many experiments with 3 cm radiation in this course could have been done with 3 cm sound or ultrasonic waves? If all such experiments in this course had used sound waves, would we have shown that light and sound waves are basically the same? Could we argue that light is a fluctuation of air pressure because we can show that sound is? We could not, and the reason why not shows the need for further arguments about electromagnetic waves.

Waves and aerials

To most people, 'electromagnetic waves' mean radio waves, so the teaching can begin with a quick look at radio sets.

There are two sorts of aerials which portable radios can have, and some portable radios have both sorts.

For receiving medium wavelengths, such as 300 metres, a coil of a few dozen turns is wound on a ferrite rod. (Ferrite is a material in which large magnetic field strengths are easily produced. One would expect an iron rod to be usable, but as it is a conductor, eddy currents in it damp out the changes of field, which are what matter. Ferrite is not a conductor.) Thirty years ago, when portable radios were larger, a big coil was used without a core.

The system of using a coil is very like using a search coil to detect an alternating magnetic field in several experiments in Unit 7, *Magnetic fields*, and it is in fact a magnetic field which is being detected.

For receiving very high frequencies like 100 MHz (or 3 metres wavelength), which are too high for ferrite, a metal rod usually less than a metre long is used. The radio amplifies voltages, which appear on the rod as if they are induced by an electric field, and in fact changes in electric field are what is being picked up.

The advantage of using the coil over the dipole is at low frequencies (or long wavelengths) when the magnetic field at any instant is uniform across the coil and the voltage induced is therefore proportional to the number of turns. (Discussion of a very large number of turns, introducing self-capacitance and a low resonant frequency, would be unprofitable.)

the yz plane, usually Oy (or $-Oy$). And the ferrite rod is best along some direction in the yz plane, almost certainly Oz (or $-Oz$). But in many localities it is hard to know what transmitters the signals come from, and in many buildings there is so much conducting material that the wave fronts are badly distorted. Also, good radio sets have a lot of automatic gain control to amplify the signal if it is weak, making the demonstration less effective. (Nearly flat batteries reduce the effectiveness of automatic gain control.) It would be best not to discuss the directional qualities of the aerials if a demonstration cannot be made fairly conclusive.

A television set with a portable u.h.f. aerial (625-line system) can also be used. For the clearest results, the vision signal should be slightly detuned.

Demonstration

8.9 A spark transmitter

1050 15 cm dipoles and oscillator (oscillator not required)

14 e.h.t. power supply

1001 galvanometer (internal light-beam)

Some of the class have investigated the propagation and interference of waves between the transmitter and receiver in the first year (Unit 4, experiment 4.1). There is no advantage in using the e.h.t. power supply and the sparks in the first year. The 1 GHz oscillator is more suitable for first year work, because its amplitude is constant and the changes in received signals during investigation of interference or polarization are not complicated by inconsistency in the transmitter.

For demonstration work, in the second year, of the spark method, the transmitted amplitude will be bigger and more constant if the ends of the rods are carefully polished. If students become interested enough to want to investigate the effects, the use of finer polishes such as are used on silver could be suggested.

The use of magnesium rods, the same size, with clean unpolished surfaces for the spark, also improves performance. A galvanometer with high current sensitivity (e.g. 100 mm/μA or more) enables signals to be detected at more impressive distances. A galvanometer resistance of thousands of ohms is no disadvantage.

This is a demonstration that a student, briefed beforehand, could well prepare and show to the rest of the class.

Reading: Hertz's account of his experiments

See Hertz, *Electric waves*, for Hertz's own account of the series of experiments which led up to his discovery of means of generating electromagnetic waves. Chapter XI, 'On electric radiation', is the easiest to read, and the apparatus described is very like that used in experiment 8.9. Hertz's introduction to the book makes interesting reading, as it describes the muddles he got into, and his slow and patient disentangling of difficulties.

Chapter IV, 'On the effect of ultra violet light upon the electric discharge', has a special interest. Here Hertz reports the accidental discovery that light from one spark can assist the passage of another. Ultimately, developments from this discovery of the photo-electric effect were to show the quantum limitations of the very electromagnetic wave theory Hertz was helping to corroborate.

Demonstration

8.9 A spark transmitter

The transmitter is very like that developed by Hertz when he made the first artificial electric waves in 1886–7, except that it is smaller. The two brass rods have a very narrow gap between them. They are raised to a high voltage and, when sparks go between them, the detector picks up a signal.

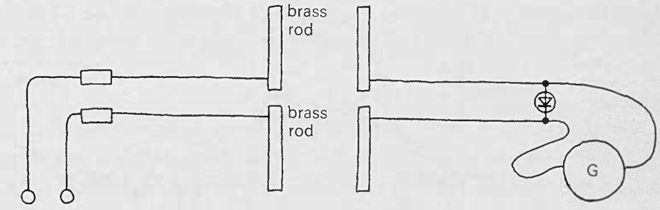


Figure 42
Spark transmitter and receiver.

Is the signal due to an electric field caused by the high voltage? We can pull the transmitter rods apart until they no longer spark, but give a bigger electric field. But then the detector doesn't pick up a signal. But when the transmitter rods are supplied with a much smaller voltage, alternating at 10^9 Hz, the detector picks up a good signal.

So it seems that a simple steady electric field is not what is going between the two aerials.

Discussion and speculation about what the electricity does on the rods, and why, should be encouraged while the equipment is still on the bench. But the speculation should be only to draw attention to the complicated behaviour of the fields nearby, not because students ought to know the details. Points which might emerge are these.

When the rods are charged just before the spark, there must be an electric field near them (figure 43). Its direction will be from one rod to the other, but the detailed shape need not concern us.

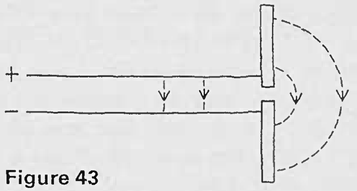


Figure 43

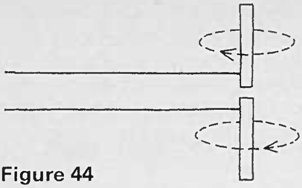


Figure 44

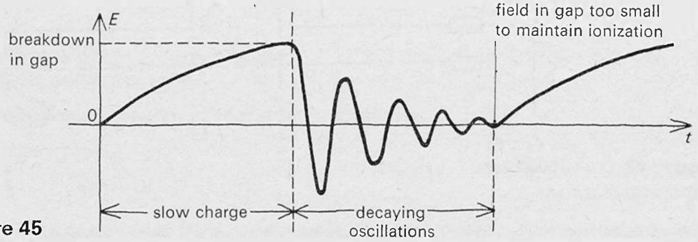


Figure 45

Points to avoid in discussion

Close to the dipole the electric and magnetic field fluctuations are not in phase and they control what happens on the dipole. They are known as an 'induction field'. Far away from the dipole the electric and magnetic fields are in phase and represent energy which has got away from the dipole which transmitted it. They are known as a 'radiation field'. There is no reason for students to hear these terms or to know the difference between them. But later work in this section is concerned only with the radiation field, and it is undesirable that, by too much attention to the dipoles, students should be in a position to get confused. Discussion of the dipoles should bring out that both electric and magnetic fields may be involved, probably at right angles to one another.

Another point is best omitted unless questions come spontaneously. What happens in the long leads between the dipoles and their sockets? The answer is that these lengths are designed to let the electricity run up and down between the dipoles and a pair of high resistors at the same frequency as it does in the dipoles. Then the sockets and plugs can be far enough from the dipole (see figure 46) not to affect the radiation much.

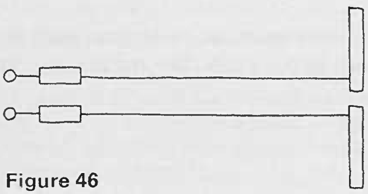


Figure 46

When there is a spark, and electricity flows from one rod to the other, it is equivalent to a current flowing along both rods. Such a current must have a magnetic field, and the magnetic field's direction must be round the rods (figure 44), at something like right angles to the electric field, which must now be diminishing.

Where there is a spark, many ions are produced, and the air is a good conductor. So the current can continue to flow easily and there is reason for it to do so, as the changing magnetic field will induce a voltage to make the current continue, instead of stopping immediately. So a p.d. and an electric field opposite to the original one will build up.

And the process repeats itself with a series of decaying oscillations (figure 45), the energy of the motion being lost by resistance to the current and by radiation of the fields, until the electric field across the gap is too small to maintain conduction at any time during the cycle. The process is like the bowing of a violin string, but few students will know enough about violins for the analogy to be useful.

Summary

The radiation from radio aerials has *wave properties*, and the waves travel at a definite speed (3×10^8 m s^{-1}) regardless of wavelength. The waves have some connection with electric and magnetic fields. Whatever they are like, the space in every room is full of them, coming from radio and television transmitters all over the world.

Marconi's work

Maxwell's prediction of electric waves was an outstanding intellectual achievement, and people immediately started looking for the waves to justify his theory. Twenty-five years later Hertz found them. But the sequel to Hertz's experiments is a striking example of how knowledge inhibits action, and teachers may like to mention it if their students accept what they are told too readily.

In 1894 Marconi, aged 20, suggested that Hertz's electric waves could be used for communication. He was also recognized for what he was – well meaning but ignorant. Men familiar with waves and with electricity explained how limited the use of electric waves must be; as they would not pass through conductors they would be propagated in straight lines with only very slight diffraction round the back of obstacles, and receivers would not be able to distinguish one transmission from another. The advice was right, and still qualitatively expresses the limitations on radio communication. But Marconi was determined to explore the idea of a telegraph without wires, and he was very skilful. And after each successive demonstration, from attic to garden, across a valley, from shore to ship, across the English channel, across the Irish Sea, more people saw that, limited and primarily a toy as it was, wireless could have uses.

In 1901 Marconi tried to send a signal across the Atlantic. The distance was ten times further than the greatest he had tried before, and the path deviated from a straight line

Demonstration

8.10 The speed of a pulse along a cable

1031 200 kHz pulse generator

1062 drum of coaxial cable

1007 double-beam oscilloscope

1051 resistor, 68 Ω *2*

1040 clip component holder *2*

This demonstration is one which it would not be hard for a student to prepare and show the class.

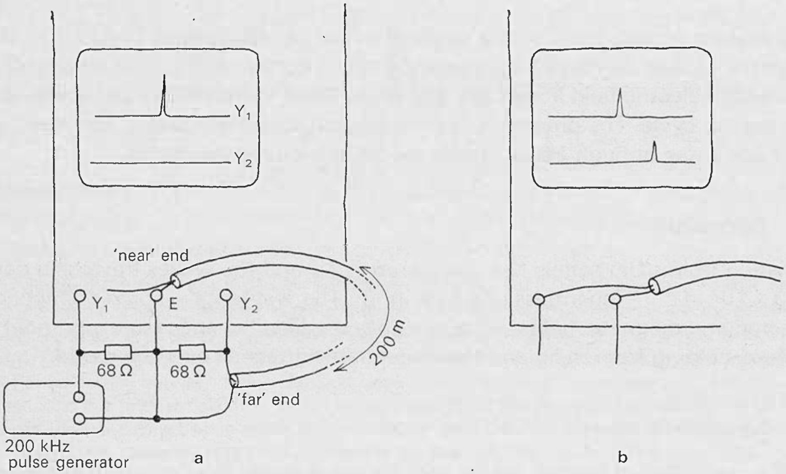


Figure 47
Two-beam display of the speed of a pulse along a cable.

The 200 kHz pulse generator should be connected across a 68 Ω resistor and across the Y_1 input of the oscilloscope. Both ends of the outer sheath of the coaxial cable should be connected to earth. The centre of the far end of the coaxial cable should be connected to the Y_2 input, across which is another 68 Ω resistor. Both inputs should be set to about 0.2 V cm^{-1}. The time base should be set to 1 cm μs^{-1}, and the pulse should be made stationary on the screen. 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.

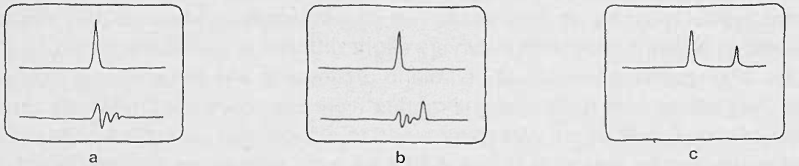


Figure 48
a and **b** Undesired oscillations in experiment 8.10.
c Single-beam oscilloscope display of pulse delay time.

Coaxial cable is preferable because it picks up little interference from outside the experiment and because it does not matter whether the cable is unwound from the drum or not. But there is a danger that the far end leading to Y_2 may pick up the pulse by induction direct from the lead to Y_1. This causes oscillations on the lower trace starting when the pulse is delivered to Y_1 (figure 48) even if the near by several hundred wavelengths. The Earth's diameter is to wireless wavelengths roughly as half a centimetre is to the wavelength of light, so Marconi's chance of transatlantic signalling could be compared with the chance that a small insect on a one centimetre steel ball would have of seeing a faint star from the wrong side of the ball. As he said later 'The experiment had involved risking at least £50 000 to achieve a result which had been declared impossible by some of the principal mathematicians of the time'. He succeeded at the first attempt. The ionosphere had reflected and partly focused the waves.

The mathematicians were right to think that transatlantic wireless ought to fail, and Marconi understood their reasoning, but it was his nature to be less impressed with arguments from established learning. The situation had been what he had hoped it would be – one in which knowledge was useless and in which willingness to gamble would be rewarded.

The speed of electricity in a circuit

The question, 'How fast does electricity in a wire travel?', has two answers: about 10^{-3} metre per second, and about 3×10^8 metres per second. The first answer may have been obtained as an estimate of the drift speed of electrons in a current-carrying wire, in Unit 2. On the other hand the working of telegraphs and telephones shows that signals, or pulses, or disturbances travel very fast. How fast?

Demonstration

8.10 The speed of a pulse along a cable

We can find this out by applying a pulse to one end of a wire and measuring the time it takes to get to the other end. Coaxial cable is convenient. The time the pulse takes to get to the other end is most simply measured with a double-beam oscilloscope, applying the original pulse to the upper trace, and the pulse that has travelled along the wire to the lower trace. The speed is so high that a very fast time base must be used, which can only be made visible by frequent repetition.

The pulse generator and the oscilloscope may be connected so that the pulse is displayed, the far end of the cable being connected to the amplifier of the other trace, without the near end of the cable being connected yet (figure 47 *a*). Then, when the near end of the cable is connected, the pulse which has been along the cable appears on the lower trace, delayed (figure 47 *b*). If students doubt that the two oscilloscope spots move simultaneously, the two inputs may be joined together by a short lead, to show that then there is no delay. Questions about the use of resistors at the ends of the cable, or about stray pick-up by leads at the far end, can be answered, but there is no reason to discuss them otherwise.

The pulse will be found to travel at between 2×10^8 and 2.5×10^8 m s^{-1}. The exact speed depends on the material filling the space between the cable's conductors, but if this were a vacuum we should find the speed to be 3×10^8 m s^{-1}, that is, the speed of light.

end has not been connected to Y_1. These oscillations are more noticeable if there are long unshielded wires connected to Y_2, and quite conspicuous if the outer sheath of the coaxial cable has not been connected to earth at the far end.

The 68 Ω resistors are needed to prevent the ends of the coaxial cable reflecting the pulse. Without them, the pulse of positive electricity on the inner conductor would go to the far end and, finding no outlet (the oscilloscope being a bad conductor), would come back to the near end where the pulse generator would also reflect it, though not so completely, and so on. The extra blips on the oscilloscope screen then need more explaining. The 'characteristic impedance' of coaxial cable is usually about 70 Ω, and a resistor of this value across the end of the cable absorbs the pulse, just as a further infinite length of cable would do if it were connected instead. The resistors are best connected directly between the oscilloscope terminals.

If only a single-beam oscilloscope is available, the speed of the pulse may be deduced from the time it takes the pulse to return, after it has been reflected from the far end with no 68 Ω resistor across it. The trace is then as in figure 48 *c*, the delay being twice as long.

E- and *B*-fields in coaxial and other conductors

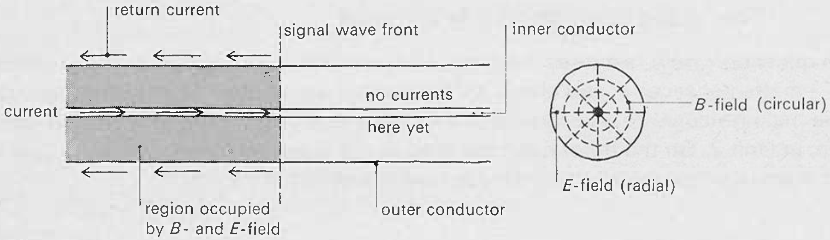


Figure 49
B- and *E*-fields propagating along a coaxial cable.

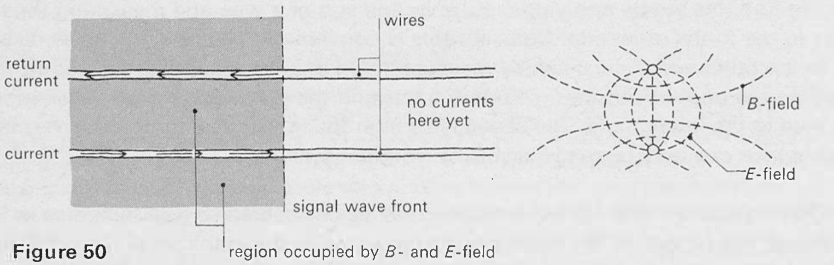


Figure 50
B- and *E*-fields propagating along a pair of wires.

Teachers may find it convenient to have the following information in mind. The *E*- and *B*-fields travelling along between the inner and outer conductors of the cable do not differ fundamentally from those of a plane electromagnetic wave in empty space such as is discussed later in this Part. The most important differences are:

1 The wave travels more slowly because it is going through polythene.
2 Neither field is parallel, or uniform over a plane perpendicular to the direction of travel.
3 The wave fronts are not very wide, their edges being bounded by conductors in which electricity travels with them.

Early measurements of the signal speed

Measurements of the speed of electrical signals along wires were made quite early in the nineteenth century, though with conflicting results, as the experimenters did not always realize that currents in one part of their wires could induce currents elsewhere. (Electromagnetic induction itself was only just being discovered and understood.)

Probably the first effective measurement of the speed of electrical pulses was made by Wheatstone in about 1830. He had spark gaps close together but connected by wires several kilometres long, so that on applying a sudden p.d. the sparks at the gaps came one after the other because of the delay in the wire. He measured the time interval with a rotating mirror, which he suggested could also be used to measure the speed of light as in experiment 4.3. Rotating mirrors were a common nineteenth century device for doing things we should now use an oscilloscope for.

Wheatstone's result was half as big again as we would today expect it to be, probably because the 13 km of wire he used was strung round and round his garden on posts, so that it was a large coil rather than a long wire.

The present understanding, illustrated by an argument which follows the next experiment (8.11), is that an electromagnetic wave travels along and around a long conductor when a signal is sent down it, setting charge into motion as it passes, and leaving behind the sluggishly drifting electrons that constitute the current. Similarly, in a masses-and-springs wave model, the speed of a massive trolley is not the same as the speed of the wave.

The early experimenters did not know that they were experimenting with electromagnetic waves. But their results stimulated others to think about what was happening. In particular, as can be seen from the quotation from Faraday on page 59, the nearness of the 'speed of electricity' and the speed of light gave rise to speculation that they might be connected in some way.

Slow electrical waves

The pulse along a cable (experiment 8.10) is far quicker than even a sound wave in steel. In the latter case, in Unit 4, *Waves and oscillations*, it was convenient to have a slow motion wave model, made of masses and springs. A slow motion electrical wave can be made too, using a row of inductors and capacitors.

If it were not impracticable, the experiment would be repeated in the single-beam form (figure 49) but *using, instead of a coaxial cable, two horizontal plates 1 metre by 200 metres separated by a* 10-millimetre gap. The speed of the pulse would then be the same as the speed of light, because the electricity forming the pulse would be that required to be at the edges of the electromagnetic wave which would run along in the space between the plates. This system is the one discussed later on.

Demonstration

8.11 Slow electrical waves

1051	electrolytic capacitor, 1000 μF 15 V *10*
1030	high inductance coil *10*
92 G	double C-core and clip *10*
1007	double-beam oscilloscope
1033	cell holder with four U2 cells
1041	potentiometer holder *with*
1051	preset potentiometer, 5 kΩ
1040	clip component holder *2* (or *12* if used to hold 1000 μF capacitors)
1051	resistor, 120 Ω *2*
1003/2	milliammeter (10 mA) *4*
1004/1	voltmeter (1 V) *4*
52 L	mounted bell-push
158	class oscilloscope *6*

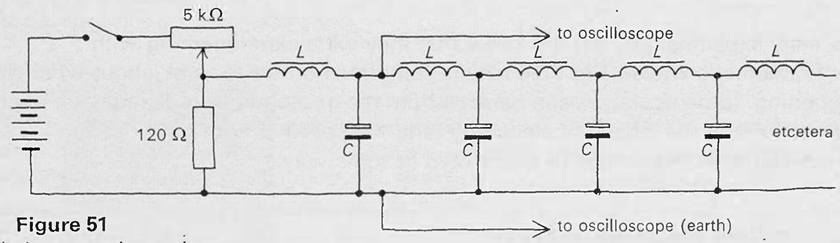


Figure 51
Chain of inductors and capacitors.

Each high inductance coil should have a C-core whose surfaces should be smooth and clean and should fit exactly so that all inductors have the same inductance. The 120 Ω resistor in figure 51 is roughly equal to the characteristic impedance of the line and it enables the pulses to be smoothly started and stopped. The 5 kΩ potentiometer should be adjusted until closing the switch causes a few tenths of a volt across the first or second capacitor. The second 120 Ω resistor can be connected across the other end of the line so that it absorbs pulses.

The inductors are each about 15 henries if the C-cores fit properly, and the speed of the pulse, $1/\sqrt{LC}$, is $1/\sqrt{15\times10^{-3}}$ or eight sections per second. So a pulse takes more than two seconds to pass along ten sections of the line and back again.

Demonstration
8.11 **Slow electrical waves**

The pulse along the cable shows a tendency to stick together as it travels. That is, voltage changes propagate at a fixed finite speed.

Now it is a property of electricity to spread itself all over a conductor, so the tendency of this small quantity of electricity to stick together as it goes along the cable is remarkable. The reason why it does this can be examined in another experiment using meters, in which the same kind of process happens much more slowly.

A pulse is shown travelling along a chain of inductors and capacitors, as in figure 51. But the wave which is seen, though instructive, is not an electromagnetic wave.

Voltmeters and milliammeters can be put anywhere in the circuit, except that the resistance of more than one milliammeter at a time would affect the current being measured. Oscilloscopes can be connected across any number of capacitors.

Class discussion, with measurements when appropriate, should bring out the following processes.

When the switch is closed, a p.d. of say 0.5 volt appears across the 120 Ω resistor and across the beginning of the 'cable'. Current flows through the first inductor, then the next, then the next, raising each capacitor successively to 0.5 volt. When the switch is opened again, the first capacitor, then the next, then the next, discharges itself to maintain the current which has been built up in the inductors. The electricity travels, in a bunch, at a finite speed along the line. (See figure 52.)

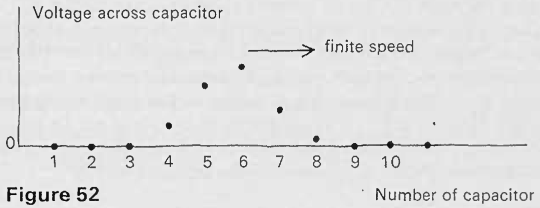


Figure 52

The fact that the back of the pulse travels as fast as the front suggests that one is looking at a propagating disturbance, not, for example, at the slow leakage of electricity onto a conductor. It might help to remove all the inductors, and show the difference.

A dramatic demonstration is obtained by connecting a class oscilloscope (item 158) across every other capacitor, with their time bases off, and viewing the motions of their spots with the room lights off. The double-beam oscilloscope can be used to show the pulse at two places.

For the best results the apparatus should be set up beforehand and the inductors carefully checked. Some C-cores may give less inductance than the average and cause spurious reflections, and they should be replaced or put at the ends of the line.

The resistance of the high inductance coils attenuates and slightly distorts pulses. Distortion is reduced, although attenuation is increased, by adding leakage resistance in parallel with the capacitors. Voltmeters, connected to show how far the pulse has got, help to do this.

A voltmeter (1 volt, item 1004/1) across each capacitor will show how the voltage travels, which will be useful later. If students examine pulses in the experiment closely and become worried about why the pulses get smaller so quickly it may help to tell them that the total length of wire in the experiment is more than a kilometre, and the wire's resistance is responsible. If students insert more resistance, for example by putting a milliameter in series with each coil, the attenuation becomes worse. Also students may regard the time for which the pulse lasts as being more impressive if they realize that a radio message to the Moon does not take so long as the reflection takes to come back in figure 51.

Reasons for looking at the slow electrical wave

This lumped *LC* system does not play an essential role in the theoretical discussion, and the wave on it is not an electromagnetic wave. So it is not a thing to be taken too solemnly. But it provides a visible system made of familiar components, in which the reason for the finite speed of a wave both has some analogy with an earlier system, and is similar in kind to a reason which might be given to explain the finite speed of electrical disturbances along a wire, namely the inductance and capacitance associated with each part of the wire.

Lumped-constant wave media

A system like the line of inductors and capacitors (or the masses-and-springs wave model) does not behave in all respects like a system in which the inductance and capacitance (or mass and elasticity) are smeared out smoothly, as is the case, for example, for the coaxial cable in experiment 8.10. Although for low frequency, long-wavelength waves, smooth and lumped systems are comparable (in terms of capacitance and inductance per metre), the lumped medium is always dispersive, whilst the smooth medium need not be. For wavelengths shorter than the repeat distance between lumps, the lumped system fails to propagate waves at all. When waves are propagated, the speed depends upon wavelength. This means that a wave will be distorted by a lumped system, but need not be by a smooth one. (If the first mass in a masses and springs model is waggled very rapidly, no wave propagates. If the switch in figure 51 is opened and closed very rapidly, no electrical wave propagates.)

For teachers who wish to refresh their memories, Feynman *et al., The Feynman lectures on physics,* Volume 2, Chapter 22, gives a good account of *LC* lines.

Use of analogy

At the risk of being wearisome, for the point has been made before, it may be worth pointing out how analogies can be exploited to see quickly what might happen in some new situation, given a knowledge of what happens in some other, partly analogous situation. Analogy is one of the intellectual tools used (with caution) by physicists and engineers.

Series-parallel duality in the mechanical–electrical analogy

It may disturb students that the springs are in series, while the capacitors are in parallel. (A similar worry may have arisen in Unit 6.) The reason is that force is to be analogous to voltage. The forces in springs in series (under steady conditions) are equal, and to have equal voltages across electrical components, one must put them in parallel.

Analogy with a wave model made of masses and springs

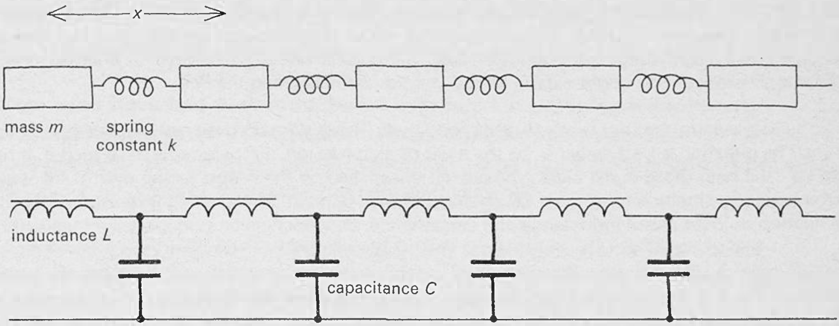


Figure 53
Analogy of an LC line with a wave model.

In Unit 4, the speed of a compression wave along a model made of masses and springs was worked out and tested (experiment 4.6). If masses m were spaced at distances x apart, using springs with spring constant k (force per unit change in length), the velocity of a compression wave was given by,

$$v = x\sqrt{k/m}.$$

Over how many pairs of masses and springs ('sections', for short) does the wave travel in a second? (v/x, since the wave covers v metres in a second and there are $1/x$ sections in a metre.) Putting $u = v/x$,

$$u = \sqrt{k/m}$$

gives the speed u in sections per second.

Now in Unit 2, a capacitor was compared with a spring, and it was seen that there is some analogy between the capacitance C and $1/k$, the compliance of the spring (or between $1/C$ and k).

In Unit 6, the same analogy was drawn, together with another, between inductance L and mass m. The two analogies were used to help understand the frequency of oscillation of a parallel LC circuit. The analogy might be worth extending to the row of inductors and capacitors, as in figure 53. In the analogy, the compression of a spring is compared with the charge on a capacitor, the tension in a spring with a p.d., and the velocity of a mass with a current. The two wave models are alike in that in the mechanical one, the velocity of a mass changes only if the springs on either side of it are differently compressed, so that there is a net force on the mass; while in the electrical one, the current through an inductor is changing only if the capacitors on either side of it are differently charged, so that there is a net p.d. across the inductor.

Students' book

Question 38 gives a step by step argument for the velocity of an electrical wave along a line of inductors and capacitors. The question, while parallel in many respects to question 24 in the Unit 4 *Students' book*, which discussed the motion of a wave along a row of masses and springs, does not exploit the analogy directly, as the argument opposite does. It argues in terms of the currents, charges, and potential differences one might expect to find at various times along the line.

It is a hard question, and not every student need try it. Those who do need not be expected to finish it easily. The question is best suited to be the basis of a discussion, after students have tried it. It need only be used with those in the class who are left unsatisfied by the simple taking over of the wave speed from one situation to another, by analogy. It does contain, however, an interesting new application of ideas about inductance and capacitance, and teachers or students might value it for that.

The argument is that, if a wave in which each capacitor charges at a current I, steady as the wave passes, to a p.d. V, each must charge for a time VC/I. If the wave travels at a 'speed' u capacitors per second, there will be a number VCu/I capacitors charging at any one time, and a total current VCu is required. Since the p.d. across a capacitor rises at the rate I/C, the p.d. between adjacent charging capacitors is I/Cu, and this is the p.d. across the inductors separating such capacitors. The current through each inductor past which the wave has gone rises at the rate Iu, because of the need to charge the capacitors beyond it over which the wave is passing at u capacitors per second. The inductance of each inductor is thus $L = 1/Cu^2$, dividing the p.d. I/Cu across it by the rate of change of current in it, Iu. Thus the 'speed' u in *sections per second* is $1/\sqrt{LC}$. Finally, if L^* and C^* are the values per metre in the chain, the speed v in *metres per second* is $1/\sqrt{L^*C^*}$.

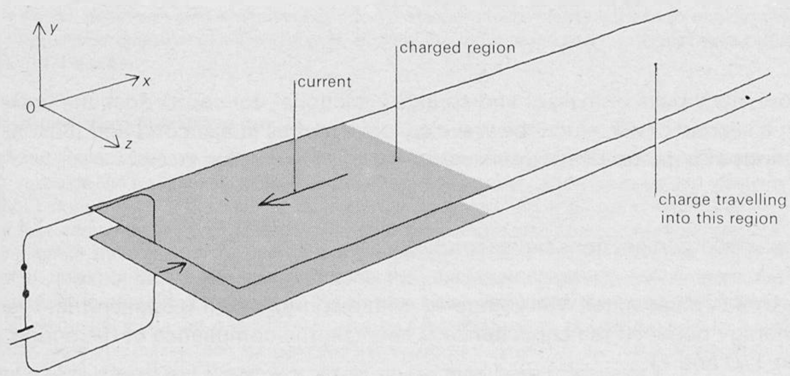


Figure 54
Two long parallel plates conveying a 'switch-on' signal.

Waves near conductors

This part is about what the inventors of electric telegraphs wanted to know: 'How fast does a signal go along a wire?' The answer is usually 3×10^8 metres per second.

The speed is mostly controlled by things which are outside the wire, where there must be changes of electric and magnetic fields accompanying changes of current. The reason for being interested in these changes is that they also happen in electromagnetic waves going through spaces without wires.

The system which is discussed is a very long pair of parallel plates, between which an electromagnetic wave travels at speed $1/\sqrt{\varepsilon_0\mu_0}$, as a charged region spreads along the plates at the same speed, following the connection of a battery to one end. The problem is like that explored for coaxial cable in experiment 8.10, but the geometry is different, using flat parallel plates, which have been discussed previously in the course.

If the analogy can be trusted, we can write down the speed of the electrical wave at once:

$$u = 1/\sqrt{LC}$$

where u is the speed in sections per second, L having been written for m, and $1/C$ for k.

The inductors used have had their inductance estimated in previous experiments in Unit 7 (7.12); it is of the order 10 H (substitute the measured value). If the capacitances are each 10^{-3} F, the speed u comes to 10 sections per second. A pulse should take about a second to traverse the ten-section line.

The finite speed of a compression wave along a row of masses and springs is the result of the time taken for a mass to acquire a velocity under the push of a spring from behind it, and the time taken for it to move to cover the displacement which compresses the next spring. Similarly, the speed of the slow electrical wave is finite because of the time taken for a current to grow in an inductor under the p.d. due to the charged capacitor behind it, and the time taken for this current to deliver charge to the next capacitor, so raising the p.d. across it. The larger the masses and the softer the springs (small k, large compression for a small force), the slower the compression wave. The larger the inductances and the larger the capacitances (much charge for a small p.d.) the slower the electrical wave.

Without detailed calculation one cannot be sure that a pulse would travel along the row without any change. It might become much wider and shallower as it goes along, but still remain recognizable. But the teacher may assert that exact calculations do predict zero distortion, with infinitesimal and 'pure' components, though there are complications if the pulse is as quick to change size as the time taken for the disturbance to go from one section to the next.

Waves near conductors

'I have now got materials for calculating the velocity of transmission of a magnetic disturbance through air founded on experimental evidence, without any hypothesis about the structure of the medium. . . . The result is that only transverse disturbances can be propagated and that the velocity is nearly that of light.'
(James Clerk Maxwell, from a letter to G. B. Stokes written in 1864.)

Experiment 8.10, with the coaxial cable, suggests that electrical signals travel along wires at speeds comparable with the speed of radio waves or of light. (Further experiments would show that the speed in experiment 8.10 is the speed of radio waves in polythene, less than that in empty space, while the speed of signals along straight wires with nothing between them is equal to the speed of radio waves, or of light, in empty space.) The argument which follows is an attempt to understand this situation better. It ends by relating the speed to the electrical constants ε_0 and μ_0, previously introduced (for steady or slowly changing fields) in Units 3 and 7 respectively. To keep things simple, it is easiest to abandon the coaxial system, good though that is in practice, and to discuss a less practical but theoretically more tractable system; a pair of very long parallel plates, as shown in figure 54.

The advantage of discussing waves associated with conductors is that the B-field may be presented as arising from currents in the conductors, so avoiding the need to introduce a 'displacement current' term, with B depending on dE/dt, as Maxwell had to for waves in free space. Yet the description given of the waves is very like that for waves in free space, and the description of the wave in the parallel plates is like that of a plane wave. Similar equations, $E = Bv$, and $B = \varepsilon_0\mu_0 E$, apply to both. An alternative argument is offered on page 84. This exploits the result $v = 1/\sqrt{LC}$, where L and C are the inductance and capacitance per metre, which may have been obtained in connection with experiment 8.11. Despite the dangers of assimilating a result for lumped components to one for distributed capacitance and inductance, this alternative may seem clearer to some students, especially those for whom the assumptions about the way the fields behave seen implausible. The alternative argument offers less by way of insight, but still produces the vital result $1/\sqrt{\varepsilon_0\mu_0}$ for the speed.

Pace of teaching

The argument about the propagation of a signal along the plates shown in figure 54 is developed from here to page 87. It is long because it is unfamiliar and has been set out in detail. It is also long because it is first developed arithmetically; a briefer algebraic version appears on page 85.

The whole argument should be presented at a fair pace, since it is more important that students see it as a whole than that they puzzle over every detail. It is not a thing to be learned by heart; rather, it is a way, which students can see, of using earlier work.

Teachers should consider whether or not to use the arithmetical version at all. The arithmetic seems to some of us to make the whole argument seem more concrete, but some may find the algebra both quicker and clearer.

Students' book

Question 39 develops an argument for the speed of a pulse between the plates, as an extension of the result $v = 1/\sqrt{LC}$. Question 40 does the same thing, following more closely the line of argument in this *Guide*.

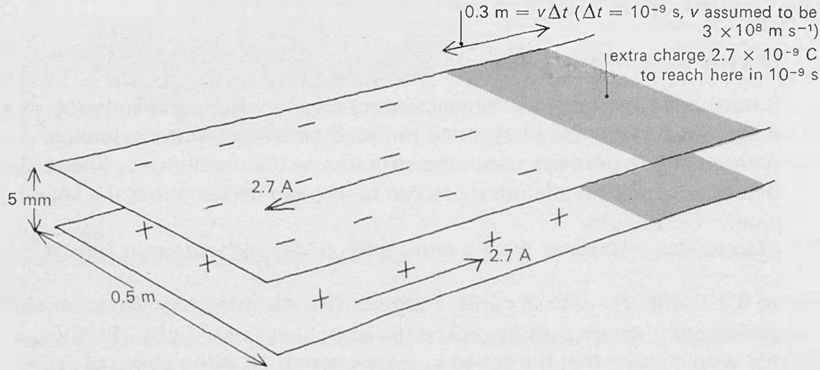


Figure 55
Numerical example of the charge spreading along the plates.

Such a pair of plates could in imagination consist of two lengths of aluminium foil 0.5 m wide and 5 mm apart. At the near end there would be a switch and a cell so that a steady voltage could be applied. Alternatively, one could apply quick pulses with the pulse generator from the coaxial experiment. When the switch is closed, what happens when a steady voltage is applied?

Can it be assumed that the voltage moves along the plates as it did in experiments 8.10 and 8.11, leaving a fairly sharp boundary between charged and uncharged parts? If so, then, by considering the electric and magnetic fields between the plates, it is possible to work out the speed at which the change in voltage travels. The argument for a radio wave in space is similar: which is why this deserves study.

The fields between the plates

What kind of field must there be between the plates, over the region where they are charged? (An electric field, more or less uniform up to the edge of the charged region directed vertically along $0y$ if the lower plate is positive. See figure 56.)

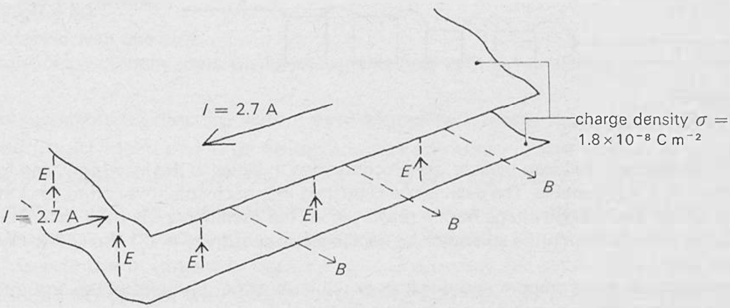


Figure 56
E- and *B*-fields between the plates.

If the battery voltage V is 10 volts, and the spacing d is 5 mm, the field E is equal to V/d, that is, 2000 volts per metre. The amount of charge on each square metre of either plate, σ, is given by,

$$\sigma = \varepsilon_0 E$$

which is about 1.8×10^{-8} coulomb per square metre ($\varepsilon_0 \approx 9 \times 10^{-12}$ C^2 N^{-1} m^{-2}).

How does the charge get to the distant regions into which it travels? (Along the plates.) Thus each plate carries a current, from the battery to the edge of the charged region, where it delivers charge to the next 'empty' bit of the plates, as indicated in figure 55. Suppose the signal does spread along the plates at the speed of light, 3×10^8 m s^{-1}. In one nanosecond (10^{-9} s), the charge to cover an area 0.3×0.5 m is delivered. Using the charge density found previously, this comes to 2.7×10^{-9} C in 10^{-9} s, requiring the surprisingly big current of 2.7 A to be flowing.

Speed of charges and speed of signal

The signal travels very fast. Where it has passed, there is a region containing slowly moving electrons. These move fast enough to form a current big enough to deliver charge to the front of the wave at the required rate. They do *not* travel as fast as the wave; indeed they usually travel some 10^{10} or more times slower.

Demonstration

8.12 Magnetic field in a flat solenoid

1037 set of solenoids

1079 flat solenoid

1039/1 search coil (axial)

1057 a.c. ammeter

59 l.t. variable voltage supply

64 oscilloscope

1000 leads

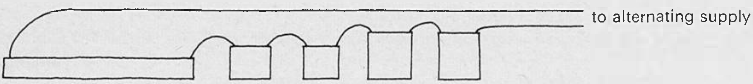


Figure 57
Comparison of fields in solenoids.

See figure 57. The solenoids, laid side by side, are all connected in series to the l.t. supply, and to the ammeter. A current of 1 A is suitable. The search coil is pushed into each solenoid in turn, and the peak to peak distance of the alternating voltage from it displayed on the oscilloscope is used as a measure, in arbitrary units, of the field within the solenoid. An oscilloscope sensitivity of 0.1 V cm^{-1} is required.

The demonstration should be as smooth and quick as possible so as not to interrupt the argument. Indeed, it may be better to defer it until the argument is completed.

It is also possible to bend a sheet of conducting material into the form shown in figure 58 *a*, and to explore the magnetic field inside it with a search coil. An alternating current at 10 kHz supplied to the sheet is suitable. The current is distributed more evenly if it is fed to the sheet via about three pairs of leads in parallel, taken to the edges and centre of the sheet.

What other field must exist near the plates, if such a current flows? (A magnetic field.) It is not hard to calculate this field, and decide what its direction must be, if the plates are compared with a flattish solenoid, as in figure 58.

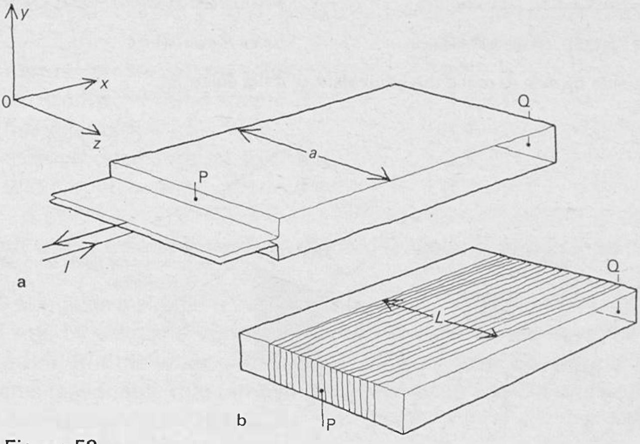


Figure 58
a Oblong solenoid with one turn.
b Oblong solenoid with many turns, giving measurable field with few amperes.

The plates resemble a solenoid with a very elongated cross-section and one single turn in a width a (figure 58). The solenoids already used in the course have more turns per metre length, but the only effect of the smaller number of turns, if the current is spread evenly over the whole distance a, is to give less magnetic field per ampere. The lack of the sides P and Q should make no difference to the magnetic field at a point well inside the solenoid because P and Q are very far away. Students may wish to make measurements of the field in an oblong solenoid so as to know about the magnetic field from experience, and they will find that the field is the same near P and Q as in the middle of the solenoid. The field in a solenoid is given by,

$$B = \mu_0 NI/L$$

if there are N turns in length L. Here there is one turn, in length a, giving

$$B = \mu_0 I/a.$$

For the half-metre wide plates, carrying 2.7 amperes, the B-field is 6.6×10^{-6} tesla, which, as it happens, is about one-tenth of the Earth's magnetic field.

Demonstration

8.12 Magnetic field in a flat solenoid

It may help students who are doubtful of the comparison of the flat plates with a flat solenoid, to see a quick demonstration of the field in such a solenoid. The main problem is getting the orientation right: the wires must run in the same direction as the *length* of the long pair of plates, 0x in figure 58, so that the open ends of the solenoid compare with the open sides of the plates, and the B-field lies along 0z.

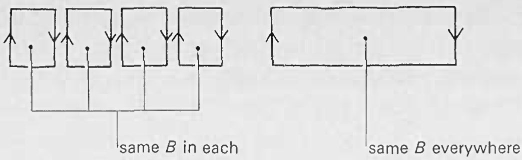


Figure 59
Putting identical solenoids side by side to make the equivalent of a flat solenoid.

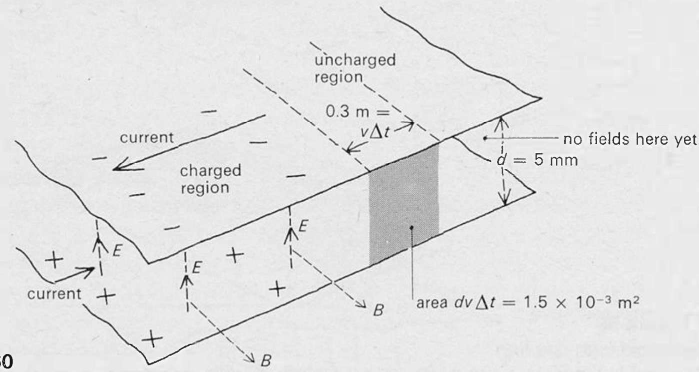


Figure 60
B-field sweeping along into new area between the plates.

The voltage across the plates

The argument on the text page again takes the view, used before in Unit 7, that the voltage needed to maintain a rate of increase of flux is equal to that rate of increase, rather than arguing, as may equally validly be done, in terms of an induced e.m.f. opposing the battery e.m.f.

The second style of argument needs care, for it may seem to suggest that at the edge of the pulse there is no voltage across the plates, two opposite voltages being supposed to cancel. Whatever the argument, there is a voltage V across the plates, equal both to $\varepsilon_0 Ed$ *and* to Bdv. The voltage across the plates is only zero where they are uncharged and where no flux is yet entering the space between them.

To show that the field in a solenoid depends only on the current and density of turns, and not on the shape, the flat solenoid can be connected in series with all the other four in the set of solenoids, as in figure 57, measuring the field at the centre of each in turn. The results may be roughly:

	***B*-field in arbitrary units**
narrow, close-wound solenoid	6.6
wide, close-wound solenoid	6.6
flat solenoid	6.4
narrow, open-wound solenoid	3.3
wide, open-wound solenoid	3.3

The slightly lower result for the flat solenoid is the consequence of its shortness.

Figure 59 suggests a simple reason for this result. If several identical square solenoids were laid side by side, and connected in series, each would have the same B-field in it as any other. But the wires which divide the solenoids carry equal opposing currents, and it seems reasonable that removing such wires to make a wide solenoid would make no difference.

Voltage needed to carry magnetic flux into new regions

Presumably the B-field does not extend, in the space between the long plates shown in figure 55 and again in figure 60, beyond the place where the current ends, which is where the charged region ends, since the current is delivering charge to that place. Thus the B-field occupies more and more of the space between the plates as the *signal sweeps along*.

If the plates have spacing $d = 5$ mm, in time $\Delta t = 10^{-9}$ second as before, the signal sweeps along an extra distance 0.3 metre, supposing the velocity to be 3×10^8 metres per second. The B-field occupies an extra area $dv\Delta t$, which is 1.5×10^{-3} square metre in the numerical example.

For magnetic flux to occupy new areas requires a voltage, equal to the rate of change of flux. This rate is given by the B-field multiplied by the extra area occupied, and divided by the time taken. B was found previously to be 6.6×10^{-6} tesla: putting in the other values, the voltage V is given by,

$$V = 6.6 \times 10^{-6} \times 1.5 \times 10^{-3}/10^{-9} \approx 10 \text{ V}.$$

This is just the battery voltage with which the calculation began. (If an exact value of ε_0 had been used, the agreement would have been better.) The only quantity imported from outside electrical theory was the speed, 3×10^8 m s^{-1}. What would have happened in the calculation if a smaller or larger speed had been assumed? If the speed were smaller, less charge would need to be conveyed in 10^{-9} second, since the area to be filled up would be smaller. Thus the current, and so also the B-field, would be smaller. Then again, this B-field would occupy a smaller area in the same time, so the voltage needed would be smaller, both on that account, and because the B-field

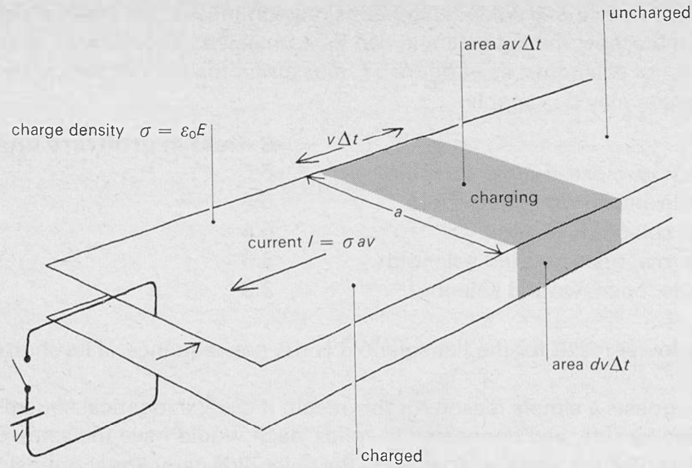


Figure 61

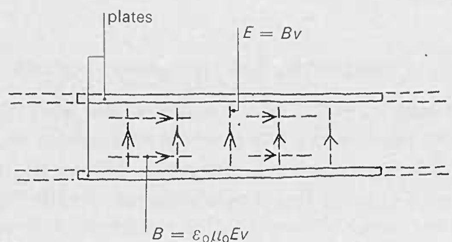


Figure 62
The E- and B-fields between the plates (plates seen end-on).

Alternative argument

If students think it reasonable (from question 38) that the velocity will be $1/\sqrt{LC}$ where C and L are the capacitance and inductance per metre, their values can be written down and substituted.

$$C = \varepsilon_0 a/d \quad (\mathrm{F\ m^{-1}}).$$

The magnetic field for current I is $B = \mu_0 I/a$. If the plates have a spacing d, a one metre length encloses area d perpendicular to B, so the flux between unit length is $\mu_0 Id/a$.

If the current changes at one ampere per second, the flux changes at the rate $\mu_0 d/a$, which gives a voltage also equal to $\mu_0 d/a$.

The voltage induced per unit rate of change of current is the self-inductance.

$$L = \mu_0 d/a \quad (\mathrm{H\ m^{-1}})$$

$$v^2 = \frac{1}{LC} = \frac{1}{(\varepsilon_0 a/d)\,(\mu_0 d/a)} = \frac{1}{\varepsilon_0\mu_0}.$$

Students' book

Question 39 develops the above argument.

would not be so large. The battery could send the fields along faster, and would doubtless do so.

It seems that the reason the speed *is* 3×10^8 m s^{-1} might be that this is the speed at which the battery voltage is just what is required to sweep along the magnetic flux created by the current needed to charge the plates to that same battery voltage, as the electric field also sweeps along the plates.

What is not clear, because the arithmetic conceals it, is where the special value 3×10^8 m s^{-1} comes from. Here is a case in which algebra is better than arithmetic, and, given the same line of argument, it is easy to follow the calculation in symbols.

Why electrical signals go at 3×10^8 metres per second

The following argument uses the results

$$\varepsilon_0 E = \sigma$$
$$B = \mu_0 I/a$$

employed on pages 79 to 81, to show now that the signal speed is related to ε_0 and to μ_0, but not to the dimensions of the plates shown in figures 55 and 56, and again in figure 61.

If the battery voltage is V, and the plate spacing is d, the electric field E between the charged part of the plates is V/d. The charge density σ is given by $\varepsilon_0 E = \sigma$.

Let the signal travel at speed v. In time Δt, if the plates have width a, the charge spreads onto area $av\Delta t$. The charge running along the plates is thus $\sigma av\Delta t$, and the current must be σav.

If there is a current I, there is a magnetic field B, equal to $\mu_0 I/a$. Inserting the value of the current, B is given by $\mu_0 \sigma v$, the width a cancelling. Inserting the value of σ in terms of E gives the important equation,

$$B = \varepsilon_0 \mu_0 E v.$$

Another equation between E and B comes from thinking about the rate of change of magnetic flux as the signal travels down the plates. As before, the edge of the signal travels a distance $v\Delta t$ in time Δt, so the B-field occupies new area $dv\Delta t$ in this time. The extra flux between the plates is $Bdv\Delta t$, and its rate of change is Bdv.

The voltage V across the plates must be sufficient to maintain such a rate of increase of flux, and will be numerically equal to Bdv. But a voltage across the plates goes with an E-field between them, with $E = V/d$. Dividing $V = Bdv$ by d gives the second important result (see figure 62),

$$E = Bv.$$

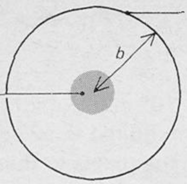


Figure 63
Coaxial cable.

Note on other geometries

For an air-cored coaxial cable, figure 63, the capacitance C and inductance L per metre are given by,

$$C = \frac{2\pi\varepsilon_0}{\ln b/a}$$

$$L = \frac{\mu_0 \ln b/a}{2\pi}$$

$$v^2 = \frac{1}{\varepsilon_0\mu_0} \quad \text{as before}$$

The speed is always $1/\sqrt{\varepsilon_0\mu_0}$ for conductors which are straight in the direction of the disturbance and are not surrounded by other material. Such a result suggests that it is the space which counts, not the wires.

Maxwell's calculation

It is best not to mention the 'displacement current' introduced by Maxwell, because it was suggested by an imaginative model of the 'ether', and used by him to cope with the fact that the magnetic field was at the time given only in terms of current distributions. With hindsight, and help from relativity, we can now see that a magnetic field distribution ought to go with a changing electric field.

It is interesting to note that Maxwell's theory developed slowly. In 1855–6 his paper 'On Faraday's lines of force' began the job of quantifying Faraday's field ideas. Not until 1861–2, did the paper 'On physical lines of force', which first obtained the speed of electromagnetic disturbances, appear. This paper argues almost entirely by analogy, employing a mechanical model of the 'ether'. In 1864 it was followed by the final paper, 'A dynamical theory of the electromagnetic field', in which Maxwell's equations, and the derivation of a wave equation, appear in something like their modern form.

Maxwell's writing was not easy to understand, nor did everyone think it altogether plausible. Not until Hertz showed that such waves could be made and detected in the laboratory, was there good cause for a general acceptance of the new view relating electricity, magnetism, and optics.

Waves in space

An argument, parallel to that given for the long plates, but for the velocity of a wave in space, is hard. It is hard because the currents in the plates are no longer there to give a reason for the existence of a magnetic field. The equation, for a plane wave,

$$B = \varepsilon_0\mu_0 E v,$$

has then to come from an equation linking spatial variations of B with the rate of change of E. Combined with $E = Bv$ the velocity follows as before. See Appendix B for more details.

These two equations between E and B can only both hold at the same time if v has one special value. Substituting in one equation from the other gives,

$$v^2 = 1/\varepsilon_0\mu_0.$$

This special speed, usually given the symbol c, depends not at all on how wide or how far apart the plates may be. It depends only on the values of the electrical constants ε_0 and μ_0. Their values are:

$$\varepsilon_0 = 8.854\times10^{-12}\ \mathrm{C^2\ N^{-1}\ m^{-2}}$$
$$\mu_0 = 4\pi\times10^{-7}\ \mathrm{N\ A^{-2}}$$

so $$c \approx 3\times10^8\ \mathrm{m\ s^{-2}}.$$

The importance of the calculation of *c*

Maxwell, between 1855 and 1864, was the first man ever to show that the speed $c = 1/\sqrt{\varepsilon_0\mu_0}$ could be understood as the speed of an electrical wave. Others had noticed that the combination of these two constants from electricity gave a speed, and it was a matter of speculation whether the nearness of the speed to the measured speed of light was a coincidence or not.

Before Maxwell, these were disconnected facts, puzzling but not illuminating. What Maxwell did was to show that the speed could be deduced from the basic laws of electric and magnetic fields. His achievement was greater than that of the calculation for the two long plates; indeed, his calculation was a different one. He showed much more: that the laws of electric and magnetic fields could be made to predict waves in empty space where there were no currents, as well as obtaining the same expression for the wave speed, that is, $1/\sqrt{\varepsilon_0\mu_0}$.

This quantity was measured electrically by Weber and Kohlrausch (1857), who obtained $3.1\times10^8\ \mathrm{m\ s^{-1}}$. As Maxwell said, 'The only use made of light in the experiment was to see the instruments.' The velocity of light had been measured directly, values available to Maxwell being:

Fizeau $3.14\times10^8\ \mathrm{m\ s^{-1}}$
Foucault $2.98\times10^8\ \mathrm{m\ s^{-1}}$
Bradley $3.08\times10^8\ \mathrm{m\ s^{-1}}$.

Before Maxwell, there were three subjects, electricity, magnetism, and optics. After Maxwell, there was one subject combining all three. And, as will appear later in this Unit, Maxwell's ideas contained the seed of a further unification, the theory of relativity, which linked dynamics into the same total pattern, so that the velocity c, previously the preserve of electromagnetism, became a constant of dynamics as well.

Getting electromagnetic waves away from conductors

Ideas of what would happen between a pair of long parallel plates have been useful so far. But always the electricity on the plates may be considered as the cause of the fields between them. Can we imagine that the fields could behave in the same way in spaces far distant from any conductor?

Waves coming out of plates

The plausibility argument given in the text at this stage about waves coming out of plates ought not in honesty to be presented as a reason for thinking that they should emerge, but as a simple picture to make sense of the experimental fact that they do.

The picture is qualitative and very limited. It could not be extended to deal adequately with reflection and transmission at the end of the plates, nor with subsidiary diffraction minima. It is dubious in explaining the direction in which the disturbance is travelling where the field is curved in figure 68. These problems need three-dimensional differential equations for their solution and are beyond the scope of sixth forms. (See Appendix B.)

Speed of the back of a pulse

The argument for the speed of a pulse between the plates seems to show that a 'rising' edge of field and charge will propagate at this velocity. Students may wonder if a 'falling' edge will propagate at the same velocity.

In effect, this has already been assumed, for it was supposed that the edge remained sharp: that is, that its profile propagates unchanged. And a question about the back of the pulse is just this problem over again – does the profile change?

So an argument, the converse of that given already, for the speed of the back of the pulse might seem to add conviction, but would be arguing in a circle.

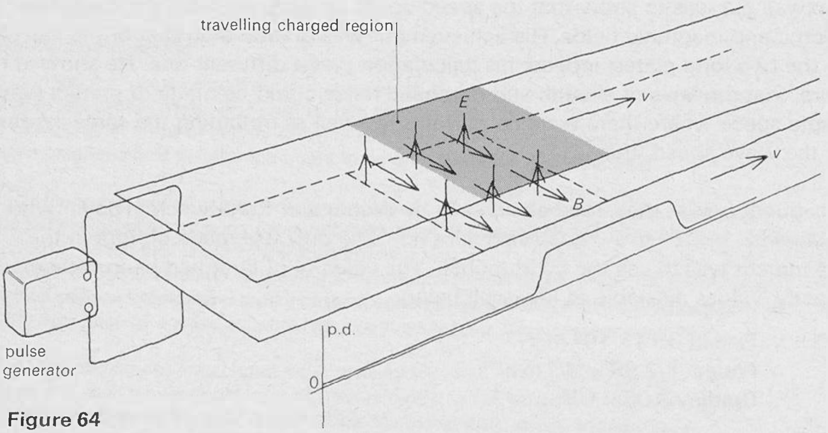


Figure 64
A short on–off pulse travelling along the plates.

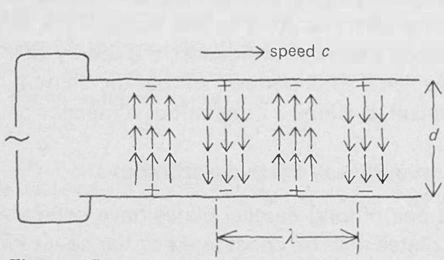


Figure 65

Microwave links often use an arrangement in which the signal travels inside a conducting metal tube ('waveguide') to the focus of a concave mirror. The signal comes out of the end of the tube, spreads out, hits the mirror, and is reflected as a nearly parallel beam towards the distant receiver. Permanent towers carrying such equipment for sound and television transmission are becoming common in Great Britain and elsewhere. The 3-centimetre transmitter (item 184/1) and wax lens (item 1014) can do the same (but a Post Office licence is needed under the Wireless Telegraphy Act of 1949 for all but the simplest educational demonstrations). The process at the end of a waveguide is complicated, but the following line of thought may make its happening seem more possible.

The following discussion suggests the kind of thing that might happen, not why it happens. A critical student might prefer an empirical attitude: radio waves do emerge from waveguides and their measured velocity is the same as that predicted above, suggesting that what comes out is not unlike what was inside.

Fields and charges in a pulse

Imagine a pulse generator connected to the end of the long parallel plates. Experience with experiments 8.10 and 8.11 suggests that the back of a pulse, produced if the switch in figure 61 were opened suddenly, will travel along as fast as the front, where the voltage comes on. Figure 64 suggests the idea.

The natural speed for a pulse to pass along the plates is c or $1/\sqrt{\varepsilon_0\mu_0}$ metres per second.

The pulse consists of

1 Equal and opposite electric charges on the two plates so that the total quantity of electricity taken with the pulse is zero.

2 A little patch of electric field E between the opposite charges on the plates.

3 A little patch of magnetic field B between the opposite currents on the plates.

Which of these parts would be absent were something like this pulse to emerge and travel away, perhaps as a radio message to the Moon? The charges would be absent, for empty space contains no conductor to carry them. (Radio waves are well called 'wireless' waves.) So it may be reasonable to think about how the charges might become unnecessary.

A way for waves to get out of the plates

What happens when the electric and magnetic fields reach the end of the plates? To guess the answer to this question we must put together ideas from several earlier parts of the course. Suppose that it is not a single pulse of E and B that runs along the space, but the result of an alternating potential difference being applied to the lefthand ends of the plates. (See figure 65.)

Lines of force

The field patterns in the descriptive sketches, figures 65 to 68, could have been represented by lines which Faraday would have called 'lines of force'. Such lines map directions in which test charges would experience forces. They do not represent physical links joining particular positive and negative charges. So we prefer the broken arrow representation used in these sketches.

It is convenient to speak, when discussing such pictures, of the piece of field which is now 'here', and which later appears 'there', and no student is likely to appreciate any cautious noises one may make. But it may be well to bear in mind that these fields are not 'things' with a conservable physical presence that can be caught hold of. Indeed, one merely has to travel parallel to the plates for some of the E-field to become B-field.

Although the E- and B-fields in such field pulses are an inseparable unity, it is simplest to draw only one of them for the purposes of this argument. But students should not suppose that the pulse is made of two distinct components which are independent of each other. See Appendix B for a further discussion.

Why figure 66 is wrong

The electric field cannot suddenly change magnitude in the y direction, and remain uniform in the xz plane, unless there is a sheet of charge over the surface where it changes size. This follows at once from Gauss's theorem, for instance.

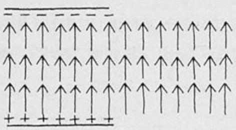


Figure 66

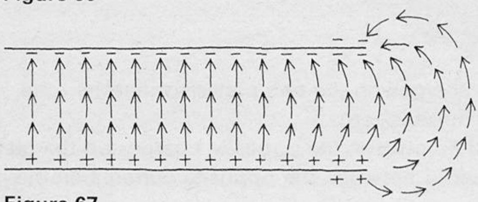


Figure 67

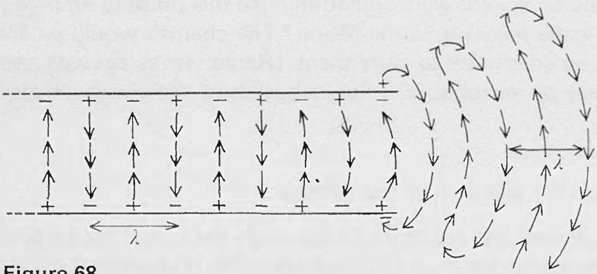


Figure 68

If d, the distance apart of the plates, is much less than λ, the 'wavelength' or distance between successive patches of the same field direction, we could never expect plane wave fronts of electromagnetic radiation to come out into open space. No sort of wave front so far studied (ripples, light, 3 cm waves) ever emerges from a narrow hole or slit without spreading out widely afterwards. If it were to do so in figure 65 the bands of electric (and magnetic) field would have edges in empty space, as in figure 66. Displays of the E-field with small insulating particles or fibres, and of the B-field with iron filings have never shown that happening. Figure 67 shows what we could expect. The positive and negative electric charges must stop moving at the end of the plates, and the familiar field patterns due to a pair of parallel plates will appear. The charge density along the plates will be nearly uniform, but changing slightly at the ends, where the field will be curved, and a little of the charge will go to the outside of the plates. Sensitive field detectors would show a field curving right round the ends of the plates. When the charge has been halted one thing that can happen, as in experiment 8.10, is that it would go back from right to left, forming a reflected pulse of the sort which was detected by the oscilloscope in that demonstration. The pulse at the end is not different from a pulse put there by a generator, so it ought to travel back.

But if d is big in comparison with λ, as in figure 68, we get the kind of situation in which nearly plane wave fronts have emerged from a wide aperture in optical diffraction experiments. And there is also a means by which an electric field can do so. For when a thin slice of downwards field reaches the end of the plates, and as before, extends round the plates to charge which has gone to the outside, it is immediately followed by a thin slice of upwards field between the plates. So the field is in the same direction above and below each plate and would be continuous across each plate but for the plates being conductors on which equal and opposite charges can be induced, top and bottom. The net charge near the ends of each plate is zero. But carrying charge is all that the plates are there for, and if they carry no charge they need not be there. It may not matter that they come to an end. One might speculate that the fields could become continuous, as they are elsewhere where there are no charges, and travel away as in figure 68.

In this way nearly plane wave fronts can be produced. The direction of travel is at right angles to both fields in the plane part of the wave front. At the edges where the wave fronts are curved they also travel outwards and contribute to the diffraction effects which are always associated with apertures. In the nearly plane parts of the wave fronts the magnetic field is at right angles both to the electric field and to the velocity. The edges of the magnetic field are like the edges of the electric field.

A wave of two fields travelling together

Electromagnetic waves do escape into space from conductors, as a matter of almost daily observation by anyone equipped with a radio set. Given that they do, the arguments which go with figures 65 to 68 may make this seem a plausible fact, and might incline one to think it likely that such a wave in space would be like the wave between the plates. Maxwell was able to argue that these things should be so, given the laws of electric and magnetic fields.

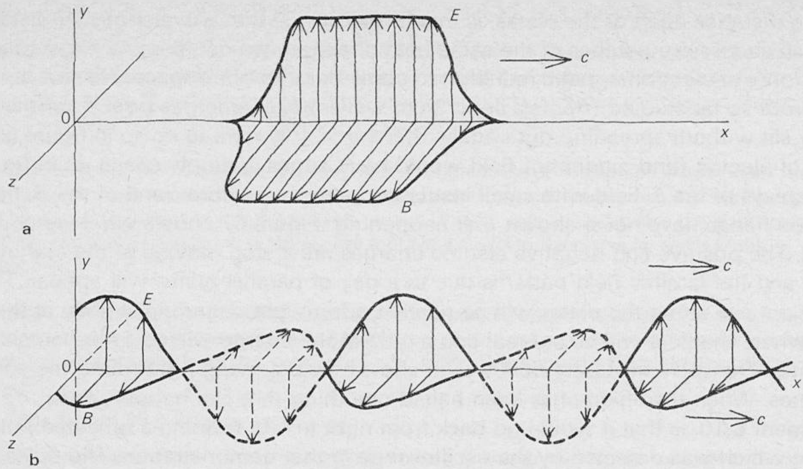


Figure 69

Further arguments about travelling fields

Pages 93 to 101, which follow, contain an optional further stage of argument about electromagnetic waves. Teachers may gain something from looking at them, and it is possible that some students may like to hear a bit more of the story. But we recommend most to pass straight to the discussion of polarization, on page 101, which takes up the description of electromagnetic waves given in the text, at a practical level. Time is short, and the main aim of supporting a description of an electromagnetic wave with experiments and argument which make it seem moderately plausible is unlikely to be furthered to any great extent by yet more argument.

Speculation and the development of new theories

The difficulty here is to make Maxwell's work seem important without doing Maxwell's theory. Some indication of how ideas have developed may be helpful, especially pointing out the separateness of '*light*', '*electricity*', and '*magnetism*' during much of the history of science.

It may help to mention the division of opinion between those (well represented by Ampère) who wanted to think only in terms of forces between things, and those (Faraday most of all) whose instincts were that the field would be itself at least as valuable a subject of thought as the things whose interactions it described. This was no issue of fact: neither party's views made any difference to the experimental facts. It was a matter of hunches about how best to invent good new ideas.

Faraday wrote in June 1852, 'it is not to be supposed . . . that speculations . . . are useless, or necessarily hurtful, in natural philosophy. They should ever be held as doubtful, and liable to error and to change, but they are wonderful aids in the hands of the experimentalist and mathematician.' (*Experimental researches in electricity*, III, Art. 3244, Dover.)

Of Faraday, Maxwell wrote: 'We are probably ignorant even of the name of the science which will be developed out of the materials we are now collecting, when the great philosopher next after Faraday makes his appearance.' (*Nature,* VIII, 1873.)

With hindsight we can name the new science – relativity – and the philosopher – Einstein.

As a result, we are led to describe electromagnetic waves as follows. Such a wave is made of two fields at once, B-field and E-field. Neither is present in the wave without the other. They are found at right angles to one another, and both are at right angles to the direction of propagation. In empty space, this bunch of fields travels at the unique speed $c = 3 \times 10^8$ m s^{-1}, where c is equal to $1/\sqrt{\varepsilon_0 \mu_0}$. Figure 69 shows what two such waves might be like, one a pulse and one a sinusoidal variation. Notice that E and B are in step, rising and falling together in free space, just as it was plausible to think that they might within a pair of long parallel plates.

No student should think that sound reasons have been given for thinking that this description is correct. The wave speed has been worked out for one special case only, and it must remain for more advanced work to show that the results are general. The directions of the fields in empty space do happen to be like those within the plates, even when there are no plates carrying charges and currents to guide the thinking about directions. Nor is the description a complete one. It suggests that the two fields, though associated, are independent of one another, for example. It would be nearer the truth to say that there is just one complex electromagnetic field. Many a good story has a sequel which one must wait a while to read.

Optional: moving fields

'What led me more or less directly to the special theory of relativity was the conviction that the electromotive force acting on a body in motion in a magnetic field was nothing else but an electric field.'
(Einstein, A., 1952, from a letter to the Michelson Commemorative Meeting of the Cleveland Physics Society.)

'According to theory it was not necessary . . . to produce special magnetic waves; the electric waves should at the same time be waves of magnetic force . . .'
(Hertz, Electric waves, *page 18. Dover.)*

In the argument about the signal travelling along a pair of parallel plates (pages 79 to 87), the E- and B-fields seemed to play a subsidiary role. The charges and currents on the plates seemed more important, and the fields appeared to be just a consequence of the presence of this moving electricity. In any case, are not fields just intellectual devices for making calculations less cumbersome, lacking any extra interesting properties of their own?

The trouble with doing research in physics is that one can't tell beforehand what is going to be important. Most people in the nineteenth century felt that the fields were not the most fundamental things to think about, but Faraday and Maxwell speculated that it might be otherwise. So, with them, we make an intellectual about-turn. We wonder whether it might be the fields that really matter, and look on the moving charges as just what runs along at the edge of the field. Such a step cannot be made by deduction; it has to be made by guesswork, by intuition, and by the 'smell of the problem'. But this is the way physicists have to work.

Optional demonstration

8.13 Moving fields

a *Moving a magnet and its magnetic field*

1001 galvanometer (internal light beam)

1040 clip component holder

1054 copper wire, 14 s.w.g., bare

92 B Magnadur magnet *20*

92 I mild steel yoke *5*

106/1 dynamics trolley *2*

1053 Plasticine

1000 leads

b *Moving a capacitor and its electric field*
add to above:

1038 Hall probe and circuit box

1025 pair of capacitor plates

1033 cell holder with U2 cells

106/1 dynamics trolley *2*

small insulating blocks about 30 mm thick to serve as plate spacers

8.13a Moving a magnet and its magnetic field

This experiment repeats one done in Unit 7, experiment 7.9a. Details are given there.

The distinction between E and B

Traditionally a magnetic field was something produced by magnets, whereas an electric field was produced by quantities of electricity. The distinction lost some of its simplicity during the nineteenth century, when magnets came to be viewed as exactly equivalent to electric currents which are moving quantities of electricity, although for most of the nineteenth century the distinction between moving and stationary things still seemed an absolute one. But a twentieth-century physics course cannot invite students to think of the frame of reference in which they do magnetic experiments as being stationary in any absolute sense. Consequently we must allow that a magnetic force (and field) may be an electric force (and field) if we observe it from another frame of reference, and vice versa.

E- and B-fields are distinguished by the force F exerted on a test charge q. If q is not moving relative to the observer, F/q is called the E-field. If q is moving, and there is an extra force F', then F'/qv is called the B-field.

Because 'not moving' means one thing to one person and a different thing to a person moving relative to him, so a field may seem to be electric to one person and magnetic to the other.

A short on–off signal going along the plates seems to be associated with a moving slab of magnetic field travelling along between the plates, as in figure 64. That is, a region of high intensity travels along the plates. It has a 'front' and a 'back', at each of which the intensity falls off. There seems also to be a moving slab of electric field, which travels in the same sense as does the magnetic field. No real 'thing' goes along, but a high intensity region goes from place to place. If the fields are the important thing, it may be possible to learn more about the wave by setting slabs of B-field and E-field into motion in other experiments.

Optional demonstration
8.13 Moving fields

8.13a Moving a magnet and its magnetic field

A pair of Magnadur magnets on a steel yoke makes a slab of B-field not unlike that which would go along between the long parallel plates of figure 64, if the switch were put on, and then off again. In Unit 7, such a slab of field was moved near a wire connected to a galvanometer, as in figure 70.

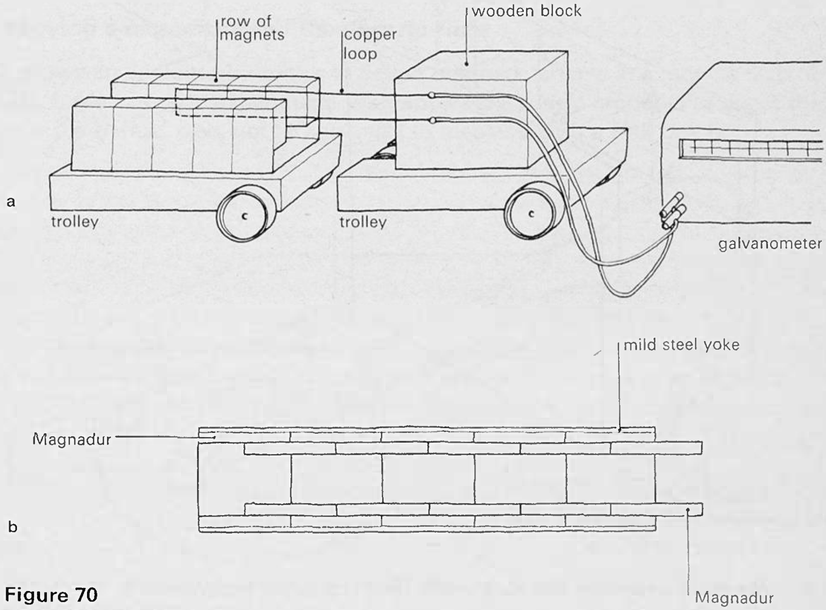


Figure 70
a A moving B-field experiment.
b Construction of magnet.

Suppose the wire is kept over one place on the bench, and the magnet is moved past it. The galvanometer deflects. Clearly, as was suggested in Unit 7, there is a force, upward or downward, on the electrons in the short vertical part of the wire.

Teachers should avoid becoming embroiled in a relativistic discussion of experiment 8.13, but there follow some notes to set the record a little straighter. The equation $E = Bv$ is exact if both E and B are measured in the same reference frame; in this case, B is the B-field one would measure by sending a charge with a small velocity through the gap of the *moving* magnet. The result is not exactly the same as when the magnet is at rest relative to the person who measures the test charge's velocity. The two differ by a factor $\sqrt{1-v^2/c^2}$. Identical remarks apply to the equation $B = \varepsilon_0 \mu_0 E v$. Students are unlikely to think that there would be a difference, and need not be told that there is.

In two reference frames, moving relative to one another, in situations of the sort discussed here there will be a pair of fields E and B in one frame and different magnitudes E' and B' in the other. But it will be true that

$$E = Bv$$

and that $E' = B'v'$

and similarly for the other equation giving B in terms of E. These are simple versions of Maxwell's equations, as explained in Appendix B. They are *not* equations for transforming from one frame to another. Maxwell's equations refer to one frame at a time, but their form stays the same when all quantities in them take whatever values they have in a new frame.

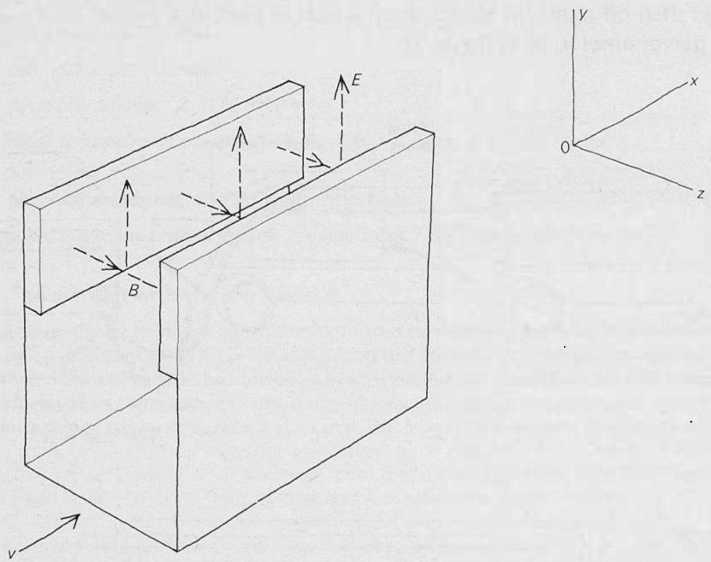


Figure 71
E-field associated with a moving B-field.

8.13b **Moving a capacitor and its electric field ('thought experiment')**

The capacitor plates, far enough apart to allow the Hall probe to be put between them, are mounted on one trolley and connected by long leads to the cell holder so that the bottom plate is positive. The Hall probe is mounted on the other trolley and its leads go to the circuit box and galvanometer. The trolleys are placed so that when either moves there is relative motion between the plates and the probe in the direction we can call Ox. The Hall probe is arranged so that it will detect magnetic fields in the direction Oz, and the electric field is in the direction Oy.

These electrons are not moving past us. What is the name of the force on a stationary charge, such as is exerted on a charged oil drop in the Millikan experiment? (An electrical force.) It is hard to set aside a prejudice in favour of thinking that a force originating in a magnet is a 'magnetic' force, but this is *not* a magnetic force.

A magnetic force is one which is exerted only on a moving charge, $F = Bqv$. The force here is given instead by $F = Eq$ where E is an upward electric field, as in figure 71. How big is it? The answer is given by moving the wire past the magnet, when, for the same relative velocity, the result is just the same. So, if v is the speed of the slab of B-field past a stationary charge,

$$F = Eq = Bqv$$
$$E = Bv.$$

The electric field is real enough, though small. If B is about 10^{-2} tesla, and v is $1\ \mathrm{m\ s^{-1}}$, E is only $10^{-2}\ \mathrm{V\ m^{-1}}$, compared with over $10^5\ \mathrm{V\ m^{-1}}$ in the Millikan experiment, and a good galvanometer is needed to observe the voltage across the few millimetres' length of the end of wire in the moving field.

8.13b **Moving a capacitor and its electric field**

Figure 72 shows an analogous attempt to detect magnetic effects in a moving slab of electric field. It is a 'thought experiment' with apparatus to help students to guess the result, if only the B-field were not far too small to measure with a Hall probe.

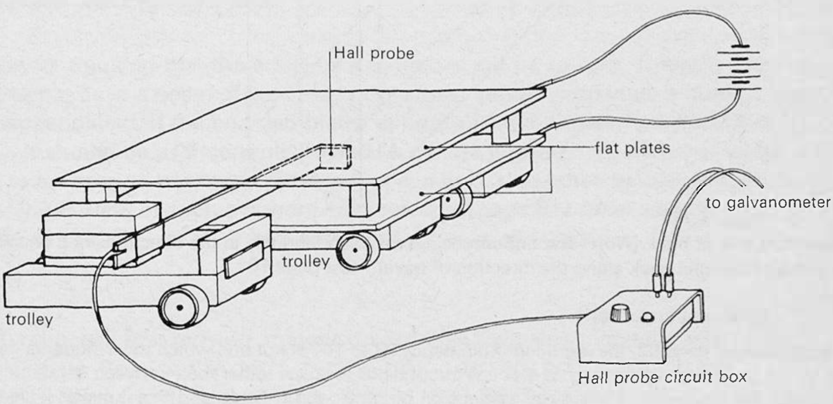


Figure 72
A moving E-field experiment (illustrative only).

Maxwell's displacement current

It is in supposing that the second of the two equations between B and E holds in empty space away from sources, that we make the move equivalent to Maxwell's introduction of the 'displacement current'.

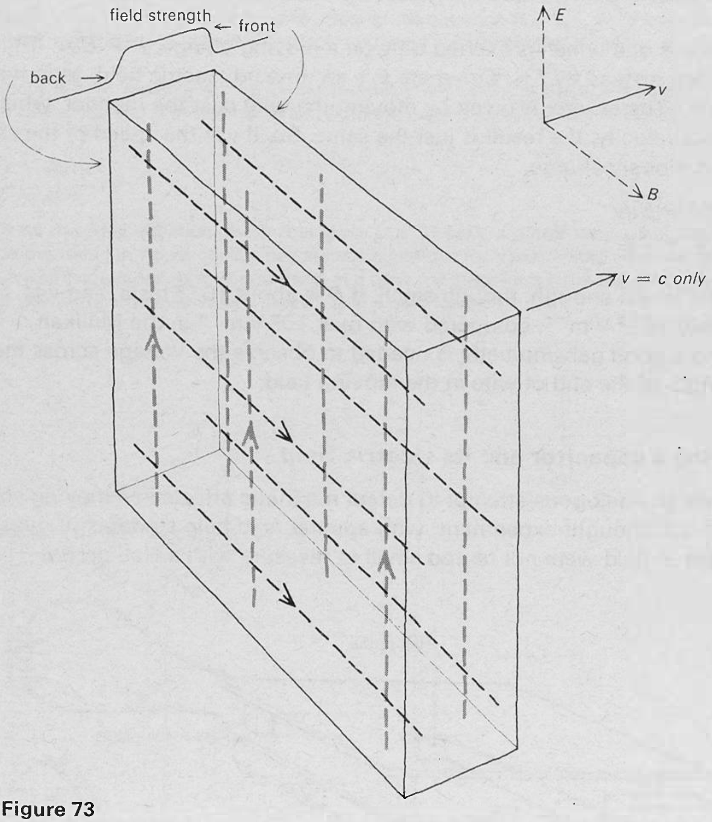


Figure 73
A travelling slab of field. (*Note.* The uniform field extends indefinitely in the directions of E and B, but has a sharp front and back along the direction of travel.) See page 101.

Optional argument

As suggested on page 92, the argument from pages 93 to 101 is not one which most students need hear about. Its virtue, if it has any, is that it is about fields in space rather than between conductors, that it suggests the intimate and necessary association of the B- and E-fields, and that it makes it clearer than might otherwise appear that c is a relative velocity. In a way, the last is the most important point, though by far the most difficult. Not only is c the velocity of the wave relative to a detecting device, but it turns out to be the *same* relative velocity regardless of any steady motion of the detector. This paradoxical state of affairs leads directly to the theory of relativity.

If the Hall probe were more sensitive, and if random magnetic fields (from machinery, thunderstorms, students' pen-knives, and so on) were smaller, the result would be about

$$B = 10^{-17} Ev$$

where v is the speed of the pair of charged plates, and so of the E-field between them, going past a fixed Hall probe. Figure 74 shows the field directions.

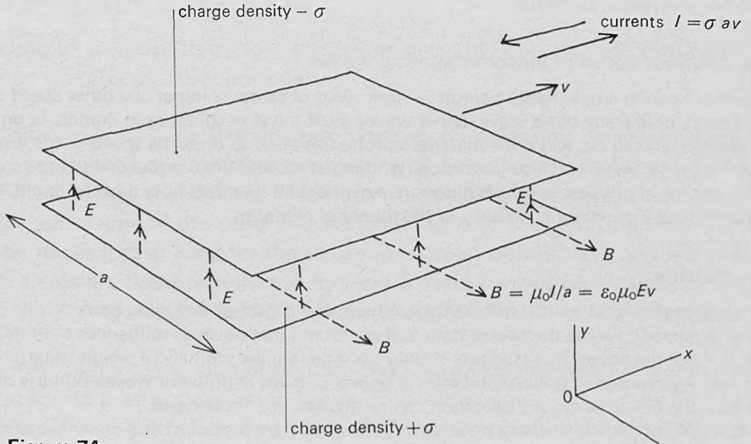


Figure 74
B-field associated with a moving E-field.

The reason for thinking that there will be a B-field is not far to seek. Charges moving past a detector, as in a beam of electrons or protons, should constitute a current and have a magnetic field. If the charged plates in figure 74 are transported sideways at velocity v, it should be as if currents flowed. A charge density σ on plates of width a amounts to a transport of charge at the rate $I = \sigma av$, in coulombs per second, or amperes. The B-field of such a current I has already (page 81) been seen to be $B = \mu_0 I/a$, so that $B = \sigma v$. Remembering that $\varepsilon_0 E = \sigma$, we obtain

$$B = \varepsilon_0 \mu_0 Ev.$$

It might be thought that moving charges do not give the same magnetic effects as currents in wires. Rowland, in 1875, took out an insurance policy against this possibility, by showing that a spinning charged disc did give a magnetic field as expected, just as if the disc were a coil carrying a current.

Moving fields

We have been careful not to imply that by a 'moving field' anything more is meant than a pattern of intensity which goes from place to place. See the note on 'Lines of force' on page 90, and Appendix B.

A field has nothing material to be caught hold of, no conservable physical presence, and can change its name and size if seen from a different reference frame. 'Lines of force' which one may be tempted to imagine as either present or not can thus be misleading in situations where, if they existed, they would be moving. So lines of force may be more trouble than they are worth in describing electromagnetic waves. If a field moves, the 'things' that go from place to place are regions of changes of intensity, like the edges of pulses described previously.

The constant speed of electromagnetic waves

An electromagnetic wave in empty space cannot be conceived of as travelling at any other speed than c. It is, in this respect, unlike any other wave. Other waves must travel in solids, or in liquids, or on ropes, and on springs, and so on, and if the material's properties alter, so does the speed of the wave in it. But an electromagnetic wave can also go through nothing at all, and there is just one unique speed, a fundamental constant of physics, at which it does it. Anyone who wonders how a *speed* could possibly be a fundamental constant is halfway to the theory of relativity.

Polarization

Teachers who omitted pages 93 to 101 will start again here. The object of including work on polarization is to illustrate how the transverse nature of electromagnetic waves influences their use and behaviour. It is not necessary that students should become familiar with Nicol prisms, quarter-wave plates, and the rest. Awareness of polarization effects or lack of them at different wavelengths is more important. Broadly, the electromagnetic spectrum can be divided into three types:

1 Long wavelengths which show polarization when they are produced artificially because the generating current had a direction.

2 Short wavelengths emitted by individual atoms so that typical sources have a random mixture of planes of polarization although devices exist for resolving and separating components.

3 Very short wavelengths for which no easy means of detecting polarization exist (page 107).

There is a lot of overlap in such a classification, but the many complications need no mention to a class. It would only waste time to discuss whether visible light should be put in the second type because of filament lamps or in the first type because of lasers, for example.

If the fields move, they move in pairs

If there is a slab of B-field moving sideways at speed v, there will be a slab of E-field going along with it at the same speed, its direction being at right angles to the B-field and to the direction of motion. If there is a slab of E-field moving sideways at speed v there will be a slab of B-field going along with it at the same speed, at right angles to the E-field and to the direction of motion.

For the first $E \doteq Bv$.
For the second $B = \varepsilon_0 \mu_0 E v$.

We are going to suppose that these equations correctly describe the two fields, even if there is no magnet or capacitor nearby.

In earlier work in Unit 7, on B-fields produced by currents and on electromagnetic induction, we might have said that the moving B-field makes the E-field, and that the moving E-field makes the B-field. This does not really make sense. *Each is necessarily there at the same time*, if the other moves. To be able to say that one *causes* the other, it would be necessary at least for the cause to happen before the thing it caused. But the pair of fields are there together. A moving B-field *cannot be thought of* without a moving E-field. A moving E-field *cannot be thought of* without a moving B-field.

Suppose then that a plane slab of uniform field might be imagined travelling past a detector at speed v, as suggested in figure 73. Neither sort of field may be imagined without the other as part of the travelling slab of fields. Not only must both fields be there, but their relative magnitudes are fixed too. From the two equations developed previously, $E/B = v$ and $B/E = \varepsilon_0 \mu_0 v$.

It follows that *only one* velocity v may be imagined for the travelling field-pairs; the velocity at which both equations balance. This velocity is just c, where

$$c^2 = 1/\varepsilon_0 \mu_0$$

and $E/B = c$ in both equations.

The remarkable consequence is that an electromagnetic wave in empty space *cannot consistently be thought of* as travelling past a detector at any speed other than $c = 3 \times 10^8 \text{ m s}^{-1}$.

Polarization

The electromagnetic wave described previously is transverse in character. The electric field may only point in any of the directions perpendicular to the velocity. (The magnetic field is then at right angles to that.) Such a wave can be polarized.

Most of the effects of polarization of light were known before the nature of light was understood, and it was from them that scientists deduced that in a light wave the displacement is perpendicular and not parallel to the direction of travel.

Demonstration

8.14 Polarization of radio waves

a Polarization of 3 cm waves

184/1 3 cm wave transmitter

184/2 3 cm wave receiver

181 general purpose amplifier

183 loudspeaker (if not part of item 181)

optional: polarization indicating grille

b Polarization of 30 cm waves

1050 15 cm dipoles and oscillator

1001 galvanometer (internal light beam)

1000 leads

8.14a Polarization of 3 cm waves

Set up the transmitter and receiver facing one another, and rotate each about the direction that joins them, as shown in figure 76.

8.14b Polarization of 30 cm waves

It is much easier to use the oscillator rather than the spark system used in experiment 8.9. The signal can be detected with a light beam galvanometer across the receiving dipole (with its diode) and switched to 'direct'. Show the effect of rotating either dipole about the line joining the dipoles. If the dipoles are crossed, a signal is received if a rod, about 0.15 m long, is held between them at an angle of some 45° to each, as in figure 75.

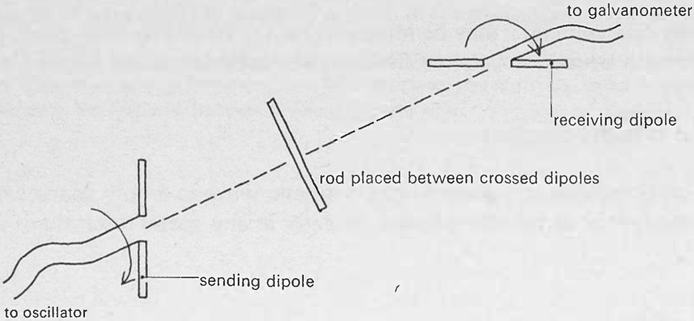


Figure 75
Polarization of 30 cm waves.

Demonstration

8.14 Polarization of radio waves

It is easy to show that the two sorts of radio wave used in the course have properties corresponding to polarization.

8.14a Polarization of 3 cm waves

The class should see that the receiver picks up radiation unless it is 'crossed' with the transmitter. Most equipment commercially available is made so that the electric vector is vertical (in the y-direction, as in the theoretical discussions in this Unit). If the receiver is at some other angle, it detects the component of radiation at that angle.

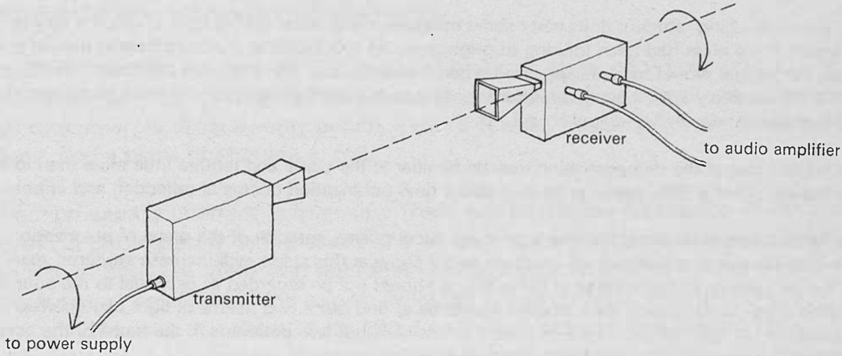


Figure 76
Polarization of 3 cm waves.

With the receiver in its best receiving position, the class can see that the polarization indicating grille will extinguish the radiation if its wires are parallel to the electric vector, and reduce it by varying amounts otherwise.

With the receiver crossed with the transmitter the class may then be asked what the effect of putting the grille with its wires at some angle parallel to neither transmitter nor receiver will be. When they have guessed, they may be shown that the grille allows a component to pass which has itself a component parallel to the receiver, so that the receiver detects microwaves although previously it did not. (This effect may also be seen with three polaroids.)

8.14b Polarization of 30 cm waves

Transmission can be shown with the dipoles parallel and crossed. A single, straight conductor 0.15 m long (or more) can be placed between the dipoles at an angle to make the receiver pick up a signal even when the dipoles are crossed.

Experiment
8.15 **Polarization of light**

1083	polarizing filter *2*
100/1	rectangular plastic tank
94 A	lamp, holder, and stand
94 B	housing shield
94 G	barrier *2*
94 H	plano-cylindrical lens, +7 D
27	transformer
	milk should be available; each experiment requires only a drop or two

The 'non-vanishing' student does carry some message about what visible light is like, if it can be polarized. If the class had been looking through polaroid spectacles at a picture framed behind polaroid sheet, the picture would have disappeared when it was rotated. The student's continued visibility shows that ordinary light, having a random mixture of planes of polarization, is more complicated than electromagnetic waves generated by a.c.

The second part of the demonstration may be familiar to the class, and require little more than to be mentioned. There is little useful to be said about how polarization occurs at reflection and in polaroids.

Any further demonstrations, such as scattering, Nicol prisms, rotation of the plane of polarization, and including the use of a laser, which teachers might show at this stage, will increase students' experience, but the acquisition of knowledge of these things should not be regarded as essential to the course. A possible small investigation for a student would be to find out if two beams of light can interfere if they are polarized at right angles. The experiment needs skill, but two polaroids in the paths of the beams of light in an arrangement like that of experiment 9.3h can be used.

Plane of polarization

The phrase 'plane of polarization' used in older books on optics should be avoided. It was defined before the nature of electromagnetic waves was understood. It is the *xz* plane in figure 69, that is, the plane including the magnetic vector and the direction of travel. Nearly all important effects of electromagnetic waves are actually more directly associated with the electric vector. The phrase 'plane of vibration' is used for the plane of the electric vector, but it is simpler to call the electric vector by its own name.

Polarization by scattering

Each pair of students should arrange a lamp and lens, with barriers as necessary, to send a bright narrow beam of light through a tank of water, as in figure 77. Tap water often contains enough suspended matter for the beam to be visible. If not, add one drop of milk, so that the beam is just visible. It is essential not to add too much, since if the light is scattered more than once before it emerges, the argument opposite will not hold, and the scattered light will be substantially unpolarized.

From the apparatus on the bench, discussion should pass to the positioning of television aerials. Some member of the class may be well informed about this. Microwave links, radar, and radio astronomy often use aerials which illustrate polarization.

Experiment
8.15 Polarization of light

Teachers who do not mind being thought temporarily insane might begin by asking a student to stand on a table, and having established that all the class can see him, get him to lie down, asking if the class can still see him. Naturally they can, but how is this to be reconciled with the results of experiments 8.14a and b? The very absurdity of the problem may help to provoke useful discussion.

Each student can then have a pair of squares of polaroid, through which to look at a lamp while rotating one or both the polaroids. It is worth looking at various reflections through one polaroid, from a shiny bench, a glass-fronted cupboard, a piece of sheet polythene, and a piece of aluminium foil.

The reflection usually changes in intensity. There can be a short discussion of the uses of polaroid in photography, sun glasses, and so on. Students can cross two polaroids and put a third between them, and note that the combination allows light to pass through, for most positions of the middle polaroid. Students have already (Unit 1, experiment 1.11) seen the effect of stretched polythene between polaroids. They could now look at, for example, some thin Cellophane wrapping material. Strips of Sellotape stuck on glass are also worth seeing, for the beautiful colours produced.

Discussion after these experiments should bring out the observable differences between waves which have a displacement at right angles to the direction of travel and those which, like sound waves, do not. Understanding the directions of the electric and magnetic fields in electromagnetic waves should make polarization effects seem natural.

The polarization of scattered light is a direct and simple consequence of light being a transverse wave, and makes a useful conclusion to the brief experience of polarization. Figure 78 illustrates how the argument could go. If unpolarized light comes in from the right to a place P where there are particles which might scatter it, along either PA or PB at right angles to the path of the light, not all the light can be scattered in these directions. Along PA, for example, the light *cannot* include any part of the original light whose oscillation direction was along PA, for then that light would be longitudinal in character. Similarly, the light scattered along PB must be polarized parallel to PA.

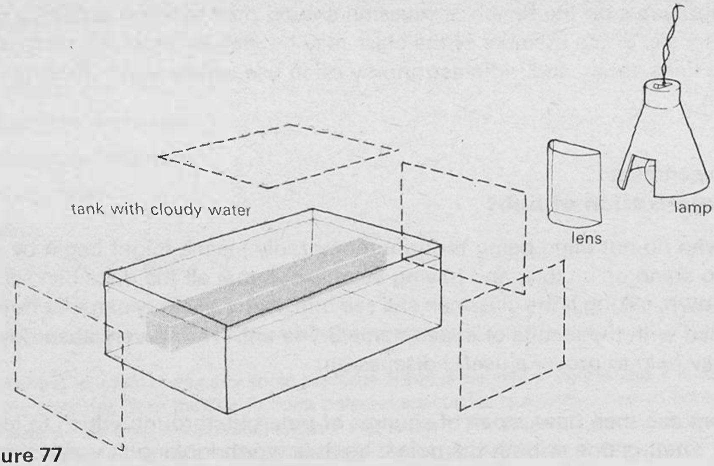


Figure 77
Scattering by cloudy water. (Broken rectangles show positions for a polaroid square.)

Figure 77 shows several positions in which a polaroid may usefully be held and rotated. In some ways, the best place is between the lamp and tank, when the failure of scattered light to emerge along one direction is a convincing support to the argument.

Note that the particles must be small so as to scatter light rather than to reflect it diffusely. Cigarette smoke in air does not polarize by scattering to any great extent for this reason.

Students' book

Questions 47 to 50 are about polarization effects. They involve some home experiments, for which students need pieces of polaroid.

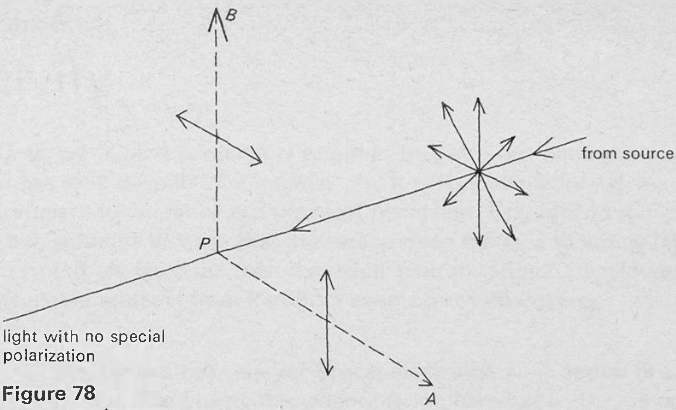


Figure 78
Polarization by scattering.

It is not hard to check these points with a beam of light going through slightly cloudy water. Students should look at the scattered beam along more than one direction, and compare the orientations of a polaroid held so as to cut off the scattered light along these directions. They should also see that the transmitted light is not polarized, and explain why. (Equal amounts of polarized light are scattered in all sideways directions.)

If, as seems probable, the scattering mechanism is the oscillation of electrons in the oscillating electromagnetic field, the direction of the electric vector in a polarized beam can be located. For light going along *PA* in figure 78, the scattering electrons must be oscillating along *PB*, which must be the direction of the *E*-field of the light going 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.

Polarized gamma rays

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, a scattered beam of gamma rays can then be scattered again from a second target, when the intensity scattered in different directions varies. This variation gives information about the polarization on scattering, and so will yield information about how charges in the nuclei respond to electromagnetic forces.

Part Four

Relativity

Time: we think that relativity ought to be given just enough time for students *to see that, despite their interest in it, it makes intellectual demands* sufficient to defeat all but the most persistent. Before a double period is over, students will see that they are in deep water, and it may be just as well to stop then. No more than two such sessions should be allowed, selecting whatever material from Part Four seems most interesting.

'. . . Take, for example, the reciprocal electrodynamic action of a magnet and a conductor. The observable phenomenon here depends only on the relative motion of the conductor and magnet. . . . Examples of this sort . . . suggest that the phenomena of electrodynamics as well as of mechanics possess no properties corresponding to the idea of absolute rest.'

(A. Einstein, 1905, 'On the electrodynamics of moving bodies,' Annalen der Physik **17**, *891–921.)*

Why include any relativity in the course?

There are conflicting and not easily resolvable arguments about whether relativity deserves a place in a sixth-form course. Some would point to the way it draws together dynamics and electromagnetism, with the law of the constancy of the speed of light being used to make it possible to treat magnetism as an effect of electric field and motion. But to do that is very difficult, and the difficulty is illustrated by the view others propose, that the 'explanation' of magnetism is illusory, in that if the two fields were not related by Maxwell's equations, there would be no relativity theory to 'explain' their relationships all over again.

It seems a strange subject, unlike any other part of physics, though this feature can be used as an argument both for including and for excluding it.

It is difficult, because it involves a fundamental revision of previous conceptions deeply rooted in common experience. Again, this feature can be used on either side of the argument.

We take a simple, if naïve view. Students do seem to have an enormous initial interest in the subject, We aim to show them that such an interest cannot be pursued far without a great deal of intellectual endeavour, but also to show, so far as possible, what sort of ideas relativity involves.

There are plenty of good books, at various levels, and we aim to let a student who thinks he might read some of them know what he is in for.

More materials are provided in this Part than there will usually be time for, and teachers should make a selection.

Students' book

Questions 51 to 56 are about relativity, and uses of ideas from relativity.

Books for students

Landau and Rumer, *What is relativity?* can be strongly recommended. It contains no mathematics.

Bondi, *Relativity and common sense* is excellent, and the mathematics is simple. It propounds the view that the speed of light *must* be constant, because the time of flight of light is the natural way to measure distances.

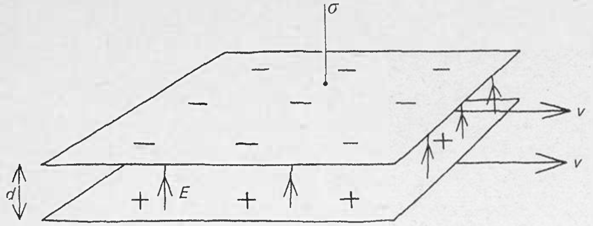


Figure 79

Trouton's and Noble's experiment

If the plates are carried along in their own plane, the moving charges might have acted like oppositely directed currents, and produced a repulsion, weakening the net attractive force between the plates and reducing the potential energy. Plates carried along at right angles to their own plane would suffer no such change in energy. Trouton and Noble tried to investigate such ideas by looking for a tendency of parallel plates to twist into the position of lowest energy when suspended by their edges on a fine wire and suddenly charged when in various orientations. No such tendency was found.

The constant speed of light

An electromagnetic wave travels at 3×10^8 m s^{-1}. Past whom? If one heads for the Sun in a spacecraft at, say, 10^8 m s^{-1}, will sunlight go past one at 4×10^8 m s^{-1} (as it would if it were like another object coming the other way), or at 3×10^8 m s^{-1}, or at some other velocity?

The modern answer is that light goes past everyone at the same velocity, 3×10^8 m s^{-1}. Why should anyone suppose anything so odd?

A number of different sorts of reason can be offered. One reason sounds too good to be true. It says that metre rules are a very bad way of measuring distances, when all one needs is a clock and a radar set, and there is force in this view, since radar methods are much used for accurate distance measurement. But the next proposal is more startling: if the speed of light is the standard of distance measurement, *of course* the speed is constant. Distances would be given as 'light microseconds' (about 300 m) or 'light nanoseconds' (about 0.3 m), so that the 'speed of light' would be simply unity – one light second in a second. This ruthless view will be too much for most students to swallow.

Another reason, which influenced Einstein, sounds subtle. The velocity of electromagnetic waves follows from the laws of electromagnetism, such as the law of induction or the production of magnetic field by currents. But these laws only mention relative motion. An induced voltage exists when a magnet moves relative to a wire, or the wire relative to the magnet, but not when both move together. Nor do we find a magnetic field near a copper rod that is being transported along with us by the Earth. To get the field, the electrons in the rod must move past the detector. It may be (and it is) that the velocity $1/\sqrt{\varepsilon_0 \mu_0}$ is a relative velocity only, but one relative to *anybody*.

A third reason for thinking that the speed of light might be constant seems like bandying terms, and depends upon having seen experiment 8.13. If $c = 1/\sqrt{\varepsilon_0 \mu_0}$ depended on the motion of the observer, so would ε_0 (μ_0 is settled by decision, when the size of the ampere is fixed). A pair of parallel plates has been used to measure ε_0, in Unit 3, as in figure 79. From the p.d. V across the plates, the field $E = V/d$ between the plates can be found. Then $E = \varepsilon_0 \sigma$, and σ, the density of charge, can be measured, giving ε_0. Experiment 8.13 suggests that sideways movement of *all* the apparatus will have no effect perceptible to an observer travelling along with the apparatus. Indeed, this is just how a student who measured ε_0 was actually moving with the Earth in its orbit. Only if this motion makes no difference to the E-field between the plates or the forces on the plates will ε_0 be a universal constant. And if ε_0 is a universal constant so is c; that is, light travels at the same speed past anybody.

Film

The PSSC film 'Frames of reference' could be good value here, though not essential. It is not concerned with special relativity, but with the motion of bodies seen from slowly moving or accelerated frames.

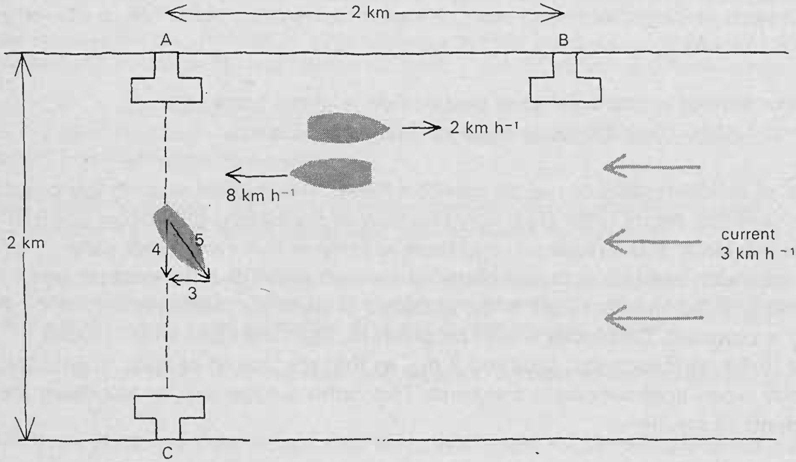


Figure 80
Journeys of a boat. The boat goes at 5 km h^{-1} in smooth water.

Demonstration

8.16 Microwave analogue of the Michelson–Morley experiment

184/1	3 cm wave transmitter
184/2	3 cm wave receiver
181	general purpose amplifier
183	loudspeaker (if not part of item 181)
1053	metal screen about 0.3 m square *2*
1053	hardboard screen about 0.3 m square

The apparatus is set up as in figure 81. The two reflectors should be about the same distance from the hardboard sheet, which partly reflects the waves. Moving either reflector in a direction at right angles to its plane will cause changes in the intensity at the receiver. The reflectors can be set for a minimum.

The 'experiment' may seem more real if the apparatus is on a trolley, able to move in the direction of the transmitter's axis.

The 'experiment' gives a null result, which is correct, but is too insensitive to give a result even if there were one. At a frequency of 10^{10} Hz, a time difference Δt of the order of, say 0.5×10^{-10} s (half a cycle) would be detected by the receiver giving a *minimum output*, compared with the time t, about 0.3×10^{-8} s for the radiation to travel out and back along arms half a metre long. $\frac{\Delta t}{t}$ is of an order 10^{-2}, and would be the same order of magnitude as $\frac{v^2}{c^2}$ if there were a detectable effect. A null result thus shows that motions of the Earth through space of more than $10^{-1}\,c$ are not detected, which is true but scarcely impressive.

Appendix C contains an account of an accurate version of the same experiment, performed by Essen.

Finally, whether or not one is impressed by these reasons, the constancy of the speed of light is an experimental fact.

If it were correct that one had to add or subtract one's own velocity from that of light, the situation would be like that of a boat travelling in moving water. Figure 80 shows some boat journeys in a river, which are analogous to the experiment done by Michelson and Morley, showing that light journeys are *not* like river travel.

The boat goes at 5 km h^{-1} in still water, so going upstream from pier A to pier B, it will travel at 2 km h^{-1} if the river flows at 3 km h^{-1}. The 2 km journey takes one hour. Coming back, the time is only $\frac{1}{4}$ h, because the boat goes at 8 km h^{-1}, with the stream. The total time is one and a quarter hours, from A to B and back, on a four kilometre journey.

If the boat is piloted across river from A to C, the pilot must head the boat upstream a little, so as to end up going straight across. The 5 km h^{-1} velocity combines with the 3 km h^{-1} velocity, to give a velocity of only 4 km h^{-1} across the stream. (The numbers have been chosen so that the vectors form a right-angled triangle with sides 5, 3, and 4.) Thus the boat takes half an hour to go the 2 km from A to C at 4 km h^{-1}, and the same time to return, making one hour for the return trip.

The upstream-downstream journey takes a quarter of an hour longer. What Michelson and Morley did was to make beams of light go out and back along paths at right angles, and look for a journey-time difference between the two beams.

Demonstration
8.16 Microwave analogue of the Michelson–Morley experiment

A microwave analogue is easy to arrange to demonstrate the principle, though not the result. Light was split into two beams at right angles, by a lightly silvered mirror, and returned back along each path and mixed to produce interference fringes. The idea was to detect changes in the fringe pattern because of the light travelling at a different velocity along one path to its velocity along the other, when one path happened to point in the direction of the Earth's travel and the other across it. If the light beams travel at different speeds they may arrive out of step instead of in step.

The result was astonishing at the time: there were no changes at all. The apparatus was slowly turned round and round so that the arms pointed in all directions, but there were no significant fringe shifts, even though shifts caused by the motion of the Earth round the Sun (30 km s^{-1}) could have been detected. It has been repeated many times, in many ways. Some recent experiments could have detected an absolute motion of only a few centimetres per second. So far as light is concerned *the Earth is not moving*! Or, to put it better, light seems to travel at the same speed however fast the platform from which you observe it is moving.

The origin of the experiment

Michelson did this experiment because, as a young man working with a group of astronomers, he saw a letter from Maxwell asking if the records could detect a shift in the velocity of light from Jupiter as it and the Earth moved relative to one another. The answer was no (the time difference expected was a few seconds over a decade), but Michelson determined to devise a method capable of detecting such a change.

R. S. Shankland, 'The Michelson–Morley experiment', reports the history and discusses the experiments simply.

Other relativistic experiments

See Appendix C for details of a number of experiments concerned with relativity, done in recent years.

The PSSC ether-drift experiment: PSSC devised an ether-drift experiment that can be done in school. See PSSC *College physics*, Chapter 30. It does involve some further theory, but a few students may find it of interest.

Modern ether-drift experiments: The extremely sharp resonance obtainable for emission and absorption of some gamma rays by nuclei locked in crystal lattices (Mossbauer effect) has been used to raise the sensivity of the experiment enormously. Recent results could have detected effects corresponding to a velocity of only centimetres per second (but they did not). See Appendix C.

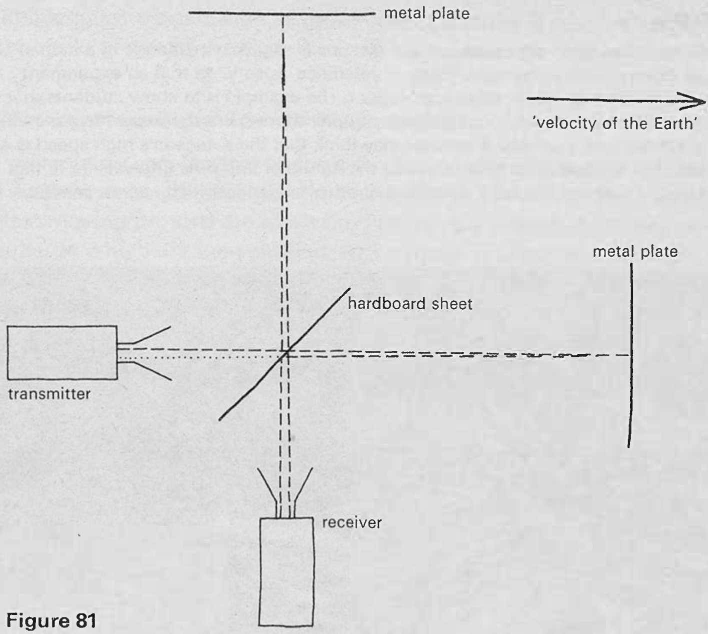


Figure 81
Microwave analogue of the Michelson–Morley experiment.

The principle of relativity

When Copernicus suggested that the Earth went round the Sun, people ridiculed the idea. Surely we would all be left behind if we stopped clinging to the ground for a moment? But gradually, people came to see that if we, the Earth, and everything on it, travelled along together, it made no difference to things happening on the Earth. This is the principle of relativity in mechanics: mechanics is the same even if you move along at a steady speed.

Einstein took the step of extending this principle to all of physics. He suggested that the only steady motion that can be detected is relative motion; that we could never know which objects in the Universe were 'not moving'. For this to be true, all experiments like measuring the velocity c of light, or measuring ε_0 and so calculating c (μ_0 is fixed by choice), would have to come out the same however the Earth moved. In this way, the velocity of light becomes another universal constant, like G.

Experiments like Michelson's and Morley's, designed to measure absolute motion directly, *must* give a null result, according to Einstein.

The rest of this Part is concerned with the strange consequences of the constancy of the speed of light.

π-mesons

The object of briefly considering lifetime effects for π-mesons is to give an example of a case in which the change of a time interval with a changed 'frame of reference' is so large that an experiment occupies ten times the space one would otherwise expect. The example is to show students that relativity is not concerned only with tiny corrections to simpler theory. But the experiment is not good simple evidence for altered time intervals. A student may think that the π-meson's high speed is as likely to have increased its lifetime as to have changed the nature of unit time intervals for it: this explanation would be the simplest but for a great deal of other evidence which cannot profitably be used here.

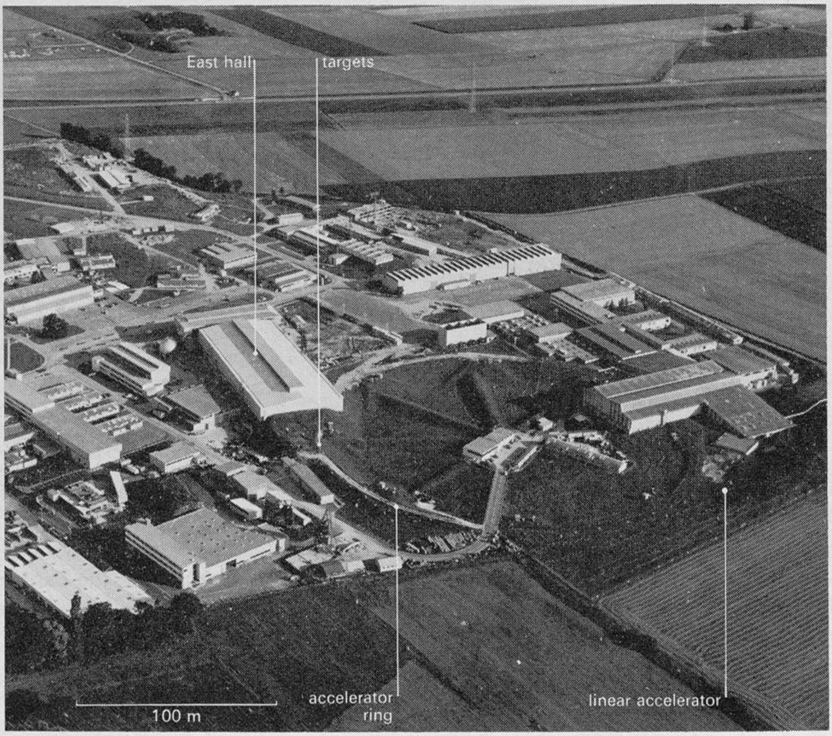


Figure 82
Aerial photograph of CERN accelerator.
Photograph, CERN.

Students' book

Figures 82 and 83 also appear in the *Students' book*. They are part of question 54, which discusses the meson lifetime problem given in the text.

Check on time dilation

Greenberg and others (*Phys. Rev. Letters*, **23**, 21, page 1267, November 1969) report accurate measurements of the number of π-mesons surviving along a beam. The observed lifetime was 2.4 times the lifetime at rest, and this factor agreed with the calculated factor $\sqrt{1-v^2/c^2}$ to within 0.4 per cent.

Before embarking on a discussion of these consequences, it may be helpful to have a practical example of one of them. In the case considered, the effect is not small.

The strange behaviour of fast particles

In recent years, large accelerators such as the proton synchrotron at CERN (the European nuclear research laboratory in Switzerland) have been built (figures 82 and 83). The fast moving protons are often fired into metal targets, from which many new particles emerge which are then studied. An example is π-mesons (thought to be a kind of nuclear 'glue') which have a measured lifetime of 2.5×10^{-8} second. They decay randomly, like radioactive nuclei, so that out of N particles, a constant fraction decay on average in each unit time interval. After 2.5×10^{-8} second, there will be N/e left ($e = 2.718\ldots$); N/e^2 after twice this time, and so on.

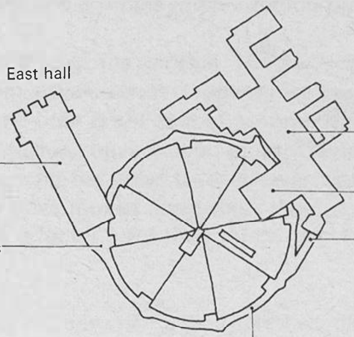


Figure 83
Plan of CERN accelerator.

The π-mesons emerge travelling at nearly the velocity of light, their velocity being measured by timing them over a distance. At 3×10^8 m s^{-1}, in 2.5×10^{-8} s, they will travel about 7.5 m. Clearly any experiment on the π-mesons must be set up close to the target, or there will be none left to observe. Yet it is not so. As may be seen from an aerial photograph, the experimental area is much further than this from the target, and it is found that there are still plenty of the original mesons left. At 150 m there should be only a fraction of the order of e^{-20} left (10^{-9}). At CERN, the distance from target to bubble chamber is about this distance.

The lifetime measurement can be confirmed, and is not in error. But it is measured at low speeds: *at high speeds the particles seem to live longer*. This was not an unexpected result discovered by accelerator designers, but had been predicted by Einstein in 1905 in his theory of relativity. It is a convenient result, for the long meson flight makes room for magnets to switch the beam from experiment to experiment, as points are used to direct trains in a railway yard. Many room-sized experiments can be set up to share the beam, which would be impossible in the space only a few metres in front of the target.

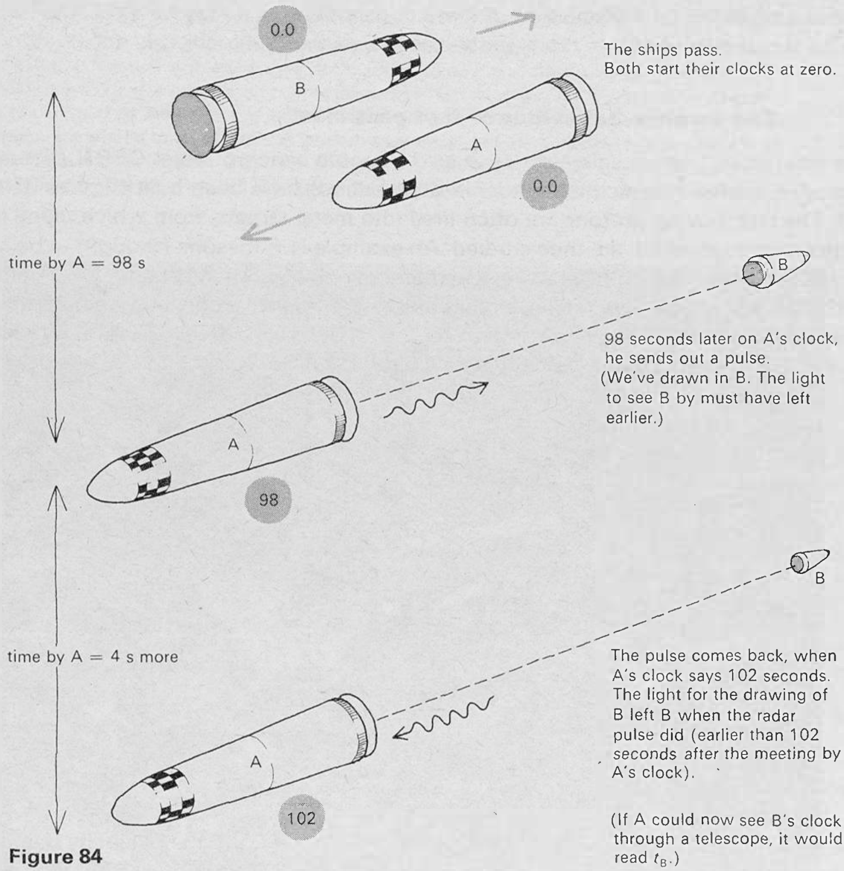


Figure 84

Students' book

Question 52 is about the ratio of t_A to t_B, following the argument given. Question 51 introduces the idea that things look different to different people.

Question 51 does not involve relativity; only that light and the pulse in the wire travel at the same speed. Einstein and the other man stood equidistant from the middle cow and sketched it in the same *position, so the sketches are simultaneous.*

Film

The PSSC film 'Time dilation', which shows a meson lifetime experiment, would be useful here or perhaps earlier.

To say that the particles live longer is to say that their 'clocks' are going slower than ours. It is this which relativity predicts. If 'time' for one person is not the same as 'time' for another, moving relative to the first, clearly our most fundamental ideas are going to need thinking through afresh.

Radar ranging

We now show how the constant velocity of light has strange consequences, in particular, the stretching of time for a moving object as seen from another object not moving with the first object, as in the π-meson example.

Because the velocity of light is constant, radar makes a good tool for measuring the distances and velocities of moving objects. It is the only practicable such tool for space journeys. We now analyse a simple radar-ranging problem.

Suppose spaceships A and B pass one another. Ninety-eight seconds from this instant, the operator on one of them, say A, sends a radar pulse towards B (figure 84). The pulse must arrive at B after a time B will record as more than 98 s, because light has a finite velocity, and the spaceships have moved apart. The larger time B records, say t_B, will depend on how long the pulse takes to overtake him: that is, on the velocity v with which they have moved apart since they passed and on the velocity c of the pulse. We will write a factor k for how much larger this time that B records is.

$$t_B = k\,98.$$

k depends only on how big v is compared with c, not on the times.

Suppose that the radar echo comes back after A records a further 4 seconds, so that he has noted a total wait of 102 s. How long did he think the pulse took to reach B? (2 s.) How far away does he calculate B was when the pulse arrived and was reflected? (6×10^8 m.) At what time will A suppose the pulse reached B? (100 s after they passed.) How fast are they moving apart? (6×10^6 m s^{-1}.) It is clear how either ship can measure the distance and speed of the other. Had B made the measurement, he must have obtained the same result, for c must be the same for both, and they must both obtain the same value for their relative velocity. The factor k must be the same for both, so that if B sends a pulse to A, after a time he records as t_B, it must arrive after a time kt_B recorded by A.

But B did send (or rather, reflect) the radar echo pulse to A, when B's clock stood at time t_B. It arrived after 102 s, measured by A.

Thus $\quad kt_B = 102$

but previously

$$t_B = k\,98.$$

What is t_B? A thinks his signal arrived after 100 s, but that will not do, for if $t_B = 100$ s we have

The quantity k

k is the Doppler factor; the ratio between time intervals at which a pair of signals (or two peaks of a wave) are received and sent. It is *not* the time dilation factor, but contains it.

$$k = \sqrt{\frac{1+v/c}{1-v/c}} = \frac{1+v/c}{\sqrt{1-v^2/c^2}} \quad \text{for bodies moving apart.}$$

(For bodies moving together, the sign of v is reversed.)

Note that k is not the same as for sound, or other waves in a material medium, nor is the usual optical Doppler shift expression $\frac{df}{f} = \frac{v}{c}$ exact.

An observer who sends a pair of pulses to a moving object will receive them back at an interval k^2 (two shifts by a factor k) times larger than when he sent them. k^2 can be measured. Then $v/c = (k^2-1)/(k^2+1)$.

Relativity arguments using k were introduced by Bondi. They have two advantages:

1. Co-ordinate transformations are avoided.
2. Arguments proceed in terms of practicable radar measurements.

Other versions appear in many books, notably:

Bondi, *Relativity and common sense.*

Further uses of k

Teachers may like to know how some more results can be produced.

1 Doppler shift for approach

$$k = \sqrt{\frac{1+v/c}{1-v/c}} \text{ for recession.}$$

If A sends a pair of signals at an interval t to B, who moves away from him at velocity v, B will receive them at an interval kt. Suppose B sends them on to C who is at rest with respect to A, so that B approaches C at velocity v. The signals leave B at interval kt, and must reach C at interval t, for signals from A to C suffer no Doppler shift. Thus the Doppler shift for signals received at C from B is $1/k$, equal to

$$\sqrt{\frac{1-v/c}{1+v/c}}.$$

2 Addition of velocities

B recedes from A at velocity u, C recedes from B at velocity v. How fast does C recede from A? If k_1 is the Doppler shift B sees on A's signals, and k_2 is the Doppler shift C sees on B's signals, then the Doppler shift C sees on A's signals is k_1k_2. (Imagine B sending A's signals straight on to C.)

If V is the velocity of C relative to A,

$$\sqrt{\frac{1+V/c}{1-V/c}} = \sqrt{\frac{1+u/c}{1-u/c}}\sqrt{\frac{1+v/c}{1-v/c}}$$

From this, after some algebra:

$$V = \frac{u+v}{1+uv/c^2}$$

Note that V never exceeds c, even if u and v both approach c. $V = c$ if $u = v = c$. Nor can v exceed c in the Doppler shift expressions which become indefinitely large or become zero as v tends to c.

$$100 = k\,98$$
and
$$k\,100 = 102$$

which do not both give the same value of k. If their time estimates agree, their velocity estimates do not.

Solving for t_B

$$t_B{}^2 = 102 \times 98 = 9996 \text{ (less than } 100^2)$$
$$t_B = 99.98 \text{ s (less than 100).}$$

If they agree on v and c, they do not agree about the time. A thinks B's clock is running slow, for it records 99.98 s while his own records 100 s. This is the time dilation effect noted for the π-mesons. Observe that it is not because B is 'really moving' that his clock runs slow compared with A's. The whole argument cuts both ways, and if B makes the same measurements he finds that A's clock is running slow! They disagree about times, but agree in that each calculates that the other's clock is running slow. A calculated the time of arrival as 100 s. He cannot now change his mind, and say it was 99.98 s, for then the light would have gone out quicker than it came back.

Were the conclusion less strange, the argument might seem more compelling. But no one has yet found any convincing flaw in it.

Algebraic form of the argument

Let the spaceships pass and separate at velocity v. A sends a radar pulse so that it arrives at B at a time A calculates to be t_A (100 s above), by splitting the difference between his outgoing and incoming pulse times.

He calculates that B must then be a distance vt_A away, and so calculates that the pulse takes time vt_A/c to cover this distance. It must have been sent and received at times A would take to be:

pulse sent after time $(t_A - vt_A/c) = t_A(1 - v/c)$ (98 s above)
pulse received back after time
$$(t_A + vt_A/c) = t_A(1 + v/c) \text{ (102 s above).}$$

Note that t_A is a *computed* intermediate time, worked out from the measured times which appear in the above expressions.

Suppose B receives and reflects the pulse at a time he records as t_B. (Note that we do not call this t_A, even though we might expect it to be equal to t_A. It turns out not to be.) Writing k for the factor by which the time one of them waits to receive the pulse is larger than the time the other waits before sending it, both taken from the moment of passing,

for A's pulse to B $\quad t_B = kt_A(1 - v/c)$
for B's echo to A $\quad kt_B = t_A(1 + v/c)$.

3 Length contraction

Consider the π-mesons (page 117) travelling many metres across the laboratory. They are their own clocks, and to a π-meson the proper time 2.5×10^{-8} s elapses before it decays (on average). It must agree with laboratory physicists about its velocity relative to them, so the 100 m or so of laboratory must look like about 7 m of distance to the mesons. Lengths in one reference frame along the line of flight of another frame must seem shorter in the moving frame. The factor $\sqrt{1-v^2/c^2}$ is the same as for time dilation.

4 The inertia of energy ('increase of mass') and $E = mc^2$

Arguments about mass, momentum, and energy are harder, for one has to make new assumptions about the conservation laws. Bondi, *Relativity and common sense*, gives a simple and crude argument. Feynman, *The Feynman lectures on physics,* Volume I, Chapter 16.6, gives a better but harder argument based on collisions, also to be found in many other books, for example, Sherwin, *Basic concepts of physics*.

The PSSC film, 'The ultimate speed', shows how the velocity of electrons accelerated to larger and larger energies does not rise above the velocity c. The quantitative results can be used to test the mass increase relationship.

Students' book

Questions 55 and 56 are about the Doppler effect. Question 55 is a simple one concerning sound. Question 56 gives red-shift data for some distant galaxies, and could lead to discussion of the expansion of the Universe.

Sources of Doppler shift photographs

Doppler shifted spectra appear in:

Sandage, 'The red-shift' *Scientific American.*

Burbidge and Hoyle, 'The problem of the quasi-stellar objects' *Scientific American.*

Useful books for teachers

All of us easily get into muddles in thinking about relativity. A good book for sorting out such confusions is Sherwin, *Basic concepts of physics*.

The Berkeley Physics Course, Volumes I (*Mechanics*) and II (*Electricity and magnetism*), are also clear, though rather abstract.

French, *Special relativity*, is very good, especially the introductory chapters which treat relativity as based on experiment rather than abstract thought. The final chapters give a good account of the relation between magnetism and relativity.

Einstein *et al., The principle of relativity,* contains a translation of Einstein's 1905 paper, which should be read once in one's life.

Doppler shift and time dilation

Note that an observer looking through a telescope at a clock (such as atoms emitting light) on a moving star does *not* see them running slow by $\sqrt{1-v^2/c^2}$. He sees a Doppler shift. If the star is moving towards him, the clock seems to run *fast*. But in all cases the size of the Doppler effect is such that it must have been produced by a slow running clock. For an oncoming star, the Doppler rise is not as big as it would be if there were no time dilation.

As before, k must be the same for both, if the velocity of light is the same for both, since k measures how long the pulse takes to overtake the other ship. If k is the same, t_A cannot be equal to t_B.

$$k = \frac{t_B/t_A}{(1-v/c)} = \frac{(1+v/c)}{t_B/t_A}.$$

$$\therefore \left(\frac{t_B}{t_A}\right)^2 = (1+v/c)(1-v/c)$$

$$\therefore t_B = t_A\sqrt{1-v^2/c^2}.$$

Note that t_B is a time *measured* by B, while t_A is a time *calculated* by A from measurements made by A.

The factor $\sqrt{1-v^2/c^2}$ is the time dilation factor. As v^2/c^2 is always positive, t_B is less than t_A. Remember that t_B is the time measured by B; t_A is the 'same' time calculated by A from his measurements. If the roles are reversed, t_A is less than t_B. In the numerical example above, $v/c = 2/100$, and

$$\sqrt{1-v^2/c^2} = \sqrt{1-4\times10^{-4}} = (1-0.000\,2) \text{ approximately.}$$

The Doppler shift

From the previous argument, the factor k is:

$$k = \frac{(1+v/c)}{t_B/t_A} = \frac{(1+v/c)}{\sqrt{1-v^2/c^2}}.$$

The factor $(1+v/c)$ is what one would expect. If a signal chases after a spaceship travelling at, say, half the signal speed, one would expect it to need half as long again to catch up. The time dilation factor $\sqrt{1-v^2/c^2}$ is the correction to this commonsense story, needed because the clock rates differ.

The factor k has a simple interpretation. Instead of sending a signal a time t_A after the spaceships pass, so that it arrives after a time kt_A, a whole series of signals can be sent. If t is the time between any pair of signals as they are sent, kt is the time between them when they arrive. Such a row of signals could be the oscillations of a radio, or a light wave of frequency f, which would arrive at a *lower* frequency f/k – intervals $k(1/f)$ – if the spaceships were moving apart. If the spaceships are coming together, the frequency rises, and is kf. This is called the Doppler effect, well known before relativity.

The Doppler effect is widely used in astronomy to find the velocities of distant stars and galaxies. It has led to the picture of the expanding Universe, in which distant galaxies seem to recede at a speed proportional to their distance. Photographs of Doppler-shifted spectra can be shown.

Demonstration
8.17 Doppler effect using sound waves

1055 whistle

1055 rubber tubing, about 1 m long, to fit whistle

Fix the whistle into one end of the tube, and blow down the other end. Then whirl the whistle round in a circle, aligned so that the whistle travels towards and away from the class as it goes round.

Demonstration
8.18 Doppler shift using microwaves

184/1 3 cm wave transmitter

184/2 3 cm wave receiver

181 general purpose amplifier

183 loudspeaker (if not part of item 181)

1053 metal screen about 0.3 m square

1053 hardboard screen about 0.3 m square

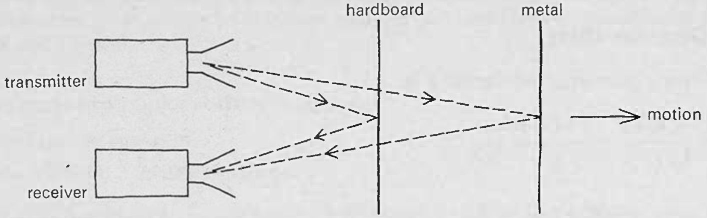


Figure 85
Doppler shift using microwaves.

The apparatus is set up as in figure 85. The transmitter should emit unmodulated microwaves. The hardboard sheet should be near the transmitter and receiver. The metal reflector should be moved as fast as possible and space should be left so that whoever moves it can get a good swing at arm's length.

Students may comment that what they hear is only a quick succession of interference fringes. This interpretation is just another way of describing the effect of beats due to the Doppler effect.

Students' book

Question 53 involves a rough estimate of the ratio of a magnetic force to an electrical force, and a comparison of the ratio with v^2/c^2.

Demonstration

8.17 Doppler effect using sound waves

A whistle swung round in a circle gives a rising and falling note, highest when the whistle travels towards the class. Any student asked to, 'make a noise like a racing car' will probably show a knowledge of the Doppler effect in practice, producing a falling tone, 'e-e-e-e-e-o-o-o-o-o-w'.

Demonstration

8.18 Doppler shift using microwaves

A receiver beside the transmitter sees microwaves partly reflected from a hardboard sheet, and also others reflected from a metal sheet behind the hardboard. If the metal sheet is moved, the waves reflected from it are shifted in frequency, and beat with those from the unmoving hardboard. The beat frequency is in the audible range for fairly rapid movements of the reflector.

Clocks, forces, and magnetism

The above analysis did not have to say what kind of clock was to be used. *All* kinds of clock should seem to run slow when moving relative to another clock. This has been confirmed in a number of cases, like decaying particles, and may well be confirmed soon for accurate crystal or atomic clocks in Earth satellites.

Consider a spring clock: a mass M on a spring of force constant K (force per unit extension), with a periodic time T,

$$T = 2\pi\sqrt{\frac{M}{K}}.$$

If T is to change by a factor $\sqrt{1-v^2/c^2}$, either M or K or both will have to change. It turns out that both do. Moving masses seem larger than their counterparts at rest in the laboratory, and forces are also altered, both by amounts of the order v^2/c^2. In particular, the force between a pair of masses joined by a spring or a pair of electrons (figure 86) will seem to be reduced by about v^2/c^2 if they move together sideways relative to the instruments or observer investigating the force. This fact links electricity and magnetism.

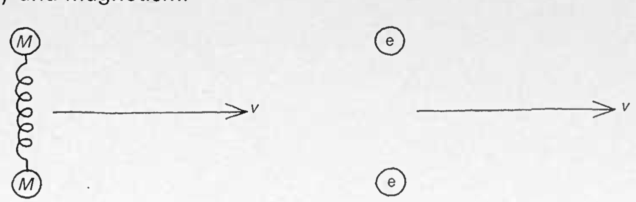


Figure 86

Forces between moving charges

One of the following books may help a teacher who wants to see how relativistic corrections to electrical forces can be interpreted as magnetic forces. Also, see Appendix B, page 148.

French, *Special relativity.*

Gibson, *Basic electricity.*

Berkeley Physics Course, Volume II, *Electricity and magnetism.*

Possible relativistic changes to the other forces

Students may reasonably think that, of the forces in figure 87, the force R_{++} between fixed positive ions won't change, but that the forces A_{+-} and A_{-+} between positive and negative charges, one of which is moving, might just as well be changed as the force R_{--} between moving electrons.

A fuller argument for the force between a pair of moving charges says that the force on a moving charge due to a fixed charge is unchanged. So the force A_{-+} of the ions in the wire on the moving electrons in the other wire is unaltered.

But the transverse force on a charge at rest due to one in motion at right angles to the line joining them is increased, by a factor $1/\sqrt{1-v^2/c^2}$.

Now the force A_{+-} is the vector sum of many such forces, but including some between charges joined by lines inclined to the length of the wire; indeed the forces between some charges are along lines nearly parallel to the velocity. Such forces between pairs of charges almost parallel to the velocity turn out to be *diminished,* not increased. It happens, by coincidence and for this geometry alone, that when all these forces, some bigger, some smaller, are added up, the net effect is that the net force A_{+-} is unchanged. The proof is given in French, *Special relativity.* So the argument based on figure 87 is correct, but only by a lucky accident.

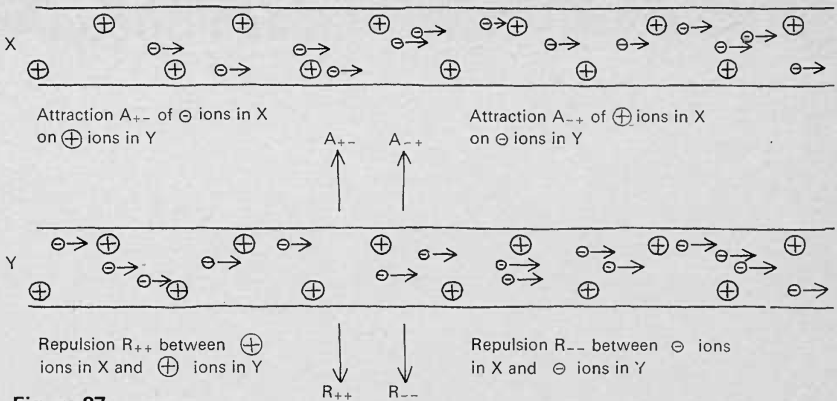


Figure 87
Forces between two current-carrying wires. (Arrows represent average drift velocity, not large but random thermal motions.)

Consider two wires, with electrons free to move among fixed positive ions. The like charges repel, the unlike ones attract, giving four counterbalancing forces as in figure 87.

If the electrons are at rest (no current), the forces balance,

$$A_{+-} + A_{-+} = R_{++} + R_{--}.$$

If the electrons move at velocity v, the force R_{--} is reduced by an amount of about $(v^2/c^2)R_{--}$. There is then a net attraction, of magnitude v^2/c^2 times the electrostatic repulsion. This is the 'magnetic' force between currents used to define the ampere. The magnetic force constant μ_0 is equal to $1/\varepsilon_0 c^2$.

It is remarkable that the velocities involved in electric currents are very small, less than one millimetre per second, for which v^2/c^2 is of the order 10^{-20}. Yet magnetic forces are the most easily observed of all relativistic effects. Relativity is not confined to unrealizably high velocities. Indeed, every electric train is driven by these tiny relativistic corrections to electrostatic forces.

Appendices

Appendix A
Phasor treatment of diffraction

It would be unfortunate for a student to encounter the adding of phasors for the first time while studying diffraction theory. The ideas of displacement, amplitude, and phase, which can be visibly illustrated for sinusoidal alternating current, are hypothetical at first in physical optics. The treatment below assumes that the student is familiar with phasor addition but not with optics, and that he knows that a.c. power delivered to a resistor is proportional to the square of the voltage amplitude.

Single-slit diffraction

The amplitude of the light, of a particular wavelength λ, arriving at a point on the infinitely distant screen from a strip, say q, can be represented by the length of a phasor. The amplitude from another equal strip, say r, is the same, but the light is not necessarily in phase with the light from q. If α is the angle representing the phase difference, the amplitude of the combination of light from q and r is given by the resultant of the two phasors when they are at angle α to one another. From four strips, p, q, r, and s, the amplitude of the combination is the vector sum of phasors p, q, r, and s, as in figure 88.

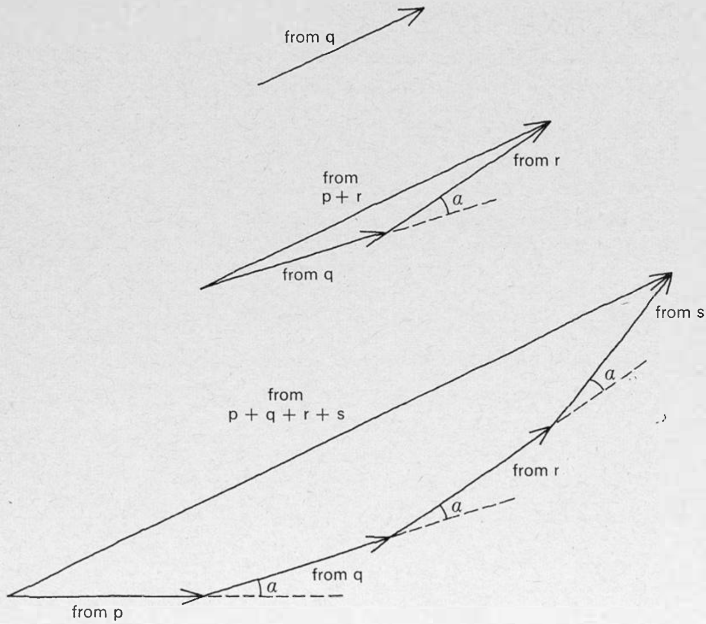


Figure 88

If light from one strip has a path length one wavelength λ different from another, what is the phase angle α between them? (2π radians.)

If light from one strip arrives a distance L behind light from another strip, what is the phase angle α between them? ($2\pi L/\lambda$ radians.)

If light from one strip arrives a distance $x \sin \theta$ behind the light from another strip, what is the phase angle α between them? ($(2\pi x \sin \theta)/\lambda$ radians.)

Suppose the width b of the slit is divided into eight strips, and we choose angle θ so that $b \sin \theta = \lambda/2$, then the width of each strip is $b/8 = \lambda/16 \sin \theta$.

So the phase difference between light from adjacent strips =

$$(\lambda/16 \sin \theta)\ (2\pi \sin \theta)/\lambda = \pi/8.$$

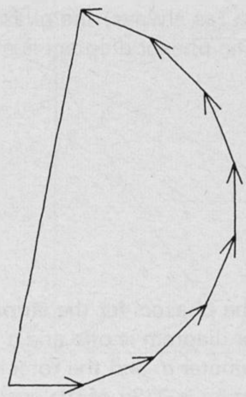
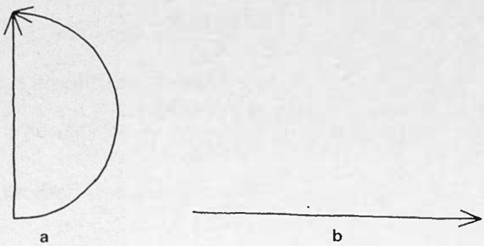


Figure 89

If we had divided the slit into 80 strips, each strip would have had one-tenth of the width, and its phasor would have been one-tenth of the length. But the angle between adjacent phasors would also have been one-tenth of what it was before. So the diagram showing their sum would have been more nearly a semi-circle, with the same 'radius' as before. The total length of phasors would be the same. If we had divided the slit into infinitesimal strips, the diagram would have been a semi-circle. (Figure 90 *a*.)



Figures 90a and 90b

What is the amplitude where $\theta = 0$? In this case the phase angles are zero, and the amplitude from the whole slit is the length of the phasors placed end to end, that is, the distance round the outside of the semi-circle in the previous case (figure 90 *b*). Now the ratio of the diameter of a circle to half its circumference is $2/\pi$. So the amplitude in the direction given by $\sin\theta = \lambda/2b$ is $2/\pi$ of the amplitude in the direction $\theta = 0$. When a.c. flows through a resistor, the rate at which energy is transformed in the resistor is proportional to the square of the voltage (or current) amplitude. Similarly with light, we may expect the intensity (which is the rate of flow of energy) to be proportional to the square of the amplitude. So the intensity is $4/\pi^2$ of the intensity at $\theta = 0$.

What is the amplitude where $\sin\theta = \lambda/b$? In this case the phasors for the strips at the extreme edges of the slit are in phase, and again (as always) the phase angle α changes linearly with distance across the slit. The phasor diagram is a complete circle, and the closing chord is zero (figure 91).

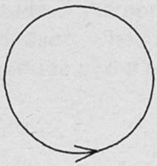


Figure 91

What is the amplitude where $\sin\theta = 3\lambda/2b$? The phasors for the strips at the extreme edges of the slit are in antiphase, but the phasor diagram is one and a half complete circles (figure 92). The closing chord is one diameter *d*, and the total length of the phasors is $3\pi d/2$. So the amplitude in this direction is $2/3\pi$ of the amplitude for $\theta = 0$, and the intensity is $4/9\pi^2$ of the intensity for $\theta = 0$, because the intensity is the square of the amplitude.

Figure 92

What is the intensity for $\sin\theta = 5\lambda/2b$? ($4/25\pi^2$ of the intensity for $\theta = 0$.) These values enable us to plot a graph of intensity against $\sin\theta$ with more certainty (figure 94).

The phasor treatment shows the more mathematical students that they can calculate the amplitude and intensity for any angle θ if they want to. They could go a little further and derive a general expression for the amplitude. One way is set out below for teachers who may be asked to discuss it.

The phasors form the arc BC, figure 93. Its length A_0 represents the amplitude for $\sin\theta = 0$. The chord BC represents the sum of the phasors, that is, the amplitude for angle θ.

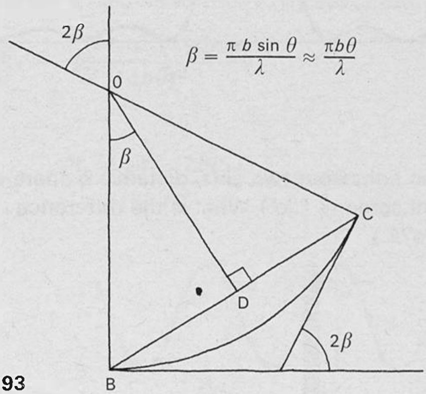


Figure 93

In the right-angled triangle OBD, $\sin(\pi b \sin\theta)/\lambda = \sin\beta = \text{BD}/\text{OB}$. OB is the radius of a circle in which arc A_0 subtends angle 2β.

Thus $\quad A_0/\text{OB} = 2\beta$

and $$\frac{\sin\beta}{\text{BD}} = \frac{2\beta}{A_0}$$

or $$2\text{BD} = A_0\frac{\sin\beta}{\beta}$$

$$= \frac{A_0\lambda\sin(\pi b\theta/\lambda)}{\pi b\theta}.$$

2BD is the amplitude, therefore

$$\text{intensity} = \frac{I_0\lambda^2\sin^2(\pi b\theta/\lambda)}{(\pi b\theta)^2} = I_0\frac{\sin^2\beta}{\beta^2},$$

where I_0 is the intensity at $\theta = 0$.

Figure 94 shows the variation of intensity with angle, expressed by this last equation.

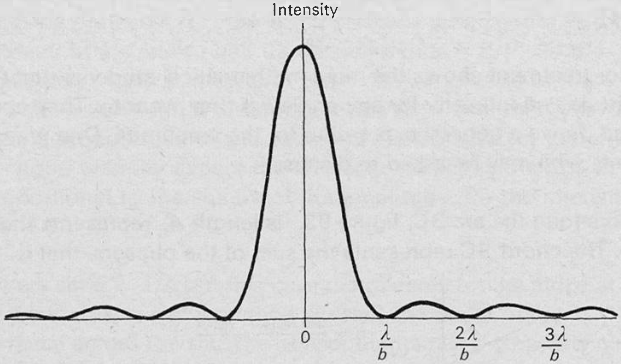


Figure 94

Two-slit interference

What is the path difference between light from two slits, distance d apart, after it has travelled at an angle θ onto a distant screen? (θd.) What is the difference in phase between light from two slits? ($2\pi\theta d/\lambda$.)

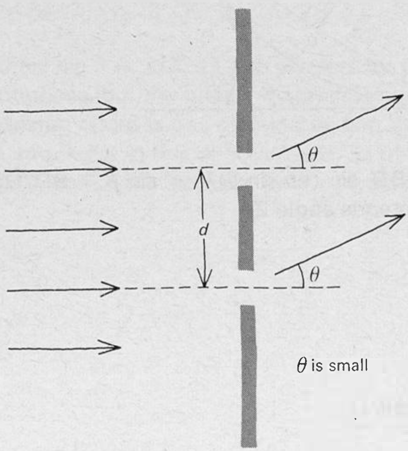


Figure 95

So we can draw two phasors of equal length A_0 to represent the light from the two slits, the angle between the phasors being $2\pi\theta d/\lambda$. What is their sum?

If A is their sum, $\cos \pi\theta d/\lambda$ is $A/2A_0$ in the triangle obtained by dropping a perpendicular (figure 96).

$$A = 2A_0 \cos \pi\theta d/\lambda,$$

and $$I = 4I_0 \cos^2 \pi\theta d/\lambda.$$

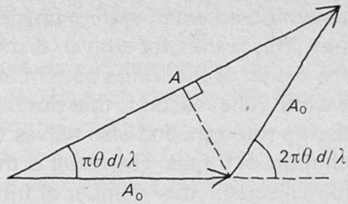


Figure 96

The graph of this expression against θ, shown in figure 97, has the shape, though not the position, of a cosine curve, and does not show the property (possessed by real fringes) of decreasing with large angles. This is because I_0 has been treated as constant when really it decreases with θ. By substituting from the previous result

$$\text{intensity} = \frac{I_0\lambda^2 \sin^2(\pi b\theta/\lambda)}{(\pi b\theta)^2},$$

for I_0, we get

$$\text{intensity} = \frac{4I_0\lambda^2 \sin^2(\pi b\theta/\lambda)}{(\pi b\theta)^2} \cos^2(\pi\theta d/\lambda)$$

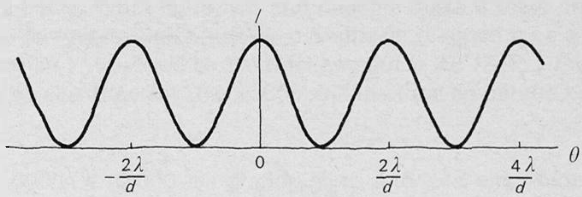


Figure 97

and a graph of this function (figure 98) shows all the properties which we observe in the actual fringes.

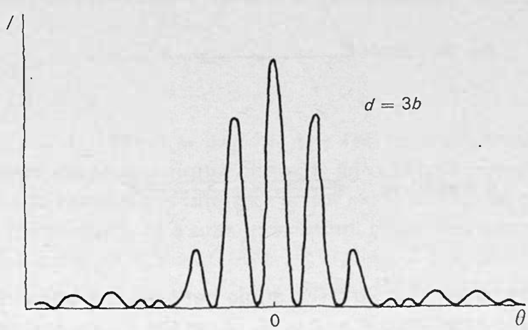


Figure 98

A practical point, which should be remembered when setting up slits to show fringes, almost comes out of these expressions. This is that, for a given distance between the slits, the maximum intensities near the centre of the fringe pattern are roughly proportional to the square of the slit width. The reason is that doubling the width of each slit doubles the light energy passing through, and also halves the area (between the first diffraction minima) on which most of it falls. The result is that if slits are very little narrower than is necessary to form a satisfactory number of fringes, the fringes are all too dim to see properly. A student who has made his slits too narrow should appreciate the reason for his failure.

Three slits

At angle θ, light from the second slit travels $d \sin \theta$ further to a distant screen than light from the first. This introduces a phase lag of $(2\pi d \sin \theta)/\lambda$ radians. So this is the angle between the phasors representing contributions from the first two slits, and also the angle between any two adjacent phasors in a grating of interval d. The total length (say A_0 or $3a$) of the three phasors represents the amplitude for $\theta = 0$ (figure 99 *a*). What is the amplitude for $\sin^{-1} \lambda/3d$? The angle between the phasors is then $2\pi d\lambda/3d\lambda$ or $2\pi/3$, and the total amplitude is zero (figure 99 *b*).

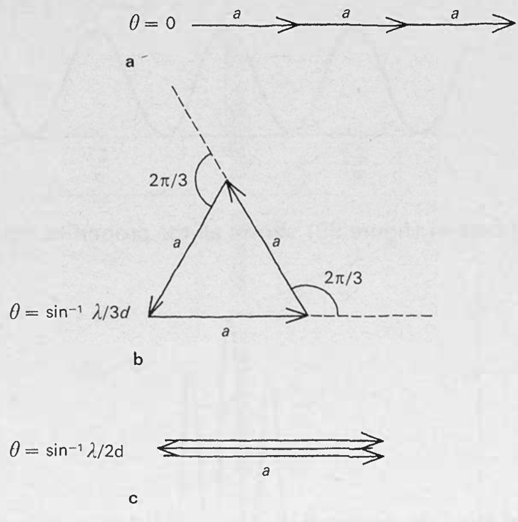


Figure 99

What is the amplitude for $\sin^{-1} \lambda/2d$? The angle between the phasors is then $2\pi d\lambda/2d\lambda = \pi$, and the amplitude $A = a$ or $A_0/3$, so the intensity is one-ninth of the intensity for $\theta = 0$ (figure 99 *c*).

This method is not limited to the particular angles with which it has been illustrated. A class which masters it should be able to apply it to four slits or ten slits without trouble. The additional sharpness of the main maxima and the reduction in size of the smaller maxima become clearer with bigger numbers of slits. A demonstration of the effects of 1, 2, 3, 4, 5, and 6 slits with a laser is particularly effective.

Diffraction grating

We consider the amplitude contributions at a distant point from the individual slits in the grating. If there are N slits, each contributing amplitude a, the total amplitude A_0 when there is no phase difference between individual contributions is Na. This occurs for $\theta = 0$, and for $\theta = \sin^{-1} n\lambda/d$. What is n? (The order of the spectrum.) What is the difference in length of path for light from adjacent slits? (Zero and $n\lambda$.) What is the angle between phasors corresponding to adjacent slits? (Zero and $2\pi n$ radians.)

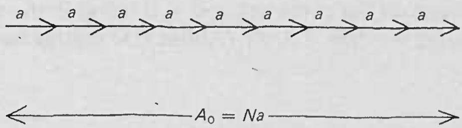


Figure 100 $N = 9$ in this case

Then we consider the phasors when the angle θ is very small, for example, the angle at which adjacent slits emit light with a path difference of $\lambda/1000$, $\theta = \sin^{-1} \lambda/1000d$. The angle between adjacent phasors is $2\pi/1000$ radians, or one-thousandth of a complete revolution. This means that one thousand adjacent phasors will make up a figure which is almost a circle. What figure will the phasors make if there are 10 000 slits in the grating? (Ten complete circles, just like the phasor diagram for a single slit as wide as the whole grating.) What will be the total amplitude A? (Zero.) For how many diffraction angles smaller than $\sin^{-1} \lambda/1000d$ will the total amplitude have been zero? (Nine.)

A path difference of $\lambda/1000$ in a 10 000-slit grating gives zero total amplitude, but what amplitudes do we get between the minima? Suppose we consider the angle halfway between $\sin^{-1} \lambda/1000d$ and the next angle which gives zero amplitude. What figure will the phasors make up? (Ten and a half complete circles.) What will be the total amplitude?

$$10.5 \times \pi A = A_0.$$

Thus $$A = 2A_0/21\pi.$$

What is the intensity? ($I = 4I_0/21^2\pi^2 = 4I_0/441\pi^2$.) The total amplitude will be almost a maximum where the phasor figure contains an odd half-circle, so between the tenth and the eleventh minima the intensity is not likely to exceed one-thousandth (roughly $4/441\pi^2$) of the intensity of a main maximum. However, nearer to the main maxima the intensities are bigger. For a 10 000-slit grating, if $\theta = \sin^{-1} \lambda/20\,000d$, what figures do all the phasors make up? (Semi-circle.) What is the total amplitude?

$$0.5 \times \pi A = A_0.$$

Therefore $$A = 2A_0/\pi.$$

What is the intensity? ($I = 4I_0/\pi^2$.)

Similarly, what is the intensity at angles $\sin^{-1} \lambda/10\,000d$ and $\sin^{-1} 3\lambda/20\,000d$? (Zero, $4I_0/9\pi^2$.)

In this way we can build up the intensity-angle graph (figure 101). When we approach the angle $\sin^{-1} \lambda/d$ we find the intensities $4I_0/\pi^2$, zero, and $4I_0/9\pi^2$ corresponding to the angles

$$\sin^{-1}(1-1/20\,000)\lambda/d, \quad \sin^{-1}(1-1/10\,000)\lambda/d,$$

and $\sin^{-1}(1-3/20\,000)\lambda/d,$

before the angle $\sin^{-1} \lambda/d$ itself, where $I = I_0$.

If the entire graph is plotted on the same scale, $\theta = 0$ being in the same place, where is the first principal maximum, $\theta = \sin^{-1} \lambda/d$? (About 100 metres beyond the righthand edge of the page.)

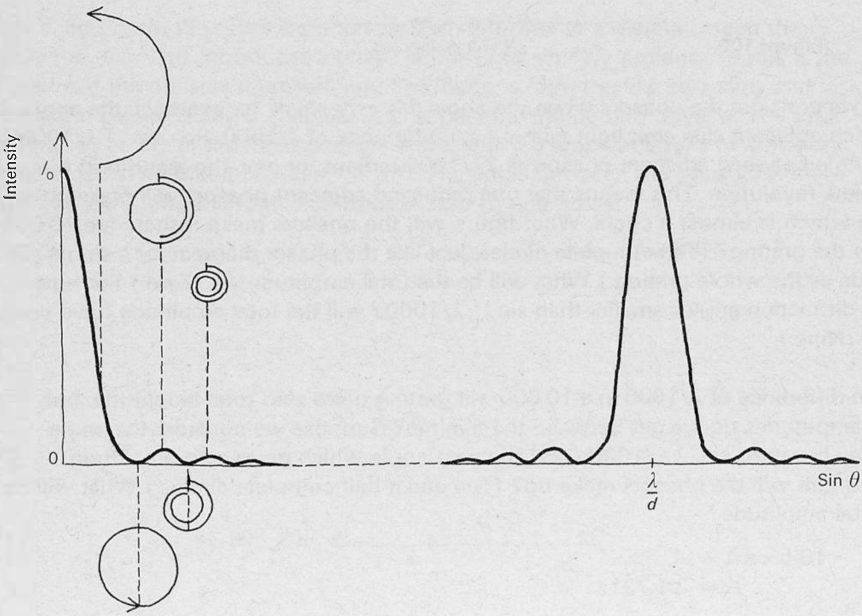


Figure 101

If the entire graph is plotted on such a scale that it is a distance 0.1 m from $\theta = 0$ to the first main maximum, how near to a main maximum is the intensity less than $I_0/1000$? (0.1 mm or less.)

The smaller maxima which are far from main maxima obviously have low intensity, but ones near main maxima are appreciable. Students may think that these smaller maxima add undesirable complication, and that a prism is simpler and therefore better. The quick commonsense answer is that the smaller maxima are negligible, but the question can be looked at another way. The smaller maxima exist because the grating's aperture is limited, and they are the single-slit diffraction pattern of the whole grating. A prism of the same aperture would show the same smaller maxima, and because a prism disperses the wave fronts of different colours by smaller angles, they matter more.

Appendix B

The problem of simplifying Maxwell's theory

> 'The present state of electrical science seems peculiarly unfavourable to speculation. . . . In order . . . to appreciate the requirements of the science, the student must make himself familiar with a considerable body of most intricate mathematics, the mere retention of which in the memory materially interferes with progress.'
>
> *Maxwell, J. C., from* The scientific papers of James Clerk Maxwell, *1890, ed. Niven, W. D., Volume I.* (Maxwell gave this as his reason for pursuing analogies and models in the study of electromagnetic waves.)

Because of its complexity, most school courses have avoided any treatment of electromagnetic wave theory. Teachers may be used to deducing a differential equation for waves from Maxwell's equations for fields in space without having thought much about how to teach a simpler version to sixth forms. This Appendix makes an attempt to bridge the gap. But familiarity with vector calculus is not needed in order to teach the course. Teachers to whom 'div' and 'curl' are unfamiliar may ignore the first of the remarks below without any direct effect on their teaching.

Maxwell's equations

Various experiments and speculations led Maxwell to think that the behaviour of electric and magnetic fields in empty space can be described by a set of equations, which in a modern form (applicable only to a space devoid of electrical charges and any other material), can be written

$$\operatorname{div} E = 0 \qquad \mathbf{1}$$

$$\operatorname{div} B = 0 \qquad \mathbf{2}$$

$$\operatorname{curl} E = -\frac{\partial B}{\partial t} \qquad \mathbf{3}$$

$$\operatorname{curl} B = \varepsilon_0 \mu_0 \frac{\partial E}{\partial t}. \qquad \mathbf{4}$$

The first two equations say that the electric and magnetic fluxes entering any small volume of space also leave that volume. The equations derive from Gauss's theorem for each field. The third equation represents the proportionality of an e.m.f. to the rate of change of magnetic flux producing it. The fourth equation represents an idea of Maxwell's to avoid some apparent inconsistencies.

The four equations imply the ability of fields to transmit waves, for algebra alone will produce a differential equation for waves from them.

Taking the curl of equation **3**:

$$\begin{aligned}\text{curl curl } E &= -\frac{\partial}{\partial t}(\text{curl } B)\\ &= -\frac{\partial}{\partial t}\left(\varepsilon_0\mu_0\frac{\partial E}{\partial t}\right) \text{ using equation 4}\\ &= -\varepsilon_0\mu_0\frac{\partial^2 E}{\partial t^2}\end{aligned}$$

But $\quad \text{curl curl } E \equiv \text{grad div } E - \nabla^2 E$

and $\quad \text{div } E = 0 \quad$ from equation 1

So $$\nabla^2 E = \varepsilon_0\mu_0\frac{\partial^2 E}{\partial t^2} \qquad \textbf{5}$$

Equation 5 is a differential equation whose solutions represent waves travelling in three dimensions at a speed c which is $1/\sqrt{\varepsilon_0\mu_0}$. The value of $1/\sqrt{\varepsilon_0\mu_0}$ is about 3×10^8 metres per second.

Restriction to a plane wave

This elegant and general derivation is not a good introduction to electromagnetic waves because it compresses so many stages of thought into such a small space. The work suggested in Unit 8 is designed to do the opposite, that is, to make clear a few physical principles even at the sacrifice of elegance and generality. It is restricted to plane waves, travelling parallel to $0x$, the electric field in the wave being parallel to $0y$ and the magnetic field parallel to $0z$. This restriction of directions is no more than the sort of thing one would do in considering a sound wave for the first time.

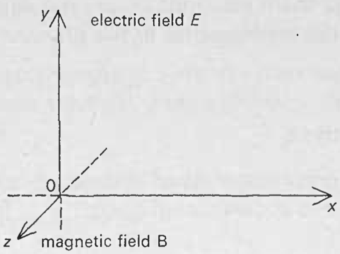


Figure 102

With these restrictions a possibility of simplifying the theory appears, if equation 3 is rewritten in the form of three equations, such as Maxwell used, to get

$$\frac{\partial E_z}{\partial y}-\frac{\partial E_y}{\partial z} = -\frac{\partial B_x}{\partial t} \qquad \textbf{6}$$

$$\frac{\partial E_x}{\partial z}-\frac{\partial E_z}{\partial x} = -\frac{\partial B_y}{\partial t} \qquad \textbf{7}$$

$$\frac{\partial E_y}{\partial x}-\frac{\partial E_x}{\partial y} = -\frac{\partial B_z}{\partial t} \qquad \textbf{8}$$

Because of the restrictions, all the terms are zero except two in the last equation,

$$\frac{\partial E_y}{\partial x} = -\frac{\partial B}{\partial t} \qquad 9$$

Similarly from equation 4,

$$-\frac{\partial B_z}{\partial x} = \varepsilon_0 \mu_0 \frac{\partial E_y}{\partial t} \qquad 10$$

The equivalent of taking the curl of equation 3 in the vector treatment is to differentiate equation 9 with respect to x,

$$\begin{aligned}\frac{\partial^2 E_y}{\partial x^2} &= -\frac{\partial}{\partial x}\left(\frac{\partial B_z}{\partial t}\right) \\ &= -\frac{\partial}{\partial t}\left(\frac{\partial B_z}{\partial x}\right)\end{aligned}$$

As we substituted equation 4, we can now substitute equation 10,

$$\begin{aligned}\frac{\partial^2 E_y}{\partial x^2} &= \frac{\partial}{\partial t}\left(\varepsilon_0 \mu_0 \frac{\partial E_y}{\partial t}\right) \\ &= \varepsilon_0 \mu_0 \frac{\partial^2 E_y}{\partial t^2}\end{aligned} \qquad 11$$

which is again a differential equation of a wave, but travelling in the direction $0x$. There is a similar wave equation for B_z.

But to use differential equations at all is still complicated and is not within easy grasp of students, who, although some of them may manipulate the equations confidently, do not see through the surface of the mathematics to the physical relationships it represents.

Use of a graphical method

A further simplification is to use a graphical method instead of calculus, and to start by supposing that a wave with E- and B-fields might exist.

Suppose the values of the electric and magnetic fields are such that, near a plane of particular x-value, a small increase Δx is associated with small decreases ΔE and ΔB in the respective fields. Then equation 9 says that because the electric field changes with x, the magnetic field must be changing with time. Then

$$\frac{\Delta E}{\Delta x} = -\frac{\Delta B}{\Delta t} \qquad 12$$

where Δt is the time taken by the magnetic field to change by ΔB.

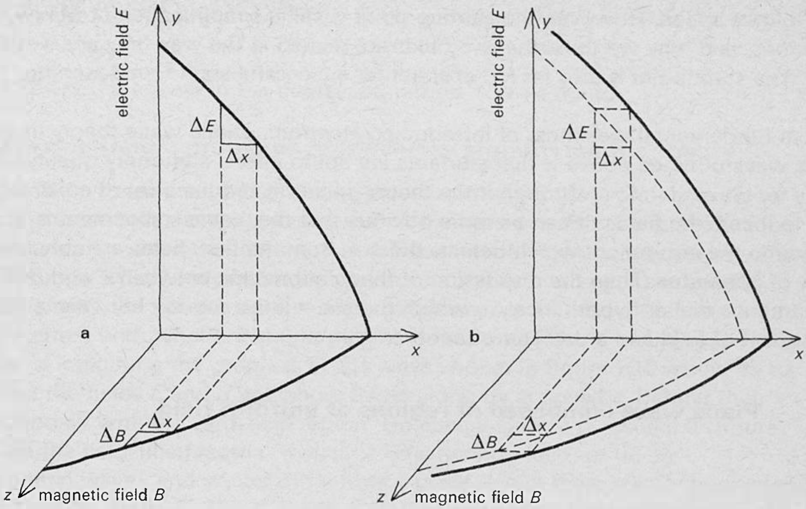


Figure 103

Now a change of the magnetic field by ΔB could be brought about by a movement of a particular value of B to a point Δx further along the x-axis, that is to say, a motion with speed $\Delta x/\Delta t$.

Similarly, from equation 10, as the magnetic field changes with x,

$$-\frac{\Delta B}{\Delta x} = \varepsilon_0\mu_0\frac{\Delta E}{\Delta t} \qquad 13$$

The electric field thus also changes in time. We now suppose that the pulse of the form shown in figure 103 *a* does travel forward a distance Δx in time Δt, keeping the form it has (see figure 103 *b*).

Then in a region Δx wide, B will change by ΔB and E by ΔE. Combining equations 12 and 13 by multiplying them together:

$$\frac{\Delta E\Delta B}{(\Delta x)^2} = \varepsilon_0\mu_0\frac{\Delta E\Delta B}{(\Delta t)^2}$$

$$\left(\frac{\Delta x}{\Delta t}\right)^2 = \frac{1}{\varepsilon_0\mu_0}$$

$$\frac{\Delta x}{\Delta t} = \frac{1}{\sqrt{\varepsilon_0\mu_0}} \qquad 14$$

The stages of this are simpler, and make it evident that an unevenness in one field must be accompanied by a changing magnitude of the other, and that the way in which the two fields achieve this reciprocally is for an unevenness in both to travel with a finite speed. However, the starting point is still a simplification of Maxwell's equations, and why we think the two fields are related in the way they are is still not clear. The standpoint is also far too abstract for successful sixth form teaching.

A more fundamental weakness of introducing electromagnetic wave theory in any of the ways outlined above is that students are apt to infer a stationary quality of space for waves to move through if the theory seems to require a set of co-ordinate axes to locate the fields. It can be more obvious that the waves' movement is in relation to the equipment which detects them if, from the first, fields are considered in terms of apparatus. Then the discussion of the relationships between E and B is tied to experiments, real or hypothetical, in which there is relative motion between a source of one sort of field and a detector of another.

Plane wave composed of regions of uniform field

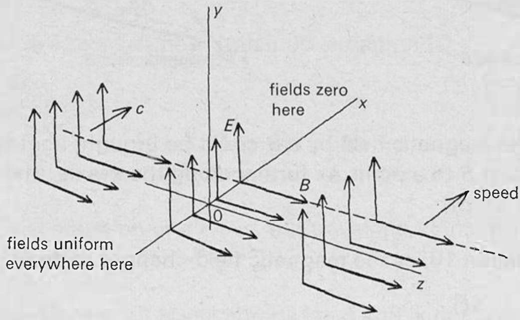


Figure 104

Figure 104 illustrates a further simplification. The wave now consists simply of a boundary, propagating at speed c along $0x$, between a region where the fields are zero, and one where there is uniform E- and B-field over all space behind the boundary. As before E is along $0y$ and B is along $0z$.

Equation 12 says that as B propagates along $0x$, occupying space where B was zero, E must change along x such that E-field appears on the boundary and

$$E = Bc \qquad \textbf{15}$$

Time changes of B go with spatial variation of E.

Equation 13 says a parallel thing about a propagating E-field, so that B appears at the boundary, where

$$B = \varepsilon_0 \mu_0 E c \qquad \textbf{16}$$

Time changes of E go with spatial variation of B.

Taking equations 15 and 16 together

$$c^2 = 1/\varepsilon_0\mu_0$$

Equations 15 and 16 appear in the course· equations 1 to 14 do not.

Moving fields: problems produced by simplifying the theory

The simplification introduced by restricting discussion to the wave shown in figure 104, which is essentially a step function, is a worthwhile one. But the simplification produces in its turn some other conceptual difficulties. (It might be fairer to say that it reveals difficulties which are present in any version of the theory, but are usually concealed beneath the smooth surface of the mathematics.) These difficulties are mainly concerned with what one might mean by a *moving field*. It would seem as if a simple way of imagining the progress of the wave shown in figure 104 would be to suppose that the fields E and B somehow travel along $0x$ at speed c, behind their moving boundary with the zero-field region. On such a view, a place like 0 (figure 104), within the field-filled region, would be 'in a moving field' or 'in an electromagnetic wave', and would differ from a place where there merely happened to be a pair of similar, static E- and B-fields. But there is *no* such difference; when one has said that at 0 there are a pair of perpendicular, uniform E- and B-fields, such that $E = Bc$, everything has been said that can be said about the fields at that place at that time.

The point is not just an academic one, for it is involved every time one switches on the electric light. After the switch is closed, a step-like wave front of B- and E-fields around the wires (accompanied by a wave front of moving charge on the wires) travels along near the wires. Ultimately, the fields become steady, and a steady current flows through the lamp (on d.c.). The simple case shown in figure 105 (the long parallel plate system discussed in the text) illustrates the point.

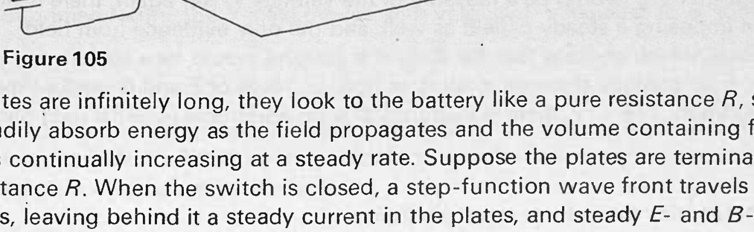


Figure 105

If the plates are infinitely long, they look to the battery like a pure resistance R, since they steadily absorb energy as the field propagates and the volume containing field energy is continually increasing at a steady rate. Suppose the plates are terminated by this resistance R. When the switch is closed, a step-function wave front travels along the plates, leaving behind it a steady current in the plates, and steady E- and B-fields between the plates. When the front reaches R, the same steady current flows in R, and energy is dissipated in R at the same rate as it was supplied to the parallel plate system while the front was propagating. The current in the plates, and the fields between the plates, retain the values they had behind the propagating wave front.

While the front is propagating, one would say that there was an energy flux through the field, at right angles to B and to E, and proportional to their product (see the Poynting vector in any advanced text). There is such an energy flux in any e.m. wave, behind the wave front. When the wave front has reached R, and the situation is an ordinary, steady-current one, which is not usually thought of in terms of an e.m. wave, energy is dissipated at the rate I^2R in the resistor. But the fields are as they were before the wave reached R, and it is still possible to regard the energy as being transported through the electromagnetic field, in this case at the same rate as when the wave was filling the space between the plates with field energy. (If the resistance is different from the characteristic resistance, there are reflected waves, and the system ends up with static fields not equal to those in the travelling wave situation. The energy I^2R is, however, still equal to the rate of transport of energy in the field, calculated from the final values of the fields.)

This result is a general one, and it is always possible to regard the energy dissipated in a circuit as arriving through the electromagnetic field around the circuit. Thus, the steady fields left behind a step-function wave front are in no respect different from the steady fields familiar from elementary treatments of E and B.

What moves in an electromagnetic wave is not the field, but a pattern of intensities. The only experimental evidence one can have that such a wave has passed is that each of a pair of recording stations distance d apart indicated changes of E and B which had a time lag d/c. Because the fields must have the ratio $E/B = c$, it seems as if one station could note the existence of the wave if it 'saw' such a pair of fields with magnitudes in this ratio. But this is wrong, for there is no difficulty in arranging to set up a pair of static fields, properly oriented, in the ratio $E/B = c$ to one another.

Somewhat similar remarks apply to the movement of the field of a magnet or a capacitor, at velocity v, discussed in the text in connection with experiment 8.13. For example, if the magnet is moved sideways at velocity v (see figure 71), there is an E-field given by

$$E = Bv.$$

It may seem that one could be sure the magnet was moving simply because the field E exists, and that E/B would be a measure of the velocity v. But again, there is no difficulty in imposing a steady E-field as well, and the only evidence from field measurements which certifies that the B-field is moving would be a set of indications from stations distance d apart, of rises and falls of E and B with a time lag d/v. If in addition, $E/B = v$, there is evidence that no additional external field has been imposed.

Relativity and electromagnetism

A careful analysis of experiment 8.13 involves relativity. One difficulty is that there are three sets of equations involved which contain similar terms in similar combinations, and the equations used in the text,

$$E = Bv$$
$$B = \varepsilon_0 \mu_0 E v,$$

are a simple version of only one set – the Maxwell equations – out of these three.

One other set of equations – the Lorentz force equations – defines E and B. In a reference frame in which a test charge moves with velocity v at right angles to a B-field, with an E-field at right angles to B, the force F is given by

$$F = Eq + Bqv.$$

In another frame, such that v becomes v', the equation

$$F' = E'q + B'qv'$$

defines B' and E' in the new frame.

The way F and v transform (q is invariant) determines how E and B transform. In this special case (fields transverse to velocities), if u is the relative velocity of the two frames, this leads to a second set of equations:

$$E' = \gamma(E + Bu)$$
$$B' = \gamma(B + Eu/c^2)$$
where $\gamma = (1 - u^2/c^2)^{-\frac{1}{2}}$.

Figure 106 illustrates these transformations. What seems to be an E-field in one frame, is part E- and part B-field in another frame.

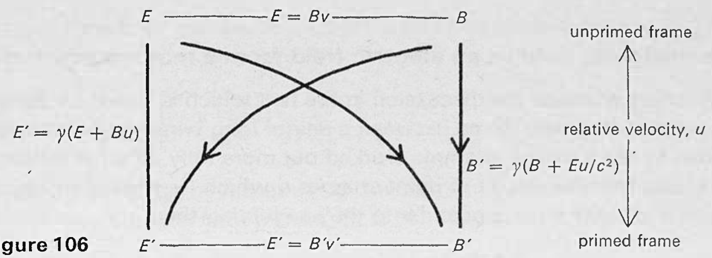


Figure 106

If in the unprimed frame, the B-field moves with velocity v transverse to B, there is an E-field

$$E = Bv.$$

This is one of Maxwell's equations (essentially curl $E = -\mathrm{d}B/\mathrm{d}t$) applied to this simple, moving, uniform field situation. From the transformations, it turns out that in the primed frame the equation becomes,

$$E' = B'v'.$$

That is, the form of the equation is identical. (This is what is meant by saying that Maxwell's equations are covariant.) It follows that in the arguments in the text about experiment 8.13, the term B must be understood to be the field of the *moving* magnet, just as E is the field detected in the case where the magnet moves. The equations

$$E = Bv$$
$$B = \varepsilon_0 \mu_0 E v,$$

used in the text, are simple forms of Maxwell's equations. They represent necessary relations between moving uniform slabs of E- and B-field if either is supposed to move. They are *not* recipes for transforming from one frame of reference to another.

It then becomes possible to argue that if such a pair of fields propagates on its own in empty space, both these necessary relations obtain together and there is only one unique velocity, $c = 1/\sqrt{\varepsilon_0 \mu_0}$, at which the fields can be supposed to propagate. The speed c, like v previously, is the velocity of propagation of both fields. It is their velocity relative to some field detector (a charge which can be given a small suitable velocity to detect B). The remarkable thing is that c seems to have the same value relative to any detector, regardless of its velocity relative to another detector. All detectors see the wave going at velocity c. This is a consequence of the covariance of the equations.

The above discussion is restricted to E and B because these are the vectors used in the course. There is no intention to suggest that in every circumstance, B and E are one analogous pair, D and H being the other. In some circumstances, it may be more natural to associate E with H, and D with B.

The magnetic field as an electric field from a moving point of view

It may help teachers who use the discussion in the text which is based on figure 87, illustrating how the 'magnetic' force between a pair of long wires might be understood within relativity, to see a similar example worked out more fully. In what follows, we consider the forces between a pair of point charges q which, as shown in figure 107, may move with a velocity v perpendicular to the line joining them.

Figure 107

This problem is discussed very clearly in French, *Special relativity*, and in Gibson, *Basic electricity*. See also the Berkeley Physics Course, Volume 2, *Electricity and magnetism*.

We shall call the charge on which acts the force which we choose to discuss, the *test* charge, and the other the *source* charge. The argument is symmetrical, and either charge may be regarded in either role.

We use one definition and two principles.

Definition The force on a test charge at rest with respect to the observer gives the 'electric' part of any force, while that force proportional to its velocity with respect to the observer is the 'magnetic' part.

Principle 1 The force on a test charge which moves past an observer, measured *by the observer,* is not different from the force on the charge if it is at rest *as long as the source charge is at rest* with respect to the observer. Evidence: electrostatic calculations correctly predict the motion of moving electrons deflected by charged plates in a cathode ray tube.

Principle 2 Force = $\mathrm{d}p/\mathrm{d}t$ where p is momentum. Observers agree about momenta at right angles to their relative velocity. But the interval $\mathrm{d}t$ is given its *smallest* value by an observer at rest with respect to a particle he is timing; all other observers obtain values for $\mathrm{d}t$ larger by a factor $1/\sqrt{1-v^2/c^2}$. So the observer at rest with respect to the particle assigns the *largest* value to such forces.

$$F_{\text{rest}} = (F_{\text{moving}})/\sqrt{1-v^2/c^2}$$

'Electric' force To an observer for whom the *source* charge is at rest and the *test* charge is moving, the force on the *test* charge is correctly given by Coulomb's Law, using *principle 1*,

$$F = \left(\frac{1}{4\pi\varepsilon_0}\right)\frac{q^2}{r^2}$$

To an observer for whom the *source* charge moves and the test charge is at rest, the same force becomes, using *principle 2*,

$$F_{\text{electric}} = \left(\frac{1}{4\pi\varepsilon_0}\right)\frac{q^2}{r^2}\times\frac{1}{\sqrt{1-v^2/c^2}}$$

We have called this F_{electric} because, using the *definition*, it is the 'electric' part of the force between two charges *both* moving past the observer.

'Magnetic' force To an observer for whom a pair of charges are at rest, the force between them is given by Coulomb's Law.

$$F = \left(\frac{1}{4\pi\varepsilon_0}\right)\frac{q^2}{r^2}.$$

To another observer, for whom *both* charges move at velocity v, the force between them will be *smaller*, using *principle 2*, and he will observe a force:

$$F_{\text{total, moving}} = F_{\text{electric}} + F_{\text{magnetic}} = \left(\frac{1}{4\pi\varepsilon_0}\right)\frac{q^2}{r^2}\sqrt{1-v^2/c^2}$$

We have called the total force $F_{electric}+F_{magnetic}$ because it is the total force between moving charges, and we want, using the *definition*, to split it into two parts. Taking the previously calculated value of $F_{electric}$, we have for the magnetic part of the force:

$$F_{magnetic} = F_{total,\ moving} - F_{electric} = \left(\frac{1}{4\pi\varepsilon_0}\right)\frac{q^2}{r^2}\left(\sqrt{1-v^2/c^2} - \frac{1}{\sqrt{1-v^2/c^2}}\right)$$

Expanding $\sqrt{1-v^2/c^2}$ and neglecting terms smaller than v^2/c^2:

$$F_{magnetic} = -\frac{v^2}{c^2}\left(\frac{1}{4\pi\varepsilon_0}\right)\frac{q^2}{r^2}$$

The minus sign arises because $F_{magnetic}$ is an attraction, reducing the value of the force *from the higher value* $F_{electric}$.

Note that as $\dfrac{1}{c^2\varepsilon_0} = \mu_0$, the above expression gives the usual result for the magnetic force between two moving charges.

Appendix C
Recent experiments in relativity

'Ether-drift' experiments L. Essen ('A new Aether-drift experiment', *Nature* **175**, 793–4, 1955) did an experiment which was analogous to the Michelson–Morley experiment, and which some teachers may prefer to describe since it did not involve interference fringes.

Essen made a hollow tube, up and down which radio waves travelled in a standing wave, making a clock of high and stable frequency. The tube was rotated about a line perpendicular to its length, and the frequency compared with that of a standard oscillator or clock, which was not rotating. Shifts in frequency as the tube rotated were looked for, but the only effect found was due to the magnetic field of the Earth, which made the tube contract slightly. No velocity greater than $\frac{1}{100}$ of the Earth's orbital velocity could be detected.

Cedarholm, J. P., Bland, G. F., Havens, B. L., and Townes, C. H. ('New experimental test of special relativity', *Phys. Rev. Letters* **I**, November 1958, page 342) compared the frequencies of two ammonia maser beams pointing in opposite directions and rotated. Result: no velocity detectable up to $\frac{1}{100}$ of the Earth's velocity.

Champeney, D. C., Isaak, G. B., and Khan, A. M. (see, for example, 'An aether-drift experiment based on the Mossbauer effect', *Physics Letters*, **7**, No. 4, 1963) used the Mossbauer effect to show that there is no detectable velocity greater than one-millionth of the Earth's velocity. This test used the sharp resonance absorption of gamma rays (falling on nuclei identical to those which emitted them) as a clock, which was again rotated.

Velocity independent of source velocity. Sadeh, B. ('Experimental evidence for the constancy of the velocity of gamma rays using annihilation in flight', *Phys. Rev. Letters* **10**, 271–3, 1963) showed that the velocity of gamma rays emitted from a stationary source was the same to within 10 per cent as the velocity of gamma rays from a moving source, by a direct timing method. (Positrons were annihilated by electrons, first at rest and then in flight.)

Alvager, T., Farley, S. J. M., Kjellman, J., and Wallin, I. ('Test of the second postulate of special relativity in the GeV region', *Physics Letters*, **12**, 260, 1964) timed the velocity of gamma rays produced by the decay of π^0-mesons; travelling at about 99.9 per cent of the velocity of light. The gamma rays still travelled at velocity c.

Time dilation. The lifetimes of unstable particles travelling at high speeds have been compared with their rest lifetime. Early trials used cosmic ray μ-mesons, and the PSSC film 'Time dilation' shows such an experiment. Later tests used π-mesons from accelerators.

Durbin, R. P., Loar, H. H., and Havens, W. W. ('The lifetime of π^+ and π^- mesons', *Phys. Rev.* **88**, 179–83, 1952) used π^+ and π^- mesons of energy 70 MeV.

Greenberg, A. J., and others ('Charged pion lifetime and a limit on a fundamental length', *Phys. Rev. Letters* **23**, 21, 1267–70, 1969) also used π-mesons and showed that the calculated time dilation factor (2.4) was correct to within 0.4 per cent.

Limiting velocity. Bertozzi, W. ('Speed and kinetic energy of relativistic electrons', *Am. J. Phys.,* **32**, 551–5, 1964, and the PSSC film 'The ultimate speed') showed that the velocity of accelerated electrons rose towards a limit c as the energy increased (maximum energy 15 MeV).

Appendix D
Use of a laser

Low power gas lasers (0.25 mW) run continuously and produce monochromatic light at an intensity which enables any diffraction effect to be visible by diffuse reflection from a screen. With a laser it is possible to show a whole class simultaneously all the phenomena which are to be discussed, under conditions where individual features of diffraction patterns, etc. can be pointed out to students. Such a method saves a great deal of teaching time, and is much better than demonstrations using conventional light sources. But it is not as effective an introduction to the subject as good individual experiments. Used in a demonstration, it does not give students the chance to change for themselves the variables which control what they see, and, in students' minds, it might associate small-scale optics with specialized equipment rather than with the situations of everyday life. Later, when students discuss quantitatively the effects they have encountered, laser demonstrations have irreplaceable value as a quick reference to the experimental observations. Even at the present cost, lasers may be considered economical in schools where several large classes can see them. The luminous intensity of lasers is high enough to endanger eyes if they are used ignorantly or carelessly, in the same way that binoculars are dangerous where the Sun may come into field of view. But with reasonable precautions lasers may be used by students. The Department of Education and Science issues administrative memoranda from time to time setting out the required rules. See Administrative memorandum 7/70 'Use of lasers in schools and other educational establishments'.

The Unit has been written with a laser in mind and there are references to its possible use, but it is not yet economical for small numbers of students and must be regarded only as a desirable luxury.

Films, books, and apparatus

List of films

16 mm film 'The velocity of gamma rays' 16 minutes, colour, sound.
No. 21.7853. Rank Audio Visual Ltd. This film was made in conjunction with the Nuffield Advanced Physics Project.

16 mm film 'Frames of reference' 26 minutes, black and white, sound.
No. 900 4141-9. Guild Sound and Vision Ltd (formerly Sound Services).

16 mm film 'Time dilation' 36 minutes, black and white, sound.
No. 900 4040-3. Guild Sound and Vision Ltd.

16 mm film 'The ultimate speed' 38 minutes, black and white, sound.
No. 900 4039-3. Guild Sound and Vision Ltd.

None of the last three is essential. 'Time dilation' is the most useful one.

Books and further reading

Page numbers of references in this *Guide* appear in bold type.

For teachers

Crawford, F. S. (1968) Berkeley Physics Course, Volume 3 *Waves.* McGraw-Hill.
Einstein, A., Lorentz, H. A., Minkowski, H., and Weyl, H. (1958) *The principle of relativity.* Dover. **122.**
Feynman, R. P., Leighton, R. B., and Sands, M. (1963) *The Feynman lectures on physics.* Volume 1. Addison-Wesley. **74, 122**
French, A. P. (1968) *Special relativity.* Nelson. **122, 126, 148.**
Gibson, W. M. (1969) *Basic electricity.* Penguin. **126, 148.**
Hertz, H. (1962) *Electric waves.* Dover. **64.**
Kittel, C., Knight, W. D., and Ruderman, M. A. (1965) Berkeley Physics Course, Volume 1. *Mechanics and relativity.* McGraw-Hill. **122.**
Purcell, E. M. (1965) Berkeley Physics Course, Volume 2. *Electricity and magnetism.* McGraw-Hill. **126, 148.**
Sherwin, C. W. (1961) *Basic concepts of physics.* Holt, Rinehart, & Winston. **122.**

For students

Textbooks

Arons, A. B. (1965) *Development of concepts of physics.* Addison-Wesley.
Baez, A. V. (1967) *The new college physics.* Freeman.
PSSC (1968) *College physics.* Raytheon. **26, 114.**
PSSC (1965) *Physics.* (Second edition.) Heath. **26.**
Rogers, E. M. (1960) *Physics for the inquiring mind.* Oxford University Press.

Further reading

Bondi, H. (1964) *Relativity and common sense.* Heinemann. **110, 120, 122.**
Burbidge, G., and Hoyle, F. (1966) 'The problem of the quasi-stellar objects.' *Scientific American* Offprint No. 305. W. H. Freeman. **36, 122.**
Landau, L. D., and Rumer, G. B. (1960) *What is relativity?* Oliver & Boyd. **110.**
Sandage, A. R. (1956) 'The red-shift.' *Scientific American* Offprint No. 240. W. H. Freeman. **122.**
Shankland, R. S. (1964) 'The Michelson-Morley experiment.' *Scientific American* Offprint No. 321. W. H. Freeman. **114.**
Smith, A. G. (1967) *Radio exploration of the Sun.* Van Nostrand.
Smith, F. Graham (1966) *Radio astronomy.* Penguin. **36.**
Tolansky, S. (1968) *Revolution in optics.* Penguin.

Apparatus

14	e.h.t. power supply	8.9
23	microscope	8.8
27	transformer	8.1, 8.3, 8.15
46/1	translucent screen	8.5, 8.8
47	illuminant	8.2
52 L	mounted bell-push	8.11
59	l.t. variable voltage supply	8.5, 8.8, 8.12
64	oscilloscope	8.7, 8.12
90	ripple tank kit	8.2
92 B	Magnadur magnet	8.13
92 G	double C-core and clip	8.11
92 I	mild steel yoke	8.13
94 A	lamp, holder, and stand	8.1, 8.3, 8.5, 8.8, 8.15
94 B	housing shield	8.5, 8.15
94 G	barrier	8.15
94 H	plano-cylindrical lens, +7 D	8.15
100/1	rectangular plastic tank	8.15
106/1	dynamics trolley	8.13
116	plane mirror	8.1
157	microphone	8.7
158	class oscilloscope	8.11
181	general purpose amplifier	8.4, 8.6, 8.14, 8.16, 8.18
183	loudspeaker	8.4, 8.6, 8.7, 8.14, 8.16, 8.18
184/1	3 cm wave transmitter	8.1, 8.4, 8.6, 8.14, 8.16, 8.18
184/2	3 cm wave receiver	8.6, 8.14, 8.16, 8.18
191/1	coarse grating	8.8
501	metre rule	8.4, 8.5, 8.7, 8.8
503–6	retort stand base rod, boss, and clamp	8.7
541/1	rheostat (10 to 15 Ω)	8.2, 8.7
1000	leads	8.4, 8.7, 8.12, 8.13
1001	galvanometer (internal light beam)	8.9, 8.13, 8.14
1003/2	milliammeter (10 mA)	8.11
1004/1	voltmeter (1 V)	8.11
1007	double-beam oscilloscope	8.10, 8.11
1009	signal generator	8.7
1014	wax lens	8.4, 8.6
1025	pair of capacitor plates	8.13
1030	high inductance coil	8.11
1031	200 kHz pulse generator	8.10
1033	cell holder with U2 cells	8.2, 8.11, 8.13
1035	pre-amplifier	8.7

	set of solenoids	8.12
1038	Hall probe and circuit box	8.13
1039/1	search coil (axial)	8.12
1040	clip component holder	8.10, 8.11, 8.13
1041	potentiometer holder	8.11
1045	diode probe for microwave experiments	8.4
1050	15 cm dipoles and oscillator	8.9, 8.14
1051	*Small electrical items*	
	resistor, 68 Ω	8.10
	resistor, 120 Ω	8.11
	preset potentiometer, 5 kΩ	8.11
	electrolytic capacitor, 1000 μF 15 V	8.11
1053	*Local purchase items*	
	card	8.1
	aluminium foil	8.1, 8.8
	cardboard 35 mm slide mounts	8.1, 8.8
	metal screen about 0.3 m square	8.4, 8.6, 8.16, 8.18
	transparent ruler with mm graduations	8.1, 8.5
	bent metal plate	8.6
	Plasticine	8.13
	hardboard screen about 0.3 m square	8.16, 8.18
1054	*Expendable items*	
	copper wire, 36 s.w.g., bare	8.1, 8.13
1055	*Small laboratory items*	
	whistle	8.17
	rubber tubing, about 1 m long, to fit whistle	8.17
1057	a.c. ammeter	8.12
1062	drum of coaxial cable	8.10
1063	multiple light source	8.1
1067	physical optics kit	8.3
1067/1 A	sheet with slit and holes	8.5
1067/1 B	holder for lens of diameter 37 mm	8.5
1067/1 C	plano-convex lens, diameter 37 mm	8.5
1067/1 H	small translucent screen	8.5
1067/1 J	holder for two halves of a razor blade to be used as a single unit	8.1, 8.5
1067/2 M	slit of variable width	8.5, 8.8
1067/2 N	set of coarse gratings	8.8
1067/2 O	set of 1, 2, 3, 4, 5, 6 parallel slits	8.5, 8.8
1067/3 R	set of three colour filters	8.1, 8.8
1067/3 S	small piece of fine black chiffon	8.8
1073	concave diffraction grating	8.8
1079	flat solenoid	8.12
1083	polarizing filter	8.15

Index

Where significant information is contained in an illustration or diagram, the page reference is italicized. In general, odd-numbered page references are to the main text, and even-numbered references are to commentary material.

Editor **Jon Ogborn**

Contributors
P.J.Black Jon Ogborn *Joint organizers, Advanced Physics*
W.Bolton R.W.Fairbrother G.E.Foxcroft Martin Harrap
John Harris A.L.Mansell A.W.Trotter

There are ten Units in the Advanced Physics course. This is the *Teachers' guide* for Unit 8, *Electromagnetic waves.* It is intended to provide whatever information and ideas are required for the day-to-day teaching of the Unit. The book begins with an Introduction setting out the purpose of the Unit, a summary of the Unit, and a list of suggested experiments. Following this, the main text consists of four Parts, 'Looking through holes', 'Spectra', 'Electric waves', and 'Relativity'. It contains teaching suggestions, details of experiments, and a commentary giving background information and other guidance. There are also Appendices on 'Phasor treatment of diffraction', 'The problems of simplifying Maxwell's theory', 'Recent experiments in relativity', and 'Use of a laser', and lists of relevant films, books, and apparatus.

Published by Penguin Education

140827188