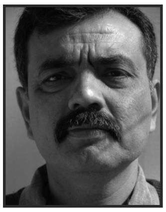
Electricity, Magnetism and Electromagnetic Theory

About the Authors



Professor Shobhit Mahajan received his Masters and Doctorate degrees in Physics from the University of California, Berkeley. He has taught at St. Stephen's College, Delhi and is currently Professor of Physics at the Department of Physics & Astrophysics, University of Delhi. He has worked in the fields of astroparticle physics, cosmology and particle physics.

Professor Mahajan has been the author of numerous articles on Science & Technology and has been a consulting editor for Down to Earth magazine. In addition, he has written more than a dozen books on popular science titles and has served as a consultant for television shows and documentary series.

Professor S Rai Choudhury received his Masters and Doctorate degrees in Physics from the University of Delhi. Subsequently, he has held Postdoctoral research and faculty positions at Moscow State University, Princeton University (USA), University of Michigan (USA), Purdue University (USA), Rockefeller University (USA), University of Melbourne (Australia). Professor Choudhury has authored many research papers in internationally recognised journals. He has been a Professor of Physics at the University of Delhi till 2006 . He was a Ramanna fellow of the Department of Science and Technology (Govt. of India) from 2006 to 2009 and since 2009 Professor of Physics at the Indian Institute of Science Education and Research (IISER) Bhopal, M.P.

Professor Choudhury has been an Associate of the International Centre of Theoretical Physics, Trieste, Italy and a Senior Associate during 1981–1987. He was elected Fellow of the Indian Academy of Sciences, Bangalore in 1986.

Professor Mahajan and Prof. Choudhury have co-authored (along with A. Beiser), *Concepts of Modern Physics,* 6E, SIE, in 2009.

Electricity, Magnetism and Electromagnetic Theory

Shobhit Mahajan
Professor of Physics
Department of Physics and Astrophysics
University of Delhi, Delhi

S Rai Choudhury
Professor of Physics
Indian Institute of Science Education and Research
Bhopal, Madhya Pradesh

Tata McGraw Hill Education Private Limited
NEW DELHI

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

Tata McGraw Hill

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

Electricity, Magnetism and Electromagnetic Theory

ISBN (13): 978-1-25-900459-9
ISBN (10): 1-25-900459-7

Vice President and Managing Director—McGraw-Hill Education: *Ajay Shukla*

Head— Higher Education Publishing and Marketing: *Vibha Mahajan*
Publishing Manager—SEM & Tech Ed.: *Shalini Jha*
Asst. Sponsoring Editor: *Smruti Snigdha*
Development Editor: *Renu Upadhyay*
Copy Editor: *Preyoshi Kundu*
Sr Production Managers: *P. L. Pandita*, *Satinder S Baveja*

Marketing Manager—Higher Education: *Vijay Sarathi*
Product Specialist: *Sachin Tripathi*
Graphic Designer—Cover: *Meenu Raghav*

General Manager—Production: *Rajender P Ghansela*
Production Manager: *Reji Kumar*

Typeset at Script Makers, 19, A1-B, DDA Market, Paschim Vihar, New Delhi 110063 and printed at
Pushp Print Services, B-39/12 A, Gali No. 1, Arjun Mohalla, Moujpur, Delhi 110053

Cover Printer: SDR

RAXLCRAHDRZRC

The McGraw·Hill Companies

The authors would like to dedicate
this book to the memory of their mothers

Contents

Preface

About the Book

Electricity, Magnetism and Electromagnetic Theory forms a part of the core curriculum of any physics course, both at the undergraduate and postgraduate levels.

The book does not assume any prior knowledge of the subject except at the secondary school level. It does however, assume some familiarity with the mathematical techniques used, in particular vector analysis and differential equations. These and other mathematical techniques used are collectively introduced and discussed in 'Mathematical Preliminaries' at the beginning of the book.

The book grew out of a need felt by us of a textbook which covers the entire range of topics in electromagnetism which are usually taught both in introductory and in more advanced courses at Indian universities. There are books available in the market which serve as excellent introductory textbooks. Some even have a comprehensive treatment for more advance level courses. Our aim is to span the entire range covered by such books and make the subject interesting, clear and instructive.

Salient Features of the Book

- Concepts have been lucidly explained through interactive pedagogy
- All important topics such as Electric field, Electric Potential, Electric fields in matter, Electrostatics in presence of charges, Charges in motion and electric currents, Magnetic field, Magnetic fields in matter, Alternating currents, Maxwell's equations, Relativity and electrodynamics, and Charges in motion have been discussed
- Stepwise problem solving approach has been adopted
- Additional information has been provided, appositely in boxes for further understanding
- Pedagogy includes

  90 Solved examples within the chapters
  120 Objective type questions
  388 Exercise problems
  314 Illustrations

Structure of the Book

The book consists of 12 chapters and one appendix. Each chapter begins with Learning Objectives – a set of concepts, which are elaborated in the respective chapter.

The book begins with a preliminary chapter on mathematics which lays the foundation for the subsequent chapters in the book. Topics like vector analysis and vector calculus, differential equations, complex variables, matrices and special functions have been summarised in this chapter .

Chapter 1 introduces basics of electrostatics. Coulomb's law, Concept of electric field, fields due to discrete and continuous charge distributions, Differential form of Coulomb's law, Gauss's law and Simple applications of Gauss's law have been discussed.

Chapter 2 introduces the concept of electric potential and its relationship to electric field; the laws of electrostatics are written in terms of the electric potential; the Laplace and Poisson equations are introduced; boundary value problems for solution of these equations have been discussed with the help of examples.

Modifications of the laws of electrostatics in the presence of polarisable but non-conducting medium is explained in **Chapter 3**. Solution of Laplace and Poisson equations in the presence of dielectrics are discussed; concept of local fields are introduced and calculated for a few situations. Microscopic theories of dielectric phenomena have been eleborated in this chapter.

Chapter 4 introduces conductors and their behaviour in the presence of electric fields along with the modification of electric potentials and fields in the presence of conductors in the neighbourhood. The concept of capacitance of conductors, typical calculations of capacitances and laws for combination of capacitances have also been covered.

Electric currents are introduced in **Chapter 5** and the basic laws governing the currents are stated. Microscopic classical theory of conduction is discussed; Kirchhoff's laws, network theorems are stated and proved. Calculation of varying currents in RC circuit is discussed.

Chapter 6 introduces the magnetic effects of currents and the basic laws in that context. Hall effect and its basic theory, Sources of magnetic fields and the laws for magnetic fields created by current sources have been discussed. Magnetic vector potential has been introduced and laws restated in terms of the potentials.

Chapter 7 discusses magnetic materials and modifications of the laws in their presence. Elementary theories of various forms of magnetism are discussed. Boundary value problems in the presence of ferromagnetic materials are discussed with some typical examples. Concept of magnetic circuits is introduced with some illustrative examples using the concept.

Phenomena of electromagnetic induction and the laws thereof are discussed in **Chapter 8**. The concept of self and mutual inductance, their coefficients and their role in working out currents in circuits are introduced along with the concept of energy stored in magnetic fields and their evaluation.

Chapter 9 is devoted to the analysis of circuits with time dependent currents. Alternating currents (AC) have been discussed in detail. Apart from the concepts of complex impedances and analysis of typical AC circuits, Phenomenon of resonance and quality factor in resonant circuits and performance of AC transformers have been discussed.

Chapter 10 discusses the shortcoming of Ampere's law, and Maxwell's introduction of displacement current. The complete set of Maxwell's equations both in free space and in the presence of matter are included. Wave solutions of Maxwell's equations are worked out and their properties discussed. Reflection and refraction at interfaces are discussed.

Chapter 11 discusses applications of Maxwell's equations for calculations of electric and magnetic fields created by moving charges . Propagation of waves in wave guides, radiation emitted by simple antenna are worked out as application of Maxwell's equations.

Chapter 12 is devoted to the special theory of relativity. The essentials of the theory—Lorentz transformations, relative nature of space and time are elaborated. Obvious consequence like length contraction, time dilation, relativistic energy and mass equivalence are presented. The transformation of electric and magnetic fields due to Lorentz transformations are presented and discussed in detail.

In addition, each chapter has several solved examples to illustrate the use of the concepts and relevant techniques. Important key words have been highlighted in bold and the important equations have been put in a box to make identification easier. Some chapters contain topics which are usually not covered in standard textbooks but are important for understanding. Such topics have been marked with the symbol ‘*’ in the headings.

Advanced topics have also been included at the end of the chapter. These topics, though interesting and important, are not usually part of any regular curricula but will equip students with the desired edge and thereby generate interest. At the end of each chapter, we have collated what we consider the key points that need to be reiterated about the topic. To help students assimilate the material in the text, there are ten multiple choice or short answer conceptual questions at the end of each chapter. For students to learn the art of problem solving, many unsolved problems have been added at the end of each chapter.

Web Supplement

Interesting web supplements are available at *http://www.mhhe.com/mahajan/emet1*

Acknowledgements

We have, over the years, benefitted from discussions with several colleagues and many students who are too numerous to be acknowledged individually. Shobhit Mahajan would like to acknowledge his immense gratitude to Dr. S.C. Bhargava, his former teacher who introduced the subject to him and showed how much fun it can be. Unfortunately, Dr. Bhargava did not live to see and comment on our efforts.
Our thanks are also due to the reviewers for their valuable suggestions.

Nataraja Sarma (Retd.)	*Bhabha Atomic Research Centre, Pune*
Doreen A. War	*St. Mary's College, North Eastern Hill University, Shillong*
Preeta Sharan	*Oxford College of Engineering, Bangalore*
S K Saha	*Indian Institute of Astrophysics, Bangalore*

Our sincere gratitude to Late **Dr. Arawind Parasnis (IIT, Kanpur)** whose efforts in reviewing the script were commendable.

We would also like to thank the editorial team of Tata McGraw Hill Education for their support and cooperation in bringing out this book.

Feedback

We would be very grateful to readers who give comments, suggestions or point out errors in the book. These can be sent by mail to the authors *shobhit.mahajan@gmail.com* and *sraichoudhury@gmail.com* so that they can be incorporated in the subsequent editions of the book.

Shobhit Mahajan
S Rai Choudhury

Publisher's Note

Do you have a feature request? A suggestion? We are always open to new ideas (the best ideas come from you!). You may send your comments to *tmh.sciencemathsfeedback@gmail.com* (don't forget to mention the title and author name in the subject line).

Guided Tour for Better Learning

Mathematical Preliminaries

Learning Objectives

- Review of results from vector analysis
- Definition and properties of solid angle
- Definition and properties of Dirac delta function
- Review of complex numbers
- Review of some results from matrix algebra
- Review of ordinary differential equations
- Definition and properties of some special functions

0.1 VECTOR ALGEBRA

0.1.1 Scalars

Scalars are physical quantities described by their magnitude. Examples of scalars are mass, temperature, electric charge, etc.

A **scalar field** is a physical quantity that varies from point to point and at each point, its value is specified by its magnitude at that point. An example of the scalar field is the gravitational or electric potential. It varies from point to point in general and at every point, one only needs to know the magnitude of the potential to completely specify it.

0.1.2 Vectors

Vectors are physical quantities that are described by their magnitude and direction. They are usually represented with an arrow over them ($\vec{a}$) or in bold face (**a**). In three dimensions, vectors can be represented by a sum of projections on any three mutually orthogonal vectors of unit magnitude. These mutually orthogonal vectors of unit magnitude are called **basis vectors**. The most common basis in Cartesian coordinates are unit vectors along the x, y and z directions and these are represented by $\hat{i}$, $\hat{j}$ and $\hat{k}$ respectively. Thus, a typical vector $\vec{A}$ can be represented by

$$\vec{A} \quad A_x\hat{i} \quad A_y\hat{j} \quad A_z\hat{k}$$

where A_x, A_y, A_z are the x, y, z components of $\vec{A}$ and are really the projections of $\vec{A}$ ... This decomposition of a vector follows from the vector law of addition, which specifi... can be added. A simple geometric way to see the vector addition law is to think of t...

Mathematical Preliminaries *summarises the important mathematical results which are used in the subsequent chapters.*

Learning Objectives *point out the main topics to be discussed in the chapter.*

Learning Objectives

- To comprehend the nature of electric charge and its properties.
- To learn about the nature of force between point charges.
- To learn about how forces due to several point charges add vectorially to give the net force and to use this fact for continuous charge distributions.
- To understand the concept of an electric field and how it leads to a local description of the electrostatic phenomenon.
- To learn how to calculate electric field due to various charge configurations.
- To understand the concept of electric lines of force and electric flux.
- To learn about Gauss's law and its use in determining the electric field for situations with symmetry.

Additional Information *has been summarised in boxes wherever necessary, to facilitate further understanding for students.*

Let us summarise what we have learnt about the boundary conditions in the presence of a conductor, whether charged or not:

> A conductor is throughout at the same potential including its surfaces which therefore are equipotential surfaces. The surfaces may carry surface charges. The electric field at the surface is always normal to the surface at every point. In the presence of any hollow cavity in a conductor, there is no electric field in the cavity if there are no free charges in the cavity and the surface of the cavity carries no surface charge. If a hollow cavity contains free charges, there is an electric field in the cavity and there are surface charges in the cavity. The electric field at the surface of the cavity is normal to the surface at every point on the surface and the conductor throughout is at the same potential.

Examples *are interspersed throughout each chapter to clarify and illustrate the concepts and principles.*

EXAMPLE 1.2 AB is a line charge of length L which is uniformly charged and carries a total charge Q. A point charge q is located at the point C, situated at a distance R along the right bisector of AB. Calculate the electric force on the point charge due to the line charge.

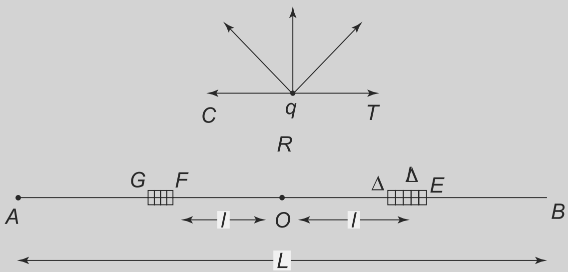


Fig. 1.4 *Example 1.2*

PROBLEM 1.3 A thin spherical shell has a radius R and is uniformly charged with a charge density σ per unit area. Calculate, by direct use of Coulomb's law, the electric force on a point charge q located at a distance r from the centre of the shell.

Problems *within the chapter will help students in self-study and analyse the extent of their understanding of practical concepts.*

Starred Sections provide additional information to the readers.

which gives us the electrostatic energy for 4 NaCl molecules as

$$U_8 \quad -\frac{e^2}{a}(5\ 08)$$

Thus, the energy per molecule is simply

$$U \quad -1\ 27\frac{e^2}{a}$$

PROBLEM 2.11 In the experiment done by Rutherford to determine the structure of nuclei, a beam of α particles which has twice the charge of proton and approximately four times as heavy, was directed towards a gold leaf. Each of the α particles had an energy of 8 MeV and some of the α particles which hit a gold nucleus, having 79 times the charge of a proton, head on, were turned backwards. What was the closest separation of the α particles and the gold nuclei? What was the potential energy of the gold nucleus and the α particle at that point.

PROBLEM 2.12 $ABCD$ is a rectangle with sides AB and CD are parallel and of length L where parallel sides AC and BD are of length $2L$. Charges $q, -q,$ q and $-q$ are placed respectively at A, B, C and D. How much work will be done in interchanging the charges at A and B.

2.6.2 Starred Section Electrostatic Binding Energy of a NaCl Molecule in a Crystal

Starting from the three point charges, one can easily track down the lines of force as shown in Fig. 1.10.

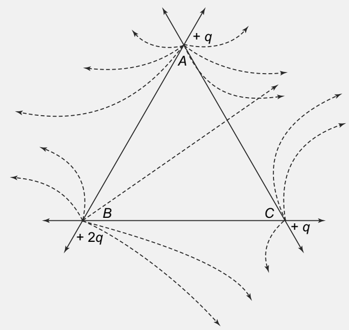


Fig. 1.10 *Example 1.6*

Illustrations *abound the text giving a more clearer understanding of concepts explained in theory.*

1.7 ADVANCED TOPICS

1.7.1 Electrostatics in Two-Dimensions

In 1884, an English schoolteacher, Edwin Abbot Abbot published a satirical novel called '*Flatland*' where he invented a two dimensional flat world. Although the novel was a satire on Victorian England, it was a pioneering effort in actually conceiving of worlds in dimensions different from the ordinary three dimensions of space that we are familiar with.

In 1979, Canadian computer scientist A.K. Dewdney published a monograph called 'Two dimensional Science & Technology' in which he considered how the various laws of physics would be in a two dimensional universe. Let us try and see how electrostatics as we know, will look like, if the universe was flat, i.e., instead of three dimensions, we only had 2.

To understand how fundamentally different things are in this fictional universe, we need to first define certain mathematical objects, which, as we shall see, behave very differently to their three dimensional counterparts.

Advanced Topics *cover concepts and principles which are not part of regular curricula but will equip students with the desired edge and thereby generate interest.*

Summary *at the end sums up all the topics discussed in the respective chapter.*

SUMMARY

- A type of material called 'conductors' have a special property that they have electrons in it that are almost 'free' and respond to electric fields by moving.
- The movement of such electrons inside a conductor make it special. There is no electric field inside it and the entire conductor is at the same potential when placed in an external steady electric field.
- The movement of free electrons results in the steady state with free charges being present only at the surface of a conductor.
- If there is a cavity inside a conductor without any charge being placed there from outside, there is no electric field there. The potential there is the same as the bulk conductor which encloses it. All free charges lie on the outer surface of such a conductor.
- When a single conductor is there carrying a charge, its potential is proportional to its charge. When several charge carrying conductors are present, their potentials are linearly related to the charges through the 'coefficients ' of potential. Conversely, the coefficients relating the charges to the potentials form a 'capacitance matrix'.

Conceptual Questions *follow after the summary in every chapter to allow students to evaluate their level of understanding of concepts.*

This is just a double Fourier series and therefore,

$$\int_0^a \int_0^b V(x, y) \sin \frac{m\pi x}{a} \sin \frac{n\pi y}{b} dxdy \qquad A_{m,n} \frac{ab}{4} \sinh(k_{m,n}c)$$

and therefore,

$$A_{m,n} \qquad \frac{4}{ab \sinh(k_{m,n}c)} \int_0^a \int_0^b V(x, y) \sin \frac{m\pi x}{a} \sin \frac{n\pi y}{b} dxdy$$

The Laplace equation, as we remarked in separable in eleven coordinate systems. In particular, the spherical and cylindrical coordinate systems are very useful since many of the problems in electrostatics have an inherent spherical or cylindrical symmetry. This we shall see in a later Chapter.

Problems *at the end of each chapter will help students in further practicing the principles explained in the respective chapter.*

PROBLEMS

Three charges of 2 nC, −5 nC and 0 2 nC are located at points $A(2, \frac{\pi}{2}, \frac{\pi}{4})$, $B(1, \pi, \frac{\pi}{2})$ and $C(5, \frac{\pi}{3}, \frac{2\pi}{3})$. Find the force on a charge the charge of 2 nC located at A.

A hydrogen molecule consists of two hydrogen atoms. An alpha particle is a helium nucleus with charge 2e. An α particle is shot through the exact centre of a hydrogen nucleus along a line perpendicular to the line, joining the two hydrogen nuclei (each with charge e) in the molecule as shown in the Fig. 1.23. The distance between the nuclei is b. Assume that the speed of the α particle is very high and so the nuclei do not move and also assume that the electric field due to the electron cloud surrounding each of the hydrogen atoms in the molecule is negligible. Where on its path does the α particle experience the maximum force?

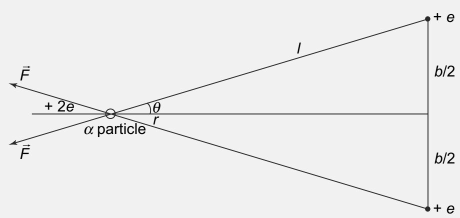


Fig. 1.24 *Problem 1.2*

Mathematical Preliminaries

Learning Objectives

- Review of results from vector analysis.
- Definition and properties of solid angle.
- Definition and properties of Dirac delta function.
- Review of complex numbers.
- Review of some results from matrix algebra.
- Review of ordinary differential equations.
- Definition and properties of some special functions.

1 VECTOR ALGEBRA

1a Scalars

Scalars are physical quantities described by their magnitude. Examples of scalars are mass, temperature, electric charge, etc.

A **scalar field** is a physical quantity that varies from point to point and at each point, its value is specified by its magnitude at that point. An example of the scalar field is the gravitational or electric potential. It varies from point to point in general and at every point, one only needs to know the magnitude of the potential to completely specify it.

1b Vectors

Vectors are physical quantities that are described by their magnitude and direction. They are usually represented with an arrow over them ($\vec{a}$) or in bold face (**a**). In three dimensions, vectors can be represented by a sum of projections on any three mutually orthogonal vectors of unit magnitude. These mutually orthogonal vectors of unit magnitude are called **basis vectors**. The most common basis in Cartesian coordinates are unit vectors along the x, y and z directions and these are represented by $\hat{i}$, $\hat{j}$ and $\hat{k}$ respectively. Thus, a typical vector $\vec{A}$ can be represented by

$$\vec{A} = A_x\hat{i} + A_y\hat{j} + A_z\hat{k} \tag{1}$$

where A_x, A_y, A_z are the x, y, z components of $\vec{A}$ and are really the projections of $\vec{A}$ on the three axis. This decomposition of a vector follows from the vector law of addition, which specifies how two vectors can be added. A simple geometric way to see the vector addition law is to think of the two vectors that need to be added as forming two sides of a triangle. Then the third side is the vector that is the sum of the two vectors. The direction of the sum is in the opposite sense of the two vectors as shown in Fig. 1.

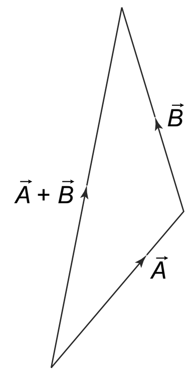


Fig. 1 *Vector addition*

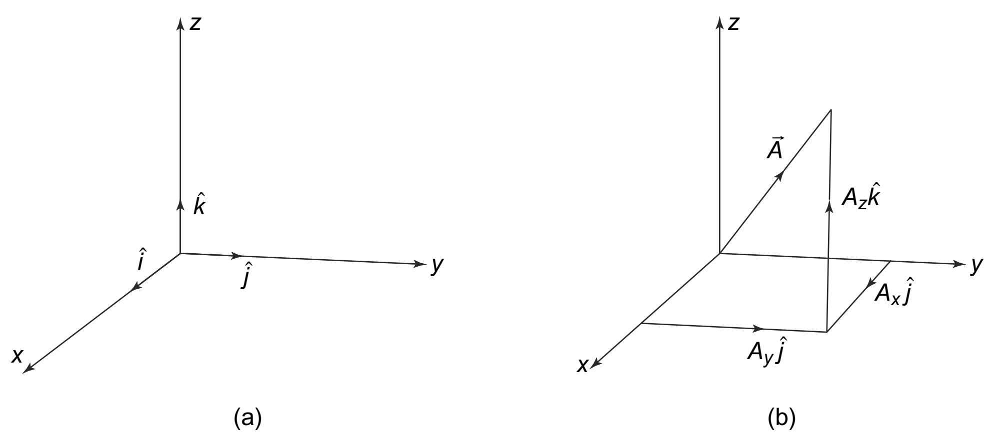


Fig. 2 *(a) Unit vectors $\hat{i}$, $\hat{j}$, $\hat{k}$ in the Cartesian system of coordinates, (b) The components of a vector $\vec{A}$ along $\hat{i}$, $\hat{j}$, $\hat{k}$*

A point P in Cartesian coordinates is represented by (x, y, z). We define a position vector as the directed distance from the origin O to P. Thus,

$$\vec{r}_P = x\hat{i} + y\hat{j} + z\hat{k} \tag{2}$$

The displacement vector is the displacement from one point to another. Hence,

$$\vec{r}_{PQ} = (x_Q - x_P)\hat{i} + (y_Q - y_P)\hat{j} + (z_Q - z_P)\hat{k} \tag{3}$$

The **unit vector along any vector** is simply given by

$$\hat{A} = \frac{\vec{A}}{|\vec{A}|} \tag{4}$$

where

$$|\vec{A}| = \sqrt{A_x^2 + A_y^2 + A_z^2}$$

is the magnitude of the vector.

In terms of the components, the sum or difference of two vectors is easily determined to be

$$\begin{aligned}\vec{A} &= A_x\hat{i} + A_y\hat{j} + A_z\hat{k}\\ \vec{B} &= B_x\hat{i} + B_y\hat{j} + B_z\hat{k}\\ \vec{A} \pm \vec{B} &= (A_x \pm B_x)\hat{i} + (A_y \pm B_y)\hat{j} + (A_z \pm B_z)\hat{k}\end{aligned}$$

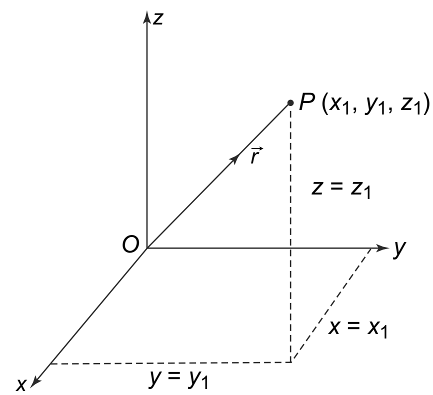


Fig. 3 *Position vector $\vec{r}$*

A **vector field** is a physical quantity which varies from point to point and at every point is completely specified by its magnitude and direction at that point. For example, the gravitational or electric field is an example of a vector field. To specify these fields, one not only needs to know the magnitude of the gravitational or electric field at that point but also the direction.

1c Dot or Scalar Product

We can define two kinds of products of vectors. The dot or scalar product is defined as

$$\vec{A} \cdot \vec{B} = \vec{B} \cdot \vec{A} = |\vec{A}||\vec{B}| \cos\theta \tag{5}$$

where θ is the angle between $\vec{A}$ and $\vec{B}$

This allows us to immediately see that

$$\begin{aligned}\hat{i} \cdot \hat{i} &= \hat{j} \cdot \hat{j} = \hat{k} \cdot \hat{k} = 1\\ \hat{i} \cdot \hat{j} &= \hat{i} \cdot \hat{k} = \hat{j} \cdot \hat{k} = 0\end{aligned}$$

and hence,

$$\boxed{\vec{A} \cdot \vec{B} = A_x B_x + A_y B_y + A_z B_z} \tag{6}$$

We can also easily see that associativity and distributivity are valid. Thus,

$$\vec{A} \cdot (\vec{B} + \vec{C}) = \vec{A} \cdot \vec{B} + \vec{A} \cdot \vec{C} \tag{7}$$

and for a scalar k,

$$(k\vec{A}) \cdot \vec{B} = k(\vec{A} \cdot \vec{B}) = \vec{A} \cdot (k\vec{B}) \tag{8}$$

1d Cross or Vector Product

For two vectors, $\vec{A}$ and $\vec{B}$, we can also define a vector product as

$$\vec{A} \times \vec{B} = |\vec{A}||\vec{B}| \sin\theta \hat{e} \tag{9}$$

where θ is the angle between $\vec{A}$ and $\vec{B}$ and $\hat{e}$ is a unit vector in the direction perpendicular to both $\vec{A}$ and $\vec{B}$ in the direction that is given by the right-hand rule.
One can then see, using this definition that

$$\hat{i} \times \hat{i} = \hat{j} \times \hat{j} = \hat{k} \times \hat{k} = 0$$

and

$$\hat{i} \times \hat{j} = \hat{k}, \hat{i} \times \hat{k} = -\hat{j}, \hat{j} \times \hat{k} = +i$$

With these, one can find an expression for $\vec{A} \times \vec{B}$ as

$$\vec{A} \times \vec{B} = \begin{vmatrix} \hat{i} & \hat{j} & \hat{k} \\ A_x & A_y & A_z \\ B_x & B_y & B_z \end{vmatrix} \tag{10}$$

Clearly

$$\vec{A} \times \vec{B} = -\vec{B} \times \vec{A}$$

For three vectors, one can define two kinds of products. One, $\vec{A} \cdot (\vec{B} \times \vec{C})$ is called the **triple scalar product**. One can easily see that

$$\vec{A} \cdot \vec{B} \times \vec{C} = \vec{B} \cdot \vec{C} \times \vec{A} = \vec{C} \cdot \vec{A} \times \vec{B} = -\vec{A} \cdot \vec{C} \times \vec{B}$$

Also, writing the vectors in terms of their Cartesian components and using the definitions of dot and cross product given above, one can see that

$$\vec{A} \cdot \vec{B} \times \vec{C} = \begin{vmatrix} A_x & A_y & A_z \\ B_x & B_y & B_z \\ C_x & C_y & C_z \end{vmatrix} \tag{11}$$

It can also be shown that the triple scalar product has a geometrical representation. It is simply the volume of the parallelepiped defined by the vectors $\vec{A}$, $\vec{B}$, $\vec{C}$.

One can also see that

$$\boxed{\vec{A} \times (\vec{B} \times \vec{C}) = \vec{B}(\vec{A} \cdot \vec{C}) - \vec{C}(\vec{A} \cdot \vec{B})} \tag{12}$$

an identity which is very useful.

2 COORDINATE SYSTEMS

2a Cartesian Coordinates

We are familiar with the Cartesian coordinate system where a point is defined by (x, y, z). Recall that any vector can be written as a sum of its components along the three axes. Thus,

$$\vec{A} = A_x\hat{i} + A_y\hat{j} + A_z\hat{k}$$

where $\hat{i}$, $\hat{j}$, $\hat{k}$ are the unit vectors along x, y, z respectively.

The range of coordinates is

$$
\begin{aligned}
-\infty < x < \infty \\
-\infty < y < \infty \\
-\infty < z < \infty
\end{aligned}
$$

The elements of length, area and volume are easily defined in Cartesian coordinates as

$$
\boxed{
\begin{aligned}
\vec{dl} &= dx\hat{i} + dy\hat{j} + dz\hat{k} \\
\vec{dS} &= dydz\hat{i} + dxdz\hat{j} + dxdy\hat{k} \\
dV &= dxdydz
\end{aligned}
}
\tag{13}
$$

These are also displayed in Fig. 4.

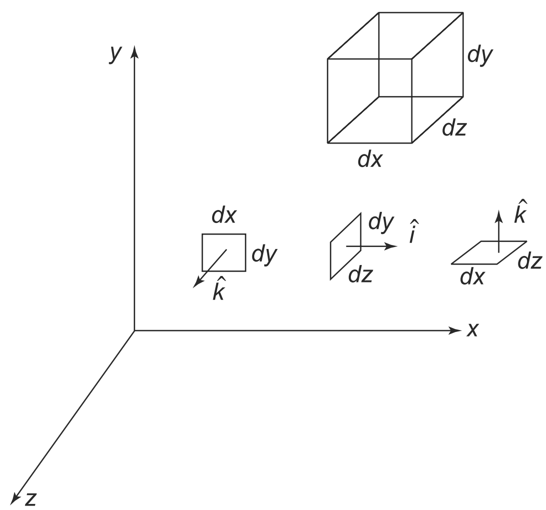


Fig. 4 *Differential element of length, area and volume in Cartesian coordinates*

2b Circular Cylindrical Coordinates

The circular cylindrical coordinate system (ρ, ϕ, z) is very convenient when dealing with problems with cylindrical symmetry. The range of the coordinates is given by

$$
\begin{aligned}
0 \leq \rho < \infty \\
0 \leq \phi < 2\pi \\
-\infty < z < \infty.
\end{aligned}
$$

Just like the Cartesian system, a vector $\vec{A}$ in this coordinate system can be represented as

$$\vec{A} = A_\rho \hat{\rho} + A_\phi \hat{\phi} + A_z \hat{z} \tag{14}$$

The magnitude of $\vec{A}$ is given by

$$|\vec{A}| = \sqrt{A_\rho^2 + A_\phi^2 + A_z^2}$$

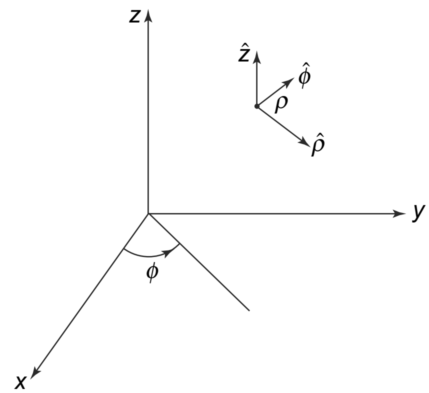


Fig. 5 *Unit vectors in the circular cylindrical system of coordinates*

This is shown in Fig. 5.

The unit vectors that we have are also orthogonal. Thus,

$$\hat{\rho} \cdot \hat{\rho} = \hat{\phi} \cdot \hat{\phi} = \hat{z} \cdot \hat{z} = 1$$

and

$$\hat{\rho} \cdot \hat{\phi} = \hat{\rho} \cdot \hat{z} = \hat{\phi} \cdot \hat{z} = 0$$

and

$$\begin{aligned} \hat{\rho} \times \hat{\phi} &= \hat{z} \\ \hat{\phi} \times \hat{z} &= \hat{\rho} \\ \hat{z} \times \hat{\rho} &= \hat{\phi} \end{aligned}$$

(ρ, ϕ, z) are related to (x, y, z) as

$$\begin{aligned} \rho &= \sqrt{x^2 + y^2} \\ \tan\phi &= \frac{y}{x} \\ z &= z \end{aligned}$$

One can easily obtain the relationship between $\hat{\rho}, \hat{\phi}, \hat{z}$ and $\hat{i}, \hat{j}, \hat{k}$ either algebraically or geometrically, as shown in Fig. 6, as

$$\begin{aligned} \hat{\rho} &= \cos\phi \;\; \hat{i} + \sin\phi \;\; \hat{j} \\ \hat{\phi} &= -\sin\phi \;\; \hat{i} + \cos\phi \;\; \hat{j} \\ \hat{z} &= \hat{k} \end{aligned} \tag{15}$$

and the inverse relations

$$\begin{aligned} \hat{i} &= \cos\phi \;\; \hat{\rho} - \sin\phi \;\; \hat{\phi} \\ \hat{j} &= \sin\phi \;\; \hat{\rho} + \cos\phi \;\; \hat{\phi} \\ \hat{k} &= \hat{z} \end{aligned} \tag{16}$$

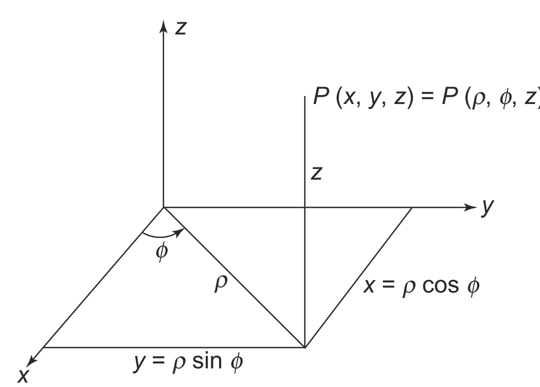


Fig. 6 *Relationship between (x, y, z) and (ρ, ϕ, z)*

With these results, one can easily obtain the transformation between a vector $\vec{A}$ in the two coordinate systems.

$$\begin{pmatrix} A_\rho \\ A_\phi \\ A_z \end{pmatrix} = \begin{pmatrix} \cos\phi & \sin\phi & 0 \\ -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} A_x \\ A_y \\ A_z \end{pmatrix} \tag{17}$$

The inverse transformations can also be obtained as

$$\begin{pmatrix} A_x \\ A_x \\ A_z \end{pmatrix} = \begin{pmatrix} \cos\phi & -\sin\phi & 0 \\ \sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} A_\rho \\ A_\phi \\ A_z \end{pmatrix} \tag{18}$$

The elements of length, area and volume are easily defined in cylindrical coordinates as

$$\boxed{\begin{aligned} \vec{dl} &= d\rho \;\; \hat{\rho} + \rho d\phi \;\; \hat{\phi} + dz \;\; \hat{k} \\ \vec{dS} &= \rho d\phi dz \;\; \hat{\rho} + d\rho dz \;\; \hat{\phi} + \rho d\phi d\rho \;\; \hat{k} \\ dV &= \rho d\rho d\phi dz \end{aligned}} \tag{19}$$

This is shown in Fig. 7.

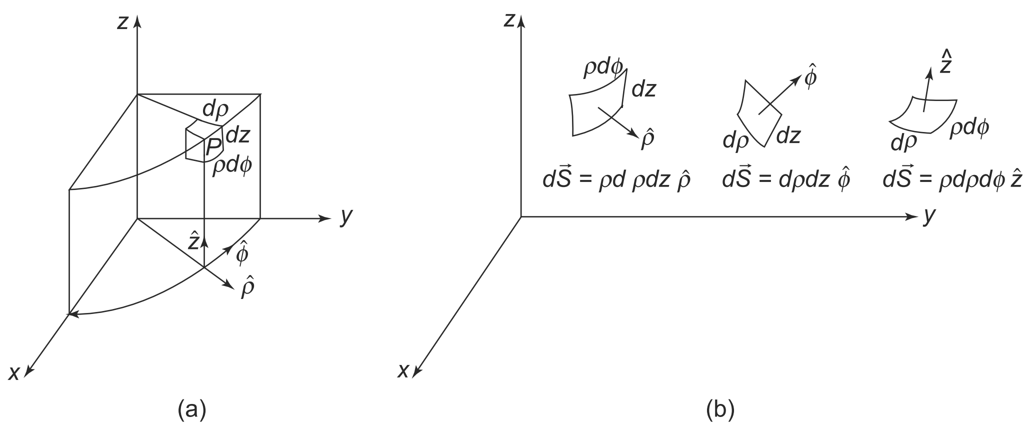


Fig. 7 *Differential element of length, area and volume in cylindrical coordinates*

2c Spherical Coordinates

The spherical polar coordinate system (r, θ, ϕ) is useful in problems with spherical symmetry. The ranges of the coordinates are given by

$$\begin{aligned} 0 \le r &< \infty \\ 0 \le \theta &\le \pi \\ 0 \le \phi &< 2\pi \end{aligned}$$

This is shown in Fig. 8.

Just like the Cartesian system, a vector $\vec{A}$ in this coordinate system can be represented as

$$\vec{A} = A_r\hat{r} + A_\theta\hat{\theta} + A_\phi\hat{\phi} \tag{20}$$

The magnitude of $\vec{A}$ is given by

$$|\vec{A}| = \sqrt{A_r^2 + A_\theta^2 + A_\phi^2}$$

The unit vectors that we have are also orthogonal. Thus,

$$\hat{r}\cdot\hat{r} = \hat{\theta}\cdot\hat{\theta} = \hat{\phi}\cdot\hat{\phi} = 1$$

and

$$\hat{r}\cdot\hat{\theta} = \hat{r}\cdot\hat{\phi} = \hat{\phi}\cdot\hat{\theta} = 0$$

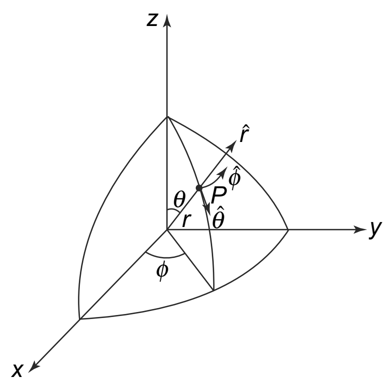


Fig. 8 *Unit vectors in the spherical coordinates*

and

$$\hat{r}\times\hat{\theta} = \hat{\phi}$$
$$\hat{\theta}\times\hat{\phi} = \hat{r}$$
$$\hat{\phi}\times\hat{r} = \hat{\theta}$$

(r, θ, ϕ) are related to (x, y, z) as

$$r = \sqrt{x^2 + y^2 + z^2}$$
$$\tan\theta = \frac{\sqrt{x^2 + y^2}}{z}$$
$$\tan\phi = \frac{y}{x} \tag{21}$$

One can easily obtain the relationship between $\hat{r}, \hat{\theta}, \hat{\phi}$ and $\hat{i}, \hat{j}, \hat{k}$ either algebraically or geometrically as

$$\hat{r} = \sin\theta\cos\phi\ \ \hat{i} + \sin\theta\sin\phi\ \ \hat{j} + \cos\theta\ \ \hat{k}$$
$$\hat{\theta} = \cos\theta\cos\phi\ \ \hat{i} + \cos\theta\sin\phi\ \ \hat{j} - \sin\theta\ \ \hat{k}$$
$$\hat{\phi} = -\sin\theta\ \ \hat{i} + \cos\phi\ \ \hat{j} \tag{22}$$

and the inverse relations

$$\hat{i} = \sin\theta\cos\phi\ \ \hat{r} + \cos\theta\cos\phi\ \ \hat{\theta} - \sin\phi\ \ \hat{\phi}$$
$$\hat{j} = \sin\theta\sin\phi\ \ \hat{r} + \cos\theta\sin\phi\ \ \hat{\theta} + \cos\phi\ \ \hat{\phi}$$
$$\hat{k} = \cos\theta\ \ \hat{r} - \sin\theta\ \ \hat{\theta} \tag{23}$$

This is depicted in Fig. 9.

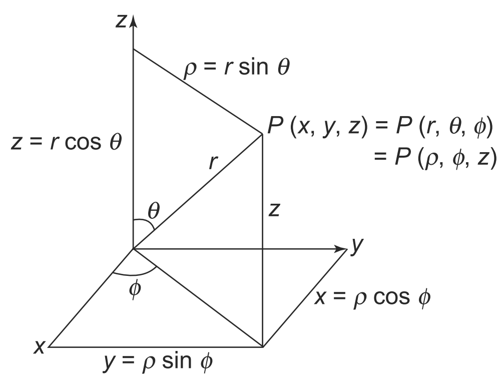


Fig. 9 *Relationship between* (x, y, z)*,* (ρ, ϕ, z) *and* (r, θ, ϕ)*)*

With these results, one can easily obtain the transformation between a vector $\vec{A}$ in the two coordinate systems.

$$\begin{pmatrix} A_r \\ A_\theta \\ A_\phi \end{pmatrix} = \begin{pmatrix} \sin\theta\cos\phi & \sin\theta\sin\phi & \cos\theta \\ -\cos\theta\cos\phi & \cos\theta\sin\phi & -\sin\theta \\ -\sin\phi & \cos\phi & 0 \end{pmatrix} \begin{pmatrix} A_x \\ A_y \\ A_z \end{pmatrix} \tag{24}$$

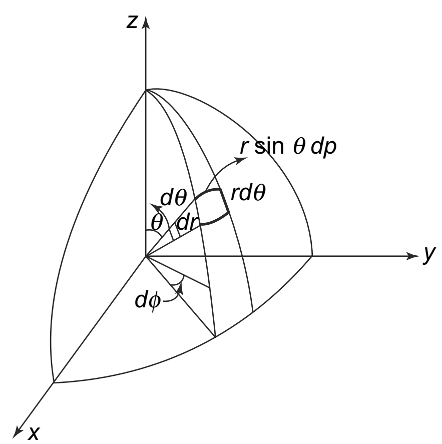
The inverse transformations can also be obtained as

$$\begin{pmatrix} A_x \\ A_x \\ A_z \end{pmatrix} = \begin{pmatrix} \sin\theta\cos\phi & \cos\theta\cos\phi & -\sin\phi \\ \sin\theta\sin\phi & \cos\theta\sin\phi & \cos\phi \\ \cos\theta & -\sin\theta & 0 \end{pmatrix} \begin{pmatrix} A_r \\ A_\theta \\ A_\phi \end{pmatrix} \tag{25}$$

The elements of length, area and volume are easily defined in spherical coordinates as

$$\boxed{\begin{aligned} \vec{dl} &= dr\hat{r} + rd\theta\hat{\theta} + r\sin\theta d\phi\hat{\phi} \\ \vec{dS} &= r^2\sin\theta d\theta d\phi\hat{r} + r\sin\theta dr d\phi\hat{\theta} + r dr d\theta\hat{\phi} \\ dV &= r^2\sin\theta dr d\theta d\phi \end{aligned}} \tag{26}$$

This is shown in Fig. 10.



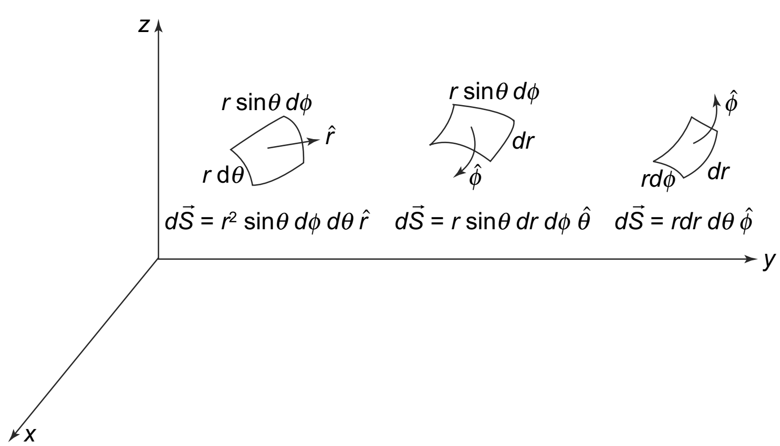


Fig. 10 *Differential element of length, area and volume in spherical coordinates*

3 VECTOR ANALYSIS

3a Gradient

Consider a scalar field $V(x, y, z)$. It is possible to construct a vector out of this field by using a vector operator called the gradient operator, which is defined in **Cartesian coordinates** as

$$\vec{\nabla} = \hat{i}\frac{\partial}{\partial x} + \hat{j}\frac{\partial}{\partial y} + \hat{k}\frac{\partial}{\partial z} \tag{27}$$

and its operation on a scalar field $V(x, y, z)$ gives us the gradient of the scalar field.

$$\vec{\nabla}\phi = \hat{i}\frac{\partial V}{\partial x} + \hat{j}\frac{\partial V}{\partial y} + \hat{k}\frac{\partial V}{\partial z}$$

The operator $\vec{\nabla}$ is a vector differential operator that can operate on the scalar field.

In **cylindrical coordinates**, we have

$$\vec{\nabla} V = \frac{\partial V}{\partial \rho}\hat{\rho} + \frac{1}{\rho}\frac{\partial V}{\partial \phi}\hat{\phi} + \frac{\partial V}{\partial z}\hat{k} \tag{28}$$

and in **spherical coordinates**

$$\vec{\nabla} V = \frac{\partial V}{\partial r}\hat{r} + \frac{1}{r}\frac{\partial V}{\partial \theta}\hat{\theta} + \frac{1}{r\sin\theta}\frac{\partial V}{\partial \phi}\hat{\phi} \tag{29}$$

3b Divergence

The gradient operator $\vec{\nabla}$ is a vector operator and we can use it to operate on a vector field. In particular, we can dot it into a vector field $\vec{A}(x, y, z)$ to get, in **Cartesian coordinates**

$$\vec{\nabla}\cdot\vec{A} = \frac{\partial A_x}{\partial x} + \frac{\partial A_y}{\partial y} + \frac{\partial A_z}{\partial z} \tag{30}$$

This quantity is called the divergence of the vector field. It is, of course, a scalar since it is made from a scalar or dot product of two vectors.

In **cylindrical coordinates**,

$$\vec{\nabla}\cdot\vec{A} = \frac{1}{\rho}\frac{\partial(\rho A_\rho)}{\partial \rho} + \frac{1}{\rho}\frac{\partial A_\phi}{\partial \phi} + \frac{\partial A_z}{\partial z} \tag{31}$$

and in **spherical coordinates**

$$\vec{\nabla}\cdot\vec{A} = \frac{1}{r^2}\frac{\partial(r^2 A_r)}{\partial r} + \frac{1}{r\sin\theta}\frac{\partial(A_\theta\sin\theta)}{\partial\theta} + \frac{1}{r\sin\theta}\frac{\partial A_\phi}{\partial\phi} \tag{32}$$

With the gradient and divergence defined, we can show that

$$\vec{\nabla}\cdot\vec{\nabla}\phi = \nabla^2\phi = \frac{\partial^2\phi}{\partial x^2} + \frac{\partial^2\phi}{\partial y^2} + \frac{\partial^2\phi}{\partial z^2} \tag{33}$$

and

$$\vec{\nabla}\cdot(\phi\vec{A}) = (\vec{\nabla}\phi)\cdot\vec{A} + \phi\vec{\nabla}\cdot\vec{A} \tag{34}$$

It is also easy to see that

$$\vec{\nabla}\cdot\vec{r} = 3 \tag{35}$$

where $\vec{r} = x\hat{i} + y\hat{j} + z\hat{k}$.

We can also generalise this to a function of r to get

$$\vec{\nabla} \cdot \vec{r} r^{n-1} = (n+2) r^{n-1} \tag{36}$$

3c Curl

The gradient operator can also operate on a vector field by taking its vector product with the vector field $\vec{A}(x, y, z)$ to give us the curl of the vector field. In **Cartesian coordinates**

$$\vec{\nabla} \times \vec{A} = \begin{vmatrix} \hat{i} & \hat{j} & \hat{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ A_x & A_y & A_z \end{vmatrix} \tag{37}$$

In **cylindrical coordinates**

$$\vec{\nabla} \times \vec{A} = \frac{1}{\rho} \begin{vmatrix} \hat{\rho} & \rho\hat{\phi} & \hat{z} \\ \frac{\partial}{\partial \rho} & \frac{\partial}{\partial \phi} & \frac{\partial}{\partial z} \\ A_\rho & \rho A_\phi & A_z \end{vmatrix} \tag{38}$$

and in **spherical coordinates** as

$$\vec{\nabla} \times \vec{A} = \frac{1}{r^2 \sin\theta} \begin{vmatrix} \hat{r} & r\hat{\theta} & r\sin\theta\hat{\phi} \\ \frac{\partial}{\partial r} & \frac{\partial}{\partial \theta} & \frac{\partial}{\partial \phi} \\ A_r & rA_\theta & r\sin\theta A_\phi \end{vmatrix} \tag{39}$$

It is easy to see that

$$\vec{\nabla} \times \vec{r} = 0$$

Some identities immediately follow from the definitions of gradient, divergence and curl.

$$\vec{\nabla} \times (\phi\vec{A}) = \phi\vec{\nabla} \times \vec{A} + (\vec{\nabla}\phi) \times \vec{A} \tag{40}$$

which allows us to see that

$$\vec{\nabla} \times \vec{r} f(r) = 0$$

One can apply the gradient operator successively to get

$$\vec{\nabla} \times \vec{\nabla}\phi$$

which is easily seen to be zero.

$$\vec{\nabla} \cdot \vec{\nabla} \times \vec{A} = \begin{vmatrix} \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ A_x & A_y & A_z \end{vmatrix} \tag{41}$$

Also

$$\vec{\nabla} \cdot \vec{\nabla} \times \vec{A} = 0$$

3d Some Other Vector Identities

For vector fields $\vec{A}(\vec{r})$ and $\vec{B}(\vec{r})$, and scalar fields $\phi(\vec{r})$ and $\psi(\vec{r})$

$$\vec{\nabla}(\phi\psi) = \phi\vec{\nabla}\psi + \psi\vec{\nabla}\phi \tag{42}$$

$$\vec{\nabla} \cdot (\vec{A} \times \vec{B}) = \vec{B} \cdot (\vec{\nabla} \times \vec{A}) - \vec{A} \cdot (\vec{\nabla} \times \vec{B}) \tag{43}$$

$$\vec{\nabla} \times (\vec{A} \times \vec{B}) = (\vec{B} \cdot \vec{\nabla})\vec{A} - (\vec{A} \cdot \vec{\nabla})\vec{B} + \vec{A}(\vec{\nabla} \cdot \vec{B}) - \vec{B}(\vec{\nabla} \cdot \vec{A}) \tag{44}$$

$$\vec{\nabla} \times (\vec{\nabla} \times \vec{A}) = \vec{\nabla}(\vec{\nabla} \cdot \vec{A}) - \vec{\nabla}^2 \vec{A} \tag{45}$$

$$\vec{\nabla}(\vec{A} \cdot \vec{B}) = \vec{A} \times (\vec{\nabla} \times \vec{B}) + \vec{B} \times (\vec{\nabla} \times \vec{A}) + (\vec{A} \cdot \vec{\nabla})\vec{B} + (\vec{B} \cdot \vec{\nabla})\vec{A} \tag{46}$$

3e Stokes' Theorem

The line integral of any vector field along the boundary C of a surface S is given by

$$\oint_C \vec{A} \cdot \vec{dl} = \iint_S (\vec{\nabla} \times \vec{A}) \cdot \vec{dS} \tag{47}$$

where the sense of the line integral is related to the outward normal of the surface area by the right-hand rule. This is because there is an inherent ambiguity in defining the direction of the contour C as is there in defining the direction of the outward normal for an open surface. For a closed surface, the outward normal is well defined.

3f Gauss's Divergence Theorem

The volume integral of the divergence of any vector field over a closed volume V is the surface integral of the vector field over any closed surface S enclosing the volume.

$$\iiint_V (\vec{\nabla} \cdot \vec{A}) dV = \iint_S \vec{A} \cdot \vec{dS} \tag{48}$$

3g Green's Theorem

For two scalar fields ϕ and ψ,

$$\iiint_V \left(\phi \nabla^2 \psi - \psi \nabla^2 \phi\right) dV = \iint_S \left(\phi \vec{\nabla} \psi - \psi \vec{\nabla} \phi\right) \cdot \vec{dS} \tag{49}$$

3h Transformation of Vectors and Tensors

We have defined vectors as quantities that have direction and magnitude. This definition, though sufficient for most part, becomes problematic in several cases. For instance, quantities like the index of refraction in certain crystals which are anisotropic, have a magnitude and are dependent on the direction of propagation of light, but these are not vectors. It is important therefore, to have a more general definition of vectors, an approach that also allows naturally to define higher order objects called tensors.

The definition of vectors is given in terms of their transformation properties under a rotation of the coordinate system. Consider a vector $\vec{r}$ in two dimensions, which has components x and y in some coordinate system. Now imagine that the coordinate system is rotated counterclockwise by an angle θ while keeping $\vec{r}$ fixed. Then in the rotated coordinate system, the vector $\vec{r}$ has components x', y' and these are related to the unrotated components by

$$\begin{aligned} x' &= \cos\theta x + \sin\theta y \\ y' &= -\sin\theta x + \cos\theta y \end{aligned}$$

This can be conveniently represented in a new notation where instead of (x, y) we use (x_1, x_2). This new notation allows the transformation equations above to be conveniently written as

$$x_i' = \sum_{j=1}^{2} a_{ij} x_j \tag{50}$$

where i takes the value 1, 2 and a_{ij} are the coefficients above. Thus, $a_{11} = \cos\theta$, $a_{12} = \sin\theta$, etc. This single relation is actually two relations for the two values of i. The index j over which the sum is taken is often referred to as the dummy index. Of course, this definition can also be generalised to any number of dimensions, though for our purposes three dimensions would suffice. Thus, any vector V, with N components, under a rotation transforms as

$$V_i' = \sum_{j=1}^{N} a_{ij} V_j \tag{51}$$

and $i = 1, 2, \cdots, N$. Frequently, the summation signs are dropped in a convention known as the *summation convention* where it is assumed that all indices which are repeated once, are to be summed over. Thus in this convention, we can write the equation above as

$$V_i' = a_{ij} V_j$$

where since j is repeated, it is understood that it is to be summed over. These transformation properties of the vector are usually taken as the definition of a vector. Thus, any set of N quantities $V_1, V_2, \cdots, V_N$ are the components of a vector in N dimension if they transform as above. The quantities a_{ij} may be written as

$$a_{ij} = \frac{\partial x_i'}{\partial x_j} \tag{52}$$

and therefore the transformation of V becomes

$$V_i' = \sum_{j=1}^{N} \frac{\partial x_i'}{\partial x_j} V_j \tag{53}$$

We can generalise this definition to objects of higher rank called **tensors**. A tensor of rank two is a two index object (just like a tensor of rank zero has no indices; that is a scalar, or a vector is a tensor of rank one, with one index) whose components transform as follows:

$$A'^{ij} = \sum_{kl} \frac{\partial x_i'}{\partial x_k} \frac{\partial x_j'}{\partial x_l} A^{kl} \tag{54}$$

In three dimensions, it is obvious that a second rank tensor can be represented by a 3×3 array whose components are

$$\mathbf{A} = \begin{pmatrix} A^{11} & A^{12} & A^{13} \\ A^{21} & A^{22} & A^{23} \\ A^{31} & A^{32} & A^{33} \end{pmatrix} \tag{55}$$

In defining the tensors, we must mention that there are actually three kinds of tensors—contravariant

A^{ij}, covariant A_{ij} and mixed A_i^j. These are, once again, defined by their transformation properties as

$$A'^{ij} = \sum_{kl} \frac{\partial x_i'}{\partial x_k} \frac{\partial x_j'}{\partial x_l} A^{kl} \tag{56}$$

$$A_j'^i = \sum_{kl} \frac{\partial x_i'}{\partial x_k} \frac{\partial x_l}{\partial x_j'} A_l^k \tag{57}$$

$$A_{ij}' = \sum_{kl} \frac{\partial x_k}{\partial x_i'} \frac{\partial x_l}{\partial x_j'} A_{kl} \tag{58}$$

Finally, a tensor for which

$$A^{ij} = A^{ji}$$

is called a *symmetric tensor* and one for which

$$A^{ij} = -A^{ji}$$

is called an *antisymmetric tensor*. Clearly, any tensor can be written as a sum of symmetric and antisymmetric parts

$$A^{ij} = \frac{1}{2}(A^{ij} + A^{ji}) + \frac{1}{2}(A^{ij} - A^{ji}) \tag{59}$$

It is important to realise that vectors by themselves are independent of the coordinate system and a vector equation is not dependent on a particular coordinate system. The coordinate system only gives us the components of the vector. **In particular, if we establish a vector equation in a coordinate system, say the Cartesian coordinates, the equation is valid in any coordinate system.**

The transformation rules for vectors and tensors do not depend on the particular vector or tensor whose transformation one is studying. That means that if we have any equality between two vectors, say $\vec{V}$ and $\vec{W}$, in one coordinate system

$$V_i = W_i$$

It is automatic that the equality holds in any transformed coordinate system:

$$V_i' = W_i'$$

The laws of physics expressed in terms of vectors are thus, independent of coordinate system.

4 SOLID ANGLE

In three-dimensions, we can define a quantity Ω called the solid angle, which can be thought of as a two-dimensional angle subtended by a three-dimensional surface at a point. It is the exact analogue of an arc subtending an angle at a point in two dimensions. A solid angle subtended at a point by a surface is a measure of how large the surface appears to an observer at the point. Its value is equal to the area of the patch on the surface of a unit sphere that the surface restricts. By this definition, the solid angle subtended by a sphere at its centre is 4π. The unit of solid angle is steradian and it is dimensionless.

In spherical polar coordinates, we can write the solid angle as

$$d\Omega = \sin\theta d\theta d\phi$$

Thus the solid angle subtended by an arbitrary surface S at a point O is the solid angle subtended by the projection of the surface on a unit sphere with O as the centre. Thus,

$$\Omega = \iint_S \frac{\vec{r}\cdot\hat{n}}{r^3} dS = \iint_S \sin\theta d\theta d\phi \tag{60}$$

where $\vec{r}$ is the position vector of a the infinitesimal area dS on the surface S, with O as the origin, and $\hat{n}$ is the outward normal to dS.

5 DIRAC DELTA FUNCTION

In one dimension, the Dirac Delta function $\delta(x)$ is defined as

$$\begin{aligned} \delta(x) &= 0 \qquad \text{if } x \neq 0 \\ \delta(x) &= \infty \qquad \text{if } x = 0 \\ \int_{-\infty}^{\infty} \delta(x) dx &= 1 \end{aligned} \tag{61}$$

It can be thought of as an infinitely peaked, infinitesimally narrow spike at $x = 0$. One can think of it as the limit of a peaked curve, say a Gaussian, which becomes narrower and narrower and higher and higher such that the area under the curve is a constant. It is strictly not a mathematical function but can be rigorously defined in terms of distributions.

The most important property of the Delta function is its use under the integral sign

$$\int_{-\infty}^{\infty} dx \delta(x) f(x) = f(0) \tag{62}$$

It is important to note that the limits of the integration do NOT have to be from $-\infty$ to ∞. It is sufficient for the limits to just enclose the domain of the Delta function.

Of course, the above can be trivially extended to be defined at any point. Thus, for instance

$$\int_{-\infty}^{\infty} \delta(x-a) f(x) = f(a) \tag{63}$$

Some of the other important properties of the Delta function can be obtained by remembering that it is used under the integral sign. Thus,

$$\delta(f(x)) = \sum_i \frac{1}{|\frac{df(x_i)}{dx}|} \delta(x - x_i) \tag{64}$$

where x_i are the simple zeroes of the function $f(x)$. Clearly, then

$$\delta(kx) = \frac{1}{|k|}\delta(x) \tag{65}$$

One can, by using integration by parts, also show that

$$x\frac{d\delta(x)}{dx} = x\delta'(x) = -\delta(x) \tag{66}$$

In more than one dimension, the Delta function is defined by merely taking the product of Delta functions in each dimension. Thus, in Cartesian coordinates

$$\delta^3(\vec{r} - \vec{R}) = \delta(x - X)\delta(y - Y)\delta(z - Z) \tag{67}$$

which vanishes everywhere, except at $\vec{r} = \vec{R}$. Here $\vec{R} = X\hat{\imath} + Y\hat{\jmath} + Z\hat{k}$.

From this, it is obvious that

$$\int f(\vec{r})\delta^3(\vec{r} - \vec{a})dV = f(\vec{a})$$

Delta functions are very useful in physics. For instance, consider the following 'paradox':

The function

$$\vec{a} = \frac{1}{r^2}\hat{r}$$

is well defined. Calculating its divergence gives us 0. Now apply the divergence theorem to this function, using a sphere centred at the origin and of radius R as the surface. Then we get

$$\iint_S \vec{a} \cdot \vec{dS} = \iint (\frac{1}{R^2})R^2 \sin\theta d\theta d\phi = 4\pi$$

The surface integral over a closed surface of $\vec{a}$ should be equal to, by the divergence theorem above, the volume integral of the divergence of $\vec{a}$ taken over the volume V enclosed by the surface S. But the volume integral

$$\iiint (\vec{\nabla} \cdot \vec{a})dV = 0$$

since the divergence is zero! What is the solution to this paradox?

The problem is really with the definition of the divergence of $\vec{a}$ at the origin. It is zero everywhere except at the origin, where it is not defined. Also, the integral of the function, over a volume which encloses the origin is a constant, 4π. Therefore, it is more accurate to write

$$\vec{\nabla} \cdot \left(\frac{\hat{r}}{r^2}\right) = 4\pi\delta^3(\vec{r})$$

This will then have the correct properties and there would be no paradox. Also, note that

$$\vec{\nabla}\left(\frac{1}{r}\right) = -\frac{\hat{r}}{r^2}$$

and thus,

$$\nabla^2 \frac{1}{r} = -4\pi\delta^3(\vec{r}) \tag{68}$$

6 COMPLEX NUMBERS

We are aware of real numbers. It is possible to define another kind of quantity called a complex number, which is of the form

$$z = x + iy \tag{69}$$

where x and y are real numbers and i is defined as

$$i = \sqrt{-1} \quad i^2 = -1$$

Thus, a complex number has two parts—a real part x and an 'imaginary' part y.

We can define the ordinary arithmetic operations with complex numbers once we define two rules:

1. A complex number $z = x + iy$ is zero if and only if both $x = 0$ and $y = 0$.
2. Complex numbers follow all rules of ordinary real algebra with the identification of $i^2 = -1$.

With these two rules, it is easy to see that we can define addition, multiplication, etc., of complex numbers. Let $z_1 = x_1 + iy_1$ and $z_2 = x_2 + iy_2$ be two complex numbers. Then

$$z_1 \pm z_2 = (x_1 \pm x_2) \pm i(y_1 \pm y_2) \tag{70}$$

$$z_1 z_2 = (x_1 x_2 - y_1 y_2) + i(x_1 y_2 + x_2 y_1) \tag{71}$$

We can also define the complex conjugate z^* of a complex number z as

$$z = x + iy$$

then

$$z^* = x - iy$$

Multiplication of a complex number with its complex conjugate gives us

$$zz^* = (x + iy)(x - iy) = x^2 + y^2$$

Frequently it is convenient to represent complex numbers geometrically. Consider a point $P(x, y)$ in the (xy) plane, as shown in Fig. 11.

We can then assign to P a complex number

$$z = x + iy$$

The x axis in the figure is called the real axis and the y axis the imaginary axis. This plane is called the **complex plane** or **Argand plane.** This representation also allows us to represent complex numbers in another way. From the figure, we can see that

$$x = r\cos\theta \qquad y = r\sin\theta \tag{72}$$

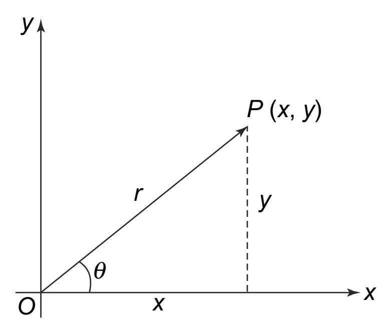


Fig. 11 *Graphical representation of a complex number*

Thus, we can write z as

$$z = r(\cos\theta + i\sin\theta) \tag{73}$$

The modulus of the complex number is then defined as the length of line OP and is given by

$$|z| = r = \sqrt{x^2 + y^2} \tag{74}$$

The angle θ is called the argument of z and is given by

$$\theta = \arctan\frac{y}{x}$$

From the infinite series expansion of $\cos\theta$ and $\sin\theta$, it is easy to see that

$$z = x + iy = r(\cos\theta + i\sin\theta) = re^{i\theta} \tag{75}$$

This form of writing complex numbers makes it very convenient to manipulate them. Thus,

$$z_1 z_2 = (r_1 e^{i\theta_1})(r_2 e^{i\theta_2}) = r_1 r_2 e^{i(\theta_1+\theta_2)} \tag{76}$$

or

$$\frac{z_1}{z_2} = \frac{r_1 e^{i\theta_1}}{r_2 e^{i\theta_2}} = \frac{r_1}{r_2} e^{i(\theta_1-\theta_2)} \tag{77}$$

Finally, using the Argand plane one can easily prove the triangle inequalities

$$|z_1| - |z_2| \le |z_1 + z_2| \le |z_1| + |z_2| \tag{78}$$

7 MATRICES

A matrix is defined as a square or a rectangular array of numbers or functions which obeys certain laws. We write a matrix, with m rows and n columns as

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & . & . & . & a_{1n} \\ a_{21} & a_{22} & . & . & . & a_{2n} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ a_{m1} & a_{m2} & . & . & . & a_{mn} \end{pmatrix} \tag{79}$$

Two matrices $\mathbf{A}$ and $\mathbf{B}$ are equal if and only if all elements are equal, i.e.,

$$\mathbf{A} = \mathbf{B} \qquad \text{if and only if} \quad a_{ij} = b_{ij} \quad \text{for all } i,\ j$$

The operations of addition and multiplication of matrices is also defined easily. Two matrices can be added

$$\mathbf{A} + \mathbf{B} = \mathbf{C} \qquad \text{where} \quad c_{ij} = a_{ij} + b_{ij} \quad \text{for all } i,\ j$$

Multiplication of two matrices is defined as

$$\mathbf{AB} = \mathbf{C} \qquad \text{where} \quad c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj} \quad \text{for all } i,\ j \tag{80}$$

that is, the $(ij)^{\text{th}}$ element of $\mathbf{C}$ is formed as a scalar product of the elements of the i^{th} row of $\mathbf{A}$ and the j^{th} column of $\mathbf{B}$. Thus, for instance, for $m = n = 3$, we have

$$c_{13} = a_{11}b_{13} + a_{12}b_{23} + a_{13}b_{33}$$

Note that matrix multiplication is **not** commutative, i.e.,

$$\mathbf{AB} \ne \mathbf{BA} \tag{81}$$

in general. A special matrix is the unit matrix $\mathbf{I}$ which is defined as

$$\mathbf{I} = \begin{pmatrix} 1 & 0 & 0 & . & . \\ 0 & 1 & 0 & . & . \\ 0 & 0 & 1 & . & . \\ . & . & . & . & . \\ . & . & . & . & . \end{pmatrix} \tag{82}$$

Clearly then, the unit matrix **I** has elements $\mathbf{I}_{ij} = \delta_{ij}$.

A null matrix **O** is one which has **all** its elements equal to zero.

Trace of a square matrix is defined as the sum of its diagonal terms. Thus,

$$\mathrm{Tr}\mathbf{AB} = \sum_i (\mathbf{AB})_{ii} \tag{83}$$

It is easy to see that even though $\mathbf{AB} \neq \mathbf{BA}$ in general,

$$\mathrm{Tr}\mathbf{AB} = \mathrm{Tr}\mathbf{BA}$$

Another quantity of great importance in physics is the determinant of a square matrix. This is defined as

$$\det\mathbf{A} = \sum_{i_1,i_2,\cdot,i_n=1}^{n} \epsilon_{i_1,i_2,\cdots,i_n} a_{1i_1}\cdots a_{ni_n} \tag{84}$$

where $\epsilon_{i_1,i_2,\cdots,i_n}$ is called the Levi–Civita symbol and is $+1$ for even permutations of $i_1, i_2, \cdots, i_n$ and -1 for odd permutations. For a 3×3 matrix, this gives us

$$\det\mathbf{A} = a_{11}(a_{22}a_{33} - a_{32}a_{23}) - a_{21}(a_{12}a_{33} - a_{32}a_{13}) + a_{31}(a_{12}a_{32} - a_{22}a_{31}) \tag{85}$$

The inverse of a matrix **A** is defined as a matrix $\mathbf{A}^{-1}$ such that

$$\mathbf{AA}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I} \tag{86}$$

We define the transpose of a matrix **A** as $\mathbf{A}^{\mathrm{T}}$ as

$$\mathbf{A}^{\mathrm{T}}_{ij} = \mathbf{A}_{ji} \tag{87}$$

that is, we interchange the rows and the columns. This operation then allows us to define symmetric and antisymmetric matrices. A matrix is said to be symmetric if

$$\mathbf{A} = \mathbf{A}^{\mathrm{T}}$$

and antisymmetric if

$$\mathbf{A} = -\mathbf{A}^{\mathrm{T}}$$

Obviously, the diagonal elements of an antisymmetric matrix are identically zero.

8 DIFFERENTIAL EQUATIONS

Differential equations play a very important role in physics. Almost all of physics is formulated in terms of differential equations. In electricity and magnetism, the basic laws are the Maxwell's equations and these are a set of differential equations. All the differential equations that we will encounter will be linear and most will be second order. Let us first define these terms.

A differential equation involves a function $f(x)$, its derivatives $\frac{df(x)}{dx}, \frac{d^2f(x)}{dx^2}, \cdots$ and x. The highest order of the derivative involved in the equation is called the **degree of the equation** and the highest power of the unknown function $f(x)$ or its derivatives is called **the order of the equation**. Thus,

$$\left(\frac{df(x)}{dx}\right)^3 + f^2(x) + x = 0 \tag{88}$$

is an equation of degree 1 and order 3. The equation

$$\frac{d^5 f(x)}{dx^5} + f(x) + 3 = 0 \tag{89}$$

is an equation of degree 5 and order 1. **Equations of order** 1 **are called linear differential equations**. Thus, the equation for the angular displacement $\theta(t)$ of a simple pendulum

$$\frac{d^2\theta(t)}{dt^2} + \frac{g}{l}\sin\theta(t) = 0 \tag{90}$$

is **not a linear equation**, but the version for small displacements is linear

$$\frac{d^2\theta(t)}{dt^2} + \frac{g}{l}\theta(t) = 0$$

A linear n^{th} degree equation has the general form

$$\boxed{P_n(x)\frac{d^n f(x)}{dx^n} + P_{n-1}(x)\frac{d^{n-1} f(x)}{dx^{n-1}} + \cdots + P_1(x)\frac{df(x)}{dx} + P_0(x) = 0} \tag{91}$$

where the functions $P_0(x), P_1(x), \cdots, P_n(x)$ can be any functions of x.

A general result for the solution of such n-degree equations is that they involve n arbitrary constants. The values of these constants are determined by imposing constraints on the solution called 'boundary conditions'.

A special class of differential equations are linear, homogenous differential equations. Consider, for instance, the equation

$$\frac{d^2 y}{dx^2} + P(x)\frac{dy}{dx} + Q(x)y = 0 \tag{92}$$

where $P(x)$ and $Q(x)$ are well-behaved functions. This equation is homogenous because each term contains the function $y(x)$ or its derivatives and it is linear because each of the quantities $y, \frac{dy}{dx}, \frac{d^2y}{dx^2}$ appears as the first power and there are no products. Of course, there can be non-homogenous equations with a term, independent of y or its derivatives on the right-hand side, like

$$\frac{d^2 y}{dx^2} + P(x)\frac{dy}{dx} + Q(x)y = F(x)$$

where $F(x)$ is called the source term. Such an inhomogeneous equation, of course, is encountered in physics a lot—Poisson's equation is an example that we would be dealing with in the next few chapters. There are several different ways of solving these equations like the Green's function method, the Laplace transform technique, etc. We shall be dealing with them later in the book. However, note that to any solution of the inhomogeneous equation, we can always add the solutions of the homogeneous equation.

The type of equation we discussed above involved a single variable x and a function thereof $f(x)$. These are called ordinary differential equations. Sometimes, there is more than one variable , $x, y, z, \cdots$ and a function $f(x, y, z, \cdots)$ dependent on these variables. The derivatives entering the partial derivatives of $f(x, y, z, \cdots)$ are with respect to $x, y, z, \cdots$. These equations are called **partial differential equations.**

As an example of the Green's function method of solving inhomogeneous equations, consider the Poisson equation

$$\nabla^2\psi = -4\pi k\rho \tag{93}$$

This is an inhomogeneous equation with a source term which is the charge density ρ. This is the equation which is satisfied as we shall see, by the electrostatic potential ψ in a region where there is a charge density ρ. This is, of course, NOT an ordinary differential equation but rather a partial differential equation which we will deal with later. But the essential nature of the Green's function solution does not depend on this nature of the differential equation. The Green's function method consists of finding a function ϕ, called the *Green's function*, such that it satisfies the following equation:

$$\nabla^2\phi = \delta^3(\vec{r}_1 - \vec{r}_2) \tag{94}$$

We can think of the function ϕ as the electrostatic potential due to a point source at $\vec{r}_2$. We now use Green's theorem and assume that the integrand falls faster than $\frac{1}{r^2}$. We can then take the volume to be at ∞ so that the surface integral goes to zero. Then

$$\iiint \phi\nabla^2\psi dV_2 = \iiint \psi\nabla^2\phi dV_2 \tag{95}$$

Using the definition of the Green's function and the Poisson equation, we get

$$-\iiint \psi(\vec{r}_2)\delta^3(\vec{r}_1 - \vec{r}_2)dV_2 = -\iiint 4\pi k\phi(\vec{r}_1, \vec{r}_2)\rho(\vec{r}_2)dV_2 \tag{96}$$

The Delta function integral can be done easily and we get

$$\psi(\vec{r}_1) = 4\pi k \iiint \phi(\vec{r}_1, \vec{r}_2)\rho(\vec{r}_2)dV_2 \tag{97}$$

We also know that

$$\nabla^2\left(\frac{1}{4\pi r_{12}}\right) = -\delta^3(\vec{r}_1 - \vec{r}_2)$$

Comparing this with the definition of the Green's function above, we get

$$\phi(\vec{r}_1, \vec{r}_2) = \frac{1}{4\pi|\vec{r}_1 - \vec{r}_2|} \tag{98}$$

and the solution to the differential equation ψ is, therefore,

$$\psi(\vec{r}_1) = k\iiint \frac{\rho(\vec{r}_2)}{|\vec{r}_1 - \vec{r}_2|}dV_2 \tag{99}$$

There are many different ways of solving ordinary linear differential equations. Perhaps the most commonly used one is the Frobenius method to get a series solution to the differential equation. Apart from this, there are other methods like the complementary function method and also methods which use integral transforms. We will not discuss these methods here.

Of particular interest to us are partial differential equations. These are differential equations where there is more than one variable. Thus, for instance, the Laplace equation in three-dimension, in Cartesian coordinates

$$\nabla^2 V(x, y, z) = \frac{\partial^2 V(x, y, z)}{\partial x^2} + \frac{\partial^2 V(x, y, z)}{\partial y^2} + \frac{\partial^2 V(x, y, z)}{\partial z^2} = 0$$

is an example of a homogenous, second order partial differential equation. The most common method of solving partial differential equations is called the Separation of Variable method. This is basically a method of splitting the partial differential equation into a set of ordinary differential equations, which can then be solved by standard methods.

We will be discussing the solution to this equation in Cartesian coordinates by the method of separation of variables in Chapter 2, Section 2.7. Instead, let us try and solve this equation in spherical coordinates which will also allow us to define a class of functions called orthogonal functions.

8a Legendre Equation

The Laplace equation in spherical polar coordinates (r, θ, ϕ) can be written as

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial V}{\partial r}\right)+\frac{1}{r^2\sin^2\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial V}{\partial\theta}\right)+\frac{1}{r^2\sin^2\theta}\frac{\partial^2 V}{\partial\phi^2}=0 \tag{100}$$

The method of separation of variables consists of looking for solution in the form of products of functions of single variables. In this case,

$$V(r,\theta,\phi)=\frac{R(r)}{r}\Theta(\theta)\Phi(\phi) \tag{101}$$

Substituting this into the Laplace equation, and multiplying by $r^2\sin^2\theta$ and dividing by Vr, we get

$$r^2\sin^2\theta\left(\frac{1}{R}\frac{d^2R}{dr^2}+\frac{1}{r^2\Theta\sin\theta}\frac{d}{d\theta}\left(\sin\theta\frac{d\Theta}{d\theta}\right)\right)+\frac{1}{\Phi}\frac{d^2\Phi}{d\phi^2}=0 \tag{102}$$

The ϕ dependence now is only in the last term and so this term must be equal to a constant, say $-m^2$. The ϕ equation thus, becomes

$$\frac{1}{\Phi}\frac{d^2\Phi}{d\phi^2}=-m^2 \tag{103}$$

This differential equation has the solutions

$$\Phi=e^{\pm im\phi} \tag{104}$$

where to ensure single valuedness, m must be an integer. Similarly, the other two equations can also be separated. Since these are independent variables, these must each be a constant. Thus,

$$\frac{d^2R}{dr^2}-\frac{l(l+1)}{r^2}R=0 \tag{105}$$

and

$$\frac{1}{\sin\theta}\frac{d}{d\theta}\left(\sin\theta\frac{d\Theta}{d\theta}\right)+\left[l(l+1)-\frac{m^2}{\sin^2\theta}\right]\Theta=0 \tag{106}$$

where l is constrained to be zero or a positive integer in order that the solution is well behaved in the entire domain $0<\theta<\pi$. The radial equation can be solved easily to get

$$R(r)=Ar^{l+1}+\frac{B}{r^l} \tag{107}$$

where A, B are constants to be determined by boundary conditions. The angular equation can be expressed in terms of another variable $x=\cos\theta$ and we get

$$\frac{d}{dx}\left[(1-x^2)\frac{d\Theta}{dx}\right]+\left[l(l+1)-\frac{m^2}{1-x^2}\right]\Theta=0 \tag{108}$$

This equation is the generalised Legendre equation with solutions known as *associated Legendre functions*. For $m=0$, the equation reduces to the Legendre equation (which would be the case if there was no ϕ dependence in the problem, or in other words, the solution would have azimuthal symmetry).

$$\frac{d}{dx}\left[(1-x^2)\frac{d\Theta}{dx}\right] + [l(l+1)]\,\Theta = 0 \tag{109}$$

This equation can be solved by the power series (Froebenius method). The solutions which are finite at $x = \pm 1$ are possible only if the series terminates and that is possible only if the constant l is zero or a positive integer. In this case, the solutions become polynomials called the Legendre polynomials, which are given by the Rodrigues' Formula

$$P_l(x) = \frac{1}{2^l l!}\frac{d^l}{dx^l}(x^2-1)^l \tag{110}$$

The first few polynomials are

$$\begin{aligned} P_0(x) &= 1 \\ P_1(x) &= x \\ P_2(x) &= \frac{1}{2}(3x^2-1) \end{aligned} \tag{111}$$

etc.

The Legendre polynomials are orthogonal polynomials in the sense that

$$\int_{-1}^{1} P_l(x)P_{l'}(x)dx = \frac{2}{2l+1}\delta_{ll'} \tag{112}$$

We have considered the case where $m = 0$ or the case where the problem has azimuthal symmetry. In general however, $m \neq 0$ and hence, the solutions to the angular part of the Laplace equation also has a m dependence. It can be shown that to have finite solutions to the generalised Legendre equation given above, in the range $-1 \leq x \leq 1$, we must have l to be zero or positive integer and m, which we had restricted to be an integer, take only values $-l, -(l-1), \cdots, 0, \cdots (l-1), l$. The solutions are called the associated Legendre functions and are defined by (for both positive and negative m)

$$P_l^m(x) = \frac{(-1)^m}{2^l l!}(1-x^2)^{m/2}\frac{d^{l+m}(x^2-1)^l}{dx^{l+m}} \tag{113}$$

One can also show that $P_l^m(x)$ and $P_l^{-m}(x)$ are proportional and

$$P_l^{-m}(x) = (-1)^m\frac{(l-m)!}{(l+m)!}P_l^m(x) \tag{114}$$

The associated Legendre functions are also orthogonal and

$$\int_{-1}^{1} P_{l'}^m(x)P_l^m(x)dx = \frac{2}{2l+1}\frac{(l+m)!}{(l-m)!}\delta_{ll'} \tag{115}$$

Frequently, it is more convenient to combine the θ and ϕ solutions and use orthonormal functions, which are called spherical Harmonics. These are, of course, linear combinations of the associated Legendre functions and the solutions to the ϕ equation and are defined by

$$Y_{lm}(\theta,\phi) = \sqrt{\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}}P_l^m(\cos\theta)e^{im\phi} \tag{116}$$

The spherical harmonics for $+m$ and $-m$ are related as

$$Y_{l,-m}(\theta,\phi) = (-1)^m Y^*_{lm}(\theta,\phi) \tag{117}$$

They are also orthonormal since

$$\int_0^{2\pi} d\phi \int_0^{\pi} \sin\theta d\theta Y^*_{l'm'}(\theta,\phi) Y_{l,m}(\theta,\phi) = \delta_{ll'}\delta_{mm'} \tag{118}$$

Some of the spherical harmonics can be easily seen to be

$$\begin{aligned}
Y_{00} &= \frac{1}{\sqrt{4\pi}} && l=0\\
Y_{11} &= -\sqrt{\frac{3}{8\pi}}\sin\theta e^{i\phi} && l=1, m=1\\
Y_{10} &= \sqrt{\frac{3}{4\pi}}\cos\theta && l=1, m=0\\
Y_{22} &= \frac{1}{4}\sqrt{\frac{15}{2\pi}}\sin^2\theta e^{2i\phi} && l=2, m=2\\
Y_{21} &= -\sqrt{\frac{15}{8\pi}}\sin\theta\cos\theta e^{i\phi} && l=2, m=1\\
Y_{20} &= \sqrt{\frac{5}{4\pi}}\left(\frac{3}{2}\cos^2\theta - \frac{1}{2}\right) && l=2, m=0
\end{aligned} \tag{119}$$

With all this, we can now write the general solution to the Laplace equation in spherical polar coordinates as

$$\boxed{V(r,\theta,\phi) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l}\left[A_{lm}r^l + B_{lm}r^{-(l+1)}\right]Y_{lm}(\theta,\phi)} \tag{120}$$

8b Bessel Equation

We have seen above how the Laplace equation in spherical polar coordinates, when separated, leads to the generalised Legendre equation for the angular part. Let us now consider the Laplace equation in cylindrical coordinates. The equation is

$$\nabla^2 V = \frac{\partial^2 V}{\partial\rho^2} + \frac{1}{\rho}\frac{\partial V}{\partial\rho} + \frac{1}{\rho^2}\frac{\partial^2 V}{\partial\phi^2} + \frac{\partial^2 V}{\partial z^2} = 0 \tag{121}$$

Using the method of separation of variables once again, we write

$$V(\rho,\phi,z) = R(\rho)\Phi(\phi)Z(z) \tag{122}$$

and we get three ordinary differential equations

$$\frac{d^2 Z}{dz^2} - k^2 Z = 0 \tag{123}$$

$$\frac{d^2\Phi}{d\phi^2} + \nu^2\Phi = 0 \tag{124}$$

and

$$\frac{d^2R}{d\rho^2} + \frac{1}{\rho}\frac{dR}{d\rho} + (k^2 - \frac{\nu^2}{\rho^2})R = 0 \tag{125}$$

The z and ϕ equations can be easily solved to get

$$Z(z) = e^{\pm kz} \tag{126}$$

$$\Phi(\phi) = e^{\pm i\nu\phi} \tag{127}$$

Once again, to make sure that the solution is single valued under a rotation of 2π in the ϕ direction, ν should be an integer. The parameter k has no such restriction, except if there are any restrictions in the boundary conditions in the z direction.

The radial equation can be put in a more convenient form by changing variables to $x = k\rho$. Then the equation becomes

$$\frac{d^2R}{dx^2} + \frac{1}{x}\frac{dR}{dx} + \left(1 - \frac{\nu^2}{x^2}\right)R = 0 \tag{128}$$

This equation is called the *Bessel equation* and the solutions are Bessel functions of order ν. Using the Froebenius method to solve this we get the two solutions to be

$$J_\nu(x) = \left(\frac{x}{2}\right)^\nu \sum_{j=0}^{\infty} \frac{(-1)^j}{j!\Gamma(j+\nu+1)} \left(\frac{x}{2}\right)^{2j} \tag{129}$$

and

$$J_{-\nu}(x) = \left(\frac{x}{2}\right)^{-\nu} \sum_{j=0}^{\infty} \frac{(-1)^j}{j!\Gamma(j-\nu+1)} \left(\frac{x}{2}\right)^{2j} \tag{130}$$

$J_\nu(x)$ and $J_{-\nu}(x)$ are called *Bessel functions of the first kind*. Here $\Gamma(\nu)$ is the Gamma function of ν.

Bessel functions of the second kind or Neumann functions are defined as

$$N_\nu(x) = \frac{J_\nu(x)\cos\nu\pi - J_{-\nu}(x)}{\sin\nu\pi} \tag{131}$$

Linear combinations of the Bessel functions of the first and second kind are called *Bessel functions* of the third kind or *Hankel functions*, and are defined by

$$H_\nu^{(1)}(x) = J_\nu(x) + iN_\nu(x) \tag{132}$$

$$H_\nu^{(2)}(x) = J_\nu(x) - iN_\nu(x) \tag{133}$$

Finally, sometimes it is useful to have the asymptotic forms of the Bessel functions for large and small values of the argument.

$$J_\nu(x) \to \frac{1}{\Gamma(\nu+1)}\left(\frac{x}{2}\right)^\nu \qquad\qquad x \ll 1 \tag{134}$$

and

$$J_\nu(x) \to \sqrt{\frac{2}{\pi x}}\cos(x - \frac{\nu\pi}{2} - \frac{\pi}{4}) \qquad\qquad x \gg 1 \tag{135}$$

1

Electric Forces

Learning Objectives

- To comprehend the nature of electric charge and its properties.
- To learn about the nature of force between point charges.
- To learn about how forces due to several point charges add vectorially to give the net force and to use this fact for continuous charge distributions.
- To understand the concept of an electric field and how it leads to a local description of the electrostatic phenomenon.
- To learn how to calculate electric field due to various charge configurations.
- To understand the concept of electric lines of force and electric flux.
- To learn about Gauss's law and its use in determining the electric field for situations with symmetry.

1.1 ELECTRIC CHARGE

Matter in its natural form is characterised by mechanical properties like its mass, density, shape and size, etc. We also know that most phenomena relating to matter depend on these properties. However, there are certain kinds of processes that cannot be explained in terms of a substance's mechanical properties. One such simple process is rubbing of the substance—it has been known for a long time that the certain behaviour of matter when it is rubbed can only be explained in terms of a new characteristic called the *electric charge*. It has also been known for quite some time that this electric charge occurs in two forms, called positive and negative. Forces that exist between the charges are repulsive between like charges and attractive between unlike ones. This kind of force between charged pieces of matter is given the name 'Electrostatic Force'.

The nature of electric charge has become more transparent with our present knowledge of the structure of matter at the sub-microscopic level. We know that matter is composed of molecules, which in turn are composed of one, two or more atoms. The number of molecules in matter is incredibly large with a gram molecule of any substance having 6.022×10^{23} molecules (Avogadro Number). Beginning early twentieth century, it has also been known that atoms are composed of equal numbers of protons and electrons which are oppositely charged, together with neutrons which are electrically neutral. The charge on the proton is called positive and that on the electron negative. This is just a matter of convention and we could have defined it the other way around. The important thing is that charges are additive, like positive and negative numbers. If a body contains an equal number of protons and electrons, it is on the whole, electrically neutral. If their numbers are different, the net charge on the body is positive if the number of protons exceed the number of electrons. In the opposite situation, the net charge is negative.

History of Electrostatics

The first recorded mention of some kind of electrostatic force is by Thales of Miletus around 600 BC when he observed that rubbing rods of amber (fossilised resin) with certain materials (like animal fur) makes it attract other objects. Thales attributed it to a magnetic force but later in the early seventeenth century, William Gilbert, the English physician investigated the phenomenon (technically called the Triboelectric effect) and concluded that the phenomenon was a different one from magnetic attraction by magnetite. He was also the one who coined the word *electricus* from *elektron*, the Greek word for amber.

Several scientists contributed to a greater understanding of the electrostatic phenomenon. These included Boyle, Guricke and Benjamin Franklin. It was the classic experiment of Franklin flying a kite in a storm which demonstrated that lightning was nothing but an electrical phenomenon. Franklin is also credited with the convention of denoting the charge on the electron as negative. It was with the French scientist Coulomb in the late eighteenth century that the study of electrostatics became mathematically rigorous. He not only discovered that charge could be of two kinds, positive and negative, but also the force law between two charges, which depended on the amount of the charges and inverse of the square of the distance between them. The SI unit of charge is named in honour of Coulomb.

Given the atomistic picture of matter, several aspects of the phenomena of electric charges become obvious. Charging neutral matter clearly involves the piece of matter losing or gaining some of these charged particles so that they have more of one type than the other. Since charging or discharging pieces of matter simply involves exchanges of the charged particles, the total charge of any closed system (i.e., a system in which all interactions are limited as being within the system with nothing leaving or entering the system) clearly does not change since none of the charged constituents of matter get destroyed or created in the process. Charge on a body can change only if the charged constituents transform into something that has a different charge. Everyday experience tells us this does not happen. Of course, our everyday experience can only tell us whether such a process happens within a time span of say months or years or even decades. **In the absence of any evidence to the contrary, it is believed that the total amount of charge of any isolated system does not change at all. This principle of 'Conservation of Charge'** has been verified to great precision and becomes a cornerstone in later theoretical developments of electromagnetism.

Experiments to detect whether electric charge is conserved focus on decay of protons or electrons into decay products that have a total charge different from the parent particle. These experiments put a limit on the time over which such a violation of the Principle of Conservation of Charge can occur. A typical experiment performed in 2002 (**H.O.Back et. al., Physics Letter B525, pages 29–40, 2002**) studies, the possibility of the following decay:

$$e^- \rightarrow \nu + \gamma$$

that is, an electron (e^-) decaying into a photon (γ) which is a quanta of electromagnetic radiation and a particle called a neutrino (ν). As we will see later, light is a form of electromagnetic radiation just as radio waves and microwaves. Quantum theory tells us that electromagnetic radiation can be thought of as being made up of quanta of energy called photons, which are electrically neutral. Neutrino is a particle that was postulated to explain certain anomalies in nuclear reactions and discovered subsequently. It is a very unusual particle since it is almost massless and carries no charge. Now since both the decay particles are electrically neutral, if this decay was observed, it would mean that a charged electron is decaying into neutral particles and hence, the charge is being destroyed. This would indicate that the principle of Conservation of Charge is being violated.

Such an experiment was carried out by H.O. Back and others in 2001. The challenging part of observing such a decay would be the observation of the neutral particles. Particles are detected usually by their interaction with other particles. Charged particles, for instance, interact with the atoms and molecules of a material, and deposit their energy in the form of tracks in the material. Neutral particles, though not experiencing electric forces, can still ionise atoms by transferring their energy. The ionised atoms emit radiation in a phenomenon called luminescence. In the experiment above, a large chamber of luminescent liquid is used to see if the photon (which will cause luminescence) in the reaction above is produced or not. The chamber was observed for about 30 days and no event (luminescence) was seen. This puts a limit on the lifetime of the electron by this reaction to be greater than 4.6×10^{26} years, which is greater than the age of the universe. This is the best limit as yet on the possibility of violation of the Principle of Charge Conservation.

We will thus adopt as a postulate (which is supported amply by empirical observations) the **Principle of Charge Conservation** as follows-

> The total electric charge, i.e., the algebraic sum of the positive and negative charge, in an isolated system does not change at any time.

An equally startling fact also emerges when we try to study the nature of charge. Since matter is electrically neutral on the whole, the charge carried by the proton must be equal in magnitude and opposite in sign to that of the electron. This is an amazing fact which we don't understand yet. But clearly, the fact that the electron and the proton, two particles which otherwise differ so much in their other properties (masses, their interactions with other submicroscopic particles, etc.), carry equal magnitude of charge is something which is a deep law of nature. By extension, anything getting charged must also have a charge equal in magnitude to an integral multiple of the basic charge carried by these particles.

This fact became even more puzzling with the subsequent discovery of other charged subatomic particles in nature, i.e., other than these two that exist in stable matter. The charged particles discovered also revealed that the charges carried by them share the same property, i.e., their charges were integral multiples of this basic unit of charge. Carefully prepared charged bodies, e.g., oil drops used in the classic Millikan's experiment, carry a net charge of a few units of this basic charge. However, in most cases of electrically charged objects that we encounter, like for example, a material charged by rubbing it with silk or the charges on the terminals of a battery, the net charge is many millions times this basic charge. For instance, an ordinary pencil cell, when fully charged has something like 10^{23} times this basic unit of charge on its terminals. Hence, for all practical purposes, when we are dealing with macroscopic charged bodies, this **'quantisation'** of the magnitude of charge plays no role. By quantisation, we mean that all charges that we encounter are simple integral multiples of a basic unit or quantum of charge (that which is carried by the electron or proton). Given that all ordinary bodies when charged have an enormous amount of charge when measured in terms of this quantum, we can safely consider the charges on these bodies as continuous and not discrete. However, it must be remembered that at the most fundamental level, charge is very much a discrete quantity and its basic quantum is the charge of an electron or proton.

Quarks

With the discovery of the electron, proton and the neutron by the middle of the twentieth century, it was thought that one could explain all submicroscopic phenomenon in terms of the interactions of these fundamental particles alone. However, by then, various experiments and observations of cosmic rays, etc., made it clear that the number of particles was significantly more than these three. The discovery of the pi meson, and a host of other sub-atomic particles motivated several scientists to look for more fundamental building blocks from which these newly discovered particles could be built.

In 1964, Murray Gell-Mann proposed the quark model, which postulated the existence of fundamental particles called quarks, which have very unusual properties. Combinations of quarks and anti-quarks, in this model, were then used to build various hadrons (subatomic particles that experience the strong nuclear force). In the original model, three kinds of quarks were postulated—the **up (u) quark, the down (d) quark and the strange (s) quark**. The up quark carried a charge which was $(+\frac{2}{3}e)$ while the down and the strange quarks carried a charge $(-\frac{1}{3}e)$ each. A quark-antiquark combination was required to build various mesons while baryons (like proton, neutron, etc.) were composed of three quarks (or their antiparticles).

Subsequently, three more kinds of quarks, the **top(t), bottom(b) and charm(c)** were postulated and discovered. With the advent of the unified theory of electromagnetic and weak interactions in the late 1960s, the six quarks were assumed to be fundamental and their interactions understood.

The basic unit of charge was thought to be the magnitude of the electron or proton charge for a long time. Laboratory experiments carried out in the 1970s indicated the existence of particles, called quarks, which carried charges of magnitude $(\frac{1}{3})$ and $(\frac{2}{3})$ of the magnitude of the electron's charge. These quarks however, do not exist singly but only in combinations carrying charges that are integral multiples of the electronic charge. We do not yet have a basic theoretical understanding of this phenomenon, usually referred to as **'quantisation of charge'**. The experimentally verified equality of the magnitude of the charge of a proton and electron does not have any obvious theoretical explanation and is a subject matter of much research even today. Later in the course of our discussion of electromagnetism, we will talk more about this subject.

1.2 COULOMB'S LAW

Forces between electrically charged bodies have been known for quite some time but it was Coulomb who discovered the basic law governing this. Two charged bodies exert forces on one another such that the forces on them are equal in magnitude but oppositely directed. Coulomb showed that it is possible to assign to every charged body a quantity that can be called 'amount of charge' such that the force F between charged bodies depends on the amount of charge and the distance between them.

Coulomb's Law is stated in terms of point charges, i.e., charged bodies that have vanishingly small size. To specify the location of a finite size object one obviously needs to specify the positions of all its constituents. Location of a point charge, on the other hand, can be described by a single position vector. Consider two such point charges A and B carrying amounts of charges q_A and q_B respectively and separated by a distance r_{AB}.

Coulomb's Law states that two point charges experience an equal and opposite force, which is directed along the line joining them and has a magnitude proportional to the product of the charges and varying inversely as the square of the distance between them:

$$F \propto \frac{q_A q_B}{r_{AB}^2} \tag{1.1}$$

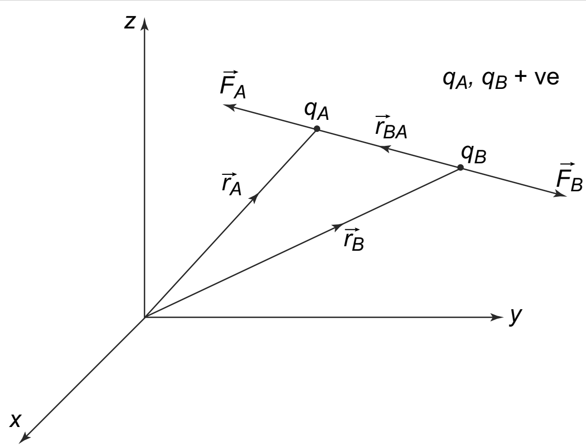


Fig. 1.1 *Two point charges experience equal and opposite forces given by Coulomb's Law*

This can be stated in terms of vectors as follows. Let $\hat{n}_{BA}$ be the unit vector along the line joining the positions of A and B and directed from B to A (Fig 1.1). Then the electrical forces on q_A and q_B, denoted respectively by $\vec{F}_A$ and $\vec{F}_B$ respectively, are given by

$$\vec{F}_A = -\vec{F}_B = \frac{kq_A q_B}{r_{AB}^2}\hat{n}_{BA} \tag{1.2}$$

where k is a positive constant. Note that the nature of the forces, attractive or repulsive, is built into the above equation. Thus, for example, if the two point charges q_A and q_B have the same sign, the force on A is away from B and this equation correctly shows that.

The value of the constant k in the equation above clearly depends on the units in which the quantity of charge is measured. The earliest system of units, called electrostatic units (esu) defined 1 esu of charge in the following way: Two point charges, each carrying 1 esu of charge and separated by 1 cm. exert a force of 1 dyne (1 dyne = 10^{-5} Newton) on each other. That would mean $k = 1$ in Eq. (1.2) above, if F is measured in dynes and the charges in esu. The system of units that is commonly used these days is the SI (short for the phrase in French 'Systeme Internationale') system of units. The unit of quantity of charge in SI units is the Coulomb, the unit of force is Newtons, unit of distance is a meter and the value of $k = 8.988 \times 10^9\ \mathrm{Nm^2C^{-2}}$. The quantity k is often written as $k = \frac{1}{4\pi\varepsilon_0}$. A discussion of systems of units in electromagnetism is given in **Appendix A**. In our discussion, we shall be using the constant k. However, in some of the solved problems, we will use SI units so that the reader is aware of their use.

1.2.1 Principle of Superposition

Coulomb's law as written above refers to forces exerted on one another by two point charges independent of everything else. That implies that electrostatic forces are additive in the following sense. Consider three point charges q_1, q_2 and q_3 at points 1, 2 and 3 respectively as shown in Fig. 1.2.

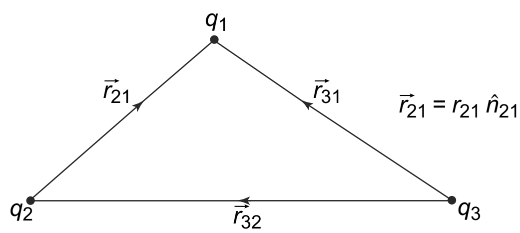


Fig. 1.2 *Three point charges q_1, q_2 and q_3 are at points 1,2 and 3 respectively. A vector joining point i to point j is denoted by $\vec{r}_{ij}$ and the corresponding unit vector is denoted by $\hat{n}_{ij}$. Obviously, $\vec{r}_{ij} = -\vec{r}_{ji}$. Thus, in the figure the vectors joining 2 to 1, 3 to 1 and 3 to 2, are $\vec{r}_{21}, \vec{r}_{31}$ and $\vec{r}_{32}$ respectively. The corresponding unit vectors are denoted by $\hat{n}_{21}, \hat{n}_{31}$ and $\hat{n}_{32}$ respectively*

If the charge at point 3 were not there, the force on q_1 would be

$$\frac{kq_1q_2}{r_{12}^2}\hat{n}_{21} \tag{1.3}$$

by the Coulomb's law stated above. Similarly, if the charge at point 2 were not there, the force on q_1 would be

$$\frac{kq_1q_3}{r_{13}^2}\hat{n}_{31} \tag{1.4}$$

If now all charges are there, the force on the charge q_1 would be the vector sum of these, i.e.,

$$k\left[\frac{q_1q_2}{r_{12}^2}\hat{n}_{21} + \frac{q_1q_3}{r_{13}^2}\hat{n}_{31}\right] \tag{1.5}$$

This principle, which shows that electric force due to charges can be added vectorially to get the total force on any charge, is called the Principle of Superposition. This principle is very convenient to calculate the total force on a point charge due to any number of point charges or due to a continuous distribution of charges. We give a few examples.

EXAMPLE 1.1 A cube of edge r has charge q placed at each corner. Find the magnitude of the electrical force on any one of the charges.

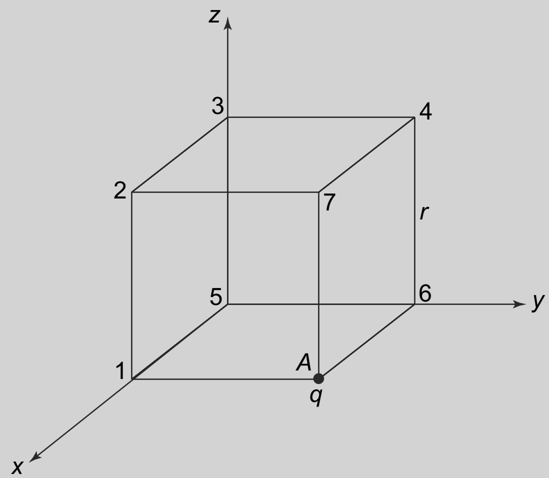


Fig. 1.3 *Example 1.1*

Solution

Let us consider the cube placed as in Fig. 1.3. Then on any one charge, the force is due to 7 other charges. Three of them are at a distance r, three of them at a distance $\sqrt{2}r$ and one at a distance $\sqrt{3}r$. By the superposition principle, the force from each charge acts independently of the other charges and the resultant force on the charge is just a vector sum of all the forces. Assuming a coordinate system as shown in the figure, we get for the the forces $\vec{F}_i$ $(i = 1, \cdots, 7)$ acting on the charge at the point A marked in the figure caused by the charges at location $i(i = 1, \cdots, 7)$ marked in the Fig. 1.3:

$$\vec{F}_1 = \frac{q^2}{4\pi\varepsilon_0 r^2}\hat{j} \tag{1.6}$$

$$\vec{F}_2 = \frac{q^2}{4\pi\varepsilon_0 r^2}\frac{1}{2\sqrt{2}}(\hat{j} - \hat{k}) \tag{1.7}$$

$$\vec{F}_7 = -\frac{q^2}{4\pi\varepsilon_0 r^2}\hat{k} \tag{1.8}$$

$$\vec{F}_6 = \frac{q^2}{4\pi\varepsilon_0 r^2}\hat{i} \tag{1.9}$$

$$\vec{F}_4 = \frac{q^2}{4\pi\varepsilon_0(2r^2)}\frac{1}{\sqrt{2}}(\hat{i} - \hat{k}) \tag{1.10}$$

$$\vec{F}_5 = \frac{q^2}{4\pi\varepsilon_0(2r^2)}\frac{1}{\sqrt{2}}(\hat{i} + \hat{j}) \tag{1.11}$$

$$\vec{F_3} = \frac{q^2}{4\pi\varepsilon_0(3r^2)}\left(-\frac{1}{\sqrt{3}}\hat{k} + \frac{1}{\sqrt{3}}(\hat{i} + \hat{j})\right) \tag{1.12}$$

With these vector forces, one can find the magnitude of the force, which is given by

$$F_{\text{resultant}} = \frac{0.261q^2}{\varepsilon_0 r^2}$$

The direction of the force is along the body diagonal, i.e., along the force F_3.

PROBLEM 1.1 Points A, B, C, D, E and F are located on a circle of radius L whose centre is O. The z-axis is along OA. Chords OB, OC, OD, OE and OF make angles of $\frac{\pi}{6}, \frac{\pi}{3}, \frac{\pi}{2}, \frac{2\pi}{3}$ and $\frac{5\pi}{6}$ respectively with OA. Point charges $q, 2q, 3q, 4q, 3q$ and $2q$ are located respectively at A, B, C, D, E and F. Calculate the electric force on the charge q at A.

PROBLEM 1.2 How many electrons should be there at the centre of the earth such that the repulsive force due to them on an electron at the surface of the earth just balances the earth's attractive gravitational force on it.

1.2.2 Continuous Charge Densities

While the Principle of Superposition can be used directly for any number of discrete point charges, it is also useful when one wants to find the force due to or on an extended charged object. Such objects are called charge distributions and are usually considered to be uniformly charged. Depending on the geometry of the object, one can assign a charge density to it.

Consider first a one-dimensional object like a line of length L. Suppose we distribute a charge Q uniformly on this line. Then, one can assign a charge density per unit length to this charged object, which is given by

$$\lambda = \frac{Q}{L}$$

This quantity is called the linear charge density. If the line has infinite length, the total charge will, of course, be infinite. We can also have non-uniform charge densities $\lambda(l)$ defined as

$$\lambda(l) = \lim_{\Delta l \to 0} \frac{\Delta Q(l)}{\Delta l}$$

where $\Delta Q(l)$ is the charge contained in the length Δl along the line at a point separated by l from some chosen origin on the line. The total charge on the line then is

$$Q = \int_{-\infty}^{\infty} \lambda(l)dl$$

Similarly, one can consider a two-dimensional surface Σ of area A on which we uniformly put a charge Q. In this case, one can assign a surface charge density σ to a two-dimensional surface which is charged. σ, **the surface charge density then is the amount of charge per unit area**.

$$\sigma = \frac{Q}{A}$$

if the charge distribution is uniform or more generally

$$\sigma(S) = \lim_{\Delta S \to 0} \frac{\Delta Q(S)}{\Delta S}$$

where $\Delta Q(S)$ is the amount of charge in a surface area ΔS at a point S on the surface Σ.

Finally, a volume V that is uniformly charged will have a volume charge density ρ defined as the charge per unit volume

$$\rho = \frac{Q}{V}$$

if the distribution is uniform or more generally

$$\rho(\vec{r}) = \lim_{\Delta V \to 0} \frac{\Delta Q(\vec{r})}{\Delta V}$$

where $\Delta Q(\vec{r})$ is the charge contained in a volume ΔV around the point $\vec{r}$.

It is important to remember that, strictly speaking, quantisation of charge forbids this kind of indefinite sub-division of charge. However, since the atomic volumes are so small, for any reasonable macroscopic object, this procedure is valid.

The strategy to find the force on a charge density due to another charge is then as follows: One considers the extended charged object (a line, surface or volume) to be made up of infinitesimal parts, each carrying some charge, which is given by the charge density times the length, area or volume respectively of the infinitesimal part. This infinitesimaly charged object now behaves like a point charge and one can then use Coulomb's Law to find the force on this from another point charge. For finding the force on the whole charged object, one can then use the Principle of Superposition and vectorially add up the forces on all the infinitesimal charges. If, on the other hand, we need to determine the force due to an extended charged object on another charge, then we do the same procedure. This vectorial addition for the infinitesimal parts of the extended body is basically an integration over the whole body. The integration is over a line, surface or a volume depending on the geometry. We illustrate this by some examples.

EXAMPLE 1.2 AB is a line charge of length L which is uniformly charged and carries a total charge Q. A point charge q is located at the point C, situated at a distance R along the right bisector of AB. Calculate the electric force on the point charge due to the line charge.

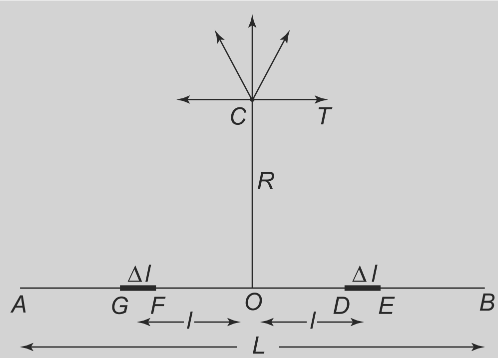


Fig. 1.4 *Example 1.2*

Solution

Consider two infinitesimally small portions of the line charge, shown as DE and FG in Fig. 1.4, symmetrically situated about O, each of length Δl with $OD = OF = l$.

The charge carried by the infinitesimal pieces are $\left(\frac{Q}{L}\right)\Delta l$, since the line charge is uniformly charged. The force due to the piece DE on the point charge q at C is

$$\overrightarrow{F}_q = -k\frac{qQ\Delta l}{L}\frac{1}{(R^2+l^2)}\hat{n}_{CE} \tag{1.13}$$

where $\hat{n}_{CE}$ is a unit vector along CE. $\hat{n}_{CE}$ has components $\frac{R}{\sqrt{(R^2+l^2)}}$ along CO and $\frac{l}{\sqrt{(R^2+l^2)}}$ along CT. Resolving the force on q due to DE along these two directions, we have:

Component of the force along CO

$$-k\frac{qQ\Delta l}{L}\frac{1}{(R^2+l^2)}\frac{R}{\sqrt{(R^2+l^2)}} \tag{1.14}$$

and the component of the force along CT

$$-k\frac{qQ\Delta l}{L}\frac{1}{(R^2+l^2)}\frac{l}{\sqrt{(R^2+l^2)}} \tag{1.15}$$

Exactly similarly the force on q due to the piece FG, will be

$$-k\frac{qQ\Delta l}{L}\frac{1}{(R^2+l^2)}\hat{n}_{CF} \tag{1.16}$$

where $\hat{n}_{CF}$ is a unit vector along CF. The component of $\hat{n}_{CF}$ along CO is the same as that of $\hat{n}_{CE}$ but its component along CT is opposite to that of $\hat{n}_{CE}$.

When we add the forces due to the two pieces, the components along CO add up but the ones along CT cancel. The force due to these two pieces on the charge q is therefore, equal to

$$-2k\frac{qQ\Delta l}{L}\frac{1}{(R^2+l^2)}\frac{R}{\sqrt{(R^2+l^2)}} \tag{1.17}$$

and is directed along CO.

The entire line charge can be divided into such pairs of infinitesimal charges with l varying from 0 to $\left(\frac{L}{2}\right)$. The total force on the charge q, by the Principle of Superposition is the sum of all these forces. This, in the limit that the infinitesimal parts are vanishingly small, is the integral over the line charge and is given by

$$F = -\int_0^{L/2}\frac{2kqQ}{L}\frac{1}{R^2+l^2}\frac{R}{\sqrt{(R^2+l^2)}}dl \tag{1.18}$$

along CO.

We can do the integral to get

$$F = -\frac{2kqQ\sin[\arctan\left(\frac{L}{2R}\right)]}{LR}$$

EXAMPLE 1.3 A circular ring of radius R is uniformly charged with centre O carries a total charge Q as shown in Fig. 1.5. Calculate the force due to the ring on a point charge q placed at a point P distance L from the centre such that OP is perpendicular to the plane of the ring.

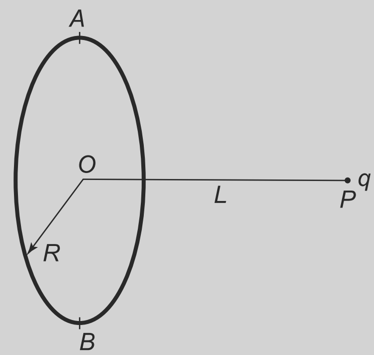


Fig. 1.5 *Example 1.3*

Solution

The charge per unit length on the ring is $\frac{Q}{2\pi R}$. Consider two infinitesimally small pieces of the ring, each of length Δl and situated at diametrically opposite points A and B on the ring. Each subtends an angle $\Delta\theta = \frac{\Delta l}{R}$ at the centre of the ring and each carries a charge $\frac{Q\Delta l}{2\pi R}$. The force on q due the piece at A is

$$\vec{F}_A = -\frac{kqQ\Delta l}{2\pi R}\frac{1}{R^2 + L^2}\hat{n}_{PA} \tag{1.19}$$

where $\hat{n}_{PA}$ is a unit vector along PA. The force due to the piece at B similarly is

$$\vec{F}_B = -\frac{kqQ\Delta l}{2\pi R}\frac{1}{R^2 + L^2}\hat{n}_{PB} \tag{1.20}$$

where $\hat{n}_{PB}$ is a unit vector along PB. We can now resolve these forces along PO and at right angles to it. The component of $\hat{n}_{PA}$ along PO is $\frac{L}{\sqrt{R^2+L^2}}$ and at right angles to PO is $\frac{R}{\sqrt{R^2+L^2}}$. The component of $\hat{n}_{PB}$ along PO is the same as that of $\hat{n}_{PA}$ but the one at right angles to PO is opposite in direction and equal in magnitude.

Hence, if we add up the two forces above, their components along PO will add up equally but will cancel in the perpendicular direction. The sum of the two forces is

$$-2\frac{kqQ\Delta l}{2\pi R}\frac{1}{R^2 + L^2}\frac{L}{\sqrt{R^2 + L^2}} \tag{1.21}$$

directed along PO. The whole circle can be divided into such pairs of infinitesimal charges and hence the total force on q directed along PO will be

$$F = -2\int_0^{\pi}\frac{kqQ}{2\pi R}\frac{1}{R^2 + L^2}\frac{L}{\sqrt{R^2 + L^2}}Rd\theta \tag{1.22}$$

$$= -\frac{kQqL}{(R^2 + L^2)^{3/2}} \tag{1.23}$$

where we have used the relation $\Delta\theta = \Delta l/R$

EXAMPLE 1.4 Find the electric force on a unit charge (called Electric Field—see next section) at a distance z along the perpendicular from the centre of a charged circular plate of radius r and surface charge density σ, using Coulomb's Law.

Solution

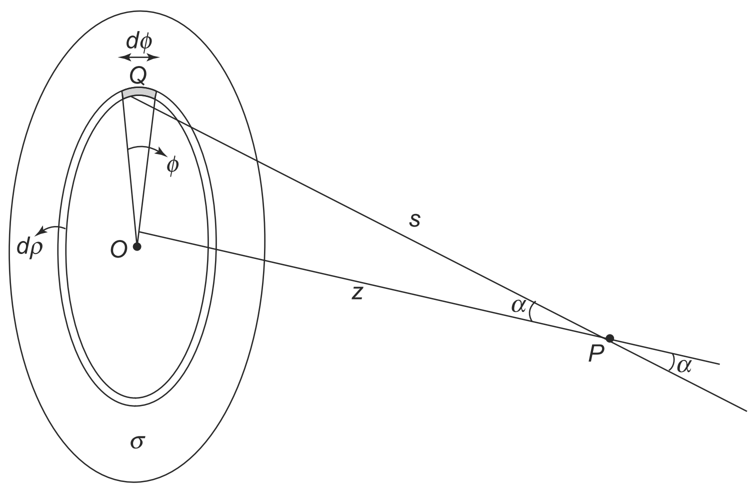


Fig. 1.6 *Example 1.4*

From Fig. 1.6, it is easy to see that one should use cylindrical coordinates. The electric field due to a small portion at a point Q of the ring at a point (ρ, ϕ) having widths $d\rho$ and $d\phi$, along the direction from Q to P is

$$dE = \frac{1}{4\pi\varepsilon_0}\frac{\sigma\rho d\phi d\rho}{(\rho^2 + z^2)} \tag{1.24}$$

This field along QP has component $\cos\alpha dE$ along the line joining P to the centre O of the plate and also a component $\sin\alpha dE$ perpendicular to it. Once again by symmetry, only the $\cos\alpha$ component will survive when we add all the contributions and so the field will be

$$dE\cos\alpha = \frac{1}{4\pi\varepsilon_0}\frac{\sigma\rho d\phi d\rho}{(\rho^2 + z^2)}\cos\alpha \tag{1.25}$$

We need to integrate this expression. Clearly, the limits of ϕ are from 0 to 2π. Writing ρ in terms of z and α, and integrating α from 0 to $\tan^{-1}(r/z)$, we get the field as

$$E = \frac{\sigma}{2\varepsilon_0}\int_0^{\tan^{-1}(r/z)} \sin\alpha\, d\alpha \tag{1.26}$$

which gives us

$$E = \frac{\sigma}{2\varepsilon_0}\left[1 - \cos(\tan^{-1}(r/z))\right] \tag{1.27}$$

This can be written as

$$E = \frac{\sigma}{2\varepsilon_0}\left[1 - \frac{z}{\sqrt{r^2 + z^2}}\right] \tag{1.28}$$

If the plate is infinite, then we get the result for the field near an infinite non-conducting plate namely

$$E = \frac{\sigma}{2\varepsilon_0} \tag{1.29}$$

Coulomb's Law was shown to be experimentally true by Coulomb, who used a torsion balance to measure the electrical force and then later by Cavendish in the late eighteenth century. Over the years, the law has been tested and better limits placed on its validity. Typically, the tests of the law determine whether the distance dependence of the force is truly of the inverse square type. That is, the tests assume that the force law between two point charges varies as $\frac{1}{r^{2+\delta}}$ and then place limits on the value of the correction parameter δ. Cavendish, who used two conducting globes to test the inverse square dependence placed an upper limit of

$$|\delta| \le 0.02$$

Cavendish's experiment was improved up by Maxwell who found that

$$|\delta| \le 5 \times 10^{-5}$$

which is a remarkable increase in precision. The limits on δ continued to be improved and in 1971, Williams, Faller and Hill (**E.R. Willams, J.E. Faller and H.A. Hill, Phys. Rev. Lett., 26, 721 (1971)**) found that

$$|\delta| = (2.7 \pm 3.1) \times 10^{-16}$$

In 1983, Crandall reported another laboratory experiment to determine the limit on δ and found that (**R.E. Crandall, *American Journal of Physics*, 51,698(1983)**.)

$$|\delta| \le 6 \times 10^{-17}$$

It turns out that the precise inverse square nature of the force between two point charges is related to the mass of the photon. If the force varies exactly as the inverse of the square of the distance, the photon mass is zero. Experiments have been done to put limits on the mass of the photon in the laboratory. Various astrophysical data have also been used to place limits on the photon mass. In all situations, it is seen that Coulomb's Law is valid to an astonishing degree of accuracy.

PROBLEM 1.3 A thin spherical shell has a radius R and is uniformly charged with a charge density σ per unit area. Calculate, by direct use of Coulomb's law, the electric force on a point charge q located at a distance r from the centre of the shell.

PROBLEM 1.4 A thin circular ring of radius R carries a uniform line charge density λ. At what distance from the centre along the axis of the ring is the magnitude of the electric field a maximum?

PROBLEM 1.5 A thin sheet of charge is in the $x - y$ plane at $z = 0$. It is infinite but not uniformly charged. A circular piece of radius R with its centre at the origin has a surface charge density $+\sigma$ while the rest of the sheet carries a surface charge density $-\sigma$. Show that the electric field vanishes at points located at $(0, 0, +\sqrt{3}R)$ and $(0, 0, -\sqrt{3}R)$.
[*Hint: The sheet can be regarded as an infinite sheet of uniform surface charge density plus a circular sheet with a different uniform surface charge density*]

1.3 ELECTRIC FIELD, LINES OF FORCE

The Coulomb force that a charge of $+1$ unit experiences at any given point is called the **Electric Field** at that point. The electric field is clearly a vector, since force is a vector and both its magnitude and direction will in general be different at different points. In this chapter, we are considering the case of all charges at rest and so the electric field at any point will also be time independent.

The definition of electric field given above seems to be nothing new—one multiplies the electric field at any point by the charge present at that point to get the force on that charge. Thus, it might appear that the introduction of the concept of an electric field is somewhat trivial. However, there are good reasons for introducing this concept. As given by Coulomb's law, the force that a charged particle experiences at a given point is caused by charges present elsewhere. This kind of phenomenon is called **'action at a distance'**. This, as we will see later on, is not compatible with Special Theory of Relativity since it implicitly assumes that the effect of charges is propagated instantaneously to arbitrary distances which is prohibited by relativity, since no information can travel faster than the speed of light. If, on the other hand, charges are thought to create 'electric fields' everywhere, a charge at any point can be thought of experiencing an electric force due to the total electric field present at that point. This way of looking at electrostatic forces makes the phenomenon **'local'**, i.e., forces on a charged particle being caused by things present (in this case, the electric field) in its immediate neighbourhood.

The electric field introduced in such a way is obviously a vector field, i.e., to every point in space, is assigned a vector which gives the direction and magnitude of the electric field. This turns out to a huge simplification and the laws of electrostatics become much simpler stated in terms of fields rather than forces on charged particles. Once the electric field is specified in a region, it is unimportant what the sources of the field are. The behaviour of charges in that region then becomes much simpler to calculate. Moreover, when we consider the more realistic situation of charges in motion so that both electric and magnetic forces are present, fields appear to be much more real and the laws in terms of them, simple. As stated above, the Special Theory of Relativity will turn out to be compatible with the 'fields' causing forces locally rather than the 'action at a distance' implied in our statement for Coulomb's Law.

Having introduced the concept of electric field, it seems natural to graphically indicate them at every point by a vector showing its direction, as shown in Fig. 1.7a.

The vectors show only the direction and since the direction changes as we move from point to point, the vectors keep changing directions. If we make these vectors infinitesimally small and join them up at neighbouring points, we obviously get continuous curves as shown in Fig. 1.7b. The tangent to such a

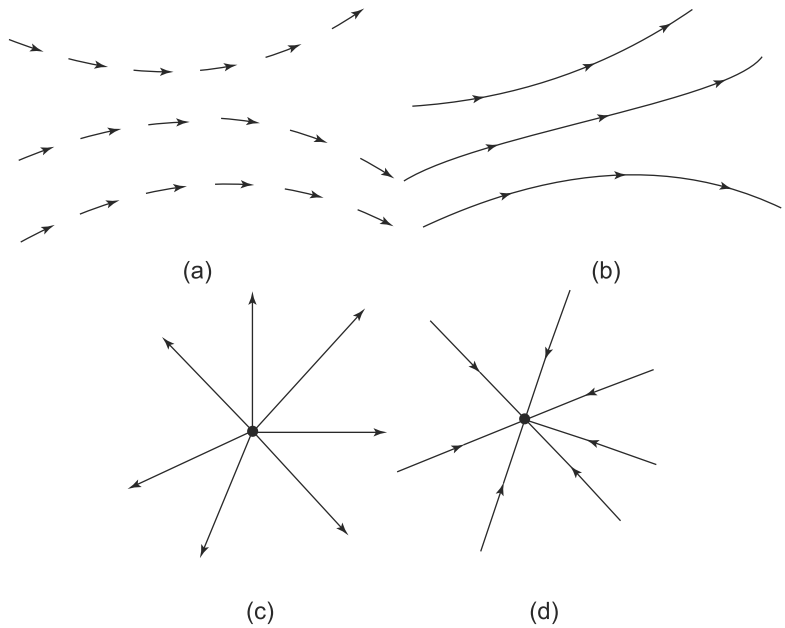


Fig. 1.7 *(a) Typical graphical representation of electric fields at various points, (b) Typical graphical representation of electric lines of force, (c) Outgoing electric lines of force from a positive point charge, (d) Electric lines of force from a negative point charge*

curve represents the direction of the electric field at that point. These lines are called the **electric lines of force**. Figures 1.7c and 1.7d show these lines of force in the presence of a single positive and negative point charge.

In the neighbourhood of a point charge, the electric field due to the point charge becomes very large compared to the field due to any other charges elsewhere. The electric field near a point charge hence, is radially outward due to a positive charge and radially inward due to a negative one. Such situations cannot arise otherwise and thus, electric lines diverging radially outward or converging radially inwards towards a point is a definite signature of a point charge being present there.

An electric field line at any point tells us about the direction of the electric field at that point. However, consider drawing a certain number of lines of force in the neighbourhood of a point charge. From Fig. 1.7c and Fig. 1.7d we see that if we follow the neighbouring lines of force and they diverge as we go along the direction of $\vec{E}$ (or opposite to it), the field gets weaker. Conversely, if the lines of force seem to converge in some direction, the electric field is getting stronger in that direction. **Thus, the density of field lines drawn is an indication of the strength of the electric field in that region, with regions of lower field having lines less dense than regions of relative higher values of field**.

EXAMPLE 1.5 A dipole consists of two charges $+q$ and $-q$ separated by a distance l as shown in Fig. 1.8(a). Draw the electric lines of force due to a dipole.

Fig. 1.8(a) *Example 1.5*

Solution

In the immediate neighbourhood of the point A, the field lines are radially outward and similarly near B, they are radially inwards. On the line joining A and B, the field is directed towards B. At any point P on the right bisector COD of the line AB, the fields due to charges at A and B are equal in magnitude, directed along PB due to charge at B and along AP due to charge at A. Adding these two, the net electric field is perpendicular to COD, directed towards the side where B is. The field lines thus, start radially outwards from point A, bend towards B as we move out from A, becoming parallel to AB on the right bisector COD. After that they continues bending towards B ultimately becoming radially inwards at B.

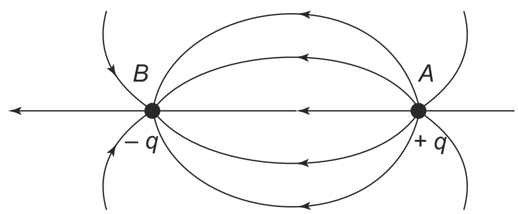


Fig. 1.8(b) *Example 1.5*

The resultant pattern is shown in Fig. 1.8(b).

EXAMPLE 1.6 Three charges $+q$, $+2q$ and $+q$ are located respectively at the three vertices A, B and C of an equilateral triangle of side l as shown in Fig. 1.9. Draw the lines of force in the region inside and on the triangle.

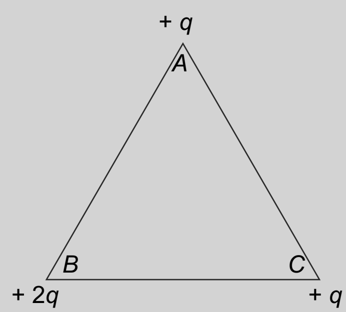


Fig. 1.9 *Example 1.6*

Solution

The direction of the electric field at some of the points is easily established. At the vertices, the field lines are all radially outwards. At the midpoint of AC, the electric field due to the two $+q$ charges cancel out, so that the field there is solely due to the $+2q$ charge and hence the field line is perpendicular to AC pointing out. At the midpoint of AB, the resultant of the field due to charge $+q$ at A and $+2q$ at B, points towards A and of net magnitude $k\frac{(2q-q)}{(l/2)^2}$. The field due to charge $+q$ at C has the a magnitude $\frac{4kq}{3l^2}$ and is perpendicular to AB, directed outwards. The resultant of all three fields thus, makes an angle to AB at the midpoint tilting towards A. The field at the midpoint of BC similarly makes an angle with

BC tilting towards C. At the centre O of the triangle, the electric fields due to charges at A and C each have a magnitude $\frac{3kq}{l^2}$ directed respectively along AO and CO. The components along OB of these add up while the components along a direction perpendicular to OB cancel. Thus, the resultant of these two fields is

$$\frac{3kq}{l^2}\cos 60 + \frac{3kq}{l^2}\cos 60 = \frac{3kq}{l^2} \tag{1.30}$$

along OB. The field due to charge $+2q$ at B is directed along BO of magnitude $\frac{3k2q}{l^2}$. Thus, the resultant of all the three fields has a magnitude $\frac{3kq}{l^2}$ directed along BO.

Starting from the three point charges, one can easily track down the lines of force as shown in Fig. 1.10.

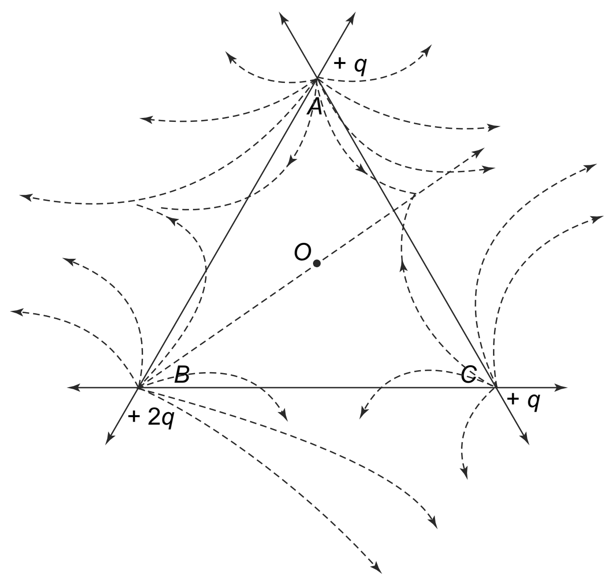


Fig. 1.10 *Example 1.6*

PROBLEM 1.6 P and Q are two similar dipoles each having point charges $+q$ and $-q$ at the ends separated by a distance d. P is along the x-axis and Q is along the y-axis and their centres are coincident. Sketch the nature of the lines of force created by the two dipoles.

1.4 ELECTRIC FLUX

A very useful concept in electrostatics is that of **Electric Flux**. In general, flux is a quantity which is associated with any kind of flow. Thus, for instance, the flow of water allows us to define a flux for the flow. Consider water flowing in a certain direction as shown in Fig. 1.11.

If the area is held along the velocity vector $\vec{v}$, i.e., is along the flow of the water, then it is clear that nothing flows through the area. On the other hand, if we hold the area such that the normal to the area

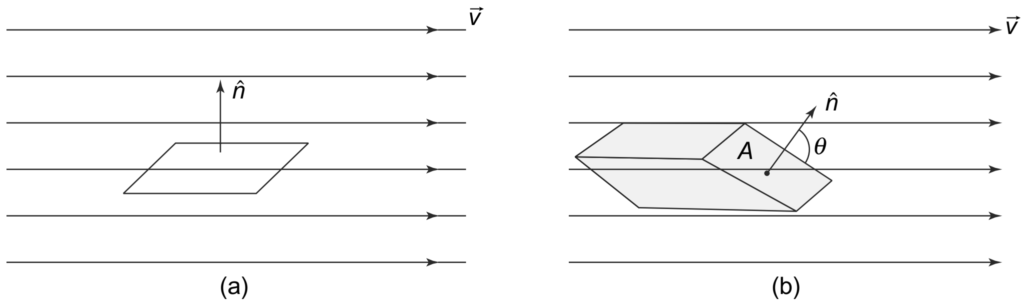


Fig. 1.11 *Water flowing with a velocity $\vec{v}$ and an area A with a normal $\hat{n}$ placed in the flow: (a) If the normal $\hat{n}$ is perpendicular to the flow, i.e., to $\vec{v}$ then no water flows through the area, (b) If $\vec{n}$ is inclined at an angle θ to the velocity vector, then the volume of water passing through A in 1 second is the quantity of water in a volume represented by the shaded area. This volume is $vA\cos\theta = \vec{v}\cdot\vec{A}$ where we have taken $\vec{A} = A\hat{n}$*

makes an angle θ with the flow of water (velocity $\vec{v}$), then a volume of water $\vec{v}\cdot\vec{A}$ passes through A in unit time. This quantity, the rate of flow of water through an area A is called the **flux** through A.

For the electric field, obviously, there is nothing that is flowing. However, we can still define an electric flux in essentially the same manner as for flowing water and this concept of flux proves to be a very useful one, as we shall see. Consider an infinitesimal area, as defined in Section 0.2.1, $d\vec{A} = dA\hat{n}$ where $\hat{n}$ is a unit vector along $d\vec{A}$. Let the electric field $\vec{E}$ be defined in the region where the infinitesimal area is located. Then, if $\vec{E}$ is the field at the location of dA, the electric flux across $d\vec{A}$ is defined as

$$\text{Flux across dA} = \vec{E}\cdot d\vec{A}$$

This flux is along $\hat{n}$. If $\vec{d}A$ is part of an area which is closed enclosing a volume, then the outward direction at any point on the surface is the direction going out of the volume at that point. The outward flux thus would be $\vec{E}\cdot\vec{d}A$ or its negative depending on whether $\hat{n}$ is in the outward direction or against it. In other cases, one is free to define which is the outward direction either $+\hat{n}$ or $-\hat{n}$.

If we want the flux across any finite areas, that can easily be obtained by integrating the flux across its infinitesimal parts. This is shown in Fig. 1.12.

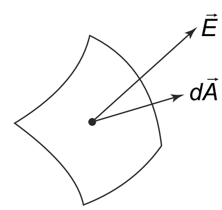


Fig. 1.12 *Electric flux through an infinitesimal area dA is defined as $\vec{E}\cdot d\vec{A}$*

There are some very interesting properties of the electric flux. If we think of flux as the rate of flow (of any vector field-water flowing, electric lines of force, etc.) across an area, then certain properties are obvious. For instance, in the case of water flowing, if there are no sources or sinks in a closed volume. Then the amount of water flowing in will equal the amount of water flowing out. For the closed surface which encloses the volume, the net flux in this case is zero. For electric fields, the situation is similar. Electric field is generated by electric charges. To see this, consider a point charge located at P outside a closed surface S enclosing a region of space which has no charges inside (Fig. 1.13).

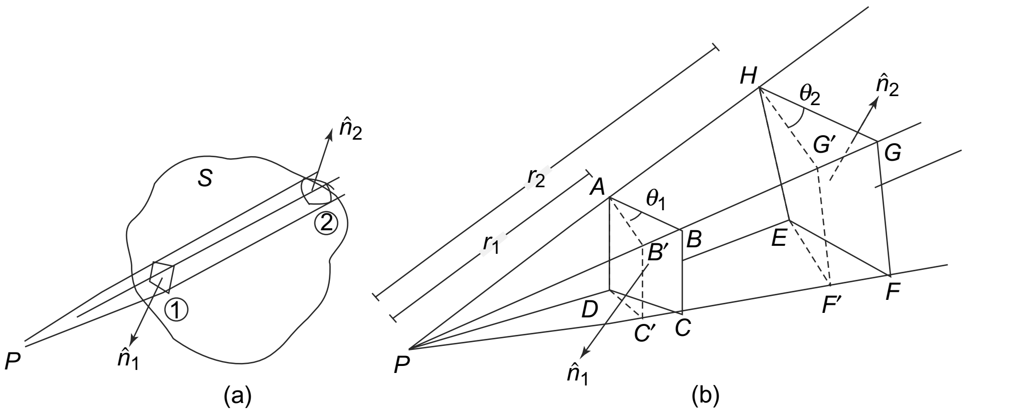


Fig. 1.13 *A point charge outside a closed surface S which encloses a charge free region: (a) A cone with P at the vertex cuts out rectangles $ABCD$ and $EFGH$ with outward normals $\hat{n}_1$ and $\hat{n}_2$, (b) The rectangles $ABC'D'$ is $ABCD$ projected normal to the direction of $\vec{E}$. Similarly, $EF'G'H$ is the projection of $EFGH$*

Draw an infinitesimal rectangle on S, $ABCD$ of area dA_1 which can be considered planar since it is infinitesimal. Join the sides of the rectangle to the point P, so that a cone is formed with P at the vertex. Extend the cone beyond the rectangle. Since S is a closed surface, the cone would cut out at least one more rectangular patch on S (Fig.(1.13)) of area dA_2. It may cut more patches, 3 or 5 or any odd number after the initial one. (Fig. 1.14).

Let us focus first on Fig. 1.13. Call the two patches 1 and 2 at distances r_1 and r_2 respectively from the point P. Let the magnitude of the electric field created by the point charge at 1 and 2 be denoted by E_1 and E_2 respectively. The two patches will generally not be normal to the direction of the electric field. We will consider flux in the outward direction, so that the component of the electric field along the normal to the patches 1 and 2 are $-E_1 \cos\theta_1$ and $E_2 \cos\theta_2$ respectively where the angles θ_1 and θ_2 are shown in Fig. 1.13(b). The outward electric flux Φ at the two patches are therefore,

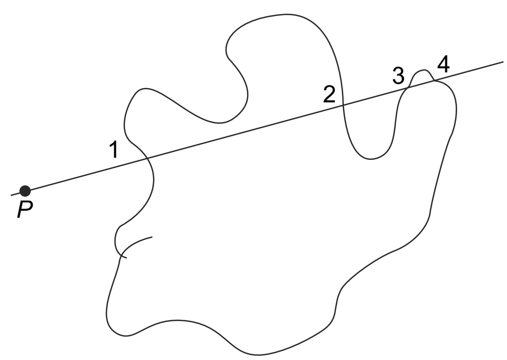


Fig. 1.14

At patch 1 and 2:

$$
\begin{aligned}
\Phi_1 &= -E_1 dA_1 \cos\theta_1 \\
\Phi_2 &= +E_2 dA_2 \cos\theta_1
\end{aligned}
\tag{1.31}
$$

Let us now project the two patches onto a plane normal to the direction of the electric field, resulting in patches $AB'C'D$ and $EF'G'H$, both normal to $\vec{E}$.

$AB'C'D$ and $HG'F'E$ are both normal to $\vec{E}$ (Fig. 13b). The area of these patches $AB'C'D$ and $HG'F'E$ are $dA_1 \cos\theta_1$ and $dA_2 \cos\theta_2$ respectively. Next, consider triangles PAB' and PHG' in Fig. 1.13. They are similar and hence, the length of the sides AB' and HG' are in the ratio $r_1 : r_2$. Similarly, the sides AD and HE are also in the same ratio. Hence, the areas of the two rectangles $AB'C'D$ and $HG'F'E$ are in the ratio $r_1^2 : r_2^2$.

On the other hand, according to Coulomb's Law, the magnitude of the electric field at patches 1 and 2, namely E_1 and E_2, are in the ratio, r_2^2/r_1^2. The two fluxes given in Eq. (1.31) are therefore, equal and opposite. The total outward flux through these two patches, which is the sum of these two therefore, vanishes.

The entire closed surface can be broken up into pairs of patches similar to the pair considered and the total outward flux through them will also vanish. If we have a surface such that the cone from P cuts out at S a total of more than two patches, 4 or 6 or some even number (Fig. 1.14), we can apply exactly the same arguments as before for these pairwise, patches 1 and 2 and patches 3 and 4 in the Fig. 1.12. They will all add upto zero as in this case.

We have considered the flux due to a single point charge in the arguments above. The total flux across any surface due to a set of point charges, however, is simply the sum of the individual fluxes due to each one of them considered separately. This is simply because the electric field due to a set of point charges is the vector sum of the electric field created by them separately, by the Principle of Superposition stated above. The arguments presented above apply to the electric flux created by each of the point charges located outside S and hence, also applies to the total flux. We therefore, have the result:

> 'The total outward flux across a closed surface due to charges present outside it, vanishes.'

PROBLEM 1.7 A hemispherical surface of radius R has its rim on the $x-y$ plane with the origin as its centre. A point charge $+q$ is present at $(0, 0, d)$ and similarly a point charge $-q$ is located at $(0, 0, -d)$. Calculate the flux through the hemispherical surface.

1.5 GAUSS'S LAW

Consider first a simple case of a single point charge q situated at the centre of a closed spherical surface of radius R (Fig. 1.15).

The electric field created by q at the surface is outward and has a magnitude $\frac{kq}{R^2}$. The surface of the spherical surface has an area $4\pi R^2$ and the outward normal is everywhere is in the direction of the electric field. Hence, the total outward electric flux across the surface is

$$\Phi = \frac{kq}{R^2} 4\pi R^2 = 4\pi kq. \tag{1.32}$$

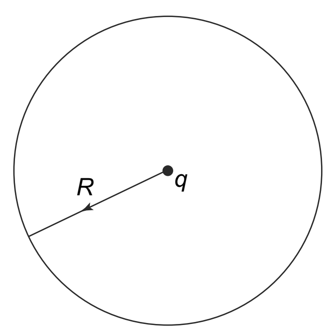


Fig. 1.15 *A point charge at the centre of a closed spherical surface*

Using the result of the previous section, we will now be able to prove that Eq. (1.32) is more generally valid for closed surface of any shape with the charges situated anywhere inside it.

Consider now an arbitrary closed surface S of any shape and let there be any number of point charges inside it. Let us focus on any one of them of, charge q. Draw a sphere, centred at q and of sufficiently small size such that there are no other charges within it (Fig. 1.16). Imagine scooping out this sphere from the region enclosed by the surface S. We are now left with a new region that has two closed surfaces namely, the (i) original surface S and (ii) the small spherical surface inside. For this region, outward flux at the small spherical surface inside it is $-\Phi$, i.e., it has the same magnitude as Φ in (1.32) but has a negative sign since the outward direction there is opposite the direction of electric field. For this new region having two closed surfaces, the charge q is outside. Hence, from the result proved in the last section, the total outward flux due to q across this region vanishes, i.e., the flux across the small spherical surface (which will have a negative flux). The outward flux across the original surface S therefore is just Φ, given by Eq. (1.32).

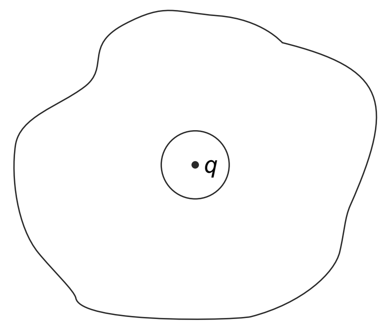


Fig. 1.16 *A point charge q inside a surface. A small spherical volume around q is scooped out*

This result holds independently for flux created for any point charge located inside it. If S encloses many point charges, the total flux across it will simply be the flux created by the individual charges. Hence, the total flux will be given by (1.32) with q replaced by the total charge. We therefore, have the result:

> The total outward electric flux across any closed surface enclosing a charge Q is $4\pi k Q$.

This important result is called **Gauss's law.**

We can now verify the statement made earlier regarding the converging and diverging of electric field lines and its relation to the magnitude of the electric field. Consider a case where the electric field lines appear to diverge from each other (Fig. 1.17).

Draw an area A_1 at a point 1 normal to the field lines. Follow **all** the lines of force crossing A_1 to another point 2 where draw another area A_2 once again normal to the field lines at 2. Now consider the closed surface formed by the two areas A_1 and A_2 and the curved surface. Since there are no charges inside the volume enclosed by this closed surface,

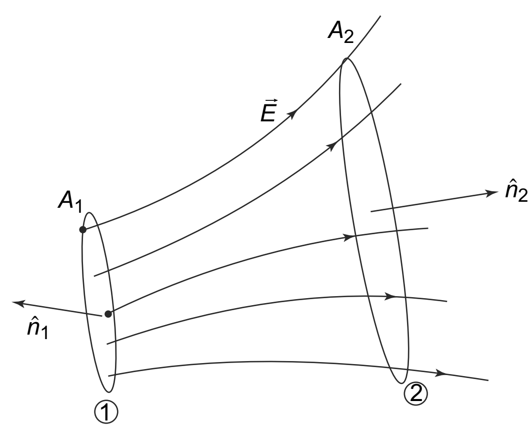


Fig. 1.17 *Field at point 1 is stronger than field at point 2*

by Gauss's Law, the net electric flux across this closed surface vanishes. The electric flux across the closed surface has three parts—the flux across A_1, flux across A_2 and the flux across the curved surface. The electric field is along the curved surface and hence the outward normal is perpendicular to the electric field at every point, which means that $\vec{E} \cdot d\vec{A} = 0$ for the curved surface. For the area A_1, the flux is $\hat{n}_1 \cdot \vec{E}_1 A_1$ and the flux across A_2 is $\hat{n}_2 \cdot \vec{E}_2 A_2$. But the normal $\hat{n}_1$ is opposite to $\vec{E}_1$ and hence, the flux across A_1 is $-E_1 A_1$ while the flux across A_2 is $E_2 A_2$. By Gauss's Law, the sum is

$$-E_1 A_1 + E_2 A_2 = 0$$

Thus, we see that if the lines are diverging, then since $A_2 > A_1$, $E_2 < E_1$. On the other hand, if the lines are converging towards 2, then $A_2 < A_1$ and consequently $E_2 > E_1$. This is what we had asserted in our discussion of electric lines of force.

Gauss's Law is very useful in determining the fields from charge distributions in certain situations. Of course, one can always use Coulomb's Law and the Principle of Superposition to determine the forces on charges (and hence, the fields) from charge distributions. However, in certain cases where there is symmetry, Gauss's Law proves to be far more convenient. Remember that Gauss's Law is valid in all situations. Whether it can be used easily to determine the electric field at any point depends on whether there is enough symmetry in the problem to be able to easily evaluate the flux integral in terms of the unknown field. Then Gauss's Law relates this to the charge inside the closed surface and allows us to easily determine the electric field. We illustrate this with some examples of the use of Gauss's Law in certain situations with symmetry.

EXAMPLE 1.7 Calculate the electric field due to an uniformly charged sphere of radius R carrying a charge Q, at a point P situated at a distance $r > R$ from its centre, as shown in Fig. 1.8.

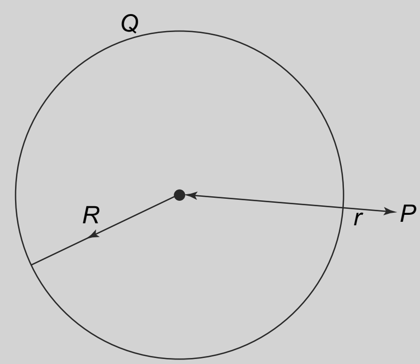


Fig. 1.18 *Example 1.7*

Solution

Draw a spherical surface of radius r with its centre. The given point obviously lies on it. We draw a straight line from P passing through the centre of the charged sphere. This line is radial. Now, imagine rotating the sphere about this line. Nothing changes, since the sphere is uniformly charged. Hence, all directions perpendicular to this radial line are equivalent. The field at P thus, cannot have a component perpendicular to the radial direction, since otherwise the field would have to single out one particular such direction. This applies equally well to any point on the spherical surface drawn and hence, the field everywhere on it is radial. Further, all points on the drawn surface are exactly similarly situated with respect to the charged sphere. Hence, its magnitude is the same everywhere on it. Call its magnitude $E(r)$. The flux across the drawn sphere then is

$$E(r) \times \text{ area of the spherical surface} = 4\pi r^2 E(r) \tag{1.33}$$

By Gauss's Law, this should be related to the total charge inside the drawn surface, which is Q,

Thus,

$$4\pi r^2 E(r) = 4\pi k Q \tag{1.34}$$

or

$$E(r) = \frac{kQ}{r^2} \tag{1.35}$$

which is the same value if the entire charge was concentrated at the centre of the sphere.

EXAMPLE 1.8 Consider an infinite region bounded by two planes separated by a distance d measured normally to the planes. The region is filled uniformly with charges with a density ρ. Calculate the electric field at a point P located at a perpendicular distance r from the planes when

(a) P is located outside and

(b) P is located inside and $r < d/2$.

This is shown in Fig. 1.19.

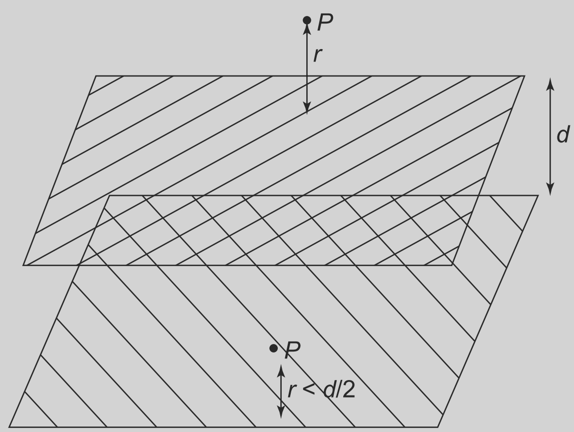


Fig. 1.19 *Example 1.8*

Solution

Draw a plane parallel to the two planes midway between them. By symmetry, the field is everywhere normal to this planes and oppositely directed on its two sides. Also, at a given perpendicular distance from this surface, the field has the same value on either side.

(a) Draw a cylinder symmetrically about the central plane drawn, such that that the two planar ends of the cylinder each have unit area and each is at a distance r from the nearest plane. On the surface of this cylinder, there is no flux across the curved surface since the field is parallel to it. Let $E(r)$ be the field at any of the two flat ends. The fields there are normal to the surface and hence, the flux across them is

$$E(r) \times \text{ area of the flat ends} = 2E(r) \tag{1.36}$$

The charge inside this area is ρd. Hence, by Gauss's Law:

$$2E(r) = 4\pi k\rho d \tag{1.37}$$

or

$$E(r) = 2\pi k\rho d \tag{1.38}$$

(b) This is exactly as in (a), except that the charge enclosed by the cylindrical closed surface is $\rho(d - 2r)$. We thus, get

$$E(r) = 2\pi k\rho(d - 2r) \tag{1.39}$$

EXAMPLE 1.9 Find the electric field inside and outside of

(a) a conducting sphere of radius a and total charge Q, given that the charge resides uniformly entirely on the surface,
(b) a sphere of radius a and total charge Q spread uniformly within it and
(c) a sphere of radius a, charge Q and charge density which varies as r^n, $(n > -3)$.

Solution

In all the three cases, the charges are completely spherically symmetric, i.e., the charge densities depend only on r and not on θ or ϕ. Hence, in all cases the electric field is along $\hat{r}$ and depends only on r. For all the three cases, we draw a sphere of radius r. The flux through this sphere will be $E(r)4\pi r^2$.

In case 1, for a point inside at distance r from the centre, by Gauss's Law this should equal $4\pi k\times$ charge inside sphere of radius r. In this example, we use SI units and take $k = \frac{1}{4\pi\varepsilon_0}$. Hence, for the conducting sphere, at a point inside, the field will vanish since all the charge resides on the surface of the conducting sphere. For a point outside, we can use a spherical Gaussian surface and find that the field is given by

$$\vec{E} = \frac{Q}{4\pi\varepsilon_0 r^2}\hat{r}$$

For the sphere with a uniform charge density, the field inside can be seen to be

$$\begin{aligned}\vec{E}(r) &= \frac{4\pi k}{4\pi r^2} \times \text{ charge inside the sphere of radius } r\\ &= \frac{4\pi k}{4\pi r^2}\left(\rho\frac{4\pi r^3}{3}\right)\\ &= \frac{\rho r}{3\varepsilon_0}\hat{r}\end{aligned}$$

and the field outside is simply as if all the charge was at the centre of the sphere

$$\vec{E} = \frac{Q}{4\pi\varepsilon_0 r^2}\hat{r}$$

For the third case, for the field inside, we take a spherical Gaussian surface centred at the centre of the sphere and radius r and get

$$E4\pi r^2 = \frac{\int \rho 4\pi r'^2 dr'}{\varepsilon_0}$$

which gives us, writing $\rho = \rho_0 r^n$,

$$E = \frac{1}{\varepsilon_0 r^2} \int \rho_0 r'^{n+2} dr'$$

where k is some constant. Integrating, we get

$$E = \frac{1}{\varepsilon_0} \frac{\rho_0 r^{n+1}}{(n+3)}$$

For the point outside, we get

$$E = \frac{\rho_0}{\varepsilon_0 r^2} \frac{a^{n+3}}{n+3}$$

EXAMPLE 1.10 A long cylinder of radius R carries a volume charge density that is proportional to the distance from the axis, given by $\rho = \chi r$. Find the electric field at a point distance $a < R$ from the axis.

Solution

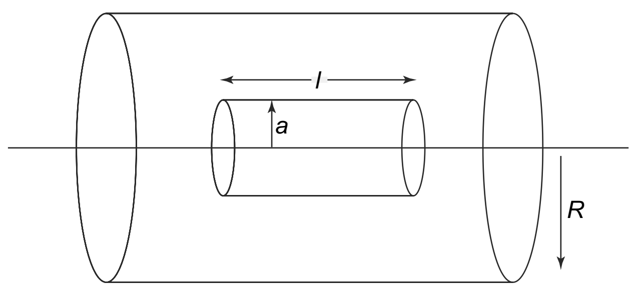


Fig. 1.20 *Example 1.10*

Since the geometry of the problem has symmetry, it is convenient to use Gauss's Law. In this case, we can draw a cylindrical Gaussian surface of length l and radius a as shown in the Fig. 1.20. Using SI units, Gauss's Law then states

$$\oint \vec{E} \cdot \vec{d}S = \frac{Q_{\text{inside}}}{\varepsilon_0}$$

where Q_{inside} is the charge enclosed by the Gaussian surface. This can easily be computed using cylindrical coordinates.

$$Q_{\text{inside}} = \iiint \rho dV = \int\int\int (\chi r) r dr d\phi dz = 2\pi\chi l \int_0^a r^2 dr = \frac{2}{3}\pi\chi l a^3$$

Symmetry also tells us that the field at the Gaussian surface must be radially outward and hence, the left-hand integral in Gauss's Law above can be easily evaluated. The only contribution thus, comes from the curved surface of the Gaussian cylinder while the ends contribute nothing. Therefore, Gauss's Law becomes

$$\iint \vec{E} \cdot dS = E \iint dS = E2\pi al$$

But this is also equal to the $\dfrac{Q_{\text{inside}}}{\varepsilon_0}$, and so we get

$$E = \frac{\chi a^2}{3\varepsilon_0}$$

and

$$\vec{E} = \frac{\chi a^2}{3\varepsilon_0}\hat{r}$$

PROBLEM 1.8 A point charge $+q$ is present at the origin O of a coordinate system. $OABCDEFG$ is a cubical surface of size L with edges OC, OG and OA along the x, y and z-axis respectively as in Fig. 1.21. Calculate the flux across the face $BEFC$ of the cubical surface. [*Hint: Consider similar cubes in the seven other quadrants; now use symmetry and Gauss's theorem*]

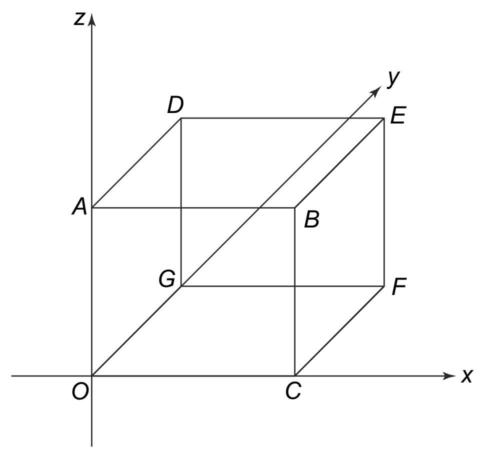


Fig. 1.21 *Problem 1.8*

PROBLEM 1.9 A spherical charge distribution has a volume charge density $\rho(r)$ given by

$$\begin{aligned}\rho(r) &= q\left(1 - \frac{5r^2}{R^2}\right) \quad r \le R \\ &= 0 \quad r > R\end{aligned}$$

Calculate the electric field at points with $r > R$ and also $r < R$. If we consider a cubical surface with length of sides $5R$ and with its centre at $r = 0$, what would be the flux across one of its faces? [*Hint: Use symmetry and Gauss's law*]

From the examples above, we see the immense usefulness of Gauss's Law for finding the electric field in a region, in cases of a high degree of symmetry. The strategy then is as follows—if the physical situation has a degree of symmetry, then first try to see if the flux integral in Gauss's Law, i.e.,

$$\iint_S \vec{E}.d\vec{S}$$

can be evaluated easily using the symmetry. This would obviously mean that the electric field's direction and magnitude are such that the integral is easily solvable. Then relating the flux integral to the charge enclosed by the surface (which is normally known) would allow us to evaluate the electric field in any region.

1.6 A DIFFERENT FORM OF GAUSS'S LAW

Gauss's Law as presented above involves surface integrals. It is also possible to present the law not involving integrals but derivatives, and this turns out to be very useful, as we shall see later on. We shall do that in this section.

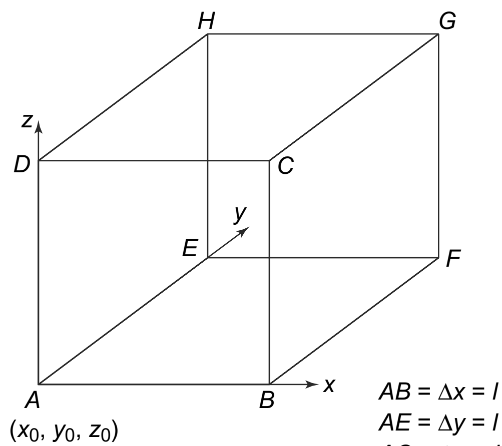


Fig. 1.22

Consider an infinitesimally small cube $ABCDEFGH$ of side l with A located at (x_0, y_0, z_0). The sides AB, AE and AD are parallel to the three axes and have infinitesimal lengths Δx, Δy and Δz, respectively. The electric charge density and the electric field inside the cube at any point with coordinates (x, y, z) are denoted by $\rho(x, y, z)$ and $E(x, y, z)$, respectively. We denote by $\hat{i}$, $\hat{j}$ and $\hat{k}$ the unit vectors respectively along x, y and z axes. This is depicted in Fig. 1.22.

The outward normal $\hat{n}$ to the face $ABCD$ is directed along the direction $-\hat{j}$. The scalar product $\vec{E} \cdot \hat{n}$ thus, is $-E_y$. All points on $ABCD$ have the same y coordinate as the point A, i.e., y_0. The outward flux across it therefore, is

$$-\iint dxdzE_y(x, y_0, z) \tag{1.40}$$

where the limits of integration are $x_0 < x < x_0 + \Delta x$ and $z_0 < z < z_0 + \Delta z$.

We can similarly calculate the outward flux across the face $EFGH$. The outward normal will now be along $\hat{j}$. All points on $EFGH$ have a y-coordinate $y_0 + \Delta y$. Hence, similar to Eq. (1.40), the outward flux across $EFGH$ is

$$\iint dx dz E_y(x, y_0 + \Delta y, z) \tag{1.41}$$

with integration limits exactly as in Eq. (1.40). The sum of the two fluxes therefore, is

$$\iint dx dz [E_y(x, y_0 + \Delta y, z) - E_y(x, y_0, z)] \tag{1.42}$$

Since Δy is infinitesimal, the quantity in square brackets is by definition

$$\left[\frac{\partial E_y(x, y_0, z)}{\partial y}\right] \Delta y$$

Furthermore, $E_y(x, y_0, z)$ in the region of integrations in (1.42) differs infinitesimally from $E_y(x_0, y_0, z_0)$ because x and z in that region differ from x_0 and z_0 by an infinitesimal amount. Hence, the net flux across the two faces $ABCD$ and $EFGH$ is

$$\left[\frac{\partial E_y(x_0, y_0, z_0)}{\partial y}\right] \Delta y \iint dx dz = \left[\frac{\partial Ey(x_0, y_0, z_0)}{\partial y}\right] \Delta y \Delta x \Delta z \tag{1.43}$$

Following exactly similar reasoning, the net outward flux across the faces $ABFE$ and $DCGH$ is given by

$$\left[\frac{\partial E_z(x_0, y_0, z_0)}{\partial z}\right] \Delta y \Delta x \Delta z$$

while across the two faces $BCGF$ and $AEHD$ it is

$$\left[\frac{\partial E_x(x_0, y_0, z_0)}{\partial x}\right] \Delta y \Delta x \Delta z$$

Adding, the net flux across all the faces of the cube is given by

$$\left[\frac{\partial E_x(x_0, y_0, z_0)}{\partial x} + \frac{\partial E_y(x_0, y_0, z_0)}{\partial y} + \frac{\partial E_z(x_0, y_0, z_0)}{\partial z}\right] \Delta y \Delta x \Delta z \tag{1.44}$$

The quantity inside the square bracket of Eq. (1.44) is something that is familiar to us from vector analysis and is called the divergence of $\vec{E}$

$$\vec{\nabla} \cdot \vec{E} = \text{div}\vec{E}(x_0, y_0, z_0) = \left[\frac{\partial E_x(x_0, y_0, z_0)}{\partial x} + \frac{\partial E_y(x_0, y_0, z_0)}{\partial y} + \frac{\partial E_z(x_0, y_0, z_0)}{\partial z}\right]$$

Also, note that

$$(\Delta x \Delta y \Delta z) = \Delta V$$

the infinitesimal volume of the cube. The net outward flux across the cube is thus,

$$\vec{\nabla} \cdot \vec{E}(x_0, y_0, z_0) \Delta V$$

The charge density ρ can vary with the coordinates. But every point inside the cube is separated by infinitesimal amounts from the coordinates of A, i.e., (x_0, y_0, z_0). Thus, neglecting infinitesimal quantities which will vanish in the limit, the charge inside the cube is

$$[\rho(x_0, y_0, z_0)\Delta V]$$

By Gauss's Law, then

$$\vec{\nabla} \cdot \vec{E}(x_0, y_0, z_0)\Delta V = 4\pi k[\rho(x_0, y_0, z_0)\Delta V] \tag{1.45}$$

The coordinates (x_0, y_0, z_0) were completely arbitrary. Hence, we obtain from Eq. (1.45)

$$\boxed{\vec{\nabla} \cdot \vec{E} = \text{div } \vec{E}(x, y, z) = 4\pi k\rho(x, y, z)} \tag{1.46}$$

Equation (1.46) is the differential form of Gauss's Law. The merit of this way of writing Gauss's Law is clear. The electric field at any point is related to the charge density at the same point. The feature of 'action at a distance' present in the original form of Coulomb's Law thus, is no longer present in this form.

PROBLEM 1.10 In a charge free region electric field lines are given to be parallel. Show that, starting from any point in that region, the strength of the electric field cannot increase or decrease as we go in the direction of the field as a consequence of Gauss's law.

PROBLEM 1.11 The electric field is given to be radial everywhere having values

$$\begin{aligned} |\vec{E}| &= E_0 e^{(\alpha(R-r))} \quad & r > R \\ &= E_0 e^{(\alpha(r-R))} \quad & r < R \end{aligned}$$

What is the charge density and the total charge giving rise to such a field?

1.7 ADVANCED TOPICS

1.7.1 Electrostatics in Two Dimensions

In 1884, an English schoolteacher, Edwin Abbot Abbot published a satirical novel called '*Flatland*' where he invented a two-dimensional flat world. Although the novel was a satire on Victorian England, it was a pioneering effort in actually conceiving of worlds in dimensions different from the ordinary three dimensions of space that we are familiar with.

In 1979, Canadian computer scientist A.K. Dewdney published a monograph called 'Two-dimensional Science & Technology' in which he considered how the various laws of physics would be in a two-dimensional universe. Let us try and see how electrostatics as we know, will look like, if the universe was flat, i.e., instead of three dimensions, we only had two.

To understand how fundamentally different things are in this fictional universe, we need to first define certain mathematical objects, which, as we shall see, behave very differently from their three-dimensional counterparts.

The concept of a vector in two dimensions is easy to think of—it is an object which will have two components instead of three. But what about operations on vectors? We are familiar with the concept of a scalar or dot product between two vectors as well as the notion of cross or vector product. How can we define these in two dimensions?

Recall that for two vectors $\vec{a}$ and $\vec{b}$, the scalar product is defined as

$$\vec{a} \cdot \vec{b} = \sum a_i b_i = a_1 b_1 + a_2 b_2 + a_3 b_3$$

where i runs from 1 to 3 in three dimensions. This same definition can be extended to two dimensions and so we have

$$\vec{a} \cdot \vec{b} = \sum a_i b_i = a_1 b_1 + a_2 b_2$$

What about the vector product? This involves some modification, since the usual vector or cross product in three dimensions is a vector that is perpendicular to **both** $\vec{a}$ and $\vec{b}$.

$$\vec{a} \times \vec{b} = |\vec{a}||\vec{b}| \sin\theta \hat{c}$$

where θ is the angle between the two vectors and $\hat{c}$ is a unit vector perpendicular to the plane containing $\vec{a}$ and $\vec{b}$. But how does one define a vector perpendicular to a plane when the universe itself is two-dimensional? For this, we use the concept of rotation of a vector. The rotation operator acts on a vector and transforms the vector into another vector, which is rotated by some angle. For instance, the rotation by an angle θ can be achieved by the following transformation:

$$\vec{a} = (a_1, a_2)$$

and

$$\mathbf{R}_\theta = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \tag{1.47}$$

and the transformed vector is given by

$$\vec{a}' = \mathbf{R}_\theta \vec{a}$$

As a special case, we consider rotation by an angle $-\pi/2$. Then

$$\vec{a}^* \equiv \mathbf{R}_{-\pi/2} \vec{a}$$

With this, we can define the vector product in two dimensions as

$$\vec{a} \otimes \vec{b} = \vec{a} \cdot \vec{b}^* = a_1 b_2 - a_2 b_1$$

With these concepts, we can also introduce the derivative or gradient operator in two dimensions as follows:

$$\vec{\partial} = \left(\frac{\partial}{\partial x}, \ \frac{\partial}{\partial y} \right) \tag{1.48}$$

and similarly, the operator ∂^* which is obtained by applying the transformation defined above to ∂

$$\vec{\partial}^* = \left(\frac{\partial}{\partial y}, \ -\frac{\partial}{\partial x} \right) \tag{1.49}$$

Now we have operations on vectors and also vector fields defined, what about the theorems of Gauss and Stokes, which we are familiar with in three dimensions?

There is obviously no concept of a surface (as in the directed surface $\vec{d}S$) that we are familiar with from three dimensions since the space itself is two-dimensional. Instead, we replace the concept of the closed surface with that of a closed contour and so we can generalise the Divergence Theorem

$$\oint_c \vec{a} \otimes \vec{d}r = \oint_c \vec{a} \cdot \hat{n} dl = \int_S (\vec{\partial} \cdot \vec{a}) ds \tag{1.50}$$

and Stokes' Theorem as

$$\int_S (\vec{\partial} \otimes \vec{a}) ds = \oint_c \vec{a} \cdot \vec{d}r \tag{1.51}$$

We assume that the conservation of charge is valid in two dimensions and a point charge q, produces an electric field $\vec{E}$, which gives rise to an electric force on a test charge q' given by

$$\vec{F} = \vec{E} q'$$

In addition, we assume that the Principle of Superposition is valid in two-dimensions and that the force between two point charges is along the line joining the two. We also assume that Gauss' Law is valid in two dimensions in the sense that the flux through a closed contour only depends on the charge contained within the contour (in analogy to the three dimension case, where it depends on the charge contained within a closed surface). Thus, we have

$$\oint_c \vec{E} \otimes \vec{d}r = \oint_c \vec{E} \cdot \hat{r} dl = \frac{Q}{\varepsilon_0} \tag{1.52}$$

This then immediately yields, assuming symmetry, Coulomb's Law in two-dimensions

$$\vec{E}(\vec{r}) = \frac{Q}{2\pi \varepsilon_0 r} \hat{r} \tag{1.53}$$

Note that the dependence on r has changed from an inverse square dependence and this brings about radical changes in electrostatics. As an example, let us consider how the field of an infinitely long line charge will look in two dimensions. This is also shown in Fig. 1.23.

In three dimensions, we know that the field at a distance r from the line is given by

$$E = \frac{Q}{2\pi \varepsilon_0 r}$$

In two dimensions, the situation is very different. To apply Gauss's Law in two dimensions, consider a line charge with a uniform charge density λ. The analogue of the closed surface that we use in three dimensions (in this case, a cylinder of radius r), is a closed contour. The electric field is perpendicular to the line charge (say, along the x-direction) and hence, applying Gauss's Law given above, we get

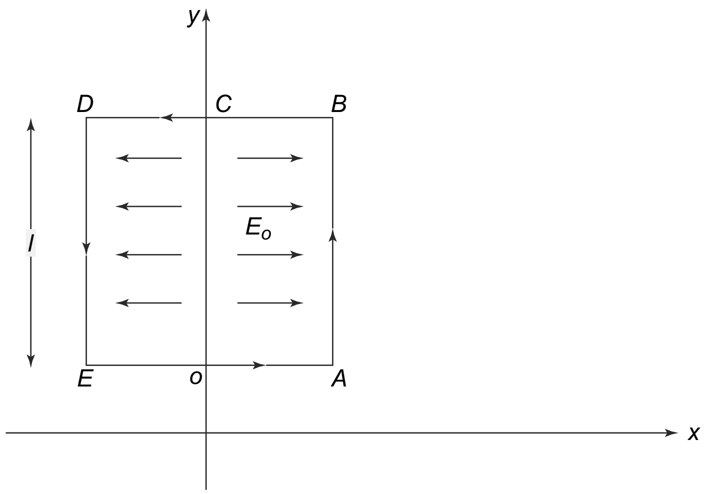


Fig. 1.23 *Line charge in two dimensions*

$$\oint_c \vec{E} \otimes \vec{d}r = (E_x dy - E_y dx) \tag{1.54}$$

$$= 2E_0 l \text{ (since } \vec{E} = E_0 \hat{i}) \tag{1.55}$$

$$2E_0 l = \frac{\lambda l}{\varepsilon_0} \tag{1.56}$$

$$E_0 = \frac{\lambda}{2\varepsilon_0} \tag{1.57}$$

Notice that the electric field is independent of the distance from the line charge. This is expected, since the electric field (Coulomb's Law) is very different in two dimensions.

What about electric potential which we will study in the next chapter? The concept of potential in electrostatics is possible because the electric field in three dimensions is conservative, i.e., the work done is independent of the path taken. Put another way, using Stoke's Theorem, we can show that this implies

$$\vec{\nabla} \times \vec{E} = 0$$

and hence,

$$\vec{E} = -\vec{\nabla}\phi$$

where ϕ is the potential. In two dimensions, since the dependence of $\vec{E}$ on r has changed, we need to first verify if the electric field is conservative, since it is only then that a potential is meaningful.

Let us consider the example of the infinite line charge that we studied above. In Fig. 1.23, the contour C is given by the rectangle $ABDE$. Now

$$\oint_c \vec{E} \cdot \vec{d}r = \int_c (E_x dx + E_y dy) \tag{1.58}$$

$$= \int_0^A E_x dx + \int_B^C E_x dx + \int_C^D E_x dx + \int_E^0 E_x dx$$
$$= 0 \tag{1.59}$$

In fact, it is possible to show that

$$\vec{\partial} \otimes \vec{E} = 0$$

and so we can define

$$\vec{E} = -\vec{\partial}\Phi$$

Applying this to the field of a point charge in two dimensions (1.53), we get

$$\Phi = -\frac{Q}{2\pi\varepsilon_0} \ln r$$

instead of the $1/r$ potential that we get in three dimensions. This change in the dependence also has profound consequences for the definition of the potential since in three dimensions, we can put the potential at infinity to be zero, which is not possible in two dimensions.

For more details, see **E Moreno and J Rivas, *Eur J. Phys,* 5 (1984)**.

SUMMARY

- Principle of conservation of charge, which states that the total electric charge, which is the algebraic sum of the positive and negative charge, in an isolated system does not change at any time.
- Quantisation of charge, which is the statement that all charges are simple integral multiples of a basic unit of charge, namely the charge of the electron.
- The basic law of force between two point charges called ‘Coulomb’s Law’. This law implies ‘action at a distance’, i.e., a charge experiences electrostatic force due to other charges removed from it.
- Principle of Superposition which states that forces due to many point charges on a given charge add up vectorially.
- The concept of electric field. A charged particle creates this field all around it. Any other charge experiences an electrostatic force due to the electric field in its immediate neighbourhood. This way of looking at electrostatic forces eliminates the ‘action at a distance’ aspect of Coulomb’s Law.
- Concept of Electric Flux across a surface. This is used to prove ‘Gauss’s Law’, which relates the total flux across a closed surface to the total charge inside it, irrespective of actual location.

CONCEPTUAL QUESTIONS

1. Two point charges $Q_1 = 1$ μC (1 $\mu C = 10^{-6}$ C)and $Q_2 = 2$ μC are placed distance $r = 1$ m apart. Which of the following statements are true?

a. The force on Q_1 is repulsive.
b. The force on Q_2 is the same magnitude as on Q_1.
c. As the distance r decreases, the force on Q_1 increases linearly.
d. The force on Q_2 is along the line joining the two charges.
e. A point charge $Q_3 = -3$ μC placed at the midpoint between the two charges experiences no net force.

2. Give some examples of quantities which are conserved in nature. Which of these quantities are also quantised?

3. a. Why do not the electric lines of force ever cross each other?
b. Can electric lines of force be all radial at a point with some going out and some going in?
c. Can electric lines of force be in the form of a closed ring?

4. The electric flux density on a spherical surface $r = r_1$ is the same for a point charge Q located at the origin and for charge Q uniformly distributed on the surface $r = r_2 (r_2 < r_1)$.
a. Yes
b. No
c. Not necessarily

5. The electric field in a region is given by

$$\vec{E} = 10\hat{e}_r + 5\hat{e}_\theta + 3\hat{e}_\phi$$

Find the electric flux passing through the surface which is bounded by the region $z \geq 0$ and $x^2 + y^2 + z^2 = 36$.

6. Suppose we have a Gaussian surface which encloses no net charge. Can we conclude only from this that the electric field is 0 at all points on the Gaussian surface? Conversely, suppose the electric field is 0 at all points on a Gaussian surface. Can we conclude that the net charge inside is 0?

7. Two infinite parallel plates carry equal but opposite uniform charge densities, $\pm\sigma$. Find the field to the left of both, to the right of both and in between them.

8. If the electric field $\vec{E}$ in a region is given by

$$\vec{E} = (2y^2 + z)\hat{i} + 4xy\hat{j} + x\hat{k}$$

find
a. the volume charge density at the point $(-1, 0, 3)$.
b. The total charge enclosed by the cube $0 \leq x \leq 1, 0 \leq y \leq 1, 0 \leq z \leq 1$.

9. Point charges 30 n C (1 nC = 10^{-9} C), -20 n C and 10 n C are located at $(-1, 0, 2)$, $(0, 0, 0)$ and $(1, 5, -1)$, respectively. The total flux leaving a cube of side 6 m centred at the origin is
a. -20 n C
b. 10 n C
c. 20 n C
d. 30 n C
e. 60 n C

10. The electric flux through a spherical surface of radius R centred on a charge Q is Φ. The electric flux through a cube of side R centred on the same charge Q is

a. Less than Φ
b. More than Φ
c. Equal to Φ
d. Cannot be determined

PROBLEMS

1. Three charges of 2 nC, -5 nC and 0.2 nC are located at points $A(2, \frac{\pi}{2}, \frac{\pi}{4})$, $B(1, \pi, \frac{\pi}{2})$ and $C(5, \frac{\pi}{3}, \frac{2\pi}{3})$. Find the force on a charge the charge of 2 nC located at A.

2. A hydrogen molecule consists of two hydrogen atoms. An alpha particle is a helium nucleus with charge +2e. An α particle is shot through the exact centre of a hydrogen nucleus along a line perpendicular to the line, joining the two hydrogen nuclei (each with charge $+e$) in the molecule as shown in the Fig. 1.23. The distance between the nuclei is b. Assume that the speed of the α particle is very high and so the nuclei do not move and also assume that the electric field due to the electron cloud surrounding each of the hydrogen atoms in the molecule is negligible. Where on its path does the α particle experience the maximum force?

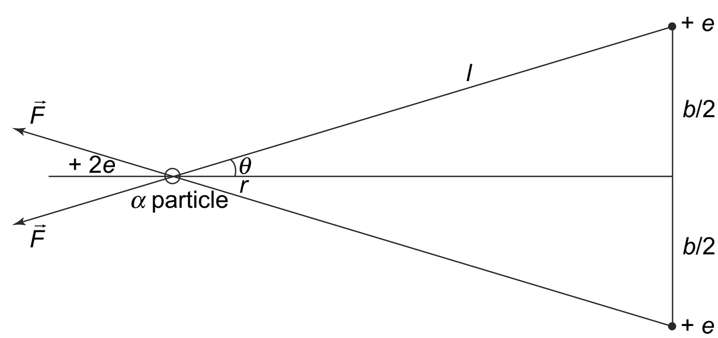


Fig. 1.24 *Problem 2*

3. A bead with charge q_1 is fixed at the end of a wire that makes an angle θ with the horizontal. Another bead of mass m and charge q_2 slides without friction on the wire. At what distance l of the sliding bead from the fixed bead, is the gravitational force on the second bead exactly balanced by the electrostatic force?

4. Four charges q_A, $q_B = -1$ C, $q_C = -1$ C and $q_D = +1$ C are located at the points $(0, 0, 0)$, $(a, 0, 0)$, $(0, a, 0)$ and $(2a, 2a, 0)$. What should be the magnitude and sign of q_A such the the force on q_D is zero?

5. Two point charges of equal mass m charge Q are suspended from a common point by two threads of negligible mass and length l. Show that at equilibrium, the angle of inclination to the vertical α is given by

$$\sin^2\alpha \tan\alpha = \frac{Q^2}{16\pi\varepsilon_0 mgl^2}$$

Further show that for α small, we have

$$\alpha = \left[\frac{Q^2}{16\pi\varepsilon_0 mgl^2}\right]^{1/3}$$

6. A thin half ring of radius R is uniformly charged with a total charge Q. What is the magnitude of the electric field at the centre of curvature of the half ring?
7. A thin non-conducting rod has charge Q and a length a. The charge is spread uniformly along the rod. Find the electric field at a point at a P distance r from the rod, along the perpendicular bisector.
8. A thin, non-conducting ring of radius r has a charge Q placed on it. Find the electric field at a point P on the axis at a distance x from the centre of the ring. At what value of x is the field the maximum?
9. A hollow spherical shell of inner radius r_1 and outer radius r_2 carries a charge density $\rho = k/r^2$. Find the electric field for $r < r_1$, $r_1 < r < r_2$ and $r > r_2$.
10. A hollow metal cylinder of radius R has a surface charge density σ on it. A long thin wire with a linear charge density λ runs through the centre of the cylinder along the axis. Find the electric field at a point r, such that
    a. $r < R$
    b. $r \geq R$.
11. A small sphere of mass m with a charge q hangs from a thread which makes an angle θ with a large, conducting charged plate. Find the charge density of the plate.
12. An electron with energy E_0 is fired directly at a large metal plate with surface charge density σ. From what distance should the electron be fired so that it just fails to strike the plate?
13. Find, by direct integration, the electric field at the centre of a hemisphere of radius r charged uniformly with a surface density σ.
14. A thin non-conducting ring of radius r has a linear charge density given by $\lambda = \lambda_0 \cos\phi$ where ϕ is the azimuthal angle. Find the electric field at a point on the axis of the ring at distance z from the centre. Take the limit $z \to 0$ to find the field at the centre of the ring.
15. A thin annular disc of inner radius a and outer radius b has a surface charge density σ. Find the electric field at a point distance z along the axis.
16. A uniform sphere of radius r_1 has a charge Q distributed uniformly throughout. It is surrounded by a thick shell carrying a charge $-Q$ and having an outer radius $2r_1$. Find the electric field as a function of r the distance from the centre of the sphere.
17. The Thomson model of the hydrogen atom is a sphere of radius R, of positive charge in which an electron of charge e is at the centre of the sphere. The total positive charge is equal to the electron charge. Find the force on the electron when it is at a distance r from the centre.

18. Given a charge distribution that is spherically symmetric and has a charge density defined by

$$\rho = \frac{\rho_0 r}{R}, \qquad 0 \le r \le R$$

and

$$\rho = 0, \qquad r > R$$

determine $\vec{E}$ everywhere.

19. A finite sheet $0 \le x \le 1$, $0 \le y \le 1$ on the $z = 0$ plane has a charge density $\sigma = xy(x^2 + y^2 + 25)^{3/2}$ nC per sq m. Find the total charge on the sheet and the electric field at (0, 0, 5).

20. A spherically symmetric charge distribution of radius R has a charge density given by

$$\rho(r) = \rho_0 \left(1 - \frac{r}{R}\right) \quad \text{for} \quad r \le R$$

and

$$\rho(r) = 0 \quad \text{for} \quad r > R$$

Find the electric field as a function of r.

21. An ion rocket emits positive cesium ions from a wedge-shaped electrode into a region described by $x >| y |$ The electric field is given by $\vec{E} = -400\hat{i} + 200\hat{j}$ kV/m. The ions are singly charged with a charge $e = -1.6 \times 10^{-19}$ C and mass $m = 2.22 \times 10^{-25}$ kg and come out with zero initial velocity. If the emission is confined to the region -40 cm $< y <$ 40 cm, find the largest value of x that can be reached.

22. Given that $\vec{E} = z\rho \cos^2 \phi \hat{e}_z$ C /m^2, calculate the total charge enclosed by a cylinder of radius 1 meter with $-2 \le z \le 2$ m.

23. A semi-infinite rod charged with a uniform charge density λ is placed along the x axis. Show that the field at point P, distance R from one end of the rod (the point $x = 0$) is always at an angle 45° to the rod and this result is independent of R.

24. A region is filled with a charge with volume charge density given by $\rho = \rho_0 e^{-\alpha r^3}$ where ρ_0 and α are positive constants. Find the electric field as a function of r. Also, find the field in the limit of $\alpha r^3 \ll 1$ and $\alpha r^3 \gg 1$.

2

Electric Potential

Learning Objectives

- To introduce the idea of electric potential and to relate it to the concept of work done.
- To comprehend the relationship between the electric field and electric potential and to calculate the electric potential for various charge configurations.
- To learn about the differential equations satisfied by the electric potential in the presence and absence of charges.
- To be able to solve the differential equations satisfied by the potential for various situations and obtain the electric potential.
- To understand the concept of potential energy in electrostatics and its relation to electric potential.
- To learn about the energy of the electric field.
- To be able to solve boundary value problems for various configurations, that is to solve Laplace and Poisson equations with boundary values specified.

2.1 LINE INTEGRAL AND CURL OF THE ELECTRIC FIELD

We have seen in the previous Chapter that the electrostatic force between two point charges is given by Coulomb's Law. This also allowed us to define the concept of a vector field, i.e., the electric field at every point in space. The extension from point charges to extended charge distributions was done using the Principle of Superposition.

Working with the electric field is not very convenient in many situations, primarily because the electric field is a vector field. It has three components and at any point the addition of electric field due to various charges involves vector addition. However, it turns out that the nature of the electrostatic force, as given by Coulomb's Law, allows us to define a scalar field called the electric potential, which is much easier to calculate in some situations. The electric potential is not just a convenient computational device but has significance in terms of its relationship to the potential energy in an electric field, as we shall see later.

So what is it about the nature of the electrostatic force, as given by Coulomb's Law, which allows this simplification? The property of the electric force (or the electric field to which it is related) which allows us to define a scalar potential is the path independence of the work done in the field. To see this, consider a point charge Q located at a point which we take as our origin and a second point charge q, which is taken from a point A to another point B, which have position coordinates $\vec{r}_A$ and $\vec{r}_B$ respectively. First consider moving q along the path $ACDB$, where the sections AC and DB are straight lines, whereas the section CD is in the form of an arc drawn with O as centre as shown in Fig. 2.1.

The force on the charge at a position $\vec{r}$ is $q\vec{E}(r)$, where $\vec{E}(r)$ is the electric field at the position $\vec{r}$ due to the charge Q, i.e.,

$$\vec{F}(\vec{r}) = \frac{kqQ}{r^2}\hat{r} \tag{2.1}$$

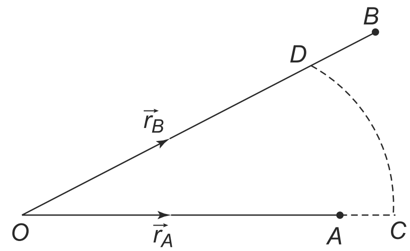


Fig. 2.1

Let us calculate the work done in moving the charge q along the path $ACDB$. Along AC, the displacement is from A to C and is along the direction of the electrostatic force (Eq. (2.1)). Hence the work done in moving the charge along this section of the path is simply

$$W_{AC} = -\int_A^C \vec{F}(\vec{r})\cdot \vec{dr} = kqQ\left[\frac{1}{r_C} - \frac{1}{r_A}\right] \tag{2.2}$$

Next, consider the displacement along the section CD. Here the displacement is perpendicular to the direction of the electrostatic force and hence no work is done, i.e., $W_{CD} = 0$. For the section DB, once again the displacement is parallel to the direction of the electrostatic force and hence the work done is similar to that found in (2.2)

$$W_{DB} = -\int_D^B \vec{F}(\vec{r})\cdot \vec{dr} = kqQ\left[\frac{1}{r_B} - \frac{1}{r_D}\right] \tag{2.3}$$

The total work done therefore in moving the charge q from A to B along the path $ACDB$ is

$$W_{ACDB} = W_{AC} + W_{CD} + W_{DB} = kqQ\left[\frac{1}{r_B} - \frac{1}{r_A}\right] \tag{2.4}$$

where we have used the fact that $r_D = r_C$ since the points C and D are on the arc of a circle centred at the origin.

Next, consider the work done in moving the charge q from A to B but along an arbitrary curve **R** as shown in Fig. 2.2.

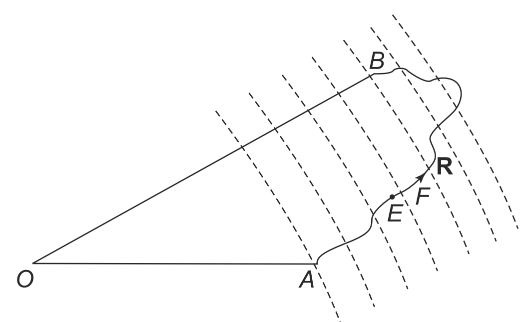


Fig. 2.2

We draw a large number N of arcs that divide the path into N small displacements. In the limit of N tending to infinity, these small sections become infinitesimal in length and they can be considered as straight lines. Consider one such section (Fig. 2.2), between two infinitesimally close points E and F having positions $\vec{r}_i$ and $\vec{r}_{i+1}$, and let the infinitesimal vector joining E and F be $\vec{\Delta}_i = \vec{r}_{i+1} - \vec{r}_i$. The work done in moving the charge through this infinitesimally small section is

$$W_R(\vec{r}_i, \vec{r}_{i+1}) = -kqQ\frac{(\vec{\Delta}_i \cdot \hat{r}_i)}{r_i^2} \tag{2.5}$$

Notice that the quantity $(\vec{\Delta}_i \cdot \hat{r}_i)$ is simply the component of the vector joining E and F along the direction of $\vec{r}_i$. But this is just the length of the line FN (Fig. 2.3), which is just $r_{i+1} - r_i$. Hence, we have

$$\vec{\Delta}_i \cdot \hat{r}_i = r_{i+1} - r_i \tag{2.6}$$

Further, since the points r_{i+1} and r_i are infinitesimally separated, we have

$$\frac{1}{r_i} - \frac{1}{r_{i+1}} = \frac{(r_{i+1} - r_i)}{r_i^2} \tag{2.7}$$

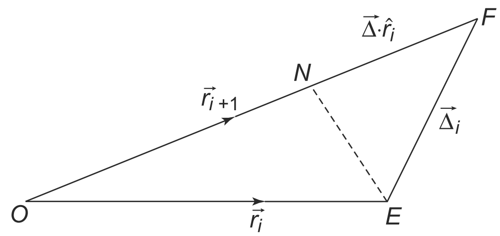


Fig. 2.3

Using Eqs. (2.5), (2.6) and (2.7), we get

$$W_R(\vec{r}_i, \vec{r}_{i+1}) = kqQ\left[\frac{1}{r_{i+1}} - \frac{1}{r_i}\right] \tag{2.8}$$

The total work done in moving the charge q from A to B is obtained by summing up the work done in moving the charge through N such paths, i.e., summing $W_R(\vec{r}_i, \vec{r}_{i+1})$ from $i = 1$ to $i = N - 1$ with $r_1 = r_A$ and $r_N = r_B$. Thus, we get

$$\begin{aligned} W_R(\vec{r}_A, \vec{r}_B) &= kqQ\left[\frac{1}{r_B} - \frac{1}{r_{N-1}} + \frac{1}{r_{N-1}} - \frac{1}{r_{N-2}} + \ldots + \frac{1}{r_2} - \frac{1}{r_A}\right] \\ &= kqQ\left[\frac{1}{r_B} - \frac{1}{r_A}\right] \end{aligned} \tag{2.9}$$

which is the same as the work done in Eq. (2.4).

Thus, we see that the work done in moving point charge from one point to another in the presence of electric field of another point charge is independent of the path that is taken for moving the charge. We have proved this in the presence of a single point charge. But as we saw in Chapter 1, the electric field due to any number of point charges is just the vector sum of the individual electric fields (Principle of Superposition) and hence this result would hold in the presence of an electric field due to an arbitrary number of point charges. Electric field due to continuous distribution of charges can also be regarded as due to an infinite number of point charges, as we had discussed in the last chapter.We thus have a general result:

> Work done in moving a charge from one position to another in the presence of electrostatic fields is independent of the path taken.

This remarkable fact about electrostatic fields and forces is not shared by some of the other types of forces of nature. The gravitational force, of course, has the same property as we know—the difference in potential energy in a gravitational field (or the work done in moving a mass from one point to another) is dependent only on the the positions of the two points and not on the path taken to go from one to another. On the other hand, in moving an object against frictional forces from one point to another, the total work done obviously depends on the path taken.

Mathematically, the path independence for electrostatic fields arises due to special nature of Coulomb's Law for such forces. To see this, consider an integral

$$\int \vec{E}(\vec{r}) \cdot d\vec{r}$$

along two paths marked PSR and PTR as shown in Fig. 2.4.

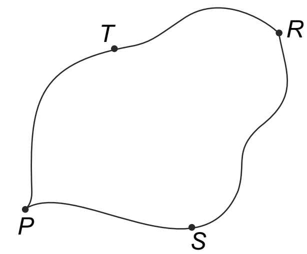


Fig. 2.4

Now the loop integral over the closed loop $PSRTP$ is given by

$$\begin{aligned}\oint_{PSRTP} \vec{E}(\vec{r}) \cdot d\vec{r} &= \int_{PSR} \vec{E}(\vec{r}) \cdot d\vec{r} + \int_{RTP} \vec{E}(\vec{r}) \cdot d\vec{r} \\ &= \int_{PSR} \vec{E}(\vec{r}) \cdot d\vec{r} - \int_{PTR} \vec{E}(\vec{r}) \cdot d\vec{r} \end{aligned} \tag{2.10}$$

where we have used the property of the line integral that it changes sign when the path of integration is reversed. The loop integral on the LHS of Eq. (2.10) would vanish if the line integrals along the paths PSR and PTR were equal.

Stokes Theorem, which we have encountered in the Chapter on Mathematical Preliminaries, allows us to convert the loop integral on the LHS of Eq. (2.10) into a surface integral

$$\oint_{PSRTP} \vec{E}(\vec{r}) \cdot d\vec{r} = \iint_A \vec{\nabla} \times \vec{E}(\vec{r}) \cdot d\vec{S} \tag{2.11}$$

where A is any surface with the loop $PSRTP$ as its boundary or periphery. Thus, if the loop integral on the LHS of Eq. (2.11) vanishes, this implies that the surface integral on the RHS vanishes. Since the loop on the LHS can be **ANY** loop and the surface on the RHS can be **ANY** surface with the loop as its boundary, this implies that $d\vec{S}$ can be in any direction and so, in general, $\vec{\nabla} \times \vec{E}$ vanishes at all points.

We can explicitly verify this result, i.e., the Curl of an electrostatic field vanishes at all points, with a collection of point charges. Consider N point charges Q_i located at points $\vec{R}_i$ where $i = 1, 2, \ldots, N$. Consider the field due to these charges at a point P located at $\vec{r}$.

$$\vec{E}(\vec{r}) = \sum_{i=1}^{N} kQ_i \frac{\vec{r} - \vec{R}_i}{(|\vec{r} - \vec{R}_i|^2)^{3/2}} \tag{2.12}$$

where we have used the fact that the unit vector along $\vec{r} - \vec{R}_i$ can be written as $\frac{\vec{r}-\vec{R}_i}{|\vec{r}-\vec{R}_i|}$. To compute $\vec{\nabla} \times \vec{E}(\vec{r})$, we use the expressions given in the Mathematical Preliminary. The x-component of Curl $\vec{E}$ is given by

$$\left[\vec{\nabla} \times \vec{E}\right]_x = \frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z}$$

and therefore, using Eq. (2.12), we get

$$\left[\vec{\nabla} \times \vec{E}\right]_x = \sum_{i=1}^{N} -kQ_i \frac{(y - Y_i)}{|r - R_i|^5}(-3(z - Z_i)) - \sum_{i=1}^{N} -kQ_i \frac{(y - Y_i)}{|r - R_i|^5}(-3(z - Z_i)) = 0 \tag{2.13}$$

We can compute the other two components and see that they too vanish. Therefore, in general, for a collection of charges, $\vec{\nabla} \times \vec{E} = 0$. Using the Principle of Superposition, one can easily verify that the same result holds for an extended charge distribution. Thus, we have a powerful result that

$$\vec{\nabla} \times \vec{E} = 0$$

The Curl of the electrostatic field vanishes at all points.

Any vector field whose Curl vanishes is called irrotational. What we have shown is that the electrostatic field produced by a single charge, a collection of charges and (using the Principle of Superposition) by a charge distribution is irrotational. It is clear from Eq. (2.11), that for irrotational vectors like the electrostatic fields, loop integrals of the vector across any closed loop vanishes. Conversely, only when a loop integral of a vector vanishes for any, can we conclude the vector is irrotational.

EXAMPLE 2.1 Consider a hypothetical situation where the electrostatic force between a charge q located at (0, 0, 0) and a charge Q located at (x, y, z) is given by

$$\vec{F} = (x\hat{i} + 2y^2\hat{j} + z^2\hat{k})$$

Consider a right angled triangle ABC such that $A(0, 0, 0)$, $B(1, 0, 0)$ and $C(1, 1, 0)$, as shown in Fig. 2.5. Calculate the work done in moving a charge Q from A to C

(a) along the straight line AB and then BC and
(b) along the straight line AC.

Are they equal? Also calculate the curl of the force.

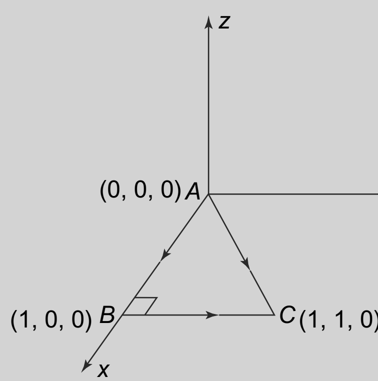


Fig. 2.5 *Example 2.1*

Solution

The unit vectors along AB, BC and AC are (1, 0, 0), (0, 1, 0) and $(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0)$, respectively. Therefore, the work done in moving along the various paths is

For path AB:

$$-\int_{AB} \vec{F} \cdot d\vec{r} = -\int_0^1 F_x dx = -\frac{1}{2}$$

For path BC:

$$-\int_{BC} \vec{F} \cdot d\vec{r} = -\int_0^1 F_y dy = -\frac{2}{3}$$

thus, the total work done in moving from A to B and then from B to C is simply $-\frac{5}{6}$.

For the path AC: The component of the force along AC is $\frac{x}{\sqrt{2}} + \frac{2y^2}{\sqrt{2}}$. But along the line AC, $x = y = \frac{l}{\sqrt{2}}$, where l is the distance from the origin. Therefore, along AC, the force is given by $\frac{l}{2} + \frac{l^2}{\sqrt{2}}$. To get the work done,we need to integrate along AC from $l = 0$ to $l = \sqrt{2}$:

$$-\int_{AC} \vec{F} \cdot d\vec{r} = -\int_0^{\sqrt{2}} \left(\frac{l}{2} + \frac{l^2}{\sqrt{2}} \right) dl = -\frac{5}{6}$$

Thus, the work done in moving the charge along the two paths is the same.

The curl of the force can be easily evaluated:

$$\vec{E} = \frac{1}{Q}(x\hat{i} + 2y^2\hat{j} + z^2\hat{k})$$

and hence $\vec{\nabla} \times \vec{E}$ is seen to be

$$\vec{\nabla} \times \vec{E} = \begin{vmatrix} \hat{i} & \hat{j} & \hat{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ E_x & E_y & E_z \end{vmatrix}$$

which gives us

$$\vec{\nabla} \times \vec{E} = \frac{1}{Q} \begin{vmatrix} \hat{i} & \hat{j} & \hat{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ x & 2y^2 & z^2 \end{vmatrix} = 0$$

EXAMPLE 2.2 We know that the electrostatic field at a point $\vec{r}$ due to a charge q placed at the origin is given by Coulomb's Law as

$$\vec{E}(\vec{r}) = kq \frac{x\hat{i} + y\hat{j} + z\hat{k}}{(x^2 + y^2 + z^2)^{3/2}}$$

If instead, the field had been

$$\vec{E}(\vec{r}) = kq \frac{x\hat{i} + 2y\hat{j} + 3z\hat{k}}{(x^2 + y^2 + z^2)^3}$$

Determine whether the work done in moving a charge in this electric field is independent of the path taken.

Solution

As we have seen above, the work done in moving a charge in an electric field is independent of the path if the electric field is irrotational or the curl of the field vanishes. The curl of the hypothetical electric field can be easily obtained as

$$\vec{\nabla} \times \vec{E} = \begin{vmatrix} \hat{i} & \hat{j} & \hat{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ E_x & E_y & E_z \end{vmatrix}$$

and hence, in our case we have ($r = (x^2 + y^2 + z^2)^{1/2}$)

$$(\vec{\nabla} \times \vec{E})_x = kq(6)\left(\frac{2yz - 3yz}{r^4}\right) \neq 0$$

$$(\vec{\nabla} \times \vec{E})_y = kq(6)\left(\frac{3zx - zx}{r^4}\right) \neq 0$$

$$(\vec{\nabla} \times \vec{E})_z = kq(6)\left(\frac{xy - 2xy}{r^4}\right) \neq 0$$

Hence, the work done in moving the charge in this field is NOT independent of the path.

EXAMPLE 2.3 $ABCD$ is a square with the diagonal given by $l = 2$ in the $x - y$ plane, as shown in Fig. 2.6. Denote by W_1 the work done when a particle is moved from A to C along the arc ADC drawn with midpoint O of AC at its centre, and by W_2 the work done in moving the particle A to C, but in this case along the straight line AC. Calculate both W_1 and W_2 when the force is $\vec{F}(x, y, z) = 5x^2(z+1)\hat{i} + 2y^2\hat{j}$ where O is the origin and OC, OD are along x- and y axes respectively. Also calculate the curl $\vec{F}(x, y, z)$ and show that the surface integral

$$\iint (\vec{\nabla} \times \vec{F}) \cdot d\vec{s}$$

taken over the semicircular surface $ADCA$ vanishes.

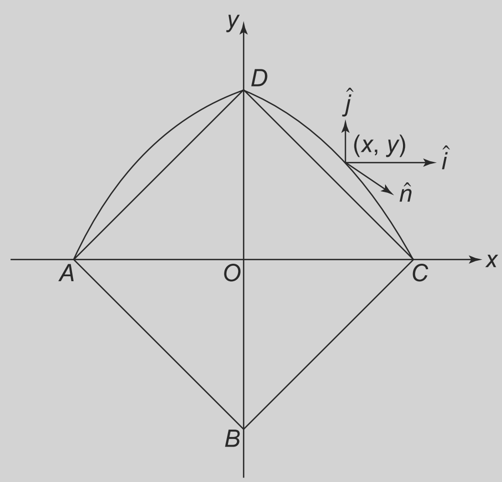


Fig. 2.6 *Example 2.3*

Solution

Let us take O as the origin. Then, the coordinates of A are $(-1, 0)$ and coordinates of C are $(1, 0)$. An infinitesimal displacement dl along AC is simply then $dx\hat{i}$. We also have $y = z = 0$ on this line. From A to C, along the straight path, the work done is therefore,

$$W_2 = -\int_{-1}^{1} (5x^2\hat{i}) \cdot (dx\hat{i}) = -\frac{10}{3}$$

The arc ADC has radius 1 and its centre is at $(0, 0)$. Hence, any point (x, y) on the arc satisfies the equation of the circle which is $x^2 + y^2 = 1$. At an infinitesimally displaced point $(x + dx, y + dy)$ we will have $(x + dx)^2 + (y + dy)^2 = 1$ also. Since dx and dy are infinitesimal, we also have $x^2 + y^2 + 2(xdx + ydy) = 1$ and therefore $(xdx + ydy) = 0$, which relates infinitesimal displacement components dx and dy along the arc ADC with $z = 0$. An infinitesimal displacement along the arc is

$$\vec{dl} = [dx\hat{i} + dy\hat{j}] = dx\left[\hat{i} - \left(\frac{x}{y}\right)\hat{j}\right]$$

using the relation between dx and dy given above. Hence, the work done in moving along the arc is

$$-\int [5x^2\hat{i} - 2y^2\hat{j}] \cdot [\hat{i} - \left(\frac{x}{y}\right)\hat{j}]dx = -\int [5x^2 + 2yx]dx = -\int [5x^2 + 2(1-x^2)^{1/2}x]dx = -\frac{10}{3}$$

which is the same as work done along the straight path.

On the semicircular region, an infinitesimal surface area is along the z-axis, i.e.,

$$\vec{ds} = ds\hat{k}$$

Hence,

$$\vec{\nabla} \times \vec{F}(x, y, z) \cdot \vec{ds} = \left[\vec{\nabla} \times \vec{F}(x, y, z)\right]_z ds$$

But

$$\left[\vec{\nabla} \times \vec{F}(x, y, z)\right]_z = \left[\frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y}\right] = 0$$

Hence,

$$\iint \vec{\nabla} \times \vec{F}(x, y, z) \cdot \vec{ds} = 0$$

PROBLEM 2.1 Can these two vector fields be possible electric fields?

$$\vec{E} = \left(\frac{1}{x+2y+3z}\right)\hat{i} + \left(\frac{2}{x+2y+3z}\right)\hat{j} + \left(\frac{3}{x+2y+3z}\right)\hat{k}$$
$$\vec{E} = \left(\frac{2}{x+2y+3z}\right)\hat{i} + \left(\frac{1}{x+2y+3z}\right)\hat{j} + \left(\frac{4}{x+2y+3z}\right)\hat{k}$$

If the vector fields are indeed possible electric fields, show explicitly that the work done in moving a unit charge from origin to the point (1, 1, 1) via the following two paths are same:

(a) along a straight line and

(a) along the x-axis to the point (1, 0, 0) and then parallel to the y-axis to (1, 1, 0) and then parallel to z-axis to (1, 1, 1).

2.2 ELECTROSTATIC POTENTIAL

Consider moving a **unit positive charge** in the presence of an electrostatic field between any three points amongst A, B and C in Fig. 2.7 with coordinates $\vec{r}_A, \vec{r}_B, \vec{r}_C$.

Let $W(\vec{r}_A, \vec{r}_B)$, $W(\vec{r}_A, \vec{r}_C)$ and $W(\vec{r}_C, \vec{r}_B)$ be the work done in moving the unit charge from A to B, from A to C and from C to B respectively. As we saw in the last section, all of these are independent of the path taken. Consider now moving the unit charge from A to C and then from C to B. In this process, we have moved the charge along the path ACB and hence the work done is $W(\vec{r}_A, \vec{r}_B)$. Thus, we have

$$W(\vec{r}_A, \vec{r}_B) = W(\vec{r}_A, \vec{r}_C) + W(\vec{r}_C, \vec{r}_B) \tag{2.14}$$

The point C (for any A,B) is obviously arbitrary and hence there is no dependence on C on the LHS of the Eq. (2.14). This implies that there should be no dependence on C on the RHS either. But this can only happen if W is of the form

$$W(\vec{r}_A, \vec{r}_C) = \phi(r_C) - \phi(r_A)$$
$$W(\vec{r}_C, \vec{r}_B) = \phi(r_B) - \phi(r_C) \qquad (2.15)$$

where $\phi(r)$ is a scalar function of the scalar r.

From Eqs. (2.14) and (2.15), we get

$$W(\vec{r}_A, \vec{r}_B) = \phi(r_B) - \phi(r_A) \qquad (2.16)$$

for any two points A and B.

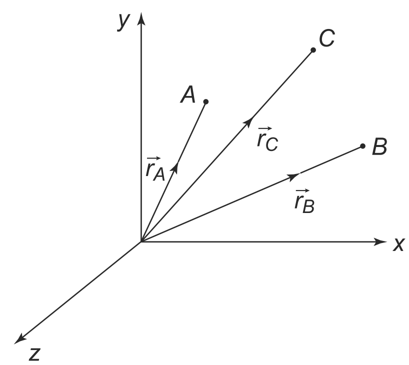


Fig. 2.7

The scalar $\phi(r)$ is called the **electrostatic potential** at the point with coordinate $\vec{r}$. It should be noted that ϕ itself is a scalar quantity, but its value at any point in general depends on the vector coordinate of the point, i.e., $\vec{r}$. However, for brevity $\vec{r}$ quite often is written simply as r in the argument without causing any confusion.

From the relationship between the potential and the work done, it is clear that only the difference in potential between two points has a physical meaning. By convention, one defines the potential to be 0 at infinity assuming of course that all charges are present at a finite distance. The work done in moving a unit charge from infinity to a point with coordinate $\vec{r}$ is by definition $W(\infty, r)$ and using the relationship between the potential and the work done, we get

$$\boxed{W(\infty, r) = \phi(r) - \phi(\infty) = \phi(r)} \qquad (2.17)$$

since by the convention mentioned above, the potential at ∞ is taken to be zero. We shall follow this convention so that the electrostatic potential at any point is the work done in bringing a unit charge from ∞ to that point. Note that this potential is not to be confused with the potential energy of the system, which we shall discuss later.

The potential $\phi(r)$ can immediately be related to the electric field at any point $\vec{E}(\vec{r})$. Let us suppose we have unit charge present at a point with coordinate $\vec{r} = (x, y, z)$ and we move it along the x-direction through an infinitesimal distance dx, i.e., to a point with coordinates $(x + dx, y, z)$. The work done in the process is $-E_x dx$. By definition, this is also the difference of potential between the final and initial points. Hence,

$$-E_x(r)dx = \phi(x + dx, y, z) - \phi(x, y, z) = \frac{\partial\phi(r)}{\partial x}dx \qquad (2.18)$$

where we have used the fact that dx is infinitesimal. Hence,

$$E_x(r) = -\frac{\partial\phi}{\partial x}$$

Similarly,

$$E_y(r) = -\frac{\partial\phi}{\partial y}$$

and

$$E_z(r) = -\frac{\partial \phi}{\partial z}$$

But these are precisely the components of the gradient function introduced in the chapter on Mathematical Preliminaries. So we have the very important result:

The electric field at any point is simply given by the negative of the gradient of the scalar potential at that point.

$$\boxed{\vec{E}(r) = -\vec{\nabla}\phi(\vec{r})} \tag{2.19}$$

This relationship between the electric field at any point (a vector associated with all points) and the scalar potential at that point will prove to be a very useful one. Recall that we had mentioned the fact that frequently dealing with the electric field is tedious because of its vector nature. The scalar potential is much easier to carry out computations with. Introducing the concept of a potential makes the problem of calculating the electric field at any point due to arbitrary collection of charges a much simpler one. This is because it is a scalar unlike the electric field which has three components that have to be calculated separately. The electric field due to a collection of charges is the vector sum of the individual electric fields (Principle of Superposition). Hence the work done in bringing a unit charge from infinity to any point is the sum of work done in the presence of the individual charges or fields separately. **The potential at any point is thus, just the algebraic sum of potentials due to the charges separately.** This leads to considerable ease in calculating the electric potential due a collection of charges and then use Eq. (2.19) to calculate the resultant electric field.

PROBLEM 2.2 An electric field has the following form

$$\vec{E} = xe^{(-\frac{x^2}{2}-\frac{y^2}{2}-\frac{z^2}{2})}\hat{i} + ye^{(-\frac{x^2}{2}-\frac{y^2}{2}-\frac{z^2}{2})}\hat{j} + ze^{(-\frac{x^2}{2}-\frac{y^2}{2}-\frac{z^2}{2})}\hat{k}$$

By explicitly calculating the work done in moving a unit charge from the point $P(x, y, z)$ to infinity by any path of your choice, obtain the value of the potential at P.

PROBLEM 2.3 A point charge q is at the point (1, 0, 0). Calculate the value of the potential at the origin by explicitly calculating the work done in bringing a unit charge from $-\infty$ along the x-axis to the origin. Verify that if instead the unit charge were brought in three steps: first, from $+\infty$ along the x-axis in the negative direction to the point $(1+\varepsilon, 0, 0)$ where ε is a small positive number $\varepsilon \ll 1$; second, from the point $(1+\varepsilon, 0, 0)$ to $(1-\varepsilon, 0, 0)$ along a semicircle of radius ε centred at (1, 0, 0) and finally from $(1-\varepsilon, 0, 0)$ to the origin along the x-axis, the sum of these three pieces of work done is the same as your earlier result.

2.2.1 Equipotential Surfaces

We are familiar with the concept of equipotential surfaces from the gravitational field of the earth- these are simply contour lines or surfaces which indicate constant value of gravitational potential that corresponds to a constant height above the surface of the earth. No work is done against gravity in moving from one point on a specific equipotential surface to another one on the same surface. This concept is equally valid in electrostatics leading to the concept of equipotential surfaces.

If one joins up all points which have the same value of electrostatic potential we get a surface which is called an equipotential corresponding to the value of the potential on it. Equipotential surfaces have certain obvious properties. For instance, two equipotential surfaces corresponding to different values of electric potential clearly can never intersect—for if they did, the points where they intersect will belong to both the surfaces, which is impossible. The potential at such points would then have two values corresponding to the values of the potential on the two surfaces, which is not possible from the definition of potential, Eq. (2.15).

A further property of equipotential surfaces also follows from their definition. Consider moving a charge through an infinitesimal distance along an equipotential surface at any point. Both the initial and final points, lying on the same equipotential surface have the same potential. Hence, by Eq. (2.16), no work is done in the process of moving the charge. On the other hand, we know that moving a charge in the presence of an electric field will entail doing work in general. However, if the electric field is normal to the displacement then no work is done. But the displacement we have taken to be along the equipotential surface. Hence, we conclude the the electric field at that point is normal to the displacement made along the equipotential surface. Further, since the direction of movement on the equipotential surface can be any direction on it, we conclude that **the electric field is always normal to the equipotential surfaces**.

Note that the feature underlying Eq. (2.15) namely that the work done in moving an object from one point to another is independent of the path is shared by many other possible force laws. We have seen that this is the case for the gravitational force field. In fact, any law for the force $F(r)$ on a particle that has the form $\vec{F}(\vec{r}) = \hat{r}\, g(r)$, where $g(r)$ is any function dependent only on the magnitude r has this property of path independence of the work done.

EXAMPLE 2.4 In a region free of charges or matter, can a spherical surface be an equipotential surface with a non-zero electric field inside it?

Solution

Since there are no charges, field lines cannot converge or diverge at any point. Starting from any point on the surface of the sphere, move inwards till we reach another point on the surface. Along the line traced, the potential will either continuously increase or decrease and hence when we reach another point on the sphere, the potential would be greater than or less than the potential at the starting point. The spherical surface is given to be equipotential and hence there is a contradiction. Hence, the answer to the question is no.

EXAMPLE 2.5 Someone claims that the arrangement shown in Fig. 2.8, depicts close-by equipotential surfaces in a region where no matter is present and charges are present only inside the innermost surface shown. The potentials on the surfaces as we move inwards starting from the outermost are shown as ϕ, -2ϕ and $-\phi$. Determine whether this is possible.

Fig. 2.8 *Example 2.5*

Solution

Starting from the outermost surface, let us move inwards along a field line. As we move, the potential will keep on increasing if we are moving against the direction of the field and will keep on decreasing if we are moving along the direction of the field since the field line cannot abruptly reverse its direction unless there is a charge present. Thus, as we follow a field line inwards from the outer surface, it cannot turn back and intersect the outer surface again, since all points on the outer surface are at the same potential. The field lines thus, will cross the middle surface. By a similar reasoning, it will reach the innermost surface. Thus, the potentials of the outermost, middle and the innermost surface must be in increasing or decreasing order. The claim thus about the relative magnitudes of the potential thus, cannot be true.

PROBLEM 2.4 Two charges $+Q$ and $-2Q$ are a distance L apart and the line joining them makes an angle of 45° with the x- axis in the $x-y$ plane.

(a) At how many points in the $x-y$ plane is the electric field zero. Find such point/points.

(b) Draw the locus of points in the $x-y$ plane which are at zero electric potential.

PROBLEM 2.5 In a charge free region of space, electric field everywhere is directed along the x-axis. Give arguments on the basis of Gauss's law and the fact that the electric field is conservative that the magnitude of the field in this region must be constant.

PROBLEM 2.6 Two infinite parallel wires separated by a distance d, lie in the $x-y$ plane parallel to the x-axis. They carry line charge densities of $+\lambda$ and $-\lambda$. Show that points having perpendicular distances r_1 and r_2 from the wires such that $r_1 = br_2$, where b is a positive constant, are equipotential. Show also that such points lie on a circle in the $x-y$ plane of radius $R = \frac{db}{|1-b^2|}$. Locate the centre of this circle.
[*Hint: Express r_1 and r_2 for points on the $x-y$ plane in terms of Cartesian coordinates.*]

2.3 POTENTIAL DUE TO A COLLECTION OF POINT CHARGES

Consider N point charges $q_1, q_2, \cdots, q_N$ located at points $P_1, P_2, \cdots, P_N$ with coordinates $\vec{r}_1, \vec{r}_2, \cdots, \vec{r}_N$. We wish to calculate the electrostatic potential at a point P with coordinate $\vec{r}$ as in Fig. 2.9.

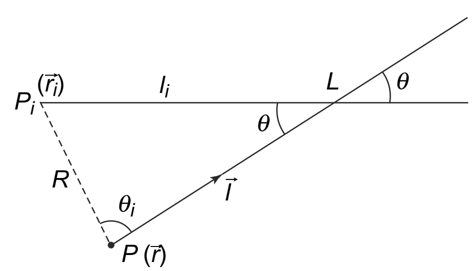


Fig. 2.9

Draw a straight line from P to infinity along any direction. Consider the field due to one of the charges, q_i at any point L (distance l from P) on this line. This has a magnitude

$$|\vec{E}_i| = \frac{kq_i}{(P_iL)^2} = \frac{kq_i}{l_i^2} \tag{2.20}$$

where l_i is the length of P_iL. The field is directed along P_iL. Let the distance of P from L be l which is related to l_i by

$$l_i^2 = l^2 + R^2 - 2lR\cos\theta_i \tag{2.21}$$

where R is the distance PP_i and is given by

$$R = |\vec{r} - \vec{r}_i|$$

and θ_i is the angle as shown in the figure.

Now consider moving a unit charge by an infinitesimal distance dl along PL. The component of the electric field $\vec{E}_i$ along the displacement is obviously $\vec{E}_i \cos\theta$ where θ is the angle as shown in the Fig. 2.9. The work done in moving the unit charge through this distance dl is therefore,

$$dW = |\vec{E}_i|\cos\theta dl = -\frac{kq_i}{l_i^2}\cos\theta dl \tag{2.22}$$

From the figure, on using the cosine rule on the triangle PP_iL we also have

$$l_i\cos\theta + R\cos\theta_i = l \tag{2.23}$$

or

$$\cos\theta = \frac{l - R\cos\theta_i}{(l^2 + R^2 - 2lR\cos\theta_i)^{1/2}}$$

This thus, gives us

$$dW = -\frac{kq_i(l - R\cos\theta_i)}{(l^2 + R^2 - 2lR\cos\theta_i)^{3/2}}dl \tag{2.24}$$

The total work done in moving the charge to infinity is simply integrating this expression from $l = 0$ to $l = \infty$. That, by definition, is simply $-\phi_i(r)$ where $\phi_i(r)$ is the potential at $\vec{r}$ due to the charge q_i. Thus, we get

$$\begin{aligned}\phi_i(\vec{r}) &= \int_0^\infty \frac{kq_i(l - R\cos\theta_i)}{(l^2 + R^2 - 2lR\cos\theta_i)^{3/2}}dl \\ &= \frac{kq_i}{R} \\ &= \frac{kq_i}{|\vec{r} - \vec{r}_i|}\end{aligned} \tag{2.25}$$

Summing over all charges, we get the **total potential at the point P due to the N charges** as

$$\boxed{\phi(\vec{r}) = \sum_{i=1}^{N}\frac{kq_i}{|\vec{r} - \vec{r}_i|}} \tag{2.26}$$

For a single charge Q, located at the origin, the potential at some point $\vec{r}$ is easily seen to be, from Eq. (2.26)

$$\boxed{\phi(r) = \frac{kQ}{r}} \tag{2.27}$$

2.3.1 Electric Dipole

We next consider a simple case of a collection of charges—that of an **electric dipole**. An electric dipole is a system of two charges, $+q$ and $-q$ separated by a distance d. This is a particularly useful arrangement of charges to investigate since, as we shall see later, several atomic level phenomenon can be modelled by thinking of atoms in certain cases as electric dipoles. If we choose as the origin to be the midpoint of the two charges, the coordinates of the $+q$ and $-q$ charges are taken respectively as $+\frac{\vec{d}}{2}$ and $-\frac{\vec{d}}{2}$. The potential due to such a system of point charges is therefore, from Eq. (2.26)

$$\phi(\vec{r}) = kq\left[\frac{1}{|\vec{r}-\frac{\vec{d}}{2}|} - \frac{1}{|\vec{r}+\frac{\vec{d}}{2}|}\right] \tag{2.28}$$

which gives us, (if θ is the angle between the vectors $\vec{d}$ and $\vec{r}$)

$$\phi(\vec{r}) = kq\left[\frac{1}{(r^2+\frac{d^2}{4}-rd\cos\theta)^{1/2}} - \frac{1}{(r^2+\frac{d^2}{4}+rd\cos\theta)^{1/2}}\right] \tag{2.29}$$

For $d << r$, we have

$$(r^2+\frac{d^2}{4}\pm rd\cos\theta)^{1/2} \approx r\left(1\pm\frac{d}{2r}\cos\theta\right) \tag{2.30}$$

Equation (2.26), in this case thus, reduces to

$$\phi(\vec{r}) = \frac{kq}{r}\left[\frac{1}{1-\frac{d}{2r}\cos\theta} - \frac{1}{1+\frac{d}{2r}\cos\theta}\right] \approx \frac{kqd}{r^2}\cos\theta \tag{2.31}$$

where we have neglected terms proportional to $(\frac{d}{r})^2$. **The vector quantity, $q\vec{d}$ is called the electric dipole moment $\vec{\mu}_e$. It is also sometimes written as $\vec{p}$.**

$$\boxed{\vec{\mu}_e = \vec{p} = q\vec{d}}$$

The electric dipole moment vector $\vec{\mu}_e$ is all that we need to know about a dipole. Now since $\vec{\mu}_e\cdot\vec{r} = \mu_e r\cos\theta$, we can rewrite Eq. (2.31) as

$$\phi(\vec{r}) = \frac{k\vec{\mu}_e\cdot\vec{r}}{r^3} \tag{2.32}$$

which is valid of course only for $\frac{d}{r} << 1$. Note that the electric potential due to the electric dipole can be evaluated knowing the electric dipole moment vector $\vec{\mu}_e$ alone.

Once we know $\phi(r)$ we can find the electric field since

$$\vec{E}(\vec{r}) = -\vec{\nabla}\phi \tag{2.33}$$

In the present case, we can use Eq. (2.33) to calculate the electric field of the dipole, using the expression for potential (Eq. (2.32)). This can be done in a straightforward manner and we quote the result

$$E_i(\vec{r}) = -\frac{k(\mu_e)_i}{r^3} + \frac{2k\vec{\mu}_e\cdot\vec{r}}{r^4}\hat{r}_i \qquad i = x, y, z \tag{2.34}$$

or

$$\vec{E}(\vec{r}) = \frac{k}{r^3}\left[3(\vec{\mu}_e \cdot \hat{r})\hat{r} - \vec{\mu}_e\right] \tag{2.35}$$

where we have used

$$\frac{\partial(\vec{\mu}_e \cdot \vec{r})}{\partial r_i} = (\vec{\mu}_e)_i$$

and

$$\frac{\partial(r^{-3})}{\partial r_i} = -3\frac{r_i}{r^4}$$

where

$$\hat{r}_i \equiv \frac{(\vec{r}_i)}{r}$$

The magnitude of $\vec{E}(\vec{r})$ is conveniently expressed in terms of θ (since $\vec{\mu}_e \cdot \hat{r} = \mu_e \cos\theta$)

$$|\vec{E}(\vec{r})| = \frac{k\mu_e}{r^3}(1 + 3\cos^2\theta)^{1/2} \tag{2.36}$$

It is interesting to note that for an isolated point charge, the electric field falls of as $\dfrac{1}{r^2}$ as we know. In the case of an electric dipole, the electric field falls off as $\dfrac{1}{r^3}$. The electric field lines can be plotted as shown in the Fig. 2.10 below, which make the distinction between the two fields even clearer.

It should be clear from the relationship between the electric field and potential, that we cannot find the electric field at a point if we ONLY know the potential at that point just as it is not possible to know $\dfrac{df(x)}{dx}$ if we only know $f(x)$ at **one** point. For finding the derivative of a function at any point, we not only need to know the value of the function at that point but also in the neighbourhood around the point.

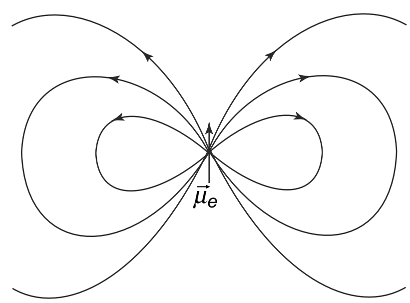


Fig. 2.10 *Field of an electric dipole*

The potential $\phi(\vec{r})$ is very often denoted by the symbol $V(\vec{r})$ or $V(x, y, z)$. The unit of electric potential is Volts (V) in honour of the eighteenth century Italian scientist Alessandro Volta (1745-1827). In some cases, where the value of the potential depends only on the magnitude of $\vec{r}$, it is written as $V(r)$. In cases when the potential does not depend on one or two of the coordinates, the potential is written only as a function of coordinates that it depends on. Thus, if the potential depends only on x and not on y and z, we write $\phi(x)$. In this case, Eq. (2.33) tells us that the components of the electric field along the y and z axes vanish, since in this case $\dfrac{\partial\phi(x)}{\partial y}$ and $\dfrac{\partial\phi(x)}{\partial z}$ vanish.

EXAMPLE 2.6 Points P_1, P_2, P_3, P_4 along the x-axis at $x = 1$ m, 2 m, 4 m and 8 m respectively. Four point charges 1C, -1C, 1C, -1C are placed at P_1, P_2, P_3 and P_4 respectively as shown in Fig. 2.11. Calculate the electrostatic potential and electric field at the origin ($x = 0$, $y = 0$, $z = 0$) and at the point ($x = 0$, $y = 1$, $z = 0$).

Solution

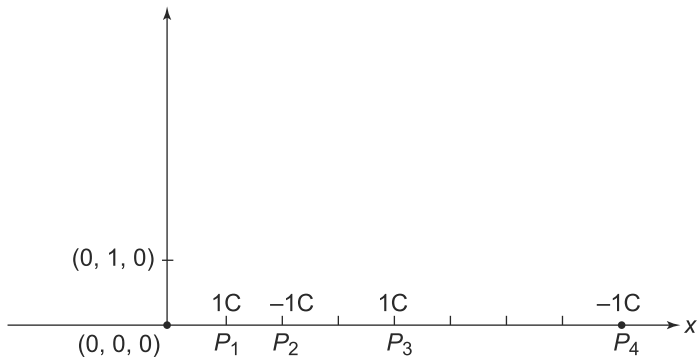


Fig. 2.11 *Example 2.6*

The point charges are located at (1, 0, 0), (2, 0, 0), (4, 0, 0) and (8, 0, 0). The electric potential at the origin is thus given by (Eq. (2.26)) as

$$V(0,0,0) = k\left[\frac{1}{1} + \left(\frac{-1}{2}\right) + \frac{1}{4} + \left(\frac{-1}{8}\right)\right] = \frac{5k}{8}V$$

At the point (0, 1, 0), similarly, using Eq. (2.26) we get

$$V(0,1,0) = k\left[\frac{1}{1} + \left(\frac{-1}{\sqrt{5}}\right) + \frac{1}{\sqrt{17}} + \left(\frac{-1}{\sqrt{65}}\right)\right]V$$

To find the electric field, we need the potential at a point (x, y, z). This is given by

$$V(x,y,z) = k\left[\frac{1}{((x-1)^2+y^2+z^2)^{1/2}} - \frac{1}{((x-2)^2+y^2+z^2)^{1/2}} + \frac{1}{((x-4)^2+y^2+z^2)^{1/2}} - \frac{1}{((x-8)^2+y^2+z^2)^{1/2}}\right]$$

The components of the electric field are thus easily evaluated to be

$$E_x(x,y,z) = -\frac{\partial V}{\partial x} = k\left[\frac{(x-1)}{((x-1)^2+y^2+z^2)^{3/2}} - \frac{(x-2)}{((x-2)^2+y^2+z^2)^{3/2}} + \frac{(x-4)}{((x-4)^2+y^2+z^2)^{3/2}} - \frac{(x-8)}{((x-8)^2+y^2+z^2)^{3/2}}\right]$$

$$E_y(x,y,z) = -\frac{\partial V}{\partial y} = k\left[\frac{(y)}{((x-1)^2+y^2+z^2)^{3/2}} - \frac{(y)}{((x-2)^2+y^2+z^2)^{3/2}} + \frac{(y)}{((x-4)^2+y^2+z^2)^{3/2}} - \frac{(y)}{((x-8)^2+y^2+z^2)^{3/2}}\right]$$

$$E_z(x,y,z) = -\frac{\partial V}{\partial z} = k\left[\frac{(z)}{((x-1)^2+y^2+z^2)^{3/2}} - \frac{(z)}{((x-2)^2+y^2+z^2)^{3/2}} + \frac{(z)}{((x-4)^2+y^2+z^2)^{3/2}} - \frac{(z)}{((x-8)^2+y^2+z^2)^{3/2}}\right]$$

At the point (0, 0, 0), the field components are thus

$$E_x = k\left[-\frac{1}{1}+\frac{2}{8}-\frac{4}{64}+\frac{8}{512}\right] = -\frac{51}{64}\text{ V/m}$$

$$E_y = E_z = 0$$

At the point (0, 1, 0), the fields are given by

$$\begin{aligned} E_x &= k\left[-\frac{1}{(2)^{3/2}}+\frac{2}{(5)^{3/2}}-\frac{4}{(17)^{3/2}}+\frac{8}{(65)^{3/2}}\right]\text{ V/m} \\ E_y &= k\left[\frac{1}{(2)^{3/2}}-\frac{1}{(5)^{3/2}}+\frac{1}{(17)^{3/2}}-\frac{1}{(65)^{3/2}}\right]\text{ V/m} \\ E_z &= 0 \end{aligned}$$

EXAMPLE 2.7 Four point charges q, $2q$, $3q$ and $4q$ are placed at the four corners of a square of length $l = \sqrt{2}$ m. (Fig. 2.12) Calculate the potential and the electric field at

(a) centre of the square and

(b) at a point P, 1 m away from the centre along a direction perpendicular to the plane of the square.

Solution

Let the square be in the $x - y$ plane with the centre as the origin with the line joining q to $3q$ as the x-axis. The coordinates of the charges q, $3q$, $2q$ and $4q$ are $(-1, 0, 0)$, $(1, 0, 0)$, $(0, 1, 0)$ and $(0, -1, 0)$. $V(x, y, z)$ hence, is

$$V(x,y,z) = kq\left[\frac{1}{((x+1)^2+y^2+z^2)^{1/2}}+\frac{3}{((x-1)^2+y^2+z^2)^{1/2}}+\frac{2}{(x^2+(y-1)^2+z^2)^{1/2}} + \frac{4}{(x^2+(y+1)^2+z^2)^{1/2}}\right]$$

Therefore, the potential at the origin is

$$V(0,0,0) = kq\left[\frac{1}{1}+\frac{3}{1}+\frac{2}{1}+\frac{4}{1}\right] = 10kq\text{ V}$$

and at the point P, the the potential is

$$V(0,0,1) = kq\left[\frac{1}{\sqrt{2}}+\frac{3}{\sqrt{2}}+\frac{2}{\sqrt{2}}+\frac{4}{\sqrt{2}}\right] = kq\frac{10}{\sqrt{2}}\text{ V}$$

The electric field at the point (x, y, z) is obtained by differentiating $V(x, y, z)$:

$$E_x(x,y,z) = kq\left[\frac{(x+1)}{((x+1)^2+y^2+z^2)^{3/2}}+\frac{3(x-1)}{((x-1)^2+y^2+z^2)^{3/2}}+\frac{2x}{(x^2+(y-1)^2+z^2)^{3/2}} + \frac{4x}{(x^2+(y+1)^2+z^2)^{3/2}}\right]$$

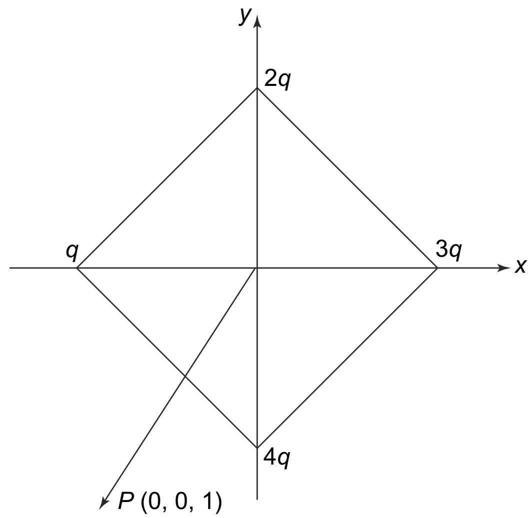


Fig. 2.12 *Example 2.7*

$$E_y(x, y, z) = kq\left[\frac{y}{((x+1)^2+y^2+z^2)^{3/2}} + \frac{3y}{((x-1)^2+y^2+z^2)^{3/2}} + \frac{2(y-1)}{(x^2+(y-1)^2+z^2)^{3/2}} + \frac{4(y+1)}{(x^2+(y+1)^2+z^2)^{3/2}}\right]$$

$$E_z(x, y, z) = kq\left[\frac{z}{((x+1)^2+y^2+z^2)^{3/2}} + \frac{3z}{((x-1)^2+y^2+z^2)^{3/2}} + \frac{2z}{(x^2+(y-1)^2+z^2)^{3/2}} + \frac{4z}{(x^2+(y+1)^2+z^2)^{3/2}}\right]$$

With these expressions, we can easily evaluate the field at any point. At the origin

$$E_x(0, 0, 0) = kq[1 - 3 + 0 + 0] = -2kq \text{ V/m}$$

$$E_y(0, 0, 0) = kq[0 + 0 - 2 + 4] = 2kq \text{ V/m}$$

$$E_z(0, 0, 0) = kq[0 + 0 + 0 + 0] = 0 \text{ V/m}$$

Similarly, at the point P

$$E_x(0, 0, 1) = kq\left[\frac{1}{2^{3/2}} - \frac{3}{2^{3/2}} + 0 + 0\right] = -\frac{kq}{\sqrt{2}} \text{ V/m}$$

$$E_y(0, 0, 1) = kq\left[0 + 0 - \frac{2}{2^{3/2}} + \frac{4}{2^{3/2}}\right] = +\frac{kq}{\sqrt{2}} \text{ V/m}$$

$$E_z(0, 0, 1) = kq\left[\frac{1}{2^{3/2}} + \frac{3}{2^{3/2}} + \frac{2}{2^{3/2}} + \frac{4}{2^{3/2}}\right] = \frac{5kq}{\sqrt{2}} \text{ V/m}$$

PROBLEM 2.7 Charges $+Q$, $-2Q$, $-4Q$ and $+4Q$ are located at (1, 0, 0), (2, 0, 0), (16, 0, 0) and (8, 0, 0) respectively. Show that all points on the spherical surface $x^2 + y^2 + z^2 = 16$ are at zero potential.

PROBLEM 2.8 Charges $+Q$, $-2Q$ and $+Q$ are located on the x-axis respectively at $x = -L$, $x = 0$ and $x = L$. Calculate the electric potential at the point $(r\cos\theta, r\sin\theta, 0)$. Show that in the limit of $r \gg L$ so that one can neglect terms of order $(\frac{L^2}{r^2})$, the angular dependence of the potential is given by $P_2(\cos\theta)$ where $P_2(\cos\theta)$ is the Legendre Polynomial of order 2.

2.4 POTENTIAL DUE TO A CONTINUOUS CHARGE DISTRIBUTION

The expression for the potential, Eq. (2.26), for a collection of point charges can be easily extended to a continuous charge distribution. Consider a continuous charge distribution with a volume charge density $\rho(\vec{r})$. An infinitesimal volume $d^3\vec{r}\,'$ at $\vec{r}\,'$ can be thought of as a point charge carrying a charge $\rho(\vec{r}\,')d^3\vec{r}\,'$ located at $\vec{r}\,'$. Summing over all such infinitesimal charges (volumes) implies, in the limit that the volume $d^3\vec{r}\,' \to 0$, an integral over $\vec{r}\,'$. Hence, in place of Eq. (2.26) we will obtain

$$\phi(\vec{r}) = \int d^3r' \frac{k\rho(\vec{r}')}{|\vec{r} - \vec{r}'|} \tag{2.37}$$

For regular shaped, uniformly charged bodies, this integral can be evaluated.

Equation (2.37) is a general expression involving charge density which is charge per unit volume at any point. Very often, for charge distributions that are restricted along a thin line or a thin surface, it is more convenient to express the potential in terms of line charge densities or surface charge densities. Thus, consider a case where the electric charges is confined to a thin line (Fig. 2.13). A point on the line can be characterised by its distance l along the line from one end or some other reference point on the line. The charge contained in a small section dl of the line is λdl where λ is the charge per unit length or the line charge density. The coordinates of a point l along the line can be written as $\vec{r}'(l)$. Analogous to

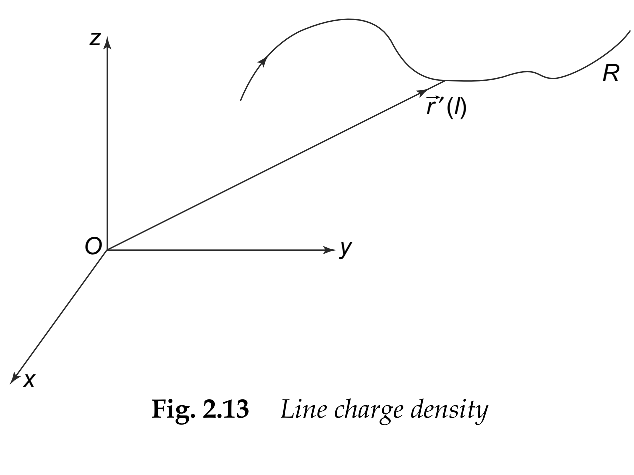


Fig. 2.13 *Line charge density*

(2.37), we can write for the potential at any point with coordinates $\vec{r}$:

$$\phi(\vec{r}) = \int \mathrm{dl} \frac{k\lambda(\vec{r}\,')}{|\vec{r} - \vec{r}\,'(l)|} \tag{2.38}$$

Charges confined to a thin two-dimensional surface can similarly be treated. For any point on such a surface, only two of the three coordinates are independent. Let us choose them to be x and y so that the third coordinate of any point on the surface is a function of x and y: $z = z(x, y)$. If we consider at a point with coordinates x, y a small element of area $dxdy$ on the surface, the charge contained in it is $\rho(x, y)tdxdy$ where t is thickness of the surface on which the charge is confined (Fig. 2.14). This can be equivalently written in terms of a surface charge density $\sigma(x, y)$ as $\sigma(x, y)dxdy$ where $\sigma(x, y) = \rho(x, y)t$ at any point (x, y) on the surface is the charge per unit area. The potential at a point with coordinate $\vec{r}$ can then be written as

$$\phi(\vec{r}) = \iint dx'dy' \frac{k\sigma(\vec{r}\,')}{|\vec{r} - \vec{r}\,'(x', y')|} \tag{2.39}$$

The integral over x and y runs over the entire surface and $\vec{r}(x'y')$ denotes the position vector of the point with coordinates (x', y') on the surface.

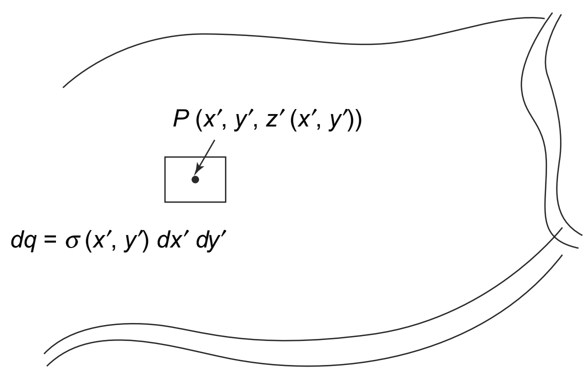


Fig. 2.14 *Surface charge density*

We have written Eq. (2.39) in terms of Cartesian coordinates, but this can be done in any coordinate system.

Equation (2.37) can be explicitly evaluated for the simple cases where the distribution is uniform over regular geometrical shapes.

2.4.1 Charged Wire

Consider first the potential at a point P with coordinates $\vec{r} = (x, y, z)$ due to a charged wire of length $2d$ placed along the z-axis with its midpoint at the origin O (Fig. 2.15).

Using Eq. (2.38), we get the potential to be

$$\phi(\vec{r}) = k \int_{-d}^{+d} \lambda(l) \frac{dl}{\left[x^2 + y^2 + (z - l)^2\right]^{1/2}} \tag{2.40}$$

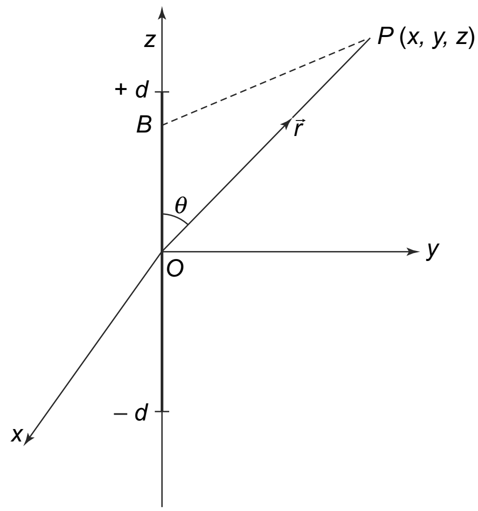


Fig. 2.15 *Charged wire*

where $\lambda(l)$ is the charge per unit length of the wire or the linear charge density, which in general, could depend on the position along the wire. In the case where $\lambda(l) = \lambda$, a constant, i.e., the case of a uniformly charged wire, we can evaluate the integral in Eq. (2.40) to get

$$\phi(\vec{r}) = -k\lambda \ln\left[\frac{(z-d) + \{x^2 + y^2 + (z-d)^2\}^{1/2}}{(z+d) + \{x^2 + y^2 + (z+d)^2\}^{1/2}}\right] \tag{2.41}$$

Several limits of this expression can be easily seen. In the limit that P is far from the wire and hence x^2, y^2 and z^2 are much greater than d^2, Equation (2.41) reduces to

$$\phi(\vec{r}) = k\lambda\frac{2d}{r} = k\frac{Q}{r} \tag{2.42}$$

since $2d\lambda = Q$, the total charge on the wire. This is the expression for the potential due to a point charge which is what the point P, far away from the wire will see if the distance is much greater than the dimensions of the wire. For the case $d = \infty$, i.e., an infinite wire, the integral diverges or is infinite. This is simply because our convention for the electrostatic potential is that the potential is 0 at ∞. Clearly for an infinite charged wire, there are charges present at ∞ and so our assumption/convention is not justified. We can circumvent this problem by working out the difference in the value of $\phi(r)$ from its value at another point.

2.4.2 Charged Ring

Next, consider the potential at a point P with coordinates (x, y, z) due to a uniformly charged circular ring of radius R carrying a charge λ per unit length. We have taken the ring to lie in the $x - y$ plane at $z = 0$ with its centre as the origin, as shown in Fig. 2.16. The coordinates of a point A on the ring are denoted by

$$\vec{r}_A = (-R\sin\theta,\, R\cos\theta,\, 0)$$

The charge in an infinitesimal section AB of the ring subtending an angle $d\theta$ at the centre is $\lambda R d\theta$.

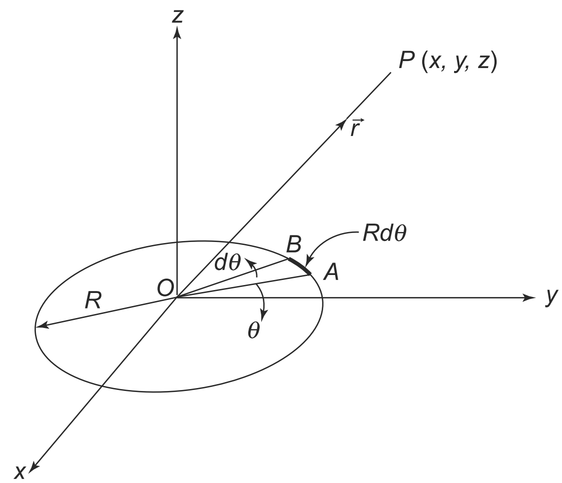


Fig. 2.16 *Charged ring*

From the general expression for the potential due to a charge distribution (Eq. (2.37) and Eq. (2.38)), the potential is given by:

$$\phi(\vec{r}) = k\lambda R \int_0^{2\pi} \frac{\mathrm{d}\theta}{|\vec{r} - \vec{r}_A|} = k\lambda R \int_0^{2\pi} \frac{\mathrm{d}\theta}{[(x + R\sin\theta)^2 + (y - R\cos\theta)^2 + z^2]^{1/2}} \tag{2.43}$$

The electric field is obtained, using Eq. (2.19) as

$$E_x = kR\lambda \int_0^{2\pi} \frac{(x + R\sin\theta)\mathrm{d}\theta}{[(x + R\sin\theta)^2 + (y - R\cos\theta)^2 + z^2]^{3/2}} \tag{2.44}$$

$$E_y = kR\lambda \int_0^{2\pi} \frac{(y - R\sin\theta)\mathrm{d}\theta}{[(x + R\sin\theta)^2 + (y - R\cos\theta)^2 + z^2]^{3/2}} \tag{2.45}$$

$$E_z = kR\lambda \int_0^{2\pi} \frac{z\mathrm{d}\theta}{[(x + R\sin\theta)^2 + (y - R\cos\theta)^2 + z^2]^{3/2}} \tag{2.46}$$

The integrals above, known as elliptic integrals, cannot in general be expressed in terms of well known functions. However, practical calculation of electric fields due to such and similar charge distributions are required for design of many devices like electrostatic ion-traps and mass spectrometers. Numerical methods thus have been developed for such calculations. See e.g **F. R. Zypman, American J. of Phys., 74, 295 (2006)**, for such calculations and discussions of their use in practise.

For special choices of the point P, these integrals can be simplified. Thus, if the point P is on the z-axis, we get

$$\begin{aligned}\phi(0,0,z) &= k\lambda R\int_0^{2\pi}\frac{d\theta}{[(R\sin\theta)^2+(R\cos\theta)^2+z^2]^{1/2}}\\ &= kR\lambda\frac{2\pi}{(R^2+z^2)^{1/2}}\\ &= kQ\frac{1}{(R^2+z^2)^{1/2}}\end{aligned} \tag{2.47}$$

since $Q = 2\pi R\lambda$ is the total charge on the ring.

The results above for a ring can be used for calculation of electrostatic potential and electric field due to a charged disc of uniform surface charge density σ. A disc of radius R can be thought of as being made up of an infinite number of circular rings of radii r varying from 0 to R and of infinitesimal thickness or width dr. The total charge on a ring of radius r is $dQ = 2\pi r dr\sigma$ and hence the charge per unit length on it is $\lambda = \sigma dr$. For each of these rings, we can use the results above. Integrating over the infinite number of rings will result in the potential and field for the disc. As in the case of a single ring, the result cannot be given in terms of simple functions. For points on the axis, however, the result is easily calculated from Eq. (2.47):

$$\phi(0,0,z) = k\int_0^R\frac{2\pi r\sigma}{(r^2+z^2)^{1/2}}dr \tag{2.48}$$

The result is

$$\phi(0,0,z) = 2kQ\frac{\left[(1+\frac{z^2}{R^2})^{1/2}-\frac{z}{R}\right]}{R} \tag{2.49}$$

in terms of total charge on the disc $Q = \pi R^2\sigma$ on the disc.

EXAMPLE 2.8 Calculate the potential due to a uniformly charged spherical shell of radius R at a point $\vec{r}$ inside and outside it.

Solution

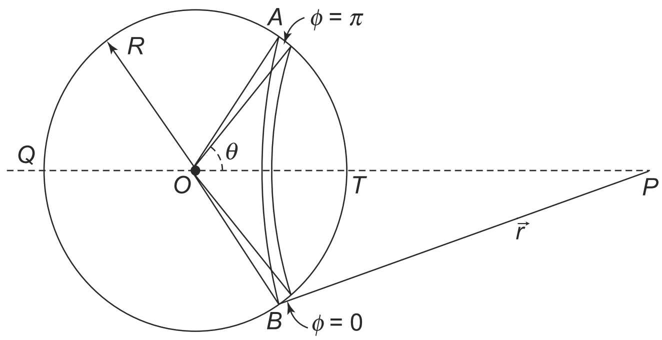
Let σ be the charge per unit area on the shell. Then the total charge on the shell is simply

$$Q = 4\pi R^2\sigma$$

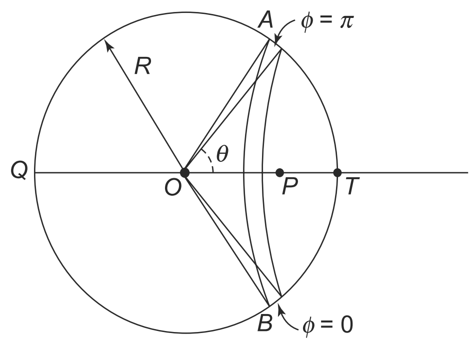
Draw a line from the centre to the surface of the shell (a radius) and also an infinitesimal solid angle $d\Omega$ around it at an angle θ with respect to the line joining the centre and the point at $\vec{r}$ where the potential is to be evaluated, as shown in Fig. 2.17. As we rotate the radius around this line, the angle θ will not change and the radius will trace out a circle on the shell. On a two-dimensional plot, points A and B are the points where the circle intersects the plane.

The area of the shell cut by the solid angle $d\Omega = \sin\theta d\theta d\phi$ is

$$dA = R^2 d\Omega$$



(a)



(b)

Fig. 2.17 *Example 2.8*

and hence, the charge on this area is

$$dQ = \sigma R^2 d\Omega$$

All points on the shell making an angle θ with OP have a distance from P equal to

$$(R^2 + r^2 - 2Rr\cos\theta)^{1/2}$$

The potential due to the charge dQ at P is given by

$$dV = \frac{k\sigma R^2 d\Omega}{(R^2 + r^2 - 2Rr\cos\theta)^{1/2}}$$

The potential due to the whole shell is obviously obtained by integrating this over θ from 0 to π and over ϕ from 0 to 2π. This is done and we get

$$V = \frac{2\pi k\sigma R^2}{Rr}\left[(R^2 + r^2 - 2Rr\cos\theta)^{1/2}\right]_{\theta=0}^{\pi}$$

The difference between the point interior to the shell and exterior to it is evident when we evaluate the limits on θ. The quantity $(R^2 + r^2 - 2Rr\cos\theta)^{1/2}$ at $\theta = \pi$ is the distance PQ which is $R + r$ in both cases. However, for $\theta = 0$, it is the distance PT which is $R - r$ for points inside the spherical shell and $r - R$ for points outside. Thus, we must take the positive root for the points outside the shell and negative root for points inside. Using this, we get

$$V(r) = \frac{2\pi k\sigma R^2}{Rr}[(R+r)+(R-r)] = \frac{kQ}{r} \qquad \text{for} \quad r > R$$

and

$$V(r) = \frac{2\pi k\sigma R^2}{Rr}[(R+r)-(R-r)] = \frac{kQ}{R} \qquad \text{for} \quad r < R$$

PROBLEM 2.9 ABC is an isosceles triangle with BC as the base of length L and with base angles $\frac{\pi}{4}$. The sides AB and AC of the triangle have a uniform line charge density of λ whereas the base BC has a uniform line charge density 2λ. Calculate the electric potential at the centroid of the triangle.

2.5 LAPLACE AND POISSON EQUATIONS

Gauss's Law, as we have seen, relates the derivative of the electric field with the charge density in the region. In its differential form, Gauss's Law reads

$$\vec{\nabla} \cdot \vec{E}(\vec{r}) = 4\pi k\rho(\vec{r})$$

We have seen in the previous chapter that given the charge density, this equation allows us to find the electric field in certain situations. The electric potential in a region is related to the electric field as we have seen above.

$$\vec{E}(\vec{r}) = -\vec{\nabla}\phi(\vec{r})$$

Clearly then, we can reformulate Gauss's Law in terms of the electric potential. This will also have the advantage that the differential equation will then be in terms of a scalar quantity, rather than the vector electric field. Substituting the relation between the electric field and potential in the differential form of Gauss's Law, we get

$$\vec{\nabla} \cdot (-\vec{\nabla}\phi(\vec{r})) = 4\pi k\rho(\vec{r}) \tag{2.50}$$

We know from the chapter on Mathematical Preliminaries that the divergence of a gradient gives us the ∇^2 operator which in rectangular Cartesian coordinates is given by

$$\vec{\nabla} \cdot \vec{\nabla} = \nabla^2 \equiv \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$$

Therefore, in terms of the electric potential, Gauss's Law can be rewritten as

$$\boxed{\nabla^2\phi = -4\pi k\rho} \tag{2.51}$$

This equation is called **Poisson's Equation**. This relates the charge density in a region to the derivative of the electric potential in that region.

For regions where there is no charge present, the right-hand side of Eq. (2.51) is obviously zero and we get

$$\boxed{\nabla^2 \phi = 0} \tag{2.52}$$

Equation (2.52) is called **Laplace's Equation**.

It may be mentioned that Poisson and Laplace equations are not unique to electrostatics. There are several branches of physics where they are encountered. For instance, the gravitational potential satisfies the same equations, of course with the the source term (right-hand side) instead of charge density, now replaced by mass or energy density which is the source of the gravitational field. In electrostatics the solutions of these equations give all the electrostatic potentials and from that the electric fields.

For Poisson equation, the electrostatic potential expressed in terms of the charge density, Eq. (2.37) is the formal solution.

$$\phi(\vec{r}) = \int d^3r' \frac{k\rho(\vec{r}')}{|\vec{r} - \vec{r}'|}$$

Of course, this is purely a formal solution. We need to evaluate the integral for the given charge density which is possible analytically only in certain special cases. We have already evaluated for some special cases (uniformly charged line, ring and a spherical shell) and have explicitly obtained the electrostatic potential.

For regions without any charges, the electric potential satisfies the Laplace equation. Solving this equation in charge free regions gives us the value of the potential in the region. However, this being a second order differential equation, we need to know the value of the potential or its first derivatives at some points to get the complete solution. In general, if we have a region without any charges and the value of the potential is known at some points or surfaces, Laplace's equation can be very useful in determining the complete solution to the problem. We will study the general electrostatic problem of solving the Poisson or Laplace equation with given boundary conditions in the Advanced Topic at the end of the chapter. Let us now illustrate the use of Laplace equation in certain situations for determining the electric potential and thus, the electric field.

EXAMPLE 2.9 For some spherically symmetric distribution of charges, the potential on the surface of a sphere of radius R is given to be V_0. Calculate the potential in the region $r > R$. Also calculate the potential for $r < R$, when no charges are present inside the sphere.

Solution

There is an obvious spherical symmetry and the potential can only be a function of r ($\phi(\vec{r}) = \phi(r)$) in the problem. Hence, it is most convenient to try and solve the Laplace Equation in terms of spherical polar coordinates. From the chapter on Mathematical Preliminaries, we know that in spherical polar coordinates (r, θ, ϕ), the ∇^2 operator has the form

$$\nabla^2 = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 \frac{\partial}{\partial r}\right) + \frac{1}{r^2 \sin^2\theta}\frac{\partial}{\partial \theta}\left(\sin\theta \frac{\partial}{\partial \theta}\right) + \frac{1}{r^2 \sin^2\theta}\frac{\partial^2}{\partial \phi^2}$$

Therefore, the Laplace equation in these coordinates is

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\phi(r)}{\partial r}\right)+\frac{1}{r^2\sin^2\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\phi(r)}{\partial\theta}\right)+\frac{1}{r^2\sin^2\theta}\frac{\partial^2\phi(r)}{\partial\phi^2}=0 \tag{2.53}$$

Since the potential is only a function of r, the last two terms vanish and we get

$$\frac{d}{dr}\left(r^2\frac{d\phi(r)}{dr}\right)=0 \tag{2.54}$$

which can easily be integrated once to get

$$r^2\frac{d\phi(r)}{dr}=A$$

or

$$\phi(r)=-\frac{A}{r}+B \qquad (r\geq R) \tag{2.55}$$

where A and B are constants.

Since there are no charges present at infinity, we can take the potential at ∞ to be zero. Then, since

$$\phi(\infty)=0=B$$

we get $B=0$. We also know the potential at $r=R$, that is V_0. Thus,

$$\phi(R)=V_0=-\frac{A}{R}$$

and hence

$$A=-V_0R$$

Therefore, we have the potential for $r\geq R$ as

$$\phi(r)=\frac{V_0R}{r} \qquad r\geq R$$

For the region $r<R$, once again we can use the Laplace's equation, since there are no charges present inside the region. We can do a similar analysis and get

$$\phi(r)=-\frac{A'}{r}+B' \qquad (r\leq R) \tag{2.56}$$

But at $r=0$, the potential must be finite. This is possible, only if $A'=0$ and thus, we get

$$\phi(r)=-B' \qquad (r\leq R) \tag{2.57}$$

Now at $r=R$, the potential is known to be V_0. Thus,

$$B'=-V_0$$

and we get

$$\phi(r)=V_0 \qquad r\leq R$$

that is the potential inside is constant and equal to the value at the surface at $r = R$. This is expected since we know that inside the spherical region, there are no charges and hence if we use Gauss's Law, we immediately see that the electric field must be zero. But the electric field is related to the gradient of the potential and hence, potential must be a constant. Since the value of the potential is V_0 at the surface, it follows that the potential at all interior points must be V_0 as we found.

There are several situations where one needs to use both Poisson's and Laplace equation. One such configuration is discussed below. This is a situation is similar to that which is used in many electrostatic copying machines.

EXAMPLE 2.10 A, B, C and D are infinite planar surfaces. A and D are at zero potential whereas the region between B and C has charges with a constant charge density ρ. The separation between A and B, between B and C and between C and D are respectively d, $2t$ and d. Calculate the electrostatic potential between A and B, between B and C and between C and D. The separation between the planer surfaces are depicted in Fig. 2.18.

Solution

Let the normal to the surfaces be along the z-axis and let the midpoint between B and C be at $z = 0$. The surfaces being infinite, all values of x and y are similar and hence, the potential depends only on z. Let $V(1)$, $V(2)$ and $V(3)$ respectively be the potential between A and B (region 1), between B and C and between C and D, respectively. The geometry of the problem demands that we use the rectangular Cartesian coordinate system. In this system, the ∇^2 operator is given by

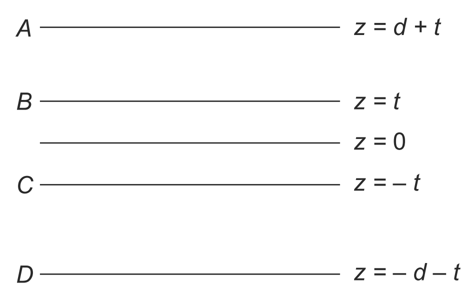


Fig. 2.18 *Example 2.10*

$$\vec{\nabla} \cdot \vec{\nabla} = \nabla^2 \equiv \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$$

Since all the potentials can only depend on the z coordinate, we get the following equations for the potentials in the three regions:

$$\frac{d^2V(1)}{dz^2} = 0 \tag{2.58}$$

$$\frac{d^2V(2)}{dz^2} = -4\pi k\rho \tag{2.59}$$

$$\frac{d^2V(3)}{dz^2} = 0 \tag{2.60}$$

These equations can be easily solved to get

$$V(1) = C_1 z + C_2 \tag{2.61}$$

$$V(2) = C_3 z + C_4 - 4\pi k\rho \frac{z^2}{2} \tag{2.62}$$

$$V(3) = C_5 z + C_6 \tag{2.63}$$

where $C_1, \cdots, C_6$ are constants.

Let us first apply the boundary conditions at $z = d + t$ and at $z = -d - t$. The potential vanishes at these surfaces and hence:

$$C_2 + C_1(d + t) = 0 \tag{2.64}$$
$$C_6 + C_5(-d - t) = 0 \tag{2.65}$$

Thus, we get

$$V(1) = C_1(z - d - t)$$

and

$$V(3) = C_5(z + d + t)$$

Next at $z = t$, $V(1)$ must equal $V(2)$ and also $\frac{dV(1)}{dz}$ must equal $\frac{dV(2)}{dz}$. So, we get

$$-C_1 d = C_3 t + C_4 - 4\pi k \frac{t^2}{2} \tag{2.66}$$
$$C_1 = C_3 - 4\pi k t \tag{2.67}$$

Similarly, equating $V(2)$ with $V(3)$ at $z = -t$ as also their derivatives, we get

$$-C_3 t + C_4 - 4\pi k \frac{t^2}{2} = C_5 d \tag{2.68}$$
$$C_3 + 4\pi k t = C_5 \tag{2.69}$$

Equations (2.66), (2.67), (2.68) and (2.69) are four linear, homogenous equations in four unknowns, C_1, C_3, C_4 and C_5. This kind of system of equations can be solved by using the standard method of determinants. However, in this case, the solution can be found by elimination. Eliminating C_1 from Eqns. (2.66) and (2.67) and C_5 from Eqns. (2.68) and (2.69), we get

$$C_3(t + d) + C_4 = 4\pi k\rho(\frac{t^2}{2} + td) \tag{2.70}$$

and

$$C_3(t + d) - C_4 = -4\pi k\rho(\frac{t^2}{2} + td) \tag{2.71}$$

which gives us

$$C_3 = 0$$

$$C_4 = 4\pi k\rho(\frac{t^2}{2} + td)$$

and

$$C_1 = -4\pi k\rho t$$

$$C_5 = 4\pi k\rho t$$

Thus, we have used the boundary conditions on the various surfaces to determine $C_1, C_2, \cdots, C_6$. This then, gives us the potential everywhere as

$$V(1) = -4\pi k\rho t(z - d - t)$$
$$V(2) = 4\pi k\rho t\left(\frac{t}{2} + td\right) - 4\pi k\rho\frac{z^2}{2}$$
$$V(3) = 4\pi k\rho t(z + d + t)$$

PROBLEM 2.10 Show that in two dimensions, the function $f(r, \theta) = Ar^n \cos(n\theta + \delta)$, where A and δ are constants and n is any positive or negative integer, satisfies Laplace's equation.

2.6 ENERGY OF THE ELECTROSTATIC FIELD

We know that any collection of masses possesses gravitational potential energy. This energy can be thought of as that which is required to assemble the masses, which are initially at infinity, to their present configuration against their mutual forces. One can similarly think of the potential energy in the gravitational field. A very similar situation exists in electrostatics where again, charges exert forces on each other. Hence, to assemble any collection of charges, which are initially at infinity, one needs to do work which is then stored as the electrostatic potential energy.

As a simple illustration of this concept, consider a uniformly charged sphere, of radius R and charge density ρ, as shown in Fig. 2.19. We want to calculate the total energy required to assemble such a charge configuration from small bits, which are initially at infinity. To do this, we bring in charges in the form of infinitesimal charged spherical shells from infinity and build up the sphere.

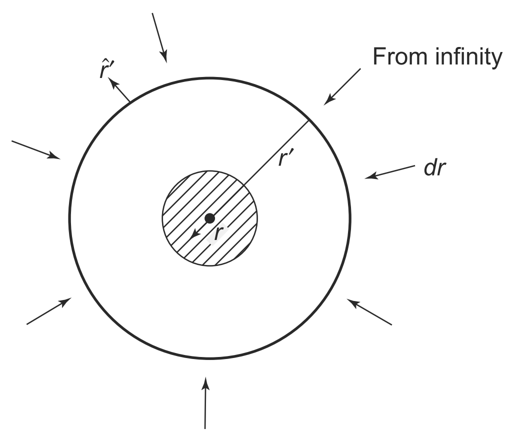


Fig. 2.19

At some intermediate stage in this building up process, the sphere has been partially built up to a radius r and another shell, of thickness dr is being brought from infinity. At this intermediate stage, the shell which is being brought from infinity is at a radial distance r'. Obviously, the final value of r' is r while the initial value is ∞; the initial value of r is 0 and the final value R. All the points on the spherical shell at r' experience an electric field due to the charged, partially built up sphere of radius r. All points in the shell are outside the uniformly charged sphere of radius r (since $r' > r$) and so they experience an electric field given by

$$\vec{E}(\vec{r}') = \left(\frac{4\pi\rho r^3}{3}\right)\frac{k}{r'^2}\hat{r}' \tag{2.72}$$

where $\hat{r}'$ is a unit vector in the radially outward direction at the shell and $\left(\frac{4\pi\rho r^3}{3}\right)$ is the charge of the partially built-up sphere. In the process of collapsing, i.e., bringing the shell from ∞ to r, all points in the shell move along the direction of -$\hat{r}'$. The charge on the shell is obviously $dq = 4\pi\rho r^2 dr$ since this shell ultimately forms a layer of thickness dr around the shell. The work done in bringing this shell from $\vec{r}'$ to $\vec{r'} - d\vec{r}'$ is (the displacement is $-dr'\hat{r}'$, since the shell is moving inwards)

$$\begin{aligned} \text{Work done} &= dq\vec{E} \cdot (\hat{r}' dr') \\ &= 4\pi\rho r^2 dr \frac{k4\pi\rho r^3}{3r'^2} dr' \end{aligned} \tag{2.73}$$

The total work done in bringing this shell from $r' = \infty$ to $r' = r$, $dW(r)$ is simply obtained by integrating Equation (2.73) from $r' = \infty$ to $r' = r$.

$$\begin{aligned} dW(r) &= k\frac{16}{3}\pi^2\rho^2 r^5 dr \int_{\infty}^{r} \frac{dr'}{r'^2} \\ &= k\frac{16}{3}\pi^2\rho^2 r^4 dr \end{aligned} \tag{2.74}$$

In the entire process of building up the sphere, we need to increase r from 0 to R. Thus, the total work done is

$$\begin{aligned} W(R) &= \int_0^R dW(r) \\ &= \int_0^R k\frac{16}{3}\pi^2\rho^2 r^4 dr \\ &= k\frac{16}{3}\pi^2\rho^2\frac{R^5}{5} \end{aligned} \tag{2.75}$$

The total charge on the sphere of radius R is $Q = \dfrac{4\pi\rho R^3}{3}$ and so we get

$$W(R) = \frac{3kQ^2}{5R} \tag{2.76}$$

This is very similar to the analogous result obtained for the gravitational energy of a sphere of radius R. Also, notice that in the limit $R \to 0$, the potential energy becomes infinite. That is, the electrostatic potential energy of a point charge Q is infinite. This again, is similar to the situation in gravity. It turns out that the self-energy of a charge being infinite is a serious problem, both at the classical as well as quantum level. Of course, a point charge was an idealisation that we used to simplify things. We will circumvent this problem altogether, by always considering continuous charge distributions of the kind used to obtain Eq. (2.76).

The uniformly charged sphere is obviously a special case of an object with enough symmetry to allow us to do an analytical calculation of the electrostatic potential energy. For a charged object of arbitrary shape, the procedure followed for calculation of potential energy for a uniformly charged sphere, cannot be repeated. For some regular shapes, it is possible to perform analytical calculation, as the following example will show.

EXAMPLE 2.11 Calculate the electrostatic potential energy of a uniformly charged thin spherical shell of radius R carrying a total charge Q.

Solution

We follow a variant of the procedure that we followed for the charged sphere. We build up the spherical shell by bringing in infinitesimal charge dq in the form of a spherical shell, all the way from infinity to the shell. We consider that the shell becomes smaller and smaller as it moves radially inwards. At any intermediate stage, the shell has a charge q and radius R. At that stage, the electric field at any point at a distance r from its centre is

$$\vec{E}(\vec{r}) = \frac{kq}{r^3}\vec{r} \tag{2.77}$$

The charge dq on the incremental shell that is being brought in experiences a force

$$\vec{F}(\vec{r}) = dq\,\vec{E}(\vec{r}) \tag{2.78}$$

The motion of each point on the shell is radially inwards, i.e., along $(-\vec{r})$ and a small displacement hence is $-d\vec{r}$. Hence, the total amount of work done in bringing the charge dq from infinity to the shell of radius R is

$$dW = -dq\int (\vec{E}(\vec{r})\cdot(-d\vec{r})) = dqkq\int_{\infty}^{R}\frac{dr}{r^2} = kq\frac{dq}{R} \tag{2.79}$$

The total work done in building up the total charge Q in these small steps of dq therefore, is

$$W = \int dW = \frac{k}{R}\int_{0}^{Q} q\,dq = \frac{kQ^2}{2R} \tag{2.80}$$

2.6.1 Potential and Potential Energy

We have seen above that it is not possible define the potential energy in a finite way for a point particle carrying a charge. In most cases involving protons and electrons, this is not a problem as far as the principle of conservation of energy is concerned. This is because, in classical physics, the number of electrons and protons does not change in any process involving the transfer or movement of charges. This is not true when one starts understanding the processes using quantum mechanics where particles can be created and destroyed, which is not the case in classical physics.

One way to look at this is that whatever the potential energy the point charged particle had acquired in the process of its being built up from infinitesimal charges at infinity, will stay with it. In any process, involving their motion or transfer, since the number of electrons and protons does not change, this potential energy does not change and can be considered as a constant factor. The only changes that we observe are the changes in the potential energy, which the charges acquire in the process of moving in the electrostatic fields created by other charges. This change is finite and well defined. When we speak of the potential energy of a set of two or more particles, it is this energy that is required to bring the particles to their final positions from infinity in the electric field of other particles. It is precisely this change in the potential energy which enters the Conservation of Energy principle- the other part, which is the energy of the individual charged point particle, remaining constant in any process and so does not matter in classical physics.

Let us find the electrostatic potential energy of a system of two point charges q_1 and q_2 located at $\vec{r}_1$ and $\vec{r}_2$, as shown in Fig. 2.20.

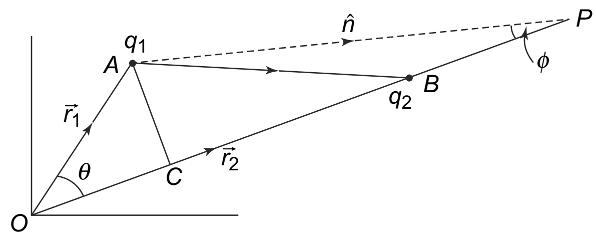


Fig. 2.20 *Electrostatic energy of two point charges*

As discussed above, the potential energy of this system of charges is simply the work done in bringing q_2 from ∞ to $\vec{r}_2$, keeping q_1 fixed at $\vec{r}_1$. Alternatively, we could keep q_2 fixed while moving in q_1 from ∞. Let us consider holding q_1 fixed at $\vec{r}_1$ and bringing in q_2 from ∞ along $-\hat{r}_2$. At a point P, a distance x from the origin along the vector $\vec{r}_2$, the electric field experienced by q_2 due to q_1 is

$$\vec{E}(x) = \frac{kq_1}{(AP)^2}\hat{n} \tag{2.81}$$

where AP is the distance as indicated in Fig. 2.20 and $\hat{n}$ is the unit vector along AP. The field can be written, using the cosine rule as

$$\vec{E}(x) = \frac{kq_1}{r_1^2 + x^2 - 2r_1 x\cos\theta}\hat{n} \tag{2.82}$$

where θ is the angle between $\vec{r}_1$ and $\vec{r}_2$. The force experienced by q_2 in this field is of course $\vec{F} = q_2\vec{E}$. If we make an infinitesimal displacement from x to $x - dx$ along PO, which is along $\hat{r}_2$, then the displacement is $-dx\hat{r}_2$. The work done in making this displacement is

$$dU = q_2|\vec{E}|(-dx)\cos\phi$$

where ϕ is the angle between PA and PO i.e., between $-\hat{n}$ and $-\hat{r}_2$. $\cos\phi$ is easily determined in terms of x by dropping a perpendicular AC from A to the line OP. Thus,

$$\cos\phi = \frac{PC}{AP} = \frac{(x - r_1\cos\theta)}{(x^2 + r_1^2 - 2xr_1\cos\theta)^{1/2}} \tag{2.83}$$

Integrating dU from $x = \infty$ to $x = r_2$, we get the potential energy of the system.

$$\boxed{\begin{aligned} U(q_1, \vec{r}_1, q_2, \vec{r}_2) &= kq_1q_2\int_{\infty}^{r_2} \frac{-dx(x - r_1\cos\theta)}{(x^2 + r_1^2 - 2xr_1\cos\theta)^{3/2}} \\ &= \frac{kq_1q_2}{r_{12}} \end{aligned}} \tag{2.84}$$

where r_{12} is the separation between the charges given by

$$r_{12} = (r_1^2 + r_2^2 - 2r_1r_2 \cos\theta)^{1/2}$$

The expression given in Eq. (2.84) is easily extended to the situation when many point charges are present. Consider a situation with N charges. The charges $q_1, q_2, \cdots, q_N$ be located at $\vec{r}_1, \vec{r}_2, \cdots, \vec{r}_N$. Starting with q_1 at $\vec{r}_1$, we bring in charges $q_2, q_3, \cdots, q_N$, one by one from infinity to their final positions $\vec{r}_2, \vec{r}_3, \cdots, \vec{r}_N$. When q_2 is brought in we get the potential energy $U(q_1, \vec{r}_1, q_2, \vec{r}_2)$ as in Eq. (2.84).

Now with q_1 and q_2 in position, we can bring in q_3 from infinity. In doing this q_3 obviously experiences a force due to both, q_1 and q_2. But we have seen in the previous chapter that the Principle of Superposition gives us that the total, net force on q_3 due to the two charges q_1 and q_2 will be just the vector sum of the individual forces, i.e., $\vec{F} = \vec{F}_1 + \vec{F}_2$ where $\vec{F}_1$ is the force on q_3 due to charge q_1 alone and $\vec{F}_2$ is the force due to q_2 alone. The work done, similarly, in moving the charge q_3 a distance $\vec{ds}$ is simply $\vec{F} \cdot \vec{ds}$, which is just $\vec{F}_1 \cdot \vec{ds} + \vec{F}_2 \cdot \vec{ds}$. What we have seen is that the total work done in moving q_3 in the combined field of q_1 and q_2 is just the sum of the work done against the forces due to q_1 and q_2 separately. Thus, the total work in bringing q_3 from ∞ to $\vec{r}_3$ is

$$U(q_1, \vec{r}_1, q_3, \vec{r}_3) + U(q_2, \vec{r}_2, q_3, \vec{r}_3)$$

To get the total potential energy of the system of three charges, we need to add this to the potential energy of the two charges q_1, q_2. Thus,

$$U_3 = U(q_1, \vec{r}_1, q_2, \vec{r}_2) + U(q_1, \vec{r}_1, q_3, \vec{r}_3) + U(q_2, \vec{r}_2, q_3, \vec{r}_3) \tag{2.85}$$

We can of course continue this process in a similar fashion and bring in q_4 which now experiences the combined field of q_1, q_2, q_3. It is straightforward to see that

$$U_4 = \sum_{\substack{j=i+1,i+2,i+3,i+4 \\ i=1,2,3,4}} U(q_i, \vec{r}_i, q_j, \vec{r}_j) \tag{2.86}$$

Continuing this for all the charges, we see that the total potential energy of a system of N charges can thus, be found to be

$$\boxed{\begin{aligned} U_N &= \sum_{\substack{j=i+1 \\ i=1}}^{\substack{j=N \\ i=N-1}} U(q_i, \vec{r}_i, q_j, \vec{r}_j) \\ &= \sum_{\substack{j=i+1,\ldots,N \\ i=1,\ldots,N-1}} \frac{kq_iq_j}{r_{ij}} \end{aligned}} \tag{2.87}$$

where r_{ij} is given by

$$r_{ij} = (r_i^2 + r_j^2 - 2r_ir_j\hat{r}_i \cdot \hat{r}_j)^{1/2}$$

$\hat{r}_i \cdot \hat{r}_j$ is the cosine of the angle between $\vec{r}_i$ and $\vec{r}_j$.

This expression for the potential energy of a collection of N charges is a general one. It is possible to evaluate this potential analytically for a few charges, which are regularly placed as the example below will illustrate.

EXAMPLE 2.12 Four equal charges q are placed at the four corners of a square of side a, as shown in Fig. 2.21. Calculate the potential energy of this system.

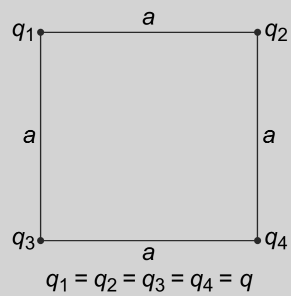


Fig. 2.21 *Example 2.12*

Solution

In this case, there are four charges in the problem, and so there will be 6 terms in Eq. (2.87). Thus,

$$\begin{aligned} U_4 &= kq^2 \sum_{\substack{j=i+1,\dots,4 \\ i=1,2,3}} \frac{1}{r_{ij}} \\ &= \frac{kq^2}{a}\left(1 + \frac{1}{\sqrt{2}} + 1 + 1 + \frac{1}{\sqrt{2}} + 1\right) \\ &= \frac{kq^2}{a}\left(4 + \sqrt{2}\right) \end{aligned}$$

since $r_{12} = r_{24} = r_{34} = r_{13} = a$ and $r_{23} = r_{14} = \sqrt{2}a$ and $q_1 = q_2 = q_3 = q_4 = q$.

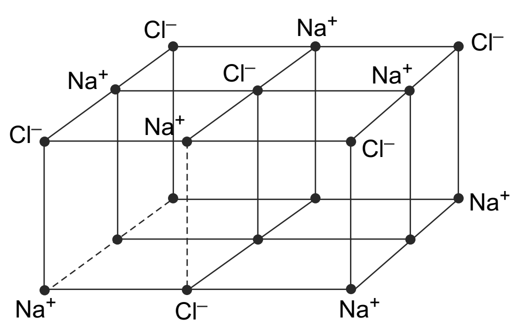


Fig. 2.22 *Sodium Chloride Crystal*

This example that we have considered of a set of charges placed in a regular arrangement (in this case, on the vertices of a square) is encountered frequently in nature. For instance, regular arrangements of charged ions exist in nature in the form of ionic crystals. The electrostatic forces, between the positive and negative ions located at regular intervals, are largely responsible for holding the ions together in the crystal. A well known example of such an ionic crystal is common salt or Sodium Chloride (NaCl). In the salt crystal, sodium (Na^+) and chloride (Cl^-) ions occupy alternate vertices of a cube of side $a = 5.64$ AU (Fig. 2.22).

The electrostatic energy of such an arrangement can in principle be calculated using Eq. (2.87). The practical problem is that even for a small crystal, the number of terms N in the sum in Eq. (2.87), which is the number of charges, is of the order of Avogadro's number or 6×10^{23}, which makes any practical computation impossible. One can try to get a flavour of such a computation by considering only a linear array.

EXAMPLE 2.13 Consider a hypothetical situation where 4 Na^+ and Cl^- ions are placed alternately along a line with equal spacing a. Calculate the electrostatic energy per NaCl molecule. (Ref: M. Elertand & E Kobeck, Journal of Chemical Education, 63, 840 (1986)).

Solution

From the Fig. 2.23, it is clear that charges with separation a, $3a$, $5a$ and $7a$ have opposite signs and hence contribute with a negative sign to the sum in Eq. (2.87) while those at $2a$, $4a$, $6a$ contribute with a positive sign. The terms involving the first Na^+ ion are

$$-\frac{e^2}{a}\left[1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\frac{1}{5}-\frac{1}{6}+\frac{1}{7}\right]$$

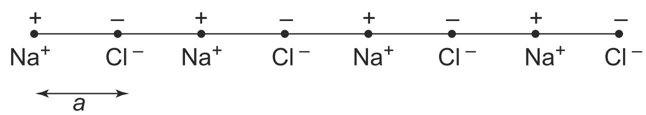


Fig. 2.23 *Example 2.13*

Now notice that the terms involving the second Cl^- ion are exactly the same except for the last term $(\frac{1}{7})$. The next contribution is from the Na^+ ion and this is also similar with the last two terms $(\frac{1}{6}$ and $\frac{1}{7})$ missing. Thus, the total electrostatic energy given by the Eq. (2.87) is simply

$$U_8 = -\frac{e^2}{a}\left[7-\frac{6}{2}+\frac{5}{3}-\frac{4}{4}+\frac{3}{5}-\frac{2}{6}+\frac{1}{7}\right]$$

which gives us the electrostatic energy for 4 NaCl molecules as

$$U_8 = -\frac{e^2}{a}(5.08)$$

Thus, the energy per molecule is simply

$$U = -1.27\frac{e^2}{a}$$

PROBLEM 2.11 In the experiment done by Rutherford to determine the structure of nuclei, a beam of α particles which has twice the charge of proton and approximately four times as heavy, was directed towards a gold leaf. Each of the α particles had an energy of 8 MeV and some of the α particles which hit a gold nucleus, having 79 times the charge of a proton, head on, were turned backwards. What was the closest separation of the α particles and the gold nuclei? What was the potential energy of the gold nucleus and the α particle at that point?

PROBLEM 2.12 $ABCD$ is a rectangle with sides AB and CD are parallel and of length L where parallel sides AC and BD are of length $2L$. Charges $+q$, $-q$, $+q$ and $-q$ are placed respectively at A, B, C and D. How much work will be done in interchanging the charges at A and B.

2.6.2 Electrostatic Binding Energy of a NaCl Molecule in a Crystal ∗

The simplified model we used for calculating the energy of the NaCl molecule, though instructive, is not useful for studying the actual properties of the ionic crystal. The actual calculation of the electrostatic binding energy per molecule in a realistic three-dimensional picture is somewhat more complicated than the simplified one-dimensional picture presented above. The number of ions is, of course, very large, or rather infinite for all practical purposes. However, the calculation of such energies is very important for the study of ionic crystals and considerable effort has gone into doing these computations.

Consider an ion, Na^+ or Cl^- in a NaCl crystal. It has a face-centred cubic symmetry, which means that the ions are located at intervals of a distance a regularly in the three directions at right angles to each other. We can take these 3 directions as x, y and z directions. Let us start from a particular ion at a certain location, which we take as the origin. The location of other ions are then (ma, na, pa) where m, n, p are integers, both positive or negative. The case when $m = n = p = 0$ is excluded since that is the starting location. In reaching another ion at such a location, the number of steps of a each is $(m + n + p)$. Since positive and negative ions alternate, we will reach an ion of the same sign as the one at the origin, if $(m + n + p)$ is even, and of opposite sign if this number is odd. The distance of that ion from the origin will be $a(m^2 + n^2 + p^2)^{1/2}$. Hence, in the expression for the electrostatic energy, Eq. (2.87), the terms involving the ion at the origin are:

$$U_1 = -\sum ke^2 \left(\frac{(-1)^{(m+n+p)}}{a(m^2 + n^2 + p^2)^{1/2}} \right) \tag{2.88}$$

where the summation is carried out over all possible values of m, n and p (both positive and negative), except $m = n = p = 0$.

The same of course will be true of any other ion. If we multiply this by the total number of ions we however will get twice the total energy since all pairs of ions would have been counted twice in the process. Thus, U_1 is the energy per ion since a molecule consists of two ions and hence U_1 is the electrostatic energy per molecule U(molecule). This is normally expressed as

$$U(\text{molecule}) = -\frac{ke^2}{a} M \tag{2.89}$$

where M is called the Madelung constant, named after the German physicist Erwin Madelung (1881-1972). From Eq. (2.88), one can see that

$$M = \sum \left(\frac{(-1)^{(m+n+p)}}{(m^2 + n^2 + p^2)^{1/2}} \right) \tag{2.90}$$

The summation is over an infinite series and therefore one has problems. We can easily see what the first few terms of the series for M look like. The lowest value of the denominator is 1 and there are six such terms corresponding to 1- the three integers taking on value $+1$ or -1 and the other two zero. The next value of the denominator is 2 and there are 12 such terms with $m = 0$, $n = 1$ or -1, $p = 1$ or -1 and interchange of these numbers that result in different values. We can go on a few more steps and the result is

$$M = -6 + \frac{12}{\sqrt{2}} - \frac{8}{\sqrt{3}} + \frac{6}{2} - \frac{24}{\sqrt{5}} + \dots \tag{2.91}$$

This series can be summed but not easily, and the result is -1.74756 (See e.g., **C. Kittel, Introduction to Solid State Physics, Wiley India, 2010**). For NaCl, the separation a between nearest ions is 2.82 A.U. If we use this value of M, we can easily estimate the energy U per molecule. $-U$ is called the lattice energy and is usually expressed as energy/mol. We get

$$-U/\text{mole} = 861 \quad \text{kJ/mol}$$

This is somewhat higher than the what is determined experimentally. The difference is not unexpected, since treating the ions as point charge as we have done is clearly an approximation because the ions themselves are charge distributions. Better estimates taking care of these are available in literature. (See for a pedagogic review: **A.D.Baker and M.D.Baker,** ***Am. J of Physics*****, 78,102 (2010)**.)

There is also an interesting and important mathematical point regarding summation of infinite series of the type that is involved in evaluating M. The sign of the terms alternate and the series for M does not have the property of what is called absolute convergence. This creates problems in numerical computation unlike absolutely convergent series.

To see this point, consider an infinite one dimensional ionic chain [discussed in **E.L.Burrows and S.F.A. Kettle,** ***Journal of Chemical Education*****, 52, page 58 (1973)**] as shown in Fig. 2.23.

Taking any ion as a reference point, the total energy can be written as

$$\left(\frac{e^2}{a}\right)2\left[1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\cdots\right]$$

This by definition is $\left(\frac{e^2}{a}\right)M$, where M is the Madelung constant. Summing up, we get $M = 2\ln 2$. However, let us rearrange the terms in the series as follows:

$$M = 2\left[1-\frac{1}{2}-\frac{1}{4}+(\frac{1}{3}-\frac{1}{6})-\frac{1}{8}+\ldots+(\frac{1}{2n-1}-\frac{1}{4n-2})-\cdots\right]$$

We get

$$M = \left[1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\cdots\right] = \ln 2$$

which is half the earlier value.

Note that all we have done is to rearrange the terms. This is an example of what is known in mathematics about summing infinite series. In an absolute convergent series, rearrangement of terms does not change the result. In a conditionally convergent series, this is not true and the sum is dependent on the way we add the terms. In the present case, the way to sum is dictated by physical arguments. We should sum over cells around an ion and then go over to bigger cells around so that the infinite series is evaluated as the limit of a finite series in the limit of the number of terms becoming very large.

2.6.3 Electrostatic Energy of Electric Field

Energy is a very useful concept in physics since on general grounds the total energy of any isolated system does not change with time. In this way of stating the energy conservation principle, any change in any part of the system at any time is necessarily accompanied at exactly the same time by a compensating change in other parts. However, different parts of the system are not at the same point and the special theory of relativity tells us that no signal can be transmitted instantaneously between the various parts of the system. We have already encountered this problem when we discussed the electric field. Recall that Coulomb's Law assumed an instantaneous action at a distance between charges interacting through the electrostatic force. To circumvent the problem of this being in violation of relativity, we introduced the concept of the electric field which was a purely local construct. Thus, just as it was necessary to introduce the concept of an electric field to rephrase Coulomb's Law so as to be consistent with relativity, one needs to think of the principle of conservation of energy as being applicable locally. Change in energy in an infinitesimal volume must be related to some energy flow across its surface.

In the context of pure electrostatics where charges do not move, discussion of assigning the energy to something or other is more or less meaningless, since everything is stationary. Nonetheless, it is a valid question to ask whether the potential energy of charges, Eq. (2.87) can be thought of as the energy of the accompanying electric field. We will attempt to answer this question here.

Let us begin with the expression for energy for a collection of charges. Eq. (2.87) can be rewritten as

$$U = \frac{1}{2} k \sum \sum \frac{q_i q_j}{r_{ij}} \tag{2.92}$$

where now the sum runs over i, j, both from 1 to N, with the only restriction that $i \neq j$. The factor $\frac{1}{2}$ comes in because in this way of writing, each pair comes in twice unlike Eq. (2.87). The last equation (Eq.(2.92)) can be rewritten in terms of the potential ϕ_i present at the location of the charge q_i

$$\phi_i(r_i) = k \sum \frac{q_j}{r_{ij}}$$

where the summation runs over $j = 1$ to N except $j = i$.

Using this expression for the potential, we get

$$U_N = \frac{1}{2} \sum q_i \phi_i \tag{2.93}$$

We have already mentioned that for calculation of potential energy we shall not deal with point charges. We will therefore, generalise the last equation to a case where instead of point charges, we have a distribution of charges with density $\rho(r)$.

Consider the situation where space is divided into two regions—Region 1 and Region II as in Fig. 2.24. Region 1 has a charge distribution while Region II is charge free. Now an element of volume dV at r has charge $\rho(r)dV$ and if $\phi(r)$ is the electrostatic potential at r, the total potential energy of the distribution of charges is

$$U_1 = \frac{1}{2} \int_1 \rho(\vec{r})\phi(\vec{r}) dV \tag{2.94}$$

where the integration is over the volume of the Region 1 where the charge distribution is located.

But we know that ρ is related to the divergence of the electric field (Gauss's Law) which in turn is

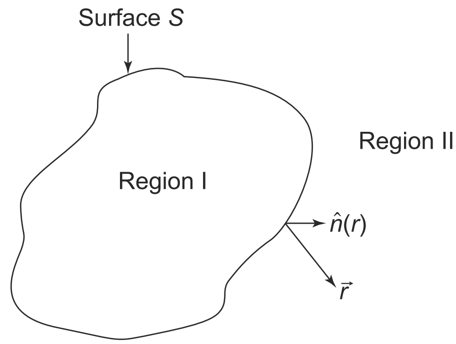


Fig. 2.24 *Charges are present in region I. $\hat{n}(r)$ is the outward drawn normal to the surface S enclosing region I at a point r on the surface. Region II has no charges. It is bounded by the surface S and the surface at infinity. For Region II, the outward drawn normal on its bounding surface S at the point r is $-\hat{n}(r)$*

related to $\phi(\vec{r})$ by its definition, $\vec{E}(\vec{r}) = -\vec{\nabla}\phi(\vec{r})$. Therefore, we get

$$\begin{aligned}\rho(\vec{r}) &= \frac{1}{4\pi k}\vec{\nabla}\cdot\vec{E}(\vec{r}) \\ &= -\frac{1}{4\pi k}\vec{\nabla}\cdot(\vec{\nabla}\phi(\vec{r})) \\ &= \frac{1}{4\pi k}(-\nabla^2\phi(\vec{r}))\end{aligned} \tag{2.95}$$

and hence

$$U_1 = \frac{1}{8\pi k}\iiint dV(-\nabla^2\phi(\vec{r}))(\phi(\vec{r})) \tag{2.96}$$

To simplify the integrand, we can use the identity

$$-\nabla^2\phi(\vec{r})\phi(\vec{r}) = -\vec{\nabla}\cdot\left[(\vec{\nabla}\phi(\vec{r}))(\phi(\vec{r})\right] + \left[\vec{\nabla}\phi(\vec{r})\right]^2$$

to get (since $\vec{E}(\vec{r}) = -\vec{\nabla}\phi(\vec{r})$)

$$-\nabla^2\phi(\vec{r})\phi(\vec{r}) = -\vec{\nabla}\cdot\left[(\vec{\nabla}\phi(\vec{r})(\phi(\vec{r}))\right] + \left[\vec{E}(\vec{r})\right]^2 \tag{2.97}$$

On using this identity in Eq. (2.96) we get

$$U_1 = \frac{1}{8\pi k}\iiint dV\left[E(\vec{r})\right]^2 - \frac{1}{8\pi k}\iiint dV\vec{\nabla}\cdot\left[(\vec{\nabla}\phi(\vec{r}))\phi(\vec{r})\right] \tag{2.98}$$

The right-hand side has two terms. The second term is a volume integral involving a divergence. Hence, we can use the divergence theorem to convert this into a surface integral over the surface bounding the volume

$$U_1 = \frac{1}{8\pi k}\iiint_1 dV\left[E(\vec{r})\right]^2 - \frac{1}{8\pi k}\iint_S dS\hat{n}(r)\cdot\left[(\vec{\nabla}\phi(\vec{r}))(\phi(\vec{r}))\right] \tag{2.99}$$

where the volume integral is over the volume marked Region 1 in Fig. 2.24 and the surface integral is over its bounding surface S shown in the figure.

Region II has no charges and hence an integral U_{II} similar to Eq. (2.94) but over the Region II is zero. If we repeat the steps leading from Eq. (2.94) to Eq. (2.99), we get

$$U_{II} = \frac{1}{8\pi k}\iiint_{II} dV\left[E(\vec{r})\right]^2 + \frac{1}{8\pi k}\iint_{S} dS\hat{n}(r)\cdot\left[(\vec{\nabla}\phi(\vec{r}))(\phi(\vec{r}))\right] \tag{2.100}$$

Note that the second term has a sign different from Eq. (2.99) because for Region II, the normal to its bounding surface S is $-\hat{n}(r)$. The second bounding surface of Region II is at infinity where the integrand vanishes and thus, the surface integral there is zero anyways.

Adding Eq. (2.99) and Eq. (2.100), we see that the second term in the right-hand side cancels. The first terms in the rhs of these equations have the same integrand but over Regions I and II. Thus, on adding they result in a volume integral over the entire space. We thus, get an expression for the energy solely as a volume integral over entire space. We drop the subscript I now and thus, the total electrostatic energy can be written as

$$\boxed{U = \frac{1}{8\pi k}\iiint dV E^2(\vec{r})} \tag{2.101}$$

where the integral is over all space. With this expression, we can see that the electrostatic energy can thus, be regarded as the energy of the electric field with an energy density given by $\frac{1}{8\pi k}E^2$. The expression for energy of the electric field was derived here under electrostatic conditions. Later in Chapter 10, we shall discuss the more general case for the energy density.

EXAMPLE 2.14 Calculate the total energy of the electric field caused by a uniformly charged sphere of radius R and charge Q.

Solution

The electric field outside the sphere is given by

$$\vec{E}_{\text{out}}(\vec{r}) = \frac{kQ}{r^2}\hat{r} \qquad\qquad r > R \tag{2.102}$$

Inside the sphere it is given by

$$\vec{E}_{\text{in}}(\vec{r}) = \frac{kQr}{R^3}\hat{r} \qquad\qquad r < R \tag{2.103}$$

The electric field is radial in both regions and has no dependence on the other coordinates θ and ϕ. The energy of the electric field is thus,

$$\begin{aligned} U &= \frac{k^2Q^2}{8\pi k}\left(\int_{r>R}\frac{d^3r}{r^4} + \int_{r<R}\frac{d^3r r^2}{R^6}\right) \\ &= \frac{k^2Q^2 4\pi}{8\pi k}\left(\int_R^\infty \frac{dr}{r^2} + \int_0^R \frac{dr r^4}{R^6}\right) \\ &= \frac{3kQ^2}{5R} \end{aligned} \tag{2.104}$$

where we have used the fact that integrand is independent of θ and ϕ and hence the angular integration gives us 4π. Note that this result is the same that we obtained as the energy required to assemble the uniformly charged sphere, Eq. (2.76) as it should be.

EXAMPLE 2.15 A spherical shell of radius R is uniformly charged with a charge $-Q$. A second shell, concentric with the first one, but with a radius $2R$, is charged uniformly to a charge Q. Calculate

(a) the potential energy of the two shells,
(b) the electric field everywhere and
(c) the total energy in the electric field created by the two shells.

Solution

1. The potential energy of the inner shell is

$$U = \frac{kQ^2}{2R}$$

After the inner sphere is in place, imagine building up the outer sphere in the following way—bringing infinitesimal quantity of charge dq from infinity to $2R$. Then the total work done in building up the charge on the outer sphere to Q will be the potential energy of the outer sphere. At the instant when the charge on the outer shell is q, the work done in bringing an additional charge dq from infinity to $r = 2R$ is given by

$$W = k(-Q+q)\int_{\infty}^{2R} \frac{-dr}{r^2} dq$$

which is

$$W = k(-Q+q)\frac{dq}{2R}$$

The total work done in assembling the charge on the outer shell is obviously then

$$\int_0^Q k(-Q+q)\frac{dq}{2R}$$

which is

$$-\frac{kQ^2}{4R}$$

The total potential energy of the two shells is therefore,

$$\frac{kQ^2}{2R} + (-\frac{kQ^2}{4R}) = \frac{kQ^2}{4R}$$

2. By Gauss's Law, the electric field is restricted to the region between the two shells. Its magnitude is easily seen to be $\frac{kQ}{r^2}$

3. The energy of the electric field is simply the integral of the energy density over all space. However, the energy density is

$$\frac{E^2}{8\pi k}$$

and is only non-zero in the region between the two spherical shells. Therefore, the total energy of the electric field is

$$\frac{1}{8\pi}\int_R^{2R}\left(\frac{kQ^2}{r^4}\right)4\pi r^2 dr = \frac{kQ^2}{4R}$$

which is what we found as the potential energy of the shells.

2.7 ADVANCED TOPICS

2.7.1 Boundary Value Problems

We have already seen how to solve an electrostatic problem, once the charge distribution everywhere is specified. One uses Coulomb's Law and the Principle of Superposition to evaluate the field everywhere. In particular, we know that the field is given by

$$\vec{E}(r) = \frac{1}{4\pi\varepsilon_0}\int \frac{d^3r'\rho(\vec{r}')(\vec{r}-\vec{r}')}{|\vec{r}-\vec{r}'|^3} \tag{2.105}$$

Once one knows the field, one can determine the potential ϕ. Alternatively, one can solve for the potential and obtain the field.

But frequently, one does not know the charge distribution everywhere and so it is not possible to use this strategy. For instance, one may be given a system of conductors where the potential is specified on each conductor but the charge distribution is not given, then the above mentioned approach is not very useful. For solving the general electrostatic problem, one needs to solve the Poisson (in the presence of charge) or Laplace (in a charge-free region) equation with the given boundary conditions. In general, one encounters two kinds of boundary conditions in electrostatics:

The first kind, also called the **Dirichlet Boundary conditions** are when one has specified the potential on a closed surface. For instance, one can imagine a system of conductors held at different potentials which then serve as boundary conditions which the solution to the Poisson or Laplace equation needs to satisfy.

The second kind of boundary conditions can be when one specifies the electric field on the surface. Recall that the electric field is the normal derivative of the potential. This kind of boundary condition is called the **Neumann boundary condition**. Thus, the two kinds of boundary conditions either specify the potential ϕ or its normal derivative on the surfaces. The essential problem then in electrostatics is to find the solutions to Poisson or Laplace equation with these boundary conditions.

In most cases, as we have seen, we have localised charge distributions or discrete point charges or charges on the surface of conductors. In these cases, the charge density at most points in space vanishes and so we need to solve the Laplace equation instead of the Poisson equation. Consider a set of conductors that are maintained at some constant potential $\phi_1, \phi_2, \cdots, \phi_n$. Then the electrostatic problem is simply finding the solution of Laplace's equation everywhere, which reduces to these potentials on the surfaces of the relevant conductors.

The solutions to the Laplace equation satisfy certain properties, which we can state in terms of two theorems.

If $\phi_1, \phi_2, \cdots, \phi_n$ are all solutions to the Laplace equation, then

$$\phi = a_1\phi_1 + a_2\phi_2 + \cdots + a_n\phi_n$$

where a_n's are arbitrary constants, is also a solution.

To prove this theorem, apply the Laplacian operator ∇^2 on ϕ. Then

$$\nabla^2\phi = a_1\nabla^2\phi_1 + a_2\nabla^2\phi_2 + \cdots + a_n\nabla^2\phi_n$$

which vanishes since each of the ϕ_n's satisfies the Laplace equation.

This is analogous to the Principle of Superposition that we used in the case of finding the fields. This result is very useful, since it allows us to superimpose two or more solutions to the Laplace equation to get another solution, which may satisfy the given boundary conditions.

The second result is the **Uniqueness Theorem.**

> The Uniqueness Theorem states that two solutions of Laplace's equation that satisfy the **same** boundary conditions can differ at the most by an additive constant. In other words, two solutions of the Laplace equation satisfying the same (i) Dirichlet boundary conditions are equal and (ii) Neumann boundary condition can differ by a constant.

To prove this theorem, let us consider a system of conductors with surfaces $S_1, \cdots, S_n$. We are looking for a solution of the Laplace equation with the specified boundary conditions on these surfaces. These surfaces are bounded by a closed surface Σ which encloses the region R, which is exterior to the surfaces. Now clearly, there is no charge density in R and so Laplace's equation is valid. The surface Σ can of course, be at infinity and thus, enclose all of space. Now assume that ϕ_1 and ϕ_2 are two solutions of the Laplace equation in R with the same boundary conditions on the surfaces $S_1, S_2, \cdots, S_n$. Note that the boundary conditions could be Dirichlet or Neumann type, i.e., we could either have specified the potential or the normal derivative on each of the surfaces.

Now consider another function ϕ, which is given by

$$\phi = \phi_1 - \phi_2$$

Clearly ϕ itself is a solution to the Laplace equation in R since

$$\nabla^2\phi = \nabla^2\phi_1 - \nabla^2\phi_2 = 0$$

Consider a vector $\phi\vec{\nabla}\phi$. Applying the divergence theorem to this vector field, we get

$$\iiint\limits_R \vec{\nabla}\cdot(\phi\vec{\nabla}\phi)dV = \iint\limits_{\Sigma+S_1+S_2+\cdots+S_n} \phi\vec{\nabla}\phi\cdot\hat{n}dS \tag{2.106}$$

since the bounding surface of R is Σ and the surfaces $S_1,\cdots,S_n$ of the conductors. The boundary conditions either specify the potential or its normal derivative (gradient) on the surfaces, $\phi_1=\phi_2$ or $\vec{\nabla}\phi_1\cdot\hat{n}=\vec{\nabla}\phi_2\cdot\hat{n}$ on the surfaces and hence, the surface integral vanishes over $S_1,\cdots,S_n$. We also take surface Σ to ∞ where the potentials and hence, the surface integral over Σ also vanishes. Therefore, we get

$$\iiint\limits_R \vec{\nabla}\cdot(\phi\vec{\nabla}\phi)dV = \iint\limits_{\Sigma+S_1+S_2+\cdots+S_n} \phi\vec{\nabla}\phi\cdot\hat{n}dS = 0$$

Therefore,

$$\iiint\limits_R \vec{\nabla}\cdot(\phi\vec{\nabla}\phi)dV = \iiint\limits_R \phi\nabla^2\phi dV + \iiint\limits_R (\nabla\phi)^2 dV = 0 \tag{2.107}$$

where we have used the vector identity

$$\vec{\nabla}\cdot(\phi\vec{A}) = \phi\vec{\nabla}\cdot\vec{A} + \vec{A}\cdot\vec{\nabla}\phi$$

But $\nabla^2\phi=0$ all over R since ϕ is a solution to the Laplace equation. Thus, we get

$$(\vec{\nabla}\phi)^2 = 0$$

everywhere in R since its integral vanishes.

What we have seen is that the gradient of ϕ vanishes everywhere in R. This implies that ϕ cannot change in the region and is a constant. In particular, its value is the same everywhere as it is on the surfaces. Now suppose that the boundary conditions are the Dirichlet boundary conditions, i.e., we have specified ϕ_1,ϕ_2 on the surfaces. If this is the case, it follows that $\phi=0$ on the surfaces. But we already have that $\phi=\phi_1-\phi_2$ cannot change in the region. Hence if it is zero on the surface, it must be zero everywhere. On the other hand, if the boundary conditions are of the Neumann kind, then we have specified the normal derivative of ϕ_1 and ϕ_2 on the surfaces and hence $\vec{\nabla}\phi\cdot\hat{n}$ is zero on the surfaces. This is only possible if ϕ is a constant, thereby proving the theorem that $\phi=\phi_1-\phi_2$ is either zero or a constant which means that ϕ_1 and ϕ_2 differ at most by an additive constant. This proves the Uniqueness Theorem.

2.7.2 Laplace Equation—Solution by Separation of Variables

The Laplace equation

$$\nabla^2\phi = 0$$

is a partial differential equation involving the three-dimensional Laplacian operator ∇^2. One of the ways of solving partial differential equations is by the method of **Separation of Variables.** We shall illustrate this method by considering the Laplace equation in rectangular coordinates. It turns out that the Laplace equation is separable in eleven coordinate systems, some of which we will be considering later.

The Laplace equation obeyed by the potential $\phi(x, y, z)$ in rectangular coordinates is given by

$$\nabla^2\phi = \frac{\partial^2\phi}{\partial x^2} + \frac{\partial^2\phi}{\partial y^2} + \frac{\partial^2\phi}{\partial z^2} = 0$$

To solve this, we assume that the solution can be represented as a product of three functions $X(x)$, $Y(y)$ and $Z(z)$. Thus,

$$\phi(x, y, z) = X(x)Y(y)Z(z)$$

Now we can substitute this ϕ in the Laplace equation to get

$$\frac{1}{X(x)}\frac{d^2X}{dx^2} + \frac{1}{Y(y)}\frac{d^2Y}{dy^2} + \frac{1}{Z(z)}\frac{d^2Z}{dz^2} = 0 \tag{2.108}$$

Each of the three terms above depend only on one component of coordinate: x, y or z, which are independent of each other. This can be true in general only if each of the three terms are constant.

$$\frac{1}{X}\frac{d^2X}{dx^2} = -k_x^2 \tag{2.109}$$

$$\frac{1}{Y}\frac{d^2Y}{dy^2} = -k_y^2 \tag{2.110}$$

$$\frac{1}{Z}\frac{d^2Z}{dz^2} = -k_z^2 \tag{2.111}$$

with the constraint that

$$k_x^2 + k_y^2 + k_z^2 = 0$$

These are of course, three ordinary differential equations and can be easily solved. The solutions are

$$X(k_x, x) \propto e^{ik_x x} \tag{2.112}$$

$$Y(k_y, y) \propto e^{ik_y y} \tag{2.113}$$

$$Z(k_z, z) \propto e^{ik_z z} \tag{2.114}$$

where of course not all the k's can be real. The solution to the Laplace equation therefore, becomes

$$\phi = e^{\pm ik_x x}e^{\pm ik_y y}e^{\pm\sqrt{k_x^2+k_y^2}z} \tag{2.115}$$

Please note that the constants k_x, k_y and k_z are completely arbitrary, with the constraint that their squares add up to zero. They will be fixed by the boundary conditions of the given problem.

As an example, let us consider a rectangular geometry so that the symmetry considerations dictate the use of the Laplace equation in rectangular coordinates.

Consider a rectangular box with dimensions (a, b, c) in the (x, y, z) directions. This is shown in Fig. 2.25. The box is such that all surfaces except one, are kept at zero potential. The surface at $z = c$ is kept at a potential $V(x, y)$. One needs to find the potential everywhere inside the box.

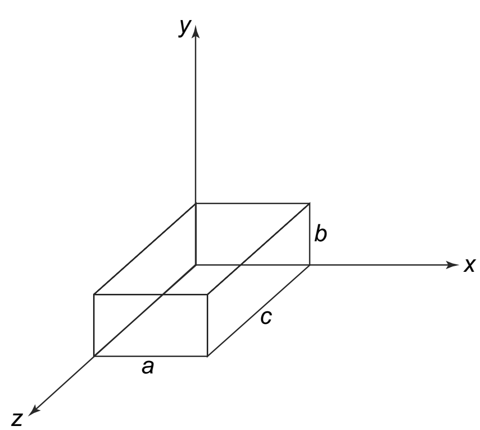


Fig. 2.25 *Laplace equation in rectangular coordinates*

There is obviously no charge inside the box and hence the potential everywhere inside satisfies the Laplace equation with the boundary conditions

$$\phi = 0 \quad \text{at} \quad x = 0, a$$
$$\phi = 0 \quad \text{at} \quad y = 0, b$$
$$\phi = 0 \quad \text{at} \quad z = 0$$

and

$$\phi = V(x, y) \quad \text{at} \quad z = c$$

Now since the periodic boundary conditions are on x and y, we should choose the periodic functions for $X(x)$ and $Y(y)$. Thus, we have

$$k_x = \frac{m\pi x}{a}$$

and

$$k_y = \frac{n\pi y}{b}$$

This gives us

$$k_z = \pm i\sqrt{\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2}$$

If we define

$$k_{m,n} = \sqrt{\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2}$$

then

$$k_z = \pm i k_{m,n}$$

Thus, we get

$$X(k_x, x) \quad \propto \quad \sin\frac{m\pi x}{a} \tag{2.116}$$
$$Y(k_y, y) \quad \propto \quad \sin\frac{n\pi y}{b} \tag{2.117}$$
$$Z(k_z, z) \quad \propto \quad e^{\pm k_{m,n} z} \tag{2.118}$$

Now, we know that $\phi = 0$ at $z = 0$ and therefore,

$$Z(k_z, z) \propto e^{k_{m,n} z} - e^{-k_{m,n} z} \propto \sinh(k_{m,n} z)$$

Thus, we have the solution as

$$\phi_{mn} \propto A_{m,n} \sin\frac{m\pi x}{a} \sin\frac{n\pi y}{b} \sinh k_{m,n} z$$

and therefore, the general solution will be a superposition of all the solutions

$$\phi(x, y, z) = \sum_{m,n} A_{m,n} \sin\frac{m\pi x}{a} \sin\frac{n\pi y}{b} \sinh k_{m,n} z$$

We still need to impose the remaining boundary condition, namely the potential at $z = c$. We thus have

$$\phi(x, y, c) = V(x, y) = \sum_{m,n} A_{m,n} \sin \frac{m\pi x}{a} \sin \frac{n\pi y}{b} \sinh k_{m,n} c$$

This is just a double Fourier series and therefore,

$$\int_0^a \int_0^b V(x, y) \sin \frac{m\pi x}{a} \sin \frac{n\pi y}{b} dx dy = A_{m,n} \frac{ab}{4} \sinh(k_{m,n} c)$$

and therefore,

$$A_{m,n} = \frac{4}{ab \sinh(k_{m,n} c)} \int_0^a \int_0^b V(x, y) \sin \frac{m\pi x}{a} \sin \frac{n\pi y}{b} dx dy$$

The Laplace equation, as we remarked in separable in eleven coordinate systems. In particular, the spherical and cylindrical coordinate systems are very useful since many of the problems in electrostatics have an inherent spherical or cylindrical symmetry. This we shall see in a later chapter.

SUMMARY

- The form of the Coulomb's Law enables one to introduce the concept of electric potential, which is a scalar quantity related to the electric field.
- Calculation of electric field in many cases gets simplified by calculating the potential first and then the field.
- The electric potential satisfies simple differential equations: Poisson's for regions where charge is present and Laplace where there is no charge.
- The charge-free Laplace equation can be solved in such regions. Effect of charge present elsewhere gets reflected in the boundary conditions of the solutions.
- Potential energy of a system of charges can be calculated and related to the electric potential. We can also think of the energy as being carried by the electric field.

CONCEPTUAL QUESTIONS

1. Suppose a uniform electric field exists in a room, such that the lines of force are horizontal and perpendicular to one wall. As one walks towards the wall from which the lines of force are emerging, is one walking towards

a. points of higher potential
b. points of lower potential
c. points of same potential, that is along an equipotential line.

2. When we place a positive charge in an electric field, it will accelerate from

a. higher to lower electric potential; lower to higher potential energy
b. higher to lower electric potential; higher to lower potential energy

c. lower to higher electric potential; lower to higher potential energy
d. lower to higher electric potential; higher to lower potential energy

3. If we move a positive charge from infinity to a point exactly mid-way between two point charges, $+q$ and $-q$, the work done in doing this is
a. positive
b. negative
c. zero
d. cannot be determined from the data given in the problem.

4. Can two equipotential surfaces intersect? If not, why not?

5. If the potential V is constant in a given region, what can be said about the electric field $\vec{E}$ in that region?

6. The work done by the force $\vec{F} = 4\hat{x} - 3\hat{y} + 2\hat{z}$ N in giving a charge of 1 nC a displacement of $10\hat{x} + 2\hat{y} - 7\hat{z}$ is
a. 103 nJ
b. 60 nJ
c. 64 nJ
d. 20 nJ

7. A point charge of 1 C is brought from infinity to a point a in free space. How much work is required to do this? If the energy of 1 J is expended in bringing a second charge of 1 C from infinity to a point b, what is the separation between the charges?

8. In a charge free region, $E_x = \alpha x$, $E_y = \beta y$. Find E_z.

9. The electric potential in a region is given by $V = 3x^2y - yz$. Which of the following statements is NOT true?
a. At point $(1, 0, -1)$, V and $\vec{E}$ vanish.
b. $x^2y = 1$ is an equipotential line on the xy plane.
c. The equipotential surface $V = -8$ passes through the point P (2, −1, 4).
d. The electric field at P is $12\hat{x} - 8\hat{y} - \hat{z}$.

10. When we say that the electric field is conservative, we do NOT mean that
a. It is the gradient of a scalar potential.
b. Its circulation is identically zero.
c. Its curl is identically zero.
d. The work done in a closed path is zero.
e. The potential difference between two points is zero.

PROBLEMS

1. Find the potential at a point P, which is at a perpendicular distance y from a uniformly charged rod of length l placed along the x-axis.

2. Find the potential at a point P, which is at a perpendicular distance z from a uniformly charged ring of radius R placed in the $x - y$ plane.
3. Find the potential at a point P, which is at a perpendicular distance z from a uniformly charged disc of radius R placed in the $x - y$ plane.
4. A thin rod is placed along the x-axis from $x = -d$ to $x = +d$. A charge Q is spread uniformly on the rod. Find the electric potential and the field at a point P on the x axis with $x > d$. Choose the electric potential to be zero at infinity.
5. An L-shaped rod extends from the origin of rectangular system to the point $(L, 0)$ on the x-axis, and from the origin to the point $(0, L)$ on the y-axis. The rod is charged uniformly with a linear charge density λ. Evaluate the electric field strength and the electric potential at points $P(a, a)$ and $Q(\frac{5a}{2}, 0)$.
6. Starting with N equal spherical drops of mercury with each drop at a potential V with respect to infinity, we combine all these small drops into a large spherical drop. What is the potential of the large drop?
7. Find the electric potential due to a uniformly charged cylinder of radius R and length L at a point on its axis, but outside it.
8. A dipole with dipole moment $p = 2qa$ is placed along the y-axis. Find the potential and the field at a point P at a distance r $(r \gg a)$ from the centre. The line joining the point P and the centre makes an angle of θ with the y-axis. What is the field for $\theta = 90°, 0°$?
9. A square sheet of side a is uniformly charged with a surface charge density σ. Show that the potential at the centre of the square is given by

$$\phi = 4k\sigma a \ln(1 + \sqrt{2})$$

10. An annular disc, with inner radius a and outer radius b has a uniform negative charge density $-\sigma$ where $\sigma > 0$. Calculate the potential at the centre of the disc (point P) assuming that the potential at infinity is zero. If an electron is released with a velocity v_0 from the centre of the disc, in the upward direction, find the velocity of the electron at infinity. Assume there are no other forces acting on the electron.
11. Find the electric field due to the following potentials:
    a. $V = \sin(x^2 + y^2 + z^2)^{1/2}$
    b. $V = \rho^2(z + 1)\sin\phi$
    c. $V = e^{-r}\sin\theta\cos 2\phi$
12. A region of space has an electric field given by $\vec{E} = E_0 x e^{-x}\hat{i}$. For what value of x is the field a maximum? What is the electric potential at this point w.r.t. the potential at $x = 0$?
13. An infinite charged sheet has a charge density $\sigma = 1.0 \times 10^{-7}$ coul/m^2. How far are the equipotential surfaces whose potentials differ by 5.0 volts?
14. The electric potential at a distance of 25 cm from a positive point charge is 10 V. What is the magnitude of the charge? Determine the radius of the equipotential surface where the potential is 20 V and 5 V.

15. A uniform electric field is give by $\vec{E} = 100\hat{i}$ kV/m. Find the potential at any point in space if the potential at the origin is zero.

16. Calculate the dipole moment of the water molecule assuming that the electrons in the molecule circulate symmetrically about the oxygen nucleus. The distance between the oxygen and hydrogen atoms is given to be 0.96×10^{-8} cm and the angle between the two oxygen-hydrogen bonds is $104°$.

17. A point charge $Q = 5$nC is located at the point A$(-3, 4, 0)$ and a line $y = 1, z = 1$ carries a charge density of $\lambda = 2$ nC/m. If $V = 0$ at O(0, 0, 0), find the potential V at a point B(5, 0, 1)?

18. Three charges, each $+q$ are placed at the three corners of a square of side a. Find the electric field and potential at the fourth corner. What is the energy required to assemble such a configuration?

19. Given that $V = \frac{10}{r^2}\sin\theta\cos\phi$, find the electric field $\vec{E}$ at $(2, \frac{\pi}{2}, 0)$. Also, determine the work done in moving a charge Q from a point A$(1, 30°, 120°)$ to B$(4, 90°, 60°)$.

20. Three point charges $Q_1 = -1$ nC, $Q_2 = 4$ nC and $Q_3 = 3$ nC are located at $A(0, 0, 0)$, $B(0, 0, 1)$ and $C(1, 0, 0)$ respectively. Find the total energy of the system.

21. A spherically symmetric charge distribution is given by

$$\rho = \rho_0\,,\, 0 \leq r \leq R$$
$$\rho = 0\,,\, r > R$$

Find the energy stored in the region $r < R$.

22. An isolated soap bubble of radius 1 cm is at a potential of 100 volts. It collapses into a drop of radius 1 mm. What is change in its electrostatic energy?

23. A sphere of radius R_1 has a uniform volume charge density ρ except for a small, spherical region of radius R_2 located at a distance a from the centre. Find the electric field and potential at the centre of the hollow sphere.

24. Electric charge is distributed in a spherical shell of inner radius R_1 and outer radius R_2 with a charge density $\rho = a + br$ where r is the distance from the centre. Find the electric field everywhere. Find the electric potential and energy density for $r < R_1$, assuming that the potential is 0 at ∞.

25. The inside of a grounded spherical metal shell (inner radius R_1 and outer radius R_2) is filled with space charge of uniform charge density ρ. Find the electrostatic energy of the system. Find the potential at the centre.

3

Electric Fields in Matter

Learning Objectives

- To describe the properties of dielectrics and distinguish them from conductors.
- To learn about the different kinds of dielectrics and their properties.
- To understand the phenomenon of polarisation at the atomic level.
- To be familiar with Coulomb's Law and Gauss's Law in the presence of a dielectric and learn about electric displacement.
- To be able to solve Laplace and Poisson equation in presence of dielectric materials and to learn about the boundary conditions on electric field and displacement at the interface of two dielectrics.
- To be familiar with the concept of the local field and average field inside the dielectric.
- To learn about the various theories to explain the observed properties of dielectrics.

3.1 INTRODUCTION

The laws of electrostatics that we have formulated in the previous Chapters, relate to phenomena where no bulk matter is present. Thus, for example, Coulomb's Law which gives us the force between two point charges, refers to a situation where there is no intervening matter between the point charges. This of course, is a simplification which is unrealistic since charges are normally found inside matter. The situation when matter is actually present is very different. What are the changes that we see in the forces between charges or the electric fields caused by them when matter is present? Are there any new phenomena which happen when matter interacts with an electric field? Do the laws that we have studied—Coulomb's Law and Gauss's Law—take a different form in the presence of matter? These are some of the issues we will investigate in this chapter.

Let us start with the structure of matter first. We know from atomic theory that an individual atom consists of a central, positively charged nucleus with negatively charged electrons around it. The basic structure of the atom though cannot be pictured in purely classical terms. The reason for this is easy to see— the electron, being oppositely charged to the nucleus, will be attracted towards the nucleus and hence fall into it if it were at rest, thereby making all matter unstable. This is obviously not the case. Hence there must be some other mechanism to keep the electron from falling into the nucleus. This was the basis of the Bohr model for the atom where we think of the electron being in a circular (or elliptical) orbit around the positively charged nucleus. This would also not work since, as we shall see later in the book, the accelerating electron will emit energy as radiation and soon will fall into the nucleus as it loses its energy. Thus, stable atoms cannot exist in a classical theory. This problem is resolved in quantum mechanics where the wave function approach provides for a stable picture of atoms. Unfortunately, the quantum mechanical formulation does not provide an exact and easy picture of the charges in the atom. However, for our purposes an approximate picture is sufficient. In this picture, we think of the atom as a central positively charged nucleus with a cloud of electrons surrounding it.

At the next level of structure in matter are the molecules. Molecules are made up of one or more atoms, either of the same or different kind. Obviously, since individual atoms in our picture, comprise of a positively charged centre surrounded by a negatively charged electronic cloud, when the atoms interact and combine, these charged entities (nuclei and electrons) will interact and the stable structure which results would have possibly distorted electronic clouds as compared to the individual atomic constituents. This interaction-induced distortion has profound consequences as we shall see later in this chapter.

To focus our attention, let us consider a simple case of where the cloud of electrons is spherically symmetric.

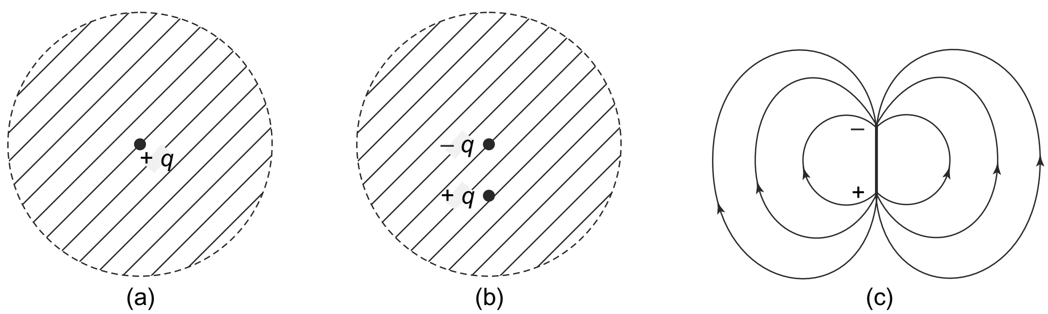


Fig. 3.1 *A spherically symmetric negatively charged cloud around a point nucleus: (a) Centre of the cloud coincides with the point nucleus, (b) The centre of the cloud is displaced from the nucleus, (c) The lines of force for case (b)*

It is easy to see that there can be two cases. If the centre of the cloud coincides with the positively charged point nucleus, as in Fig. 3.1(a), then outside the cloud, there is no electric field since for outside points, a uniformly charged sphere behaves as if all the charge is concentrated at the centre and the net charge of this sphere is zero. For points inside the spherical cloud, as we go in from the spherical surface, the field due to the negatively charged cloud decreases while that due to the positively charged nucleus increases. The latter dominates and becomes in fact, very intense close to the nucleus.

The second case is when the centre of the negatively charged cloud does not precisely coincide with the positively charged nucleus as shown in Fig. 3.1(b). In this case, for points outside the spherical cloud, the charge distribution behaves as a dipole (assuming of course that for points outside, the uniformly charged sphere behaves as if all the charge is concentrated at the centre). The electric lines of force are shown in Fig. 3.1(c). Though for points outside, the charge distribution behaves like a dipole, for points inside, it is the superposed field due to two unequal charges separated by a distance. The charges are unequal because as we go inside from the surface, the negative charge which contributes to the field at that point is only that part of the charge in the cloud which is interior to the point as we know from

Gauss's Law. Thus, it is only at the surface of the spherical configuration and outside it that the full negative charge will contribute to the field. As we go inside the sphere towards the centre, depending on the charge distribution of the negative charge within the spherical cloud, the charge will decrease. This configuration of unequal charges can be thought of as a dipole plus some residual extra positive charge.

In either case, as we move towards the centre, we experience an intense electric field. However, as we go away from the nucleus, for points outside the electron cloud, in the first case, we experience no field at all while in the second case we experience a dipole type of field. Thus, purely on the basis of their electronic configurations, we seem to have two types of molecules. **These two types of molecules are called non-polar and polar molecules.** Non-polar molecules are characterised by zero dipole moment while polar molecules have a net dipole moment. Common polar molecules are water molecule, sulphur dioxide molecule and wax molecule. Non-polar molecules are the inert gases, nitrogen, oxygen, methane, etc.

What happens when we place molecules in an external electric field? Clearly, since the molecules have charges (though their net charge might be zero), we expect them to interact with the electric field. The external electric field will act in opposite directions on the positively charged nucleus and the negatively charged electron cloud. In particular, if we had a spherically symmetric configuration in the absence of the external electric field (with the centre of the electron cloud coinciding with the positively charged nucleus), this would be distorted with the centres of charges now being displaced. Thus, the external electric field will change the shape of the spherically symmetric molecule. For us, the relevant issue is that this distortion would induce a dipole moment in the molecule, since now we have equal and opposite charges displaced from each other. Though we have considered this effect in molecules which had no dipole moment in the absence of the external electric field, it is easy to see that a similar effect would be present for molecules which do have an inherent dipole moment. In this case, the net dipole moment would be a resultant to the two dipole moments—the inherent dipole moment which comes from the distribution of the electronic charge in the molecule and the induced dipole moment coming from the effect of the external electric field.

3.2 BULK MATTER: DIELECTRICS AND CONDUCTORS

Matter, as we have seen, can be thought of as a collection of molecules which themselves are combinations of atoms. Ordinarily we know that matter exists in three phases—gas, liquid and solid. Of course, at some temperatures, there can be a coexistence of these three phases. These three phases or states of matter have different properties as far as their interaction with the electric field is concerned.

Molecules in a gas are sufficiently apart from each other and so their electron clouds do not overlap. When an external electric field is introduced, each individual molecule responds to it in a way which depends on the strength of the external field and the molecule's charge distribution. In short, there are no collective effects amongst the molecules of a gas. As we saw above, the external field induces a dipole moment in the molecule. This induced dipole moment is different for polar and non-polar molecules. For non-polar molecules, this induced dipole moment is the net dipole moment since there is no intrinsic dipole moment that a non-polar molecule has. In the case of polar molecules however, the net dipole moment is a combination of the induced dipole moment in the presence of an external electric field and the intrinsic dipole moment that the polar molecule possesses. The important thing is that in either case, the molecules retain their charge neutrality and hence cannot serve as carriers of charge. Materials which have this property, i.e., having no free charge carriers are called **insulators**. As we have seen, gases fall under this category.

Liquids and solids, on the other hand, have molecules which are in some kind of arrangement. An individual molecule in a liquid or a solid, unlike the molecule in a gas, is not totally free. In solids, the arrangement of molecules is very regular over many molecules. We see even more regularity in crystalline solids where the regular arrangement persists through out the whole sample. In liquids, on the other hand, though there is order or regularity, the scale over which we see such a regularity in the arrangement of molecules is much smaller. In either case, because the molecules are in close proximity to each other, we see drastic changes in the electronic orbits of the molecules. The electrons sense the electric field in their neighbourhood, and this electric field is not just caused by the nucleus of its atom, but also the nuclei of other atoms nearby. Their behaviour in this complicated field due to the charge distribution of their own molecule as well as all the other nearby molecules has been studied extensively. We shall not be focussing on this aspect. For our purposes, it suffices to categorise materials into two categories based on their behaviour in an external electric field. These two categories are called **dielectrics and conductors**.

Dielectrics are materials where the electrons are spread out over an array of molecules but are all bound to them. The interaction between the molecules is such that the electrons are not free to move. In an external electric field, the electrons and the nuclei of course experience a force but since they are bound to the molecules, the only effect of the electric field on them is to change the orientation of the molecules or stretch them. In insulators or dielectrics, there is no transport of charge in the presence of an external electric field. However, this change in the orientation and/or the shape of the molecules has a profound effect on the electric field inside the dielectric. We will see later in this chapter, that the net electric field inside the dielectric is the combination of the external field and the electric field of the distorted charge distribution in the molecules of the dielectric.

Conductors, on the other hand are materials where some of the electrons in the material are spread out over the whole sample as free particles. If an external, steady electric field is applied to such a material, the 'free' electrons will move because of the electrostatic force that they experience. Can this process go on indefinitely? Obviously this cannot be the case since the charges will move and redistribute themselves such that the electric field produced by the redistributed charges acts in the opposite direction to the external field. At some point, the two fields will be exactly equal and opposite and so there wont be any net field inside the conductor. Once the net electric field inside the conductor vanishes, the charges inside the conductor do not experience any net force and so remain stationary in their positions. **Thus, at equilibrium or in steady state, there is no net electric field inside a conductor.**

This characteristic of conductors also implies that the conductor is an equipotential. To see this, note that the definition of the electric potential, namely $\vec{E} = -\vec{\nabla}\phi$, tells us that in a region of vanishing electric field, the gradient of the potential is zero. If the gradient of a scalar function vanishes in a region, it must be a constant. Hence, the potential everywhere on the conductor is the same making it equipotential. Further, Gauss's Law tells us that if there is a non-zero charge density, it leads to an electric field. Inside a conductor, this implies that there cannot be any local accumulation of charges since that would then lead to a non-zero electric field. The accumulation will occur only on the surfaces of the conductor where the surface charge density will be such that the net field **inside** the conductor is zero. That is to say, in the presence of an external electric field, the charges in a conductor will be redistributed on the surface in such a way that as soon as one enters inside the conductor, the net field, that is the field due to both the external sources as well as the redistributed charges is zero. The presence of this surface charge on conductors leads to certain interesting consequences as we will study in the next Chapter. Metals of course, are the most prominent examples of conductors.

As an example, consider a conductor in the form of an infinite slab with plane faces and an electric field applied externally perpendicular to the faces. Charges redistribute themselves in the conductor such that finally negative and positive surface charges appear on the two flat faces as shown in Fig. 3.2(a). These surface charges are of magnitude so as to create an electric field inside the conductor which is exactly opposite of the applied external field, so that the total field inside the conductor is zero. When the conductor does not have as simple a shape as this example, the distribution of charge is more complicated. Thus, in the conductor shown in Fig. 3.2(b), charges on the surfaces are somewhere positive and somewhere negative and not uniform.

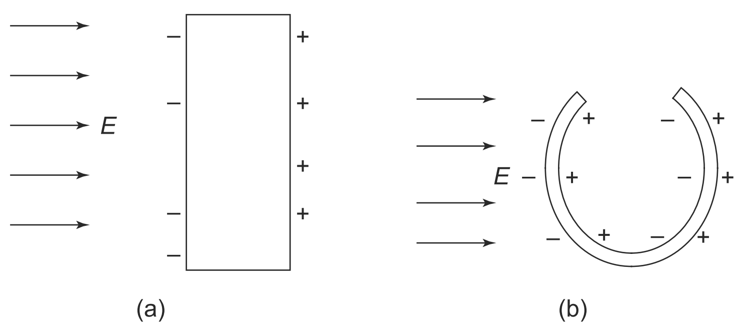


Fig. 3.2 *Charges on the surface due to an external electric field $\vec{E}$: (a) on a conductor slab with flat faces perpendicular to $\vec{E}$, (b) On a U - shaped conductor*

It is important to remember that the above statements and our other discussion in this chapter will focus only on the electrostatic aspects of insulators and conductors. If the external field is not static and changes with time, or if the charges on the surface of the conductor are redistributed continuously, then electric fields will be present inside the conductor. This is in fact exactly what happens when we connect a conductor to the ends of a battery where the battery is continuously supplying charges. We will study the non-electrostatic aspects of conductors when we talk about current electricity later in the book.

3.3 DIELECTRICS

Dielectrics form an important class of materials. As noted above, dielectrics can be made up of polar or non-polar molecules. The mechanism of interaction between the molecules and the external field is different for polar and non-polar molecules. We will consider these separately.

3.3.1 Non-polar Dielectrics

Consider first a dielectric made up of non-polar molecules. In this, as we have seen, the electrons and protons in the molecules are bound together. Further, these molecules have no net dipole moment in the absence of an external field. The exact nature of the bonds can only be understood through quantum theory but for our purpose, we can use a model wherein we think of the electron and proton being bound by a stretched spring. When an external electric field $\vec{E}$ is applied, the positive and negative charges in the molecule experience forces along and against its direction respectively. The positive charges (the protons) are much heavier than the negative charges (the electrons) and so we can assume they do not move very much. The total electric force on the negative charges will be

$$\vec{F}_{\mathrm{E}} = -|q|\vec{E}$$

where $-|q|$ is the total negative charge in a molecule. The negative charges will be stretched relative to the positive ones along the direction opposite to that of the electric field because of this. Let the displacement vector from the negative to the positive charge be $\vec{d}$ and the effective spring constant of the model be σ. Then a displacement of $\vec{d}$ will cause a restoring force in the spring which is given by

$$\vec{F}_{\mathrm{R}} = \sigma\vec{d}$$

Thus at equilibrium

$$\sigma\vec{d} = |q|\vec{E}$$

The displacement $\vec{d}$ creates a electric dipole moment

$$\vec{p}_0 = |q|\vec{d} = \left(\frac{q^2}{\sigma}\right)\vec{E}$$

This way of writing $\vec{p}_0$ is obviously model dependent since we have assumed the force between the electron and proton to be like a restoring force of a stretched spring. The important thing in this expression is that the induced dipole moment is proportional to the applied electric field $\vec{E}$ as it should be. The proportionality constant in this relationship is a quantity which obviously depends on the nature and properties of the material. In the case of the simplified model above, it depends on the spring constant σ. In general, $\vec{p}_0$ we can write as

$$\boxed{\vec{p}_0 = \alpha\frac{1}{4\pi k}\vec{E}} \tag{3.1}$$

Here α, called the **atomic polarisability** is a quantity which depends on the nature of the forces in the molecules of the material. It basically is a measure of the response of the molecule to an external electric field.

Multiplying by N, the number of molecules per unit volume, we get a quantity which is now characteristic of the bulk material. This is the **polarisation** $\vec{P}$ which is defined as

$$\boxed{\vec{P} = \left(\frac{Nq^2}{\sigma}\right)\vec{E} = \frac{\chi}{4\pi k}\vec{E}} \tag{3.2}$$

where $\chi = \alpha N$ is called the **susceptibility of the dielectric**. The factor of $4\pi k$ in the definition of α and χ has been introduced for convenience.

3.3.2 Polar Dielectrics

Let us now see what happens in the case of polar molecules in the presence of an external electric field. In such a situation, polar molecules too will be stretched like non-polar molecules. However, there is a significant difference—in the case of polar molecules, there is already a permanent dipole moment which is there even in the absence of the external field. It turns out that the effect of the induced dipole moment is much smaller than the permanent dipole moment. That is, numerically, the induced polarisation is very small as compared with the polarisation arising out of the permanent dipole moment that these molecules have. We will see that shortly after we compare the numerical values of χ of these two types of molecules actually determined from experiments.

Another interesting fact is that the permanent dipole moments of the polar molecules, in the absence of an external electric field, will not be aligned in the same direction throughout a macroscopic sample. They will be in random directions with all directions equally populated. The average polarisation (which we recall, is a macroscopic quality unlike the atomic polarisability α which is a microscopic quality) over any sample will thus vanish. Now when an external electric field is applied the situation is different. The positive and negative charges experience equal and opposite forces due to the field. There is thus, no net force on the molecules but only a torque of magnitude

$$qdE \sin\theta$$

where θ is the angle between the direction of the dipole moment and the external applied field (Fig. 3.3).

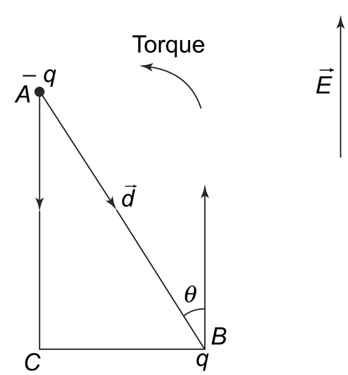


Fig. 3.3 *AB is a polar molecule with A at the negative centre of charge and B the positive one separated by a distance d. The force on A is $-qE$ and on B it is $+qE$. There is thus, no net force but only a torque on the molecule is qE.(distance BC) $= qEd\sin\theta$ in the counterclockwise sense*

If the torque was the only effect of the electric field, the dipoles will ultimately fall into the position of stable equilibrium which is at $\theta = 180^\circ$ and the polarisation vectors of the molecules will be along $\vec{E}$. In terms of potential energy of the dipoles in the electric field, this orientation will be the one of lowest potential energy. At any other angle θ, the potential energy $U(\theta)$ is given by

$$U(\theta) - U(\theta = 180^\circ) = \text{work done in changing the orientation from } \theta = 180 \text{ to } \theta = \theta$$

$$= \int_{\theta}^{\pi} q \;\; d \;\; E \sin\theta' \mathrm{d}\theta'$$
$$= qEd(1 + \cos\theta) \tag{3.3}$$

Now, if the molecules in a sample were all at rest, which can be expected only at absolute zero temperature, all of them would be in the state of lowest potential energy, i.e., at $\theta = 180^\circ$, and their polarisation vectors will all be parallel to $\vec{E}$. At finite temperatures however, this would not be true since the kinetic theory tells us that the molecules have a distribution of energies and will be constantly exchanging energy with each other. We know however, from the principles of statistical mechanics, that under such circumstances the relative number of molecules at a given energy decreases with energy. In other words, there would be more molecules in states of lower energy than in states of higher energy. In the present case then, at finite temperatures the maximum number of molecules will be in a state of $\theta = 180^\circ$ in which orientation their dipole moments will be parallel to $\vec{E}$. As θ decreases from this value, the relative number of molecules starts decreasing. At the other end, when the value of θ is zero, the number of molecules will be the minimum at which position the dipole moment will be antiparallel to $\vec{E}$.

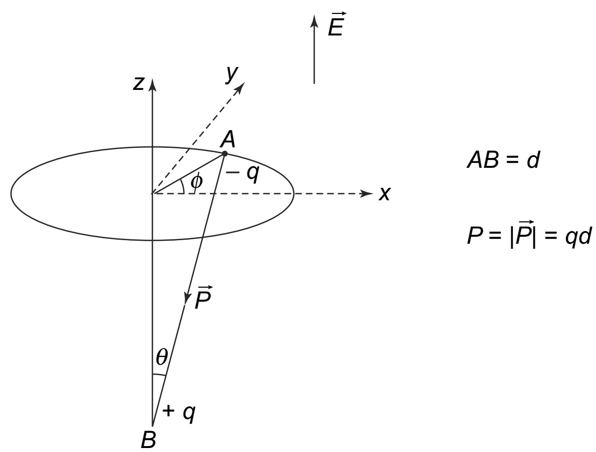


Fig. 3.4 *A dipole BA making an angle θ with the direction of $\vec{E}$ with an azimuthal angle ϕ. The dipole moment is along AB and its components are $(p \sin\theta \cos\phi,\ p \sin\theta \sin\phi,\ -p \cos\theta)$*

Consider now the dipoles at a given value of θ (Fig. 3.4). Dipoles with a given θ can have any azimuthal angle ϕ ranging from 0 to 2π. The energy however, depends only on θ, so that as discussed above, the number of molecules at various values of ϕ for a given θ would be the same since they have the same energy. If the electric field is taken to be along the z-direction (Fig. 3.4), the components of the dipole moment of a molecule with a given value of θ and azimuthal angle ϕ will be $(\mathbf{p} \sin\theta \cos\phi, \mathbf{p} \sin\theta \sin\phi, -\mathbf{p} \cos\theta)$. Averaging over ϕ, we get zero for the first two components and $-p \cos\theta$ for the z- component which is along $\vec{E}$. This is positive for θ between 90 and 180 degrees and negative for θ between 90 degrees and zero. Molecules with $90^\circ < \theta < 180^\circ$ have less energy than those with $0^\circ < \theta < 90^\circ$ and thus will be more in number. If we now add up the moments of all the molecules in a sample, we will get a net moment along the direction of the electric field. We thus, have for the polarisation vector $\vec{P}$,

$$\boxed{\vec{P} = \frac{\chi}{4\pi k}\vec{E}} \tag{3.4}$$

where χ is the susceptibility of the material. Note that the polarisation $\vec{P}$ will be zero in the absence of the field so that the the polarisation arising because of the electric field, Eq. (3.4) can be considered as arising out of charge separation due to the electric field just as in the case of non-polar molecules.

3.4 COULOMB'S LAW IN A DIELECTRIC

As we have seen above, the charges do move over short distances in a dielectric to form dipoles. The amount of charge moving across any surface is obviously related to the polarisation vector $\vec{P}$. This allows us to write a modified form of Coulomb's Law (or Gauss's Law which is derived from it) for dielectric materials since the form of Gauss's Law that we considered in Chapter 1 is not applicable in this case.

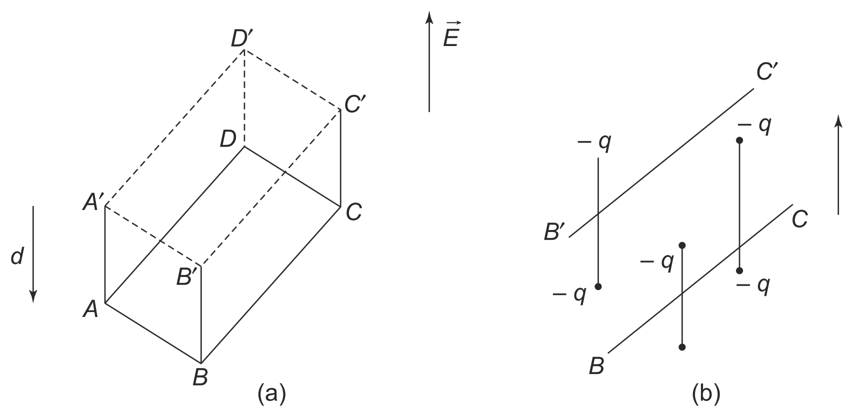


Fig. 3.5 *Changes moving to form diploes: (a) ABCD is an infinitesimal surface of area dS with an outward normal $\hat{n}$. $\vec{E}$ is vertically up. In the presence of the electric field, one can assume that the positive charges remain fixed while the negative charges move through a distance d in the direction opposite to $\vec{E}$. $A'B'C'D'$ is parallel to ABCD in the direction of $\vec{E}$ and removed by a distance d, (b) Lateral view. Charges between BC and $B'C'$ would cross*

Consider a small area $ABCD$ with area dS and the outward normal $\hat{n}$ as shown in Fig. 3.5. In the presence of an external electric field E, the charges in the molecule will move. The positive charges, as before, we can consider to be fixed while the negative charges move a distance $\vec{d}$ in a direction opposite to the direction of $\vec{E}$. Consider an area $A'B'C'D'$ parallel to $ABCD$ and removed a distance d along the direction of $\vec{E}$. All the negative charges contained in the volume bounded by $ABCD$ and $A'B'C'D'$ would have crossed $ABCD$ under the influence of $\vec{E}$. The volume element bounded by $ABCD$ and $A'B'C'D'$ is $dS\vec{d}\cdot\hat{n}$. If N is the number of molecules per unit volume, then the net charge crossing $ABCD$ along the outward normal direction is

$$N(-q)dS\vec{d}\cdot\hat{n} \tag{3.5}$$

After a displacement $\vec{d}$ of the negative charges, the dipole moment of a molecule will be $\vec{p}_0 = -q\vec{d}$, where $-\vec{d}$ is the displacement from the negative to the positive charge. The polarisation $\vec{P}$ is by definition, $N\vec{p}_0$ and thus, the charge crossing the surface dS from Eq. (3.5) is given by

$$dQ_{\text{pol}} = (\vec{P} \cdot \hat{n})dS$$

$$\text{charge density at the surface} = dQ_{\text{pol}}/dS = \vec{P} \cdot \hat{n} \tag{3.6}$$

Equation (3.6) allows us to calculate the net charge that leaves a volume due to the effect of an external electric field. Integrating Eq. (3.6) across the bounding surface of the volume gives us the net charge leaving the volume

$$\boxed{Q_{\text{pol}} = \int dQ_{\text{pol}} = \iint (\vec{P} \cdot \hat{n})dS = \iiint \vec{\nabla} \cdot \vec{P}dV} \tag{3.7}$$

where we have used the Divergence Theorem (Gauss's Divergence Theorem) to convert the surface integral to a volume integral of the divergence.

Note that the dielectric as a whole is electrically neutral and consequently, when a charge dQ_{pol} flows out of dS, a corresponding amount $-dQ_{\text{pol}}$ stays inside.

Consider first a case where there are no free charges. Then, we know that $\vec{\nabla} \cdot \vec{E} = 0$ by Gauss's law. But $\vec{P} = \chi\vec{E}$ and hence $\vec{\nabla} \cdot \vec{P} = 0$. This implies, because of Eq. (3.7) that Q_{pol} vanishes. This is easily visualised by taking the simples case of a constant electric field Fig. 3.6.

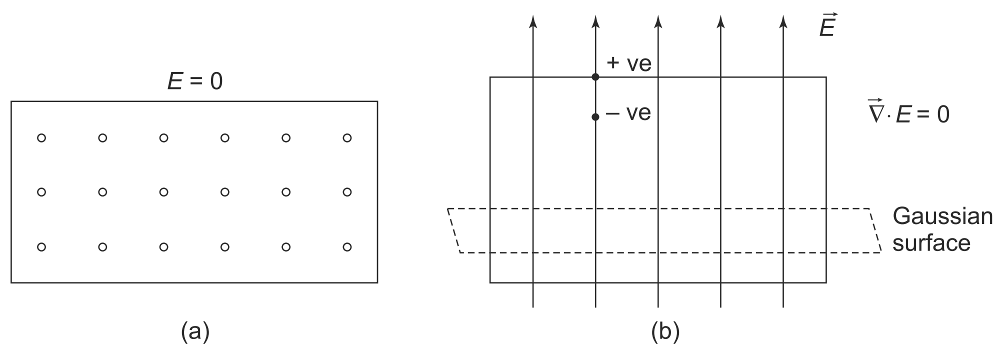


Fig. 3.6 *Charge flow due to polarisation in the case when* $\vec{\nabla} \cdot \vec{E} = 0$ *(a) The dots represent molecules, (b) In the presence of a constant uniform electric field, the negative charges in the dipole, which are represented by tiny dumbbells, are pulled down. Across any Gaussian surface inside the dielectric, as much enters the top surface as exits from the bottom one*

In the presence of the electric field, the negative charges which are more mobile being lighter, are pulled down in the direction opposite to the electric field. However, if we consider a Gaussian surface inside the dielectric (Fig. 3.6(b)), as much charge would be pulled inside its top surface as exits from the bottom one. Hence, the net charge flow due to polarisation for such cases is zero.

The situation is different when there are free charges present in the dielectric. In this case, $\vec{\nabla} \cdot \vec{E} \neq 0$. To see this, consider a simple case of a uniformly charged sphere with charge density $\rho_{\text{free}} \neq 0$ inside a dielectric.

Across the Gaussian surface G, charges would be pulled in without any compensating charge flowing out unlike the case where $\vec{\nabla} \cdot \vec{E} = 0$. Thus, inside G, we will not only have free charges but also a polarisation charge $-Q_{\text{pol}}$ which is given by Eq. (3.7).

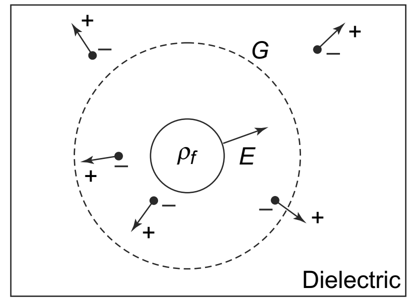


Fig. 3.7 *A spherically symmetric charged sphere inside a dielectric*

In general, applying Gauss's theorem to any Gaussian surface in a dielectric, we get

$$\begin{aligned}\iiint \vec{\nabla} \cdot \vec{E} dV &= 4\pi k Q_{\text{inside the surface}} \\ &= 4\pi k (Q_{\text{free}} - Q_{\text{pol}}) \\ &= 4\pi k \left(\iiint \rho_{\text{free}} dV - Q_{\text{pol}} \right) \\ &= 4\pi k \left(\iiint (\rho_{\text{free}} - \vec{\nabla} \cdot \vec{P}) dV \right) \end{aligned} \tag{3.8}$$

since the charge inside the surface is $-Q_{\text{pol}}$. Now using Eq. (3.2) that is

$$\vec{P} = \frac{1}{4\pi k} \chi \vec{E}$$

and substituting in Eq. (3.8), we get

$$\iiint dV \vec{\nabla} \cdot (\vec{E} + \chi \vec{E}) = \iiint dV (1 + \chi) \vec{\nabla} \cdot E = 4\pi k \iiint \rho_{\text{free}} dV \tag{3.9}$$

The quantity

$$1 + \chi = K$$

is called the **dielectric constant of the material**. Since Eq. (3.9) is valid for any arbitrary Gaussian surface, we can equate the integrands on both sides to get

$$(1 + \chi) \vec{\nabla} \cdot \vec{E} = 4\pi k \rho_{\text{free}} \tag{3.10}$$

or alternately

$$\boxed{\vec{\nabla} \cdot \left(\frac{\vec{E}}{4\pi k} + \vec{P} \right) = \rho_{\text{free}}} \tag{3.11}$$

The quantity

$$\frac{\vec{E}}{4\pi k} + \vec{P} = \vec{D}$$

is called the **displacement vector**.

Thus, Gauss's law for a dielectric can be written as

$$\boxed{\vec{\nabla} \cdot \vec{D} = \rho_{\text{free}}}$$

We can also define another quantity ε, called the **permittivity of the material** as

$$\boxed{\vec{D} = \varepsilon \vec{E}}$$

Clearly

$$\varepsilon = \frac{1}{4\pi k}(1+\chi) = \frac{K}{4\pi k}$$

For vacuum, $K = 1$ and hence, ε reduces to $\frac{1}{4\pi k}$. This is normally referred to as ε_0 and is called **permittivity of the vacuum.** The quantity K is called the relative permittivity or dielectric constant.

Displacement Vector

The term displacement in the context of Electricity and Magnetism was introduced by James Clerk Maxwell (1831–1879). The term obviously was used since the vector $\vec{P}$ refers to charges being displaced from their normal positions under the action of an external electric field. In his book, *A Treatise on Electricity & Magnetism* (1873), Maxwell calls the entire combination $\left(\frac{\vec{E}}{4\pi k} + \vec{P}\right)$ as displacement. Its value, relative to its value in the vacuum (where $\vec{P} = 0$) is called 'specific inductive capacity of the dielectric' or briefly, dielectric constant. In our notation, it is just $(1+\chi)$. The quantity $\vec{D} = \left(\frac{\vec{E}}{4\pi k} + \vec{P}\right)$ is also written as $\varepsilon\vec{E}$ where ε is the permittivity of the dielectric. The history of using this particular word is still debated and it is not clear who used it first. (See, D.E. Neuenschwander, American Journal of Physics, 66(8), 655, (1998); A.Sihvola, American Journal of Physics, 68 (8), 692 (2000)).

It is important to note that in quantum theory, we come across a concept of 'vacuum polarisation'. This term has nothing to do with the permittivity of the vacuum, as is normally used in Classical Electrodynamics. Permittivity in classical theory is the constant ε entering the relation between $\vec{D}$ and $\vec{E}$ arising out of polarisation of the molecules of the medium. When there is no medium, the constant is called ε_0 and called the permittivity of vacuum. This is merely a nomenclature and no polarisation is involved. In quantum theory, the term refers to fluctuations in the vacuum which has measurable effects. Though the names are somewhat similar, the two concepts are vastly different and should not be confused with each other. The quantum phenomena has no analogue in classical physics.

Table 3.1 *Dielectric constants of common materials*

Material	K
Vacuum	1
Glass	5-10
Mica	3-6
Water	80.4
Air	1.008
Oxygen	1.005
Nitrogen	1.007
Helium	1.0006

The values of dielectric constants of materials have been measured and are tabulated in standard tables. Since this is an inherent property of the material, it depends on external conditions like temperature. Table 1 below gives the dielectric constants of some common materials.

Note that the susceptibility of non-polar materials like O_2, N_2 etc is very small leading to small dielectric constants or permittivity since they are related by the factor of $4\pi k$.

EXAMPLE 3.1 Make an order of magnitude estimate of the electric susceptibility of hydrogen gas at STP from the known parameters of hydrogen.

Solution

We know that the hydrogen atom, in its ground state has a binding energy of 13.6 eV. Here eV stands for electron Volts which is a unit of energy relevant for atomic systems. It is the energy gained by an electron when it accelerates through a potential difference of 1 V. In terms of the more well known unit for energy, Joules, we have

$$1 \text{ eV} = 1.6 \times 10^{-19} \text{ J}$$

This relationship between eV and Joules makes it evident why we need to use a different unit for the energy of atomic systems. The orbit of the hydrogen atom has a radius of 0.51 Angstroms. The potential energy is -27.2 eV. Classically, there cannot exist stable orbits for the electron around the hydrogen nucleus as have seen above since the accelerating electron would rapidly lose energy as radiation and fall into the nucleus. For our purposes, we assume that the negative and positive charge is separated by a distance a and there is an electric force

$$\frac{ke^2}{a^2}$$

between them. Then, assuming our simplified model of the dipole being considered as a dumbbell with a spring in between, the 'spring constant' is clearly

$$\frac{\text{force}}{a}$$

The external electric field, causes an additional displacement which will simply be the force divided by the spring constant. Thus, this extra displacement would be

$$d = eE\frac{a}{ke^2/a^2} = \frac{Ea^3}{ke}$$

This displacement would create an atomic dipole moment given by

$$p = ed = \frac{Ea^3}{k}$$

Multiplying this by twice the Avogadro's number N_A and dividing by the molar volume, and taking note of the fact that the susceptibility is defined with a factor of $4\pi k$ as in Eq. (3.2), we get

$$\chi = \frac{8\pi N_A a^3}{V} \sim 1.0 \times 10^{-4}$$

which is only about 30% of the value in the tables. This is a remarkable agreement given that our assumptions are obviously very simplistic.

PROBLEM 3.1 A solid cylindrical rod made of a dielectric of dielectric constant $K = 3$ is of radius R. Its axis is along the x-axis and the end plates are at $x = 0$ and $x = L$. The polarisation in the dielectric is along the axial direction and has the value $|\vec{P}| = \varepsilon_0 \left(1 + \frac{x^2}{L^2}\right)$ statvolt/m. Calculate

(a) the electric field $\vec{E}$ and the charge density inside the dielectric,
(b) the surface charges on the cylinder and
(c) the volume density of polarisation charge.

PROBLEM 3.2 A dielectric material of dielectric constant K is in the form of a cube of size L. The centre of the cube is at the origin and the edges are along the x, y and z-axis. The polarisation vector $\vec{P}$ in the cube is given by $\vec{P}(x, y, z) = 2x^2\hat{i} + 3y^2\hat{j} + 7z^2\hat{k}$. Calculate the total bound charge inside as well as on the surface of the cube.

PROBLEM 3.3 A circular disk made of material of dielectric constant K has a radius R and thickness d. The plane faces of the disk are parallel to the $x - y$ plane. The polarisation of the disk is uniform of value $\vec{P} = P_0\hat{k}$. $\phi(z)$ denotes the potential at points along the axis at a distance z from the midpoint between the two faces on its axis. Determine the electric potential difference $\phi(z) - \phi(z = 0)$ for $z > \frac{d}{2}$.

Water as a Solvent—Electrostatics at Work

Water is a very unusual compound. It is an excellent solvent and this makes it essential for life as we know it. Living cells have a very high proportion of water which makes possible the transport of not only nutrients and chemicals but also oxygen and other gases. Water molecule, as we know is made up of two atoms of hydrogen and one atom of oxygen. What is it about water which accounts for its unique properties? Can we relate its being an excellent solvent to its electrical properties? The answer turns out to be yes.

The water molecule has a very high permanent dipole moment which gives water a very high relative permittivity (around 80). This in turn accounts for some of the exceptional properties of water as an excellent solvent. It also makes it possible for ionic crystals, like common salt (NaCl), to dissolve readily in water as also the phenomenon that salt crystals sprinkled on frozen ice melt it.

A water molecule has a shape which is given in Fig. 3.8

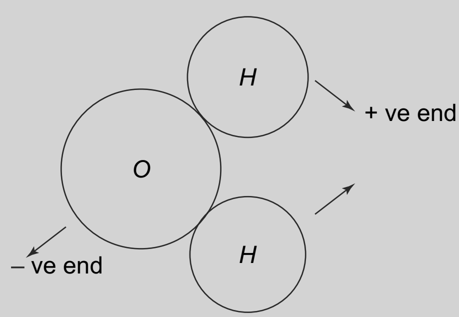


Fig. 3.8 *Water molecule*

The molecule as a whole is of course neutral but the neutrality does not exist throughout in the distribution of charges in the molecule. The hydrogen regions of the water molecule are positively charged while the oxygen region in the middle is negatively charged leading to a large electric dipole moment in the neutral molecule. In solid ice, the water molecules bond with other water molecules in regular patterns which are repeated over many molecules while in the liquid state, though there is some bonding, it is much weaker and limited in size leading to liquid water being able to flow easily.

We know from statistical thermodynamics that in any assembly of molecules, states which have the least energy are preferred in the sense that they would be the most heavily populated while states with higher energies would be sparsely populated in relative terms. NaCl crystal has a regular arrangement of Na^+ and Cl^- ions as shown in Fig. 3.9(a).

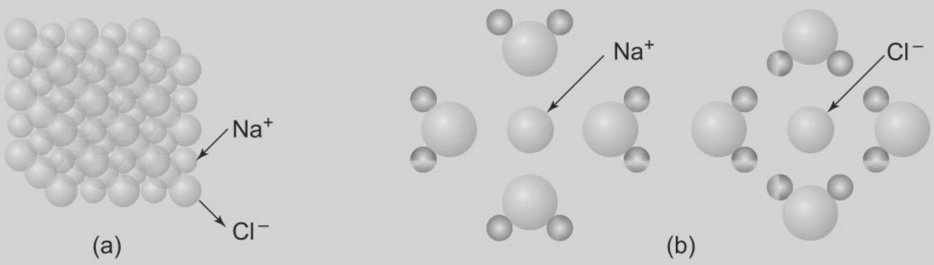


Fig. 3.9 *(a) Sodium Chloride crystal (b) Sodium Chloride in water*

Now when sodium chloride is dissolved in water, the positively charged Na^+ ion is attracted towards the negative end of the water molecule. More than one water molecule 'binds' with Na^+ and this results in a 'hydrated' Na^+ ion. Since there is an attractive force between the Na^+ ions and the negative (O) end of the water molecule, the resulting configuration has a much lower energy compared to when the Na^+ and water molecules are separated. The hydrated Na^+ ion is preferred at ordinary temperatures. In a similar fashion, the Cl^- ion gets attracted to the positive end (H) end of the water molecule resulting in hydrated Cl^- ions. These hydrated ions thus have a structure like a point charge surrounded by a dielectric. As we have seen above, the effect of the dielectric is to reduce the electric field inside it. The chlorine ion therefore, cannot get close enough to the hydrated Na^+ ion to combine with it to form NaCl. A consequence of this is that in aqueous solution, it is the hydrated ions rather than the Na^+ or Cl^- ions which move freely. This, as we shall see later, allows salt solution to conduct electricity.

A very similar phenomenon occurs when salt is added to crystalline ice. The Na^+ and Cl^- combine with water to form hydrated ions. The hydrated ions have reduced attraction to each other compared to water molecules. Thus, at temperatures where water molecules exist in solid, crystalline form, the hydrated ions stay in liquid form. This is the reason why salt is used to 'melt' snow on the roads during the winter season.

PROBLEM 3.4 In a dielectric material, we have $E_y = 10$ V/m and the polarisation vector $\vec{P} = \frac{1}{10\pi}\left(2\hat{i} - \hat{j} + 4\hat{k}\right)$. Find

(a) χ and
(b) $\vec{D}$

PROBLEM 3.5 A dielectric sphere has a $K = 5$ and is of radius 100 cm. A point charge $q = 10^{-12}$ C is place at its centre. Calculate the surface density of the polarisation charge on the surface of the sphere.

3.5 LAPLACE AND POISSON'S EQUATIONS: BOUNDARY CONDITIONS IN THE PRESENCE OF DIELECTRICS

We have already seen that the basic electrostatic problem is the solution of the Laplace and/or Poisson equation with the specified boundary conditions. The equations for the electric potential, when solved, allow us to calculate the electric field everywhere for a given set of boundary conditions. In the presence of dielectrics, the equations are

$$\nabla^2\phi(r) = 0 \qquad \text{(no free charges present)}$$

and

$$\nabla^2\phi(r) = -\frac{\rho(\vec{r})}{\varepsilon} \qquad (\rho(\vec{r}) \text{ is the free charge density at } \vec{r})$$

These equations are second order partial differential equations and therefore for a complete solution, we need to have two boundary conditions. These would be the value of ϕ and its derivative at a surface or a point. We are presently concerned with dielectrics and so we need to know what the boundary conditions are at the interface of the dielectric and another dielectric or vacuum. In particular, we need to know how ϕ, $\vec{E}$, $\vec{D}$ etc change at the interface.

3.5.1 Boundary Conditions

Consider an interface of two dielectrics with permittivities ε_1 and ε_2 in regions I and II with an interface as shown in Fig. 3.10.

We assume that there are no free charges on the interface. Consider a small section AB of the interface which can be considered planar (Fig. 3.10(a)). Now consider a pill-box with two faces of area A and of infinitesimally small width (Fig. 3.10(b)). The pill-box is placed on the section AB such that the faces are parallel to the section AB and one face is in region I while the other is in region II (Fig. 3.10(b)). Let us apply Gauss's Law for dielectrics to this pill-box. The displacements in the two regions are $\vec{D}_1$ and $\vec{D}_2$.

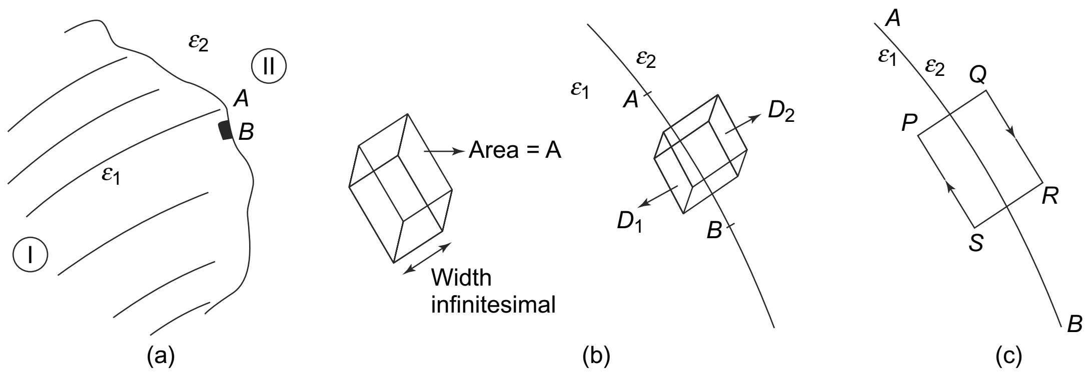


Fig. 3.10 *Interface of two dielectrics with permitives: (a) Dielectrics in two regions I and II, (b) Enlarged version of the interface, (c) Closed path at the interface*

Now to apply Gauss's law, we need to determine the flux through the pill-box. The flux across the infinitesimal sides can be neglected. For the faces with area A, the flux across them is

$$\iint_{S_2} \vec{D}_2 \cdot d\vec{S} = D_2^{\perp} A$$

and

$$\iint_{S_1} \vec{D}_1 \cdot d\vec{S} = D_1^{\perp} A$$

in opposite directions. Here $D_1^{\perp}$ and $D_2^{\perp}$ are the components of the displacement vectors $\vec{D}_1$ and $\vec{D}_2$ normal to the surface in the two regions I and II and S_1 and S_2 are the two faces of the pill-box parallel to AB on opposite sides of the interface. Since there are no free charges on the interface, Gauss's Theorem gives us

$$\iint_{S} \vec{D} \cdot d\vec{S} = 0$$

and hence,

$$D_2^{\perp} A - D_1^{\perp} A = 0 \tag{3.12}$$

or, since $A \neq 0$,

$$\boxed{D_2^{\perp} = D_1^{\perp}} \tag{3.13}$$

Now since the two faces of pill-box are parallel to AB, we can conclude that

"Normal component of the displacement vector $\vec{D}$ is continuous across the interface when no free charges are present".

Since the displacement vector and the electric field are related ($\vec{D} = \varepsilon\vec{E}$), the normal components of the electric field $E_1^{\perp}$ and $E_2^{\perp}$ will then satisfy

$$\boxed{\varepsilon_1 E_1^{\perp} = \varepsilon_2 E_2^{\perp}} \tag{3.14}$$

We thus, have the conditions for the normal components of the displacement vector and electric field across an interface of two dielectrics. What about the tangential or the components parallel to the interface?

To find the conditions on the tangential components of the electric field, consider a rectangle $PQRS$ as shown in Fig. 3.10(c) such that PQ and RS are of infinitesimal length. The sides QR and SP are of length L and are parallel to the surface element AB but on different sides of the boundary. We will then have

$$\oint_{\text{PQRS}} \vec{E} \cdot \vec{dl} = 0 \tag{3.15}$$

since we know that $\vec{E} = -\vec{\nabla}\phi$. The contribution to the loop integral of the sides PQ and RS will vanish since these are of infinitesimal length. We thus get from Eq. (3.15),

$$\oint_{\text{PQRS}} \vec{E} \cdot \vec{dl} = \int_{QR} \vec{E}_2 \cdot \vec{dl} + \int_{SP} \vec{E}_1 \cdot \vec{dl} = 0$$

But

$$\int_{QR} \vec{E}_2 \cdot \vec{dl} = -E_2^{||} L$$

and

$$\int_{SP} \vec{E}_1 \cdot \vec{dl} = E_1^{||} L$$

where $E_1^{||}$ and $E_2^{||}$ are the tangential components (parallel to the interface directed from S to P) of $\vec{E}_1$ and $\vec{E}_2$ respectively and L is infinitesimal.

$$E_1^{||} L - E_2^{||} L = 0$$

and hence,

$$\boxed{E_1^{||} = E_2^{||}} \tag{3.16}$$

We thus, get the boundary condition on the tangential component of $\vec{E}$ as

> "The tangential components of the electric field $\vec{E}$ are continuous across an interface if no free charges are present."

Since both the normal and tangential components of $\vec{E}$ are present across the interface, the scalar potential ϕ is continuous across the interface but because of Eq. (3.14), the normal derivative is not continuous.

We thus, have the boundary conditions on the displacement vector and the electric field across a boundary of a dielectric. To solve for electrostatic problems in the presence of dielectrics, we need to solve the Laplace and/or Poisson equations ensuring that our solutions respect the above mentioned boundary conditions on $\vec{D}$ and $\vec{E}$. We illustrate this procedure with a few examples.

EXAMPLE 3.2 A point charge q is placed at the centre of a dielectric sphere of radius R. Find the electric potential both inside and outside the sphere.

Solution

The problem of finding the potential inside and outside the sphere is one of solving the Laplace and Poisson's equations. Since the problem has inherent spherical symmetry, it is convenient to use spherical polar coordinates. The Laplacian operator ∇^2 in this coordinate system is given by

$$\nabla^2 = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right) + \frac{1}{r^2 \sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2}{\partial\phi^2} \tag{3.17}$$

In our case, there is nothing in the problem to distinguish one angular direction (θ, ϕ) from another and hence there is no dependence on the angular coordinates. Thus, $\phi(r, \theta, \phi) = \phi(r)$. In this case, the Laplacian simplifies and only the first term, with the radial dependence is relevant.

In the region $r > R$, since there is no charge, we get the Laplace equation as

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\phi(r)}{\partial r}\right) = 0 \tag{3.18}$$

which can be written as

$$\frac{1}{r}\frac{d^2(r\phi(r))}{dr^2} = 0 \tag{3.19}$$

This equation has a solution

$$r\phi(r) = Ar + B \qquad \text{(where } A \text{ and } B \text{ are constants)}$$

Thus, we get

$$\phi(r) = A + \frac{B}{r} \tag{3.20}$$

Our usual convention on the electric potential is that it vanishes at ∞ and thus $A = 0$. The solution thus, becomes

$$\phi(r) = \frac{B}{r} \tag{3.21}$$

The constant B in this solution of course needs to be determined by the boundary conditions.

For the region $r < R$, the presence of the free charge demands that we use Poisson's equation. The region inside has no charge apart from a point charge q located at the centre of the sphere or at the origin. Thus, the charge density $\rho(r)$ which enters the Poisson's equation can be represented by $\rho(r) = q\delta(\vec{r})$ since the $\delta(\vec{r})$ ensures that the charge density is non-zero only at $\vec{r} = 0$. The Poisson equation then reads, (once again dropping the angular derivatives, since there is no dependence on θ and ϕ),

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\phi(r)}{\partial r}\right) = -\frac{q}{\varepsilon}\delta(\vec{r}) \tag{3.22}$$

This is an inhomogeneous differential equation and we can add to its solution, the solution of the corresponding homogenous differential equation which in this case is the Laplace equation. A particular solution to Eq. (3.22) can be obtained by the usual Green's function method (See Mathematical Preliminaries). The Green's function for the Laplacian operator is defined by

$$\nabla^2 G(\vec{r} - \vec{r}') = \delta^3(\vec{r} - \vec{r}') \tag{3.23}$$

whose solution is given by

$$G(\vec{r} - \vec{r}') = -\frac{1}{4\pi}\frac{1}{|\vec{r} - \vec{r}'|} \tag{3.24}$$

The general solution to Eq. (3.22) for $r < R$ is thus,

$$\phi(r) = \int d^3r' G(\vec{r} - \vec{r}')\left(-\frac{q}{\varepsilon}\right)\delta(r') + f(r') \tag{3.25}$$

where $f(r)$ is the solution to the Laplace equation (Eq. (3.19)). But we have already got a solution for the Laplace equation, given in Eq. (3.20) that is

$$f(r) = C + \frac{D}{r} \qquad \text{(C, D are constants)} \tag{3.26}$$

Using this in Eq. (3.25), and doing the integration we get

$$\phi(r) = \frac{q}{4\pi\varepsilon}\frac{1}{r} + C + \frac{D}{r} \qquad (0 < r < R) \tag{3.27}$$

Since we are now considering the region $0 < r < \infty$, regularity of $f(r)$ at $r = 0$ requires $D = 0$. The coefficient of the $(\frac{1}{r})$ term in $\phi(r)$ is determined uniquely by solution of Poisson's equation to be $\frac{q}{4\pi\varepsilon}$.

To find the other constants, we now use the boundary conditions. We equate $\varepsilon\frac{d\phi}{dr}$ from Eq. (3.21) and Eq. (3.27) at $r = R$ and get

$$B = kq \tag{3.28}$$

Finally, if we demand that $\phi(r)$ be continuous across the boundary when we approach it from either outside or inside, we get

$$C = \frac{kq}{R}\left(1 - \frac{1}{K}\right)$$

The general solution therefore, becomes

$$\phi(r) = \frac{kq}{r} \qquad (r > R)$$

$$\phi(r) = \frac{q}{4\pi\varepsilon}\frac{1}{r} + C = kq\left(\frac{1}{Kr} + \frac{1 - \frac{1}{K}}{R}\right) \qquad (0 < r < R) \tag{3.29}$$

Thus, the electric field is a factor K smaller inside the dielectric as compared with its value outside at the interface. This is what we had expected since the presence of an external electric field (created by the free charge) in the dielectric induces a polarisation which acts to decrease the electric field inside the dielectric.

It would have been noted that we could have used Gauss's Law to solve the above problem because of the inherent symmetry. This is true. However, we tried to solve it using Laplace's and Poisson's equation only to illustrate how the equations are solved and the boundary conditions are used. The next two examples will be solved using Gauss's Theorem but in the last example, which cannot be solved using Gauss's Theorem, we shall again use the Laplace and Poisson equations, in a way to similar to this one.

EXAMPLE 3.3 A thin sheet of charge, infinite in extent, of surface charge density σ is parallel to the $x - y$ plane. An infinite dielectric slab of thickness t is placed below the charged sheet at a perpendicular distance d. Calculate the field inside and outside the dielectric as also the surface charge density on the dielectric.

Solution

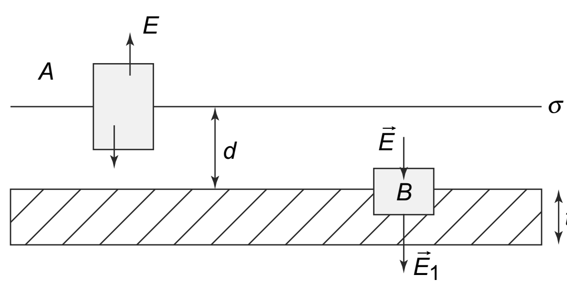


Fig. 3.11 *Example 3.3*

We will be using Gauss's law to determine the field quantities in the problem. Note that the field everywhere is in the z-direction since both the sheet and the dielectric are infinite and parallel to the $x - y$ plane. Given this symmetry, it is obvious that the Gaussian surface should be chosen whose surfaces are either perpendicular or parallel to the direction of the field. Consider first the Gaussian surface A in Fig. 3.11. This surface, in the shape of a pillbox has the two flat surfaces on either side of the infinite sheet of charge. By symmetry, the field at the two surfaces which are in the $x - y$ direction is the same, perpendicular to the surfaces. If these surfaces are taken to be of unit area, then Gauss's law gives us

$$2E = 4\pi k\sigma$$

or

$$E = 2\pi k\sigma \tag{3.30}$$

This is the expected result of the field near an infinite sheet with charge density σ. Now consider another Gaussian surface B, again in the shape of a pillbox, but now with one flat surface inside the dielectric and one outside. Now applying Gauss's law to these flat surfaces (the contributions of the curved surface is again zero since the field is in the z-direction), we get

$$-\frac{E}{4\pi k} + \varepsilon E_1 = 0$$

where E_1 is the field inside the dielectric and the R.H.S. vanishes since there are no free charges inside the Gaussian surface. Hence,

$$E_1 = \frac{\sigma}{2\varepsilon} = \frac{2\pi k\sigma}{K} \tag{3.31}$$

This is, as expected, lower by a factor of $1/K$ than the field outside the dielectric.

The polarisation P can be easily obtained as

$$P = \frac{\chi}{4\pi k} E_1 = \frac{\chi\sigma}{2K}$$

which is also the surface charge density of the dielectric as we saw in Eq. (3.6) and Eq. (3.7).

EXAMPLE 3.4 A dielectric slab of dielectric constant K extends from $z = 0$ to $z = -\infty$ with the top face of the slab along the $x - y$ plane, as shown in Fig. 3.12. A point charge q is placed at a point $z = z_0$ above the top face of the slab. Find the electric field just below and just above P where P is the point directly below the point charge on the surface of the dielectric.

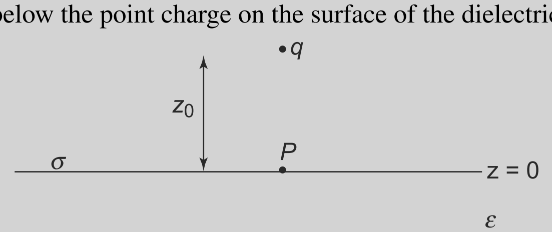


Fig. 3.12 *Example 3.4*

Solution

The slab in question is infinite along the $x - y$ plane. If it was finite, there would be surface charges on all the faces which will determine the field at P. In our case, only the surface charges on the top face will determine the electric field at P. Also, since the charges are all left-right symmetric around P, the field would be vertical. Call the fields E and E_1, just above the surface above P and at P respectively, both in the downward direction. By using the boundary condition, Eq. (3.13), we get:

$$E_1 = \frac{E}{K}$$

The polarisation $\vec{P}$ is simply

$$P = \frac{\chi}{4\pi k} E_1$$

which is also the surface charge density σ on the dielectric surface. Note that the field E_1 is less than E (since $K > 1$) because of the surface charge σ on the dielectric. σ in the present case is not uniform throughout the surface. However, in the present case, the field at the point P just below the surface due both to the charge q and the surface charge is perpendicular to the surface. Call this field E_s which is perpendicular to the surface going out of the surface on both sides. If we now consider a pill box like B as in Fig. 3.11 of infinitesimal area dA at the flat surfaces and apply Gauss's law to it, we get just like Eq. (3.30) above:

$$2E_s dA = 4\pi k\sigma dA$$

or

$$E_s = 2\pi\sigma k$$

This field subtracts from the field due to the free charge q inside the dielectric and adds to it outside. Hence,

$$E_1 = \frac{kq}{z_0^2} - \frac{4\pi k}{2}\left(\frac{\chi E_1}{4\pi k}\right) \tag{3.32}$$

since

$$P = \sigma = \frac{\chi}{4\pi k}E_1$$

Thus, we get

$$E_1(1 + \frac{\chi}{2}) = \frac{kq}{z_0^2} \tag{3.33}$$

Hence,

$$E_1 = \frac{kq}{z_0^2}\frac{1}{1 + \frac{\chi}{2}}$$

and

$$E = E_1(1 + \chi) = \frac{kq}{z_0^2}\frac{1 + \chi}{1 + \frac{\chi}{2}} \tag{3.34}$$

EXAMPLE 3.5 A uniform electric field $\vec{E} = E\hat{z}$ exists in the z-direction throughout space, as shown in Fig. 3.13. A sphere of dielectric constant ε and of radius R is introduced with its centre at the origin. Calculate the electric field outside and inside the sphere.

Solution

Since we have an electric field in the z-direction and the sphere is spherically symmetric, we have axial symmetry in the problem, i.e., the potential is unchanged if we rotate about the z-axis. The essential geometry of the problem is spherical and so we use spherical polar coordinates. In this, because of the axial symmetry, it is easy to see that the potential, ϕ is a function only of r and θ that is, $\phi = \phi(r, \theta)$.

Consider first the region $r > R$. There is no charge in this region and hence the potential satisfies the Laplace equation.

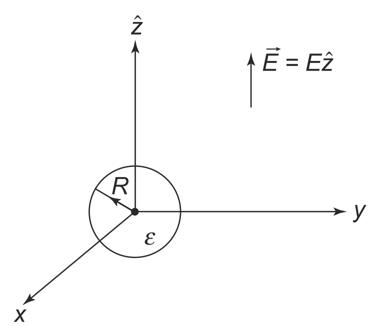


Fig. 3.13 *Example 3.5*

Using the Laplacian in spherical polar coordinates, we have:

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\phi(r,\theta)}{\partial r}\right)+\frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\phi(r,\theta)}{\partial\theta}\right)=0 \qquad (r>R) \tag{3.35}$$

The usual method to solve such a partial differential equation is by separation of variables which we described in the Advanced Topic in Chapter 2. We write

$$\phi(r,\theta)=R(r)\Theta(\theta)$$

and divide Eq. (3.35) by $\phi(r,\theta)$. We get

$$\frac{1}{R(r)}\frac{1}{r^2}\left(\frac{d}{dr}(r^2\frac{dR(r)}{dr})\right)+\frac{1}{\Theta(\theta)}\frac{1}{r^2\sin\theta}\left(\frac{d}{d\theta}(\sin\theta\frac{d\Theta(\theta)}{d\theta})\right)=0 \tag{3.36}$$

We can now multiply by r^2 to get

$$\frac{1}{R(r)}\left(\frac{d}{dr}(r^2\frac{dR(r)}{dr})\right)=-\frac{1}{\Theta(\theta)}\frac{1}{\sin\theta}\left(\frac{d}{d\theta}(\sin\theta\frac{d\Theta(\theta)}{d\theta})\right) \tag{3.37}$$

Notice that Eq. (3.37) has the following property- the left hand side is a function of r alone while the right hand side is a function of θ only. Thus, the two sides must be equal to a constant, h.

$$\frac{1}{R(r)}\left(\frac{d}{dr}(r^2\frac{dR(r)}{dr})\right)=h \tag{3.38}$$

and

$$-\frac{1}{\Theta(\theta)}\frac{1}{\sin\theta}\left(\frac{d}{d\theta}(\sin\theta\frac{d\Theta(\theta)}{d\theta})\right)=h \tag{3.39}$$

Rewriting Eq. (3.39) in terms of $z=\cos\theta$, we get

$$\frac{d}{dz}\left[(1-z^2)\frac{d\Theta(z)}{dz}\right]+h\Theta(z)=0 \tag{3.40}$$

This equation is Legendre's equation (see Mathematical Preliminaries) and is well known to have non-singular or regular solutions in the interval $-1<z<1$ with $h=l(l+1)$ where l is zero or a positive integer.

The solutions are denoted by $P_l(z)$ and the first few are given by

$$P_0(z)=1$$

$$P_1(z)=z$$

$$P_2(z)=\frac{3z^2-1}{2}$$

Since $h=l(l+1)$, we can label the solutions to the radial equation (Eq. (3.38)) as $R_l(r)$ and this gives the general solution to the Laplace equation (Eq. (3.35) as

$$\phi(r,\theta)=\sum_{l=0}^{\infty}A_l R_l(r)P_l(z) \tag{3.41}$$

where A_ls are constants.

Now Eq. (3.38) reads

$$\left(\frac{d}{dr}(r^2\frac{dR_l(r)}{dr})\right) = l(l+1)R_l(r) \tag{3.42}$$

Equation (3.42) has two solutions.

$$\begin{aligned} R_l(r) &= \text{ constant } r^l \\ R_l(r) &= \text{ constant } r^{-l-1} \end{aligned} \tag{3.43}$$

The most general solution thus to the Laplace equation becomes

$$\phi(r,\theta) = \sum_{l=0}^{\infty}(B_l r^l + C_l r^{-l-1})P_l(z) \tag{3.44}$$

To determine the constants B_l and C_l, we need to impose the boundary conditions. Let us first consider the boundary conditions at infinity. Since all the induced charges are at finite values of r, the field at infinity would not change by the introduction of the sphere at a finite distance. Thus, at $r = \infty$, we have

$$\vec{E} = E\hat{k}$$

and since $\vec{E} = -\vec{\nabla}\phi$, we get

$$\phi(r,\theta) = -Ez = -Er\cos\theta \qquad\qquad (r \to \infty) \tag{3.45}$$

Imposing this condition on the general solution (Eq. (3.44)), we get

$$B_1 = -E$$

and

$$B_l = 0 \qquad l \neq 1$$

Thus, we get

$$\phi(r,\theta) = -Er\cos\theta - \sum_{l=0}^{\infty}(C_l r^{-l-1})P_l(z) \tag{3.46}$$

The C_l's of course are yet undetermined and we would need the boundary conditions at $r = R$ to fix them.

For the region $r < R$, we see that there are no free charges and hence the relevant equation satisfied by the potential is still the Laplace equation. The general solution will be identical in form to Eq. (3.44) and we get

$$\phi(r,\theta) = \sum_{l=0}^{\infty}(B_l' r^l + C_l' r^{-l-1})P_l(z) \qquad\qquad 0 < r < R \tag{3.47}$$

where the constants B_l' and C_l' are different from the corresponding unprimed ones. At the origin, the potential must be finite since there is no point charge. Thus, all the C_l's must vanish in Eq. (3.47).

At $r = R$, we know that the potential $\phi(r, \theta)$, must be continuous across the boundary. Hence, from Eqs. (3.46) and (3.47), we get

$$-ER\cos\theta - \sum_{l=0}^{\infty}(C_l R^{-l-1})P_l(z) = \sum_{l=0}^{\infty}(B'_l R^l)P_l(\cos\theta) \tag{3.48}$$

We also know that the normal component of the displacement vector is continuous across a boundary at $r = R$. That is, the normal component of $\vec{D} = \varepsilon\vec{E} = -\varepsilon\vec{\nabla}\phi$ is continuous at $r = R$. This implies that, at $r = R$, $\frac{\partial\phi}{\partial r}$ when we approach from the outside, should be K times $\frac{\partial\phi}{\partial r}$ when approached from inside the sphere. Thus,

$$-E\cos\theta - \sum_{l=0}^{\infty} C_l(-l-1)R^{-l-2}P_l(\cos\theta) = K\sum_{l=0}^{\infty}(lB'_l R^{l-1})P_l(\cos\theta) \tag{3.49}$$

Both Eq. (3.48) and Eq. (3.49) involve sums over l. We now use the orthogonality properties of the Legendre polynomials.

$$\int_{-1}^{1} P_l(\cos\theta)P_m(\cos\theta)d(\cos\theta) = \frac{2}{2l+1}\delta_{lm} \tag{3.50}$$

Multiplying Eqs. (3.48) and (3.49) by $P_m(\cos\theta)$ and integrating over $\cos\theta$ from -1 to $+1$, we get

$$-\int_{-1}^{1} ER\cos\theta P_m(\cos\theta)d(\cos\theta) - C_m R^{-m-1}\left(\frac{2}{2m+1}\right) = \left(\frac{2}{2m+1}\right)B'_m R^m \tag{3.51}$$

and

$$-\int_{-1}^{1} E\cos\theta P_m(\cos\theta)d(\cos\theta) - C_m R^{-m-2}(-m-1)\left(\frac{2}{2m+1}\right) = K\left(\frac{2}{2m+1}\right)mB'_m R^{m-1} \tag{3.52}$$

The remaining integral in these two equations can be easily done by remembering that

$$\cos\theta = P_1(\cos\theta)$$

and therefore,

$$\int_{-1}^{1} \cos\theta P_m(\cos\theta)d(\cos\theta) = \int_{-1}^{1} P_1(\cos\theta)P_m(\cos\theta)d(\cos\theta) = \frac{2}{3}\delta_{m1}$$

Now multiplying Eq. (3.51) by $(-m-1)$ and Eq. (3.52) by R and subtracting the two, we get

$$-\frac{2}{3}(-m-2)ER\delta_{m1} = \left(\frac{2}{2m+1}\right)R^m B'_m(-m-1-Km) \tag{3.53}$$

Thus, we get

$$B'_1 = -\frac{3E}{2+K} \tag{3.54}$$

and

$$B'_m = 0 \quad m \neq 1 \tag{3.55}$$

Substituting these values into Eq. (3.48), we get

$$-\frac{2}{3}ER\delta_{m1} - C_m R^{-m-1}\left(\frac{2}{2m+1}\right) = -\left(\frac{2}{2m+1}\right)\frac{3E}{2+K}R\delta_{m1} \tag{3.56}$$

Therefore, we get

$$C_m = 0 \qquad\qquad m \neq 1$$

and

$$C_1 = -ER^3\left(\frac{K-1}{K+2}\right) \tag{3.57}$$

With all the constants determined, we now have the potential for all regions. Substituting in the general expressions, we get

$$\phi(r,\theta) = -Er\cos\theta + ER^3\left(\frac{K-1}{K+1}\right)r^{-2}\cos\theta \qquad\qquad r > R \tag{3.58}$$

and

$$\phi(r,\theta) = -\frac{3E}{2+K}r\cos\theta \qquad\qquad r < R \tag{3.59}$$

The field pattern in the presence and absence of the dielectric sphere is given below in Fig. 3.14. It is important to remember that this complete analytical solution to the Laplace equation was possible because of the symmetry of the problem (the sphere is completely spherically symmetric). In general, the solutions of the Laplace equation (and hence the potential and the field) can only be obtained numerically.

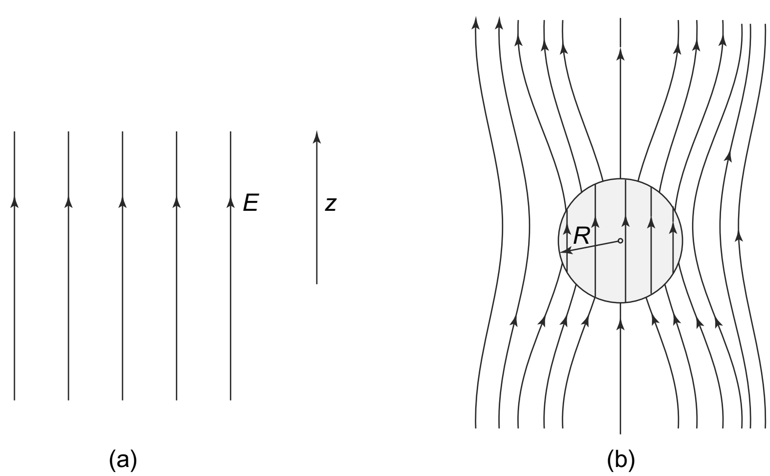


Fig. 3.14 *Example 3.5: (a) A dielectric sphere placed in an uniform electric field. The electric field $\vec{E}$ is parallel in the absence of the sphere, (b) Distorted field lines when a dielectric sphere is introduced. Near the sphere, the field lines bend towards the sphere*

PROBLEM 3.6 The plane $x = 0$ marks the boundary of two dielectrics with dielectric constants $K_1 = 10$ and $K_2 = 40$. The electric field next to the interface in Medium 1 ($K = 10$) is given by

$$\vec{E} = 10\hat{\imath} + 25\hat{\jmath} + 5\hat{k}$$

Find the electric field everywhere in space.

PROBLEM 3.7 A cylindrical conductor of radius 5 cm carries a uniform surface charge density of 100 μCm^{-2}. The cylinder is embedded in an infinite medium with dielectric constant $K = 10$. Determine $\vec{D}$ and $\vec{E}$ just above the conductor in the medium. Also determine the bound surface charge density per unit length on the surface of the medium next to the cylinder.

PROBLEM 3.8 Two infinite dielectric slabs with their faces parallel to the $x - y$ plane are separated by a gap of width d. The polarisation in both the slabs are uniform and same equal to $\vec{P} = P_0(\hat{\imath} + \hat{k})$. Calculate the electric field in the gap.

PROBLEM 3.9 A thin hollow spherical shell radius R with its centre at the origin has charge Q uniformly distributed on it. It is immersed in a dielectric medium with non-uniform dielectric constant $K(r) = 2 + \frac{R}{r}$. Determine the electric field as a function of r for $r > R$.

PROBLEM 3.10 The region $z < 0$ for all values of x and y is at a potential V_0 and similarly the region $z > a$ (with $a > 0$) is at zero potential for all x and y. Starting from Laplace equation, calculate the electric field at all points.

3.6 AVERAGE AND LOCAL FIELDS

3.6.1 Average and Local Fields

As we have seen above, and as is expected, the electric field inside a dielectric will change drastically across molecular distances. The reason for this is basically the localisation of the charge densities. Therefore, it makes little sense to talk of electric field (which, you would remember is a vector field and hence is defined at each point) as such. However, a meaningful concept is the average electric field $\vec{E}$ which is the 'average' over many molecular distances. This average field should be such that the line integral of $\vec{E}$, that is $\int \vec{E} \cdot d\vec{l}$ along a line (contour) should equal the potential difference between the ends of the line (contour). It is also important to note that the displacement vector $\vec{D}$ which we introduced above, also is related to this average field $\vec{E}$ as the surface integral of $\vec{D}$ over any Gaussian surface is related to the **free** charge inside the surface.

In a sense then, neither $\vec{E}$ nor $\vec{D}$ have any relevance at microscopic scales and are only meaningful as averages. However, it is clear that determining the electric field acting at the site of the individual molecules is very important since the molecules obviously only feel the field at their site (which is usually due to both the free charges as well as the induced dipoles) and respond to it. Neither $\vec{E}$ nor $\vec{D}$ correspond to this and we shall be using the symbol $\vec{E}_{\text{local}}$ for this local field.

3.6.2 Local Field Calculations

Definitions: Before we attempt to calculate the local electric field at the site of a molecule, we need to define it somewhat more precisely. Imagine a dielectric in which all molecules, except one, are frozen. The single molecule which is not frozen is scooped out. This scooping out of the molecule will leave a 'cavity' in the dielectric whose shape is not well defined. This is because, even before the molecule was scooped out, there were empty spaces in the dielectric and thus what the shape of the cavity is, is not easily determined. Let us however try out various shapes and see what it gives us. Assume that the average electric field is uniform as is $\vec{D} = \varepsilon \vec{E}$.

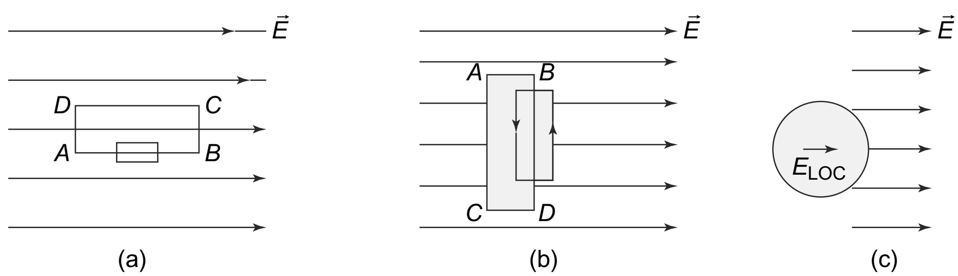


Fig. 3.15 *Three configurations of cavities inside a dielectric for calculation of the local field inside the dielectric: (a) The cavity is a pill box with infinitesimal breadth and length parallel to the electric field. A rectangle ABCD is such that one of the long sides (parallel to E) lies within and one outside the pillbox, (b) The cavity is a pill box again but now with the longer sides perpendicular to the electric field direction. Once again the rectangle has one of its longer sides within and one outside the pill box, (c) The cavity is a spherically shaped*

Figure 3.15 shows the three cavities that one can consider to be left when a molecule is removed from the dielectric. Consider Fig. 3.15(a). The points A and D and B and C are infinitesimally apart and so the potential is the same at these pairs of points. However, the potential difference between A and B is simply

$$\int_A^B \vec{E} \cdot d\vec{l} = E_{\text{local}} AB$$

Similarly, the potential difference between C and D is

$$\int_C^D \vec{E} \cdot d\vec{l} = E \quad CD$$

Now these must be equal and since $AB = CD$ as we have taken the potential to be the same at A and D and B and C. Thus, we get

$$\vec{E}_{\text{local}} = \vec{E} \tag{3.60}$$

Thus, in the case of this pill box shaped cavity left on removing a molecule, the local field $\vec{E}_{\text{local}}$ is equal to the average field $\vec{E}$.

For the cavity as in Fig. 3.15(b), the infinitesimal sides are along $\vec{E}$ and the longer sides are perpendicular to it. Let us consider a closed surface in the shape of the rectangle with the longer sides just inside and just outside the cavity. Since there are no free charges inside this closed surface, the surface integral of $\vec{D}$ should vanish as is seen from Gauss's Law for dielectrics, Eq. (3.11). The surface integral of the surface below AC and BD vanishes since the electric field is perpendicular to the normal to the surface (the direction of $d\vec{S}$). Thus, we get

$$D_{\text{in}} \times A = D_{\text{out}} \times A = 0 \tag{3.61}$$

where D_{in} is the value of $\vec{D}$ below CD and D_{out} the value just below AB and A is the area of the surfaces. From the relationship between $\vec{D}$ and $\vec{E}$, we know that outside the cavity $D_{\text{out}} = \dfrac{KE_{\text{out}}}{4\pi k}$ while inside it is $D_{\text{in}} = \dfrac{E_{\text{in}}}{4\pi k}$. But E_{out} is the field outside the cavity, which by definition is the average field $\vec{E}$. Similarly, E_{in} is the local field $\vec{E}_{\text{local}}$. Thus, we get

$$\begin{aligned}\vec{E}_{\text{local}} &= K\vec{E}\\ &= (1+\chi)\vec{E}\\ &= \vec{E} + 4\pi k\vec{P}\end{aligned} \tag{3.62}$$

In this case, we see that the local field at the site of the molecule is different from the average field because of the induced charges.

Cavities (a) and (b) did not need an elaborate calculation of the field because of their geometry. However, the spherical cavity in Fig. 3.15(c) requires a detailed calculation.

One simple way to do this is to see that the average field in the case of a spherical cavity would simply be the vector sum of the local field $\vec{E}_{\text{local}}$ and the field produced by a uniformly polarised sphere that we scooped out in the process of creating the cavity. Thus,

$$\vec{E} = \vec{E}_{\text{local}} + (\text{field created by the polarised sphere}) \tag{3.63}$$

The field created by a uniformly polarised sphere is most easily calculated by considering it as being two uniformly charged spheres, carrying equal and opposite charges, with their centres displaced along $\vec{P}$ as shown in Fig. 3.16(a).

If V is the volume of the spheres, then the polarisation is simply $NV\vec{p} = V\vec{P}$ where $\vec{p}$ is the polarisation of a single molecule which is simply the charge times the distance ($\vec{p} = q\vec{d}$). The total positive (or negative) charge in the sphere is $\pm Q = \pm NVq$. Once we take the entire polarisation to be due to these two displaced spheres, one can simply find the electric field due to them and add them. For the positively charged sphere, the field at an internal point T is caused by a sphere of radius r whose charge is concentrated at the centre. Thus,

$$\vec{E}_+ = \hat{n}_+ \frac{k}{r^2} \times (\text{charge inside the sphere of radius } r) \tag{3.64}$$

The charge in the sphere can be easily calculated and we get

$$\vec{E}_+ = \hat{n}_+ \frac{kQ}{r^2}\frac{4\pi r^3}{3V} = \hat{n}_+ \frac{4\pi kQr}{3V} \tag{3.65}$$

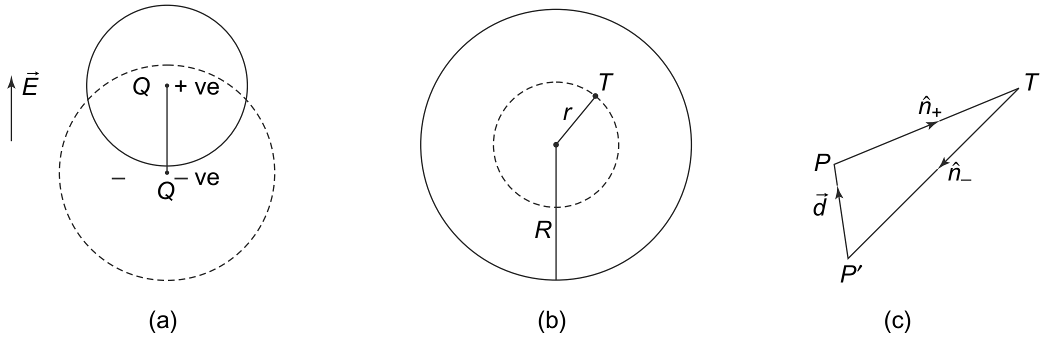


Fig. 3.16 *Electric field due to a uniformly charged sphere: (a) The dipoles can be considered as being formed from two spherical, charged spheres with charge $+Q$ and $-Q$, and their centres displaced by d such that $Q\vec{d} = V\vec{P}$ where $\vec{P}$ is the polarisation, (b) At any point T, the field due to either of the charged spheres will be will be due to a sphere of radius r, (c) The electric field due to the positively charged sphere will be along TP while that due to the negatively charged sphere will be along TP'*

Similarly, the field due to the negatively charged sphere will be

$$\vec{E}_- = \hat{n}_- \frac{kQ}{r^2}\frac{4\pi r^3}{3V} = \hat{n}_- \frac{4\pi kQr}{3V} \tag{3.66}$$

Adding the two gives us the field due to a uniformly polarised sphere as

$$\vec{E}_+ + \vec{E}_- = \frac{4\pi kQ}{3V}(-\vec{d}) \tag{3.67}$$

since $r(\hat{n}_+ + \hat{n}_-) = -\vec{d}$ from Fig. 3.16(c). Using $Q\vec{d} = V\vec{P}$, we get

$$\text{electric field due to the polarised sphere} = -\frac{4\pi k}{3}\vec{P} \tag{3.68}$$

The local field is therefore, from Eq. (3.63)

$$\vec{E}_{\text{local}} = \vec{E} + \frac{4\pi k}{3}\vec{P} \tag{3.69}$$

which we see is in between the value obtained for the case of two pill box shaped cavities considered above.

3.6.3 Claussius— Mosotti Relation

In our initial discussion of the phenomenon of polarisation in a dielectric at the atomic level, we had tacitly assumed that the polarisation is due to the average field $\vec{E}$. This is obviously not strictly correct, as seen above, since the molecule will actually experience a field at its location ($\vec{E}_{\text{local}}$) rather than $\vec{E}$ and hence the polarisation will be due to $\vec{E}_{\text{local}}$. Thus, Eq. (3.1) should strictly read

$$\vec{p}_0 = \frac{1}{4\pi k}\alpha\vec{E}_{\text{local}} \tag{3.70}$$

We however, see a problem immediately- as we saw above, defining the local field as the field inside a scooped out cavity region, depends on the shape of the cavity which of course, is unknown! There is of course, no way to know the shape of the cavity which allows us to calculate the local field. However, it seems reasonable to expect that for isotropic gases and liquids, the local field would be that inside a spherical shaped cavity at the site of the individual molecule. Using this assumption, we will try to find a relationship between the dielectric constant (or more strictly the permittivity ε which is related to the dielectric constant K by a constant factor) and the atomic polarisability α defined in Eq. (3.1).

If N is the number of molecules per unit volume and $\vec{p}_0$ is the dipole moment of each molecule, then $\vec{P} = N\vec{p}_0$. Using the expression (Eq. (3.69) for $\vec{E}_{\text{local}}$, we get

$$\begin{aligned} \vec{P} &= N\vec{p}_0 \\ &= \chi \frac{\vec{E}}{4\pi k} \\ \vec{P} &= \frac{N\alpha}{1 - \frac{N\alpha}{3}} \frac{\vec{E}}{4\pi k} \end{aligned} \tag{3.71}$$

Since

$$\vec{P} = \varepsilon \frac{\vec{E}}{4\pi k}$$

we get

$$\boxed{\begin{aligned} K &= 1 + \chi \\ &= 1 + \frac{N\alpha}{1 - N\alpha/3} \\ \text{or} \\ \alpha &= \frac{3}{N}\left(\frac{K-1}{K+2}\right) \end{aligned}} \tag{3.72}$$

This relation is called the **Clausius-Mosotti relation**, which was given in the middle of the nineteenth century by Mosotti and independently by Clausius. The importance of this relationship is that it allows us to relate a purely microscopic quantity, namely the atomic polarisability α to a macroscopic quantity ε which is a bulk property of the material. In other words, while α depends on the structure and electrical properties of an individual molecule, ε is dependent on the arrangement of these molecules in the material.

3.7 ADVANCED TOPICS

3.7.1 Multipole Expansion and Multipole Moments

The electrostatic potential created by a charge distribution $\rho(\vec{r})$ at a point P with coordinate $\vec{R}$ is given by

$$\phi(\vec{R}) = k \iiint\limits_V \frac{\rho(\vec{r}')}{|\vec{R} - \vec{r}'|} d^3r' \tag{3.73}$$

where the integral is over the entire volume where $\rho \neq 0$. We choose the origin to be in the region of the charge distribution so that range of $\vec{r}'$ is of the order of the dimensions of the charge distribution, say r_C. If P is far away from the distribution, so that $|\vec{R} - \vec{r}'| \approx R$, then $\phi(\vec{R})$ reduces to $\frac{kQ}{R}$ where Q is the total charge of the distribution, that is

$$Q = \iiint_V \rho(\vec{r}')d^3r'$$

It is frequently very useful to have an expansion of $\phi(\vec{r}')$ in powers of $(\frac{r_C}{R})$ beyond this zeroth order approximation. The multipole expansion is precisely such an expansion. Expanding the denominator of Eq. (3.73) in powers of $(\frac{r'}{R})$,

$$\begin{aligned} |\vec{R} - \vec{r}'|^{-1} &= (R^2 + r'^2 - 2\vec{R}\cdot\vec{r}')^{-1/2} \\ &= \frac{1}{R}\left[1 + (\hat{R}\cdot\hat{r}')\left(\frac{r'}{R}\right) + \frac{3(\hat{R}\cdot\hat{r}')^2 - 1}{2}\left(\frac{r'}{R}\right)^2 \right. \\ &\qquad \left. + \frac{5(\hat{R}\cdot\hat{r}')^3 - 3(\hat{R}\cdot\hat{r}')}{2}\left(\frac{r'}{R} + \cdots\right)\right] \end{aligned} \tag{3.74}$$

where $\hat{R} = \frac{\vec{R}}{R}$ and $\hat{r}' = \frac{\vec{r}'}{r'}$ and we have rearranged the binomial expansion of the denominator in terms of powers of $(\frac{r'}{R})$.

Note that the second term in the expansion is precisely the dipole moment term that we have already discussed. The first term is called the monopole moment while the third term in the expansion is called the quadrupole moment. Substituting Eq. (3.74) in Eq. (3.73) we get

$$\phi(R) = \frac{kQ}{R} + k\iiint d^3r'\hat{r}'\cdot\hat{R}\left(\frac{r'}{R^2}\right)\rho(\vec{r}') + \frac{k}{2}\iiint d^3r'\rho(\vec{r}')\left(3\hat{r}'_i\cdot\hat{r}'_j - \delta_{ij}\right)\cdot\hat{R}_i\hat{R}_j\left(\frac{r'^2}{R^3}\right) + \cdots \tag{3.75}$$

This expression for the potential has some interesting properties. Note that for a spherically symmetric charge distribution, i.e., a spherically symmetric $\rho(\vec{r}')$, the potential only depends on Q and R since

$$\iiint d^3r' \;\; \rho(\vec{r}")\hat{r}'_i = \iiint d^3r' \;\; (3\hat{r}'_i\hat{r}'_j - \delta_{ij})\rho(\vec{r}') = 0$$

The first two terms of Eq. (3.75) are familiar as the potential due to a charge Q and that due to a dipole. The quantity

$$Q_{ij} = \iiint d^3r'\rho(\vec{r}')(3\hat{r}'_i\hat{r}'_j - \delta_{ij})r'^2$$

is called the quadrupole moment of the charge distribution. This quantity has interesting properties.

First, observe that $\sum_i Q_{ii} = 0$ (because $\sum_i 3(\hat{r}'_i\hat{r}'_i)r'^2 = 3r'^2$ and $\sum_i \delta_{ii}r'^2 = 3r'^2$). The elements of Q_{ij} are such that $Q_{ij} = Q_{ji}$ that is to say that the elements are symmetric. This is obvious since $3\hat{r}'_i\cdot\hat{r}'_j - \delta_{ij} = 3\hat{r}'_j\cdot\hat{r}'_i - \delta_{ji}$. Now, if we think of Q_{ij} as a 3×3 matrix, then a matrix with these properties is called a symmetric, traceless matrix. The actual nature of the quantity Q_{ij} is that of a symmetric, traceless tensor of rank 2. (See Mathematical Preliminaries for details).

If we consider the quadrupole moment as a 3×3 matrix with the properties given above, then it is easy to see that it has only 5 independent component (a 3×3 matrix has 9 components. Since the trace or the sum of its diagonal terms is zero, that reduces the number of independent components by 1. Also, it is symmetric and thus three more are reduced and so we get $9 - 1 - 3 = 5$ independent components.)

Note that by definition, the value of Q_{ij} is dependent on the choice of origin. Thus, if we shift the origin such that

$$r'_i \to t_i = r'_i + d_i$$

then

$$Q_{ij} = \iiint d^3r'\rho(r')(3r'_ir'_j - r'^2\delta ij) \to Q'_{ij} = \iiint d^3t\rho'(t)(3t_it_j - t^3\delta_{ij}) \tag{3.76}$$

We have written $\rho'(t)$, the charge density as a function of t. A given point with coordinate $\vec{r'}$ and charge density $\rho(r')$, after the shift of origin has coordinates $\vec{t}$ and charge density $\rho'(t)$. Therefore, $\rho'(t) = \rho(r')$.

In general, $Q_{ij} \neq Q'_{ij}$ since

$$\begin{aligned}
Q'_{ij} &= \iiint d^3t\rho'(t)(3t_it_j - t^2\delta_{ij}) \\
&= \iiint d^3t\rho'(t)\left[3(r'_i + d_i)(r'_j + d_j) - (r'^2 + d^2 + 2r'd)\delta_{ij}\right] \\
&= \iiint d^3r'\rho(r')(3r'_ir'_j - r'^2)\delta_{ij} + 3d_id_jQ + 3d_ip_j + 3d_jp_i - d^2\delta_{ij}Q - 2\delta_{ij}dp \\
&= Q_{ij} + (3d_id_j - d^2\delta_{ij})Q + (3d_ip_j + d_jp_i - 2\delta_{ij}dp)
\end{aligned} \tag{3.77}$$

Here Q is the total charge and p_i is the i^{th} component of the dipole moment. Thus, we see that $Q_{ij} = Q'_{ij}$ only if the total charge Q and the dipole moment $\vec{p}$ vanish. This is not usually the case in nature for instance in atoms and nuclei, which is what would concern us. This ambiguity in the definition of the quadrupole moment is usually resolved by choosing the centre of mass as the origin to define the quadrupole moment.

Let us try and calculate the quadrupole moment of a simple charge distribution.

EXAMPLE 3.6 Calculate the quadrupole moment of a uniformly charged ellipsoid, defined by $\frac{z^2}{a^2} + \frac{x^2}{b^2} + \frac{y^2}{b^2} = 1$

Solution

Let the charge density of the ellipsoid by ρ. Then the total charge, or the monopole moment is simply

$$Q = \rho \iiint d^3r = \frac{4\pi}{3}\rho ab^2$$

The quadrupole moment is given by

$$Q_{ij} = \rho \iiint d^3r(3r_ir_j - r^2\delta_{ij})$$

where the limits of the integration are

$$-a \leq z \leq a$$

$$-b\sqrt{1-\frac{z^2}{a^2}} \leq x \leq b\sqrt{1-\frac{z^2}{a^2}}$$

and

$$-b\sqrt{1-\frac{z^2}{a^2}-\frac{x^2}{b^2}} \leq y \leq b\sqrt{1-\frac{z^2}{a^2}-\frac{x^2}{b^2}}$$

Because of the symmetry, the integral over $r_i r_j$ vanishes unless $i = j$. Therefore,

$$Q_{ij} = 0 \qquad\qquad \text{if } i \neq j$$

The non-vanishing terms of the quadrupole tensor can be obtained by writing

$$Q_{11} = \rho \iiint d^3r(3x^2 - r^2) = \rho \int dy \int dz \int dx(2x^2 - y^2 - z^2)$$

$$Q_{22} = \rho \iiint d^3r(3y^2 - r^2) = \rho \int dx \int dz \int dy(2y^2 - x^2 - z^2)$$

$$Q_{11} = \rho \iiint d^3r(3z^2 - r^2) = \rho \int dy \int dx \int dz(2z^2 - y^2 - x^2)$$

which can be easily evaluated with the above limits to get

$$Q_{11} = Q_{22} = -\frac{1}{5}Q(a^2 - b^2)$$

and

$$Q_{33} = \frac{2}{5}Q(a^2 - b^2)$$

3.7.2 Moments in Spherical Basis

Instead of using Cartesian vectors like r_i, it is much more convenient to express the multipole moments of a charge distribution in terms of a spherical basis. In particular, this representation of multipole moments is extremely useful in the study of atomic and nuclear systems.

To do this, we first expand

$$\frac{1}{|\vec{R} - \vec{r}'|} = \sum_{l=0}^{\infty} \frac{r'^l}{R^{l+1}} P_l(\cos\gamma) \tag{3.78}$$

where the P_l's are the Legendre polynomials and γ is the angle between $\hat{R}$ and $\hat{r}'$. We also use the relationship between Legendre Polynomials and Spherical Harmonics

$$P_l(\cos\gamma) = \left(\frac{4\pi}{2l+1}\right) \sum_{m=-l}^{+l} Y_{lm}(\theta, \phi) Y^*_{lm}(\theta', \phi') \tag{3.79}$$

where the angles $\theta, \phi, \theta', \phi'$ are defined by the components of $\hat{R}$ and $\hat{r}'$.

$$\hat{R} = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\phi)$$

and

$$\hat{r}' = (\sin\theta' \cos\phi', \sin\theta' \sin\phi', \cos\phi')$$

Substituting Eqs. (3.78) and (3.79) in Eq. (3.73), we get

$$\begin{aligned}\phi(\vec{R}) &= k \iiint d^3r' \rho(\vec{r}') \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \left(\frac{4\pi}{2l+1}\right) \\ &= k \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \left(\sqrt{\frac{4\pi}{2l+1}}\right) \frac{Y_{lm}(\theta,\phi)}{R^{l+1}} Q_{lm} \end{aligned} \tag{3.80}$$

where Q_{lm} are the multipole moments defined by

$$Q_{lm} = \left(\sqrt{\frac{4\pi}{2l+1}}\right) \iiint d^3r' \rho(r') r'^l Y^*_{lm}(\theta', \phi') \tag{3.81}$$

with m going from $-l$ to l.

The $l = 0$ moment is just the total charge Q since

$$Y_{00} = \frac{1}{\sqrt{4\pi}}$$

The $l = 1$ moments for $m = -1, 0, 1$ are the three components of the dipole moment vector in spherical basis. The $l = 2$ term is the quadrupole term with components $m = -2, -1, 0, 1, 2$ which are the five independent components of Q_{2m} in spherical basis.

3.7.3 Polar Molecules and Langevin–Debye Theory

In this chapter, we have mostly talked about the effect of an external electric field on a dielectric. Dielectrics, as we know, are substances which differ from conductors in that they have no free charges. Hence, it is possible to establish an electric field inside a dielectric. In the presence of an external field, the positive and negative charges in the molecule of the dielectric substance get displaced in opposite directions and the dielectric is said to polarised. In other words, the molecules of the dielectric have an induced dipole moment because of the displacement of the centre of the positive and negative charges. The dipole moment thus induced produces an electric field which obviously adds vectorially to the external field to produce the net field inside and outside of the dielectric.

There are some dielectrics which have an inherent dipole moment. A molecule which is made up of similar atoms can never have a permanent dipole moment since the atoms cannot be arranged in an asymmetrical way. Typically, molecules which are made up of two or more dissimilar atoms can possess a permanent dipole moment. Water is one such polar molecule. In the water molecule, the lines joining the centre of the oxygen atom to the centres of the hydrogen atoms makes an angle of 104°. Thus, each pair of oxygen and hydrogen atom has a dipole moment and both these dipole moments, $\vec{p}_1$ and $\vec{p}_2$ add vectorially to give a net dipole moment $\vec{p}$ whose magnitude is 6.1×10^{-30} C m. When such a molecule is placed in an external electric field $\vec{E}$, it experiences a torque as we have seen in Chapter 2. This torque tends to align the dipole with the field. However, the molecules in the dielectric also experience thermal motion and this can tend to destroy this alignment of dipoles. In equilibrium, at any given temperature, some net alignment therefore exists in the presence of an electric field.

Recall that the polarisation is defined as

$$\vec{P} = \frac{1}{\Delta v} \sum \vec{p}_m$$

where $\vec{p}_m$ is the molecular dipole moment and the summation extends over all molecules in a volume Δv. At any given temperature, the dipoles $\vec{p}_m$ are aligned at random and hence their summation vanishes. However, in the presence of an external field, the dipoles align themselves and if there is complete alignment, we get

$$\vec{P} = N \vec{p}_m$$

per unit volume (containing N molecules).

As the temperature increases, the random motion tends to destroy this alignment. Now we know from Statistical Mechanics that at a temperature T, in a large sample of molecules with a distribution of energies, the number with a particular energy W is proportional to

$$\sim e^{-W/kT}$$

where k is the Boltzmann constant.* For instance, in a perfect gas, the molecules obey the Maxwellian distribution of velocities and hence the probability of finding a molecule with velocity v is proportional to

$$\sim e^{-mv^2/2kT}$$

However, the polar molecule gas that we are considering is not perfect gas since there are interactions between the molecules and hence there is potential energy contribution to the total energy and the proportionality factor thus, becomes

$$\sim e^{-\frac{mv^2}{2kT}} e^{-\frac{W_p}{kT}}$$

where W_p is the potential energy of the molecule.

We have already seen in Chapter 2 that the potential energy of a dipole, with dipole moment $\vec{p}$ in an electric field $\vec{E}$ (this is the net electric field experienced by the molecule, which includes contributions from the external field and the field due to all other dipoles in the material) is given by

$$W_p = -\vec{p} \cdot \vec{E} = -pE \cos\theta$$

The effective component of the dipole moment is from this, the component along the field direction, $p \cos\theta$. The average value of this in this distribution can be found by

$$< p \cos\theta >= \frac{\iint p \cos\theta e^{pE \cos\theta/kT} d\Omega}{\iint e^{pE \cos\theta/kT} d\Omega}$$

*Throughout this section, k denotes the Boltzmann constant and should not be confused with k in Coulomb's Law.

where $d\Omega$ is the element of the solid angle. Assuming azimuthal symmetry, we can do the ϕ integral in the solid angle and get

$$< p\cos\theta >= \frac{\int_0^\pi p\cos\theta\sin\theta e^{pE\cos\theta/kT}d\theta}{\int_0^\pi e^{pE\cos\theta/kT}\sin\theta d\theta}$$

Substituting

$$z = \frac{pE}{kT}$$

and

$$t = \frac{\vec{p}\cdot\vec{E}}{kT} = z\cos\theta$$

we get

$$< p\cos\theta > = \frac{p/z\int_{-z}^{z} e^t t\,dt}{\int_{-z}^{z} e^t dt} \tag{3.82}$$

$$= p\left(\coth z - \frac{1}{z}\right) \tag{3.83}$$

$$= p\left(\coth\frac{pE}{kT} - \frac{kT}{pE}\right) \tag{3.84}$$

Equation (3.82) is known as the Langevin equation and was derived first by Langevin in the context of magnetic dipoles. This is depicted in Fig. 3.17.

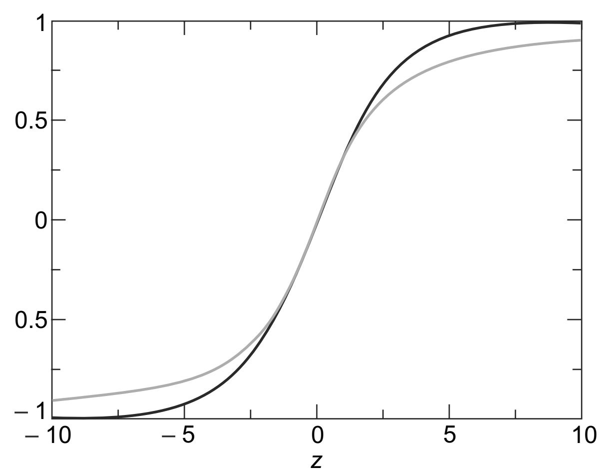


Fig. 3.17 *Langevin function*

What is plotted on the graph is $\frac{<p\cos\theta>}{p}$ vs z as the light line. The dark line is the graph for $\tanh(\frac{z}{3})$. We see that there is a saturation at large values of z or E which is what we expect since for large E we expect total alignment. For small values of z, the curve is linear and this is the region of interest at ordinary temperatures. For instance, in the water molecule, we saw that the dipole moment is 6.1×10^{-30} C m.

Therefore, even with a high field intensity of 10^7 volts/meter, z is of the order of 10^{-3}. Thus we expect the dielectric to behave linearly in this region.

Expanding the exponential and retaining terms upto order z^3, we get

$$< p\cos\theta > \approx \frac{1}{3}pz = \frac{p^2E}{3kT}$$

But $< p\cos\theta >$ is the average dipole moment and therefore, the total polarisation of the dielectric can be obtained as

$$\vec{P} = N < p\cos\theta > = \frac{Np^2E}{3kT}$$

where N is the number density of the molecules.

This allows us to read the polarisability α (defined as the dipole moment per unit field) to be

$$\alpha = \frac{p^2}{3kT}$$

This result has been obtained ignoring the effect of the induced dipoles. In general therefore, the total polarisability is given by

$$\boxed{\alpha = \alpha_0 + \frac{p^2}{3kT}}$$

This is known as the Langevin–Debye equation.

We can also relate this to the dielectric constant K which we recall is related to the electric susceptibility χ_e by

$$K = 1 + \chi_e$$

where χ_e is defined by the relation between the polarisation $\vec{P}$ and the electric field $\vec{E}$

$$\vec{P} = \varepsilon_0\chi_e\vec{E}$$

We have also seen in the Chapter that the local field, or the field which acts on an individual dipole is given by

$$\vec{E} = \vec{E}_{\text{ext}} + \frac{\vec{P}}{3\varepsilon_0}$$

where $\vec{E}_{\text{ext}}$ is the external field and the second term is the effect of all other dipoles in the material.

Therefore, we get, using these results

$$\vec{P} = N\left(\alpha_0 + \frac{p^2}{3kT}\right)\vec{E}$$

which gives us

$$\varepsilon = \left(\frac{K+2}{3}\right)E_{\text{ext}}$$

or

$$\frac{K-1}{K+2} = \frac{N}{3\varepsilon_0}\left(\alpha_0 + \frac{p^2}{3kT}\right)$$

This can be put in a more conventional form by using the fact that

$$N = \frac{\rho}{M} N_A$$

where ρ is the mass density, M is the molecular weight and N_A is the Avogadro number to get

$$\boxed{\alpha_M = \frac{N_A}{3\varepsilon_0}\left(\alpha_0 + \frac{p^2}{3kT}\right)}$$

This is known as the Debye equation.

3.7.4 Polarisability Tensor

We have seen in the Chapter that the dipole moment induced by an external field in an atom or molecule is usually expressed as

$$\vec{p} = \alpha \vec{E}$$

where α is the atomic or molecular polarisability. This relationship of course assumes that the dipole moment is induced in the direction of the external electric field and is proportional to the applied field. However, it is seen that in the case of molecules, the situation is very different. Because of the structure of the molecule, it frequently happens that the molecule is more easily polarised along some directions than others. For instance, the carbon dioxide molecule has a polarisability which is twice along the axis of the molecule, i.e. along the line joining the carbon and two oxygen atoms, than in directions orthogonal to the axis. In this case, we can resolve the external field in the two directions, along and perpendicular to the molecular axis and thus, get

$$\vec{p} = \alpha_{||}\vec{E}_{||} + \alpha_{\perp}\vec{E}_{\perp}$$

In general, in more complicated molecules, the relation between $\vec{p}$ and $\vec{E}$ is replaced by

$$\begin{aligned} p_x &= \alpha_{xx}E_x + \alpha_{xy}E_y + \alpha_{xz}E_z \\ p_y &= \alpha_{yx}E_x + \alpha_{yy}E_y + \alpha_{yz}E_z \\ p_x &= \alpha_{zx}E_x + \alpha_{zy}E_y + \alpha_{zz}E_z \end{aligned} \tag{3.85}$$

The α_{ij} are nine components and form the **polarisability tensor**. Their values basically tell us how the molecule behaves in the presence of an external field along any direction. It is in some ways like the tensor in elasticity, which tells us how the material will change in the presence of external stress along any direction.

In a later Chapter, we will see that light is an electromagnetic wave and is basically oscillating electric and magnetic fields which travel as waves. Thus, light when it falls on a material and interacts with it, interacts because of the electric field. The polarisation of light caused by certain materials can be related to the polarisability tensor given above.

SUMMARY

- The electrons in a dielectric are not free to move inside the material when an external electric field is applied.
- In some materials, the molecules have charge distribution such that an electric dipole moment is present in the molecules even in the absence of any external electric field. In other materials, dipole moment develops on when an external electric field is applied.
- Gauss's Law is modified in a dielectric because of dipole moments created by the electric fields.
- At the interface between two dielectrics, the normal component of the displacement vector and the tangential component of the electric field remain continuous.

CONCEPTUAL QUESTIONS

1. Despite charged particles being present in a dielectric, they do not freely move when an external electric field is applied. This is because:
  a. of collisions of these particles when they start moving.
  b. they are bound tightly into overall neutral objects.
  c. the nuclei of dielectric materials are very heavy,
  d. the electric field inside a dielectric vanishes in the presence of any external electric field.

2. Different dielectrics have different values of dielectric constant. This is because:
  a. their crystals have different arrangement of atoms.
  b. their electronic charge distributions around the nucleus is different.
  c. the dielectric constant of different dielectrics are in the ratio of the number of electrons in an atom in them.
  d. dielectric constant is inversely proportional to atomic mass.

3. In a non-polar dielectric, in the presence of an external electric field, the induced dipole moments of the molecules
  a. are all aligned in the same direction.
  b. are randomly arranged.
  c. are in all directions but with a preferential direction.
  d. lead to a temperature independent dielectric constant.

4. The electric field that one deals with in writing down the laws of electrostatics in a dielectric is
  a. the value of the field at any point inside a dielectric
  b. its average value averaged over the entire sample.
  c. its average value averaged over a region with many molecules.
  d. the value at a point after imagining a region being scooped out around the point.

5. At an interface between free space and a dielectric of dielectric constant 5, the normal component of the $\vec{D}$ vector in free space is 10 C/m^2 and the tangential component of the $\vec{E}$ field in the dielectric is 100 V/m. Find their counterparts in the other media.

6. In Question 5, what is the bound surface charge density?

7. A dielectric material contains 2×10^{19} polar molecules per cubic meter, each of dipole moment 1.8×10^{-27} C/m. Assuming all dipoles are aligned to a field of $\vec{E} = 10^5 \hat{x}$ V/m, find $\vec{P}$.

8. Region I $(z < 0)$ has dielectric constant 2.5 while region II $(z > 0)$ has dielectric constant 4. If $\vec{E}_1 = -30\hat{x} + 50\hat{y} + 70\hat{z}$, find $\vec{D}_2$.

9. A slab of dielectric with dielectric constant 2.4 has $V = 300z^2$V. Find $\vec{D}$ and ρ_{free}.

10. The polarisation vector $\vec{P} = x\hat{x} + y\hat{y} + x\hat{z}$ in a cube of side b with its centre at the origin. Find the bound volume charge density.

PROBLEMS

1. Consider an anisotropic medium where the dielectric constant or relative permittivity is dependent on the direction. Given that the permittivity tensor is

$$\varepsilon = \varepsilon_0 \begin{bmatrix} 4 & 1 & 1 \\ 1 & 4 & 1 \\ 1 & 1 & 4 \end{bmatrix} \tag{3.86}$$

Obtain the value of D for $\vec{E} = 10\hat{i} + 19\hat{j}$ V/m

2. A thin rod of cross-section area A is along the x-axis from $x = 0$ to $x = L$. The polarisation on the rod is along its length and is given by $P_x = ax^2 + b$ where a, b are constants. Calculate the value of the surface and volume-bound charge densities at each end of the wire. Show that the total bound charge is zero.

3. Two homogenous isotropic dielectrics meet on the plane $z = 0$, as shown in Fig. 3.18. For $z \geq 0$, $\varepsilon_1 = 4$ while for $z \leq 0$, $\varepsilon_2 = 3$. A uniform electric field $\vec{E}_1 = 5\hat{i} - 2\hat{j} + 3\hat{k}$ kV/m is there for $z \geq 0$. Find the value of the electric field $\vec{E}_2$ inside the dielectric 2 for $z \leq 0$ and the angles $\vec{E}_1$ and $\vec{E}_2$ make at the interface.

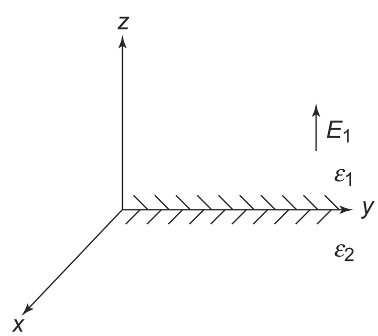


Fig. 3.18 *Problem 3*

4. Consider two conducting concentric spherical shells, of radii a and b where $b > a$ maintained at a potential difference of V_0, as shown in Fig. 319. The potential at the outer sphere is zero while the potential on the inner sphere is V_0. Determine V, $\vec{E}$ in the region between the spheres. If the region between the spheres has a dielectric constant K, determine the charge induced on the shells.

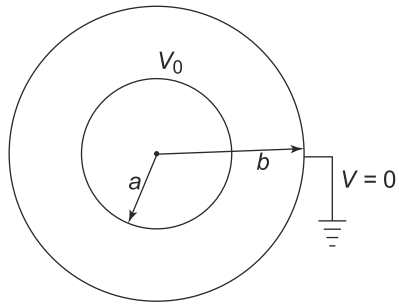


Fig. 3.19 *Problem 4*

5. A dielectric spread all over has variable permittivity given by $\varepsilon(r) = \varepsilon_0(1 + 0.1r)$, r being the distance from a certain point in the dielectric. A tiny spherically charged conductor carrying a charge 1 C is placed at $r = 0$. Calculate the electric field at all points.

6. Electrometer is a device to measure potential which uses the principle of the balancing of forces between the two plates of a parallel plate capacitor with an external weight. The guard plate is suspended from one of the arms of a balance and the force is balanced by a known weight mg to get the potential between the two plates of the capacitor. Show that the potential difference between the two plates, $V_1 - V_2$ is given by

$$V_1 - V_2 = \left[\frac{2mgd^2}{\varepsilon_0 A}\right]^{1/2}$$

where A is the area of the plates and d the distance between them. This is shown in Fig. 3.20

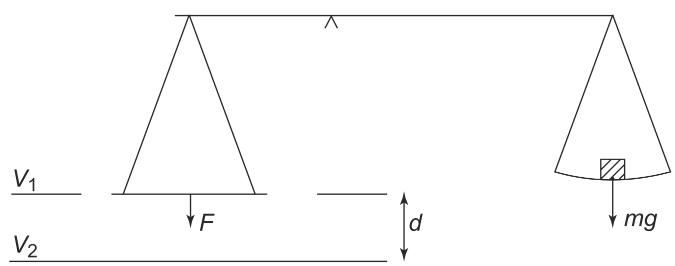


Fig. 3.20 *Problem 6*

7. Helium gas at some temperature and pressure contains 5×10^{25} atoms in a cubic meter. In an electric field $E = 10$ kV/m, the atoms in the gas suffer an average electron shift of 10^{-18}m. What is the dielectric constant of helium?

8. A conducting sphere of radius 10 cm is centred at the origin and embedded in a dielectric with $\varepsilon = 2.5\varepsilon_0$. If the surface charge density on the sphere is 4nC/m^2, find the electric field at the point (−3 cm, 4 cm,12 cm).

9. In a dielectric material, with $\varepsilon = 2.4\varepsilon_0$ and $V = 300z^2$ Volts, find

a. $\vec{D}$
b. $\vec{P}$
c. ρ and ρ_{pol}

10. Two homogenous dielectric regions, region I with ($\rho \leq 4$ cm) and region II with ($\rho \geq 4$ cm) have dielectric constants 3.5 and 1.5. If $\vec{D}_2 = 12\hat{\rho} - 6\hat{\phi} + 9\hat{z}$ nC/m^2, find $\vec{E}_1$ and $\vec{D}_1$.

11. A dielectric cylinder of radius a is polarised along its length. The cylinder is placed along the z axis from $z = -\frac{L}{2}$ to $z = \frac{L}{2}$. Assuming uniform polarisation, calculate the electric field due to the polarisation at a point on the z axis both inside and outside the cylinder.

12. Two large parallel metal sheets separated by a distance d have between them a dielectric which carries charge. The charge density in the dielectric is not uniform but given by $\rho(x) = \rho_0\left(1 + \frac{x}{d}\right)$ where x is the perpendicular distance from the plate at higher potential. One of the plates is at zero potential and the other at $+V$ volts. Solve the Poisson equation to determine the electric field at all points in the dielectric and outside the plates.

13. A dielectric sphere of radius R and of permittivity ε has free charges of uniform density ρ inside it. Calculate the electric field inside and outside the sphere. If the density instead of being uniform varied as $\rho(r) = \rho_0\left(1 + \frac{r}{R}\right)$, where r is the distance from the centre of the sphere, how would your answer change?

14. A dielectric disk of permittivity ε, radius R and thickness t has a uniform polarisation $\vec{P}$ parallel to its axis. Calculate the electric field on the axis of the disk at a distance d from the centre of the disk.

15. A sphere of radius R is made of a material of relative permittivity ε_r. It is polarised with a uniform polarisation $\vec{P}$ along the polar axis. Calculate the electric field at a point outside the sphere.

16. Two large metallic plates, each of area A are parallel and separated by a distance d. The plates are at potentials V_1 and V_2. A dielectric slab of the same area but of thickness $t < d$ sits atop the plate with the potential V_1 and the space above the dielectric slab is air. Find the electric field in the dielectric, in the air gap above the dielectric and the charge densities on the plates and on the surface of the dielectric.

17. A metallic spherical shell has an outer radius of 20 cm and inner radius of 18 cm. The region between the outer and inner surfaces is filled with a dielectric of relative permittivity ε_r. The dielectric is uniformly charged with a charge density ρ_0 and there is also a charged metallic sphere carrying a charge q inside, concentric with the shell. Calculate

a. the electric field inside, in the shell and outside and
b. the polarisation on the surfaces of the shell.

18. A horizontal parallel plate capacitor in circular shape with radius 1 m has a separation $d = 2$ cm between the plates. The upper plate carries a charge $Q = 0.1$ C and the lower plate $-Q$. A dielectric material of total volume 10^{-3} m^3 and of relative permittivity $\varepsilon_r = 1.5$ is present in the gap between the plates. The dielectric obviously does not fill up the entire space between the plates. Calculate the electric field

a. when the dielectric is in the form of a slab on the lower plate, on the upper plate and at a point just below the upper plate and just above the lower plate
b. when the dielectric is moulded in the form of a cylinder with its axis along the line joining the centre of the plates, at any point not in the dielectric.

4

Conductors and Capacitance

Learning Objectives

- To understand the behaviour of conductors in an electric field.
- To learn about the method of images and its use in finding the potential in the presence of conductors.
- To comprehend the relationship between potential of and charge on various conductors in a system of conductors and define the concept of capacitance.
- To learn about the capacitance in terms of the energy stored in electric fields.
- To be able to find the capacitance of simple geometrical arrangements of conductors.
- To be able to derive the resultant capacitance for a system of capacitors in various combinations.
- To be aware of different ways in which capacitors are constructed to maximise the capacity to store electric charge and energy.

4.1 INTRODUCTION

In the previous chapter, we studied one class of materials called dielectrics or insulators. These materials were characterised by the property that they do not have any free charges. In the presence of electric fields, they develop a polarisation, the strength of which for a given electric field is characterised by the dielectric constant. In this chapter, we turn to study another class of materials called conductors. **Conductors are materials that unlike dielectrics have at least one electron in every molecule essentially free to move within the material.** These electrons thus are effectively not attached to any molecule, unlike the situation in dielectrics or insulators as we saw in the previous chapter. We will study the electrical properties of conductors in this chapter.

Classically, there is no way to understand why there are some materials with free charges (electrons) and why some have no free charges. However, quantum theory allows us to have a reasonable explanation for this phenomenon. We know that in quantum theory electrons are bound to the nucleus in definite way. For instance, in Bohr's Theory of the atom, the electrons are envisaged to be in orbits which obey the quantisation rules for angular momentum. We know Bohr's Theory was not quite accurate in explaining the atomic phenomenon. Quantum mechanics replaced Bohr's Theory and is able to explain most of the phenomena that we see at the atomic and molecular level. In quantum mechanics, the electrons are in one of the many allowed stationary 'wave functions'. In either way of looking, one thing is common—in states with very low binding energy the electron is much farther away from its parent nucleus as compared with those in tightly bound states. This applies to an individual atom.

In an assembly of atoms, the situation is more complex. The atomic nuclei are separated by distances of the order of 1 Angstrom. The tightly bound core electrons move around with distances from their parent nucleus much smaller than this so that they do not feel the influences of charges from the neighbouring nuclei. Electrons with very small binding energies on the other hand, are to found at distances away from their parent nuclei which is comparable to the distances between neighbouring nuclei. Therefore, they feel the effect of more than one nuclei. The usual picture of Bohr orbits of electrons going around a single nucleus thus, is no longer valid for these electrons. In the language of quantum theory, the wave functions of such electrons cannot be thought to be localised around any single nucleus. Their wave functions are spread out over the entire macroscopic sample. The Swiss physicist Felix Bloch (1905–1983) first worked out the correct quantum mechanics of electrons in a 'lattice' of atoms. The wave function of electrons that results from such analysis are called 'Bloch waves', which are spread over all the nuclei in the sample. The classical picture corresponding to these is of having 'free electrons' in conductors, a model that was used extensively before the quantum basis was established. The existence of free electrons in conductors leads to many phenomena and in this chapter, we will study electrostatics in the presence of conductors just as the last chapter dealt with electrostatics with dielectrics.

We start by answering an obvious question: If some of the electrons in a conductor are free, why don't they come out of the conductor? The answer to this is based on the simple fact that though there are free electrons in a conductor, the conductor as a whole is electrically neutral. When a free electron moves within the bulk conducting material, it is surrounded on all sides by other nuclei. Thus, there is no 'net' force on it due to Coulomb attraction. But imagine such an electron approaching the surface of the material. The sample being on the whole neutral, the negative charge of the electron near the surface has necessarily a balancing 'net' positive charge in the sample, which is called the 'image charge'. As the free electron approaches the surface of the sample, the balancing image charge is no longer symmetrically situated with respect to the electron. It is located on the side that the sample is, and thus, exerts an electrostatic force on it pulling it back into the sample. The electron thus is unable to escape. This is the situation ordinarily in a conductor. However, by this logic, if the electrons within a sample were sufficiently energetic, they could escape the attraction of the image charges and come out of the sample. This is indeed true. The electrons could be energised by heating the sample or by the use of very strong fields. If, for instance, the sample was heated enough, some electrons will gain enough kinetic energy to overcome the restraining force and escape from the sample. This effect is called **thermionic emission** and was used extensively in vacuum tubes or thermionic valves which were devices used in electronics before the advent of semiconductors.

4.2 CONDUCTORS IN AN ELECTRIC FIELD

Dielectrics, in an external electric field experience a polarisation as we saw in the previous chapter. What is the effect of an electric field on conductors? Conductors, we know, have free charges and in addition they may be carrying any charge placed in them from outside. How do they interact with an external electric field?

To see this, consider a piece of a conducting material of arbitrary shape in the presence of an electric field. The conductor may be carrying charge in which case, the net electric field will be due to its own charge and the external electric field. The conductor has both free and bound electrons. The electrons that are bound to their parent nuclei will react exactly as in dielectrics forming dipole electric moments. The free electrons on the other hand, will start moving under the external electric field. When such a free electron reaches the surface, it will stay there since as we have just seen it cannot leave the surface under normal conditions. Can this process continue indefinitely? The answer is no, since as more and more free electrons reach the surface, they will form a layer of surface charge. This layer of surface charge will create an electric field of its own, which will act on all charged particles just like the external electric field. We can thus expect the migration of free electrons to stop when the surface charge created just balances the external electric field. **Thus, under equilibrium conditions for a steady electric field, there will be no 'net' electric field in the body of the conductor, the external field and the one created by the surface charge cancelling each other.** There being no electric field in the body of a conductor, we can conclude that in electrostatics, all points in a conductor are at the same potential.

What about the surface of a conductor? Here the situation is different. Clearly, the electrons accumulated there do not move out because of the balancing effect of the external electric field and the attraction of the net positive charge left behind in the conductor. The electric field at the surface is thus necessarily normal to the surface at every point. (Fig. 4.1(a)). The electric field can only have a normal component (normal to the surface) since any tangential component would cause the free electrons to move along the surface. This will continue till such time the surface charge distribution is such that the net electric field in the tangential direction is zero. **Thus, we see that under electrostatic conditions, the only electric field on the surface of a conductor is in the normal direction.**

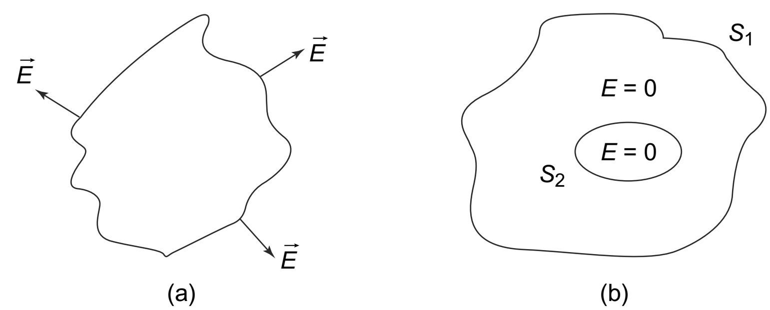


Fig. 4.1 *(a) A conductor in an electric field. The field everywhere is normal to the surface. The field in the body of the conductor is zero and the potential the same everywhere in it, (b) A conductor which has a hollow cavity inside. The cavity can have no electric field in it and is at the same potential as the rest of the body of the conductor*

Whatever we have reasoned out in the last two paragraphs continues to be true even if the conductor carries any free charge. The body of the conductor would continue to be all at the same potential. The free charge would create an electric field which would cause the free electrons in the body of the conductor to move. Just as in the case of external electric field, at equilibrium there would be no resultant electric field in the body of the conductor and net charge densities would be present at the surface only. The surface would be at the same potential as the body and thus would be an equipotential surface. The electric field at the surface would therefore be normal to the surface pointing outwards.

What about a hollow conductor like the one shown in Fig. 4.1(b) with an outer surface S_1 and a cavity inside enclosed by the inner surface S_2? There would not be any electric field in the body of the conductor as we have seen above. The charges in such a conductor must necessarily reside on the surfaces. The electric field lines starting from S_1 cannot penetrate inside into the conductor since there is no electric field there. That also means that the inner surface S_2 is at the same potential as S_1 since no work will be done if charge is transported from S_1 to S_2, there being no electric field in the intervening region. Also, there cannot be any surface charge on the inner surface S_2 which can be reasoned out as follows. If there are surface charges on S_2, then as argued, field lines starting from those surface charge cannot penetrate the body of the conductor and hence must go inside the cavity region as shown in Fig. 4.2.

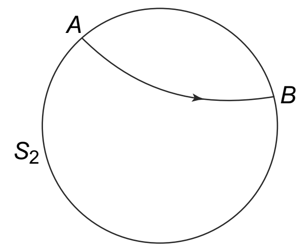


Fig. 4.2 *Cavity inside a conductor. If charges existed on the surface S_2, then field lines starting from some point A must pass through the cavity and end at another point on S_2, namely B*

Now consider any point A on the surface S_2 from where a field line starts and enters the cavity or where a field line from the cavity ends. There being no charges inside the cavity, the other end of the field line must end at some point B on the surface S_2 itself. But that would mean that the integral

$$\int_A^B \vec{E} \cdot \vec{dl} > 0$$

since along the field line A to B, $\vec{E}$ and $\vec{dl}$ are in the same direction. But

$$\int_A^B \vec{E} \cdot \vec{dl} = \phi(A) - \phi(B)$$

that is the integral is the difference in potential between the points A and B. But these points are at the same potential since the entire surface is an equipotential. Thus, the surface S_2 cannot carry any charge and consequently there is no electric field inside the cavity in a conductor and all points in the cavity are at the same potential as the conductor.

The fact that in a cavity inside a conductor, the potential is a constant is actually a trivial example of the Uniqueness Theorem of solutions of Laplace equation which we studied in Chapter 2, Advanced Topic. Recall that the Uniqueness Theorem states that two solutions of Laplace's equation that satisfy the **same** boundary conditions can differ at most by an additive constant. Clearly, $\phi =$ constant is a solution and if the constant equals the potential of the conductor, the boundary conditions are satisfied. If this is the case, the uniqueness theorem tells us that this is the only solution.

But what if the cavity contained some free charges? These charges in the cavity will of course, create a field inside the cavity. But, once again, the free electrons inside the conductor will move in a way to migrate to the surface. This would be in such a way such that the resultant field inside the body of the conductor, which will thus, be the vector sum of the field created by the free charges inside the cavity and the surface charges on the surfaces of the conductor and the cavity, would vanish.

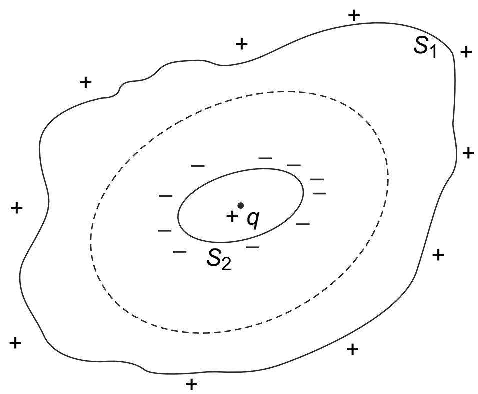


Fig. 4.3 *Charge $+q$ inside a cavity in a conductor. Induced surface charges will appear on the surfaces S_1 and S_2. In this case, S_2 will have negative charges and S_1 positive charges*

Let us summarise what we have learnt about the boundary conditions in the presence of a conductor, whether charged or not:

> A conductor is throughout at the same potential including its surfaces which therefore are equipotential surfaces. The surfaces may carry surface charges. The electric field at the surface is always normal to the surface at every point. In the presence of any hollow cavity in a conductor, there is no electric field in the cavity if there are no free charges in the cavity and the surface of the cavity carries no surface charge. If a hollow cavity contains free charges, there is an electric field in the cavity and there are surface charges in the cavity. The electric field at the surface of the cavity is normal to the surface at every point on the surface and the conductor throughout is at the same potential.

4.2.1 Surface Charges ∗

We have seen that for conductors, there exists a layer of surface charge, formed from the free charges in the conductor which accumulate there under the action of an external field. We have implicitly assumed that this layer of surface charge is a geometrical plane, i.e., it is a true two-dimensional surface with no thickness. This is an assumption which we need to justify.

On the face of it, it seems that this assumption of the surface charge being a geometrical plane cannot be true since a real conductor will obviously have irregularities of the size of atoms or over interatomic distances. These irregularities on the surface would be of roughly 1 Angstrom size. To get some idea of the thickness of the surface layer of the charge, we can use a model in which the free electrons move about as a gas while the positive ions left behind are fixed in position in their lattice sites (Fig. 4.4(a)). Consider copper as the material of the conductor and consider a cube of copper of size 1 m so that there

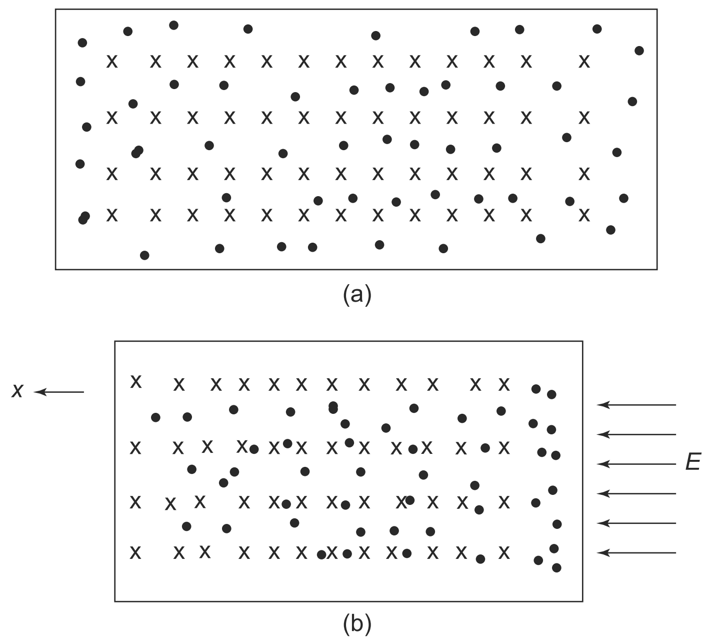


Fig. 4.4 *(a) The conductor is modeled here as comprising of positive ions (marked as X) in fixed positions with electrons marked as dots freely roaming inside all over, (b) In the presence of an electric field electrons are pushed towards the right. The rightmost layer of electrons see positive charges only towards the left. The force on these negatively charged electrons due to the electric field is balanced by the attractive force due to the positive ions on their left*

will be $N = 8.6 \times 10^{28}$ atoms and as many free electrons assuming each atom of copper contributes one free electron. If we imagine the atoms to be arranged in a cubic lattice with the atoms separated from their nearest neighbour by a distance d (which is also of the order of an Angstrom.), then there will be $1/d$ sheets of positive ions each having a surface density of charge eNd. If we neglect thermal motion of the free electrons, the free electron charge density would be homogeneous in the absence of an external field.

Now imagine an external uniform, electric field $\vec{E}$ perpendicular to the surface and along the x-direction (Fig. 4.4(b)). The free electron cloud would move along the $-x$ direction leaving bare, positively charged layers at the farther end of the conductor. This would ultimately tend to an equilibrium situation wherein the electric field created by the bare positively charged ion layer at one end and the layer of free electrons that have moved to the other end annuls the external electric field in the body of the cube. To see if this is the case, we need to have an estimate of the field created by the layers of charge. This can be done easily in the following way.

Consider for example, a situation in which the free electron has moved out such that only one mono-atomic layer of possible ions is created at the opposite end. This positively charged sheet of charge would create an electric field of magnitude $(4\pi k)eNd$ opposite to the direction of $\vec{E}$. For $d = 1$ Angstrom, if we equate the this field to the external field, we get an electric field of the order of 10^{11} volts/meter. Thus, for ordinary values of electric fields in general use, the net displacement of the free electron cloud beyond the ions would be much smaller. Of course, as we have mentioned in the beginning of this chapter, the fields that we are talking about are fields averaged over several atoms, which is of the order of a few Angstroms. Thus, the thickness of the surface charges, i.e., the distance by which the free electron cloud spills over the body and thus also the thickness of the layer of positive charges at the opposite end, are limited to a few interatomic distances.

The detailed calculation of such fields of course has to be quantum mechanical in nature and involves working out the minimum energy configuration taking into account the external potential created by the electric field, thermal motions, inter-electron interactions and the interaction of the electrons with the background electric field created by the positive ion lattice. These calculations have been done (**W. Kohn and L. J. Sham Phys. Rev., 140, A1173(1965); N. D. Lang and W. Kohn, *Phys Rev B7*, 3541(1973)**) and the results confirm our general result above.

One can also ask another question about the nature of the surface charge on a conductor. We have already seen that there is no electric field inside a conductor. Suppose we place a conductor in an external field. We saw above that the free charges move to create a surface layer of charge which ensures that the body of the conductor has no field. What if we now withdraw the external electric field? Do the surface charges now move instantaneously? If not, what is the order of magnitude of the time taken for the the surface charges to move in response to this sudden withdrawal of the external field? This is obviously strictly speaking, not a question of electrostatics since by definition, charges will be in motion. But it is instructive to try and understand whether, on the sudden withdrawal of the external electric field, the surface charges which exist on the surface of a conductor (as we saw above) vanish instantaneously or whether there is a time delay.

As we saw above, the free electrons inside the conductor in the presence of an external electric field are not distributed homogenously. Obviously, when the field is withdrawn the free electron density becomes homogenous as it should be in a conductor. The movement of electrons constitutes an electric current, as we shall see in the following chapters. The current has a current density, a vector quantity denoted usually by $\vec{j}$. Any inhomogeneity in the free electron charge distribution (and hence in the total charge distribution in the conductor) $\delta(r, t)$ creates an electric field $\delta E(r)$ such that, by Gauss's Law we have

$$\vec{\nabla} \cdot \delta\vec{E}(r) = 4\pi k\delta(r)$$

On the other hand, the field δE creates a current density $j(r)$, such that by Ohm's Law (which we shall discuss in the following chapters)

$$\sigma\delta\vec{E}(r, t) = \vec{j}(r, t)$$

where σ is called the conductivity of the material. Taking the divergence of this equation, and using the equation for divergence of $\vec{E}$ above, we get

$$\sigma 4\pi k\delta(r, t) = \vec{\nabla} \cdot \vec{j}(r, t)$$

Now consider an infinitesimal volume δV of the conductor. The total outflow of charge from this volume is obviously related to the current density (since the current is defined as the rate of flow of charge) and is given by $\vec{\nabla} \cdot \vec{j}(r,t)\delta V$. But by charge conservation, this is also the rate of change of charge in that volume with a negative sign, that is, $-\delta V \frac{d\delta(r,t)}{dt}$. Equating the two we get

$$4\pi k\sigma\delta(r,t) = \vec{\nabla} \cdot \vec{j}(r,t) = -\frac{d\delta(r,t)}{dt}$$

This equation can be easily solved and we get

$$\delta(r,t) = e^{-4\pi k\sigma t}\delta(r,0)$$

This solution for the time dependence of the inhomogeneity in the charge distribution $\delta(r,t)$ tell us that the charge distribution is smoothed out with a characteristic time scale of $\frac{1}{4\pi k\sigma}$. For typical values of σ, we get the time scale to be extremely short, of the order of 10^{-16} seconds. Thus, we can, for all practical purposes think of the change in the surface charge density when the external field goes to zero, as instantaneous.

To summarise, it is a reasonably correct assumption that in electrostatics, the surface charge density is at a surface in the geometrical sense and also that the creation and destruction of such charges in the surface in response to external electric fields is instantaneous.

4.2.2 Surface Charges and Curvature

We now know that the surface of a conductor, in the presence of an electric field, develops a surface charge density which is created almost instantaneously. How is this charge density distributed on the surface? Is it uniform? These are obviously important questions because the surface charge density would have a contribution to the field external to the conductor.

For a conducting sphere of radius R carrying a total charge $+q$ which migrates to its surface, the answer is clear. By symmetry, the charge will be uniformly distributed over its surface resulting in a surface charge density $\frac{q}{4\pi R^2}$. In general however, the surface charge density is not uniform and to get some general results about it, we first work out an important connection between surface charge and the electric field just outside a conductor.

Consider an infinitesimal portion $ABCD$ of area δa of the surface of a conductor, which can be considered planar. We draw an infinitesimally thin pill box of flat faces parallel to $ABCD$. One of the faces is inside the conductor and the other outside it (Fig. 4.5). Since there is no electric field inside the conductor, there is no flux through the face inside. The flux through the sides is infinitesimal. The field just outside ABCD is normal to the surface and hence normal to the face of the pillbox which is outside. The flux outwards through it then is $E\delta a$, where E is the field created by the surface charge just outside it. Hence by Gauss's Theorem:

$$E\delta a = 4\pi k(\sigma\delta a)$$

where σ is the surface charge density. Thus,

$$E = 4\pi k\sigma \tag{4.1}$$

For the spherical case, this gives us

$$E = 4\pi k\sigma = 4\pi k\left(\frac{q}{4\pi R^2}\right) = \frac{kq}{R^2}$$

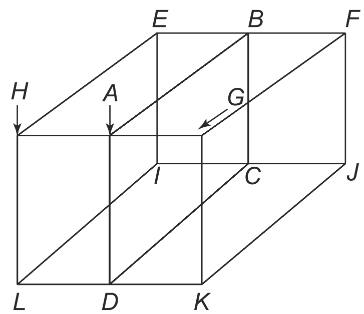


Fig. 4.5 *ABCD is an infinitesimal portion of a conductor to its left. Its area is δa.EFGHIJKL is a pill box with EF, IJ, HG and LK infinitesimal. The face EILH is inside the conductor and the face GFJK outside, both parallel to ABCD shifted by infinitesimal amount to its left and right*

which is just the field that would be created if the conducting sphere were not there. Let us now use the relationship between surface charge density and electric field in more general cases.

In general, let us consider two neighbouring equipotential surfaces that have slightly different potentials ϕ_1 and ϕ_2 and one of the surfaces enclosing the other. (Fig. 4.6). The work done in carrying a unit charge along two paths shown BA and CD of lengths l_A, l_D respectively will be product of these lengths with the fields E_A and E_D at the points A and D respectively. The work done along either paths will be $d\phi$. Hence,

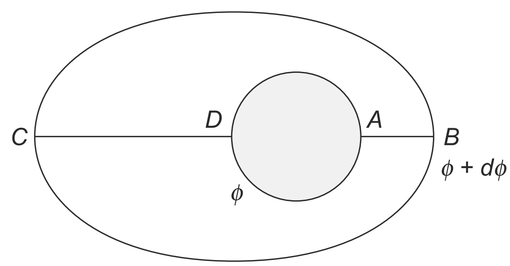


Fig. 4.6 *Two equipotential surfaces: the one at ϕ is totally enclosed by the one at $\phi + d\phi$. AB and CD are two lines joining the two surfaces of lengths l_A and l_D respectively. If E_A and E_D are the electric fields at A and D respectively, the work done in carrying a unit charge from the inner to outer surface along the paths BA and CD are $(E_A l_A)$ and $(E_D l_D)$ respectively. They must be equal and equal to $d\phi$*

$$E_A l_A = E_D l_D \tag{4.2}$$

Now since l_A is much smaller than l_D, it follows that $E_A > E_D$. If the equipotential surfaces were actual conductors we can easily relate the two fields to the surface charge densities at the two points from Eq. (4.1) and hence, we have

$$\sigma_A >> \sigma_D \tag{4.3}$$

We thus, have a general result: If we consider neighbouring equipotential surfaces, electric fields at points on the equipotential surfaces separated by smaller normal distances have bigger electric fields than at points which have a larger normal separation between them.

Consider now a pear shaped conductor which carries a charge q (Fig. 4.7). As shown above, the conductor will be at the same potential throughout. The surface will be an equipotential surface and the electric field at the surface will be normal to the surface at every point on it.

Very close to the surface of the conductor, an equipotential surface will be similar to the surface of the conductor. However at distances much larger than the size of the conductor, the equipotential surface will be almost spherical with the conductor at its centre since all charges in the conductor will be almost equidistant from such a spherical surface. Thus, as we advance from very large distances to regions close to the conductor, the equipotential surfaces changes from spherical shape to the shape of the conducting surface as shown in Fig. 4.7. Clearly, the separation between the equipotential surfaces is much smaller at the tipping end of the pear shaped conductor as compared with the more rounded end. From what we have just established in the last para, the electric field at the sharper end of the conductor will be much larger than the value at the flatter end. The relationship, Eq. (4.3), between curvature and surface charge density hence, tells us: **On a surface of a conductor carrying charge which resides necessarily on its surface, the surface charge density is higher at points of larger curvature compared to points of lesser curvature.** Thus, if the conductor has a sharp point, the field outside it will be very intense, sometimes sufficiently large to cause ionisation of the molecules of air near it.

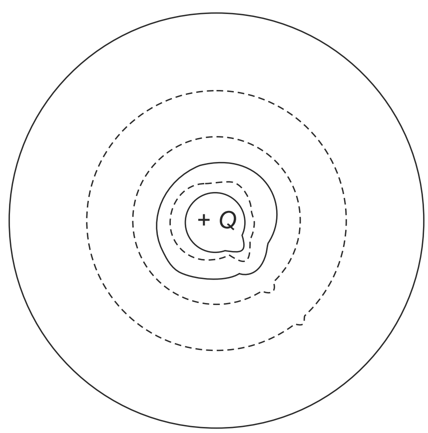


Fig. 4.7 *A pear-shaped conductor carrying charge. The equipotential surfaces at large distances from the conductor are spherical and as we go in towards the conductor, they start resembling the surface of the conductor becoming almost exactly like its surface very close to it. The equipotential surfaces are more closely packed at the sharper end of the conductor than at the flatter end*

What we have seen above is a general statement. Is it possible to have an exact mathematical relationship between the curvature of the surface of the conductor and the surface charge density on it? It turns out that the exact relation between surface charge density and the curvature cannot be put in a simple mathematical form. In general, the problem amounts to solving Laplace's equation outside the conductor with the boundary condition that the potential on the surface of the conductor is the same everywhere and the gradient of the potential (which is the negative of the electric field) is normal to the surface on the conductor. Once the potential is known, the normal electric field and hence the surface charge density is known. We were able to do it easily for a spherical surface since the surface of the sphere corresponds to one of the spherical coordinates that we used to solve the Laplace equation. For a slight deviation from the exact spherical shape, to say a spheroidal shape given by

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{b^2} = 1 \tag{4.4}$$

it is still possible to do this, though much more complicated. If one solves Laplace's equation in spheroidal coordinates, one gets

$$\sigma(x, y, z) = \frac{q}{4\pi ab^2}\left(\frac{x^2}{a^4} + \frac{y^2 + z^2}{b^4}\right)^{-1/2} \tag{4.5}$$

which reduces to the spherical result of $\sigma = \dfrac{q}{4\pi a^2}$ for a spherical surface in which $a = b$. The charge distribution in Eq. (4.5) is not uniform. Thus, if we take $a \gg b$, at the sharper ends $x^2 = a^2$, $y = z = 0$ and $\sigma = \frac{q}{4\pi b^2}$. This is much larger than the density at the flat middle portion: $x = 0$, $y^2 + z^2 = b^2$ where $\sigma = \frac{q}{4\pi ab}$. We will not derive this result here which can be found in various texts. (See for example, **'Electrodynamics of Continuous Media', L.D. Landau, E.M. Lifshitz and L.P. Pitaveskii, 2005; Kun Mu Lui, American Journal of Physics, 55, 849 (1987)**).

We now work out a few examples of solving Laplace equation when conductors are present.

EXAMPLE 4.1 A uniform electric field $\vec{E} = E\hat{z}$ exists in the z-direction throughout space. A conducting sphere of radius R with its centre at the origin, is placed in the electric field. Calculate the electric field outside the conductor.

Solution

This problem is similar to Example 3.5 considered in the previous Chapter but differs from it regarding the boundary conditions at $r = R$.

For $r > R$, we have to solve the Laplace equation. The boundary conditions at $r = \infty$ is the same as in Example 3.5, so that we can use Eq. (3.46), for the potential in the present example as it is. Thus:

$$\phi(r, \theta) = -Er\cos\theta + A - \sum_{l=0}^{\infty}(C_l r^{-l-1} P_l(z)) \quad (r > R) \tag{4.6}$$

We have included a constant A that can be set by the choice of the value of the potential at ∞. We had taken this as zero in Eq. (3.46) by choosing $\phi(r)$ at ∞ to be $-Er\cos\theta$.

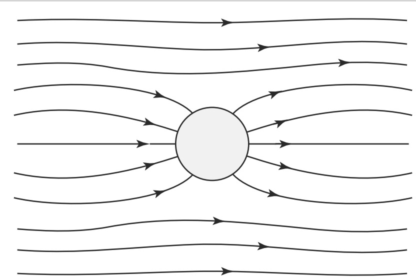


Fig. 4.8 *A conducting sphere placed in an uniform electric field. The conductor will change the electric field everywhere in such a way that the surface of the sphere becomes an equipotential surface*

There is one more constraint that exists in the problem because of symmetry. The electric field before the sphere was introduced was odd under the transformation $\theta \to \pi - \theta$ i.e $E(r, \theta) = -E(r, \pi - \theta)$.

The sphere was completely symmetric and thus, this property will be retained even after the introduction of the sphere. In particular, putting $\theta = \frac{\pi}{2}$, we have

$$E\left(r, \frac{\pi}{2}\right) = -E\left(r, \pi - \frac{\pi}{2}\right) = -E\left(r, \frac{\pi}{2}\right) = 0 \tag{4.7}$$

Thus, along the direction $\theta = \frac{\pi}{2}$ there is no electric field. This means that $\left(\frac{\partial\phi(r,\theta)}{\partial r}\right)_{\theta=\frac{\pi}{2}} = 0$. Thus, we have a boundary condition that

$$\phi(r, \pi/2) = \text{constant} = \phi_0 \quad (\text{for all } r) \tag{4.8}$$

In the present case the boundary condition at $r = R$ is far simpler than in Example 3.5. At $r = R$, the potential for all values of θ must be the same, since $r = R$ is the surface of a conductor and hence an equipotential surface. Let us call the potential of the conductor ϕ_0.

We also know that the $P_l(z)$ are orthogonal (Eq. (3.50)), which implies that in Eq. (4.6), the coefficients of $P_l(z)$ must match term by term on both sides. Applying this property to Eq. (3.50), Chapter 3 at $r = R$ and using the fact that

$$P_0(z) = 1$$
$$P_1(z) = z$$

we get

$$\begin{aligned} \phi(R, \theta) = \phi_0 &= A - C_0/R \\ 0 &= -ER - C_1/R^2 \\ 0 &= C_l R^{-l-1} \qquad\qquad (l > 1) \end{aligned}$$

Thus, we get

$$\phi(r, \theta) = \phi_0 + C_0\left(\frac{1}{R} - \frac{1}{r}\right) - \left(r - \frac{R^3}{r^2}\right) E\cos\theta \tag{4.9}$$

Now using Eq. (4.8), that is putting $\theta = \frac{\pi}{2}$, we get

$$C_0 = 0$$

Therefore,

$$\phi(r, \theta) = \phi_0 - \left(r - \frac{R^3}{r^2}\right) E\cos\theta \tag{4.10}$$

With this potential, we can easily find the the tangential and normal (radial) components of the electric field at the surface $r = R$. The tangential component

$$E_\theta = -\frac{1}{r}\frac{\partial\phi}{\partial\theta} = 0 \qquad\qquad (\text{ at } r = R)$$

and the radial component

$$E_r = 3E\cos\theta \tag{4.11}$$

Knowing the electric field at the surface, we can easily compute the surface charge density using the relationship between the surface charge density and the electric field near the conducting surface,

$$\sigma(r = R, \theta) = \frac{E}{4\pi k} = \frac{3}{4\pi k} E \cos\theta \tag{4.12}$$

Note the interesting fact- the top half of the conducting sphere$(0 < \theta < \frac{\pi}{2})$ is positively charged while the bottom half $(\frac{\pi}{2} < 0 < \pi)$ is negatively charged.

EXAMPLE 4.2 An infinite metallic cylinder of cross-sectional radius R is placed in a uniform electric field E such that the the electric field is at perpendicular to the axis of the cylinder. Calculate the electric field at all points after the cylinder is placed. This is depicted in Fig. 4.9.

Solution

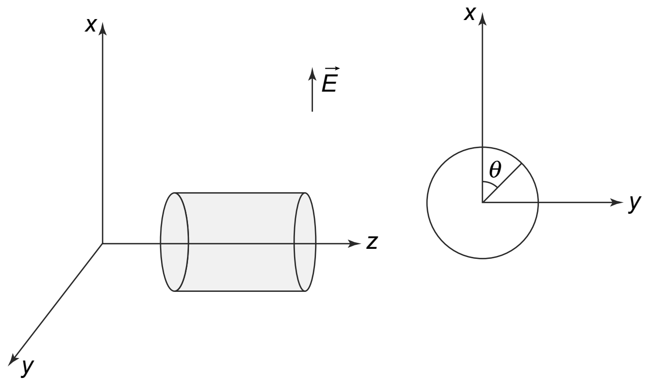


Fig. 4.9 *A metallic cylinder with its axis perpendicular to an external uniform electric field. We use two-dimensional polar coordinates r, θ in the $x - y$ plane*

Let the axis of the cylinder be the z-axis and the electric field be in the x-direction. Since the cylinder is infinite all values of z are similar in all respects and hence, the potential ϕ is a function of x and y only that is $\phi(x, y)$. The symmetry of the problem further dictates that instead of using Cartesian coordinates x, y, we use two-dimensional plane polar coordinates r, θ in the $x - y$ plane. The Laplace equation in these coordinates is given by

$$\nabla^2\phi = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial\phi}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2\phi}{\partial\theta^2} = 0 \tag{4.13}$$

with the boundary conditions

$$\phi(R, \theta) = \phi_0 \qquad (\phi_0 = \text{constant}) \tag{4.14}$$

and

$$\frac{\partial\phi}{\partial\theta} = 0 \qquad (\text{at } r = R) \tag{4.15}$$

The first of these boundary conditions is defining the potential on the surface of the cylinder (at $r = R$) and the second one is ensuring that the electric field at the surface is normal to the surface, i.e., only in the direction of $\hat{r}$ and has no component in the tangential ($\hat{\theta}$) direction. As discussed above, we use the separation of variables technique to solve this partial differential equation. We write

$$\phi(r,\theta) = F(r)G(\theta) \tag{4.16}$$

Substituting this in Eq. (4.13), multiplying by r^2 and dividing throughout by $\phi(r,\theta)$, we get

$$\frac{r}{F(r)}\frac{d}{dr}\left(r\frac{dF}{dr}\right) + \frac{1}{G(\theta)}\frac{d^2G(\theta)}{d\theta^2} = 0 \tag{4.17}$$

The first term is independent of θ and the second term is independent of r and hence they both must individually be equal to a constant. Thus, we have

$$\frac{r}{F(r)}\frac{d}{dr}\left(r\frac{dF}{dr}\right) = m^2 \tag{4.18}$$

and

$$\frac{1}{G(\theta)}\frac{d^2G(\theta)}{d\theta^2} = -m^2 \tag{4.19}$$

where m is a constant. The θ equation is easily solved and we get

$$G(\theta) = C_1 \cos m\theta + C_2 \sin m\theta \tag{4.20}$$

where C_1 and C_2 are constants. Now we know that $G(\theta)$ is a periodic function of θ with a period 2π (because the points $\theta = 0$ and $\theta = 2\pi$ are the same on the cylinder), the constant m must be either integer or zero.

The r equation, Eq. (4.18) can be written as

$$\frac{d^2F(r)}{dr^2} + \frac{1}{r}\frac{dF(r)}{dr} - \frac{m^2}{r^2}F(r) = 0 \tag{4.21}$$

This is a second order differential equation with coefficients which are singular at $r = 0$. The general solution therefore, is

$$F(r) = r^m \qquad \text{and} \qquad r^{-m} \tag{4.22}$$

for $m = 0$, $F(r) = \ln r$.

Knowing the solutions for $F(r)$ and $G(\theta)$ allows us to write the most general solution of the Laplace equation in this situation

$$\phi(r,\theta) = \sum_{m=0,1,2,\cdots} f \ln r + \left[a_m r^m \cos m\theta + b_m r^{-m} \cos m\theta + c_m r^m \sin m\theta + d_m r^{-m} \sin m\theta\right] \tag{4.23}$$

where $f a_m$, b_m, c_m and d_m are constants which will be determined by the boundary conditions. Let us first take the boundary condition at $r = \infty$. We know that

$$\phi(r,\theta) = -Er\cos\theta \qquad (r \to \infty)$$

since the external electric field is taken to be in the x-direction and at infinity, this would remain unchanged by the cylinder. Therefore, we get

$$c_m = 0, \qquad (\text{ for all } m)$$

and

$$a_1 = -E, \qquad a_m = 0, \text{ for all } m \neq 1 \qquad \text{and } f = 0$$

Next, we know that at $r = R$, the potential is a constant ϕ_0 for all θ. Thus, we have

$$b_1 = a_1 R^3 = ER^3$$

and

$$b_0 = \phi_0$$

with

$$b_m = 0 \qquad \text{for all } m \neq 0, 1$$

and

$$d_m = 0 \qquad \text{for all } m$$

Thus, we get

$$\phi(r, \theta) = -Er\cos\theta\left(1 - \frac{R^2}{r^2}\right) + \phi_0 \tag{4.24}$$

We can also use the symmetry argument now. Before the introduction of the cylinder

$$\phi(r, \theta) = -\phi(r, \pi - \theta)$$

Since the cylinder was uncharged, this symmetry won't change with its introduction. Thus, at $\theta = \frac{\pi}{2}$, we must have

$$\phi\left(r, \frac{\pi}{2}\right) = -\phi\left(r, \frac{\pi}{2}\right) = 0 \tag{4.25}$$

Using this in Eq. (4.24) obviously implies $\phi_0 = 0$ and we get

$$\phi(r, \theta) = -Er\cos\theta\left(1 - \frac{R^2}{r^2}\right) \tag{4.26}$$

The radial field at $r = R$ is given by

$$E_r(R, \theta) = -\left(\frac{\partial\phi(r, \theta)}{\partial r}\right)_{r=R} = 2E\cos\theta \tag{4.27}$$

Using this we can also determine the surface charge density on the surface of the conductor as

$$\sigma(r, \theta) = \frac{E}{4\pi k} = -\frac{2E}{4\pi k}\cos\theta \tag{4.28}$$

We have seen above an example of solving the Laplace equation in spherical polar and cylindrical coordinates. We now illustrate the use of Cartesian coordinates for solving Laplace's equation. This example is particularly relevant since it pertains to a geometry which is very widely used in devices called parallel plate capacitors, which we will study later in the Chapter.

EXAMPLE 4.3 Two infinite flat metal plates carrying charges q_1 and q_2 per unit area are kept parallel to each other at a perpendicular distance d, as displayed in Fig. 4.10. Calculate the potential at all points.

Solution

Let the plates be kept at right angles to the z-direction, and let the origin be on the top plate as shown in Fig. 4.10. The plates are infinite along the $x - y$ plane and hence by symmetry the potential ϕ is a function of z alone, i.e., $\phi(x, y, z) = \phi(z)$.

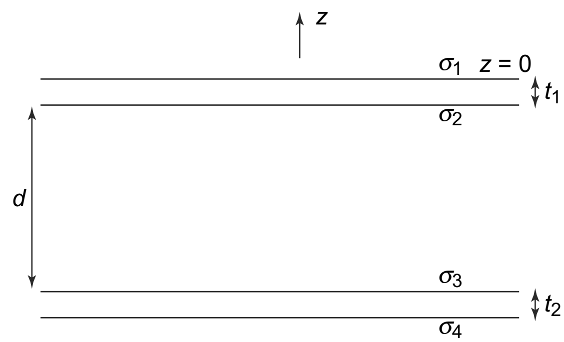


Fig. 4.10 *Parallel conducting plates kept at a distance d between them. The plates are of thickness* t_1, t_2

Let us first consider the potential, ϕ_1 arising out of the topmost layer of charge at $z = 0$. The charge density is $\sigma_1\delta(z)$ since the surface of the plate is confined to the plane $z = 0$ and thus, potential ϕ_1 due to this will satisfy Poisson's equation

$$\frac{d^2\phi_1(z)}{dz^2} = (-4\pi k)\sigma_1\delta(z) \tag{4.29}$$

Note that in general, if ϕ is a solution to the Poisson's equation

$$\nabla^2\phi = -4\pi k\rho$$

then $\phi + \phi_2$ is also a solution, where ϕ_2 is a solution to the Laplace equation

$$\nabla^2\phi_2 = 0$$

Thus, for example, for Eq. (4.29) above

$$\frac{d^2(\phi_1(z) + \phi_2(z))}{dz^2} = \frac{d^2\phi_1(z)}{dz^2} = (-4\pi k)\sigma_1\delta(z) \tag{4.30}$$

However, when one is trying to find the potentials (and hence fields) due to a given charge distribution, our solution must obviously vanish when the charge goes to zero. Hence in such cases, we don't need to add the solution of the Laplace equation to the original solution.

We know that there are no external fields present. Hence, the solution to this equation which will be suitable will be one which vanishes if $\sigma_1 = 0$. Furthermore, the Eq. (4.29) is symmetric with respect to z and hence the solution must respect this symmetry. It is easy to check that the solution with these properties is

$$\phi_1(z) = (-4\pi k)\frac{\sigma_1 z\varepsilon(z)}{2} \tag{4.31}$$

where $\varepsilon(z)$ is the sign function which has the following properties: It is $+1$ for z positive, and -1 for z negative. Further, its derivative is simply $2\delta(z)$ where $\delta(z)$ is the Dirac Delta function. That this function satisfies the Poisson Eq. (Eq. (4.30) and has the required properties of symmetry, etc., can be easily checked.

The problem above has three more surfaces the solutions to which will be similar to the one we have just found. The three surfaces are at $z = -t_1$, $z = -t_1 - d$ and $z = -t_1 - d - t_2$. Their charge densities are σ_2, σ_3 and σ_4 respectively. Thus, the surface charge densities are given by

$$\sigma_2\delta(z + t_1),\ \sigma_3\delta(z + t_1 + d),\ \sigma_4\delta(z + t_1 + d + t_2)$$

Since these are of the same form as the charge density considered for the topmost surface at $z = 0$, we can write the solution to the Poisson's equation for each of them in a similar fashion. The most general solution to the Poisson equation then, will be a superposition of these solutions and is hence, given by

$$\begin{aligned}\phi(z) &= -\tfrac{1}{2}(4\pi k)\,[\sigma_1 z\varepsilon(z) + \sigma_2(z + t_1)\varepsilon(z + t_1) + \sigma_3(z + t_1 + d)\varepsilon(z + t_1 + d) \\ &\qquad + \sigma_4(z + t_1 + d + t_2)\varepsilon(z + t_1 + d + t_2)]\end{aligned} \tag{4.32}$$

We define,

$$\sigma_1 + \sigma_2 = q_1$$

and

$$\sigma_3 + \sigma_4 = q_2$$

which are the charge densities on the top and bottom plates respectively.

Next, we need to impose the boundary conditions. First, we know that in between the two surfaces of each of the plates, the potential is a constant since the plates themselves are conducting. This means that the potential cannot be a function of z. Also, inside the top plate, $z < 0$ but $z + t_1$, $z + t_1 + d$, $z + t_1 + d + t_2$ are all positive. Collecting the terms proportional to z in the potential, we get

$$\begin{aligned}-\sigma_1 &= \sigma_2 + \sigma_3 + \sigma_4 = \sigma_2 - q_2 \\ -\sigma_1 - \sigma_2 - \sigma_3 &= \sigma_4 = -q_1 - \sigma_3\end{aligned} \tag{4.33}$$

Thus, we have

$$\begin{aligned}\sigma_1 &= \frac{1}{2}(q_1 + q_2) \\ \sigma_2 &= \frac{1}{2}(q_1 - q_2) \\ \sigma_3 &= \frac{1}{2}(q_2 - q_1) \\ \sigma_4 &= \frac{1}{2}(q_2 + q_1)\end{aligned} \tag{4.34}$$

Since we know the general solution, we can now evaluate the potential in the various regions. Let us start from the region above the top plate. Here $z > 0$ and all terms inside the square bracket in Eq. (4.32) are positive since ε function has positive argument and we get

$$\phi(z) = -\frac{4\pi k}{2}\,(\sigma_1 z + \sigma_2(z + t_1) + \sigma_3(z + t_1 + d) + \sigma_4(z + t_1 + d + t_2)) \tag{4.35}$$

For the region in between the two surfaces of the top plate, the first term is negative ($z < 0$) though all other terms are positive and we get

$$\phi(z) = -\frac{4\pi k}{2}\left(-\sigma_1 z + \sigma_2(z+t_1) + \sigma_3(z+t_1+d) + \sigma_4(z+t_1+d+t_2)\right)$$

$$\phi(z) = -\frac{4\pi k}{2}\left(+(t_1+t_2)(q_1+q_2) + q_2 d\right) \tag{4.36}$$

where we have substituted the expressions for the σ's from Eq. (4.34).

In the empty space between the two plates, the first two terms are negative since $z < 0$ and also $(z+t_1) < 0$ and so we get

$$\phi(z) = -\frac{4\pi k}{2}\left(-\sigma_1 z - \sigma_2(z+t_1) + \sigma_3(z+t_1+d) + \sigma_4(z+t_1+d+t_2)\right)$$

$$\phi(z) = -\frac{4\pi k}{2}\left(-z(q_1-q_2) + (t_1+t_2)(q_1+q_2)/2 + t_1(q_1-q_2) + q_2 d\right) \tag{4.37}$$

In between the two surfaces of the lower plate, the first three terms are negative and we get

$$\phi(z) = -\frac{4\pi k}{2}\left(-\sigma_1 z - \sigma_2(z+t_1) - \sigma_3(z+t_1+d) + \sigma_4(z+t_1+d+t_2)\right)$$

$$\phi(z) = -\frac{4\pi k}{2}\left(+(t_1+t_2)\frac{(q_1+q_2)}{2} + q_2 d\right) \tag{4.38}$$

and finally in the region below the second plate, all terms are negative and we get

$$\phi(z) = -\frac{4\pi k}{2}\left(-\sigma_1 z - \sigma_2(z+t_1) + \sigma_3(z+t_1+d) - \sigma_4(z+t_1+d+t_2)\right)$$

$$\phi(z) = -\frac{4\pi k}{2}\left(-z(q_1+q_2) - (t_1+t_2)\frac{(q_1+q_2)}{2} + q_2 d\right) \tag{4.39}$$

Now we have the potentials in all regions and so it is easy to find the electric field. First of all, note that the electric field $-\frac{\partial\phi}{\partial z}$ vanishes between the two surfaces of the upper and lower plates since these are conductors and there is no electric field within a conductor. Above the top plate, taking $-\frac{\partial\phi}{\partial z}$, we get the the electric field

$$\frac{4\pi k}{2}(q_1+q_2)$$

Below the lower plate, repeating the same procedure with the relevant potential, the value is

$$-\frac{4\pi k}{2}(q_1+q_2)$$

In the empty space between the two plates, the electric field is given by

$$\frac{4\pi k}{2}(q_2-q_1)$$

An interesting fact emerges from these expressions of the field everywhere. **Consider a special case of this, the situation in which $q_1 = -q_2 = q$. Then, from the above expressions, we notice that the electric field is zero above the top plate and below the bottom plate and is only restricted to the region between the two plates where it is simply $4\pi kq$. As we shall see later in the chapter, this is the widely used example of the parallel plate capacitor where two conducting plates are**

placed parallel to each other and carry equal and opposite charges. The field then is confined to the region between the plates.

We shall come back to this later.

PROBLEM 4.1 Two grounded plane electrodes of width a are placed parallel to each other at $y = 0$ and $y = b$ in the $x - z$ plane. The plane at $x = 0$ is kept at potential V_1 while that at $x = a$ is kept at V_2. Find the potential everywhere in between the planes.

PROBLEM 4.2 Two conducting spheres of radii a and b, $b > a$ are kept at potentials V_1 and V_2. Find the potential at all points between the spheres.

PROBLEM 4.3 Repeat Problem 4.2 for the case of two cylinders with the outer cylinder at $V = 0$.

4.3 METHOD OF IMAGES

If we have a charge distribution, $\rho(\vec{r})$ in a certain region and nothing else between the region and infinity, then the potential due to the charge distribution is given by the solution to the Poisson equation

$$\nabla^2\phi(\vec{r}) = -4\pi k\rho(\vec{r}) \tag{4.40}$$

with the boundary condition that ϕ goes to zero as $r \to \infty$. However, if there are conductors present then the solution changes. This would be expected since the presence of the conductor in the region will lead to the presence of induced (surface) charges. Thus, in the region that excludes the conductor, the potential still satisfies Eq. (4.40) but $\phi(\vec{r})$ will need to satisfy additional boundary conditions, namely that it should be constant on the surface of the conductor. In general, the solution will be different from the one where no conductors are present. An analytical solution to this problem is not possible in general. However, in specific cases, a particular trick, first introduced by Lord Kelvin works. In this, one is able to express the difference between the two potentials (with and without the conductor) as the potential created by ‘image’ charges.

Consider a region or regions in space where there are NO conductors. If the conductors are all finite then obviously there is only one such connected region outside all conductors which we will call as Region I. For infinite conductors like an infinite plane sheet, there may be more than one such region. In Region I, the solution to Eq. (4.40) can be written as:

$$\phi(\vec{r}) = \phi_f(\vec{r}) + \phi_c(\vec{r}) \tag{4.41}$$

where $\phi_f(\vec{r})$ is the solution to Eq. (4.40) in the absence of conductors and $\phi_c(\vec{r})$ is the potential created by the charges on the conductors. $\phi(r)$ should satisfy the boundary condition that it should be equipotential on every conductor present at the interface of Region 1 and conductors. Now since both $\phi(\vec{r})$ and $\phi_f(\vec{r})$ satisfy Eq. (4.40), clearly $\phi_c(\vec{r})$ satisfies

$$\nabla^2\phi_c(\vec{r}) = 0 \tag{4.42}$$

in Region I. Now consider any charge present **outside** Region I. It will create a potential in Region I which will automatically satisfy Eq. (4.42). The Method of Images is basically identifying a charge or charges outside of Region I, which have the property that the resulting $\phi(\vec{r})$ satisfies the required boundary conditions. By the uniqueness theorem, this resulting $\phi(r)$ is also a solution of Laplace equation in Region 1 in the presence of conductors at interfaces. These charge or charges outside of Region I are called 'image' charge(s). It should be clear that there are no physical image charges but only a mathematical construct to solve a problem with some given configuration of charge densities and conductors, etc.

To understand the concept, let us consider a few examples

EXAMPLE 4.4 A point charge q is located in front of an infinite conducting plane at a perpendicular distance d, as shown in Fig. 4.11. What is the potential and electric field in the region outside the conductor?

Solution

Let the charge be located at the origin and the conducting plate at $x = -d$. The potential which satisfies Eq. (4.40) is given by

$$\phi(\vec{r}) = \frac{kq}{\sqrt{x^2 + y^2 + z^2}} + \phi_c(\vec{r}) \tag{4.43}$$

where $\phi_c(\vec{r})$ is such that $\phi(\vec{r})$ on the plane $x = -d$ is a constant.

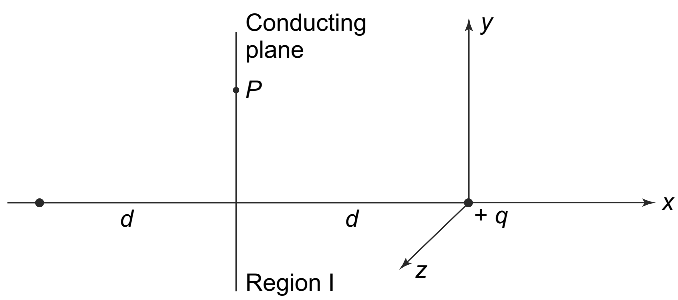


Fig. 4.11 *Example 4.4: A point charge q at a distance d from a conducting plane. We need to find the potential and electric field in the region that is to the right of the conducting plane*

The image charge which can create such a potential is fairly obvious—it will be a charge $-q$, located at the location of the mirror image of q on the left of the plane, i.e., at a point $(-2d, 0, 0)$. On the surface of the infinite conducting plane, at a point $(-d, y, z)$, this image charge will create a potential

$$\phi_c = -\frac{kq}{\sqrt{d^2 + y^2 + z^2}}$$

With such a charge, the potential at any point on the surface will then be given by a superposition of the two potentials, namely

$$\phi(x = -d, y, z) = \frac{kq}{\sqrt{d^2 + y^2 + z^2}} - \frac{kq}{\sqrt{d^2 + y^2 + z^2}} = 0 \tag{4.44}$$

and thus, satisfies the given boundary condition. Thus, at any point (x, y, z) in Region I, the potential is given by

$$\phi(x, y, z) = kq\left[\frac{1}{\sqrt{x^2 + y^2 + z^2}} - \frac{1}{\sqrt{(x + 2d)^2 + y^2 + z^2}}\right] \tag{4.45}$$

since the location of the image charge is $(-2d, 0, 0)$. The field at any point in Region I can be evaluated since $\vec{E} = -\vec{\nabla}\phi$ and we get

$$\begin{aligned}\vec{E}(\vec{r}) &= -\vec{\nabla}\phi(\vec{r})\\ &= -kq\left[\hat{i}\left(\frac{-x}{(x^2 + y^2 + z^2)^{3/2}} + \frac{x + 2d}{((x + 2d)^2 + y^2 + z^2)^{3/2}}\right)\right.\\ &\quad + \hat{j}\left(\frac{-y}{(x^2 + y^2 + z^2)^{3/2}} + \frac{y}{((x + 2d)^2 + y^2 + z^2)^{3/2}}\right)\\ &\quad \left.+ \hat{k}\left(\frac{-z}{(x^2 + y^2 + z^2)^{3/2}} + \frac{z}{((x + 2d)^2 + y^2 + z^2)^{3/2}}\right)\right]\end{aligned} \tag{4.46}$$

On the surface of the conductor, i.e. at $x = -d$, the field is

$$\vec{E}(-d, y, z) = \hat{i}(kq)\frac{(-2d)}{(d^2 + y^2 + z^2)^{3/2}} \tag{4.47}$$

which as expected is along the x axis and hence is normal to the conducting plane. The value of $\vec{E}$ at the conducting plane also allows us to evaluate the surface charge density σ since

$$\sigma = \frac{1}{4\pi k}(E_x(-d, y, z)) = \frac{-2qd}{4\pi}\frac{1}{(d^2 + y^2 + z^2)^{3/2}} \tag{4.48}$$

which is also as expected, maximum at $y = z = 0$ and falls off as we move away from this point.

Knowing the surface charge density, we can also easily calculate the total induced charge. To do this, we need to integrate σ over the whole conducting surface. Let

$$y^2 + z^2 = t^2$$

then

$$dydz = 2\pi t dt$$

where we have performed the angular integration. The total charge induced on the plane will be

$$q_c = \int_0^\infty\int_0^\infty \sigma dydz = -qd\int_0^\infty \frac{t}{(d^2 + t^2)^{3/2}}dt = -q \tag{4.49}$$

which is exactly what we expect. The induced charge on the surface of the conductor will obviously also create a force between the conductor and the charge q. The force, on an infinitesimal area $dydz$ on the conductor will be simply

$$\sigma(y, z)\vec{E}(-d, y, z)dydz$$

Total force is given by

$$\begin{aligned}\vec{F} &= \int_0^\infty\int_0^\infty \sigma(y, z)\vec{E}(-d, y, z)dydz\\ &= \hat{i}\int_0^\infty\int_0^\infty \frac{-2qd}{4\pi(d^2 + y^2 + z^2)^{3/2}}\frac{-2dkq}{(d^2 + y^2 + z^2)^{3/2}}dydz\end{aligned}$$

$$= \frac{kd^2q^2}{2\pi}\hat{i}\int_0^\infty \frac{2\pi t\,dt}{(d^2+t^2)^3}$$

$$= \frac{kq^2}{4d^2}\hat{i} \tag{4.50}$$

This is exactly the force between the charge q and its image charge $-q$ which is located at a distance of $2d$ from it.

There are not too many examples where one can find such a simple 'image' charge(s). The next example is another example of this.

EXAMPLE 4.5 A charge $+q$ is placed in front of, and at a distance of d from the centre of a conducting sphere of radius R. Find the potential and the electric field everywhere outside the sphere.

Solution

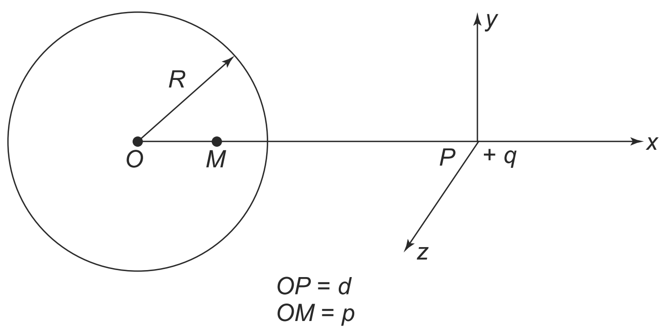


Fig. 4.12 *Example 4.5: A charge $+q$ in front of a conducting sphere. The potential is required outside the sphere in Region I*

As in the last example, the potential is given by

$$\phi(\vec{r}) = \phi_f(\vec{r}) + \phi_c(\vec{r}) \tag{4.51}$$

where

$$\phi_f(\vec{r}) = \frac{kq}{\sqrt{x^2+y^2+z^2}}$$

The potential $\phi_c(\vec{r})$ is to be determined such that the potential ϕ should satisfy the boundary condition that it is constant on the surface of the sphere

$$(x+d)^2 + y^2 + z^2 = R^2$$

To find the image charge q_c, we first note that if it exists, it must, by symmetry, lie on the line OP in the Fig. 4.12. Let its position be such that it is a distance p from the centre of the sphere, towards the point charge $+q$, i.e. to the right of the centre in the figure. Then in Region I, we can write the potential as

$$\phi(\vec{r}) = \phi_f(\vec{r}) + \phi_c(\vec{r}) = k\left[\frac{q}{\sqrt{x^2+y^2+z^2}} + \frac{q_c}{\sqrt{(x+d-p)^2+y^2+z^2}}\right] \tag{4.52}$$

Thus, on the surface of the sphere,

$$\begin{aligned}\phi(x, y, z)|_{\text{sphere}} &= k\left[\frac{q}{\sqrt{x^2 + R^2 - (x+d)^2}} + \frac{q_c}{\sqrt{(x+d-p)^2 + R^2 - (x+d)^2}}\right] \\ &= k\left[\frac{q}{\sqrt{R^2 - d^2 - 2xd}} + \frac{q_c}{\sqrt{R^2 + p^2 - 2p(x+d)}}\right] \quad (4.53)\end{aligned}$$

By inspection, if we choose $p = \frac{R^2}{d}$ and $q_c = -\frac{qR}{d}$, we see that

$$\phi(x, y, z)|_{\text{sphere}} = 0 \quad (4.54)$$

Hence, an image charge $q_c = -\frac{qR}{d}$, placed at a distance of $\frac{R^2}{d}$ to the right of the centre of the sphere, will make the surface of the sphere equipotential with potential equal to zero. In case the sphere is maintained at some non-zero potential V, then an extra image charge $Q = VR$ would need to be placed at the centre of the sphere to make the potential of the surface to be V which will be due to the charge q and the two image charges.

The potential at any point outside the sphere would be given by Eq. (4.52) with the values of q_c and p given above.

PROBLEM 4.4 A point charge q is placed at a distance a from a circular metal disc at zero potential. The disc has a radius b. Show that the total induced charge on the disc is given by

$$Q = -\left(\frac{2q}{\pi}\right)\arctan\left(\frac{b}{a}\right)$$

PROBLEM 4.5 A point charge q is placed at a distance r_1 from the centre of a grounded spherical conductor of radius r_0, ($r_0 > r_1$). Find the potential at an external point and determine the position and magnitude of the image charge.

4.4 CAPACITANCE OF CONDUCTORS

4.4.1 Coefficients of Potential

Consider a spherical conductor of radius R_1 carrying a charge Q_1. The conductor, as we have seen, is an equipotential surface and hence, has the same potential V on all points. This potential is simply the amount of work done in bringing a unit charge from infinity (where, by convention, we choose the potential to be zero) to the sphere. Thus,

$$V_1 = k\int_{R_1}^{\infty}\frac{Q_1}{r^2}dr = k\frac{Q_1}{R_1} \quad (4.55)$$

In general for an isolated conductor, we will have a proportionality relation like Eq. (4.55) i.e., $V_1 = \frac{Q_1}{C}$, where the constant C is called the capacitance of the conductor. For a spherical conductor of radius R_1, we thus have $C = \frac{R_1}{k}$, and in general the capacitance of a conductor would be a function of its shape

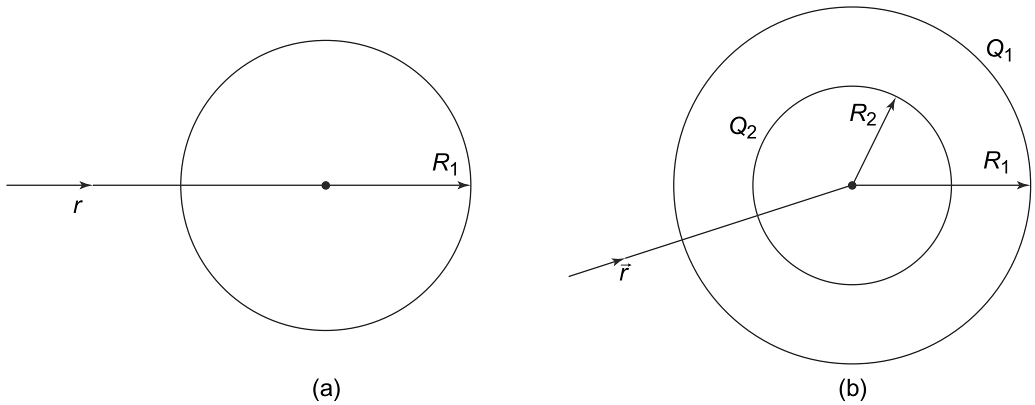


Fig. 4.13 *(a) A unit charge taken from ∞ to the surface of the charged spherical conductor. The work done is given by Eq. (4.55), (b) The same process done with a pair of concentric spheres of radii R_1, R_2 with $R_1 > R_2$. The work done now changes to Eq. (4.56)*

and size. The unit of Capacitance in SI units is 1 Farad = 1 Coulomb / 1 Volt. This turns out to be a huge unit for ordinary sized conductors and so sub-units like pf (picofarad = 10^{-12} F) or μF (microfarad = 1 μF = 10^{-6} F) are more commonly used. The dimension of capacitance is length.

In our calculation of the work done, we have implicitly assumed that the electric field is solely due to the charge Q_1 and there are no other charges present anywhere else. Now suppose a second charged conductor is present, as shown in Fig. 4.13(b) which is a sphere of radius R_2 carrying charge Q_2, concentric with the original sphere. If we now bring the charge from infinity to the first sphere, along any radial direction, then the work done will be

$$\begin{aligned} V_1' &= k\int_{R_1}^{\infty} dr \left(\frac{Q_1 + Q_2}{r^2}\right) \\ &= k\left(\frac{Q_1}{R_1} + \frac{Q_2}{R_1}\right) \end{aligned} \tag{4.56}$$

which is different from V_1.

In general, the potential of any conductor with a certain amount of charge is dependent on the presence of other charges elsewhere. A general relationship, of the type Eq. (4.56) can be reasoned as follows:

Let there be n conductors, carrying charges $Q_1, Q_2, \cdots, Q_n$ respectively and let their potentials be $V_1, V_2, \cdots, V_n$. Then at any point in space, the electric field will obviously depend on the shapes, sizes and locations of these charged conductors as well as the amount of charge on each. But for any configuration, the electric field at any point will be, by the Principle of Superposition, a sum of n terms with the i^{th} term being proportional to Q_i. Hence, the work done bringing a unit charge from ∞ to the i^{th} charged conductor will be similarly a sum of n terms

$$V_i = p_{i1}Q_1 + p_{i2}Q_2 + \cdots + p_{in}Q_n \tag{4.57}$$

An equivalent argument to arrive at Eq. (4.57) is by considering the solution to the Laplace equation. The solution to the Laplace equation, as we have seen, gives us the value of the potential at all points outside the conductors (i.e., in regions where there is no charge). The boundary condition that we use to solve the Laplace equation is that at very large distances (at ∞), the potential approaches

$$k\frac{\sum Q_i}{r}$$

Now notice that if the charges are all increased by some factor λ, the potential is simply λ times the earlier potential since the equation is linear. Thus, the potential everywhere must be a linear function of the charges.

The coefficients p_{ij} would depend on the relative location of the charges (since the field depends on this) but not on the charges themselves. Thus, in the example of the concentric spheres above

$$\begin{aligned} p_{11} &= \frac{k}{R_1} \\ p_{12} &= \frac{k}{R_1} \end{aligned} \tag{4.58}$$

The coefficients p_{ij}'s are called the coefficients of potential. They depend on the detailed configuration of the conductor. It is also clear that they also change if there are dielectrics present since the presence of dielectric materials changes the electric field at any point.

4.4.2 Potential Energy of a Set of Charged Conductors

In Chapter 2, we have already seen how to get the total electrostatic energy of a collection of charges in terms of the potential at those charges. We can do a similar exercise with a set of n charged conductors in terms of the coefficients of potential.

Imagine the charges being built up in infinitesimal steps of $(d\lambda)Q_i$, i.e., a charge of this amount is brought from infinity to the i^{th} conductor and this process is repeated till the charge on the conductor is finally Q_i. Thus, at any intermediate stage, the charge on the i^{th} conductor would be λQ_i where λ will be 0 initially and 1 when the charge is Q_i. At any intermediate stage, the potential of the i^{th} conductor would be

$$U_i(\lambda) = \lambda \sum_j p_{ij} Q_j \qquad\qquad i = 1, 2, \cdots, n \tag{4.59}$$

Therefore, when a charge $d\lambda Q_i$ is brought from ∞ to the i^{th} conductor, and this process is done for all the charges, the work done is

$$dW(\lambda) = \lambda \sum_i U_i(\lambda) d\lambda Q_i \tag{4.60}$$

and hence the total electrostatic energy or the work done in assembling this configuration of conductors to their respective charges is

$$\boxed{W = \int_{\lambda=0}^{1} dW(\lambda) = \frac{1}{2}\sum_{i,j=1}^{n} p_{ij} Q_i Q_j} \tag{4.61}$$

One property of the coefficients of potential p_{ij}s can be immediately recognised. Let all the Q_is be positive. Then the potential energy of each pair is positive since the like charges repel each other. Thus, in Eq. (4.61), the p_{ij} must be positive for all i, j.

Equation (4.57) is a set of linear equations relating the n quantities V_is to n quantities Q_is. These can be reversed and we can get the Q_is in terms of V_is. This is what we do next.

4.4.3 Capacitance Matrix and Capacitors

Equation (4.57) can be inverted and instead of expressing V_is in terms of Q_is, we can do the reverse and get Q_is in terms of V_is. These again will be linear functions

$$Q_i = \sum C_{ij} V_j \tag{4.62}$$

where C_{ij} **are elements of a matrix called the capacitance matrix**. For a single conductor, $n = 1$ this reduces to $Q_1 = C_{11} V_1$, which is a direct proportionality relation. **Equation (4.62) with $C_{11} = C$, is referred to usually as the capacitance of the conductor.** It should be remembered that V is the potential of the conductor relative to infinity.

The following example will illustrate the magnitude of capacitances in relation to size.

EXAMPLE 4.6 What is the capacitance of a metal ball of the size of the earth, i.e., a metallic sphere of radius 6400 km ?

Solution

From Eq. (4.55), we know that the capacitance of an isolated sphere of radius R is

$$C = \frac{R}{k}$$

Putting in the values for $R = 6400$ km and $k = 8.9 \times 10^9$ Nm^2/C^2, we get $C \approx 0.7$ μF.

Thus, conductors of size in meters or centimeters have capacitances in nanofarads (10^{-9} F) or picofarads (10^{-12} F).

Note that just like the coefficients of potential, the coefficients of capacitance are symmetric, that is $C_{ij} = C_{ji}$. As an example, consider the case of $n = 2$, i.e., 2 conductors with charges and potentials Q_1, Q_2, V_1, V_2. Then Eq. (4.57) for this case gives us

$$\begin{aligned} V_1 &= p_{11} Q_1 + p_{12} Q_2 \\ V_2 &= p_{21} Q_1 + p_{22} Q_2 \end{aligned} \tag{4.63}$$

Inverting this, we get

$$\begin{aligned} Q_1 &= \frac{p_{22} V_1 + p_{21} V_2}{p_{11} p_{22} - p_{12} p_{21}} \\ Q_2 &= \frac{-p_{12} V_1 + p_{11} V_2}{p_{11} p_{22} - p_{12} p_{21}} \end{aligned} \tag{4.64}$$

If one were to use matrix notation, then in general

$$\mathbf{C} = |C_{ij}| = \mathbf{P}^{-1}$$

where $\mathbf{P}$ is the matrix of p_{ij}s.

The energy W, given by Eq. (4.61) can also be obtained in terms of C_{ij}s instead of p_{ij}s. To do this, let us use a matrix notation. The collection of charges Q_i, $i = 1, 2, \cdots, n$ can be thought of as a column vector of n rows or a $n \times 1$ matrix.

$$\mathbf{Q} = \begin{pmatrix} Q_1 \\ Q_2 \\ . \\ . \\ Q_n \end{pmatrix}$$

Similarly, we can think of the n potentials $V_1, V_2, \cdots, V_n$ as a column vector

$$\mathbf{V} = \begin{pmatrix} V_1 \\ V_2 \\ . \\ . \\ V_n \end{pmatrix}$$

Then

$$\begin{aligned} \mathbf{W} &= \frac{1}{2}\mathbf{Q}^T\mathbf{P}\mathbf{Q} \\ &= \frac{1}{2}\mathbf{V}^T\mathbf{C}^T\mathbf{P}\mathbf{C}\mathbf{V} \\ &= \frac{1}{2}\mathbf{V}^T\mathbf{C}\mathbf{V} \end{aligned} \tag{4.65}$$

Thus, we have

$$\boxed{\mathbf{W} = \frac{1}{2}\sum_{i,j=1}^{n}\mathbf{C}_{ij}V_iV_j}$$

where we have used the fact that $\mathbf{C}$ is symmetric and hence $\mathbf{C}^T = \mathbf{C}$ by definition. Also, we have used the fact that $\mathbf{C} = \mathbf{P}^{-1}$ and hence $\mathbf{CP} = \mathbf{I}$, the unit matrix.

4.4.4 Capacitors with Equal Charges on the Plates

A special case of a set of conductors is one in which there are only conductors carrying equal and opposite charges $\pm Q$. This is called a capacitor or a condensor. Consider a case where there are two conductors carrying equal and opposite charge $\pm Q$. Then Eq. (4.62) becomes

$$\begin{aligned} Q &= C_{11}V_1 + C_{12}V_2 \\ -Q &= C_{21}V_1 + C_{22}V_2 \end{aligned} \tag{4.66}$$

Such an arrangement of a pair of conductors implies that all electric field lines starting from or ending on one of the conductor necessarily end on or start from the other.

A pair of concentric hollow metallic spheres of different sizes R_1 and R_2 with $R_1 > R_2$, with the inner one held by some mechanical device is an example of such an arrangement. It is instructive to calculate the coefficients of capacitance for such an arrangement. Since by Gauss's Law, there will be no electric field outside the spheres, clearly the potential of the outer sphere with potential V_2 must have $V_2 = 0$ by definition. C_{12} and C_{22} thus drop out of Eq. (4.66) and we get

$$Q = C_{11}V_1 = -C_{12}V_1 \tag{4.67}$$

Which tells us that $C_{11} = -C_{12}$. The value of C_{11} is easily determined. The electric field in the region between the spheres, with a charge Q on the inner sphere is

$$E(r) = \frac{kQ}{r^2}$$

directed outwards. Since the outer sphere is at zero potential, we get

$$0 - V_1 = -(kQ)\int_{R_2}^{R_1} d^3r'\frac{1}{r'^2}$$

$$V_1 = 4\pi kQ\left(\frac{1}{R_2} - \frac{1}{R_1}\right) \tag{4.68}$$

Note that C_{22} does not enter the calculation and the capacitance for this capacitor is defined as

$$C = \frac{Q}{V_1 - V_2} = \frac{Q}{V_1} = \frac{1}{4\pi k}\frac{R_2R_1}{R_1 - R_2} \tag{4.69}$$

A more practical capacitor widely used is a pair of large parallel plates held close to each other and separated by a distance d. If d is much greater than the size of the plates, we can consider the plates infinite in size and assume the electric field to be confined to the region between the plates. The arrangement here is completely symmetric between the two plates so that $C_{11} = C_{22}$. From Eq. (4.66), we get

$$C_{12} = C_{21} = -C_{11}$$

and hence

$$Q = C_{11}(V_1 - V_2)$$

The plane midway between the plates is at zero potential since the field there is normal to it and one can go to infinity along this plane without any work done. Thus, if we call V_1 the potential of the positively charged plate and V_2 of the negatively charged plate, then $V_1 = -V_2 = \frac{V}{2}$ where V is the potential difference between the plates, that is

$$V = (V_1 - V_2)$$

The capacitance C is defined by

$$Q = C(V_1 - V_2) = CV$$

Thus, in this case C is simply $C = C_{11} = C_{22}$.

The expression for energy also simplifies in this case. We get from Eq. (4.65)

$$W = \frac{1}{2}\left(C_{11}V_1^2 + C_{22}V_2^2 + C_{12}V_1V_2 + C_{21}V_2V_1\right)$$

$$= \frac{1}{2}CV^2 \tag{4.70}$$

Thus, for a given V, the energy stored in a capacitor grows with value of capacitance. The energy stored in capacitors usually found in electronic circuits is thus, very tiny and for practical applications as storage devices, capacitors with capacitances orders of magnitude higher than these have been designed.

We can understand the expression for the energy as follows: Imagine that we are building up the charge on the capacitor in an infinitesimal fashion. What this means is that we carry a charge $(d\lambda)Q$ from the negatively charged conductor to the positively charged one when the conductors have charges $-\lambda Q$ and λQ and a potential difference of $\frac{\lambda Q}{C}$. After we have transferred this infinitesimal charge, their charges would be $-(\lambda + d\lambda)Q$ and $(\lambda + d\lambda)Q$, Initially, $\lambda = 0$ and when the conductors are charged fully to their final configuration, $\lambda = 1$. Initially, since $\lambda = 0$, there are no charges on the conductors, hence no electric field and therefore, no energy. Thus, the energy needed to build up the charge from 0 to $\pm Q$ or from $\lambda = 0$ to $\lambda = 1$ is

$$\begin{aligned} W &= \int_{\lambda=0}^{1} d\lambda Q \frac{\lambda Q}{C} \\ &= \frac{1}{2}\frac{Q^2}{C} \\ &= \frac{1}{2} C V^2 \end{aligned} \tag{4.71}$$

This is thus, the energy stored in the charged capacitor.

The property of a capacitor to able to store energy when charged makes it a very useful device in electric and electronic circuits. When discharged this energy appears elsewhere in the circuit and we shall, later in the book, study such circuits.

One of the commonest configuration of a capacitor is the parallel plate capacitor. It consists of two planar conducting plates of area A, identical in shape and size which are placed parallel to each other such that corresponding points on their surfaces and the edges have a separation d perpendicular to the plane of the conductors (Fig. 4.14).

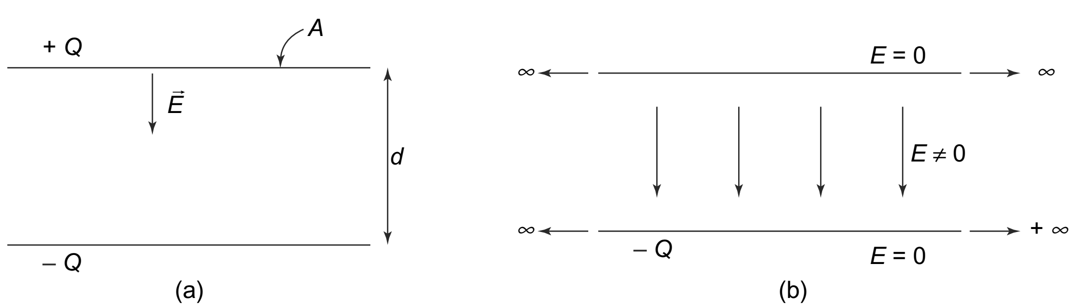


Fig. 4.14 *A parallel plate capacitor. The top and bottom plates carry charges $+Q$ and $-Q$ respectively. The electric field lines run from the positive plate to the negative plate. (b) In the limit where the area A and charge Q go to infinity with the surface charge density $\sigma = \frac{Q}{A}$ remaining constant, the field is restricted to the gap between the plates only*

The two plates being conductors will be both equipotential surfaces. As we have just established, by symmetry, their potentials will be equal and opposite, $\frac{V}{2}$ for the positive and $-\frac{V}{2}$ for the negative plate so that the potential difference between them is V. The calculation of the electric field involves solution of the Laplace equation outside the plates with the boundary condition of the potential being constant at the plates. This is a situation where the solution cannot be written down in an analytic form. However, if we go to the limit where the area A and the charge Q both go to infinity with the surface charge density, $\sigma = \frac{Q}{A}$ remaining constant, the problem is easily solved.

For the surface of a conductor, we have already calculated the electric field at the surface to be $\vec{E} = 4\pi k\sigma\hat{n}$ where $\hat{n}$ is the outward normal to the surface. In the present case, charges on the two plates will be there on the inner faces facing each other with no surface charge on the outer faces. This is easily seen by applying Gauss's Theorem. By symmetry, the electric fields, wherever they are, will always be normal to the plates and hence, constant as we follow the lines of force. Suppose there are fields, E', present above the top plate in the upward direction. By symmetry, we have the same field E' below the bottom plate in the downward direction. Consider a pill box as shown in Fig. 4.15 with unit cross-section area. The electric flux across the cylindrical surfaces vanishes since the field is parallel to this surface. The flux through the end pieces adds up to $2E'$ (remember the end surfaces are unit area and the direction of the outward normal is the same as the direction of the field). By Gauss's Theorem, then

$$2E' = 4\pi k \times \text{total charge enclosed} = 0$$

since the two plates are equally but oppositely charged. Hence, $E' = 0$. Since $E' = 0$, there are no surface charge on the outer surfaces of the plates (since E' and the surface charge density of the outer faces are related by a factor $4\pi k$).

The surface charges on the inner faces will create a field $E = (4\pi k)\sigma$ in the gap between the plates. Hence, the potential difference V between the plates is:

$$V = Ed = (4\pi k)\sigma d$$

The capacitance therefore, is:

$$C = \frac{Q}{V} = \frac{\sigma A}{4\pi k\sigma d} = \frac{1}{4\pi k}\frac{A}{d}$$

Note that strictly speaking, this goes to ∞ since we have derived this result in the limit $A \to \infty$.

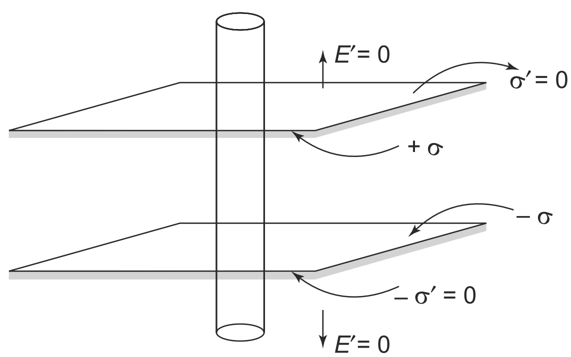


Fig. 4.15

We have assumed in the derivation above that the gap between the plates contains nothing. If instead a dielectric material of dielectric constant K, fills up the gap, the electric field in the gap would be reduced by a factor K, as we saw in the previous Chapter. The capacitance in that case would be

$$\boxed{C = \frac{Q}{V} = \frac{\sigma A}{(4\pi k\sigma d/K)} = \frac{1}{4\pi k}\frac{KA}{d}}$$

Thus, capacitances of capacitors increase if a dielectric is present in the gap. This fact is used widely in practical uses of capacitors to increase their capacitances. Typical substances used are glass ($\frac{K}{4\pi k} \sim 3-10$), mica ($\frac{K}{4\pi k} \sim 5-7$), metal oxides ($\frac{K}{4\pi k} \sim 6-10$), paper ($\frac{K}{4\pi k} \sim 2.5-3.5$) which can be used for relatively low voltage (a few volts to a few tens of volts). Ceramics have much higher K, sometimes as high as a thousand and therefore, ceramic filled capacitors are used in high voltage situations.

We can estimate the rough order of magnitudes of capacitances of parallel plate capacitors as in the following example:

EXAMPLE 4.7 Calculate the capacitance of a square shaped parallel plate capacitor whose plate size is 1 square cm and the separation between the plates is 1 mm.

Solution

We know that

$$C = \frac{1}{4\pi k}\frac{A}{d}$$

which in the present case gives us

$$C = \frac{1}{4\pi \times 9 \times 10^9}\frac{10^{-4}}{10^{-3}} = 9 \times 10^{-13}\text{F (approximately)}$$

We can also estimate the electrostatic force experienced by the plates of a parallel plate capacitor. The charge on either plate is σA and the electric field in the gap is $E = 4\pi k\sigma$. The force experienced by either plate is therefore,

$$\boxed{F = 4\pi k\sigma^2 A}$$

This force has to be balanced by non electrostatic forces for the plates to remain fixed.

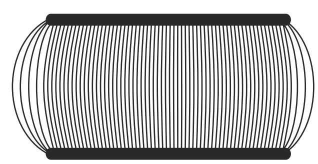


Fig. 4.16 *A typical image of electric field lines in a parallel plate capacitor obtained by having light silk thread lining up along them*

In practice, the last equation which assumed fields existing only in the gap between the plates, is still a very good approximation when d is much smaller than the linear size of the plates: $d << A^{1/2}$. Fig. 4.16 shows a typical representation of the nature of the fields of a finite size parallel plate capacitor, which clearly shows that such an approximation is reasonably good. The field does bulge outwards at the edges, which causes a 'edge correction' to the previous result. The edge correction can actually be approximately estimated.

4.4.5 Estimate of Edge Correction of a Parallel Plate Capacitor *

An approximate estimate of the correction due to finite size of a parallel plate capacitor can be made as follows. We consider a parallel plate capacitor with circular plates of radius R. Consider the plane midway between the plates of the capacitor. By symmetry, it is at zero potential.

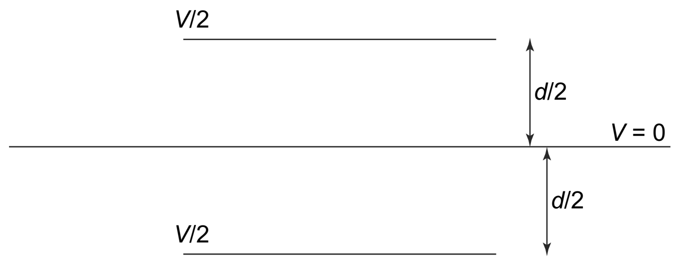


Fig. 4.17 *Zero potential plane*

If $R \to \infty$, then the field lines would be completely confined to within the capacitor plates. For finite R, the electric field lines bulge out of the plates and the charge accumulates at the edges as well as on the upper surface of the plates as in the figure below (Fig. 4.18).

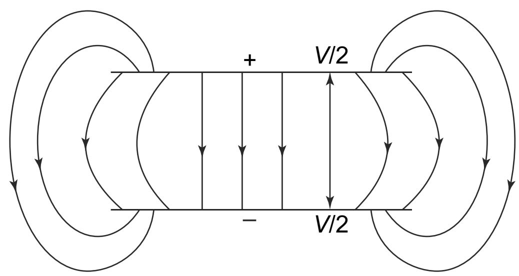


Fig. 4.18 *The case of a finite sized capacitor*

By symmetry, the surface midway between the plates is at zero potential. Also, away from the plates, in between the two plates, the field lines are straight and the value of the electric field is simply $E = V/d$. The surface charge density thus, in these regions is

$$\sigma = \frac{E}{4\pi k} = \frac{V}{4\pi kd}$$

Towards the edges, the field lines are curved of course. If we neglect the fact that the zero potential surface is removed by a distance $d/2$ from the surfaces and consider it as adjoining, then the field lines look like Fig. 4.19.

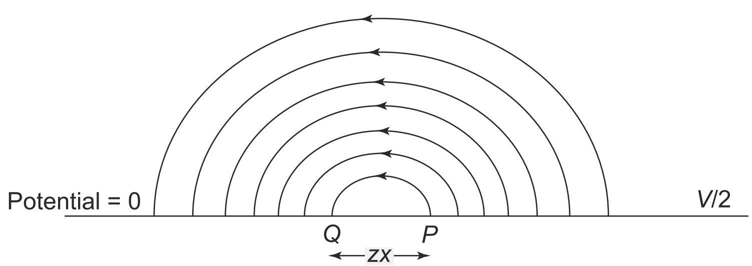


Fig. 4.19 *Approximate field lines between two adjacent surfaces at different potentials*

Now let us assume that the field lines joining the surface of the plates (at potential $V/2$) and the median plane with potential $V = 0$ are semi-circular. This can be proved by solving the Laplace equation for small values of the distance from the edge of the plate, of the point P in Fig. 4.19. **(S.J.N. Shaw, Physics of Fluids, 13, 1933, (1970), L.D. Landau & Lifshitz, 'Electrodynamics of Continuous Media', pg 19, G. Kirchhoff, M Deutsch Akad Wiss Berlin, 144 (1877))**. Then, a reasonable estimate of the electric field at a distance x from the edge of the plate is

$$E \times \text{length of the semicircle joining surface to the median plane} = \frac{V}{2}$$

$$E(x) = \frac{V}{2\pi x} \tag{4.72}$$

This expression is obviously not valid for very small values of x, since we have neglected distances of the order of d. We will assume this value of $E(x)$ to be valid for $R > x > d$ where R is the radius of the circular plates. Again, this expression cannot be true for values of x comparable to R since the field drops rapidly near the edges. Not too much of an error is made in assuming that this field is valid for $d \leq x \leq R$. The surface charge density corresponding to this value of $E(x)$ is given by

$$\sigma = \frac{E}{4\pi k} = \frac{V}{8\pi^2 kx} \tag{4.73}$$

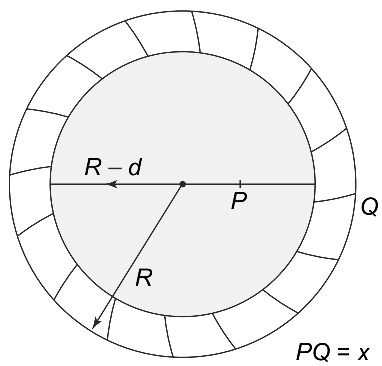


Fig. 4.20 *Edge effects and additional charge on a circular parallel plate capacitor*

and hence, this charge density contributes to the charge on the plate to the extent of

$$\begin{aligned} Q_{\text{edge}} &= \int_0^{R-d} \frac{V}{8\pi^2 k} \frac{2\pi r dr}{R - r} \\ &= \frac{V}{4\pi k} \int_0^{R-d} \frac{r dr}{R - r} \\ &= \frac{V}{4\pi k} \left[-(R-d) - R \ln\left(\frac{d}{R}\right) \right] \\ &\approx \frac{V}{4\pi k} R \left[\ln\left(\frac{R}{d}\right) - 1 \right] \end{aligned} \tag{4.74}$$

where in the last line we have assumed that $R \gg d$.

If this is the 'extra' contribution to the charge, the capacitance then becomes

$$\begin{aligned} C &= \frac{Q + Q_{\text{edge}}}{V} \\ &= \frac{1}{4\pi k}\left[\frac{A}{d} + R\ln\left(\frac{R}{d}\right)\right] \\ &= \frac{1}{4\pi k}\frac{\pi}{d}\left[R^2 + \frac{Rd}{\pi}\ln\left(\frac{R}{d}\right)\right] \\ &\approx \frac{1}{4\pi k}\frac{\pi}{d}\left[(R + \frac{d}{2\pi}\ln\left(\frac{R}{d}\right))^2\right] \end{aligned} \tag{4.75}$$

As a general rule then, we can think of the edge effect increasing the radius R to an effective radius R_{eff} which is $R + \frac{d}{2\pi}\ln\left(\frac{R}{d}\right)$.

A more exact formula was obtained by Kirchhoff in 1877 and is given by

$$C = \frac{A}{4\pi kd} + \frac{R}{4\pi k}\left(\ln\left(\frac{16\pi R}{d}\right) - 1\right) \tag{4.76}$$

This is close to the result we have obtained above in an approximate sense numerically.

EXAMPLE 4.8 A capacitor consists of two cylindrical infinite plates which are coaxial with radii a, b with $a > b$, as shown in Fig. 4.21. The plates carry a charge of Q per unit length. The gap between the two cylinders is filled with a dielectric of dielectric constant K. Find the capacitance of such a capacitor.

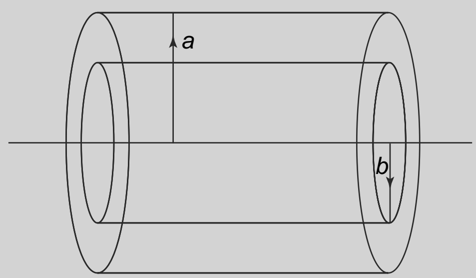


Fig. 4.21 *Example 4.8: A cylindrical capacitor*

Solution

To find the capacitance, we need to find the potential difference between the two cylinders as a function of the charges (equal and opposite) on the two cylinders. To do this, we use Gauss's Law to determine the electric field in the space between the cylinders and then use the relationship between potential and electric field to determine the potential.

By symmetry, at any point P between the cylinders, at a distance r from the axis ($b < r < a$), the field will be radially outwards and will only depend on r. Thus, choosing a Gaussian surface as a cylinder of unit length, of radius r, coaxial with the two cylinders, we apply Gauss's Law

$$\begin{aligned} 2\pi r E(r) &= \frac{kQ}{K} \\ E(r) &= \frac{kQ}{2\pi r K} \end{aligned} \tag{4.77}$$

But the potential difference between the two cylinders is related to the electric field as

$$V = -\int_a^b \vec{E} \cdot d\vec{l}$$

$$= -\frac{kQ}{2\pi K} \int_a^b \frac{dr}{r}$$

$$= \frac{kQ}{2\pi K} \ln\left(\frac{a}{b}\right) \tag{4.78}$$

Obviously, the capacitance is infinite since the total charge QL is infinite as $L \to \infty$. But the capacitance per unit length can be determined

$$\frac{C}{L} = \frac{Q}{V} = \frac{2\pi K}{k \ln\left(\frac{a}{b}\right)} \tag{4.79}$$

PROBLEM 4.6 Consider a parallel plate capacitor with area A and plate separation d. The dielectric constant of the dielectric inside the capacitor increases linearly from K_1 at the plate at $x = 0$ to K_2 at $x = d$. Find the capacitance of the arrangement.

PROBLEM 4.7 The space between two concentric conducting shells of radius $r = 10$ cm and $r = 15$ cm is filled with a dielectric material of dielectric constant $K = 5$. The shells are maintained at a potential difference of 100 V. Find the capacitance of the system and also the charge density on the shell at $r = 15$ cm.

PROBLEM 4.8 A spherical capacitor has an inner conductor of radius r_1 carrying a charge Q. The inner conductor is at zero potential. The outer conductor contracts from an initial radius r_2 to r_3. Show that the work done by the electric field is

$$W = \frac{Q^2(r_2 - r_3)}{8\pi \varepsilon r_2 r_3}$$

PROBLEM 4.9 A cylindrical capacitor with inner radius r_1 and outer radius r_2 is filled with a dielectric with $\varepsilon = \varepsilon_0 \frac{\alpha}{r^2}$ where α is a constant. Determine the capacitance of the capacitor.

4.4.6 Capacitors with Unequal Charges on the Plates

In the last sub-section, we have analysed capacitors with equal and opposite charges on the plates. This of course, is the most common configuration. However, for clarity and completeness, let us consider the case of a capacitor with unequal charges on the plates. This is shown in Figs. 4.22(a) and (b). We shall be using an infinite parallel plate capacitor to discuss this, though the results will be more general.

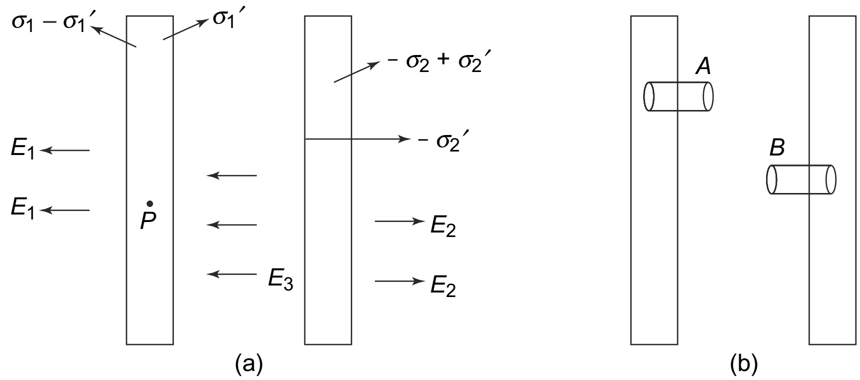


Fig. 4.22 *(a) Parallel plate capacitors with surface charge densities σ_1 and $-\sigma_2$ on the two plates. The plates are drawn thick so that the charge densities can be seen on the individual faces of each plate, (b) The various Gaussian pill boxes which can be drawn*

Since the plates are infinite, we work with the surface charge densities

$$\sigma_1 = \frac{Q_1}{A}$$
$$-\sigma_2 = -\frac{Q_2}{A}$$

Note that $A \to \infty$ but the charge density remains finite. The charge densities on the faces of the plates are shown in Fig. 4.22a. We draw two pill box shaped Gaussian surfaces A and B as shown in Fig. 4.22b. Applying Gauss's Theorem to these and remembering that the field inside the conductor vanishes, we get

$$-E_3 = (4\pi k)\sigma_1'$$
$$E_3 = (4\pi k)(-\sigma_2') \tag{4.80}$$

Hence, $\sigma_1' = \sigma_2'$ and the charges on the inner surfaces of the plates are equal and opposite.

Now consider a point P **inside** the left plate. The field at this point vanishes. The field is caused by a surface charge $(\sigma_1 - \sigma_1')$ on the left and $\sigma_1', -\sigma_2', (-\sigma_2 + \sigma_2')$ on its right. Thus, we must have

$$\sigma_1 - \sigma_1' = \sigma_1' - \sigma_2' - \sigma_2 + \sigma_2' \tag{4.81}$$

or since $\sigma_1' = \sigma_2'$, we get

$$\sigma_1' = \frac{(\sigma_1 + \sigma_2)}{2} = -\sigma_2'$$
$$\sigma_1 - \sigma_1' = \frac{(\sigma_1 - \sigma_2)}{2} = -(\sigma_2 - \sigma_2') \tag{4.82}$$

A capacitor with unequal charges on the two plates does not occur naturally in electrical circuits. Usually, an uncharged capacitor is charged by connecting its two plates to the two ends of a battery or another electrical power source. Whatever charge the source deposits on one plate, an equal and opposite amount of charge is deposited automatically on the other plate. The power source thus, does not supply any net charge to the capacitor and hence $\sigma_1 = \sigma_2$.

4.4.7 Combinations of Capacitors

We have already discussed that a capacitor is a system of two conductors such that there are no other conductors. This condition ensures that the potential difference between the plates is only related to the charges on the plates. There may, however, be situations in which even though one or more capacitors are present in the vicinity, and yet, the charge on each of those capacitors is still only related to the potential difference between its own plates. This is obviously possible only if the field in the gap between the plates of each of the capacitors is only caused by the charges on their own plates. This situation is possible only approximately in practice.

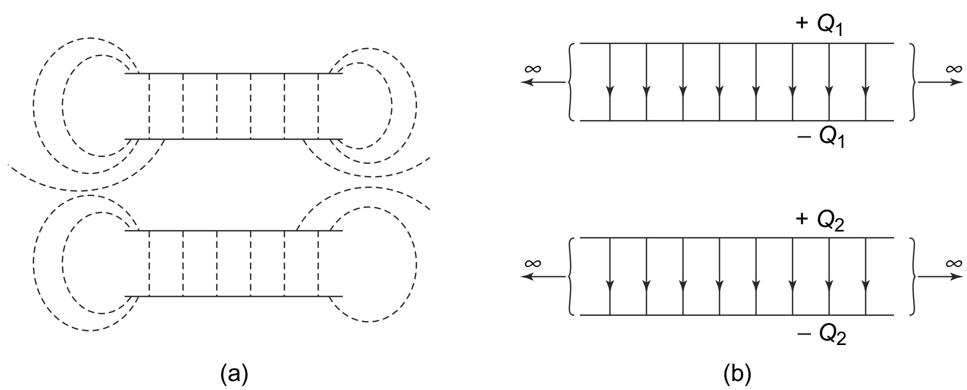


Fig. 4.23 *More than one capacitor: (a) The field at any point is caused by the charges on the plates of either capacitor. There may be points where the field due to one capacitor might dominate but in general both will be present, (b) Two parallel plate, idealised capacitors where the respective fields are confined only to the gap between the plates and hence are only dependent on the charges on the individual plates.*

Given such idealised capacitors, with their individual charge-voltage relationships, we can think of two basic ways of combining them.

Parallel Combination C_1 and C_2 are the capacitances of two capacitors carrying charges Q_1 and Q_2 respectively (Fig. 4.24(a)).

Now if the two capacitors are joined in such a way that their positive plates are at the same potential and similarly the negative plates of the two are at that same potential, then we have a configuration shown in Fig. 4.24(b). The potential difference between the positive and negative plates is V and so we can think of the combination as a single capacitor with capacitance

$$C = \frac{Q_1 + Q_2}{V} = C_1 + C_2 \tag{4.83}$$

This combination is called a parallel combination of capacitors. Obviously, this can be easily generalised to any number of capacitors connected in parallel Fig. 4.25.

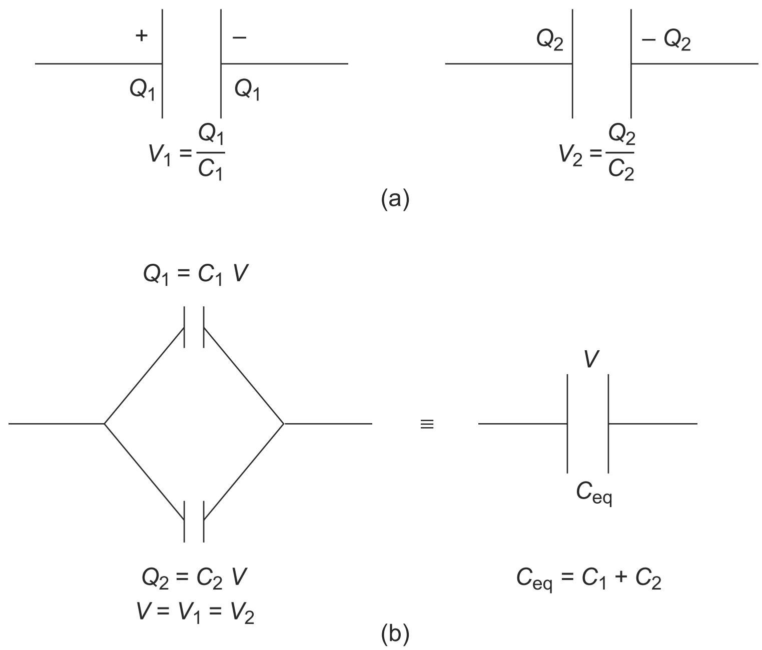


Fig. 4.24 *Parallel combination of capacitors: (a) Two capacitors with capacities C_1 and C_2 in isolation, (b) A parallel combination of these*

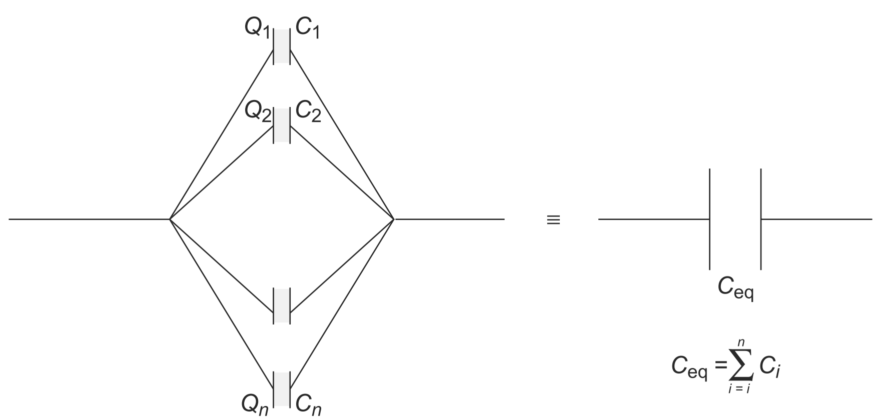


Fig. 4.25 *Parallel combination of n capacitors. The equivalent capacitance of such a combination is the sum of individual capacities*

It is easy to see that in this case, the combination can be considered to be equivalent to a single capacitor with a capacitance given by

$$\boxed{C_{\text{eff}} = \frac{Q_1 + Q_2 + \cdots + Q_n}{V} = C_1 + C_2 + \cdots + C_n} \tag{4.84}$$

The fact that capacitances add up when capacitors are connected in parallel is put to practical use when designing capacitors with large capacitances. The schematic sketch shown in Fig. 4.26, shows a design for one such combination.

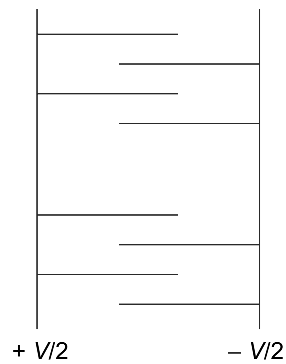


Fig. 4.26 *An array of parallel plate capacitors in parallel. The total charge on the positive potential plate is the sum of charges on the plates connected to it and similarly on the negative side*

It is easy to see that the equivalent configuration in terms of $C_{\rm eff}$ has the same energy as the original capacitors. The total energy stored in the individual capacitors is given by

$$\boxed{\frac{1}{2}C_1V^2 + \frac{1}{2}C_2V^2 + \cdots + \frac{1}{2}C_nV^2 = \frac{1}{2}C_{\rm eff}V^2}$$

Series Combination Consider now the same two capacitors as above, but connected as shown in Fig. 4.27.

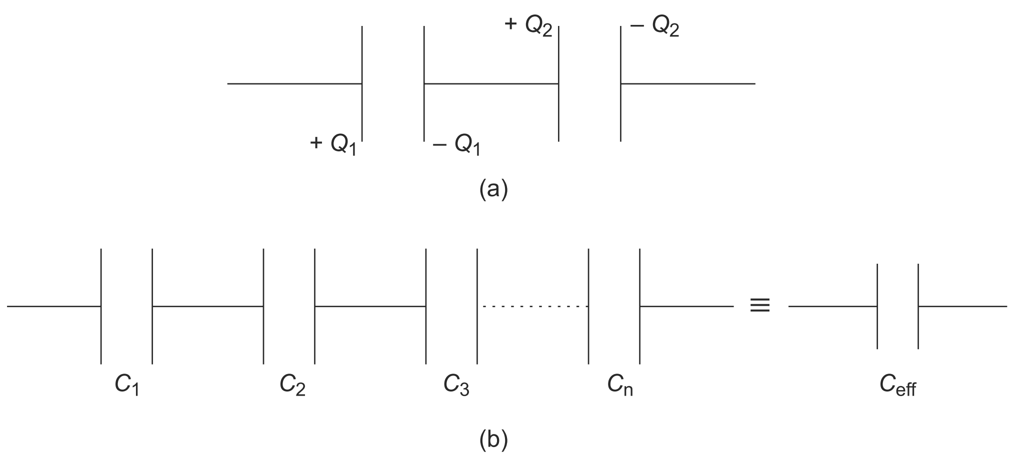


Fig. 4.27 *Series combination of capacitors: (a) Two capacitors carrying charges $\pm Q_1$ and $\pm Q_2$ in series, (b) Series combination of n capacitors. The equivalent capacitance of such a combination is the reciprocal of the sum of the reciprocals of the individual capacitances*

Unlike the parallel combination, in this case, the negative plate of one capacitor is connected with the positive plate of the other one. Consider first the case when Q_1 is not equal to Q_2. The two inner plates will be at the same potential and the combination of the two will carry a charge $(Q_2 - Q_1)$. The whole arrangement thus, will be of three conductors carrying charges $+Q_1$, $(Q_2 - Q_1)$ and $-Q_2$ at different potentials. Thus, this combination cannot be replaced, as far as charges and potentials are concerned by a single capacitor carrying equal and opposite charges on its plates.

However, in practice the case where the charges are equal, i.e., $Q_1 = Q_2$ is the case which is most often encountered. This is because capacitors are usually charged from a neutral state, by connecting the plates to a battery. In the case of a series combination shown above, the charge flows into the two outer plates of the series combination. Let us analyse this case.

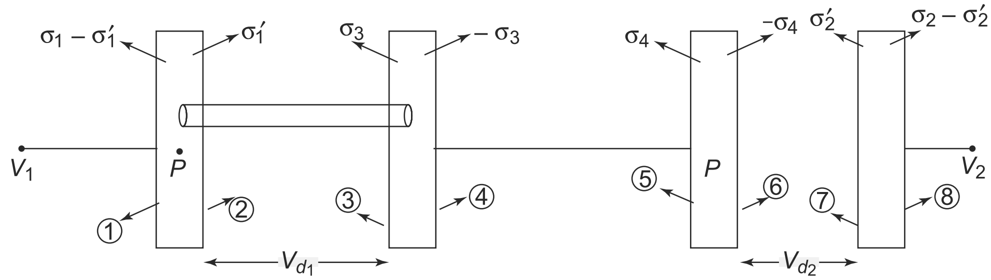


Fig. 4.28 *Series combination of 2 capacitors. The surface charge densities are as shown*

The various surface charge densities are shown in Fig. 4.28. There is no electric field inside the plates since they are conducting. Let us consider two Gaussian surfaces A, B, in the shape of unit area pill boxes as shown in the Fig. 4.28. Since there is no flux across A and B, the charge densities on faces 2 and 3 and also on faces 6 and 7 must be equal and opposite. Thus,

$$\begin{aligned}\sigma_1' &= -\sigma_3 \\ \sigma_2' &= \sigma_4 \end{aligned} \tag{4.85}$$

Now consider a point P in the left plate of the first capacitor. The electric field due to faces 2 and 3 and also faces 6 and 7 cancel. Thus, the electric field there is due to face 1 to the right and faces 4, 5, 8 to the left. Therefore,

$$\begin{aligned}0 &= 4\pi k(\sigma_4 - \sigma_1') - 4\pi k(-\sigma_3) - 4\pi k(\sigma_4) - 4\pi k(\sigma_2 - \sigma_2') \\ &= 4\pi k(\sigma_1 - \sigma_2 - 2\sigma_1') \\ &= 8\pi k(\sigma_1 - \sigma_1') \qquad (\text{since } \sigma_1 = -\sigma_2)\end{aligned} \tag{4.86}$$

Thus, we see that $\sigma_1 = \sigma_1'$. Similarly one can show that $\sigma_2 = \sigma_2'$. The outer faces thus, have no charge and all the charge is confined to the inner faces 2 and 7 of the free end plates of the capacitors. We thus, have

$$C_1 V_{d_1} = \sigma_1 A \tag{4.87}$$

and

$$C_2 V_{d_2} = -\sigma_2 A = \sigma_1 A \tag{4.88}$$

where A is the area of the plates.

The potential difference between the terminals of the battery is as shown

$$V = V_1 - V_2 = V_{d_1} + V_{d_2} = \sigma_1 \left(\frac{1}{C_1} + \frac{1}{C_2} \right) \tag{4.89}$$

The relation between the potential difference across the end plates and the charge on the capacitors can thus, be written as

$$C_{\text{eff}} V = \sigma_1 \tag{4.90}$$

where

$$C_{\text{eff}} = \frac{C_1 C_2}{C_1 + C_2} \tag{4.91}$$

A series combination of two capacitors is thus, equivalent to an effective single capacitance given by

$$\boxed{\frac{1}{C_{\text{eff}}} = \frac{1}{C_1} + \frac{1}{C_2}} \tag{4.92}$$

The generalisation to n capacitors connected in series is obvious from this expression and we have

$$\boxed{\frac{1}{C_{\text{eff}}} = \frac{1}{C_1} + \frac{1}{C_2} + \cdots + \frac{1}{C_n}} \tag{4.93}$$

Just as in the parallel combination, the sum of the energies stored in the individual capacitors is equal to the energy stored in the equivalent capacitor. The total energy stored is given by

$$\begin{aligned} \frac{1}{2}C_1 V_1^2 + \frac{1}{2}C_2 V_2^2 + \cdots + \frac{1}{2}C_n V_n^2 &= \frac{1}{2}\frac{Q^2}{C_1} + \frac{1}{2}\frac{Q^2}{C_2} + \cdots + \frac{1}{2}\frac{Q^2}{C_n} + \\ &= \frac{1}{2}\frac{Q^2}{C_{\text{eff}}} \end{aligned} \tag{4.94}$$

These basic expressions for the equivalent capacitance in the parallel and series combination of capacitors can be used to analyse more complicated combinations of the capacitors, as the example below will illustrate.

EXAMPLE 4.9 Determine the effective capacitance of the combination of capacitances as shown in Fig. 4.29.

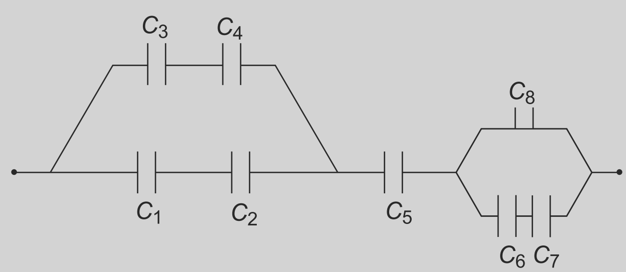


Fig. 4.29 *Example 4.9: Combinations of capacitors*

Solution

As a first step, we combine C_3 and C_4 into an effective capacitance

$$C_{34} = \frac{C_3 C_4}{C_3 + C_4}$$

since they are in series. Similarly, we combine C_1 and C_2 which are in series into

$$C_{12} = \frac{C_1 C_2}{C_1 + C_2}$$

Now capacitances C_{12} and C_{34} are in parallel and therefore the effective capacitance of this combination is

$$C_{1234} = C_{12} + C_{34}$$

Next, we combine C_6 and C_7 into

$$C_{67} = \frac{C_6 C_7}{C_6 + C_7}$$

and combine this effective capacitance with C_8 to get

$$C_{678} = C_{67} + C_8$$

The configuration that we have now is as

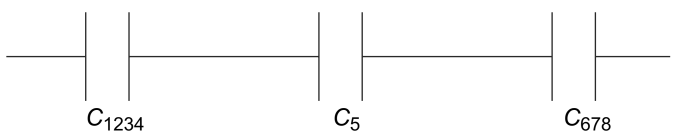


Fig. 4.30 *Combinations of capacitors*

These capacitors, C_{1234}, C_{678} and C_5 are in series and therefore, the effective capacitance is simply

$$\frac{1}{C_{\text{eff}}} = \frac{1}{C_{1234}} + \frac{1}{C_5} + \frac{1}{C_{678}}$$

or

$$\frac{1}{C_{\text{eff}}} = \frac{1}{\frac{C_1 C_2}{C_1+C_2} + \frac{C_3 C_4}{C_3+C_4}} + \frac{1}{C_5} + \frac{1}{C_8 + \frac{C_6 C_7}{C_6+C_7}}$$

PROBLEM 4.10 A capacitor of 1 μF is fully charged by connecting it to a battery of 12 V and then disconnected. Now the capacitor is connected to an uncharged capacitor of capacitance C and the potential difference across it is found to be 4 V. Find C and also the energy stored in the second capacitor.

PROBLEM 4.11 A capacitor of capacitance C_1 is charged to V_1 V and a second capacitor of capacitance C_2 is charged to V_2 V. The positive plate of each is connected to the negative plate of the other one. What is the final charge on the capacitor with capacitance C_1?

PROBLEM 4.12 Ten parallel plate capacitors are connected in series. The distance between the plates of the first capacitor is d while that of the second one is $2d$, third one is $3d$ and so on. All capacitors have the same area. Find the equivalent capacitance of the arrangement in terms of the capacitance of the first capacitor.

4.4.8 Electrolytic Capacitors and Supercapacitors

Ordinarily, capacitors were made by taking two metallic plates and having a dielectric layer between them. The capacitance, as we have seen depends directly on the area of the plates and the dielectric constant of the dielectric and depends inversely on the separation of the plates. There are of course, limits to how large one can make the plates as there are for the dielectric constants of the materials used as dielectrics in the capacitors. One can think of making the separation between the plates very small and thus increasing the capacitance while still retaining the small size which makes it convenient. However, there are limits to this too, since beyond a point, if the two oppositely charge carrying metallic plates are brought closer, the electric field becomes very intense and there is a electrical breakdown with charge flowing from one plate to the other. The dielectric becomes conducting at this point and we do not have a capacitor anymore in the sense that we have been discussing above. This thus limits the value of the capacitance one can have with such capacitors.

Electrolytic capacitors overcome this limitation by replacing one of the plates with an electrolyte, which is also conducting. In such a capacitor, shown in Fig. 4.31, typically we have a metallic plate (which is usually aluminum) coated with a layer of the dielectric (in this case, aluminum oxide). The plate and the layer are then immersed in an electrolyte, which serves as a second plate. Now the advantage of

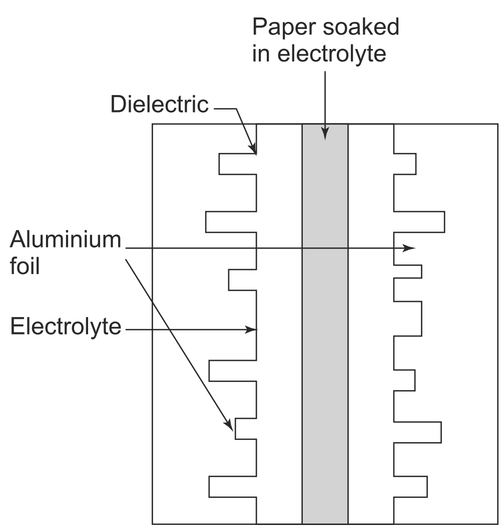


Fig. 4.31 *A typical sketch of an electrolytic capacitor. The surfaces are roughened to increase the effective surface area and hence the capacitance. Capacitances of the order of farads can be obtained using electrolytic capacitors*

the electrolyte is that the liquid electrolyte can indeed come very close to the coated plate. This then gives us an effective separation of the 'plates' which is very small and hence a very large value of the capacitance of such a capacitor.

Recently, another kind of capacitors have become more common—these are called supercapacitors. A supercapacitor is like an ordinary capacitor in that it has two plates which are usually made of activated charcoal. Charcoal being a porous substance has a much larger effective surface area which increases the capacitance. The activated charcoal plates are immersed in an electrolyte. The charged plates polarise the electrolyte with ions migrating towards the plates of opposite charges. Unlike in conventional capacitors where charges accumulate on the plates, in supercapacitors, the charges collect between the surface of the conducting plates and the electrolyte.

Ultimately, we have the ions which have migrated to the conducting plates, forming an electric double layer with the separation of each layer being the order of Angstroms. The schematic sketch of a supercapacitor is shown in Fig. 4.32. The large effective area (because of the use of activated charcoal or any other porous substance) and extremely small separation makes the capacitance of such capacitors extremely large. It is not uncommon to see supercapacitors with capacitances of the order of tens or even thousands of farads. These large capacitances and hence an ability to store huge amounts of energy have made supercapacitors being used in electrical cars and wind turbines, etc., where electrical energy needs to be stored.

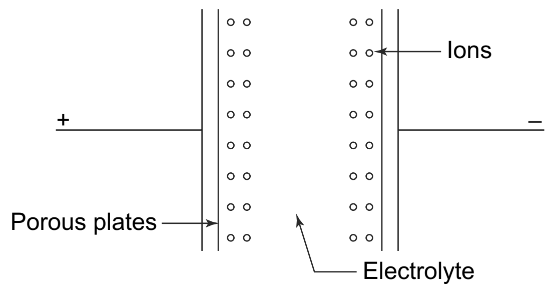


Fig. 4.32 *Schematic sketch of a supercapacitor*

SUMMARY

- A type of material called 'conductors' have a special property that they have electrons in it that are almost 'free' and respond to electric fields by moving.
- The movement of such electrons inside a conductor make it special. There is no electric field inside it and the entire conductor is at the same potential when placed in an external steady electric field.
- The movement of free electrons results in the steady state with free charges being present only at the surface of a conductor.
- If there is a cavity inside a conductor without any charge being placed there from outside, there is no electric field there. The potential there is the same as the bulk conductor which encloses it. All free charges lie on the outer surface of such a conductor.
- When a single conductor is there carrying a charge, its potential is proportional to its charge. When several charge carrying conductors are present, their potentials are linearly related to the charges through the 'coefficients ' of potential. Conversely, the coefficients relating the charges to the potentials form a 'capacitance matrix'.

- A pair of conductors carrying equal and opposite charges is usually called a ‘capacitor’. They are useful devices for storing energy.

CONCEPTUAL QUESTIONS

1. When an external electric field is applied to a conductor
   a. all electrons in it move freely
   b. all the electrons attached to some of the atoms move
   c. some of the outer electrons of all atoms move
   d. it is the core electrons of atoms which move

2. In the presence of an steady electric field, the free electrons in a conductor
   a. are always moving due to the field
   b. move to the surface
   c. move in circular orbits
   d. create additional electric field inside, increasing the strength of the applied field.

3. If there is a cavity inside a conductor and an external electric field is applied, which of the following statements is correct?
   a. The free electrons will be continuously moving inside the cavity.
   b. The value of the electric field in the cavity would be the same as the external electric field.
   c. The potential in the cavity will be the same at all points but that need not be the potential in the conductor which encloses the cavity.
   d. If no external charge is present in the cavity, free charges may be present in the inner surface between the cavity and the rest of the conductor.

4. If many charge carrying conductors are there separated from each other
   a. Their potentials will all be necessarily equal
   b. The electric field inside the conductors need not vanish
   c. The free electrons can move between the conductors
   d. The charges on each them must be the same
   e. Surface charges will be present in all the conductors

5. A parallel plate capacitor stores twice as much charge with a given dielectric than with air. The susceptibility of the dielectric is
   a. 0
   b. 1
   c. 2
   d. 3
   e. 4

6. A metal sphere of radius a has a charge Q on its surface. If the surface of the sphere is at a potential V, what is the energy stored by the sphere?

7. A point charge +Q is placed inside a neutral, hollow, spherical conductor. As the charge is moved around inside, the electric field

a. is zero and does not change
b. is non-zero but does not change
c. is zero when centred but changes
d. is non-zero and changes

8. A parallel-plate capacitor, disconnected from a battery, has plates with equal and opposite charges, separated by a distance d. Suppose the plates are pulled apart until separated by a distance $D > d$. How does the final electrostatic energy stored in the capacitor compare to the initial energy?

a. The final stored energy is smaller
b. The final stored energy is larger
c. Stored energy does not change

9. A parallel plate capacitor is charged to a total charge Q and the battery removed. A slab of material with dielectric constant K is inserted between the plates. The charge stored in the capacitor

a. increases
b. decreases
c. stays the same
d. Cannot say from the information given.

10. In Question 9, what happens to the energy stored in the capacitor?

a. Increases
b. Decreases
c. Stays the same
d. Cannot say from the information given

PROBLEMS

1. A parallel plate capacitor with the plate separation d is charged to carry a charge Q and a potential V and then disconnected from the charge source. The plate separation is now halved to $\frac{d}{2}$.What happens to the charge on the plates, the electric field, the potential and the energy stored in the field?

2. Consider two concentric thin, conducting shells separated by a vacuum. The radius of the inner shell is a and the outer shell is b. A charge $+Q$ is placed on the inner shell and a charge $-Q$ on the outer shell.

a. What is the electric field everywhere?
b. What is the capacitance of such an arrangement?

c. Assume that the outer radius is increased from b to $2b$ while keeping the charges constant. What is the change in potential energy?

3. Two coaxial conducting cylinders of radius a and b, $(a < b)$ and length h are charged so that the outer surface of the inner cylinder has a charge $+Q$ and the inner surface of the outer cylinder has charge $-Q$.

a. Calculate the field in the region $a < r < b$.
b. Calculate the potential difference between the two cylinders.
c. Calculate the capacitance of such an arrangement.
d. Find the energy density at any point between the cylinders. Find the total energy in a cylindrical shell of radius r, thickness dr and length h and integrate it to obtain the total electrostatic energy in the system. Compare it with the energy obtained by using the capacitance and voltage found in part (b) and (c).

4. A parallel-plate capacitor has fixed charges $+Q$ and $-Q$. The separation of the plates is then doubled.

a. By what factor does the energy stored in the electric field change?
b. How much work must be done if the separation of the plates is doubled from d to $2d$? The area of each plate is A.

Consider now a cylindrical capacitor with inner and outer radii a and b, respectively.

a. Suppose the outer radius b of a cylindrical capacitor is doubled, but the charge is kept constant. By what factor would the stored energy change?
b. Repeat (c), assuming the voltage remains constant.

5. A dielectric slab of thickness b and dielectric constant K is placed between the plates of a parallel plate capacitor of area A and plate separation d. A potential difference of V_0 is applied when there is NO dielectric present. The battery is disconnected and then the dielectric slab is inserted. Calculate

a. the capacitance before the slab is inserted
b. the free charge
c. the electric field in the gap
d. the electric field in the dielectric

With $A = 100$ sq cm., $d = 1$cm, $b = 0.5$ cm, $V_0 = 100$ volts and $K = 7.0$, calculate the above quantities.

6. Consider a parallel plate capacitor of capacitance C which is connected to a battery of EMF V and charged fully. The battery is then disconnected and the plates are pulled apart a further distance x such that the potential difference between the plates changes by a factor of 4.

a. Did the potential increase or decrease?
b. How did the field change by pulling apart the plates?
c. How did the energy stored in the capacitor change in this process?
d. A dielectric of dielectric constant K is now introduced to completely fill the space between the plates. How does the energy change?
e. Find the volume of the capacitor?

7. A parallel-plate capacitor is constructed by filling the space between two square plates with blocks of three dielectric materials, as in Fig. 4.33. Assume that $l \gg d$ and find the capacitance of the device in terms of the plate area A and d, K_1, K_2, and K_3.

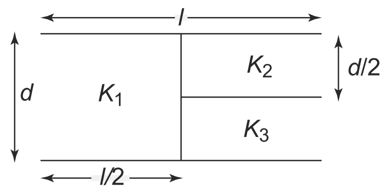


Fig. 4.33 *Problem 7*

8. Consider two nested spherical shells (Fig. 4.34)—the inner one with inner radius a and outer radius b and the outer one with inner radius c and outer radius d. The spheres are initially uncharged. Suppose that the spheres are floating i.e., their net charge remains fixed. Now introduce a charge $+Q$ at the centre of the inner spherical shell. Determine the charge on each of the four surfaces of the spherical shells as well as the electric field and potential everywhere as a function of r, the distance from the common centre. Take the zero of potential to be infinity.

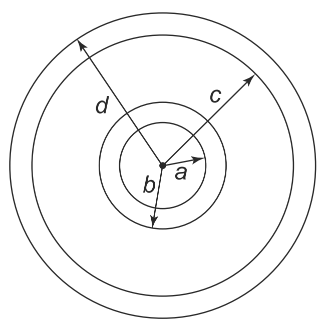


Fig. 4.34 *Problem 8*

9. A capacitor C_1 is charged to a potential difference V_0. The charging battery is then removed and the capacitor is connected to an uncharged capacitor C_2 in parallel as in Fig. 4.35.

a. What is the final potential difference across the combination of two capacitors?

b. What is the initial and final stored energy?

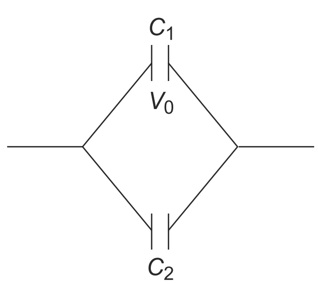


Fig. 4.35 *Problem 9*

10. Consider a parallel plate capacitor which is filled with polystyrene with $K = 2.55$ as in Fig. 4.36. The distance between the plates is 1.5 mm and the field intensity is 10 kV/m. Find D, P, the surface charge density of free charges on the plates, the potential difference between the plates.

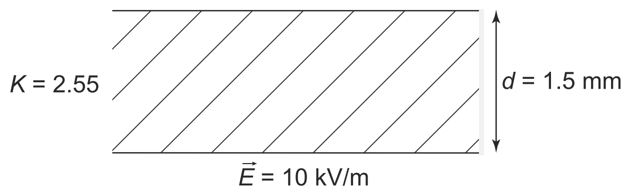


Fig. 4.36 *Problem 10*

11. In Fig. 4.37, all capacitors are of equal capacitance and the voltage of the battery is 120 V. The capacitors are initially uncharged. Consider two cases:

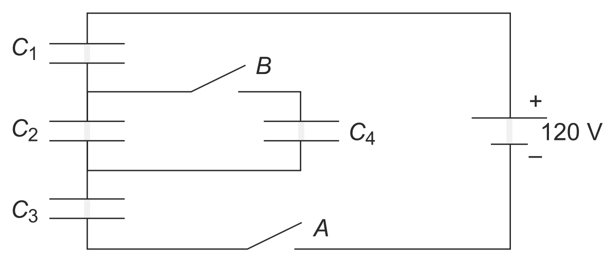


Fig. 4.37 *Problem 11*

Case I: (a) Switch B is open and switch A is closed and then opened after C_1, C_2, C_3 are completely charged. Determine the potential difference across each capacitor.
(b) Now switch B is closed. What is the potential difference across each capacitor now?
Case II: (a) Switch A is kept open and switch B is closed. What is now the potential across each capacitor?
(b) Switch A is now closed. What is the potential in this case now?

12. Consider two conducting concentric spherical shells, of radii a and b where $b > a$ maintained at a potential difference of V_0. The potential at the outer sphere is zero while the potential on the inner sphere is V_0. This calculation of the capacitance of two concentric spherical shells can also be done if instead of the potential difference between the two shells, we know the charges on the shells. In this case, show that we obtain the same value of the capacitance namely

$$C = \frac{Q}{V_0} = \frac{K}{k}\frac{1}{\frac{1}{a} - \frac{1}{b}}$$

13. A solid metallic sphere of radius r_1 is placed concentrically inside a conducting spherical shell of inner radius r_2 and outer radius r_3. A charge $+q_1$ is placed on the inner sphere and a charge

$-q_2$ is placed on the outer conductor and the space between the sphere and the shell is filled with a dielectric of dielectric constant K. Find $\vec{E}$, $\vec{D}$, $\vec{P}$ everywhere.

14. A parallel plate capacitor with plates of length b and width w and separation between the plates d is filled with a solid dielectric of permittivity ε (Fig. 4.38). The plates have a constant potential difference V. If the dielectric is now pulled along the length such that a length x remains between the plates, find the force pulling the dielectric block back into the capacitor.

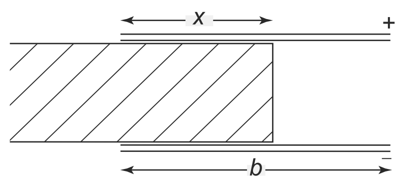


Fig. 4.38 *Problem 14*

15. A metal sphere of radius r_2 has a uniform surface charge distribution with a total charge Q. The permitivitty of the surrounding region is $\varepsilon = \varepsilon_0 \left(1 + \frac{r_2}{r}\right)$. Show that the potential in the surrounding dielectric is

$$V(r) = \frac{Q}{4\pi\varepsilon_0 r_2} \ln(1 + \frac{r_2}{r})$$

16. Consider a cylindrical capacitor with radii a and b and length L (Fig. 4.39). Suppose the space inside the capacitor is filled with an inhomogeneous dielectric with a dielectric constant given by $K = \frac{10+\rho}{\rho}$. Find the capacitance.

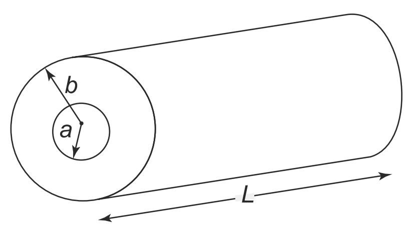


Fig. 4.39 *Problem 16*

17. In a parallel-plate capacitor with a separation d between the plates of area A, dielectric in between is not of uniform permittivity. It varies as $\varepsilon = \varepsilon_0 \varepsilon_r \left(1 + \frac{x}{d}\right)\left(1 + \frac{2x}{d}\right)$, where x is the distance from one of the plates. Calculate the capacitance of the capacitor. Determine the volume and surface polarisation charges if the plate at $x = 0$ is charged with a charge $Q(Q > 0)$, and the other with $-Q$.

18. A spherical capacitor consisting of a thin outer spherical shell of radius R_1 and an inner one of radius R_2, has two dielectrics in the space between them as shown in Fig. 4.40. In the lower half the permittivity is ε_2 and similarly, the upper half has a dielectric of permittivity ε_1. Calculate the electric field and displacement vectors at points within and outside the capacitor and the capacitance of the capacitor.

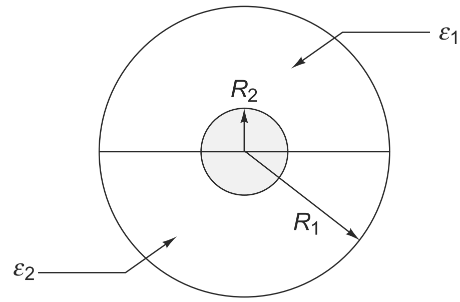


Fig. 4.40 *Problem 18*

19. Consider the following geometries:

a. Two, large flat conducting plates of area A separated by a distance d

b. Two concentric, conducting spheres of radii a, b, with $b > a$

c. Two coaxial, conducting cylinders of length L and radii a, b with $b > a$ and $L >> b$.

For each of these cases, calculate the total electrostatic energy in terms of the equal and opposite charges $\pm Q$ placed on the two conductors. Sketch the energy density of the electrostatic field as a function of the appropriate linear coordinate.

20. Two semi-infinite conducting planes at $\phi = 0$ and $\phi = \phi_0$ are separated by an infinitesimal insulating gap where they meet as in Fig. 4.41. The first plane is maintained at $V = 0$ and the second one at $V = V_0$. Find the potential and electric field everywhere between the planes.

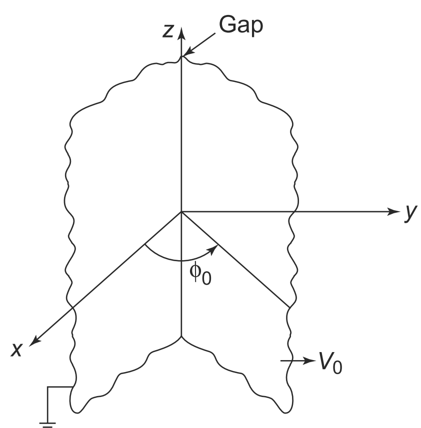


Fig. 4.41 *Problem 20*

21. Two conducting cones of infinite extent with $\theta = \theta_1$ and $\theta = \theta_2$ are separated by an infinitesimal gap at $r = 0$ (Fig. 4.42). If the first cone is maintained at $V = 0$ and the second one at $V = V_0$, then find the potential and the electric field between the cones.

22. An infinite line charge with charge density ρ is located at a distance h from a grounded conducting plane at $z = 0$ as shown in Fig. 4.43. Find the potential and electric field for $z > 0$ and also the induced surface charge.

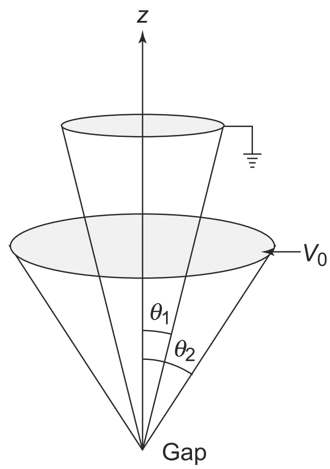


Fig. 4.42 *Problem 21*

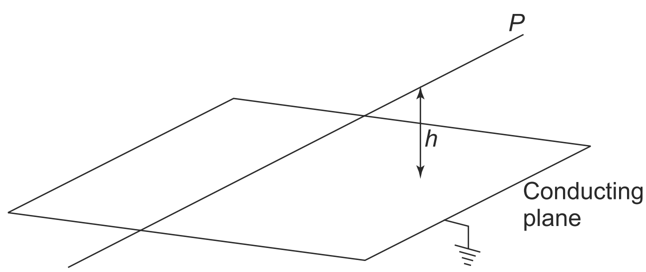


Fig. 4.43 *Problem 22*

23. A parallel plate capacitor of very large area A has the region between the plates completely filled up with a mixture of two dielectric liquids. $\frac{2}{3}$ part of the volume is occupied with one of permittivity ε_1 and density ρ_1 and the rest by one with permittivity ε_2 and density $\rho_2 < \rho_1$. The densities are mass densities. Consider three positions

a. the plates horizontal with the heavier liquid below and the lighter one on top at rest
b. the same horizontal position, but imagine there is no gravity and the liquid is stirred violently and
c. the plates and the gap vertically aligned with the heavier liquid below and the lighter one on top due to buoyancy.

Calculate the capacitance in all three cases.

24. A parallel plate capacitor of large area A and a separation d between the plates has a dielectric of permittivity ε between the plates. Initially, there are charged $+Q$ and $-Q$ on the plates and the dielectric is fully in between the plates. The dielectric is then pulled so that only half the area of the plates have the dielectric between them. Calculate the amount of work done.

5

Electric Currents

Learning Objectives

- To be able to understand the concept of electric current and current density in terms of motion of charges.
- To learn about the way in which a steady current is produced using an electrolytic cell.
- To comprehend the simple model for resistivity and to calculate the resistance of conductors with simple geometries.
- To learn about the heating effect of current.
- To be able to use Kirchhoff's laws to calculate the effective resistance for various combinations of resistances.
- To learn about Network theorems and use them to find out the currents, voltages, etc., in various circuits.
- To comprehend the concept of a variable current and learn about the RC circuit.

5.1 CURRENT DENSITY

In the previous chapters, we have studied the behaviour of electric charges under various conditions. However, in all our discussion, we restricted ourselves to charges which are stationary or in other words, only studied electrostatic phenomena. Clearly, since charges experience a force in an electric field, they would obviously also be in motion in certain circumstances. It is this study of charges in motion that we now turn to in this chapter.

Electric charges in motion constitute an electric current. Although we know that charge particles move under the influence of an electric field, not all motion is necessarily due to the presence of electric fields. Thus, for instance, cosmic rays consist of charged particles and their bombardment of the earth from outer space constitutes a current. Similarly, if charged particles are falling under the influence of a gravitational field (for instance, the charged drops in the Millikan's Oil drop experiment), then these constitute an electric current.

We have already seen that the laws of electrostatics which govern the behaviour of static charges are formulated in terms of the relationships between fields and charges and charge densities. These quantities are insufficient to describe charges in motion and we need another quantity. For studying electric currents, we use the basic measure of currents which is called current density, $\vec{j}$.

Consider a set of charges in motion around a point P with a coordinate $\vec{r}$. Let n_- and n_+ be the number density of negative and positive charges at P respectively. Similarly let $\vec{v}_-$ and $\vec{v}_+$ be the respective velocities of these charges at P. This is shown in Fig. 5.1.

Consider now an infinitesimal surface ΔS with outward normal $\hat{n}$ around P (Fig. 5.2). Charges which are in motion will cross the surface area and we want to find out which of the charged particles will cross the surface in a time Δt

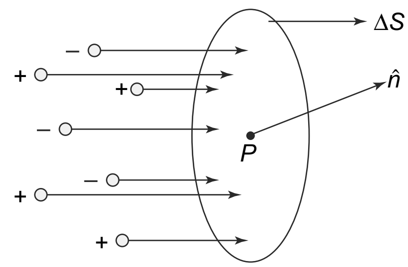


Fig. 5.1 *Charges, both positive and negative in motion around a point P*

Particle A shown in Fig. 5.2, travels a distance $v\Delta t$ along the direction of its velocity vector in time Δt. Consequently, its perpendicular distance from the surface is reduced by $(\vec{v}\cdot\hat{n})\Delta t$ but that is not enough to make it cross the surface. Particle B on the other hand, also travels a distance $(\vec{v}\cdot\hat{n})\Delta t$ but that is enough for it to cross the surface. Imagine a pill box with ΔS as one of the surfaces with the other surface been drawn parallel to it but removed by a distance $(\vec{v}\cdot\hat{n})\Delta t$. Now all the particles inside the pill box would, like B, cross the surface ΔS in time Δt while particles outside it, like A, will not be able to cross. The number of particles inside the pill box would simply be the number density times the volume of the pill box itself. The volume of the pill box is simply $\Delta S(\vec{v}\cdot\hat{n})\Delta t$. There are $n_-(\Delta S(\vec{v}\cdot\hat{n})\Delta t)$ negatively charged and $n_+(\Delta S(\vec{v}\cdot\hat{n})\Delta t)$ positively charged particles inside the pill box which will cross the surface in time Δt. Thus, the amount of charge crossing the surface ΔS in the direction of $\hat{n}$ is

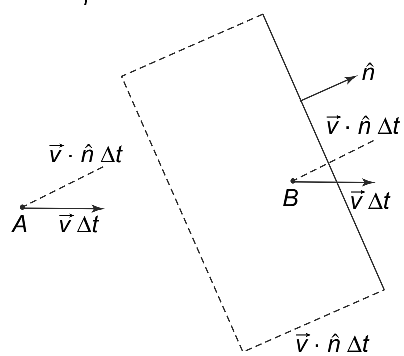


Fig. 5.2 *Charges crossing the surface area ΔS in time Δt. Particle A does not cross the surface while particle B does*

$$\Delta Q(\hat{n}) = e\Delta S(n_+(\vec{v}_+\cdot\hat{n}) - n_-(\vec{v}_-\cdot\hat{n})\Delta t) \tag{5.1}$$

We had of course implicitly assumed in Fig. 5.2 that the velocity $\vec{v}$ is such that the charges cross the surface ΔS from left to right. However, if $\vec{v}$ is such that $\vec{v}\cdot\hat{n}$ is negative, then the charges will cross the surface in the other direction. A charge $+q$ crossing along $-\hat{n}$ is the same as a charge $-q$ crossing along $\hat{n}$. Thus, Eq. (5.1) is equally valid in this case. Now this equation can be rewritten as

$$\boxed{\vec{j}(\vec{r}) = e(\vec{v}_+(\vec{r})n_+ - \vec{v}_-(\vec{r})n_-)} \tag{5.2}$$

The rate at which charge flows in the direction of $\hat{n}$ across ΔS is given by

$$\frac{\Delta Q(\hat{n})}{\Delta t} = \Delta S(\vec{j}\cdot\hat{n}) \tag{5.3}$$

The vector quantity $\vec{j}(\vec{r})$ in Eq. (5.2) is called the current density vector at $\vec{r}$. Equation (5.2) can of course also be written in terms of the charge density $\rho(\vec{r})$. Since

$$\begin{aligned}\rho(\vec{r}) &= \rho_+(\vec{r}) + \rho_-(\vec{r})\\ \rho_+(\vec{r}) &= en_+\\ \rho_-(\vec{r}) &= -en_-\end{aligned}$$

Equation (5.2), for the case of charge density instead of point charges, is thus easily seen to be equivalent to

$$\boxed{\vec{j}(\vec{r}) = \vec{v}_+(\vec{r})\rho_+(\vec{r}) + \vec{v}_-(\vec{r})\rho_-(\vec{r})} \tag{5.4}$$

The current density $\vec{j}(\vec{r})$, though a useful quantity, is sometimes not adequate to describe current. We need to introduce another quantity, called the total current I which is defined as follows: we consider a surface of finite area S with points with coordinates $\vec{r}$ on it. The total current I across S such that at every point the current is along the locally defined normal $\hat{n}(\vec{r})$ at that point is then

$$\boxed{I = \iint\limits_S \vec{j} \cdot \hat{n}(\vec{r}) dS} \tag{5.5}$$

where the integral is over the surface S.

The current density $\vec{j}(\vec{r})$ can, in general be time dependent. It also satisfies another important property. Consider an arbitrary volume V enclosed by a surface S as in Fig. 5.3.

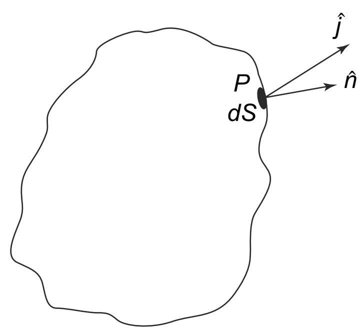


Fig. 5.3 *Charge flow across the bounding surface S of a volume V. $\vec{j}$ is the current density at the infinitesimal surface element as shown. $\hat{n}$ is the normal to the surface at that point. The rate charge flow through dS is $(\vec{j} \cdot \hat{n})dS$*

If we consider an infinitesimal surface element dS on S with normal $\hat{n}$ and coordinates $\vec{r}$ where the current density is $\vec{j}(\vec{r}, t)$, then the outflow of charge across dS along $\hat{n}$ in time Δt is

$$\text{outflow of charge} = (\vec{j}(\vec{r}, t) \cdot \hat{n}) dS \Delta t \tag{5.6}$$

Integrating over the entire surface will give us the net outflow of charge from V across the surface S

$$\Delta Q_{\text{out}} = \Delta t \iint\limits_S (\vec{j}(\vec{r}, t) \cdot \hat{n}) dS \tag{5.7}$$

The total charge enclosed by V is obviously related to the charge density ρ as

$$Q = \iiint\limits_V \rho(\vec{r}, t) d^3\vec{r} \tag{5.8}$$

Equation (5.7) can be rewritten using the Divergence theorem as

$$\Delta Q_{\text{out}} = \Delta t \iiint\limits_V \vec{\nabla} \cdot \vec{j}(\vec{r}, t) d^3\vec{r} \tag{5.9}$$

We know that conservation of charge implies that charge cannot be created or destroyed. Hence, ΔQ_{out}, the charge flowing out of the volume V must equal $-\Delta Q$, the negative of the charge inside V. Thus,

$$\Delta t \iiint_V \vec{\nabla} \cdot \vec{j}(\vec{r}, t) d^3\vec{r} = -\iiint_V \Delta\rho(\vec{r}, t) d^3\vec{r} \tag{5.10}$$

Taking the limit that Δt and $\Delta\rho$ are infinitesimal, we get

$$\frac{\Delta\rho}{\Delta t} = \frac{\partial\rho}{\partial t}$$

With this, Eq. (5.10) can be rewritten as

$$\iiint_V \left(\vec{\nabla} \cdot \vec{j}(\vec{r}, t) + \frac{\partial\rho(\vec{r}, t)}{\partial t}\right) d^3\vec{r} = 0 \tag{5.11}$$

Since the volume over which the integration is being done is totally arbitrary, it follows that the integrand must vanish. Hence,

$$\boxed{\vec{\nabla} \cdot \vec{j}(\vec{r}, t) + \frac{\partial\rho(\vec{r}, t)}{\partial t} = 0} \tag{5.12}$$

This equation is called the **Equation of Continuity.** In the case of the flow of current in conductors, there is no accumulation or reduction of charges inside the conductor and hence $\frac{\partial\rho}{\partial t}$ vanishes. In this case, we get

$$\boxed{\vec{\nabla} \cdot \vec{j}(\vec{r}, t) = 0} \tag{5.13}$$

The Equation of Continuity, in both forms (Eqs. (5.12 and 5.13)), plays a very important role in electromagnetism, as we shall see later.

PROBLEM 5.1 The cross section of a long rod, placed with its length along the z-axis, is rectangular with the larger sides (AB and CD in the Fig. 5.4) of length $2a$ along the x-axis and the shorter ones (AC and BD) of length $2b$ along the y-axis. The current is flowing along the length of the rod. The current density $\vec{j}(x, y)$ is not uniform. It has a maximum magnitude j_0 at the axis of the rod passing through the centre of the rectangle O, which is the origin. The current density decreases linearly with the square of distance from O becoming zero at the four corners A, B, C and D. Express the current density as a vector and obtain the total current flowing through the rod.

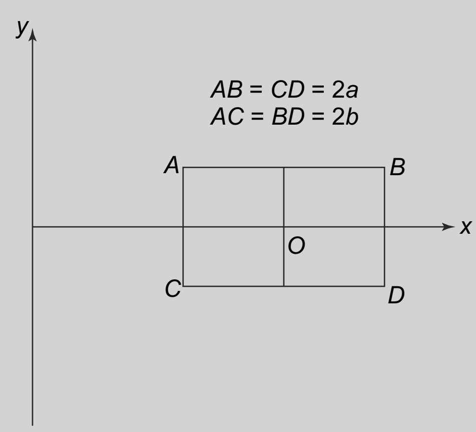


Fig. 5.4 *Problem 5.1*

Answer:

$$\vec{j}(x, y) = j_0 \left(1 - \frac{x^2 + y^2}{a^2 + b^2}\right) \hat{k}$$

Total current

$$\int |\vec{j}(x, y)| dx dy = \frac{8ab}{3} j_0$$

PROBLEM 5.2 A thick ring has an inner radius R_1 and outer radius R_2. The ring is placed with its axis exactly in the $x - y$ plane at $z = 0$ and the centre at the origin. A current is flowing through the the ring. The current density is maximum at the central axis equal to $\vec{j}$ and decreases uniformly to zero at periphery of the ring. Express the current density as a vector and obtain the total current flowing through the ring.

PROBLEM 5.3 A rod of length L is uniformly charged with a line charge density of λ. Its centre is at the origin and it is rotating in the $x - y$ plane about the z-axis with its centre as the pivot with a constant angular speed ω. Calculate the vector expression for the current density.
[*Hint: obtain the charge density first at time t; it obviously contains Delta function*]

5.2 ELECTROLYTIC CELLS

For charges to be in motion, we need to have an electric field or, which is the same thing, a potential difference which is maintained. The first attempts to build a device which constantly supplies charges to an external conductor used the chemical properties of certain substances.

We know that chemical reactions between atoms and molecules take place through the electrons in the outermost shells, the so called valence electrons. Elements can be divided broadly into two categories- **electropositive elements are those which readily give up one or more electrons and electronegative elements which easily take in electrons.** These definitions are, of course, qualitative but one can devise a quantitative scale or index ordering these elements such for any given pair, one knows which one is more electronegative than the other. In general, metals are the most electropositive elements and halogens (chlorine, bromine etc) are the most electronegative. Table 5.1 shows the electronegativity of common elements.

Table 5.1 *Electronegativity of common elements (Pauling Scale)*

Element	Electronegativity
Hydrogen	2.20
Carbon	2.55
Fluorine	3.98
Copper	1.90
Zinc	1.65
Chlorine	3.16
Silver	1.93
Gold	2.54

The electropositive or electronegative nature of substances leads to an interesting phenomenon. When, for instance, two conductors of different electropositive (or electronegative) indices are brought into contact, because of their differing propensities to accept or donate electrons, charge will flow from one to another. This flow of charge, over time, will lead to a build up of a potential difference between the two conductors, much like a capacitor. The first device to supply charges to an external circuit, a Voltaic cell, exploits this property. It was the Italian scientist Alessandro Volta (1745–1827) who built the first such device called the Voltaic pile.

The basic or simplest cell is a primary cell in which two plates (typically metallic) called electrodes are immersed in an electrolyte. Electrolytes are usually ionic compounds in aqueous solution. We have already seen in Chapter 3 that water is an excellent solvent because of the nature of its molecules. The ionic compounds in an aqueous solution break up into their ionic components. These ionic components then are free to move in the solution. In the presence of an electric field, these ionic components will experience a force and move and create a current, just like electrons in a conductor. To understand the basic principles of a primary cell, let us consider the type which is most commonly used.

In the most commonly used type of a primary cell, the electrodes are a rod of Zinc (Zn) and another of Copper (Cu). These electrodes are immersed in an electrolyte which is a dilute solution of sulphuric acid (H_2SO_4). The sulphuric acid breaks up into positively charged $2H^+$ and negatively charges SO_4^{--} ions. The zinc electrode has a Zn^{++} ion and two electrons. The Zinc ion combines readily with the SO_4^{--} ion via the reaction

$$Zn^{++} + SO_4^{--} \rightarrow ZnSO_4$$

to form neutral $ZnSO_4$. The zinc plate, having shed the positive Zn^{++} ion, now becomes negatively charged. Can this process continue indefinitely? The answer is clearly no, since the Zn^{++} ion which has been released from the plate will repel other Zn^{++} ions. In addition, the positive ions which are released will get attracted back by the now negatively charged zinc plate. At equilibrium, the number of zinc ions released by the plate equal the number of ions attracted back. The potential difference between the negatively charged zinc electrode and the electrolyte at equilibrium can be measured and is known to be -0.62 V.

On the other hand, the copper plate being more electronegative, easily releases two electrons from the outermost shell of the atoms. These electrons combine with the $2H^+$ ions present in the electrolyte to form neutral hydrogen. The copper plate is positively charged since it is losing electrons and at equilibrium, has a potential difference of 0.46 V relative to the electrolyte. Thus, at equilibrium, the copper plate, called the anode is at a potential $1.08(1.08 = 0.46 - (-0.62))$ V relative to the Zinc plate which is called the cathode. **This potential difference is called the electromotive force or EMF of the primary cell.** This name, though inappropriate, has been used since electrolytic cells were first introduced.

We now have an arrangement of a primary cell, with two electrodes (called the anode and the cathode) which have a potential difference of 1.08 V between them. If the two electrodes are now connected with a conductor, then an electric field will be set up inside the conductor since its two ends will be at different potential. We have already seen that in such a situation, an electric current will flow in the conductor, with the electrons moving from the negatively charged cathode (zinc plate) to the positively charged anode (copper plate). This current is carried by the negatively charged electrons which are moving from the zinc plate to the copper plate, though the direction of the current (because of the negative charge on the electrons), will be from the anode to the cathode which is also the direction of the electric field. The zinc plate thus, loses some electrons and the magnitude of its potential falls. The equilibrium which had existed between the zinc plate and the electrolyte is disturbed and to restore it, more zinc ions move to the electrolyte where they combine with the SO_4^{--} ions. For every one SO_4^{--} ion which recombines with the zinc ion, there are left behind $2H^+$ ions which move to the copper plate (anode). The copper plate has now more electrons than at equilibrium since excess electrons have flowed from the zinc plate through the external conductor which is connected to both the zinc and copper plates. The hydrogen ions combine with the electrons at the anode to form neutral hydrogen, thereby trying to restore equilibrium.

In this process, a cycle is completed whereby electrons flow from zinc to the copper plate through the external conductor and hydrogen ions, which are formed near the zinc plate in the electrolyte solution, move towards the copper plate to take in the electrons and form neutral hydrogen. The external conductor thus, has a current flowing from the copper to the zinc plate (and electrons in the opposite direction). This process continues till the zinc plate, which is releasing zinc into the electrolyte, gets depleted completely as also when the H_2SO_4 gets completely exhausted.

Thus, a primary cell basically is a device which creates a current in the external circuit. This external current produces energy in the form of Joule heating, which we will study later in the Chapter. The movement of the ions within the electrolyte maintains the potential difference between the two plates which ultimately drives the current. Note that the direction of the current is from the anode to the cathode externally and from the cathode to the anode (through the H^+ ions) inside the cell. The ultimate energy source is the chemical energy released in the reaction

$$Zn^{++} + SO_4^{--} \rightarrow ZnSO_4$$

which causes the ions to flow and hence cause an electric current in the external circuit also. **Thus, a primary cell is a device to convert chemical energy into electrical energy.**

A device commonly used is **a combination of primary cells called a battery** (Fig. 5.5).

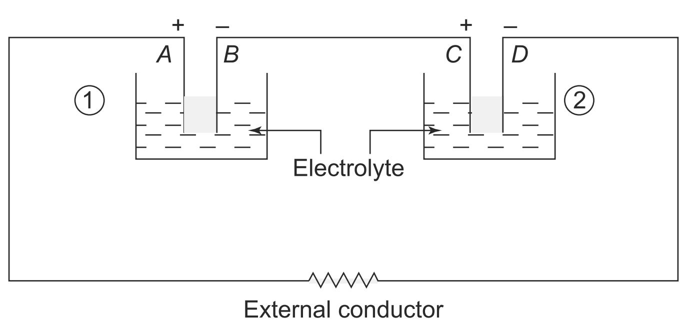


Fig. 5.5 *A battery containing two primary cells*

If there was no external conductor, then the EMF between A and B of primary cell 1 will be $\mathcal{E}_1$. Similarly, the potential difference between C and D of primary cell 2 will be $\mathcal{E}_2$. If B and C are connected as shown, then they will be at the same potential and hence, the potential difference between A and D

will be $\mathcal{E}_1 + \mathcal{E}_2$. Thus, we have an arrangement where the effective EMF of the device is the sum of the EMFs of the individual primary cells. Such a device is called a battery.

Since the zinc plate and the electrolyte are both getting degraded constantly in the primary cell described above, it is clear that a primary cell has a finite life and cannot be resurrected once the components are depleted. Several choices of electrolytes and plate materials have been used to devise cells of different EMFs as well as of extended lifetime. Thus, for example, one of the earliest primary cells called the Leclanche cell uses ammonium chloride as an electrolyte, carbon as anode and zinc as the cathode.

In some cells, called dry cells, the electrolyte is used not in the form of an aqueous solution but instead in the form of a paste or a gel. In a lithium cell, for example, we have manganese dioxide as the anode, lithium as the cathode and a salt like lithium perchlorate in a solvent like propylene carbonate as the electrolyte.

A particularly useful variant of the primary cell is a secondary cell. This is a device in which the chemical reaction which the electrolyte undergoes can be reversed by passing a current through it in the opposite direction. These kind of cells can thus, be charged by connecting them to an external source of current which is made to pass in the reverse direction. After charging, the electrolyte gets more or less restored to its original condition and the device can be used again.

5.3 CURRENTS IN CONDUCTORS: OHM'S LAW

An external electric field causes a flow of charge in a conductor. The flow of charge of current is described by a vector quantity called the current density $\vec{j}(\vec{r})$. It seems reasonable to assume that there ought to be a relationship between the external electric field and the current density. Indeed, such a relationship exists and is known as Ohm's Law.

We have already seen that when a conductor is placed in an electric field, the free charges in the conductor move in such a way to the surface of the conductor such that in equilibrium, there is no net electric field in the body of the conductor. Let us suppose, we arrange to put a charge $+Q$ and a charge $-Q$ on either ends of a conductor (Fig. 5.6).

Immediately, an electric field will be established from $+Q$ to $-Q$. As a result, the free electrons inside the conductor will react to the electric field and almost instantaneously annul the charges on either side. This situation will thus, result in no net field inside the conductor as also no flow of charge except instantaneously. Now suppose an external device is connected to the two ends of the conductor such that it replenishes the charges at the ends A and B, in such a way that the charges remains constant or even vary with time in a definite way (Fig. 5.6(b)). In this situation, an electric field will persist inside the conductor. As a result of this electric field, the free electrons inside the conductor will experience a force and thus move, causing a current to flow.

The relationship between the electric field inside a conductor and the current density which it causes, is a law named after the German physicist, Georg Ohm (1789–1854) and is called **Ohm's Law.**

$$\boxed{\vec{j}(\vec{r}) = \sigma \vec{E}(\vec{r})} \tag{5.14}$$

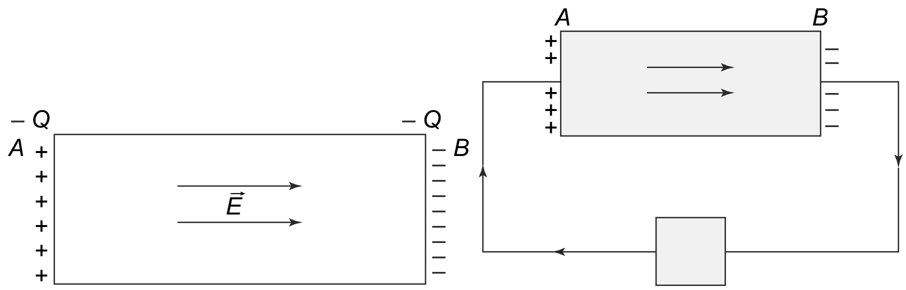


Fig. 5.6 *Currents in conductors (a) Charges $+Q$ and $-Q$ on either ends of the conductor. Left to itself, the free electrons in the conductor will move to neutralise these charges. Thus, the current in the conductor will be instantaneous, (b) An external device exists which transfer $+$ charges from B to A even as the free electrons try to neutralise them. In this situation, an electric field can be maintained inside the conductor and there is a steady flow of charges within the conductor*

where σ is a constant called electrical conductivity of the conductor. It depends on the nature of the material as well as the temperature.

Ohm's Law is obeyed by a wide range of materials over a large range of values of $\vec{E}$. The range of values of σ is also huge, very small for insulators and very high for good conductors (Table 5.2).

Table 5.2 *Conductivity (σ) of common materials*

Material	$(1/\sigma)\Omega m$
Copper	1.68×10^{-8}
Gold	2.44×10^{-8}
Carbon	$5 - 8 \times 10^{-4}$
Glass	$10^{10} - 10^{14}$
Air	$1.3 - 3.3 \times 10^{16}$
Quartz	7.5×10^{17}
Sea Water	0.2

Ohm's Law, however, has limitations and there are many materials like semiconductors where it is, not obeyed. We will not be considering such materials in this chapter.

5.3.1 Ohm's Law: Microscopic Understanding

If one considers the phenomenon of conduction of a current in a conductor at the microscopic level, it seems that Ohm's Law, or the relationship between the current density and the electric field inside a conductor is a bit intriguing. Think of a region in a conductor where $\vec{E}$ is a constant. By Ohm's Law (Eq. 5.14), this would imply that $\vec{j}$ is a constant. Now the current density $\vec{j}$ is related to the velocity of the charge carriers (Eq. 5.2) and hence this implies that a constant electric field should give us a constant velocity. This simple-minded interpretation is obviously in contradiction with the basic laws of motion which tell us that a constant force (the constant electric field gives us a constant force on the charges) should lead to a constant acceleration and NOT a constant velocity! Thus, understanding Ohm's Law at the microscopic level is not straightforward and we need to develop a consistent theory or model to explain it.

One of earliest theories based on classical physics was the **Free Electron Theory** proposed by Paul Drude in 1900. Basically, it is an application of the kinetic theory of gases which had proved to be very successful in thermodynamics. The theory assumes that

a. A conductor is made up of fixed atoms and free electrons. Each fixed atom contributes a certain number of free electrons which are free to move inside the conductor, unlike the fixed atoms. In the presence of an electric field, the electrons experience a force and their motion is governed by Newton's laws of motion.
b. The free electrons, in their motion within the conductor, collide with the fixed atoms. These collisions occur with an average time between collisions τ. τ is called the relaxation time. At any given time t, the last collision occurs $(t - t_{\text{last}})$, i.e., at a time t_{last} before t. It can be shown that if the collisions occur randomly, then the average value of t_{last} is also τ.
c. After each collision, the electron loses its memory of the time previous to the collision. The momentum after the collision has no relationship with the electron's momentum prior to the collision. Drude further assumed that the electron velocities after the collision are like those of particles in thermal equilibrium at a temperature T, which is the temperature of the conductor. A consequence of this assumption is that the average velocity of the electrons after their last collision is zero.

With these assumptions, consider an electron which emerges after a collision with a velocity $\vec{v}_0$. In the presence of an electric field, it experiences an acceleration

$$\vec{a} = -\frac{e\vec{E}}{m}$$

where m is the electron mass. After a collision, consider an amount of time t_{last} spent by an electron without any further collision. The velocity of the electron at such a time would be

$$\vec{v} = \vec{v}_0 - \frac{e\vec{E}}{m} t_{\text{last}}$$

Let us now calculate the average velocity of the electrons at a given time t_0. At this instant, different electrons will have a different value of t_{last}, since the last collision for each electron occurred at a time $t_0 - t_{\text{last}}$, which is different for different electrons. We know that, by our assumptions, the **average value of** $\vec{v}_0$, the velocity with which the electrons emerged after their last collision, is zero. Thus, the average value of the velocity $\vec{v}$ is

$$\begin{aligned} < \vec{v} > &= < \vec{v}_0 > - \frac{e\vec{E}}{m} < t_{\text{last}} > \\ &= -\frac{e\vec{E}}{m}\tau \end{aligned} \tag{5.15}$$

Now using Eq. (5.2) and assuming there are no charge carriers with positive charges (the atoms being fixed in this model), the current density is given by

$$\vec{j} = \left(\frac{e^2 n_- \tau}{m}\right)\vec{E} \tag{5.16}$$

Comparing this with Ohm's Law (Eq. 5.14), we see that in Drude's free electron theory, the conductivity can be expressed as

$$\sigma = \left(\frac{e^2 n_- \tau}{m}\right) \tag{5.17}$$

In the case of electrolytes, where we have seen that there are both positive and negative charge carriers (positive and negative ions), we can also apply this model. In this case, in general the masses of the positive and negative ions, m_+ and m_- respectively would be different. Also the relaxation times, τ_+ and τ_- would be in general different and the conductivity would be

$$\sigma = \frac{n_- q^2 \tau_-}{m_-} + \frac{n_+ q^2 \tau_+}{m_+} \tag{5.18}$$

where q is the magnitude of the charge on the ions. Eq. (5.17) is mostly applicable to currents in conductors with m as the electron mass. Equation (5.18) applies for example for the current flowing through the electrolyte in a cell or battery. The net resistance it produces in the electrolyte is called the internal resistance of the cell or battery.

Thus we see that, in the Free Electron model, it is possible to understand electric current in conductors at the microscopic level. Further, it allows us to express a macroscopic quantity σ in terms of the microscopic parameters of the material, namely the density of charge carriers and their masses as well as the relaxation times.

PROBLEM 5.4 A potential difference of 100 V is maintained between the ends of a copper wire 5 m in length. The mean time between collisions is 3×10^{-14} s. Determine the average velocity of the free electrons in the wire.

PROBLEM 5.5 An aluminum wire of radius 1 mm carries a current 1 mA. Find the current density as well as the average velocity of the electrons in the wire. Given that the density of aluminium is 2700 kg/m^3 and 1 mole of the metal weighs .02698 kg. Assume that each atom of aluminum contributes 1 conduction electron.

PROBLEM 5.6 The particle accelerator at Stanford accelerates a beam of electrons with 2×10^{14} electrons per second. The potential through which the electrons are accelerated is 2×10^{10} V. Calculate the current in the beam.

5.4 RESISTORS AND RESISTIVE CIRCUITS

If we consider a current carrying conductor such that the current density inside it is uniform and the end surfaces are equipotential, then Ohm's Law can be put in a more convenient form. Consider the conductors shown in Fig. 5.7.

The conductor in Fig. 5.7(a) has uniform current density inside. Its end surfaces are at potentials V_+ and V_- and the current density and the electric field are constant and in the direction of its length. The

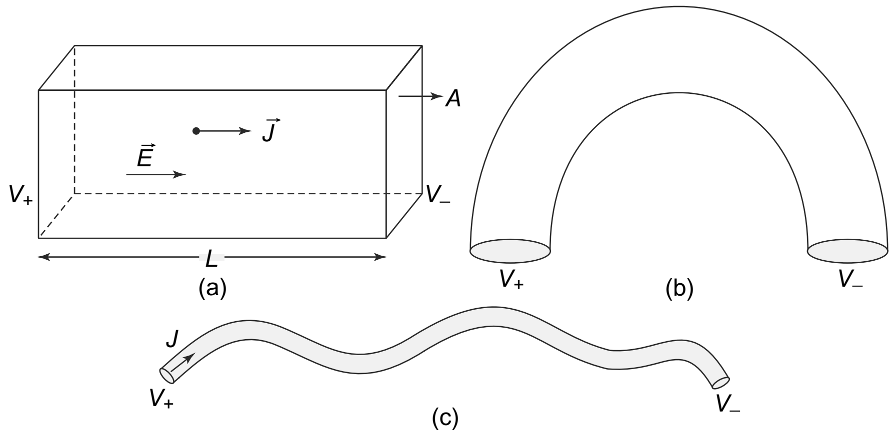


Fig. 5.7 *(a) A slab-shaped conductor of length L and cross-sectional area A. The current and the electric field are both parallel to the length of the conductor, (b) A thin bent wire of circular cross section. The current is not the same at all points, (c) A very thin bent wire. The current density can be considered constant everywhere inside the wire*

total current flowing across the conductor is hence,

$$\iint_{A} \vec{j} \cdot d\vec{S} = jA = I \tag{5.19}$$

The potential difference between the two ends is simply

$$V_d = V_+ - V_-$$

But this is related to the electric field inside it as

$$V_d = \int \vec{E} \cdot d\vec{l} = EL \tag{5.20}$$

Now we can use Ohm's Law (Eq. (5.14)) ($\vec{j} = \sigma \vec{E}$) to get

$$\begin{aligned} I &= jA \\ &= \sigma E A \\ &= \frac{\sigma A}{L} V_d \end{aligned} \tag{5.21}$$

This equation is usually written as

$$\boxed{V_d = \left(\frac{L}{\sigma A}\right) I = RI} \tag{5.22}$$

Here

$$\boxed{R = \left(\frac{L}{\sigma A}\right)}$$

is called the resistance R of the conductor. It is a proportionality constant between the current I and the voltage difference causing the current V_d. Of course, it is related to the conductivity as above but it also depends on the dimensions of the conductor. The unit of resistance is Ohm and is represented by Ω. We can also define the inverse of the resistance, called the conductance which is given by

$$\boxed{K = \frac{1}{R} = \frac{\sigma A}{L}}$$

Although almost universally used, one needs to exercise a bit of caution that we use Eq. (5.22) only in cases where our assumptions leading to the relationship between I and V_d are satisfied. In deriving Eq. (5.22), we have assumed that the current density is constant within the conductor and the electric field lines between the two end surfaces at potentials V_+ and V_- are all of equal length. This is not always the case. For example, Fig. 5.7(b) shows a case of a thick conductor where there is no symmetry and hence, the current density and the electric field are not parallel at all points within it. In such cases, Ohm's Law in the form given above is not valid. However, if the lateral dimensions of the conductor are very small compared to its length, the Eq. (5.22) is approximately valid. This is the case for electric cables commonly used in homes and laboratories.

EXAMPLE 5.1 A uniform thick metal plate of rectangular cross-section is bent into a semi-circular shape as shown in Fig. 5.8. The inner and outer radii are b and a respectively, the thickness t and conductivity of the metal σ. The two ends are equipotential at potentials V_1 and V_2. Find the resistance of the plate.

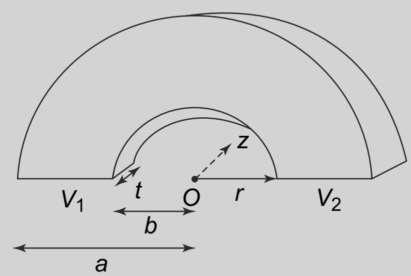


Fig. 5.8 *Example 5.1*

Solution

The geometry of the problem is not something which allows us to apply Ohm's Law in a direct fashion. Remember, to find the resistance, we need to find the current I and relate it to the potential difference between the two ends. However, let us try and use the basic equations satisfied by the electric potential V and try to find out the current density, current and electric field—these will then obviously allow us to find the current.

We take the origin of the coordinate system as shown in the figure, namely the centre of the semicircles. It is most convenient to use cylindrical coordinates in this situation. With this origin, the cylindrical coordinates for the plate extend from $0 \le \phi \le \pi$, $b \le r \le a$ and $0 \le z \le t$. The scalar potential in the plate is $V(r, \phi, z)$. Since there are no charges inside the plate, the potential satisfies the Laplace equation and we have

$$\nabla^2 V(r, \phi, z) = 0$$

which in cylindrical coordinates gives us

$$\frac{1}{r}\left(\frac{\partial}{\partial r} r \frac{\partial V}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 V}{\partial \phi^2} + \frac{\partial^2 V}{\partial z^2} = 0 \tag{5.23}$$

There will be no current in the direction of r and z and so we can write the Laplace equation as

$$\frac{1}{r^2}\frac{\partial^2 V}{\partial \phi^2} = 0 \tag{5.24}$$

which has the solutions

$$V(\phi) = A\phi + B \qquad \text{(A, B are integration constants)} \tag{5.25}$$

To determine A and B we apply boundary conditions. We know that at $\phi = 0$, $V(0) = V_1$ and therefore, we get $B = V_1$. Similarly, at $\phi = \pi$, $V = V_2$. Thus

$$\begin{aligned} V_2 &= A\pi + V_1 \\ A &= \frac{V_2 - V_1}{\pi} \\ A &= \frac{V_d}{\pi} \end{aligned} \tag{5.26}$$

The potential thus, is

$$V(r, \phi, z) = \frac{V_d}{\pi}\phi + V_1$$

Knowing the potential allows us to find the electric field easily. Remember that $\vec{E} = -\vec{\nabla}V$ and therefore, since the potential has only a ϕ dependence, we get

$$\begin{aligned} \vec{E} &= -\vec{\nabla}V(r, \phi, z) \\ &= -\hat{\phi}\frac{1}{r}\frac{\partial V}{\partial \phi} \\ &= -\hat{\phi}\frac{V_d}{\pi r} \end{aligned} \tag{5.27}$$

Knowing the electric field, one can use Ohm's Law to determine the current density.

$$\vec{j} = \sigma\vec{E} = -\hat{\phi}\sigma\frac{V_d}{\pi r} \tag{5.28}$$

The total current flowing from the end with $V = V_2$ to one with $V = V_1$, i.e., in the direction of $-\phi$ will be the integral over the appropriate surface. It is easy to see that

$$\begin{aligned} I &= \int_b^a dr \int_0^t dz \frac{\sigma V_d}{\pi r} \\ &= \frac{\sigma V_d t}{\pi} \ln\left(\frac{a}{b}\right) \end{aligned} \tag{5.29}$$

Now we know I and so can easily determine its relationship with the potential difference V_d and get the resistance R as

$$R = \frac{V_d}{I} = \frac{\pi}{\sigma t}\frac{1}{\ln\left(\frac{a}{b}\right)} \tag{5.30}$$

This result is interesting. Consider a wire of thin cross-section, $a \approx b$, i.e., $a = b + \delta$. In this case,

$$\ln\left(\frac{a}{b}\right) \approx \ln\left(1 + \frac{\delta}{b}\right) \approx \frac{\delta}{b}$$

and the resistance is approximately

$$R \approx \frac{\pi b}{\sigma t \delta}$$

But $\pi b = L$, the length of the plate while $t\delta = A$, the cross-sectional area. The resistance thus, becomes

$$\boxed{R = \frac{L}{\sigma A}}$$

which is the same as for a flat plate.

EXAMPLE 5.2 Consider a cone-shaped resistor whose ends are two spherical surfaces, with radii r_1 and r_2 concentric with each other. The two end surfaces are equipotential at potentials V_1 and V_2, thereby giving a potential difference of $V_d = V_2 - V_1$ between them. This is shown in Fig. 5.9. The material of the resistor has a conductivity σ. Calculate the resistance of the resistor.

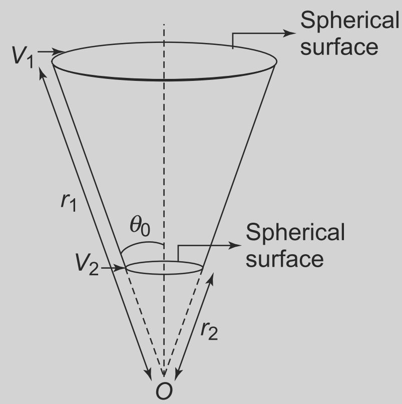


Fig. 5.9 *Example 5.2: A conical conductor with spherical end-surfaces. θ_0 is the opening angle of the cone*

Solution

Once again, we use the same strategy as in the previous example. We solve the Laplace equation to get the potential inside the conductor and then relate it to the electric field and use Ohm's Law. In this case, since the geometry is such, we use spherical polar coordinates with O as the origin. By symmetry, the potential can only be a function of r and therefore, one can write

$$V(r, \theta, \phi) = V(r)$$

Inside the conductor, there are no charges and hence the potential satisfies the Laplace equation. In spherical polar coordinates, we get

$$\nabla^2 V(r, \theta, \phi) = \left[\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 \frac{\partial}{\partial r}\right) + \frac{1}{r^2 \sin\theta}\frac{\partial}{\partial \theta}\left(\sin\theta \frac{\partial}{\partial \theta}\right) + \frac{1}{r^2 \sin^2\theta}\frac{\partial^2}{\partial \phi^2}\right] V(r, \theta, \phi) = 0 \quad (5.31)$$

and since V depends only on r, we get

$$\left[\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 \frac{\partial}{\partial r}\right)\right] V(r) = 0 \quad (5.32)$$

Hence,

$$r^2 \frac{\partial}{\partial r} V(r) = \mathrm{A} = \text{constant} \tag{5.33}$$

or

$$V(r) = -\frac{A}{r} + B \tag{5.34}$$

where B is a constant. We have the boundary conditions given that the two spherical surfaces are equipotential, or $V(r_1) = V_1$ and $V(r_2) = V_2$. Thus,

$$\begin{aligned} -\frac{A}{r_1} + B &= V_1 \\ -\frac{A}{r_2} + B &= V_2 \end{aligned} \tag{5.35}$$

Solving these equations, we get

$$\begin{aligned} A &= V_d \left(\frac{r_1 r_2}{r_2 - r_1} \right) \\ B &= V_2 - V_d \left(\frac{r_1}{r_1 - r_2} \right) \end{aligned} \tag{5.36}$$

Thus, the potential is given by

$$V = -\frac{A}{r} + B = -\frac{V_d}{r} \left(\frac{r_1 r_2}{r_2 - r_1} \right) + V_2 - V_d \left(\frac{r_1}{r_1 - r_2} \right)$$

Knowing the potential, allows us to calculate the electric field inside the resistor since $\vec{E} = -\vec{\nabla} V$ and hence, we have

$$\vec{E} = -\hat{r} \frac{A}{r^2} \tag{5.37}$$

The current density is related to the electric field and the conductivity as we have seen above and so

$$\vec{j} = -\sigma \frac{A}{r^2} \hat{r} \tag{5.38}$$

With the current density, we can find the current. This is obviously along the $\hat{r}$ direction and is given by

$$\begin{aligned} I &= \int_0^{\theta_0} (\vec{j} \cdot \hat{r}) 2\pi r^2 \sin\theta d\theta \\ &= -(2\pi\sigma A)(1 - \cos\theta_0) \\ &= -2\pi\sigma A(1 - \cos\theta_0) \\ &= 2\pi\sigma(1 - \cos\theta_0) V_d \left(\frac{r_1 r_2}{r_1 - r_2} \right) \end{aligned} \tag{5.39}$$

We now have the required relationship between the current and the potential difference and so the resistance is simply, by Ohm's Law, the ratio of the potential difference and the current

$$R = \frac{V_d}{I} = \frac{1}{2\pi\sigma(1 - \cos\theta_0)} \left(\frac{r_1 - r_2}{r_1 r_2} \right) \tag{5.40}$$

PROBLEM 5.7 A metallic bar of conductivity σ is bent to form a flat 90° sector of thickness d with inner radius r_1 and outer radius r_2. Show that the resistance between the vertical curved faces at r_1 and r_2 is

$$R = \frac{2\ln\frac{r_2}{r_1}}{\sigma\pi d}$$

PROBLEM 5.8 A circular disc, made of material of conductivity σ is of thickness d has a radius r_1. The disc has a hole in the centre of radius r_2. Find the resistance between the hole and the rim of the disc.

PROBLEM 5.9 A cylindrical conductor of cross-sectional area A is made of a material whose resistivity depends on the distance from the axis as $\rho = \frac{a}{r^2}$ where a is a constant. Find the resistance per unit length of the conductor.

PROBLEM 5.10 A typical electrolytic cell has for the total surface area of its two plates $S = 0.1\ \text{m}^2$. The distance between the plates is 0.6 mm. Find the approximate internal resistance of the battery, if the resistivity of the electrolyte $\rho = \frac{1}{\sigma} = 0.02$ ohm m.

PROBLEM 5.11 A spherical shell with its centre at the origin is of radius R_1 and is maintained at a constant potential V_1 . It is completely enclosed by a concentric spherical shell which is maintained at a constant potential V_2 and the space between the two shells is filled with a conducting material of conductivity σ. Calculate

(a) the current density in the conducting material,

(b) the total current between the shells and

(c) the total resistance of the conducting material.

5.5 JOULE HEATING

Let us consider the model of conduction discussed above, that is the Free Electron Model, in terms of the energies. A charged particle in an electric field experiences a force and hence is accelerated. In this process, it gains kinetic energy and this gain in kinetic energy of a charge q as it travels from point $\vec{r}_1$ to $\vec{r}_2$ is exactly equal to the loss in its electrostatic potential energy, $q\left[\phi(\vec{r}_1) - \phi(\vec{r}_2)\right]$ where $\phi(\vec{r})$ is the electric potential.

In the model of conduction of electric current that we have discussed above, recall that the electron gains kinetic energy in between the collisions with the fixed atoms. During the collision, it transfers this energy to the fixed atoms. This transfer of kinetic energy to the atoms is the mechanism whereby the electrons, on the average do not gain any kinetic energy and move with a constant velocity, called the **drift velocity**, despite the presence of the electric field causing the current flow. **The energy transferred**

to the fixed atoms appears as heat in the conductor and this phenomenon is called Joule heating of a current carrying conductor.

As we have argued above, the gain in the kinetic energy of the atoms and hence the heat energy must be equal to the difference in the electrostatic potential energy when the electron travels from one point to another since energy has to be conserved. We can determine the amount of heat generated.

Consider an infinitesimal surface ΔS normal to the flow of current or $\vec{j}$ at some point $\vec{r}$ as shown in Fig. 5.10.

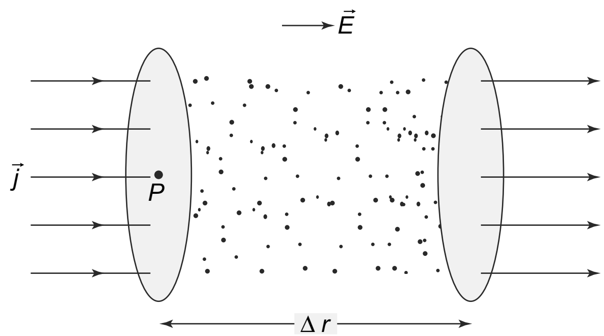


Fig. 5.10 *Charge crossing an infinitesimal area ΔS , normal to $\vec{j}$ at some point $\vec{r}$ to reach a similar area at a distance Δr away. $\vec{j}$, $\vec{E}$ and $\vec{r}$ are all parallel while ΔS is normal to $\vec{j}$*

In time Δt, a charge $j\,\Delta t\,\Delta S$ would have crossed ΔS to reach a similar area removed by a displacement $\Delta\vec{r}$ parallel to $\vec{j}(\vec{r})$. Thus, the loss in electrostatic potential energy of the charges crossing ΔS in time Δt is

$$\text{charge} \times \text{difference in potential} \times \text{time}$$

The charge we have already seen to be

$$j\,\Delta t\,\Delta S$$

The change in potential is simply

$$\left[\phi(\vec{r}) - \phi(\vec{r} + \Delta\vec{r})\right]$$

Thus, we get

$$\begin{aligned}\text{Loss of electrostatic potential energy} &= (\vec{j}\cdot\Delta\vec{S})\left[\phi(\vec{r}) - \phi(\vec{r}+\Delta\vec{r})\right]\Delta t \\ &= (\vec{j}\cdot\Delta\vec{S})\Delta t(\vec{E}\cdot\Delta\vec{r}) \\ &= j\,\Delta S\Delta t\,E\,\Delta r \qquad (5.41)\end{aligned}$$

since $\vec{j}$, $\Delta\vec{S}$, $\vec{E}$, $\Delta\vec{r}$ are all parallel. Now $\Delta S\Delta r$ is simply the volume ΔV of a cylinder with bases comprising of the two areas ΔS and of height Δr. The electrons crossing ΔS during time Δt collide with the atoms inside this volume and generate heat. This heat energy, ΔH is exactly equal to the loss of electrostatic potential energy. Hence, using Ohm's Law, ($\vec{j} = \sigma\,\vec{E}$) we can rewrite Eq. (5.41) as

$$\Delta H = \Delta V\,\Delta t\sigma\,E^2 \qquad (5.42)$$

where ΔH is the heat generated by the electric current. Thus, the rate of Joule heating, per unit volume, per unit time is given by

$$\boxed{\frac{\Delta H}{\Delta V \Delta t} = \sigma E^2}$$

Now if V is the potential difference between the two ends of a conductor of length L and uniform cross-sectional area A, then the power density is simply

$$p = \sigma \frac{V^2}{L^2}$$

and the total power lost by the conductor is

$$P = \sigma \frac{AV^2}{L}$$

But we know that

$$R = \frac{L}{\sigma A}$$

is the resistance R. Thus, the total power lost as heat by the conductor is given by

$$P = \frac{V^2}{R}$$

which is usually written as

$$\boxed{P = I^2 R}$$

Thus, we have the fact that the rate of dissipation of heat varies as the square of the current through a linear conductor.

PROBLEM 5.12 In the circuit shown in Fig. 5.11, a battery with internal resistance r is connected to three identical resistors of resistance R. What is the value of R such that the power generated in the circuit is maximum?

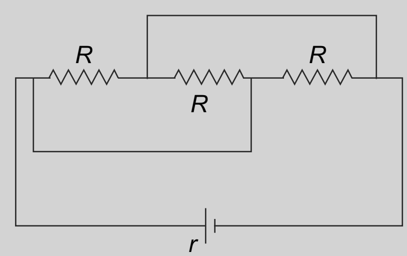


Fig. 5.11 *Problem 5.12*

5.6 COMBINATION OF RESISTORS

5.6.1 Resistors in Series

We have seen how a resistor, which obeys Ohm's Law behaves when a potential difference is maintained across its ends. Just as in the case of capacitors, it is possible to combine resistors together in various configurations. We first consider a combination called a series combination.

Consider a current carrying conductor and surfaces which cut across the conductor. We know that there cannot be any accumulation of charge within the conductor and hence currents flowing through these surfaces must be equal. Let us see if we can show this using the formulation above.

Consider a cylindrical volume V with flat surfaces A and B which cut across a resistor as in Fig. 5.12(a).

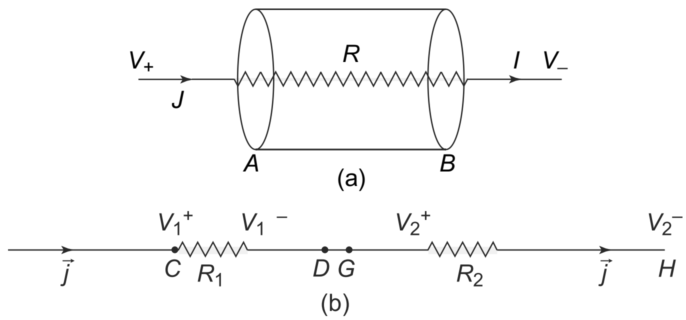


Fig. 5.12 *Combination of resistors (a) A cylindrical-shaped volume V whose flat surfaces A and B cut across a resistor R, (b) Two resistors R_1 and R_2 in series. $V_1^+, V_1^-, V_2^+, V_2^-$ are the potentials at C, D, G, H respectively*

We have already seen, from the Equation of Continuity, that the condition that there are no accumulated charges implies that the divergence of the current density $\vec{j}$ vanishes. Since $\vec{\nabla} \cdot \vec{j}(\vec{r}) = 0$, we also have

$$\iiint_V \vec{\nabla} \cdot \vec{j}(\vec{r}) d^3 r = 0 \tag{5.43}$$

Applying the Divergence Theorem, we can convert the volume integral of the divergence of $\vec{j}$ into a surface integral of $\vec{j}$ over a surface enclosing the volume and get

$$\iint_S \vec{j}(\vec{r}) \cdot d\vec{S} = 0 \tag{5.44}$$

where S is the entire surface enclosing the volume V. There is no current through the surface except through the faces marked A and B in Fig. 5.12. Further, the direction of $\vec{j}$ is along the outward normal at B and opposite to the outward normal at A. Thus, Eq. (5.44) becomes

$$-\iint_A \vec{j}(\vec{r}) \cdot d\vec{S} + \iint_B \vec{j}(\vec{r}) \cdot d\vec{S} = 0 \tag{5.45}$$

But these two integrals are just the currents flowing across the surfaces in question. Thus, we have the important result:

The current flowing across all surfaces that cut across the conductor is equal. This current is called the current through the conductor.

We will be using this result in our discussion of resistive networks.

In particular, consider a combination of two resistors R_1 and R_2 as shown in Fig. 5.12(b). The end D of one is connected to the end G of the other, leaving the ends C and H as free ends or terminals of the combination. **Such a combination is called resistances in series**.

Just as in Fig. 5.12(a), it is easy to see that the current flowing across any surface that cuts across R_1 is the same as that flowing across any surface that cuts across R_2. To see this, consider a volume V once again but this time the end surfaces A and B be such that one cuts across R_1 and the other across R_2. Equation (5.45) then implies that the current, I through the two resistors are equal.

Now because the ends D and G are connected, they must be at the same potential and hence V_1^- and V_2^+ are equal. Applying Ohm's law, we get

$$\begin{aligned} V_1^+ - V_1^- &= IR_1 \\ V_2^+ - V_2^- &= IR_2 \end{aligned} \tag{5.46}$$

Adding these two equations and using the fact that the potentials V_1^- and V_2^+ are equal, we get

$$V_1^+ - V_2^- = I(R_1 + R_2) \tag{5.47}$$

But this potential difference is the potential difference V_d across the free ends of the combination and hence the current through the combination is related to the potential difference by

$$V_d = I(R_1 + R_2) = IR_{\text{eq}} \tag{5.48}$$

where R_{eq} is the equivalent resistance of the combination given by

$$\boxed{R_{\text{eq}} = R_1 + R_2} \tag{5.49}$$

that is, resistances simply add up when we combine them in series. This result can of course, be easily generalised to a combination in series of any number of resistances as in Fig. 5.13.

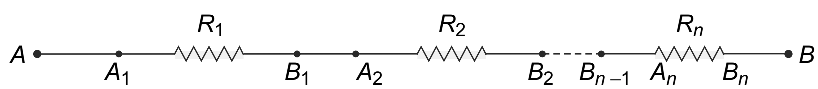


Fig. 5.13 *Series combination of n resistances $R_1, R_2, \cdots, R_n$. The ends of the resistances are connected in such a way so as to leave A and B free as the free terminals*

The equivalent resistance of such a combination is clearly

$$\boxed{R_{\text{eq}} = R_1 + R_2 + \cdots + R_n} \tag{5.50}$$

This is a very useful result and allows us in certain situations to determine the equivalent or effective resistance of combinations of resistors. However, one has to be careful of where one can use such a formulation. The next example illustrates an incorrect use of the rule for combining resistors in series to get the equivalent resistance.

EXAMPLE 5.3 Consider the resistor shown in below in Fig. 5.14, made of a material of conductivity σ in the shape below. What is the resistance?

Solution

At first sight, one can think of finding the total resistance by cutting the resistor into thin slices as shown in Fig. 5.14(b) and considering them in series and then using the rule for combining resistors in series. To do this, one would of course, be integrating over the variable x rather than summing up the resistors. Let us see if this is correct.

The radius of the circular disc at a distance x from the vertex grows linearly with x

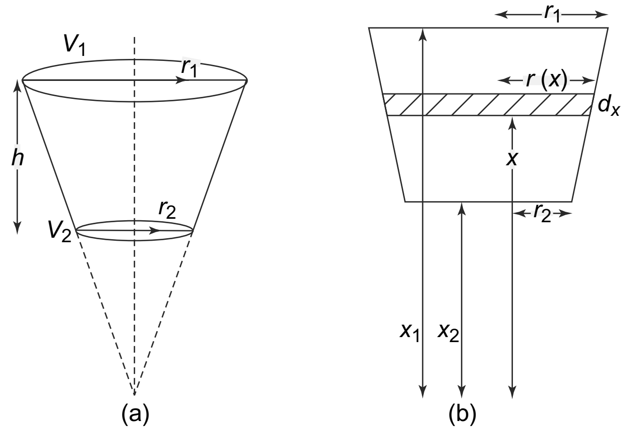


Fig. 5.14 *(a) A resistor in the shape of a section of a cone with flat end plates. The dimensions are shown in the figure. V_1 and V_2 are the potentials at the end plates, (b) Cutting the cone into discs with faces parallel to the end faces of the cone*

$$r(x) = x \tan\theta$$

At the lower and upper surfaces r has values r_1 and r_2 and hence

$$r_1 = x_1 \tan\theta$$

$$r_2 = x_2 \tan\theta$$

and therefore,

$$\tan\theta = \frac{r_1 - r_2}{h}$$

since $h = x_1 - x_2$. Now if we use the expression for the resistance of a conductor with constant cross-section area, i.e.,

$$R = \frac{L}{\sigma A}$$

for a disc of thickness dx, at x, we get

$$\begin{aligned} dR &= \frac{dx}{\sigma\pi r(x)^2} \\ &= \frac{h^2}{\sigma\pi x^2 (r_1 - r_2)^2} dx \end{aligned} \tag{5.51}$$

and hence, the total resistance is given by

$$\begin{aligned} R &= \int_{x_2}^{x_1} \frac{h^2}{\sigma\pi x^2 (r_1 - r_2)^2} dx \\ &= \frac{h^2}{\sigma\pi r_1 r_2} \end{aligned} \tag{5.52}$$

This result is not quite correct because of an incorrect interpretation of the expression for the resistance of a slab of constant cross section. A central assumption in arriving at that expression was that the electric field and hence the current density is constant and parallel everywhere in the slab. in our case, this is not true. Of course, if r_1 and r_2 are not much different, the Eq. (5.52) is approximately true.

The current at the central axis is, by symmetry parallel to the axis. But at the edges, the current must be parallel to the edges since there is no current normal to the edges.

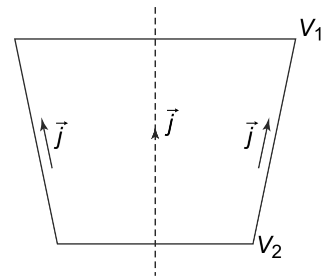


Fig. 5.15 *Example 5.3: The current density in the conical shaped resistor*

The correct way to solve this would be to find the electric field everywhere inside the resistor and from that determine the current density everywhere. Since there are no free charges inside a conductor, we can solve the Laplace equation inside the conductor subject to the boundary conditions that the potentials at the upper and lower surfaces are V_1 and V_2, as shown in Fig. 5.15. Another boundary condition needs to be imposed in this case and that is that the electric field $\vec{E} = -\vec{\nabla} V$ is parallel to the edges since it cannot have a component normal to the edges.

With these boundary conditions, the Laplace equation cannot be solved analytically and one has to numerically solve the equation. A calculation of this kind done by Romano & Price (***American Journal of Physics, 64, 1150(1996)***) gives us a value for the resistance a little higher than Eq. (5.52). They divide the resistor into N tapered slabs and compute the value numerically of each slab and then add it up. For instance, for $N = 4$ and $\frac{r_2}{h} = \frac{1}{2}$ and $\frac{r_1}{h} = 1$, we get a numerical value which is 2% higher than the one obtained from Eq. (5.52).

5.6.2 Circuit with a Resistor and a Battery

The series combination result that we have obtained above, is also applicable in the case when a resistor is connected to a battery with a finite internal resistance. Figure 5.16 shows a resistor of resistance R connected to a battery of EMF $\mathcal{E}$. A current I flows from the positive terminal of the battery to its negative terminal through R.

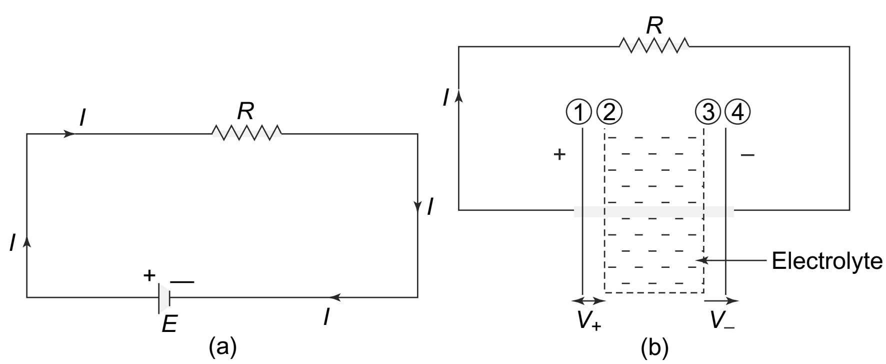


Fig. 5.16 *(a) A resistor R across a battery of EMF $\mathcal{E}$, (b) The same circuit but showing the elements and current flow within the battery. As we have seen, the positive plate has a potential V_+ higher than the electrolyte immediately surrounding it and the negative plate has a potential V_- lower than the electrolyte around it. The current within the electrolyte thus flows from the negative to the positive terminal*

As we have discussed, the current I flows from the positive to the negative terminal of the battery in the outside circuit, through R. Inside the battery though, the same current I flows from the negative terminal to the positive terminal. This will ensure that the positive plate maintains a constant potential V_+ higher than the electrolyte around it while the negative plate has a constant potential V_- lower than the electrolyte around it. Thus,

$$\begin{aligned} V_+ &= V_1 - V_2 \\ V_- &= V_3 - V_4 \end{aligned} \tag{5.53}$$

where V_1, V_2, V_3, V_4 are the potentials at the surfaces marked 1, 2, 3, 4 in Fig. 5.16(b).

Applying Ohm's Law across R, we get

$$V_1 - V_4 = IR \tag{5.54}$$

Similarly, the electrolyte also has a resistance r which is usually called the internal resistance of the battery. Applying Ohm's Law once again, we get

$$V_3 - V_2 = Ir \tag{5.55}$$

Combining Eqs. (5.53), (5.54) and (5.55), and using $\mathcal{E} = V_+ - V_-$, we get

$$I(r + R) = (V_1 - V_2 + V_3 - V_4) = \mathcal{E} \tag{5.56}$$

Thus, when a resistor R is connected to a battery, the circuit behaves as if it is a series combination of the external resistance R and the internal resistance of the battery r with a combined resistance of $R + r$.

5.6.3 Kirchhoff's First Law and Resistors in Parallel

In many situations, the terminals of resistors and batteries may be connected to a single point called a node as in Fig. 5.17.

The current flowing in the circuit elements are in general, different. Circuit elements in this subsection refer to resistors and batteries, though later on we will include other devices too. However, the currents, though different, satisfy a condition which is known as Kirchhoff's First Law and which is really a restatement of the conservation of charge.

> **Kirchhoff's First Law:** The algebraic sum of the currents flowing into a node is zero.

The important word in the statement is algebraic. If I denotes the current through a circuit element in one direction, then $-I$ is the current in the opposite direction.

The proof of this law is along the same line as we have seen above.

Consider a closed surface S that cuts through all the circuit elements that are connected to the node in question as in Fig. 5.17(a). Now since there is no charge accumulated inside of S, we have

$$\vec{\nabla} \cdot \vec{j}(\vec{r}) = 0$$

inside of S. On using the divergence theorem, we get

$$\iint_S \vec{j}(\vec{r}) \cdot d\vec{S} = \iiint_V \vec{\nabla} \cdot \vec{j}(\vec{r}) d^3\vec{r} = 0 \tag{5.57}$$

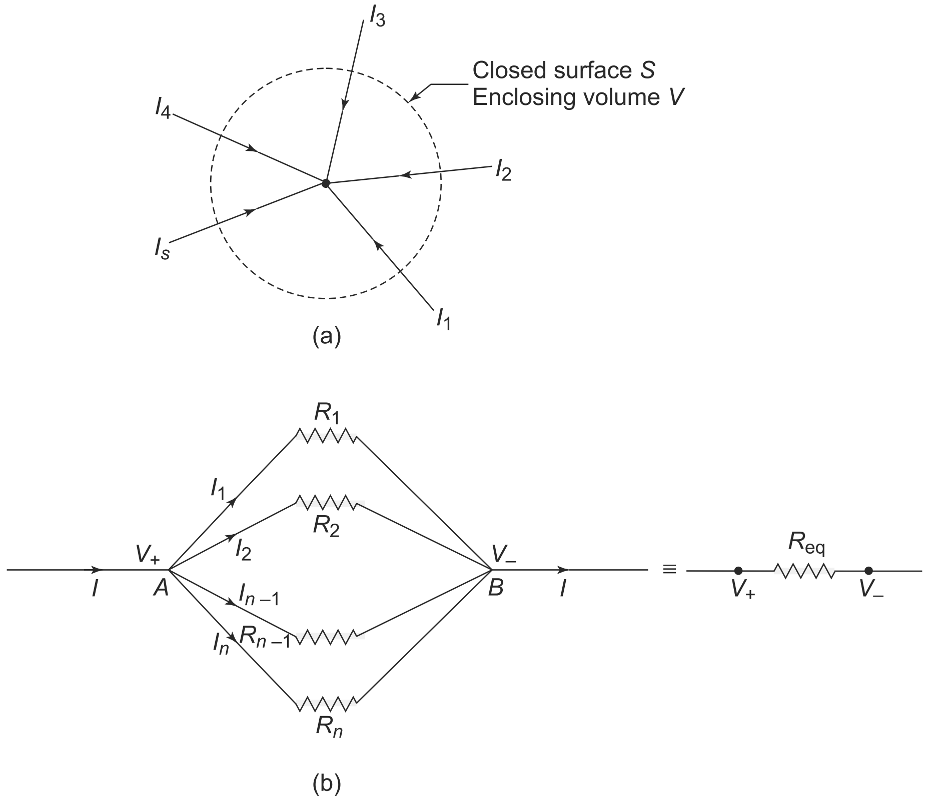


Fig. 5.17 *(a) Currents $I_1, I_2, \cdots, I_n$ flowing in n circuit elements that are all connected to a common node. A positive value of the current indicates current flowing into the node while a negative value indicates current flowing out from the node (b) n resistors connected in parallel across a battery*

where the surface integral is over S and V is the volume enclosed by S. On S, the current is zero except over areas ΔS_i $(i = 1, 2, \cdots, n)$ where the circuit elements intersect S. Hence,

$$\boxed{\iint_S \vec{j}(\vec{r}) \cdot d\vec{S} = \sum_{i=1}^{n} \iint_{\Delta S_i} \vec{j}(\vec{r}) \cdot d\vec{S} = -\sum_{i=1}^{n} I_i = 0} \tag{5.58}$$

where we have used the fact that

$$\iint_{\Delta S_i} \vec{j}(\vec{r}) \cdot d\vec{S} = -I_i$$

which is the outgoing current through the i^{th} circuit element.

Equation (5.58) is Kirchhoff's First Law which we have seen is simply a restatement of the Principle of Conservation of Charge.

Kirchhoff's First Law can be applied in many circumstances. As an example, consider another combination of resistors. A number of resistors are connected to the terminals of a battery in such a way that each resistor has its ends connected to the same pair of nodes A and B as in Fig. 5.17(b). **Such a combination of resistors is called a parallel combination.** In the figure, the node A is at a potential V_+ while the node B is at a potential V_-. Also, I is the current entering the node A while $I_1, I_2, \cdots, I_n$ are currents all leaving the node and flowing through the resistors $R_1, R_2, \cdots, R_n$ respectively. Applying Kirchhoff's Law to the node A, we get

$$I - I_1 - I_2 \cdots - I_n = 0 \tag{5.59}$$

Applying Ohm's Law to the n resistors $R_1, R_2, \cdots, R_n$, we get

$$V_d \equiv V_+ - V_- = I_i R_i \qquad (i = 1, 2, \cdots, n) \tag{5.60}$$

Combining these two equations, we get

$$I = \sum_{i=1}^{n} I_i = \sum_{i=1}^{n} \frac{V_d}{R_i} = V_d \left(\sum_{i=1}^{n} \frac{1}{R_i} \right) \tag{5.61}$$

Now

$$I = \frac{V_d}{R_{\text{eq}}}$$

and thus the current and potential difference between the nodes A and B is as if there was an equivalent resistance R_{eq}

$$\boxed{\frac{1}{R_{\text{eq}}} = \frac{1}{R_1} + \frac{1}{R_2} + \cdots + \frac{1}{R_n}} \tag{5.62}$$

Equation (5.62) is the law of combining resistances in parallel combination which should be compared with the law for combining resistances in series combination (Equation (5.50)).

With these two rules or laws of combining resistances in series and parallel, one can calculate the currents in more complicated circuits, as the example below illustrates.

EXAMPLE 5.4 In the circuit shown below, in Fig. 5.18, the battery has an internal resistance r. Calculate the current I in the circuit.

Solution

The circuit is clearly a combination of circuit elements connected in various series and parallel combinations. Let us start with R_1 and R_2. These two are in parallel and hence, their equivalent resistance R_{eq}^{12} is easily seen to be

$$R_{\text{eq}}^{12} = \frac{R_1 R_2}{R_1 + R_2}$$

Now this equivalent combination is in series with the resistance R_3 and hence the equivalent resistance of these is simply

$$R_{\text{eq}}^{123} = R_3 + R_{\text{eq}}^{12}$$

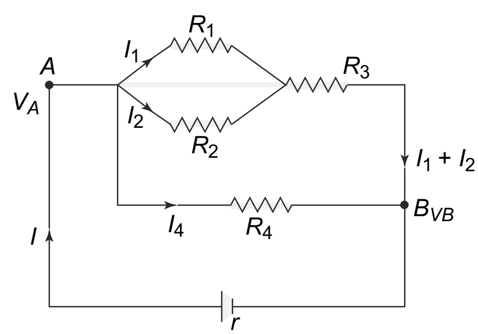


Fig. 5.18 *Example 5.4*

Also, R_{eq}^{123} is in parallel combination with R_4 and hence the equivalent resistance of this combination is R_{eq}^{1234} which is given by

$$R_{\text{eq}}^{1234} = \frac{R_{\text{eq}}^{123} R_4}{R_{\text{eq}}^{123} + R_4}$$

We thus, have an equivalent way of looking at the circuit as shown below in Fig. 5.19.

The current I can be easily worked out with Ohm's Law as

$$I = \frac{\mathcal{E}}{R_{\text{eq}}^{1234} + r}$$

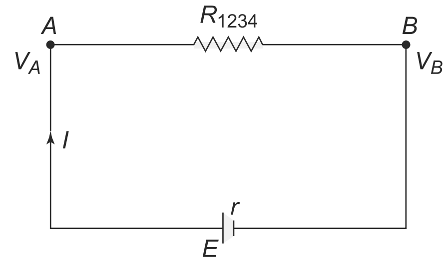


Fig. 5.19 *Example 5.4*

The currents I_1, I_2 and I_4 can now be found. We know that the potential difference between A and B , V_{AB} is simply related by Ohm's Law to the equivalent resistance R_{eq}^{1234} as

$$V_{AB} \equiv V_A - V_B = I\,R_{\text{eq}}^{1234}$$

But V_{AB} is also equal to $I_1 R_1$, $I_2 R_2$ as well as $I_4 R_4$. Thus,

$$I_1 = I \frac{R_{\text{eq}}^{1234}}{R_1}$$

$$I_2 = I \frac{R_{\text{eq}}^{1234}}{R_2}$$

$$I_4 = I \frac{R_{\text{eq}}^{1234}}{R_4}$$

In more complicated circuits, this simple reduction of combinations to parallel and series combinations may not work. A more systematic way of finding the currents in circuits uses what is known as Kirchhoff's Second Law.

PROBLEM 5.13 In the circuit in Figure 5.20, find the potential between points A and B given that $R_1 = 100$ Ohms, $R_2 = 200$ Ohms, $\mathcal{E}_1 = 50$ V and $\mathcal{E}_2 = 20$ V.

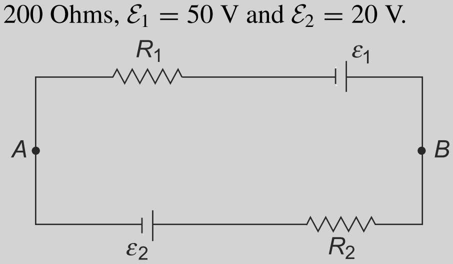


Fig. 5.20 *Problem 5.13*

5.6.4 Kirchhoff's Second Law

Kirchhoff's Second Law is a relationship between the potential drops across various circuit elements like batteries and resistors in a closed loop. The law states that:

> In any loop, i.e., a closed path in a circuit, the sum of
> (1) the product of currents (I_i) along the path and resistances (R_i) for each circuit element and
> (2) the potential drop along the path for every battery in the path
> is zero.

The proof of the law is straightforward. Consider a closed path with n resistances ($R_1, R_2, \cdots, R_n$) and m batteries with EMFs ($\mathcal{E}_i$). Let us start from any node A. Now as we go along the loop, the potential drop across resistor R_i is, by Ohm's Law given by $I_i R_i$. Similarly, when the circuit element in the loop is a battery, the potential drop across it is $\mathcal{E}_i$. The path is a closed one and so it closes and ends at A. The total potential drop as we go along the loop must be zero (since the start and end of the loop is at the same point, in this case A) and therefore,

$$\boxed{\sum_{i=1}^{n} I_i R_i + \sum_{i=1}^{m} \mathcal{E}_i = 0} \tag{5.63}$$

This is precisely Kirchhoff's Second Law, which is basically a statement of conservation of energy.

The usefulness of these Kirchhoff's Laws in analysing circuits will be illustrated in the following example.

EXAMPLE 5.5 **Wheatstone's Bridge** Calculate the currents through and the potential difference across each of the resistances in the circuit below. Assume that the battery has zero internal resistance.

Solution

This is the circuit which is commonly used in the laboratory for determining the resistance of resistors. It was first used in the middle of the nineteenth century.

The nodes and the currents in the circuit are as shown in Fig. 5.21. Applying Kirchhoff's First law at nodes 1, 2 and 4, we get

$$\begin{aligned} I_3 &= I - I_1 \\ I_5 &= I_1 - I_2 \\ I_4 &= I - I_2 \end{aligned} \tag{5.64}$$

There are six currents in the circuit : I_1, I_2, I_3, I_4, I_5 and I. There are three relations (Eq. (5.64) between them and hence only 3 out of six currents are independent. We shall take them to be I, I_1 and I_2.

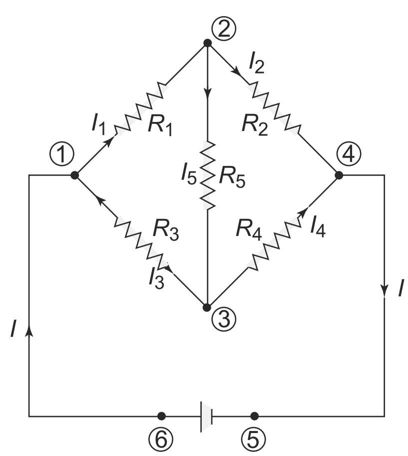


Fig. 5.21 *Wheatstone's Bridge*

Let us now apply Kirchhoff's Second Law to the loops

$$1 \to 2 \to 3 \to 1$$
$$2 \to 4 \to 3 \to 2$$
$$1 \to 3 \to 4 \to 5 \to 6 \to 1$$

and using Eq. (5.63), we get

$$\begin{aligned} I_1 R_1 + (I_1 - I_2)R_5 - (I - I_1)R_3 &= 0 \\ I_2 R_2 - (I - I_2)R_4 - (I_1 - I_2)R_5 &= 0 \\ (I - I_1)R_3 + (I - I_2)R_4 - \mathcal{E} &= 0 \end{aligned} \tag{5.65}$$

This set of equations can be written as a matrix equation

$$\begin{pmatrix} R_1 + R_3 + R_5 & -R_5 & -R_3 \\ -R_5 & R_2 + R_4 + R_5 & -R_4 \\ -R_3 & -R_4 & R_3 + R_4 \end{pmatrix} \begin{pmatrix} I_1 \\ I_2 \\ I \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \mathcal{E} \end{pmatrix} \tag{5.66}$$

or

$$\mathbf{RI} = \mathcal{E} \tag{5.67}$$

where $\mathbf{R}$ is a 3×3 matrix, $\mathbf{I}$ and $\mathcal{E}$ are 3×1 matrices.

The solution of Eq. (5.67) can be obtained by standard techniques and we get

$$I_1 = \mathbf{det} \begin{pmatrix} 0 & -R_5 & -R_3 \\ 0 & R_2 + R_4 + R_5 & -R_4 \\ \mathcal{E} & -R_4 & R_3 + R_4 \end{pmatrix} / \mathbf{detR} \tag{5.68}$$

$$I_2 = \mathbf{det} \begin{pmatrix} R_1 + R_3 + R_5 & 0 & -R_3 \\ -R_5 & 0 & -R_4 \\ -R_3 & \mathcal{E} & R_3 + R_4 \end{pmatrix} / \mathbf{detR} \tag{5.69}$$

and

$$I = \mathbf{det} \begin{pmatrix} R_1 + R_3 + R_5 & -R_5 & 0 \\ -R_5 & R_2 + R_4 + R_5 & 0 \\ -R_3 & -R_4 & \mathcal{E} \end{pmatrix} / \mathbf{detR} \tag{5.70}$$

The condition that no current passes through the resistance R_5 i.e. $I_5 = 0$ is known as the null condition. This condition is used frequently in laboratories to determine the resistance.

$I_5 = 0$ implies $I_1 = I_2$ and expanding the determinants in Eqs. (5.68, 5.69, 5.70) it is easy to see that this implies

$$\boxed{\frac{R_1}{R_2} = \frac{R_3}{R_4}}$$

This is the condition for a balanced Wheatstone Bridge. This condition is used to determine the values of unknown resistances to a high degree of accuracy.

PROBLEM 5.14 In the circuit given in Fig. 5.22, $R_1 = 30$ Ohms, $R_2 = 60$ Ohms and $R_3 = 200$ Ohms. The battery has an emf of $\mathcal{E} = 120$ V. Calculate the equivalent resistance and the current flowing through the resistance marked A in the figure.

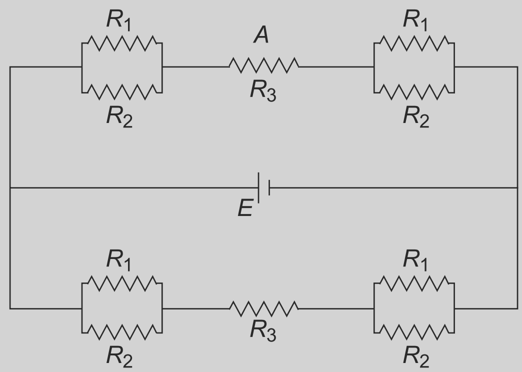


Fig. 5.22 *Problem 5.14*

5.7 NETWORKS

A network is defined as a set of resistors and batteries connected together. When one draws a general network on a two dimensional surface, one would, in general find lines crossing each other. Such networks are called 'non-planar'. Figure 5.23, is an example of a non-planar network and we will not be considering them. We would restrict our discussion to **planar networks which can be defined as those where no lines cross each other**, like for instance the Wheatstone's Bridge in the example above. A general method to calculate the currents in all parts of a planar network given the resistances and the EMFs is outlined below.

We first define a loop in the context of a planar network. **A loop is simply a closed, oriented path through resistors and batteries in the network.** Thus, for instance, the path $1 \to 2 \to 3 \to 1$ is a loop in the Wheatstone Bridge circuit above (Fig. 5.21). A particular circuit element may of course belong to more than one loop. Currents in the network, through batteries and resistors, can be written in terms of n independent loop currents, where n is smaller than or at best equal to the number of currents in the network for simple networks.

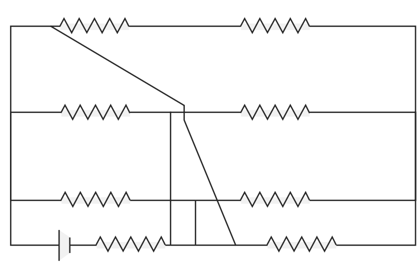


Fig. 5.23 *Example of a non-planar network*

For planar networks, this number n is easy to find. When a planar network is drawn on paper, there will be 'holes' which will be bordered by resistors and batteries or lines joining them. Thus, in Fig. 5.21, there are only three independent loop currents in terms of which all currents can be obtained. Thus, in Fig. 5.24, the loop currents let us call I_1, I_2, I_3 .

Thus, the current from the node 3 to 8 is simply $I_1 - I_2$ where the negative sign for I_2 indicates it runs from $8 \to 3$. It is easy to verify that Kirchhoff's First Law is automatically satisfied when all currents are written in terms of loop currents.

Once we have all the currents in the network in terms of loop currents, we can use Kirchhoff's Second Law for the loops to get n linear equations with n variables, i.e. the loop currents and also the EMFs in the circuit. These can be solved to get the loop currents and hence all the currents in the network.

This method of solving networks is best illustrated by an example.

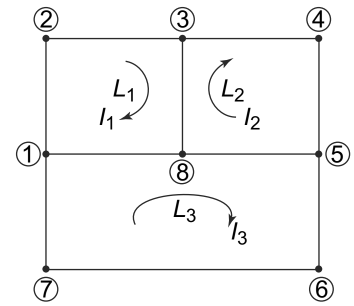


Fig. 5.24 *Loop currents for a wheatstone bridge circuit*

EXAMPLE 5.6 Calculate the currents through the resistances in the circuit below. Assume that the batteries have zero internal resistance.

Solution

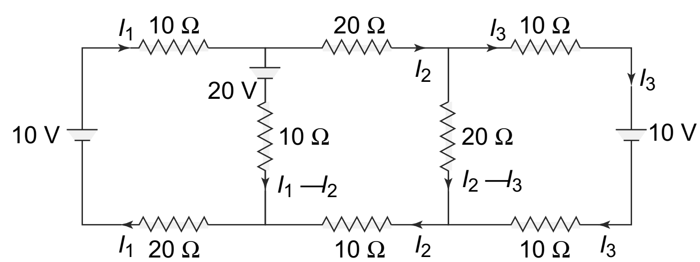


Fig. 5.25 *Example 5.6*

The battery EMFs and the resistances are given in the Fig. 5.25. Let I_1, I_2, I_3 be the loop currents in the three loops. The currents in the batteries and resistances in terms of the loop currents are also indicated in the figure. Using Kirchhoff's Second Law on the three loops, we get

$$\begin{aligned} 10I_1 + 10(I_1 - I_2) + 20I_1 &= -10 \\ 20I_2 + 20(I_2 - I_3) + 10I_2 - 10(I_1 - I_2) &= 20 \\ 10I_3 + 10I_3 - 20(I_2 - I_3) &= -10 \end{aligned} \tag{5.71}$$

which can be written in matrix form as

$$\mathbf{RI} = \mathcal{E} \tag{5.72}$$

with

$$\mathbf{R} = \begin{pmatrix} 40 & -10 & 0 \\ -10 & 60 & -20 \\ 0 & -20 & 40 \end{pmatrix} \tag{5.73}$$

$$\mathbf{I} = \begin{pmatrix} I_1 \\ I_2 \\ I_3 \end{pmatrix} \tag{5.74}$$

and

$$\mathcal{E} = \begin{pmatrix} -10 \\ 20 \\ -10 \end{pmatrix} \tag{5.75}$$

Inverting **R**, we get

$$\mathbf{R}^{-1} = \begin{pmatrix} 0.02632 & 0.00526 & 0.00263 \\ 0.00526 & 0.02105 & 0.01053 \\ 0.00263 & 0.01053 & 0.03026 \end{pmatrix} \tag{5.76}$$

and hence

$$\mathbf{I} = \mathbf{R}^{-1}\mathcal{E} = \begin{pmatrix} -0.184 \\ 0.263 \\ -0.118 \end{pmatrix} \tag{5.77}$$

Thus, we see that once we follow the general prescription for writing all the currents in terms of the loop currents and then using Kirchhoff's Second law, any planar network can be analysed. Of course, for more complicated networks, one needs to invert larger matrices and that is not always an easy task.

PROBLEM 5.15 Find the potential difference between the plates of the capacitor C in the Fig. 5.26. Given that $\mathcal{E}_1 = 40$ V, $\mathcal{E}_2 = 10$ V, $R_1 = 100$ Ohms, $R_2 = 200$ Ohms and $R_3 = 300$ Ohms. Assume zero internal resistance of the batteries.

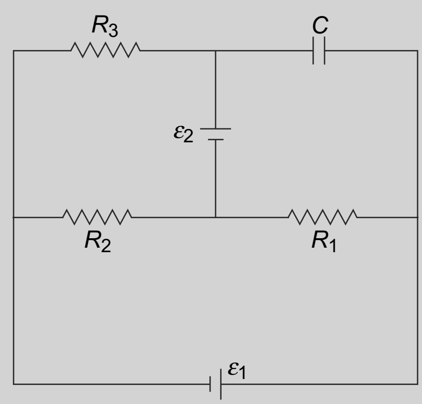


Fig. 5.26 *Problem 5.15*

5.8 NETWORK THEOREMS

We have discussed above the general method of calculating the currents (and hence the potential differences) across resistors in a planar network. In general, the current through any branch of a network depends not just on the resistances and EMFs in that branch but also the resistances and EMFs in other branches. There are however, two general theorems which simplify the calculations immensely. These are Thevenin's Theorem and Norton's Theorem.

5.8.1 Thevenin's Theorem

For the kind of networks we have been discussing, **Thevenin's Theorem** can be stated as follows:

'Any circuit, consisting of batteries and resistors and having two terminals, can be replaced by an equivalent circuit with a resistor R_{th} connected to a voltage source V_{th}. This equivalent circuit is called the Thevenin equivalent.'

Figure 5.27 below illustrates the theorem.

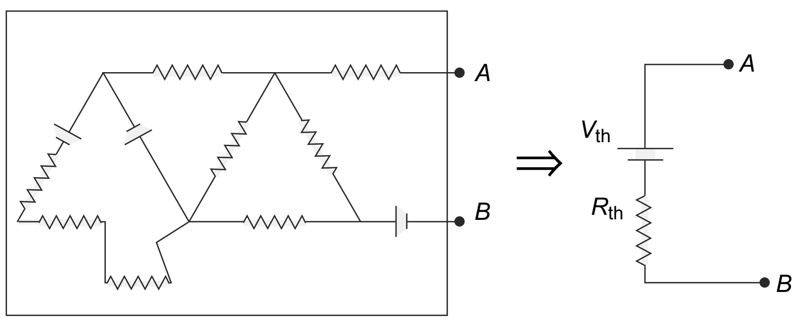


Fig. 5.27 *Thevenin equivalent circuit*

Note that the box on the left-hand side can be any complex circuit with two terminals A and B. As far as the terminals A and B are concerned, the complex circuit in the box is just equivalent to the Thevenin equivalent. Thus, if we connect a load across A and B, the current through it and the potential across it are determined by the Thevenin equivalent circuit and the details of the complex network as unimportant. The theorem also allows us to find the values of V_{th} and R_{th}.

1. V_{th} is the value of the voltage or potential difference between A and B in an open circuit, i.e., without connecting anything in between A and B.
2. R_{th} is simply $\frac{V_{th}}{I}$ where I is the value of the current through AB when the terminals A and B are short circuited, i.e., connected with a zero resistance.

We shall prove Thevenin's theorem by referring to the simple circuit and then showing that it can be easily generalised to more complicated, arbitrary circuits. Consider the circuit given in Fig. 5.28.

The circuit consists of three loops with loop currents I_1, I_2 and I. The loop with the current I connects the terminals A and B through the load resistance R. We can write down the Kirchhoff's relations for

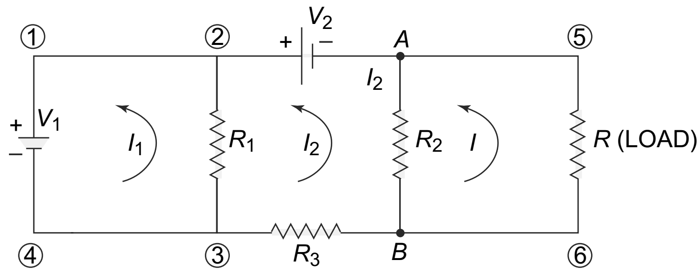


Fig. 5.28 *A simple network*

these three loops.

$$\textbf{Loop 32143} \qquad (R_1)I_1 + (0)I_2 + (0)I = -V_1 \tag{5.78}$$

$$\textbf{Loop BA23B} \qquad (-R_1)I_1 + (-R_1 + R_2 + R_3)I_2 + (-R_2)I = +V_2 \tag{5.79a}$$

$$\textbf{Loop 65AB6} \qquad (0)I_1 + (-R_2)I_2 + (R_2)I = -IR \tag{5.79b}$$

Note that in writing these loop equations, in the last equation, we have written the IR term on the RHS so that the LHS has no terms containing the load resistance R. These equations can be written in matrix form as

$$\begin{pmatrix} R_1 & 0 & 0 \\ -R_1 & R_2 + R_3 + R_1 & -R_2 \\ 0 & -R_1 & R_2 \end{pmatrix} \begin{pmatrix} I_1 \\ I_2 \\ I \end{pmatrix} = \begin{pmatrix} -V_1 \\ +V_2 \\ -IR \end{pmatrix} \tag{5.80}$$

or

$$\mathbf{RI} = \mathbf{E} \tag{5.81}$$

It is easy to see that this form of the matrix equation will be true for any complicated network, the only difference being that the size of the matrices will be bigger if there are more loops. For a network of n loops, $\mathbf{R}$ will be a $(n \times n)$ matrix while $\mathbf{I}$ and $\mathbf{E}$ will be $(n \times 1)$ matrices.

Consider now a general circuit of the type shown in Fig. 5.27 with a load R connected between A and B, through which a circuit I flows from A to B and call the potential difference between A and B as V with this load.

We can now invert the matrix equation, Eq. (5.81) to get

$$\mathbf{I} = \mathbf{R}^{-1}\mathbf{E} \tag{5.82}$$

and thus,

$$\mathbf{I} = (\mathbf{I})_n = (\mathbf{R}^{-1}\mathbf{E})_n \tag{5.83}$$

Now notice that $\mathbf{R}^{-1}$ contains other resistances but not R. Also $\mathbf{E}$ is linear in IR and other EMFs in the circuit. Thus, the RHS of Eq. (5.83) contains terms linear in IR and a term independent of I. Thus, we can write

$$I = \alpha + \beta IR = \alpha + \beta V \tag{5.84}$$

where V is the potential difference across R and is given by IR. In Eq. (5.84), α and β do not depend on either I or R.

We can easily identify α and β. Putting $I = 0$ the voltage V becomes the open circuit voltage V_{OC} and thus

$$\alpha = -\beta V_{\text{OC}} \tag{5.85}$$

Next, putting $V = 0$, $I = I_{\text{SC}}$ the short circuit current. Thus,

$$I_{\text{SC}} = \alpha \tag{5.86}$$

Hence, we get

$$I = I_{\text{SC}} - \frac{I_{\text{SC}}}{V_{\text{OC}}} V \tag{5.87}$$

Now calling

$$V_{\text{OC}} = V_{\text{Th}}$$

and

$$I_{\text{SC}} = \frac{V_{\text{Th}}}{R_{\text{Th}}}$$

we get

$$V = V_{\text{Th}} - I R_{\text{Th}} \tag{5.88}$$

The current I through the resistance R (the load) from A to B is thus,

$$I(R + R_{Th}) - V_{Th} = 0 \tag{5.89}$$

Equation (5.89) is precisely the $I - R$ relationship when R is connected across A and B in the Thevenin equivalent circuit Fig. 5.27, where I is the current from A to B through R. This proves the Thevenin theorem.

The importance of the Theorem lies in the fact that as the load R across the terminals A and B of the network is varied, the current through it can be calculated from the simple Thevenin equivalent circuit, however complicated be the network.

5.8.2 Norton's Theorem

The second network theorem, Norton's Theorem is based on the idea of a constant current source, something we have not encountered so far. We have of course, seen constant voltage sources like a battery which produce a constant voltage across their terminals irrespective of the circuit they are placed in. This statement is strictly not true since any battery will have some internal resistance and thus, depending on the current passing through it, will have a potential drop because of the internal resistance. However, in most cases, the internal resistance of the battery is negligible and hence we can think of batteries as constant voltage sources. A constant current source is complimentary to a constant voltage source and it produces a constant current through itself irrespective of the circuit. Norton's Theorem is depicted in Fig. 5.30.

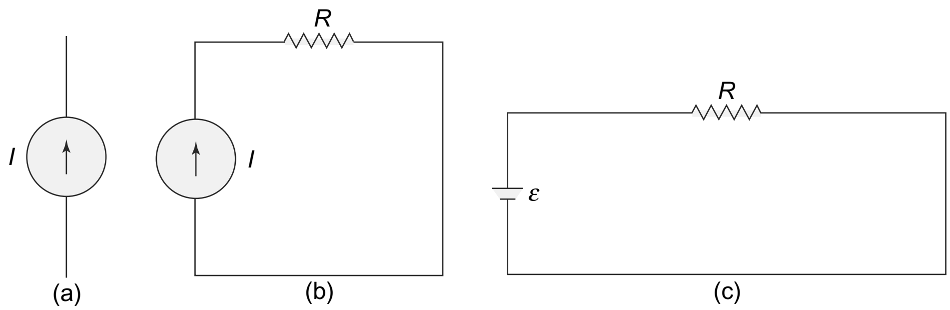


Fig. 5.29 *(a) The symbol of a constant current source, (b) In any circuit, as in a simple one shown here, the current through R will be I irrespective of R, while the voltage across it will change since the voltage is IR, (c) For a constant voltage source like a battery, the voltage across R will be constant while the current through it, V/R will change with R*

Norton's Theorem states that any network with resistors and batteries with two terminals can be replaced by a current source $I_N = I_{SC}$ with a resistance $R_N = R_{\text{Th}}$ in parallel for calculating the current through and voltage across any load R connected across the two terminals.

The proof is contained in the equation for Thevenin theorem.

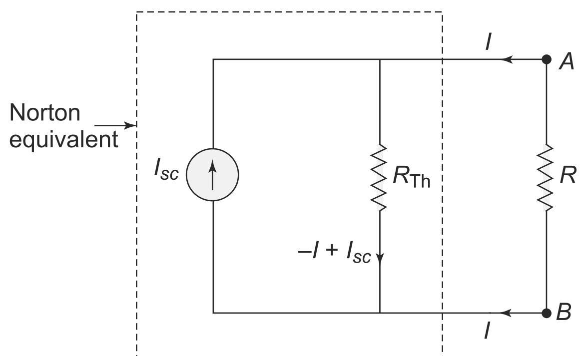


Fig. 5.30 *A network for the purposes of calculating the current through a load R can be replaced by its Norton equivalent*

The short circuit current I_{SC} is when $R = 0$ and is given by

$$I_{\text{SC}} = \frac{V_{\text{OC}}}{R_{\text{Th}}}$$

Hence, from Eqn. (5.89),

$$I = \frac{V_{\text{Th}}}{R + R_{\text{Th}}} = I_{\text{SC}} \frac{R_{\text{Th}}}{R + R_{\text{Th}}} \tag{5.90}$$

This is precisely what the Norton equivalent circuit, shown in Fig. 5.30 will give. The resistances R and R_{Th} are in parallel and hence their equivalent resistance is simply

$$R_{\text{eff}} = \frac{R_{\text{Th}} R}{R_{\text{Th}} + R}$$

The voltage difference between the terminals A and B is

$$V_A - V_B = I_{\mathrm{SC}} \frac{R_{\mathrm{Th}} R}{R_{\mathrm{Th}} + R}$$

The current I is therefore

$$\frac{V_A - V_B}{R} = I_{\mathrm{SC}} \frac{R_{\mathrm{Th}}}{R_{\mathrm{Th}} + R}$$

which is the same as in Eq. (5.90)

EXAMPLE 5.7 In the circuit as shown in Fig. 5.31, a load of 2 Ω is connected across A and B. Calculate the current through it. Also, calculate the Thevenin equivalent voltage and resistance and verify your result using the Thevenin equivalent circuit.

Solution

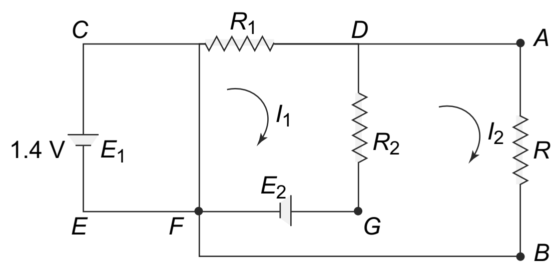


Fig. 5.31 $R_1 = 2\ \Omega$, $R_2 = 0.5\ \Omega$, $R = 1\ \Omega$, $E_1 = 1.4V$, $E_2 = 0.35V$

The circuit has two loops which we take as $CDGEC$ and $DABFGD$ with loop currents I_1 and I_2, respectively. Then, applying Kirchhoff's second law gives us two equations

$$I_1 R_1 + R_2(I_1 - I_2) + E_2 - E_1 = 0 \tag{5.91}$$

and

$$I_2 R + R_2(I_2 - I_1) - E_2 = 0 \tag{5.92}$$

These can be easily solved to give

$$I_1 = \frac{R(E_1 - E_2) + E_1 R_2}{R(R_1 + R_2) + R_1 R_2} \tag{5.93}$$

and

$$I_2 = \frac{E_2}{R + R_2} + \frac{R_2}{R_2 + R} \frac{R(E_1 - E_2) + E_1 R_2}{R(R_1 + R_2) + R_1 R_2} \tag{5.94}$$

Putting in the values of the various resistances and EMFs, we get

$$I_1 = 0.5\ \mathrm{A} \tag{5.95}$$

and

$$I_2 = 0.4\ \mathrm{A} \tag{5.96}$$

The Thevenin equivalent circuit, is given by

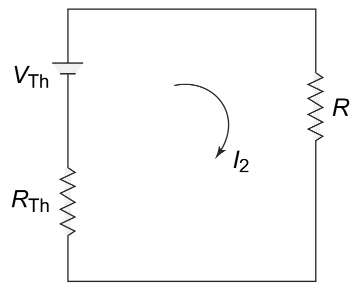


Fig. 5.32 *Example 5.7: Thevenin equivalent circuit*

In this V_{Th} is the value of I_2R as $R \to \infty$ which, from Eq. (5.94) is simply

$$V_{Th} = E_2 + \frac{(E_1 - E_2)R_2}{R_1 + R_2} = 0.56 \text{ V} \tag{5.97}$$

The Thevenin equivalent resistance R_{Th} is simply $\frac{V_{Th}}{I_{SC}}$ where I_{SC} is the value of I_2 as $R \to 0$.

$$I_{SC} = \frac{E_2}{R_2} + \frac{E_1}{R_1} = 1.4 \text{ A} \tag{5.98}$$

and therefore,

$$R_{Th} = 0.4\ \Omega \tag{5.99}$$

Thus, the current in the Thevenin equivalent is

$$I_2 = \frac{0.56}{1 + 0.4} \text{ A} = 0.4\text{A} \tag{5.100}$$

which agrees with Eq. (5.96).

EXAMPLE 5.8 In the circuit in Fig. 5.33, obtain the current through a resistor $R = 1\ \Omega$ connected across terminals A and B. Also, obtain the Norton equivalent circuit and verify your result.

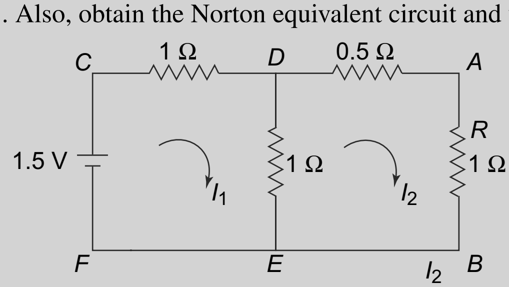


Fig. 5.33 *Example 5.8*

Solution

This network has two loops, which we take as $CDEFC$ and $ABEDA$ carrying loop currents I_1 and I_2. Kirchhoff's loop equations give us

$$I_1 1 + (I_1 - I_2)1 - 1.5 = 0 \tag{5.101}$$

$$I_2R + (I_2 - I_1)1 + I_2(0.5) = 0 \tag{5.102}$$

For $R = 1\Omega$, these are easily solved to get

$$I_1 = 1.125 \text{ A} \qquad\qquad I_2 = 0.375 \text{ A} \tag{5.103}$$

The Norton equivalent circuit as shown in Fig. 5.34 is

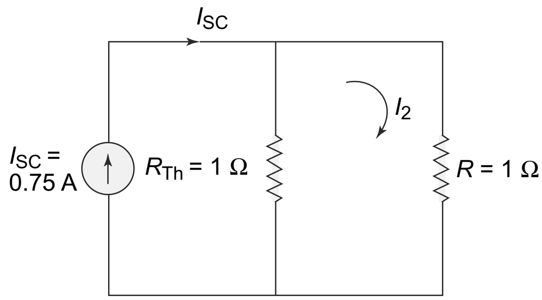


Fig. 5.34 *Example 5.8: Norton equivalent circuit*

The short-circuit current is obtained by setting $R = 0$ in Eq. (5.101) and Eq. (5.102) for I_2 and we get

$$I_{SC} = I_2(R = 0) = 0.75 \text{ A} \tag{5.104}$$

The open circuit voltage V_{OC} is I_2R for $R \to \infty$ and we get

$$V_{OC} = 0.75 \text{ A}$$

and therefore,

$$R_{Th} = \frac{V_{OC}}{I_{SC}} = 1\ \Omega$$

For $R = 1\ \Omega$, the current I_2 is $\frac{I_{SC}}{1+1} = 0.375$ A which agrees with Eq. (5.103).

5.9 VARYING CURRENTS

5.9.1 RC Circuit

We have been discussing currents that are steady till now. This is obviously not always that case—we will have circuits where the current is time-dependent. For instance, a circuit in which a capacitor is present has a time-dependent current. A capacitor, being an insulator (the plates are filled with air or some dielectric) of course doesn't allow any current to flow through it. However, its presence in the circuit may make the currents through the other elements of the circuit time dependent.

We have thought of capacitors as two plates or conductors which are separated by a dielectric or air and which carry equal and opposite charge. In general, in such a device, the field would not be confined only to the region in between the plates. If such a capacitor is present in an electrical circuit, then our usual analysis of potentials (for instance, Ohm's Law) would need to take into account the potential produced by the charges on the plates. However, in most cases in practice, this effect is rather small since the electric field decreases rapidly as we go away from the plates.

In our analysis of circuits, we thus consider 'ideal capacitors' which are defined as those in which the electric field produced by the charges on the plates is confined only to the region between the plates. A toroidal capacitor and a cylindrical capacitor, shown in Fig. 5.35 are good approximations of an 'ideal capacitor'. The potential difference across the terminals is $V = \frac{Q}{C}$.

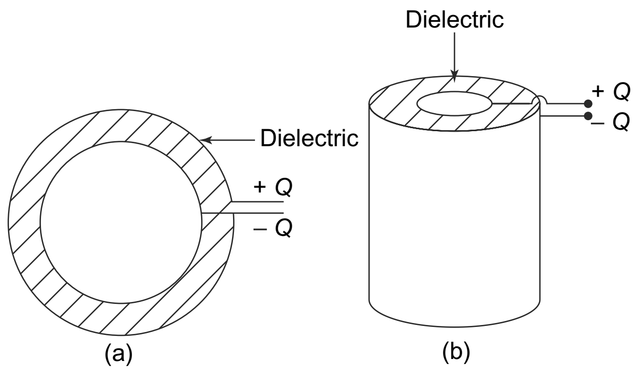


Fig. 5.35 *Approximately ideal capacitors (a) A section of a toroidal capacitor, (b) A cylindrical capacitor*

To illustrate this phenomenon, consider the circuit in Fig. 5.36 below.

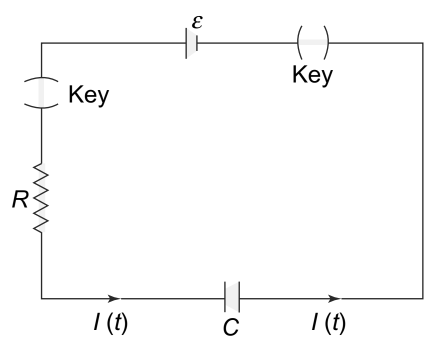


Fig. 5.36 *When the keys are out, the capacitor plates are not connected to any charge source. When the keys are inserted, the plates are connected to the positive and negative terminals of the battery and charge flows creating a time-dependent current through the resistance R*

We assume that the capacitor plates initially carry no charge. When the keys in the circuit are out, the capacitor plates are not connected to any charge source and nothing happens. When the keys are inserted, charge flows from the battery terminals to the capacitor plates and this flow of charge constitutes a current. The current lasts till the capacitor plate connected to the positive terminal of the battery achieves the same potential as the terminal and is in equilibrium. Similarly, the negative terminal and the plate connected to it are at the same potential.

The time dependence of this process of charging the capacitor is easy to work out. Let the charge on the capacitor at any time t be $Q(t)$. When the keys are inserted at time $t = 0$, $Q(0) = 0$, as we have defined. The rate at which the capacitor plates are charging is $\frac{dQ(t)}{dt}$ and this is also the current flowing in the circuit. The potential drop across the resistor R, at any time t is therefore, given by Ohm's Law to be $R\frac{dQ(t)}{dt}$. The potential across the capacitor is $V(t) = \frac{Q(t)}{C}$ where C is the capacitance of the capacitor. Assuming that the internal resistance of the battery is negligible, we can use Kirchhoff's law for the circuit along with the capacitor, to get

$$R\frac{dQ(t)}{dt} + \frac{Q(t)}{C} = \mathcal{E} \tag{5.105}$$

This equation is solved by defining

$$Q' = Q(t) - \mathcal{E}C$$

Equation (5.105) in terms of Q' reads

$$RC\frac{dQ'(t)}{dt} + Q' = 0 \tag{5.106}$$

which has a solution

$$Q'(t) = Q'(0)e^{-\frac{t}{RC}} \tag{5.107}$$

But

$$Q'(0) = Q(0) - \mathcal{E}C = -\mathcal{E}C$$

since $Q(0) = 0$. Thus, the solution for the charging case is

$$\boxed{Q(t) = \mathcal{E}C(1 - e^{-\frac{t}{RC}})} \tag{5.108}$$

The quantity $RC = \tau$ is called the time constant of the circuit. It is easy to see that it has dimensions of time.

The same method can be used to describe the discharging of a charged capacitor through a resistor.

Consider the circuit in Fig. 5.37.

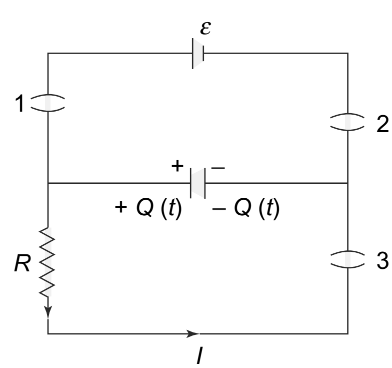


Fig. 5.37 *Discharging a capacitor. Before time $t = 0$, the keys 1 and 2 are in and key 3 is out. At time t=0, key 1 and 2 are taken out and key 3 inserted*

The capacitor is fully charged at time $t = 0$. The charge on the capacitor is simply $Q(0) = C \times V = C\mathcal{E}$. The equations are exactly as the ones we got for the charging circuit except that at $t = 0$, $Q(0) = EC$ and hence,

$$Q'(0) = Q(0) - \mathcal{E}C = 0$$

At time $t = 0$ the keys 1 and 2 are taken out and key 3 inserted. Current will now flow through the lower part of the circuit in the figure above. This process is exactly the same as the one considered for charging the capacitor except that there is no battery. Thus, in our equations, we set $E = 0$ for $t > 0$.

$$R\frac{dQ(t)}{dt} + \frac{Q(t)}{C} = 0 \tag{5.109}$$

which has a solution for the discharging case as

$$\boxed{Q(t) = \mathcal{E}Ce^{-\frac{t}{RC}}} \tag{5.110}$$

since $Q(0) = \mathcal{E}C$.

5.9.2 More Complex Capacitor Circuits

The circuits considered above for charging and discharging of a capacitor were both one loop circuits and hence, easy to analyse. In more complicated circuits with capacitors, the loop analysis with currents is no longer applicable and hence we need to develop other methods as the example shows.

EXAMPLE 5.9 In the circuit in Fig. 5.38, obtain the current in the capacitor branch when the switch S is switched on at time $t = 0$.

Solution

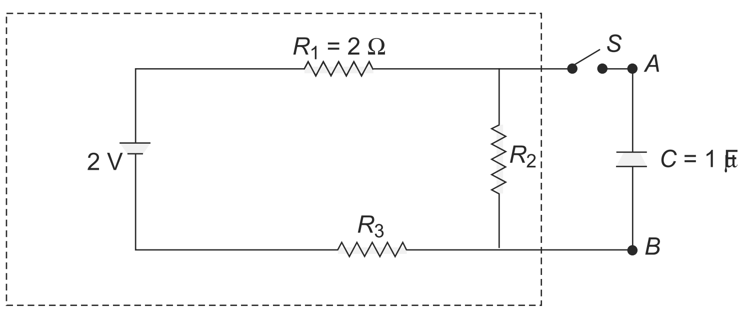


Fig. 5.38 *Example 5.9*

This is a two-loop network with a capacitor, as shown in Fig. 5.38. To solve it, consider removing the capacitor and working out the Thevenin equivalent of the remaining network indicated with a dashed line above. The Thevenin equivalent circuit, as shown in Fig. 5.39 below is

The Thevenin voltage, which is the voltage across R_2 is simply

$$V_{\text{Th}} = \frac{2(0.5)}{(2 + 3 + 0.5)} = 0.18 \text{ V}$$

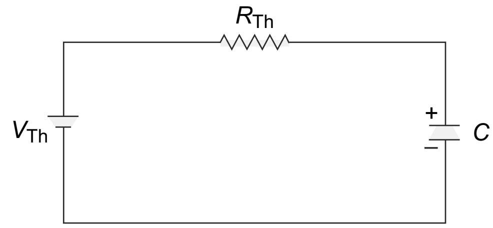


Fig. 5.39 *Example 5.9: Thevenin equivalent circuit*

The short-circuit current I_{SC}, which is the value of I when the capacitor is replaced by a zero resistance conductor is

$$I_{\text{SC}} = 0.4 \text{ A}$$

Since the parallel combination of R_2 and 0 will ensure that no current will pass through R_2, we get

$$R_{\text{Th}} = \frac{V_{\text{Th}}}{I_{\text{SC}}} = 0.45 \ \Omega$$

The Thevenin equivalent circuit is thus, the usual one loop circuit with a capacitor charging. The potential difference across the plates of the capacitor, (the potential difference being zero at $t = 0$) is

$$V_{\text{d}} = V_{\text{Th}}(1 - e^{-\frac{t}{R_{\text{Th}}C}})$$

and the current is

$$I = -C\frac{dV_d}{dt} = -\frac{V_{\text{Th}}}{R_{\text{Th}}}e^{-\frac{t}{R_{\text{Th}}C}} = 0.4e^{-\frac{t}{R_{\text{Th}}C}}$$

The example above could be dealt with using Thevenin's theorem since there was only one capacitor in the circuit. For cases involving capacitors in more than one branch of the circuit this method will not work, nor will the general network analysis using Kirchhoff's laws. The mathematical methods needed to solve such circuits with DC sources are more complicated and beyond the scope of this book.

PROBLEM 5.16 Consider the circuit in Fig. 5.40. The switch S is initially open and a steady current is flowing in the circuit. The switch is closed at time $t = 0$. Find the charge on the and the voltage of the capacitor as a function of time. What is the energy stored on the capacitor when it is fully charged?

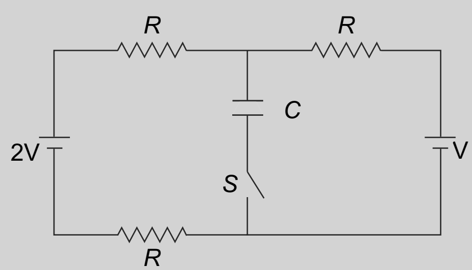


Fig. 5.40 *Problem 5.16*

PROBLEM 5.17 In Problem 5.16, after the capacitor is fully charged, we remove the batteries and replace them with resistanceless wires. Find the charge on the capacitor as a function of time, assuming we have reset the clock to $t = 0$ at the time when we have removed the batteries. Compute also the energy dissipated in the resistors.

5.10 ATMOSPHERIC ELECTRICITY *

Earth's atmosphere is one enormous laboratory of electrical phenomena, both static electrical phenomena as well as electrical currents. Since there are enormous variations in density, electrical conductivity and other parameters as we go up in the atmosphere starting from the surface of the earth, atmosphere has been divided into various layers. Characteristic phenomena happen in the various layers depending upon their electrical properties. The boundaries of these layers are not very sharp and are shown in Fig. 5.41

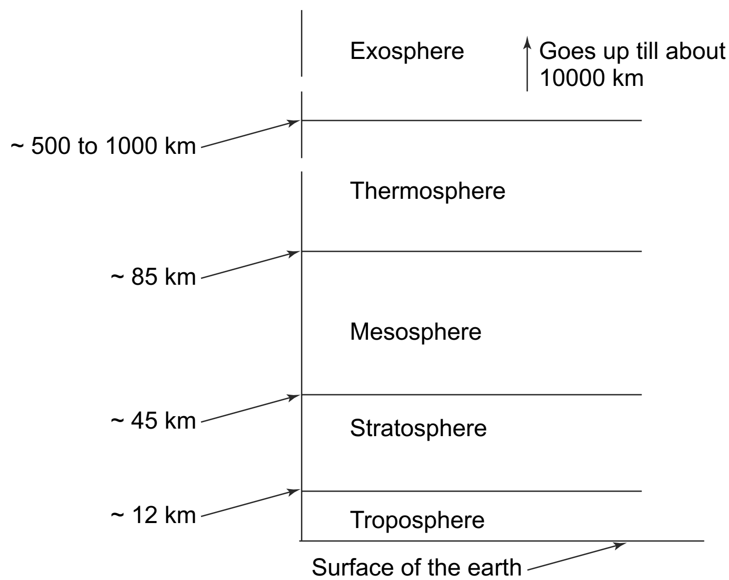


Fig. 5.41 *Various layers of the earth's atmosphere with approximate and average heights of the various layers. The ionised part of the atmosphere, called the ionosphere, starts from the bottom of the thermosphere and goes on in its upper layer well into the exosphere*

The most visible atmospheric electrical phenomena occur in the troposphere where most of the atmospheric air is present. Lightning is the most spectacular phenomenon that occurs in this layer.

The molecules of the air get electrically charged by a variety of processes like friction between masses of air, radiation from outside the atmosphere and other processes. The charge in the air can be of either sign but is mostly positive with the molecules losing electrons. These charges in the atmosphere induce opposite charges in the earth so that between them and the surface of the earth there is always an electric field. The electric field is enormous, something like 100 V/m. Air being non-conducting, in normal circumstances, this field causes no currents. However, there are occasions when the water vapour in the atmosphere condenses into water droplets. These water droplets, much denser than air, get charged as they moves through the atmosphere. Coalescing of two or more droplets results in bigger droplets. Sometimes thus, we get a condensed accumulation of electric charges in these drops and that creates very high electric field between the water drops and the earth's surface.

Every material has a limit such that when an applied electric field in that medium exceeds the limit, there is electrical breakdown of the medium and current flows through it. 'Lightning' is one such phenomenon when there is an electrical breakdown of the atmosphere resulting in huge currents flowing between the atmosphere and the earth for a short time till the charges gets neutralised. The potential difference when lightning occurs can be very high like 10^8 volts and the total current runs into thousand of amperes. It is also true that the flow of current between the atmosphere is not restricted to lightning events but occurs under fair weather conditions too. The average fair weather current has been estimated at about 10^3 Amperes. This arises mostly as an average combined effect of thousands of small lightning flashes that occurs which mostly are not spectacular and go unnoticed.

The layer above the troposphere is the stratosphere which is home to the so-called ozone-layer. The stratosphere does not have any direct electrical phenomenon associated with it. But this layer is primarily responsible for absorbing the harmful ultraviolet components of solar radiation thus making life on the surface of the earth possible. Presence of the ozone layer is a delicate balance between its formation from oxygen and breakup. Man-made emissions of fluorocarbon compounds disturb this delicate balance. Indications are that the ozone layer is already depleting which will lead to catastrophic consequences if allowed to go unchecked.

The next higher layer, the mesosphere has too thin a density of air to absorb the sun's radiation or show any significant electrical activity. The air there however, is able to offer substantial viscous resistance to incoming meteorites which sometimes burn up there, leaving tell-tale trails visible from the surface of the earth.

The next layer, the thermosphere contains the ionosphere, where free charged particles exist and thus, considerable electrical activity takes place. The ionisation of neutral atoms and molecules in the ionosphere is caused by very energetic ultraviolet rays which are a part of the solar radiation. The phenomenon of 'aurora' in the polar regions of the earth is a phenomenon of electrical discharge from the ionosphere. But the most significant phenomenon in the ionosphere occurs in relation to propagation of electromagnetic waves through it. Because of the sizable concentration of charged particles, it becomes transparent to high frequency electromagnetic waves. Radio waves are reflected by it but shorter waves used extensively in modern times for telecommunications pass though it.

SUMMARY

- Electric charges in motion constitute an electric current. Electric current is described by a current density $\vec{j}(\vec{r})$.

- For steady currents, conservation of charge implies that the divergence of the current density vanishes.
- To maintain an electric current in a conductor, a potential difference needs to be maintained across the ends of the conductor. This is usually supplied by a battery or a source which normally converts chemical energy to electrical energy.
- The current density in a conductor and the electric field are related by Ohm's Law. The proportionality constant is called the conductivity of the material.
- Circuits or combinations of resistors and batteries can be analysed using Kirchhoff's Laws, which are basically statements of Conservation of Charge and Conservation of Energy as applied to the circuits.
- For complicated circuits or networks, Thevenin and Norton's theorems can be used.
- Circuits with capacitors have a time-dependent current. The time constant is related to the product of resistance and capacitance.

CONCEPTUAL QUESTIONS

1. An ideal battery is hooked to a light bulb with wires. A second identical light bulb is connected in parallel to the first light bulb. After the second light bulb is connected, the current from the battery compared to when only one bulb was connected is
   a. higher
   b. lower
   c. the same

2. An ideal battery is hooked to a light bulb with wires. A second identical light bulb is connected in series to the first light bulb. After the second light bulb is connected, the current from the battery compared to when only one bulb was connected is
   a. higher
   b. lower
   c. the same

3. When a current flows in a wire of length L and cross-sectional area A, the resistance of the wire is
   a. proportional to A; inversely proportional to L
   b. proportional to both A and L
   c. proportional to L; inversely proportional to A
   d. inversely proportional to both L and A

4. An uncharged capacitor is connected to a battery, resistor and switch. The switch is initially open but at $t = 0$ it is closed. A very long time after the switch is closed, the current in the circuit is
   a. nearly zero

b. at a maximum and decreasing
c. nearly constant but non-zero
d. None of the above

5. A wire carries a steady current I. How is the current j density affected if
a. the length is doubled
b. the area is doubled
c. the length is doubled and the area is halved

6. The relationship

$$R = \frac{L}{\sigma A}$$

for a thin wire is
a. always true
b. always false
c. not necessarily true in all cases

7. A copper wire and an aluminium wire of the same length and cross section have the same potential difference across them. Do they carry the same current?

8. An ideal battery is hooked to a light bulb with wires. A second identical light bulb is connected in series with the first light bulb. After the second light bulb is connected, the light (power) from the first bulb (compared to when only one bulb was connected)
a. is four times higher
b. is twice as high
c. is the same
d. is half as much
e. is $\frac{1}{4}$ as much

9. An ideal battery is hooked to a light bulb with wires. A second identical light bulb is connected in parallel with the first light bulb. After the second light bulb is connected, the light (power) from the first bulb (compared to when only one bulb was connected)
a. is four times higher
b. is twice as high
c. is the same
d. is half as much
e. is $1/4$ as much.

10. The current in a simple series circuit is 5 A. When an additional resistance of 2 Ω is inserted, the current drops to 4 A. What is the resistance of the original circuit?

PROBLEMS

1. For $\vec{j} = \frac{1}{r^3}(2\cos\theta\hat{r} + \sin\theta\hat{\theta})$, calculate the current $\vec{I}$ passing through a hemispherical shell of radius 20 cm.

2. Resistance of a round, long wire of diameter 3 mm is 4.04 Ω/km. A current of 40 A flows through the wire. Find the conductivity of the wire and the current density.

3. A wire of resistance 6 Ω is drawn out into a new length which is three times the original length. Assuming that the resistivity and density of the wire is unchanged in the process, find the resistance of the new wire.

4. Two batteries of EMF $\mathcal{E}$ and internal resistance r are connected in parallel across a resistor R. For what value of R is the power delivered to the resistor a maximum and what is the value of this maximum power?

5. What is the equivalent resistance between the points x and y for the circuits shown in Fig. 5.42? All resistances are of 1 Ω each.

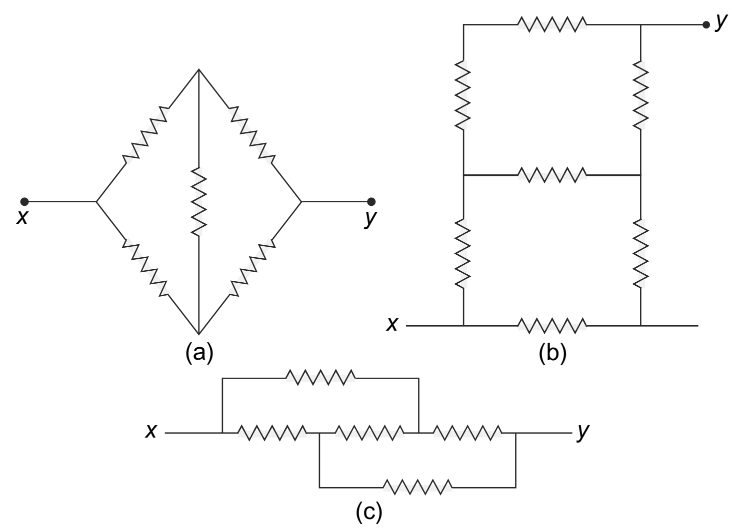


Fig. 5.42 *Problem 5*

6. Two thin metallic spherical shells A and B of radii R_A and R_B with $R_A > R_B$ respectively, are concentric. The outer one is maintained at a potential of 50 volts and the inner one at 100 volts. The space between the two is filled with conducting matter of resistivity $\rho = 2.6 \times 10^{-8}$ ohm-m. Calculate the total current flowing between the shells.

7. A capacitor of capacitance C is fully charged in a CR circuit with a battery of emf $\mathcal{E}$ and a resistor of resistance R. The emf of the battery suddenly falls to a value $\frac{\mathcal{E}}{4}$. Determine after how much time will the charge on the capacitor reduce to half its starting value.

8. Show that the units of the time constant of a circuit with a resistance R and a capacitance C are indeed the units of time.

9. A resistor $R = 3.0 \times 10^6$ Ω and a capacitor with $C = 1.0$ μF are connected to a source of emf with $\mathcal{E} = 4.0$ volts. At 1 seconds after the connection is made, find the rate at which the capacitor is charging, the rate at which heat is being dissipated in the resistor and the energy being delivered by the source.

10. The operating voltage of a 5 watt bulb is 25 volts between its terminals. The bulb has a resistance R connected in parallel across its terminals. The combination is connected in series with a resistor of resistance of 75 ohms and a DC power source of emf 75 volts and of internal resistance 8 ohms. Determine the value of R such that the bulb operates at its designated voltage across its terminals.

11. Two resistors of resistances R_1 and R_2 are connected in series with a battery of internal resistance r . With R_2, r and the emf of the battery fixed, what should be the value of R_1 so that the power dissipated in it is a maximum?

12. An infinite network of resistors in the form of a ladder is connected between terminals A and B which are at potentials V_A and V_B respectively. What is the value of the current I shown in the Fig. 5.43?

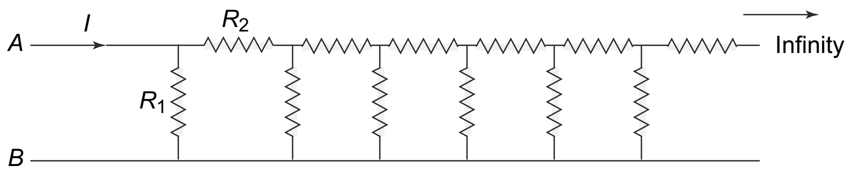


Fig. 5.43 *Problem 12*

13. Six equal resistors of resistance R each are connected to form a hexagon. Then six more resistors are connected between each of the vertices and the centre of the hexagon. Find the equivalent resistance between opposite vertices and also the equivalent resistance between adjacent vertices.

14. A Wheatstone Bridge circuit (Fig. 5.21) is made with substituting a galvanometer with resistance R_g for the resistance between points 2 and 3. The bridge is balanced with an unknown resistance R_4 by varying R_2. Find the current I through the galvanometer for the off-balance situation. We define the sensitivity of the bridge as $S = CR_2\left(\frac{\partial I}{\partial R_2}\right)_0$ where the subscript refers to the derivative has to be evaluated at balance and C is some galvanometer constant. Show that

$$S = \frac{C\mathcal{E}}{R_1 + R_2 + R_3 + R_4 + R_g(1 + \frac{R_3}{R_4})(1 + \frac{R_2}{R_1})}$$

15. Find the magnitude and direction of the current flowing in resistance R in the circuit shown in Fig. 5.44. Take $\mathcal{E}_1 = 1.5$ V, $\mathcal{E}_2 = 3.7$ V and $R_1 = 10\ \Omega$, $R_2 = 20\ \Omega$. You can take the internal resistances of the batteries to be negligible.

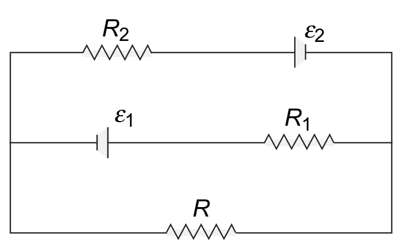


Fig. 5.44 *Problem 15*

16. In the circuit shown in Fig. 5.45, we have $\mathcal{E}_1 = 1.5$ V, $\mathcal{E}_2 = 2.0$ V, $\mathcal{E}_3 = 2.5$ V and $R_1 = 10$ Ω, $R_2 = 20\ \Omega$ and $R_3 = 30\ \Omega$. You can take the internal resistances of the batteries to be negligible.

a. Find the current through the resistance R_1.
b. Find the potential between the points A and B.

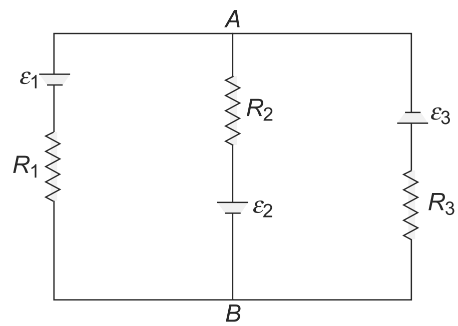


Fig. 5.45 *Problem 16*

17. In the circuit shown in Fig. 5.46, the resistance between the points A and B is 300 Ω. The long resistance is tapped at one-third points. What is the equivalent resistance between X and Y? If the potential difference between X and Y is 320 volts, find the potential difference between B and C.

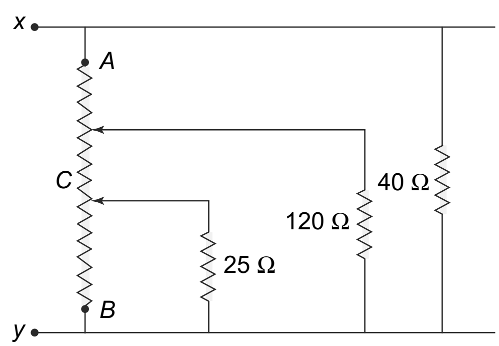


Fig. 5.46 *Problem 17*

18. In Fig. 5.47, find the energy in the 3 μF capacitor.

19. In the circuit in Figure 5.48, a current I_0 divides in some fashion amongst two resistance R_1 and R_2 in parallel. Show that the condition that $I_0 = I_1 + I_2$ and the requirement of minimum power dissipation leads to the same current values as one would get by analysing the circuit using Kirchhoff's laws.

20. A copper wire 10 km in length and 1.3 mm in diameter is connected to an emf source $\mathcal{E} = 24$ V. Determine the resistance of the wire, the current through the wire, the current density and the power dissipated as heat in the wire. The resistivity of copper is $1.72 \times 10^{-8}\,\Omega$m.

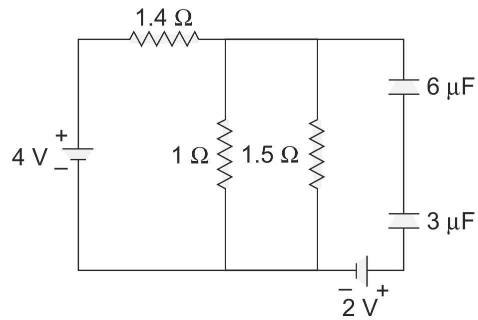


Fig. 5.47 *Problem 18*

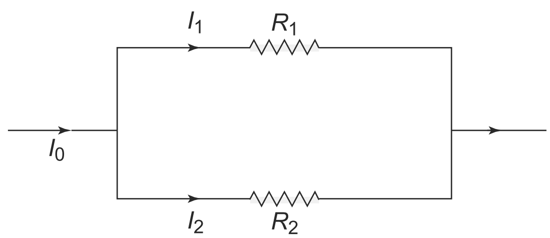


Fig. 5.48 *Problem 19*

21. In the circuit shown in Fig. 5.49, the two batteries have emf $\mathcal{E}_1$ and $\mathcal{E}_2$ with internal resistances r_1 and r_2 respectively. R is the load resistance. With all other quantities kept constant, what should be the value of R such that maximum Joule heat is produced in it?

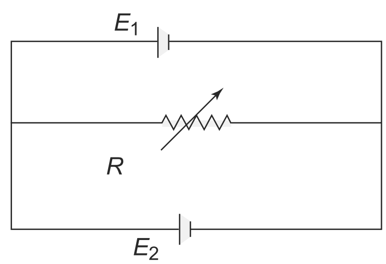


Fig. 5.49 *Problem 21*

22. The plates of a capacitor of capacitance $C = 5$ microfarad with initial charges $+Q$ and $-Q$ on them are connected with each other at time $t = 0$ through a resistor $R = 1$ megaohm. Initially, the gap between the plates was air but at time $t = 1$ second, the gap was instantaneously filled with oil of relative permittivity 5. Calculate the potential difference across the plates of the capacitor at time $t = 2$ sec.

23. A capacitor consists of two horizontal large parallel plates with a separation d and with a dielectric of relative permittivity ε_r in between. At time $t = 0$, the capacitor is fully charged and in a circuit in series with a resistor of resistance R and a battery of emf $\mathcal{E}$. The plates at $t = 0$ start moving vertically, the upper one with a speed v_1 and the lower one with a speed $v_2 < v_1$. Will there be a current through R? If so, calculate its value at times $t > 0$.

6
Magnetic Forces and Fields

Learning Objectives

- To be able to understand the force on a charged particle in a magnetic field.
- To learn about the motion of charged particles in various combinations of electric and magnetic fields.
- To learn about the Hall effect and its importance.
- To be able to explain the forces and torques on a current carrying loop in uniform and non-uniform magnetic fields.
- To learn about the sources and properties of the magnetic field and use Ampere's law to calculate fields for various configurations.
- To comprehend the idea of a magnetic loop as a magnetic dipole.
- To be able to understand the concept of magnetic vector potential and its relationship to the magnetic field.

6.1 MAGNETIC FORCE ON A CURRENT CARRYING CONDUCTOR

The history of magnetism is fairly old—naturally occurring substances called lodestones (which today we refer to as permanent magnets) were the first materials in which magnetic forces were observed. Till the early nineteenth century, the study of magnetism was separate from the study of electrical phenomenon and it was only in 1820 that H. Oersted and Ampere discovered the connection between these two phenomena. They discovered that magnetic forces are exerted over electric currents and also between two current carrying conductors. The work of these scientists as also of many others after them, led to a unified study of electromagnetism in which the electrical and magnetic phenomena are united.

Current carrying conductors experience a force when placed in a magnetic field. Consider a conductor carrying a steady current I along its length in a magnetic field $\vec{B}$. Then the force on an element $\vec{dl}$ of the conductor due to the magnetic field is given by

$$\boxed{d\vec{F} = I(\vec{dl} \times \vec{B})} \tag{6.1}$$

The force law is illustrated using the right-hand rule in Fig. 6.1.

In SI units, since the units of force, current and length are defined, this equation also defines the SI unit of magnetic field $\vec{B}$ which is called a Tesla, after the Serbian-American inventor Nikolai Tesla (1856–1943).

This empirical result for the force on a current carrying element in a magnetic field also allows us to determine the force experienced by a charge in a magnetic field. Recall that current is nothing but a flow of charge. Let λ be the charge density in the conductor which is carrying a current I. Then an element dl of the conductor carries a charge of

$$dq = \lambda dl$$

The velocity, $\vec{v}$ of the charge is related to the current by

$$\vec{v} dq = I d\vec{l}$$

Using Eq. (6.1), we get the force on the moving charge to be

$$\boxed{d\vec{F} = dq(\vec{v} \times \vec{B})} \tag{6.2}$$

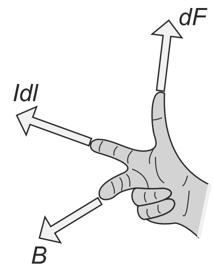


Fig. 6.1 *Right-hand rule for magnetic force on a current. The three fingers are at right angles to each other*

Thus, on any charge Q, in an electric and magnetic field, the total force will be the vector sum of the electric force (given by Coulomb's Law) and the magnetic force as in Eq. (6.2)

$$\boxed{\vec{F} = Q\left[\vec{E} + (\vec{v} \times \vec{B})\right]} \tag{6.3}$$

This force is called the Lorentz force on the charge Q.

It is interesting that the two kinds of forces have very different properties. Consider the displacement $d\vec{r}$ of the charge in an electric and magnetic field. The work done, by definition will be

$$\begin{aligned} dW &= -\vec{F} \cdot d\vec{r} \\ &= -Q(\vec{E} \cdot d\vec{r}) - Q\left(\frac{d\vec{r}}{dt} \times \vec{B}\right) \cdot d\vec{r} \\ &= Q d\phi \end{aligned} \tag{6.4}$$

where $d\phi$ is the electric potential difference across the displacement $d\vec{r}$ and the second term vanishes because of the properties of the vector triple product. This is a very interesting result. Motion of a charge in an electric field entails doing some work which, as we have seen before, is related to the difference in the electric potential of the electric field. **The magnetic field on the other hand, does not do any work on a charged particle in motion since the force exerted by the magnetic field is always perpendicular to the displacement.**

From Chapter 2, we also know that the conservative nature of the electric field or the fact that the work done is independent of the path taken, allows us to define a scalar potential which is related to the electric field by the gradient operator. For a magnetic field, this is not the case. We cannot define a scalar potential in the usual way. Later in this chapter, we will introduce a different kind of potential called the vector potential related to the magnetic field.

6.2 MOTION OF CHARGED PARTICLES IN ELECTRIC AND MAGNETIC FIELDS

The motion of a charged particle in electric and magnetic fields depends on the magnitude and orientation of the two fields. For a general configuration, this motion would be fairly complicated. We will discuss some special cases.

6.2.1 Motion in a Constant Electric Field

Let us consider first a constant electric field, like for instance the one present in between the plates of a parallel plate capacitor. Let the constant electric field be in the z-direction, i.e., $\vec{E} = E\hat{z}$. Then the equation of motion of a charge Q can be written as

$$m\frac{d^2\vec{r}}{dt^2} = \vec{F} = Q\vec{E}$$

For the special case of the electric field in the z-direction, we have thus

$$\begin{aligned} m\frac{d^2x}{dt^2} = m\frac{d^2y}{dt^2} &= 0 \\ m\frac{d^2z}{dt^2} &= QE \end{aligned} \tag{6.5}$$

These equations are second order, ordinary differential equations and thus can be solved if we know the initial position and velocity. Let the initial conditions on the position and velocity be

$$\begin{aligned} \vec{r}(0) &= (x(0), y(0), z(0)) \\ \vec{v}(0) &= (v_x(0), v_y(0), v_z(0)) \end{aligned}$$

Then the solution to Eq. (6.5) is

$$\begin{aligned} x(t) &= x(0) + v_x(0)t \\ y(t) &= y(0) + v_y(0)t \\ z(t) &= z(0) + v_z(0)t + \left(\frac{QE}{2m}\right)t^2 \end{aligned} \tag{6.6}$$

This is the general solution to the equations of motion for this configuration of the electric field. Let us take a specific initial condition of the particle being at the origin at $t = 0$, i.e., $\vec{r}(0) = 0$. We also take the initial velocity to be in the $x - z$ plane, i.e., $v_y(0) = 0$. With these initial conditions, we see that

$$\begin{aligned} x(t) &= v_x(0)t \\ y(t) &= 0 \\ z(t) &= v_z(0)t + \left(\frac{QE}{2m}\right)t^2 \end{aligned}$$

This is a trajectory of a parabola (Fig. 6.2(a)) as expected. The particle's motion in the direction with no force (x-direction) is without any acceleration. In the z-direction, the electric force causes the acceleration. The motion is like a particle projected with a velocity in a uniform gravitational field. The motion of the particle continues to be confined to the $x - z$ plane.

We next consider the case of a charged particle in a constant magnetic field.

6.2.2 Motion in a Constant Magnetic Field

A constant magnetic field will exert a force on a charged particle given by Eq. (6.3) with $\vec{E} = 0$. Let us take the direction of the constant magnetic field to be along the z-direction, i.e., $\vec{B} = B\hat{z}$. The equation

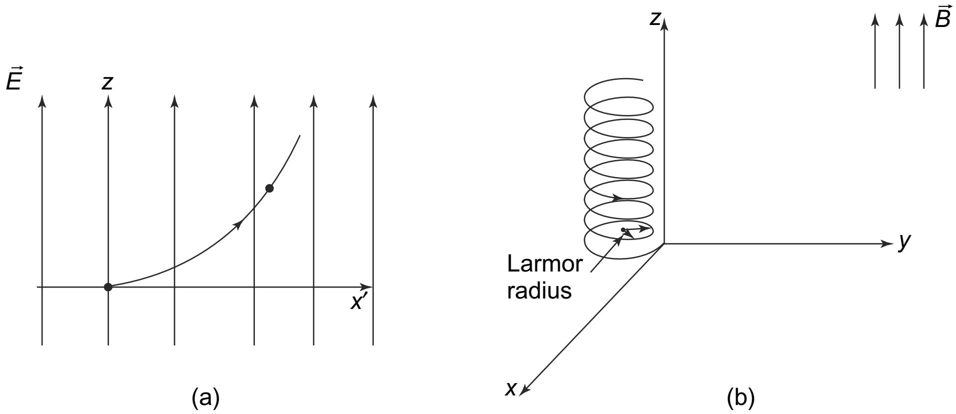


Fig. 6.2 *(a) Trajectory of a particle in the $x - z$ plane in the presence of a uniform Electric field in the z direction. (b) The same for a charged particle in a uniform magnetic field*

of motion can be easily written as

$$m\frac{d^2\vec{r}}{dt^2} = Q\left(\frac{d\vec{r}}{dt} \times \vec{B}\right) \tag{6.7}$$

Resolving it into its three components, we get

$$m\frac{d^2x}{dt^2} = \frac{QB}{m}\frac{dy}{dt} \tag{6.8}$$

$$m\frac{d^2y}{dt^2} = -\frac{QB}{m}\frac{dx}{dt} \tag{6.9}$$

$$m\frac{d^2z}{dt^2} = 0 \tag{6.10}$$

since as we have seen, there is no force in the direction of the magnetic field.

The z equation can be easily solved to get

$$z(t) = z(0) + v_z(0)t \tag{6.11}$$

where $z(0)$ is the initial position and $v_z(0)$ is the the initial velocity in the z-direction.

To solve the x and y equations of motion, we define a new variable

$$x_+ \equiv x(t) + iy(t)$$

In terms of x_+, the equations become

$$\frac{d^2x_+}{dt^2} = -i\frac{QB}{m}\frac{dx_+}{dt}$$

If we define

$$\omega \equiv \frac{QB}{m}$$

then the equation reads

$$\frac{d^2x_+}{dt^2} = -i\omega\frac{dx_+}{dt} \tag{6.12}$$

The quantity $\omega = \frac{QB}{m}$ has dimensions of frequency and is called the Larmor frequency. We will see later that this combination of charge, mass and magnetic field arises in many situations. Equation (6.12) is easily solved and we get

$$\frac{dx_+}{dt} = Ce^{-i\omega t} \tag{6.13}$$

where C is a constant which can be complex. Writing

$$C = |C|e^{i\delta}$$

we get

$$\frac{dx_+}{dt} = |C|e^{-i\omega t+i\delta} \tag{6.14}$$

which has a solution

$$x_+(t) = \frac{|C|e^{-i\omega t+i\delta}}{-i\omega} + D \tag{6.15}$$

where D is an integration constant. In general, once again, the integration constant D can be complex. We write

$$D = D_\mathrm{r} + iD_\mathrm{i}$$

Substituting this in Eq. (6.15) and taking the real and imaginary parts of this equation, we get

$$x(t) = -\frac{|C|}{\omega}\sin(-\omega t + \delta) + D_\mathrm{i} \tag{6.16}$$

and

$$y(t) = \frac{|C|}{\omega}\cos(-\omega t + \delta) + D_\mathrm{r} \tag{6.17}$$

There are, as expected four integration constants since each of the two equations for x and y are second order differential equations. We have chosen these constants to be $|C|$, δ, D_r and D_i, which are of course related to the initial values of position and velocity in the x and y directions. To determine the integration constants, we need to impose initial conditions. We take the the particle to be at the origin at $t = 0$ i.e. $x(0) = y(0) = z(0) = 0$. Then

$$x(t) = -\frac{|C|}{\omega}[\sin(-\omega t + \delta) - \sin\delta] \tag{6.18}$$

and

$$y(t) = +\frac{|C|}{\omega}[\cos(-\omega t + \delta) - \cos\delta] \tag{6.19}$$

From Eqs. (6.18) and (6.19), we can see that

$$\frac{dx(t)}{dt} = |C|\cos(-\omega t + \delta) \tag{6.20}$$

and

$$\frac{dy(t)}{dt} = |C| \sin(-\omega t + \delta) \tag{6.21}$$

These equations can be inverted and we can get the integration constant $|C|$ and δ in terms of the initial velocity components $\dot{x}(0)$, $\dot{y}(0)$.

$$|C| = \left(\dot{x}(0)^2 + \dot{y}(0)^2\right)^{\frac{1}{2}} \equiv v_\perp(0) \tag{6.22}$$

and

$$\tan \delta = \frac{\dot{y}(0)}{\dot{x}(0)} \tag{6.23}$$

where we have defined a velocity $v_\perp$ as the velocity in the direction perpendicular to the magnetic field. In our case, the magnetic field is in the z direction, so $v_\perp$ is in the $x - y$ plane or

$$v_\perp(0) = \left(\dot{x}(0)^2 + \dot{y}(0)^2\right)^{\frac{1}{2}}$$

This solves the problem completely. Now knowing the initial velocities, $\dot{x}(0)$, $\dot{y}(0)$, we can obtain the values of the position $\vec{r}(t)$ of the particle at any time.

The solution, Eqs. (6.18) and (6.19) have some interesting properties. Note that the time average values of $x(t)$ and $y(t)$ from Eqs. (6.18) and (6.19) are

$$\overline{x} = \frac{|C|}{\omega} \sin \delta \tag{6.24}$$

and

$$\overline{y} = +\frac{|C|}{\omega} \cos \delta \tag{6.25}$$

It is clear from Eqs. (6.18) and (6.19) that $x(t)$ and $y(t)$ oscillate around $\overline{x}$ and $\overline{y}$ such that

$$\begin{aligned}(x(t) - \overline{x})^2 + (y(t) - \overline{y})^2 &= \frac{|C|^2}{\omega^2} \\ &= \left(\frac{mv_\perp(0)}{|Q|B}\right)^2 \end{aligned} \tag{6.26}$$

In the $x - y$ plane then, the particle moves in a circular orbit. However, there is an initial velocity in the z direction and that velocity remains unchanged causing a uniform motion in the z direction. Thus, the trajectory of the particle, is in the form of a helix. The axis is the line $z = \dot{z}(0)t$, $x = \overline{x}$, $y = \overline{y}$ with the helical spiral around this line having a radius of

$$\boxed{a = \frac{mv_\perp(0)}{|Q|B}} \tag{6.27}$$

This radius is called the Larmor radius. Figure 6.2(b) is the sketch of this helix for $\delta = 0$. Note that for very high values of B, the Larmor radius shrinks to zero. The charged particle almost follows the magnetic field lines, which in our case is along the z-direction. If the initial velocity of the charged particle is orthogonal to the magnetic field, i.e., in this case $\dot{z}(0) = 0$, then the helix collapses to a circle in the $x - y$ plane. The particle then goes around in a closed, circular trajectory in the $x - y$ plane.

We consider two examples of motion of charged particles in electric and magnetic fields.

EXAMPLE 6.1 An proton with speed $v = 3.0 \times 10^6$ ms^{-1} enters the gap between the plates of a parallel plate capacitor, midway between the plates with its velocity parallel to the plates. The plates have a separation $d = 1$ cm, the potential difference between the plates is $V_+ - V_- = 100$ V and the length of the plates $AB = CD = 30$ cm. A fluorescent screen is placed at a distance D_0 from the further end of the plates. Determine the distance below the midpoint of the screen where the proton strikes the screen. Assume that the field is restricted to inside the capacitor and is normal to the plates.

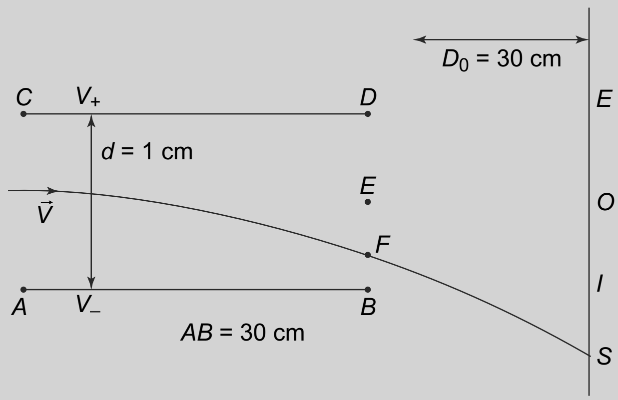


Fig. 6.3 *Example 6.1*

Solution

The initial velocity of the proton is at right angles to the electric field in between the capacitor plates. Hence, the velocity in the direction of the initial velocity remains unchanged. This implies that the proton takes a time

$$t_1 = \frac{AB}{v} = \frac{0.3}{3 \times 10^6} \text{ secs} = 10^{-7} \text{ secs}$$

to traverse the capacitor.

The electric field between the plates is

$$E = \frac{V}{d} = \frac{100}{0.01} = 10^4 \text{ V/m}$$

In the direction of the electric field, the proton experiences an acceleration of

$$a = \frac{eE}{m} = \frac{1.6 \times 10^{-19} \times 10^4}{1.6 \times 10^{-27}} = 10^{12} \text{ ms}^{-2}$$

The initial velocity in the direction of the electric field is zero and hence, after a time t_1, the displacement due to this acceleration will be

$$d_1 = \frac{1}{2}\frac{eE}{m}t_1^2 = 0.5 \text{ cm}$$

and the transverse velocity in this direction will be

$$v_1 = \frac{eE}{m}t_1 = 10^5 \text{ ms}^{-1}$$

At time t_1, the proton is at the position F as shown. After this point, there is no electric field and hence there is no transverse acceleration on the proton. The longitudinal (in the direction of the initial velocity) velocity continues to be v. Hence, the time taken for the proton to reach the screen would be

$$t_2 = \frac{D_0}{v} = 10^{-7} \text{ sec}$$

Thus, the additional displacement caused in the transverse direction is simply

$$d_2 = v_1 t_2 = \frac{eE}{m} t_1 t_2 = 1.0 \text{ cm}$$

Thus, the total displacement of the proton from its initial path in the transverse direction when it hits the screen is

$$d_1 + d_2 = 1.5 \text{ cm}$$

EXAMPLE 6.2 *ABCDEFGH* is a cube of side L and KL is a slit of width d in the face *EFGH* such that $EK = KF = GL = LH = \frac{L}{2} - \frac{d}{2}$. The face *ACHE* is open and a stream of charged particles of charge $+q$ and velocity v is normal to it. The inside of the cube has a magnetic field B in the direction normal to the faces *ABFE* and *DGHC*. Determine what fraction of incident particles pass through the slit. Given that $L = 10$ cm, $B = 10^{-1}$ T, $d = 2$ cm, $m = 1.6 \times 10^{-19}$ kg, $v = 3 \times 10^6$ ms^{-1} and $q = 1.6 \times 10^{-11}$ C.

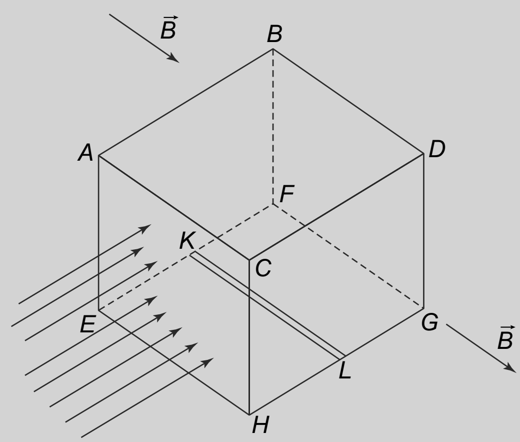


Fig. 6.4 *Example 6.2*

Solution

Let E be the origin with EH, EF and EA as the x, y and z axes respectively. Consider a particle at a height z_0 above the bottom face, that is the face with the slit. The magnetic field in the y direction will make the particle bend into an arc of radius

$$R = \frac{mv}{qB} = 0.3 \text{ m}$$

in the $y - z$ plane. The particles at the point of entry into the face $ABCD$ have coordinates $(x_0, 0, z_0)$ and their velocities are in the y-direction. Due to the Lorentz force, Eq. (6.3), the x-coordinate will remain unchanged but the particles will bend downwards. The centre of the arc will be at a distance R below z_0. The trajectory for this particle therefore, is given by

$$y^2 + (z - z_0 + R)^2 = R^2$$

When it hits the bottom face at $z = 0$, the y coordinate is given by

$$y_0^2 = R^2 - (R - z_0)^2 = 2Rz_0 - z_0^2$$

The only particle which will pass through the slit will be whose y-coordinate satisfies

$$\left(\frac{L-d}{2}\right) \leq y_0 \leq \left(\frac{L+d}{2}\right)$$

Thus, the values of z_0 for which the particles will pass through the slit will be

$$z_{0L} \leq z_0 \leq z_{0R}$$

where

$$2z_{0L}R - z_{0L}^2 = \frac{(L-d)^2}{4}$$

and

$$2z_{0R}R - z_{0R}^2 = \frac{(L+d)^2}{4}$$

Putting in the numbers and solving, we get

$$z_{0L} = 0.27 \text{ cm}$$

and

$$z_{0R} = 0.61 \text{ cm}$$

Only particles in the strip $0.61 - 0.27 = 0.34$ cm on the face $ACHE$ will pass through the slit. This represents a fraction

$$\frac{0.34}{10} = 0.034$$

or 3.4% of the particles incident on the face.

PROBLEM 6.1 A deuteron is travelling in a straight line along the x-direction with a speed 2×10^7 m/sec. At a point P it enters a region of uniform magnetic field of strength 0.2 w/m^2 (1 Tesla = 1weber m^{-2}) along the z-direction. How much will be the displacement in the y -direction from the original position P after it has travelled a distance of 1 cm in the x-direction? After how much time would the particle's velocity be instantaneously in the y-direction?

PROBLEM 6.2 At the equator, the strength of the earth's magnetic field is approximately 0.3×10^{-4} webers/m^2. If a piece of copper wire of 1 m length and cross-sectional area of 3×10^{-6} m^2 is to be kept afloat in a horizontal position perpendicular to the direction of the magnetic field, how much current needs to be passed through the wire. Assume $g = 10$ ms^{-2} and density of copper to be 9×10^3 kg/m^3.

PROBLEM 6.3 A rod of length L and cross-sectional area A has a current I flowing through it. It is pivoted at its centre which is at the origin and it is rotating with a constant angular speed ω in the $x - y$ plane about the z-axis. A uniform magnetic field is present all over in the z-direction of magnitude B. What will be the stress in the rod at a distance $x < \frac{L}{2}$ as a function of time?

6.2.3 Motion in Crossed Electric and Magnetic Fields

We have seen how a charged particle behaves in the presence of an electric field as well as in a magnetic field. The trajectories, a parabola in the electric field case and a helical path in the magnetic field case arise from the different nature of the two forces. The next obvious question we need to address is the behaviour of the particle in the presence of both, magnetic and electric fields.

We already know, that in the presence of both electric and magnetic fields, the motion of the charged particle is governed by the Lorentz Force Equation (Eq. 6.3). Using this force equation in the equations of motion, we can in principle, solve to get the trajectory of the particle. However, for fields with arbitrary magnitudes and orientations, the equations cannot be solved. For certain special cases, the force equation can be solved and some interesting trajectories obtained.

Let us first consider the case with uniform $\vec{E}$ and $\vec{B}$ fields present in the region $y > 0$. Further, let the fields be orthogonal to each other, that is, crossed. Let the electric field be along the z-axis, i.e., $\vec{E} = E\hat{z}$ and the magnetic field along the x-axis, i.e., $\vec{B} = B\hat{x}$. A charged particle of charge $+q$ and mass m, is initially at the origin $\vec{r}(0) = 0$ with a velocity $\vec{v} = v_0\hat{y}$ in the y direction. Assume $\frac{E}{B} > v_0$ so that the particle starts going up at $t = 0$. The analysis for the case $\frac{E}{B} < v_0$ is similar.

The equation of motion for the particle is given by

$$\vec{F} = q(\vec{v} \times \vec{B} + \vec{E})$$

which when resolved into components with the given fields, gives us

$$m\frac{d^2z}{dt^2} = qE - qB\frac{dy}{dt} \tag{6.28}$$

$$m\frac{d^2y}{dt^2} = qB\frac{dz}{dt} \tag{6.29}$$

$$m\frac{d^2x}{dt^2} = 0 \tag{6.30}$$

Notice that at $t = 0$, the particle is at the origin and the velocity in the x-direction is zero. Thus, Eq. (6.30) shows that $x(t) = 0$ for all t. The motion thus is restricted to the $y - z$ plane. Defining

$$y' = y - \frac{E}{B}t \tag{6.31}$$

we can combine Eqs. (6.28) and (6.29) to get

$$\frac{d^2(z + iy')}{dt^2} = \frac{iqB}{m}\frac{d(z + iy')}{dt} \tag{6.32}$$

This equation can be solved immediately to get

$$\frac{d(z+iy')}{dt} = A_1 e^{\frac{iqBt}{m}} \tag{6.33}$$

where A_1 is an integration constant. To determine the integration constant, we use the initial conditions given. At $t = 0$,

$$\frac{d(z+iy')}{dt} = i\frac{dy'}{dt} = i\left(\frac{dy}{dt} - \frac{E}{B}\right) = i\left(v_0 - \frac{E}{B}\right)$$

and thus,

$$A_1 = -i\left(\frac{E}{B} - v_0\right)$$

Now integrating Eq. (6.33), we get

$$z + iy' = -\frac{iA_1}{qB/m} e^{i\frac{qBt}{m}} + A_2 \tag{6.34}$$

where A_2 is the second integration constant. Since at $t = 0$, the particle is at the origin, we get

$$A_2 = \frac{iA_1}{qB/m}$$

and thus,

$$z + iy' = \frac{iA_1}{qB/m}\left(1 - e^{-\frac{iqBt}{m}}\right) \tag{6.35}$$

Separating real and imaginary parts and writing

$$\frac{qB}{m} = \omega$$

the Larmor frequency, we get

$$\boxed{z = \frac{\frac{E}{B} - v_0}{\omega}(1 - \cos\omega t)} \tag{6.36}$$

and

$$\boxed{y = \frac{E}{B}t - \left(\frac{E}{B} - v_0\right)\frac{\sin\omega t}{\omega}} \tag{6.37}$$

The velocity along the y-direction is thus,

$$\dot{y} = \frac{E}{B} - \left(\frac{E}{B} - v_0\right)\cos\omega t \tag{6.38}$$

If one takes the average of this, we get $u = \frac{E}{B}$. **This is called the drift velocity**. Thus, while y increases linearly as u, z oscillates between 0 and $\frac{2\left(\frac{E}{B} - v_0\right)}{\omega}$. This kind of trajectory is shown in Fig. 6.5. Taking

$$R = \frac{E/B - v_0}{\omega}$$

Equations (6.36) and (6.37) can be written as

$$(z - R)^2 + \left(y - \frac{E}{B}t\right)^2 = R^2 \tag{6.39}$$

This is an equation of a circle but one with its centre which changes with time. The coordinates of the centre are

$$\left(0, \frac{E}{B}t, R\right)$$

Thus, in a crossed electric and magnetic field, the particle traces a path which is called a cycloid.

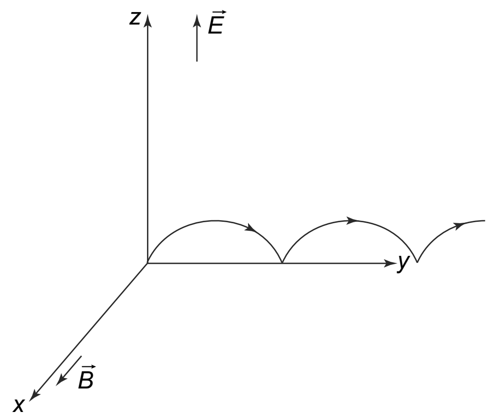


Fig. 6.5 *Trajectory of a charged particle when the electric and magnetic fields are orthogonal to each other. The particle is initially at the origin. The trajectory shown is for $\frac{E}{B} > v_0$. For $\frac{E}{B} < v_0$ the trajectory will go down at the origin and curve upwards*

Let us now see what the trajectories are for some special values of the parameters. Note that for $v_0 = \frac{E}{B}$, Eqs. (6.36) and (6.37) give us $z(t) = 0$ while y increases linearly with time with its initial speed v_0. **This fact is used in a device called the velocity selector.**

Consider a beam of particles with a distribution of velocities in the y-direction, entering a region of crossed E and B fields at $y = 0$. Particles with $\frac{E}{B} > v_0$ will trace out a cycloid path with positive z and those with $\frac{E}{B} < v_0$ a path with negative z. Particles with $\frac{E}{B} = v_0$ will not be affected at all. Hence, if a barrier parallel to the $x - z$ plane is introduced at some value of y, with a small aperture at $x = z = 0$ then only those particle with $\frac{E}{B} = v$ will pass through the aperture. This device can thus select particles with a particular velocity out of a beam of particles with a distribution of velocities.

Our analysis above has one limitation—the trajectory of the particle is a cycloid, no matter what the value of $\frac{E}{B}$ is. In particular, if $\frac{E}{B} > c$, the velocity of light, our analysis implies that the average speed of the particle is more than c. This is in violation of the principle of special relativity. Clearly, the non-relativistic equations of motion used by us are not valid when one is dealing with velocities close to the speed of light. A more detailed analysis shows that the results above are valid only when $\frac{E}{B}$ is much less than the speed of light.

6.2.4 The Cyclotron

The interaction of charged particles with electric and magnetic fields and their motion in them is used in several devices. In particular, particle accelerators are based on this interaction. The first such accelerator was the cyclotron which was built by E.O.Lawrence and M.S. Livingstone in the early 1930s. It has been extensively used to study elementary particles and their properties and interactions. Though particle accelerators have advanced technologically significantly since the time the cyclotron was invented, it is instructive to study the cyclotron since its principles are used even in more advanced accelerators.

The basic principle of the cyclotron is that a charged particle is accelerated in an electric field. This by itself was known and had been used for a long time, in, for instance, X-ray tubes, etc. The ingenuity in the cyclotron design was to use the interaction of the charged particle with a magnetic field. Recall that the magnetic field does not increase the kinetic energy of the particle but only changes the direction of motion. Using a combination of electric fields (to accelerate the charged particles) and magnetic fields in a ingenious way, Lawrence was able to build a compact accelerator which could accelerate particles to high energies. Let us see how this was done.

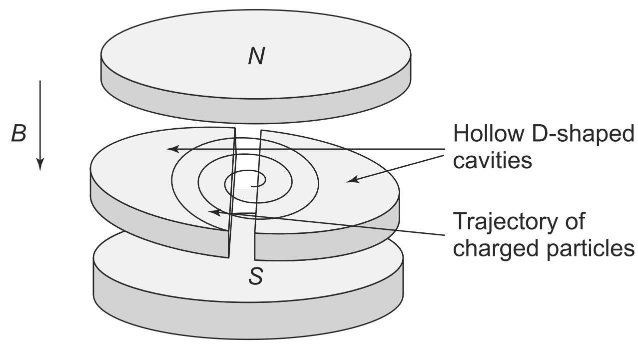


Fig. 6.6 *The Cyclotron: The magnetic field in the cavities is vertical and a varying electric field is applied in the gaps. The electric field is suitably timed so as to accelerate the charged particles when they enter the gap*

The device consists of two D-shaped cavities placed horizontally with their straight edges separated by a gap. A vertical, steady magnetic field is applied such that the D-shaped cavities have a uniform magnetic field in them which is at right angles to their plane surfaces. Charged particles are injected at the centre of the gap between the D shaped cavities. The initial velocity, v_0 of the charged particles is in the horizontal direction, i.e., in the plane of the D-shaped cavities. The perpendicular magnetic field exerts a force which is also in the horizontal direction, but orthogonal to the direction of the initial velocity.

$$F = qvB \tag{6.40}$$

The charged particles, under the influence of this magnetic force start moving in a circular trajectory with radius R such that the magnetic force provides the centripetal force for the circular motion.

$$qvB = \frac{mv^2}{R}$$
$$R = \frac{mv}{qB} \tag{6.41}$$

where m is the mass of the particle. The time taken, $T_{\frac{1}{2}}$ by the particle to complete one semicircle in one of the D shaped cavities is thus,

$$T_{\frac{1}{2}} = \frac{\pi R}{v} = \frac{\pi m}{qB} \tag{6.42}$$

Note that this is independent of v and R. As the particle enters the gap between the D-shaped cavities, it is accelerated by an electric field of the appropriate sign as shown in the Fig. 6.6. The particle is accelerated while it is traveling in the gap (the only region where there is an electric field) and renters the other D-shaped cavity but this time with a higher energy and velocity than with which it had emerged from the other D-shaped cavity. It reaches the other end of the cavity, traveling in a semi circular path because of the action of the magnetic field, exactly after a time $T_{\frac{1}{2}}$ though it has a higher energy and velocity! The particle is travelling in a direction opposite to its initial velocity when it emerges out of the D-shaped cavity in to the gap once again. So the direction of the electric field is reversed and it once again accelerates the particle. The electric field thus has to have the correct frequency to coincide the change in its direction exactly with the entry of the particle in the gap. This process continues and with each round, the particle moves in a circular path with larger and larger R since $R \propto v$. It ultimately emerges out of the cavity with an energy much higher than its initial energy.

This simple picture that is presented above was the basic principle of the original cyclotron. It was soon realised that this picture is only valid in a certain domain. The crucial thing in the principle is that the frequency of the electric field (which is required to accelerate the particles in the gap) is independent of the velocity and radius of the particle or its path. This is clear from Eq. (6.42). However, this expression is valid only in the non-relativistic domain. When the particle velocities are comparable to c, the speed of light, this equation is not valid. Hence, the electric field frequency required for repeated acceleration is dependent on the energy of the particle when the particles reach velocities comparable to c. This is possible with another device, called a synchrotron which was developed soon after the cyclotron.

After the development of the cyclotron, many different kinds of particle accelerators have been developed with increasing energies. These extremely high energies are required to probe the structure of matter at smaller and smaller distances. Currently, the highest energy accelerating machine is called the Large Hadron Collider (LHC) at CERN near Geneva. It can accelerate protons such that their energies are almost 7,500 times their rest mass energy. The basic principle remains the same in all these accelerators—particles are accelerated in an electric field and their trajectories are modified using a magnetic field.

6.2.5 Motion of Charged Particles in Non-uniform Magnetic Fields

We have studied the motion of charged particles in uniform magnetic fields. However, the magnetic fields encountered in most situations are not uniform and hence the analysis developed above is strictly not valid. The solution to the problem of the motion of a charged particle in a general non-uniform magnetic field is a very complicated one and we shall not be considering it. However, there are certain kinds of fields, which, though non-uniform, allow us to use the above analysis in an approximate way.

One such situation is where the magnetic field is non-uniform but only over distance scales which are much larger than the Larmor radius (Eq. (6.27)). In this case, some of the analysis can be used in an approximate fashion over small regions where the field can be considered to be uniform. One such example is when the magnetic field lines converge towards a particular direction as shown in the Fig. 6.7.

The magnetic field that we consider has a dominant z-component and a much smaller x-component. Further, the field increases as z increases. A charged particle having a velocity v_z along z and $\vec{v}_\perp =$

chap6 Samples May 2, 2012 16:11

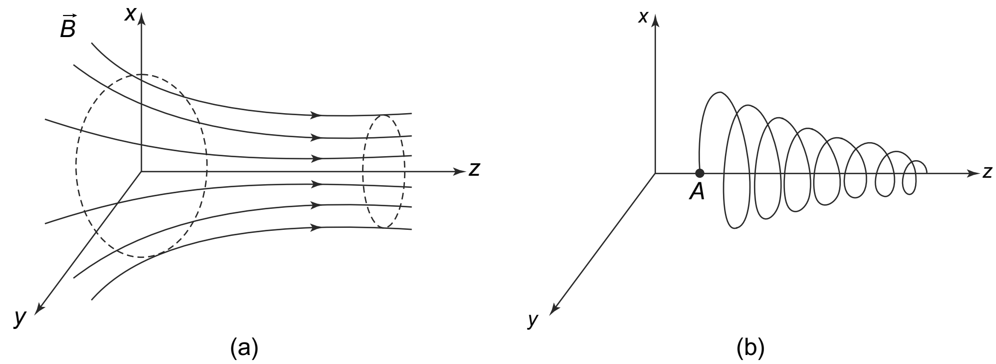


Fig. 6.7 *(a) Non-uniform field. As z increases, B_z increases but the magnitude of the transverse component to $\hat{z}$, $B_\perp$ decreases, (b) The particle traces out a helical path along the z axis. As z increases, the radius of the helix decreases*

(v_x, v_y) in the transverse direction at some point, will locally move along a helical path as discussed above. This is because locally, the particle will experience a uniform field since the field is changing (is non-uniform) over a distance much larger than the Larmor radius ($a = \frac{mv_\perp}{|q|B}$). The radius of the helix is

$$r = \frac{m|v_\perp|}{qB_z}$$

where B_z is the component of the magnetic field along the z direction. The particle at the point A in the Fig. 6.7 will also move along the z-direction with velocity v_z. The energy of the particle is given by

$$E = \frac{1}{2}m(v_z^2 + v_\perp^2)$$

which is not changed by its motion in the magnetic field.

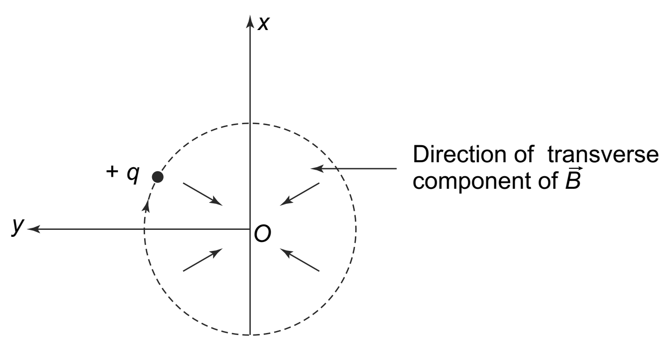


Fig. 6.8 *Profile of the trajectory. The z-direction is coming out of the plane of the paper as is the longitudinal magnetic field. The direction of the magnetic force $q(\vec{v} \times \vec{B})$ due to the transverse components of $\vec{B}$ is along the $-z$ direction and hence it slows down the velocity along z-direction*

If the field was constant, along the z-direction, the positively charged particle would trace a trajectory whose projection on the $x - y$ plane is as shown in Fig. 6.8. However, the field is not uniform and has a transverse component. The force on the particle due to this transverse component will be along the $-z$-direction and hence if the particle had an initial velocity v_z along the z-direction, this would decrease as the particle moves along the z-direction.

Now, since the energy of the particle cannot change in a magnetic field, any change in v_z will be compensated by a corresponding change in $v_\perp$. In this case, as v_z decreases, $v_\perp$ will increase and as the particle moves forward in the z-direction, it would slow down. The Larmor radius

$$a = \frac{mv_\perp}{qB_z}$$

will decrease and the helical path will start to look like what is shown in Fig. 6.7(b).

In the configuration of the magnetic field shown, if the pattern of the magnetic field extends along the z-axis sufficiently, there will come a time when v_z vanishes. The deceleration along z-axis will still be there and that means the particle will start moving backwards from that point. The trajectory of the particle will thus retrace back from that point. Such a magnetic field thus acts like a **'magnetic mirror'**. A more interesting configuration for the magnetic field would be like Fig. 6.9 where the B_z starts tapering off also in the negative z-direction at some value of z. The motion of the particle in the left half of the field would be the same as the right half reflected in the $x - y$ plane. The particle on its way back after reflection from the right end enters the left half ultimately, where it again gets reflected at some point in the left half. This to and fro motion obviously continues indefinitely. Such a device is called a **'magnetic bottle'** which essentially confines charged particles within a finite volume.

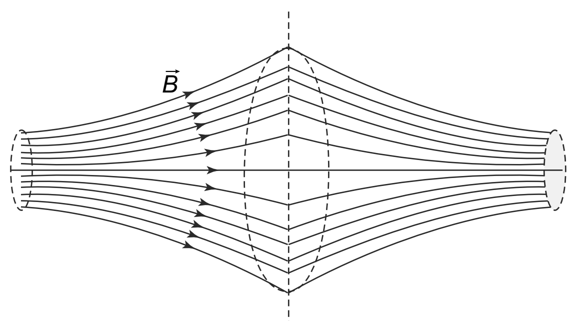


Fig. 6.9 *Magnetic confinement bottle. The magnetic field becomes stronger as one goes from the centre towards the right or left. The charged particles execute a helical motion as they go to and fro from one end to the other*

Aurora and Van Allen Belts

The confinement of charged particles in certain kinds of magnetic fields also occurs in nature. For instance, the magnetic field of the earth, shown in Fig. 6.10 is an example of a non-uniform field. The field is the strongest near the magnetic poles and falls as we go far from them. This field configuration is very similar to the one we have studied above and so, in a similar manner the Earth's magnetic field will trap charged particles.

The source of the charged particles is mostly the Sun, which spews them out constantly in the form of solar wind. Most of these particles, on encountering the earth's magnetic field are deflected away from the earth, but some of these do get trapped. This trapping leads to some spectacular natural phenomena.

The formation of Auroras at the poles (Aurora Borealis near the northern pole and Aurora Australis near the South Pole) is one such phenomenon. The trapped particles near the poles, where the field is the strongest, collide with the molecules in the atmosphere and produce radiation which is what is seen as the northern or southern lights at very high latitudes.

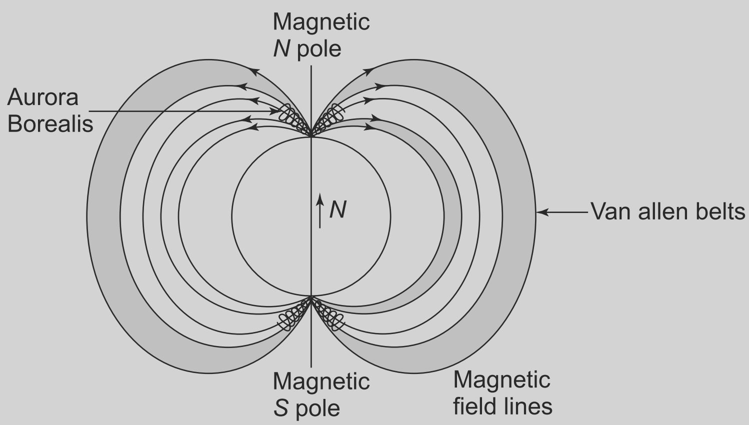


Fig. 6.10 *Particle confinement in the earth's magnetic field. The shaded portions are the Van Allen belts. The inner one is about 500–6000 kilometres above the earth while the outer one is about 15000–30000 kilometres above*

At much higher altitudes, the trapped particles lead to the formation of regions called Van Allen belts. These torus-shaped regions, at very high altitudes were discovered in the 1958 by James Van Allen in rocket-based surveys of the atmosphere. Interestingly, similar regions of trapped charged particles have been discovered around other planets in the solar system.

6.3 HALL EFFECT

The Hall effect, discovered by E.J.Hall in 1879 is a direct manifestation of the forces on charged particles in motion by magnetic fields. Its importance lies in the fact that the experimental measurements done by Hall to measure the tiny effect preceded the discovery of the electron by almost a decade or more.

Let us consider a conductor placed in a magnetic field. An electric current is passed through the conductor in a direction which is perpendicular to the magnetic field. Now, as we have seen, current is basically the flow of charge. The charge carriers in motion, will experience a magnetic force in the presence of the magnetic field. This force, given by Eq. (6.3) would be perpendicular to both the direction of the current and the magnetic field. The charge carriers will move under the influence of the force. However, the motion would be short-lived. This is because the charge carriers would soon reach the edges of the conductor and accumulate there being unable to cross the barrier presented to them.

This accumulation of charge on one side of the conductor would create an electric field which would be in a direction which is opposite to the magnetic force (not the magnetic field which is perpendicular to the force). At equilibrium, the electric force caused by the field due to the accumulated charges and the magnetic force would balance each other and there would be no further movement of the charge carriers in a direction perpendicular to the magnetic field.

The presence of the electric field due to the accumulated charge carriers on one side of the conductor will obviously lead to the a voltage perpendicular to the direction of the current. This voltage was measured by Hall and is called **Hall Voltage**. The Hall effect is shown in Fig. 6.11.

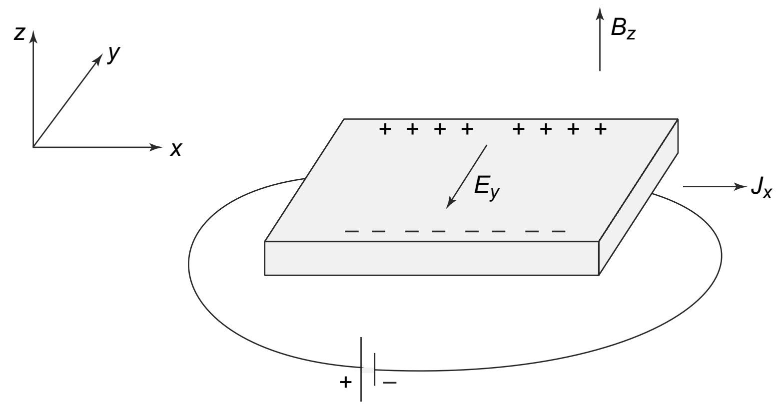


Fig. 6.11 *Hall effect: An electric field in the x-direction in the conductor creates a current density j_x in the x-direction. B_z is a uniform magnetic field in the z-direction over the conductor. Charges are piled up at the two edges in the y direction creating an electric field E_y as shown. At equilibrium current in the y-direction, j_y is zero*

The Hall effect as described above can be easily understood in terms of the free electron or Drude's model for conduction of electric current in conductors. Recall that according to this model, conduction in most conductors is due to the negatively charged electrons carrying a charge $-e$ where $e > 0$. If n is the number density of these electrons and v is their drift velocity along the x-axis due to an applied electric field (or voltage), then the current density is in the x-direction and is given by

$$j_x = \sigma E_x \tag{6.43}$$

where σ is the conductivity and is given by

$$\sigma = \frac{ne^2\tau}{m}$$

Here m is the mass of the electron and τ is the collision time.

Now the charges in motion along the x-direction will experience a force due to the magnetic field $\vec{B} = B_z\hat{k}$. This force will be

$$\begin{aligned}\vec{F} &= (-e)(\vec{v} \times \vec{B})\\ &= (-e)(v_x\hat{i} \times B_z\hat{k})\end{aligned}$$

$$= evB_z\hat{j} \tag{6.44}$$

As we have seen above, this force will result in an accumulation of charges at the edges in the y-direction. Hence, an electric field E_y will be created in the y-direction which will result in a force on the electrons given by

$$F_E = (-e)E_y$$

At equilibrium, the two forces, both of which are in the y-direction must add to zero. So we have

$$(-e)E_y + evB_z = 0 \tag{6.45}$$

or

$$E_y = vB_z \tag{6.46}$$

But

$$v = -\frac{j_x}{ne}$$

and so we have

$$E_y = -\frac{j_x B_z}{ne} \tag{6.47}$$

The Hall voltage is given by $V_H = dE_y$, where d is the width of the sample in the y-direction. We can define a quantity called the **Hall coefficient,** R_H as

$$R_H = \frac{E_y}{j_x B_z} \tag{6.48}$$

Then Eq. (6.47) can be written as

$$\boxed{R_H = -\frac{1}{ne}} \tag{6.49}$$

This result is very surprising: the Hall coefficient (Eq. (6.48)), which can be experimentally measured by measuring magnetic field strength B_z, current and voltage, provides information on the charge carrier, density n! For negatively charged carriers n is a negative number. Since this derivation of the Hall coefficient is based on the classical Drude model of electrical conduction, experimental verification of Eq. (6.49) can be used to test the model.

Experimentally, it is seen that Eq. (6.49) is not quite accurate except in some metals and for very low magnetic field strengths. This should not be surprising since the model on which it is based is classical in nature while the correct way of explaining conduction is based on quantum theory.

We now give an example which gives some idea of the magnitude of the Hall coefficient.

EXAMPLE 6.3 A conductor, of rectangular shape has the dimensions as shown in Fig. 6.12. The voltage across AD is 10 mV and the conductivity of the material is $\sigma = 6 \times 10^7$/ohm m. A current I flows in the direction as shown in the conductor. The conductor is placed in a magnetic field of magnitude 1 Tesla in the direction as shown in the figure. Calculate the Hall voltage and Hall coefficient given that the carrier density is 8×10^{28} m^{-3}.

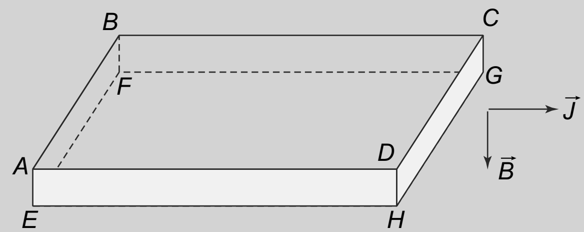


Fig. 6.12 *Example 6.3* $AB = 20$ cm, $AE = 1$ mm, $AD = 1$ m

Solution

The electric field can be calculated from the voltage across AD as

$$E = \frac{10^{-3}}{1}\frac{\text{V}}{\text{m}} = 10^{-3}\ \text{V/m}$$

With this field, the current density can be determined

$$j = \sigma E = 6 \times 10^7 \times 10^{-3} = 6 \times 10^4\ \text{A/m}^2$$

and the Hall coefficient

$$R_H = -\frac{1}{8 \times 10^{28} \times 1.6 \times 10^{-19}} \sim -8 \times 10^{-11}\ \text{m}^3/\text{C}$$

The electric field in the direction AB is

$$E_y = |R_H jB| = 8 \times 10^{-14} \times 6 \times 10^4 \times 1 = 4.8 \times 10^{-9}\ \text{V/m}$$

And hence, the Hall voltage is

$$V_H = E_y \times 0.2 = 0.96 \times 10^{-9}\ \text{V}$$

6.4 FORCE ON A CURRENT CARRYING LOOP

6.4.1 Force and Torque on a Current Carrying Loop in a Uniform Field

We have seen how a current carrying conductor, when placed in a magnetic field experiences a force. If the current carrying conductor is in the form of a loop, then the geometry of the loop makes the interaction with the magnetic field more complicated. We will first see what happens in the case the loop is a rectangular loop and then generalise to a general geometry.

Consider a planar loop $OPQR$ carrying a current I. As shown in Fig. 6.13, we take the plane of the loop to be the $x - y$ plane and O as the origin.

A uniform magnetic field $\vec{B} = B_x\hat{i} + B_y\hat{j} + B_z\hat{k}$ is present throughout. Consider two infinitesimal elements of the loop, each of length dl on the two opposite arms, OP and QR of the loop. The current flowing in these two elements, I, is equal but in opposite directions and the forces on these elements are

$$d\vec{F}_1 = Idl(\hat{i} \times \vec{B}) \qquad (6.50)$$

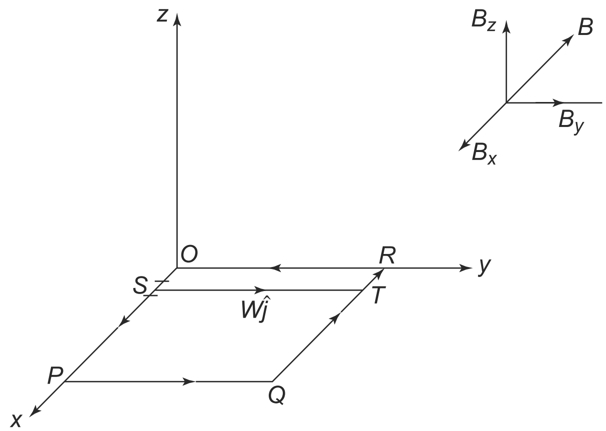


Fig. 6.13 *A rectangular loop in a uniform magnetic field B. 1 and 2 are infinitesimal elements located at S and T having positions $OS = \vec{r}_1$ and $OT = \vec{r}_2$. $\overrightarrow{TS}$ is parallel to $\hat{j}$*

and

$$d\vec{F}_2 = Idl(-\hat{i} \times \vec{B}) \tag{6.51}$$

It is obvious that these forces add up to zero. Now we can break up the entire loop into such pairs of elements with equal and opposite currents and thus the **net force on the loop in a uniform magnetic field is zero**.

However, since the loop is an extended object with equal and opposite forces acting on different parts, this will lead to a torque. For the two elements above, the torques are

$$d\vec{\tau}_1 = Idl\vec{r}_1 \times (\hat{i} \times \vec{B}) \tag{6.52}$$

and

$$d\vec{\tau}_2 = -Idl\vec{r}_2 \times (\hat{i} \times \vec{B}) \tag{6.53}$$

But

$$(\vec{r}_1 - \vec{r}_2) = -W\hat{j}$$

where $W\hat{j}$ is the line TS joining T to S. Thus, the sum of these two torques is simply

$$d\vec{\tau} = d\vec{\tau}_1 + d\vec{\tau}_2 = -IdlW\hat{j} \times (\vec{i} \times \vec{B}) = -IdlWB_y\hat{i} \tag{6.54}$$

Adding up all the torques on infinitesimal elements along *OP* and *QR*, we get

$$\vec{\tau}_1 = -ILWB_y\hat{i} \tag{6.55}$$

We can do a similar exercise for the two arms OR and PQ and get

$$\vec{\tau}_2 = ILWB_x\hat{j} \tag{6.56}$$

Thus, the total torque on the loop is

$$\vec{\tau} = \vec{\tau}_1 + \vec{\tau}_2 = ILW(B_x\hat{j} - B_y\hat{i}) \tag{6.57}$$

We can now define the area vector of the loop. This is a vector whose magnitude is the area of the loop and direction is the outward normal.

$$\vec{A} = WL\hat{k}$$

With this, the total torque on the loop can be written in a compact form as

$$\boxed{\vec{\tau} = I(\vec{A} \times \vec{B})} \tag{6.58}$$

We further define a quantity called the **magnetic dipole moment associated with the current carrying loop.**

$$\boxed{\vec{\mu} = I\vec{A}}$$

In terms of the dipole moment, the total torque thus, becomes

$$\boxed{\vec{\tau} = (\vec{\mu} \times \vec{B})} \tag{6.59}$$

The fact that the total torque on a current carrying loop is dependent on the dipole moment of the loop seems to indicate that this result, though derived for a rectangular loop, might be valid for an arbitrary shape. This is indeed the case. To see this, consider a loop as in Fig. 6.14.

We divide the loop into a large number of rectangular loops which in the limit are infinite in number. The arms next to the boundary carry a current I while the inner rectangles have a current I going one way and the same current going in the opposite direction. Thus, the only arms where there is a net current are the outermost arms which in the limit of the number of rectangles going to infinity become the boundary of the loop of arbitrary shape. For each of the inner rectangles, the result for the net torque applies. The normal to all the loops (which is the direction of the area vector) is in the same direction. Thus, the torque due to the rectangles adds up and we get the torque as in Eq. (6.59) except that now the area entering the dipole moment is the area of the loop.

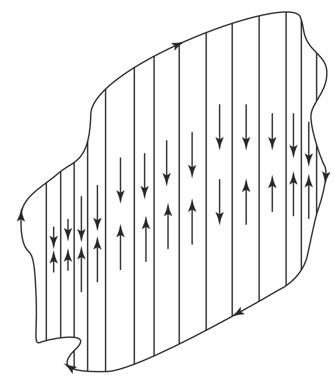


Fig. 6.14 *A loop of arbitrary shape divided into rectangles. In the limit of the number of rectangles becoming infinite, the outermost arms form the loop and the net current in all but the outermost arms is zero*

It is interesting to note that the expression for the torque on a magnetic dipole is very similar to the torque on an electric dipole, with the magnetic field replaced by the electric field and the magnetic dipole moment by the electric dipole moment. The magnetic dipole is made of a current carrying loop while an electric dipole was made of two equal and opposite charges separated by a distance. Whether we can conceive of magnetic dipoles in a similar fashion, i.e., made up of two equal and opposite magnetic 'charges', is something we shall explore in later chapters.

6.4.2 Moving Coil Galvanometer

We have seen in Section 6.4.1 that a current carrying loop in a uniform magnetic field experiences no net force. Instead, it experiences a torque which is given by Eq. (6.59). Recall that the magnetic moment of a loop is directly proportional to the current in the loop. These two facts can be used to construct devices which detect or measure currents in circuits. Any such device, which is used to measure or detect a current is called a **galvanometer**. The most common type of galvanometer is the moving coil galvanometer.

The basic construction of a moving coil galvanometer is shown in Figs. 6.15 and 6.16.

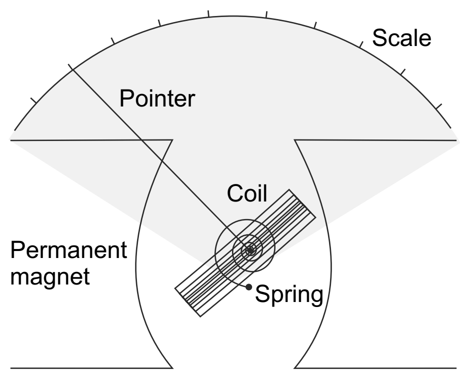


Fig. 6.15 *The basic construction of a pivot type moving coil galvanometer*

The source of the magnetic field is usually a permanent magnet. A current carrying coil C, which typically consists of several turns of insulated copper wire is wound on a rectangular frame. The coil is either suspended with a conducting wire or strip or mounted on a pivot with a spring, so that it can rotate in the region between the poles of the permanent magnet freely. In the case of a suspended coil, a small mirror is attached to the suspension wire while in the case of a mounted coil, a pointer is attached to the coil itself.

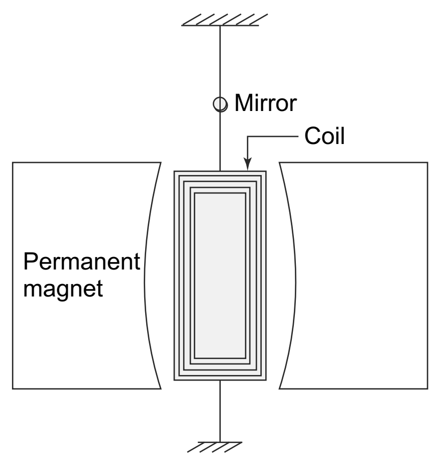


Fig. 6.16 *The basic construction of a suspended type moving coil galvanometer*

When a current I passes through the coil, it experiences a deflecting torque which is given by

$$\tau = NIBab \sin\theta \tag{6.60}$$

where N is the number of turns of wire in the coil, a and b are the length and breadth of the rectangular coil and θ is the angle between the normal to the coil and the magnetic field. In the case of the arrangement in the moving coil galvanometer, this angle is always 90°. Now under the influence of this deflecting torque, the coil, which is free to rotate, starts rotating. As soon as the coil rotates, the suspension wire or the spring exerts a restoring torque which is proportional to the angle of rotation. At some point, the deflecting and the restoring torques are equal and the coil comes to a rest at an angle ϕ to the initial equilibrium position. In the case of the suspension wire, this angle of deflection is usually measured by the motion of a beam of light which shines on the mirror attached to the suspension wire. The incident light on the mirror gets deflected by an angle 2ϕ and this deflection is usually observed on a scale placed at a distance of 1 meter from the galvanometer. In the case of the pivoted coil, this deflection is directly measured by the motion of the pointer which moves through an angle ϕ.

When the coil comes to rest under the influence of two opposing torques, the deflecting torque caused by the interaction of the magnetic field and the current and the restoring torque caused by the elasticity of the suspension wire, we see that

$$NIabB = \kappa\phi \tag{6.61}$$

where κ is the restoring torque per unit angle of the wire or the spring. Thus, we see that by measuring ϕ and knowing all the other parameters, one can measure the current passing through the galvanometer coil.

6.4.3 Ballistic Galvanometer

A galvanometer, as we saw above is an instrument which is frequently used to measure or detect currents. Since current is simply the rate of flow of charge, it seems natural that one should be able to use the same instrument, with suitable modifications for measurement of charge. Recall that

$$Q = \int I dt$$

and so, if we can make an integrator with a moving coil galvanometer described in the previous subsection, we should be able to measure the charge. Such a device is called a **Ballistic Galvanometer**. Basically, a ballistic galvanometer is really a modified moving coil galvanometer where the moving part (in this case the coil) has a large moment of inertia. This ensures that the time period for oscillation is large and measurements can be made easily. We now describe the working of a basic ballistic galvanometer.

A moving coil of N turns placed in a magnetic field and suspended by a suspension wire will experience a deflecting torque when a current I is passed through it given by Eq. (6.60) with $\theta = \pi/2$ as we saw above. Recall that the torque for rotational motion is like force for linear motion. Hence, we have

$$\text{angular momentum} = \int_0^T \tau dt = \int_0^T NIABdt = NABQ \tag{6.62}$$

But the angular momentum for the coil is also given by $I_m\omega$ where I_m is the moment of inertia and ω is the frequency of oscillation. Thus, we have

$$NBAQ = I_m\omega \tag{6.63}$$

and the kinetic energy of the coil is $\frac{1}{2}I_m\omega^2$. If κ is the restoring couple per unit deflection, then the work done in twisting the suspension through an angle ϕ is given by

$$\int_0^\phi \kappa\phi d\phi = \frac{1}{2}\kappa\phi^2 \tag{6.64}$$

This work done should be equal to the kinetic energy of the coil

$$\frac{1}{2}I_m\omega^2 = \frac{1}{2}\kappa\phi^2$$

$$\omega^2 = \frac{\kappa}{I_m}\phi^2 \tag{6.65}$$

But from Eq. (6.63),

$$\omega^2 = \frac{B^2 N^2 A^2 Q^2}{I_m^2}$$

and so, we have

$$\frac{B^2 N^2 A^2 Q^2}{I_m^2} = \frac{\kappa}{I_m}\phi^2$$

$$Q^2 = \frac{\kappa^2}{B^2 N^2 A^2}\frac{I_m}{\kappa}\phi^2 \tag{6.66}$$

The time period of oscillation T is given by

$$T = 2\pi\sqrt{\frac{I_m}{\kappa}}$$

and therefore,

$$\boxed{Q = \frac{\kappa T}{2\pi BNA}\phi} \tag{6.67}$$

Thus, we have a relationship between the charge passing through the galvanometer and the deflection angle ϕ in terms of the galvanometer parameters.

The theory presented above is obviously for an idealised galvanometer—one in which the coil executes undamped simple harmonic motion. In reality, there is always some damping present in the system. These are primarily due to air damping and a form of damping called electromagnetic damping which we shall study about when we study electromagnetic induction. Whatever the causes of the damping are, the net result is of course that the deflection with each successive oscillation becomes less and less. If $\phi_1, \phi_2, \cdots$ are the successive values of the deflection, we have

$$\frac{\phi_1}{\phi_2} = \frac{\phi_2}{\phi_3} = \cdots = d$$

where d is called the decrement. It is frequently convenient to use the logarithm of the decrement and define it as a quantity called logarithmic decrement λ

$$\lambda = \ln d$$

Thus, we have

$$\frac{\phi_1}{\phi_2} = \frac{\phi_2}{\phi_3} = \cdots = e^{\lambda}$$

For a whole vibration or oscillation (recall that ϕ_1 is the deflection on one side, followed by ϕ_2 on the other side or half an oscillation later) the decrement is thus,

$$\frac{\phi_1}{\phi_3} = e^{2\lambda}$$

or for a quarter of the oscillation it is

$$e^{\frac{\lambda}{2}}$$

Thus, if the observed deflection is ϕ_1, the actual deflection in the absence of damping would be ϕ and the two are related by

$$\phi = \phi_1 e^{\frac{\lambda}{2}}$$

Thus, if we were to take into account damping, the Eq. (6.67) needs to be replaced by

$$\boxed{Q = \frac{\kappa T}{2\pi BNA}\phi_1 e^{\frac{\lambda}{2}} = \frac{\kappa T}{2\pi BNA}\phi_1\left(1+\frac{\lambda}{2}\right)} \tag{6.68}$$

where we have used the fact that for a relatively undamped galvanometer, λ will be small and hence

$$e^{\frac{\lambda}{2}} \approx \left(1+\frac{\lambda}{2}\right)$$

If a ballistic galvanometer is used to compare charges, then of course one does not need to determine the constants in Eq. (6.68) since they will cancel out. However, if one needs to measure charge, then one would have to calibrate the galvanometer first with some known quantity of charge passing through it. This is usually done by means of a known capacitor which is charged to a fixed, known voltage and then allowed to discharge through the galvanometer. The deflection and the known charge is then used to find the constants and the galvanometer can then be used to measure charges.

6.4.4 Force on a Current Loop in a Non-uniform Magnetic Field

As we have seen, a current loop in a uniform field experiences no net force. Instead, there is a torque which it experiences which depends on a quantity called the magnetic dipole moment. In this way, it is like an electric dipole in a uniform electric field. Indeed, as we have seen, the current loop can be thought of as a magnetic dipole. The situation is of course, different when we place the loop in a non-uniform field. Just like an electric dipole experiences a net force in a non-uniform field, we expect that a net force would act on the dipole (current loop) in such a situation. To see this, let us consider an infinitesimal, rectangular loop carrying a current I in a non-uniform magnetic field.

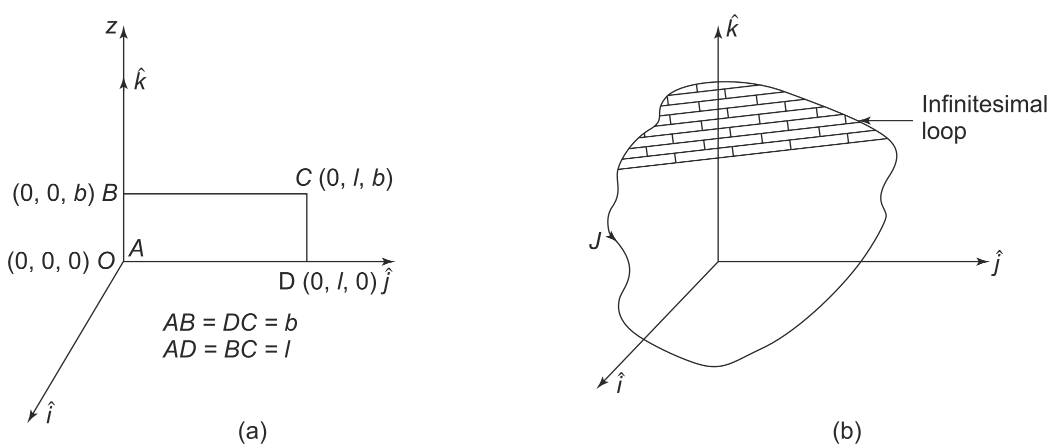


Fig. 6.17 *(a) An infinitesimal current loop in a non-uniform field, (b) A finite loop, which can be thought of as being made of many infinitesimal loops*

Let us take the loop $ABCD$ to be in the $y-z$ plane as shown in Fig. 6.17(a). Now the forces on the four sides of the rectangle can be easily evaluated using Eq. (6.1). The side DC is along the z-direction and thus $\vec{dl} = \hat{k}dz$. Therefore,

$$\vec{F}_{DC} = \int_{DC} I\vec{dl} \times \vec{B} = I\int_0^b dz\hat{k} \times \vec{B}(0,l,z) \tag{6.69}$$

where $\vec{B}(0,l,z)$ is the magnetic field along DC.

Similarly, the force along AB is given by

$$\vec{F}_{AB} = \int_{DC} I\vec{dl} \times \vec{B} = -I\int_0^b dz\hat{k} \times \vec{B}(0,0,z) \tag{6.70}$$

where the $-$ sign is due to the fact that the current is along $-\hat{k}$ direction. Adding the two, we get

$$\vec{F}_{DC} + \vec{F}_{AB} = I\int_0^b dz\hat{k} \times \left[\vec{B}(0,l,z) - \vec{B}(0,0,z)\right] \tag{6.71}$$

Now since we are taking an infinitesimal rectangle, both l, b are infinitesimal. Hence, we can write

$$\vec{B}(0,l,z) \approx \vec{B}(0,0,z) + \left(\frac{\partial \vec{B}}{\partial y}\right)_{(0,0,z)} l \tag{6.72}$$

and hence,

$$\vec{F}_{DC} + \vec{F}_{AB} = Il\int_0^b dz\hat{k} \times \frac{\partial \vec{B}}{\partial y} \approx Ilb\hat{k} \times \left(\frac{\partial \vec{B}}{\partial y}\right)_{(0,0,0)} \tag{6.73}$$

where we have written

$$\left(\frac{\partial \vec{B}}{\partial y}\right)_{(0,0,z)} \approx \left(\frac{\partial \vec{B}}{\partial y}\right)_{(0,0,0)}$$

and neglected the higher order infinitesimals. Similarly, we can compute the total force on the arms AD and BC.

$$\begin{aligned}\vec{F}_{AD} + \vec{F}_{BC} &= I\int_0^l \hat{j} \times \vec{B}(0,y,0) - I\int_0^l \hat{j} \times \vec{B}(0,y,b) \\ &\approx \ -Ilb\hat{j} \times \left(\frac{\partial \vec{B}}{\partial z}\right)_{(0,0,0)} \end{aligned}\tag{6.74}$$

We can express this in a slightly different and more compact way. Noting that the magnetic dipole moment of the current loop is

$$|\vec{\mu}| = Ilb$$

and

$$\vec{\mu} = \mu\hat{i}$$

we get

$$\vec{F} = \vec{F}_{AD} + \vec{F}_{BC} + \vec{F}_{DC} + \vec{F}_{AB}$$

$$\begin{aligned}\vec{F}_{AD} + \vec{F}_{BC} + \vec{F}_{DC} + \vec{F}_{AB} &= \mu\left[\frac{\partial}{\partial y}(-B_y\hat{i}, B_x\hat{j}, 0) - \frac{\partial}{\partial z}(-B_z\hat{i}, 0, -B_x\hat{k})\right] \\ &= \mu\left[-\hat{i}\left(\frac{\partial B_y}{\partial y} + \frac{\partial B_z}{\partial z}\right) + \hat{j}\frac{\partial B_x}{\partial y} + \hat{k}\frac{\partial B_x}{\partial z}\right] \\ &= \mu\left[\hat{i}\frac{\partial B_x}{\partial x} + \hat{j}\frac{\partial B_x}{\partial y} + \hat{k}\frac{\partial B_x}{\partial z}\right] \qquad (6.75)\end{aligned}$$

where we have used the fact that

$$\vec{\nabla}\cdot\vec{B} = \frac{\partial B_x}{\partial x} + \frac{\partial B_y}{\partial y} + \frac{\partial B_z}{\partial z} = 0$$

a fact which we shall prove later. Hence,

$$\frac{\partial B_y}{\partial y} = -\frac{\partial B_x}{\partial x} - \frac{\partial B_z}{\partial z}$$

For the loop that we have used, the magnetic dipole moment $\vec{\mu}$ has a direction perpendicular to the loop which is in the $y - z$ plane. Also, given the direction of the current, the direction of the magnetic dipole moment vector is along the $\hat{i}$ direction.

$$\vec{\mu} = \mu\hat{i}$$

Thus,

$$\mu B_x = \vec{\mu}\cdot\vec{B}$$

for our loop. Also

$$\vec{\nabla} = \hat{i}\frac{\partial}{\partial x} + \hat{j}\frac{\partial}{\partial y} + \hat{k}\frac{\partial}{\partial z}$$

Thus, we have

$$\mu\left[\hat{i}\frac{\partial B_x}{\partial x} + \hat{j}\frac{\partial B_x}{\partial y} + \hat{k}\frac{\partial B_x}{\partial z}\right] = \vec{\nabla}(\vec{\mu}\cdot\vec{B})$$

and Eq. (6.75) can be written as

$$\boxed{\vec{F} = \vec{\nabla}(\vec{\mu}\cdot\vec{B})} \qquad (6.76)$$

A magnetic dipole in a non-uniform field will experience a net force only when the component of $\vec{B}$ along the direction of the magnetic dipole moment is non-uniform.

For a current carrying loop of finite size, one obviously cannot use the approximations we have used to arrive at Eq. (6.76). However, we can subdivide the finite sized loop into an infinite, infinitesimal loops as shown in Fig. 6.17(b). The force on each of these infinitesimal loops is of course as given by Eq. (6.76). The total force on the finite loop is then obtained by integrating over all the infinitesimal loops. As an illustration, consider the loop as shown in Fig. 6.17. The finite loop is in the $y - z$ plane and for the infinitesimal loop in the Fig. 6.17(b), the magnetic moment is

$$\vec{\mu} = I dz dy\hat{i}$$

and hence the force on it is

$$d\vec{F} = I\vec{\nabla}(dzdyB_x(0, y, z))$$

and the total force on the finite loop is given by

$$\vec{F} = I\int dzdy\vec{\nabla}B_x(0, y, z)$$

where the integration is over the area of the finite loop.

EXAMPLE 6.4 $PQRS$ is a square loop carrying a current I in the sense shown in Fig. 6.18. It is suspended by wire from the top such that it can rotate around the wire as the axis. The moment of inertia of the loop around the axis is T and the magnitude of the magnetic moment of the loop is M. In the initial position of the loop as shown in the figure, the wire is unwound exerting no torque on the loop but does exert a torque on the loop equal to $\tau\theta$, where θ is the angle of rotation of the loop around the wire from that position. At time $t = 0$, the loop is at rest in the position shown. A uniform magnetic field B is switched on. Calculate

(a) the time period of the resulting oscillations and
(b) the position around which the oscillations take place for values of $(MB/T) \gg 1$ and $(MB/T) \ll 1$. You may make the small angle approximation, wherever valid.

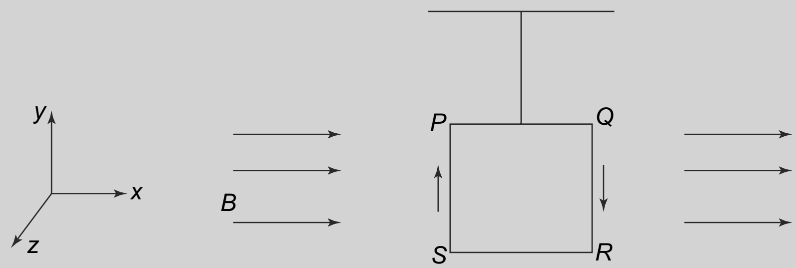


Fig. 6.18 *Example 6.4: Torque on a loop*

Solution

In Fig. 6.19, the angle between PQ and B is θ. M is the magnetic moment perpendicular to the loop. The angle between M and B is $\left(\frac{\pi}{2} - \theta\right)$. Initially, θ is zero.

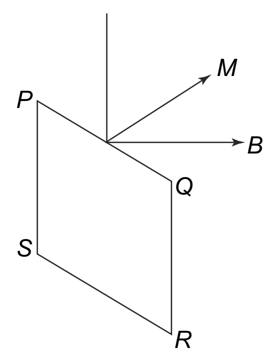


Fig. 6.19 *Example 6.4*

The torque exerted by the wire at angle θ is $\tau\theta$ which is trying to decrease the value of θ. The forces of the magnetic field on the sides PQ and RS are in the vertical y- direction and cancel out. The forces on the sides PS and QR result in a torque of magnitude $MB\sin\left(\frac{\pi}{2} - \theta\right)$ in the direction increasing the value of θ. Hence, the equation of motion of the loop for rotation around the axis is:

$$T\frac{d^2\theta}{dt^2} = -\tau\theta + MB\sin\left(\frac{\pi}{2} - \theta\right) = -\tau\theta + MB\cos\theta$$

The energy at an angle of rotation θ is

$$W = \frac{T}{2}\left(\frac{d\theta}{dt}\right)^2 + \frac{1}{2}\tau\theta^2 - MB\sin\theta$$

(a) At the initial position, $W = 0$ since both θ and $\frac{d\theta}{dt}$ are zero.

The loop will turn back again when $\frac{d\theta}{dt}$ is zero. This happens at $\theta = \theta_t$ where θ_t is given by

$$\frac{1}{2}\tau\theta_t^2 - MB\sin\theta_t = 0$$

Let us now consider the two cases. When $\frac{MB}{\tau} \ll 1$

$$\theta_t = \frac{2MB}{\tau}$$

since $\sin\theta_t \approx \theta_t$

When $\frac{MB}{\tau} \gg 1$, θ_t is almost π. Thus, writing

$$\theta_t = \pi - \beta, \qquad \beta \ll 1$$

we get

$$\frac{1}{2}\tau(\pi^2 - 2\pi\beta) - MB\beta = 0$$

and therefore,

$$\beta = \frac{\tau\pi^2}{2(\tau\pi - MB)}$$

The loop therefore, oscillates between $\theta = 0$ and $\theta = \theta_t$.

(b) For $\frac{MB}{\tau} \ll 1$, the values of θ are always small. Hence, we can replace $\cos\theta$ by 1 in the equation of motion and get

$$T\frac{d^2\theta}{dt^2} = -\tau\theta + MB$$

We introduce a new variable ϕ given by

$$\phi = \theta - \frac{MB}{\tau}$$

The equation of motion thus, becomes

$$\frac{d\phi^2}{dt^2} = -\frac{\tau}{T}\phi$$

This is the equation of a simple harmonic oscillator and hence, the time period is

$$t_{\text{osc}} = 2\pi\sqrt{\frac{T}{\tau}}$$

6.5 SOURCES AND PROPERTIES OF MAGNETIC FIELD

6.5.1 Sources of Magnetic Field

Historically, magnetic fields were always thought to be created by permanent magnets like lodestone. However, with the work of Oersted, who discovered that current carrying conductors experience a force in a magnetic field and Ampere, who discovered that current carrying conductors themselves produce magnetic fields, this view changed. We now understand that all magnetic fields are produced by currents.

Even in a permanent magnet, it is the currents at the atomic level which produce the magnetic fields. We shall study this phenomenon later. Thus, just as the laws for electric fields are expressed in terms of charges and charge densities (the sources for the electric field), the laws for magnetic fields are expressed in terms of currents and current densities.

In this chapter, we shall restrict ourselves to steady currents only. Steady currents of course cannot exist in isolation since the continuity equation is valid if there are no charges accumulating anywhere.

$$\vec{\nabla} \cdot \vec{j}(\vec{r}) = 0 \tag{6.77}$$

where $\vec{j}(\vec{r})$ is the the current density. This equation obviously implies that electric currents cannot abruptly end at any point. For a constant current I flowing along the x-axis in a conductor, we have

$$\vec{j}(\vec{r}) = I\delta(y)\delta(z)\hat{i} \tag{6.78}$$

and the current is thus

$$\vec{I} = \int \vec{j}(\vec{r})dxdydz$$

which, with the properties of Delta functions gives us

$$\vec{I} = I\hat{i}$$

As in the electric field case, where a form of Coulomb's Law gives us the electric field at any point produced by a charge, we need a law which will determine the magnetic field at any point $\vec{r}$ due to a current density $\vec{j}(\vec{r}')$ at the point $\vec{r}'$. **This law is called the Biot–Savart law.**

$$\boxed{d\vec{B}(\vec{r},\vec{r}') = \left(\frac{\mu_0}{4\pi}\right)\frac{\vec{j}(\vec{r}') \times (\vec{r}-\vec{r}')}{|\vec{r}-\vec{r}'|^3}} \tag{6.79}$$

The total magnetic field at the point $\vec{r}$ is obviously the integral of $d\vec{B}(\vec{r},\vec{r}')$ over $\vec{r}'$.

$$\boxed{\vec{B}(\vec{r}) = \left(\frac{\mu_0}{4\pi}\right)\iiint d^3r'\frac{\vec{j}(\vec{r}') \times (\vec{r}-\vec{r}')}{|\vec{r}-\vec{r}'|^3}} \tag{6.80}$$

For a wire carrying a current I along the x-axis, $\vec{j}(\vec{r})$ as given by Eq. (6.78), the last equation gives

$$\vec{B}(\vec{r}) = \int dx'd\vec{B}(\vec{r},x')$$
$$d\vec{B}(\vec{r}) = \frac{\mu_0}{4\pi}I\frac{\hat{i}\times\vec{r}}{|\vec{r}-x'\hat{i}|^3}$$

The appearance of the quantity $\left(\frac{\mu_0}{4\pi}\right)$ in these expressions needs to be explained. This is a constant, very much like $k = \frac{1}{4\pi\varepsilon_0}$ that we encountered in the discussing Coulomb's Law. In S.I. units, $\left(\frac{\mu_0}{4\pi}\right) = 10^{-7}\text{NA}^{-2}$. The value of $\left(\frac{\mu_0}{4\pi}\right)$, which is called the **permeability of free space** appears arbitrary at this stage. In this, it is like the value of the constant ε_0 whose value appears arbitrary. Actually, the values are chosen to make the laws compatible with our choice of units for charge and current. In addition, as we will see later in the book, the two constants, $\left(\frac{\mu_0}{4\pi}\right)$ and ε_0 are related to another fundamental constant of nature, c the velocity of light.

To calculate the magnetic field due to a charged particle of charge q moving with a velocity $\vec{v}$, we can use the same strategy as we used in arriving at Eq. (6.2). Taking the location of the particle at the origin, we replace $I\hat{i}dx'$ by $q\vec{v}$ in the expression for $d\vec{B}(\vec{r})$. The magnetic field due to the charged particle is thus,

$$B(\vec{r}) = \frac{\mu_0}{4\pi} q \frac{\vec{v} \times \vec{r}}{|\vec{r}|^3}$$

The expressions we have derived for the magnetic field due to currents and charges at any given point without any reference to time. Just like Coulomb's Law in electrostatics, this implies 'action at a distance'. As we shall later, this is in violation of the results of the Special Theory of Relativity. The correct results will be discussed later in this book where it will turn out that the results of this chapter are approximately valid when charges move with speeds much less than the speed of light.

We now illustrate the use of Biot-Savart law with some examples.

EXAMPLE 6.5 Find the magnetic field at the centre of a square of length L carrying a current I.

Solution

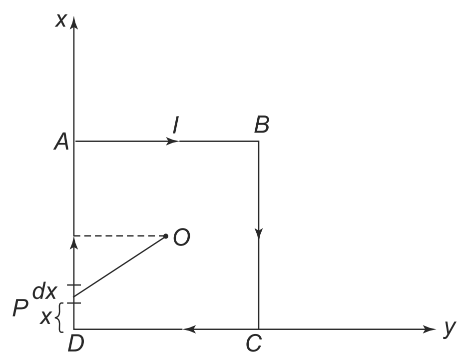


Fig. 6.20 *Example 6.5: Magnetic field at the center of a square of length L carrying a current I in the clockwise direction. The square is in the $z = 0$ plane*

From the Fig. 6.20

$$AB = AD = BC = CD = L$$
$$PD = x$$
$$(OP)^2 = \left(\frac{L}{2} - x\right)^2 + \left(\frac{L}{2}\right)^2$$

Take D to be the origin and square in the $x - y$ plane with $z = 0$. The z direction is into the plane. Consider first the side DA. An element of length dx located at P creates a magnetic field at O in the z direction given by

$$d\vec{B} = \left(\frac{\mu_0}{4\pi}\right) I \frac{dx}{OP^2}\hat{k}$$

and hence, the magnetic field due to the section DA is

$$\vec{B} = \int d\vec{B} = \left(\frac{\mu_0}{4\pi}\right) \int_0^L I \frac{dx}{\left(\frac{L}{2} - x\right)^2 + \left(\frac{L}{2}\right)^2} \hat{k} = \left(\frac{\mu_0}{4\pi}\right) I \frac{\pi}{L} \hat{k}$$

The field due to the other sections is identical to this and hence the total field at O due to the square is

$$\vec{B} = \mu_0 I \frac{1}{L} \hat{k}$$

EXAMPLE 6.6 Calculate the magnetic field due to a straight wire of length L carrying a current I at a point at a perpendicular distance d from the wire. This is shown in Fig. 6.21.

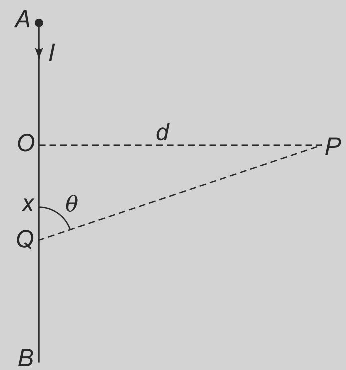


Fig. 6.21 *Example 6.6: Magnetic field due to a finite wire*

Solution

Consider a small section of the wire of length dx a distance x from O. The field due to this element at P is perpendicular to the plane of the paper and its magnitude is given by Biot-Savart law as

$$dB = \left(\frac{\mu_0}{4\pi}\right) I \frac{dx(QP)\sin(\pi - \theta)}{(QP)^3} = \left(\frac{\mu_0}{4\pi}\right) I \frac{dx}{(x^2 + d^2)} \frac{d}{(x^2 + d^2)^{1/2}}$$

Therefore, the total magnetic field at P due to the wire is simply the integral over the whole wire, or

$$B = \int dB = \left(\frac{\mu_0}{4\pi}\right) Id \int_{-l_1}^{l_2} \frac{dx}{(x^2 + d^2)^{3/2}} = \left(\frac{\mu_0}{4\pi}\right) I \frac{1}{d} \left[\frac{l_2}{(l_2^2 + d^2)^{1/2}} + \frac{l_1}{(l_1^2 + d^2)^{1/2}}\right]$$

Note that if $l_1, l_2 \to \infty$, i.e. the wire becomes infinite, the result for the magnetic field becomes

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

We will derive this result later using Ampere's Law.

In obtaining the result, we have assumed that the two ends of the wire A and B, are on either side of the point O. If both the ends were on the same side of O, the result would of course be different since the limits of the integral would go from l_1 to l_2. In that case, we would have

$$B = \left(\frac{\mu_0}{4\pi}\right) Id \int_{l_1}^{l_2} \frac{dx}{(x^2+d^2)^{3/2}} = \left(\frac{\mu_0}{4\pi}\right) I \frac{1}{d}\left[\frac{l_1}{(l_1^2+d^2)^{1/2}} - \frac{l_2}{(l_2^2+d^2)^{1/2}}\right]$$

EXAMPLE 6.7 Calculate the magnetic field of a circular loop of radius R carrying a current I at a point distance r from its centre.

Solution

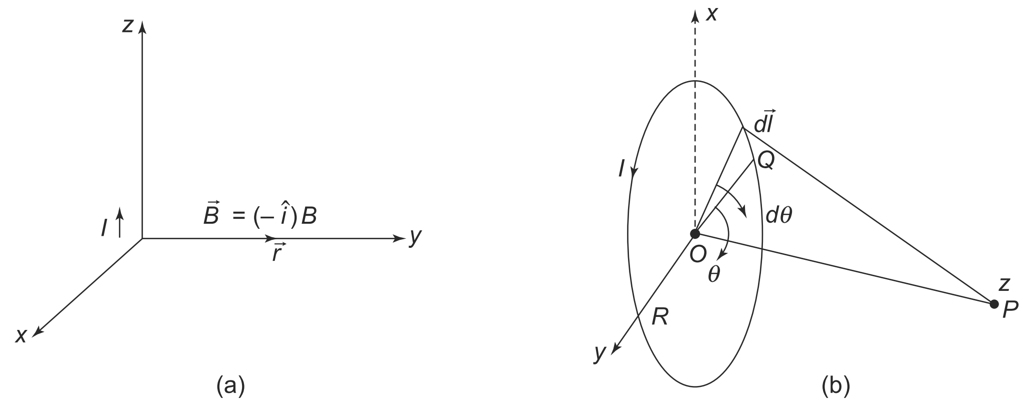


Fig. 6.22 *Example 6.7: (a) Direction of magnetic field at a point on y-axes due to a small thin conductor carrying a current in z direction. (b) Circular current carrying conductor of radius R taken to be in the $x - y$ plane with the point $z = 0$ as the centre. The point P is on the z-axis at a distance r from the centre*

In the figure,

$$\begin{aligned}
\overrightarrow{OP} &= (0, 0, r)\\
\overrightarrow{QP} &= (-R\sin\theta, -R\cos\theta, r)\\
\overrightarrow{dl} &= (\cos\theta, -\sin\theta, 0)Rd\theta\\
\overrightarrow{OQ} &= (R\sin\theta, R\cos\theta, 0)
\end{aligned}$$

Now consider a small element of the conductor subtending an angle $d\theta$ at the centre. Its length is $dl = Rd\theta$. The distance QP is

$$QP = (R^2 + r^2)^{1/2}$$

and hence the magnetic field due to the current carrying element dl at P is

$$\overrightarrow{dB} = \left(\frac{\mu_0}{4\pi}\right) I \frac{\overrightarrow{dl} \times \overrightarrow{QP}}{(r^2 + R^2)^{3/2}}$$

But

$$\vec{dl} \times \vec{QP} = (-r\sin\theta, -r\cos\theta, -R)Rd\theta$$

and hence,

$$\overrightarrow{dB} = \left(\frac{\mu_0}{4\pi}\right) I \frac{(-r\sin\theta, -r\cos\theta, -R)Rd\theta}{(r^2+R^2)^{3/2}}$$

The total magnetic field due to the loop will be obtained by integrating over the whole loop or for $\theta = 0$ to $\theta = 2\pi$. The x and y components will vanish by integrating and so we will get

$$\vec{B}(\vec{r}) = B\hat{k} = \left(\frac{\mu_0}{4\pi}\right) \int_0^{2\pi} \frac{-IR^2 d\theta}{(r^2+R^2)^{3/2}} \hat{k} = -\frac{\mu_0}{2} I \frac{R^2}{(R^2+r^2)^{3/2}} \hat{k} \tag{6.81}$$

where the $-$ve sign has come because we have taken the current in a particular direction.

EXAMPLE 6.8 Show that the magnetic force between two charged particles is much weaker than the electrostatic force between them.

Solution

Consider two charged particles of charges q_1 and q_2 moving with velocities $\vec{v}_1$ and $\vec{v}_2$ respectively. If $\vec{r}$ is the separation between the particles, the magnetic field created by q_1 at q_2 is

$$\vec{B} = \frac{\mu_0}{4\pi} q_1 \frac{\vec{v_1} \times \vec{r}}{|r|^3}$$

The force on q_2, from the Lorentz force expression is thus,

$$\vec{F}_{\text{mag}} = \frac{\mu_0}{4\pi} \frac{q_1 q_2}{r^2} (\vec{v}_1 \times (\vec{v}_2 \times \hat{r}))$$

The electrostatic force has a magnitude

$$F_{\text{elec}} = \frac{kq_1q_2}{r^2}$$

where

$$k = \frac{1}{4\pi\varepsilon_0}$$

Hence, the ratio of magnetic to electrostatic force is

$$\frac{F_{\text{mag}}}{F_{\text{elec}}} = \frac{\frac{\mu_0}{4\pi}\frac{q_1q_2}{r^2}}{\frac{q_1q_2}{4\pi\varepsilon_0 r^2}} |\vec{v}_1 \times (\vec{v}_2 \times \hat{r})|$$

The product $\mu_0\varepsilon_0$ has a value c^2. Hence the ratio has a maximum value $\left(\frac{v_1v_2}{c^2}\right)$ which is very small for ordinary speeds.

PROBLEM 6.4 A very thin disk made of a dielectric material has a radius R. It carries a uniform surface charge density σ. The plane of the disk is parallel to the $x-y$ plane and it is rotating around the z-axis passing through its centre at an angular speed of ω. What is the magnitude and direction of the magnetic field created at the centre of the disk?

PROBLEM 6.5 A square current loop $ABCD$ has sides of length L. The sides AB and AD are along the x- and y-axis and point A is at the origin. The loop carries a current I in the direction along $ABCD$. Evaluate the magnetic field vector at points $P(\frac{L}{4}, \frac{3L}{4})$, $Q\left(\frac{L}{2}, \frac{L}{2}\right)$ and $R\left(\frac{3L}{4}, \frac{L}{4}\right)$.

PROBLEM 6.6 A current I flows through a square loop of side L. Find the magnetic field at a distance R from its centre along a line passing through the middle points of two parallel sides of the square.

PROBLEM 6.7 A loop in the form of a regular n-sided polygon lies inscribed inside a circle of radius R. A current I flows through the loop. Calculate the magnetic field at a distance D from the centre of the circle along a line perpendicular to the plane of the circle and passing through its centre. Show that in the limit $n \to \infty$, your result resembles the well known result.

PROBLEM 6.8 A square slab of sides length L and thickness t is made of a conducting material. The faces of the slab are parallel to the $x - y$ plane and a current I in the x-direction is flowing through the slab. Obtain the magnetic field at a point P a distance D above one of the faces on the line perpendicular to the faces of the slab passing through the centres of the squares.
[*Hint: take filaments of current symmetrically together*]

PROBLEM 6.9 A circular current carrying coil of radius R is in the $x - y$ plane with its centre at $z = 0$. The magnetic field at a point P on the z-axis at a distance z above the centre is well known. Using that expression and the fact that $\vec{\nabla} \cdot \vec{B} = 0$, obtain the value of B_x and B_y at a point Q away from the axis the same value of z as P but removed by small values of x and y. Neglect quadratic and higher powers of x and y.
[*Hint: Use symmetry of the problem around the z-axis.*]

6.5.2 Helmholtz Coil

If we plot the magnetic field due to a circular coil carrying a steady current at a point on the axis, we see that the magnetic field varies quite rapidly with distance (Fig. 6.23(a)).

Clearly, if we can set up two parallel, coaxial and identical coils, then the field in the region between the coils would be relatively flat with distance. This is precisely the arrangement in a device called Helmholtz coil which consists of two identical coils, of radius R, separated by a distance b. As we now show, if $b = R$, then the combined field for the two coils is the flattest around the midpoint of the coils.

The field at a distance x from the midpoint of the two coils is given by the superposition of the fields due to the two coils. From Fig. 6.23(b), it is clear that

$$B(x) = B_1 + B_2 = \frac{\mu_0 R^2 I}{2}\left[\frac{1}{[(x-b)^2 + R^2]^{3/2}} + \frac{1}{[(x+b)^2 + R^2]^{3/2}}\right] \tag{6.82}$$

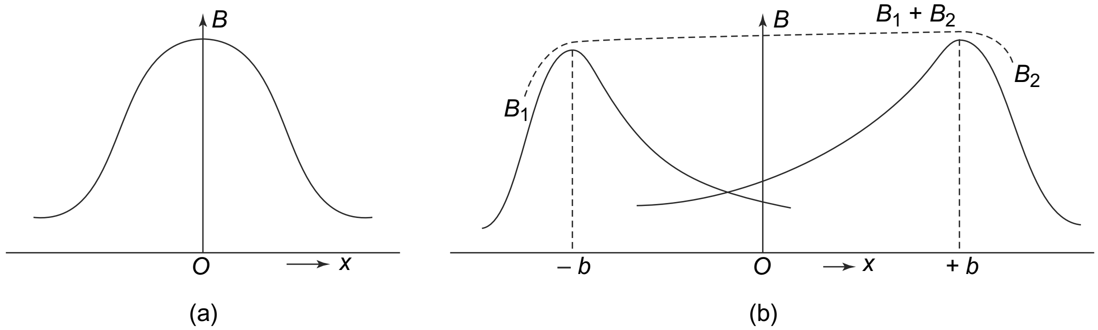


Fig. 6.23 *Helmholtz coil: (a) Magnitude of the magnetic field at a point distance d from the centre of a current carrying coil plotted against d, (b) Magnetic field B at the centre of two identical coils separated by a distance b. O is the midpoint of the coils and B_1 and B_2 are the fields due to the two coils*

From this expression, we see that $B(x)$ is an extremum at $x = 0$. We can compute the derivative and confirm this

$$\frac{dB(x)}{dx} = \left(-\frac{3}{2}\right)\frac{\mu_0 R^2 I}{2}\left[\frac{2(x-b)}{[(x-b)^2+R^2]^{5/2}} + \frac{2(x+b)}{[(x+b)^2+R^2]^{5/2}}\right] \tag{6.83}$$

Clearly

$$\left(\frac{dB(x)}{dx}\right)_{x=0} = 0$$

This is not surprising. We expected the field to be an extremum at the midpoint. What is interesting about the arrangement is that for certain distances between the coils, the second derivative also vanishes, as we now show.

$$\frac{d^2B(x)}{dx^2} = 2\left(-\frac{3}{2}\right)\frac{\mu_0 R^2 I}{2}\left[\frac{2(x-b)^2(-\frac{5}{2})}{[(x-b)^2+R^2]^{7/2}} + \frac{2(x+b)^2(-\frac{5}{2})}{[(x+b)^2+R^2]^{7/2}}\right.$$
$$\left. + \frac{1}{[(x-b)^2+R^2]^{5/2}} + \frac{1}{[(x+b)^2+R^2]^{5/2}}\right] \tag{6.84}$$

For the second derivative to vanish, i.e., for the curve to be flat at $x = 0$, we must have

$$\left(\frac{d^2B(x)}{dx^2}\right)_{x=0} = -3\frac{\mu_0 R^2 I}{2}\left[\frac{-8b^2+2R^2}{(b^2+R^2)^{7/2}}\right] \tag{6.85}$$

Thus, for $b = \frac{R}{2}$ will give us the first and the second derivatives to vanish at $x = 0$. This implies that if the coils are separated by a distance equal to their common radius, the field at the midpoint of the two coils will be very flat and not vary much with distance. This arrangement is used to produce a constant magnetic field over a region in many experiments.

6.5.3 Properties of $\vec{B}$

In earlier Chapters we have studied the properties of electric fields. These relate to the relationship that the divergence and curl of the electric field with charge densities. For instance, the divergence of the electric field is given by

$$\vec{\nabla} \cdot \vec{E} = k\rho(\vec{r})$$

Thus, for a point charge, the electric field was divergenceless everywhere, except at the location of the charge itself, where the divergence was non-zero. The electric field lines, we have seen, are continuous and converge or diverge from point charges.

We need to do the same exercise for the magnetic field and see how the divergence and curl of the magnetic field relates to the current densities which, as we have seen, are the sources of the magnetic field.

Let us first find out the divergence of the magnetic field. Using (Eq. (6.80)), we find that

$$\begin{aligned} \vec{\nabla} \cdot \vec{B}(\vec{r}) &= \vec{\nabla}_r \cdot \vec{B}(\vec{r}) \\ &= \left(\frac{\mu_0}{4\pi}\right) \iiint d^3r' \vec{\nabla}_r \cdot \frac{\vec{j}(\vec{r}') \times (\vec{r} - \vec{r}')}{|\vec{r} - \vec{r}'|^3} \end{aligned} \tag{6.86}$$

The operator $\vec{\nabla}_r$ acts only on functions of $\vec{r}$ and not on those of $\vec{r}'$. We use the identity

$$\vec{A} \cdot (\vec{B} \times \vec{C}) = \vec{B} \cdot (\vec{C} \times \vec{A})$$

on Eq. (6.86) to get

$$\vec{\nabla}_r \cdot \vec{B}(\vec{r}) = \left(\frac{\mu_0}{4\pi}\right) \iiint d^3r' \vec{j}(\vec{r}') \cdot \left[\vec{\nabla}_r \times \frac{(\vec{r} - \vec{r}')}{|\vec{r} - \vec{r}'|^3}\right] \tag{6.87}$$

Let us look at the the quantity in the square brackets and try and evaluate it. To do this, we shall use the index notation as explained in the Chapter on Mathematical Preliminaries. Recall that, in the index notation, we can write the cross-product of two vectors as

$$(\vec{A} \times \vec{B})_i = \varepsilon_{ijk} A_j B_k$$

where ε_{ijk} is the totally antisymmetric three index object and A_j and B_k are the j and k components of $\vec{A}$ and $\vec{B}$. i, j, k take values 1, 2, 3. Using this, we can rewrite

$$\begin{aligned} \left[\vec{\nabla}_r \times \frac{(\vec{r} - \vec{r}')}{|\vec{r} - \vec{r}'|^3}\right]_i &= \varepsilon_{ijk} \frac{\partial}{\partial x^j} \left(\frac{x_k - x'_k}{|\vec{r} - \vec{r}'|^3}\right) \\ &= \varepsilon_{ijk} \left[\frac{\delta_{jk}}{|\vec{r} - \vec{r}'|^3} - \frac{2(x_k - x'_k)(x_j - x'_j)}{|\vec{r} - \vec{r}'|^5}\right] \\ &= 0 \end{aligned} \tag{6.88}$$

where the last result follows from the fact that the quantity ε_{ijk} is completely antisymmetric under an exchange of the indices i, j, k while the quantity in the square brackets is symmetric (δ_{jk} and $(x_k - x'_k)(x_j - x'_j)$ are obviously symmetric under interchange of k and j). Since the product of an antisymmetric and symmetric object vanishes, this is zero. Thus, we have

$$\boxed{\vec{\nabla} \cdot \vec{B}(\vec{r}) = 0} \tag{6.89}$$

The divergence of the magnetic field is zero.

This is a very significant result. For the electric field, we know that the divergence is given by Gauss's Law and in the presence of charges it is non-zero. In terms of lines of force, for the electric case, we saw that lines of force begin and end on charges. In the magnetic case, the divergence is zero and hence, the magnetic lines of force cannot end or emanate from any point.

What about the curl of the magnetic field? Recall that in the case of the electric field, the static electric field can be written in terms of a gradient of a scalar potential, ϕ and since the curl of a gradient vanishes, the curl of the electric field is zero. This, as we saw, implies that the electric field is conservative, i.e., the work done in moving from one point to another is independent of the path taken but only depends on the end points. For the magnetic field, the situation is very different.

We know that

$$\vec{B}(\vec{r},\vec{r}') = \left(\frac{\mu_0}{4\pi}\right)\iiint d^3r'\,\frac{\vec{j}(\vec{r}')\times(\vec{r}-\vec{r}')}{|\vec{r}-\vec{r}'|^3}$$

and hence,

$$\vec{\nabla}_r\times\vec{B}(\vec{r}) = \left(\frac{\mu_0}{4\pi}\right)\iiint d^3r'\vec{\nabla}_r\times\frac{\vec{j}(\vec{r}')\times(\vec{r}-\vec{r}')}{|\vec{r}-\vec{r}'|^3} \tag{6.90}$$

Here $\vec{\nabla}_r$ is the operator, which only operates on the coordinates r and not r'. Using the identity

$$\vec{A}\times(\vec{B}\times\vec{C}) = \vec{B}(\vec{A}\cdot\vec{C}) - (\vec{A}\cdot\vec{B})\vec{C}$$

we get

$$\vec{\nabla}_r\times\vec{B}(\vec{r}) = \left(\frac{\mu_0}{4\pi}\right)\iiint d^3r'\left[\vec{j}(\vec{r}')\left[\vec{\nabla}_r\cdot\frac{\vec{r}-\vec{r}'}{|\vec{r}-\vec{r}'|^3}\right] - (\vec{j}(\vec{r}')\cdot\vec{\nabla}_r)\left[\frac{\vec{r}-\vec{r}'}{|\vec{r}-\vec{r}'|^3}\right]\right] \tag{6.91}$$

where we have used the fact that $\vec{\nabla}_r$ doesn't act on $\vec{j}(\vec{r}')$.

Let us look at the second term of above expression. We know that

$$\frac{\partial f(\vec{r}-\vec{r}')}{\partial x_i} = -\frac{\partial f(\vec{r}-\vec{r}')}{\partial x_i'}$$

where x_i, $i = 1, 2, 3$ could be any of the coordinates x, y or z. So the i^{th} component of the second term is

$$\begin{aligned}
-\left[\left(\frac{\mu_0}{4\pi}\right)\iiint d^3r'(\vec{j}(\vec{r}')\cdot\vec{\nabla}_r)\left(\frac{\vec{r}-\vec{r}'}{|\vec{r}-\vec{r}'|^3}\right)\right]_i &= -\left(\frac{\mu_0}{4\pi}\right)\iiint d^3r'\,j_k(\vec{r}')\frac{\partial}{\partial x_k}\left(\frac{x_i-x_i'}{|\vec{r}-\vec{r}'|^3}\right)\\
&= \left(\frac{\mu_0}{4\pi}\right)\iiint d^3r'\,j_k(\vec{r}')\frac{\partial}{\partial x_k'}\left(\frac{x_i-x_i'}{|\vec{r}-\vec{r}'|^3}\right)\\
&= \left(\frac{\mu_0}{4\pi}\right)\iiint d^3r'\left[\frac{\partial}{\partial x_k'}\left(j_k(\vec{r}')\frac{x_i-x_i'}{|\vec{r}-\vec{r}'|^3}\right)\right.\\
&\quad\left.-\left(\frac{\partial}{\partial x_k'}j_k(\vec{r}')\right)\frac{x_i-x_i'}{|\vec{r}-\vec{r}'|^3}\right]
\end{aligned} \tag{6.92}$$

The second term in the square brackets vanishes because

$$\frac{\partial}{\partial x_k'} j_k(\vec{r}') = \vec{\nabla}_r' \cdot \vec{j}(\vec{r}') = 0$$

by the equation of continuity. The first term is a volume integral of a divergence, which we can convert to a surface integral over a surface encompassing the volume by using the divergence theorem. Since the integral over the volume is over all space, the surface over which the surface integral is taken lies at infinity.

$$\left(\frac{\mu_0}{4\pi}\right) \iiint d^3r' \left[\frac{\partial}{\partial x_k'}\left(j_k(\vec{r}')\frac{x_i - x_i'}{|\vec{r}-\vec{r}'|^3}\right)\right] = \left(\frac{\mu_0}{4\pi}\right) \iint dS \left(\hat{n}_k \cdot j_k(\vec{r}')\frac{x_i - x_i'}{|\vec{r}-\vec{r}'|^3}\right) \tag{6.93}$$

Here $\hat{n}_k$ is a unit normal in the outward direction on the surface at infinity. The integrand in general will vanish at $r' = \infty$ and hence the first term is also zero. We therefore see that the second term of Eq. (6.91) is zero. From Eq. (6.91), we get

$$\vec{\nabla} \times \vec{B}(\vec{r}) = \left(\frac{\mu_0}{4\pi}\right) \iiint d^3r' \vec{j}(\vec{r}') \left[\vec{\nabla}_r \cdot \frac{x_i - x_i'}{|\vec{r}-\vec{r}'|^3}\right] \tag{6.94}$$

But

$$\frac{x_i - x_i'}{|\vec{r}-\vec{r}'|^3} = -\left(\vec{\nabla}_r\right)_i \left(\frac{1}{|\vec{r}-\vec{r}'|}\right)$$

as can be easily verified by direct differentiation. Hence

$$\begin{aligned}\left(\vec{\nabla}_r\right)_i \cdot \frac{x_i - x_i'}{|\vec{r}-\vec{r}'|^3} &= -\vec{\nabla}_r \cdot \left(\vec{\nabla}_r \left(\frac{1}{|\vec{r}-\vec{r}'|}\right)\right) \\ &= -\nabla_r^2 \left(\frac{1}{|\vec{r}-\vec{r}'|}\right) \\ &= 4\pi \delta^3(\vec{r}-\vec{r}')\end{aligned} \tag{6.95}$$

as we have seen in the chapter on electrostatics. Therefore, using Eqs. (6.94) and (6.95), we get

$$\begin{aligned}\vec{\nabla} \times \vec{B}(\vec{r}) &= \mu_0 \iiint d^3r' \vec{j}(\vec{r}')\delta^3(\vec{r}-\vec{r}') \\ &= \mu_0 \vec{j}(\vec{r})\end{aligned} \tag{6.96}$$

The curl of the magnetic field is given by

$$\boxed{\vec{\nabla} \times \vec{B}(\vec{r}) = \mu_0 \vec{j}(\vec{r})}$$

The divergence and curl of the magnetic field have properties which are opposite to those of the electric field.

6.5.4 Current Loop as a Magnetic Dipole

In our discussion above, we have already identified a quantity $\vec{\mu}$ as the magnetic dipole moment of a current carrying loop. What about the field due to this current carrying loop or magnetic dipole? To see this, consider Eq. (6.81) for the magnetic field at the axis of a current carrying circular loop. In the limit

$r \gg R$, i.e., at points far away from the loop (in comparison to its radius), the magnetic field becomes

$$\vec{B}(\vec{r}) \approx -\left(\frac{\mu_0}{4\pi}\right)\frac{2I\pi R^2}{r^3}\hat{k} \tag{6.97}$$

We have already identified the quantity $IA = I\pi R^2 = \mu$ as the magnetic dipole moment of the current carrying loop. In terms of μ, the magnetic field in Eq. (6.97) becomes

$$\vec{B}(\vec{r}) = -\left(\frac{\mu_0}{4\pi}\right)\frac{2\mu}{r^3}\hat{k} \tag{6.98}$$

But this result for the magnetic field is exactly parallel to the result we obtained for the electric field of an electric dipole at far-off distances from an electric dipole. This is shown in Fig. 6.24.

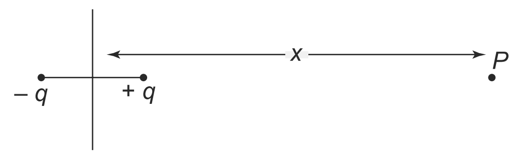


Fig. 6.24 *Field due to an electric dipole of dipole moment $p = qa$ at a distance r*

The electric field due to the dipole, at a point P a distance $r \gg a$ from the dipole is given by

$$\begin{aligned}\vec{E}(\vec{r}) &= -kq\left[\frac{1}{(r-\frac{a}{2})^2} - \frac{1}{(r+\frac{a}{2})^2}\right]\hat{k} \\ &\approx \ -k\frac{2p}{r^3}\hat{k} \end{aligned} \tag{6.99}$$

Equation (6.99) is exactly like Eq. (6.98) if we identify $\mu \to p$ and $\left(\frac{\mu_0}{4\pi}\right) \to k$.

We thus, have the result that a small current carrying loop behaves like a magnetic dipole with the dipole moment in a direction normal to the plane of the loop and of magnitude $\mu = IA$.

6.6 AMPERE'S LAW

Equation (6.96) which gives us the curl of a magnetic field was derived from the Biot-Savart law. Written in this form, this is called the **Ampere's Law.**

$$\boxed{\vec{\nabla} \times \vec{B}(\vec{r}) = \mu_0 \vec{j}(\vec{r})} \tag{6.100}$$

Frequently, it is used in its integral form which is easy to obtain from the differential form Eq. (6.100). Consider a closed loop L. Then, by Stokes' Theorem

$$\begin{aligned}\oint_L \vec{B}(\vec{r}) \cdot d\vec{l} &= \iint_S d\vec{S} \cdot \left[\vec{\nabla} \times \vec{B}(\vec{r})\right] \\ &= \mu_0 \iint_S \vec{j}(\vec{r}) \cdot d\vec{S} \end{aligned} \tag{6.101}$$

where the surface S is any surface whose boundary is the closed loop L, as shown in Fig. 6.25(a).

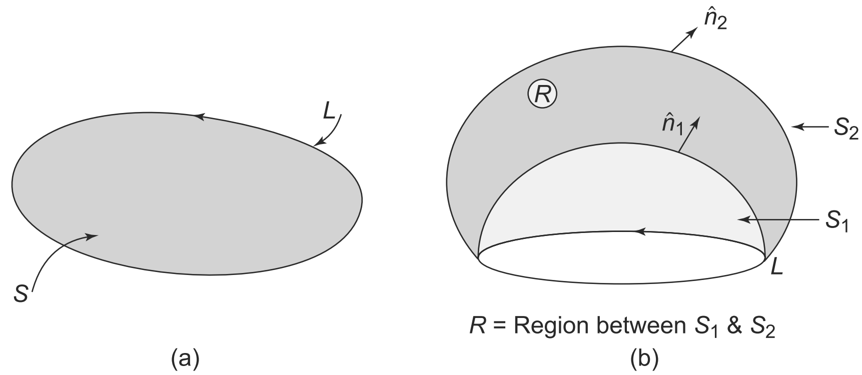


Fig. 6.25 *Currents in Ampere's law: (a) An arbitrary loop L which forms the boundary of a surface S, (b) Two surfaces S_1 and S_2 both of which have the loop L as their boundaries*

By the definition of the current I

$$\iint_S \vec{j}(\vec{r}) \cdot d\vec{S} = I$$

where I is the total current flowing across the surface S. Thus, we have

$$\boxed{\oint_L \vec{B}(\vec{r}) \cdot d\vec{l} = \mu_0 I} \tag{6.102}$$

Equation (6.102) is the integral form of Ampere's Law, which is also called the circuital law.

It might be argued that the surface S that we chose, is **any** surface having L as its boundary. So how can we be sure that the current through all such surfaces would be the same? It is indeed so for steady currents, i.e., those currents which satisfy

$$\vec{\nabla} \cdot \vec{j} = 0$$

To see that this is indeed the case, consider two surface S_1 and S_2 both of whom have the loop L as the boundary (Fig. 6.25(b)). The outward normals to these surfaces are $\hat{n}_1$ and $\hat{n}_2$ respectively. Now let us consider the region R between S_1 and S_2. Then for steady currents,

$$\begin{aligned} \int_R \vec{\nabla} \cdot \vec{j}(\vec{r}) &= \iint_{S_2} \vec{j}(\vec{r}) \cdot dS\hat{n}_2 - \iint_{S_1} \vec{j}(\vec{r}) \cdot dS\hat{n}_1 \\ &= 0 \end{aligned} \tag{6.103}$$

The second integral has a negative sign since for the region R between S_1 and S_2, the outward normal on S_1 is opposite to $\hat{n}_1$. Thus we see that

$$\iint_{S_2} \vec{j}(\vec{r}) \cdot dS\hat{n}_2 = \iint_{S_1} \vec{j}(\vec{r}) \cdot dS\hat{n}_1$$

and hence the current through any surface whose boundary is L is the same for steady currents.

Another ambiguity in Eq. (6.102) concerns the direction of the current flow I. The rule is as follows:

The current I is in the direction of the movement of a right hand screw when rotated along the direction of the line integral along L. This follows from Stokes' Theorem and our convention for the direction of the vector product in general and the vector product in Biot-Savart law in particular.

Ampere's Law doesn't obviously carry any more information than the Biot-Savart law. But just like Gauss's Law, which follows from Coulomb's Law and the nature of the electric field makes the calculation of the electric field very simple in certain situations of symmetry, Ampere's Law is also very useful in determining the magnetic field in certain situations with symmetry. We illustrate this with an example.

EXAMPLE 6.9 Calculate the magnetic field at any point due to an infinitely long wire carrying a current I. This is depicted in Fig. 6.26.

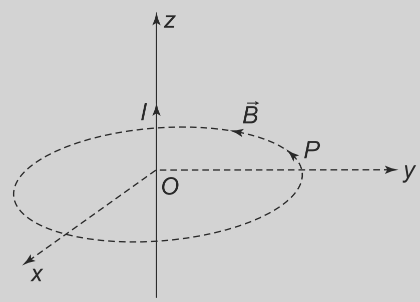


Fig. 6.26 *Example 6.9*

Solution

We take the wire to be along the z axis and passing through the origin. The point P is located on the y axis at a distance r from the origin. Consider a circular loop in the $x-y$ plane passing through P. By symmetry, the magnetic field $\vec{B}$ at all points on the loop is the same. $\vec{B}(\vec{r})$ is directed along $(\vec{j} \times \vec{r})$ which in this case is along the $(-x)$ direction at the point P. At P and everywhere on the loop $\vec{B}$ is along the loop. Hence, by Ampere's Circuital Law

$$\oint \vec{B}(\vec{r}) \cdot d\vec{l} = 2\pi r B(r) = \mu_0 I$$

and thus,

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

PROBLEM 6.10 A very long cylinder of radius a has a volume charge of density ρ. Find the expression for the magnetic flux density vector inside as well as outside of the cylinder if the cylinder is rotating around its axis with an angular velocity ω.

6.6.1 Force between Two Current Carrying Wires

The last result for the magnetic field due to an infinite current carrying wire has an immediate application for calculation of force between two current carrying wires. The Biot-Savart law tells us that a current carrying coil creates a magnetic field in its neighbourhood. We have earlier discussed the force that a current carrying coil experiences due to the presence of the magnetic field. Thus, there will be a force between two current carrying coils. The calculation of this force would in general involve the calculation of the field due to either coils and then use Eq. (6.1) to calculate the force on each other. The result obviously depends on the relative positions and orientations of the two coils and cannot be expressed in a mathematically simple form. For the special case of two infinite parallel current carrying wires separated by a distance r we can calculate this force.

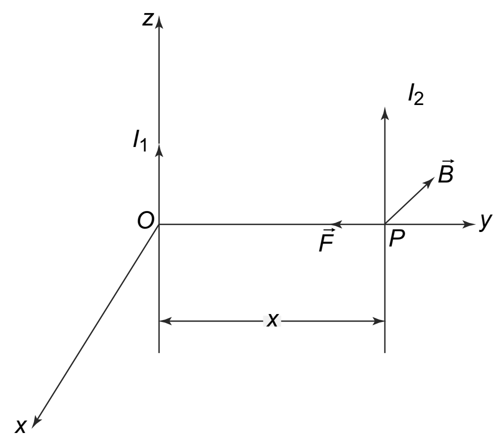


Fig. 6.27 *Force between two current carrying wires. The magnetic field due to I_1 at P is in the $-x$ direction. The force on the wire carrying I_2 is along $\vec{F}$. If I_1 and I_2 are in opposite directions, the direction of $\vec{F}$ is reversed*

With the geometry as shown in Fig. 6.27 the field due the current I_1 at the point P shown in the figure is in the negative x direction of magnitude

$$B = \frac{\mu_0}{2\pi r} I_1$$

The current I_2 is in the z-direction. By Eq. (6.1), the force on the wire carrying current I_2 is along $\hat{z} \times -\vec{x} = -\hat{y}$ direction. The force on I_2 is thus towards I_1. By Eq. (6.1), its magnitude for a length dl of the wire is proportional to dl. If F denotes the force per unit length, the force on the length dl is, by Eq. (6.1)

$$F dl = dl \frac{\mu_0}{2\pi r} I_1 I_2$$

It is easy to see similarly that the force on I_1 due to magnetic field created by I_2 is of the same magnitude and directed towards I_2. If the sense of the two currents are opposite, the direction of the force changes but the magnitude is the same as in the earlier case of currents in the same direction.

We thus, arrive at the result that two infinite parallel current carrying wires attract each other if the currents are in the same direction and repel if the directions are opposite.

The magnitude of the force per unit length in either case is

$$F = \frac{\mu_0}{2\pi r} I_1 I_2$$

This last equation is the one accepted SI system of units as defining an Ampere in terms of mechanical quantities with $\frac{\mu_0}{4\pi} = 10^{-7}$ NA^{-2}. One Ampere is that current flowing in each of two parallel infinite wires distance r apart, such that the force on a length r on each of the wires is 2×10^{-7} N.

6.6.2 Solenoid

We have seen how a circular coil carrying a current I produces a magnetic field. We now generalise this result to a solenoid. A solenoid is basically an arrangement of N coils, tightly wound on a hollow cylinder. In effect, this can be thought of as a large number of similar, circular, coaxial coils close to each other. We can use the result for the field due to a circular coil and generalise it to this case.

Recall that the field due to a circular coil of radius R, carrying a current I at a point on the axis, distance r from the centre is given by (Eq. (6.81)).

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

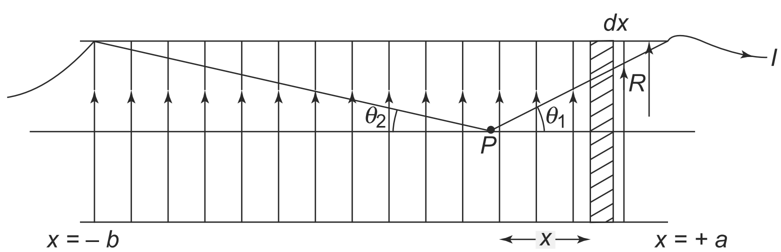


Fig. 6.28 *A solenoid with N turns per unit length of wire wound on a cylinder*

Consider a point P on the axis of a solenoid. We first find the field at P due to a section of the solenoid of length dx which has Ndx turns or coils. The field due to this section will be

$$dB = \frac{\mu_0 I (Ndx)}{2} \frac{R^2}{(R^2 + x^2)^{3/2}} \tag{6.104}$$

The field therefore due to the whole solenoid at the point P will simply be (the solenoid extends from $x = -b$ to $x = a$ as in Fig. 6.28)

$$B = \frac{\mu_0 I N R^2}{2} \int_{-b}^{a} \frac{dx}{(R^2 + x^2)^{3/2}} \tag{6.105}$$

Writing

$$\cot\theta = \frac{x}{R}$$

we get

$$B = \frac{\mu_0 I N}{2} \int_{\pi-\theta_2}^{\theta_1} (-\sin\theta) d\theta = \frac{\mu_0 I N}{2} (\cos\theta_1 + \cos\theta_2) \tag{6.106}$$

where θ_1 and θ_2 are indicated in Fig. 6.28. The field for a finite solenoid is shown in Fig. 6.29.

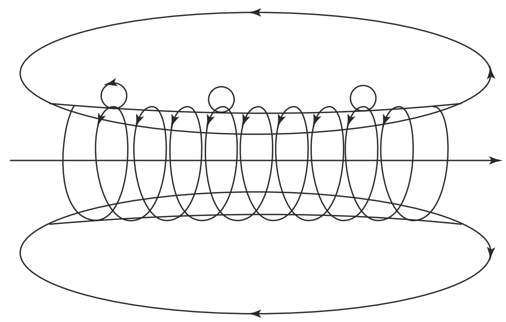


Fig. 6.29 *Magnetic field lines due to a finite solenoid. The small loops indicate the field lines due to individual current loops. The solid lines indicate the field due to the whole solenoid. The field at any point on the axis, inside and outside the solenoid is along the axis*

The above result for the $\vec{B}$ field on the axis of an finite solenoid, Eq. (6.106), can be immediately generalised to an infinite solenoid. In this case, both θ_1 and θ_2 are zero and hence we have the result

$$B = \mu_0 I N \tag{6.107}$$

This is the field on the axis. However, it turns out that for an infinite solenoid, this is the field at any point inside the solenoid as we now show. This can be most easily seen by using symmetry and Ampere's Law.

Let z be the direction of the axis of the infinite solenoid. Then the field at every point on the z-axis must be the same. Further, since there is nothing to distinguish one orientation of the solenoid from another (with respect to rotations about the z axis), the magnetic field must remain the same because of rotations. This immediately rules out any radial component of the field. To see this, consider a cylindrical surface coaxial with the solenoidal axis. If the field had any radial component, the flux across the cylindrical surface would be non-zero. But that implies that

$$\iint \vec{B} \cdot d\vec{S} \neq 0$$

But by Divergence Theorem, the surface integral of $\vec{B}$ over a surface is equal to the volume integral of the divergence of $\vec{B}$ over the volume enclosed by the surface.

$$\iint \vec{B} \cdot d\vec{S} = \iint \vec{\nabla} \cdot \vec{B} dV = 0$$

since $\vec{\nabla} \cdot \vec{B} = 0$. Thus, it is impossible for the field to have any radial component.

Next, consider circular loops inside and outside the solenoid. We want to evaluate the line integral of $\vec{B}$ along these closed paths. Notice that the tangential component of $\vec{B}$, B_ϕ must be the same at all points on the circle by symmetry, as shown in Fig. 6.30(b). Hence, the loop integral over a circle of radius r is simply

$$\oint \vec{B} \cdot d\vec{l} = B_\phi(r) 2\pi r \tag{6.108}$$

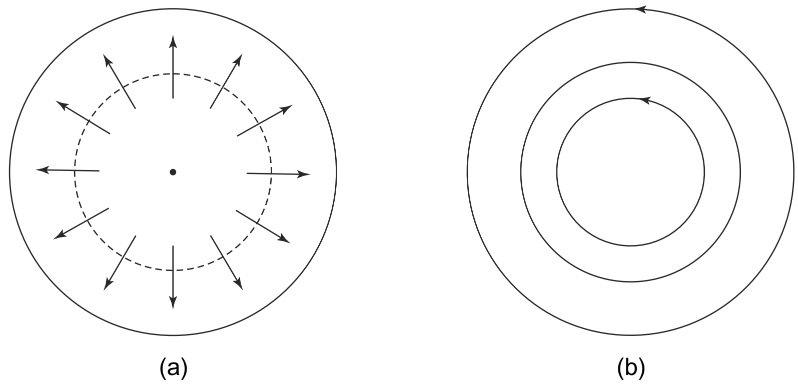


Fig. 6.30 *(a) Radial magnetic field is ruled out since $\vec{\nabla} \cdot \vec{B} = 0$, (b) Line integral of $\vec{B}$ along any closed circular loop, inside or outside, must vanish since there is no current flowing through it*

But we know that this line integral, by Ampere's law is related to the net current flowing inside the loop used to evaluate the integral. The net current flowing inside the loop is zero and hence, this must be zero. Thus,

$$B_\phi = 0 \quad \text{for all } r \tag{6.109}$$

We have thus shown that radial and azimuthal components of the magnetic field $\vec{B}$ are zero everywhere for an infinite solenoid. We now use Ampere's Law to determine the z component of the field inside and outside the infinite solenoid. Consider three closed paths as shown in Fig. 6.31.

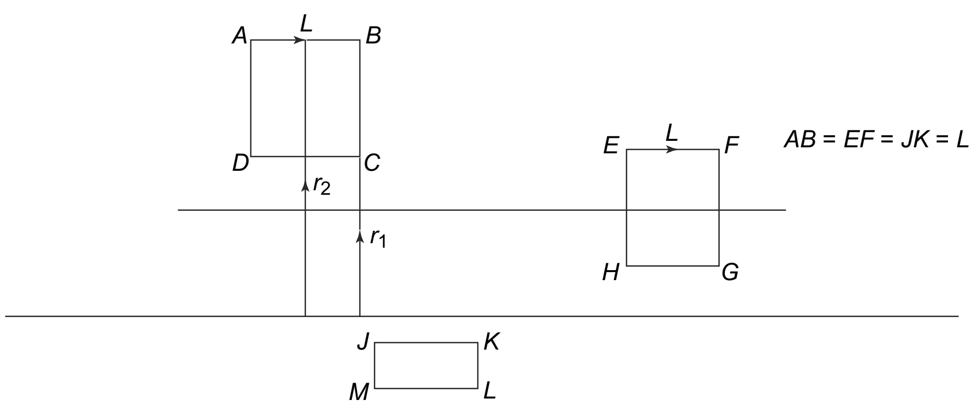


Fig. 6.31 *Rectangular paths for Ampere's Law. ABCD is totally outside the infinite solenoid, EFGH is partially inside and JKLM is totally inside the solenoid*

We evaluate the loop integral,

$$\oint \vec{B} \cdot \vec{dl}$$

For the path $ABCD$ the loop integral is zero since by Ampere's Law, this must be proportional to the current enclosed by the closed path which in this case is zero. But the line integral is the sum of contributions from the sections AB, BC, CD, DA. The contributions from the sections AD and BC vanish since these paths are orthogonal to the direction of the magnetic field which we have seen above has only a z component. Thus, since the loop integral vanishes, we have

$$\oint_{ABCD} \vec{B} \cdot \vec{d}l = 0 = \int_{AB} B_z(r_2)dl - \int_{DC} B_z(r_1)dl = \left[B_z(r_2) - B_z(r_1)\right] L \tag{6.110}$$

where we have denoted by $B_z(r)$ the z component of the field at a distance r from the axis. Thus, we have

$$B_z(r_2) = B_z(r_1)$$

The magnetic field outside the solenoid is the same at all points, irrespective of the distance from the axis. But, the field at $r = \infty$ must be zero since all currents are at infinite distance from there. Hence, the field at **all** points outside the solenoid must be zero. This is also obvious from the field lines drawn for the finite solenoid (Fig. 6.29). The field lines turn back from the ends. But if the ends are at infinity, for an infinite solenoid, then they don't turn back at any finite point.

Next, consider the loop $JKLM$ which is totally inside the solenoid. Once again, the loop integral is zero since the loop does not enclose any current. The contribution from the sections KL and MJ is zero since the field is along the z direction. We thus, have

$$\oint_{JKLM} \vec{B} \cdot \vec{d}l = 0 = \int_{JK} B_z(r_2)dl - \int_{ML} B_z(r_1)dl = \left[B_z(r_2) - B_z(r_1)\right] L \tag{6.111}$$

Once again, this shows that the magnetic field at **all** points inside the solenoid is equal.

Finally, consider the loop $EFGH$. In this case, a current of INL passes through the loop since there are N turns per unit length and each carries a current I. The sides FG and HE do not contribute as before. There is no field at EF since all of it is outside the solenoid. The field on GH is the constant field B_z^{in} inside the solenoid. Thus, we have

$$\oint_{EFGH} \vec{B} \cdot \vec{d}l = \int_{GH} B_z^{\text{in}} dl = B_z^{\text{in}} L = \mu_0 NIL \tag{6.112}$$

and so

$$\boxed{B_z^{\text{in}} = \mu_0 NI} \tag{6.113}$$

This is the same result we obtained earlier Eq. (6.107). To summarise, **for an infinite solenoid, the magnetic field only exists within the solenoid. At any point within the solenoid, the field is constant and is along the z direction, given by Eq. (6.113).**

A common variant of the cylindrical solenoid is the toroidal ring in which the solenoid is bent into a thick circular ring or torus as in Fig. 6.32(a).

In this geometry, the field lines of the solenoid (which, as we know, are parallel to the axis of the solenoid) get bent into circular shape. However, the toroid is obviously finite. It has an axis of symmetry, shown as the z axis in Fig. 6.32(a). It is easy to see that the magnetic field everywhere will be the same when we rotate around this axis of symmetry.

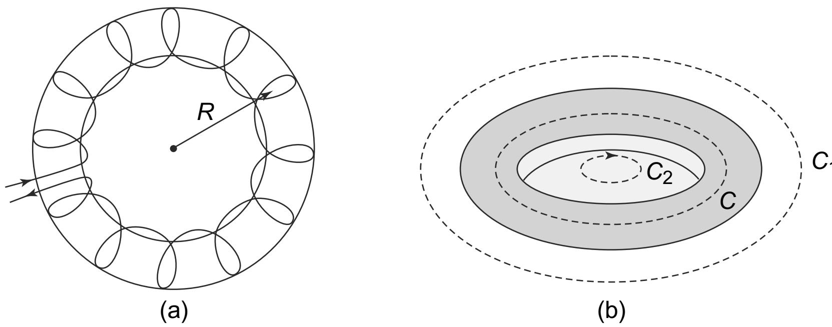


Fig. 6.32 *Toroidal ring in cylindrical solenoid: (a) A toroidal ring of radius R, (b) Lateral view*

To find the magnetic field outside the toroid, consider a circular loop C_1 concentric with the centre of the torus of radius R_1 where $R_1 > R$. Now the magnetic field, if there is any, must be the same at all points on the loop since there is azimuthal symmetry. On the other hand, the current enclosed by the loop C_1 is zero. Hence, by Ampere's law

$$\oint_{C_1} \vec{B} \cdot d\vec{l} = B_\phi(2\pi R_1) = 0 \tag{6.114}$$

and hence, the magnetic field would be zero outside the torus. The same argument would apply to points inside the inner surface of the torus where the magnetic field is zero. For points inside the torus, consider a loop C inside the core of the torus. Using Ampere's Law, we get

$$\oint_{C} \vec{B} \cdot d\vec{l} = B_\phi(2\pi R) = \mu_0 N I(2\pi R) \tag{6.115}$$

and hence,

$$\boxed{B_\phi = \mu_0 N I}$$

We thus, have the result that **for a torus, the magnetic field is confined to the inside of the torus. Inside the torus, the field has only an azimuthal component and is given by $\mu_0 N I$ where N is the number of turns per unit length and I is the current in the coil.**

6.7 MAGNETIC VECTOR POTENTIAL

In the case of the electric field, we saw that its conservative nature (the fact that the work done in moving a charge in an electric field is independent of the path taken but only dependent on the end points) allowed us to define a scalar potential ϕ. Alternatively, this is also related to the fact that

$$\vec{\nabla} \times \vec{E}(\vec{r}) = 0$$

for electrostatic fields. Since the curl of a gradient of a scalar field is by definition zero, this allowed us to write

$$\vec{E}(\vec{r}) = -\vec{\nabla}\phi(r)$$

In the case of a magnetic field, the curl of a magnetic field, in general is non-zero as we have seen above. Hence, it is not possible, in general to define a scalar potential for magnetic fields.

However, unlike the electric field lines, magnetic field lines are continuous and this implies that

$$\vec{\nabla} \cdot \vec{B} = 0$$

From vector analysis, we know that, the divergence of a curl is always zero and hence we can in general, write

$$\boxed{\vec{B}(\vec{r}) = \vec{\nabla} \times \vec{A}(\vec{r})} \tag{6.116}$$

where $\vec{A}(\vec{r})$ **is a vector field called the magnetic vector potential.** For any given $\vec{B}(\vec{r})$, one can always find a $\vec{A}(\vec{r})$ such that Eq. (6.116) is always satisfied.

6.7.1 Existence of the Vector Potential

Although it is true that the magnetic field is divergenceless and hence can be written in general as a curl of a vector, the question still remains about how do we know whether such a vector $\vec{A}$ would always exist so as to satisfy Eq. (6.116)? It turns out, that this is guaranteed by a theorem in vector analysis known as Helmhotz Theorem.

Helmohotz Theorem states that any vector field $\vec{v}(\vec{r})$ can be decomposed into a sum of two parts—one part has zero divergence and is called solenoidal while the other part has zero curl and is called irrotational.

The proof can be seen as follows:

Let

$$\vec{\nabla} \cdot \vec{v}(\vec{r}) = \vec{D}(\vec{r}) \tag{6.117}$$

and

$$\vec{\nabla} \times \vec{v}(\vec{r}) = \vec{C}(\vec{r}) \tag{6.118}$$

Then we can express $\vec{v}(\vec{r})$ as

$$\vec{v}(\vec{r}) = -\vec{\nabla}\phi(\vec{r}) + \vec{\nabla} \times \vec{W}(\vec{r}) \tag{6.119}$$

where

$$\phi(\vec{r}) = \frac{1}{4\pi} \iiint d^3r' \frac{\vec{D}(\vec{r}')}{|\vec{r} - \vec{r}'|} \tag{6.120}$$

and

$$\vec{W}(\vec{r}) = \frac{1}{4\pi} \iiint d^3r' \frac{\vec{C}(\vec{r}')}{|\vec{r} - \vec{r}'|} \tag{6.121}$$

To see this, consider that Green's function of the operator ∇^2 which we encountered in an earlier chapter.

$$\nabla^2 G(\vec{r}, \vec{r}') = \delta^3(\vec{r} - \vec{r}') \tag{6.122}$$

and hence

$$G(\vec{r}, \vec{r}') = -\frac{1}{4\pi} \frac{1}{|\vec{r} - \vec{r}'|} \tag{6.123}$$

Taking the divergence of Eq. (6.119), we get

$$\vec{\nabla} \cdot \vec{v}(\vec{r}) = \vec{D}(\vec{r}) = -\vec{\nabla} \cdot \vec{\nabla}\phi = -\nabla^2\phi \tag{6.124}$$

since

$$\vec{\nabla} \cdot \vec{\nabla} \times \vec{W}(\vec{r}) = 0$$

The solution of Eq. (6.124) is Eq. (6.120) as can be easily verified using the Green's function in Eq. (6.123).

Similarly, taking the curl of Eq. (6.119), we get

$$\vec{\nabla} \times \vec{v}(\vec{r}) = \vec{C}(\vec{r}) = -\nabla^2\vec{W}(\vec{r}) + \vec{\nabla}\left[\vec{\nabla} \cdot \vec{W}(\vec{r})\right] \tag{6.125}$$

where we have used

$$\vec{\nabla} \times (-\vec{\nabla}\phi) = 0$$

and

$$\vec{\nabla} \times (\vec{\nabla} \times \vec{W}) = \vec{\nabla}(\vec{\nabla} \cdot \vec{W}) - \nabla^2\vec{W}$$

We need to find the solution to Eq. (6.125). We will now show that the solution to this is indeed Eq. (6.121).

To see this, note that the first term on the RHS of Eq. (6.125) obviously gives us $\vec{C}(\vec{r})$ using Eq. (6.122) and Eq. (6.123). If we can show that the second term on the RHS of Eq. (6.125) vanishes, we would have proved Helmhotz Theorem. For the second term, we see that

$$\vec{\nabla}_r(\vec{\nabla}_r \cdot \vec{W}(\vec{r})) = \frac{1}{4\pi}\iiint d^3r'\vec{\nabla}_r\left[\vec{\nabla}_r \cdot \left(\frac{\vec{C}(\vec{r}')}{|\vec{r}-\vec{r}'|}\right)\right] \tag{6.126}$$

The $\vec{\nabla}_r$ inside the bracket above only acts on $\frac{1}{|\vec{r}-\vec{r}'|}$ and hence, we can replace it with $-\vec{\nabla}_{r'}$. Eq. (6.126) then reads

$$\vec{\nabla}_r(\vec{\nabla}_r \cdot \vec{W}(\vec{r})) = -\frac{1}{4\pi}\vec{\nabla}_r\iiint d^3r'\left[\vec{\nabla}_r' \cdot \left(\frac{\vec{C}(\vec{r}')}{|\vec{r}-\vec{r}'|}\right)\right] \tag{6.127}$$

The integral in Eq. (6.127) is just the integral of a divergence. Hence, using Gauss's Divergence Theorem, we get

$$\iiint d^3r'\left[\vec{\nabla}_r' \cdot \left(\frac{\vec{C}(\vec{r}')}{|\vec{r}-\vec{r}'|}\right)\right] = \iint d\vec{S} \cdot \left(\frac{\vec{C}(\vec{r}')}{|\vec{r}-\vec{r}'|}\right) \tag{6.128}$$

where the surface integral is over a surface encompassing the volume, which in this case is a surface at infinity. If we assume $\vec{v}(\vec{r})$ to vanish at infinity, it is clear that then $\vec{C}(\vec{r})$ will also vanish at infinity. In this case then, the surface integral will be zero and thus,

$$\vec{\nabla}_r(\vec{\nabla}_r \cdot \vec{W}(\vec{r})) = 0$$

Let us recapitulate what we have proved- we have shown that any vector field $\vec{v}(\vec{r})$ can be written as Eq. (6.119) where the functions $\phi(\vec{r})$ and $\vec{W}(\vec{r})$ can be explicitly evaluated as in Eq. (6.120) and Eq. (6.121). Thus we have established Helmhotz Theorem.

We see that any magnetic field $\vec{B}(\vec{r})$can be expressed in terms of a curl of a vector potential $\vec{A}(\vec{r})$. However, a moment's reflection will make it clear that the choice of $\vec{A}$ is not unique. Recall that in the case of the electric field and potential, we could always add a constant term to the electric potential ϕ without changing the electric field

$$\vec{E}(\vec{r}) = -\vec{\nabla}\phi(\vec{r}) = -\vec{\nabla}(\phi(\vec{r}) + (\text{constant})) = -\vec{\nabla}\phi(\vec{r}) \tag{6.129}$$

For the magnetic case, the freedom is actually much greater. Since the magnetic field is the curl of the a vector potential, we can add any term whose curl is zero to the vector potential without changing the magnetic field. In general, we know that the curl of a gradient vanishes and so we can always add the gradient of a scalar field $\chi(\vec{r})$ to the vector potential $\vec{A}(\vec{r})$.

$$\vec{A}(\vec{r}) \rightarrow \vec{A}'(\vec{r}) = \vec{A}(\vec{r}) + \vec{\nabla}\chi(\vec{r}) \tag{6.130}$$

and hence,

$$\begin{aligned}\vec{B}(\vec{r}) &= \vec{\nabla} \times \vec{A}'(\vec{r}) \\ &= \vec{\nabla} \times \vec{A}(\vec{r}) + \vec{\nabla} \times \vec{\nabla}\chi(\vec{r}) \\ &= \vec{\nabla} \times \vec{A}(\vec{r})\end{aligned} \tag{6.131}$$

since

$$\vec{\nabla} \times (\vec{\nabla}\chi(\vec{r})) = 0$$

The transformations Eq. (6.129) and Eq. (6.130) are known as gauge transformations.

There is another question which arises in the introduction of the vector potential. In the case of the electric field, introducing a scalar potential $\phi(\vec{r})$ in terms of a vector electric field $\vec{E}(\vec{r})$ led to a simplification of the equations since now the problem was formulated in terms of a single unknown ϕ instead of three unknown components of the electric field. For the magnetic vector potential, this convenience is not there since both the magnetic field and the associated vector potential are vectors. However, the magnetic vector potential allows for a much larger freedom of choosing the vector potential because of gauge transformations. One can exploit the gauge transformations to simplify things. Let us see how this can be done.

Consider the Ampere's Circuital law Eq. (6.100) and rewrite it in terms of the vector potential $\vec{A}(\vec{r})$.

$$\vec{\nabla} \times \vec{B}(\vec{r}) = \mu_0 \vec{j}(\vec{r})$$

gives us, in terms of $\vec{A}$

$$\begin{aligned}\vec{\nabla} \times (\vec{\nabla} \times \vec{A}(\vec{r})) &= \mu_0 \vec{j}(\vec{r}) \\ \vec{\nabla}\left[\vec{\nabla} \cdot \vec{A}(\vec{r})\right] - \nabla^2 \vec{A}(\vec{r}) &= \mu_0 \vec{j}(\vec{r})\end{aligned} \tag{6.132}$$

We now use the freedom that is allowed by the gauge transformation (Eq. (6.130)). We choose the scalar function $\chi(\vec{r})$ such that

$$\vec{A}'(\vec{r}) = \vec{A}(\vec{r}) + \vec{\nabla}\chi(\vec{r})$$

has zero divergence.

$$\vec{\nabla}\cdot\vec{A}'(\vec{r}) = \vec{\nabla}\cdot\vec{A}(\vec{r}) + \vec{\nabla}\cdot\vec{\nabla}\chi(\vec{r}) = 0 \tag{6.133}$$

Such a $\chi(\vec{r})$ can always be found since using Eq. 6.133), we see that

$$\nabla^2\chi(\vec{r}) = -\vec{\nabla}\cdot\vec{A}(\vec{r})$$

or

$$\chi(\vec{r}) = -\frac{1}{4\pi}\iiint d^3r'\,\frac{\vec{\nabla}\cdot\vec{A}(\vec{r}')}{|\vec{r}-\vec{r}'|} \tag{6.134}$$

where we have used the Green's function for ∇^2 as given in Eq. (6.122). If $\vec{\nabla}\cdot\vec{A}(\vec{r}) \neq 0$, then Eq. (6.134) gives us $\chi(\vec{r})$ such that $\vec{\nabla}\cdot\vec{A}'(\vec{r}) = 0$. This of course assumes that $\vec{\nabla}\cdot\vec{A}(\vec{r})$ goes to zero at ∞ faster than $\frac{1}{r^2}$ so that the integral converges. We now drop the prime from the divergenceless vector potential and call it $\vec{A}(\vec{r})$ with $\vec{\nabla}\cdot\vec{A}(\vec{r}) = 0$ without any loss of generality. With this choice of $\vec{A}$, Eq. (6.132) becomes

$$\boxed{\nabla^2\vec{A}(\vec{r}) = -\mu_0\vec{j}(\vec{r})} \tag{6.135}$$

which is remarkably similar to the Poisson's equation we encountered in electrostatics for the scalar potential $\phi(\vec{r})$.

$$\nabla^2\phi(\vec{r}) = -4\pi k\rho(\vec{r}) = -\frac{1}{\varepsilon_0}\rho(\vec{r}) \tag{6.136}$$

Thus, it turns out that using the freedom of gauge transformations allows us to replace Ampere's law by a much simpler Poisson's equation for the vector potential $\vec{A}(\vec{r})$. We shall also see later when we consider time varying currents, that equations with the vector potential are much easier to solve than those involving magnetic fields. We now evaluate the vector potential for some representative magnetic fields. But before that, we try to see if we can write a general form of the vector potential.

It turns out that for magnetic fields which are generated by currents, a general form of the vector potential exists. Consider the expression for the magnetic field at $\vec{r}$ due to a current density $\vec{j}$ given by Eq. (6.80).

$$\vec{B}(\vec{r},\vec{r}') = \left(\frac{\mu_0}{4\pi}\right)\iiint d^3r'\,\frac{\vec{j}(\vec{r}')\times(\vec{r}-\vec{r}')}{|\vec{r}-\vec{r}'|^3} \tag{6.137}$$

We know that

$$\left(\frac{\vec{r}-\vec{r}'}{|(\vec{r}-\vec{r}')|^3}\right) = -\vec{\nabla}_r\left(\frac{1}{|\vec{r}-\vec{r}'|}\right) \tag{6.138}$$

Hence,

$$\vec{B}(\vec{r}) = -\left(\frac{\mu_0}{4\pi}\right)\iiint d^3r'\,\vec{j}(\vec{r}')\times\vec{\nabla}_r\left(\frac{1}{|\vec{r}-\vec{r}'|}\right) \tag{6.139}$$

But

$$\vec{j}(\vec{r}')\times\vec{\nabla}_r\left(\frac{1}{|\vec{r}-\vec{r}'|}\right) = -\vec{\nabla}_r\times\left(\frac{\vec{j}(\vec{r}')}{|\vec{r}-\vec{r}'|}\right) \tag{6.140}$$

since $\vec{j}(\vec{r}')$ is not differentiated by $\vec{\nabla}_r$. This can be seen as follows. Using the index notation that we introduced earlier, for any vector field $\vec{A}(\vec{r}')$,

$$\left[\vec{A}(\vec{r}') \times \vec{\nabla}_r f(\vec{r}, \vec{r}')\right]_i = \varepsilon_{ijk} A_j(\vec{r}') \frac{\partial}{\partial x^k}(f(\vec{r}, \vec{r}'))$$

which is equal to

$$-\varepsilon_{kji} \frac{\partial}{\partial x^k} A_j(\vec{r}') f(\vec{r}, \vec{r}')$$

since ε_{ijk} is completely antisymmetric under change of indices. Therefore,

$$(\vec{A}(\vec{r}') \times \vec{\nabla}_r f(\vec{r}, \vec{r}'))_i = -\varepsilon_{kji} \frac{\partial}{\partial x^k} A_j(\vec{r}') f(\vec{r}, \vec{r}') = -\left[\vec{\nabla}_r \times (\vec{A} f)\right]_i$$

Thus, we get

$$\vec{B}(\vec{r}) = \left(\frac{\mu_0}{4\pi}\right) \vec{\nabla}_r \times \iiint d^3r' \frac{\vec{j}(\vec{r}')}{|\vec{r} - \vec{r}'|} \tag{6.141}$$

But

$$\vec{B}(\vec{r}) = \vec{\nabla} \times \vec{A}(\vec{r})$$

and hence, the form of vector potential is simply

$$\boxed{\vec{A}(\vec{r}) = \left(\frac{\mu_0}{4\pi}\right) \iiint d^3r' \frac{\vec{j}(\vec{r}')}{|\vec{r} - \vec{r}'|}} \tag{6.142}$$

As always, this $\vec{A}$ is not unique since the freedom given by gauge transformations always allows us to add a gradient of a scalar field to this vector potential without changing the magnetic field derived from it.

We now illustrate with some examples.

EXAMPLE 6.10 Calculate the vector potential for a uniform magnetic field $\vec{B}$ in the z-direction.

Solution

We have that

$$\vec{B} = B\hat{k}$$

With this field, and the definition of the vector potential, we thus have

$$(\vec{\nabla} \times \vec{A})_x = \frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z} = 0$$
$$(\vec{\nabla} \times \vec{A})_y = \frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x} = 0$$
$$(\vec{\nabla} \times \vec{A})_z = \frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y} = B$$

$\vec{A}$ would thus be proportional to $\vec{B}$ and the simplest choice would be where $\vec{A}$ is a linear function of the coordinates $\vec{r}$. Since $\vec{A}$ is a vector, there are two choices

$$\vec{A} = C_1 B \vec{r}$$
$$\text{or} \quad \vec{A} = C_2(\vec{B} \times \vec{r})$$

where C_1, C_2 are constants. It is easy to check that $\vec{A} = C_1 B\vec{r}$ gives us a zero magnetic field since $\vec{\nabla} \times \vec{r} = 0$. Thus, we have a possible choice for $\vec{A}$ for this $\vec{B}$ as

$$\vec{A} = C_2(\vec{B} \times \vec{r})$$

or

$$\begin{aligned} A_x &= -C_2 By \\ A_y &= C_2 Bx \\ \text{and} \quad A_z &= 0 \end{aligned}$$

On substituting in the definition of $\vec{A}$, we get

$$C_2 = \frac{1}{2}$$

and hence

$$\vec{A} = \frac{1}{2}(\vec{B} \times \vec{r})$$

is a possible choice for the vector potential which gives us the given magnetic field. This, as we have seen before, is not unique since we can always add a gradient of a scalar field to it without changing the magnetic field. Thus, for example, we could add

$$\vec{A}' = \vec{A} + C_3\vec{\nabla}(xy) = (-\frac{1}{2}By, \frac{1}{2}Bx, 0) - C_3(y, x, 0)$$

By choosing $C_3 = \frac{1}{2}B$, we could get

$$\vec{A}' = (0, B_x, 0)$$

and by choosing $C_3 = -\frac{1}{2}B$, we could get

$$\vec{A}' = (-B_y, 0, 0)$$

Thus, there is an ambiguity in the definition of $\vec{A}$.

EXAMPLE 6.11 Calculate the vector potential for the magnetic field generated by an infinite thin straight wire carrying a current I shown in Fig. 6.33.

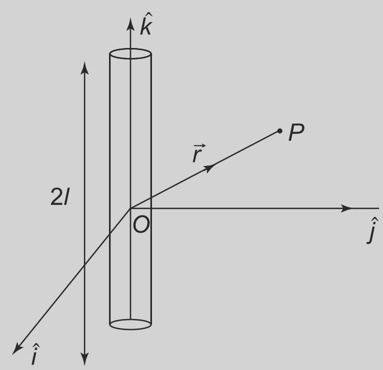


Fig. 6.33 *Example 6.11: The vector potential for the magnetic field generated by an infinite thin straight wire carrying a current I*

Solution

We take the direction of the wire to be along the z-axis. Though we have to find the vector potential for the magnetic field of an infinite wire, we will perform the calculation for a finite wire, extending from $-l$ to $+l$ and then take the limit of $l \to \infty$. Further, the wire would be considered to be of circular cross section with radius a.

Now

$$\begin{aligned}\vec{j}(\vec{r}) &= \frac{I}{\pi a^2}\hat{k} \quad -l < z < l, 0 < x^2 + y^2 < a \\ &= 0 \quad \text{otherwise}\end{aligned} \tag{6.143}$$

It is easy to see that

$$\iint d\vec{S} \cdot \vec{j}(\vec{r}) = I\hat{k} \quad \text{for } -l < z < l$$

But

$$|\vec{r} - \vec{r}'|^2 = (x - x')^2 + (y - y')^2 + (z - z')^2$$

Since we would be taking the limit of an infinitely thin wire, $a \to 0$, we can write

$$|\vec{r} - \vec{r}'|^2 \approx (x)^2 + (y)^2 + (z - z')^2$$

From the general expression for $\vec{A}$, we have

$$\vec{A}(\vec{r}) = \left(\frac{\mu_0}{4\pi}\right) \iiint dx'dy'dz' \frac{\vec{j}(\vec{r}')}{|\vec{r} - \vec{r}'|}$$

The integrals over dx', dy' are trivial and give us πa^2. Thus

$$\vec{A}(\vec{r}) = \left(\frac{\mu_0}{4\pi}\right) I \int_{-l}^{+l} dz' \frac{\vec{j}(\vec{r}')}{\sqrt{(z - z')^2 + \rho^2}}\hat{k}$$

where

$$\rho^2 = x^2 + y^2$$

This give us

$$\vec{A}(\vec{r}) = \left(\frac{\mu_0}{4\pi}\right) I\hat{k} \ln\left[(z' - z) + \sqrt{(z' - z)^2 + \rho^2}\right]_{-l}^{+l}$$

or

$$\vec{A}(\vec{r}) = \left(\frac{\mu_0}{4\pi}\right) I\hat{k} \ln\left[\frac{(l - z) + \sqrt{(l - z)^2 + \rho^2}}{(-l - z) + \sqrt{(l + z)^2 + \rho^2}}\right]$$

This is the result for a wire of cross section a and length $2l$. To find the actual vector potential, we need to take the limit $l \to \infty$ keeping z finite. On doing this, we obtain

$$\vec{A}(\vec{r}) = \left(\frac{\mu_0}{4\pi}\right) I\hat{k} \ln\left[\frac{2(l - z) + \frac{\rho^2}{2(l-z)}}{\frac{\rho^2}{2(l+z)}}\right]$$

$$\approx \left(\frac{\mu_0}{4\pi}\right) I\hat{k} \ln\left[1 + \frac{4l^2}{\rho^2}\right]$$

It is easy to check that this $\vec{A}(\vec{r})$ gives us

$$B_x = -\left(\frac{\mu_0}{2\pi}\right) I \frac{y}{\rho^2}$$
$$B_y = +\left(\frac{\mu_0}{2\pi}\right) I \frac{x}{\rho^2}$$
$$B_z = 0$$

and hence,

$$B = |\vec{B}| = (B_x^2 + B_y^2 + B_z^2)^{1/2} = \left(\frac{\mu_0}{2\pi}\right) \frac{I}{\rho}$$

which is exactly the result we obtained earlier from the Biot-Savart Law for the magnetic field of an infinite wire.

6.7.2 Multipole Expansion for the Magnetic Vector Potential

Equation (6.142) is remarkably similar to Eq. (3.73) (Advanced Topic, Chapter 3) for the electrostatic case. In the electrostatic case, we saw that the potential could be expanded in terms of multipole moments Eq. (3.78). A similar exercise can be carried out for the magnetic vector potential since the forms are very similar.

Consider a current distribution $\vec{j}(\vec{r}')$ in the form of a current loop which is localised in a region of size much smaller than the distance r at which we wish to evaluate the vector potential. We already have seen that

$$\frac{1}{|\vec{r} - \vec{r}'|} = \sum_l \frac{r'^l}{r^{l+1}} P_l(\cos\theta') \tag{6.144}$$

where θ' is the angle as shown in Fig. 6.34.

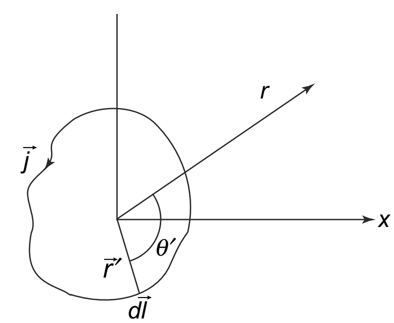


Fig. 6.34 *A small current loop with current $\vec{I}$. The current flows across the loop and hence, the current density $\vec{j}(\vec{r}')$ is non-zero only on the loop*

The vector potential in Eq. (6.142) can thus be written as

$$\vec{A}(\vec{r}) = \left(\frac{\mu_0}{4\pi}\right)\sum_l \frac{1}{r^{l+1}}\left[\int d^3r' r'^l P_l(\cos\theta')\vec{j}(\vec{r}')\right] \tag{6.145}$$

The quantity in the square brackets in Eq. (6.145) are the magnetic multipole moments or order l of this current distribution. Just like in the electrostatic case, the contribution of the l^{th} moment to $\vec{A}(\vec{r})$ goes as $\frac{1}{r^{l+1}}$. This therefore, allows us a useful expansion for large values of r as in the electrostatic case.

Let us start with the lowest term $l = 0$. It involves

$$\int d^3r' \vec{j}(\vec{r}') \tag{6.146}$$

Since $\vec{j}(\vec{r}')$ is non-zero only on the current loop, we can write it as

$$\iiint d^3r' \vec{j}(\vec{r}') = \oint d\vec{l}' I$$

where $d\vec{l}'$ is an infinitesimal displacement vector along the loop as shown in Fig. 6.34. But

$$\iiint d^3r' \vec{j}(\vec{r}') = \oint d\vec{l}' I = 0 \tag{6.147}$$

since the loop is closed. Thus, we see that the $l = 0$ term, or what is called the monopole term is absent in the multipole expansion. Let us look at the $l = 1$ term now. This will, in analogy to the electrostatic case, give us the magnetic dipole term. Thus,

$$\begin{aligned}\vec{\mu} &= \iiint d^3r' r' P_l(\cos\theta')\vec{j}(\vec{r}') \\ &= \vec{I}\oint r'\cos\theta' d\vec{l}' \end{aligned} \tag{6.148}$$

Now we know that

$$\cos\theta' = \hat{r}\cdot\hat{r}'$$

and hence, the integral becomes

$$\vec{\mu} = \vec{I}\oint \vec{r}'\cdot\hat{r}\, d\vec{l}'$$

But

$$\oint \vec{r}'\cdot\hat{r}\, d\vec{l}' = -\hat{r}\cdot\vec{a}$$

where $\vec{a}$ is the area of the loop. Thus, the dipole contribution to the vector potential in Eq. (6.145) becomes

$$\vec{A}_{\text{dipole}}(\vec{r}) = \left(\frac{\mu_0}{4\pi}\right)\frac{\vec{\mu}\times\vec{r}}{r^3} \tag{6.149}$$

where

$$\vec{\mu} = I\vec{a}$$

With this vector potential, one can easily calculate the magnetic field due to a dipole as

$$\vec{B}(\vec{r}) = \left(\frac{\mu_0}{4\pi}\right)(\vec{\nabla} \times \vec{A}_{\text{dipole}}) = \left(\frac{\mu_0}{4\pi}\right)\frac{3(\vec{\mu}\cdot\hat{r})\hat{r} - \vec{\mu}}{r^3}$$

If the dipole is located not near the origin but at a coordinate $\vec{r}'$ then obviously Eq. (6.149) becomes

$$\vec{A}_{\text{dipole}}(\vec{r}) = \left(\frac{\mu_0}{4\pi}\right)\frac{\vec{\mu}(\vec{r}') \times (\vec{r} - \vec{r}')}{|r - r'|^3} \tag{6.150}$$

Finally, if the dipoles are distributed in a volume with a dipole moment density $\vec{M}(\vec{r}')$, then the vector potential becomes

$$\vec{A}_{\text{dipole}}(\vec{r}) = \left(\frac{\mu_0}{4\pi}\right)\iiint d^3r'\,\frac{\vec{M}(\vec{r}') \times (\vec{r} - \vec{r}')}{|r - r'|^3} \tag{6.151}$$

PROBLEM 6.11 Verify by explicit evaluation that the magnetic field due to a magnetic dipole at a point removed from it has zero divergence and zero curl.

SUMMARY

- Current carrying elements experience a force in the presence of a magnetic field
- Charges in motion experience a force in magnetic fields. The force due to a magnetic field on a charge in motion acts in a direction perpendicular to both the velocity or the charge and the magnetic field. Consequently, the kinetic energy of the charged particle does not change in a magnetic field.
- A current carrying conductor placed in a magnetic field experiences an accumulation of charges transverse to the direction of current, which is called the Hall effect.
- A current carrying loop behaves like a magnetic dipole. A magnetic dipole experiences a torque in a uniform field and a force in a non-uniform field.
- Magnetic fields are produced by current carrying elements and they are given by the Biot-Savart Law, which relates the current density to the magnetic field.
- Divergence of the magnetic field is always zero. The curl of the magnetic field is related to the current density by Ampere's Law.
- The magnetic field can be written in terms of a vector potential whose curl gives the magnetic field.

CONCEPTUAL QUESTIONS

1. A conducting slab has current to the right. A $\vec{B}$ field is applied out of the page. Due to magnetic forces on the charge carriers, the bottom of the slab is at a higher electric potential than the top of the slab. Can the sign of the charge carriers in the slab be determined from this information alone? If yes, what is the sign?

2. A Helmholtz coil is a frequently used configuration of two circular coils of the same radius with a common axis separated by some distance. The configuration is used in many applications where a relatively uniform magnetic field is required over a small region in space. A magnetic dipole of dipole moment $\vec{\mu}$ is placed in the Helmhotz coil. Would it experience a net force only, a net torque only, or both?
3. Two parallel wires are carry a current I in the same direction. The wires will be
   a. attracted, since likes attract
   b. repelled, since likes repel each other
   c. experience no force
   d. they will experience a torque
4. Two straight parallel wires carry currents $2I$ and I in opposite directions. The magnetic force acting on the first wire is such that
   a. it will be repelled from the second wire
   b. it will be attracted to the second wire
   c. will remain in its position
   d. None of the above
5. Which of these is NOT a characteristic of a static magnetic field?
   a. It has vanishing divergence
   b. It has vanishing curl
   c. It has no sinks and sources
   d. Magnetic flux lines are always closed.
6. Two identical coaxial circular coils carry the same current I but in opposite directions. The magnitude of $\vec{B}$ at a point on the axis midway between the two coils is
   a. zero
   b. the same as produced by one coil
   c. twice that produced by one coil
   d. half that produced by one coil.
7. A charged particle enters a region with a magnetic field but does not experience any force. What can be concluded about the magnetic field?
8. A thin wire is bent into a semi-circular loop with two ends as straight lines. The magnetic field at the point P the centre of the semicircle is equal to the field of
   a. a semicircle
   b. a semicircle plus the field of a long straight wire
   c. a semicircle minus the field of a long straight wire
   d. None of the above
9. The expression

$$\oint \vec{B} \cdot d\vec{S}$$

   a. is equal to the magnetic work done around a closed path
   b. is equal to the current through an open surface bounded by the closed path

c. is always zero
d. None of the above

10. If a number of dipoles are randomly scattered through space, after a while they
a. attract each other and move closer
b. repel each other and move apart
c. will be stationary
d. None of the above

PROBLEMS

1. A charged particle of mass $m = 2$ kg and charge $q = 3$ C starts at a point $A(1, -2, 0)$ with a velocity $\vec{v}_0 = 2\hat{i} + 3\hat{k}$ m/s in an electric field given by $\vec{E} = (12\hat{i} + 10\hat{j})$ V/m. At time $t = 2$ s, find the
a. the acceleration and velocity of the particle
b. its position.
Determine the trajectory of the particle subsequently.

2. A charged particle of mass $m = 2$ kg and charge $q = 1$ C is at the origin. It is given an initial velocity of $\vec{v}_0 = 3\hat{j}$ m/s and it travels in a region with a uniform magnetic field $\vec{B} = 10\hat{k}$ T. Find the velocity, acceleration and also the trajectory of the particle.

3. Mass Spectrometer: A mass spectrometer is a device routinely used to analyse gas samples by measuring the abundance of particles of different masses. The device consists of chamber where the gas sample is placed and an electric discharge creates ions from the neutral gas molecules. An ion of mass m and charge q is produced at rest and then accelerated through a potential difference ΔV in this chamber. The ion then enters a selector chamber which has crossed $\vec{E}$ and $\vec{B}$ fields, the $\vec{B} = \vec{B}_1$ field being adjustable. Only ions of a selected velocity leave the chamber through a slit. The emerging ions then enter a region where there is a magnetic field $\vec{B}_2$ perpendicular to the plane of the paper. The ion trajectory is bent and they are detected by a detector placed at a distance x from the slit through which they entered the region of $\vec{B}_2$.
a. What is the magnetic field $\vec{B}_1$ needed to ensure that the particles move straight through the selector chamber?
b. What is the mass of the particle hitting the detector at a distance x?
Express your answers in terms of $|\vec{E}|, x, q, m, |\vec{B}_2|$ and ΔV.

4. A copper wire with density $\rho = 8960$ kgm^{-3} is made into a circular loop of radius 0.5 m. The cross-sectional area of the wire is 10 μ m. A voltage of 0.01 V is applied to the wire. When the coil is placed in a magnetic field of magnitude 0.25 T, what is the maximum angular acceleration experienced by the loop? Assume the loop rotates about an axis through its diameter.

5. A long rod of circular cross section of radius R has a volume charge density ρ_0. Find the magnetic field inside as well as outside of the cylinder if the cylinder rotates around its axis with an angular velocity ω.

6. Obtain the magnetic field due to a current I_0 flowing through a hollow cylindrical conductor of inner radius R_1 and outer radius R_2.

7. A rectangular slab of length l, breadth b and thickness t carries an uniform current density $\vec{j}$ parallel to the length of the slab. Find the value of the magnetic field at points inside and outside the slab.

8. Coaxial Transmission Line: Consider an infinite transmission line consisting of two concentric cylindrical conductors with their axes along the z-axis. The inner cylinder has a radius a while the outer cylinder has an inner radius b and a thickness t. The cylinders carry a current I in opposite directions. Assume that the current is uniformly distributed in the conductors. Find the magnetic field at all points.

9. A flat strip of copper of width a and negligible thickness carries a current I. Find the magnetic field at a point P, at right angles to the strip, at a distance R from the centre of the strip.

10. A thin spherical shell of radius R is charged uniformly with surface charge density σ. It is rotating around a diameter with an angular speed ω. Calculate the vector potential at points inside and outside the shell.

11. In a region, the magnetic vector potential $\vec{A}$ is given by $\vec{A} = -\frac{\rho^2}{4}\hat{z}$ in appropriate units (Tm^{-1}). Calculate the magnetic flux passing through a surface defined by $\phi = \frac{\pi}{2}, 1 \le \rho \le 2, 0 \le z \le 5$ m?

12. A copper wire of diameter d carries a current density $\vec{j}$. The wire is at the earth's equator where the earth's magnetic field is horizontal and points north and has a magnitude B_0. The wire lies in a plane parallel to the earth's surface in the east-west direction. Given that the density of copper is ρ_c what is the value of $\vec{j}$ for which the wire would just levitate? Find the current for $\rho_c = 8.9 \times 10^3$ kgm^{-3}, $B_0 = 0.5 \times 10^{-4}$ T, $g = 9.8$ msec^{-2}

13. Two parallel coaxial circular loops, each of radius a and each carrying a current I_0 in the same sense are a distance a apart. Show that at the midpoint between the loops, on their common axis, the first, second and third derivatives (with respect to the distance from the centre of any one of the coils) of the axial magnetic field vanish.

14. Four current elements are shown in Fig. 6.35 at four corners of a square of size L. Calculate the force on each one due to each of the others.

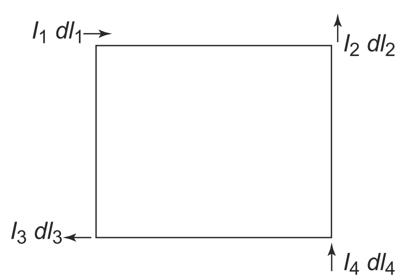


Fig. 6.35 *Problem 15*

15. Consider a coil with N turns, radius R carrying a current I. An arrangement with a spring and a magnet is used to determine the magnetic field due to such a coil in various situations. The magnet has a dipole moment μ and mass m while the spring constant of the spring is k. The current in the coil is in the counterclockwise direction and the spring arrangement is placed above the coil along the axis.

a. What is the force on the dipole when the current is switched on in the coil?
b. With the current on, the brass rod is lifted until the disk magnet is sitting a distance z_0 above the top of the coil. Now the current is turned off. Does the disk magnet move up or down? Find the displacement Δz to the new equilibrium position of the disk magnet.
c. At what height is the force on the dipole the largest?
d. What is the force when the field is a maximum?

16. A rectangular loop carrying a current I_2 is placed close to a straight conductor carrying a current I_1 as in Fig. 6.36. Calculate the magnetic force on the loop.

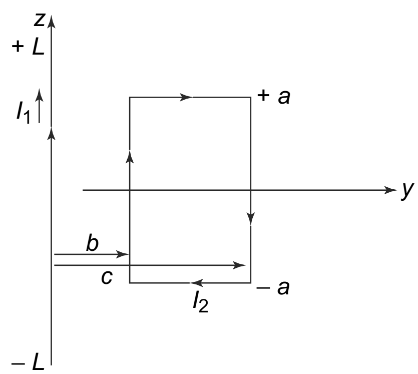


Fig. 6.36 *Problem 16*

17. A wire is bent as shown in Fig. 6.37 and lies in the $x-y$ plane. The wire carries a current I and the magnetic field in the region is $\vec{B} = B\hat{k}$. Determine the force on the wire.

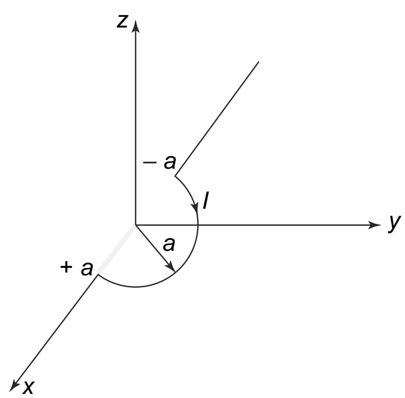


Fig. 6.37 *Problem 17*

18. A hollow cylindrical conductor of radii a and b carries a current I uniformly spread over its cross-section as in Fig. 6.38. Show that the magnetic field inside the body of the conductor is given by

$$B = 2\left(\frac{\mu_0}{4\pi}\right)\frac{I}{(b^2 - a^2)}\frac{r^2 - a^2}{r}$$

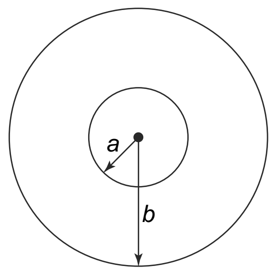


Fig. 6.38 *Problem 18*

19. A circular copper loop of radius 10 cm carries a current 15 A. At its centre is placed a second loop of radius 1 cm having 50 turns and carrying a current of 1 A. Find the magnetic field due to the larger loop at its centre. Also find the torque on the smaller loop. Assume that the planes of the two loops are perpendicular to each other and the induction $\vec{B}$ due to the larger loop is uniform over the smaller loop.

20. A wire in the form of a regular polygon of n sides is just enclosed by a circle of radius a. If the current in the wire is I, show that the magnetic induction at the centre is

$$B = 2\left(\frac{\mu_0}{4\pi}\right)\frac{In}{a}\tan\left(\frac{\pi}{n}\right)$$

Show that in the limit $n \to \infty$, this result gives us the result for a circular loop.

21. A 10 cm wide, 10 cm long and 1 cm thick copper strip carries a current of 100 A. It is placed at right angles in a uniform magnetic field of 1.75 T. Determine the Hall voltage and the Hall electric field intensity. Also determine the electric field responsible for the current in the strip. The conductivity of copper is 5.8×10^7 SI units and there are 8.5×10^{28} free electrons in copper per cubic meter.

7

Magnetic Properties of Matter

Learning Objectives

- To learn about the various kinds of magnetic materials and their properties.
- To understand the behaviour of the magnetic materials in terms of atomic level phenomenon.
- To learn about the way to look at magnetisation in terms of bound current densities.
- To understand the need for the introduction of the magnetic field $\vec{H}$.
- To comprehend the various kinds of magnetic phenomena in bulk matter.
- To learn about ferromagnetism and hysteresis.
- To understand the splitting of spectral lines in the presence of a magnetic field.
- To learn about boundary conditions on magnetic field vectors at an interface and solve boundary value problems.
- To understand the concept of magnetic circuits.

7.1 MAGNETIC MATERIALS

7.1.1 Introduction

In a previous chapter, we discussed the effects of an electrostatic field on matter. We saw that in the presence of certain kinds of substances called *dielectric substances*, the electric field is modified. We also saw that the dielectric substances themselves interact with the external electric field and get polarised. This, was because the external electric fields modify the atomic level dipoles and create induced dipole moments. The electric polarsation $\vec{P}$, a quantity which is related to the amount of dipole moment in a unit volume, is proportional to the applied electric field $\vec{E}$ and the proportionality constant is characteristic of the material and is called the electric susceptibility χ.

$$\vec{P} = \chi \vec{E}$$

What happens when we place certain materials in an external magnetic field? Do they interact with the external magnetic field and modify it? If they do, in what ways and by what mechanism is this achieved? These are some of the questions which we seek to answer in this Chapter.

On the face of it, it seems clear that there cannot be an exact parallel of the phenomenon of polarisation for the magnetic case. The reason for this is that unlike electric charges, there are no magnetic charges. However, we have seen that current loops behave like magnetic dipoles. We also know that at the atomic level, the atomic charges, specifically the electrons are not stationary, but in constant motion. And since charges in motion constitute currents, we expect all materials to have some kind of current loops at the atomic level. These current loops, as we have seen, produce magnetic fields and also interact with external magnetic fields. We thus expect that an external magnetic field will bring about changes in these currents at the atomic level and hence there would be some interaction between an external magnetic field and the material. This, it is reasonable to assume, might cause some kind of magnetic effects in bulk materials. This is indeed the case as we shall see in this chapter, though matter shows a much wider range of behaviour under the influence of magnetic fields than with electric fields.

One class of materials exhibit a property very similar to that of polarisation in dielectrics. These materials, called **paramagnetic**, have some permanent magnetic dipole moment at the atomic level and in the presence of an external field, the dipoles tend to align themselves with the applied field. The net dipole moment thus is in the same direction as the applied field. As we will see later, in paramagnetic materials, not all atomic dipoles are aligned in the direction of the magnetic field except under certain very specific and extreme conditions of very high field and very low temperatures. Examples of paramagnetic materials are aluminium, platinum, chromium, etc.

Another class of materials have a magnetic dipole polarisation in a direction opposite to the applied field. These have no counterpart in the electrostatic case. These materials are called **diamagnetic** materials. At the atomic level, diamagnetic materials do not possess a net dipole moment of their own. Further, the effect of diamagnetism is a very weak effect, much weaker than the paramagnetic effect. It is present in all materials but in those materials which are paramagnetic, the much stronger paramagnetic effect is observed while the weaker diamagnetic effects are masked. Examples of materials which exhibit diamagnetism are bismuth, copper, gold, water, etc.

The third class of materials are possibly the most well known. These are iron, cobalt, nickel, and their compounds which exhibit a phenomenon called **ferromagnetism.** When an external magnetic field is applied to these ferromagnetic materials, they develop magnetic dipole moments like the paramagnetic materials but to a much larger extent. What is more interesting is that these substances, very often retain their magnetic moments even after the external magnetic field is withdrawn. The earliest magnetic material discovered by humans, namely lodestone, belongs to this category of materials. Permanent magnets of the kind we are familiar with also belong to this category of materials. These attract anything which has iron in it as is well known.

From what we have described as the characteristics of the different kinds of magnetic materials, and without knowing the details of how such a variety of magnetic effects can be produced at the atomic level, we can immediately infer a very interesting fact. We have seen that diamagnetic materials have their dipole moments aligned opposite to the applied field while paramagnetic materials get aligned in the direction of the field. Now assume we place both these kinds of materials in a non-uniform magnetic field as in Fig. 7.1 The field becomes weaker towards the right.

From the previous chapter, we know that a magnetic dipole experiences a force when placed in a non-uniform magnetic field.

$$\vec{F} = \vec{\nabla}(\vec{\mu} \cdot \vec{B}) \tag{7.1}$$

In the present case, the field is stronger towards the left of the figure and hence, the gradient of the magnetic field is towards the left. For a paramagnetic material, we have seen, that the magnetic dipole

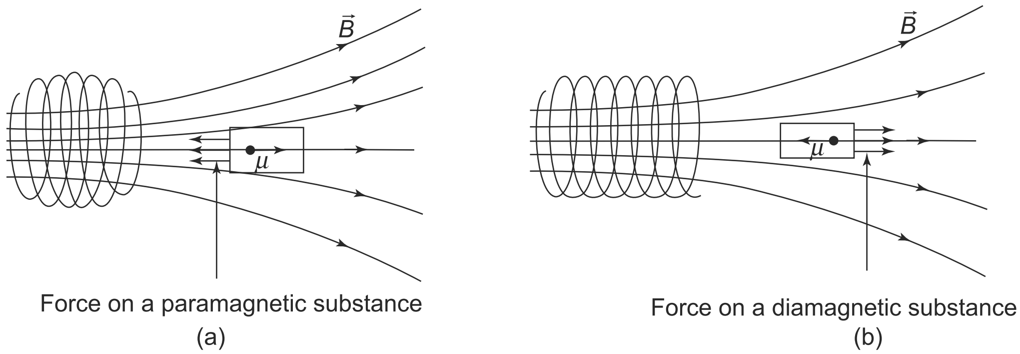


Fig. 7.1 *In (a) the sample is attracted towards the region of the stronger field while in (b) the sample experiences a force in the direction towards the region of the weaker field: (a) Forces on a paramagnetic material in a non-uniform field, (b) Forces on a diamagnetic material in a non-uniform field*

moment $\vec{\mu}$ is aligned to the direction of the magnetic field. Thus, the force experienced by such a material will be towards the direction of the gradient of $\vec{B}$. In our case, the paramagnetic material will thus move towards the left of the figure where the field is stronger.

For the diamagnetic material on the other hand, the magnetic dipole is aligned opposite to the direction of $\vec{B}$. Hence, the material will experience a force in the direction of the weaker magnetic field which in the case above is towards the right of the figure. Thus, if this magnetic field was created by a permanent magnet, a paramagnetic material will be attracted towards the magnet while a diamagnetic material will be repelled by such a magnet.

Recall that in the case of dielectrics, we could define a quantity called polarisation $\vec{P}$ which was a bulk property of the material. It is possible to define a similar quantity for magnetic materials and is called **Magnetisation** $\vec{M}$. The magnetisation vector $\vec{M}$ is just the magnetic dipole moment per unit volume of any material. If this is induced by a magnetic field $\vec{B}$, then it is found experimentally that for sufficiently weak magnetic fields, the magnetisation is proportional to the magnetic field.

By analogy with the corresponding phenomenon in electrostatics, one is tempted to write a direct proportionality relation between $\vec{M}$ and $\vec{B}$

$$\vec{M} = \frac{\chi_m}{\mu_0}\vec{B}$$

where χ_m is a constant. As we will see later in this chapter, this is not the standard way of writing the linear relationship between the two quantities. The magnetic field $\vec{B}$ is itself caused not only by external currents but due to magnetisation in the material itself also. As it stands, this equation above is valid only when $\chi_m \ll 1$. The complete relationship will appear later.

7.2 MAGNETISM AT THE ATOMIC LEVEL

We have seen that the various magnetic effects in materials are due to the interactions of electrons at the atomic level. The behaviour of electrons in an atom is governed by the laws of quantum mechanics. Quantum Theory has been successful in explaining almost all the observed phenomenon associated with atoms and molecules. However, for our purposes to understand the phenomena of magnetism at the atomic level, a much simpler model of the atom, namely Bohr's Model, is sufficient.

7.2.1 Understanding Paramagnetism

Recall that Bohr's model postulates that the negatively charged electrons go around a fixed, positively charged nucleus in well defined circular orbits. Negatively charged electrons moving in circular orbits will obviously constitute a current and thus, we can think of their motion as a current flowing in a loop. We have already seen in the previous chapter, that a current loop behaves like a magnetic dipole- both in terms of it creating a magnetic field and also in its interaction with an external magnetic field. Thus, for instance, if $\vec{\mu}$ is the magnetic dipole moment of the current loop generated by the motion of the electron around the nucleus, then, in an external magnetic field $\vec{B}$, it will experience a torque given by

$$\vec{\tau} = \vec{\mu} \times \vec{B} \tag{7.2}$$

This, we have seen, is exactly analogous to the electric dipole in an external electric field.

The torque experienced by the current loop (magnetic dipole) because of the external magnetic field has an effect on the energy of the electron. Since this torque depends on the angle between the direction of $\vec{\mu}$ and $\vec{B}$, we can see that it vanishes when $\vec{\mu}$ and $\vec{B}$ are parallel or antiparallel. In these cases, the torque is zero. It turns out that the electron has the minimum energy configuration when the magnetic moment is aligned with the external magnetic field and maximum when the moment is anti-aligned or opposite to the direction of $\vec{B}$. Since all systems in nature tend to lie in the lowest possible energy state if allowed, this implies that the atoms would be such that their magnetic moments are aligned with the magnetic field. This is the basic explanation, at the atomic level of the phenomenon of paramagnetism. Of course, any bulk material is a collection of atoms and so any explanation at the atomic level does not necessarily explain bulk properties. We will come back to this later in this chapter.

7.2.2 Understanding Diamagnetism

Another effect of the magnetic field on the atomic orbit is evident if we consider the fact that the result Eq. (7.2) for the torque assumes a rigid current loop with a constant current. In the case of atomic orbits this assumption is not strictly true. The electron in an atomic orbit is not strictly a rigid current loop and in fact, its velocity (and hence the current flowing in the 'loop') changes in the presence of an external magnetic field. To see this, consider for definiteness, an electron of mass m moving in an anticlockwise circular orbit of radius R with an angular frequency ω_0 in the absence of a magnetic field (Fig. 7.2(a)). Then, if the nucleus has a charge Ze, we will have

$$m\omega_0^2 R = k\frac{Ze^2}{R^2} \tag{7.3}$$

since the electrostatic attraction of the electron to the nucleus is the centripetal force responsible for the circular motion.

Consider now introducing a magnetic field in the z-direction. For this, note first that the process of introduction of a magnetic field is a time-dependent process. This is obvious since there is always a finite time Δt in which the field has to go from being zero to $\vec{B}_0$. During this interval of time, the magnetic

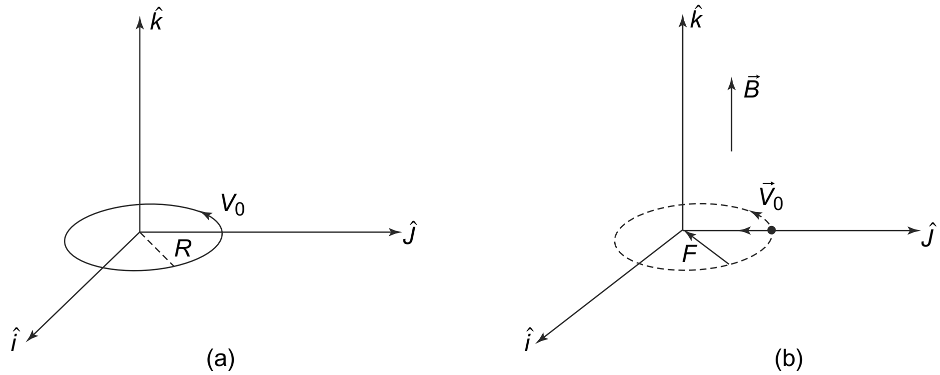


Fig. 7.2 *Electron charge in a circular orbit: (a) An electron of charge $-|e|$ in a circular orbit when $\vec{B} = 0$, (b) An electron of charge $-|e|$ in a circular orbit when $\vec{B} \neq 0$*

field is time dependent, $\vec{B}(t)$. As we will see later in the book, any time dependent magnetic field results in a changing magnetic flux associated with a circuit and this gives us an induced electromotive force in the circuit. This discovery by Faraday, of the generation of an induced electromotive force $\mathcal{E}$ (voltage) whenever the magnetic flux Φ associated with the circuit changes is now called Faraday's Law

$$\mathcal{E} = -\frac{d\Phi}{dt} \tag{7.4}$$

Here Φ is the magnetic flux across any surface defined as in the electric case as

$$\Phi = \iint \vec{B} \cdot d\vec{S}$$

The negative sign in Eq. (7.4) gives us the direction of the induced EMF $\mathcal{E}$. It is always such that the change in flux causing it is opposed by the magnetic flux caused by the induced EMF itself. This is called Lenz's law which we shall also come across in a later Chapter.

To study the effect of the induced EMF $\mathcal{E}$, consider a situation in which the orbit of the electron is in the $x - y$ plane and the magnetic field is in the z direction as shown in Fig. 7.2(b).

Initially, at time $t = 0$, there is no magnetic field and the electron moves in a counter clockwise direction in a circular orbit of radius R with a speed v_0 in the $x - y$ plane. Now consider a point P in the $x - y$ plane and located inside the circular loop. The motion of the electron in the loop, at location A causes a magnetic field at P which is given by

$$-\left(\frac{\mu_0}{4\pi}\right) e\vec{v}_0 \times (\overrightarrow{AP})$$

which is along the $-z$ direction. Thus, by Lenz's Law, the introduction of the external magnetic field should have the effect of increasing the magnitude of this magnetic field, that is by speeding up the electron during the time Δt, which is taken by the external magnetic field to go from $\vec{B} = 0$ to its final value $\vec{B} = \vec{B}$. For each orbit of the electron, will gain an energy of

$$-|e|\mathcal{E}$$

and so during a time interval Δt, it will gain energy ΔE given by

$$\Delta E = -|e|\mathcal{E}\frac{\Delta t}{T} \tag{7.5}$$

where where T is the period of revolution of the electron in the circular orbit. The force on the electron due to the magnetic field, which will be radial, in general can change both the speed of the electron as well as the radius. We make the assumption that the radius does not change and we will justify this very shortly. With this assumption,

$$-\mathcal{E}\Delta t = -\frac{\Delta\Phi}{\Delta t}\Delta t = B\pi R^2 = \frac{Bv_0RT}{2}$$

since the change in the flux is equal to $B \times$ Area of the loop and hence the change in energy is

$$\Delta E = \frac{|e|B\pi R^2}{T} \tag{7.6}$$

The speed would thus, change from v_0 to $v' = v_0 + \Delta v$ such that

$$\frac{m}{2}(v')^2 - \frac{m}{2}v_0^2 = \frac{|e|BRv_0}{2} \tag{7.7}$$

Now assuming that $\Delta v = v' - v_0$ is small, we can write

$$\Delta v = \frac{eRB}{2m} \tag{7.8}$$

Given the velocity change as given by Eq. (7.8), we will now justify the assumption of the radius of the orbit not changing, which we used in deriving it.

The magnetic field exerts a force

$$F = -|e|(\vec{v}' \times \vec{B})$$

directed towards the centre of the circular orbit, on the electron. Suppose we assume that this changed the radius to R_f. Then the forces on the electron moving in this new orbit give us

$$\frac{mv'^2}{R_f} = k\frac{Ze^2}{R_f^2} + |e|v'B \tag{7.9}$$

We can write

$$\frac{mv'^2}{R_f} = \frac{mv_0^2}{R_f} + \frac{2mv_0\Delta v}{R_f}$$

where once again we have used the fact the Δv is small. Now for Δv, we can use the expression in Eq. (7.8) by simply replacing R by R_f. This can be done in our approximation since any change in R will be first order in B and we are neglecting higher order terms. Thus, we get

$$\frac{mv_0^2}{R_f} = k\frac{Ze^2}{R_f^2} + |e|v'B - \frac{|e|v'R_fB}{R_f} = \frac{kZe^2}{R_f^2} \tag{7.10}$$

In the situation that we considered the speed of the electron increased. If the orbit was in the opposite direction, Eq. (7.8) would get a negative sign in the right hand side, i.e., the electron would have slowed down. Equation (7.10) shows that R_f satisfies the same equation as without the magnetic field and hence our assumption of the orbit's radius not changing stands justified. The net effect of the magnetic field on the atomic orbit is simply to either speed it up or slow it down, depending on the orientation of the magnetic field without changing the radius of the orbit.

The change in the speed of the electron in its orbit has an effect on the magnetic moment that is associated with it. Let the magnetic moment of the atom without the magnetic field be $\vec{\mu}$. Then

$$\vec{\mu} = -|e|\pi R^2 \frac{v_0}{2\pi R}\hat{k} = -|e|\frac{v_0 R}{2}$$

The magnetic moment is in the $-\hat{k}$ direction, as shown in Fig. 7.2. Now with the introduction of the magnetic field, the speed increases to $v_0 + \Delta v$ and hence, the magnetic moment changes by an amount

$$\Delta\vec{\mu} = -|e|\frac{\Delta v R}{2}\hat{k} = -\frac{e^2 R^2}{4m}\vec{B} \tag{7.11}$$

Notice the sign in Eq. (7.11). The change in the magnetic moment is in the opposite direction to $\vec{B}$. We had considered in the discussion above an orbit in the anticlockwise direction. However, the fact that $\Delta\vec{\mu}$ opposes $\vec{B}$ is not just an artifact of the fact that we have taken particular orientations of the orbital motion. Suppose the motion is in a clockwise direction. The orbit would now slow down and the magnetic moment would be along the $\hat{k}$ direction. Thus, $\Delta\vec{\mu}$ would be in the $\hat{k}$ direction but this time the value of $\Delta\vec{\mu}$ would be negative since the electron would have slowed down!

The whole explanation above is based on an orbit which is in a plane perpendicular to $\vec{B}$. However, a similar calculation can be made for the case when the orbit is in any general direction as in Fig. 7.3.

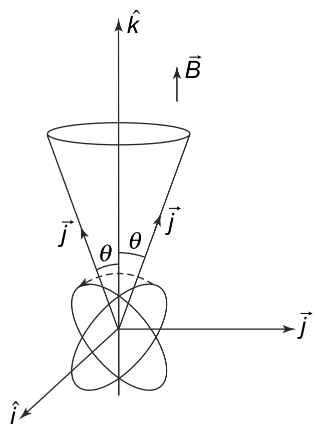


Fig. 7.3 *Orbit of an electron with angular momentum $\vec{J}$ at an angle θ with respect to the applied magnetic field $\vec{B}$. The orbit precesses around $\vec{B}$ with a fixed θ*

If we carry out this computation, the result for the change in the magnetic moment of the orbiting electron is very similar to Eq. (7.11) with R^2 now replaced by the average value of $(x^2 + y^2)$ of the orbit. However, there is an important difference in the motion of the electron. Unlike the orbit in Fig. 7.2, the orbit in this case precesses around the direction of $\vec{B}$ which we have taken to be along $\hat{k}$. This is very much like a rolling wheel which is not vertical when moving in a circular path on the ground.

We have thus seen that an external magnetic field has two distinct effects on the orbits of the atomic electrons. The first effect is that it gives rise to an interaction energy between the magnetic field and the magnetic moment caused by the orbiting electron. This interaction energy makes the orbits which are aligned with $\vec{B}$ more probable since these orientations have a lower energy. Secondly, the external magnetic field changes the magnetic moment of the orbiting electrons (magnetic moment caused by the current loop associated with the moving electrons in their orbits). This change in the magnetic moment is in a direction opposite to the external magnetic field.

These phenomena are universal and applicable to all materials. However, when paramagnetic effects are present, they usually dominate over the diamagnetic effects. Since experimentally one sees only the combination of these effects, such materials do not exhibit diamagnetism experimentally.

The picture above is what happens at the level of the individual atom. Bulk matter of course contains a huge number of atoms and/or molecules whose dynamics might be somewhat different. We now consider the atomic level effects in bulk matter.

7.2.3 Bohr Magneton and Gyromagnetic Ratio *

In quantum mechanics, the natural unit of expressing the magnetic dipole moment is called the Bohr magneton. In SI units, it is defined by

$$\mu_B = \frac{eh}{4\pi m_e}$$

where e is the electronic charge, m_e is the mass of the electron and h is the Planck's constant. It turns out that the magnitude of the orbital magnetic dipole moments for atoms is in the range of a few Bohr magnetons while the spin magnetic moment of an electron is exactly 1 Bohr magneton. The value of one Bohr magneton in SI units is 9.27×10^{-24} JT^{-1}.

Another important quantity is the gyromagnetic ratio. This is defined as the ratio of the magnetic dipole moment to the orbital angular momentum. This quantity can be defined classically. Consider a charged ring rotating about one of its symmetry axis. The ring carries a charge q and has a radius r. The magnetic moment is simply

$$\mu = IA = \frac{qv}{2\pi r}\pi r^2 = \frac{q}{2m}mvr = \frac{q}{2m}L$$

where L is the angular momentum i.e. $L = mvr$. Thus, in this case we find that the gyromagnetic ratio is simply $\frac{q}{2m}$.

The quantity gyromagnetic ratio plays a very important role in physics. An isolated electron, that is, one that is not in orbit, would classically be expected to have no angular momentum. However, quantum mechanics tells us that an electron has an intrinsic angular momentum because of electron spin which is a purely quantum phenomenon. The gyromagnetic ratio due to this spin angular momentum is usually written as

$$\gamma_e = -\frac{e}{2m_e}g_e$$

where g_e is the correction due to relativity and quantum mechanics. It turns out that

$$g_e = 2\left(1 + \frac{\alpha}{2\pi} + \cdots\right)$$

where $\alpha = \frac{1}{137}$, the fine structure constant. An accurate measurement of g_e and its concurrance with the theoretically predicted value is one of the greatest triumphs of quantum physics.

EXAMPLE 7.1 Under conditions of maximum magnetisation, the dipole moment per unit volume of cobalt is 1.5×10^5 Am^{-1}. Assuming that this magnetisation is due to completely aligned electrons, how many such electrons are there per unit volume? How many aligned electrons are there per atom? The density of cobalt is 8.9×10^3 kgm^{-3} and the atomic mass is 58.9 g/mole.

Solution

The dipole moment of 1 m^3 of cobalt is equal to 1.5×10^5 Am^2. Each aligned electron contributes a dipole moment equal to 1 Bohr magneton and hence, each aligned atom contributes, 9.27×10^{-24} Am^2. The number of aligned electrons is thus, equal to

$$N = \frac{1.5 \times 10^5}{9.27 \times 10^{-24}} = 1.6 \times 10^{28}$$

The number of atoms in 1 m^3 of cobalt can be determined from the atomic mass and density. We get the number of atoms per cubic meter to be

$$N_{\text{atoms}} = \frac{8.9 \times 10^3}{58.9 \times 10^{-3}} N_A$$

where N_A is the Avogadro's number. Thus, we see that

$$N_{\text{atoms}} = 9.1 \times 10^{28}$$

and so there are only 0.18 aligned electrons per atom.

7.3 MAGNETISATION AS BOUND CURRENTS

We have seen above that at the atomic level, an external magnetic field affects matter. This is the case both when the substance has some permanent magnetic dipole moment (paramagnetic substances) as well as for diamagnetic substances where the atomic magnetic dipole moment is created by the external field which then aligns in a direction opposite to the applied field. An interesting fact is that the magnetic moment created for diamagnetic substances or the change in the magnetic moment for paramagnetic substances are both proportional to the applied field.

In bulk matter, just as in the case of dielectrics in electrostatics, we expect that this interaction with the atomic magnetic dipole moments will create an average magnetisation or a magnetic moment density. It is reasonable to expect this to be

$$\vec{M} = \sum_i N_i < \vec{\mu}_i > \tag{7.12}$$

where N_i is the average number of i^{th} type of molecules per unit volume each of which carries a magnetic dipole moment $\vec{\mu}_i$. Just as in the case of the polarisation $\vec{P}$ in the electrostatic case, the magnetisation $\vec{M}$ will create its own magnetic field at every point. In a volume element d^3r', the magnetic moment is

$$\vec{M}(\vec{r}')d^3r'$$

A physical picture of the magnetisation discussed in the last section can be presented in terms of currents in the material induced by the external magnetic field. As we have discussed, at the atomic and molecular level, an external magnetic field creates tiny loops of currents. In materials where such current loops are already present, the magnetic field changes the value of the magnetic moment associated with the loops. In bulk matter, these atomic loops belonging to neighouring atoms or molecules will overlap as shown in Fig. 7.4.

If the magnetisation is uniform, which will happen if the external magnetic field is uniform, the currents in the individual tiny loops will all be the same with the same sense. That will mean that at any point, the current from the loop on one side will be exactly equal but opposite in direction from the loop on the other side. The net result in the bulk at any point will be no net current. That will be so at all points except at the boundary where there will be atoms or molecules on one side only. At the boundary thus the currents in the tiny atomic current loops will add up to form a net current on the surface of the sample. If the magnetisation is not uniform, then of course the cancellation between the loop currents of neighbouring molecules in the bulk will not be total. We will thus, have a net current in the bulk as well.

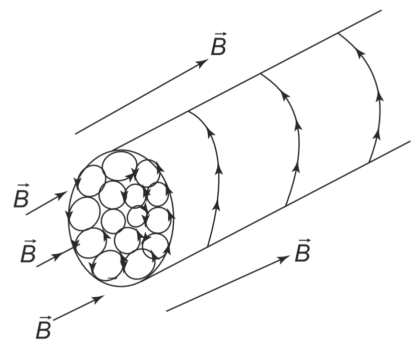


Fig. 7.4 *Atomic loops in bulk matter*

The physical picture presented above can be realised mathematically as follows. The magnetisation contribution to the vector potential can be written as

$$\vec{A}_M(\vec{r}) = \left(\frac{\mu_0}{4\pi}\right)\int \frac{\vec{M}(\vec{r}') \times (\vec{r}-\vec{r}')}{|\vec{r}-\vec{r}'|^3} d^3r'$$

This we have already seen to be equal to

$$\vec{A}_M(\vec{r}) = \left(\frac{\mu_0}{4\pi}\right)\int d^3r'\vec{M}(\vec{r}') \times \vec{\nabla}'\left(\frac{1}{|\vec{r}-\vec{r}'|}\right)$$

Partial integration gives us

$$\vec{A}_M(\vec{r}) = \left(\frac{\mu_0}{4\pi}\right)\int d^3r'\left[\frac{\vec{\nabla}' \times \vec{M}(\vec{r}')}{|\vec{r}-\vec{r}'|} - \vec{\nabla}' \times \left(\frac{\vec{M}(\vec{r}')}{|\vec{r}-\vec{r}'|}\right)\right] \tag{7.13}$$

Thus, we see that the contribution to the vector potential from the magnetisation is the sum of two terms.

We can now use Gauss's Divergence Theorem to convert one of the volume integrals into a surface integral, i.e.,

$$\int d^3r'\vec{\nabla}' \times \left(\frac{\vec{M}(\vec{r}')}{|\vec{r}-\vec{r}'|}\right) = -\iint \left(\frac{\vec{M}(\vec{r}')}{|\vec{r}-\vec{r}'|}\right) \times d\vec{S}' \tag{7.14}$$

To see this, recall that any vector can be written as

$$\vec{V} = \sum_i (\vec{V} \cdot \hat{n}_i)\hat{n}_i$$

where $\hat{n}_i$ is the unit vector in the i^{th} direction, i taking values $1, 2, 3$. With this, consider that i^{th} component of

$$\int d^3r'\vec{\nabla}' \times \left(\frac{\vec{M}(\vec{r}')}{|\vec{r} - \vec{r}'|}\right)$$

This gives us

$$\begin{aligned}\left[\int d^3r'\vec{\nabla}' \times \left(\frac{\vec{M}(\vec{r}')}{|\vec{r} - \vec{r}'|}\right)\right]_i &= \int d^3r'\vec{\nabla}' \cdot \left(\left(\frac{\vec{M}(\vec{r}')}{|\vec{r} - \vec{r}'|}\right) \times \hat{n}_i\right) \\ &= \iint d\vec{S}' \cdot \left(\left(\frac{\vec{M}(\vec{r}')}{|\vec{r} - \vec{r}'|}\right) \times \hat{n}_i\right) \\ &= -\iint \left(\left(\frac{\vec{M}(\vec{r}')}{|\vec{r} - \vec{r}'|}\right) \times d\vec{S}'\right) \cdot \hat{n}_i \end{aligned} \tag{7.15}$$

Thus, we have

$$\vec{A}_m(\vec{r}) = \left(\frac{\mu_0}{4\pi}\right) \int d^3r' \left[\frac{\vec{\nabla}' \times \vec{M}(\vec{r}')}{|\vec{r} - \vec{r}'|}\right] + \left(\frac{\mu_0}{4\pi}\right) \iint \left(\frac{\vec{M}(\vec{r}') \times \hat{n}}{|\vec{r} - \vec{r}'|}\right) dS' \tag{7.16}$$

We can write this as

$$\vec{A} = \left(\frac{\mu_0}{4\pi}\right) \int d^3r' \frac{\vec{j}_{bV}}{|\vec{r} - \vec{r}'|} + \left(\frac{\mu_0}{4\pi}\right) \int \frac{\vec{j}_{bS}}{|\vec{r} - \vec{r}'|} dS' \tag{7.17}$$

where

$$\boxed{\vec{j}_{bV} = \vec{\nabla} \times \vec{M}}$$

is the bound volume current density and

$$\boxed{\vec{j}_{bS} = \vec{M} \times \hat{n}_S}$$

is the bound surface current density. Here $\hat{n}_S$ is the unit vector such that $d\vec{S} = \hat{n}_S dS$. Notice that in case the magnetisation is uniform, the bound volume current density vanishes and the only effect is due to the bound surface current density above.

The picture that emerges is thus as follows—a magnetic material, apart from the free volume and surface current density ($\vec{j}_{fV}$ and $\vec{j}_{fS}$ respectively), would also have in general, a bound volume and bound surface current density. Thus, the total volume density is

$$\vec{j}_V = \vec{j}_{fV} + \vec{j}_{bV}$$

and the total surface current density is

$$\vec{j}_S = \vec{j}_{fS} + \vec{j}_{bS}$$

PROBLEM 7.1 An magnetised sphere of radius R has its centre at the origin. It has a non-uniform magnetisation $\vec{M}(x, y, z)$ given by $\mu_0\vec{M}(x, y, z) = \left(7\left(\frac{y}{R}\right)^2 + \left(\frac{x}{R}\right)^2\right)\hat{i}$ T. Determine the volume and surface magnetisation densities and the bound currents.

7.4 MAGNETIC FIELDS $\vec{B}$ AND $\vec{H}$ IN A MEDIUM

We saw above that the effect of magnetisation is basically to establish two kinds of bound currents in the magnetic material—a volume-bound current $\vec{j}_{bV}$ and a surface-bound current $\vec{j}_{bS}$. These were related to the magnetisation by

$$\vec{j}_{bV} = \vec{\nabla} \times \vec{M}$$

and

$$\vec{j}_{bS} = \vec{M} \times \hat{n}_S$$

Now suppose the magnetic material, in addition to these bound currents also has a free current density j_f. Thus, the total current density can be written as

$$\vec{j} = \vec{j}_f + \vec{j}_b \tag{7.18}$$

With this, Ampere's Law that we encountered in the previous chapter thus, becomes

$$\vec{\nabla} \times \vec{B} = \mu_0\vec{j} = \mu_0\vec{j}_f + \mu_0(\vec{\nabla} \times \vec{M}) \tag{7.19}$$

This can be conveniently written as

$$\vec{\nabla} \times \left(\frac{1}{\mu_0}\vec{B} - \vec{M}\right) = \vec{j}_f \tag{7.20}$$

We define a quantity $\vec{H}$ as

$$\vec{H} = \frac{1}{\mu_0}\vec{B} - \vec{M} \tag{7.21}$$

With this, Ampere's Law as applicable for magnetic materials reads

$$\boxed{\vec{\nabla} \times \vec{H} = j_f} \tag{7.22}$$

Note that for the case of vacuum, i.e., with $\vec{M} = 0$, $\vec{B} = \mu_0\vec{H}$, we get the familiar form of Ampere's Law. The important thing about Eq. (7.22) is that it relates the magnetic field $\vec{H}$ to the **free** current density $\vec{j}_f$. This is important since the free current density is the only thing that we can control. Hence, in the design of systems to produce a magnetic field, we use this relationship to determine $\vec{H}$.

We thus, have two magnetic fields : $\vec{B}$ which is the magnetic field we had earlier introduced to describe magnetic forces on currents and which satisfies Ampere's Law and now a new field $\vec{H}$. The latter is identical to $\vec{B}$ in free space but in the presence of matter includes magnetisation effects. This pair of fields is very similar to the pair ($\vec{E}$, $\vec{D}$) for the corresponding electric case. $\vec{B}$ is similar to $\vec{E}$ while $\vec{H}$ is similar to $\vec{D}$. There is however, no universally accepted nomenclature of the two fields, unlike their electric counterparts. We shall continue to refer to $\vec{B}$ as the magnetic field and refer to $\vec{H}$ as the magnetic field $\vec{H}$ with a clear understanding that the latter takes into account magnetisation effects just as $\vec{D}$ in electrostatics takes into account polarisation effects.

PROBLEM 7.2 A very long cylindrical pipe made of material of permeability μ has an inner radius R and thickness t. It encloses completely an equally long solid metallic cylinder of radius $r < R$, of permeability μ_1 and which carries a current J_f parallel to its axis uniformly along its cross section. Calculate the values of the fields $\vec{B}, \vec{H}$ and the magnetisation $\vec{M}$ inside the inner cylinder, between the cylinders, inside and outside the outer cylinder.

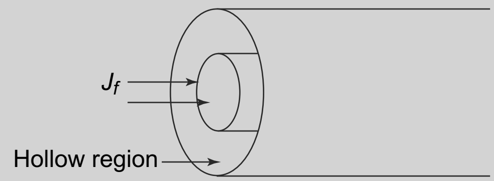


Fig. 7.5 *Problem 7.2*

7.5 DIAMAGNETISM AND PARAMAGNETISM IN BULK MATTER

As we have seen above, individual atoms or molecules, if they have a permanent magnetic dipole moment, will align themselves with the magnetic field direction since that has a lower energy than any other configuration. Bulk matter of course, contains many atoms and/or molecules. It is legitimate to expect that all the molecules which have a permanent dipole moment will align with an external magnetic field since this will be the lowest energy state individually and hence, for the whole sample of bulk matter. However, we have already seen in the case of dielectrics that in a collection of atomic electric dipoles, not all the dipoles can be in the state of minimum energy. This is because at any finite temperature there are always thermal fluctuations which are responsible for this energy spread. At any given temperature, statistical mechanics tells us that the probability of a given molecule having an energy W is proportional to the Boltzman factor

$$P(W) = \exp(-\frac{W}{kT}) \tag{7.23}$$

where T is the temperature of the assembly of molecules or atoms. Here k is the Boltzman constant.

In the case of electric polarisation, the potential energy of an atomic dipole moment $\vec{p_0}$ in an electric field $\vec{E}$ is $U = -\vec{p_0} \cdot \vec{E}$. Similarly, in the magnetic case, the potential energy of a dipole moment $\vec{\mu}$ in a magnetic field $\vec{B}$ is $U = -\vec{\mu} \cdot \vec{B}$. The analysis that we presented in Chapter 3 for the dielectrics in an electric field can thus, be carried over for the magnetic case with the replacement $\vec{p_0} \to \vec{\mu}$. The effective magnetic field inside the material is H as we have seen above.

In the discussion given in Section 3.7.3 of Chapter 3, we have already discussed the case of the electric dipoles. We could relate the polarisation vector $\vec{P}$ to the atomic dipole moment $\vec{p}$ and the external electric field $\vec{E}$ as

$$\vec{P} = \frac{Np^2\vec{E}}{3kT}$$

where N is the number density of the molecules and T is the temperature.

Proceeding in exactly the same way, with the magnetic field instead of the electric field and the magnetic dipole we can get the expression for the magnetisation (which is simply the total magnetic dipole moment per unit volume) $\vec{M}$ for a bulk material at temperature T

$$\boxed{\vec{M} = \frac{N\mu^2}{3kT}\vec{H}} \tag{7.24}$$

Following the analogy with electrostatics, we can define the susceptibility χ_m as

$$\vec{M} = \chi_m \vec{H}$$

and thus,

$$\chi_m = \frac{N\mu^2}{3kT} \tag{7.25}$$

Equation (7.24) is called Curie's Law and was first derived by Langevin. According to Langevin's Theory, the magnetic susceptibility χ_m thus, is positive for molecules having a permanent magnetic moment, i.e., paramagnetic materials. The susceptibility χ_m is dimensionless and its values are shown in Table 7.1. We see that strongly paramagnetic substances have values of $\chi_m \sim 10^{-4}$ or even greater.

Table 7.1 *Magnetic susceptibility (χ_m) of materials*

Material	χ_m
Bismuth	-1.66×10^{-4}
Helium	-9.85×10^{-10}
Oxygen	3.73×10^{-7}
Aluminum	2.2×10^{-5}
Nitrogen	-4.3×10^{-7}

Let us now see what happens for a diamagnetic substance. The expression for the induced magnetic moment Eq. (7.11) for a molecule will lead, by the above argument, to negative values of magnetic susceptibility for bulk matter also. If the orientation of the molecules in bulk matter is random, which is what we expect at any temperature, then the average value of $(x^2 + y^2)$ for a molecule is $\frac{2}{3}R^2$ where R is its radius. Thus, if there are N molecules per unit volume, Equation (7.11) with R^2 replaced by $\frac{2}{3}R^2$ will give us a susceptibility for **diamagnetic materials** as

$$\boxed{\chi_m = -\frac{Ne^2}{6m}\sum_i R_i^2} \tag{7.26}$$

where the summation is over all electrons in the molecule.

Two things should be noted about the expression for the susceptibility of diamagnetic materials, Eq. (7.26). Firstly, it is universal and does not depend on whether the molecules have any permanent magnetic moment or not. Secondly, the susceptibilities, unlike the paramagnetic case (Eq. (7.25), are independent of temperature.

Thus we have the following situation- all substances are diamagnetic and hence have susceptibilities given by Eq. (7.26). However, materials which also possess a permanent magnetic moment, i.e.

paramagnetic materials will in addition have paramagnetic contribution to the susceptibility. Thus the total magnetic susceptibility of a paramagnetic material will be the sum of the two contributions. It turns out, that for strongly paramagnetic materials, the susceptibility due to the paramagnetic contribution is more than two orders of magnitude more than the much weaker diamagnetic susceptibility. Hence, in materials with paramagnetic effects, it is almost impossible to measure the diamagnetic contribution experimentally.

Another important point is about the validity of Langevin theory. Langevin theory is based on a classical as opposed to a quantum mechanical understanding of the atoms. Classically, the parameters entering the expressions, Eq. (7.26) and Eq. (7.25) cannot be determined. However, the simple Langevin theory given here does help explain some of the major features which are observed experimentally. For instance, diamagnetic substances have susceptibilities which are independent of temperature is borne out experimentally by a large class of diamagnetic materials. Further, the temperature dependence of paramagnetic susceptibility is another feature which is observed experimentally.

Classically, the picture presented above is actually not quite correct. This was first shown by van Leeuwen in 1919 though Bohr had also pointed it out earlier. The result is formalised in a theorem which goes by the name of Bohr-van Leeuwen Theorem. This theorem states that classical statistical mechanics actually forbids the phenomenon of paramagnetism and diamagnetism and we actually need relativity and quantum mechanics to explain the existence of these magnetic phenomenon. In other words, the theorem implies that if one applies classical statistical mechanics properly to materials in bulk, then the magnetic susceptibility of both paramagnetic and diamagnetic materials is necessarily zero! Basically, this result arises out of the fact that a magnetic field acting on a charged particle through the Lorentz force does not change its energy. In equilibrium, the distribution of particles in the different momentum states is given solely by the Boltzman factor $exp(-E/kT)$ and hence, this is not affected by the magnetic field. The average value of currents in an assembly of particles in equilibrium is not affected by the magnetic field. Since currents cause magnetic moments, the average magnetic properties of an assembly of charged particles are therefore not expected to change when a magnetic field is present.

We had earlier remarked that the Langevin theory, though not quite correct, does explain some of the experimentally observed phenomenon. What is the reason for this partial success then of the Langevin theory of magnetic susceptibilities? The reason actually lies in the assumptions that are made in Langevin theory. For instance, for paramagnetic substances, Langevin theory assumes that all molecules in the bulk material have the same 'permanent' magnetic dipole moment. Earlier in the chapter, we have already seen that the orbiting electron gives rise to a magnetic dipole moment which is proportional to the angular momentum. Since Langevin theory assumes that all molecules have the same magnetic moment, it would imply that there is only one value of the moment and hence, the angular momentum that is possible. In classical statistical mechanics this is not possible since angular momentum, like linear momentum can take all possible values.

In the case of diamagnetic materials, the dipole moment is proportional to the radius. Once again, classically, this can take all possible values. van Leeuwen's theorem corrects these assumptions and applies classical statistical mechanics in the correct fashion. On doing this, as we remarked earlier, the theorem predicts that the phenomenon of paramagnetism and diamagnetism should not exist within the domain of classical physics.

Of course, we also know that classical statistical mechanics has a limited domain of validity and one should really apply quantum mechanics to atomic systems. On doing this, it is found that we get the correct theory. We shall not be discussing the quantum theory of para and diamagnetism in this book. The basic difference in applying quantum mechanics is that in quantum mechanics, angular momentum cannot have continuous values but instead is quantised. When we average over all values of angular momentum in an assembly of molecules, like the one which exists in bulk matter, the result of this averaging is very different from that obtained using classical statistical mechanics. The averaged angular momentum then can be used to find the magnetic susceptibilities and one indeed finds non-zero, finite values as predicted by experiment.

The fact that for both para- and diamagnetic materials, the magnetisation is proportional to the applied magnetic field and the constant of proportionality is the magnetic susceptibility, also allows us to define another quantity called the relative permeability, μ_r. Customarily, the relationship between $\vec{H}$ and $\vec{M}$ is written as

$$\vec{M} = \chi_m \vec{H} \tag{7.27}$$

where χ_m is the magnetic susceptibility. This linear relationship between $\vec{H}$ and $\vec{M}$ is obeyed by all paramagnetic and diamagnetic materials, though χ_m is positive for paramagnetic materials and negative for diamagnetic materials. But we also have a relationship between $\vec{B}$, $\vec{M}$ and $\vec{H}$.

$$\vec{H} = \frac{\vec{B}}{\mu_0} - \vec{M}$$

Thus, combining this, we get

$$\vec{B} = \mu_0(\vec{H} + \vec{M}) = \mu_0(1 + \chi_m)\vec{H} = \mu\vec{H} \tag{7.28}$$

where

$$\mu = \mu_0(1 + \chi_m)$$

is called the permeability of the material. We can also define a relative permeability,

$$\mu_r \equiv \frac{\mu}{\mu_0} = 1 + \chi_m$$

In the absence of any matter, we have $\mu = \mu_0$ which is called the permeability of free space.

7.6 FERROMAGNETISM

The oldest known form of magnetic materials are ferromagnetic materials. The lodestone discovered in ancient times is an example of a ferromagnetic material. The permanent magnets that we know of (of the kind used in magnetic compasses, for instance) are made from ferromagnetic materials. Ferromagnetism as a phenomenon is qualitatively different from the magnetic phenomena that we have discussed so far, namely para and diamagnetism.

Recall that though the explanation for para- and diamagnetism was radically different (in one case, it being caused by the magnetic moments induced by the external magnetic field and in the other, due to the inherent magnetic moment of the atoms); in both cases, the magnetisation $\vec{M}$ was proportional to the external magnetic field. This proportionality constant was of course the magnetic susceptibility χ_m. This is not the case with ferromagnetic materials at all. Ferromagnetic materials have a built-in magnetisation even in the absence of an external magnetic field. This magnetisation of course, changes when an external magnetic field is introduced. In addition, the magnetisation in ferromagnetic materials is much larger in magnitude than in para- and diamagnetic materials.

There is another interesting property of ferromagnetic materials which differs from the para- or diamagnetic materials. The magnetisation in ferromagnetic materials is not linear with the applied external magnetic field. In fact, it increases till a certain value of the external magnetic field and then reaches a saturation value, beyond which it does not change.

Two other properties of ferromagnetic materials stand out—one, the magnetisation of these materials disappears when the material is heated beyond a certain temperature called the **Curie temperature.** The Curie temperature is different for different ferromagnetic materials, but exists in all of them. Below the Curie point or Curie temperature, the ferromagnetic material is ferromagnetic and above it, it is paramagnetic.

Another property of ferromagnetism is that its magnetisation is history-dependent. This is illustrated in Fig. 7.6.

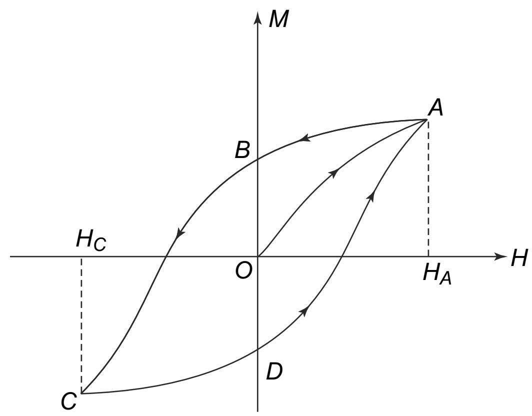


Fig. 7.6 *Hysteresis loop*

In the figure, we plot the magnetisation M against H, the applied external magnetic field. One starts at the point O where both M and H vanish. Then as H is increased, the magnetisation follows a path OA till it reaches the saturation value of M at point A where the value of the applied field is H_A. Now if the field is decreased, instead of going back along the path AO, the magnetisation follows the path ABC to reach the saturation point at C where the magnetic field is H_C. Now when the field H is increased, the magnetisation follows the curve CDA. This behaviour continues as the field goes between H_A and H_C. This loop is called the **Hysteresis loop.** The word hysteresis means 'to lag' in Greek. Thus, for ferromagnetic materials $\vec{M}$ is not proportional to $\vec{H}$ and we cannot in general, write $\vec{M} = \chi_m \vec{H}$ except for very weak $\vec{H}$ in a sample without permanent magnetisation. For situations where $\vec{M}$ and $\vec{H}$ are parallel, one normally defines an incremental permeability $\mu(\vec{H})$ in general, as

$$\mu(\vec{H}) = \mu_0(1 + \chi_m(\vec{H})) = \frac{dB}{dH}$$

For usual ferromagnetic materials χ_m can be very high, sometimes of the order of 10^6, so that the magnetisation term $\vec{M}$ dominates the $\vec{B} = \mu_0(\vec{H} + \vec{M})$.

Since magnetic moments have potential energy in a magnetic field, the process of magnetisation costs energy which is supplied by the agent (a battery or any other current source) supplying the current to produce the magnetic field. If the magnetisation is reversible then of course, along a cycle the total energy supplied would vanish. In the case of the hysteresis cycle, this is not true. Hence, if a piece of ferromagnetic material is subject to a magnetic field that keeps on changing between a positive and a negative value with the material developing a hysteresis loop, then of course, some energy will be spent during each cycle which will appear as heat. We shall discuss this more quantitatively later in Chapter 8.

The remarkable thing about the hysteresis loop is that if the external field is withdrawn at any point, there is still some residual magnetisation left in the ferromagnetic sample. This is indicated by the points B and D in the Fig. 7.6. The residual magnetisation has either a positive or a negative sign (indicating two opposite directions) depending on whether the field was withdrawn or became $H = 0$ in one half (AB) or the other half (CD) of the hysteresis loop.

Ferromagnetic materials are mostly used either to increase the magnetic flux of a current circuit or as a source of magnetic field (in the form of permanent magnets). In case we wish to make a permanent magnet, the hysteresis loop allows us to determine how it can be done. We can place the ferromagnetic material in an external magnetic field and increase the magnetic field till the magnetisation reaches the saturation value. Now when the magnetic field is withdrawn or taken to zero, some residual magnetisation will remain in the ferromagnetic material.

Thus we see that ferromagnetism is a qualitatively different phenomenon than either para- or diamagnetism. It turns out that there is no satisfactory classical theory of explaining the origin of ferromagnetism at the atomic or molecular level. The quantum theory of ferromagnetism is fairly complicated and we won't discuss it here. However, a simplified version of the same can be used to explain some of the properties that we have seen above of ferromagnetic materials.

Atoms have electrons, which, as we know, apart from having an orbital angular momentum, also have an intrinsic quality called the spin. The electron spins of various atoms interact in a ferromagnet through a complicated interaction known as exchange interaction. The net result of this interaction is that it is energetically favourable for the spins of neighbouring atoms to align in the same direction. A region in the sample (typically containing many thousands of atoms) has all the spins aligned in one direction. Such a region is called a **magnetic** or **Weiss domain.** The domains thus, have a sizeable magnetic moment (from the spins) even in the absence of an external magnetic field. A typical macroscopic sample of a ferromagnetic material would consist of thousands of such domains. The direction or orientation of the magnetic moment of each domain would be random and hence, the net magnetic moment (magnetisation) of the whole sample would average to zero.

The existence of magnetic domains can explain some of the properties of ferromagnetic materials that we saw above. Thus, for example, when an external magnetic field $\vec{H}$ is applied to a ferromagnetic sample (whose net magnetisation is zero), the magnetic domains, with magnetic moment $\vec{\mu}_m$ would acquire a potential energy given by

$$E_{\text{P.E}} = -\vec{\mu}_m \cdot \vec{H}$$

Thus, it will be energetically favourable for the magnetic domains to align with the magnetic field. Of course, at any finite temperature, the thermal motion would not allow all the domains to all align with the magnetic field. But statistically, at any temperature, more domains would be aligned in the direction of the applied field than would be if there was no field. We will thus observe a non-zero magnetisation with an external field. Now as the field is increased, more and more domains would align with the field, thus resulting in an increase of M as H increases. At some point, all the domains would be aligned to the external field. After this, any increase in the magnitude of $\vec{H}$ would not result in any increase in the magnetisation since almost all the domains are already aligned to the magnetic field. This would then result in the saturation that we observed in the Hysteresis loop.

The explanation of other properties of ferromagnets, like the hysteresis loop and the presence of Curie temperature cannot be explained in on the basis of simple physical models like the one we have just considered. Curie temperature, for instance is an example of a 'phase transition', much like the transition of a liquid to a vapour when heated. It has certain unique characteristics like a discontinuity at the phase transition in the values of certain physical parameters. A discussion of a detailed theory of ferromagnetism is beyond the scope of this book.

In the example above for a permanent magnet, there were no free currents and of course, the linear relationship between $\vec{B}$ and $\vec{H}$, Eq. (7.27), is not valid. However, in cases where a free current exists that creates a magnetic field and ferromagnetic material obeying equation Eq. (7.27) is present, its presence affects the magnetic field everywhere substantially. This kind of situation is there typically in a solenoidal or toroidal current carrying coil with an iron core whose value of χ_m is 10^4 or even higher, resulting in increase of the magnetic flux through the coil by orders of magnitude. We analyse such situations in a subsequent section.

PROBLEM 7.3 In a ferromagnetic material like iron, the number of atoms in one cubic meter is approximately 8×10^{28} and the saturation value of the magnetic field $|\vec{B}|$ is around 2 Tesla. Estimate the magnetic moment of an iron atom.

7.7 ZEEMAN EFFECT

We know that elements, under certain conditions emit light of specific frequencies. For instance, when sodium is heated, two characteristic frequencies are emitted at wavelengths $\lambda_1 = 5890$ Å and $\lambda_2 = 5896^\circ$A. Around the turn of the twentieth century, a Dutch physicist, Pieter Zeeman noticed that when elements are subjected to an external magnetic field, the characteristic frequencies emitted by them get split into two or more frequencies. In other words, the specific spectral lines emitted by a particular element (i.e., light at a specific fixed wavelength) gets split into two lines (or more) or two (or more) different wavelengths. What is more, this splitting is proportional to the strength of the magnetic field applied. This phenomenon is called **Zeeman Effect**.

Soon after the observation of this splitting of spectral lines by Zeeman, H. Lorentz gave a theory based on classical physics to explain the observed phenomenon. In classical electrodynamics, as we shall see later in this book, a charged particle undergoing oscillations at some frequency will emit radiation at that frequency. Consider an electron in an orbit around the nucleus with a frequency of oscillation ω_0. By classical electrodynamics, it will emit radiation of frequency ω_0 which is the characteristic frequency of that atom. The equation of motion would simply be

$$m\frac{d^2\vec{r}}{dt^2} = -m\omega_0^2\vec{r} \tag{7.29}$$

Now suppose we introduce an external magnetic field $\vec{B}$ in the z-direction, i.e., $\vec{B} = B\hat{k}$. The electron will experience an additional force which is given by the Lorentz force in Chapter 6. The modified equation of motion will be, in component form

$$\begin{aligned} m\frac{d^2z}{dt^2} &= -m\omega_0^2 z \\ m\frac{d^2x}{dt^2} &= -m\omega_0^2 x - eB\dot{y} \\ m\frac{d^2y}{dt^2} &= -m\omega_0^2 y + eB\dot{x} \end{aligned} \tag{7.30}$$

We can define

$$X_\pm \equiv x \pm iy$$

and thus, in terms of $X_\pm$, the x and y equations can be written as

$$m\frac{d^2X_\pm}{dt^2} = -m\omega_0^2 X_\pm \pm ieB\dot{X}_\pm \tag{7.31}$$

The solutions to these equations can be written down. The z equation is easy to solve and we get

$$z(t) = z_0 \sin(\omega_0 t + \delta) \tag{7.32}$$

Similarly the $X_\pm$ equation can be solved to get

$$X_\pm(t) = X_\pm^0 \exp(i\omega_\pm t) \tag{7.33}$$

where

$$\begin{aligned} \omega_\pm &= \left(\omega_0^2 \pm \frac{eB}{m}\omega_\pm\right)^{1/2} \\ &\simeq \ \omega_0 \pm \frac{eB}{2m} \end{aligned} \tag{7.34}$$

where we have assumed that the second term is a small correction.

The oscillations in the z direction are at the original characteristic frequency ω_0. The solution for $X_\pm$ represents circular motion. Writing

$$X_\pm^0 = R_\pm \exp(i\delta_\pm)$$

with $R_\pm$ real, we get

$$X_\pm = R_\pm \exp(i\omega_\pm t + i\delta_\pm) \tag{7.35}$$

Hence, we have two kinds of oscillations—oscillations with frequency ω_+

$$\begin{aligned} x(t) &= R_+ \cos(\omega_+ t + \delta_+) \\ y(t) &= R_+ \sin(\omega_+ t + \delta_+) \end{aligned} \tag{7.36}$$

which represent anticlockwise circular motion in the $x - y$ plane at a frequency ω_+. In addition, there is a clockwise motion at a frequency ω_- given by

$$\begin{aligned} x(t) &= R_- \cos(\omega_- t + \delta_-) \\ y(t) &= -R_- \sin(\omega_- t + \delta_-) \end{aligned} \tag{7.37}$$

We thus, have three oscillations—an oscillation in the z direction at the unchanged characteristic frequency ω_0, and two oscillations at frequencies $\omega_{\pm} = \omega_0 \pm \frac{eB}{2m}$. Thus, in classical electrodynamics, the original line will now be split into three lines in the presence of an external magnetic field $\vec{B} = B\hat{k}$—the original line at the same frequency ω_0 and two lines on either side of the original line, separated from the original line by an amount $\Delta\omega = \frac{eB}{2m}$.

This simple classical theory of Lorentz then seems to explain the observations of Zeeman effect. However, we know that the basic assumption of classical electrodynamics that electrons in orbit around the nucleus emit radiation continuously is incorrect since that would make all atoms unstable. This is because the radiating electrons will lose energy and quickly fall into the nucleus. These and other difficulties with the classical theory can be addressed in a proper quantum theory of the radiating electron. If one does a proper quantum mechanical calculation of an atom in a magnetic field, one gets a complete theory of Zeeman effect, which explains all the detailed observations too. We shall not be discussing the quantum theory in this book.

7.8 BOUNDARY VALUE PROBLEMS WITH $\vec{B}$ AND $\vec{H}$

7.8.1 Boundary Conditions

We are already familiar with boundary conditions on the vectors $\vec{D}$ and $\vec{E}$ at the interface of two regions with different permittivities in electrostatics. A very similar situation obtains in the presence of magnetic fields. The vectors $\vec{B}$ and $\vec{H}$ also obey certain conditions at the interface of two regions with different permeabilities μ. In the absence of any free currents, we know the differential equations that the field $\vec{B}$ satisfies. These are

$$\begin{aligned} \vec{\nabla} \times \vec{H} &= 0 \\ \vec{\nabla} \cdot \vec{B} &= 0 \end{aligned} \tag{7.38}$$

Now consider an interface of two different media with different permeabilities as shown in Fig. 7.7.

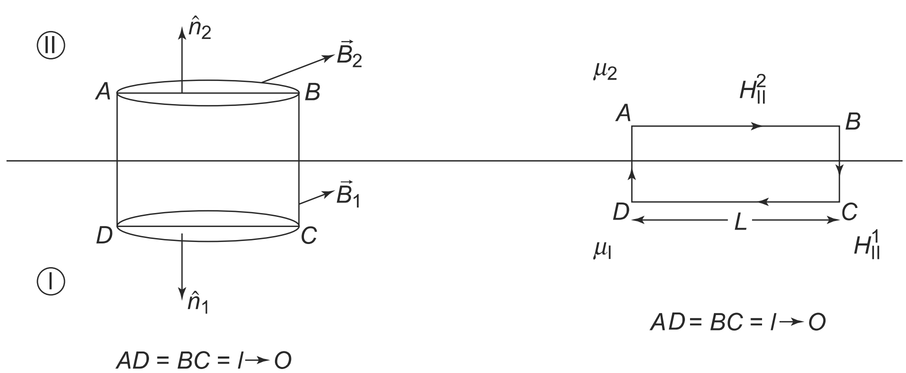


Fig. 7.7 *Boundary conditions across the interface of two regions I and II with permeabilities μ_1 and μ_2. $\hat{n}_1$ and $\hat{n}_2$ are the two outgoing unit vectors normal to the interface in the two regions. $H_{||}^1$ and $H_{||}^2$ are the tangential components of $\vec{H}$ in the two regions*

In the first part of the Fig. 7.7, we show a pill box of vanishing thickness with flat ends. The two flat ends of the pill box are in the two different regions, on either side of the interface. Applying the Divergence theorem to the closed surface of the pill box, we have

$$\iiint (\vec{\nabla} \cdot \vec{B}) dV = \iint \vec{B} \cdot d\vec{S} \tag{7.39}$$

where the surface integral is over the closed surface of the pill box. But $\vec{\nabla} \cdot \vec{B} = 0$ and so we get

$$\iint \vec{B} \cdot d\vec{S} = 0 \tag{7.40}$$

The closed surface of the pill box comprises of the two flat ends and the curved surface which we have taken to be vanishing. Thus, the contribution to the surface integral is vanishing from the curved part. The integral becomes

$$\iint \vec{B} \cdot d\vec{S} = A \left(\hat{n}_1 \cdot \vec{B}^1 + \hat{n}_2 \cdot \vec{B}^2 \right) = 0 \tag{7.41}$$

where A is the area of flat ends of the pill box. If we define

$$\hat{n}_1 \cdot \vec{B}^1 = B_\perp^1$$
$$\hat{n}_2 \cdot \vec{B}^2 = B_\perp^2$$

as the normal components of the vector $\vec{B}$ in the two regions, and knowing that $\hat{n}_1 = -\hat{n}_2$, we get

$$\boxed{B_\perp^1 = B_\perp^2} \tag{7.42}$$

Thus, we see that

> The normal component of the magnetic field $\vec{B}$ is continuous across an interface of two different media of different permeabilities.

Next, we consider the second part of Fig. 7.7. In this, we construct a closed rectangular path $ABCD$, with the long edges (AB and CD) on either side of the interface. The sides of the rectangle (AD and BC), perpendicular to the interface are of vanishing length. We apply Stokes' Theorem to the vector $\vec{H}$ on this path.

$$\iint (\vec{\nabla} \times \vec{H}) \cdot d\vec{S} = \oint_{ABCD} \vec{H} \cdot d\vec{l} = 0 \tag{7.43}$$

since $\vec{\nabla} \times \vec{H} = 0$ in the absence of free currents. But

$$\oint_{ABCD} \vec{H} \cdot d\vec{l} = L \left(H_{||}^1 - H_{||}^2 \right) = 0 \tag{7.44}$$

where L is the length of AB and CD. We have taken the contributions from the sides AD and BC to be zero since these are of vanishing length. This gives us the boundary condition on $\vec{H}$.

$$\boxed{H_{||}^1 = H_{||}^2} \tag{7.45}$$

or

The tangential component of the field $\vec{H}$ is continuous across the interface of two different media of different permeabilities.

PROBLEM 7.4 At a boundary interface between air and a ferromagnetic material, the magnetic field $\vec{B}$ is seen to be making an angle of 45 degrees with the normal. If the magnetic susceptibility of the ferromagnetic material is (0.5×10^3), how strong is the magnitude of $\vec{B}$ inside the ferromagnetic material as compared to its value outside?

7.8.2 Boundary Value Problems and Magnetic Scalar Potential

Now that we have the boundary conditions on the vectors $\vec{B}$ and $\vec{H}$, we are ready to solve boundary value problems in the presence of magnetic material. Recall the situation in electrostatics. The basic differential equations were the Laplace and Poisson's equation (in the absence of charges and in the presence of charges) for the electric scalar potential ϕ. These partial differential equations were solved with the given boundary conditions in the problem. A very similar situation obtains with magnetic fields.

The basic equation we have is for the magnetic vector potential $\vec{A}$, introduced in Chapter 6. This is

$$\nabla^2 \vec{A}(\vec{r}) = -\mu_0 \vec{j}(\vec{r}) \tag{7.46}$$

where we need to add to the free current density $\vec{j}_f$, the bound current densities as discussed in Eq. (7.18). This equation is of course-very similar to the Poisson's equation except that it is for a vector quantity $\vec{A}$ rather than for a scalar potential ϕ and hence is more complicated.

In the case where there are no free currents, the situation simplifies considerably. Then $\vec{j}(\vec{r}) = 0$ and hence, Eq. (7.39) tells us that

$$\vec{\nabla} \cdot \vec{B} = 0 \qquad\qquad \vec{\nabla} \times \vec{H} = 0 \tag{7.47}$$

Further, in magnetic materials which are linear, (i.e., diamagnetic, paramagnetic and ferromagnetic in the linear region below saturation)), the quantities $\vec{B}$ and $\vec{H}$ are proportional to each other and given by

$$\vec{B} = \mu \vec{H}$$

with μ being the permeability, the equations simplify further. In these materials, we have

$$\vec{\nabla} \times \vec{H} = 0$$

and

$$\vec{\nabla} \cdot \vec{H} = 0$$

Thus, the vector $\vec{H}$ can be written as the gradient of a scalar

$$\vec{H} = -\vec{\nabla}\phi_m \tag{7.48}$$

where ϕ_m is called the **magnetic scalar potential.** Since $\vec{\nabla} \cdot \vec{H} = 0$, we then have

$$\boxed{\nabla^2 \phi_m = 0} \tag{7.49}$$

We therefore see that for linear materials, in the absence of free currents, the magnetic scalar potential satisfies the Laplace equation, just the like electric scalar potential ϕ.

More generally, since we know

$$\vec{B} = \mu_0(\vec{H} + \vec{M})$$

where $\vec{M}$ is the magnetisation, we can write

$$\vec{\nabla} \cdot \vec{B} = \mu_0 \vec{\nabla} \cdot (\vec{H} + \vec{M}) = 0 \tag{7.50}$$

and therefore,

$$\vec{\nabla} \cdot \vec{H} = -\vec{\nabla} \cdot \vec{M} \tag{7.51}$$

But $\vec{H} = -\vec{\nabla}\phi_m$ and hence, we have

$$\nabla^2 \phi_m = \vec{\nabla} \cdot \vec{M} \tag{7.52}$$

Thus, we see that this equation is exactly like the Poisson's equation, with $\vec{\nabla} \cdot \vec{M}$ replacing the charge density. The solution to the Poisson's equation, as we saw in electrostatics is

$$\phi_m(\vec{r}) = -\frac{1}{4\pi} \int d^3r' \frac{\vec{\nabla}' \cdot \vec{M}(\vec{r}')}{|\vec{r} - \vec{r}'|} \tag{7.53}$$

The integral in Eq. (7.53) runs over all space. In cases where the magnetisation is restricted to a finite volume, the integral will run over that volume but in that case since the magnetisation $\vec{M}$ is discontinuous across the boundary (at $r = R$), the divergence of $\vec{M}$ will be a Delta function. At the surface, we can write for the normal component of $\vec{M}$,

$$\vec{M} \cdot \hat{n} = 0 \quad \text{if} \quad r > R$$

and

$$\vec{M} \cdot \hat{n} = \vec{M} \cdot \hat{n} \quad \text{if} \quad r < R$$

The three-dimensional integral, over d^3r' in Eq. (7.53) can be written as an integral over angular variables and the radial variable in spherical polar coordinates. The divergence of the radial component will be

$$-(\vec{M} \cdot \hat{n})\delta(r - r_{\text{surface}})$$

With this, one can do the radial integration in Eq. (7.53) and we get a contribution

$$\iint \frac{(\vec{M} \cdot \hat{n})}{|\vec{r} - \vec{r}'|} dS$$

where the integral is over the surface bounding the volume which is magnetised.

Since the equations satisfied by the magnetic vector potential ϕ_m are like those satisfied by the electric scalar potential ϕ, we can use the techniques we developed in electrostatics to solve the problems with magnetic materials. We illustrate this with the example of a bar magnet.

EXAMPLE 7.2 Find the magnetic field vectors $\vec{B}$ and $\vec{H}$ due to a bar magnet with a uniform magnetisation $\vec{M}$.

Solution

We have a situation where there are no free currents. Hence, the magnetic fields can be obtained from the magnetic scalar potential ϕ_m which satisfies the Poisson's equation Eq. (7.52). Now since the magnetisation is uniform in the bar magnet, we have that

$$\vec{\nabla}' \cdot \vec{M}(\vec{r}') = 0$$

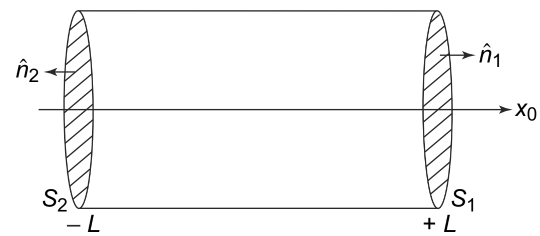


Fig. 7.8 *Example 7.2: Bar magnet*

at all points within the magnet, except at the end points. At the end plate S_1, $\vec{M}$ is zero outside (free space) and a constant inside. At the end plate S_2, it is the same situation reversed. Further, Eq. (7.53) which involves a volume integral involving the divergence of $\vec{M}$. The contribution therefore comes only from the surface region and the integration of $\frac{\vec{\nabla}' \cdot \vec{M}(\vec{r}')}{|\vec{r}-\vec{r}'|}$ over the surface thickness at any point simply gives $\left[-\frac{\vec{n}\cdot\vec{M}}{|\vec{r}-\vec{r}'|}\right]$ at that point where $\vec{n}$ is the local normal to the surface. The surface integral will only get non-zero contribution from the two end plates and so we have

$$\phi_m(\vec{r}) = \frac{1}{4\pi} \iint_{S_1} \frac{\hat{n}_1 \cdot \vec{M}(\vec{r}')}{|\vec{r}-\vec{r}'|} \cdot d\vec{S} + \frac{1}{4\pi} \iint_{S_2} \frac{\hat{n}_2 \cdot \vec{M}(\vec{r}')}{|\vec{r}-\vec{r}'|} \cdot d\vec{S} \tag{7.54}$$

where $\hat{n}_1$ and $\hat{n}_2$ are as shown in the Fig. 7.8. If $\hat{n}_1$ and $\hat{n}_2$ are perpendicular to $\vec{M}$, the two integrals for scalar potential ϕ_m in Eq. (7.54) are exactly the same as for electric potential due to uniformly charged disc. These integrals cannot be expressed in terms of elementary functions. In the limit that the point $\vec{r}$ is very far away from the magnet, we have

$$\frac{1}{|\vec{r}-\vec{r}'|} = \frac{1}{r} + \frac{\vec{r}' \cdot \vec{r}}{r^3} + \cdots \tag{7.55}$$

Notice that $\hat{n}_1 = -\hat{n}_2$ and so the contribution of the first term in Eq. (7.55) to the magnetic scalar potential ϕ_m in Eq. (7.54) cancels out. Further, notice that the difference of $\vec{r}'$ for the two ends S_1 and S_2 is precisely the dimensions of the bar magnet along the vector $\vec{M}$., i.e.,

$$\vec{r}'_{S_1} - \vec{r}'_{S_2} = \vec{d}$$

Thus, the contribution of the second term in Eq. (7.55) to the Eq. (7.54) is the dipole moment contribution since

$$\hat{n}_2 \cdot \vec{M} = M$$

$\hat{n}_2$ being parallel to $\vec{M}$. Defining

$$\vec{m} = M d\vec{A}$$

as the magnetic dipole moment of the bar magnet, we have the dipole contribution to be

$$\phi_m^{\text{dipole}}(\vec{r}') = \frac{\vec{m} \cdot \vec{r}}{4\pi r^3} \tag{7.56}$$

Using Eq. (7.56) in Eq. (7.48) will give us the dipole contribution to the magnetic field $\vec{H}$. This expression Eq. (7.56) is of course, valid for r much larger than values of r', i.e., when we are considering magnetic fields far away from the dipole. For other values of r, Eq. (7.54) has to be solved numerically. This leads to a $\vec{H}$

$$\vec{H} = -\vec{\nabla}\phi_m = \frac{3\vec{m}\cdot\hat{r} - \vec{m}}{r^3}$$

For any value of $\vec{r}$ we can of course use the solution to Eq. (7.54). Knowing $\vec{H}$, we can determine $\vec{B}$ since

$$\vec{B} = \mu_0(\vec{H} + \vec{M})$$

Furthermore, the boundary conditions, Eq. (7.42) and Eq. (7.45), can be used to evaluate the changes in the values of $\vec{B}$ and $\vec{H}$ across the interfaces of the magnet and free space.

Outside the magnet $\vec{M} = 0$ and hence, $\vec{B} = \mu_0\vec{H}$. Just inside the magnet as we cross the boundary, $H_{||}$, the tangential component of $\vec{H}$ and $B_\perp$, the normal component of B are continuous. If we assume that $\vec{M}$ is normal to the end faces, we get:

$$B_{||}^{\text{inside}} = B_{||}^{\text{outside}}$$

since M has no tangential component.

$$H_\perp^{\text{inside}} = \frac{B_\perp^{\text{inside}}}{\mu_0} - M = \frac{B_\perp^{\text{outside}}}{\mu_0} - M = H_\perp^{\text{outside}} - M$$

Thus, the normal component of the field H, suffers a discontinuity. The lines of $\vec{B}$ and $\vec{H}$ are shown in the Figs. 7.9(a) and Fig. 7.9(b).

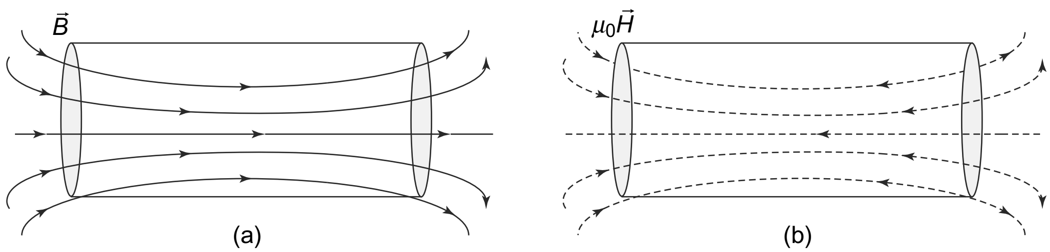


Fig. 7.9 *Magnetic field for a bar magnet (a) Field lines of $\vec{B}$ field, (b) Field lines of $\mu_0\vec{H}$. Notice the discontinuity at the end plates*

We shall consider another example of a boundary value problem in the Advanced Topic at the end of the chapter.

PROBLEM 7.5 A sphere of radius R made of material of permeability μ has a tiny magnetic dipole of moment $m\hat{z}$ at its centre. Calculate the magnetic fields $\vec{B}$ and $\vec{H}$ both inside and outside the sphere.

PROBLEM 7.6 A spherical shell of thickness t has an outer radius R and is uniformly magnetised in the z-direction with the magnetisation given by $\vec{M} = |\vec{M}|\hat{k}$. What is the magnetic scalar potential along the z-axis?

7.9 MAGNETIC CIRCUITS

7.9.1 Magnetic Flux

We are already familiar with the notion of the electric flux associated with an electric field. Recall that this is defined as

$$\Phi_e = \iint_S \vec{E} \cdot d\vec{S}$$

for any surface S. It was this quantity which is related to the charge enclosed by the surface S in Gauss's Law. For magnetic fields, we can define, in complete analogy a quantity called magnetic flux, Φ_m as

$$\Phi_m = \iint_S \vec{B} \cdot d\vec{S} \tag{7.57}$$

It turns out that this quantity is of great importance in the study of magnetic fields, as we shall see later in this Section as well as in the next Chapter. The units of magnetic flux are Webers. One property of the magnetic flux can immediately be glanced. Recall that

$$\vec{\nabla} \cdot \vec{B} = 0$$

Now if we use the Divergence Theorem, we see that for a **closed** surface S

$$\Phi_m = \iint_S \vec{B} \cdot d\vec{S} = \iiint \vec{\nabla} \cdot \vec{B} dV = 0 \tag{7.58}$$

The net flux through a closed surface is always zero. This means that there are no magnetic sources or sinks and magnetic flux lines always form closed loops.

7.9.2 Magnetic Circuits

The general problem of magnetism is to determine the vectors $\vec{B}$ and $\vec{H}$ for a given distribution of magnetisation $\vec{M}$. This is in general, a very complicated problem and cannot, in most practical cases, be solved analytically. However, in cases where the magnetic flux is confined to some region, one can speak of a magnetic circuit and solve the problem approximately.

The concept of magnetic circuit is best illustrated through an example of a toroidal coil wrapped around a ferromagnetic core. We consider a ring-shaped core made of a magnetic material with a very high permeability μ (typically a ferromagnetic substance) around which are wound N turns of a current carrying wire. The current in the coils is I.

The winding on the ring-shaped core is very closely packed. In this case then, it is a very good approximation to assume that the magnetic field created by the coil is confined only to the core without

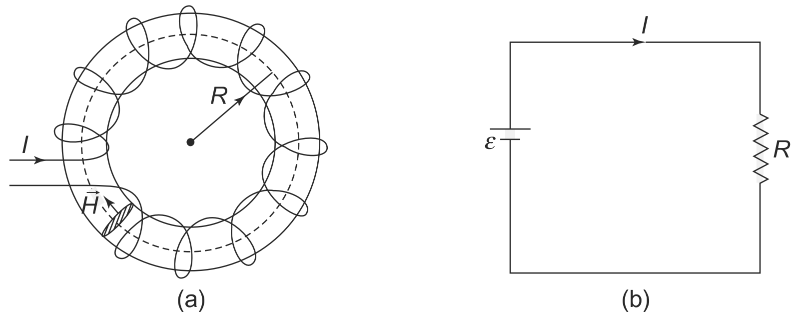


Fig. 7.10 *A magnetic circuit (a) A toroid with N turns or wire carrying current I wound on a high permeability magnetic material in a ring shape, (b) A parallel electric circuit*

any leakage to the outside region. In this case, we can apply Ampere's Law to a closed, circular path inside the ring shaped core and obtain

$$\oint H dl = NI \tag{7.59}$$

since $\vec{H}$ and $\vec{dl}$ are in the same direction. Further, at all points on the circular path, $\vec{H}$ will be constant and so we get

$$H = \frac{NI}{2\pi R} \tag{7.60}$$

It is also convenient to express $\vec{H}$ at each point along the path, in terms of the magnetic flux Φ_m. But the flux through the core is simply

$$\Phi_m = BA = \mu HA = \mu \frac{NIA}{2\pi R} \tag{7.61}$$

where A is the cross-sectional area of the core. We can rewrite Eq. (7.61) as

$$\boxed{\Phi_m = \frac{NI}{\left(\frac{L}{\mu A}\right)}} \tag{7.62}$$

This way of writing Φ_m reminds us of the Ohm's law which relates the EMF $\mathcal{E}$ in a circuit with the resistance R as

$$\text{(current)} = \frac{\text{EMF}}{\text{Resistance}} \tag{7.63}$$

or

$$I = \frac{\mathcal{E}}{R} \tag{7.64}$$

Equations (7.62) and (7.64) are similar if we define a quantity called the *magnetomotive force* (mmf) in analogy to EMF as

$$\text{mmf} = NI \tag{7.65}$$

and a quantity $\mathcal{R}$ called reluctance in analogy to resistance as

$$\mathcal{R} = \left(\frac{L}{\mu A}\right) \tag{7.66}$$

Then the flux equation Eq. (7.62) becomes

$$\boxed{\Phi_m = \frac{\text{mmf}}{\mathcal{R}}} \tag{7.67}$$

Equation (7.67) is the magnetic equivalent of Ohm's Law for magnetic circuits. The magnetic flux lines are called magnetic circuits in analogy to electric current. In fact, the analogy can be taken even further. Note that the reluctance, $\mathcal{R}$ and resistance R have a similar dependence on the dimensions of the sample. Thus, we know that for a thin, conducting bar of length L, cross-section area A and conductivity σ, the resistance is given by

$$R = \frac{L}{\sigma A}$$

Note the similarity between this and Eq. (7.66). For magnetic circuits, the permeability μ plays the same role as conductivity σ for electric circuits. In fact, the analogy to electric circuits goes even further. Consider a case where the circuit has a combination of reluctances. One such case could be the same toroidal circuit as above but instead of the core being made of a single material, it could be two materials with permeabilities μ_1 and μ_2 as shown in Fig. 7.11.

Here the core consists of a length L_1 of permeability μ_1 and a length L_2 of permeability μ_2 with $L_1 + L_2 = L = 2\pi R$. Now Ampere's Law (Eq. (7.59)) applied to a similar circular path in the core gives us

$$\oint H dl = NI \tag{7.68}$$

Though H is the same along the path as before, we now have in one part

$$B = \mu_1 H$$

and in the second part

$$B = \mu_2 H$$

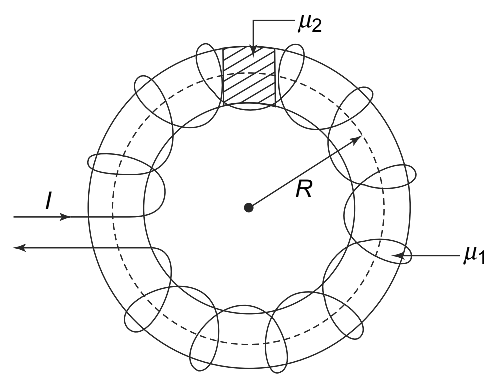


Fig. 7.11 *Two materials with different permeabilities*

Hence, we get

$$\Phi_m \left[\frac{L_1}{\mu_1 A} + \frac{L_2}{\mu_2 A}\right] = NI \tag{7.69}$$

Comparing this equation with Eq. (7.62) we see that the effective reluctance is simply

$$\mathcal{R} = \left[\frac{L_1}{\mu_1 A} + \frac{L_2}{\mu_2 A}\right] \tag{7.70}$$

which is just the sum of two reluctances. This is similar to what we will get for the effective resistance for a series combination of two resistances in the electric case. This analogy can of course, be extended to a combination of more than two kinds of materials with different permeabilities. However, there is a word of caution.

It might be thought that one can use the same argument when there is an air gap in the core with the air gap being considered as a material with permeability μ_0. This would however not be possible in all cases. Recall that the underlying logic of treating magnetic flux lines as electric current was that they are confined to the ferromagnetic core. This is a very good assumption in the case of ferromagnetic materials since these have very high permeabilities and hence B is much larger than in air and so the flux tends to concentrate into ferromagnetic materials and is confined to it. In the case of air, $B = \mu_0 H$ and there is no reason for the lines of $\vec{B}$ to be confined to the air gap. In fact, there would always be some leakage, i.e., flux lines coming out of the air gap. If the leakage is small, then of course, expressions like Eq. (7.70) are good approximations even in the presence of air gaps. The amount of leakage depends on the geometry of the air gap as well as on the permeabilities of the materials used.

Electrical currents also obey Kirchhoff's First law, which is simply a statement of conservation of charge as we have seen in a previous chapter. This law states that the algebraic sum of the currents at any point in a circuit is always zero, or as much current leaves a node in a circuit as enters it. A similar situation prevails for magnetic circuit where we can show that the magnetic flux lines obey a similar constraint. This is because, Kirchhoff's first law follows from the statement that for steady currents

$$\vec{\nabla} \cdot \vec{j} = 0$$

In magnetic circuits, we have a similar situation where

$$\vec{\nabla} \cdot \vec{B} = 0$$

To see this, consider a high permeability material so that the magnetic flux lines are confined to it (Fig. 7.12(a)).

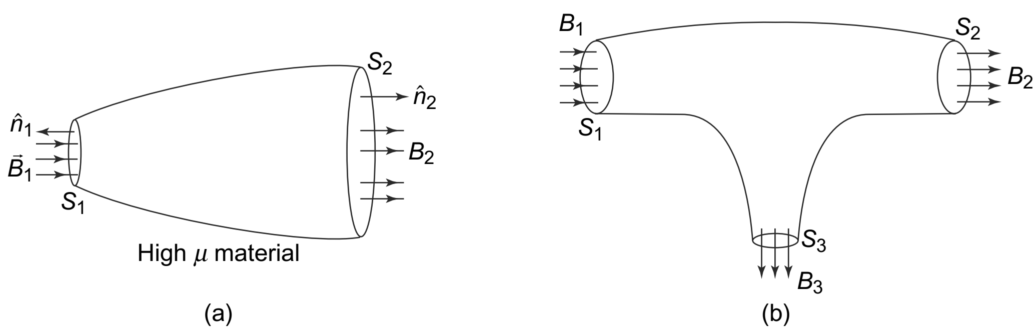


Fig. 7.12 *A magnetic circuit (a) Flux across S_1 and S_2 are equal, (b) Algebraic sum of outward flux across S_1, S_2 and S_3 must vanish*

Consider the integral

$$\iiint (\vec{\nabla} \cdot \vec{B}) dV$$

for the region shown. Then we have

$$0 = \iiint (\vec{\nabla} \cdot \vec{B}) dV = \iint_{S_1} (\hat{n}_1 \cdot \vec{B}_1) dS_1 + \iint_{S_2} (\hat{n}_2 \cdot \vec{B}_2) dS_2 \tag{7.71}$$

where we have used the Divergence Theorem to convert the volume integral to a surface integral over the closed surface bounding the volume. The contribution from the part of the surface other than the end plates vanishes since we can take $\vec{B}$ to be parallel to it and hence, the scalar product with the outward normal would vanish. We see, therefore, that as much flux enters S_1 as leaves from S_2. We can extend this to any general geometry in the circuit as shown in Fig. 7.12(b). Here the algebraic sum of the outward flux through S_1, S_2 and S_3 must vanish. This is an exact parallel to the Kirchoff's First Law for circuits, as we have seen.

We now illustrate the use of magnetic circuits in a situation where series and parallel combinations of reluctance are used.

EXAMPLE 7.3 Obtain the relationship between the magnetic flux and current for the arrangement shown in Fig. 7.13(a). The thickness of the arms may be taken to be small and uniform so that the bending of the flux lines at the corners could be ignored.

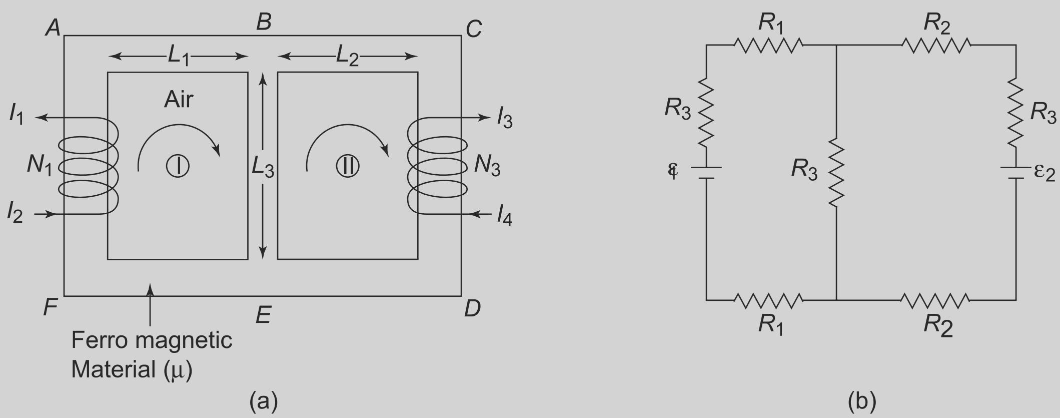


Fig. 7.13 *Example 7.3*

Solution

Let Φ_{m1} and Φ_{m2} be the fluxes in the clockwise direction in Loops 1 and 2 as shown in Fig. 7.13(a). The fluxes through AB, EF and FA are thus Φ_{m1} and fluxes through BC, DE and CD are Φ_{m2}. The flux through BE is then $\Phi_{m1} - \Phi_{m2}$. Applying Ampere's Law now to Loop 1

$$\begin{aligned} N_1 I_1 &= \oint_{FA} \vec{H} \cdot d\vec{l} + \oint_{AB} \vec{H} \cdot d\vec{l} + \oint_{BE} \vec{H} \cdot d\vec{l} + \oint_{EF} \vec{H} \cdot d\vec{l} \\ &= \frac{\Phi_{m1} L_3}{A\mu} + \frac{\Phi_{m1} L_1}{A\mu} + \frac{(\Phi_{m1} - \Phi_{m2}) L_3}{A\mu} + \frac{\Phi_{m1} L_1}{A\mu} \\ &= \frac{1}{A\mu} [\Phi_{m1}(2L_1 + 2L_3) + \Phi_{m2}(-L_3)] \end{aligned} \tag{7.72}$$

Similarly, for Loop 2 we have

$$\begin{aligned} N_3 I_3 &= \oint_{BC} \vec{H}\cdot d\vec{l} + \oint_{CD} \vec{H}\cdot d\vec{l} + \oint_{DE} \vec{H}\cdot d\vec{l} + \oint_{EB} \vec{H}\cdot d\vec{l} \\ &= \frac{\Phi_{m2}L_3}{A\mu} + \frac{\Phi_{m2}L_2}{A\mu} + \frac{(\Phi_{m2}-\Phi_{m1})L_3}{A\mu} + \frac{\Phi_{m2}L_2}{A\mu} \\ &= \frac{1}{A\mu}\left[\Phi_{m2}(2L_2+2L_3) + \Phi_{m1}(-L_3)\right] \end{aligned} \tag{7.73}$$

Now, defining reluctances as

$$\mathcal{R}_{AB} = \mathcal{R}_{EF} = \frac{L_1}{A\mu} = \mathcal{R}_1$$

$$\mathcal{R}_{CD} = \mathcal{R}_{BE} = \mathcal{R}_{AF}\frac{L_3}{A\mu} = \mathcal{R}_3$$

$$\mathcal{R}_{BC} = \mathcal{R}_{DE} = \frac{L_2}{A\mu} = \mathcal{R}_2$$

Equations (7.72 and 7.73) read

$$\begin{aligned} N_1 I_1 &= \Phi_{m1}(2\mathcal{R}_1 + 2\mathcal{R}_3) + \Phi_{m2}(-\mathcal{R}_3) \\ N_3 I_3 &= \Phi_{m2}(2\mathcal{R}_2 + 2\mathcal{R}_3) + \Phi_{m1}(-\mathcal{R}_3) \end{aligned} \tag{7.74}$$

The equivalent electrical circuit, shown in Fig. 7.13(b) has identical equations if we replace reluctances by resistances and EMF by the magnetomotive force (mmf). The solutions to these equations can be written as

$$\begin{aligned} \Phi_{m1} &= \frac{N_1 I_1 (2\mathcal{R}_2 + 2\mathcal{R}_3) + N_3 I_3 \mathcal{R}_3}{4(\mathcal{R}_1+\mathcal{R}_3)(\mathcal{R}_2+\mathcal{R}_3) - \mathcal{R}_3^2} \\ \Phi_{m2} &= \frac{N_1 I_1 \mathcal{R}_3 + N_3 I_3 (2\mathcal{R}_1 + 2\mathcal{R}_3)}{4(\mathcal{R}_2+\mathcal{R}_3)(\mathcal{R}_1+\mathcal{R}_3) - \mathcal{R}_3^2} \end{aligned} \tag{7.75}$$

7.10 ADVANCED TOPIC

7.10.1 Magnetic Field due to a Uniformly Magnetised Sphere

As an example of boundary value problem, we have already considered the magnetic fields due to a cylindrical bar magnet in Example 7.2. We now discuss the problem of a uniformly magnetised sphere of radius R with uniform magnetisation $\vec{M}$.

Let us take the direction of $\vec{M}$ to be the z-direction. The problem has obvious azimuthal symmetry, that is, there is no change when we rotate around the z-axis. Hence, the magnetic field vectors $\vec{B}$, $\vec{H}$ or the magnetic scalar potential ϕ_m (NOT to be confused with the magnetic flux, which is Φ_m) cannot depend on the variable ϕ. Further, note that the magnetisation is uniform within the sphere. Thus, the quantity $\vec{\nabla}'\cdot\vec{M}(\vec{r}') = 0$ everywhere inside the sphere and is only non-zero at the surface of the sphere. We can write the scalar potential as in Eq. (7.53) as

$$\phi_m(r,\theta) = \frac{MR^2}{4\pi}\int\int \frac{\cos\theta'}{|\vec{r}-\vec{r}'|}\sin\theta' d\theta' d\phi' \tag{7.76}$$

Note that θ is the polar angle of the vector $\vec{r}$ while θ' is the angle between $\vec{r}$ and $\vec{r}'$. To evaluate this integral, we use the expansion for the denominator as

$$\begin{aligned}\frac{1}{|\vec{r}-\vec{r}'|} &= \sum_{l=0}^{\infty}\frac{r^l}{r'^{l+1}}P_l(\hat{r}\cdot\hat{r}') \quad \text{for} \quad |\vec{r}| < |\vec{r}'| \\ &= \sum_{l=0}^{\infty}\frac{r'^l}{r^{l+1}}P_l(\hat{r}\cdot\hat{r}') \quad \text{for} \quad |\vec{r}| > |\vec{r}'| \end{aligned} \tag{7.77}$$

where $P_l(\cos\theta)$ are the Legendre Polynomials of order l that we have discussed in the previous chapters.

Thus, for points outside the sphere, where $r > r'$, we have

$$\phi_m(r,\theta) = \frac{MR^2}{4\pi}\sum_{l=0}^{\infty}\frac{R^l}{r^{l+1}}\int \sin\theta'\cos\theta' P_l(\hat{r}\cdot\hat{r}')d\theta' d\phi' \tag{7.78}$$

For the purposes of evaluating this integral, we can take the direction of $\hat{r}$ along the z axis. Then we get $\hat{r}\cdot\hat{r}' = \cos\theta'$ and $\vec{M}\cdot\hat{n} = (\cos\theta\cos\theta' + \sin\theta\sin\theta'\cos\phi')$. Therefore,

$$\begin{aligned}\phi_m(r,\theta) &= \frac{MR^2}{4\pi}\int\frac{\sin\theta'(\cos\theta\cos\theta' + \sin\theta\sin\theta'\cos\phi')}{(r^2 + r'^2 + 2rr'\cos\theta)^{\prime 1/2}}d\theta' d\phi' \\ &= \frac{1}{3}\frac{MR^3}{r^2}\cos\theta \quad \text{for} \quad r > R \\ &= \frac{1}{3}Mr\cos\theta \quad \text{for} \quad r < R \end{aligned} \tag{7.79}$$

We can determine the fields in the two regions from this scalar potential. We get

For $r > R$

$$\begin{aligned}\vec{H} &= -\vec{\nabla}\phi_m(r,\theta) \\ H_r &= \frac{1}{3}MR^3\left(\frac{2}{r^3}\right)\cos\theta \\ H_\theta &= \frac{1}{3}MR^3\left(\frac{1}{r^3}\right)\sin\theta \end{aligned} \tag{7.80}$$

In this region, $\vec{B} = \mu_0\vec{H}$.

For $r < R$

$$\begin{aligned}\phi_m(r,\theta) &= \frac{1}{3}Mz \\ \vec{H} &= -\vec{\nabla}\phi_m = -\frac{1}{3}M\hat{z} \\ \vec{B} &= \mu_0(\vec{H}+\vec{M}) = \mu_0(\frac{2}{3}M\hat{z}) \end{aligned} \tag{7.81}$$

since in this region $\vec{M} = M\hat{z}$.

The field lines in the two regions are shown in Fig. 7.14.

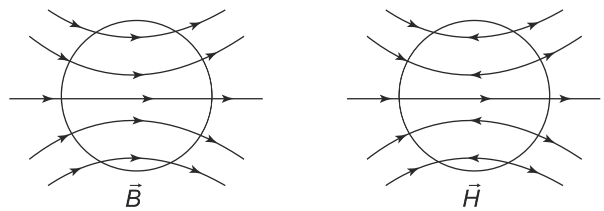


Fig. 7.14 *Field lines in the presence of a uniformly magnetised sphere*

SUMMARY

- There are three kinds of materials with magnetic properties—paramagnetic, diamagnetic and ferromagnetic materials.
- Magnetic properties are due to the atomic magnetic dipole moments created by or induced in electrons in atoms.
- Paramagnetic materials are those with permanent atomic magnetic dipole moments. In a non-uniform magnetic field, they experience a force towards regions of stronger field.
- Diamagnetic effects are present in all materials. These are caused by the induced dipole moments in an external magnetic field. Diamagnetic materials experience a force towards weaker field regions in a non-uniform field. Diamagnetism is a much weaker effect.
- Paramagnetic susceptibility depends on temperature and the relationship is given by Curie's Law.
- Ferromagnetic materials are those which can have a permanent magnetic moment. The magnetisation is not proportional to the applied magnetic field.
- Ferromagnetic materials show hysteresis or a retention of memory of the applied magnetic field.
- Ferromagnetism cannot be explained satisfactorily on the basis of classical physics. However, a simple model of magnetic domains, caused by spin exchange interaction, explains some of the observed features.
- The Zeeman effect is the splitting of a spectral line into three lines when the source of the spectral line is placed in a magnetic field.
- The magnetic field vectors $\vec{H}$ and $\vec{B}$ satisfy certain conditions at the interface of two magnetic media.
- In situations when there are no free currents, one can introduce a magnetic scalar potential which satisfies the Poisson equation.
- In certain situations, the magnetic flux in materials can be treated as the equivalent of electric current. In situations where there is no magnetic flux leakage, one can use the concept of magnetic circuits to find the relationship between the currents and flux.

CONCEPTUAL QUESTIONS

1. A multilayer coil of 2000 turns is 20 mm long and has a thickness of 5 mm of winding. If the coil carries a current of 5 mA, the mmf generated is

a. 10 A turns
b. 500 A turns
c. 200 A turns
d. None of the above

2. Which of these formulas are wrong?
a. $B_1^{\perp} = B_2^{\perp}$
b. $B_2 = \sqrt{(B_2^{||})^2 + (B_2^{\perp})^2}$
c. $H_1 = H_1^{\perp} + H_1^{||}$.

3. Which of these statements are not true of ferromagnetic materials?
a. They have a large χ_m
b. They have a fixed value of μ_r
c. The energy loss is proportional to the area of the hysteresis loop
d. They lose their non-linear property above Curie temperature

4. Which of the following will not increase the overall strength of a magnetic field produced by a current carrying coil wound over a cardboard tube?
a. Higher current in the coil
b. Inserting an iron core in the coil
c. Increasing the number of turns in the coil
d. Use a shorter cardboard tube

5. Magnetic dipoles in materials
a. are a different form of charged particles
b. arise out of currents created by charged particles
c. must be zero for all atoms.

6. Paramagnetism and diamagnetism
a. Can be simultaneously present
b. are both present necessarily in all matter
c. both resemble a weak form of ferromagnetism

7. Spectral lines in atomic spectrum sometimes get affected by an external magnetic field (Zeeman effect). This is because
a. the radiation emitted interacts with the magnetic field
b. the energy levels of the electrons in atom get affected
c. magnetic fields change the properties of atomic nuclei.

8. The normal components of the magnetic field suffer no discontinuity at an interface but the longitudinal component does. This is because
a. their strengths are different
b. presence of surface magnetic moments
c. the normal component does not get affected by magnetisation of the material.

9. In a ferromagnetic material, the field $\vec{H}$ is
a. much weaker than $\vec{M}$

b. is not affected by magnetisation
c. is caused only be external currents.

10. Why are magnetic properties of exactly the corresponding electric ones?
a. Because there are no magnetic charges
b. Because magnetic forces are weak
c. Charged particles are not affected by magnetic fields.

PROBLEMS

1. An orbiting electron is moving in a circular orbit of radius r. Derive the relationship between the orbital angular momentum and the orbital magnetic dipole moment.

2. If the electron in Problem 1 is placed in a magnetic field $\vec{B} = 2.0$ T, at right angles to the plane of the orbit, calculate the change in magnetic moment.

3. A proton has a magnetic moment due to its spin. The magnetic moment is $1.4 \times 10^{-26}\text{Am}^2$. Calculate the electric field and the magnetic field at a distance of 1 Angstrom from the proton measured along the spin axis.

4. Find the force between two magnetic dipoles, $\vec{\mu}_1$, $\vec{\mu}_2$ aligned end on end at a distance r from each other.

5. An infinite solenoid with n turns per unit length and current I is filled with a magnetic material of susceptibility χ_m. Find the magnetic field inside the solenoid.

6. The magnetic field intensity around a perfect cylindrical conductor of radius 10 cm is $10\rho^{-1}\hat{\phi}$ A/m. What is the surface current density on the conductor?

7. In a magnetic material, the B field is 1.2 T when $H = 300$ A/m. When the H is increased to 1500 A/m, the B field is 1.5 T. What is the change in the magnetisation vector?

8. Region $0 \le z \le 2$ m is filled with an infinite slab with $\mu_r = 2.5$. If the magnetic field is $\vec{B} = 10y\hat{i} - 5x\hat{j}$ T in the slab, find $\vec{j}_f$ and $\vec{M}$.

9. A permanent magnet is made out of a cylindrical iron rod magnetised uniformly in the direction of its axis with $\vec{M}$. The rod is of radius r and is bent in the form of a ring of radius $R \gg r$, with a small gap of length $d \ll r$. The permeability of iron is μ. Calculate the magnetic field $\vec{H}$ in the gap and inside the body of the magnet.

10. A long cylindrical rod of radius R made of a paramagnetic material has a current flowing through it parallel to its axis. The magnitude of the current density at a distance r from the axis is $j = j_0\frac{r}{R}$, where j_0 is a constant. Calculate the magnetic field $\vec{H}$ inside and outside the rod. Would your results change if the rod was made of a ferromagnetic instead of paramagnetic material?

11. A long cylindrical solenoid of radius R having n windings per unit length has a current I flowing through it. A solid iron cylinder of radius $\frac{R}{2}$ is present coaxially inside the solenoid. There are windings on the rod exactly like the solenoid and the current through these windings have the same value and sense as the one through the outer solenoid. Calculate the value of both the magnetic field vectors in all regions inside and outside.

12. A tiny magnetic dipole of moment $\vec{m} = m\hat{k}$ is present at the centre of a sphere of radius R made of iron of relative permittivity 500. Calculate the magnetic field at a distance $\frac{R}{2}$ above the centre of the sphere along the z-axis.

13. A long cylindrical rod of radius R is magnetised. The magnetisation direction is parallel to the axis of the cylinder but non-uniform having a magnitude $M = M_0\left(\frac{r}{R}\right)$ at a radial distance r from the axis. Calculate the magnetic field due to the rod at points inside and outside. You can neglect end effects.

14. In Fig. 7.15 N and S are the end plates of a magnet bent in a circular form with a gap between N and S. The field lines are as shown and the field B may be assumed to be constant without any fringing around the corners. P is a paramagnetic rod with a susceptibility χ_P and has a cross-sectional area a. Calculate the force on P

a. when it is fully above the pole pieces
b. when the lower end of P is between the pole pieces and the upper end above it and
c. when the lower end is below and the upper end of P is above the pole pieces.

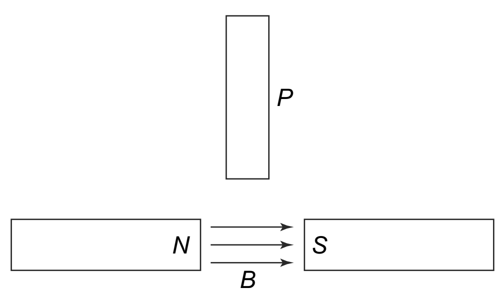


Fig. 7.15 *Problem 14*

15. The magnetic circuit in Fig. 7.10(a) has $R = 10$ cm and a cross section radius of 1 cm. The core is made of a material with $\mu = 1000\mu_0$ and the coil has 200 turns. Calculate the amount of current which will produce a flux of 0.5 Weber in the core.

16. In the magnetic circuit given in Fig. 7.16, calculate the current in the coil which will produce a magnetic flux density of 1.5 T in the air gap assuming the $\mu = 50\mu_0$ and that all the branches have a cross-sectional area of 10 square cm.

17. If $\vec{M} = \frac{k}{a}\left(-y\hat{i} + x\hat{j}\right)$, where k is a constant, in a cube of side a, find $\vec{I}_b$.

18. For a boundary between two magnetic media with permeabilities μ_1 and μ_2, show that

$$\frac{\tan\theta_1}{\tan\theta_2} = \frac{\mu_1}{\mu_2}$$

where θ_1 and θ_2 are the angles the fields make with the normal to the interface.

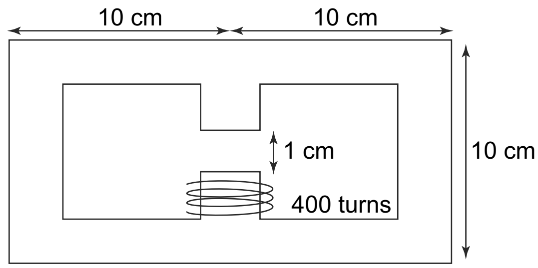


Fig. 7.16 *Problem 16*

19. The plane $z = 0$ separates air ($z \geq 0$) from iron ($z \leq 0$, $\mu = 200\mu_0$). given that

$$\vec{H} = 10\hat{i} + 15\hat{j} - 3\hat{k}$$

in air, find $\vec{B}$ in iron and the angle it makes with the interface.

20. In the magnetic circuit given in Fig. 7.17, assuming the core has $\mu = 1000\mu_0$ and a cross-section area of 4 square cm, determine the flux density in the air gap.

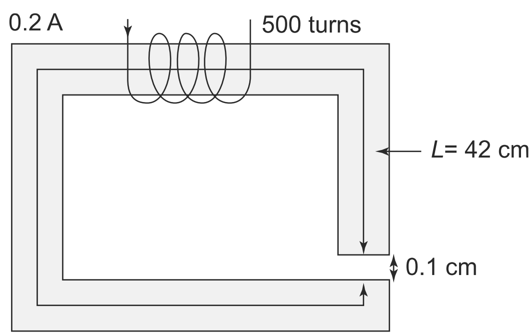


Fig. 7.17 *Problem 20*

8

Electromagnetic Induction

Learning Objectives

- To learn about the physical phenomenon associated with Faraday's Law.
- To understand the physical meaning of the ways in which emf can be induced as per Faraday's Law.
- To learn about Lenz's law and apply it in different situations.
- To comprehend the concept of energy stored in a magnetic field.
- To learn about the concept of self- and mutual inductance.
- To understand the reciprocity relation between mutual inductances.
- To relate the energy stored in inductances to the energy stored in magnetic field.
- To learn about the relationship between self- and mutual inductances and the coefficient of coupling between circuits.

8.1 FARADAY'S EXPERIMENTS

In Chapter 6, we saw that electric currents produced magnetic fields, a fact that was first noticed by Oersted and Ampere. They found that the passage of an electric current produces a magnetic field. In the early part of the nineteenth century, Michael Faraday (1791–1867) began a series of experiments to investigate the reverse effect, namely to see if magnetic fields could produce electric currents. With his experiments, Faraday was able to show that magnetic fields indeed can produce an electric current. However, he found that an electric current is produced **only if** the magnetic field at the location of the coil changed with time. There was no current produced if the magnetic field was not changing with time. This result, he confirmed, was true irrespective of whether the magnetic field was produced by an electric current or a permanent magnet. This was the discovery of the phenomenon of **electromagnetic induction**, a discovery which is responsible for not only the age of electricity but also finally led to the unification of electricity and magnetism in the work of Maxwell.

More correctly, what Faraday discovered were two apparently unrelated phenomenon. To see this, let us consider a schematic representation of Faraday's experimental arrangement as shown in Fig. 8.1. The arrangement consists of a loop which has a magnetic field $\vec{B}$ at right angles to it.

Faraday observed that

(a) With a constant magnetic field, if the loop is moved forward or backward, such that the magnetic flux associated with the loop changes, there is a current in the loop and the lamp lights up.

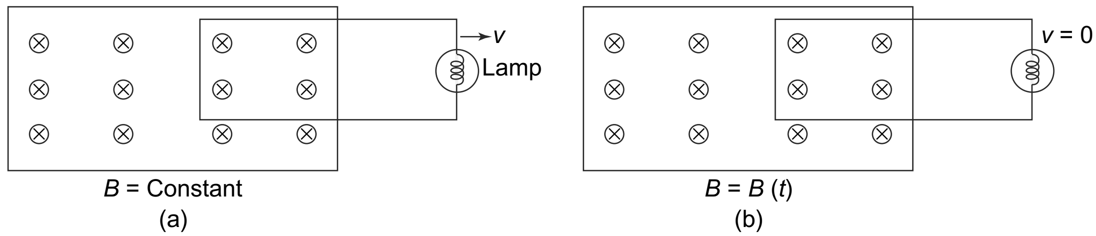


Fig. 8.1 *Schematic Diagram of Faraday's Experiment—a magnetic field exists at right angles to a conducting loop connected to a resistor in the form of a lamp. The magnetic field, going into the paper, is restricted to the rectangle as shown in the figure: (a) Magnetic field constant but the loop moves with a velocity v, (b) Loop stationary but the magnetic field changes with time,* $\vec{B}(t)$

(b) If the strength of the magnetic field is changed, keeping the loop stationary, and thereby once again, the magnetic flux associated with the loop is changed, there is an electric current in the loop and the lamp lights up.

These observations are remarkable since both of them seem to imply that the current is produced only if the magnetic flux associated with the loop is changing with time. Furthermore, the current lasts only for as long as the flux is changing. Thus, for example, in Fig. 8.1(a), if the loop is fully inside the region of constant magnetic field, moving it backward or forward while keeping the whole loop in the region of the uniform magnetic field, does not produce any electric current since the magnetic flux associated with the loop is not changing.

Faraday not only discovered the phenomenon of electromagnetic induction but also formulated the laws that govern this phenomenon.

8.2 FARADAY'S LAWS

Consider a loop in which an electric field $\vec{E}$ exists. We have already seen that the presence of the electric field leads to a current if a conductor is present in the loop. We can construct a loop integral

$$\oint \vec{E} \cdot d\vec{l}$$

where the integral is over the loop. Since this is a closed loop, we need to specify the sense of the integral. The sense of the integral can be either clockwise or anticlockwise. This integral, as we have seen earlier, is called the Electromotive force $\mathcal{E}$. It is of course, equal to the amount of work done in taking a unit charge around the loop along the path defined in the loop integral and in the same direction as the sense of the loop integral.

Faraday's discovery changes the properties of the electric field. In electrostatics, the electric field was conservative. That would imply that the line integral of $\vec{E} \cdot d\vec{l}$ over any closed loop would be zero. The existence of a current in a closed loop observed by Faraday implies that an electric field exists in a closed loop all along the same direction. The loop integral of $\vec{E} \cdot d\vec{l}$ over such a loop thus does not vanish. The integral is just the amount of work done in carrying a unit charge around the loop, called the electromotive force. **Thus, in the presence of time varying magnetic field, the electric field is no longer conservative.**

Clearly, for any such loop, with a specified sense, one can also define a magnetic flux Φ associated with it as

$$\boxed{\Phi = \iint_S \vec{B} \cdot d\vec{S}} \tag{8.1}$$

Note that we have dropped the subscript m from the notation for the magnetic flux. In this expression for the magnetic flux associated with the loop, S is any surface and $d\vec{S}$ is an element of the surface S in the outward direction. Outward in this context is of the right-hand screw rule. A right-handed screw turned in the sense of the given loop, moves out of the surface in the outer direction by definition.

Faraday's Law is a relationship between the two quantities—the EMF $\mathcal{E}$ and the magnetic flux Φ. It states that a change in magnetic flux associated with a loop induces an EMF in the loop which is given by the rate of change of the magnetic flux. That is

$$\boxed{\mathcal{E} = -\frac{d\Phi}{dt}} \tag{8.2}$$

The relationship is valid irrespective of whether the flux change is brought about by a change in the magnetic field, or the motion of the conductor in a constant magnetic field or both. We shall discuss the significance of the negative sign shortly.

PROBLEM 8.1 The plane of a circular loop is perpendicular to the direction of the magnetic field. The magnetic field has a magnitude $|\vec{B}| = 0.5$ T. The field goes to zero at a constant rate in $t = 0.5$ s and it is seen that the induced voltage in the loop is 0.62 V. Calculate the radius of the loop.

PROBLEM 8.2 A square coil having 10 turns measures 0.5 m on one side and has a resistance of 5 Ohms. It is placed in a magnetic field that makes an angle of 45° with the plane of the loops. The magnitude of the field varies as $B = 3t^3$ where t is in seconds and B is in Teslas. Calculate the induced current in the coil at $t = 5$ seconds.

8.2.1 Motional EMF

As we saw above, Faraday's law is valid as long as the magnetic flux associated with a conductor is changing. This change can be brought about by either the motion of the conductor in a constant magnetic field, or by the changing of a magnetic field with a stationary conductor or both these things happening simultaneously. In the case, where the induced emf is caused by the motion of a conductor in a constant magnetic field, the emf is called **Motional EMF**.

In this case, Faraday's Law above can be seen as a manifestation of the Lorentz force that we know is experienced by a charge in motion in a magnetic field. To see this, consider a uniform magnetic field

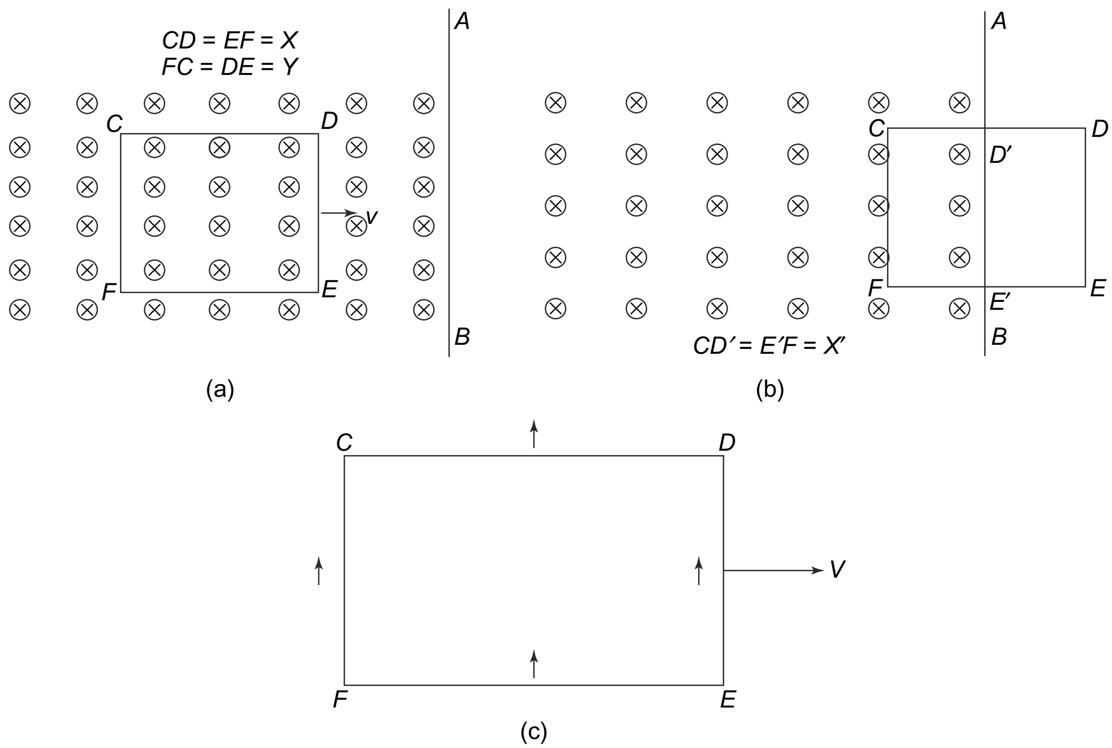


Fig. 8.2 *Motional EMF. Cross indicates magnetic field going into the plane of the paper. The field is uniform to the left of AB and zero to the right of it. $CDEF$ is a rectangular conducting loop with its outer normal parallel to $\vec{B}$. The loop is moving to the right with a speed v (a) Loop totally in the region of uniform magnetic field. The flux through the loop does not change, (b) Loop partially out of the region of uniform and constant $\vec{B}$. The flux through the loop decreases with time, (c) Direction of Lorentz force $\vec{v} \times \vec{B}$ on a positive charge*

$\vec{B}$ going into the plane of the paper as shown in Fig. 8.2(a). The magnetic field only exists to the left of the plane AB and going into the plane of the paper.

$CDEF$ is a rectangular conducting loop which is moving with a speed v towards the right. The magnetic flux associated with the loop is simply B times the area of the loop overlapping with the region where the magnetic field is non-zero. As long as the loop is totally to the left of the plane AB (Fig. 8.2(a)), the flux associated with the loop is simply $B \times \text{Area} = Ba$ where $a = X \times Y$. However, when the loop is partially to the right of AB (Fig. 8.2(b)), then the flux through the loop is

$$\Phi = BYX'$$

This flux is changing with time since X' is changing with time and $\frac{dl'}{dt} = -v$ and hence the rate of change of flux associated with the loop is

$$\frac{d\Phi}{dt} = BY\frac{dX'}{dt} = -BYv \tag{8.3}$$

Now as the loop moves, the charges in the conductor move with it and since these are charges in motion in the presence of a magnetic field, they will experience a Lorentz force. The force experienced by a unit positive charge is thus,

$$\vec{F} = \vec{v} \times \vec{B}$$

and the direction is as shown in the Fig. 8.2(c).

As long as the loop is completely to the left of the plane AB (Fig. 8.2(a)), charges in the various sections of the loop will experience a force due to the magnetic field. However, charges in the sections CD and EF cannot move out of the conductor. The charges in the sections FC and ED will move to accumulate at the end points under the action of the Lorentz force. The accumulated charges will set up an electric field which will stop any further flow of charges. Thus, there will be no net force on the electron and hence no current will flow in the loop.

Now consider the situation shown in Fig. 8.2(b). There will be no force on the charges in the section DE since there is no magnetic field in that region now. Charges in the sections CD and EF once again cannot move outside the conductor and hence, remain there. Charges in the section FC will move with the end C having a higher potential as a result of the motion of the charges. Current thus, will flow along *CDEF*. The work done in carrying a unit positive charge along the loop $FCDEF$ (which is the emf $\mathcal{E}$) will thus, get a contribution only from the section FC and hence, we have

$$\mathcal{E} = \int_{FCDEF} \vec{F} \cdot d\vec{l} = \int_{F}^{C} \vec{F} \cdot d\vec{l} = BvY \tag{8.4}$$

Thus, we see that the $\mathcal{E}$ in Eq. (8.4) and $\frac{-d\Phi}{dt}$ in Eq. (8.3) are identical.

There is one aspect of motional emf that must be understood. Consider again the situation shown in Fig. 8.2(b). For the observer who is observing that the loop is moving to the right with a velocity $\vec{v}$, the Lorentz force, caused by the magnetic field, is the reason for the current in the loop which will cause, e.g., a light bulb in the loop to light up. But, now consider another observer who is moving with the loop so that in his frame of reference the loop is at rest. There is no Lorentz force now for him but he will also observe the current in the loop and the bulb lighting up. What caused the current? Obviously, the electrons in the loop must be acted upon by some electric field. Indeed, this is true. As we will see later, the theory of relativity will lead to that. If there is a stationary magnetic field but no electric field as observed by one observer, relativity theory will tell us that for a second observer moving relative to the first, not only will the magnetic field change but he will observe an electric field as well. For the observer moving with the loop, it is this electric field that will drive the electrons in the loop to produce the current. Electric and magnetic fields thus, are not absolute observer independent concepts and their role in any phenomenon are very much dependent on the state of motion of the observer.

8.2.2 Faraday's Law in Differential Form

Consider the situation in which the loop is stationary but the magnetic field is changing with time. In this case too, Faraday's laws tell us that there will be an induced emf. This effect is a totally new effect and cannot be explained in terms of the forces on charges as above. In this case, with the loop being stationary, we can write Faraday's laws in an elegant form.

Taking a fixed loop, we can write, using Stokes' Theorem

$$\oint \vec{E} \cdot d\vec{l} = \iint_S (\vec{\nabla} \times \vec{E}) \cdot d\vec{S} \tag{8.5}$$

Hence, according to Faraday's laws

$$\iint_S (\vec{\nabla} \times \vec{E}) \cdot d\vec{S} = \mathcal{E} = -\frac{d}{dt} \iint_S \vec{B} \cdot d\vec{S} \tag{8.6}$$

Since the loop and the associated surface in the line and surface integrals above are totally arbitrary, the integrands in Eq. (8.6) must be equal. Thus, we get

$$\vec{\nabla} \times \vec{E} = -\frac{d\vec{B}}{dt} \tag{8.7}$$

The differential form of Faraday's law thus, states

> The curl of the electric field is equal to the negative of the rate of change of the magnetic field.

8.2.3 Line Integral of the Electric Field and its Relationship with EMF *

In our discussion of induced EMF, we have taken it as the integral over a closed loop of the electric field. But this integral, in electrostatics and even in current electricity vanishes since the electric field is conservative as we saw in the earlier chapters. Thus, for instance consider a simple electric circuit shown in Fig. 8.3, consisting of a battery with zero internal resistance and a resistor R.

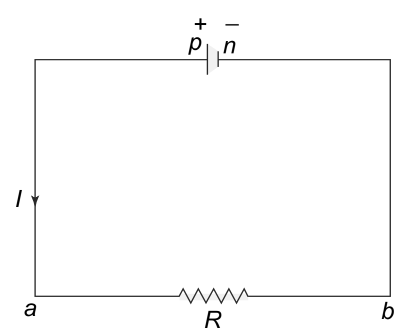


Fig. 8.3 *A simple electric circuit. The battery has zero internal resistance. The potentials at points p and n are V_p and V_n respectively and the EMF $\mathcal{E} = V_p - V_n$*

We know that the line integral of $\vec{E} \cdot \vec{dl}$ along the loop $pabnp$ will be zero

$$0 = \oint_{pabnp} \vec{E} \cdot \vec{dl} = \int_{pabn} \vec{E} \cdot \vec{dl} + \int_{np} \vec{E} \cdot \vec{dl} \tag{8.8}$$

Note that in the external circuit $pabn$, $\vec{E}$ and $\vec{dl}$ and I are in the same direction. Without any magnetic fields, the electric field is conservative and the complete closed loop integral above vanishes. What is called the emf is the integral of $\vec{E} \cdot \vec{dl}$ over the incomplete external circuit which is not a closed loop.

Thus, the integral over the external circuit, i.e., the first term in Eq. (8.8) is simply $V_p - V_n$ where V_p and V_n are the potentials of the positive and negative terminals of the battery.

For the second term in the above expression, $\vec{E}$ and $\vec{dl}$ are in opposite directions since the current flows from the negative to the positive terminal within the battery. This integral is thus, the negative of the work done by the battery, or $-\mathcal{E}$. Hence, we have

$$0 = \oint_{pabnp} \vec{E} \cdot \vec{dl} = -\mathcal{E} + (V_p - V_n) \tag{8.9}$$

or

$$\mathcal{E} = (V_p - V_n) \tag{8.10}$$

It is this $\mathcal{E}$ which is referred to as the EMF.

In the case of a loop in a varying magnetic field, $\vec{E}$ and $\vec{dl}$ are in the same direction and hence the loop integral of $\vec{E}$ over a closed loop is not zero. $\vec{E}$ cannot therefore, be a conservative field in this case and therefore, cannot be defined simply as $\vec{E} = -\vec{\nabla}\phi$ as before.

EXAMPLE 8.1 A conducting bar can slide freely over two conducting rails as shown in Fig. 8.4. Calculate the induced voltage in the bar if

(a) the bar is stationed at $y = 8$ cm and $\vec{B} = 4\cos(10^6 t)\hat{z}$ milli Tesla and
(b) the bar slides with a velocity $\vec{v} = 20\hat{y}$ m/s and $\vec{B} = 4\hat{z}$ milli Tesla.

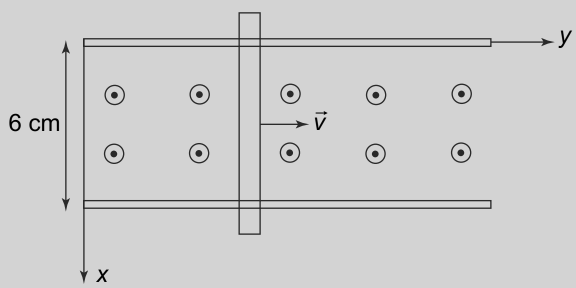


Fig. 8.4 *Example 8.1*

Solution

(a) This is the case where the flux change is being caused by the changing magnetic field. Thus, by Faraday's Law, we have

$$\mathcal{E} = -\int \frac{d\vec{B}}{dt} \cdot \vec{dS} = \int_{y=0}^{0.08} \int_{x=0}^{0.06} 4(10^3)\sin(10^6 t)\,dx\,dy = 19.2\sin(10^6 t) \text{ V}$$

(b) In this case, the EMF is due to the motion of the coil in a constant magnetic field and we have

$$\mathcal{E} = \int (\vec{v} \times \vec{B}) \cdot \vec{dl} = \int_{x=l}^{0} (20\hat{y} \times .004\hat{z}) \cdot dx\hat{x} = -vBl = -4.8 \text{ mV}$$

EXAMPLE 8.2 A rectangular loop of sides of lengths w and l is free to rotate about one of the sides of length w which is along the z-axis and at $t = 0$, it is on the $y - z$ plane as shown in Fig. 8.5. A uniform magnetic field $\vec{B} = B_0\hat{i}$ is present all over. If the loop is rotating with angular frequency ω about the z-axis, find the induced EMF across the loop.

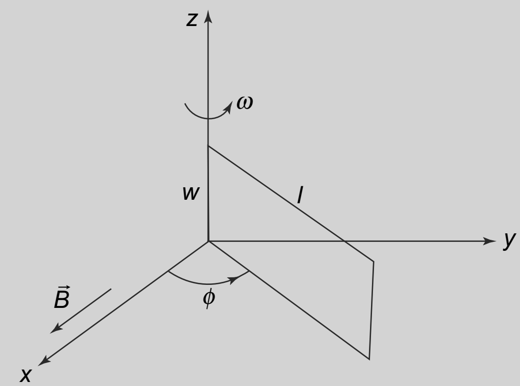


Fig. 8.5 *Example 8.2*

Solution

The induced EMF is motional EMF since the magnetic field is constant. Thus,

$$\mathcal{E} = \int (\vec{v} \times \vec{B}) \cdot \vec{dl}$$

here

$$\vec{dl} = dz\hat{z}$$
$$\vec{v} = \rho \frac{d\phi}{dt}\hat{\phi} = \rho\omega\hat{\phi}$$
$$\rho = l$$

We have used cylindrical polar coordinates above. We need to transform $\vec{B}$ into cylindrical coordinates. Doing this, we get

$$\vec{B} = B_0\hat{x} = B_0(\cos\phi\hat{\rho} - \sin\phi\hat{\phi})$$

and hence

$$\vec{v} \times \vec{B} = -\rho\omega B_0 \cos\phi\hat{z}$$

With this, we can evaluate

$$(\vec{v} \times \vec{B}) \cdot \vec{dl} = -\rho\omega B_0 \cos\phi dz$$

and the induced EMF to be

$$\mathcal{E} = \int_{z=0}^{w} (-l\omega B_0 \cos\phi)dz = -lwB_0 \cos\phi$$

We still need to determine ϕ. To do this, note that

$$\omega = \frac{d\phi}{dt}$$

and so

$$\phi = \omega t + C$$

where C is a constant. We know that at $t = 0$, the loop is in the $y - z$ plane. Thus, $C = \frac{\pi}{2}$. Putting it together, we get

$$\mathcal{E} = -lwB_0 \cos(\omega t + \frac{\pi}{2})$$

PROBLEM 8.3 A conducting rod of length 75 cm is made to slide over two parallel metal rods as shown in Fig. 8.6. The crosses indicate a magnetic field going into the plane of the paper of strength 0.1 T. The two resistors $R_1 = 50$ Ohms and $R_2 = 100$ Ohms are attached as shown. The rod moves with a uniform speed of $v = 10$ m/s. Calculate the current through the two resistors, the power delivered in each resistor and the force required to keep the rod moving a uniform speed.

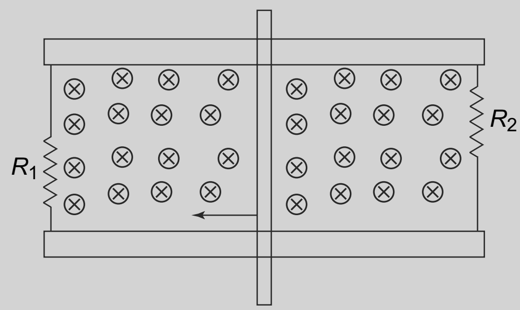


Fig. 8.6 *Problem 8.3*

PROBLEM 8.4 A long straight wire carrying a current I is placed along the y-axis as shown in Fig. 8.7. A square loop of side 2 cm with a resistance of 0.1 Ohms is placed 1 cm away from the wire. The loop moves along the $+x$ direction with a speed $v = 1$ cm/s. Find the current induced in the loop.

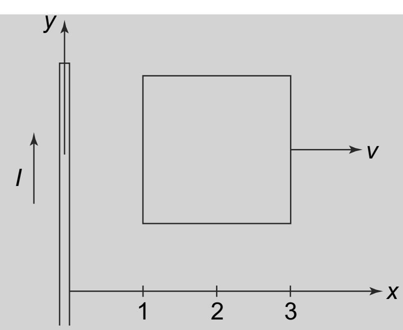


Fig. 8.7 *Problem 8.4*

8.2.4 Flux Linkage and Faraday's law

Faraday's law, as we have seen above, relates the rate of change of magnetic flux linked to a closed loop to the induced EMF in the loop. Clearly then, the concept or definition of the flux linked to a loop is of great relevance and we need to elaborate on this. For a closed loop, there is usually no ambiguity in defining this quantity. For instance, consider a closed, directed loop C, as shown in Fig. 8.8(a).

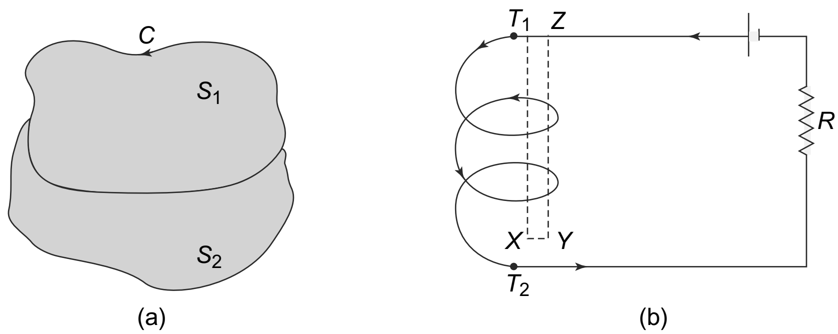


Fig. 8.8 *(a) A closed, directed loop C. S_1 and S_2 are open surfaces both with C as their boundary or edge. Flux across both the surfaces is the same, (b) A solenoidal circuit*

The flux linked with C across surface S_1 can be defined as

$$\Phi = \iint_{S_1} \vec{B} \cdot d\vec{S} \tag{8.11}$$

Here $d\vec{S}$ direction is in the sense defined before. A right-handed screw, rotated along C moves in the direction of $d\vec{S}$. Now suppose we take another surface S_2, which, like S_1 is open but has C as its edge like S_1. Then the flux linked can also be defined as

$$\Phi = \iint_{S_2} \vec{B} \cdot d\vec{S} \tag{8.12}$$

The fact that the two integrals for the flux are equal follows from the fact that magnetic flux lines are always closed. Hence, any flux lines emerging from S_2 must also pass through S_1 since these lines have no sources or sinks. Another way to look at it is that the surface integral of $\vec{B}$ over a **closed** surface vanishes. And since the flux over the closed surface is the sum of the fluxes over the two open surfaces, these fluxes are equal (the sense of $d\vec{S}$ is opposite for the two surfaces and hence the negative sign is taken care of). Equivalently, this surface integral over a closed surface being related to the volume integral of the divergence of $\vec{B}$ through the divergence theorem, allows us to conclude that $\vec{B}$ is divergence free.

The situation is slightly more complicated when one considers a loop that is not closed. For instance, consider a solenoid as in Fig. 8.8(b). The coils, which in reality are tightly bound, are shown separated for clarity. T_1 and T_2 are the two ends of the coil and the solenoid is in the $x - y$ plane. A current driven by a battery flows through the solenoid and the resistance R as shown. The current flows from T_1 to T_2 and the magnetic field due to it is in the z direction as per our discussion of the solenoid magnetic field in Chapter 6.

The flux linked with the solenoid is easiest to calculate if we take the loop to be going from T_1, X, Y, Z, battery, R, T_2, T_1. The part of the surface which is inside the solenoid is the only part through which magnetic field passes since we are considering an infinite solenoid where the field is restricted to be within the solenoid along the axis. In the limit of close packing of the solenoidal coils, the total flux through the solenoid is simply the flux through each coil multiplied by the number of turns in the solenoid.

8.3 LENZ'S LAW

We saw above that Faraday's law gives us the relationship between the EMF induced in a conductor and the rate of change of magnetic flux associated with it. However, the negative sign in Eq. (8.2) is something we need to understand. The negative sign gives us the sense of the induced current and the sign of the induced EMF along a closed loop.

Consider for example, a circular loop as shown in Fig. 8.9(a).

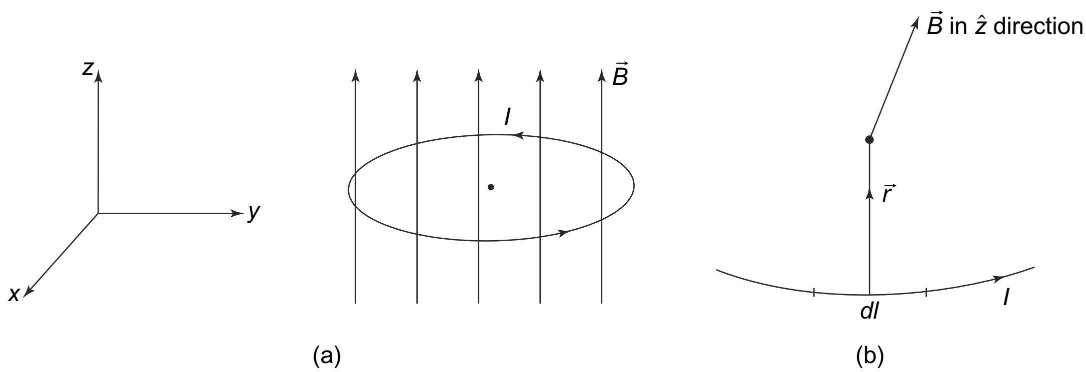


Fig. 8.9 *Lenz's Law (a) A circular loop in the $x - y$ plane. The magnetic field is in the z direction, (b) The direction of the field created by a section dl in the plane of the loop, due to a current I flowing in it*

If the magnetic field decreases, then the induced EMF $\mathcal{E} = \oint \vec{E} \cdot \vec{dl}$ will be positive in the sense of the loop as shown in Fig. 8.9(a). The current I resulting from this EMF will be positive. The magnetic field B_I produced by this induced current will have a direction as shown in Fig. 8.9(b). The magnetic field produced due to the induced current is thus in the direction of $\vec{B}$. It tries therefore to increase the magnetic field (and hence the flux) across the loop when the external magnetic field decreases. In other words, the induced magnetic field tries to oppose the change which caused it in the first place. We can check easily that the same situation is obtained if the external magnetic field is increasing. In that case, the magnetic field due to the induced current will try to weaken the external magnetic field. This property of the induced EMF is called Lenz's Law. Although we have discussed only changes in flux caused by increase or decrease of the magnetic field with time, this Law is valid for any change in flux, whether caused by a change in $\vec{B}$ or a change in the effective area.

> Lenz's Law states that the induced current caused by a change in magnetic flux would always be in a direction such that the flux it produces opposes the change.

As an illustration, consider a rectangular coil in a magnetic field which is coming out of the plane of paper as shown in Fig. 8.10.

The magnetic field decreases with time. According to Faraday's Law, the decreasing magnetic field will cause a change in flux associated with the rectangular loop and hence, induce an EMF and hence, a

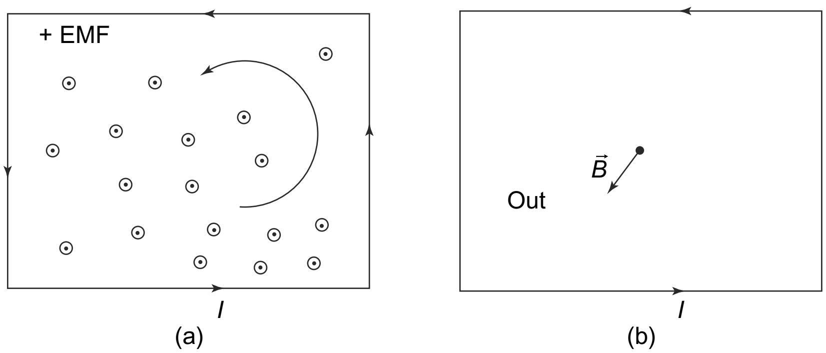


Fig. 8.10 *Lenz's Law in a rectangular field (a) A rectangular coil in a magnetic field coming out of the plane of the paper. The magnetic field is decreasing with time, (b) Direction of the induced current. Inside the loop, the magnetic field due to the induced current is coming out of the paper*

current in the closed loop. The field due to the induced current at points inside the rectangle comes out of the plane of paper. Or, to put it another way, this field due to the induced current tries to increase the flux associated with the rectangular coil since the cause of the induced EMF is the decreasing flux due to the decreasing external magnetic field.

Lenz's Law follows from the Principle of Conservation of Energy. Recall that Lenz's Law states that the direction of the induced current is such that it opposes the very change that causes it. Suppose that was not the case and the induced current actually was in a direction such that it reinforces the change causing it. Then, whatever the change that caused the induced current in the first place (increase or decrease in flux) would get bigger because of the induced current. This in turn, would cause a larger induced current which would further reinforce the change and so on. This process would continue and violate the principle of conservation of energy. To see this, consider the case of a magnet moving towards a coil. The change in the flux associated with the coil, because of the moving magnet would cause an induced current in the coil. If the induced current was reinforcing the change (the motion of the magnet) then the magnet would move faster and faster without any apparent source of energy. This would clearly be violating the Principle of Conservation of Energy.

8.3.1 Illustrations of Lenz's Law

We now consider some simple situations to illustrate Lenz' Law and its relationship with the Principle of Conservation of Energy.

Consider a rectangular loop $ABCD$ with a resistor R as in Fig. 8.11(a). The loop is in the $x - y$ plane and is made to move in the $+x$ direction. There is a magnetic field $\vec{B}$ in the $+z$ direction. The magnetic field increases with increasing x. As the loop moves, the flux associated with it will change since the magnetic field at different locations along the x direction is different. This change in magnetic flux will induce an EMF and a current in the loop. The magnitude of the induced EMF will be given by Faraday's law and the direction by Lenz's Law. The current induced in the loop will cause a generation of Joule heating in the resistor R. Clearly, there is some energy spent which is going into heat energy in the resistor. Given that energy is always conserved, what is the source of this energy?

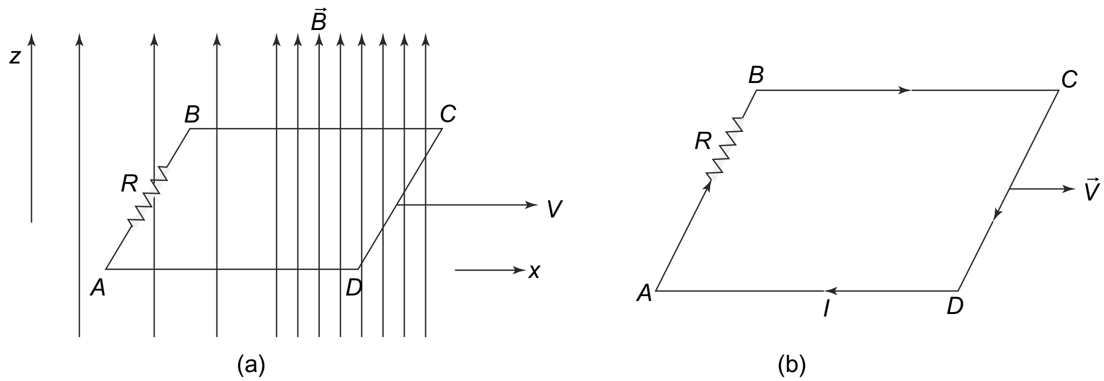


Fig. 8.11 *Rectangular loop in the resistor (a) A rectangular coil ABCD in the $x - y$ plane perpendicular to a magnetic field. The magnetic field increases with increasing x and the coil moves in the $+x$ direction, (b) Direction of the induced current*

Lenz's law will tell us the direction of the EMF (and hence the current through the loop). It is as shown in Fig. 8.11(b). We also know that when we have a current in a conductor in presence of a magnetic field, there is a force experienced by the conductor which is given by

$$d\vec{F} = I(d\vec{l} \times \vec{B})$$

Given the direction of the induced current and the magnetic field, one can easily see that the magnetic force on sides BC and DA are equal and opposite and hence, cancel each other. The magnetic force on CD is in the $-x$ direction while that on AB is in the $+x$ direction. Furthermore, the two are not equal since the magnetic field is increasing in the $+x$ direction. Thus, the force on CD will be larger in magnitude than that on AB. The resultant force is in the $-x$ direction and in the absence of any counteracting force, will make the loop move in this direction. However, in the given case, there is some external force which is making the loop move in the $+x$ direction with a constant speed v. Clearly, to achieve this, the external agency or force needs to do work against the force on the loop which we have seen is in the $-x$ direction. It is the work done by the external agency which is responsible for the energy generated by Joule heating in the resistor. If the external force is removed at some point, the loop will eventually come to a stop and the loss in kinetic energy of the loop will show up as the Joule heating in the resistor.

As another example of the illustration of Lenz's law, consider a metallic ring on top of a ferromagnetic material as shown in Fig. 8.12.

When the current through the coil which is wound around the ferromagnetic material is increased, we see a curious phenomenon. The metallic ring actually jumps in the upwards direction. This is easy to understand using Lenz's Law. The magnetic field caused by the coil around the ferromagnetic material

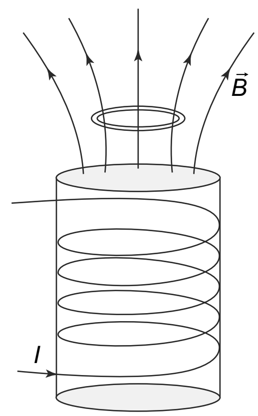


Fig. 8.12 *A block of ferromagnetic material magnetised by a coil wound around it through which a current is passed. The lines of force are as shown. The metallic ring on top jumps up when the current (and hence the $\vec{B}$ field caused by it) is increased in the coil*

decreases with distance in the upward direction. Now when the current through the coil is increased, the magnetic field generated at the location of the metallic ring also increases, thereby causing a change in flux and hence an induced current in the ring. The direction of the induced current is such that it opposes the very change which is causing it. In this case, the change is an increase in flux (caused by an increase in magnetic field which in turn, is caused by an increase in current through the coil). Hence the direction of the induced current is such that it opposes it and wants to decrease it. The only way for the flux through the ring to decrease is for it to move to a region of lower magnetic field (and hence lower magnetic flux) which is in the upward direction.

PROBLEM 8.5 A conducting rod of mass m and resistance r is placed on a rectangular frame of width b. The frame is placed vertically in a magnetic field of magnitude B and direction perpendicular to the plane of the frame. Neglecting friction, find the expression for the terminal velocity of the rod assuming it starts from rest.

PROBLEM 8.6 A helicopter is hovering over earth in a magnetic field of 0.3×10^{-4} T. The rotors of the helicopter are 5 meters long and are made of aluminium. They rotate around the hub, with an angular speed of 10, 000 rpm. Calculate the potential between the end of the rotor and the hub.

8.4 ENERGY IN A MAGNETIC FIELD

In the previous chapters, we have discussed the energy in an electric field. This arises because whenever we separate positive and negative charges from a neutral object, say an atom, then an electric field is created because of the separation. This separation obviously requires work because the unlike charges are attracted to each other by Coulomb force. The work done to bring about this separation is precisely the work stored in the electric field thus created, as we saw earlier.

The situation with magnetic fields is somewhat different because of the different nature of the magnetic force itself. For instance, unlike electrical forces which change the energy of charged particles, magnetic forces do not change the energy of charges. This is because of the nature of the Lorentz force which is perpendicular to the velocity

$$\vec{F} = q(\vec{v} \times \vec{B})$$

Hence, the rate of change in energy

$$\frac{dW}{dt} = \vec{F} \cdot \vec{v} = 0$$

However, in the process of creating a magnetic field, or even changing it, electric fields are induced as per Faraday's Law. These induced electric fields, like all electric fields can of course do work in the presence of currents or charges. Thus, when currents change to cause a changing magnetic field, induced electric fields are created which interact with the currents to do work. At the end of all this, what has finally been created is a magnetic field and just like in the case of an electric field, one can think of the work done (by the induced electric field) to be stored as energy of the magnetic field.

Let us try to calculate what exactly is this work. Consider a current density $\vec{j}(\vec{r})$ in the presence of an induced electric field $\vec{E}$ which itself is caused by the changing magnetic field. To evaluate the amount of work done by the induced electric field, we need to know the charge. dQ, the charge in a volume d^3r travels in a time dt a distance $\vec{d}$ along the direction of current density such that

$$dQ\vec{d} = \vec{j}(\vec{r})d^3r\,dt \tag{8.13}$$

The work done by the induced electric field on this charge is therefore,

$$dW = \vec{E} \cdot \vec{d}\,dQ$$

The total work done is thus,

$$\int dW = \iiint \vec{j}(\vec{r}) \cdot \vec{E}d^3r\,dt \tag{8.14}$$

This work done is the negative of the change in the magnetic energy W_B in time dt and hence, we have

$$\frac{dW_B}{dt} = -\iiint \vec{j}(\vec{r}) \cdot \vec{E}d^3r \tag{8.15}$$

But we know that

$$\vec{\nabla} \times \vec{B} = \mu_0 \vec{j}$$

and so we get

$$\begin{aligned}\frac{dW_B}{dt} &= -\frac{1}{\mu_0}\iiint (\vec{\nabla} \times \vec{B}) \cdot \vec{E}d^3r \\ &= \frac{1}{\mu_0}\iiint \left[\vec{\nabla} \cdot (\vec{E} \times \vec{B})\right] d^3r - \frac{1}{\mu_0}\iiint \vec{B} \cdot (\vec{\nabla} \times \vec{E})d^3r \end{aligned} \tag{8.16}$$

where we have used partial integration. The first term is a volume integral of a divergence and so we can use the divergence theorem to convert it into a surface integral. Remember that the volume integral is over all space and hence the surface integral is over a surface at infinity. We have, therefore,

$$\iiint \left[\vec{\nabla} \cdot (\vec{E} \times \vec{B})\right] d^3r = \iint_S (\vec{E} \times \vec{B}) \cdot d\vec{S} \tag{8.17}$$

where the surface integral is over a surface at infinity. For currents that are finite, the magnetic field at infinity falls as $\frac{1}{r^2}$ and the induced electric field will also fall to zero at infinity. Thus, the surface integral vanishes and so the volume integral vanishes. We therefore, have

$$\begin{aligned} \frac{dW_B}{dt} &= -\frac{1}{\mu_0} \iiint \vec{B} \cdot (\vec{\nabla} \times \vec{E}) d^3r \\ &= \frac{1}{\mu_0} \iiint \vec{B} \cdot (\frac{d\vec{B}}{dt}) d^3r \\ &= \frac{1}{2\mu_0} \iiint \frac{dB^2}{dt} d^3r \end{aligned} \tag{8.18}$$

We see that the energy stored in the magnetic field is therefore,

$$\boxed{E_B = \frac{1}{2\mu_0} \iiint B^2 d^3r} \tag{8.19}$$

Alternatively, we can say that the magnetic field has an energy density of

$$\frac{B^2}{2\mu_0} \tag{8.20}$$

Notice that this is exactly like the energy density expression of the electric field derived in Chapter 2, i.e.,

$$U = \frac{1}{8\pi k} \iiint dV\, E^2(r)$$

The expression for magnetic energy density has been derived using Ampere's law, i.e.,

$$\vec{\nabla} \times \vec{B} = \mu_0 \vec{j}$$

In Chapter 10, where we shall see that Ampere's Law has to be modified. We shall then, derive the expression for magnetic energy using the modified form of Ampere's Law.

As an illustration of this result, let us consider a solenoid with a total of N turns. The cross-sectional area of the solenoid is A and the length of the solenoid L, with L being large. The current in the solenoid is increased from 0 to I_0 and so the final magnetic field inside the solenoid is given by

$$B_0 = \frac{\mu_0 N I_0}{L} \tag{8.21}$$

Now consider any intermediate time between the time when the current is 0 and when it is I_0. At this time t, let the current be $I(t)$. At that time, the flux linked with the solenoid will be

$$\phi(t) = B(t)A = \frac{\mu_0 N I(t) A}{L}$$

As this flux is time-dependent, there will be an EMF induced in the solenoid which is given by

$$\mathcal{E} = -\frac{d\phi(t)}{dt} = -\frac{\mu_0 N A}{L} \frac{dI(t)}{dt}$$

The work done against the back EMF in time dt is therefore, given by

$$\text{Work done} = -\mathcal{E}I(t)dt$$

Thus, the total work done is

$$\begin{aligned} W &= -\int_0^t \mathcal{E}I(t)dt \\ &= \int_0^t \frac{\mu_0 NA}{L}\frac{dI(t)}{dt}I(t)dt \\ &= \frac{\mu_0 NA}{2L}I_0^2 \end{aligned} \tag{8.22}$$

This is the magnitude of the energy stored in the magnetic field too. The total volume of the solenoid is AL. Hence, the energy density in the magnetic field is

$$\begin{aligned} E_B &= \frac{1}{AL}\frac{\mu_0 NA}{2L}I_0^2 \\ &= \frac{B_0^2}{2\mu_0} \end{aligned} \tag{8.23}$$

where we have used the expression for the magnetic field inside the solenoid Eq. (8.21). Thus, we see that for the solenoid, the energy density of the energy stored in the magnetic field is indeed given by Eq. (8.23).

PROBLEM 8.7 An MRI machine has a superconducting magnet in a shape of a solenoid of diameter 0.75 m and length 2 meters. The inside of the solenoid has a uniform magnetic field of 2 T. Calculate the energy density of the magnetic field and the total energy of the solenoid.

8.5 SELF- AND MUTUAL INDUCTANCE

The association of magnetic flux with a circuit and its relationship with the current which produces the magnetic field and hence the flux, allows us to define two circuit parameters which are of immense practical use. Consider two loops C_1 and C_2 as shown in Fig. 8.13.

A current I_1 is passed through C_1 and this creates a magnetic field in all space. Magnetic flux lines thus pass through both C_1 and C_2. Though the actual direction and strength of the magnetic field would depend on the detailed geometry of the loops, we do know that the magnetic field would be proportional to the current I_1. Since the flux through the two loops is proportional to the magnetic field (though, once again, the exact relationship depends on the geometry of the loops, their orientation etc), we can think of a relationship between the flux associated with the two loops and the current I_1.

Let ϕ_{11} and ϕ_{21} be the flux associated with C_1 and C_2, respectively due to the current I_1 in C_1. Then we can write

$$\phi_{11} \propto I_1$$

$$\phi_{21} \propto I_1$$

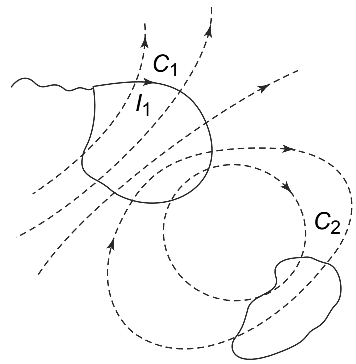


Fig. 8.13 *Mutual and self-inductance: C_1 and C_2 are two loops. A current I_1 passes through C_1 and a magnetic field is created in the whole region. Both the loops have a magnetic flux which is created by this field, associated with them*

or

$$\begin{aligned} \phi_{11} &= L_1 I_1 \\ \phi_{21} &= M_{21} I_1 \end{aligned} \tag{8.24}$$

The constants L_1 and M_{21} in Equation (8.24) are called the **coefficient of self-inductance** of C_1 and the **coefficient of mutual inductance** between C_2 and C_1 respectively.

Notice that the self-inductance L_1 as defined by Eq. (8.24) is necessarily positive. This is because when a current I_1 flows through C_1, the outward flux is defined in exactly the same sense as the magnetic field created by the current. Thus if the current I_1, is positive, then by our convention for outward direction, the flux ϕ_{11} is positive. This is not true for mutual inductance M_{21} since the relative signs of the flux in C_2 and current in C_1 depend on the orientation and geometry of the two loops. **In the SI system of units, the unit of both self- and mutual inductance is Henry.**

We have considered a current flowing in C_1 above. Similarly, a current I_2 flowing in C_2 will cause a flux ϕ_{22} and ϕ_{12} in C_2 and C_1 respectively. We will therefore, have

$$\begin{aligned} \phi_{22} &= L_2 I_2 \\ \phi_{12} &= M_{12} I_2 \end{aligned} \tag{8.25}$$

Once again, the exact values of the coefficients of self- and mutual inductance depend on the geometry and the orientation of the loops. These are not possible to evaluate in general but in certain specific, idealised geometries, it is possible to evaluate the coefficients. However, before we attempt to do this, we prove a remarkable result called the 'Reciprocity relation' between M_{21} and M_{12}.

8.5.1 Reciprocity Relation between Mutual Inductances

The magnetic field, at any point $\vec{r}$, due to a current I_1 in a loop C_1 is given by the Biot–Savart Law as we saw in Chapter 6. The field is

$$\vec{B}_1(\vec{r}) = \left(\frac{\mu_0}{4\pi}\right) I_1 \oint_{C_1} \frac{d\vec{l}_1 \times (\vec{r} - \vec{r}_1)}{|\vec{r} - \vec{r}_1|^3} \tag{8.26}$$

Alternatively, we can find the vector potential $\vec{A}(\vec{r})$ and get the magnetic field from it. The vector potential due to the current I_1 in C_1 would be

$$\vec{A}_1(\vec{r}) = \left(\frac{\mu_0}{4\pi}\right) I_1 \oint_{C_1} \frac{d\vec{l}_1}{\vec{r} - \vec{r}_1} \tag{8.27}$$

We also know that the flux through C_2 due to the magnetic field $\vec{B}_1$ is given by

$$\phi_{21} = \iint_{S_2} \vec{B}_1(\vec{r}_2) \cdot d\vec{S} \tag{8.28}$$

where S_2 is any open surface having C_2 as its boundary. But this flux can be expressed in terms of the vector potential since

$$\vec{B}_1 = \vec{\nabla} \times \vec{A}_1$$

We thus, have

$$\begin{aligned}\phi_{21} &= \iint_{S_2} \vec{B}_1(\vec{r}_2) \cdot d\vec{S} \\ &= \iint_{S_2} (\vec{\nabla} \times \vec{A}_1(\vec{r}_2)) \cdot d\vec{S} \\ &= \oint_{C_2} \vec{A}_1(\vec{r}_2) \cdot d\vec{l}\end{aligned} \tag{8.29}$$

where we have used Stokes' Theorem to convert the surface integral of a curl of a vector into a line integral of the vector. Substituting for $\vec{A}_1$ from Eq. (8.27), we get

$$\begin{aligned}\phi_{21} &= \left(\frac{\mu_0}{4\pi}\right) I_1 \oint_{C_1} d\vec{l}_1 \oint_{C_2} \frac{d\vec{l}_2}{|\vec{r}_2 - \vec{r}_1|} \\ &= M_{21} I_1\end{aligned} \tag{8.30}$$

We thus have a formal expression for M_{21} which is $\left(\frac{\mu_0}{4\pi}\right)$ times a double integral. The integral is completely symmetric betweeen C_1 and C_2. Hence if we do the same exercise for I_2 passing through C_2 to get ϕ_{12}, we will have the same integral. Thus,

$$\boxed{M_{21} = M_{12}} \tag{8.31}$$

EXAMPLE 8.3 Magnetic Monopoles: In nature, magnetic monopoles have not been found. However, there have been experiments done to detect them. If magnetic monopoles did exist, the magnetic field at the point with coordinate $\vec{r}$ created by a monopole of 'magnetic charge' m present at the origin would be

$$\vec{B}(\vec{r}) = \frac{\mu_0}{4\pi} \frac{m}{r^2} \hat{r}$$

If such a monopole from infinity crosses a loop of area A, self-inductance L and of zero resistance, starting from infinity and finally going to infinity, calculate the current that would be established in the loop.

Solution

For the time that the monopole is moving, the magnetic flux Φ through the loop will be changing. At time t, the EMF would be

$$-\frac{d\Phi(t)}{dt}$$

Hence, if $i(t)$ is the current through the loop at time t, then

$$L\frac{di(t)}{dt} = -\frac{d\Phi(t)}{dt}$$

The final current established in the resistanceless loop therefore is

$$\int \frac{di(t)}{dt} = -\frac{1}{L}\Phi$$

where $\Phi = \mu_0 m$, since during the passage of the monopole through the loop from infinity at top side of the loop to infinity at the bottom side, the entire flux passes through the loop.

Such an experiment was actually performed by B.Cabrera in 1982 (**Phys. Rev.Lett., 48,1378(1982)**). His result was positive but not conclusive and there have been no verification of his results since then.

EXAMPLE 8.4 Calculate the induced electric field at a point, distance x from an infinite straight wire which has a current $i(t) = I_0 \sin(\omega t)$ that starts flowing from time $t = 0$.

Solution

Let the current be along the z-axis and the point in the $z - x$ plane at a perpendicular distance x from the wire. The magnetic field due to the current in the wire at a distance from it is (see Example 6.6) is in the y-direction of magnitude, given by

$$B = \frac{\mu_0 I(t)}{2\pi x}$$

By symmetry, since the wire is infinite, the induced electric field at any point is parallel to the wire, i.e., $\vec{E}$ has only a z-component. By the differential form of Faraday's Law

$$[\vec{\nabla} \times \vec{E}]_y = -\frac{\partial E_z}{\partial x} = -\frac{\mu_0 I_0 \omega \cos(\omega t)}{2\pi x}$$

Using the expression for Curl of a vector in Cartesian coordinates, we obtain

$$E_z = \frac{\mu_0 I_0 \omega \cos(\omega t)}{2\pi} \ln(x) + \text{constant}$$

We can determine the value of the constant from the boundary condition given that the current and hence also the induced electric field) is zero for $t < 0$ to get

$$E_z = \frac{\mu_0 I_0 \omega(\cos(\omega t) - 1)}{2\pi} \ln(x)$$

EXAMPLE 8.5 An infinite cylindrical solenoid of radius r and of n turns/m is completely enclosed coaxially by another cylindrical solenoid of radius $R > r$, having N turns/m. Calculate the self-inductance per unit length of the two solenoids and also their mutual inductance per unit length.

Solution

The magnetic field due to either of the solenoids is restricted to the region within the solenoid itself. Let us suppose that the inner solenoid carries a current I_1 and the outer one a current I_2. The value of the magnetic field in the inner solenoid is $\mu_0 n I_1$ and due to the outer one is $\mu_0 N I_2$. Hence, in the notation of Eqs. (8.24 & 8.25), calling d the length of the solenoids

$$\phi_{11} = (\mu_0 n I_1) d n \pi r^2$$
$$\phi_{22} = (\mu_0 N I_2) d N \pi R^2$$
$$\phi_{21} = (\mu_0 n I_1) d N \pi r^2$$
$$\phi_{12} = (\mu_0 N I_2) d n \pi r^2$$

Hence, we have, for the inner solenoid

$$L_{\text{inner}}/d = (\mu_0 n^2) \pi r^2$$

and for the outer solenoid

$$L_{\text{outer}}/d = (\mu_0 N^2) \pi R^2$$

For the mutual inductance per unit length

$$M/d = (\mu_0 n N) \pi r^2$$

PROBLEM 8.8 A short coil, of radius $R = 20$ cm has 20 turns. It surrounds a long solenoid with radius $r = 15$ cm and containing 100 turns having a resistance of 5 Ohms. The current in the short solenoid is increased at a constant rate from zero to 5 A in 10 seconds. Calculate the induced current in the long solenoid.

PROBLEM 8.9 A small circular coil of radius r is totally inside a much larger concentric circular coil of radius $R \gg r$ and the two are coplanar. Derive an approximate expression for their mutual inductance.

PROBLEM 8.10 Two circular concentric and coplanar coils of radius r_1 and r_2 lie with the distance between their centres $d \gg r_1$ and $d \gg r_2$. Derive an approximate expression for their mutual inductance.

8.5.2 Inductance and Magnetic Energy

The energy stored in the magnetic field for a coil, as in the solenoid case, can also be expressed in terms of the coefficients of self- and mutual inductance.

Let us first consider the case of a current being built up in a coil because of an external energy source like a battery. Let the current go from its initial value 0 to its final value I_0. At some intermediate time t, the current in the coil is $I(t)$. At that instant, the induced EMF in the coil is given by

$$\mathcal{E} = -\frac{d\phi}{dt} = -L\frac{dI(t)}{dt}$$

The battery has to force the current against this induced EMF since the induced EMF is, according to Lenz's Law opposing the change which in this case is the increase in the current (and hence the magnetic field and flux). The rate of doing work is thus,

$$\frac{dW}{dt} = LI(t)\frac{dI(t)}{dt} \tag{8.32}$$

and therefore, the total work in building the current from 0 to I_0 is given by

$$\boxed{W = \int_0^{I_0} LI(t)\frac{dI(t)}{dt}dt = \frac{1}{2}LI_0^2} \tag{8.33}$$

This is the energy stored in the magnetic field of the coil. Since L is always positive, this is a positive energy.

The reciprocity relation above can also be seen as a consequence of energy considerations. To see this, consider two coils with coefficient of self-inductance L_1 and L_2 respectively and in which the current builds up from 0 to I_1 and I_2 respectively. First, let us take the current in coil 2 to be zero and build up the current in the first coil. This requires an amount of work given by Eq. (8.34) or the energy stored in coil 1 is

$$E_1 = \frac{1}{2}L_1 I_1^2 \tag{8.34}$$

Now, with the current in coil 1 fully established at its final value I_1, we build up the current in coil 2 from 0 to I_2. At any intermediate stage, the current in coil 2 is $I_2(t)$. At this point, the induced EMF in coil 2 is $\mathcal{E}_2 = L_2 dI_2(t)/dt$. This is not all since the change the current in coil 2 also induces an EMF of $\mathcal{E}_1 = M_{21}dI_2/dt$ in coil 1. The rate of work done by the source of energy (battery) in establishing the current in coil 2 is thus

$$\frac{dW}{dt} = L_2 I_2 \frac{dI_2(t)}{dt} + M_{21} I_1 \frac{dI_2(t)}{dt} \tag{8.35}$$

The total work done by the source, including the work done to establish I_1 in coil 1 is therefore

$$\begin{aligned} W_1 &= E_1 + \int_0^{I_2} L_2 I_2 \frac{dI_2(t)}{dt}dt + \int_0^{I_2} M_{21} I_1 \frac{dI_2(t)}{dt}dt \\ &= E_1 + E_2 + M_{21} I_1 I_2 \end{aligned} \tag{8.36}$$

where

$$E_2 = \frac{1}{2}L_2 I_2^2$$

is the energy stored in coil 2 if there was no coil 1.

Now suppose we reverse the process that is, build up the current in coil 2 first from 0 and then once it reaches I_2 build up the current in coil 1 to I_1. The work done or the energy stored then will be

$$W_2 = E_1 + E_2 + M_{12} I_1 I_2 \tag{8.37}$$

The total energy stored in the arrangement should of course be independent of the order in which the currents are built up and so we get the reciprocity relation

$$M_{12} = M_{21}$$

The total energy stored in the coils, $W_1 = W_2$ should be obviously positive.

$$W = W_1 = W_2 = \frac{1}{2}L_1 I_1^2 + \frac{1}{2}L_2 I_2^2 + M I_1 I_2 \geq 0 \tag{8.38}$$

We know already that L_1 and L_2 are always positive. However $M_{21} = M_{12} = M$ can be positive or negative depending on the orientation and the geometry of the arrangement. This gives us a relationship between the self- and mutual inductances. Let us write W in Equation (8.38) as

$$W = \frac{1}{2}\left(\sqrt{L_1}I_1 \pm \sqrt{L_2}I_2\right)^2 + \left(M \mp \sqrt{L_1 L_2}\right) I_1 I_2 \tag{8.39}$$

If we take the upper signs, the first term is positive definite being a square or zero (when $\sqrt{L_1}I_1 = -\sqrt{L_2}I_2$). In the case where the first term is zero, I_1 and I_2 have opposite signs. Thus, to keep W positive, $(M - \sqrt{L_1 L_2})$ must be negative. Thus, we have

$$M < \sqrt{L_1 L_2}$$

Similarly, if we take the lower sign, we will get

$$M > -\sqrt{L_1 L_2}$$

This allows us to define a parameter to measure the coupling between the two coils. k, **the coefficient of coupling** is defined as

$$k = \frac{|M|}{\sqrt{L_1 L_2}}$$

Clearly, from the restrictions on M, $-1 \leq k \leq 1$. The coefficient k can be thought of as measuring the strength of flux linkage between the two coils. For $k = 0$, there would be no flux linkage. An example of this would be two infinite solenoids placed side by side. The infinite solenoid has its field confined to inside the solenoid and hence, there is no flux linkage between the two solenoids.

On the other hand, $k = 1$ would signify complete linkage of the flux between two coils. An example of this could be two identical finite solenoids placed end on end with their axis aligned and no gap between the solenoids. In this case, all the flux of one solenoid passes through the second one and there is complete linkage.

8.6 CIRCUITS WITH INDUCTANCES

In Chapter 5, we saw that introducing a capacitor in a circuit with a battery and a resistance leads to a current which is time-dependent. We also noted that it is possible to introduce a capacitor in more complicated circuits and use the network theorems to analyse them. In this chapter, we have come across a new circuit element called an inductance. What are the effects on the current when one introduces an inductance L in a circuit with a battery of EMF $\mathcal{E}$ and a resistor R? This is what we will study now.

8.6.1 Ideal Inductance

Before we attempt to study circuits with inductances, we need to be sure that our usual methods of analysis, namely where we use electric potentials in loops etc (Ohms's Law and Kirchhoff's Laws) is valid. We have already seen above that the line integral over a closed loop,

$$\oint \vec{E} \cdot \vec{dl} \neq 0$$

whenever a time-varying magnetic field is present. Thus, in the presence of a time varying magnetic field, it is not possible to write $\vec{E} = -\vec{\nabla}\phi$ where ϕ is the electric potential since the electric field is non-conservative. We had met a similar situation in Chapter 5 when we introduced capacitors into circuits. There, we had introduced the concept of 'ideal capacitors' or those for whom the electric field is totally confined to the region in between the plates carrying the charge. We can do a similar idealisation for inductors.

An 'ideal inductor' is defined as one in which the magnetic field generated by the action of currents is restricted to the inside of the inductor and a small region outside it. A closely wound toroidal coil, with a little gap for the leads carrying the current is possibly the closest approximation to it. Now since in an 'ideal inductor' the magnetic field is confined to inside the inductor (or a small region outside it) we can talk meaningfully about the electric potential outside the inductor, since everywhere else the electric field is conservative. This would then allow us to use all the methods and analysis of circuit analysis to analyse circuits with inductors. In addition to this, we also make the assumption that an ideal inductor has zero resistance to currents. This, as we shall see, simplifies analysis of circuits considerably.

As an example, consider the circuit in Fig. 8.14.

The dashed lines define a box which serves as an ideal inductor and the two points 1 and 2 act as the ends of the ideal inductor. The magnetic field is confined to within this box and so outside of this, one can use electric potential and carry out the analysis.

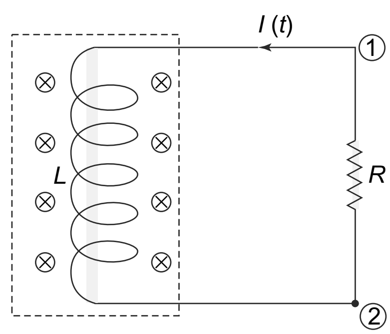


Fig. 8.14 *Ideal Inductor*

In this circuit, let the time-dependent current through L be $I(t)$ and the flux due to this be $\Phi(t)$. Then taking a loop which goes from 1 to 2 through L and then back from 2 to 1 through the resistance R, we have

$$\text{EMF} = \oint \vec{E} \cdot d\vec{l} = -\frac{d\Phi(t)}{dt} = -L\frac{dI(t)}{dt} \tag{8.40}$$

However, EMF is simply the amount of work done carrying a unit charge around the loop. The inductor is an ideal one and hence offers zero resistance and no work is done in moving a charge through it. Moreover, the points 1 and 2 are outside the ideal inductor and hence there is no magnetic field there. Thus, we can define them to be at potentials V_1 and V_2. The work done in carrying a unit positive charge from 2 to 1 is by definition

$$W = V_2 - V_1$$

Therefore, we have

$$0 + V_2 - V_1 = -L\frac{dI(t)}{dt} \tag{8.41}$$

or

$$V_1 - V_2 = L\frac{dI(t)}{dt}$$

8.6.2 LR Circuits

As already remarked, introduction of capacitors in purely resistive circuits had the effect of introducing a time-dependence (which is characteristic of the circuit depending on the values of C and R) in potential differences and currents. A similar thing happens when one introduces an inductance in a purely resistive circuit carrying a direct or steady current.

Consider the circuit in Fig. 8.15

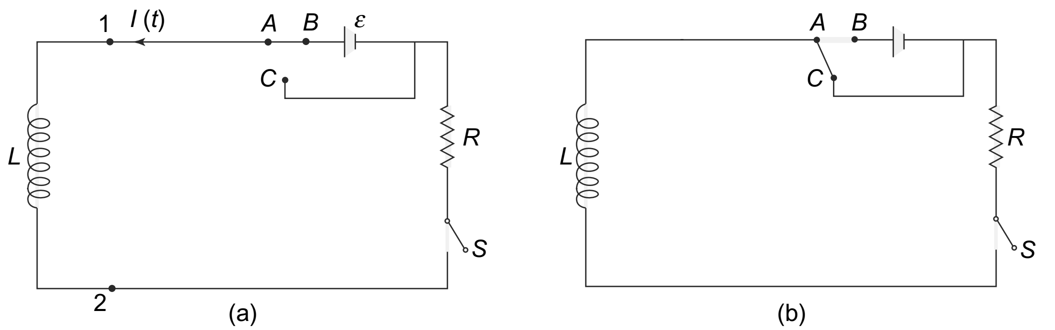


Fig. 8.15 *LR circuit: (a) An LR circuit. The switch S is closed. The two way switch ABC is connected in such a way that the battery is in the circuit, (b) The two way circuit ABC is now connected in such a way that the battery is not included in the circuit*

An ideal inductor of inductance L is connected to a resistance R and a battery of EMF $\mathcal{E}$ which has zero internal resistance (Fig. 8.15(a)). The two way switch ABC is connected as AB so that the battery is in the circuit. As we have already seen, the potential difference between points 1 and 2 is given by

$$V_1 - V_2 = L\frac{dI(t)}{dt}$$

and hence, for the loop containing the inductance, battery and resistor, we have

$$L\frac{dI(t)}{dt} + IR = \mathcal{E} \tag{8.42}$$

This is a first order differential equation which can be easily solved. Writing it as

$$\frac{d}{dt}\left(I(t)e^{(Rt/L)}\right) = \frac{\mathcal{E}}{R}e^{(Rt/L)} \tag{8.43}$$

and integrating, we get

$$I(t)e^{(Rt/L)} = \frac{\mathcal{E}}{L}\frac{e^{(Rt/L)}}{R/L} + C \tag{8.44}$$

where C is an integration constant. At time $t = 0$, the current is zero. Hence, we get

$$C = -\frac{\mathcal{E}}{R}$$

Putting it in Eq. (8.44), we have for the current in the resistor $I(t)$,

$$\boxed{I(t) = \frac{\mathcal{E}}{R}\left(1 - e^{-(Rt/L)}\right)} \tag{8.45}$$

Notice that as $t \to \infty$, the current goes to $I(t) \to I_0 = \frac{\mathcal{E}}{R}$. This is the value of the current if there was no inductance in the circuit. We can understand this on the basis of Lenz's Law. When the current is growing from an initial value $I(0) = 0$, the induced EMF in the inductance is in such a direction so as to oppose this growth. The current thus, instead of approaching this value instantaneously (as it would if there was no inductance), approaches it asymptotically. The constant $\tau = \frac{L}{R}$ is called the **time constant of the LR circuit**. In time $t = \tau$, the current has reached $(1 - \frac{1}{e})$ of its final, asymptotic value given by $\frac{\mathcal{E}}{R}$.

Now suppose that after a long time when the current has reached its asymptotic value, we change the two-way switch ABC from its position AB to AC. Of course, the current as we saw above, only becomes I_0, its asymptotic value at $t = \infty$. However, given the exponential dependence on time, we can define long time as a time much larger than the time constant for the circuit. From Eq. (8.45) it is clear that the current would then be very close to I_0. When we switch the two-way key to the AC position, the battery is no longer in the circuit. The circuit at this instant, say $t = 0$ has a current I_0. In a similar way to the charging circuit, we can write down the potential equation as

$$L\frac{dI(t)}{dt} + IR = 0 \tag{8.46}$$

since there is no EMF in the circuit now. This equation is easily integrated to get

$$I(t) = C_1 e^{(-Rt/L)} \tag{8.47}$$

where C_1 is an integration constant. At time $t = 0$, the current in the circuit is, as we have seen, I_0. Thus,

$$C_1 = I_0 = \frac{\mathcal{E}}{R}$$

The current thus, is given by

$$\boxed{I(t) = \frac{\mathcal{E}}{R}e^{(-Rt/L)}} \tag{8.48}$$

We see that once again, even after the battery is removed, the current does not immediately go to zero. Instead, it falls exponentially and only asymptotically reaches zero. We can understand this too using Lenz's Law. If there was no inductance in the circuit, the current would of course have fallen to zero as soon as we removed the battery. With the inductance, when we remove the battery, the current starts falling and hence this produces an induced EMF in the inductance. The induced EMF, according to Lenz's Law opposes the change and hence the current does not immediately go to zero. Instead, it approaches zero asymptotically. Once again, the time constant $\tau = \frac{L}{R}$ is a measure of the time taken for the current to go to zero. In time $t = \tau$, the current reaches $\frac{1}{e}$ of is initial value I_0. Clearly, if L is large, this time would be large as the induced EMF effect would be large.

We can plot $I(t)$ as a function of time for both the cases as in Fig. 8.16.

8.6.3 Energy Considerations in LR Circuit

In a purely resistive circuit, that is, one with a battery of EMF $\mathcal{E}$ and a resistance R, the battery supplies the energy which gets dissipated in the resistor as Joule heat. Suppose the current is I_0. Then we have

$$\mathcal{E}I_0 = I_0^2 R$$

Things are different of course, when an inductance L is present in the circuit. In this case, we know that the inductance stores magnetic energy which also has to be supplied by the battery, since there is no

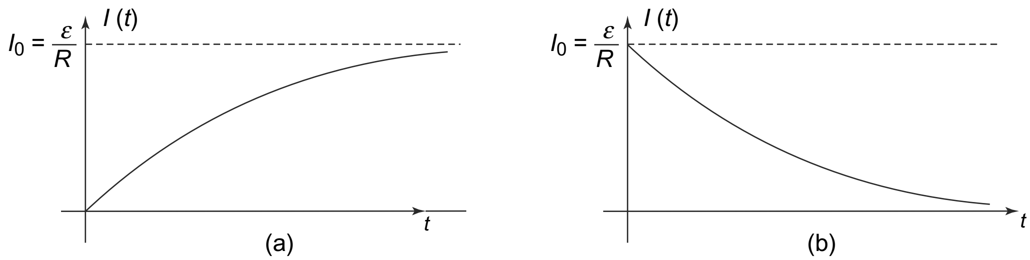


Fig. 8.16 *Current vs time for an LR circuit: (a) Current vs time for an LR circuit when the circuit after the circuit is switched on with the battery in the circuit, (b) Current vs time for an LR circuit with the battery removed at time $t = 0$ when the asymptotic current I_0 is flowing in the circuit*

other source of power in the circuit. Let us first consider the rise of the current case when the current goes from $I(0) = 0$ to some value $I(T)$ at some time $t = T$.

The Joule heat dissipated in the resistor can be found as

$$\begin{aligned} W_J &= \int_0^T I^2(t) R dt \\ &= \int_0^T \left[\frac{\mathcal{E}^2}{R} \left(1 - 2e^{(-Rt/L)} + e^{-2Rt/L}\right) \right] dt \\ &= \frac{\mathcal{E}^2}{R} \left[t + \frac{2L}{R} e^{(-Rt/L)} - \frac{L}{2R} e^{(-2Rt/L)} \right]_0^T \\ &= \frac{\mathcal{E}^2}{R} \left[T + \frac{2L}{R}(e^{(-RT/L)} - 1) - \frac{L}{2R}(e^{(-2RT/L)} - 1) \right] \end{aligned} \tag{8.49}$$

The total energy stored in the inductor is easily calculated as

$$\begin{aligned} W_L &= \frac{1}{2} L [I(T)]^2 \\ &= \frac{1}{2} L \left[\frac{\mathcal{E}^2}{R^2} \left(1 - 2e^{(-RT/L)} + e^{-2RT/L}\right) \right] \\ &= \frac{\mathcal{E}^2}{R} \left[\frac{L}{2R} - \frac{L}{R} e^{(-RT/L)} + \frac{L}{2R} e^{(-2RT/L)} \right] \end{aligned} \tag{8.50}$$

The total energy therefore, spent in this circuit to bring the current from 0 to $I(T)$ is thus, $W_J + W_L$.

$$\begin{aligned} W &= W_J + W_L \\ &= \frac{\mathcal{E}^2}{R} \left[T - \frac{L}{R}(1 - e^{(-RT/L)}) \right] \end{aligned} \tag{8.51}$$

We can easily check that this is indeed the work done or the energy supplied by the battery. Thus, the energy supplied by the battery, W_B in this time is

$$W_B = \int_0^T \mathcal{E} I(t) dt$$

$$= \frac{\mathcal{E}^2}{R} \int_0^T \left[1 - e^{(-Rt/L)}\right] dt$$

$$= \frac{\mathcal{E}^2}{R} \left[t + \frac{L}{R} e^{(-Rt/L)}\right]_0^T$$

$$= \frac{\mathcal{E}^2}{R} \left[T - \frac{L}{R}(1 - e^{(-RT/L)})\right] \tag{8.52}$$

Clearly, we can see that

$$W_B = W_J + W_L$$

as we expect.

Similarly, we can check that energy is conserved when the current falls from the maximum I_0 (at time $t = 0$) to some value $I(T)$ at time $t = T$. There is no battery to supply the energy now. However, the inductor, as we have seen, stores energy in its magnetic field. This is the energy which is the source of the energy dissipated as Joule heat in the resistor.

To see this, recall that the current, with the battery removed has a time-dependence

$$I(t) = \frac{\mathcal{E}}{R} e^{(-Rt/L)}$$

Thus, the Joule heat dissipated in the resistor in time T is given by

$$W_J = \int_0^T I^2(t) R dt$$

$$= \int_0^T \left[\frac{\mathcal{E}^2}{R} e^{(-2Rt/L)}\right] dt$$

$$= \frac{\mathcal{E}^2}{R} \frac{L}{2R} \left(1 - e^{(-2RT/L)}\right) \tag{8.53}$$

This energy, as we have noted, has to come from the decrease in the magnetic energy stored in the inductor. At time $t = 0$, the energy stored in the inductor is $\frac{1}{2} L I_0^2$ since the current is I_0. At time $t = T$, the energy stored is $\frac{1}{2} L I(T)^2$. The difference in these energies is thus,

$$W_L = \frac{1}{2} L \left(I_0^2 - I(T)^2\right)$$

$$= \frac{1}{2} L \frac{\mathcal{E}^2}{R^2} \left[1 - e^{(-2RT/L)}\right]$$

$$= W_J \tag{8.54}$$

Thus, in both cases, we see that the energy is conserved and the total balance of energy is maintained. The ultimate source of energy is of course the battery (the chemical energy in which is converted to electrical energy). This energy, during the growth phase in an LR circuit, is partly dissipated as Joule heat in the resistor and partly gets stored as magnetic energy in the inductor. When the battery is removed, at any instant the decrease in the magnetic energy of the inductor shows up as the Joule heat in the resistor.

EXAMPLE 8.6 A particle of charge $+e$ is moving with a speed $\frac{c}{5}$. Compare the densities of magnetic and electric field energy created by it at a distance of 1 m from the particle at right angles to the direction of motion.

Solution

Let the particle be at the origin and direction of motion of the particle be along the z-direction. The magnitude of the electric and magnetic fields created by the charged particle at a point P with polar coordinates $(1, \theta, \phi)$ are

$$E = \frac{1}{4\pi\varepsilon_0} e$$

and

$$B = \frac{\mu_0}{4\pi} \frac{ec\sin(\theta)}{5}$$

The energy density of the magnetic field at P is thus,

$$W_M = \frac{\mu_0}{32\pi^2} \frac{e^2c^2}{25} = 3.67 \times 10^{-31} \text{ J/m}^3$$

The energy density of the electric field at P is

$$W_E = \frac{1}{32\varepsilon_0\pi^2} e^2 = 9.16 \times 10^{-30} \text{ J/m}^3$$

EXAMPLE 8.7 In a LR circuit (Fig. 8.15), the values of L, R and E are, $L = 1$ millihenry, $R = 0.02$ ohm, $E = 6$ volts and there is no internal resistance of the battery. The battery is connected at time $t = 0$. At time T the current reaches 50% of its terminal value. Calculate at time T, the amount of energy stored in the inductor and the amount of energy produced as Joule heat.

Solution

For the values of L and R, the time constant is

$$\tau = \frac{L}{R} = 0.05 \text{ sec}$$

Since the current grows to half its terminal value at time $t = T$, we have

$$\exp(-T/0.05) = 0.5$$

and the current at time T

$$i(T) = \frac{E}{2R} = 150 \text{ A}$$

The energy stored in L is therefore,

$$\frac{L(i(T))^2}{2} = 11.25 \text{ J}$$

The Joule heat produced till time T is

$$W_J = 180 \times [0.05 + 0.1 \times (0.5 - 1) - 0.025 \times (0.25 - 1)] = 3.375 \text{ J}$$

PROBLEM 8.11 There are two stationary loops with mutual inductance M. The current in one loop is varied as $I_1(t) = at$ where a is a constant. Find the time dependence of the current $I_2(t)$ in the second loop whose self-inductance is L_2 and resistance is R.

PROBLEM 8.12 Current in an LR circuit, with $R = 5$ Ohms and $L = .5$ H, is increasing at a rate of 4 A/s. Calculate the potential difference across the circuit when the current is 2 A.

PROBLEM 8.13 Consider a square loop of side a placed at a distance of r from an infinitely long current carrying wire. Show that the mutual inductance between the two will be

$$M_{12} = \frac{\mu_0 a}{2\pi} \ln\left[\frac{a+r}{r}\right]$$

PROBLEM 8.14 Consider a coaxial cable with an inner conductor of radius a and outer conductor of radius b. The outer conductor has negligible thickness and the current is uniformly distributed inside the inner conductor. Show that the self-inductance per unit length is

$$L = \frac{\mu_0}{2\pi}\left[\frac{1}{4} + \ln\left(\frac{b}{a}\right)\right]$$

Induction Coil

An interesting application of the phenomenon of inductance is the induction coil. This is a device which is very useful for producing very high voltage pulses. Apart from the laboratory, it forms an essential part of an internal combustion engines where it is used to produce the spark to ignite the petrol and air mixture.

The simplest induction coil consists of a primary and secondary coils, tightly wound around a laminated ferromagnetic core (Fig. 8.17). The primary coil has very few turns, of the order of about a dozen while the secondary coil has many thousands of turns. The primary coil is connected to a battery through a circuit called an interrupter circuit. The interrupter circuit is basically a ‘make-and-break’ circuit which has an iron head S in the figure, placed close to but not touching the laminated core. The iron head is attached to a flexible strip which allows it to move. When it is in the position as shown in the figure, the current flows through the primary. When it is attracted to the core, the circuit breaks and no current flows in the primary. Again when the current stops flowing in the primary, the magnetisation of the core goes to zero and the iron head reverts back to its original position, making the circuit again and so on. This process allows the primary circuit to be made and be broken repeatedly.

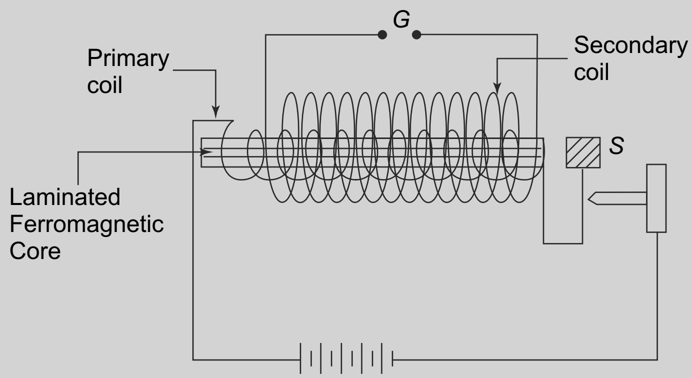


Fig. 8.17 *Induction coil*

As long as a steady current is flowing in the primary, there is no change in the magnetic flux and hence no induced EMF in the secondary coil. But suppose now that the primary circuit is suddenly made or broken, the current changes from zero to its maximum value or from the maximum value to zero abruptly. This sudden change in the current in the primary, induces an EMF in the secondary coil. Since the coupling between the two coils is almost complete (because they are tightly wound and also because of the ferromagnetic core) and the number of turns in the secondary is very large, the EMF induced in the secondary can be very high. If we create an air gap in the secondary circuit, the EMF can be large enough for the breakdown to be caused and a spark will occur in the gap. This is what happens in the spark plug of an automobile.

As already stated, the interrupter circuit is responsible for the repeated making and breaking of the primary circuit and thereby causing an induced EMF in the secondary circuit. Though there is, in both cases, an induced EMF generated in the secondary circuit, it turns out that the break in the circuit happens much more abruptly and suddenly as compared to the make in the circuit. This is because, as we have seen, when a voltage is introduced in a circuit with an inductor and a resistance (in this case, the inductance and resistance of the primary coil), the current builds up only over time and not suddenly. Since the induced EMF is proportional to the rate of change of flux linked with a circuit, the EMF induced is much smaller than the break part of the cycle. The secondary coil therefore has a very high induced voltage across it during the break part of the cycle and this can be used to generate a spark in the air gap.

As an interesting historical fact, Hienrich Hertz in 1887 designed an apparatus to generate electromagnetic waves which, as we shall see in a later chapter, were predicted by Maxwell as a consequence of a unified theory of electromagnetism. Hertz's experiments used an induction coil to generate a spark and the production of electromagnetic waves. This is not surprising—one might have noticed the disturbance in a car radio for instance when a motorcycle passes by. The disturbance is caused by the radio waves being produced in the spark plugs of the motorcycle engine.

8.7 ADVANCED TOPIC

8.7.1 Induced Current and Faraday's Law

We have seen that Faraday's Law in the differential form, namely

$$\vec{\nabla} \times \vec{E} = -\frac{d\vec{B}}{dt} \tag{8.55}$$

gives us the direction of the curl of the induced electric field in the conductor. This is of course, valid whether there is a current flowing in the conductor or not. However, knowing the direction of $\vec{E}$ does not determine uniquely the direction of the induced current. In addition, knowing the curl of $\vec{E}$ does not determine $\vec{E}$ uniquely since one can always add a gradient of a scalar function to $\vec{E}$ without changing curl of $\vec{E}$. Both these facts become relevant when there is a conductor through which the change of magnetic flux can produce a current, as we now show.

Consider a circular region in the $x - y$ plane in a time dependent magnetic field $\vec{B}(t)$. Let the magnetic field vary with time at a constant rate, i.e.,

$$\vec{B}(t) = bt \tag{8.56}$$

where b is a constant. The field is into the plane of the paper as shown in Fig. 8.18(a).

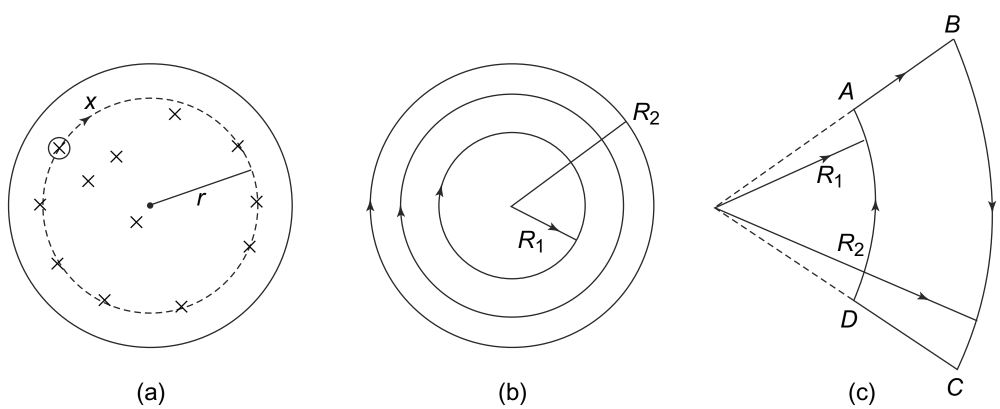


Fig. 8.18 *A circular region with a changing magnetic flux due to a time dependent magnetic field: (a) The sense of the induced electric field in the azimuthal direction, (b) Region bounded by two arcs of radius R_1 and R_2, (c) A conducting wire placed along the loop $ABCD$*

Obviously, there is complete azimuthal symmetry in the problem. The azimuthal electric field at a distance r from the centre can be easily calculated by using a circle of radius r as the path in the line integral for $\vec{E}$ since $\vec{E}$ is in the azimuthal direction and is constant at a given r. We therefore, get

$$\oint \vec{E} \cdot \vec{dl} = 2\pi r E_\phi(r) = -\pi r^2 b \tag{8.57}$$

or

$$E_\phi(r) = -\frac{br}{2} \tag{8.58}$$

Now suppose we place a circular conductor in the circular region, then an induced current will flow through the conductor in the same sense as E_ϕ. This is true, no matter what is the radius of the circular conductor.

But suppose we now place a conducting wire in the form of a loop $ABCD$ as in Fig. 8.18(c). A very curious thing happens now. In the outer arc, the sense of the induced current is in the same direction as the induced electric field but in the inner arc, the sense of the induced current is opposite to that of E_ϕ! What is more, there would also be a current in the radial sections of the loop where there should have been no induced electric field to cause it.

The reason for this curious phenomenon is actually simple. In general, when an electric field is applied to a conductor, and current flows as a result, surface charges develop on the conductor. These surface charges alter the electric field inside the conductor and at equilibrium, the current in the conductor is in the same direction as the applied electric field. The electric field created by the surface charges is conservative and so the resultant combined electric field (external plus that due to the surface charges) still obeys the flux rule and also Eq. (8.55). However, the actual value of the electric field cannot be obtained from the solution of the equation which satisfies the boundary conditions in the presence of a conductor. Thus, while the line integral over a closed loop of the combined electric field is determined by the flux rule (Eq. (8.55)), the actual value of the electric field at every point inside the conducting loop is not determined by it.

SUMMARY

- Faraday's experiments show that whenever there is a change in flux associated with a conductor, an EMF is induced in the circuit.
- The change in flux can be due to change of magnetic field or area or both.
- Faraday's Law tells us that the strength of the induced EMF in a circuit is equal to the rate of change of magnetic flux associated with the circuit.
- Lenz's Law, which is simply a restatement of conservation of energy gives us the sign of the induced EMF as that which always opposes the change causing it.
- Faraday's Law in differential form relates the curl of the electric field to the negative of the rate of change of the magnetic field.
- The magnetic field, like the electric field, has an energy density.
- The relationship between magnetic flux associated with a circuit and the current allows us to define geometric quantities called coefficients of self- and mutual inductance.
- The energy stored in the magnetic field for a coil or a circuit can be written in terms of the coefficients of inductance.

CONCEPTUAL QUESTIONS

1. The magnetic field through a wire loop is pointed upwards and increasing with time. The induced current in the coil is

   a. clockwise if seen from the top

b. counterclockwise when seen from the top

2. A rectangular wire loop is pulled through a uniform B field penetrating its top half. The induced current and the force and torque on the loop are:

a. Current clockwise (CW), Force Left, No Torque
b. Current CW, no force, torque rotates counter clockwise (CCW)
c. Current CCW, force left, no torque
d. Current CCW, no force, torque rotates CCW
e. No current, force or torque.

3. A circuit in the form of a rectangular piece of wire is pulled away from a long wire carrying current I. The induced current in the rectangular circuit is

a. clockwise
b. counterclockwise
c. There is no induced current

4. When you insert the iron core in an inductance coil what happens?

a. B Increases so L does too
b. B Decreases so L does too
c. B Increases so L decreases
d. B Decreases so L increases

5. Two conducting coils which are identical in all respects, except coil 2 has a gap in it or is split along its circumference, are placed in a uniform magnetic field that decreases at some constant rate. If the plane of the coils is perpendicular to the field, which of the following statements, if any is true?

a. An EMF is induced in both coils
b. An EMF is induced only in the split coil
c. Equal joule heating occurs in both coils
d. Joule heating does not occur in either coil

6. The flux through each turn of a 100 turn coil is $(t^3 - 2t)$ milli Wb where t is in seconds. The induced EMF at $t = 2$ seconds is

a. 1 V
b. $-$ 1 V
c. 4 mV
d. 0.4 V

7. A very long solenoid consisting of N turns has radius R and length $d(d \gg R)$. Suppose the number of turns is halved keeping all the other parameters fixed. The self-inductance

a. remains the same
b. doubles in value
c. halves in value
d. becomes four times as small
e. becomes four times as large

8. An aluminium disc is dropped vertically through a region with a uniform magnetic field. The velocity of the disc
   a. remains the same as with no magnetic field
   b. increases as it passes through the magnetic field
   c. decreases as it passes through the magnetic field
   d. Cannot say from the information given
9. A 1000 turn closely and tightly wound toroid has an inductance of 20 mH when the current is 2.5 A. Find the magnetic flux inside the toroid.
10. Two coupled coils, of self-inductance 20 mH and 80 mH have a mutual inductance of 16 mH. Find the coefficient of coupling between the coils.

PROBLEMS

1. A copper strip of length L is pivoted on one end and rotates freely with an angular velocity ω in a uniform magnetic field $\vec{B}$ in the z direction as shown in Fig. 8.19. Find the induced EMF between the two ends of the strip.

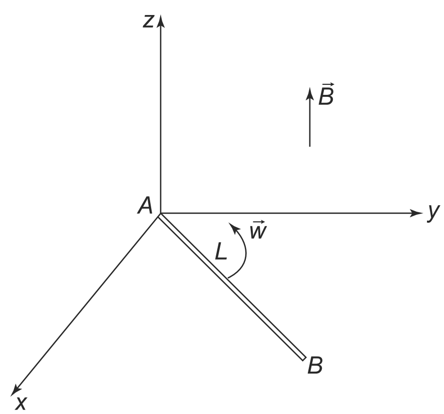


Fig. 8.19 *Problem 1*

2. A rectangular coil having N turns is rotating with a uniform angular velocity ω in a uniform magnetic field $\vec{B}$ as shown in Fig. 8.20. Find the induced EMF in the coil using the concept of motional EMF as well as Faraday's law and show that the two are the same.
3. A circular copper disk of diameter 10 cm rotates at 1800 revolutions per minute about an axis passing through its centre and at right angles to the disk. A uniform magnetic field $\vec{B}$ of magnitude 1 T is perpendicular to the disk. Find the potential difference between the axis and the rim of the disk.
4. A rigid rod made of metal of length $L = 1$ m. It is rotating in a horizontal plane about an axis through one end and perpendicular to the length of the rod. A uniform magnetic field in the

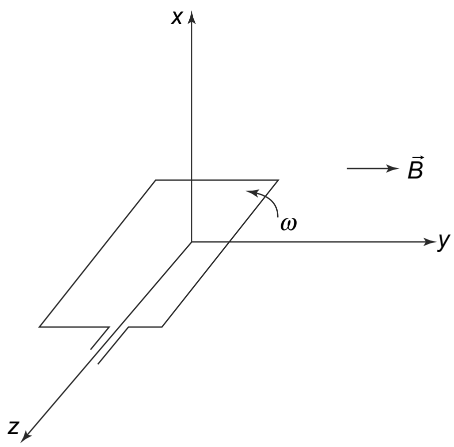


Fig. 8.20 *Problem 2*

vertical direction is present of strength 1 $\mathrm{wb/m^2}$. If the time period of rotation of the rod is 5 sec, calculate the EMF induced between the two ends of the rod.

5. A square wire of length l, mass m and resistance R slides down without friction on parallel conducting rails as shown in Fig. 8.21. The rails are connected to each other by a bar of zero resistance. The plane of the rails makes an angle θ with the horizontal and a uniform magnetic field $\vec{B}$ exists everywhere. Find the steady state velocity of the wire.

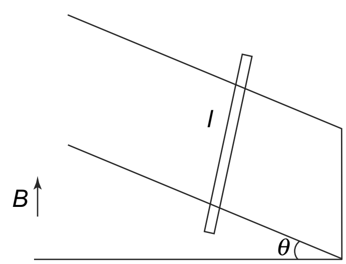


Fig. 8.21 *Problem 5*

6. A rectangular loop of dimensions 20 cm by 10 cm is made of aluminium wire of radius 1.2 mm. The loop is placed in a magnetic field which is increasing at a rate of 40 T/s. What is the induced current in the loop, given that the conductivity of aluminium is 3.57×10^7 S/m.

7. A copper disc of 20 cm radius is rotating at a speed of 1200 revolutions per minute about its axis in a uniform magnetic field. If the field makes an angle of 30° with the axis of the disc, find the induced EMF between the rim and the axis of the disc.

8. A hollow massless paper cylinder is filled with a dielectric material of permittivity ε. A spatially uniform but time-dependent magnetic field parallel to the axis of the cylinder $\vec{B}(t)$, $B = B_0 \cos^2(\omega t)$. Find the induced polarisation in the dielectric.

9. A toroidal coil of$N$ turns is wound on a non-magnetic material. If the mean radius of the coil is b and the cross-sectional radius is a show that the self-inductance of the coil is

$$L = \mu_0 N^2(b - \sqrt{B^2 - a^2})$$

10. Calculate the mutual inductance between a very long, current carrying conductor and a square loop of side a placed at a distance b from it.

An inductor of inductance L and a resistor of resistance R is connected in series across a battery. The battery however, is discharging very fast and its voltage is changing with time as $\mathcal{E}(t) = \mathcal{E}_0 e^{(-\alpha t)}$. Calculate the current in the circuit as a function of time.

11. An inductor consists of two very thin conducting cylindrical shells, one of radius a and one of radius b, both of length h. Assume that the inner shell carries current I out of the page, and that the outer shell carries current I into the page, distributed uniformly around the circumference in both cases. The z-axis is out of the page along the common axis of the cylinders and the r-axis is the radial cylindrical axis perpendicular to the z-axis.

a. Find the magnetic energy density as a function of r for $a < r < b$.
b. Calculate the self-inductance of this inductor.

12. A uniform magnetic field $\vec{B}$ is changing at a constant rate $\frac{dB}{dt}$. A mass m of copper is drawn into a wire of radius r and made into a circular loop of radius R. The density of copper is δ and resistivity ρ. Assuming that the field is perpendicular to the loop, show that the induced current in the loop is

$$I = \frac{m}{4\pi\rho\delta}\frac{dB}{dt}$$

13. Consider a circular loop connected to a resistor R in series. The loop is placed in a region with magnetic flux perpendicular to the plane of the loop and directed into the paper. The flux through the loop changes as

$$\phi = 6t^2 + 7t + 1$$

What is the magnitude of the induced EMF in the loop at time $t = 2$ seconds?

14. An arrangement of two coils as shown in Fig. 8.22 is set up. What is the direction of the current in the resistor r when

a. Switch S is closed
b. Coil 2 is moved closer to coil 1 and
c. Resistance R is decreased.

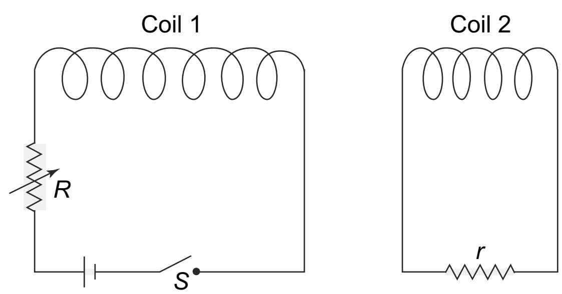


Fig. 8.22 *Problem 14*

15. An infinite wire carries a current I in the z direction as shown in Fig. 8.23. A rectangular loop of wire of side l is connected to a voltmeter and moves with a velocity $\vec{u}$ radially outwards from the wire. Find the reading in the voltmeter.

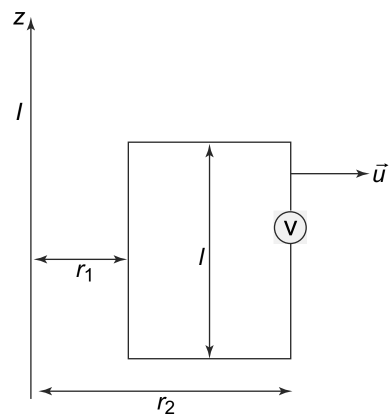


Fig. 8.23 *Problem 15*

16. Two infinite parallel wires are separated by a distance d carry equal currents I in opposite directions where I increases at a rate of $\frac{dI}{dt}$. A square loop of wire of length d lies in the plane of the wires at a distance d from one of the wires, as shown in Fig. 8.24. Find the magnitude and the direction of the induced EMF in the loop.

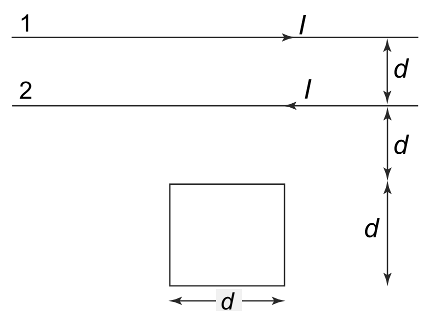


Fig. 8.24 *Problem 16*

17. A very long air core solenoid 2 cm in diameter has two coils wound over each other. The inner coil has 400 turns per meter and the outer coil has 4000 turns per meter. What is the mutual inductance of the coils? When the inner coil carries a current $0.5\cos(220t)$ A, what is the induced EMF per unit length of the outer coil?

18. A constant force is applied to a sliding wire of mass m which starts from rest. The wire is of resistance R and moves in a region of constant magnetic field B. Assume that the self-inductance of the wire and the friction of contacts is negligible, calculate the current through the resistor R as a function of time.

19. In the arrangement shown in Fig. 8.25, on the right is a ferromagnetic rod on which is wound a closed loop of a coil of resistance R. On the left is an electrical circuit with a switch, a resistor of resistance R, a battery of EMF $\mathcal{E}$ and a set of coils fixed near the rod on the right with the axis of the coils along the length of the rod. Determine whether the current in the loop is along $PQSRP$ or $PRSQP$ or is zero for the following cases

a. The switch is suddenly taken off

b. $\mathcal{E}$ is suddenly increased

c. with the switch in the off position, the polarity of the battery is reversed and then the switch is put on

d. With the switch on, the coil is brought nearer to the rod.

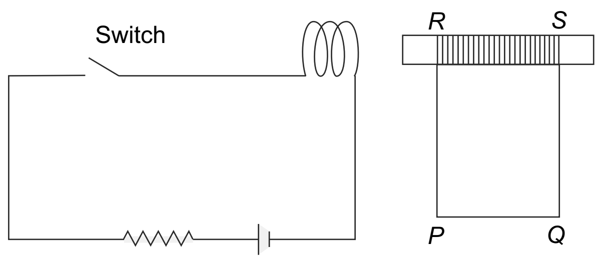


Fig. 8.25 *Problem 19*

9

Time Dependent Circuits and Alternating Currents

Learning Objectives

- To learn about ideal inductors, resistors and capacitors and the application of Kirchhoff's Laws.
- To learn about the oscillating LC circuit.
- To study the LCR circuit with a direct current source and learn about the various phenomenon of decaying currents and damped oscillating currents in the circuit.
- To understand the application of Kirchhoff's Laws to alternating current circuits.
- To learn about complex currents, voltages and impedances for various circuit elements.
- To be familiar with series and parallel combinations of impedances.
- To introduce the power factor for alternating currents with impedances and its relationship with the phase angle.
- To learn about LR and LCR circuits with alternating current sources and the phenomenon of resonance and Q-factor.
- To understand the functioning of a transformer as an example of a inductively coupled circuit with an alternating current source.

9.1 INTRODUCTION

We have already introduced time varying, as opposed to steady, currents in earlier chapters. For instance, in Chapter 5, we saw that a circuit consisting of a resistor and a capacitor has a characteristic time dependent behaviour which depends on the values of the resistance and capacitance. In Chapter 8, we introduced a circuit with an inductance and a resistance and saw another example of time dependent currents. In this chapter, we shall study general circuits with all the three elements (resistance, capacitance and inductance) and their various combinations. In addition, we shall also study alternating current circuits. Alternating current is the kind of current which is usually produced by generators and is used in the electrical supply. Typically, it has a sinusoidal time dependence.

Before we study various circuits with the above mentioned circuit elements, it is important to note that the resistors, capacitors and inductors we talk about are 'ideal' resistors, capacitors and inductors. Thus, an ideal resistor is a pure resistor and has no inductance. An ideal capacitor, as we have seen in Chapter 5 is one which has its field restricted to the region between the plates and an ideal inductor is one which has zero resistance. In all these cases, we have seen that it is possible to use the concept of electric potential and so the machinery of analysing circuits that we developed in Chapter 5 (Ohm's Law, network theorems, Kirchhoff's Laws etc) can be used to analyse the circuits containing these idealised circuit elements. The symbols for these three circuit elements are shown in Fig. 9.1

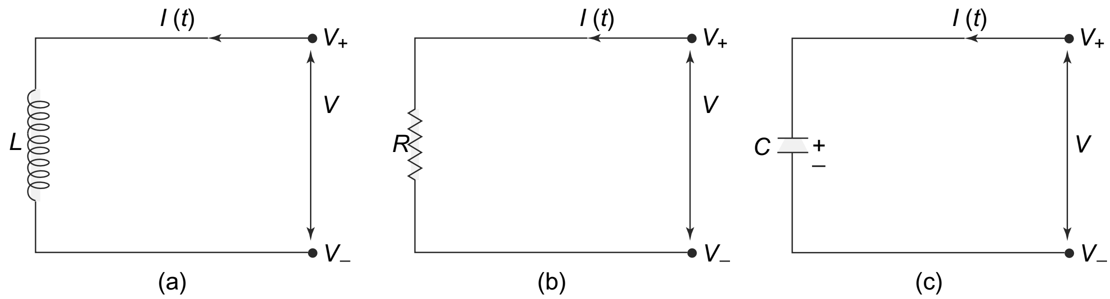


Fig. 9.1 *Symbols for ideal circuit elements. The V in each case is $V = V_+ - V_-$ where $+$ and $-$ refer to upper and lower terminals. The current in each case flows from $+$ to $-$ through the circuit elements in question: (a) Ideal inductor, (b) Ideal (Ohmic) resistor, (c) Ideal capacitor*

For an inductor, as we saw in Chapter 8 (Eq. (8.41)), the relationship between current in the inductance and the potential difference, with the current flowing from $+$ to $-$ through the inductor

$$V = V_+ - V_- = L\frac{dI(t)}{dt} \tag{9.1}$$

For the capacitor, the relationship between the current and the potential difference was discussed in Chapter 5. We have

$$V = V_+ - V_- = \frac{Q}{C} \tag{9.2}$$

and

$$I(t) = \frac{dQ(t)}{dt}$$

as usual.

For the resistor, the relationship is simply the Ohm's Law

$$V = V_+ - V_- = I(t)R \tag{9.3}$$

The important thing about these relationships (Eqs. (9.1), (9.2), (9.3)) is that the relationship is linear in each case. This means that changing V by some factor χ changes $I(t)$ by the same factor. This leads to an enormous simplification in the analysis of the circuits as we shall see. These circuit elements are also called **linear circuit elements**.

In general however, there exist many circuit elements, particularly in electronic circuits where this linear relationship does not hold. A diode (a device which allows current to flow in only one direction) for instance has a non-linear relationship between the potential and the current in general. However, some of the concepts and tools that we develop to analyse linear circuits can be used for electronic circuits with non-linear elements too. But it is important to remember that we cannot take the analysis of linear circuits directly to more general circuits.

9.2 *LC* CIRCUIT

We have already studied the behaviour of circuits with resistance and capacitance (RC circuits) as well as those with an inductance and resistance (LR circuits). Let us now see what happens when we have an inductance and a capacitance in a circuit called the LC circuit. Consider first the simple circuit in Fig. 9.2.

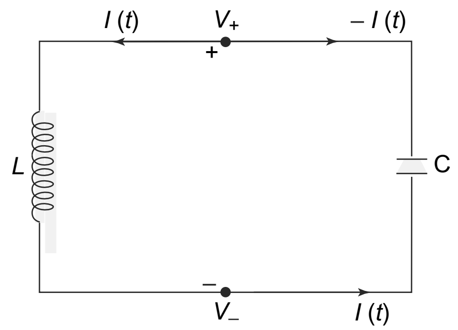


Fig. 9.2 *An LC circuit. V_+ and V_- are the potentials at the points marked + and − respectively. The current $I(t)$ flows through L from + to − and a current of $-I(t)$ flows through the capacitor from + to −*

The current-potential relationship for both these circuit elements we have already obtained in Equations (9.1) and (9.2). Thus, we have the following relationships

$$V = V_+ - V_- = L\frac{dI(t)}{dt}$$

$$V = \frac{Q}{C}$$

$$-I(t) = \frac{dQ(t)}{dt} \tag{9.4}$$

Equating the potentials, we get

$$\frac{Q}{C} = L\frac{dI(t)}{dt}$$

$$= -L\frac{d^2Q(t)}{dt^2} \tag{9.5}$$

or

$$\boxed{\frac{d^2Q(t)}{dt^2} + \omega^2 Q(t) = 0} \tag{9.6}$$

This is the equation which the charge on the capacitor obeys. Here of course

$$\omega^2 = LC$$

This equation is a second order, linear differential equation which is familiar to us from a simple harmonic oscillator oscillating with a frequency ω. There are several standard methods of solving this equation. Let us write

$$Q(t) = e^{\lambda t}$$

where λ is a constant. Then, substituting in Eq. (9.6), we get

$$\lambda^2 + \omega^2 = 0 \tag{9.7}$$

or

$$\lambda = \pm i\omega$$

The general solution to the equation can thus be written as

$$Q(t) = Ae^{i\omega t} + Be^{-i\omega t} \tag{9.8}$$

where A and B are constants which are in general complex. Since $Q(t)$, being the charge at any instant t on the capacitor is obviously real, we also have

$$A = B^* = \frac{C}{2}e^{i\delta}$$

where we have now two real constants C and δ which will be determined by initial conditions. Thus, the solution is given by

$$\boxed{Q(t) = C\cos(\omega t + \delta)} \tag{9.9}$$

To determine C and δ, we shall use initial conditions. Let us consider the case where at time $t = 0$, the capacitor carries a charge Q_0 and no current is flowing in the circuit. Then

$$\begin{aligned} Q_0 &= C\cos\delta \\ 0 &= -C\sin\delta \end{aligned} \tag{9.10}$$

This gives us $\delta = 0$ and $C = Q_0$. Thus, the solution is

$$\begin{aligned} Q(t) &= Q_0\cos(\omega t) \\ I(t) &= -Q_0\omega\sin(\omega t) \end{aligned} \tag{9.11}$$

The solutions clearly show that the charge on the capacitor and the current in the circuit oscillate with a time period

$$\boxed{T = \frac{2\pi}{\omega} = \frac{2\pi}{\sqrt{LC}}}$$

The behaviour of the the charge on the capacitor and the current in the circuit parallels that of a simple harmonic oscillator. Recall that in a simple harmonic oscillator, the displacement (say of a simple pendulum) and the velocity are out of phase as are the charge and the current in this circuit. Furthermore, in a simple pendulum, if we start with some initial displacement of the bob with zero velocity, the energy of the pendulum at $t = 0$ is purely potential. This potential energy goes into kinetic energy of the bob as the velocity increases and the displacement decreases and goes back into potential energy when the pendulum swings to extreme point on the other side. The total energy which is the sum of the potential and kinetic energy of course remains the same (equal to the potential energy of the bob in its initial position).

A very similar situation exists in the LC circuit. The energy stored in the capacitor, at any time t is given by

$$E_C = \frac{1}{2}\frac{Q(t)^2}{C} = \frac{Q_0^2}{2C}\cos^2 \omega t \tag{9.12}$$

This of course varies with time being equal to $\frac{Q_0^2}{2C}$ at $t = 0$. On the other hand, the energy in the inductor also varies with time and at time t is given by

$$E_L = \frac{1}{2}LI(t)^2 = \frac{L}{2}Q_0^2\omega^2 \sin^2 \omega t = \frac{Q_0^2}{2C}\sin^2 \omega t \tag{9.13}$$

where we have used the fact that $\omega^2 = \frac{1}{LC}$. Thus, we see that

$$E_C + E_L = \frac{Q_0^2}{2C} = E$$

is the total energy which is time independent. Just like in the pendulum case, the initial potential energy given to the bob is the total energy of the system, here too, the initial energy of the charged capacitor is the total energy of the system. This energy goes from the energy in the capacitor to the energy in the inductor as is evident from the time dependencies of Eqs. (9.12) and (9.13) but the total energy remains constant, as is expected from conservation of energy.

The energy considerations above are, of course, valid only if there are no loses in the LC circuit. This is indeed the case if both L and C are ideal. Exactly like a simple pendulum, in the ideal case with zero frictional losses, the pendulum can go on oscillating forever, in this case, the circuit can keep oscillating at the frequency ω forever. In the case of a real pendulum of course, there are frictional losses which ultimately dissipate all the energy and the motion stops. In the case of a circuit, the only dissipative circuit element is an ohmic resistance where energy would get dissipated as Joule heat. Since there is no resistance in the ideal LC circuit, the oscillations do not die out.

PROBLEM 9.1 A circuit, with a switch which is in the off position initially, has a capacitor of capacitance C, an inductor of inductance L connected in series with a battery V of EMF $\mathcal{E}$ and zero internal resistance. The capacitor is initially charged with a charge Q_0 and at time $t = 0$, the switch is turned to the on position. Calculate the current in the circuit and the charge on the capacitor as a function of time for $t > 0$.
[*Hint: The equation in this circuit is very similar to Eq. (9.5) with V adding to or subtracting from* $\frac{Q}{C}$]

9.3 *LCR* CIRCUITS

As we mentioned in the last section, a circuit with ideal L and C circuit elements has no dissipative component and hence the oscillations set up in it can continue forever. This is not the case when there is a resistive component in the circuit which dissipates energy. Let us consider a circuit with an inductance, a capacitance and a resistance in series. Each of the components are assumed to be ideal. Thus, for instance, the inductance will have zero resistance while the resistance will have zero inductance, etc.

Such a circuit is called an LCR circuit. We shall consider the case when there is a battery connected to the circuit and when there is no battery.

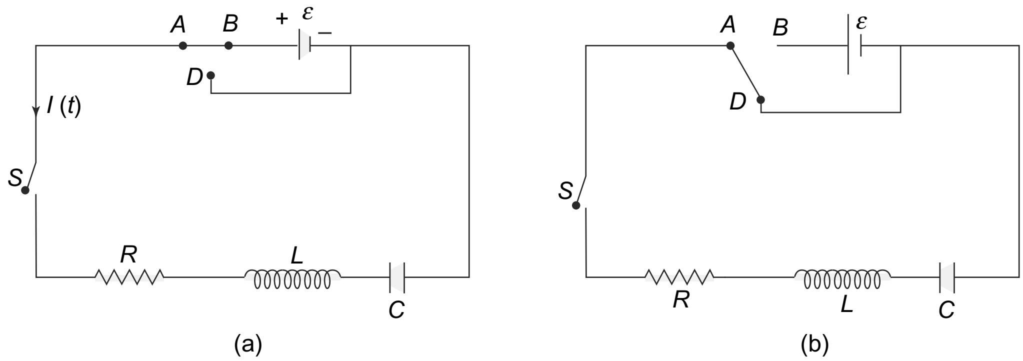


Fig. 9.3 *LCR circuit (a) An LCR circuit with a battery, (b) An LCR circuit with the battery removed*

Consider first the circuit with the battery with EMF, $\mathcal{E}$ in place, Fig. 9.3(a). In this case, we have already seen that each of the circuit elements being ideal, we can apply the rules relating potentials and currents. In particular, Kirchhoff's laws are valid for this circuit provided we take care to use the relevant potential differences as mentioned in Eqs. (9.1), (9.2) and (9.3). Thus, if we take the current direction as shown and go around the loop in the anti-clockwise direction, we get, using Kirchhoff's Laws

$$I(t)R + L\frac{dI(t)}{dt} + \frac{Q(t)}{C} = \mathcal{E} \tag{9.14}$$

Here, as usual

$$I(t) = \frac{dQ(t)}{dt}$$

We differentiate this equation once more w.r.t t and get

$$\boxed{R\frac{dI(t)}{dt} + L\frac{d^2I(t)}{dt^2} + \frac{I(t)}{C} = 0} \tag{9.15}$$

This is the basic differential equation governing the charging LCR circuit. To solve this equation, we define

$$\omega^2 = \frac{1}{LC}$$

and

$$f = \frac{R}{L} > 0$$

The equation thus, becomes

$$\frac{d^2I(t)}{dt^2} + f\frac{dI(t)}{dt} + \omega^2 I(t) = 0 \tag{9.16}$$

This equation is very similar to a simple harmonic oscillator equation except that this one has an extra term proportional to the rate of change of current and a parameter f. It is actually the equation which we encounter for a simple harmonic oscillator with a viscous drag. Recall that for the case of displacement of a simple harmonic oscillator in a viscous medium, there is a term proportional to the velocity or rate of change of displacement. The extra term is precisely of that same form together with a 'drag' coefficient f.

The method of solving this equation is again standard and is what we used for the case of an LC circuit. We put

$$I(t) = e^{\lambda t} \tag{9.17}$$

and substitute in Eq. (9.14). On doing this, the equation reduces to an algebraic equation

$$\lambda^2 + f\lambda + \omega^2 = 0 \tag{9.18}$$

which can be easily solved to get

$$\lambda_{\pm} = \frac{-f \pm \sqrt{f^2 - 4\omega^2}}{2} \tag{9.19}$$

Clearly, depending on whether $4\omega^2 > f^2$ or $4\omega^2 \leq f^2$, we have two kinds of solutions.

Case A: $f^2 \geq 4\omega^2$. **Case of damped currents** In this case, from Eq. (9.19), we see that there are two solutions, λ_+ and λ_-, both of which are real and negative. Further, if $4\omega^2 = f^2$, then $\lambda_+ = \lambda_-$. Thus, the solution for the differential equation reads

$$I(t) = Ae^{\lambda_+ t} + Be^{\lambda_- t} \tag{9.20}$$

where A and B are constants to be determined by initial conditions or values of the current and its derivative w.r.t time at $t = 0$. Suppose the capacitor was uncharged at time $t = 0$ and the current in the circuit is also zero. That is

$$Q(0) = I(0) = 0$$

In that case, from Eqs. (9.20) and (9.14), we get

$$A + B = 0 \tag{9.21}$$

and

$$L(\lambda_+ A + \lambda_- B) = \mathcal{E} \tag{9.22}$$

This gives us

$$A = -B \tag{9.23}$$

and

$$A = \frac{\mathcal{E}}{L\sqrt{f^2 - 4\omega^2}} \tag{9.24}$$

The current in the circuit at any time t is thus,

$$\boxed{I(t) = \frac{\mathcal{E}}{L\sqrt{f^2 - 4\omega^2}} \left(e^{\lambda_+ t} - e^{\lambda_- t}\right)} \tag{9.25}$$

Note that since λ_+ and λ_- are both real and negative, the current dies out. The charge on the capacitor, at any time t is given by

$$\begin{aligned}
Q(t) &= \int_0^t I(t')dt' \\
&= \frac{\mathcal{E}}{L\sqrt{f^2-4\omega^2}} \int_0^t \left(e^{\lambda_+ t'} - e^{\lambda_- t'}\right) dt' \\
&= \frac{\mathcal{E}}{L\sqrt{f^2-4\omega^2}} \left[\frac{1}{\lambda_+}\left(e^{\lambda_+ t} - 1\right) - \frac{1}{\lambda_-}\left(e^{\lambda_- t} - 1\right)\right] \qquad (9.26)
\end{aligned}$$

For $t \to \infty$, the charge on the capacitor goes to

$$\begin{aligned}
Q(t=\infty) &= \frac{\mathcal{E}}{L\sqrt{f^2-4\omega^2}} \left[\frac{1}{\lambda_-} - \frac{1}{\lambda_+}\right] \\
&= \frac{\mathcal{E}}{L\sqrt{f^2-4\omega^2}} \left[\frac{\lambda_+ - \lambda_-}{\lambda_+\lambda_-}\right] \\
&= \frac{\mathcal{E}}{L\sqrt{f^2-4\omega^2}} \left[\frac{\sqrt{f^2-4\omega^2}}{\omega^2}\right] \\
&= \frac{\mathcal{E}}{L} LC \\
&= \mathcal{E}C \qquad (9.27)
\end{aligned}$$

The charge on the capacitor is related to the voltage across its plates by the relation $Q = CV$. Thus, we see that asymptotically the capacitor gets charged such that the potential difference across its plates become equal to the emf of the battery.

Now suppose, after the capacitor is fully charged, we remove the battery from the circuit as shown in Fig. 9.3(b). The equation governing the potentials in the circuit now can be easily written with $\mathcal{E} = 0$. It is

$$I(t)R + L\frac{dI(t)}{dt} + \frac{Q(t)}{C} = 0 \qquad (9.28)$$

which once again, gives us

$$\boxed{R\frac{dI(t)}{dt} + L\frac{d^2I(t)}{dt^2} + \frac{I(t)}{C} = 0} \qquad (9.29)$$

The solution to this equation is of course, of the same form as we found above in the charging case but now with different constants, since clearly the initial conditions are different in this case.

$$I(t) = A'e^{\lambda_+ t} + B'e^{\lambda_- t} \qquad (9.30)$$

with

$$\lambda_\pm = \frac{-f \pm \sqrt{f^2-4\omega^2}}{2} \qquad (9.31)$$

For determining the constants A' and B', we use the initial conditions. At $t = 0$, the current is zero but the capacitor is fully charged and so we have

$$Q(0) = C\mathcal{E}$$

and

$$I(0) = 0$$

Putting it into Eqs. (9.29) and (9.28) we get

$$L\frac{dI(t)}{dt}|_{t=0} = -\frac{\mathcal{E}C}{C} = -\mathcal{E} \tag{9.32}$$

Putting the solution Eq. (9.30) with these boundary conditions, we get

$$A' + B' = 0 \tag{9.33}$$

and

$$L(\lambda_+ A' + \lambda_- B') = -\mathcal{E} \tag{9.34}$$

We therefore, have the current given by

$$\boxed{I(t) = \frac{-\mathcal{E}}{L\sqrt{f^2 - 4\omega^2}}\left(e^{\lambda_+ t} - e^{\lambda_- t}\right)} \tag{9.35}$$

The charge on the capacitor at any time t can be obtained by integrating the expression for $I(t)$ but remembering that $Q(0) = \mathcal{E}C$.

$$\begin{aligned} Q(t) &= \mathcal{E}C + \int\limits_0^t I(t')dt' \\ &= \mathcal{E}C - \mathcal{E}L\sqrt{f^2 - 4\omega^2}\left[\frac{1}{\lambda_+}\left(1 - e^{\lambda_+ t}\right) - \frac{1}{\lambda_-}\left(e^{1-\lambda_- t}\right)\right] \end{aligned} \tag{9.36}$$

Note that the current falls with time (Eq. (9.35)) and the charge as $t \to \infty$ goes to

$$Q(t \to \infty) = \mathcal{E}C - \mathcal{E}L\sqrt{f^2 - 4\omega^2}\left[\frac{\lambda_- - \lambda_+}{\lambda_- \lambda_+}\right] = 0 \tag{9.37}$$

The capacitor, which is initially fully charged to its maximum value ($\mathcal{E}C$) with the given voltage ($\mathcal{E}$) thus discharges completely to have a zero charge asymptotically as $t \to \infty$. There are, as in the charging case, no oscillations of the current or the charge.

Case B: $4\omega^2 > f^2$ **Case of damped oscillations**

For the second case, we consider the situation where

$$4\omega^2 > f^2$$

or

$$4\frac{1}{LC} > \left(\frac{R}{L}\right)^2$$

In this case

$$\lambda_\pm = \frac{-f \pm \sqrt{f^2 - 4\omega^2}}{2}$$

are clearly complex. Hence, we write them as

$$\lambda_\pm = \frac{-f \pm \sqrt{f^2 - 4\omega^2}}{2} = -\frac{f}{2} \pm i\alpha \tag{9.38}$$

where α is now real and given by

$$2\alpha = \sqrt{4\omega^2 - f^2}$$

With this substitution, our solution for $I(t)$ then reads

$$I(t) = e^{-ft/2}\left[A_1 e^{i\alpha t} + B_1 e^{-i\alpha t}\right] \tag{9.39}$$

Immediately, we can see that since $I(t)$ is real,

$$A_1 = B_1^*$$

We write

$$A_1 = \frac{A_0}{2} e^{i\delta}$$

where A_0 and δ are real, we get

$$I(t) = A_0 e^{-ft/2} \cos(\alpha t + \delta) \tag{9.40}$$

Note that we still have two integration constants A_0 and δ which need to be determined by initial conditions.

Again, let us first consider the case of the charging circuit, i.e. the circuit in which the battery is included. In this case, the initial conditions are

$$I(0) = 0$$

and

$$Q(0) = 0$$

Repeating the same steps as in the case of damped currents, we get

$$A_0 \cos(\delta) = 0$$

and

$$\begin{aligned} L\left(\frac{dI(t)}{dt}\right)_{t=0} &= LA_0\left[e^{-ft/2}\left(-\frac{f}{2}\cos(\alpha t + \delta) - \alpha \sin(\alpha t + \delta)\right)\right]_{t=0} \\ &= -LA_0\left(\frac{f}{2}\cos\delta + \alpha \sin\delta\right) \\ &= \mathcal{E} \end{aligned} \tag{9.41}$$

Now these conditions on A_0 and δ can be satisfied with

$$\delta = \frac{\pi}{2}$$

and

$$A_0 = -\frac{\mathcal{E}}{L\alpha}$$

The current for the charging circuit is given by

$$I(t) = \frac{\mathcal{E}}{L\alpha} e^{-ft/2} \sin(\alpha t) \tag{9.42}$$

The current thus oscillates with a frequency α but its amplitude decreases exponentially as $e^{-ft/2}$. The behaviour is shown in Fig. 9.4(a).

The charge on the capacitor at any time, can be obtained from the current. Note that at $t = 0$, $Q(0) = 0$. Thus, we have

$$Q(t) = \int_0^t I(t')dt' \tag{9.43}$$

Clearly, since $I(t')$ oscillates with a fixed frequency but decreasing amplitude, $Q(t)$ also oscillates. Asymptotically, that is for very large times, as $t \to \infty$,

$$Q(t \to \infty) = \int_0^\infty I(t')dt' = \mathcal{E}C$$

The capacitor thus, gets charged fully only asymptotically.

Finally, what happens in this case (i.e. $4\omega^2 > f^2$) when the battery is removed. This situation is very similar to the overdamped case, though now with $\mathcal{E} \to 0$ and of course, the initial condition of the charge on the capacitor being the full charge $Q(0) = \mathcal{E}C$. The current can be written down with these substitutions as

$$I(t) = -\frac{\mathcal{E}}{\alpha L} e^{-ft/2} \sin(\alpha t) \tag{9.44}$$

and the charge on the capacitor

$$Q(t) = \mathcal{E}C - \int_0^t \frac{\mathcal{E}}{\alpha L} e^{-ft'/2} \sin(\alpha t')dt' \tag{9.45}$$

The current and the charge thus oscillate just as the charging case though in this case, as $t \to \infty$, $Q(t) \to 0$. The variation of current with time is shown in Fig. 9.4(b).

To summarise then, a circuit with an inductor, a capacitor and a resistor behaves like a damped oscillator. Depending on the relative values of L, C and R, the behaviour of the current changes. For the overdamped case, that is when $f^2 \geq 4\omega^2$ or alternatively $R^2 \geq \frac{4L}{C}$ we get the current falls off without any oscillations and dies out. The charge on the capacitor goes to its fully charged value with the EMF $\mathcal{E}$. When the battery is removed in this case, once again the current does not oscillate but gets damped out. The charge on the capacitor also goes to zero asymptotically in a long time. This is understandable since in the case where $R^2 \geq \frac{4L}{C}$, the resistive component is larger than the component which leads to oscillations and hence one sees an overdamped motion.

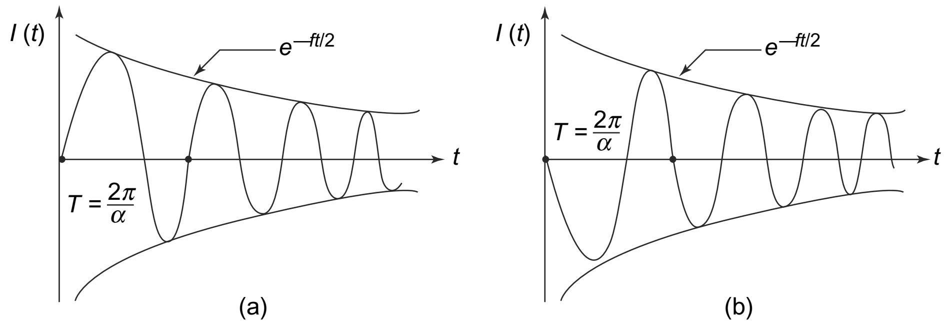


Fig. 9.4 *Transient currents in an LCR circuit (a) Behaviour of transient current in an LCR circuit with the battery, (b) Behaviour of transient current in an LCR circuit with the battery removed*

In case where $f^2 < 4\omega^2$ the reverse is true and we get oscillations. The current in the circuit oscillates with a frequency $\alpha = \sqrt{4\omega^2 - f^2}$ but with an amplitude which falls off exponentially with time. The same is true when the battery is removed from the circuit. Once again, we see that since the resistive element is smaller than the component which causes oscillations, we expect this behaviour. Of course, if $R = 0$, we would get oscillations of constant amplitude like the LC circuit.

PROBLEM 9.2 A series LCR circuit with a battery of emf $V = 2.5$ Volts and zero internal resistance is initially switched off and the capacitor has no charge on it. The values of the circuit elements are: $L = 5$ mH, $C = 8\mu$F and $R = 4$ ohms. At time $t = 0$, the circuit is switched on. What are the charge on the capacitor and the current in the circuit at time $t > 0$? Calculate approximately the time at which the capacitor will be charged to half its final value.

PROBLEM 9.3 In the circuit shown in Fig. 9.5, there is no initial charge on C and the switch is turned on time $t = 0$. Calculate the currents in all the branches for time $t > 0$.
[*Hint: the currents through C and L add up to the current through R; at the same time the potential difference across L and C are equal. Use these facts to write an equation involving one unknown*]

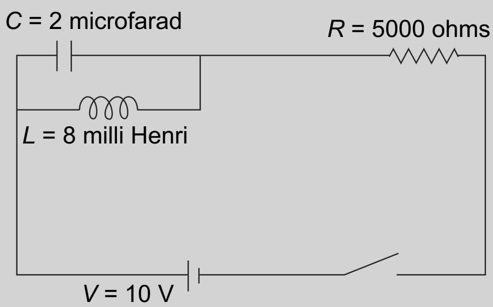


Fig. 9.5 *Problem 9.3*

9.4 ALTERNATING CURRENTS

In all our discussions about charges in motion until now, we have dealt with EMF sources which are time independent. We discussed circuit analysis with such EMF sources which were, typically resistive circuits. We also considered varying currents for circuits with non-resistive elements like inductors and capacitors. **Electrical circuits with EMF sources that are time independent are called direct current or DC circuits.** It turns out that there is another category of currents which are in fact much more important in practical life and these are called alternating currents. Alternating currents are those which vary with time with a given frequency. For instance, the power supply to our houses is normally alternating current supply at 50 Hz (in some countries, 60 Hz too). We shall study the reason why it is more practical and convenient to transmit power over long distances with alternating current later in this chapter. However, before we attempt to do analysis of circuits with alternating currents, we need to ascertain whether our analysis of circuits with direct currents (Chapter 5) is valid for the currents that vary with time.

9.4.1 Kirchhoff's Laws for Alternating Currents

In our analysis of circuits and power sources of direct current, we saw that Kirchhoff's laws, which were basically statements of charge conservation and energy conservation, were extremely useful in analysing the currents and potential differences. We now examine whether these laws are also applicable in cases where the currents and potential differences vary with time at a given frequency.

Kirchhoff's first law, namely that the algebraic sum of all the currents entering a node or any point in the circuit is zero was, as mentioned earlier, a statement of conservation of charge. To put it another way, it was a statement that in any volume, charge cannot accumulate. However, in a case where currents and potential differences are time dependent, charges do tend to accumulate temporarily at various points in the circuit. However, as we saw in the case of electric fields in conductors, this temporary accumulation of charges inside the conductor tends to get smoothed out very fast. Thus, if the frequency of the alternating current is such that the time scale for variation is much larger than the time scales at which the charges get smoothed out in the conductors, then we can assume that there is no charge accumulation in the circuit. When we have capacitors in the circuit, of course, for currents changing with time, charges do accumulate on the plates. But the charges on the plates of an ideal capacitor are always equal and opposite, even when they are changing with time. That means that whatever charge is deposited on one of the plates, an equal amount of charge flows out of the other plate. Thus, as long as we are not considering points inside the capacitor, we have the same current flowing through the plates: current going into one plate is exactly the current going out of the other plate. In this respect, the two plates are just like the terminals of a resistor. We will thus, be free to use Kirchhoff's first law with this understanding in analysing circuits with capacitors as well.

Kirchhoff's Second Law is a statement about the potential drop across a closed loop in the circuit. For direct current circuits, with pure resistive elements, we saw in Chapter 5 that this was the sum of potential drops across the resistors (IR) and the potential drops or increases across the plates of the battery. For circuits with inductors and capacitors and in the presence of alternating currents, the situation is a bit different. These elements lead to a time-dependent behaviour of the current as we have seen. However, for most circuits at ordinarily used frequencies, the time required for electric fields to propagate through various elements in a circuit is very small compared with typical time constants in the circuit. This would imply that relationships between potential differences and currents in a circuit would have an instantaneous relationship. We have also seen (Chapters 5 and 8) that for ideal capacitors and inductors, it is possible to define a potential across their terminals. Thus at any instant of time, we have a situation no different from steady currents case for relating the currents and potential differences. Krichhoff's Second Law, which, as we have mentioned, is an expression of energy conservation, would thus be valid at all times, even with time dependent currents and potential differences.

In the light of this, we see that for a circuit containing all the three circuit elements, namely resistors, inductors and capacitors, in addition to the potential drop across the resistor (which is given by Ohm's Law to be IR), we will also have a potential drop $\frac{Q}{C}$ across the capacitor and $L\frac{dI(t)}{dt}$ across the inductors. Now for an alternating current circuit, Kirchhoff's Second Law then says that for a closed loop, at any instant, the sum of these potential drops across the circuit elements and the potential drop across the EMF sources vanishes. The potential drops across the various circuit elements and the EMF thus are related by:

$$V_R = IR$$

$$V_C = \frac{Q(t)}{C}$$

$$V_L = L\frac{dI(t)}{dt}$$

$$\mathcal{E} = V_R + V_C + V_L = IR + \frac{Q(t)}{C} + L\frac{dI(t)}{dt}$$

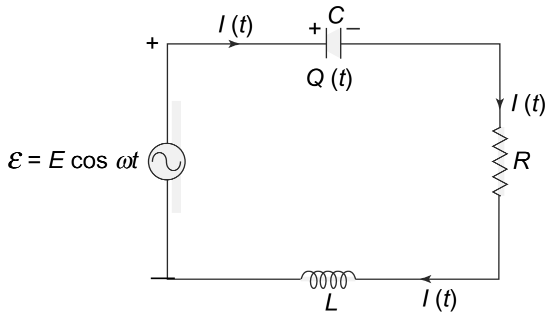


Fig. 9.6 *Kirchhoff's Second Law for A.C circuits. The sum of potential drops across a closed loop is zero*

This is depicted in Fig. 9.6.

Thus, we see that both of Kirchhoff's Laws can be applied to alternating current circuits with resistive, inductive and capacitative elements provided we take into account the appropriate potential drops across all these elements and the power sources.

9.4.2 Phasor Diagrams

Quantities varying sinusoidally with time with a frequency ω are often conveniently represented as **'phasors'**. A phasor corresponding to a physical quantity varying with time like $\cos(\omega t + \delta_1)$ is represented by a directed line starting at the origin of a two-dimensional plane. The length of the line is proportional to the amplitude of the quantity and it makes angle $(\omega t + \delta_1)$ with the x-axis.

Thus, in Fig. 9.7(a), a quantity $V_1 = V_{10}\cos(\omega t + \delta_1)$ is represented by a line OP whose length is proportional to V_{10} and a quantity $V_2 = V_{20}\cos(\omega t + \delta_2)$ is represented by another line OQ whose length is proportional to V_{20}. The sum of these two quantities is easily seen to be OR in Fig. 9.7(b). We

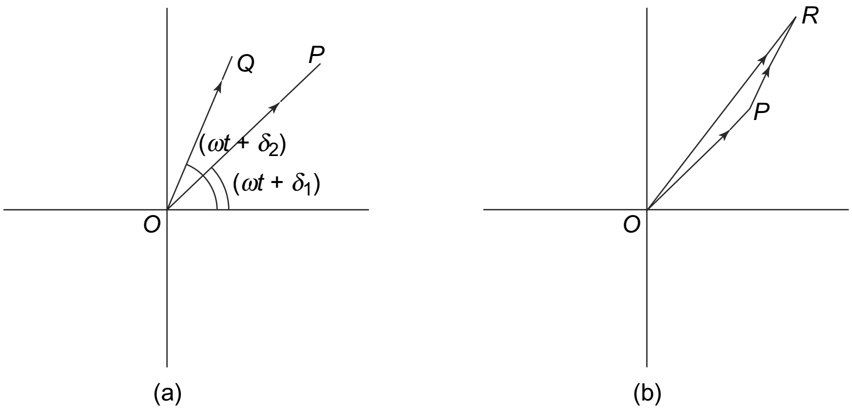


Fig. 9.7 *Phasor diagram*

can also represent the derivatives with respect to time of quantities since taking a derivative w.r.t time is simply multiplying with a factor of ω and increasing the angle of the argument by $\frac{\pi}{2}$. Thus,

$$\frac{dV_1}{dt} = -V_{10}\omega \sin(\omega t + \delta_1) = \omega V_{10} \cos\left(\omega t + \delta_1 + \frac{\pi}{2}\right)$$

Notice that as t changes, all the directed lines will rotate. However, since all directed lines rotate by the same amount, the relative angles between them remain unchanged with changing t. Very often, the lines are drawn for a definite t, say $t = 0$ and since only the relative phases are of importance it doesnt really matter.

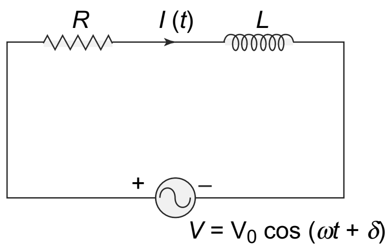


Fig. 9.8 *LR circuit*

In simple cases, the solutions to differential equations involving physical quantities can be easily read off from the phasor diagram. As an example, consider the LR circuit in Fig. 9.8.

Let the current in the circuit be represented by

$$I(t) = I_0 \cos(\omega t + \delta_i)$$

The potential difference across the resistance R is then

$$V_R = IR$$

and across the ideal inductance is

$$V_L = L\frac{dI}{dt}$$

Thus, the current I for a given voltage V is given by

$$V = IR + L\frac{dI}{dt} \tag{9.46}$$

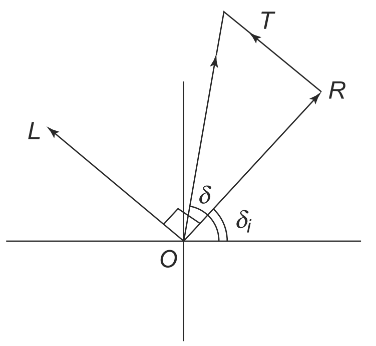


Fig. 9.9 *Phasor diagram for a LR circuit—OR and OL are added vectorially to get OT*

The phasors corresponding to I, V_R, V_L are shown in Fig. 9.9. V_R has an angle δ_i while V_L has an angle $(\delta_i + \frac{\pi}{2})$. As mentioned above, we have set $t = 0$ in the phasor diagram above.

OR and OL are the phasors corresponding to V_R and V_L respectively. Their sum OT must equal $V = V_0\cos(\omega t + \delta)$. Thus, the angle made by OT with the x axis is δ. The magnitude of OT is given by the magnitude of the vector obtained by vectorially adding the vectors (directed lines) OR and OL. Thus,

$$V_0 = OT = \sqrt{OR^2 + OL^2} = \frac{I_0}{\sqrt{R^2 + \omega^2 L^2}}$$

From the diagram it is clear that

$$\tan(\delta - \delta_i) = \frac{TR}{OR} = \frac{\omega L}{R}$$

This gives us the phase difference between the voltage $V = V_0\cos(\omega t + \delta)$ and the current $I = I_0\cos(\omega t + \delta_i)$.

Representing currents and voltages by phasors does not immediately lead to any mathematical convenience for solving for circuit equations given by Kirchhoff's laws. If however, we regard the plane of phasor diagrams as the complex plane, circuit equations involving time varying currents and non-resistive circuit elements become much easier to solve.

To do this, we replace all voltages and currents with complex quantities. The physical voltages and currents are of course the real parts of the corresponding complex quantities. Thus, in the example above, the voltage becomes $V_c = V_0 e^{i(\omega t+\delta)}$ and the current becomes $I_c = I_0 e^{i(\omega t+\delta_i)}$. We can rewrite this in a different way as

$$\begin{aligned} V_c &= V_{c0} e^{i\omega t} \\ V_{c0} &= V_0 e^{i\delta} \\ I_c &= I_{c0} e^{i\omega t} \\ I_{c0} &= I_0 e^{i\delta_i} \end{aligned}$$

The complex potential differences across the resistance and the inductance can be written as

$$\begin{aligned} V_{Rc} &= I_c R \\ V_{Lc} &= L\frac{dI_c}{dt} \end{aligned}$$

Now with these complex quantities we rewrite the Eq. (9.46) as

$$V_c = V_{Rc} + V_{Lc} = RI_c + L\frac{dI_c}{dt} \tag{9.47}$$

It is clear from Eq. (9.47) that the real parts of this satisfy the Eq. (9.46). However, this equation is much easier to solve. The reason for this is that the time-dependence becomes easier to deal with since it is always through a multiplicative factor $e^{i\omega t}$. Thus, differentiation means multiplying with a factor of $i\omega$ and integration means dividing by a factor of $i\omega$. Using these results, we see that the Eq. (9.47) becomes (the exponential time-dependence being the same, cancels out)

$$V_{c0} = RI_{c0} + (i\omega L)I_{c0} \tag{9.48}$$

This is an algebraic equation, unlike Eq. (9.47) and hence, easier to solve, though it has complex quantities. Once we solve for I_{c0}, we can get the physical current as

$$I = \mathrm{Re}\,(I_{c0}e^{i\omega t})$$

and similarly the voltages across the resistance and inductance, V_R and V_L.

We have of course, used a simple circuit to illustrate phasors and the use of complex variables to convert the differential equation involved in solving the circuit equations. However, for any complicated circuit, the basic features which enabled us to do that, are always present. The equations, one or more than one, are equations involving derivatives (when L is present) or integration (when C is present) with all real coefficients involving R, L or C. They thus can be replaced by a complex equation like Eq. (9.47) such that the physical currents and voltages are the real parts of the corresponding complex quantities. Then, just like the simple case above, derivatives can be replaced by multiplication by $i\omega$, integration by division by $i\omega$. The equations, all become algebraic equation just as in the simple case above. Once the equation is solved, physical currents, voltages involved can be extracted as the real parts of the corresponding currents and voltages.

We will make use of this formalism below to solve for voltages and currents in various circuits with alternating current below. Very often the subscript c for complex currents and voltages that we have introduced is dropped, it being understood that the physical quantities are the real parts of the quantities involved.

9.4.3 Impedances and Reactances

As we have already discussed, in circuits where the elements are purely resistive, the relationship between current I through and the potential difference V across any element is simple and linear

$$V = IR$$

This makes the analysis of these circuits relatively simple. If we include time dependent currents and also other elements like inductors and capacitors, then the relationship is not as simple as we have already seen. However, there is a simple mathematical manipulation which can be done which allows us to treat circuits with alternating currents and also inductive and capacitative elements in the same fashion as circuits with purely resistive elements.

Consider the circuit shown in Fig. 9.10.

We have the relationship between the current at any time in the circuit and the charge on the capacitor at that time

$$\begin{aligned} V(t) &= V_0\cos(\omega t) \\ Q(t) &= CV(t) \\ I(t) &= \frac{dQ(t)}{dt} = C\frac{dV(t)}{dt} \end{aligned} \tag{9.49}$$

In this expression, I, V and Q are all real. However, instead of this, let us consider another equation

$$I_c(t) = \frac{dQ_c(t)}{dt} = C\frac{dV_c(t)}{dt} = C\frac{dV_0e^{i\omega t}}{dt} \tag{9.50}$$

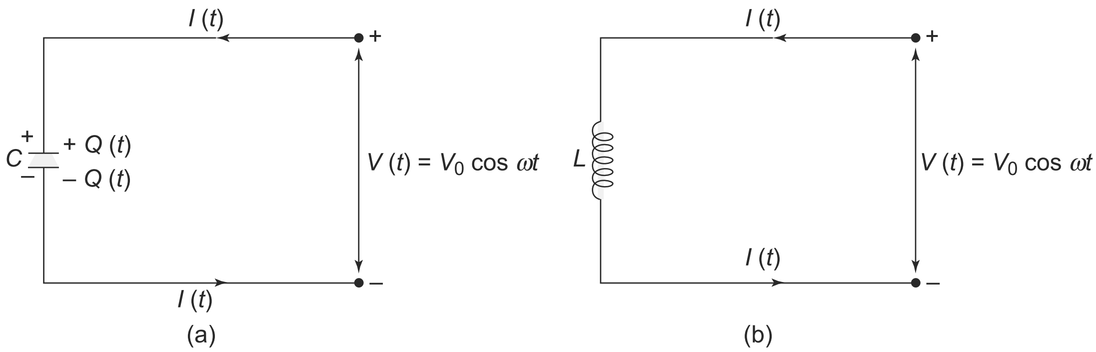


Fig. 9.10 *Circuit elements with alternating current (a) Ideal capacitor, (b) Ideal inductor*

where now I_C, Q_c and $V_c = V_0 e^{i\omega t}$ are all complex. Notice that if we take the real part of Eq. (9.50), we get Eq. (9.49).

$$I(t) = \text{Re}\, I_c(t)$$
$$Q(t) = \text{Re}\, Q_c(t)$$
$$V(t) = \text{Re}\, V_c(t) = \text{Re}\, V_0 e^{i\omega t}$$

This simple mathematical trick actually proves to be very useful in analysing circuits with inductive and capacitative elements and alternating current sources. We thus, have the following strategy to analyse circuits with alternating currents.

Take the potential difference across the source to be complex, $V_C = V_0 e^{i\omega t}$, use Kirchhoff's Laws and solve for complex currents and complex potentials across the branches of the circuits and then take the real part of the currents and potential differences to get the actual currents and potential differences. Thus, in Eq. (9.50), the complex current $I_c(t)$ and complex charge $Q_c(t)$ have the same time dependence as the complex potential of the source. Thus,

$$I_c(t) = I_c^0 e^{i\omega t}$$
$$Q_c(t) = Q_c^0 e^{i\omega t}$$

where I_c^0 and Q_c^0 are still complex. Now if we substitute this in Eq. (9.50), we get

$$I_c^0 e^{i\omega t} = C V_0 (i\omega) e^{i\omega t} \tag{9.51}$$

or

$$\boxed{V_0 = \frac{1}{i\omega C} I_c^0} \tag{9.52}$$

This relationship between the potential across the capacitor and the current in the circuit is very similar to the relationship we get for a pure resistance. Carrying out a similar analysis for a resistance, we see that the current $I_R(t) = I_R^0 e^{i\omega t}$ and potential drop across the resistance are related by

$$V_R^0 = I_R^0 R \tag{9.53}$$

The relationship is similar if we think of $\frac{1}{i\omega C}$ playing the role of the resistance. **This complex quantity is called the reactance of the capacitor Z_C.**

$$Z_c = \frac{1}{i\omega C} \tag{9.54}$$

Note that since ω and C are real, the reactance of the capacitor is purely imaginary! This has an important physical significance. Recall that the physical currents and voltages will be given by taking the real part of the corresponding complex quantities. Hence,

$$V(t) = \mathrm{Re}\,(V^0 e^{i\omega t}) = V^0 \cos\omega t \tag{9.55}$$

and

$$\begin{aligned} I_C(t) &= \mathrm{Re}\,(I_c(t)) \\ &= \mathrm{Re}\left(\frac{V_C^0 e^{i\omega t}}{1/i\omega C}\right) \\ &= \left(\frac{-1}{1/\omega C}\right) V^0 \sin\omega t \\ &= \left(\frac{-1}{1/\omega C}\right) V^0 \cos\left(\omega t + \frac{\pi}{2}\right) \end{aligned} \tag{9.56}$$

In a similar fashion, the current through the resistor would be

$$I_R(t) = \mathrm{Re}\left(\frac{V^0 e^{i\omega t}}{R}\right) = \frac{V^0}{R} \cos\omega t \tag{9.57}$$

From Eq. (9.57), we see that the current in the resistor is in phase with the voltage across it, both of them varying as $\cos(\omega t)$. However, for a capacitor, the current, though at the same frequency (ω) as the voltage across it, is ahead of the voltage by a phase angle of $\frac{\pi}{2}$. The amplitude of the current through it is the same as would be through a resistor of resistance $\frac{1}{\omega C}$. Thus, we see that a reactance in the case of a capacitor is just like a resistance with its imaginary nature reflecting the change of phase between the current through a capacitor and the voltage across it.

We can carry out a similar analysis for the case of an inductor in an alternating current circuit, Fig. 9.10(b). Proceeding exactly as in the capacitor case, we have

$$V(t) = V^0 \cos\omega t = L\frac{dI(t)}{dt} \tag{9.58}$$

In terms of complex quantities, this is

$$V_L(t) = V_0 e^{i\omega t} = L\frac{dI_c(t)}{dt} \tag{9.59}$$

where we can obtain the Eq. (9.58) by taking the real part of Eq. (9.59). Writing

$$I_c(t) = I_c^0 e^{i\omega t}$$

we get

$$V_0 e^{i\omega t} = (i\omega L) I_c^0 e^{i\omega t} \tag{9.60}$$

or

$$I_c^0 = \frac{V_0}{(i\omega L)} \tag{9.61}$$

This is similar to the relationship between I and R for a resistance, except that the resistance R is replaced by $i\omega L$. Thus,

$$\begin{aligned} I(t) &= \text{Re}\, I_c(t) \\ &= \text{Re}\left(\frac{V_0}{i\omega L}e^{i\omega t}\right) \\ &= \left(\frac{V_0}{\omega L}\right)\sin\omega t \\ &= \left(\frac{V_0}{\omega L}\right)\cos\left(\omega t - \frac{\pi}{2}\right) \end{aligned} \tag{9.62}$$

The quantity

$$Z_L = i\omega L$$

is called the reactance of the inductor. Its magnitude, ωL plays the role of the resistance while its imaginary nature signifies the fact that the current through the inductor lags behind the voltage across it by a phase angle of $\frac{\pi}{2}$.

We thus, have the following results. In an alternating current circuit with a voltage $V(t)$, the resistor acts in precisely the same way as it does in a direct current circuit. However, a capacitance has a reactance $Z_c = \frac{1}{i\omega C}$ which implies that the current through a capacitor is ahead of the voltage across it by $\frac{\pi}{2}$. Similarly, an inductance in the circuit has a reactance given by $i\omega L$ and in this case, the current lags the voltage by $\frac{\pi}{2}$.

It is important to remember that the phase relationship between the current and the voltage is the relationship between the current through the circuit element (capacitor or inductance) and the voltage across the same element. This voltage or potential difference across a capacitor or an inductance, may, in general, have a phase relationship with potentials in other parts of the circuit as we shall see.

PROBLEM 9.4 Calculate the current $i(t)$ shown in the circuit in Fig. 9.11. On the extreme right is a constant current source supplying a constant current $4\cos(200t)$.

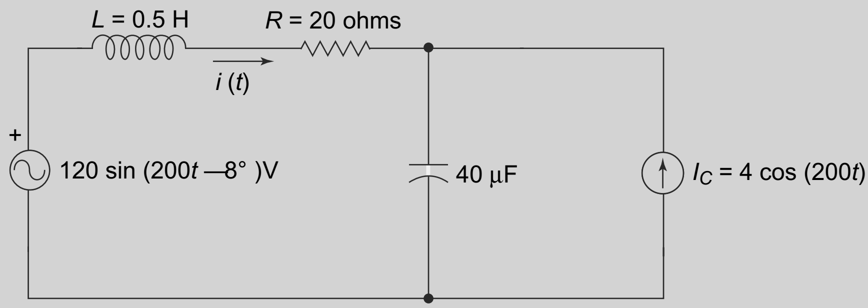


Fig. 9.11 *Problem 9.4*

PROBLEM 9.5 In the circuit in Fig. 9.12, obtain the conditions on the circuit elements so that the potentials at points P and Q remain the same for any value of E and ω.

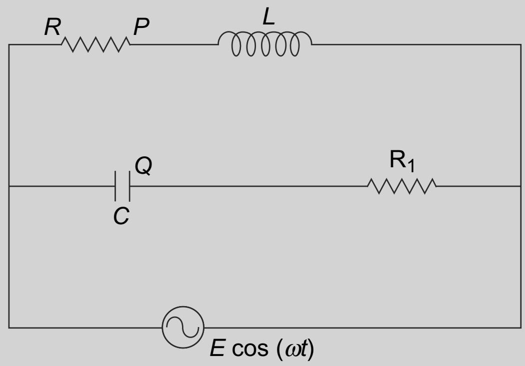


Fig. 9.12 *Problem 9.5*

PROBLEM 9.6 In the circuit shown in Fig. 9.13 calculate the potential difference between points 1 and 2 marked in the figure.

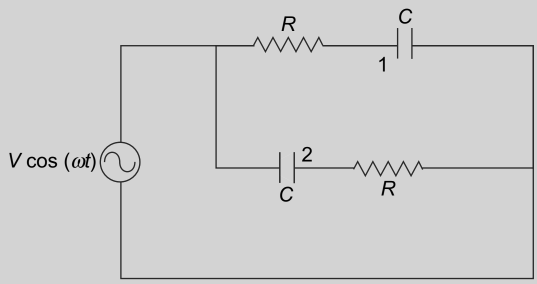


Fig. 9.13 *Problem 9.6*

9.4.4 Complex Voltages and Currents: Kirchhoff's Laws

In the last section, we saw that when we analyse alternating current circuits with non-resistive elements, a useful mathematical trick is to consider complex voltages and currents in the circuit. The non-resistive circuit elements then have complex reactances which we saw above lead to a phase shift between the current and the voltages. We have also earlier remarked that Kirchhoff's laws are valid for time varying or alternating currents provided the time scale of variation is much larger than the time taken for the charges to redistribute in a conductor. The question then arises whether Kirchhoff's laws for circuit analysis are valid for complex currents and voltages? This is a question we now address.

We restrict our discussion at first to a circuit where there is a single power source $V_1 = V_0 \cos \omega t$. Consider an arbitrary circuit with loops and nodes with this power source. The currents in any branch of such a circuit will obviously oscillate with the same frequency ω as the voltage. However, in general, the current in any branch may or may not be in phase with the voltage. This we have already seen is the case whenever there is, for instance, a capacitance or an inductance in the circuit. Therefore, in general, the currents in the branches will be $I = I_0 \cos(\omega t + \delta)$ where δ could be in general different for different branches.

Consider a particular node in such a circuit as in Fig. 9.14.

In this node, n currents, $I_1, I_2, \cdots, I_n$ enter the node and they are all real. These are of the form

$$I_{1i} = I_{1i}^0 \cos(\omega t + \delta_i) \qquad i = 1, 2, \cdots, n$$

Here δ_i is the phase difference between the current I_{1i} and V_1. In this case, the current is ahead of the voltage. Now Kirchhoff's First Law, as we saw is valid for time varying currents and so we have

$$\sum_{i=1}^{n} I_{1i} = 0 \tag{9.63}$$

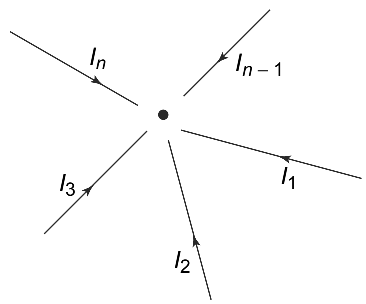


Fig. 9.14 *Currents entering a node*

Now suppose we replace this source V_1 by another source $V_2 = V_0 \sin\omega t = V_0 \cos(\omega t - \frac{\pi}{2})$. Then all the currents will shift in phase by the same amount $-\frac{\pi}{2}$. The currents at the node above will therefore, be

$$I_{2i} = I_{1i}^0 \cos(\omega t - \frac{\pi}{2} + \delta_i)$$

These currents, once again, being real and time varying will obey Kirchhoff's First law at the node, i.e.,

$$\sum_{i=1}^{n} I_{2i} = 0 \tag{9.64}$$

Note that Eqs. (9.63) and (9.64) will be valid for **all** nodes in the circuit. Furthermore, the currents I_{1i} and I_{2i} are related to V_1 and V_2 by **identical linear equations**. Thus, by definition of linearity, if we replace the voltage sources by

$$V = V_1 + iV_2 = V_0 e^{i\omega t}$$

the corresponding currents will be

$$I_i = I_{1i} + iI_{2i} = I_{1i}^0 e^{i(\omega t + \delta_i)}$$

From Eqs. (9.63) and (9.64), we will have

$$\sum I_i = 0$$

We therefore, have a remarkable, though expected result. The complex currents entering any node in a circuit obey Kirchhoff's First Law since the real and imaginary parts separately obey the Kirchhoff's law.

We can carry out a similar analysis for Kirchhoff's Second Law too. Consider a circuit, once again with a single voltage source given by $V = V_1 = V^0 \cos\omega t$. Across any closed loop, the voltage drops across any branch and the currents in the branch will all be real and they will be oscillating with a frequency ω. The voltage drops across any branch, though oscillating at ω will, in general, not be in phase with the power source. Let the potential drop across the i^{th} branch be

$$V_{1i} = Vi^0 \cos(\omega t + \delta_i)$$

Over a closed loop then we will have, by Kirchhoff's Second Law,

$$\sum_i V_{1i} = \sum_i V_i^0 \cos(\omega t + \delta_i) \tag{9.65}$$

The summation is over all the branches of the closed loop under consideration. Now if the source is replaced by another source $V = V_2 = V^0 \sin(\omega t) = V^0 \cos(\omega t - \frac{\pi}{2})$ then the voltage across the i^{th} branch will be

$$V_{2i} = V_i^0 \cos\left(\omega t - \frac{\pi}{2} + \delta_i\right)$$

and Kirchhoff's second law for any closed loop will be

$$\sum_i V_{2i} = \sum_i V_i^0 \sin(\omega t + \delta_i) = \sum_i V_i^0 \cos\left(\omega t - \frac{\pi}{2} + \delta_i\right) = 0 \tag{9.66}$$

Now, since both these sources obey the Kirchhoff's second law, any complex power source

$$V = V^0 e^{i\omega t} = V^0(\cos\omega t + i\sin\omega t)$$

will result in a voltage drop across the i^{th} branch given by

$$V_i = V_{1i} + iV_{2i} = V_i^0 e^{i\omega t + i\delta_i}$$

These complex V_i will obey exactly the same equations, namely Kirchhoff's Second Law Eqs. (9.65 and 9.66).

We illustrate this with an example.

EXAMPLE 9.1 In the circuit below, calculate the currents in the branches shown.

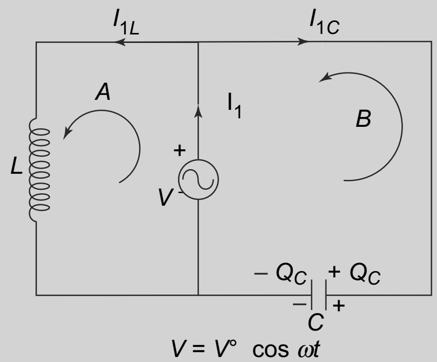


Fig. 9.15 *Example 9.1*

Solution

From Fig. 9.15, it is clear that

$$I_1 = I_{1L} + I_{1C}$$

and

$$I_{1C} = \frac{dQ_C}{dt}$$

We can also apply Kirchhoff's Second Law to get

$$L\frac{dI_{1L}}{dt} = V^0 \cos \omega t$$
$$\frac{Q_c}{C} = V^0 \cos \omega t$$

Solving the differential equation for I_{1L} we get

$$I_{1L} = \frac{V^0}{\omega L} \sin(\omega t) + A$$

where A is a constant to be determined by initial conditions. Assuming that $I_{1L} = 0$ at $t = 0$ or when the power source is switched on, we get $A = 0$. We also have

$$I_{1C} = -\omega V^0 C \sin(\omega t) + C_1$$

Once again, if we assume that the current through the capacitor is zero at $t = 0$, the constant C_1 vanishes. We, thus, have

$$I_1 = I_{1L} + I_{1C} = V^0 \sin(\omega t)\left[\frac{1}{\omega L} - \omega C\right]$$

We now replace V by $V^0 \sin(\omega t)$ and repeat the steps above. The currents are now I_2, I_{2L} and I_{2C} and

$$I_2 = I_{2L} + I_{2C}$$

The current in the inductor can be found by solving the differential equation and we get

$$I_{2L} = -\frac{V^0}{\omega L} \cos(\omega t) + C_2$$

The constant C_2 is to be determined by the initial condition at $t = \frac{\pi}{2\omega}$, i.e., demanding that $I_{2L}(t = \frac{\pi}{2\omega}) = 0$. Thus, we get $C_2 = 0$. Similarly, solving for I_{2C}, we get

$$I_{2C} = V^0 \omega C \cos(\omega t) + C_3$$

and the constant $C_3 = 0$. Therefore the current I_2 is given by

$$I_2 = I_{2L} + I_{2C} = V^0 \cos(\omega t)\left[-\frac{1}{\omega L} + \omega C\right] = V^0 \sin(\omega t - \frac{\pi}{2})\left[\omega C - \frac{1}{\omega L}\right]$$

If we now consider a complex power source

$$V = V^0 e^{i\omega t}$$

the complex currents will be

$$I_C = I_{1C} + iI_{2C} = (i\omega C)V^0 e^{i\omega t}$$
$$I_L = I_{1L} + iI_{2L} = \frac{-iV^0}{\omega L} e^{i\omega t}$$

$$I = I_C + I_L = V^0 e^{i\omega t}\left[i\omega C + \frac{1}{i\omega L}\right]$$

It is easy to check with these complex currents and voltages that the loop equations are satisfied. For loop A, the drop in voltage across the inductance is given by $I_L Z_L$ where Z_L is the reactance for the inductor.

$$I_L Z_L = \frac{-iV^0}{\omega L} e^{i\omega t}\,(i\omega L) = V^0 e^{i\omega t}$$

For loop B

$$-I_C Z_C = -(i\omega C)V^0 e^{i\omega t}\frac{1}{i\omega C} = -V^0 e^{i\omega t}$$

These are precisely the Kirchhoff loop equations

Finally, we need to consider circuits with multiple alternating current power sources, all of which may or may not be in phase. Let one of the EMF's be $V^0\cos(\omega t)$. Then the other EMFs, say the i^{th} one will be $V_i\cos(\omega t + \delta_i)$. Of course $V^0\cos(\omega t)$ is the real part of $V^0 e^{i\omega t}$ while $V_i\cos(\omega t + \delta_i)$ is the real part of $V_i e^{(i\omega t + i\delta)}$. Note that the factor $e^{i\omega t}$ will be present in all of the voltages and currents. This can be factored out and we will be left with the potential drops across the inductive and capacitative and resistive elements, one voltage $V^0\cos(\omega t)$ and the other voltages $V_i\cos(\omega t + \delta_i)$. It is important to realise that these are all complex EMFs and hence the phase is very important. We illustrate this with an example of a circuit with multiple EMF's.

EXAMPLE 9.2 In the circuit below, shown in Fig. 9.16, calculate the currents in the branches shown. It is given that the two source EMFs V_1 and V_2 are both sinusoidal with the same amplitude. However, V_2 is $\frac{\pi}{2}$ ahead of V_1. We also know that $\omega^2 = \frac{1}{LC}$ and $R = \frac{\omega L}{2}$

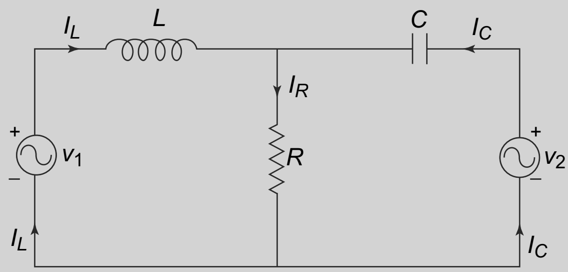


Fig. 9.16 *Example 9.2*

Solution

We take

$$V_1 = V^0\cos\omega t$$
$$V_2 = V^0\cos(\omega t + \frac{\pi}{2})$$

In complex notation, these can be written as

$$V_1 = V^0 e^{i\omega t}$$
$$V_2 = V^0 e^{i\omega t} e^{i\frac{\pi}{2}}$$

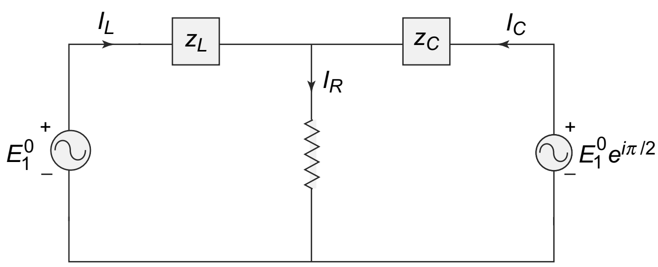


Fig. 9.17 *Example 9.2*

where we have dropped the subscript c for complex voltages. The circuit, with these complex voltages, currents and impedances (reactances) for the circuit elements looks like the one shown in Fig. 9.17 where we have used E_1^0 instead of V^0 and factored out $e^{i\omega t}$ from all voltages. The impedances Z_L and Z_C are as before, $Z_L = i\omega L$ and $Z_C = \frac{1}{i\omega C}$.

The loop equations can be written down easily with these complex voltages and currents and impedances as

$$V^0 = E_1^0 = I_L Z_L + R(I_L + I - C)$$
$$V^0 e^{i\frac{\pi}{2}} = E_1^0 e^{i\frac{\pi}{2}} = Z_C I_C + R(I_C + I_L)$$

In our case, we have

$$Z_L = i\omega L$$
$$Z_C = \frac{1}{i\omega C} = -i\omega L = -Z_L = -Z$$

With this, the loop equations can be solved easily to give

$$Z(I_C + I_L) = E_1^0(1 - e^{i\frac{\pi}{2}}) = E_1^0(1 - i)$$
$$2R(I_L + I_C) + Z(I_L - I_C) = E_1^0(1 + e^{i\frac{\pi}{2}}) = E_1^0(1 + i)$$

These equations can be solved easily and using the values of Z and R, we get

$$I_L = \frac{E_1^0}{\omega L}(\frac{1}{2} - 2i)$$
$$I_C = \frac{E_1^0}{\omega L}(i - \frac{3}{2})$$

The physical currents are therefore, reintroducing the factor of $e^{i\omega t}$,

$$I_L = \mathrm{Re}(I_L e^{i\omega t}) = \frac{E_1^0}{\omega L}\left[\frac{\cos\omega t}{2} + 2\sin\omega t\right]$$
$$I_C = \mathrm{Re}(I_C e^{i\omega t}) = -\frac{E_1^0}{\omega L}\left[\frac{3\cos\omega t}{2} + \sin\omega t\right]$$

9.4.5 Impedances in Series and Parallel

We have already seen in Chapter 5 that the series and parallel combination laws (which are consequences of Kirchhoff's laws) for resistors are useful in analysing circuits. These can be generalised when capacitors and inductors are included in the circuit as we have also seen. For complex currents and voltages, we can also establish similar laws for parallel and series combination of impedances, i.e., resistors, capacitors and inductors.

Let us first consider the **series combination**. Consider two circuit elements in series as in Fig. 9.18(a). These could be resistors, capacitors or inductors in general. The voltage to current ratio is denoted by Z for each of them. Thus, for a pure resistor, $Z = R$ while for a capacitor, $Z = \frac{1}{i\omega C}$ and for an inductor, $Z = i\omega L$. Since the two elements are in series, the same current I flows in both of them. Let the complex voltage across Z_1 be V_1 while that across Z_2 be V_2. Then

$$V_1 = Z_1 I$$
$$V_2 = Z_2 I \tag{9.67}$$

The total potential difference across the combination is obviously

$$V = V_1 + V_2 = (Z_1 + Z_2)I$$

We see that the voltage to current ratio thus, is simply the sum of the two impedances $Z_1 + Z_2$ exactly as the effective resistance of a series combination of two resistors is the sum of the two resistances. The effective impedance of the circuit is therefore,

$$Z = Z_1 + Z_2$$

This can be trivially extended to a series combination of n impedances and we get the

$$\boxed{Z = Z_1 + Z_2 + \cdots + Z_n} \tag{9.68}$$

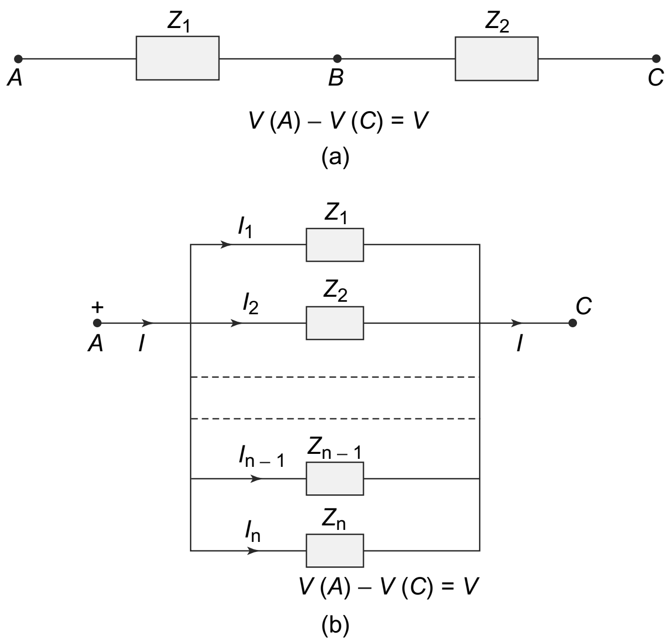


Fig. 9.18 *Combination of Impedances: (a) Series combination of impedances, (b) Parallel combination of impedances*

We next consider a parallel combination of impedances as shown in Fig. 9.18(b). Here we have a set of n impedances, $Z_1, Z_2, \cdots, Z_n$ connected in parallel. Let the currents through these impedances be $I_1, I_2, \cdots, I_n$. These currents will in general be complex. Now since each of the impedance is connected between the same two points, A and C, they have the same voltage across them. Therefore,

$$\begin{aligned} V &= I_1 Z_1 \\ V &= I_2 Z_2 \\ \vdots &= \vdots \\ V &= I_n Z_n \end{aligned} \tag{9.69}$$

Hence, we get for the total current I

$$I = I_1 + I_2 + \cdots + I_n = V\left(\frac{1}{Z_1} + \frac{1}{Z_2} + \cdots + \frac{1}{Z_n}\right) \tag{9.70}$$

The effective impedance therefore, is

$$Z = \frac{V}{I} = \left[\left(\frac{1}{Z_1} + \frac{1}{Z_2} + \cdots + \frac{1}{Z_n}\right)\right]^{-1} \tag{9.71}$$

or

$$\boxed{\frac{1}{Z} = \frac{1}{Z_1} + \frac{1}{Z_2} + \cdots + \frac{1}{Z_n}} \tag{9.72}$$

Although the form of the effective impedances for the parallel and series combinations is the same as that of combinations of resistances, the complex nature of the impedances leads to certain results which are not quite obvious. As an example, consider the two circuits given in Fig. 9.19. One is a series combination of an inductance L and a capacitance C while the other is a parallel combination of the same two elements.

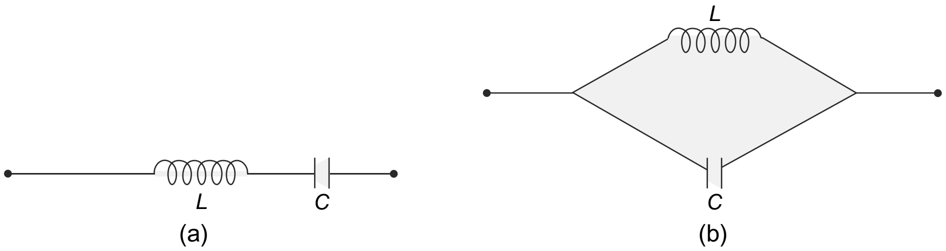


Fig. 9.19 *Series and parallel combinations (a) L and C in series (b) L and C in parallel*

For the series combination, the equivalent impedance is

$$Z_S = Z_L + Z_C = i\left(\omega L - \frac{1}{\omega C}\right) \tag{9.73}$$

and for the parallel combination,

$$\frac{1}{Z_P} = \frac{1}{Z_L} + \frac{1}{Z_C} = i\left(\omega C - \frac{1}{\omega L}\right) \tag{9.74}$$

Now consider a situation where the frequency of the alternating current is $\omega^2 = \frac{1}{LC}$. We see a very curious thing—the series combination gives us zero impedance while the parallel combination gives us infinite impedance! At this frequency, the parallel combination will not allow any current to flow. Thus, we see that even though the form of the rules for combining impedances in series and parallel are similar to that of resistances, the complex nature of the impedances introduces subtlety which can lead to non-intuitive results.

9.4.6 Power in Alternating Current Circuits

We have discussed the phenomenon of Joule heating in a conductor or a pure resistor. For a steady current, we have seen that the power P dissipated in a resistor R is constant in time and is given by

$$P = I^2 R = VI$$

where I is the current flowing in the resistor and V is the voltage across it.

For an alternating current, this power of course will vary with time. Consider an impedance Z through which flows a current $I(t)$ and across which there is a voltage $V(t)$. These are, as we have seen given by

$$I(t) = \mathrm{Re}\,(Ie^{i\omega t})$$
$$V(t) = \mathrm{Re}\,(Ve^{i\omega t})$$
$$Z = \frac{V}{I}$$

Here, V and I can in general be complex quantities. Let us write these as

$$V = V_0 e^{i\delta_V}$$

and

$$I = I_0 e^{i\delta_I}$$

where now V_0 and I_0 are real quantities and being magnitudes of the current and voltage, are positive. The instantaneous power dissipated in the impedance is now time-dependent and can be written as

$$\begin{aligned} P(t) &= V(t)I(t) \\ &= V_0 \cos(\omega t + \delta_V) I_0 \cos(\omega t + \delta_I) \end{aligned} \tag{9.75}$$

The instantaneous power given by Eq. (9.75) is real but can be of either sign depending on the arguments of the cosine functions. This is of course, just a reflection of the fact that in general, when an alternating current passes through circuit elements, the current and the voltage are not in phase. Averaging over a whole cycle or period of $T = \frac{2\pi}{\omega}$, we get the average power dissipated in the circuit as

$$\begin{aligned} P =< P(t) > &= \frac{1}{T}\int_0^T P(t)dt \\ &= \frac{V_0 I_0}{2}\cos(\delta_V - \delta_I) \\ &= \frac{1}{2}\mathrm{Re}(VI^*) \end{aligned} \tag{9.76}$$

Note that the average power dissipated depends on the angle $\theta \equiv \delta_V - \delta_I$, that is the phase difference between the voltage across the impedance and the current through it. **The quantity $\cos\theta$ is called the power factor for the impedance.**

Power factor is an important quantity for alternating circuits. For a purely resistive circuit element, $\theta = 0$ and so the power factor is 1. On the other hand, for a purely inductive or capacitative circuit element, the voltage and current have a phase difference $\theta = \frac{\pi}{2}$ and so the power factor is zero.

9.5 *LR* CIRCUITS WITH ALTERNATING CURRENTS

In Chapter 8, we have already analysed the behaviour of a circuit with a pure inductance L and a pure resistance R joined in series with a direct current source. The current in the circuit is transient and decays with a characteristic time constant which depends on the value of the resistance and the inductance. We now consider a similar circuit but this time with an alternating current source as in Fig. 9.20(a).

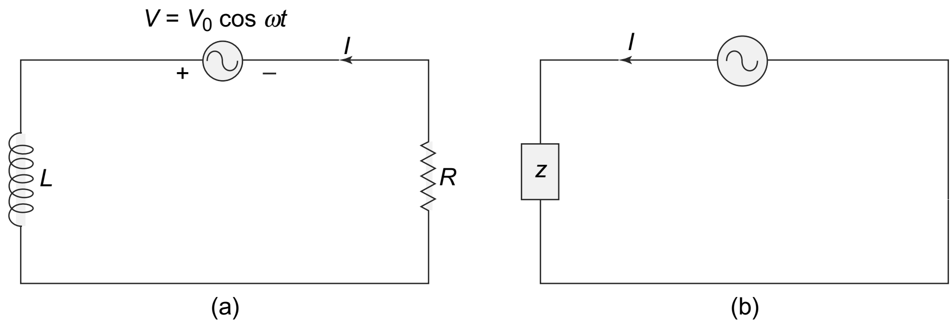


Fig. 9.20 *(a) LR circuit with alternating current source, (b) Equivalent Circuit*

The alternating current source has a voltage

$$V(t) = V_0 \cos \omega t$$

Kirchhoff's laws for the loop gives us then

$$L\frac{dI(t)}{dt} + I(t)R = V_0 \cos \omega t \tag{9.77}$$

This differential equation is similar to the one we encountered for an LR circuit with a direct current source with an important difference. The equation is what is called an inhomogeneous equation because of the presence of the source term which is a function of t. It is easy to see that to any solution of this equation, we can add a solution to the homogenous equation (the equation obtained by putting the source term to be zero). The general solution to the inhomogeneous equation is therefore, given by

$$I(t) = I_P(t) + I_H(t) \tag{9.78}$$

where $I_P(t)$ is a solution (called a particular solution) of the inhomogeneous equation and $I_H(t)$ is the solution to the homogenous equation. The particular solution is proportional to the source term while $I_H(t)$ satisfies

$$L\frac{dI_H(t)}{dt} + I_H(t)R = 0 \tag{9.79}$$

This homogenous equation is exactly the one which we have already analysed with direct currents and hence, the solutions are identical to the those already obtained. The particular solution $I_P(t)$ is of course something we need to obtain. This solution will be oscillatory in nature since as we have noted, the particular solution is in general proportional to the source term which is oscillatory. We use the method of complex currents and voltages as outlined above.

As we have seen, we can write $I_P(t)$ as

$$I_P(t) = \mathrm{Re}(I_P e^{i\omega t}) \tag{9.80}$$

where I_P is in general, complex. We need to determine I_P which will have a magnitude and a phase. We see that the circuit has two circuit elements, an inductance L and a resistance R in series. The combined impedance of these two then is simply

$$Z = i\omega L + R \tag{9.81}$$

and the current I_P has a magnitude I_0 given by

$$I_0 = \frac{V_0}{|Z|} = \frac{V_0}{\sqrt{\omega^2 L^2 + R^2}} \tag{9.82}$$

and a phase given by

$$\theta = \arctan(-\frac{\omega L}{R})$$

The particular solution is therefore, from Eq. (9.80),

$$I_P(t) = \mathrm{Re}(I_P e^{i\omega t}) = \frac{V_0}{\sqrt{\omega^2 L^2 + R^2}} \cos(\omega t + \theta) \tag{9.83}$$

To this solution, we must add the solution to the homogenous equation and then ensure that the initial conditions imposed on the currents are satisfied. The solution to the homogenous equation is transient which dies out. This particular solution is of course oscillatory in nature with a frequency ω, the same as that of the source but shifted in phase by an angle θ which depends on the values of L and R. Note that, as expected, if $L = 0$, i.e., the circuit is purely resistive, there is no phase shift of the current with respect to the source voltage. On the other hand, if $R = 0$, i.e., the circuit is purely inductive, the phase shift is $\frac{\pi}{2}$, which is what we expect for a pure inductance, as we saw earlier.

PROBLEM 9.7 In the circuit in Fig. 9.21, the capacitor is initially uncharged and both the switches are in the off position. At time $t = 0$, switch $S1$ is switched on. At time $t = 0.1$ msec, switch $S1$ is switched off and at time $t = 0.2$ msec switch $S2$ is switched on. Calculate the charge on the capacitor at time $t = 0.3$ msec.

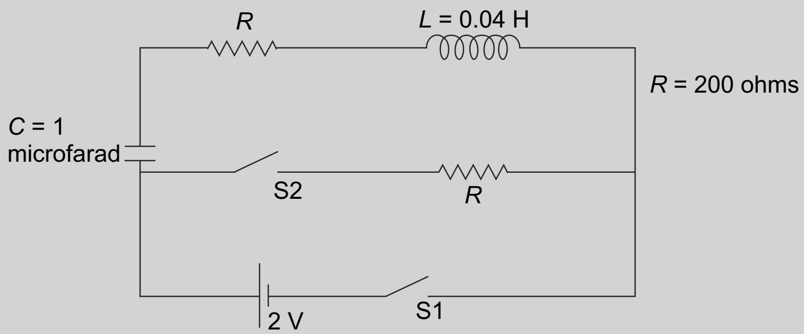


Fig. 9.21 *Problem 9.7*

9.6 *LCR* CIRCUITS WITH ALTERNATING CURRENTS

9.6.1 Series *LCR* Circuit

We next turn to a series combination of an inductor, capacitor and a resistor connected to an alternating current source. The circuit with and without a direct current source has already been discussed earlier. The circuit we are considering is given in Fig. 9.22(a) with its equivalent circuit given in Fig. 9.22(b).

We can once again write the differential equation governing the current in the circuit using Kirchhoff's laws. The equation will be an inhomogeneous equation with a source term ($V_0 e^{i\omega t}$) exactly as in the

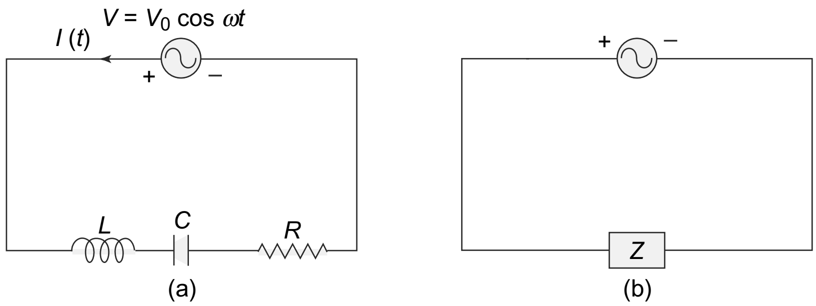


Fig. 9.22 *(a) LCR circuit with alternating current source, (b) Equivalent circuit*

LR circuit above. The current is transient decaying at a rate given by the values of R, L and C. The solution to the homogenous equation, i.e., with $V_0 = 0$, we have already considered. We now try to find the particular solution, which we expect to be oscillatory. The particular solution to the inhomogeneous equation is obtained by once again considering complex currents and voltages. We write the current to be

$$I_P(t) = \mathrm{Re}\,(I_0 e^{i\omega t})$$

where once again I_0 will be complex in general.

The effective impedance of the circuit of L, R and C in series can be easily written as

$$Z = R + i\left(\omega L - \frac{1}{\omega C}\right) \tag{9.84}$$

We can therefore, write the magnitude of the current I_0 as

$$|I_0| = \frac{V_0}{|Z|} = \frac{V_0}{\sqrt{R^2 + (\omega L - \frac{1}{\omega C})^2}} \tag{9.85}$$

and the phase as

$$\delta = \arctan\left(-\frac{\omega L - \frac{1}{\omega C}}{R}\right) \tag{9.86}$$

The oscillatory part of the current in the circuit is therefore, given by

$$\boxed{I_P(t) = \mathrm{Re}(I_0 e^{i\omega t}) = \frac{V_0}{\sqrt{R^2 + (\omega L - \frac{1}{\omega C})^2}} \cos(\omega t + \delta)} \tag{9.87}$$

It is instructive to plot the magnitude of the current as a function of the frequency of the alternating current source, Fig. 9.23.

The curve is a bell-shaped curve which peaks at a value of $\omega = \omega_R = \frac{1}{\sqrt{LC}}$. **This value of the frequency is called the resonance frequency of the circuit.**

$$\omega_R = \frac{1}{\sqrt{LC}}$$

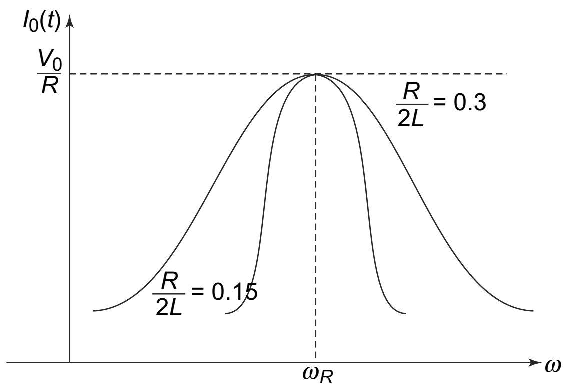


Fig. 9.23 *Plot of current versus frequency ω in a series LCR circuit. The sharpness of the curve increases as $\frac{R}{2L}$ becomes smaller*

The current has a maximum value at this frequency and is given by $\frac{V_0}{R}$. This is expected since at this frequency, the impedance effects of the inductance and the capacitance cancel out exactly and we are left with a purely resistive impedance. The sharpness of the response curve depends on the values of R, L and C.

The behaviour of the current in the series LCR circuit as a function of the frequency is identical to the response of the amplitude of a driven, damped mechanical oscillator. For such an oscillator, the amplitude is a maximum when the driving frequency equals the natural frequency of the oscillator. In this case too, the natural frequency of the circuit is $\frac{1}{\sqrt{LC}}$ as we have seen above. When the driving frequency, i.e., the frequency of the alternating current source equals this, we get resonant behaviour and the amplitude of the current is a maximum.

EXAMPLE 9.3 Figure 9.24 shows an *LCR* circuit with an alternating current source. It is given that $V_{\text{rms}} = 100$ V, $V_0 = 141$ V, $L = 20$ milli Henry, $C = 20$ microFarad and $R = 20$ ohms with the frequency $f = 400$ Hz. Find the values of X_L, X_C, Z, I_{rms}, V_{rms}^L, V_{rms}^C, V_{rms}^R

Solution

$$X_L = 2\pi \times 400(20 \times 10^{-3}) = 50.3 \text{ ohms}$$

$$X_C = \frac{1}{2\pi \times 400 \times (20 \times 10^{-6})} = 19.9 \text{ ohms}$$

$$Z = \sqrt{20^2 + (50.3 - 19.9)^2} = 36.4 \text{ ohms}$$

$$I_{\text{rms}} = \frac{1}{\sqrt{2}}\left(\frac{141}{36.4}\right) = 2.75 \text{ A}$$

$$V_{\text{rms}}^L = 2.75 \times 50.3 = 138.3 \text{ V}$$

$$V_{\text{rms}}^C = 2.75 \times 19.9 = 54.7 \text{ V}$$

$$V_{\text{rms}}^R = 2.75 \times 20 = 55 \text{ V}$$

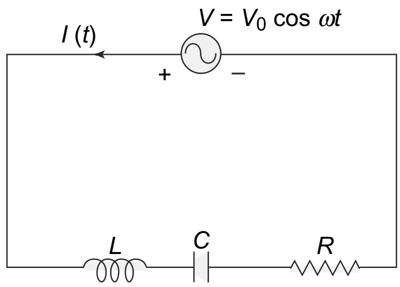


Fig. 9.24 *Example 9.3: Series LCR circuit with alternating currents*

Note that the sum of the rms voltages do not add up $V_{\rm rms}$ - this is because the rms voltages across the various elements are not in phase.

9.6.2 Q-factor: Sharpness of Response

The response of the LCR circuit to an alternating current, as we have seen is a bell-shaped curve which peaks at the resonant frequency ω_R. The curve also depends on the values of $\frac{R}{2L}$ since the sharpness of the peak of the curve is more for lower values of this quantity. Let us try to quantify this.

To study the sharpness of resonance, let us expand the quantity

$$\omega L - \frac{1}{\omega C}$$

in the neighbourhood of the resonant frequency $\omega_R = \frac{1}{\sqrt{LC}}$. We write

$$\omega = \omega_R + \Delta\omega$$

where $\frac{\Delta\omega}{\omega_R} \ll 1$. Then

$$\begin{aligned}\omega L - \frac{1}{\omega C} &= (\omega_R + \Delta\omega)L - \frac{1}{(\omega_R + \Delta\omega)C} \\ &\simeq \ \Delta\omega\left(L + \frac{1}{C\omega_R^2}\right) \\ &= 2L\Delta\omega \end{aligned} \tag{9.88}$$

where we have used the fact that $\omega_R = \frac{1}{\sqrt{LC}}$. Recall that

$$I_0 = \frac{V_0}{Z} = \frac{V_0}{\sqrt{R^2 + (\omega L - \frac{1}{\omega C})^2}} \tag{9.89}$$

and so in the neighbourhood of the resonance, we have

$$I_0 = \frac{V_0}{\sqrt{R^2 + 4\Delta\omega^2 L^2}} \tag{9.90}$$

Now suppose

$$\Delta\omega = \Delta\omega_0 = \pm\frac{R}{2L}$$

that is, we go to a frequency $\omega_R \pm \Delta\omega_0$. The current at this frequency, as in Eq. (9.90) is then $\frac{V_0}{\sqrt{2}R}$ that is it falls to a value $\frac{1}{\sqrt{2}}$ of its peak value at resonance. We therefore, can use $\Delta\omega_0$ as a measure of the sharpness of the resonance. In practice, we define a dimensionless quantity **Q-factor,** Q as

$$\frac{1}{Q} \equiv \frac{2\Delta\omega_0}{\omega_R} = \left(\frac{R}{\omega_R L}\right) \tag{9.91}$$

The higher the value of Q, the sharper is the resonance. We can of course, use a similar definition of Q-factor for any oscillator.

9.6.3 Phase Relationship between I and V. Power Transfer

The current $I(t)$ in a series LCR circuit is, as we have seen above, not in phase in general, with the voltage $V(t)$. Figure 9.25(a) shows the phase relationship between the two. We also show the behaviour of the phase difference δ as a function of the frequency ω in Fig. 9.25(b). Note that the phase difference vanishes for $\omega = \omega_R$ the resonance frequency. Also, the shape of the curves depends on the value of R and L or the Q-factor. The higher the Q factor, the flatter the curve.

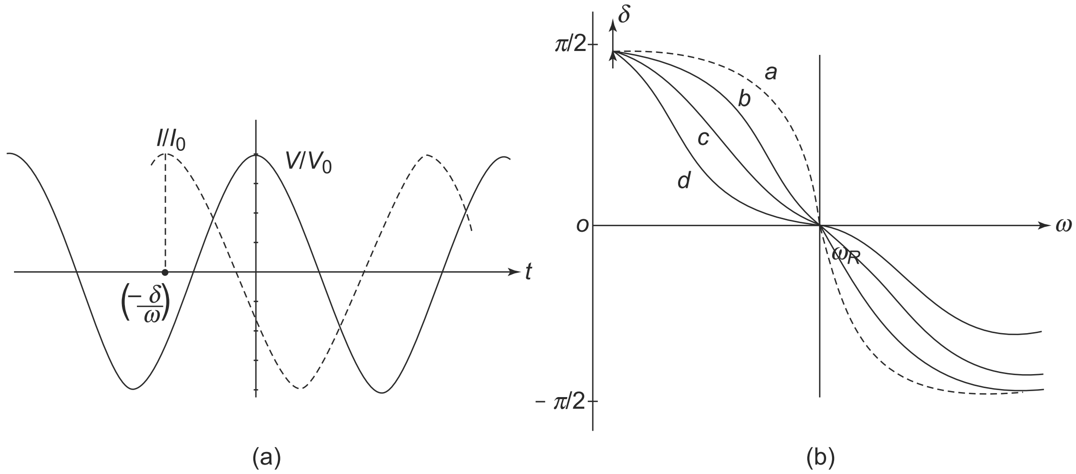


Fig. 9.25 *Phase relationship (a) Phase relationship between I and V in a series LCR circuit, (b) Phase difference δ vs ω. Curve (a) is for $R \to 0$. As we increase R, curves take the form (b) and (c). Curve (d) is when $R \gg \omega L$*

The power averaged over a time period for any circuit was calculated in Eq. (9.76). For an LCR circuit that we are considering, the average power dissipated in the effective impedance Z is thus

$$< P(t) >= \frac{V_0|I_0|}{2}\cos(\delta) \tag{9.92}$$

At resonance, we have already seen that $\delta = 0$ and hence the power supplied by the source is a maximum at resonance. This fact is of immense importance in the design of practical resonant circuits with alternating currents.

9.6.4 Parallel *LCR* Circuit and Antiresonance

We now consider a circuit in which an inductance L, a capacitance C and a resistance R are connected in parallel to an alternating current source Fig. 9.26(a). We will see that this circuit will exhibit an interesting phenomenon of anti-resonance.

Let the real currents in the three circuit elements are $i_L(t)$, $i_C(t)$ and $i_R(t)$ with the total current of course being $i(t) = i_L(t) + i_C(t) + i_R(t)$. If I_L, I_C and I_R are the complex currents through the three elements, we then have

$$i_L(t) = \text{Re}\, I_L$$

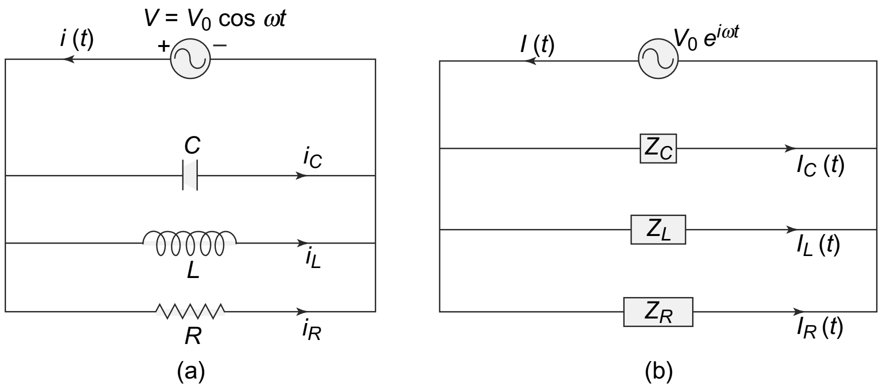


Fig. 9.26 *Parallel LCR circuit: (a) A parallel combination of L,C and R in an alternating current circuit, (b) Equivalent circuit with complex currents*

$$i_C(t) = \text{Re}\, I_C$$
$$i_R(t) = \text{Re}\, I_R$$

Since all three elements have the same potential $V(t) = V_0 \cos \omega t$ across them, the currents through them can be written down easily as

$$I_a(t) = \frac{V(t)}{Z_a}$$

where a is L, C or R. Thus, we get

$$\begin{aligned} I_L &= \frac{V_0 e^{i\omega t}}{i\omega L} \\ I_C &= \frac{V_0 e^{i\omega t}}{1/(i\omega C)} \\ I_R &= \frac{V_0 e^{i\omega t}}{R} \end{aligned} \tag{9.93}$$

The total complex current $I(t) = I_L(t) + I_C(t) + I_R(t)$ is thus, given by

$$\begin{aligned} I &= V_0 e^{i\omega t}\left[\frac{1}{R} + i\left(\omega C - \frac{1}{\omega L}\right)\right] \\ &= V_0 e^{i\omega t + i\theta}\left[\frac{1}{R^2} + \left(\omega C - \frac{1}{\omega L}\right)^2\right]^{1/2} \end{aligned} \tag{9.94}$$

where

$$\theta = \arctan\left(\frac{\omega C - \frac{1}{\omega L}}{\frac{1}{R}}\right) \tag{9.95}$$

The real current $i(t)$ is simply the real part of $I(t)$ and is given by

$$i(t) = V_0\left[\frac{1}{R^2} + \left(\omega C - \frac{1}{\omega L}\right)^2\right]^{1/2} \cos(\omega t + \theta) \tag{9.96}$$

The real currents in the three circuit elements can be also found out as

$$
\begin{aligned}
i_L &= \text{Re}\, I_L(t) = \frac{V_0}{(\omega L)}\sin(\omega t)\\
i_C &= \text{Re}\, I_C(t) = -V_0(\omega C)\sin(\omega t)\\
i_R &= \text{Re}\, I_R(t) = \frac{V_0}{R}\cos(\omega t)
\end{aligned}
\tag{9.97}
$$

The phases of the three currents are different from each other and also from the total current. If we plot the total real current i as a function of ω, we see an interesting feature (Fig. 9.27).

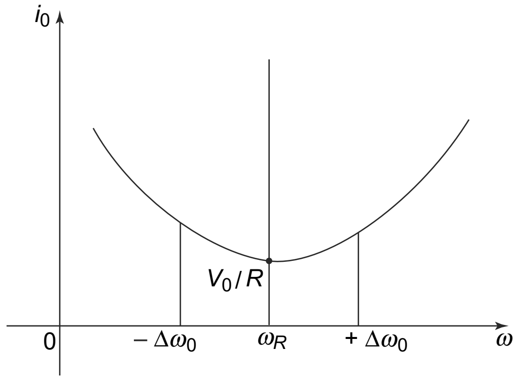


Fig. 9.27 *Plot of current amplitude i_0 versus frequency ω in a parallel LCR circuit. At $\omega = \omega_R = \frac{1}{\sqrt{LC}}$ the current is minimum at a value of $\frac{V_0}{R}$. At values of $\omega \pm \Delta\omega_0$ where $\Delta\omega_0 = \frac{1}{2CR}$ the current rises to $\sqrt{2}$ times the minimum value*

At a frequency $\omega_R = \frac{1}{\sqrt{LC}}$, we have the opposite effect of a series LCR circuit. In a series circuit, we saw the phenomenon of resonance at this frequency and the current was a maximum. Here, we see an anti-resonance and the current is a minimum! However, as with a series LCR circuit, here too the sharpness of the response curve depends on the relative value of $(\omega C - \frac{1}{\omega L})$ and $\frac{1}{R}$. Thus, for $\omega = \omega_R + \Delta\omega_0$ where $\frac{\Delta\omega_0}{\omega_R} \ll 1$, we can write

$$
\begin{aligned}
\omega C - \frac{1}{\omega L} &= (\omega_R + \Delta\omega_0)C - \frac{1}{L}(\omega_R + \Delta\omega_0)^{-1}\\
&\simeq \ \Delta\omega_0\left(C + \frac{1}{\omega_R^2 L}\right)\\
&= 2\Delta\omega_0 C
\end{aligned}
\tag{9.98}
$$

With this, we can easily find the current amplitude to be

$$
\begin{aligned}
i_0 &= -V_0\left[\frac{1}{R^2} + 4(\Delta\omega)^2 C^2\right]^{1/2}\\
&= (i_0)_{\min}\left[1 + 4(\Delta\omega)^2 C^2 R^2\right]^{1/2}
\end{aligned}
\tag{9.99}
$$

where $(i_0)_{\min} = \frac{V_0}{R}$.

When $\Delta\omega = \frac{1}{2CR}$, we see that

$$i_0 = \sqrt{2}(i_0)_{\min}$$

that is the current amplitude goes up by $\sqrt{2}$ at this value of ω as compared to the anti-resonant frequency ω_R. At this point, the quantity $\frac{2\Delta\omega}{\omega_R}$ is

$$\begin{aligned}\frac{2\Delta\omega}{\omega_R} &= \frac{1}{RC}\sqrt{LC}\\ &= \frac{1}{R}\sqrt{\frac{L}{C}}\end{aligned} \tag{9.100}$$

This quantity is defined to be the inverse of the Q-factor for the parallel LCR circuit. Thus,

$$Q_{\text{parallel}LCR} = R\sqrt{\frac{C}{L}} \tag{9.101}$$

This expression is the inverse of the Q factor for the series LCR circuit given in Eq. (9.91).

PROBLEM 9.8 A series LCR circuit is to be designed using a 10 mH inductor, a capacitor C and a resistor R with an AC source $V = V_0 \cos(\omega t)$. The inductor however is **not** resistanceless and has a resistance of 70 Ohms. The design of the circuit calls for a resonant frequency of 20 KHz and a bandwidth $\Delta\omega$ not higher than 2.4 KHz.

(a) What value of capacitance should be used?

(b) What is the maximum value of the resistance R?

(c) When the maximum resistance R found in part (b) is used, what is the voltage across R?

PROBLEM 9.9 In a circuit in Fig. 9.28, a resistor of resistance R and an inductor of inductance L is in series; this series combination is parallel to another combination - the latter one being a series combination of a resistor of the same value of resistance R and a capacitor of capacitance C. Obtain the condition under which the total impedance of the whole combination would be frequency independent.

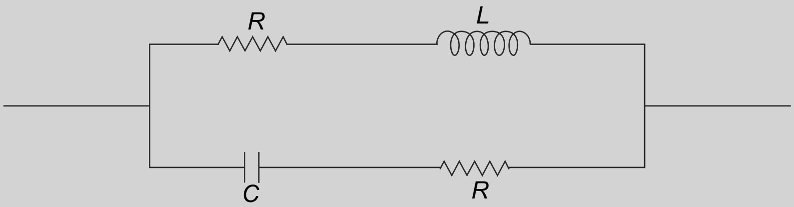


Fig. 9.28 *Problem 9.9*

PROBLEM 9.10 A non-ideal inductor L has an inductance of 4μH and a resistance of 45 Ohms. What is the phase angle ϕ at a supply voltage of frequency of 500 cycles /sec? What should be the value of the capacitance in parallel with the inductor such that the whole combination behaves like a pure resistor at the same frequency ? With such a capacitor connected, if the frequency of the supply is increased by 50 cycles, would the phase of the impedance of the combination exceed the value ϕ of the inductor by itself ?

9.7 MUTUAL INDUCTANCE AND TRANSFORMERS

9.7.1 Mutual Inductance in Alternating Current Circuits

We have already studied the phenomenon of mutual inductance between two inductances. This happens when the flux in one is linked to the other and therefore, a change in flux in one of them, causes an induced EMF in the other one. The induced EMF is proportional to the rate of change of of current in the source coil. In an alternating circuit, a similar phenomenon is seen. Frequently, an inductance in an alternating current circuit is influenced by other inductances in the circuit or in a nearby circuit. The mutually inductive coupling between such inductances then leads to a coupling of the circuit equations for the two circuits in case the inductances are in different circuits.

When we have mutual inductances in the circuits, we also need to bear in mind another fact—mutual inductance M can take positive as well as negative values. This is unlike the self-inductance L of a coil which can take only positive values. M on the other hand, depends on the relative orientation of the two coils as well as the sense of the currents in the two coils. If i_1 and i_2 are the currents in the two coils with mutual inductance M, then the EMF induced in each of the coils due to the current change in the other coil is $M\frac{di_2}{dt}$ and $M\frac{di_1}{dt}$. Thus, we see that we need to specify the sense of the two currents i_1 and i_2 to get the sign of M. A standard convention to specify the sign of the mutual inductance term in a circuit is the following:

Put a dot at one end of both the coils. Then if the current enters the dotted terminal of coil number 1 with current i_1, the polarity of the $M\frac{di_1}{dt}$ term in coil no. 2 is positive at the dotted terminal. In case current i_1 was leaving at the dotted terminal, the polarity of the $M\frac{di_1}{dt}$ term in coil no. 2 is negative. Figure 9.29 illustrates this convention.

Once we have done this, the analysis of the circuit(s) can be done in a way which is very similar to what we have done in the LCR and other circuits.

As an example, consider the circuit in Fig. 9.30.

The circuit, as shown has two loops—$KABDGHK$ and $KABFGHK$. We can apply Kirchhoff's laws to get

$$\begin{aligned} i &= i_1 + i_2 \\ V &= \frac{Q}{C} + L\frac{di_1}{dt} + M\frac{di_2}{dt} + iR \\ V &= L\frac{di_2}{dt} + M\frac{di_1}{dt} + iR \end{aligned} \tag{9.102}$$

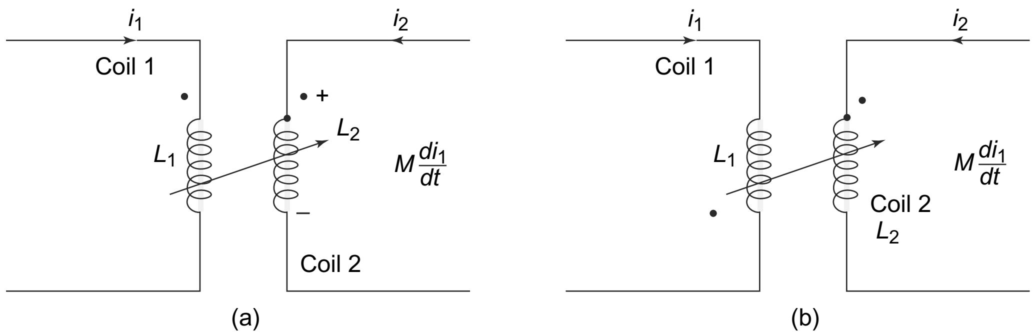


Fig. 9.29 *Illustrating the convention for mutual inductance (a) Current i_1 in coil 1 is entering the dotted terminal and hence the dotted terminal in the coil 2 has a positive polarity for the $M\frac{di_1}{dt}$ term, (b) Current i_1 in coil 1 is leaving the dotted terminal and hence the the dotted terminal in coil 2 has negative polarity for the mutual inductance term*

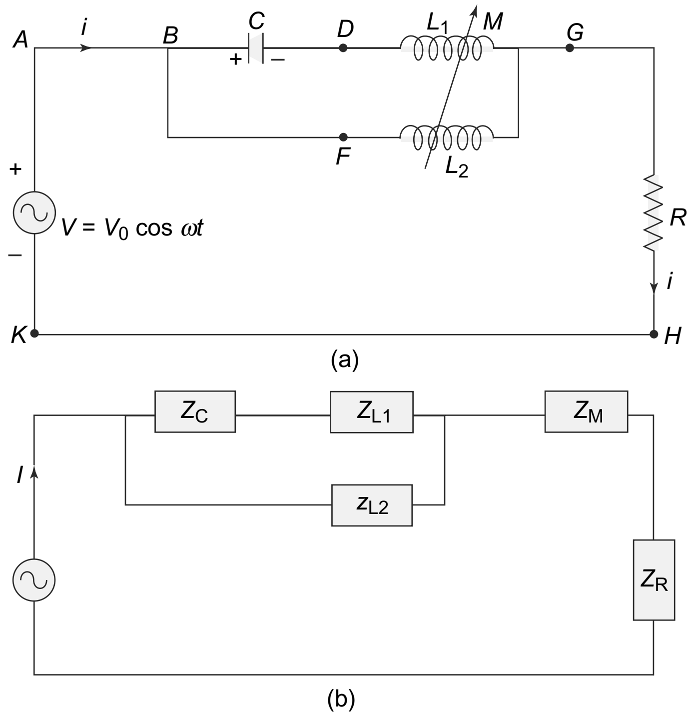


Fig. 9.30 *(a) An alternating current circuit with two inductors L_1 and L_2 and a mutual inductance coupling between them. The coupling is indicated by an arrow across both of them, (b) Equivalent Circuit with complex currents*

We would only be interested in the oscillatory components of the currents, i.e., the components which will survive once the transients have died out. The equivalent circuit is shown in Fig. 9.30(b) where $Z_M = i\omega M$. We can rewrite the loop equations in terms of the complex impedances Z_C, Z_L and Z_M and the complex currents I_1 and I_2 where

$$i_1 = \text{Re}\, I_1 = \text{Re}\,(I_1^0 e^{i\omega t})$$
$$i_2 = \text{Re}\, I_2 = \text{Re}\,(I_2^0 e^{i\omega t})$$

The loop equations are therefore, given by

$$V_0 = Z_C I_1^0 + Z_{L1} I_1^0 + Z_M I_2^0 + I_1^0 R + I_2^0 R$$
$$V_0 = Z_{L2} I_2^0 + Z_M I_1^0 + I_1^0 R + I_2^0 R \tag{9.103}$$

Solving these for I_1^0 and I_2^0, we get

$$I_1^0 = V_0 \frac{Z_M - Z_{L2}}{[(Z_M + R)^2 - (Z_C + Z_{L1} + R)(Z_{L2} + R)]} \tag{9.104}$$

and

$$I_2^0 = V_0 \frac{Z_M - Z_C - Z_{L1}}{[(Z_M + R)^2 - (Z_C + Z_{L1} + R)(Z_{L2} + R)]} \tag{9.105}$$

We can find the real currents in the circuit by taking the real part of $I_1^0 e^{i\omega t}$ and $I_2^0 e^{i\omega t}$.

PROBLEM 9.11 In the circuit shown in Fig. 9.31, calculate the currents when the mutual inductance M corresponds to 0, +1 and −1.

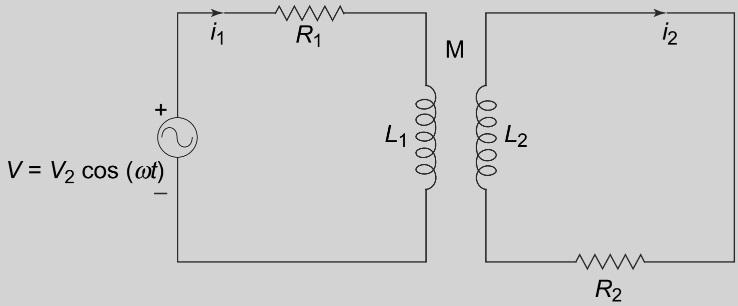


Fig. 9.31 *Problem 9.11*

9.7.2 Transformers

A particularly useful device which uses the phenomenon of mutual inductance is a transformer. These devices are used in a variety of appliances as well as form an essential part of the electrical power distribution network. A transformer is basically a device which transfers electrical energy from one circuit to another, using the inductive coupling between the two circuits.

The pair of inductively coupled circuits in a transformer are called the primary and secondary circuits. The primary circuit has an EMF source (an alternating current source) and an inductance L_1 and an impedance Z_1. The secondary circuit has no source of EMF in it but only an inductor L_2 and a load impedance Z_L. The two inductances L_1 and L_2 are coupled with a mutual inductance M between them. The schematic representation of a transformer is shown in Fig. 9.32.

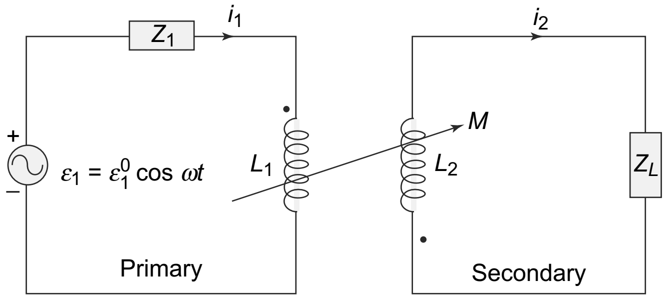


Fig. 9.32 *Transformer. Two inductances L_1 and L_2 are coupled with a mutual inductance M between them*

To analyse the transformer circuit, we introduce complex currents I_1 and I_2 for the primary and secondary circuits respectively, such that the real currents i_1 and i_2 are given by

$$i_1 = \mathrm{Re}\, I_1 = \mathrm{Re}(I_1^0 e^{i\omega t})$$
$$i_2 = \mathrm{Re}\, I_1 = \mathrm{Re}(I_2^0 e^{i\omega t})$$

The source EMF is given by

$$\mathcal{E} = \mathcal{E}_1^0 \cos \omega t$$

The loop equations can be written down for the primary and secondary circuits as

$$\begin{aligned} \mathcal{E}_1^0 &= I_1^0 Z_1 + (i\omega L_1) I_1^0 - i\omega M I_2^0 \\ 0 &= I_2^0 Z_L + (i\omega L_2) I_2^0 - i\omega M I_1^0 \end{aligned} \tag{9.106}$$

Note that for the secondary circuit, the mutual inductance term acts like a source of EMF. These equations can be solved once we know $\mathcal{E}_1^0$, Z_1, Z_L, L_1, L_2 and M.

A particularly useful case is when the EMF in the secondary becomes a multiple of the EMF in the primary loop. This happens when $Z_1 = 0$ and also the inductive coupling between the two circuits is 1. Then

$$\begin{aligned} I_2^0 &= -\frac{-i\omega M I_1^0}{(Z_L + i\omega L_2)} \\ I_1^0 &= \frac{\mathcal{E}_1^0}{\left(i\omega L_1 + \frac{M^2\omega^2}{Z_L + i\omega L_2}\right)} \\ &= \frac{\mathcal{E}_1^0 (Z_L + i\omega L_2)}{\omega^2 L_1 L_2 + \omega^2 M^2 + i\omega L_1 Z_L} \end{aligned} \tag{9.107}$$

For the case of complete coupling between the two inductive elements

$$M^2 = L_1 L_2$$

and therefore, we have

$$\begin{aligned} V_2^0 &= I_2^0 Z_L \\ &= \frac{i\omega M Z_L}{Z_L + i\omega L_2} \frac{Z_L + i\omega L_2}{i\omega L_1 Z_L} E_1^0 \\ &= \left(\frac{M}{L_1}\right) E_1^0 \end{aligned} \tag{9.108}$$

Consider now the case when the two coils in the primary and secondary circuits are wound over a ferromagnetic core and hence the magnetic flux over each turn of one also crosses each turn of the other coil. In that case, if n_1 and n_2 are the number of turns of the primary and secondary coils respectively, then clearly

$$L_1 \propto n_1^2$$
$$L_2 \propto n_2^2$$

and therefore,

$$M \propto n_1 n_2$$

with the same proportionality constant. In this case, the ratio of the output voltage across the load impedance Z_L and the input EMF is simply related by

$$\boxed{\frac{V_2^0}{\mathcal{E}_1^0} = \frac{n_2}{n_1}} \tag{9.109}$$

The ratio $\frac{n_2}{n_1}$ is called the turns ratio of the transformer. We see that if it is greater than 1, the output voltage is more than the input EMF. In this case we have a **'step-up'** transformer. If the turns ratio is less than 1, then we have a **'step-down'** transformer. Thus, by suitably arranging the ratio of the number of turns in the primary and secondary coils, we can get any desired potential difference across the load Z_L.

The primary and secondary currents can also be computed in this special case. They are

$$\begin{aligned} I_1^0 &= \mathcal{E}_1^0 \left(\frac{L_2}{L_1 Z_L} + \frac{1}{i\omega L_1}\right) \\ I_2^0 &= \mathcal{E}_1^0 \left(\frac{M}{L_1 Z_L}\right) \end{aligned} \tag{9.110}$$

There is a interesting thing about these currents. Note that the second term in the primary current I_1^0 is not in phase with the applied EMF $\mathcal{E}_1^0$ since Z_L is real for resistive loads. Thus, the real voltage and currents in the first loop are

$$\mathcal{E}_1 = \mathcal{E}_1^0 \cos\omega t \tag{9.111}$$

while

$$\begin{aligned} i_1 &= \mathrm{Re}\, I_1 \\ &= \mathrm{Re}\,(I_1^0 e^{i\omega t}) \\ &= \mathcal{E}_1^0 \left(\frac{L_2}{L_1 Z_L}\right) \cos\omega t + \frac{\mathcal{E}_1^0}{\omega L - 1} \sin\omega t \end{aligned} \tag{9.112}$$

The current therefore, has two terms, one in phase with the applied EMF and one out of phase with it. This has an interesting consequence on the average power. The average power in the primary circuit is defined by

$$\begin{aligned} P_{\text{in}} &= \frac{\omega}{2\pi} \int_0^{2\pi\omega} i_1 \mathcal{E}_1 \\ &= \frac{(\mathcal{E}_1^0)^2}{2} \left(\frac{L_2}{L_1 Z_L} \right) \end{aligned} \tag{9.113}$$

where the second term (which is out of phase) of the current i_1 does not contribute in the integral.

Similarly, we can compute the output power.

$$\begin{aligned} P_{\text{out}} &= \frac{\omega}{2\pi} \int_0^{2\pi\omega} (I_2^0)^2 Z_L \\ &= \frac{(\mathcal{E}_1^0)^2}{2} \frac{M^2}{L_1^2 Z_L} \\ &= \frac{(\mathcal{E}_1^0)^2}{2} \left(\frac{L_2}{L_1 Z_L} \right) \end{aligned} \tag{9.114}$$

Thus, the output and input power is the same as it should be since there is, in this special case, no resistive loss in the primary circuit ($Z_1 = 0$ in this case).

It is sometimes convenient to take the magnitude of only the first term in the expression for I_1^0 (Eq. (9.110)) which is in phase with $\mathcal{E}_0$ and calling this the effective current I_1^0(eff). Then

$$\frac{|I_1^0(\text{eff})|}{|I_2^0|} = \frac{L_2}{M} = \frac{n_1}{n_2} \tag{9.115}$$

which we see, is just the inverse of the turns ratio.

SUMMARY

- An LC circuit shows oscillatory behaviour with a frequency given by $\frac{1}{\sqrt{LC}}$. The energy in the system oscillates between the energy stored in the capacitor and the inductor.
- An LCR circuit has either damped currents or damped oscillations depending on the values of L,C and R.
- Kirchhoff's Laws are valid for alternating currents or time dependent currents of a particular frequency if we take into account the potential drops across the inductive, capacitative and resistive components.
- Inductive and capacitative elements in a circuit can be thought of as having complex impedances. Their presence in the circuit leads to a phase shift between the voltage and the current.
- Circuits with alternating current sources and impedances can be analysed using Kirchhoff's Laws with complex currents and voltages.

- Impedances in series and parallel behave in exactly the same manner as resistors in series and parallel combinations.
- The average power in an alternating current circuit depends not only on the magnitude of the current and the voltage but also on the phase between them. This is quantified by a quantity called the power factor which is the cosine of the phase difference.
- A series LCR circuit with an alternating current exhibits resonance at a frequency given by the natural frequency of the LC circuit. The sharpness of resonance is measured by a quantity called the Q-factor which depends on the resistance and the inductance. A parallel LCR circuit exhibits the phenomenon of anti-resonance.
- Transformers are devices which use the inductive coupling between two inductances to change the EMF of an alternating current. They are extensively used in power distribution.

CONCEPTUAL QUESTIONS

1. In an LC circuit, at the time the current in the circuit is a maximum, the
  a. charge on the capacitor is zero
  b. charge on the capacitor is maximum
  c. magnetic field is zero
  d. electric field is a maximum
  e. None of the above

2. In an oscillating LC circuit which has zero resistance, what determines the amplitude and the frequency?

3. You are given an inductor of inductance 10 mH and two capacitors of 5 and 2 μ F capacity. What resonant frequencies can be obtained with these elements?

4. A circuit consists of an alternating current source with a variable ω connected with a resistor R and a black box which contains either an inductor, or a capacitor or both. At a frequency $\omega = \omega_0$, the current and voltage are in phase. The black box contains
  a. Only an inductor
  b. Only a capacitor
  c. Both an inductor and a capacitor
  d. Can not say with the information provided

5. An inductance L is suddenly connected to a battery and a resistance R. At what time will the battery deliver one-half of its steady state current?

6. A circuit has an inductance of 10 mH, a capacitance of 0.1 μ F and a resistance of 1000 Ohms. Is this circuit periodic?

7. In the circuit in Question 6, we apply an EMF of 10^5 V. What is the final charge on the capacitor?

8. A circuit with an EMF at 60 Hz has a resistance of 2 Ohms and an inductance of 10 mH. What is the power factor?

9. What is the value of the capacitance that needs to be introduced in the circuit in Question 8 to make the power factor unity?

10. Inductances of 2 and 5 mH are connected in parallel and the combination is placed in series with a 10 Ohm resistance. The whole combination is now connected to a EMF with amplitude 100 V and frequency 1000 cycles per second. Find the angle of the lag in the current through the resistance and the inductance.

PROBLEMS

1. At the moment depicted in the *LC* circuit the current is non-zero and the capacitor plates are charged (as shown in the Fig. 9.33). The energy in the circuit is stored

a. only in the electric field and is decreasing
b. only in the electric field and is constant
c. only in the magnetic field and is decreasing
d. only in the magnetic field and is constant
e. in both the electric and magnetic field and is constant
f. in both the electric and magnetic field and is decreasing

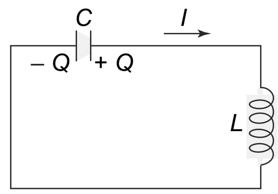


Fig. 9.33 *Problem 1*

2. In a freely oscillating *LC* circuit, (no driving voltage), suppose the maximum charge on the capacitor is $Q_{\max}$. Assume the circuit has zero resistance.

a. In terms of the maximum charge on the capacitor, what value of charge is present on the capacitor when the energy in the magnetic field is three times the energy in the electric field.
b. How much time has elapsed from when the capacitor is fully charged for this condition to arise?
c. If the resistance is non-zero, will the natural frequency of oscillation compared to the natural frequency of the ideal LC circuit (with zero resistance) increase, stay the same or decrease?

3. An inductor having inductance L and a capacitor having capacitance C are connected in series. The current in the circuit increase linearly in time as described by $I = Kt$. The capacitor initially has no charge. Determine

a. the voltage across the inductor as a function of time
b. the voltage across the capacitor as a function of time
c. the time when the energy stored in the capacitor first exceeds that in the inductor

4. A toroidal coil has N turns, and an inner radius a, outer radius b, and height h. The coil has a rectangular cross section shown in Fig. 9.34.

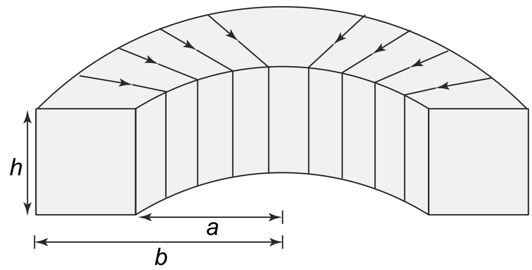


Fig. 9.34 *Problem 4*

The coil is connected via a switch S_1, to an ideal voltage source with electromotive force $\mathcal{E}$. The circuit has total resistance R. Assume all the self-inductance L in the circuit is due to the coil. At time $t = 0$ S_1 is closed and S_2 remains open.

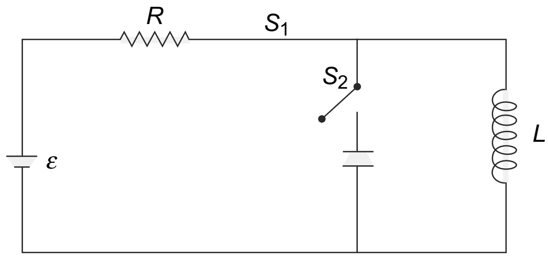


Fig. 9.35 *Problem 4*

a. When a current I is flowing in the circuit, find an expression for the magnitude of the magnetic field inside the coil as a function of distance r from the axis of the coil.
b. What is the self-inductance L of the coil?
c. What is the current in the circuit a very long time ($t \gg LR$) after S_1 is closed?
d. How much energy is stored in the magnetic field of the coil a very long time ($t \gg LR$) after S_1 is closed?
   Assume now that a very long time ($t \gg LR$) after the switch S_1 was closed, the voltage source is disconnected from the circuit by opening S_1, and by simultaneously closing S_2 the toroid is connected to a capacitor of capacitance C. Assume there is negligible resistance in this new circuit.
e. What is the maximum amount of charge that will appear on the capacitor?
f. How long will it take for the capacitor to first reach a maximal charge after S_2 has been closed?

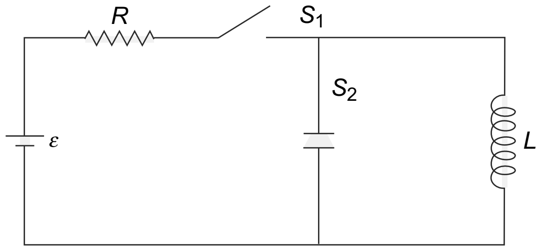


Fig. 9.36 *Problem 4*

5. A series combination of R and L is put in parallel with a series combination of R and C. Find the relationship between R, L and C such that the impedance is independent of frequency.

6. A series LCR circuit has the values $L = 0.1$ henry, $C = .001$ farad and $R = 1$ ohm. The AC source has the EMF $\mathcal{E} = 50(\cos(50t) + \cos(60t))$. Calculate the average power consumed in the circuit.

7. A capacitor initially uncharged and of capacitance C in series with a resistor of resistance R_0 is connected to an AC source of EMF $\mathcal{E} = 50\sin(\omega t)$. After a time period $t_1 = \frac{\pi}{4\omega}$ the AC source stops and in its place a resistor of resistance R_1 appears so that after $t = t_1$, we have a series CR circuit with a resistance value $(R_0 + R_1)$ without any power source. Calculate the current in the circuit at time $t = 2t_1$.

8. A capacitor of capacitance C initially carrying a charge q is discharged by connecting its plates through an inductor of inductance L and a resistor of resistance R. It goes through 200 oscillations by the time the charge on the capacitor falls to half its initial value. Calculate the Q-value of the circuit.

9. A capacitor of capacitance C carries a charge Q. At time $t = 0$, it is connected across a resistor. Owing to being heated the resistance of the resistor does not remain constant but changes like $R(t) = R_0(1 + \frac{t}{t_0})$, where R_0 and t_0 are constants. Calculate how much time will elapse for the charge on C to go down to half its starting value. Given $CR_0 = 1$ millisec., $t_0 = 2$ millisec.

10. A circuit consisting of a inductor of inductance L and a resistor of resistance R is connected in series across a power source. The power source consists of an AC power supply $\mathcal{E} = 2\cos(\omega t)$ V in series with a DC battery of voltage $V_0 = 2$ volts. Calculate the current in the circuit.

11. Determine the both the phase and amplitude of the current through the circuit shown in Fig. 9.37. The circle represents an AC source $\mathcal{E} = \mathcal{E}_0 \cos(\omega t)$.

12. A 60 Hz transformer with a turns ratio of 2 : 1 has a primary inductance of 100 H and a D.C resistance of 20 Ohms. The coupling coefficient is 1. If 1000 volts is across the primary calculate the current in the primary when the secondary circuit is open and also when a load of 20 ohms is put in the secondary circuit.

13. Consider the circuit given in Fig. 9.38. Determine the voltage across the capacitor at steady state given that the source voltage is $120\cos(1000t)$ V and the coefficient of coupling is unity.

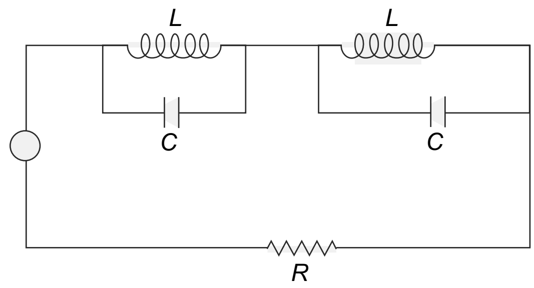


Fig. 9.37 *Problem 11*

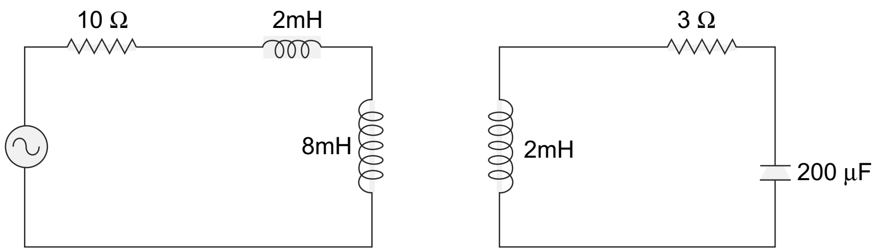


Fig. 9.38 *Problem 13*

14. A rectangular loop of wire with dimensions l and w is released at $t = 0$ from rest just above a region with a magnetic field B_0 (Fig. 9.39). The loop has a resistance of R and self inductance L and mass m. Consider the loop when its upper edge is in the zero field region.

a. Ignoring the self-inductance but not the resistance, calculate the current in the loop and velocity of the loop as functions of time.

b. Ignoring the resistance but not the self-inductance find the current in the loop and the velocity of the loop as functions of time.

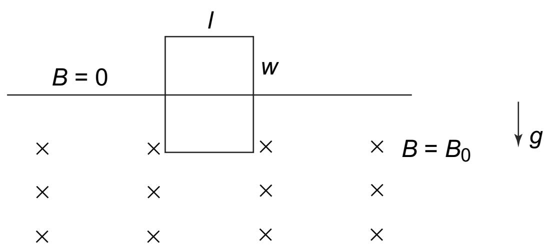


Fig. 9.39 *Problem 14*

15. In the circuit in Fig. 9.40, the resistance of L is negligible and the switch is open initially with zero current. Find the heat dissipated in R_2 when the switch is closed and remains closed for a long time. Also find the heat dissipated in R_2 when the switch is opened after having being closed for a long time and is kept open for a long time.

16. In the circuit in Fig. 9.41, find the complex impedance between the two terminals. What is the condition for the impedance to become infinite?

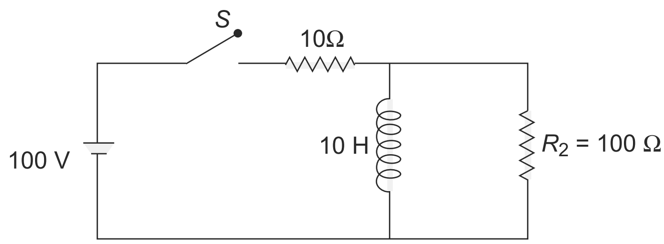


Fig. 9.40 *Problem 15*

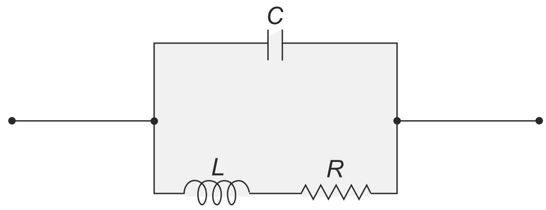


Fig. 9.41 *Problem 16*

17. In the circuit in Fig. 9.42, the coupling coefficient for mutual inductance for L_1 and L_2 is unity. Find the instantaneous current delivered by the alternating current source. What is the current when the alternating current source frequency is equal to the resonant frequency of the secondary circuit?

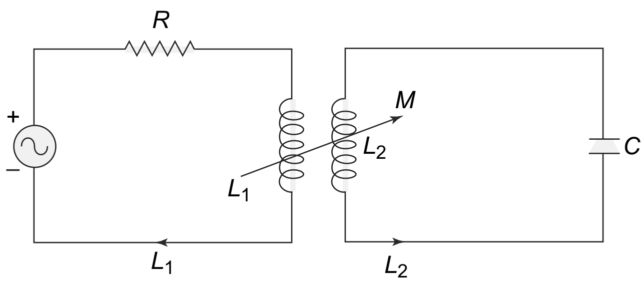


Fig. 9.42 *Problem 17*

18. A resonant circuit has a parallel plate capacitor C and an inductor of N turns on a toroid. All the linear dimensions of the capacitor and the inductor as reduced by a factor of 10 while the number of turns of the toroid remain the same. By what factor is the capacitance, the inductance and the resonant frequency changed?

19. A long cylindrical solenoid of length L and of radius R has another solenoid of the same length but of radius $R/2$. The windings are connected such that the current in the outer and inner loops are in opposite directions. Neglecting end effects, calculate the self-inductance of the arrangement.

10

Maxwell's Equations

Learning Objectives

- To understand the need for modifying Ampere's Law for time-varying currents and charges.
- To learn about the displacement current term and how it makes Ampere's Law consistent for time dependent currents.
- To comprehend how a time-varying current in a physical situation illustrates the need for displacement current.
- To learn about Maxwell's Equations without dielectrics and in linear polarisable media.
- To revisit the concept of energy in electrical and magnetic fields in light of Maxwell's equations and Poynting vector.
- To comprehend the existence of electromagnetic waves as a natural result of Maxwell's equations.
- To learn about dispersion and its relationship with the properties of the medium.
- To understand the phenomena of reflection and refraction at interfaces.
- To learn about reflection and transmission coefficients and polarisation of electromagnetic waves.

10.1 REVISITING AMPERE'S CIRCUITAL LAW AND MAXWELL'S MODIFICATION

We have seen in the previous chapters, that an electric current is a source of a magnetic field. We have also discussed the relationship between the electric current and the magnetic field it generates namely, Ampere's Law. This is given by

$$\vec{\nabla} \times \vec{B}(\vec{r}, t) = \mu_0 \vec{j}(\vec{r}, t) \tag{10.1}$$

where $\vec{j}(\vec{r}, t)$ is the current density. Although this relationship gives us the magnetic field due to steady currents, it was Maxwell who first pointed out that it breaks down in case of time dependent current densities.

The inconsistency of Ampere's Law can be easily seen from the Eq. (10.1) itself. Taking the divergence of both sides of the equation, we get

$$\vec{\nabla} \cdot (\vec{\nabla} \times \vec{B}(\vec{r}, t)) = \mu_0 \vec{\nabla} \cdot \vec{j}(\vec{r}, t) \tag{10.2}$$

The left-hand side of the equation, being a divergence of a curl vanishes identically. So we have

$$\vec{\nabla} \cdot \vec{j}(\vec{r}, t) = 0$$

as a consequence of Ampere's Law. Gauss's Law relates the electric field and the charge density at the same point in space at the same time and is valid when both these quantities are function of time. However, we know from the Equation of Continuity, which itself is a reformulation of the Principle of Conservation of Charge, that

$$\vec{\nabla} \cdot \vec{j}(\vec{r}, t) + \frac{\partial \rho}{\partial t} = 0 \tag{10.3}$$

Clearly, in the presence of time-dependent charge densities, the divergence of the current density is non-vanishing. However, Ampere's Law demands that the divergence of $\vec{j}$ must vanish. Thus, there is a contradiction, as was pointed out by Maxwell.

The contradiction is also easily seen in a simple physical situation with time-dependent charge densities. Consider a charged conductor C_1 with a charge Q. At time $t = 0$, C_1 is connected to another, uncharged conducting sphere C_2 with a thin wire.

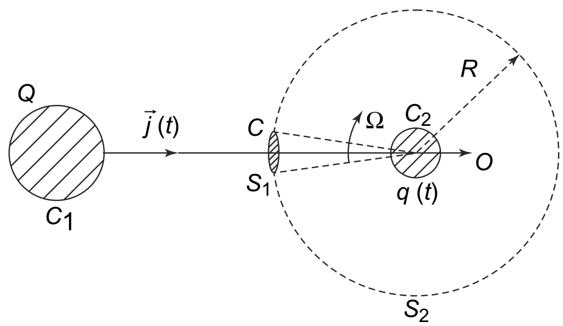


Fig. 10.1 *Conductor C_1 is charged to a charge Q and connected with a thin wire to an uncharged spherical conductor C_2 at time $t = 0$. Charge flows from C_1 to C_2. C is a loop around the connecting wire and S_1 and S_2 are spherical surfaces both with centre O, the centre of C_2 and both have C as its edge. S_1 subtends a solid angle of Ω at O and S_2 subtends a solid angle of $4\pi - \Omega$*

Since the two conductors are at different potentials, charge flows from the initially charged one, C_1 to the initially uncharged one C_2. This motion of charge constitutes a current I in the connecting wire. Any current creates a magnetic field and thus, we have a magnetic field $\vec{B}$ around the wire. Consider a closed loop C around the wire. Applying Ampere's circuital law to this loop, we get

$$\begin{aligned} \oint_C \vec{B} \cdot d\vec{L} &= \iint_S (\vec{\nabla} \times \vec{B}) \cdot d\vec{S} \\ &= \mu_0 \iint_S \vec{j} \cdot d\vec{S} \end{aligned} \tag{10.4}$$

where S is any surface which has the closed loop C as its boundary. Let a surface S_1 be one such surface (Fig. 10.1). Then since the connecting wire passes through the surface S_1, we have

$$\oint_C \vec{B} \cdot d\vec{L} = \mu_0 \iint_{S_1} \vec{j} \cdot d\vec{S} = \mu_0 I \tag{10.5}$$

where I is the total current at some point of time t. Now consider another surface S_2 which also has the closed loop C as its boundary. This surface has no current passing through it and hence the surface integral of the current density in Eq. (10.5) vanishes. So we get

$$\oint_C \vec{B} \cdot d\vec{L} = \mu_0 \iint_{S_2} \vec{j} \cdot d\vec{S} = 0 \tag{10.6}$$

Thus, we see that there is a contradiction between these two ways of finding the line integral of the magnetic field. Clearly, the value of the line integral cannot depend on what surface one chooses around the contour. This simple physical example illustrates the inadequacy of Ampere's Law, as we have studied in situations where the charge density is time-dependent.

The problem was solved by Maxwell in 1862, who introduced a new term in Ampere's Law to remedy the situation. This term is called the **'displacement current'**. The modified version of Ampere's Law is thus consistent even with time-dependent charge densities. But its importance is much more than just removing an inconsistency. It turns out that this modification essentially laid the foundations for the discovery of electromagnetic waves as we know them. We shall see that later in this chapter.

Let us first consider the case where there are no electrically or magnetically polarisable materials present. Then Gauss's Law states that

$$\vec{\nabla} \cdot \vec{E}(\vec{r}, t) = 4\pi k \rho(\vec{r}, t) = \frac{1}{\varepsilon_0} \rho(\vec{r}, t) \tag{10.7}$$

where we have also written the right-hand side in SI units with ε_0, the permittivity of free space. We can also write the Equation of Continuity as

$$\begin{aligned} 0 &= \vec{\nabla} \cdot \vec{j}(\vec{r}, t) + \frac{\partial \rho(\vec{r}, t)}{\partial t} \\ &= \vec{\nabla} \cdot \left(\vec{j} + \varepsilon_0 \frac{\partial \vec{E}}{\partial t} \right) \end{aligned} \tag{10.8}$$

Notice that we now have a quantity $\vec{j} + \varepsilon_0 \frac{\partial \vec{E}}{\partial t}$ whose divergence is identically zero. Therefore, if we can use this in Ampere's Law instead of $\vec{j}$ alone, the inconsistency that we encountered could be circumvented. This is precisely what Mawxwell did. Calling

$$\vec{j}_{\text{tot}} \equiv \vec{j} + \varepsilon_0 \frac{\partial \vec{E}}{\partial t}$$

and substituting in Ampere's Law (Eq. (10.1)), we get

$$\boxed{\vec{\nabla} \times \vec{B}(\vec{r}, t) = \mu_0 \vec{j}_{\text{tot}}(\vec{r}, t)} \tag{10.9}$$

Clearly now, if we take the divergence of both sides, we do get a consistent result unlike the original form of Ampere's Law. The total current density $\vec{j}_{\text{tot}}(\vec{r}, t)$ is the sum of the normal conduction current density $\vec{j}(\vec{r}, t)$ and a new 'displacement current' term $\vec{j}_D = \varepsilon_0 \frac{\partial \vec{E}}{\partial t}$ which is the contribution from a time dependent electric field.

What about the physical example we have considered above of two conductors connected with a wire? In this case too, we can see that the inconsistency is removed by using this modification. Let the charge on C_2 at a time t be $q(t)$. The electric field then, on both S_1 and S_2 is simply

$$\vec{E}(t) = \frac{1}{4\pi\varepsilon_0}\frac{q(t)}{R^2}\hat{n}$$

where $\hat{n}$ is the outward unit normal. If we now compute the integral of the total current $\vec{j}_{\text{tot}}$ on S_1, we get (remembering that $d\vec{S}$ is opposite to $\vec{E}$ on S_1)

$$\begin{aligned}
\iint_{S_1} \vec{j}_{\text{tot}} \cdot d\vec{S} &= \iint_{S_1} \left(\vec{j} + \varepsilon_0 \frac{\partial \vec{E}}{\partial t}\right) \cdot d\vec{S} \\
&= I - \frac{\varepsilon_0}{R^2}\frac{\dot{q}(t)}{4\pi\varepsilon_0} \iint_{S_1} |d\vec{S}| \\
&= \frac{dq(t)}{dt} - \frac{1}{4\pi R^2}\frac{dq(t)}{dt} R^2 \Omega \\
&= \frac{dq(t)}{dt}\left(1 - \frac{\Omega}{4\pi}\right) \qquad (10.10)
\end{aligned}$$

where Ω is the solid angle subtended by S_1 at the centre of C_2. Repeating the same for S_2, we get

$$\begin{aligned}
\iint_{S_2} \vec{j}_{\text{tot}} \cdot d\vec{S} &= \iint_{S_2} \left(0 + \varepsilon_0 \frac{\partial \vec{E}}{\partial t}\right) \cdot d\vec{S} \\
&= \frac{\varepsilon_0}{R^2}\frac{\dot{q}(t)}{4\pi\varepsilon_0} \iint_{S_2} |d\vec{S}| \\
&= \frac{1}{4\pi R^2}\frac{dq(t)}{dt}(4\pi)R^2(1 - \frac{\Omega}{4\pi}) \\
&= \frac{dq(t)}{dt}\left(1 - \frac{\Omega}{4\pi}\right) \qquad (10.11)
\end{aligned}$$

since the surface area of S_2 is $4\pi R^2\left(1 - \frac{\Omega}{4\pi}\right)$. We see that the integral now, with this modification is identical to the one obtained for S_1. Thus, the inconsistency is removed with Maxwell's modification of Ampere's Law.

To summarise, in the presence of time-dependent charge densities, Ampere's Law gives rise to an inconsistency. This can be remedied by adding a displacement current term to Ampere's Law and with this modification, Ampere's Law reads

$$\boxed{\vec{\nabla} \times \vec{B}(\vec{r}, t) = \mu_0 \vec{j}_{\text{tot}}(\vec{r}, t) = \mu_0 \left(\vec{j} + \varepsilon_0 \frac{\partial \vec{E}}{\partial t}\right)} \qquad (10.12)$$

PROBLEM 10.1 The magnetic field in free space is given by $\vec{B} = B_0 \sin(\omega t - \alpha z)$ where α is a constant. Determine the displacement current density and the electric field intensity.

PROBLEM 10.2 A parallel plate capacitor has circular plates of radius 5 mm and separated by 1 mm. The region between the plates has air. What is the displacement current between the plates when the charge is increasing at a rate of 10×10^{-6} C per second?

10.2 MAXWELL'S EQUATIONS WITHOUT DIELECTRICS

With this modification to Ampere's Law, we are now ready to write the fundamental equations of electromagnetism. These are **Gauss's Law** for electrostatic fields which is really equivalent to Coulomb's Law, as we have seen. This give us the relationship between the electric field and the charge densities.

$$\vec{\nabla} \cdot \vec{E} = \frac{\rho}{\varepsilon_0}$$

Gauss's law relates the electric field and the charge density at the same point in space at the same time and is valid when both these quantities are function of time. For time-dependent electric fields, we have **Faraday's Law** which tells us the relationship between the derivative (Curl) of the electric field and the rate of change of magnetic field.

$$\vec{\nabla} \times \vec{E} = -\frac{\partial \vec{B}}{\partial t}$$

For the magnetic field, we know that it is divergenceless that is, magnetic lines of force have no source or sink. Thus,

$$\vec{\nabla} \cdot \vec{B} = 0$$

Finally, we have the **Ampere's Law** which is a relationship between the magnetic field and the current. With Maxwell's modification this reads,

$$\vec{\nabla} \times \vec{B} = \mu_0 \vec{j}_{\text{tot}} = \mu_0 \left(\vec{j} + \varepsilon_0 \frac{\partial \vec{E}}{\partial t} \right)$$

These set of four equations contain all the information we need to define electric and magnetic fields at any point in space. Once the charge densities ρ and current densities $\vec{j}$ are known, we can determine the electric and magnetic fields. These set of four equations are collectively known as **Maxwell's Equations without dielectrics.**

$$\boxed{\begin{aligned} \vec{\nabla} \cdot \vec{E} &= \frac{\rho}{\varepsilon_0} \\ \vec{\nabla} \times \vec{E} &= -\frac{\partial \vec{B}}{\partial t} \\ \vec{\nabla} \cdot \vec{B} &= 0 \\ \vec{\nabla} \times \vec{B} &= \mu_0 \left(\vec{j} + \varepsilon_0 \frac{\partial \vec{E}}{\partial t} \right) \end{aligned}} \qquad (10.13)$$

10.3 MAXWELL'S EQUATIONS IN POLARISABLE MEDIA

We have found the modification that is required by Ampere's Law to make it consistent for time-dependent current densities in free space above. In the presence of media in general, the modifications required are somewhat different. This is because the continuity equation satisfied by the current and charge densities is applicable only to free electrons or charges. However, we know that in the presence of media, the magnetic field for instance, is dependent on free and bound currents.

To find the relevant modification in the presence of polarisable media, we start with Ampere's Law as applicable in presence of media. This is given by

$$\vec{\nabla} \times \vec{H}(\vec{r}, t) = \vec{j}_f(\vec{r}, t) \tag{10.14}$$

where $\vec{j}_f$ is the free current density only. The free current density satisfies the continuity equation given by

$$\vec{\nabla} \cdot \vec{j}_f(\vec{r}, t) + \frac{\partial \rho_f(\vec{r}, t)}{\partial t} = 0 \tag{10.15}$$

where ρ_f is the charge density of free carriers. In addition to this, in presence of polarisable media, Gauss's Law also gets modified and we have

$$\vec{\nabla} \cdot \vec{D}(\vec{r}, t) = \rho_f(\vec{r}, t) \tag{10.16}$$

Combining Eqs. (10.15) and (10.16), we get

$$\vec{\nabla} \cdot \left(\vec{j}_f(\vec{r}, t) + \frac{\partial \vec{D}(\vec{r}, t)}{\partial t} \right) = 0 \tag{10.17}$$

From this, it is clear that in a medium, the total current $\vec{j}_{\text{tot}}$ is divergenceless and is given by

$$\vec{j}_{\text{tot}}(\vec{r}, t) = \vec{j}_f(\vec{r}, t) + \frac{\partial \vec{D}(\vec{r}, t)}{\partial t} \tag{10.18}$$

With this, the Maxwell's equations in presence of polarisable medium can be written as

$$\boxed{\begin{aligned} \vec{\nabla} \cdot \vec{D} &= \rho \\ \vec{\nabla} \times \vec{E} &= -\frac{\partial \vec{B}}{\partial t} \\ \vec{\nabla} \cdot \vec{B} &= 0 \\ \vec{\nabla} \times \vec{H} &= \left(\vec{j} + \frac{\partial \vec{D}}{\partial t} \right) \end{aligned}} \tag{10.19}$$

10.4 ELECTROMAGNETIC ENERGY DENSITY

In the previous chapters, we have discussed the concept of energy density of the electric and magnetic fields. Recall that for the electric field, we invoked the concept of the electric potential and related it to the work done in moving a charge in an electric field and therefore, to the energy density of the electric field itself. For the magnetic field, we used Ampere's Law to derive an expression for the energy density in the magnetic field. What we had obtained was

$$U_{\text{electric}} = \frac{1}{8\pi k} E^2(r) = \frac{\varepsilon_0 E^2(r)}{2}$$

for the electric field energy density and

$$U_{\text{magnetic}} = \frac{1}{2\mu_0} B^2(r)$$

for the magnetic field.

However, it turns out that the premises or assumptions on which these expressions were based are not strictly true for time dependent fields. Thus, for instance, the energy density of a electric field was derived using the concept of electric potential which can only be defined if the electric field is conservative in nature or equivalently, if the line integral of the electric field over a closed loop vanishes. We have seen that Faraday's Law shows that the electric field is not strictly conservative in the presence of time dependent magnetic fields. Therefore, the line integral over a closed loop does not vanish and hence the underlying assumptions for our derivation of energy density no longer seem to be valid.

Similarly, we used Ampere's Law for deriving the energy density of the magnetic field. We have seen above that for time dependent currents, Ampere's Law is inconsistent and needs to be modified. The introduction of displacement current term modifies the Ampere's law and we need to re-examine the whole derivation of energy density.

In this section, we re-examine the expressions for electric and magnetic energy densities. We shall be doing it in the presence of a medium with electric and magnetic polarisabilities. The case for vacuum or free space is obviously a special case of this.

Consider a volume V with a surface S. There exists a conduction current (density) $\vec{j}(\vec{r}, t)$ in the volume. Then $\rho(\vec{r}, t)d^3\vec{r}$ is the charge inside a volume element d^3r and $\vec{j}(\vec{r}, t)d^3\vec{r}$ is the rate of transport of this charge in the direction of $\vec{j}(\vec{r}, t)$. Therefore, the rate at which work is done by the electric field in this volume is

$$\frac{dW}{dt} = \iiint_V d^3\vec{r}\,\vec{j}(\vec{r}, t) \cdot \vec{E} \tag{10.20}$$

Note that writing the rate of work done in this way does not involve any reference to the electric potential. This, as we noted above, is because for time dependent currents or charge densities, the electric field is not strictly conservative in nature and hence the concept of a scalar potential (related to the electric field by the gradient operator) is not defined. For static field, this expression obviously equals that obtained earlier.

The work done by the electric field manifests itself as some other form of energy, which is typically mechanical in nature. For instance, for an isolated charge, this work done increases the kinetic energy of the charged particle. For a charge moving inside a conductor, this work manifests as the Joule heat produced by collisions of the charge with the lattice.

Let us try and understand the importance of Eq. (10.20). Using Maxwell's equations, we can rewrite the expresion by substituting for the current density. We know from the modified form of Ampere's Law that

$$\vec{j}(\vec{r}, t) = \vec{\nabla} \times \vec{H} - \frac{\partial \vec{D}}{\partial t}$$

Substituting in Eq. (10.20), we get

$$\frac{dW}{dt} = \iiint_V d^3\vec{r} \left(\vec{\nabla} \times \vec{H} - \frac{\partial \vec{D}}{\partial t} \right) \cdot \vec{E} \tag{10.21}$$

We now use the identity

$$\vec{\nabla} \cdot (\vec{E} \times \vec{H}) = (\vec{\nabla} \times \vec{E}) \cdot \vec{H} - \vec{E} \cdot (\vec{\nabla} \times \vec{H}) \tag{10.22}$$

to get

$$\frac{dW}{dt} = \iiint\limits_V d^3\vec{r} \left[(\vec{\nabla} \times \vec{E}) \cdot \vec{H} - \frac{\partial \vec{D}}{\partial t} \cdot \vec{E} - \vec{\nabla} \cdot (\vec{E} \times \vec{H}) \right] \tag{10.23}$$

But Faraday's Law tells us that

$$\vec{\nabla} \times \vec{E} = -\frac{\partial \vec{B}}{\partial t}$$

and so we get

$$\boxed{-\frac{dW}{dt} = \iiint\limits_V d^3\vec{r} \left[\vec{\nabla} \cdot (\vec{E} \times \vec{H}) + \frac{\partial \vec{D}}{\partial t} \cdot \vec{E} + \frac{\partial \vec{B}}{\partial t} \cdot \vec{H} \right]} \tag{10.24}$$

Equation (10.24) tells us that the negative of the work done by the field is the sum of three terms. Let us see if we can interpret them. The first term

$$\iiint\limits_V d^3\vec{r} \vec{\nabla} \cdot (\vec{E} \times \vec{H})$$

can be written as a surface integral using the divergence theorem.

$$\iiint\limits_V d^3\vec{r} \vec{\nabla} \cdot (\vec{E} \times \vec{H}) = \iint\limits_S \vec{P} \cdot d\vec{S} \tag{10.25}$$

where we have called

$$\vec{E} \times \vec{H} = \vec{P}$$

the **Poynting vector.** $\vec{P}$ thus represents the energy flow across the surface S of the volume under consideration. The direction of the flow is orthogonal to both $\vec{E}$ and $\vec{H}$. This fact will prove to be crucial when we study electromagnetic waves.

The second term turns out to be what we have already seen, the energy density in an electric field. To see this, let us assume we have a linear media that is

$$\vec{D} = \varepsilon \vec{E}$$

Then we can write the second term as

$$\iiint\limits_V d^3\vec{r} \frac{\partial \vec{D}}{\partial t} \cdot \vec{E} = \frac{\varepsilon}{2} \frac{\partial}{\partial t} \iiint\limits_V d^3\vec{r} \vec{E}^2 \tag{10.26}$$

This implies that the energy density of the electric field is

$$W_E = \frac{\varepsilon}{2} \vec{E}^2$$

Note that for free space, $\varepsilon = \varepsilon_0$ and we get the expression for the electric field energy density obtained before.

The third term is the rate of work done in changing the magnetic field in the volume V. We have already seen that the linear relationship between $\vec{B}$ and $\vec{H}$ is not as common in magnetic materials as the linear relationship between $\vec{D}$ and $\vec{E}$ for the electric case. Thus, in general we cannot interpret the third term as the rate of change of magnetic energy. However, for linear media, we know that

$$\vec{B} = \mu \vec{H}$$

and therefore, we can rewrite the third term as

$$\begin{aligned} \iiint\limits_V d^3\vec{r}\frac{\partial \vec{B}}{\partial t} \cdot \vec{H} &= \frac{1}{2}\frac{\partial}{\partial t}\iiint\limits_V d^3\vec{r}(\vec{H} \cdot \vec{B}) \\ &= \frac{1}{2\mu}\frac{\partial}{\partial t}\iiint\limits_V d^3\vec{r}\vec{B}^2 \end{aligned} \tag{10.27}$$

Thus, we see that in the case of linear magnetic media, we can interpret this as the energy density of a magnetic field. This is the same as the expression obtained earlier.

For ferromagnetic materials, the linear relation between $\vec{B}$ and $\vec{H}$ breaks down beyond a point when the field is increased. Thus, if a piece of ferromagnetic material is subjected to an applied magnetic field created by a current and the resultant magnetisation exhibits a hysteresis loop (see Fig. 7.5), the net work done per cycle is:

$$W_{\text{cycle}} = \iint d^3r\,dt\,\vec{H} \cdot \frac{d\vec{B}}{dt} = \iiint \vec{H} \cdot \vec{B} d^3r \tag{10.28}$$

Thus, an amount of energy per cycle equaling the area of the hysteresis loop would be generated per cycle and lost as heat. This becomes a serious problem in devices like transformers which use ferromagnetic materials in oscillating magnetic fields.

We now illustrate the relative magnitudes of electric and magnetic energy densities for the case of ordinarily encountered currents and charge densities.

EXAMPLE 10.1 Two circular, conducting plates of radius $R = 1$ m are placed parallel to each other at a distance $d = 1$ m. The plates are maintained at a potential $\pm V_0$ Volts. A thin wire of resistance $R_e = 1$ Ohms connects the centres of the two plates and the wire carries a current I. The wire is in the form of a thin cylinder of radius 1 cm. If V_0 is constant in time and maintained at 10 Volts, calculate the electric and magnetic energies stored in the volume between the plates. You may ignore edge effects in the parallel plate capacitor for calculating the electric field and also assume that the wire connecting the plates is infinite in length for calculating the magnetic field.

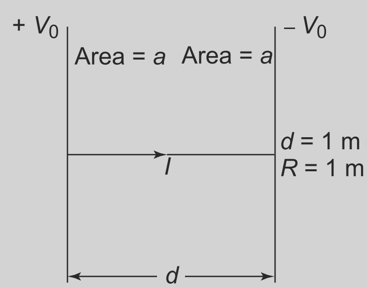


Fig. 10.2 *Example 10.1*

Solution

The potential difference between the plates is $2V_0$ and hence the electric field in between the plates is given by

$$E = \frac{2V_0}{d}$$

The total electric energy in the volume enclosed between the two plates is thus,

$$W_{\mathrm{E}} = \frac{1}{2}\varepsilon_0 E^2 \pi R^2 d = \frac{2\varepsilon_0 V_0^2 \pi R^2}{d}$$

We can also calculate the magnetic field at a distance r, $r > 1$ cm, from the wire using Ampere's Law and a closed circular loop around the wire. For that, we need to know the current which can be found by Ohm's Law, since

$$I = \frac{V}{R_e} = \frac{2V_0}{R_e}$$

and

$$B 2\pi r = \mu_0 I = \mu_0 \frac{2V_0}{R_e}$$

or

$$B = \frac{\mu_0 V_0}{\pi r R_e}$$

Thus, the magnetic energy in the volume is

$$\begin{aligned} W_{\mathrm{M}} &= \frac{1}{2\mu_0} \int_{1\mathrm{cm}}^{R} B^2 (2\pi r) d\, dr \\ &= \frac{\mu_0 V_0^2 d}{\pi {R_e}^2} \int_{1\mathrm{cm}}^{R} \frac{dr}{r} \\ &= \frac{\mu_0 V_0^2 d}{\pi R_e^2} \ln\left(\frac{R}{1\mathrm{cm}}\right) \end{aligned}$$

Putting in the values, we get

$$W_{\mathrm{E}} = 5.5 \times 10^{-9} \quad \text{Joules m}^{-3}$$

and the magnetic energy density as

$$W_{\mathrm{M}} = 1.84 \times 10^{-4} \quad \text{Joules m}^{-3}$$

PROBLEM 10.3 The electric field intensity in a source-free dielectric medium is given by

$$\vec{E} = E_0 (\sin(ax - \omega t) + \sin(ax + \omega t))\, \hat{j} \quad \text{V/m}$$

Calculate the magnetic field intensity as well as the time averaged energy densities in the electric and magnetic fields.

PROBLEM 10.4 Non-relativistic electrons are accelerated through a potential difference V and form a round beam with a current I. Find the Poynting vector at a point outside the beam, at a distance r from the axis

10.5 ELECTROMAGNETIC WAVES

Maxwell's Equations for free space that is without dielectrics but with charges and currents, (Eq. (10.13)) and in presence of media (Eq. (10.19)) reveal an extraordinary symmetry. The symmetry is between how electric and magnetic fields are related. Note that a time-dependent electric field can give rise to a magnetic field while a time dependent magnetic field can act as a source of electric field. This is also true in the absence of charge and current densities as can be easily seen.

$$\boxed{\begin{aligned}\vec{\nabla}\cdot\vec{E} &= 0\\ \vec{\nabla}\times\vec{E} &= -\frac{\partial\vec{B}}{\partial t}\\ \vec{\nabla}\cdot\vec{B} &= 0\\ \vec{\nabla}\times\vec{B} &= \mu_0\left(\varepsilon_0\frac{\partial\vec{E}}{\partial t}\right)\end{aligned}} \tag{10.29}$$

The symmetry between the electric and magnetic fields is manifest above. This fact was noticed by Maxwell himself after he introduced the displacement current term since it is only with the introduction of that term does the symmetry emerge. These equations, which are a set of differential equations, have a non-trivial solution even in the absence of any currents and charges. This remarkable fact, that even without sources (charges and currents) the time dependent electric and magnetic fields can give rise to each other leads to the phenomenon of electromagnetic waves as we shall see.

Let us consider free space with no charges and currents, i.e., $\vec{j} = \rho = 0$. The equations then take the form of Eq. (10.29). Taking the curl of the curl of $\vec{B}$, we get

$$\vec{\nabla}\times(\vec{\nabla}\times\vec{B}) = \vec{\nabla}(\vec{\nabla}\cdot\vec{B}) - \nabla^2\vec{B} = -\nabla^2 B \tag{10.30}$$

since $\vec{\nabla}\cdot\vec{B} = 0$. But Maxwell's equation tell us that

$$\vec{\nabla}\times(\vec{\nabla}\times\vec{B}) = \mu_0\varepsilon_0\frac{\partial(\vec{\nabla}\times\vec{E})}{\partial t} \tag{10.31}$$

But

$$\vec{\nabla}\times\vec{E} = -\frac{\partial\vec{B}}{\partial t}$$

and hence, we get

$$\vec{\nabla}\times(\vec{\nabla}\times\vec{B}) = \mu_0\varepsilon_0\frac{\partial^2\vec{B}}{\partial t^2} \tag{10.32}$$

or

$$\left(\nabla^2\vec{B} - \mu_0\varepsilon_0\frac{\partial^2\vec{B}}{\partial t^2}\right) = 0 \tag{10.33}$$

We can rewrite this as

$$\boxed{\left(\nabla^2 \vec{B} - \frac{1}{c^2}\frac{\partial^2 \vec{B}}{\partial t^2}\right) = 0} \tag{10.34}$$

where $c^2 = \frac{1}{\mu_0 \varepsilon_0}$. Proceeding in exactly the same manner, but this time starting from the Maxwell's equation for $\vec{\nabla} \times \vec{E}$ instead, we get

$$\boxed{\left(\nabla^2 \vec{E} - \frac{1}{c^2}\frac{\partial^2 \vec{E}}{\partial t^2}\right) = 0} \tag{10.35}$$

Equations of the kind (10.34) and (10.35) are partial differential equations. They are of the form of a wave equation. The reason these are called wave equations is because any function $\vec{f}(u)$, of a variable u defined as $u \equiv (\vec{r} \cdot \hat{n} - ct)$ satisfies it. In this, $\hat{n}$ is an arbitrary unit vector. This can easily be seen since

$$\frac{1}{c^2}\frac{\partial^2 \vec{f}(u)}{\partial t^2} = \vec{f}''(u)$$

$$\nabla^2 \vec{f}(u) = \vec{f}''(u) \tag{10.36}$$

and thus, $\vec{f}(u)$ satisfies Eq. (10.35). Solutions of Eqs. (10.34) and (10.35) in terms of $\vec{E}(\vec{r}, t)$ and $\vec{B}(\vec{r}, t)$ which depends on $\vec{r}$ and t only through the combination $\vec{r} \cdot \hat{n} - ct$ represents a wave moving in the direction $\hat{n}$ with a velocity c. This will become evident if we take an example.

Let the electric field $\vec{E}$ be given by a function $\vec{f}(u)$ with $\vec{f}$ having only a y component and let $\hat{n}$ be along the x direction as in Fig. 10.3. The profile of $E_y(x, t)$ vs x represents some pattern which moves unchanged as time goes on. This is very much like a wave on the water surface.

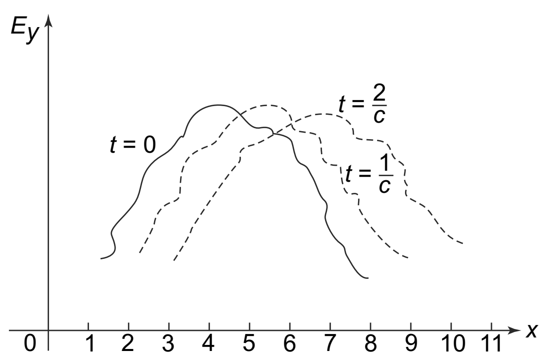


Fig. 10.3 *The profiles of the y component of the electric field given by $E_y = f(x - ct)$ at various times. The pattern moves as a whole with time*

Thus, we see that Maxwell's equations have wave solutions which move with a velocity c. But we know that

$$\boxed{c = \frac{1}{\sqrt{\mu_0 \varepsilon_0}}}$$

and since the values of μ_0 and ε_0 are known, we get $c = 3.0 \times 10^8$ m/s. This number was already known in Maxwell's time as the speed of light. This led Maxwell to make his historic observation about the nature of light as

'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'. (**On Physical Lines of Force' (1862). In W. D. Niven (ed.), The Scientific Papers of James Clerk Maxwell** *(1890), Vol. 1, 500.*)

These undulations or waves are called electromagnetic waves.

Apart from the speed of the electromagnetic waves, which we have seen is a constant in free space, there are several other properties which we can deduce immediately. If we write

$$\vec{B} = \vec{f}_B(\vec{r} \cdot \hat{n} - ct) \tag{10.37}$$

and

$$\vec{E} = \vec{f}_E(\vec{r} \cdot \hat{n} - ct) \tag{10.38}$$

we can use Maxwell's equations to relate these. Thus, we know that

$$\vec{\nabla} \times \vec{E} = -\frac{\partial \vec{B}}{\partial t} = -c\vec{f}'_B(\vec{r} \cdot \hat{n} - ct) \tag{10.39}$$

But

$$\vec{\nabla} \times \vec{E} = \hat{n} \times \vec{f}'_E(\vec{r} \cdot \hat{n} - ct) \tag{10.40}$$

But these equations and relations are valid for all times and at all points. Thus, we see that

$$\vec{f}_B(\vec{r} \cdot \hat{n} - ct) = -\frac{1}{c}\hat{n} \times \vec{f}'_E(\vec{r} \cdot \hat{n} - ct) \tag{10.41}$$

or

$$\boxed{\vec{B} \cdot \vec{E} = \vec{f}_B \cdot \vec{f}_E = 0} \tag{10.42}$$

The electric and magnetic field vectors in an electromagnetic wave are orthogonal to each other.

We can also use that Maxwell's equations giving the divergence of the electric and magnetic field vectors

$$\vec{\nabla} \cdot \vec{E} = 0$$
$$\vec{\nabla} \cdot \vec{B} = 0$$

to get

$$\boxed{\hat{n} \cdot \vec{f}_E(\vec{r} \cdot \hat{n} - ct) = 0} \tag{10.43}$$

and

$$\boxed{\hat{n} \cdot \vec{f}_B(\vec{r} \cdot \hat{n} - ct) = 0} \tag{10.44}$$

Thus, not only are the electric and magnetic fields orthogonal to each other, they are both orthogonal to the direction of propagation. **The direction of the electric field is called the polarisation direction.**

Since, as we have seen, this is perpendicular to the direction of propagation ($\vec{k}$), there are two linearly independent polarisation directions. Any general polarisation is then a superposition of these two polarisations.

We can now summarise what we have learnt about the electromagnetic waves in free space.

> The solutions to Maxwell's equations in free space, in the absence of charge and current densities are travelling waves. These waves travel with a velocity of light. The oscillations are of the electric and magnetic fields. The electric and magnetic fields are orthogonal to each other and to the direction of propagation. The waves are thus transverse in nature.

We have derived all these results for electromagnetic waves in free space using the Maxwell's equations Eq. (10.13). If instead, we used the Maxwell's equations in the presence of matter, Eqs. (10.19), then we see that the situation is very similar. For linearly polarisable media, $\vec{D} = \varepsilon\vec{E}$ and $\vec{B} = \mu\vec{H}$. If we substitute these in Eq. (10.19) and then go through the steps leading from Eq. (10.30) to Eq. (10.35), we get the wave equation in the presence of a medium, which is exactly like Eqs. (10.34) and Eq. (10.35) with the substitution $\varepsilon_0 \to \varepsilon$, $\mu_0 \to \mu$, $c \to c_m = \frac{1}{\sqrt{\mu\varepsilon}}$.

Thus, all the results we have obtained above remain valid with these substitutions. In particular, we get that electromagnetic waves in matter travel with a velocity

$$c_m = \frac{1}{\sqrt{\mu\varepsilon}} = \frac{c}{\sqrt{\mu_r\varepsilon_r}} \tag{10.45}$$

The speed of electromagnetic waves in matter is thus different from vacuum and the ratio of the two speeds is given by

$$\boxed{\frac{c_m}{c} = \frac{1}{\sqrt{\mu_r\varepsilon_r}} = \frac{1}{n}} \tag{10.46}$$

$n = \frac{c}{c_m}$, the ratio of the speed of light in vacuum to that in the medium is called the **refractive index of the medium**. Thus, we have a remarkable fact that the refractive index of a material is determined by its relative permittivity and relative permeability.

EXAMPLE 10.2 An electric field in free space has the form.

$$\vec{E}(\vec{r}, t) = E_0\hat{x}\sin^3(z - ct)$$

(a) Can this field be an electromagnetic wave?

(b) If so, determine its frequency or frequencies and

(c) The magnetic field associated with it.

Solution

(a) For an electric or magnetic field to be an electromagnetic wave, it should satisfy the wave equation given in Eq. (10.34 or 10.35). If we substitute the field given into the wave equation for $\vec{E}$, we get

$$\begin{aligned}\nabla^2 \vec{E} &= \frac{\partial^2 \vec{E}}{\partial z^2} \\ &= E_0 \hat{x} \frac{\partial^2}{\partial u^2} \left(\sin^3 u\right) \quad (u = z - ct)\end{aligned}$$

and

$$\frac{1}{c^2} \frac{\partial^2 \vec{E}}{\partial t^2} = E_0 \hat{x} \frac{\partial^2}{\partial u^2} \left(\sin^3 u\right) \tag{10.47}$$

Thus, we see that this electric field satisfies the wave equation and hence it is an electromagnetic wave.

(b) To determine the frequency, we need to write the electric field in a way from where the frequency can be read out, i.e., in the canonical way in which we have represented waves. This can easily be done and we can write

$$\vec{E} = E_0 \hat{x} \frac{1}{4} (3 \sin u - \sin(3u)) = \frac{E_0}{4} \hat{x} \left[3 \sin(z - ct) - \sin(3z - 3ct)\right]$$

Thus, we see that the given wave is a superposition of two waves of frequencies c and $3c$.

(c) To determine the magnetic field associated with this electromagnetic wave, we can use the relationship between the electric and magnetic fields in an electromagnetic wave. Alternately, we know that the relationship between $\vec{E}$ and $\vec{B}$ should be consistent with Maxwell's equations. Thus,

$$\vec{\nabla} \times \vec{E} = \hat{y} E_0 \frac{\partial}{\partial z} \left(\sin^3(z - ct)\right) = \hat{y} E_0 \frac{\partial \sin^3 u}{\partial u}$$

But this should be equal to $-\frac{\partial \vec{B}}{\partial t}$ according to Faraday's laws. Thus, we have

$$\vec{B} = -\hat{y} \frac{E_0}{c} \sin^3(z - ct)$$

PROBLEM 10.5 If the magnetic field intensity in free space of a plane wave is given by

$$\vec{H} = 100 \cos(3 \times 10^4 t + \alpha z) \hat{i} \quad \text{A/m}$$

find the $\vec{E}$ field and the time averaged power flow per unit area.

10.5.1 Monochromatic Waves and Polarisation

We have seen above that electromagnetic waves are transverse vibrations of the electric and magnetic field which travel with a velocity which depends on the electrical and magnetic properties of the medium in which they propagate. A particular form of electromagnetic waves are plane waves of a particular frequency. The electric field for such a wave can be written as

$$\vec{E}(\vec{r}, t) = \vec{E}_0 \cos(\vec{k} \cdot \vec{r} - \omega t + \delta) \tag{10.48}$$

where the direction of propagation is along $\vec{k} = k\hat{n}$ with k related to ω by $k = \frac{\omega}{c}$. But we know that c is related to the permittivity and permeability and hence $k = \omega\sqrt{\varepsilon_0\mu_0}$ for propagation in vacuum. For propagation in a medium, the relation becomes $k = \frac{\omega}{c_m} = \omega\sqrt{\mu\varepsilon}$, as given in Eq. (10.46).

The amplitude of the wave is $\vec{E}_0$ which is orthogonal to $\hat{n}$ that is

$$\vec{E}_0 \cdot \hat{n} = 0$$

Like in any electromagnetic wave, the electric and magnetic fields are not independent but are related by Maxwell's equations as

$$\vec{\nabla} \times \vec{E} = -\frac{\partial \vec{B}}{\partial t}$$

and so we get

$$\vec{B}(\vec{r}, t) = \left(\frac{\vec{k} \times \vec{E}_0}{\omega}\right)\cos(\vec{k} \cdot \vec{r} - \omega t + \delta) \tag{10.49}$$

Clearly, the magnetic field is orthogonal to both $\vec{E}$ and $\hat{n}$ as it should be.

$$\vec{B} \cdot \vec{E} = 0$$
$$\hat{n} \cdot \vec{B} = 0$$

Assuming that the electric field is in the x direction, the magnetic field will be in the y direction and the direction of propagation will be the z direction. Such kinds of waves are said to be **plane polarised**. The plane containing the magnetic field vector $\vec{H}$ and the propagation vector $\vec{k}$ is customarily said to be the **plane of polarisation**. However, nowadays it is more common to refer to the plane of vibration which is the plane defined by $\vec{E}$ and $\vec{k}$. Clearly, any plane polarised wave can be considered to be the sum of two components which are perpendicular to each other and in phase. Thus, as we shall see, we can resolve the vector $\vec{E}$ into two mutually orthogonal components, say $\vec{E}_1$ and $\vec{E}_2$.

If we add two plane polarised waves which are not in phase, then the situation is different. With the two components, say $\vec{E}_1$ and $\vec{E}_2$ which are not in phase will lead to the resultant $\vec{E}$ which will describe an ellipse in the plane of polarisation. Such a wave is called **elliptically polarised**.

Finally, if the two components $\vec{E}_1$ and $\vec{E}_2$ are exactly 90° out of phase then the ellipse will reduce to a circle and we get **circularly polarised wave**. The handedness of polarisation depends on whether the vectors $\vec{E}$ and $\vec{H}$ are rotating clockwise or anticlockwise to a person who is facing the $-\vec{k}$ direction. Clockwise rotation gives us right-handed polarisation and anti-clockwise, left-handed polarisation.

We have seen that $\vec{E}_0$ is orthogonal to $\vec{k}$. Therefore, in the plane orthogonal to $\vec{k}$, we can resolve $\vec{E}_0$ into two orthogonal components in the direction of the two unit vectors $\hat{\varepsilon}_k^1$ and $\hat{\varepsilon}_k^2$ which have the property that

$$\begin{aligned}\hat{\varepsilon}_k^1 \cdot \hat{\varepsilon}_k^2 &= 0\\ \hat{\varepsilon}_k^1 \cdot \hat{n} &= 0\\ \hat{\varepsilon}_k^2 \cdot \hat{n} &= 0\end{aligned} \tag{10.50}$$

This allows us to write $\vec{E}$

$$\vec{E}_k = \sum_{r=1}^{2} \hat{\varepsilon}_k^r E_{k0}^r \cos(\vec{k} \cdot \vec{r} - \omega t + \delta) \tag{10.51}$$

where we have denoted the electric field as $\vec{E}_k$ to remind ourselves that this decomposition is for a particular $\vec{k}$.

These explicit forms of the field in terms of the two components in the plane orthogonal to the direction of propagation allows us to easily compute the electric and magnetic energy densities using Eqs. (10.26, 10.27). In free space,

$$\begin{aligned} W_E &= \frac{\varepsilon_0}{2}|\vec{E}(\vec{r}, t)|^2 \\ &= \frac{\varepsilon_0}{2}\vec{E}_0^2 \cos^2(\vec{k} \cdot \vec{r} - \omega t + \delta) \\ &= \frac{\varepsilon_0}{2}\left(|E_{k0}^1|^2 + |E_{k0}^2|^2\right) \cos^2(\vec{k} \cdot \vec{r} - \omega t + \delta) \end{aligned} \tag{10.52}$$

Similarly, using the expression for the $\vec{B}$ field, we get

$$\begin{aligned} W_B &= \frac{1}{2\mu_0}|\vec{B}(\vec{r}, t)|^2 \\ &= \frac{1}{2\mu_0}\frac{k^2}{\omega^2}\left(|E_{k0}^1|^2 + |E_{k0}^2|^2\right) \cos^2(\vec{k} \cdot \vec{r} - \omega t + \delta) \\ &= W_E \end{aligned} \tag{10.53}$$

since $\frac{k^2}{\omega^2} = \frac{1}{c^2} = \mu_0 \varepsilon_0$. The total energy density in the electromagnetic wave is thus

$$W = W_E + W_B = 2W_E$$

Thus, we see that in an electromagnetic wave, the electric and magnetic energy densities are equal at all times, though both vary with time and space. The average densities are

$$< W_E >=< W_B >= \frac{\varepsilon_0}{4}\left(|E_{k0}^1|^2 + |E_{k0}^2|^2\right) \tag{10.54}$$

and the average energy density for the wave is

$$< W >=< W_E > + < W_B >= \frac{\varepsilon_0}{2}\left(|E_{k0}^1|^2 + |E_{k0}^2|^2\right) \tag{10.55}$$

For any given volume V, energy escapes at a rate given by the Poynting vector, as we saw above in Eq. (10.25). Thus, the rate of energy flow per unit area is simply $|\vec{P}|$. In free space, $\vec{H} = \frac{\vec{B}}{\mu_0}$ and so we get

$$\begin{aligned} |\vec{P}| &\equiv |\vec{E} \times \vec{H}| \\ &= \frac{1}{\mu_0}|\vec{E} \times \vec{B}| \\ &= \frac{1}{\mu_0}\frac{k}{\omega}|\vec{E}|^2 \\ &= \sqrt{\frac{\varepsilon_0}{\mu_0}}\left(|E_{k0}^1|^2 + |E_{k0}^2|^2\right) \cos^2(\vec{k} \cdot \vec{r} - \omega t + \delta) \\ &= cW \end{aligned} \tag{10.56}$$

In a medium, the energy density of the electric field is

$$W_m = \left(\frac{\varepsilon}{2}\right)\left(|E^1_{k0}|^2 + |E^2_{k0}|^2\right)\cos^2(\vec{k}\cdot\vec{r} - \omega t + \delta) \tag{10.57}$$

We also have $c_m = \frac{\omega}{k} = \frac{c}{n_{m0}}$ where

$$n_{m0} = \sqrt{\frac{\mu\varepsilon}{\mu_0\varepsilon_0}}$$

is the refractive index of the medium with respect to the vacuum. The Poynting vector thus, in the medium is

$$\vec{P}_m = \left(\frac{1}{\mu}\right)(\vec{E}\times\vec{B}) \tag{10.58}$$

and its magnitude is

$$|\vec{P}_m| = c_m W_m$$

This is the standard relation between energy density and rate of flow of energy per unit area or intensity. Consider electromagnetic waves propagating in the z direction. Then, if we consider a cylindrical volume with its axis parallel to the z axis (Fig. 10.4) and of length c. The end faces of the cylinder are of unit area. Now the energy contained within this volume is obviously cW and all of this will flow out of the right end face (in the $+z$ direction) in one second.

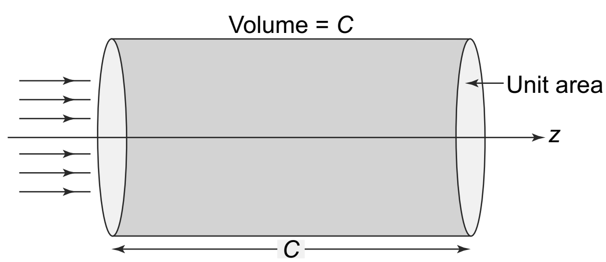


Fig. 10.4 *Flow of energy for electromagnetic waves. The direction of propagation is z direction. In 1 second, all the energy in the shaded volume would flow out of the surface on the right*

We have considered above monochromatic waves or waves with a definite $\vec{k}$ and ω. These are extremely useful in studying waves in general. The tool for doing this is the Fourier integral decomposition which allows for the decomposition of a general wave into its plane wave components. Thus, using the Fourier decomposition, we can study general electromagnetic waves since each plane wave component can be analysed as above and the general wave is simply a superposition of these different components.

EXAMPLE 10.3 The intensity of sunlight received on earth is approximately 1.3×10^3 watts/sq.m. Compare the average magnitude of the magnetic field with the earth's magnetic field at the equator.

Solution

The intensity of radiation is given by

$$I = c\left[\frac{\varepsilon_0}{2} < \vec{E}^2 > + \frac{1}{2\mu_0} < \vec{B}^2 >\right] = \frac{c}{\mu_0} < \vec{B}^2 >$$

since the two terms are equal. This gives us

$$< \vec{B}^2 >= 5 \times 10^{-12} \quad \text{T}^2$$

or

$$< \vec{B} > \sim 2.2 \times 10^{-6} \quad \text{T}$$

In comparison, the earth's magnetic field near the equator is roughly

$$|\vec{B}_e| \sim 2 \times 10^{-5} \quad \text{T}$$

PROBLEM 10.6 The electric field intensity of a wave is given by

$$\vec{E} = 3\cos(\omega t - \beta x - 45^\circ) + 4\sin(\omega t - \beta x + 45^\circ)\hat{k} \quad \text{V/m}$$

Find the polarisation of the wave.

PROBLEM 10.7 In empty space, a magnetic field is given by

$$\vec{B} = B_0 e^{ax}\hat{k}\sin(ky - \omega t)$$

Calculate the value of $\vec{E}$ and the speed of such a field.

10.5.2 Fourier Integral Decomposition *

Consider a function $f(x)$ defined over $-\infty < x < \infty$. Further, assume that the function and its derivative are both piecewise continuous in this domain and the integral

$$\int_{-\infty}^{\infty} |f(x)| dx$$

converges. Fourier's integral theorem states that for such a function

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dk f(k) e^{ikx} \tag{10.59}$$

where

$$f(k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dx f(x) e^{-ikx} \tag{10.60}$$

$f(k)$ is called the Fourier transform of $f(x)$. For a given $f(x)$, the Fourier transform is unique.

For any two functions $f(x)$ and $g(x)$, which admit Fourier decomposition

$$\int_{-\infty}^{\infty} f(x)g(x)dx = \int_{-\infty}^{\infty} f(k)g(-k)dk \tag{10.61}$$

where $f(k)$ and $g(k)$ are Fourier transforms of $f(x)$ and $g(x)$ respectively.

The Fourier transform of $f^*(x)$ is easily seen to be $f^*(-k)$. To see this consider

$$\begin{aligned} f^*(x) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dk f^*(k) e^{-ikx} \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dk f^*(-k) e^{ikx} \end{aligned} \tag{10.62}$$

Thus, if $f(x)$ is real, $f^*(-k) = f(k)$.

Choosing $g(x)$ to be $f^*(x)$ and therefore the Fourier transform of $f^*(x)$ is $g(k) = f(-k)$, we see that

$$\begin{aligned} \int_{-\infty}^{\infty} |f(x)|^2 dx &= \int_{-\infty}^{\infty} f(k)g(-k)dk \\ &= \int_{-\infty}^{\infty} |f(k)|^2 dk \end{aligned} \tag{10.63}$$

This identity is very useful since it allows us to resolve integrated total intensities into Fourier components with $f(x)$ as the amplitude.

We have till now only considered $f(x)$ as a scalar function. Generalisation to the case where $f(x)$ is a vector function $\vec{f}(x)$ is easy. We just do the procedure above for each component independently and get

$$\vec{f}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dk \vec{f}(k) e^{ikx} \tag{10.64}$$

Thus, we see that the Fourier transform of $\vec{f}(x)$ becomes a vector $\vec{f}(k)$.

Further, we have only considered $\vec{f}(x)$ to be a function of only one variable. Suppose the function is a function of more than one variable like the coordinate $\vec{r}$, then the Fourier transform is obtained once again by taking the decomposition for each of the independent variables separately

$$\begin{aligned} \vec{f}(\vec{r}) &= \left(\frac{1}{\sqrt{2\pi}}\right)^3 \iiint dk_x dk_y dk_z \vec{f}(k_x, k_y, k_z) e^{i(k_x x + k_y y + k_z z)} \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^3 \iiint d^3k \vec{f}(\vec{k}) e^{i\vec{k}\cdot\vec{r}} \end{aligned} \tag{10.65}$$

Equation (10.63) can be generalised to

$$\iiint_{-\infty}^{\infty} |\vec{f}(\vec{r})|^2 d^3\vec{r} = \iiint_{-\infty}^{\infty} |\vec{f}(\vec{k})|^2 d^3\vec{k} \tag{10.66}$$

We can apply the Fourier decomposition to electric and magnetic fields satisfying the wave equation. We write

$$\vec{E}(\vec{r}, t) = \left(\frac{1}{2\pi}\right)^{3/2} \int d^3\vec{k}\,\vec{E}(\vec{k}, t)e^{i\vec{k}\cdot\vec{r}} \tag{10.67}$$

and

$$\vec{B}(\vec{r}, t) = \left(\frac{1}{2\pi}\right)^{3/2} \int d^3\vec{k}\,\vec{B}(\vec{k}, t)e^{i\vec{k}\cdot\vec{r}} \tag{10.68}$$

The Maxwell equation

$$\vec{\nabla} \times \vec{E} = -\frac{\partial \vec{B}}{\partial t}$$

relates $\vec{E}(\vec{k}, t)$ to $\vec{B}(\vec{k}, t)$

$$\vec{k} \times \vec{E}(\vec{k}, t) = \omega\vec{B}(\vec{k}, t)$$

Inserting these in the wave equations Eqs. (10.34,10.35), we get

$$\begin{aligned} \left(k^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)\vec{E}(\vec{k}, t) &= 0 \\ \left(k^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)\vec{B}(\vec{k}, t) &= 0 \end{aligned} \tag{10.69}$$

These equations can be easily solved to get

$$\begin{aligned} \vec{E}_i(\vec{k}, t) &= E_i(\vec{k})e^{-i\omega t} \\ &= E_{io}(\vec{k})e^{-i\omega t + \delta^E_{ik}} \\ B_i(\vec{k}, t) &= B_i(\vec{k})e^{-i\omega t} \\ &= B_{io}(\vec{k})e^{-i\omega t + \delta^B_{ik}} \end{aligned} \tag{10.70}$$

where we have factored out the phases of the complex quantities $E_i(\vec{k})$ and $B_i(\vec{k})$ as

$$\begin{aligned} E_i(\vec{k}) &= E_{io}(\vec{k})e^{i\delta^E_{ik}} \\ B_i(\vec{k}) &= B_{io}(\vec{k})e^{i\delta^B_{ik}} \end{aligned}$$

with $E_{io}(\vec{k})$ and $B_{io}(\vec{k})$ as real.

Here $\omega = kc$ and $\vec{B}_0(k) = \frac{\vec{k}\times\vec{E}_0(k)}{\omega}$ as we saw above. Substituting in the Fourier integral, we get for the physical electric and magnetic fields

$$\begin{aligned} E_i(\vec{r}, t) &= \left(\frac{1}{2\pi}\right)^{3/2} \int d^3\vec{k} \left|E_{io}(\vec{k})\right| \cos(\vec{k}\cdot\vec{r} - \omega t + \delta^E_{ik}) \\ B_i(\vec{r}, t) &= \left(\frac{1}{2\pi}\right)^{3/2} \int d^3k \left|B_{io}(\vec{k})\right| \cos(\vec{k}\cdot\vec{r} - \omega t + \delta^B_{ik}) \end{aligned} \tag{10.71}$$

where we have taken the real part of the expression, since we know that the physical fields are real. Thus, the general solution to the wave equation is a superposition of monochromatic plane waves considered earlier in Eqs. (10.48) and (10.49).

One can see the usefulness of Fourier decomposition considered above, if we look at the energy expression. The total energy of the electromagnetic wave is

$$W = \int d^3\vec{r}\left[\frac{\varepsilon_0}{2}|\vec{E}(\vec{r},t)|^2 + \frac{1}{2\mu_0}|\vec{B}(\vec{r},t)|^2\right] \tag{10.72}$$

We can use Eq. (10.64) and get

$$W = \int d^3\vec{k}\left[\frac{\varepsilon_0}{2}|\vec{E}_k(\vec{r},t)|^2 + \frac{1}{2\mu_0}|\vec{B}_k(\vec{r},t)|^2\right] \tag{10.73}$$

Thus, we see that energy of the electromagnetic wave becomes the integral over the energies of individual component waves of different k values.

The Fourier decomposition will prove to be very useful when we study the propagation of electromagnetic waves in dispersive media.

10.6 DISPERSION IN A MEDIUM

We have studied solutions to Maxwell's equations in free space as well as linear polarisable media. These media are characterised by a linear relationship between the polarisation vector $\vec{P}$ and the applied electric field $\vec{E}$, as we saw in Chapter 3. For static fields, we saw that the polarisation was primarily due to the applied electric field moving or displacing the positive and negative charges in a molecule. For such fields, in equilibrium conditions, it is reasonable to expect that the displacement and hence the induced dipole moments are proportional to the applied electric field. However, when the electric field is changing with time, as is the case with electromagnetic waves, we need to see if the assumptions hold true. We will see that indeed, things are a bit more complicated for time dependent fields and lead to interesting phenomena.

Consider an electric field which varies sinusoidally with time

$$\vec{E}(t) = \vec{E}_0\cos(\omega t)$$

A particle with charge e and mass m will experience a force in such an electric field

$$\vec{F}_E = e\vec{E}(t)$$

As soon as the charge moves, there will be a restoring force which would try to resist the change in the position of the charge. This restoring force, for small values of the displacement $\vec{r}(t)$ will be proportional to $\vec{r}$. We can write this restoring force therefore, as

$$\vec{F}_{\text{res}} = -m\omega_0^2\vec{r}(t)$$

There is yet another effect which comes into play when charges move in a medium in the presence of an applied electric field. This would be a viscous or frictional force which comes about because of the effect of other charges and molecules in the neighbourhood. These viscous forces would in general be proportional to the speed of the particle and would oppose the motion. Thus, we may model it as

$$\vec{F}_{\text{vis}} = -\gamma \frac{\vec{d}r}{dt}$$

Under the combined effect of these forces—the force due to the applied electric field, the restoring force and the viscous forces, the equation of motion of the particle would be

$$m\frac{d^2\vec{r}}{dt^2} = e\vec{E}_0\cos(\omega t) - m\omega_0\vec{r} - \gamma\frac{\vec{d}r}{dt} \tag{10.74}$$

The simplest way to solve this equation for a driven, damped oscillator is to write it in terms of a complex $\vec{r}$ or $\vec{r}_C$. In terms of $\vec{r}_C$, the equation becomes

$$m\frac{d^2\vec{r}_C}{dt^2} = e\vec{E}_0 e^{i\omega t} - m\omega_0\vec{r}_C - \gamma\frac{d\vec{r}_C}{dt} \tag{10.75}$$

It is clear that $\text{Re}\,\vec{r}_C$ will satisfy Eq. (10.74). We drop the subscript with the understanding that the real part of $\vec{r}_C$ will be the solution or the physical solution. In the complex form (Eq. (10.75)), the solution can be written down as

$$\vec{r} = \frac{e\frac{\vec{E}_0}{m}e^{-i\omega t}}{\omega_0^2 - \omega^2 - \frac{i\omega\gamma}{m}} \tag{10.76}$$

This displacement, as we have seen in Chapter 3, gives rise to a molecular dipole moment $\vec{p} = e\vec{r}$. Assuming N molecules per unit volume, the polarisation $\vec{P}$ which is simply the dipole moment per unit volume becomes

$$\vec{P} = \frac{Ne^2\frac{\vec{E}_0}{m}e^{-i\omega t}}{\omega_0^2 - \omega^2 - \frac{i\omega\gamma}{m}} \tag{10.77}$$

Remember that what we are calling $\vec{r}$ is actually $\vec{r}_C$ and to find the physical quantities we need to take the real part. Thus, the physical polarisation becomes

$$\text{Re}\,\vec{P} = \frac{Ne^2\frac{\vec{E}_0}{m}\cos(\omega t + \delta)}{\left[(\omega_0^2 - \omega^2)^2 + \frac{\gamma^2\omega^2}{m^2}\right]^{1/2}} \tag{10.78}$$

where

$$\tan\delta = \frac{\frac{\gamma\omega}{m}}{\omega_0^2 - \omega^2}$$

This is a remarkable result since it shows that the physical polarisation which results from the electric field varying sinusoidally ($\vec{E} = \text{Re}\,\vec{E}_0 e^{i\omega t} = \vec{E}_0\cos(\omega t)$) is not in phase with the source that is the electric field itself. In terms of the real, physical field and polarisation, the direct proportionality relation does not hold for this sinusoidally varying fields. However, in terms of the complex field and polarisation, the direct relation holds and we have

$$\vec{P} = \varepsilon_0\chi\vec{E} \tag{10.79}$$

Here χ is complex and is given by

$$\chi = \frac{Ne^2}{m\varepsilon_0}\frac{1}{(\omega_0^2 - \omega^2) - i\frac{\gamma\omega}{m}} \tag{10.80}$$

Clearly the fact that χ and hence, $\varepsilon = \varepsilon_0(1+\chi)$ are complex for $\gamma \neq 0$ is a reflection of the fact that the displacement $\vec{r}$ and $\vec{E}$, the sinusoidal electric field causing it are not in phase.

An interesting fact emerges from Eq. (10.80)- consider a medium in which the viscous term is small, that is one can then drop the term $i\frac{\gamma\omega}{m}$ in the denominator of Eq. (10.80). Then obviously ε is real. However, even the real ε is a function of the frequency of the electromagnetic field ω. This leads to an interesting phenomenon. Recall that the velocity of light in a medium and hence its refractive index is a function of ε. The frequency dependence of ε thus implies that the refractive index (or the speed of light) is different for different frequencies. **This is the well known phenomenon of dispersion of light**, the kind which we see when white light passes through a prism for instance.

Figure 10.5 shows the qualitative behaviour of the real and imaginary parts of $\varepsilon(\omega) = \varepsilon(1+\chi(\omega))$ where $\chi(\omega)$ is given by Eq. (10.80).

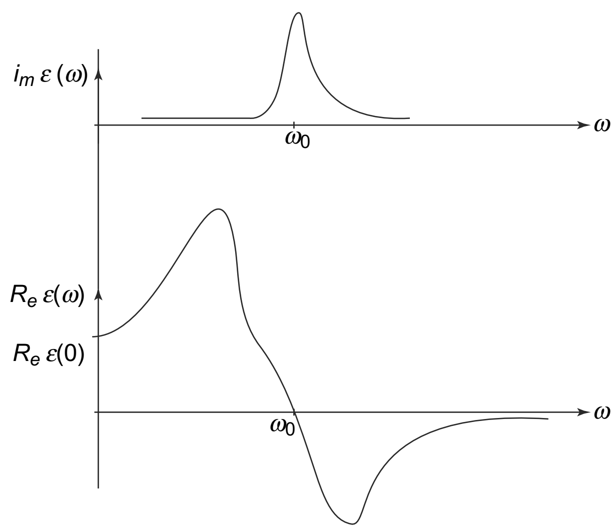


Fig. 10.5 *Behaviour of Re $\varepsilon(\omega)$ and Im $\varepsilon(\omega)$ as functions of ω in a medium. ω_0 is a constant and is a measure of the spring constant restoring force of the molecule ($m\omega_0^2\vec{r}$) which causes the charge to vibrate in the presence of the sinusoidal electric field. The y-intercept of Re $\varepsilon(\omega)$ is the static dielectric constant $\varepsilon(0) = 1 + \frac{Ne^2}{m\varepsilon_0}\frac{1}{\omega_0^2}$*

The presence of an imaginary part of $\varepsilon(\omega)$ leads to a damping of electromagnetic waves. To see this, consider a solution of the wave equation for an electromagnetic wave moving in the $+z$ direction.

$$\vec{E}(\vec{r},t) = \vec{E}_0 e^{i(kz-\omega t+\delta)} \tag{10.81}$$

with the physical field being the real part of the electric field given in Eq. (10.81). The wave number k and the frequency ω are related through the velocity of light

$$k = \omega\sqrt{\mu\varepsilon} \tag{10.82}$$

Clearly when ε becomes complex as we saw above in a medium with non-zero viscous parameter γ, then the wave number also becomes complex.

$$k = k_r + ik_i \tag{10.83}$$

Substituting this in Eq. (10.81), we see that the physical electric field then becomes

$$\text{Re}\,\vec{E}(\vec{r},t) = \vec{E}_0 e^{-k_i z}\cos(k_r z - \omega t + \delta) \tag{10.84}$$

Thus Eq. (10.84) is an electromagnetic wave propagating along the $+z$ axis with a speed $c_m = \frac{\omega}{k_r} = \text{Re}\,\frac{1}{\sqrt{\mu\varepsilon}}$. However, its amplitude gets attenuated by a factor $e^{-k_i z}$ are it travels. The energy density of such a wave in this medium also decreases with z since the energy density is proportional to $\vec{E}^2$. The energy lost is dissipated in the medium. Thus, we see that a **complex permittivity ε leads to an attenuation of electromagnetic waves**.

10.6.1 Normal and Anomalous Dispersion

Let us recapitulate what we have seen above: the dielectric constant ε is related to the complex susceptibility χ by

$$\varepsilon = \varepsilon_0(1+\chi)$$

The susceptibility χ is given by Eq. (10.80). For frequencies $\omega \ll \omega_0$, the dielectric constant $\varepsilon(\omega)$ exhibits an interesting and characteristic dependence on ω. Expanding $\varepsilon(\omega)$ in powers of $\left(\frac{\omega}{\omega_0}\right)$, we get

$$\begin{aligned}\frac{\text{Re}\,\varepsilon}{\varepsilon_0} &= 1 + \left(\frac{Ne^2}{m\varepsilon_0}\right)\frac{(\omega_0^2-\omega^2)}{(\omega_0^2-\omega^2)^2+\frac{\gamma^2\omega^2}{m^2}} + \cdots \\ &= \left(1+\frac{Ne^2}{m\varepsilon_0\omega_0^2}\right) + \left(\frac{Ne^2}{m\varepsilon_0\omega_0^2}\right)\left(\frac{\omega^2}{\omega_0^2}\right)\left[1-\frac{\gamma^2}{m^2\omega_0^2}\right]\gamma\end{aligned} \tag{10.85}$$

The refractive index of the medium,

$$n = \frac{c}{c_m} = \text{Re}\sqrt{\frac{\varepsilon}{\varepsilon_0}}$$

in this limit is thus given by (neglecting terms of order $\left(\frac{\omega^4}{\omega_0^4}\right)$ and higher)

$$n = A + \frac{B}{\lambda^2} \tag{10.86}$$

where

$$\begin{aligned}A &= \left[1+\frac{Ne^2}{m\varepsilon_0\omega_0^2}\right]^{1/2} \\ B &= \frac{2\pi^2 c_m^2 Ne^2}{m\varepsilon_0\omega_0^4 A^{1/2}}\left[1-\frac{\gamma^2}{m\omega_0^2}\right]\end{aligned} \tag{10.87}$$

Here λ is the wavelength which is given by

$$\lambda = \frac{2\pi c_m}{\omega}$$

Equation (10.86) is called the **Cauchy formula for normal dispersion.**

From the expression for $\varepsilon(\omega)$ as also Fig. 10.5, we can see that when ω is near ω_0, both $\mathrm{Re}\,\varepsilon(\omega)$ and $\mathrm{Im}\,\varepsilon(\omega)$ undergo very rapid and drastic changes. This sudden change in $\varepsilon(\omega)$ and hence the refractive index n, is called **'anomalous dispersion'.** The real part of $\varepsilon(\omega)$ rises to a high value just before $\omega = \omega_0$ and then goes to ε_0 at ω_0 as can be seen from Eq. (10.85). As we increase ω, the value becomes less than ε_0. The imaginary value of $\varepsilon(\omega)$ peaks at $\omega = \omega_0$ underlying the fact that there is maximum absorption and therefore dissipation of energy at the 'natural' frequency of the molecules or charges in the medium. This is very much like a forced oscillator being driven near its natural frequency.

As we saw, at values of $\omega > \omega_0$, $\varepsilon(\omega)$ becomes less than ε_0. Since $n = \frac{\varepsilon}{\varepsilon_0}$, this implies that the refractive index n becomes less than one, or alternatively, the speed of light in the medium, c_m for this range of frequencies is more than the speed of light in vacuum, c. This is NOT in contradiction with the principles of relativity as we shall see since this speed or velocity is the 'phase or wave' velocity of an infinite wave train of definite wave number k as in Eq. (10.81). There is no information transmitted by such an infinite train. On the other hand, a wave packet, which has a finite extent in space travels with a 'group velocity'.

10.7 PHASE AND GROUP VELOCITIES

The monochromatic wave solution for the Maxwell's equations, Eq. (10.48)

$$\vec{E}(\vec{r}, t) = \vec{E}_0 \cos(\vec{k} \cdot \vec{r} - \omega t + \delta)$$

travelling with a speed $c_m = \frac{\omega}{k} = \sqrt{\mu\varepsilon}$ in a medium, spans the entire space. The speed c_m is called the **'phase velocity'**. It is the speed with which a disturbance at a certain phase in the wave travels. Thus, as in Fig. 10.6, a wave with the electric vector in the z-direction is travelling in the x-direction which extends to infinity on both sides. A peak representing a value of phase $(n + \frac{1}{2})\pi$ for a given integer n, moves by an amount $(c_m \Delta t)$. For a wave travelling in the x-direction, a profile of a component of $\vec{E}$ namely E_z at a given time $t = 0$ and at a later time is shown in Fig. 10.6. The wave moves by an amount $c_m \Delta t$ along the x-axis.

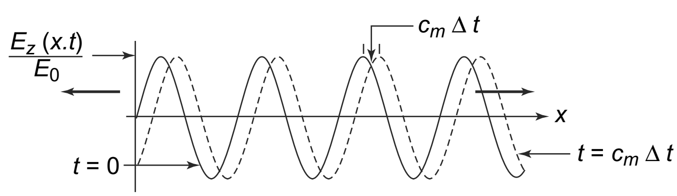


Fig. 10.6 *A monochromatic wave travelling in the x-direction with speed c_m. The wave moves bodily by an amount $c_m \Delta t$ in time Δt*

An infinite sinusoidal wave repeats itself indefinitely all over and thus cannot carry any information which obviously is localised. In vacuum or in a medium with constant value of ε, more general wave solutions of the Maxwell's equations may easily be obtained by an arbitrary superposition of solutions like Eq. (10.48) with different ω. The weight factors for different ω can be tailored to reproduce any desired profile at a given time. Thus, consider propagation in the x-direction with the electric vector in the z-direction. We use the complex formalism and write a monochromatic wave (Eq. (10.48)) as

$$E_z(\vec{r}, t) = E_0 e^{(ikx - i\omega t + \delta)} = E_0 e^{(ikx - i\frac{kt}{\sqrt{\mu\varepsilon}} + \delta)} \tag{10.88}$$

Since $\frac{\omega}{k} = c_m = \frac{1}{\sqrt{\mu\varepsilon}}$. Eq. (10.88) is a solution of the wave Eq. (10.35), with $\mu_0 \to \mu$, $\varepsilon_0 \to \varepsilon$ for any value of k. Hence, an superposition of Eq. (10.88) with different k with arbitrary weight factor is also a solution of the wave equation since the wave equation is a homogeneous and linear differential equation. These solutions are called **'wave -packets'**. We write such a superposition as

$$E_z(x,t) = \frac{1}{\sqrt{2\pi}} \int E_z(k) e^{(ikx - i\frac{kt}{\sqrt{\mu\varepsilon}})} dk \tag{10.89}$$

The solution, Eq. (10.89), is is in the form of a Fourier integral. Choosing various forms of the weight factor leads to various profiles of E_z. The weight factor can be related to the profile of $E_z(x,t)$ at any given time. Taking this time as $t = 0$ for convenience, we have by the Fourier inversion formula

$$E_z(k) = \frac{1}{\sqrt{2\pi}} \int E_z(x, t = 0) e^{(-ikx)} dx \tag{10.90}$$

If ε is a constant not dependent on ω and hence, not on k in Eq. (10.89), c_m is not dependent on k and hence, we can write Eq. (10.89) as

$$E_z(x,t) = \frac{1}{\sqrt{2\pi}} \int E_z(k) e^{(ik(x - c_m t))} dk \tag{10.91}$$

$E_z(x,t)$ thus is a function of the combination $(x - c_m t)$. Whatever profile was created for $E_z(x, t = 0)$ by one's choice of the weight factor $E_z(k)$ in Eq. (10.90), thus exists also at later times with x shifted by an amount $(c_m t)$. For constant values of ε and μ and hence of c_m, we thus have a wave like solution of the Maxwell's equation which can have any profile by suitable choice of the weight function and the whole pattern propagates with a speed c_m without altering its shape. Thus, for instance, choosing a Gaussian weight function $E_z(k)$, as an example

$$E_z(k) = e^{-k^2\sigma^2} \quad \sigma = \text{constant} \tag{10.92}$$

we get, from Eq. (10.91),

$$E_z(x,t) = e^{[\frac{-(x - c_m t)^2}{4\sigma^2}]} \tag{10.93}$$

Figure 10.7 shows the profile of $E_z(x,t)$ at $t = 0$ and at a later time, $t = t_1$. Such a solution thus represents a pulse-like solution of Maxwell's equation for constant values of ε and μ and hence the same c_m for all k values. It propagates retaining its shape with a speed c_m. Other such finite size solutions can similarly be obtained by making the appropriate choice of the weight function, using Eq. (10.90). Thus, the peak or any other marked point in the profile propagates with the same speed c_m as the bulk of the wave-packet. This has come about because of the linear relationship assumed between ω and k

$$\omega = k c_m \tag{10.94}$$

with, ε, μ and hence $c_m = \frac{1}{\sqrt{\mu\varepsilon}}$ all constants.

Similarly, a finite wave train of finite length which is typically the form of light emitted by matter also can be considered as superposition of monochromatic waves. Using complex formalism, such a wave train of wave number k_0 has the form

$$\begin{aligned} E_z(x,t) &= E_0 e^{(ik_0 x - i c_m k_0 t)} \quad 0 < x < L \\ &= 0 \quad x < 0 \quad \text{or} \quad x > L \end{aligned} \tag{10.95}$$

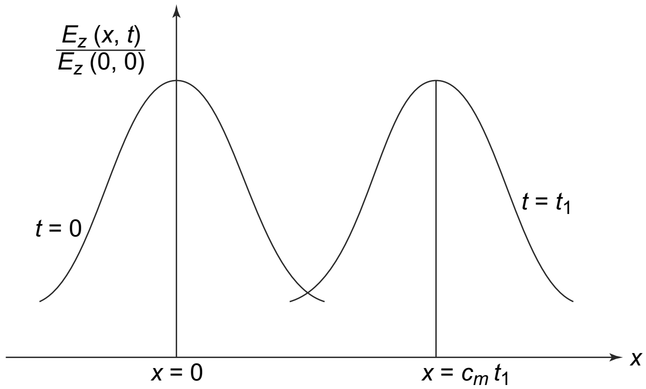


Fig. 10.7 *Profiles at two different times $t = 0$ and $t = t_1$ of a pulse like solution of the Maxwell's equation in a medium with constant μ and ε and hence, a constant c_m for all values of k. The pulse propagates completely retaining its shape with a speed c_m*

The weight function $E_z(k)$ for such a wave train is, from Eq. (10.89)

$$\begin{aligned} E_z(k) &= \frac{1}{\sqrt{2\pi}} E_0 \int_0^L e^{(ik_0x)} e^{(-ikx)} dx \\ &= E_0 L \frac{\sin\beta}{\beta} e^{i\frac{(k_0-k)L}{2}} \end{aligned} \tag{10.96}$$

where $\beta = \frac{(k_0-k)L}{2}$.

The shape of Re $E_z(x,t)$ and $|\frac{E_z(k)}{(E_0L)}|$ is shown in Fig. 10.8. As with the earlier example, the wave trains moves with its shape maintained with speed c_m

In dense media, ε starts depending significantly on ω leading to the phenomenon of dispersion which we have discussed. The propagation of wave packets becomes different from the propagation of monochromatic waves. We will now discuss this phenomenon restricting ourselves to ε values which are frequency dependent but real. Complex values of ε do occur in nature, e.g., in conductors which we will discuss in the next chapter.

The relation $\omega = kc_m$ which holds in non-dispersive medium is replaced in a dispersive medium by a more general relation

$$\omega = \omega(k) \tag{10.97}$$

where the functional form of the function $\omega(k)$ depends on the medium. Consider in a dispersive medium a wave packet

$$E_z(x,t) = \frac{1}{\sqrt{2\pi}} \int E_z(k) e^{(ikx - i\omega(k)t)} dk \tag{10.98}$$

For definiteness, let us assume that at a given time, say $t = 0$, this wave packet resembles a finite wave train like the one shown in Fig. 10.8(a) in the central region having a wave number k_0. At $t = 0$,

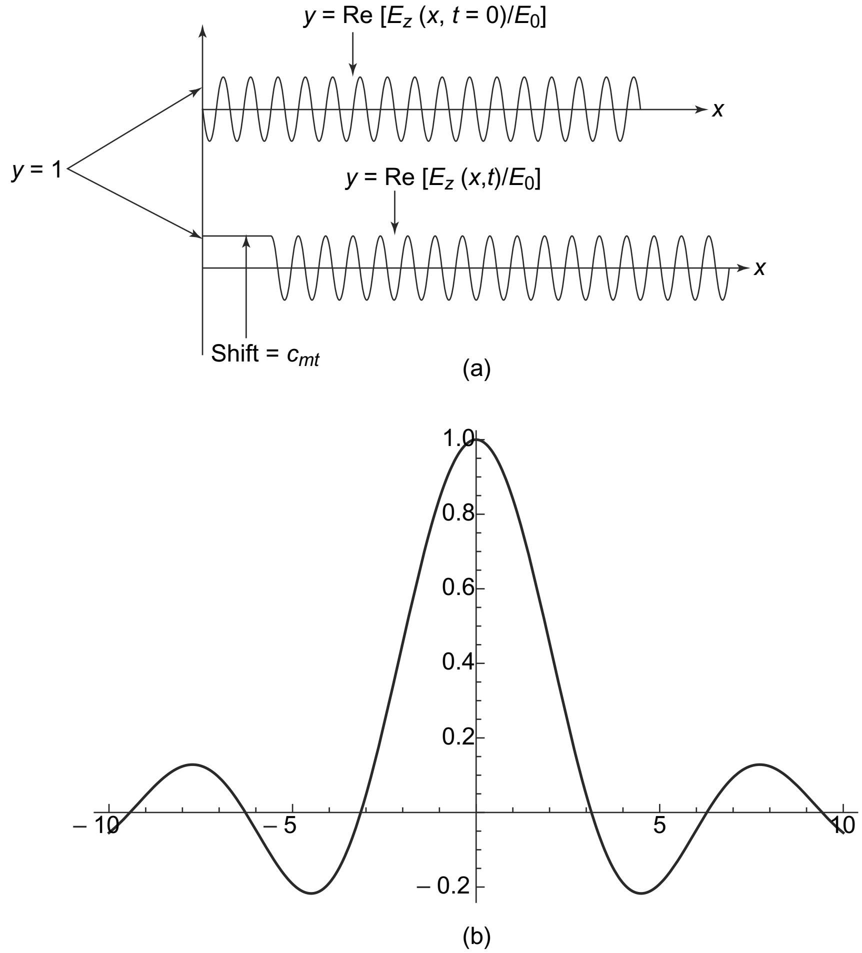


Fig. 10.8 *(a) A wave train of finite length Eq. (10.95). The y-axis is Re $E_z(x, 0)$, (b) The profile of the modulus of the weight function $|E_z(k)|$ (Eq. (10.96)). The y-axis is $\frac{|E_z(k)|}{E_0 L}$ and the x-axis is $\frac{(k-k_0)L}{2}$*

Equation (10.98) will read

$$E_z(x, t = 0) = \frac{1}{\sqrt{2\pi}} \int E_z(k) e^{(ikx)} dk \tag{10.99}$$

which will resemble the shape shown in Fig. 10.8(a) vanishing for $x < 0$ and $x > L$.

As shown in Fig. 10.8(b), the weight function $E_z(k)$ then is peaked around $k = k_0$ with the result that the contribution to the integral in Eq. (10.98) will dominantly come from the region around $k = k_0$. That being so, we can perform an expansion of $\omega(k)$ around $k = k_0$

$$\omega(k) = \omega(k_0) + (k - k_0)\omega'(k_0) + \text{terms of order} \quad (k - k_0)^2 \quad \text{and higher} \tag{10.100}$$

and retain only the first two terms in Eq. (10.100) for evaluation of the integral in Eq. (10.98). The result is

$$E_z(x,t) = \frac{1}{\sqrt{2\pi}} e^{(-i\omega_0 t + ik_0\omega'(k_0)t)} \int E_z(k) e^{(ik(x-\omega'(k_0)t))} dk \tag{10.101}$$

Comparing Eqs. (10.101 and 10.99), we see that but for the factor $e^{(-i\omega(k_0)t+ik_0\omega'(k_0)t)}$, we see $E_z(x,t)$ would be $E_z(x-\omega'(k_0)t, 0)$. Hence, but for the exponential factor, $E_z(x,t)$ would be a finite wave train that we started with at $t=0$ as in Eq. (10.99) with the position shifted by an amount $\omega'(k_0)t$. Thus, including the factor $e^{(-i\omega_0 t+ik_0\omega'(k_0)t)}$, we get from Eq. (10.101 and 10.99)

$$E_z(x,t) = e^{(-i\omega_0 t + ik_0\omega'(k_0)t)} E_z(x-\omega'(k_0)t, 0) \tag{10.102}$$

The exponential factor simply changes the phase of the sinusoidal pattern when we take the real part. The factor $E_z((x-\omega'(k_0)t), 0)$ is the original profile of size L but now shifted to the region $0 < (x-\omega'(k_0)t) < L$ at time t. We thus, have the result that the pattern effectively propagates with a velocity c_g. This is called the '**group velocity**'

$$c_g = \left(\frac{d\omega}{dk}\right)_{k=k_0} \tag{10.103}$$

As in our discussion on dispersion, the dispersive properties of a medium are normally expressed with ε as a function of ω

$$\varepsilon = \varepsilon(\omega)$$

Since $\omega = kc_m = \frac{kc}{\sqrt{\mu_r\varepsilon_r}} = \frac{kc}{n(\omega)}$ where $n(\omega)$ is the refractive index of the medium relative to vacuum, we get

$$\begin{aligned} \frac{d\omega}{dk} &= \frac{c}{n(\omega)} - kc\frac{\frac{dn(\omega)}{dk}}{n^2(\omega)} \\ &= \frac{c}{n(\omega)} - \frac{kc}{n^2(\omega)}\frac{dn(\omega)}{d\omega}\frac{d\omega}{dk} \end{aligned} \tag{10.104}$$

which finally gives

$$c_g = \frac{d\omega}{dk} = \frac{c}{n + \omega\frac{dn}{d\omega}} \tag{10.105}$$

evaluated at the central value of ω.

In the region of normal dispersion $\frac{dn(\omega)}{d\omega} > 0$. Since $n(\omega) > 1$ being ratio of speed of light in vacuum relative to that in the medium, we see from Eq. (10.105) that the group velocity is less than c. This would be satisfactory since as we mentioned before, we need wave packets to send signals. The wave packets would travel with velocity c_g and that being less than c seems to be in consonance with the theory of relativity. In the approximation we are working, namely neglecting higher order terms in Eq. (10.100), the wave packet retains its shape as it propagates. When the function $\omega(k)$ is such that such approximation is not justified, the shape of the wave packet may change.

The example below illustrates the phenomenon of change of shape.

EXAMPLE 10.4 An electromagnetic pulse travelling in the x-direction in a medium with the electric vector in the z-direction, has at time $t = 0$, the shape

$$E_z(x, t = 0) = E_0 e^{(-\frac{x^2}{4})}$$

The angular frequency $\omega(k)$ has the form $\omega(k) = k + ak^2$ (a = constant). Compare its shape at time $t = 0$ with its shape at time $t = 1$ for
(i) $a = 0$ and
(ii) $a = 1$.

Solution

The electric field initially is Gaussian, with a width = 2 as in Fig. 10.9(a). We can express $E_z(x, t)$ as in Eq. (10.98)

$$E_z(x, t) = \frac{1}{\sqrt{2\pi}} \int E_z(k) e^{(ikx - ikt - iak^2 t)} dk$$

Putting $t = 0$, we obtain from the inverse Fourier integral

$$E_z(k) = \frac{1}{\sqrt{2\pi}} \int E_z(x, t = 0) e^{(-ikx)} dx = \sqrt{2} E_0 e^{(-k^2)}$$

Using this value of $E_z(k)$, we get

$$E_z(x, t = 1) = \frac{1}{\sqrt{\pi}} E_0 \int e^{(-k^2)} e^{(ikx - ik - iak^2)} dk$$

The integral is easily evaluated. The physical value of the electric field is of course the real part of the expression above and is given by

For $a = 0$

$$E_z(x, t = 1)_{\text{physical}} = E_0 e^{\frac{(-(x-1)^2}{4}}$$

This is exactly the profile at $t = 0$ shifted in x by 1. This is the case of no dispersion and the wave and group velocities both are $\frac{\omega}{k} = 1$ (Fig. 10.9(b))

For $a = 1$

$$E_z(x, t = 1)_{\text{physical}} = 2^{-1/4} E_0 e^{(-\frac{(x-1)^2}{8})} \cos\left[\frac{(x-1)^2}{8} - \frac{\pi}{8}\right]$$

The shape thus changes (Fig. 10.9(c)). At $t = 1$, the envelope remains almost gaussian with width increased to ($\sqrt{8}$). The pulse is centred around $x = 1$ so that the group velocity is still = 1 in this case. This is so because the central value of k is $k_0 = 0$. The derivative $\frac{d\omega}{dk} = 1 + 2ak$ is hence = 1 at $k = 0$. However, there are regions in dispersion curve Fig. 10.5 where ε decreases with ω so that $\frac{dn(\omega)}{d\omega}$ is negative as also the possibility that the phase velocity $c_m > c$ implying $n < 1$. From Eq. (10.105) this implies the possibility that group velocity may under suitable circumstances be greater than c. Such possibilities actually are not totally theoretical but have actually been realised experimentally using laser pulses. (See e.g, **M.W.Mitchell and Chiao, American J. of Physics, 66. 14(1997)** for their and other experiments for 'superluminal' propagation of pulses). There is no violation of special relativity in such propagation. The pulses that are sent do not have sharp front edge, whose superluminal propagation would certainly

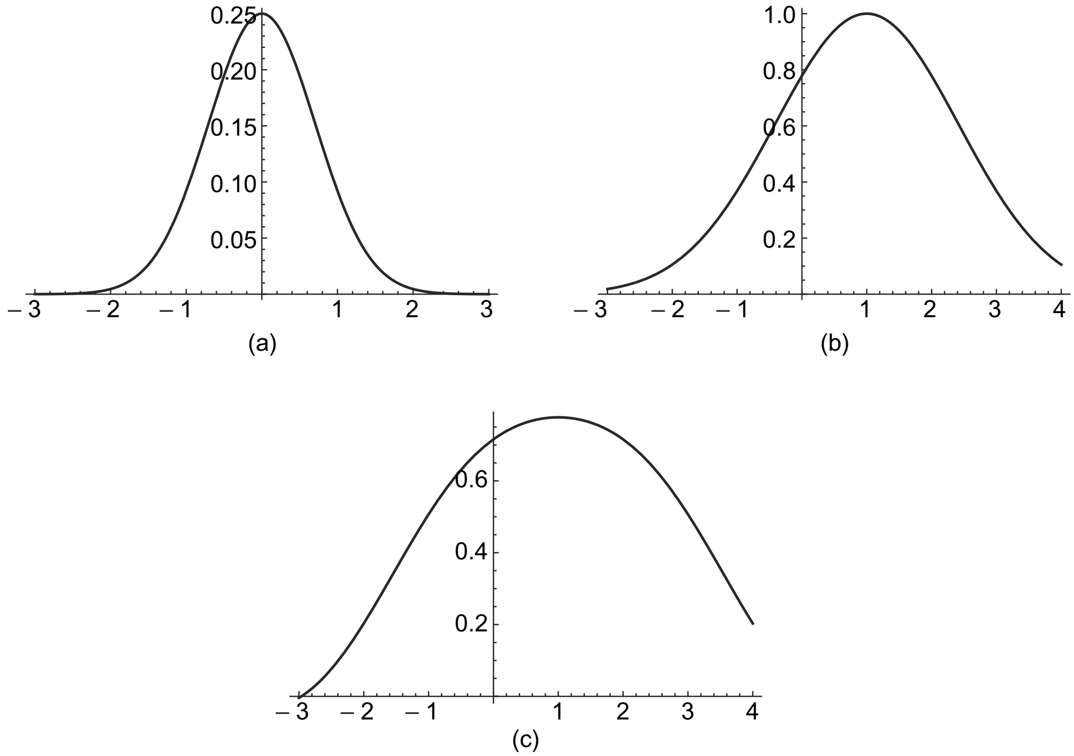


Fig. 10.9 *Example 10.4: (a) the profile of the wave packet at time $t = 0$. The y-axis is $E_z(x, 0)/E_z(0, 0)$ plotted as a function of x, (b) the profile of the same wave packet at a later time $t = 1$ for $a = 0$. There is no dispersion in this case and the wave packet moves with unchanged shape. The group and wave velocities both are =1 in this case. The y-axis is $E_z(x, 0)/E_0$ plotted as a function of x, (c) The profile of the same wave packet at time $t = 1$ for $a = 1$. There is dispersion now and shape of the wave packet has changed. The group velocity is still =1, since the central frequency in this wave packet is at $k_0 = 0$ and $\frac{d\omega}{dk} = 1 + 2ak$, which is =1 at $k = k_0 = 0$. The y-axis is $E_z(x, 0)/E_0$ plotted as a function of x*

be in violation of relativity. The receiving end of the setup in such 'superluminal' propagation with its special dispersion properties, simply 'reshapes' a pulse after the front edge reaches it. The peak of the pulse then reappears at the receiving end before the peak of the incident pulse reaches the receiving end. No violation of 'causality' thus occurs.

10.8 ELECTROMAGNETIC WAVES AT INTERFACES: REFLECTION AND REFRACTION

Till now, we have discussed the solutions of Maxwell's equations, in the form of electromagnetic waves in media which are infinite in extent and homogenous in nature. Such media are characterised by fixed values of ε and μ which are constant throughout the space. In particular, there were no boundaries and hence, no need to impose boundary conditions. In a situation where different regions of space have different values of ε and μ, we would clearly need to adopt a different strategy. This is because the solutions to Maxwell's equations in the different regions would be solutions corresponding to the relevant ε and μ for that region. However, at the interfaces of the regions, Maxwell's equations demand that certain boundary conditions on the field vectors be satisfied. This condition constrains the solutions in the different regions. The general solution to this problem of solutions at arbitrary interfaces is not easy to solve analytically. However, for simple cases of interfaces, an analytical solution is possible and we discuss such a simple case below of two regions with different ε and μ separated by an infinite plane which then serves as the interface.

Consider a planar infinite interface forming the $x - y$ plane at $z = 0$, as in Fig. 10.10.

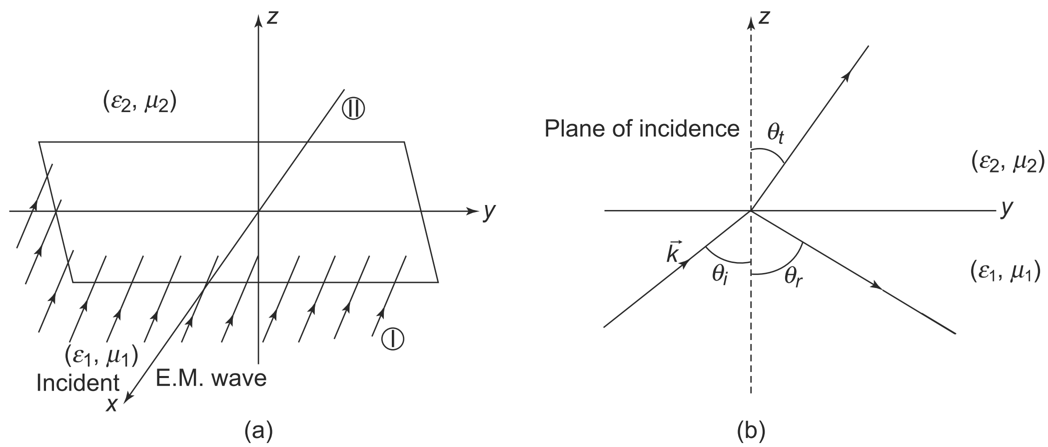


Fig. 10.10 *Electromagnetic wave at a planar interface (a) A plane interface in the $x - y$ plane at $z = 0$ separating two regions, Region I and Region II. Region II, for $z > 0$ has values (ε_2, μ_2) and Region I ($z < 0$) has values (ε_1, μ_1) for the electrical permittivity and magnetic permeability respectively. Electromagnetic waves are incident on the interface from Region I, (b) At any point on the interface, the plane of incidence containing the incident wave direction at that point and the normal to the interface at that point*

The medium above, i.e., for $z > 0$ has electrical permittivity ε_2 and magnetic permeability μ_2 while the medium below has ε_1 and μ_1. The wave velocities in medium I (Region I) are $c_1 = \frac{1}{\sqrt{\mu_1 \varepsilon_1}}$ and in medium II (Region II) $c_2 = \frac{1}{\sqrt{\mu_2 \varepsilon_2}}$. The ratio $\frac{c_1}{c_2} = \frac{n_2}{n_1}$ defined as n_{21} is the ratio of the refractive index of the second medium to that of the first. We consider electromagnetic waves of frequency ω and wave number k incident on the interface from Region I.

Figure 10.10(b) shows the side view at any point on the interface. The plane containing the normal at that point and the incident wave vector at that point defines a plane called **the plane of incidence**. If there was no change in the medium at the interface, then the incident wave would of course propagate across the interface without any change. But with a different medium across the interface, with different values of ε and μ, the wave will have a different speed across the interface. And these waves in the two media should also satisfy the boundary conditions at the interface as demanded by Maxwell's equations. This we shall see, puts constraints on the nature of propagation.

The boundary conditions on the interface for the electric and magnetic field vectors has already been discussed in Chapters 3 and 7. At any point on the interface, we should have

$$B_{\perp}^{I} = B_{\perp}^{II} \tag{10.106}$$

$$\varepsilon_1 E_{\perp}^{I} = \varepsilon_2 E_{\perp}^{II} \tag{10.107}$$

$$H_{\parallel}^{I} = H_{\parallel}^{II} \tag{10.108}$$

$$E_{\parallel}^{I} = E_{\parallel}^{II} \tag{10.109}$$

Superscripts I, II refer to the two media. The subscript $\perp$ refers to the normal or the z component of the respective fields at the interface while the subscript $\parallel$ refers to the tangential or the component of the fields in the $x - y$ plane of the respective fields.

We take the initial wave vector $\vec{k}_i$ to have y and z components. The electric field of the incident wave on the interface in Region I is therefore,

$$\begin{aligned}\vec{E}_i(\vec{r}, t) &= E_i \hat{\varepsilon}_i(\vec{k}_i) \cos(\vec{k}_i \cdot \vec{r} - \omega_i t + \delta_i) \\ \vec{k}_i &= k_i(0, \sin\theta_i, \cos\theta_i)\end{aligned} \tag{10.110}$$

Here E_i is the amplitude of $\vec{E}_i$ and $\hat{\varepsilon}_i(\vec{k}_i)$ is a unit vector normal to $\vec{k}_i$ that is $\hat{\varepsilon}_i(\vec{k}_i) \cdot \vec{k}_i = 0$. Also $k_i = \frac{\omega}{c_1} = \frac{\omega_i}{c}\sqrt{\mu_1 \varepsilon_1}$. The subscript i indicates the incident wave. There is also a transmitted wave in Region II and a reflected wave in Region I. The electric fields for these can be written as

$$\begin{aligned}\vec{E}_t(\vec{r}, t) &= E_t \hat{\varepsilon}_t(\vec{k}_t) \cos(\vec{k}_t \cdot \vec{r} - \omega_t t + \delta_t) \\ \vec{k}_t &= k_t(0, \sin\theta_t, \cos\theta_t)\end{aligned} \tag{10.111}$$

and

$$\begin{aligned}\vec{E}_r(\vec{r}, t) &= E_r \hat{\varepsilon}_r(\vec{k}_r) \cos(\vec{k}_r \cdot \vec{r} - \omega_r t + \delta_r) \\ \vec{k}_r &= k_r(0, \sin\theta_r, \cos\theta_r)\end{aligned} \tag{10.112}$$

where E_t and E_r are the amplitudes of $\vec{E}_t$ and $\vec{E}_r$ respectively. The relations between k_t and ω_t and k_r and ω_r follow similar to those between k_i and ω_i as

$$\begin{aligned}k_t &= \frac{\omega_t}{c_2} = \frac{\omega_t}{c}\sqrt{\mu_2 \varepsilon_2} \\ k_r &= \frac{\omega_r}{c_1} = \frac{\omega_r}{c}\sqrt{\mu_1 \varepsilon_1}\end{aligned}$$

The unit vectors $\hat{\varepsilon}_t(\vec{k}_t)$ and $\hat{\varepsilon}_r(\vec{k}_r)$ satisfy $\hat{\varepsilon}_t(\vec{k}_t) \cdot \vec{k}_t = 0$ and $\hat{\varepsilon}_r(\vec{k}_r) \cdot \vec{k}_r = 0$.

The magnetic fields can easily be written down since, as with all electromagnetic waves, the electric and magnetic fields are related by Eq. (10.49)

$$\vec{B}(\vec{r}, t) = \left(\frac{\vec{k} \times \vec{E}_0}{\omega} \right) \cos(\vec{k} \cdot \vec{r} - \omega t + \delta)$$

and so we get

$$\vec{B}_i(\vec{r}, t) = \left(\frac{\vec{k}_i \times E_i \hat{\varepsilon}_i}{\omega_i} \right) \cos(\vec{k}_i \cdot \vec{r} - \omega_i t + \delta_i) \tag{10.113}$$

$$\vec{B}_t(\vec{r}, t) = \left(\frac{\vec{k}_t \times E_t \hat{\varepsilon}_t}{\omega_t}\right) \cos(\vec{k}_t \cdot \vec{r} - \omega_t t + \delta_t) \tag{10.114}$$

$$\vec{B}_r(\vec{r}, t) = \left(\frac{\vec{k}_r \times E_r \hat{\varepsilon}_r}{\omega_r}\right) \cos(\vec{k}_r \cdot \vec{r} - \omega_r t + \delta_r) \tag{10.115}$$

For a given $\vec{E}_i(\vec{r}, t)$, the values of $|\vec{E}_r(\vec{r}, t)|$, $|\vec{E}_t(\vec{r}, t)|$, $\vec{k}_t$, $\vec{k}_r$, ω_t, ω_r, δ_t, δ_r will be determined by the boundary conditions, i.e., Eqs. (10.106, 10.107, 10.108 and 10.109). Let us try to impose the boundary conditions on the interface that we are considering.

At the interface, $\vec{r}$ has only x and y components since the interface is in the $x - y$ plane. Also note that in the boundary conditions we need to be careful that the fields in Region I are the sums of the fields in the incident and reflected waves while the fields in Region II are those in the transmitted wave alone. Further, the boundary conditions need to be satisfied for all values of x and y (the interface being an infinite plane) and at all times. A necessary condition for this to be satisfied is that the phases of the three fields in Eqs. (10.110,10.111 and 10.112) are equal. Thus we get

$$\omega_i = \omega_t = \omega_r = \omega \tag{10.116}$$

$$\delta_i = \delta_t = \delta_r = \delta \tag{10.117}$$

$$\vec{k}_i \cdot \vec{r} = \vec{k}_t \cdot \vec{r} = \vec{k}_r \cdot \vec{r} \tag{10.118}$$

at the interface. At the interface, $\vec{r}$ has only x and y components. The vectors $\vec{k}_i$, $\vec{k}_t$, $\vec{k}_r$ have y and z components. In terms of the angles θ_i, θ_r and θ_t, we therefore, get from Eqs. (10.118) and the definition of k,

$$\vec{k}_i \cdot \vec{r} = \frac{\omega}{c_1} \sin\theta_i \tag{10.119}$$

$$\vec{k}_t \cdot \vec{r} = \frac{\omega}{c_2} \sin\theta_t \tag{10.120}$$

$$\boxed{\vec{k}_r \cdot \vec{r} = \frac{\omega}{c_1} \sin\theta_r} \tag{10.121}$$

Using Eq. (10.118) we get

$$\boxed{\theta_i = \theta_r} \tag{10.122}$$

and

$$\frac{\sin\theta_i}{\sin\theta_t} = \frac{c_1}{c_2} = n_{21} \tag{10.123}$$

Equation (10.122) is easily recognised to be the Law of Reflection which states that the angle of incidence is the same as the angle of reflection. Equation (10.123) is Snell's Law of Refraction which relates the angle of incidence and angle of refraction to the refractive index of the media.

The boundary conditions that we have considered, not only give us the laws of reflection and refraction but also determine the relative amplitudes of the reflected and transmitted waves. To see this, we rewrite the boundary conditions without the space-time-dependent cos term which, as we have seen, cancels out, being the same across the boundary (Eqs. (10.116–10.118). We therefore, get

$$E_i \frac{\left[\vec{k}_i \times \hat{\varepsilon}_i(\vec{k}_i)\right]_z}{\omega} + E_r \frac{\left[\vec{k}_r \times \hat{\varepsilon}_r(\vec{k}_r)\right]_z}{\omega} = E_t \frac{\left[\vec{k}_t \times \hat{\varepsilon}_t(\vec{k}_t)\right]_z}{\omega} \tag{10.124}$$

and

$$\varepsilon_1 E_i \left[\hat{\varepsilon}_i(\vec{k}_i)\right]_z + \varepsilon_1 E_r \left[\hat{\varepsilon}_r(\vec{k}_r)\right]_z = \varepsilon_2 E_t \left[\hat{\varepsilon}_t(\vec{k}_t)\right]_z \tag{10.125}$$

$$\frac{E_i}{\mu_1}\left[\vec{k}_i \times \hat{\varepsilon}_i(\vec{k}_i)\right]_{x,y} + \frac{E_r}{\mu_1}\left[\vec{k}_r \times \hat{\varepsilon}_r(\vec{k}_r)\right]_{x,y} = \frac{E_t}{\mu_2}\left[\vec{k}_t \times \hat{\varepsilon}_t(\vec{k}_t)\right]_{x,y} \tag{10.126}$$

and

$$E_i \left[\hat{\varepsilon}_i(\vec{k}_i)\right]_{x,y} + E_r \left[\hat{\varepsilon}_r(\vec{k}_r)\right]_{x,y} = E_t \left[\hat{\varepsilon}_t(\vec{k}_t)\right]_{x,y} \tag{10.127}$$

Equations (10.124–10.127) look complicated and hard to solve for the relative amplitudes. They can in fact be solved independently for the two independent directions of the incident electric field.

PROBLEM 10.8 A uniform plane wave in air is incident normally on a dielectric medium with $\varepsilon = 3\varepsilon_0$ and $\mu = \mu_0$. The incident wave is given by

$$\vec{E}_i = 20\cos(wt - z)\hat{j} \text{ V/m}$$

Find the wavelength of the incident and transmitted wave as well as the incident magnetic field.

PROBLEM 10.9 A 30 MHz uniform plane wave is given by

$$\vec{H} = 10\sin(wt + \beta x)\hat{k} \text{ mA/m}$$

It is traveling in a region ($x \geq 0$) with $\varepsilon = 10\varepsilon_0$ and $\mu = 5\mu_0$. At $x = 0$, there is an interface with free space. Calculate the reflected and transmitted magnetic fields.

10.8.1 Polarisation of E.M Waves

The polarisation direction of the incident wave is given by the vector $\hat{\varepsilon}_i(\vec{k}_i)$ which, as we have seen, is orthogonal to $\vec{k}_i$. We can choose two independent $\hat{\varepsilon}_i(\vec{k}_i)$ such that any arbitrary $\varepsilon_i(\vec{k})$ is a linear combination of these two. A convenient choice of the two polarisations is to take one of them in the plane of incidence and the other perpendicular to it as in Fig. 10.11.

From the Figures it is clear that

$$\vec{k}_i = k_i\,[\sin\theta_i\hat{y} + \cos\theta_i\hat{z}]$$
$$\vec{k}_t = k_t\,[\sin\theta_t\hat{y} + \cos\theta_t\hat{z}]$$
$$\vec{k}_r = k_r\,[\sin\theta_r\hat{y} - \cos\theta_r\hat{z}]$$

Choosing two independent polarisations in this way has the advantage that in both cases, the electric fields in both the transmitted and reflected waves are also in the same plane, that is either in the plane of incidence or perpendicular to it. To see this, consider first the case where the electric field is in the

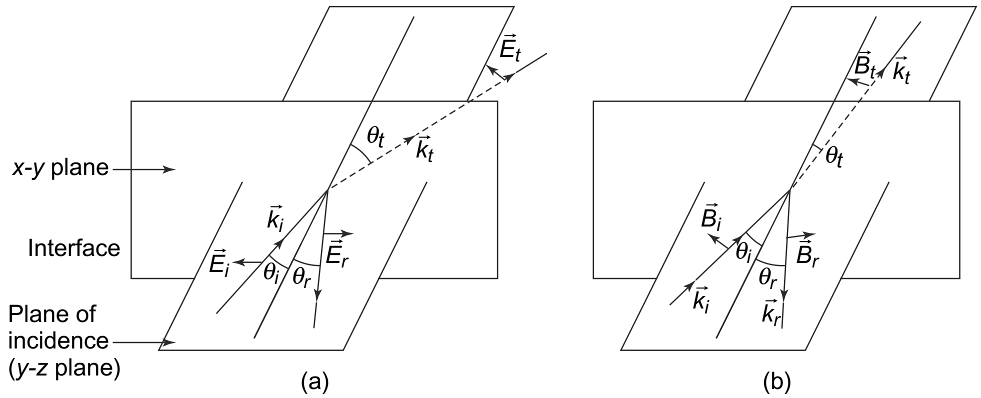


Fig. 10.11 *Two independent directions of polarisations of an EM wave (a) All electric fields are in the plane of incidence, (b) All magnetic fields are in the plane of incidence*

plane of incidence (Fig. 10.11(a)). The magnetic field for the incident wave in this case would be normal to the plane of incidence, i.e., in the x-direction. Now suppose, in this case, E_t and E_r were to have non-vanishing x- components. That is

$$\vec{E}_t = E_t\hat{x} \tag{10.128}$$

and

$$\vec{E}_r = E_r\hat{x} \tag{10.129}$$

The magnetic fields for the transmitted and reflected wave will then be in the plane of incidence and will be given by

$$\vec{B}_t = \frac{(\vec{k}_t \times \vec{E}_t)}{\omega} = \frac{E_t k_t}{\omega}(\cos\theta_t\,\hat{y} + \sin\theta_t\hat{z}) \tag{10.130}$$

and

$$\vec{B}_r = \frac{(\vec{k}_r \times \vec{E}_r)}{\omega} = \frac{E_r k_r}{\omega}(-\cos\theta_r\,\hat{y} + \sin\theta_r\hat{z}) \tag{10.131}$$

Recall that in this case we are considering, $\vec{E}_i$ does not have any x-component and hence $\vec{B}_i$ does not have any y or z components. Now if we apply the boundary conditions on the x component of the electric field and the y component of the magnetic field, (Eqs. (10.124–10.127), we get

$$E_t = E_r \tag{10.132}$$

and

$$\frac{E_t k_t \cos\theta_t}{\omega\mu_2} = -\frac{E_r k_r \cos\theta_r}{\omega\mu_1} \tag{10.133}$$

For non-vanishing E_t and E_r, these two equations, Eqs. (10.132) and (10.133) demand that

$$\frac{k_t \cos\theta_t}{\mu_2} = -\frac{k_r \cos\theta_r}{\mu_1}$$

This is not possible since the two terms have opposite signs. Hence, we can conclude that $E_t = E_r = 0$. This implies that if the incident electric field is in the plane of incidence, so will the electric fields in the transmitted and reflected waves.

Next, consider the case where the electric field of the incident wave is perpendicular to the plane of incidence. In this case, the magnetic field of the incident wave will be in the plane of incidence, Fig. 10.11(b). We now repeat the above arguments, but this time for the magnetic fields in the transmitted and reflected waves, which we assume to have non-vanishing x components.

$$\vec{B}_t = B_t \hat{x}$$
$$\vec{B}_r = B_r \hat{x}$$

The electric fields for the transmitted and reflected waves are therefore,

$$\vec{E}_t = \frac{(\vec{k}_t \times \vec{B}_t)}{\omega\varepsilon_2} = -\frac{(B_t k_t)}{\omega\varepsilon_2}(\cos\theta_t \hat{y} + \sin\theta_t \hat{z}) \tag{10.134}$$

and

$$\vec{E}_r = \frac{(\vec{k}_r \times \vec{B}_r)}{\omega\varepsilon_1} = -\frac{(B_r k_r)}{\omega\varepsilon_1}(-\cos\theta_r \hat{y} + \sin\theta_r \hat{z}) \tag{10.135}$$

since

$$\vec{k}_t = k_t(\sin\theta_t \hat{y} + \cos\theta_t \hat{z})$$
$$\vec{k}_r = k_r(\sin\theta_r \hat{y} - \cos\theta_r \hat{z})$$

The boundary condition demanding the continuity of the normal component of $\vec{D}$ across the interface then gives us

$$\frac{B_t k_t}{\omega}\sin\theta_t = \frac{B_r k_r}{\omega}\sin\theta_r \tag{10.136}$$

which reduces to

$$B_t = B_r \tag{10.137}$$

since

$$\frac{k_t}{\omega} = \frac{1}{c_2}$$
$$\frac{k_r}{\omega} = \frac{1}{c_1}$$

and

$$\frac{c_1}{c_2} = \frac{\sin\theta_i}{\sin\theta_t} = \frac{\sin\theta_r}{\sin\theta_t} = n_{21}$$

from Snell's Law and the law of reflection.

Continuity of the y-component of the electric field gives us

$$\frac{B_t\cos\theta_t}{c_2\varepsilon_2} = -\frac{B_r\cos\theta_r}{c_1\varepsilon_1} \tag{10.138}$$

Once again, if B_t and B_r are non-vanishing and equal (Eq. (10.137), then the two sides of Eq. (10.138) have different signs and cannot be equal. Thus $B_t = B_r = 0$.

But from Eqs. (10.134,10.135), $\vec{E}_t$ and $\vec{E}_r$ therefore, cannot have any y or z components. We thus see that in both cases, if the electric field of the incident wave is in the plane of incidence (or perpendicular to it), the electric fields of the reflected and transmitted wave is also in the plane of incidence (or perpendicular to it). We now analyse these two independent polarisation directions separately.

10.8.2 Reflection and Transmission Coefficient: Incident Wave Polarised in the Plane of Incidence

Let us first consider the case of the incident wave having its electric field in the plane of incidence, Fig. 10.11(a). Then

$$\begin{aligned}\hat{\varepsilon}_i &= \cos\theta_i\,\hat{y} - \sin\theta_i\,\hat{z}\\ \hat{\varepsilon}_t &= \cos\theta_t\,\hat{y} - \sin\theta_t\,\hat{z}\\ \hat{\varepsilon}_r &= \cos\theta_r\,\hat{y} + \sin\theta_r\,\hat{z}\end{aligned} \tag{10.139}$$

Using these expressions for the unit vectors in Eq. (10.126) and taking the y component, we get

$$\frac{k_i(E_i - E_r)}{\mu_1} = \frac{k_t E_t}{\mu_2} \tag{10.140}$$

From Eq. (10.127), we get

$$E_i\cos\theta_i + E_r\cos\theta_r = E_t\cos\theta_t \tag{10.141}$$

We can solve these equations for the ratios $\frac{E_r}{E_i}$ and $\frac{E_t}{E_i}$ to get

$$\boxed{\begin{aligned}\left(\frac{E_r}{E_i}\right)_{\parallel} &= \frac{\frac{\cos\theta_t}{\mu_1} - \frac{n_{21}}{\mu_2}\cos\theta_i}{\frac{\cos\theta_t}{\mu_1} + \frac{n_{21}}{\mu_2}\cos\theta_i}\\ \left(\frac{E_t}{E_i}\right)_{\parallel} &= \frac{2\frac{\cos\theta_i}{\mu_1}}{\frac{\cos\theta_t}{\mu_1} + \frac{n_{21}}{\mu_2}\cos\theta_i}\end{aligned}} \tag{10.142}$$

We have used the fact that

$$\frac{k_i}{k_t} = \frac{c_2}{c_1} = \frac{1}{n_{21}}$$

where n_{21} is the refractive index of medium 2 with respect to medium 1. We have also indicated a subscript $\parallel$ with the ratios of amplitudes to remind ourselves that this is the case when the incident electric field is in the plane of incidence.

10.8.3 Reflection and Transmission Coefficient: Incident Wave Polarised Perpendicular to the Plane of Incidence

Let us first consider the case of the incident wave having its electric field perpendicular to the plane of incidence, Fig. 10.11(b). Then, as we saw above, the electric field vector of all the three, the incident, reflected and transmitted waves will be in the x direction. Thus,

$$\hat{\varepsilon}_i = \hat{x}$$

$$\hat{\varepsilon}_t = \hat{x}$$
$$\hat{\varepsilon}_r = \hat{x}$$

Now using Eq. (10.126)

$$E_i + E_r = E_t \tag{10.143}$$

and from Eq. (10.125), we get

$$\frac{k_i E_i \cos\theta_i}{\mu_1} - \frac{k_r E_r \cos\theta_r}{\mu_1} = \frac{k_t E_t \cos\theta_t}{\mu_2} \tag{10.144}$$

Once again, we can easily solve these for the rations $\frac{E_r}{E_i}$ and $\frac{E_t}{E_i}$ to get

$$\boxed{\begin{aligned} \left(\frac{E_r}{E_i}\right)_\perp &= \frac{\frac{\cos\theta_i}{\mu_1} - \frac{n_{21}\cos\theta_t}{\mu_2}}{\frac{\cos\theta_i}{\mu_1} + \frac{n_{21}\cos\theta_t}{\mu_2}} \\ \left(\frac{E_t}{E_i}\right)_\perp &= \frac{\frac{2\cos\theta_i}{\mu_1}}{\frac{\cos\theta_i}{\mu_1} + \frac{n_{21}\cos\theta_t}{\mu_2}} \end{aligned}} \tag{10.145}$$

Here the subscript $\perp$ indicates that the electric fields are perpendicular or normal to the plane of incidence. Equations (10.145 and 10.142) are called **Fresnel's equations**.

The intensities of the transmitted and reflected waves for a given incident wave can be easily calculated using Eqs. (10.142 and 10.145) and Eq. (10.143). The Poynting vectors for the incident, reflected and transmitted waves will give us the rate of flow of energy. These can be computed from the above expressions and we get

$$|\vec{P}_i| = c_1\varepsilon_1|\vec{E}_i|^2 \cos^2(\vec{k}_i \cdot \vec{r} - \omega t + \delta) \tag{10.146}$$
$$|\vec{P}_r| = c_1\varepsilon_1|\vec{E}_r|^2 \cos^2(\vec{k}_r \cdot \vec{r} - \omega t + \delta) \tag{10.147}$$
$$|\vec{P}_t| = c_2\varepsilon_2|\vec{E}_t|^2 \cos^2(\vec{k}_t \cdot \vec{r} - \omega t + \delta) \tag{10.148}$$

The time average values of the Poynting vectors, $< P >$ are given by replacing the $\cos^2$ factors by $\frac{1}{2}$. Thus

$$< |\vec{P}_i| >= \frac{1}{2} c_1\varepsilon_1 |\vec{E}_i|^2$$
$$< |\vec{P}_i| >= \frac{1}{2} c_1\varepsilon_1 |\vec{E}_r|^2$$
$$< |\vec{P}_i| >= \frac{1}{2} c_2\varepsilon_2 |\vec{E}_t|^2$$

The P's represent energy flow per unit time across unit area normal to the direction of propagation which of course are different for three waves. If we consider energy flow normal to the interface going from medium 1 to medium 2, the time averaged energy flow per unit time per unit area for the incident, transmitted and reflected waves respectively are $\cos(\theta_i) < P_i >, \cos(\theta_t) < P_t >$ and $-\cos(\theta_r) < P_r >$.

The ratios

$$\mathbf{R} \equiv \frac{\cos(\theta_r) < |\vec{P}_r| >}{\cos(\theta_i) < |\vec{P}_i| >}$$

and

$$\mathbf{T} \equiv \frac{\cos(\theta_t) < |\vec{P}_t| >}{\cos(\theta_i) < |\vec{P}_i| >}$$

are called the **coefficients of reflection and transmission** respectively.

For the case when the electric vector lies in the plane of incidence, using Eq. (10.142)

$$\mathbf{R} = \left[\frac{r_1 - r_2}{r_1 + r_2}\right]^2 \tag{10.149}$$

$$\mathbf{T} = \left[\frac{4r_1 r_2}{(r_1 + r_2)^2}\right] \tag{10.150}$$

where

$$r_1 = \frac{\cos\theta_t}{\cos\theta_i}$$

and

$$r_2 = \frac{\mu_1 c_1}{\mu_2 c_2}$$

It is easy to check that

$$\mathbf{R} + \mathbf{T} = 1$$

We show in Fig. 10.12, a typical variation of **R** and **T** as a function of the incident angle for the air-glass interface. Here $\mu_1 = \mu_0, \varepsilon_1 = \varepsilon_0, n_{21} = \frac{\sqrt{\mu_2 \varepsilon_2}}{c} = 1.5$. In Fig. 10.12, the electric field is in the plane of incidence.

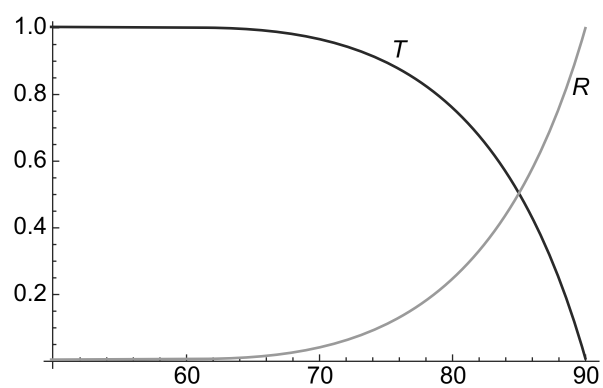


Fig. 10.12 *Plot of Reflection (R) and Transmission (T) coefficients for air-glass interface with refractive index of glass taken as 1.5. The x-axis is the angle of incidence in degrees, for the case of electric field in the plane of incidence. R vanishes at the Brewster angle, about 58 degrees in this case. The sum of R and T is 1. Below 50 degrees T stays virtually at 1.0 and R very small*

The reflection and transmission coefficients for the case when the electric vector is normal to the plane of incidence can similarly be worked out using Eq. (10.145) and this is left as an exercise.

EXAMPLE 10.5 A plane polarised beam in air is incident in a plane interface with a dielectric of refractive index $\sqrt{2}$ and relative permeability $\mu_r = 1$. The incident beam has an angle of incidence of 45° and is plane polarised at an angle of 45° from the normal to the plane of incidence. Determine if the the reflected or transmitted beams suffer any phase change at the interface. Also determine the angle between the plane of polarisation of the reflected and the incident beam.

Solution

The angle of incidence is 45°and since the refractive index is $\sqrt{2}$, the angle of refraction is 30°. Further, the incident beam is plane polarised which implies that the normal and planar components of the amplitude of the incident beam are equal and in phase.

The reflection coefficient of the normal component is

$$\frac{\frac{1}{\sqrt{2}} - \sqrt{\frac{3}{2}}}{\frac{1}{\sqrt{2}} + \sqrt{\frac{3}{2}}} = -0.26$$

The reflection coefficient of the planar component is

$$\frac{-1 + \frac{\sqrt{3}}{2}}{1 + \frac{\sqrt{3}}{2}} = -0.07$$

Thus, both the normal and planar reflected beam suffer a phase change of 180° since the reflection coefficients for both of them is negative. However, the amplitudes of the normal and planar reflected beams are no longer equal. The plane of polarisation therefore, is at an angle θ with respect to the normal such that

$$\tan\theta = \frac{0.07}{0.26}$$

In the incident beam the plane of polarisation was at an angle of 45°from the normal. Hence, the angle between the incident and reflected planes of polarisation is

$$\frac{\pi}{4} - \theta$$

10.8.4 Brewster's Angle

From the discussion above, it is clear that the amplitudes of the reflected and transmitted waves depend on the angle of incidence θ_i. In particular, for the case where the incident electric fields are in the plane of incidence, we get a curious effect. Consider the expression for $\frac{E_r}{E_i}$ from Eq. (10.142). Here we see that there exists an incident angle where the ratio vanishes, or, the reflected amplitude goes to zero. This happens when

$$\frac{n_{21}}{\mu_2}\cos\theta_i = \frac{1}{\mu_1}\cos\theta_t \tag{10.151}$$

But θ_i and θ_t are related by Snell's Law of refraction by

$$\frac{\sin\theta_i}{\sin\theta_t} = n_{21}$$

and so we get

$$\frac{n_{21}}{\mu_2}\cos\theta_i = \frac{1}{\mu_1}\sqrt{\frac{n_{21}^2 - \sin^2\theta_i}{n_{21}^2}} \tag{10.152}$$

This equation can be solved to get the value of θ_i for which the reflected amplitude is zero or there is no reflected wave. The angle θ_i for which this happens is called **Brewster's angle**. For the case where the magnetic permeability on both sides of the interface is the same, $\mu_2 = \mu_1 = \mu$, the expression simplifies and we get

$$\tan\theta_i = n_{21}$$

For such a θ_i,

$$\tan\theta_t = \frac{1}{n_{21}}$$

as can be seen from Snell's Law. Thus, $\theta_i + \theta_t = \frac{\pi}{2}$ or the angle of incidence and refraction (transmission) are complementary, which can be interpreted geometrically in this case.

SUMMARY

- Ampere's Law is inconsistent when we consider time dependent currents and charge densities.
- To make Ampere's Law consistent, we need to introduce a displacement current term which is proportional to the rate of change of electric field.
- With the modification to Ampere's Law, we can write the complete set of four equations, called Maxwell's equations, which together with the equation of continuity, contain all the information about the electric and magnetic fields and their relation to charge and current densities.
- The energy density of electric and magnetic fields also is modified because of modification of Ampere's Law and also Faraday's Law. Rate of flow of energy through a surface is related to a quantity called Poynting vector, which is given by the cross-product of electric and magnetic fields.
- Maxwell's equations in free space or in polarisable media imply the existence of electromagnetic waves. Electromagnetic waves are transverse waves. They are disturbances of the electric and magnetic field. The electric and magnetic fields are orthogonal to each other and each is also orthogonal to the direction of propagation.
- The speed of electromagnetic waves depends on the medium in which they travel through the electric permittivity and magnetic permeability. In free space, the speed is the speed of light. In a medium, it is less than the speed of light. The ratio of the speed of electromagnetic waves in a medium and free space is called the refractive index of the medium. It is related to the relative permeability and permittivity of the medium.

- The energy density of the electric and magnetic fields in an electromagnetic wave are equal.
- The frequency dependence of electrical permittivity of media leads to the phenomenon of dispersion where the refractive index and hence the speed of the electromagnetic wave in the medium is different for different frequencies.
- The laws of reflection and refraction of electromagnetic waves at the interface of two media follow naturally from Maxwell's equations and the boundary conditions on the electric and magnetic field vectors at the interface.

CONCEPTUAL QUESTIONS

1. The electric field of an electromagnetic wave is given by

$$E_x = E_z = 0$$
$$E_y = 30\cos\left((2\pi)\times 10^8 t - \frac{2\pi}{3}x\right)$$

Here E is in Volts/m and x in m. Find the frequency of the wave and the direction of the magnetic field.

2. The electric field intensity of a uniform plane wave in free space is given by $\vec{E} = E_0\cos(\omega t + 6z)\hat{z}$ V/m. Find the wavelength and the magnetic field intensity.

3. The electric field intensity of a wave in free space is given by $\vec{E} = 4\sin((2\pi)\times 10^7 t - 0.8x)\hat{z}$ V/m. Find the time average power carried by the wave.

4. Which of the following functions do NOT satisfy the wave equation?
a. $\cos^2(y + 5t)$
b. $\sin\omega(10z + 5t)$
c. $\cos(5y + 2x)$
d. $\sin x \cos t$

5. The electric field of a wave in free space is given by

$$\vec{E} = 10\cos(10^7 t + kz)\hat{y} \quad \text{V/m}$$

Which of the following statements are true?
a. The wave propagates along $\hat{y}$.
b. The wavelength of the wave is 188.5 m.
c. The amplitude of the wave is 10 V/m.
d. The wave number is 0.33 radians/m.

6. A polarised wave is incident from air to a medium with $\varepsilon = 2.6\varepsilon_0$ and $\mu = \mu_0$ at Brewster's angle. Find the transmission angle.

7. A light bulb puts out 100 W of radiation. Assuming a plane wave, calculate the time-average intensity of radiation from this light bulb at a distance of one meter and the maximum values of electric and magnetic fields at this same distance from the bulb?

8. A uniform plane wave in air with

$$\vec{E} = 8\cos(\omega t - 4x - 3z)\hat{y} \quad \text{V/m}$$

is incident on a dielectric slab ($z \geq 0$) with $\mu = \mu_0$ and $\varepsilon = 2.5\varepsilon_0$. Find the polarisation of the incident wave.

9. The electric field of a plane wave is given by

$$\vec{E}(z,t) = \hat{y}E_0 \sin(kz + \omega t)$$

The magnetic field is given by

a. $\vec{B}(z,t) = \hat{x}B_0 \sin(kz + \omega t)$
b. $\vec{B}(z,t) = -\hat{x}B_0 \sin(kz + \omega t)$
c. $\vec{B}(z,t) = \hat{z}B_0 \sin(kz + \omega t)$
d. $\vec{B}(z,t) = -\hat{z}B_0 \sin(kz + \omega t)$

10. A parallel plate capacitor has circular plates with radius r distance d apart. What is the displacement current when the rate of charge accumulation on the plates is $\frac{Q}{t}$?

PROBLEMS

1. A parallel plate capacitor has a plate area 5 cm^2 and a plate separation of 3 mm. A voltage of $50 \sin 10^3 t$ V is applied to the plates. Calculate the displacement current assuming that $\varepsilon = 2\varepsilon_0$ for the dielectric between the plates.

2. Compare the rms values of $\vec{j}$ and $\frac{\partial \vec{D}}{\partial t}$ in seawater for frequencies f of

a. 60 Hz
b. 10 kHz
c. 100 MHz or
d. 10 GHz

Assume that for seawater the permittivity is $10\varepsilon_0$ and the conductivity is 5 in SI units.

3. Find the magnitude and direction of the Poynting vector at the surface of a long straight wire of resistance per unit length R and carrying a direct current I. The wire has a circular cross section with a radius b.

4. The electric field of an electromagnetic wave is given by the superposition of two waves

$$\vec{E} = E_0 \cos(kz - \omega t)\hat{i} + E_0 \cos(kz + \omega t)\hat{i}$$

a. What is the associated magnetic field $\vec{B}(x, y, z, t)$.
b. What is the energy per unit area per unit time (the Poynting vector S) transported by this wave?
c. What is the time average of the Poynting vector? Briefly explain your answer.

5. A parallel plate capacitor with circular plates of radius a and the distance between the plates h is being charged. Show that the rate at which the energy flows into the capacitor is equal to the rate at which the stored electric energy increases.

6. Calculate the Poynting vector at the surface of the sun if the power radiated by the Sun is 3.8×10^{26} watts while its radius is 7×10^8 m. If the average distance between the Sun and the earth is 1.5×10^{11} m calculate the amplitude of the electric and magnetic field vectors on the surface of the earth. Assume the radiation is a plane wave.

7. An ideal capacitor of capacitance C, is charged with a charge Q. What is the flux of the Poynting vector through a surface

a. fully enclosing the capacitor and

b. enclosing only one plate of the capacitor.

Assume that the charge $Q(t)$ is constant in time. Repeat the calculations for both cases if the charge in time as $Q(t) = Q_0 \cos(\omega t)$ with $\omega = 100$ Hz.

8. Show that the electric field $\vec{E} = 100 \sin(10^8 t + \frac{x}{\sqrt{3}})$ is a valid solution to the wave equation in a dielectric medium. Find the dielectric constant (assuming the permeability is μ_0) and the velocity of propagation.

9. An electromagnetic wave travels along the z direction in a dielectric medium with $\varepsilon = 9\varepsilon_0$. It strikes another medium with $\varepsilon = 4\varepsilon_0$ at $z = 0$. If the maximum value of the electric field of the incoming wave is 0.1 V/m at the interface and the angular frequency 300×10^6 radians/s, determine the reflection and transmission coefficients. Also determine the power densities of the incident, transmitted and reflected waves.

10. Show that

$$E_x(z, t) = F_x(t + \sqrt{\mu\varepsilon} z)$$

satisfies the wave equation in a dielectric medium. Find the vector $\vec{H}$ and the Poynting vector.

11. The electric field of an plane, monochromatic electromagnetic wave travelling in a medium is

$$\vec{E} = E_0 \cos(\beta z - \omega t)\hat{x} + E_0 \sin(\beta z - \omega t)\hat{y}$$

where $\beta = \omega\sqrt{\mu\varepsilon}$ and E_0 is a constant. Find the corresponding $\vec{H}$ and the Poynting vector.

12. In free space

$$\vec{H} = \rho(\sin\phi\hat{\rho} + 2\cos\phi\hat{\phi})\cos(4 \times 10^6 t)$$

Find the displacement current $\vec{J}_D$ and the electric field $\vec{E}$.

13. In free space

$$\vec{H} = 0.2\cos(\omega t - \beta x)\hat{z}$$

Find the total power passing through a square plate of side 10 cm on the plane $z + x = 1$.

14. A linearly polarised plane wave travelling in air is incident on a planar interface with a dielectric on the other side. The electric vector of the incident wave makes an angle ϕ with respect to the plane of incidence. Determine the various states of polarisation of the reflected and transmitted waves as ϕ is varied between 0 and $\frac{\pi}{2}$.

15. A plane polarised beam is incident on a air-water interface. The angle of incidence of the wave is $\frac{\pi}{4}$ and the incident electric field makes an angle of $\frac{\pi}{4}$ with the plane of incidence. A circular loop of unit area is held, on the water side of the interface with its normal aligned along the normal to the interface such that the refracted beam passes through the loop. Given that for water $\mu = \mu_0$ and $\varepsilon = 80\varepsilon_0$ and that the electric field of the incident beam has an rms value of 100 mV/m, calculate the EMF generated in the loop.

16. The values of the constants A and B in Eq. (10.86) for crystalline quartz are approximately: $A = 0.53$ and $B = 7000\ \text{nm}^2$. A wave is incident on air-quartz interface at an angle of incidence $\frac{\pi}{4}$. Its electric vector is in the plane of incidence and has the value

$$\vec{E}(\vec{r}, t) = \vec{E}_0 \sin^3(\omega t - k\vec{r} \cdot \hat{n})$$

with $\omega = ck$, $k = \frac{2\pi}{\lambda}$, $\lambda = 700$ nm. Determine the nature and angles of refraction for such a wave.

17. A beam of light in air is incident in a plane glass slab of refractive index 1.5 and of $\mu = \mu_0$. The incident light is plane polarised with its electric vector making an angle of $\frac{\pi}{3}$ with the plane of incidence. Determine whether the reflected and transmitted beams will be either elliptically polarised or plane polarised. If the light is elliptically polarised determine the angle that the major axis of the electric field makes with the plane of incidence. If the light is plane polarised, determine the angle the electric field makes with the plane of incidence.

18. A uniform plane wave in air has

$$\vec{E} = 8\cos(\omega t - 4x - 3z)\hat{y}$$

This wave is incident on a dielectric slab ($z \geq 0$) with μ_0, $\varepsilon = 2.5\varepsilon_0$. Find the

a. polarisation of the incident wave
b. angle of incidence
c. reflected electric field
d. transmitted $\vec{H}$ field

19. A circularly polarised electromagnetic wave, with the amplitudes of both the transverse components of the electric field 10 mV/m propagating in vacuum is incident on a plane surface that absorbs the wave completely. The angle of incidence of the wave is $\frac{\pi}{4}$. Calculate the amount of energy received by the surface in a day.

20. In a dielectric medium with $\varepsilon = 9\varepsilon_0$ and $\mu = \mu_0$, plane wave with

$$\vec{H} = 0.2\cos(10^9 t - kx - \sqrt{8}kz)\hat{y}$$

is incident on an air boundary at $z = 0$. Find the Brewster angle.

11

Applications of Maxwell's Equations

Learning Objectives

- To write down Maxwell's equations in presence of conductors and note that the charge density decays exponentially with time.
- To find the propagating wave solutions of Maxwell's equations and see that the fields are transverse to the direction of propagation.
- To comprehend how attenuation of the EM waves occurs in conductors and learn about skin depth.
- To learn about EM waves in conducting pipes known as wave guides.
- To determine the fields that exist in an infinite, hollow conducting wave guide.
- To learn about the existence of the various modes which exist for a waveguide.
- To understand the wave propagation in a rectangular hollow wave guide and energy transmission in such a wave guide.
- To comprehend the concept of electromagnetic potentials and the gauge freedom to choose a convenient gauge for the potentials.
- To learn about the potential and field due to a charge in motion.
- To determine the fields due to an accelerated charge, an oscillating dipole and the presence of EM radiation in such cases.

In the previous Chapter, we had considered solutions of Maxwell's equations for cases where there are no free charges or currents. In this Chapter, we consider a few cases where currents and/or charges are present in some form and try to see what the solutions of Maxwell's equations look like.

11.1 MAXWELL'S EQUATIONS IN THE PRESENCE OF CONDUCTORS

Whenever we have conductors present with an electric field $\vec{E}$, the assumption of no currents is not true. This is obvious from Ohm's Law

$$\vec{j} = \sigma \vec{E} \tag{11.1}$$

where $\vec{j}$ is the current density and σ is the conductivity of the conductor. We see that there is always a $\vec{j}$ whenever we have a $\vec{E}$ in a conductor. Maxwell's equations for this case can be written down as

$$\boxed{\vec{\nabla} \cdot \vec{D} = \rho} \tag{11.2}$$

$$\vec{\nabla} \cdot \vec{B} = 0 \tag{11.3}$$
$$\vec{\nabla} \times \vec{E} = -\frac{\partial \vec{B}}{\partial t} \tag{11.4}$$
$$\vec{\nabla} \times \vec{H} = \vec{j} + \frac{\partial \vec{D}}{\partial t} \tag{11.5}$$

We can express the last equation in terms of $\vec{E}$ since $\vec{D} = \varepsilon \vec{E}$ as

$$\vec{\nabla} \times \vec{H} = \sigma \vec{E} + \varepsilon \frac{\partial \vec{E}}{\partial t} \tag{11.6}$$

The first thing we see in a conductor is that free charge densities cannot be sustained. This can be easily seen if we write the equation of continuity

$$\vec{\nabla} \cdot \vec{j} = -\frac{\partial \rho}{\partial t} \tag{11.7}$$

Combining this with Ohm's Law Eq. (11.1) and Equation for the divergence of $\vec{D}$ above, we get

$$\sigma \vec{\nabla} \cdot \vec{E} + \frac{\partial \rho}{\partial t} = \sigma \frac{\rho}{\varepsilon} + \frac{\partial \rho}{\partial t} = 0 \tag{11.8}$$

This is a differential equation for ρ and has a simple solution

$$\rho(\vec{r}, t) = \rho(\vec{r}, 0) e^{-\frac{\sigma t}{\varepsilon}} \tag{11.9}$$

Thus, we see that the charge density ρ decays with time with a characteristic time which depends on σ and ε. Under steady state conditions we can therefore, assume that $\rho = 0$. This implies that out of the four Maxwell's equations, all the equations except that for divergence of $\vec{D}$ are the same as for a dielectric medium. For the divergence of $\vec{D}$, in a dielectric, we have a non-zero ρ while for a conductor the charge density, as we saw above, vanishes in steady state.

11.2 PROPAGATING WAVE SOLUTIONS

Consider an infinite, homogenous conducting medium where the electromagnetic wave propagates along the z-direction. The field variables are assumed to have the form

$$\begin{aligned} \vec{E} &= \vec{E}(z, t) \\ \vec{B} &= \vec{B}(z, t) \end{aligned} \tag{11.10}$$

We will try to find monochromatic solutions to the wave equation in the same way as we did for the dielectric case in Chapter 10.

We have seen above that the charge density ρ can be set to zero in the steady state (Eq. (11.9)). Thus, we have, from Eq. (11.2)

$$\vec{\nabla} \cdot \vec{E} = 0 \tag{11.11}$$

Now since $\vec{E}$ is a function of z alone and not of x, y (Eq. (11.10), this also implies

$$\frac{\partial E_z}{\partial z} = 0$$

Thus, E_z cannot depend on z. However, we must show that even when $\rho = 0$, the constant value of E_z cannot be time dependent. To see this, we take the curl of Eq. (11.4), and use Eqs. (11.5) and (11.7)

$$\begin{aligned}\vec{\nabla} \times \vec{\nabla} \times \vec{E} &= \vec{\nabla}(\vec{\nabla} \cdot \vec{E}) - \nabla^2 \vec{E} \\ &= -\nabla^2 \vec{E} \end{aligned} \tag{11.12}$$

since $\vec{\nabla} \cdot \vec{E} = 0$. But

$$\vec{\nabla} \times \vec{\nabla} \times \vec{E} = \vec{\nabla} \times \left[-\frac{\partial \vec{B}}{\partial t} \right]$$

and hence we get

$$\begin{aligned}\vec{\nabla} \times \vec{\nabla} \times \vec{E} &= -\nabla^2 \vec{E} \\ &= -\vec{\nabla} \times \left[\frac{\partial \vec{B}}{\partial t} \right] \\ &= -\mu \vec{\nabla} \times \left[\frac{\partial \vec{H}}{\partial t} \right] \\ &= -\mu \left[\frac{\partial^2 \varepsilon \vec{E}}{\partial t^2} + \sigma \frac{\partial \vec{E}}{\partial t} \right] \end{aligned} \tag{11.13}$$

where we have used

$$\begin{aligned}\vec{B} &= \mu \vec{H} \\ \vec{D} &= \varepsilon \vec{E} \\ \vec{\nabla} \times \vec{H} &= \vec{j} + \frac{\partial \vec{D}}{\partial t}\end{aligned}$$

and

$$\vec{j} = \sigma \vec{E}$$

If we now take the z component of Eq. (11.13), since E_z does not depend on z, the left-hand side vanishes. Thus, we get

$$0 = \mu\varepsilon \frac{\partial^2 E_z}{\partial t^2} + \mu\sigma \frac{\partial E_z}{\partial t} \tag{11.14}$$

This equation can be easily solved and we get the solution

$$\frac{\partial E_z}{\partial t} = \left[\frac{\partial E_z}{\partial t} \right]_{t=0} e^{-\frac{\sigma t}{\varepsilon}} \tag{11.15}$$

and therefore, $E_z(t)$ can be found to be

$$E_z(t) = \frac{\partial E_z}{\partial t} \frac{1}{-\sigma/\varepsilon} + \text{constant} \tag{11.16}$$

We see that E_z cannot have any time dependence at $t \to \infty$ because of the exponential factor. It can at best be a constant. Since these kind of constant fields have no role to play in wave propagation, we can set them to be zero. Thus, in our analysis, we set $E_z = 0$. A similar analysis this time with $\vec{\nabla} \times \vec{H}$ can be done to show that we can, without any loss of generality, set $H_z = 0$.

Recall that we have taken the wave to be propagating in the z direction and have shown that for wave propagation, the z- component, which is the longitudinal component along the direction of propagations plays no role. We need to focus on the role of the transverse components of the field vectors $\vec{E}$ and $\vec{H}$.

11.2.1 Propagating Wave Equation for Transverse Components

We shall now discuss the components of $\vec{E}$ and $\vec{H}$ transverse to the direction of propagation of the wave. We will continue to use the vector sign on them with the understanding that only the transverse components are non-vanishing. To obtain the wave equation satisfied by these components, we proceed with the Maxwell equations in the same fashion as we did for dielectric medium in Chapter 10. We take the equation for the curl of $\vec{E}$ and take the curl of the equation and use the equations for the curl of $\vec{B}$. Thus

$$\begin{aligned}\vec{\nabla} \times \vec{\nabla} \times \vec{E} &= \vec{\nabla}(\vec{\nabla} \cdot \vec{E}) - \nabla^2 \vec{E} \\ &= -\mu \vec{\nabla} \times \frac{\partial \vec{H}}{\partial t} \\ +\nabla^2 \vec{E} &= +\mu\sigma \frac{\partial \vec{E}}{\partial t} + \mu\varepsilon \frac{\partial^2 \vec{E}}{\partial t^2}\end{aligned} \tag{11.17}$$

Similarly, we get

$$\nabla^2 \vec{B} = \mu\sigma \frac{\partial \vec{B}}{\partial t} + \mu\varepsilon \frac{\partial^2 \vec{B}}{\partial t^2} \tag{11.18}$$

We consider now an infinite, homogenous conducting medium where the transverse components of $\vec{E}$ and $\vec{H}$ as well as ρ and $\vec{j}$ depend only on z and t. Thus,

$$\begin{aligned}\vec{E}(\vec{r}, t) &= \vec{E}(z, t) \\ \vec{B}(\vec{r}, t) &= \vec{B}(z, t)\end{aligned} \tag{11.19}$$

so that we can use

$$\vec{\nabla} = \frac{\partial}{\partial z} \hat{z}$$

Using this in Eq. (11.17), we get

$$\frac{\partial^2 \vec{E}(z, t)}{\partial z^2} - \mu\varepsilon \frac{\partial^2 \vec{E}(z, t)}{\partial t^2} - \mu\sigma \frac{\partial \vec{E}(z, t)}{\partial t} = 0 \tag{11.20}$$

Now writing

$$\vec{E}(z, t) = \vec{E}(k, t) e^{ikz}$$

and

$$\vec{B}(z, t) = \vec{B}(k, t) e^{ikz}$$

and substituting in Eq. (11.20), we get

$$-k^2\vec{E}(\vec{k},t) - \mu\varepsilon\frac{\partial^2\vec{E}(k,t)}{\partial t^2} - \mu\sigma\frac{\partial\vec{E}(k,t)}{\partial t} = 0 \tag{11.21}$$

and a similar equation for $\vec{B}$ as

$$-k^2\vec{B}(\vec{k},t) - \mu\varepsilon\frac{\partial^2\vec{B}(k,t)}{\partial t^2} - \mu\sigma\frac{\partial\vec{B}(k,t)}{\partial t} = 0 \tag{11.22}$$

Equations (11.21 and 11.22) have solutions

$$\vec{E}(\vec{k},t) = \vec{E}(\vec{k},0)e^{i\omega t} \tag{11.23}$$

and

$$\vec{B}(\vec{k},t) = \vec{B}(\vec{k},0)e^{i\omega t} \tag{11.24}$$

with

$$-k^2 + \mu\varepsilon\omega^2 + i\mu\sigma\omega = 0 \tag{11.25}$$

Since we are only interested in propagating solutions, ω is real and hence k is complex. We write

$$k = k_r + ik_i$$

with

$$k_r = \omega\sqrt{\mu\varepsilon}\,\mathrm{Re}\left[1 + \frac{i\sigma}{\omega\varepsilon}\right]^{1/2} \tag{11.26}$$

and

$$k_i = \omega\sqrt{\mu\varepsilon}\,\mathrm{Im}\left[1 + \frac{i\sigma}{\omega\varepsilon}\right]^{1/2} \tag{11.27}$$

Note that here k_i, k_r are the imaginary and real parts of k and NOT the wave numbers for incident and reflected waves as in Chapter 10. Since the propagating $\vec{E}$ and $\vec{B}$ fields have a e^{ikz} dependence, the presence of an imaginary part of k implies that there is an attenuation factor in the amplitude as the wave propagates along the z direction.

$$\vec{E}(z,t) = \vec{E}(\vec{k},0)e^{i(k_r z - \omega t)}e^{-k_i z} \tag{11.28}$$

and

$$\vec{B}(z,t) = \vec{B}(\vec{k},0)e^{i(k_r z - \omega t)}e^{-k_i z} \tag{11.29}$$

Recall that the electric and magnetic field vectors, $\vec{E}$ and $\vec{B}$ are both transverse to the direction of propagation since we have seen above that the longitudinal components can be taken to be zero. Using the form of the field vectors in Eqs. (11.28) and (11.29), we get

$$\vec{\nabla} \times \vec{E}(z,t) = (-ik_r + k_i)\left[E_y(z,t)\hat{x} - E_x(z,t)\hat{y}\right] \tag{11.30}$$

and

$$\vec{\nabla} \times \vec{B}(z,t) = (-ik_r + k_i)\left[B_y(z,t)\hat{x} - B_x(z,t)\hat{y}\right] \tag{11.31}$$

Now using

$$\vec{\nabla} \times \vec{E} = -\frac{\partial\vec{B}}{\partial t}$$

we can write

$$i\omega B_x(z,t) = (-ik_r + k_i)E_y(z,t) \tag{11.32}$$

and

$$i\omega B_y(z,t) = (ik_r - k_i)E_x(z,t) \tag{11.33}$$

Note that the ratios $\frac{B_x}{E_y}$ and $\frac{B_y}{E_x}$ are complex. This is a reflection of the phase difference between the two vectors. We can write

$$E_x(k,0) = |E_{kx}|e^{i\delta_x^E} \tag{11.34}$$

and

$$E_y(k,0) = |E_{ky}|e^{i\delta_y^E} \tag{11.35}$$

we get, using Eqs. (11.32 and 11.33)

$$B_x(k,0) = -\frac{\sqrt{k_r^2 + k_i^2}}{\omega}|E_{ky}|e^{i(\delta_y^E + \delta_k)} \tag{11.36}$$

and

$$B_y(k,0) = \frac{\sqrt{k_r^2 + k_i^2}}{\omega}|E_{kx}|e^{i(\delta_x^E - \delta_k)} \tag{11.37}$$

with

$$\tan\delta_k = \frac{k_i}{k_r} \tag{11.38}$$

The physical electric and magnetic fields, given as usual by the real parts of the complex quantities are thus out of phase as can be seen by taking the real parts of Eqs. (11.34–11.37).

$$\text{Re}\, E_x(z,t) = |E_{kx}|e^{-k_i z}\cos(k_r z - \omega t + \delta_x^E) \tag{11.39}$$

$$\text{Re}\, E_y(z,t) = |E_{ky}|e^{-k_i z}\cos(k_r z - \omega t + \delta_y^E) \tag{11.40}$$

$$\text{Re}\, B_x(z,t) = -\frac{\sqrt{k_r^2 + k_i^2}}{\omega}|E_{ky}|e^{-k_i z}\cos(k_r z - \omega t + \delta_y^E + \delta_k) \tag{11.41}$$

$$\text{Re}\, B_y(z,t) = \frac{\sqrt{k_r^2 + k_i^2}}{\omega}|E_{kx}|e^{-k_i z}\cos(k_r z - \omega t + \delta_x^E - \delta_k) \tag{11.42}$$

11.2.2 Reflection and Transmission at Dielectric-Conductor Interface

In the discussion above, we saw that the propagation of electromagnetic waves in a conducting medium is accompanied by an attenuation of the amplitude of the waves. One way to understand this is to think of the wave number k being complex

$$k = k_r + ik_i$$

with the real and imaginary parts as given in Eqs. (11.26 and 11.27). In this case, the usual relation between the wave number k, frequency ω and the electrical permittivity ε and the magnetic permeability μ which is valid in a dielectric, namely

$$k = \sqrt{\mu\varepsilon}\omega$$

is replaced by

$$k = \omega\sqrt{\mu}\sqrt{\varepsilon_c} \tag{11.43}$$

where ε_c is the complex permittivity given by

$$\varepsilon_c = \varepsilon + \frac{i\sigma}{\omega} \tag{11.44}$$

In the previous Chapter, we have considered the behaviour of electromagnetic waves at the interface of two dielectrics. We saw that there is a phenomenon of reflection and transmission or refraction in the case of two dielectrics with different permittivities. A similar analysis can be performed when one of the media is conducting and the interface thus is that between a dielectric and a conductor. Thus, for instance, if Region II is filled with a conducting medium, then the permittivity for that region, ε_2 needs to be taken as complex as

$$\varepsilon_{2c} = \varepsilon_2 + \frac{i\sigma}{\omega} \tag{11.45}$$

where σ is the conductivity of the medium and ω is the angular frequency of the incident wave at the interface.

Apart from taking the permittivity to be complex, while considering the behaviour of electromagnetic waves at a dielectric-conducting interface, we need to be careful of another aspect. The boundary conditions that we had taken for the dielectric and dielectric interface need to be reexamined for their validity. In the presence of free surface charge density σ_f and free surface current density $\vec{j}_f$, the boundary conditions on the field vectors are

$$H_1^{\parallel} - H_2^{\parallel} = \vec{j}_f \times \hat{n}_{21} \tag{11.46}$$

$$B_1^{\perp} - B_2^{\perp} = 0 \tag{11.47}$$

and

$$\varepsilon_1 E_1^{\perp} - \varepsilon_2 E_2^{\perp} = \sigma_f \tag{11.48}$$

$$E_1^{\parallel} - E_2^{\parallel} = 0 \tag{11.49}$$

where $\hat{n}_{21}$ is the normal from Region II to Region I, from the conducting medium towards the dielectric.

One can of course, go ahead with these boundary conditions on the field vectors at the interface and calculate the reflection and transmission coefficients just as in the case of the dielectric - dielectric interface that we considered in the previous chapter. However, this turns out to be a formidable mathematical exercise and we shall not be attempting it here. (For details, see for example, **"Electromagnetic Theory" by J.A. Stratton**).

However, if we go to the limit of $\sigma \to \infty$, i.e., for a perfect conductor, things become a bit easier. This is because in this limit, there is no electric field inside the conductor because of Ohm's Law. Recall that Ohm's Law states

$$\vec{j} = \sigma \vec{E}$$

and so, if $\sigma \to \infty$, then $\vec{E}$ must necessarily be zero for $\vec{j}$ to be finite.

In the limit of infinite conductivity or a perfect conductor, things become easier because there is no accumulation of surface currents in the conductor. This is because the field inside the conductor is zero. Thus, the right-hand side of Eq. (11.46) vanishes.

As an aside, it is important to note that when $\sigma \to \infty$, then there is no transmitted wave into the conductor. This is because for infinite conductivity, the skin depth (which we shall study in the next section) is zero and hence no waves are transmitted into the conductor if the incident wave is falling on the interface from the dielectric side.

Thus, for a perfect conductor with $\sigma \to \infty$, both the electric and magnetic fields vanish inside the conductor. To see how the boundary conditions are satisfied in the case of infinite conductivity, we first take σ to be finite and then take the limit of $\sigma \to \infty$. We take the wave number in the conducting medium, k_t as complex and write it as

$$k_t = k_{tr} + ik_{ti}$$

Also note that in the limit $\sigma \to \infty$, the real and imaginary parts of k_t can be written as

$$k_t = \omega\sqrt{\mu_2\varepsilon_2}\left(1 + \frac{i\sigma}{\omega\varepsilon_2}\right)^{1/2}$$

and therefore, in the limit

$$k_{tr} = k_{ti} = \omega\sqrt{\frac{\sigma\mu_2}{2\omega}}$$

With the complex wave number, the formalism we have worked out in Chapter 10 (Section 10.7), for the dielectric-dielectric interface can be taken over provided we use complex quantities. In the interface shown in Fig. 10.6, we take Region II to be conducting. We write the incident, transmitted and reflected electric and magnetic fields, E_i, E_r, E_t, B_i, B_r, B_t as complex quantities with the understanding that the physical fields are the real parts of these complex quantities. Thus, the fields, as in Chapter 10, Section 10.7 (Eqs. (10.109–114)) are given by

$$\vec{E}_i(\vec{r}, t) = |\vec{E}_i|\hat{\varepsilon}_i(\vec{k}_i)e^{i(\vec{k}_i\cdot\vec{r}-\omega_i t+\delta_i)} \tag{11.50}$$

$$\vec{E}_t(\vec{r}, t) = |\vec{E}_t|\hat{\varepsilon}_t(\vec{k}_t)e^{i(\vec{k}_t\cdot\vec{r}-\omega_t t+\delta_t)} \tag{11.51}$$

$$\vec{E}_r(\vec{r}, t) = |\vec{E}_r|\hat{\varepsilon}_r(\vec{k}_r)e^{i(\vec{k}_r\cdot\vec{r}-\omega_r t+\delta_r)} \tag{11.52}$$

with

$$\vec{k}_t = k_t(0, \sin\theta_t, \cos\theta_t)$$

being a complex vector.

The magnetic fields corresponding to these electric fields can be easily written down as

$$\vec{B}_i = \frac{\vec{k}_i \times \vec{E}_i}{\omega_i} \tag{11.53}$$

$$\vec{B}_t = \frac{\vec{k}_t \times \vec{E}_t}{\omega_t} \tag{11.54}$$

$$\vec{B}_r = \frac{\vec{k}_r \times \vec{E}_r}{\omega_r} \tag{11.55}$$

Note that $\vec{k}_t$ is a complex vector. Thus, in Eq. (11.51) the exponential factor will give us

$$e^{i(\vec{k}_r \cdot \vec{r} - \omega_r t + \delta_r)} = e^{-\vec{k}_{ti} \cdot \vec{r}} e^{i(\vec{k}_{tr} \cdot \vec{r} - \omega_t t + \delta_t)}$$

We see from this that there is as expected an attenuation. Further, the arguments leading to Eqs. (10.116) and (10.117) hold good and so we have

$$\omega_i = \omega_t = \omega_r$$

and

$$\delta_i = \delta_t = \delta_r$$

However, in Eq. (10.117), the quantity $\vec{k}_t \cdot \vec{r}$ is now to be replaced by $\vec{k}_{tr} \cdot \vec{r}$ since the phase of the transmitted wave is governed by the exponential with $\vec{k}_{tr}$. The velocity c_2 in the conductor would thus, be

$$c_2 = \frac{\omega_t}{\mathrm{Re}\, k_t} = \frac{\omega_t}{k_{tr}} = \frac{1}{\sqrt{\mu_2 \varepsilon_2} \mathrm{Re} \left[1 + \frac{i\sigma}{\varepsilon_2 \omega}\right]^{1/2}} \tag{11.56}$$

With these modifications, the Eqs. (10.121) and (10.122) become

$$\theta_i = \theta_r \tag{11.57}$$

and

$$\frac{\sin\theta_i}{\sin\theta_t} = \frac{c_1}{c_2} = \frac{c}{\sqrt{\mu_1 \varepsilon_1}} \sqrt{\mu_2 \varepsilon_2} \mathrm{Re} \left[1 + \frac{i\sigma}{\omega \varepsilon_2}\right]^{1/2} \approx \frac{c}{\sqrt{\mu_1 \varepsilon_1}} \sqrt{\frac{\mu_2 \sigma}{2\omega}} \tag{11.58}$$

Thus, we see that the law of reflection remains unchanged for conductors. However, for the law of refraction, in the limit $\sigma \to \infty$ we get $\sin\theta_t \to 0$. This implies that the propagation in a conductor is attenuated by the factor $e^{-k_{ti} r}$ is along the normal direction.

The velocity of propagation inside the conductor, c_2 is given by

$$c_2 = \frac{\omega}{k_{tr}} = \sqrt{\frac{2\omega}{\mu\sigma}} \tag{11.59}$$

Thus, c_2 goes to zero in the limit of $\sigma \to \infty$, $n_2 = \frac{c}{c_2}$ and hence, $n_{21} = \frac{n_2}{n_1}$ goes to ∞.

Consider first the case where the electric vector of the incident and reflected waves are perpendicular to the plane of incidence. Then there is no component of the electric vector perpendicular to the interface as can be seen in Fig. 10.10. With the perpendicular component of the electric vector not being there, the surface charge density also must vanish by the boundary conditions Eq. (11.48). We therefore, have boundary conditions no different from those for reflection and transmission between two dielectrics which we have already discussed in the last chapter. Thus, with $n_{21} \to \infty$, the ratio $\frac{E_r}{E_i} \to 1$ and $\frac{E_t}{E_i} \to 0$.

A similar result also follows for the case where the electric vector is in the plane of incidence. Thus, we see that there is no transmission and indeed total reflection in the case of a perfect conductor. This is the reason why, for example, metallic coating are put in the back side of a glass mirror enabling the entire incident light to be reflected and very little transmitted.

11.2.3 Skin Depth

Each of the expressions in Eqs. (11.39–11.42) has an exponential factor, $e^{-k_i z}$. This indicates that the amplitude of the fields, as the wave progresses along the z direction in the conducting media, attenuates. We can define a distance d by

$$d = \frac{1}{k_i} \tag{11.60}$$

which is called the **Skin Depth** of the conductor. For every distance d that the wave traverses in the conductor, its amplitude (the magnitude of the transverse electric and magnetic fields) decreases by a factor of $\frac{1}{e}$. This is also called penetration depth sometimes.

The skin depth varies inversely as k_i. From Eq. (11.27), this implies that the skin depth varies inversely as the square root of the conductivity σ. Thus, very good conductors have a very small skin depth. This also leads to an interesting phenomenon whereby very high frequency electromagnetic radiation, for instance microwaves, have a very small skin depth in most conductors and thus, the fields are restricted to a very small region near the surface of the conductors.

PROBLEM 11.1 Show that for reflection of a wave traveling in vacuum at an interface with a perfect conductor, the electric field of the reflected wave has the same amplitude but has a phase difference of π relative to the incident one.

11.3 PROPAGATION OF E.M WAVES IN PRESENCE OF CONDUCTORS—WAVE GUIDES

In our discussion of electromagnetic waves, whether in vacuum or in a dielectric medium, we have not imposed any boundary conditions on the electric and magnetic fields at any finite point in space. This is because we have considered infinite homogenous medium or vacuum and the propagation was in the medium or vacuum throughout. When conductors are present, the electromagnetic waves inside them are highly attenuated, as we have seen above. For electromagnetic waves propagating outside the conductors, the presence of the conductors does have an effect on the fields since there are definite boundary conditions which need to be imposed on the fields at the interface surfaces. Such interfacing surfaces are called wave guides and we study the properties of the electromagnetic waves in a vacuum or dielectrics in the presence of such surfaces or wave guides.

The conductors we consider are ideal conductors, i.e., those having infinite conductivity σ. Now by Ohm's Law, $\vec{j} = \sigma \vec{E}$, an infinite conductivity implies a vanishing electric field inside the conductor. Furthermore, Faraday's Law

$$\vec{\nabla} \times \vec{E} = -\frac{\partial \vec{B}}{\partial t}$$

implies, for such a case, a time-independent or static field. Since we will be disregarding static fields in our consideration of electromagnetic waves, we can take the magnetic field to be zero inside the perfect or ideal conductor. Thus, we see that the conductor will have vanishing $\vec{E}$ and $\vec{B}$ inside it.

The boundary conditions at the interface of a conductor and a dielectric have already been derived. The electric field and the magnetic field inside the conductor is zero and hence, we have the conditions as

$$E_{||} = 0 \qquad\qquad B_{\perp} = 0 \tag{11.61}$$

The normal component of $\vec{E}$ and the tangential component of $\vec{B}$ will not have similar boundary conditions because there might be surface charge and surface currents at the interface.

Consider that a perfect conductor is present all along the z-axis. To visualise, this could be, for example, an infinite hollow cylinder with its axis along the z-axis, as in Fig. 11.1.

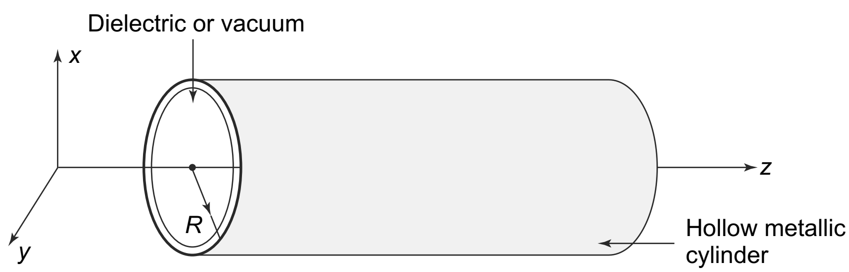


Fig. 11.1 *A wave guide in the shape of a metallic cylinder with vacuum or a dielectric inside. For a perfect metal, a wave propagating along the z-axis is confined to the hollow region inside the cylinder and has to satisfy the appropriate boundary conditions at the inner surface of the cylinder*

Consider now a wave propagating along the z-direction all along the hollow space with a wave number k so that $c_m k = \omega$, since the hollow region inside the wave guide contains a dielectric. If there is no dielectric then c_m will be just c. The electric and magnetic fields inside the wave guide will have the form:

$$\begin{aligned} \vec{E} &= \vec{E}_0(x, y)e^{i(kz-\omega t)} \\ \vec{B} &= \vec{B}_0(x, y)e^{i(kz-\omega t)} \end{aligned} \tag{11.62}$$

For an electromagnetic wave propagating in free space along the z-direction, the amplitudes $\vec{E}_0$ and $\vec{B}_0$ in Eq. (11.62) will not have any dependence on the transverse coordinates x and y. In the present case, the electric and magnetic fields have to satisfy boundary conditions at specified values of x and y representing the inner surface of the wave guide. Thus, in the example shown in Fig. 11.1, boundary is at $x^2 + y^2 = R^2$. The amplitudes $\vec{E}_0$ and $\vec{B}_0$ will thus, be dependent on the transverse coordinates x and y for wave guides.

There being no free charges or currents inside the wave guide, $\vec{E}$ and $\vec{B}$ will satisfy the usual homogeneous Maxwell's equations in a dielectric medium:

$$\begin{aligned}
\vec{\nabla} \cdot \vec{E} &= 0 \\
\vec{\nabla} \cdot \vec{B} &= 0 \\
\vec{\nabla} \times \vec{E} &= -\frac{\partial \vec{B}}{\partial t} \\
\vec{\nabla} \times \vec{B} &= \frac{1}{c_m^2} \frac{\partial \vec{E}}{\partial t}
\end{aligned} \tag{11.63}$$

The $\vec{E}$ and $\vec{B}$ in the dielectric satisfy the wave equation whose solutions represent travelling waves with a speed $c_m = \frac{1}{\sqrt{\mu\varepsilon}}$.

$$\left(\nabla^2 - \frac{1}{c_m^2}\frac{\partial^2}{\partial t^2}\right)\vec{E} = 0 \tag{11.64}$$

and

$$\left(\nabla^2 - \frac{1}{c_m^2}\frac{\partial^2}{\partial t^2}\right)\vec{B} = 0 \tag{11.65}$$

Given the form of the fields in Eq. (11.62), this reduces to

$$\begin{aligned}
\left(k^2 - \nabla_t^2 - \frac{\omega^2}{c_m^2}\right)\vec{E} &= 0 \\
\left(k^2 - \nabla_t^2 - \frac{\omega^2}{c_m^2}\right)\vec{B} &= 0
\end{aligned} \tag{11.66}$$

We can also use the Curl equations in the Maxwell's equations to express the transverse components (that is the components of $\vec{E}$ and $\vec{B}$ perpendicular to the interface) in terms of the longitudinal component. To do this, consider the Curl of $\vec{E}$

$$\vec{\nabla} \times \vec{E} = -\frac{\partial \vec{B}}{\partial t} = -\mu\frac{\partial \vec{H}}{\partial t}$$

Writing it out in the components we get

$$\frac{\partial E_z}{\partial y} - ikE_y = i\omega\mu H_x \tag{11.67}$$

$$-\frac{\partial E_z}{\partial x} + ikE_x = i\omega\mu H_y \tag{11.68}$$

$$\frac{\partial E_x}{\partial y} - \frac{\partial E_y}{\partial x} = i\omega\mu H_z \tag{11.69}$$

Similarly, using the expression for Curl of $\vec{H}$, we get

$$\frac{\partial H_z}{\partial y} - ikH_y = -i\omega\varepsilon E_x \tag{11.70}$$

$$-\frac{\partial H_z}{\partial x} + ikH_x = -i\omega\varepsilon E_y \tag{11.71}$$

$$\frac{\partial H_x}{\partial y} - \frac{\partial H_y}{\partial x} = -i\omega\varepsilon E_z \tag{11.72}$$

It is convenient to rearrange the above equations to write the transverse components in terms of the longitudinal components.

$$\left(\frac{\omega^2}{c_m^2} - k^2\right) E_x = ik\left(\frac{\partial E_z}{\partial x} + c_m \frac{\partial B_z}{\partial y}\right) \tag{11.73}$$

$$\left(\frac{\omega^2}{c_m^2} - k^2\right) B_x = ik\left(\frac{\partial B_z}{\partial x} - \frac{1}{c_m}\frac{\partial E_z}{\partial y}\right) \tag{11.74}$$

$$\left(\frac{\omega^2}{c_m^2} - k^2\right) E_y = ik\left(\frac{\partial E_z}{\partial y} - c_m \frac{\partial B_z}{\partial x}\right) \tag{11.75}$$

$$\left(\frac{\omega^2}{c_m^2} - k^2\right) B_y = ik\left(\frac{\partial B_z}{\partial y} + \frac{1}{c_m}\frac{\partial E_z}{\partial x}\right) \tag{11.76}$$

These equations tell us that if the wave equations (Eq. (11.64) and (11.65) are solved for the z-component of the fields, the transverse components are immediately obtained from them. Thus, the solution of the wave guide equation reduces to finding the components of the fields only along the z-direction subject to their satisfying the boundary conditions on the interface.

We will be restricting our discussion throughout to wave guides in the form of an infinite hollow pipe along the z direction. The pipe could be filled with a homogenous dielectric too. Furthermore, the cross section of the pipe in the $x - y$ plane is uniform throughout the length of the pipe. For such wave guides, it is common to classify the solutions to the wave equation in the following way:

(a) **TEM mode:** This is the transverse electric and magnetic mode. It is a solution where the fields are only transverse, i.e., $E_z = H_z = 0$.

(b) **TE mode:** Here $E_z = 0$ but $H_z \neq 0$.

(c) **TM mode:** Here $H_z = 0$ but $E_z \neq 0$.

Note that these are the only possible independent modes. One can of course have hybrid electromagnetic waves in the wave guide but these will then be superpositions of the TE and TM nodes. In these hybrid modes, all the components of electromagnetic fields are present.

11.3.1 TEM Mode

This is the mode in which there are only transverse electric and magnetic fields. It is easy to show that this kind of mode is not possible in the kind of wave guides that we are considering, namely, hollow, infinite pipes in the z direction with a uniform cross section.

To see this consider the divergence of $\vec{E}$. Since $E_z = 0$ for such a mode, inside the wave guide but outside the conductors, we have

$$\frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} = 0 \tag{11.77}$$

Similarly, since there is no magnetic field along the z direction, Faraday's laws tells us that the z component of the curl of $\vec{E}$ must be zero. That is

$$\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} = 0 \tag{11.78}$$

Thus, the electric field vector $\vec{E}_0(x, y)$ in Eq. (11.62) has zero divergence and zero curl if $E_{0z} = 0$. Similarly, we can see that if $\vec{B}_{0z} = 0$ then the vector $\vec{B}_0(x, y)$ has zero divergence and curl. Now, in Section 6.7.1 we have already seen that any vector can be written as a sum of a solenoidal part (divergenceless) and an irrotational (curl is zero) part (Helmhotz theorem). Since for $\vec{E}_0$ and $\vec{B}_0$ both these parts are zero when $E_{0z} = B_{0z} = 0$, we can conclude that $\vec{E}_0 = \vec{B}_0 = 0$. **Thus, we see that in an open wave guide of the kind we are considering, TEM modes are not possible.**

11.3.2 TE Mode

In these kinds of modes, we have vanishing longitudinal electric field, i.e., $E_z = 0$ but $H_z \neq 0$. We can seek solutions of Eq. (11.66) of the form

$$H_z = X(x)Y(y)e^{i(kz-\omega t)} \tag{11.79}$$

Substituting in Eq. (11.66) and dividing throughout by H_z we get

$$k^2 - \frac{\omega^2}{c^2} = \frac{X(x)''}{X(x)} + \frac{Y(y)''}{Y(y)} \tag{11.80}$$

Let us consider the right-hand side. The first term is only a function of x while the second term is only a function of y. They add up to the left-hand side which is independent of both x and y. Thus, the two terms on the right-hand side must individually be independent of x and y. That is

$$\begin{aligned} X(x)'' &= -(k_x^H)^2 X(x) \\ Y(y)'' &= -(k_y^H)^2 Y(y) \end{aligned} \tag{11.81}$$

These equations are easily solved to get

$$\begin{aligned} X(x) &= X_0^H \sin(k_x^H x + \delta_x^H) \\ Y(y) &= Y_0^H \sin(k_y^H y + \delta_y^H) \end{aligned} \tag{11.82}$$

where X_0^H, Y_0^H, δ_x^H and δ_y^H are constants. Thus, we can write the magnetic field as

$$H_z = X_0^H Y_0^H \sin(k_x^H x + \delta_x^H) \sin(k_y^H y + \delta_y^H) e^{i(kz-\omega t)} \tag{11.83}$$

The constants of course need to be determined by applying boundary conditions on the field vectors. We shall do this for specific wave guides.

11.3.3 TM Mode

In this mode, $E_z \neq 0$ and $H_z = 0$. We proceed in exactly the same way as for the TE mode and get

$$E_z = X_0^E Y_0^E \sin(k_x^E x + \delta_x^E) \sin(k_y^E y + \delta_y^E) e^{i(kz-\omega t)} \tag{11.84}$$

11.3.4 Rectangular Waveguide

We can now determine the constants for the TE and TM modes above for a specific geometry. Let us consider a wave guide which is infinite in the z direction and has a rectangular cross section as shown in Fig. 11.2.

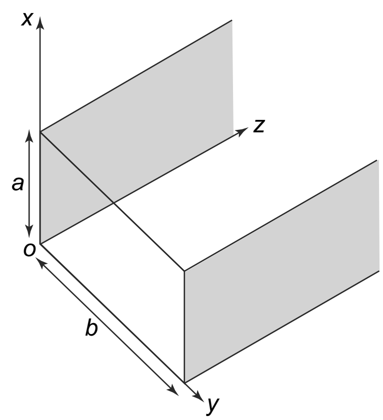


Fig. 11.2 *An infinite wave guide along the z direction with a rectangular cross section of length a along the x direction and b along the y direction*

For such a wave guide, we can impose the boundary conditions and therefore, try to find the constants above and hence the fields in both the possible modes.

(i) At $x = 0$ and $x = a$.
In this case, the tangential direction is along the y direction and the normal direction is along the x direction. Thus, the boundary conditions give us (Eq. (11.61)),

$$E_y(x = 0, a) = 0 \tag{11.85}$$

and

$$H_x(x = 0, a) = 0 \tag{11.86}$$

everywhere in the wave guide, i.e., for all z.

(ii) At $y = 0$ and $y = b$, the tangential direction is along the x direction while the normal direction is along the y axis. Thus, the boundary conditions here are

$$E_x(y = 0, b) = 0 \tag{11.87}$$

and

$$H_y(y = 0, b) = 0 \tag{11.88}$$

for all values of z in the wave guide. We can apply these conditions to the two modes now.

For the **TE mode**:

We will use these boundary conditions for the rectangular wave guide geometry to determine the allowed values of k_x^H and k_y^H. Using the Maxwell's equation for the Curl of $\vec{H}$, we get

$$\frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x} = -i\omega\varepsilon E_y \tag{11.89}$$

Here we have used the time dependence of the $\vec{E}$ field and the relationship of $\vec{D}$ and $\vec{E}$. Now applying the boundary conditions on $\vec{E}$ and $\vec{H}$ (Eqs. (11.85 and 11.86)), we get

$$\frac{\partial H_z}{\partial x} = 0 \quad \text{at } x = 0, a \tag{11.90}$$

But we know the form of H_z, i.e., Eq. (11.83). So applying Eq. (11.90) to that we get

$$\cos(\delta_x^H) = 0 \tag{11.91}$$

and

$$\cos(k_x a + \delta_x^H) = 0 \tag{11.92}$$

Thus,

$$\delta_x^H = \frac{\pi}{2}$$

and

$$k_x = \frac{m\pi}{a} \quad m = 0, 1, 2, 3, \cdots$$

Further, we can use the Curl of $\vec{H}$ equation again to get

$$\frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} = -i\omega\varepsilon E_x \tag{11.93}$$

Using the boundary conditions (Eqs. (11.87 and 11.88) we get

$$\frac{\partial H_z}{\partial y} = 0 \quad y = 0, b \tag{11.94}$$

Once again using the form of H_z in Eq. (11.83) we get

$$\cos(\delta_y^H) = 0 \tag{11.95}$$

and

$$\cos(k_y b + \delta_y^H) = 0 \tag{11.96}$$

We get

$$\delta_y^H = \frac{\pi}{2}$$

and

$$k_y = \frac{n\pi}{b} \quad n = 0, 1, 2, 3, \cdots$$

With these values of k_x, k_y, δ_x^H and δ_y^H, we get

$$\boxed{H_z = H_z^0 \cos\left(\frac{m\pi x}{a}\right)\cos\left(\frac{n\pi y}{b}\right) e^{i(kz-\omega t)}} \tag{11.97}$$

where

$$H_0 = X_0^H Y_0^H$$

and

$$m, n = 0, 1, 2, 3 \cdots$$

The H_z that we have obtained in Eq. (11.97) is called the TE_{mn} mode. Conventionally, for rectangular waveguides, the longer of the two sides is taken to be parallel to the x axis, and so, $a > b$.

Let us analyse Eq. (11.97). First of all note that the TE_{00} mode, though allowed by the values of k_x and k_y is absent. This can be seen easily. Choose a closed curve along the rectangular cross section of the wave guide. Now in the TE mode, since there is no longitudinal component of the electric field, it is everywhere tangential to the closed curve. Hence, it vanishes because of the boundary conditions. This implies that the EMF along this closed curve also vanishes since

$$\text{EMF} = \oint \vec{E} \cdot \vec{dl}$$

Now for this TE_{00} mode, H_z will be independent of x and y since $m = n = 0$. Further, it would vary like $e^{i\omega t}$ with time. Faraday's Law tells us that since the induced EMF is zero, the rate of change of flux, i.e., $B_z \times$ (cross-section area) will vanish. This implies that H_z^0 must vanish. Thus, we see that for this mode, the z component of both $\vec{E}$ and $\vec{H}$ vanishes. But we have already argued above that such a mode, which is essentially a TEM mode cannot exist in these kinds of wave guides. Hence, at least one of the integers, m, n must be non-zero for a TE mode to exist in a rectangular waveguide.

The wave equation, Eq. (11.66) gives us a relationship between ω and k, which in this case becomes

$$\frac{\omega^2}{c_m^2} - k^2 - \left(\frac{m\pi}{a}\right)^2 - \left(\frac{n\pi}{b}\right)^2 = 0 \tag{11.98}$$

Since k is real and hence $k^2 \geq 0$, we get

$$w \geq c_m^2 \pi^2 \left(\frac{m^2}{a^2} + \frac{n^2}{b^2}\right)^{1/2} \equiv \omega_{mn}^{TE} \tag{11.99}$$

ω_{mn} obviously increases with increasing values of m or n. Equation (11.99) means that at any particular value of ω only certain values of m, n which satisfy the inequality are excited in the wave guide. Since we have chosen $a > b$, the lowest value of ω_{mn} can be

$$\omega_{10}^{TE} = \frac{c_m \pi}{a} \tag{11.100}$$

The next higher value of ω_{mn} can be either ω_{11} or ω_{20}. This will depend on the relative value of a and b. This leads to an interesting result. Waves in such a rectangular wave guide, with ω above ω_{10} but below the next higher value of ω_{mn}, i.e., either ω_{20} or ω_{11} have a unique mode of propagation, i.e., the TE_{10} mode.

We can carry out an analogous calculation of E_z for the TM mode. This time though, we will require E_z to vanish at $x = 0, a$ and at $y = 0, b$. Putting this in, we will get

$$E_z = E_z^0 \sin\left(\frac{m\pi x}{a}\right) \sin\left(\frac{n\pi y}{b}\right) e^{i(kz-\omega t)} \quad m, n = 1, 2, 3, \cdots \tag{11.101}$$

Once again, the frequency can be determined as

$$\frac{\omega^2}{c_m^2} - k^2 - \left(\frac{m\pi}{a}\right)^2 - \left(\frac{n\pi}{b}\right)^2 = 0 \tag{11.102}$$

and using the same arguments we get

$$w \geq c_m^2 \pi^2 \left(\frac{m^2}{a^2} + \frac{n^2}{b^2}\right)^{1/2} \equiv \omega_{mn}^{TM} \tag{11.103}$$

Since m, n cannot have the value 0, the lowest mode possible in the TM case is thus,

$$\omega_{11}^{TM} = c_m^2 \pi^2 \left(\frac{1}{a^2} + \frac{1}{b^2} \right)^{1/2} \tag{11.104}$$

We see therefore, that the lowest mode in the TM case is higher by a factor of $\left(1 + \frac{a^2}{b^2}\right)^{1/2}$ than the lowest mode in the TE case, i.e., ω_{10}^{TE}.

EXAMPLE 11.1 A rectangular wave guide, infinite along the z-axis has dimensions along the x and y axes respectively of 4.0 and 2.0 cm. How many TE modes can propagate along the waveguide at a value of angular frequency $\omega = 10^9$ Hz? If one wanted to excite only the TE_{11} mode amongst all TE modes to propagate, what ranges of frequencies are allowed?

Solution

For the dimensions given, we know that

$$\omega_{mn} = \left(\frac{c\pi}{4}\right) \left[m^2 + 4n^2\right]^{1/2} = 10^9$$

Hence, the allowed values of m, n are

$$[m^2 + 4n^2] < \left(\frac{13.3}{\pi}\right)^2 \simeq 18$$

which tells us that the values of (mn) possible are 10, 01, 11, 20, 02, 12, 21, 30, 31.

The TE_{11} mode has a frequency

$$\omega_{11} \simeq 5.26 \times 10^8$$

The allowed frequency just below this is for the 01 and 20 modes, which is 4.71×10^8 Hz. The allowed frequency just above that of the 11 mode is that of the TE_{21} mode, which is 6.66×10^8 Hz. Thus the allowed range of frequencies for which only the TE_{11} mode will be excited is

$$4.71 \times 10^8 < \omega < 6.66 \times 10^8 \quad \text{Hz}$$

PROBLEM 11.2 A square wave guide of size b has air inside. What is the maximum value of b such that the TE_{10} mode for propagation of wave with wave number k goes through but not the TM_{11} or TE_{11}.

11.3.5 Energy Transmission

Once we have determined the electric and magnetic fields, the computation of the energy transmitted in the wave guide is easily done. We shall do this for the TE_{10} mode since other modes are similar.

The real part of the magnetic field H_z in Eq. (11.97) is given by

$$(H_z)_r = H_z^0 \cos\left(\frac{m\pi x}{a}\right) \cos\left(\frac{n\pi y}{b}\right) \cos(kz - \omega t) \tag{11.105}$$

There is no longitudinal electric field in the TE mode, i.e., $E_z = 0$. The real, transverse components of $\vec{E}$ and $\vec{H}$ can be determined from Eqs. (11.73–11.76). These are

$$
\begin{aligned}
(E_x)_r &= \frac{\mu\omega H_z^0}{(\frac{\omega^2}{c_m^2} - k^2)} \left(\frac{n\pi}{b}\right) \cos\frac{m\pi x}{a} \sin\frac{n\pi y}{b} \sin(kz - \omega t) \\
(E_y)_r &= -\frac{\mu\omega H_z^0}{(\frac{\omega^2}{c_m^2} - k^2)} \left(\frac{m\pi}{a}\right) \sin\frac{m\pi x}{a} \cos\frac{n\pi y}{b} \sin(kz - \omega t) \\
(B_x)_r &= \frac{\mu k H_z^0}{(\frac{\omega^2}{c_m^2} - k^2)} \left(\frac{m\pi}{a}\right) \sin\frac{m\pi x}{a} \cos\frac{n\pi y}{b} \sin(kz - \omega t) \\
(B_y)_r &= \frac{\mu k H_z^0}{(\frac{\omega^2}{c_m^2} - k^2)} \left(\frac{n\pi}{b}\right) \cos\frac{m\pi x}{a} \sin\frac{n\pi y}{b} \sin(kz - \omega t)
\end{aligned}
\tag{11.106}
$$

The rate of energy flow, per unit area is given as we have seen, by the Poynting vector $\vec{S}$.

$$\vec{S} = (\vec{E})_r \times (\vec{H})_r \tag{11.107}$$

However, both $\vec{E}$ and $\vec{B}$ are oscillatory with either a sine or a cosine dependence. For calculating the time averages, we use the following results.

$$< \sin^2(kz - \omega t) >=< \cos^2(kz - \omega t) >= \frac{1}{2} \tag{11.108}$$

and

$$< \sin(kz - \omega t)\cos(kz - \omega t) >= 0 \tag{11.109}$$

Here $< \cdots >$ denotes the time average. Notice that the transverse components of $\vec{S}$ have vanishing time averages because of Eq. (11.109) (recall that the real parts of H_z and E_x have different time dependences). This is expected since the electric and magnetic fields are propagating inside the wave guide and hence, there is no energy transmission through the conducting surfaces.

The energy flow is thus, as expected along the z direction. Taking the time average of S_z we get

$$< S_z >= \frac{1}{2}\frac{\mu\omega k(H_z^0)^2}{\left[\frac{\omega^2}{c_m^2} - k^2\right]^2}\left[\left(\frac{n\pi}{b}\right)^2 \cos^2\frac{m\pi x}{a} \sin^2\frac{n\pi y}{b} - \left(\frac{m\pi}{a}\right)^2 \sin^2\frac{m\pi x}{a} \cos^2\frac{n\pi y}{b}\right] \tag{11.110}$$

For the TE_{10} mode, $m = 1, n = 0$ and so we get

$$< S_z >= \frac{1}{2}\frac{\mu\omega k(H_z^0)^2}{\left[\frac{\omega^2}{c_m^2} - k^2\right]^2}\left[-\left(\frac{\pi}{a}\right)^2 \sin^2\frac{\pi x}{a}\right] \tag{11.111}$$

Using Eq. (11.106), we can rewrite this as

$$< S_z >= \frac{1}{2}\frac{(E_y^0)^2 k}{\omega\mu} \sin^2\frac{\pi x}{a} \tag{11.112}$$

where E_y^0 is the maximum value of the y component of the electric field in the wave guide.

The total energy transmitted in unit time, W_T is the integral of this rate over the cross section. Thus,

$$W_T = \int_0^a dx \int_0^b dy < S_z >$$

$$= \frac{b}{2}|E_y^0|^2 \frac{k}{\omega\mu} \int_0^a dx \sin^2 \frac{\pi x}{a}$$

$$= \frac{A}{4}|E_y^0|^2 \frac{k}{\omega\mu} \tag{11.113}$$

where A is the area of the cross section $A = a \times b$. The quantity $\frac{k}{\omega\mu}$ is the magnitude of the ratio of B_x and E_y. **It is called the impedance of the waveguide**.

PROBLEM 11.3 A rectangular wave guide of size 6 cm by 4 cm with air inside has a TE_{10} mode of electromagnetic wave travelling through it with its electric field parallel to the 4 cm side. The frequency of the wave is high at 2×10^9 Hz and the power transmitted through the wave guide is 0.6 megawatt. Calculate the value of the electric field inside the waveguide.

11.4 ELECTROMAGNETIC POTENTIALS

We saw in Chapter 2 that introducing the scalar potential $\phi(\vec{r})$ in the case of electrostatics, allows us to formulate the whole of electrostatics in terms of a scalar function rather than the vector field $\vec{E}(\vec{r})$. This led to a considerable simplification in calculation of the electric fields for a charge configuration. We usually had to solve the Laplace or Poisson's equations for the given charge distribution and determine the electric scalar potential solution which respected the boundary conditions. By taking the gradient of this scalar potential, one could determine the electric field at any point.

We have also seen that in the presence of time varying magnetic fields, the line integral of the electric field over a closed loop,

$$\oint \vec{E} \cdot d\vec{l}$$

is not equal to zero in general. Hence a straightforward definition of a scalar potential is not possible. However, as we shall now see, Faraday's Law, which related the Curl of the electric field to the rate of change of the magnetic field allows us to define a scalar potential which is valid in more general cases.

Recall that the vector potential $\vec{A}(\vec{r})$ was related to the magnetic field by

$$\vec{B}(\vec{r}, t) = \vec{\nabla} \times \vec{A}(\vec{r}, t) \tag{11.114}$$

On the other hand, Faraday's Law states that

$$\vec{\nabla} \times \vec{E}(\vec{r}, t) = -\frac{\partial \vec{B}}{\partial t} \tag{11.115}$$

Combining Eqs. (11.114) and (11.115), we get

$$\vec{\nabla} \times \left(\vec{E} + \frac{\partial \vec{A}}{\partial t} \right) = 0 \tag{11.116}$$

Thus, we have a quantity whose Curl vanishes and by the result of Section 6.7.1, we can express it as a gradient of a scalar function $\phi(\vec{r}, t)$.

$$\vec{E} + \frac{\partial \vec{A}}{\partial t} = -\vec{\nabla}\phi \tag{11.117}$$

The quantity $\phi(\vec{r}, t)$ defined in this way is called the **electromagnetic scalar potential.** Obviously from the definition of $\phi(\vec{r}, t)$, if the magnetic fields present are time-independent, $\frac{\partial \vec{A}}{\partial t} = 0$ and the electromagnetic scalar potential reduces to the usual electrostatic scalar potential.

We have already discussed the arbitrariness in the definition of the vector potential (Section 6.7.1). Recall that a transformation of $\vec{A}$ to $\vec{A}'$ with

$$\vec{A} \to \vec{A}' = \vec{A} + \vec{\nabla}\chi \tag{11.118}$$

leaves the magnetic field unchanged since the magnetic field is the curl of the vector potential which remains unchanged with the addition of a gradient of an arbitrary scalar χ. This kind of transformation, as we have seen, is called a gauge transformation. What about its effect on the electric field? Clearly, if χ is time-dependent, then this transformation will change the electric field unless a compensating term is added to it to take care of the extra term coming from the $\frac{\partial \vec{A}}{\partial t}$ term in Eq. (11.117). Therefore, when we make a transformation of $\vec{A}$ as in Eq. (11.118) we need to also make a transformation of ϕ as

$$\phi \to \phi' = \phi - \frac{\partial \chi}{\partial t} \tag{11.119}$$

It is easy to check that if we perform **both** the transformations (Eqs. (11.118 and 11.119)), together, the electric and magnetic fields will remain unchanged. **These are called general gauge transformations of the potentials ϕ and $\vec{A}$ which leave the physical quantities, namely the electric and magnetic fields unchanged.**

With the definitions of the electric and magnetic fields in terms of the vector and scalar potential, we see that two of Maxwell's equations are automatically satisfied.

$$\vec{\nabla} \cdot \vec{B} = 0$$

since $\vec{B} = \vec{\nabla} \times \vec{A}$ and divergence of a curl vanishes. Faraday's Law is also automatically satisfied since we have used it to define the electromagnetic potential in Eq. (11.117). The other two equations, namely, Ampere's Law and Gauss's Law need to be formulated in terms of the potentials $\vec{A}$ and ϕ.

Considering the fields in vacuum, we have Gauss's Law as

$$\begin{aligned} \vec{\nabla} \cdot \vec{E} &= \frac{\rho}{\varepsilon_0} \\ &= \vec{\nabla} \cdot \left(-\vec{\nabla}\phi - \frac{\partial \vec{A}}{\partial t} \right) \end{aligned}$$

$$-\varepsilon_0 \nabla^2 \phi - \varepsilon_0 \frac{\partial}{\partial t}(\vec{\nabla} \cdot \vec{A}) = \rho \tag{11.120}$$

Similarly, Ampere's Law (with the Maxwell modification) gives us

$$\vec{\nabla} \times \vec{B} = \vec{\nabla}(\vec{\nabla} \cdot \vec{A}) - \nabla^2 \vec{A}$$

$$= \mu_0 \vec{j} + \mu_0 \varepsilon_0 \frac{\partial \vec{E}}{\partial t}$$

$$= \mu_0 \vec{j} + \mu_0 \varepsilon_0 \left(-\vec{\nabla} \frac{\partial \phi}{\partial t} - \frac{\partial^2 \vec{A}}{\partial t^2} \right)$$

$$\vec{\nabla}(\vec{\nabla} \cdot \vec{A}) - \nabla^2 \vec{A} = \mu_0 \vec{j} - \mu_0 \varepsilon_0 \vec{\nabla} \frac{\partial \phi}{\partial t} - \mu_0 \varepsilon_0 \frac{\partial^2 \vec{A}}{\partial t^2} \tag{11.121}$$

Equations (11.120) and (11.121) look very formidable as differential equations to be solved to obtain ϕ and $\vec{A}$. However, we can use the freedom of gauge transformations to simplify them since recall that gauge transformations on potentials leave the physical fields unchanged.

Under the gauge transformations (Eqs. (11.118) and (11.119)), the quantity

$$\vec{\nabla} \cdot \vec{A} + \frac{1}{c^2} \frac{\partial \phi}{\partial t}$$

transforms as

$$\left(\vec{\nabla} \cdot \vec{A} + \frac{1}{c^2} \frac{\partial \phi}{\partial t} \right) \rightarrow \left(\vec{\nabla} \cdot \vec{A}' + \frac{1}{c^2} \frac{\partial \phi'}{\partial t} \right)$$

$$\left(\vec{\nabla} \cdot \vec{A}' + \frac{1}{c^2} \frac{\partial \phi'}{\partial t} \right) = \left(\vec{\nabla} \cdot \vec{A} + \frac{1}{c^2} \frac{\partial \phi}{\partial t} \right) + \left(\nabla^2 \chi - \frac{1}{c^2} \frac{\partial^2 \chi}{\partial t^2} \right) \tag{11.122}$$

We now choose χ such that the right-hand side of the equation vanishes, i.e.,

$$\left(\nabla^2 \chi - \frac{1}{c^2} \frac{\partial^2 \chi}{\partial t^2} \right) \equiv \Box \chi = -\left(\vec{\nabla} \cdot \vec{A} + \frac{1}{c^2} \frac{\partial \phi}{\partial t} \right) \tag{11.123}$$

The χ which satisfies this equation will then be given by

$$\chi = -\int d^3 \vec{r}' \int dt' G(\vec{r}, \vec{r}', t, t') \left(\vec{\nabla} \cdot \vec{A}(\vec{r}', t') + \frac{1}{c^2} \frac{\partial \phi(\vec{r}', t')}{\partial t} \right) \tag{11.124}$$

where $G(\vec{r}, \vec{r}', t, t')$ is the Green's function for the Box operator or the d'Alermbetian, $\Box$, that is

$$\Box G(\vec{r}, \vec{r}', t, t') = \delta^3(\vec{r} - \vec{r}')\delta(t - t')$$

We will not do an explicit calculation of the Green's function of the box operator. In the next subsection when solving the equation for ϕ involving the Box operator, we will write down the solution and then verify that the equation is satisfied.

Note that the χ given in Eq. (11.124) ensures that $\vec{A}'$ and ϕ' in Eq. (11.122) satisfy the equation

$$\left(\vec{\nabla} \cdot \vec{A}' + \frac{1}{c^2} \frac{\partial \phi'}{\partial t} \right) = 0$$

Choosing the χ in this fashion is choosing a specific gauge. In this case, the gauge is called the **Lorentz gauge.** We will henceforth, choose this gauge and since we will be working only in this gauge, we shall drop the primes. The ϕ and $\vec{A}$ that we choose will always satisfy

$$\left(\vec{\nabla} \cdot \vec{A} + \frac{1}{c^2} \frac{\partial \phi}{\partial t} \right) = 0$$

In the Lorentz gauge, the Eqs. (11.120) and (11.121) simplify to

$$\Box\phi(\vec{r},t) = -\frac{\rho(\vec{r},t)}{\varepsilon_0} \tag{11.125}$$

and

$$\Box\vec{A}(\vec{r},t) = -\mu_0\vec{j}(\vec{r},t) \tag{11.126}$$

Note that these equations, unlike the wave equations we derived for $\vec{E}$ and $\vec{B}$ in the absence of charge and current densities, are inhomogeneous.

However, in the case when $\rho(\vec{r},t)$ and $\vec{j}(\vec{r},t)$ become independent of time, these equations become Poisson's equations, which we have solved earlier. We can follow the same techniques to solve Eqs. (11.125) and (11.126) in that case. We will however, write the general solution for ϕ and $\vec{A}$ and then show that they satisfy these equations.

11.4.1 Retarded Potentials

The solution to Eq. (11.125) can be formally written as

$$\begin{aligned}\phi(\vec{r},t) &= -\int G(\vec{r},\vec{r}',t,t')\left(\frac{\rho(\vec{r}',t')}{\varepsilon_0}\right)d^3\vec{r}'dt' \\ &= \frac{1}{4\pi\varepsilon_0}\int d^3\vec{r}'\frac{\rho(\vec{r}',t')}{|\vec{r}-\vec{r}'|}\end{aligned} \tag{11.127}$$

Here

$$t' = t - \frac{|\vec{r}-\vec{r}'|}{c}$$

is called the **retarded time**. To check that Eq. (11.127) indeed is a solution to Eq. (11.125), we make use of the result

$$\nabla^2\frac{1}{|\vec{r}-\vec{r}'|} = -4\pi\delta^3(\vec{r}-\vec{r}') \tag{11.128}$$

Consider first the time derivative in the Box operator,

$$\Box = -\frac{1}{c^2}\frac{\partial^2}{\partial t^2} + \nabla^2$$

In Eq. (11.127), the time t enters the solution only through t'. Thus, operating by the time part of the box operator on $\phi(\vec{r},t)$, we get

$$\frac{1}{c^2}\frac{\partial^2}{\partial t^2}\phi(\vec{r},t) = \frac{1}{4\pi\varepsilon_0}\int d^3\vec{r}'\frac{\frac{1}{c^2}\frac{\partial^2}{\partial t^2}\rho(\vec{r}',t')}{|\vec{r}-\vec{r}'|} \tag{11.129}$$

We next find the spatial part of the Box operator, i.e., ∇^2.

To do this, let us first find the gradient of the potential.

$$\vec{\nabla}\phi(\vec{r},t) = \frac{1}{4\pi\varepsilon_0}\int d^3\vec{r}'\left[\left(\vec{\nabla}\rho(\vec{r}',t')\right)\frac{1}{|\vec{r}-\vec{r}'|} + \rho(\vec{r}',t')\vec{\nabla}\left(\frac{1}{|\vec{r}-\vec{r}'|}\right)\right]$$

$$= \frac{1}{4\pi\varepsilon_0}\int d^3\vec{r}'\left[\left(-\frac{1}{c}\right)\left(\frac{\partial\rho}{\partial t}\right)\vec{\nabla}(|\vec{r}-\vec{r}'|)\frac{1}{|\vec{r}-\vec{r}'|} + \rho\vec{\nabla}\left(\frac{1}{|\vec{r}-\vec{r}'|}\right)\right]$$

$$= \frac{1}{4\pi\varepsilon_0}\int d^3\vec{r}'\left[-\frac{1}{c}\frac{\partial\rho}{\partial t}\frac{(\vec{r}-\vec{r}')}{|\vec{r}-\vec{r}'|^2} - \rho\frac{(\vec{r}-\vec{r}')}{|\vec{r}-\vec{r}'|^3}\right] \tag{11.130}$$

Note that we have used the fact that there are two instances of dependance on $\vec{r}$ in the integrand. Once, through the denominator, $\frac{1}{|\vec{r}-\vec{r}'|}$ and once through the retarded time t', since $t' = t - \frac{|\vec{r}-\vec{r}'|}{c}$. We have also used the fact that

$$\vec{\nabla}|\vec{r}-\vec{r}'| = \frac{\vec{r}-\vec{r}'}{|\vec{r}-\vec{r}'|}$$

and

$$\vec{\nabla}\frac{1}{|\vec{r}-\vec{r}'|} = -\frac{\vec{r}-\vec{r}'}{|\vec{r}-\vec{r}'|^3}$$

Also note that differentiation with time t is the same as that with time t' since $t' = t - \frac{|\vec{r}-\vec{r}'|}{c}$ and $|\vec{r}-\vec{r}'|$ is independent of t.

We can now take the divergence to get the spatial part of the Box operator. Taking the divergence

$$\nabla^2\phi(\vec{r},t) = \frac{1}{4\pi\varepsilon_0}\int d^3\vec{r}'\left[-\frac{1}{c}\frac{\vec{r}-\vec{r}'}{|\vec{r}-\vec{r}'|^2}\cdot\vec{\nabla}\left(\frac{\partial\rho}{\partial t}\right) - \frac{1}{c}\left(\frac{\partial\rho}{\partial t}\right)\vec{\nabla}\cdot\frac{\vec{r}-\vec{r}'}{|\vec{r}-\vec{r}'|^2}\right.$$
$$\left. -\frac{\vec{r}-\vec{r}'}{|\vec{r}-\vec{r}'|^3}\cdot\vec{\nabla}\rho - \rho\vec{\nabla}\cdot\left(\frac{\vec{r}-\vec{r}'}{|\vec{r}-\vec{r}'|^3}\right)\right]$$
$$= \frac{1}{4\pi\varepsilon_0}\int d^3\vec{r}'\left[-\frac{1}{c}\frac{\vec{r}-\vec{r}'}{|\vec{r}-\vec{r}'|^2}\cdot\left(-\frac{1}{c}\frac{\partial^2\rho}{\partial t^2}\frac{\vec{r}-\vec{r}'}{|\vec{r}-\vec{r}'|}\right) - \frac{1}{c}\left(\frac{\partial\rho}{\partial t}\right)\frac{1}{|\vec{r}-\vec{r}'|^2}\right.$$
$$\left. +\frac{1}{c}\frac{\partial\rho}{\partial t}\frac{1}{|\vec{r}-\vec{r}'|^2} - \rho(4\pi\delta^3(\vec{r}-\vec{r}'))\right] \tag{11.131}$$

Here we have used the fact that

$$\vec{\nabla}\frac{\partial\rho}{\partial t} = -\frac{1}{c}\frac{\partial^2\rho}{\partial t^2}\vec{\nabla}\frac{1}{|\vec{r}-\vec{r}'|} = -\frac{1}{c}\frac{\partial^2\rho}{\partial t^2}\frac{\vec{r}-\vec{r}'}{|\vec{r}-\vec{r}'|^2}$$

and

$$\vec{\nabla}\cdot\left(\frac{\vec{r}-\vec{r}'}{|\vec{r}-\vec{r}'|^2}\right) = \frac{1}{|\vec{r}-\vec{r}'|^2}$$

Also, we already know that

$$\vec{\nabla}\cdot\left(\frac{\vec{r}-\vec{r}'}{|\vec{r}-\vec{r}'|^3}\right) = 4\pi\delta^3(\vec{r}-\vec{r}')$$

Putting it all together, we notice that the two terms proportional to $\frac{\partial\rho}{\partial t}$ cancel in Eq. (11.131). We finally get, using Eq. (11.127)

$$\nabla^2\phi(\vec{r},t) = \frac{1}{4\pi\varepsilon_0}\int d^3\vec{r}'\left[\frac{1}{c^2}\frac{\partial^2\rho}{\partial t^2}\frac{1}{|\vec{r}-\vec{r}'|} - 4\pi\rho\delta^3(\vec{r}-\vec{r}')\right]$$
$$= \frac{1}{c^2}\frac{\partial^2\phi}{\partial t^2} - \frac{1}{\varepsilon_0}\rho(\vec{r},t) \tag{11.132}$$

Thus, we see that this choice of ρ satisfies the Eq. (11.125).

We can carry out a similar exercise for $\vec{A}$ in Eq. (11.126) and get

$$\vec{A}(\vec{r}, t) = \frac{\mu_0}{4\pi} \int d^3\vec{r}' \frac{\vec{j}(\vec{r}', t')}{|\vec{r} - \vec{r}'|} \tag{11.133}$$

We now have solutions ϕ and $\vec{A}$ for the inhomogeneous wave equations. We next use these to determine the electric and magnetic fields for some simple cases.

11.4.2 Field due to a Charged Particle in Motion

Consider a particle of charge e in motion along a trajectory $\vec{r} = \vec{r}_p(t)$. For such a particle, we can write the charge density as

$$\rho(\vec{r}, t) = e\delta^3(\vec{r} - \vec{r}_p(t)) \tag{11.134}$$

With this charge density, we can write the potential ϕ as in Eq. (11.127)

$$\phi(\vec{r}, t) = \frac{e}{4\pi\varepsilon_0} \int d^3\vec{r}' \frac{\delta^3(\vec{r}' - \vec{r}_p(t'))}{|\vec{r} - \vec{r}'|} \tag{11.135}$$

where as always $t' = t - \frac{|\vec{r} - \vec{r}'(t')|}{c}$. The integration in Eq. (11.135) is not straightforward, since we the argument of the Delta function is NOT $\vec{r}'$. So we need to change variables so that the integration can be done easily with the Delta function. This will involve finding the Jacobian for the change of variables. Let us denote

$$\vec{R} = \vec{r}' - \vec{r}_p(t')$$

Note that $r_p(t')$ is a function of r' since it is a function of t', which is a function of r'. Let

$$J_{ij} \equiv \frac{\partial R_i}{\partial r'_j}$$

where R_i is the i^{th} component of $\vec{R}$ and $i = 1, 2, 3$. Now if we differentiate $\vec{R}$ w.r.t $\vec{r}'$, we get

$$\begin{aligned} J_{ij} &= \frac{\partial(r'_i - r_{pi}(t')}{\partial r'_j} \\ &= \delta_{ij} - \frac{\partial r_{pi}(t')}{\partial t'} \frac{\partial t'}{\partial r'_j} \\ &= \delta_{ij} - v_i \frac{-1}{c} \frac{\partial(|\vec{r} - \vec{r}'|)}{\partial r'_j} \\ &= \delta_{ij} - \beta_i \frac{r_j - r'_j}{|\vec{r} - \vec{r}'|} \\ &= \delta_{ij} - \beta_i \frac{r_j - r_{pj}}{|\vec{r} - \vec{r}_p(t' = t - \frac{|\vec{r} - \vec{r}_1|}{c})|} \end{aligned} \tag{11.136}$$

We have used the fact that $\vec{v} = \frac{\partial \vec{r}_p}{\partial t}$ and that the differentiation with time t' is the same as that with time t since $t' = t - \frac{|\vec{r} - \vec{r}'|}{c}$ and $|\vec{r} - \vec{r}'|$ is independent of t. Also $\vec{\beta} = \frac{\vec{v}}{c}$. Finally, in the last line, we have put $\vec{r}' = \vec{r}_p$ because the numerator in Eq. (11.135) has a Delta function which forces this.

We now simplify our notation and use

$$\hat{n} = \frac{\vec{r} - \vec{r}_p}{|\vec{r} - \vec{r}_p|}$$

where $\vec{r}_p$ is evaluated at $t' = t - \frac{|\vec{r}-\vec{r}_p|}{c}$. Then in terms of $\hat{n}$, Eq. (11.136) becomes

$$J_{ij} = \delta_{ij} - \vec{\beta} \cdot \hat{n} \tag{11.137}$$

We still need to evaluate the integral in Eq. (11.135). To do this, we use the usual rules for integration of Delta functions and get

$$\phi(\vec{r}, t) = \frac{e}{4\pi\varepsilon_0} \frac{1}{\det|J||\vec{r} - \vec{r}_p|} \tag{11.138}$$

To evaluate $\det|J|$, we define a quantity

$$b_{ij} = \beta_i n_j$$

Then

$$\begin{aligned} \det|J| &= (1 - b_{11})[(1 - b_{22})(1 - b_{33}) - b_{23}b_{32}] + b_{12}[b_{21}(1 - b_{33}) - b_{23}b_{31}] - b_{13}[b_{21}b_{32} - (1 - b_{22})b_{31}] \\ &= 1 - \vec{\beta} \cdot \hat{n} \end{aligned} \tag{11.139}$$

The scalar potential thus, for the charged particle in uniform motion is given by

$$\phi(\vec{r}, t) = \frac{e}{4\pi\varepsilon_0} \frac{1}{|\vec{r} - \vec{r}_p(t_r)|} \frac{1}{(1 - \vec{\beta} \cdot \hat{n})_{t_r}} \tag{11.140}$$

where t_r is the retarded time

$$t_r = t - \frac{|\vec{r} - \vec{r}_p(t_r)|}{c}$$

In an analogous way, we can determine the vector potential $\vec{A}$ as

$$\vec{A}(\vec{r}, t) = \frac{\mu_0}{4\pi} \frac{ec\vec{\beta}}{|\vec{r} - \vec{r}_p(t_r)|(1 - \vec{\beta} \cdot \hat{n})} = \frac{1}{c}\vec{\beta}\phi(\vec{r}, t) \tag{11.141}$$

The $\phi(\vec{r}, t)$ and $\vec{A}(\vec{r}, t)$ are called **Lienard-Wiechert potentials**.

The relations between the potentials and the fields can then be used to calculate the electric and magnetic fields produced by a charged particle in motion. However, in the case of the particle being in uniform motion, the fields can be calculated far more easily by using transformation equations which we shall encounter in the next Chapter. In the case of accelerated particles, the fields which are obtained using Eqs. (11.140 and 11.141) show a new phenomenon. We shall consider this next. We shall however, be confining our discussion to the case where the accelerated charge is instantaneously at rest. The more general case can be obtained by transformations.

PROBLEM 11.4 An electron is moving with a constant speed v in a circular path of radius R in the $x - y$ plane. The centre of the circle is the origin of coordinates. Calculate the Lienard-Wiechert potentials at a point a at the point $(0, 0, z)$.

PROBLEM 11.5 An electron is moving with a speed $v = c/3$ along the x-direction and at time $t = 0$ it is at the origin O. A point P is at a distance of 1 km from the origin such that OP makes an angle of $\frac{\pi}{4}$ with the x-axis. How much will the error be if at $t = 0$, the scalar potential is calculated using the usual non-relativistic formula.

11.4.3 Field due to an Accelerated Charge

We now consider the case of an accelerated charge which is instantaneously at rest. This implies that the velocity of the charge vanishes while the acceleration is non-zero. ($\vec{\beta} = 0$, $\dot{\vec{\beta}} \neq 0$). Let us first consider the magnetic field. The magnetic field $\vec{B}$ can be obtained from the vector potential in Eq. (11.141). Note that β depends on r through t_r, the retarded time. We have the vector potential,

$$\vec{A}(\vec{r}, t) = \frac{\mu_0}{4\pi} \frac{ec\vec{\beta}}{|\vec{r} - \vec{r}_p(t_r)|(1 - \vec{\beta} \cdot \hat{n})} = \frac{1}{c}\vec{\beta}\phi(\vec{r}, t)$$

which will give us $\vec{B}$ by

$$\vec{B} = \vec{\nabla} \times \vec{A}$$

Note that in the limit, $\beta \to 0$, the only term that will survive in taking the curl of the vector potential will be the one that differentiates β. Thus, we need $\vec{\nabla} \times \vec{\beta}$. To do this, it is simplest to use the index notation as before. We write

$$\begin{aligned}
\left(\vec{\nabla} \times \vec{\beta}\right)_i &= \varepsilon_{ijk} \frac{\partial \beta_k}{\partial r_j} \\
&= \varepsilon_{ijk} \partial_j \beta_k \\
\partial_j \beta_k &= \frac{\partial \beta_k}{\partial t_r} \frac{\partial t_r}{\partial r^j} \\
&= -\frac{n_j}{c} \dot{\beta}_k
\end{aligned} \tag{11.142}$$

where

$$\hat{n} = \frac{\vec{r} - \vec{r}_p(t_r)}{|\vec{r} - \vec{r}_p(t_r)|} \equiv \frac{\vec{R}}{|\vec{R}|}$$

Thus,

$$\vec{B} = \vec{\nabla} \times \vec{A} = -\frac{\mu_0 ec}{4\pi} \hat{n} \times \dot{\vec{\beta}}$$

In a similar fashion, we can determine the electric field $\vec{E}$.

$$\begin{aligned}
\vec{E} &= -\vec{\nabla}\phi - \frac{\partial \vec{A}}{\partial t} \\
&= -\left(\frac{e}{4\pi\varepsilon_0}\right)\left[\vec{\nabla}\frac{(\vec{\beta} \cdot \hat{n})}{R}\right] - \frac{\mu_0 ec}{4\pi R}\dot{\vec{\beta}}
\end{aligned}$$

$$= \left(\frac{e}{4\pi\varepsilon_0}\right)\left[\frac{\hat{n}}{c}\frac{\dot{\vec{\beta}}\cdot\hat{n}}{R}\right] - \frac{\mu_0 ec}{4\pi R}\dot{\vec{\beta}}$$

$$= \left(\frac{e}{4\pi\varepsilon_0}\right)\left[\frac{\vec{R}\times(\vec{R}\times\dot{\vec{\beta}})}{cR^3}\right] \tag{11.143}$$

Thus, in the limit, $\vec{\beta} \to 0$,

$$\vec{B}(\vec{r},t) = \frac{\hat{n}\times\vec{E}(\vec{r},t)}{c} \tag{11.144}$$

Let us consider the region where the point $\vec{r}$ is far away. In this case, $\vec{R} \to \vec{r}$ and the Poynting vector is

$$\vec{S} = \vec{E}\times\vec{H} = \frac{1}{\mu_0 c}(\vec{E}\times(\hat{n}\times\vec{E})) \tag{11.145}$$

Clearly, far away from the particle, the vectors, $\vec{E}$, $\vec{H}$ and $\hat{n}$ are mutually orthogonal. Thus, the Poynting vector is

$$\vec{S} = S\hat{n}$$
$$= \left(\frac{e}{4\pi\varepsilon_0}\right)^2 \frac{\sin^2\theta}{c^5 r^2 \mu_0}|\vec{a}|^2 \tag{11.146}$$

where θ is the angle between $\vec{r}$ and the acceleration $\vec{a} = c\dot{\vec{\beta}}$.

The power P, which is the total energy flowing past a sphere of radius r per unit time is therefore, given by

$$\boxed{\begin{aligned} P &= \int Sr^2 d\Omega \\ &= \frac{e^2\mu_0}{6\pi c}|\vec{a}|^2 \end{aligned}} \tag{11.147}$$

Equation (11.147) is **Larmor's formula for a power radiated by an accelerated charg**e. Though we have derived this result with the assumption that the particle is instantaneously at rest ($\vec{\beta} \to 0$, $\dot{\vec{\beta}} \neq 0$), a similar analysis can be done with a particle having both instantaneous velocity and acceleration. The essential elements, namely the dependence of power on the second power of the charge and the second power of the acceleration remain unchanged.

EXAMPLE 11.2 Using the Lienard-Wierchert fields, discuss the time-averaged power radiated per unit solid angle in non-relativistic motion of a particle with charge e moving
(a) along the z-axis with instantaneous position $z(t) = a\cos\omega_0 t$ and
(b) in a circle of radius R in the $x - y$ plane with a constant angular velocity ω_0.
Determine the total power radiated assuming that the particle is instantaneously at rest.

Solution

Since the particle is instantaneously at rest, the electric field is given by

$$\vec{E} = \frac{e}{c}\left[\frac{\hat{n}\times(\hat{n}\times\dot{\vec{\beta}})}{R}\right]_{ret}$$

From this, we can find the instantaneous power radiated per unit solid angle as

$$\frac{dP}{d\Omega} = \frac{e^2}{4\pi c^3}|\dot{\vec{v}}|^2 \sin^2\Theta$$

where Θ is the angle between $\dot{\vec{v}}$ and $\hat{n}$. Now in our case

$$z(t) = a\cos\omega_0 t$$

and thus,

$$|\dot{\vec{v}}|^2 = (-\omega_0^2 a\cos\omega_0 t)^2 = a^2\omega_0^4\cos^2\omega_0 t$$

The time average

$$< \cos^2\omega_0 t >= \frac{1}{2}$$

and so we get

$$\left\langle\frac{dP}{d\Omega}\right\rangle = \frac{e^2 a^2\omega_0^4\sin^2\Theta}{8\pi c^3}$$

The total power radiated is just the integral over the solid angle of the instantaneous power radiated

$$P = \int\frac{dP}{d\Omega}d\Omega$$

Putting in the expressions gives us

$$P = \frac{e^2 a^2\omega_0^4}{4\pi c^3}\cos^2\omega_0 t(2\pi)\int_0^\pi \sin^2\Theta\sin\Theta d\Theta$$

The integral is easily done as

$$\int_0^\pi \sin^2\Theta\sin\Theta d\Theta = \int_{-1}^1 (1-x^2)dx = \frac{4}{3}$$

and so we get

$$P = \frac{2e^2 a^2\omega_0^4}{3c^3}\cos^2\omega_0 t$$

which is the Larmor formula. For the time averaged total power, we get

$$\langle P\rangle = \frac{e^2 a^2\omega_0^4}{3c^3}$$

(b) We choose coordinates such that the observer's eye is in the x-z plane. Then

$$\hat{n} = \sin\theta\hat{i} + \cos\theta\hat{k}$$

For the circular path

$$\vec{X}(t) = R(\cos\omega_0 t\hat{i} + \sin\omega_0 t\hat{j})$$

and

$$\dot{\vec{\beta}} = \frac{d^2(\vec{X}(t)/c)}{dt^2} = \frac{-\omega_0^2 R}{c}(\cos\omega_0 t\hat{i} + \sin\omega_0 t\hat{j})$$

But

$$\frac{dP}{d\Omega} = \frac{e^2}{4\pi c}|\hat{n}\times(\hat{n}\times\dot{\vec{\beta}})|^2$$

Now

$$\hat{n}\times(\hat{n}\times\dot{\vec{\beta}}) = \hat{n}\times\left[-\sin\theta\sin\omega_0 t\hat{k} - \cos\theta\cos\omega_0 t\hat{j} + \cos\theta\cos\omega_0 t\hat{i}\right]\frac{\omega_0^2 R}{c}$$

and therefore,

$$\hat{n}\times(\hat{n}\times\dot{\vec{\beta}}) = \frac{\omega_0^2 R}{c}\left[\cos^2\theta\cos\omega_0 t\hat{i} + (\cos^2\theta + \sin^2\theta)\sin\omega_0 t\hat{j} - (\sin\theta\cos\theta\cos\omega_0 t)\hat{k}\right]$$

and

$$\frac{dP}{d\Omega} = \frac{e^2\omega_0^4 R^2}{4\pi c^3}\left[\cos^4\theta\cos^2\omega_0 t + \sin^2\omega_0 t + \cos^2\theta\sin^2\theta\cos^2\omega_0 t\right]$$

Taking the time average, we get

$$\left\langle\frac{dP}{d\Omega}\right\rangle = \frac{e^2\omega_0^4 R^2(1+\cos^2\theta)}{8\pi c^3}$$

since the time average of $\cos^2$ and $\sin^2 = \frac{1}{2}$. Now the total power is just the integral over solid angles, which we know is

$$\int d\Omega(1+\cos^2\theta) = 2\pi\int_{-1}^{1} dx(1+x^2) = \frac{16\pi}{3}$$

and so we get

$$\langle P\rangle = \frac{2e^2\omega_0^4 R^2}{3c^3}$$

PROBLEM 11.6 An electron travelling with a speed $v \ll c$ is brought to rest in time T by a constant deceleration. Assuming that during the deceleration period the speed was zero, calculate the total electromagnetic energy radiated during the time interval T.

11.5 OSCILLATING CHARGES AND CURRENTS—DIPOLE RADIATION

In the previous section we considered how to get the electric and magnetic fields due to time dependent sources (charges and currents). In this section, we shall study the case of oscillating charges and currents and see what kinds of electromagnetic fields are generated.

We shall restrict our attention to sources whose time dependence is sinusoidal. In the language of complex variables that we have used before, the charge and current densities then can be written as

$$\vec{j}(\vec{r}, t) = \vec{j}(\vec{r})e^{-i\omega t} \tag{11.148}$$

and

$$\rho(\vec{r}, t) = \rho(\vec{r})e^{-i\omega t} \tag{11.149}$$

As before, the physical charge and current densities are the real parts of the complex quantities. Thus,

$$\vec{j}_r = \mathrm{Re}\,(\vec{j}(\vec{r}, t))$$
$$\rho_r = \mathrm{Re}\,(\rho(\vec{r}, t))$$

A little care is needed before using the form of time dependence given above in Eqs. (11.148 and 11.149). The space integral of $\phi(\vec{r}, t)$ at any given time t is the total charge of the distribution which obviously has a time independent value. Integrating Eq. (11.149) however, leads to a value showing the exponential time dependence. We will thus, use the form given in Eq. (11.149) only when integrals over the charge density with some other factor are involved. In the language of multipole moments that we encountered while discussing the solution to the electric potential, only when we consider higher order moments (and not the monopole moment,which is the integral of the charge density alone) will be take the above time dependent form of ϕ.

Substituting these charge and current densities in Eqs. (11.127) and (11.133), we get

$$\phi(\vec{r}, t) = \frac{1}{4\pi\varepsilon_0}\int d^3r'\,\frac{\rho(\vec{r}', t')}{|\vec{r} - \vec{r}'(t')|} \tag{11.150}$$

and

$$\vec{A}(\vec{r}, t) = \frac{\mu_0}{4\pi}\int d^3r'\,\frac{\vec{j}(\vec{r}', t')}{|\vec{r} - \vec{r}'(t')|} \tag{11.151}$$

where $t' = t - \frac{|\vec{r}-\vec{r}'|}{c}$ is as usual the retarded time.

Equations (11.150) and (11.151) allow us in principle, to calculate the fields at any point at any time for any given charge and current distribution. For a class of situations where the oscillating currents and charges as given in Eqs. (11.148) and (11.149), are restricted in space over a limited region and one is interested in obtaining the fields at points far away from that region, approximate solutions can be obtained analytically. In more general cases, it is not always possible to obtain analytical solutions to the problem.

Let us consider one such case. Notice that the integrals in Eqs. (11.150) and (11.151) are over $\vec{j}(\vec{r}', t')$ and $\phi(\vec{r}', t')$. Now t' is a function of **both** $\vec{r}$ and $\vec{r}'$. But $\vec{r}'$ is restricted in the integrals in Eqs. (11.150) and (11.151) to the source regions since the charge or current densities are vanishing outside some finite region. Thus, for $\vec{r}$ such that $|\vec{r}| \gg |\vec{r}'|$ we can make the following approximation

$$|\vec{r} - \vec{r}'| \sim \vec{r} - \frac{\vec{r}\cdot\vec{r}'}{|\vec{r}|} \tag{11.152}$$

In this case therefore, the exponential factor $e^{-i\omega t'}$ in the integrals will become

$$e^{-i\omega t'} = e^{(ikr-\omega t)}e^{-i\vec{k}\cdot\vec{r}'} \tag{11.153}$$

where $\vec{k} = \frac{\omega\hat{n}}{c}$ and $\hat{n} = \frac{\vec{r}}{|\vec{r}|}$. Also notice the $|\vec{r}'|$ is restricted to d, the size of the source distribution. We will consider values of k which are much larger than $\frac{1}{|\vec{r}|}$. In this case, we have

$$|\vec{r} - \vec{r}'| \sim r - \frac{\vec{r} \cdot \vec{r}'}{|\vec{r}|} = r\left(1 - \frac{\hat{n} \cdot \vec{r}'}{|\vec{r}|}\right) \sim r \tag{11.154}$$

With these approximations, one can write the expressions for the potentials ϕ and $\vec{A}$ as

$$\phi(\vec{r}, t) = \frac{1}{4\pi\varepsilon_0}\frac{e^{ikr}}{r}\int d^3r'\phi(\vec{r}', t)e^{-i\vec{k}\cdot\vec{r}'} \tag{11.155}$$

and

$$\vec{A}(\vec{r}, t) = \frac{\mu_0}{4\pi}\frac{e^{ikr}}{r}\int d^3r'\vec{j}(\vec{r}', t)e^{-i\vec{k}\cdot\vec{r}'} \tag{11.156}$$

In these expressions for the potentials, notice that whenever the form of the charge and current densities given in Eqs. (11.148) and (11.149) can be used, the time dependent exponential factor in them will combine the exponential factor outside the integrals in Eqs. (11.155) and (11.156) to give an overall factor $e^{(ikr-i\omega t)}$. The factor $\frac{e^{ikr-i\omega t}}{r}$ is characteristic of a spherical wave emanating from the origin that is inside the region of the sources $\vec{j}(r)$ and $\rho(r)$ of size d. The wave however, is not isotropic since inside the integral the factor $e^{-i\vec{k}\cdot\vec{r}'}$ depends on the direction of r through $\vec{k}$ as we saw above.

We shall treat kd as a small parameter. In terms of wavelength λ, we know that $k = \frac{2\pi}{\lambda}$ and therefore, $(kd) \ll 1$ implies that the size of the sources is small as compared to the wavelength. We expand the exponential as

$$e^{-i\vec{k}\cdot\vec{r}'} = \sum_{n=0}^{\infty}\frac{(-i\vec{k}\cdot\vec{r}')^n}{n!} \tag{11.157}$$

When Eq. (11.157) is used in Eqs. (11.155) and (11.156), we get an infinite series. The various terms in the series are referred to as the monopole, dipole, quadrupole terms since they involve these moments of the charge and current distribution as in the Multipole expansion of the potential in Chapter 3.

We shall only be discussing the lowest term in the series though this will reveal some general properties of the whole expansion.

The lowest term for the potential, ϕ in Eq. (11.155) is the $n = 0$ term. This is

$$\phi(\vec{r}, t) = \frac{1}{4\pi\varepsilon_0}\int d^3r'\frac{\rho(\vec{r}', t)}{r} = \frac{1}{4\pi\varepsilon_0}\frac{q}{r} \tag{11.158}$$

Thus, the lowest term for the scalar potential gives us a static field which is like that of a charge. This is expected since remember we are dealing with distances much larger than the dimensions of the charge distribution and thus the distribution appears like a point charge.

For the vector potential, the $n = 0$ term in Eq. (11.156) gives us

$$\vec{A}(\vec{r}, t) = \frac{\mu_0}{4\pi}\frac{e^{ikr-\omega t}}{r}\int d^3r'\vec{j}(\vec{r}', t) \tag{11.159}$$

We can convert the integral over the current density $\vec{j}$ in Eq. (11.159) into one over the charge density using the equation of continuity.

$$\vec{\nabla}' \cdot \vec{j}(\vec{r}') + \frac{\partial \rho(\vec{r}', t)}{\partial t} = 0 \tag{11.160}$$

But we know the time dependence of the charge and current densities since these are assumed to be varying sinusoidally with time. Thus, this equation reduces to

$$(i\omega)\rho(\vec{r}') = \vec{\nabla}' \cdot \vec{j}(\vec{r}') \tag{11.161}$$

We now use integration by parts to write

$$\int d^3r' \vec{j}(\vec{r}') = -\int d^3r' \vec{r}'(\vec{\nabla}' \cdot \vec{j}) \tag{11.162}$$

and using Eq. (11.161) we get

$$\vec{A}(\vec{r}, t) = -\frac{\mu_0}{4\pi}\frac{e^{ikr-i\omega t}}{r}\int d^3r'(i\omega)\vec{r}'\rho(\vec{r}') \tag{11.163}$$

We can rewrite this in terms of the electric dipole moment which was defined as

$$\vec{p} = \int d^3r' \vec{r}' \rho(\vec{r}')$$

The vector potential thus, becomes

$$\vec{A}(\vec{r}, t) = -\frac{\mu_0}{4\pi}\frac{e^{ikr-i\omega t}}{r}(i\omega)\vec{p} \tag{11.164}$$

The fields which arise from this term in the expansion are thus, referred to as the **electric dipole radiation**.

We next try to compute the fields resulting from this potential. The magnetic field $\vec{B}$ is

$$\begin{aligned}\vec{B}(\vec{r}, t) &= \vec{\nabla} \times \vec{A}(\vec{r}, t) \\ &= \frac{\mu_0}{4\pi}\frac{e^{ikr-i\omega t}}{r}\left[1 - \frac{1}{ikr}\right](-\omega)(\vec{k} \times \vec{p}) \\ &\sim \frac{\mu_0}{4\pi}\frac{e^{ikr-i\omega t}}{r}(-\omega)(\vec{k} \times \vec{p})\end{aligned} \tag{11.165}$$

We have dropped the $\frac{1}{ikr}$ term in the final expression since we are considering only far away from the sources, i.e., at points where $\frac{1}{kr} \ll 1$. **This region is called the 'radiation zone'.**

What about the electric field? We can obtain the electric field easily without any direct evaluation. This is because in deriving Eqs. (11.155) and (11.156), we have used the Lorentz gauge condition. In this case, the scalar and vector potentials are related by

$$\frac{1}{c^2}\frac{\partial \phi(\vec{r}, t)}{\partial t} + \vec{\nabla} \cdot \vec{A}(\vec{r}, t) = 0 \tag{11.166}$$

But the time dependence of the scalar and vector potentials are both sinusoidal. Hence, differentiating with respect to t just brings down a factor of $(-i\omega)$ and we get

$$\frac{-i\omega}{c^2}\phi(\vec{r}, t) + \vec{\nabla} \cdot \vec{A}(\vec{r}, t) = 0 \tag{11.167}$$

or

$$\phi(\vec{r}, t) = -\frac{ic^2}{\omega}\vec{\nabla} \cdot \vec{A} \tag{11.168}$$

But

$$\vec{E}(\vec{r},t) = -\vec{\nabla}\phi(\vec{r},t) - \frac{\partial \vec{A}(\vec{r},t)}{\partial t} \tag{11.169}$$

or

$$\vec{E}(\vec{r},t) = \frac{ic^2}{\omega}\vec{\nabla}(\vec{\nabla}\cdot\vec{A}(\vec{r},t)) + i\omega\vec{A}(\vec{r},t) \tag{11.170}$$

As we saw above, in the radiation zone, a differentiation with respect to r is equivalent to multiplying by (ik). Specifically, if we differentiate with respect to the i^{th} component of r, i.e., r_i, we get a factor of ik_i. This implies that we can write $(-k^2\vec{A}(\vec{r},t))$ as $\nabla^2\vec{A}(\vec{r},t)$. Thus, Eq. (11.170) becomes

$$\begin{aligned}
\vec{E}(\vec{r},t) &= \frac{ic^2}{\omega}\vec{\nabla}(\vec{\nabla}\cdot\vec{A}(\vec{r},t)) + \nabla^2\vec{A}(\vec{r},t) \\
&= \frac{ic^2}{\omega}\vec{\nabla}\times\vec{\nabla}\times\vec{A}(\vec{r},t) \\
&= \frac{ic^2}{\omega}\vec{\nabla}\times\vec{B}(\vec{r},t) \\
&= -c\hat{n}\times\vec{B}(\vec{r},t)
\end{aligned} \tag{11.171}$$

where we have used the vector identity

$$\vec{\nabla}\times\vec{\nabla}\times\vec{A} = \vec{\nabla}(\vec{\nabla}\cdot\vec{A}) - \vec{\nabla}^2\vec{A}$$

and also the fact that $c = \frac{\omega}{k}$. We have also replaced the operator $\vec{\nabla}$ acting on $\vec{A}$ in the radiation zone by $i\vec{k}$.

We see therefore, that the vectors $\vec{E}$, $\vec{B}$ and $\hat{n}$ are mutually orthogonal in the radiation zone. The Poynting vector $\vec{S}$ can be easily determined as

$$\begin{aligned}
\vec{S} &= \vec{E}_r \times \vec{H}_r \\
&= -\frac{c}{\mu_0}\left[\hat{n}\times \mathrm{Re}\,\vec{B}\right]\times \mathrm{Re}\,\vec{B} \\
&= \frac{c}{\mu_0}|\mathrm{Re}\,\vec{B}|^2\hat{n} \\
&= S\hat{n}
\end{aligned} \tag{11.172}$$

Substituting the value of $\vec{B}$ from Eq. (11.165), we get

$$S = \frac{\mu_0 p^2\omega^4}{16\pi^2 c}\cos^2(kr-\omega t)\frac{\sin^2\theta}{r^2} \tag{11.173}$$

where θ is the angle between $\hat{n}$ and $\vec{p}$.

From Eq. (11.173) we see that S oscillates with time. The time average of $\cos^2(kr-\omega t)$ is $\frac{1}{2}$ and hence, the time averaged S is

$$< S >= \frac{\mu_0 p^2\omega^4}{32\pi^2 c}\frac{\sin^2\theta}{r^2} \tag{11.174}$$

Notice the typical dipole dependence on the angle θ. This came about because we used the lowest power of $(\vec{k}\cdot\hat{n})$. Higher powers of this quantity that will come in while evaluating higher multipoles will thus involve higher powers both of k and $\cos(\theta)$. This would in turn, give rise to higher powers of ω and $\cos(\theta)$ than in Eq. (11.173) and involve higher multipole moments.

The total power radiated out for the electric dipole fields that we are considering is given by integrating $(\vec{S}\cdot\hat{n})$ over the surface of sphere of radius r. We get for time averaged power $< P >$

$$< P >= \int (< \vec{S} > \cdot\hat{n}) r^2 d\Omega = \frac{\mu_0 p^2 \omega^4}{12\pi c} \tag{11.175}$$

As an example of dipole radiation, let us consider a simple case of a centrefed linear antenna. The antenna is of length d which is much smaller than the wavelength of the wave. (Typically, radio waves are of order $1-10$ meters) The antenna is oriented along the z axis from $z=\frac{d}{2}$ to $z=-\frac{d}{2}$. The current is fed from a small gap in the centre and is in the same direction in each half of the antenna. The maximum value of the current is I_0 at the centre and falls linearly to zero at the ends. Thus,

$$I(z)e^{-i\omega t} = I_0\left(1-\frac{2|z|}{d}\right)e^{-i\omega t} \tag{11.176}$$

But the continuity equation gives us

$$i\omega\rho = \vec{\nabla}\cdot\vec{j}$$

and hence the linear charge density ρ is constant along the arms and is given by

$$\rho(z) = \pm\frac{2iI_0}{\omega d} \tag{11.177}$$

The dipole moment is parallel to the z axis and is given by

$$p = \int_{-d/2}^{d/2} z\rho(z)dz = \frac{iI_0 d}{2\omega} \tag{11.178}$$

The total power radiated is given by Eq. (11.175) and since it is proportional to p^2, we see that the quantity which determines the power output in this approximation is $kd=\frac{2\pi}{\lambda}d$. For fixed input current, the power radiated increases as ω^2.

PROBLEM 11.7 Figure 11.3 is an insulated ring of radius R in the $x-z$ plane. Only two quadrants of the ring carry charge: the first quadrant with a line charge density λ and the third one with line charge density $-\lambda$. If the ring is rotated around the y-axis with an angular speed ω, estimate the amount of power of the electromagnetic radiation.

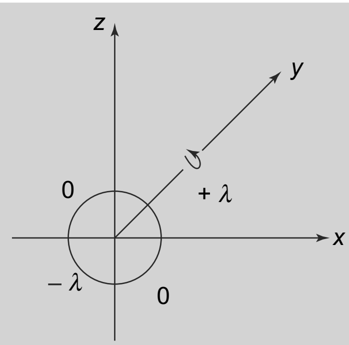


Fig. 11.3 *Problem 11.7*

SUMMARY

- In the steady state, Maxwell's equations in a conductor are identical to that in a dielectric except that the divergence of $\vec{D}$ vanishes since there is no free charge density.
- Maxwell's equations in a conducting medium allow for propagating wave solutions. In these waves, the longitudinal component of the fields plays no role.
- The transverse electric and magnetic fields in the propagating wave in a conducting medium are attenuated as they travel. They are also out of phase with each other.
- Electromagnetic waves in a conducting media have a skin depth or a distance over which the amplitude gets attenuated by $\frac{1}{e}$. The skin depth depends on the frequency of the wave and the conductivity of the material.
- The presence of conducting surfaces even when the electromagnetic waves travel in a dielectric or vacuum, impose certain boundary conditions on the field vectors of the wave.
- The boundary conditions imposed by the presence of conductors imply there are three kinds of modes—Transverse Electric and Magnetic (TEM), Transverse Electric (TE) and Transverse Magnetic (TM) in such a situation.
- For an infinite, hollow pipe with conducting surfaces, only TE and TM modes are allowed.
- For a rectangular cross section of the infinite hollow pipe or wave guide, TE_{00} mode is not allowed. The lowest TM mode is higher than the lowest TE mode.
- For a rectangular cross section of the infinite hollow pipe or wave guide, TE_{00} mode is not allowed. The lowest TM mode is higher than the lowest TE mode.
- We can use the gauge freedom to define electromagnetic potentials ϕ and $\vec{A}$ which satisfy inhomogeneous wave equations.
- The formal solutions of these inhomogeneous equations in terms of the charge and current densities for a charge in motion gives us Lienard Wierchert potentials which can be used to determine the electric and magnetic fields.
- An oscillating charge gives rise to electromagnetic radiation

CONCEPTUAL QUESTIONS

1. The dominant mode in a rectangular waveguide is

a. TE_{11}
b. TE_{10}
c. TM_{11}
d. TE_{00}

2. In a rectangular waveguide, $a = 2b$. The cutoff frequency for TE_{02} mode is 12 GHz. The cutoff frequency for TM_{11} mode is

a. 3 GHz
b. $3\sqrt{5}$ GHz
c. 12 GHz

d. $6\sqrt{5}$ GHz

3. An empty rectangular wave guide with $a = 2$ cm and $b = 1$ cm operates at 10 GHz. Determine the mode of propagation in the waveguide.

4. What is the ratio of skin depth in copper at 1000 Hz to that at 100 MHz?

5. For the TE_{30} mode, which of the following filed components exist?

a. E_x
b. E_y
c. E_z
d. H_x
e. H_x

6. An electron oscillates in a simple harmonic potential with an angular frequency of $\omega \approx 10^{15}$ rad/sec and amplitude $A = 10^{-8}$ cm. Find the energy radiated per cycle.

7. For the electron in Question 6, how long will the electron take to radiate half its energy?

8. Radiation emitted by an antenna has the angular distribution characteristic of dipole radiation when

a. the wavelength is long compared to the dimensions of the antenna
b. the wavelength is small compared to the dimensions of the antenna
c. when the antenna has a specific shape
d. all of the above.

9. An antenna at an airport has a maximum dimension of 3 m and operates at 100 MHz. An aircraft landing at the airport is 0.5 km away. The aircraft is

a. In the far field zone of the antenna
b. Not in the far field zone
c. Cannot say from the information given.

10. A radio station radiates power of 50 kW at 100 MHz. Give an estimate of the electric field at a distance of 10 km from the radio station.

PROBLEMS

1. A 2 cm-square waveguide operates at 12 GHz in the TM_{11} mode. Determine the cutoff frequency.

2. Calculate the dimensions of an air filled rectangular waveguide for which the cut-off frequency for the TM_{11} and TE_{03} modes are both equal to 12 GHz.

3. In an air-filled rectangular waveguide, the cut-off frequency of a TE_{10} mode is 5 GHz and the cut-off frequency for the TE_{01} mode is 12 GHz. Find the dimensions of the waveguide and the cut-off frequency for the next higher TE mode.

4. An air filled rectangular waveguide has a TE mode operating at 6 GHz for which

$$E_y = 5\sin\left(\frac{2\pi x}{a}\right)\cos\left(\frac{\pi y}{b}\right)\sin(\omega t - 12z)\ \text{V/m}$$

Find the mode of operation and the cutoff frequency.

5. The phase and group velocity of the TE_{mn} mode in a wave guide are given by

$$v_{ph} = \frac{\omega}{k} = \frac{c}{\sqrt{1 - \frac{f_c^2}{f^2}}}$$

and

$$v_g = c\left(\sqrt{1 - \frac{f_c^2}{f^2}}\right)$$

where f_c is the cutoff frequency. In a rectangular wave guide of dimensions $a = 2b = 2$ cm and length 10 cm, two signals in the form of wave packets are traveling, both in the TE_{10} mode. The central frequency of the first signal is 12 GHz and of the second one is is 10 GHz. Of the two waves, find the time required for the faster one to travel from one end to the other.

6. What would be limiting value of the transverse size of a square wave guide that would transmit a wave of wavelength λ_0 in the TE_{10} mode but not in the TE_{11} or TM_{11} modes?

7. A signal consists of frequencies in the vicinity of a frequency f_1 in the form of a Gaussian with width σ. It propagates along a rectangular waveguide as a TE_{10} mode. If the width of the signal is increased from time to time, show that its shape as it travels will keep on changing.

8. Calculate the total power radiated by a dipole of length 5 m at a frequency of 400 kHz when the rms value of the current at the centre of the dipole is 2 A.

9. If the real and imaginary parts of the permittivity of a particular material are $\varepsilon = \varepsilon_0(1.1(1+i))$, and $\mu = 2\mu_0$, what would be the skin depth of the material at an angular frequency ω?

10. An electric dipole of length 50 cm is in free space. If the maximum current is 25 A and frequency 10 MHz, determine the electric and magnetic fields in the far-field (radiation) zone.

11. An atom with atomic polarisability $\alpha(\omega)$ is placed in an electric field

$$\vec{E} = E_0 e^{i(kx-\omega t)}$$

Find the electric and magnetic fields in the radiation zone and also the energy radiated per unit solid angle.

12. Two equal point charges $+Q$ oscillate along the z axis with their displacements given by

$$z_1 = z_0\sin(\omega t) \quad z_2 = -z_0\sin(\omega t) \quad x_1 = x_2 = y_1 = y_2 = 0$$

The radiation field is observed at a point $\vec{r}$ where $r \gg \lambda \gg z_0$. Find the electric and magnetic field at $\vec{r}$. Also determine the total radiated power and compare it with that for dipole radiation.

13. A charged particle of charge q and mass m is released from rest and falls in a gravitational field. Determine the fraction of its potential energy lost as radiation as a function of the distance it falls.

14. Bohr's model for the hydrogen atom assumes electrons circling the proton in circular orbits being held in orbit by the Coulomb force between the nucleus and the electron. Suppose the radius of the orbit is a. How long will it take for the electron to fall into the nucleus if we assume that the velocity is non-relativistic, i.e., $v \ll c$?

15. A charged particle approaches a conducting plane. The charge and its image form a dipole and hence, radiation is emitted due to the changing dipole moment. Calculate the total radiated power as a function of the distance of the charge from the conducting plane.

12

Relativity and Electrodynamics

Learning Objectives

- To review the concept of Galilean Relativity.
- To comprehend the need for a change in Galilean transformations because of Michleson–Morley experiment.
- To learn about the Postulates of Special Relativity.
- To understand Lorentz transformations between inertial frames and show that they are lead to the constancy of the speed of light.
- To study the consequences of Lorentz transformations namely simultaneity, causality, time dilation, length contraction and law of addition of velocities.
- To learn about how to extend the concept of momentum and energy to relativity.
- To comprehend the consequences of the new definitions of energy in relativity.
- To study the transformation properties of electric and magnetic fields under Lorentz transformations.
- To learn about the invariant combinations of electric and magnetic fields.

The Theory of Relativity of Einstein has been hailed as one of the greatest intellectual achievements of humankind. Relativity profoundly changed the way we understand space, time and motion. Human beings have conjectured about space, time and motion ever since we know. However, the meaning and understanding of these concepts changed drastically with the propounding of Theory of Relativity by Einstein. It is not that relativity was invented by Einstein—Galileo and Newton had already put the science of mechanics on a sound, mathematical footing which was consistent with all known observations. We first discuss the concept of relativity as it was understood before Einstein's seminal achievement.

12.1 GALILEAN RELATIVITY

The concept of relativity for Galileo and Newton was intrinsically linked with motion. Any discussion of motion must start with an understanding of space and time. Ever since Aristotle, both space and time were considered absolute. Another requirement for describing motion is a choice of a frame of reference. This, as we understand intuitively is basically the local surroundings with respect to which motion, or a change in position is defined. Thus, for instance, we say that a car is moving with respect to the stationary ground. Obviously, the choice of which frame of reference to use to describe motion is a matter of convenience in everyday life.

An important class of frames of reference are the inertial frames. These are the ones in which Newton's laws of mechanics are valid. Of course, with this definition of inertial frames, the question of whether a frame is inertial or not becomes a matter of observation and experiment. Given the importance of inertial frames in Newtonian mechanics, we need to have a way of determining which frames are inertial independent of Newton's laws. This is not a very easy task, though a good working definition is the frame attached to the earth or in uniform motion with respect to the 'fixed' distant stars. However, the crucial thing is that once we agree on an inertial frame, then all other frames which are moving with a constant velocity with respect to such a frame are also inertial. This becomes clear when one considers the transformation of the coordinates of a point particle between two such frames.

Consider an inertial frame S with its Cartesian system of coordinates (x, y, z) and another frame S' with its own coordinates, (x', y', z'), as shown in Fig. 12.1. One assumption which is central to Newtonian physics is that time is universal and absolute. This implies that any interval of time has the same magnitude in all inertial frames. Specifically, with their clocks synchronised with identical zero of time, any event which is observed in S and S' occurs at the same time for both of them. Assume that at time $t = 0$, the origins of both the frames coincide and their axes overlap. The frame S' is moving with respect to S with a velocity v along the x axis.

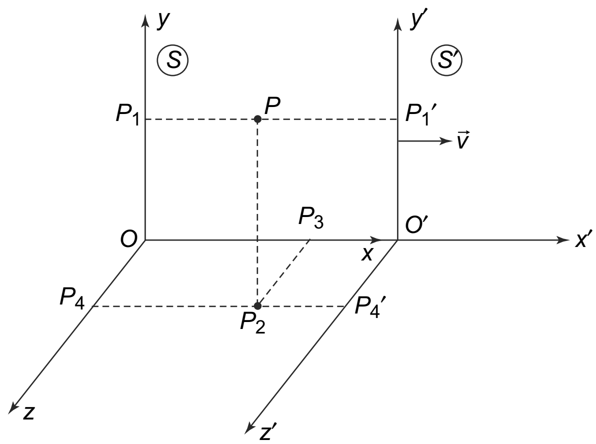


Fig. 12.1 *Two inertial frames S and S' with their axes parallel. The frames coincide when $t = 0$. S' is moving with a velocity v w.r.t. S along the x direction. PP_1 is the perpendicular drawn from a point P to the y axis, PP_2 is perpendicular to the $x - z$ plane while P_2P_3 and P_2P_4 are perpendicular to the x and z axes respectively*

O', the origin of S' with coordinates (0, 0, 0), at time t when observed from S has coordinates given by $(vt, 0, 0)$.

Consider now an 'event' which happens at a point P in S at some time t. This event could be the decay of a particle, the position of a particle at that time, etc. **Basically anything which needs a specification of the three spatial and one time coordinate for its description is called an event.** The coordinates of this event in S are therefore, given by

$$\begin{aligned} \text{Time} &= t \\ x\text{-coordinate} &= OP_3 = x \\ y\text{-coordinate} &= PP_1 = y \\ z\text{-coordinate} &= OP_4 = z \end{aligned} \tag{12.1}$$

The same event happens in S' at a time t'. However, by our assumption of absolute time, $t = t'$. However, the spatial coordinates of the event in S' are different and are given by

$$\begin{aligned} \text{Time } t' &= t \\ x'\text{-coordinate} &= O'X = OO' - OX = x' = x - vt \\ y'\text{-coordinate} &= PP_1' = y' = y \\ z'\text{-coordinate} &= O'P_4' = z' = z \end{aligned} \tag{12.2}$$

as is clear by simple geometry. This set of relations in Eqs. (12.1 and 12.2) can be summarised as

$$\boxed{\begin{aligned} t' &= t \\ x' &= x - vt \\ y' &= y \\ z' &= z \end{aligned}} \tag{12.3}$$

The above relations are called **Galilean transformations** between inertial frames.

We have considered a particle at rest. If the particle is in motion, then clearly both $\vec{r}(x, y, z)$ and $\vec{r}'(x', y', z')$ are time dependent. However, from Eq. (12.3) is follows that

$$\frac{d^2\vec{r}}{dt^2} = \frac{d^2\vec{r'}}{dt^2}$$

since the term vt has vanishing second derivative with respect to time. Now a particle that has no forces acting on it in S, has by Newton's Laws

$$\frac{d^2\vec{r}(t)}{dt^2} = 0$$

and as argued above, also has

$$\frac{d^2\vec{r'}(t')}{dt'^2} = 0$$

This proves what we anyways started out with as the definition of inertial frames. That Newton's First law is valid in both the inertial frames.

Now consider two particles of masses m_1 and m_2 with coordinates $\vec{r}_1$ and $\vec{r}_2$ exerting a force on each other. If the force is translation invariant, then the potential function, $V(\vec{r}_1, \vec{r}_2)$ in S will have the form $V(\vec{r}_1, \vec{r}_2) = V(\vec{r}_1 - \vec{r}_2)$, that is the potential is a function only of the displacement between the two. If the potential function is the same in S', i.e., $V(\vec{r}) = V(\vec{r'})$ then from Eq. (12.3) it follows that

$$V'(\vec{r'}_1 - \vec{r'}_2) = V(\vec{r'}_1 - \vec{r'}_2) = V(\vec{r}_1 - \vec{r}_2)$$

This means that Newton's Second Law in S reads

$$m_1 \frac{d^2\vec{r}_1}{dt^2} = -m_2 \frac{d^2\vec{r}_2}{dt^2} = -\frac{\partial V(\vec{r}_1 - \vec{r}_2)}{\partial \vec{r}_1} \tag{12.4}$$

Using the result above and using the transformations between the two coordinates, we get in S'

$$m_1 \frac{d^2\vec{r'}_1}{dt'^2} = -m_2 \frac{d^2\vec{r'}_2}{dt'^2} = -\frac{\partial V'(\vec{r'}_1 - \vec{r'}_2)}{\partial \vec{r'}_1} \tag{12.5}$$

We see therefore that Newton's Second law is also valid in both the inertial frames, provided we relate them using the transformation of coordiantes given in Eq. (12.3).

An important consequence of the Galilean transformations is that the velocities, unlike the accelerations, in the two inertial frames are not equal. If $\vec{u} = \frac{d\vec{r}}{dt}$ is the velocity in S and $\vec{u}' = \frac{d\vec{r'}}{dt'}$ is the velocity of a particle in S' then, in our case of the two frames in relative motion along the x axis, it follows that

$$\begin{aligned} (\vec{u}')_x &= (\vec{u})_x - (\vec{v})_x \\ (\vec{u}')_y &= (\vec{u})_y \\ (\vec{u}')_z &= (\vec{u})_z \end{aligned} \tag{12.6}$$

where $\vec{v} = (v, 0, 0)$. Thus, we see that the component of the velocity along the direction of relative motion between the two frames transforms by an amount equal to the relative velocity of the frames while the orthogonal components of the velocity are unaffected. This is the commonplace effect that we notice when observing moving objects from a moving car or train. Though we have considered velocities of particles, it is clear that the same logic applies to travelling waves. This is the origin of the well known Doppler effect for sound waves where the wave with a velocity c_w in S and c'_w in S' are related by

$$\vec{c}'_w = \vec{c}_w - \vec{v} \tag{12.7}$$

12.2 LORENTZ TRANSFORMATIONS

We have been discussing the laws of electromagnetism in the previous chapters. However, nowhere did we mention the frames of reference in which these laws would be valid. In addition, after Maxwell's discovery of the displacement current and its leading naturally to the concept of electromagnetic waves, there arose another problem. Recall that electromagnetic waves in free space travel with a velocity c which is related to the electrical and magnetic properties of free space by

$$c = \frac{1}{\sqrt{\mu_0 \varepsilon_0}}$$

In nature, one was familiar with waves much before the concept of electromagnetic waves came in. But all these waves, like, for example sound waves, required a material medium to propagate in. The speed of these waves could thus, be specified with respect to the undisturbed medium. Clearly, it seemed conceptually necessary after Maxwell to think of an all-pervading medium called 'aether' in which these waves would propagate. The velocity of light that we found would then be the velocity in a frame in which the 'aether' was at rest. The electromagnetic wave itself (the variation of electric and magnetic fields as we saw) would be thought of as vibrations of the aether. This is what the situation was till the latter half of the nineteenth century.

The first serious attempt to test the 'aether' hypothesis was the famous Michleson–Morley experiment. Michelson and Morley attempted to compare the speed of light along two perpendicular directions. The basic idea was that since the earth is in motion in some particular direction relative to the 'aether' the two velocities should be different since by the argument given above, the velocities along the motion of the earth should either add or subtract while those in the perpendicular direction should not change. We shall not discuss the experiment in detail here but the experiment attempted using a clever arrangement to observe this difference using optical interferometry even though the expected difference in the velocities was very tiny.

What Michelson and Morley found was that there was no significant difference in the two velocities. Given the sensitivity of their equipment, this put a limit on v_{aether}, the velocity of the earth through the 'aether' at less than 8 km/second. This is to be compared to the velocity of light which at that time was thought to be around 250, 000 km/sec. Since the time of Michelson and Morley, this experiment has been refined enormously. Thus, for example, Muller et al. (**Phys. Rev. Lett., 91, 020401, 2003**) performed the experiment with a cryogenic optical resonator and found that difference in the two velocities in different directions, Δc as a ratio to the speed of light is

$$\frac{\Delta c}{c} \leq (2.6 \pm 1.7) \times 10^{-15}$$

This is many orders of magnitude better than the result obtained by Michelson and Morley in 1887. A review of the status of the experimental verification of the constancy of the velocity of light till the 1960s is given in **R.S. Shankland, Amer. Journ. of Physics, 32,16 (1964)**.

This was the state of affairs when Einstein, in 1905 wrote the first of his celebrated papers, titled, 'On the Electrodynamics of Moving Bodies'. In this paper, he questioned the validity of Galilean relativity and also the existence of the 'aether' as a medium required for the propagation of electromagnetic waves. As an aside, it is interesting to note that the original paper of Einstein was in German and the first English translation was made available in 1920 by the famous Indian astrophysicist M.N. Saha!.

In this famous paper, Einstein proposed two postulates which were to be the basis of the Theory of Special Relativity.

Postulate 1. The laws of physics are the same in all inertial frames.
Postulate 2. The speed of light is the same in all inertial frames.

It is interesting to note that in this paper, Einstein makes no reference to the Michelson–Morley result. In fact, his postulating the constancy of the speed of light was motivated by the necessity of synchronising clocks separated by some distance.

One consequence of the postulates is that Galilean transformations that we discussed above (Eq. (12.3)) are not consistent with the Second Postulate. The correct transformations which are consistent with this postulate of constancy of speed of light were worked out by Einstein and are given by (again for the case of S and S' in motion relative to each other along the x axis)

$$\begin{aligned} x' &= \gamma(x - vt) \\ y' &= y \end{aligned}$$

$$\boxed{\begin{aligned} z' &= z \\ t' &= \gamma\left(t - \frac{vx}{c^2}\right) \end{aligned}} \tag{12.8}$$

Here c is the speed of light in vacuum and the quantity γ is defined as

$$\gamma = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}$$

These transformation equations are called **Lorentz transformations.** Although we have given the transformations for the case when the frame S' is in motion along the x axis, for motion along a general direction with a velocity $\vec{v} = v\hat{n}$, the coordinate component along the direction of motion $(\vec{r} \cdot \hat{n})$ will change like the x transformation above while the orthogonal components $(\vec{r} - \vec{r} \cdot \hat{n}\hat{n}\ )$ will not change. The time transformation above will have instead of $\frac{vx}{c^2}$, $\frac{v\vec{r}\cdot\hat{n}}{c^2}$. The important thing to notice is that these transformations bring about a fundamental change in our conception of time. Time is no longer, as was supposed for millenia, absolute in nature. Rather, it was now a relative quantity like space and would change from one inertial frame to another.

We can see that the Lorentz transformations, will indeed make the speed of light in vacuum independent of the inertial frame. To see this, consider two points in S which are a distance Δx apart. A beam of light sent from one point to another in S takes a time

$$\Delta t = \frac{\Delta x}{c}$$

Consider now observing the process of the beam travelling between the two points as observed in S'. The distance $\Delta x'$ between the first point from where the beam is sent at a certain time and the point where it is received after a time interval $\Delta t'$ is given by

$$\Delta x' = \gamma\,(\Delta x - v\Delta t) = \gamma\left(\Delta x - \frac{v}{c}\Delta x\right) \tag{12.9}$$

But in S' the time $\Delta t'$ is also different and is given by

$$\Delta t' = \gamma\left(\Delta t - \frac{v\Delta x}{c^2}\right) = \frac{\gamma}{c}\left(\Delta x - \frac{v}{c}\Delta x\right) \tag{12.10}$$

Thus, $\Delta t' = \frac{\Delta x'}{c}$ thereby implying that the speed of the light beam in S' is also c.

EXAMPLE 12.1 A meter stick at rest is placed at an angle of 45° with the x-axis in its rest frame. The stick moves with a velocity $\vec{v} = \frac{c}{2}\hat{i}$. Find the length of the meter stick as observed from the laboratory frame.

Solution

Let the coordinates of the rod AB of length l aligned at an angle of θ with the x-axis, in its rest frame be

$$x'_A = 0,\ y'_A = 0$$

$$x'_B = l\cos\theta,\ y'_B = l\sin\theta$$

In the lab frame, let the coordinates be x_A, y_A, x_B, y_B. The primed and the unprimed coordinates are related by a Lorentz transformation

$$x'_A = \gamma(x_A - vt)$$
$$x'_B = \gamma(x_B - vt)$$
$$y'_A = y_A$$
$$y'_B = y_B$$

which gives us

$$\begin{aligned} x'_B - x'_A &= l\cos\theta \\ &= \gamma(x_B - x_A) \\ x_B - x_A &= \frac{l\cos\theta}{\gamma} \\ y_B - y_A &= l\sin\theta \end{aligned}$$

Therefore, the length of the moving rod as observed from the laboratory frame is given by

$$\begin{aligned} l' &= \sqrt{(x_B - x_A)^2 + (y_b - y_A)^2} \\ &= \sqrt{\frac{l^2\cos^2\theta}{\gamma^2} + l^2\sin^2\theta} \\ &= l\sqrt{1 - \frac{v^2}{c^2}\cos^2\theta} \end{aligned}$$

But $l = 1$, $v = \frac{c}{2}$ and $\theta = 45°$. Putting it all in we get

$$l' = \frac{1}{2}\sqrt{\frac{7}{2}}$$

EXAMPLE 12.2 A tube of length L is filled with a transparent dielectric of refractive index n. The tube is at rest in a particular inertial frame S along the x-axis. In S, a light signal from one end to the other takes 1 sec to reach. As observed in a frame S' moving relative to S with a speed $v = \frac{3c}{5}$ along the x-axis, how long does the signal along the x direction take to reach from one end of the tube to the other?

Solution

Let us first consider the frame S.

In this frame, let the x coordinates of the ends be x_1 and x_2. Let the time the signal be sent at time t_1 from the end x_1. Then the time the signal is received at x_2 is

$$t_2 = t_1 + \frac{Ln}{c}$$

since the velocity of light in the transparent dielectric will be $\frac{c}{n}$.

Now let us consider the frame S'. In this, the coordinates of the ends of the tube and the times at which the signal starts from one end and is received at the other end are related to x_1, x_2, t_1, t_2 by Lorentz transformations. These can be easily found to be

$$x'_1 = \gamma(x_1 - vt_1)$$
$$t'_1 = \gamma(t_1 - \frac{vx_1}{c^2})$$

$$x_2' = \gamma(x_2 - vt_2)$$
$$t_2' = \gamma\left(t_2 - \frac{vx_2}{c^2}\right)$$

Hence, the time taken by the signal in S' to reach from one end to the other is

$$\begin{aligned} t_2' - t_1' &= \frac{5}{4}\left[(t_2 - t_1) - \frac{v}{c^2}(x_2 - x_1)\right] \\ &= \frac{5}{4}\left[\frac{Ln}{c} - \frac{3L}{5c}\right] \\ &= \frac{5L}{4c}\left(n - \frac{3}{5}\right) \\ &= \frac{5}{4}\left(1 - \frac{3}{5n}\right) \end{aligned}$$

since $\frac{Ln}{c} = 1$.

12.3 CONSEQUENCES OF LORENTZ TRANSFORMATIONS

Lorentz transformation in Eq. (12.8) change in a very fundamental ways some of our well established beliefs and notions about space and time. One is always conditioned to think of time as something that is absolute. This of course, as we have mentioned, is in-built into Newtonian physics. Einsteinian Relativity forces us to rethink these notions.

The first thing to notice is that Lorentz transformations make both the space coordinates as well as time relative quantities, i.e., they are dependent on the inertial frame that one chooses to measure them in. However, in any one inertial frame, one can always synchronise clocks. To see this, consider some inertial frame and two points A and B in it. Let the distance between A and B be r_{AB}. Now if a light pulse is sent from A at time t_A as measured by the clock at A then it will be received at B at a time $t_A + \frac{r_{AB}}{c}$. Thus, the clock at B can be adjusted to this value and then the clocks at A and B will be synchronised. With synchronisation done, in any particular inertial frame the time variable measured at all points is the same. **In a given frame, two events occurring at the same time are called simultaneous**.

12.3.1 Relativity of Simultaneity

Simultaneity of events is one of the things we have a common sense understanding of. This understanding and sense is mostly based on Newtonian concept of absolute time and we shall see that with time itself becoming relative, the concept of what is simultaneous changes profoundly. Consider the common example of a lightning flash and the sound of thunder. We know that during a thunderstorm, we see the lightning flash first and then after some time hear the clap of thunder. Since there is a time delay in our observing of these events, we naturally say that these are not simultaneous events. What we mean is that the two events, our seeing the flash and our hearing the report, as seen by our clocks were not simultaneous.

But now consider another observer, also in our frame of reference but separated from us and sitting in the clouds where the lightning and thunder are originating. For this observer, the two events, the lightning flash and the sound of thunder happen simultaneously! Thus, it would seem that simultaneity is a relative concept **even** within a single frame. However, this is not true as will be clear if we think about which two events one is considering.

For the observer in the clouds, the two events are indeed, the lightning flash and the sound of thunder happening in the cloud and these certainly are seen at the same instant. For the observer on the ground, the two events that he considers not happening at the same time are, one, the lightning flash arriving at his location and two, the sound report arriving at his location. These are not simultaneous. If the observer on the ground were to consider the same two events as observed by observer in the cloud, namely the flash and the report originating, he too will agree that these were simultaneous. Thus, in deciding whether two events are simultaneous, one should be careful as to what exactly are these two events.

Let us consider another example which will illustrate this fact. Imagine a street with street lights $A, B, C, \cdots$ in a line separated by a distance L from each other, as shown in Fig. 12.2. The lights have clocks attached to them which are all synchronised. Suppose that the lights are all programmed to go off at sunrise, say at 6 am. An early morning walker when he passes the light at A at 6 am will see that the light A goes off at 6 am, the light B goes off at a time which is $\frac{L}{c}$ later, the light at C goes off at time $\frac{2L}{c}$ later than 6 am etc. For this observer at A using the clock at A the lights going off are NOT simultaneous. The events that he is referring to are the events of signals from the various lights reaching him not to the lights going off at their respective locations. However, if the same observer were to see what time was recorded at the location of all the lights, i.e. at $B, C, \cdots$, he will notice that all of them indeed record that the lights go off at 6 am precisely. Thus, the events of the lights going off as recorded by clocks at the location of the events are simultaneous.

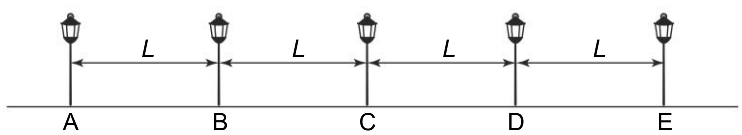


Fig. 12.2 *Street lights with synchronised clocks. The lights are separated by a distance L along a line*

We can summarise as follows. Any event is characterised by a space coordinate $\vec{r}$ corresponding to where it is happening and a time coordinate t indicating when it is happening, in any inertial frame. We repeat: t is the time when the event happens at $\vec{r}$, which may not be the time when that event is observed elsewhere in that frame. Two such events in any frame are called simultaneous if they happen at the same time irrespective of their location in that frame.

What about the concept of simultaneity for different inertial observers? Consider an observer in a frame S' moving with a speed v w.r.t to the stationary street lights along the x direction or the direction of the street. We take the time $t = t' = 0$ at 6 am and assume that the observer in S' is at the street light A at $t' = 0$.

In our original frame S which is stationary with respect to the lights, the coordinates of the event of lights getting switched off are $(x = x_A, t = t_A)$ for A, $(x = x_B, t = t_B)$ for B and so on. Clearly, $t_A = t_B = t_C = \cdots = 0$ since all of them are programmed to go off at 6 am, which we have chosen to be the zero of our time. The same event for an observer in S' will have coordinates $(x' = x'_A, t' = t'_A)$

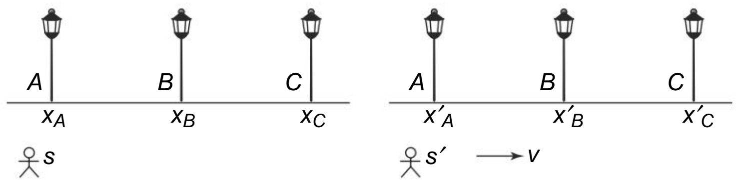


Fig. 12.3 *Observation of lights by two observers, one in S and one in S′ moving with a speed v with respect to the lights*

for A, $(x' = x'_B, t' = t'_B)$ for B and so on. These are related to the coordinates in S by the Lorentz transformations

$$\begin{aligned} x' &= \gamma(x - vt) \\ t' &= \gamma\left(t - \frac{vx}{c^2}\right) \end{aligned} \tag{12.11}$$

Thus,

$$\begin{aligned} t'_A &= \gamma\left(t_A - \frac{vx_A}{c^2}\right) = -\frac{\gamma v x_A}{c^2} \\ t'_B &= \gamma\left(t_B - \frac{vx_B}{c^2}\right) = -\frac{\gamma v x_B}{c^2} \end{aligned}$$

and so on. Clearly, $t'_A \neq t'_B \neq t'_C$, etc. The two events in S' are therefore, NOT simultaneous.

Thus, whereas simultaneity has an absolute meaning for observers in a single inertial frame, i.e., two events are simultaneous or not is the same for all observers in a single inertial frame, this is not true when we consider observers in different inertial frames. $\Delta t = 0$ in a frame S does not automatically imply that $\Delta t' = 0$ in S'.

EXAMPLE 12.3 The space-time coordinates of an event can be written as (x, y, z, t). Let the coordinates of the an event E_1 be $(a, 0, 0, a)$ and the coordinates of another event E_2 be $(3a, 0, 0, 2a)$. Show that you can always make a Lorentz transformation to an inertial frame where the events E_1 and E_2 occur at the same time. What will be the velocity of the inertial frame where the events are simultaneous? Is it possible to make a Lorentz transformation where the two events occur at the same position?

Solution

Let's make a Lorentz transformation from S to S' which is moving with velocity v in x-direction relative to S.

The coordinates of E_1 are

$$(x_1, y_1, z_1, ct_1) = (a, 0, 0, ac)$$

and the coordinates of E_2 are

$$(x_2, y_2, z_2, ct_2) = (3a, 0, 0, 2ac)$$

in S.

The spatial separation of the two events is $\Delta x = x_2 - x_1 = 2a$ and the temporal separation of the two events is $\Delta t = (t_2 - t_1) = ac$ in S.

Now, the separations in S' will be

$$\Delta x' = \gamma\,(\Delta x - v\Delta t)$$
$$\Delta t' = \gamma\left(\Delta t - \frac{v\Delta x}{c^2}\right)$$

Since $\Delta t > 0$ in S, the event E_2 follows event E_1 in S. However, since $\Delta t'$ can be made zero in S, the events E_2 and E_1 can be simultaneous in S. For this, we need to set

$$\Delta t = \frac{v\Delta x}{c^2} \tag{12.12}$$

or

$$v = \frac{c^2\Delta t}{\Delta x} \tag{12.13}$$

If $\Delta x < c\Delta t$, $v > c$ which is not possible. In our case, $\Delta x > c\Delta t$. So, the required velocity is physical. Its value is $v = \frac{c}{2}$ in our case.

It is easy to see that we cannot make $\Delta x' = 0$ as this will require $v > c$, which is not possible.

PROBLEM 12.1 Person A sees two events as simultaneous in his reference frame. These events are

Event 1 occurring at the coordinates $(0, 0, 0)$ at time $t = 0$

Event 2 occurring at the coordinates $(500\ m, 0, 0)$

Person B moves past person A along the x axis with a velocity 0.999 c $\hat{i}$. Which event occurred first in Person B's frame?

12.3.2 Time Dilation

Consider a clock at rest in the frame S. Suppose we measure an interval of time between two events in S by this clock; these two events could, for example, be the minute hand reaching 1 and then 2—an interval of five minutes in S. Since the clock is at rest in S, its space coordinates do not change. Thus, $\Delta x = \Delta y = \Delta z = 0$. Now, suppose we measure the same interval of time in the moving frame S'. The clocks at rest in S' will measure the interval as $\Delta t'$. But

$$\begin{aligned}\Delta t' &= \gamma\left(\Delta t - \frac{\Delta x v}{c^2}\right)\\ &= \gamma\Delta t\\ &> \ \Delta t\end{aligned} \tag{12.14}$$

This is the phenomenon of **Time Dilation.** An interval of time, say an hour recorded by a clock in S appears as longer than an hour as recorded by the clock at rest in the frame moving w.r.t. S.

Time dilation is verified experimentally in various experiments. A very striking one is the observation of subatomic particles called μ-mesons produced by cosmic rays. Cosmic rays from outer space impinge on the earth's atmosphere and collide and interact with the molecules in the atmosphere, some 6 kilometres above sea level. One of the particles produced in this interaction is a μ-meson. A μ-meson is an unstable particle and it is known that it has a half life of about 2.2×10^{-6} seconds. In this time, the particle decays into other particles. Now since the maximum speed with which the μ-mesons can travel would be $c = 3.0 \times 10^{10}$ cm/s, one expects that all the μ-mesons produced at the height of 6 kilometers would have decayed by the time they reach the laboratory at sea level since the maximum distance they can travel in their lifetime is $2.2 \times 10^{-6} \times 3.0 \times 10^{8} = 660$ m. However, when one detects these particles in the laboratory, one finds considerably less attenuation in their flux which means that the numbers being detected is a sizable and definitely not insignificant fraction of those being produced at the top of the atmosphere, at a height of 6 kilometres.

This paradoxical situation admits a ready explanation in terms of time dilation. The μ-mesons are produced with velocities very close to the velocity of light. In this case, the γ factor in Equation (12.14) can be large. Thus, for instance for a μ-meson with $v = 0.99c$, $\gamma \sim 7$. This implies that the lifetime of the μ-meson as observed by the observer in the laboratory (which is moving with respect to the μ-meson) is 7 times more than the lifetime observed by an observer sitting on the μ-meson. Thus, at this velocity, the distance that the μ-meson can travel before decaying becomes ~ 4 kilometres. A sizable fraction of the flux thus may be expected at sea level, some six kilometers below the top of the atmosphere.

A popular paradox in relativity, known as the twin paradox is also connected with the phenomenon of time dilation. The paradox can be illustrated as follows.

Consider two twins, say **J** and **M** who are at rest in the two inertial frames S and S' respectively. S' is moving with a speed v along the x axis. At $t = t' = 0$, the twins celebrate their 60^{th} birthday in their respective frames crossing each other at $x = x' = 0$.

At this time, that is at $t = t' = 0$, **J** sees **M** fly off for five years, turn back and return to him at $t = 10$ years. Thus, at the time of reunion now, **J** is 70 years old. However, this interval of time, starting at $t = 0$ and ending at $t = 10$ years corresponds to $t' = 0$ and $t' = \gamma t$. Since $\gamma > 1$, and hence $t' < t$, this means that the **M** sitting in S' would have aged less than 10 years.

So far, there is no paradox. However, the reasoning above which led to **M** being younger after his trip is based on assuming that **J** is at rest and **M** is moving. From the point of view of **M** however, the situation is just the opposite. He sees himself at rest and **J** moving away and returning after 10 years. By the same reasoning, **J** should be younger than **M** after the round trip! Both these statements cannot be correct.

The solution to this paradox lies in realising that the situation of **J** and **M** are not symmetrical. The key lies in realising that our analysis of time dilation which results from Lorentz transformation is valid only for inertial frames. In the twin paradox, in the frame of **J**, **M** must have experienced some acceleration, no matter how small, when he turned back after 5 years. This makes S' a non-inertial frame from the point of view of **J**. Since Lorentz transformations is between inertial frames, we cannot use the argument above. Velocity is relative but acceleration with respect to an inertial frame is an absolute concept. And accelerated or non-inertial frames need concepts beyond special relativity (specifically Einstein's Theory of General Relativity) to be understood and analysed.

PROBLEM 12.2 A clock moves along the x-axis at a speed of 0.3c and reads zero as it passes the origin. What time does it read (to an observer stationary w.r.t the clock) as it passes the 90 m mark on the axis? What is the time difference with a stationary clock synchronised at the origin?

PROBLEM 12.3 An unstable high-energy particle enters a detector and leaves a track 1 mm long before it decays. Its speed relative to the detector was 0.99 c. What is its proper lifetime? That is, how long would it have lasted before decay had it been at rest with respect to the detector?

12.3.3 Length Contraction

Another consequence of the Lorentz transformation is the phenomenon of length contraction. Consider a rod at rest along the x axis in a frame S. If x_1 and x_2 are the coordinates of the two ends of the rod measured in S, (not necessarily at the same time since the rod is at rest in S), then the length of the rod L_0, called the rest length since the rod is at rest in S is given by

$$L_0 = \Delta x = x_2 - x_1 \tag{12.15}$$

Now consider measuring the length of the rod in a frame S' which is as usual moving with a speed v along the x axis w.r.t S. Then the location of one end x'_1 is measured at some time t' and the location of the other end, x'_2 is measured at the same time t'. The difference $x'_2 - x'_1 = \Delta x' = L$ is the length of the rod as measured in S'. Since we are measuring the two end points, the two measurements in S' would correspond to measurements in S made at two different times giving the location of the end points. Since the rod is at rest in S it does not matter if the the measurements of the end points were made at different times since it would give us $L_0 = \Delta x = x_2 - x_1$. But the Lorentz transformations relate the two as

$$\begin{aligned} \Delta x' &= x'_2 - x'_1 = \gamma(\Delta x - \Delta t v) \\ \Delta t' &= t'_2 - t'_1 = \gamma\left(\Delta t - \frac{v\Delta x}{c^2}\right) \end{aligned} \tag{12.16}$$

But $\Delta t' = 0$ and hence,

$$\Delta t = \frac{v\Delta x}{c^2}$$

while $\Delta x = L_0$. Thus, we get

$$\begin{aligned} L &= \Delta x' \\ &= \gamma\left[\Delta x - v\left(\frac{v\Delta x}{c^2}\right)\right] \\ &= \gamma[L_0 - \frac{v^2}{c^2}L_0] \\ &= \sqrt{1 - \frac{v^2}{c^2}}L_0 \end{aligned} \tag{12.17}$$

Since the square root factor is less than unity, we can say that the length of the rod would be smaller in a frame when it is in motion than the length at rest.

Note that it is the length in the direction of motion which is contracted. That is, in our case, the x direction. The lengths in the perpendicular directions, namely y and z in this case are unaffected and are the same when measured in the two frames. This is because the Lorentz transformations (Eq. (12.8)) show that the coordinates in the directions orthogonal to the direction of motion are not changed.

One may think that some more sophisticated measurement of the length might get around this. This is not the case. Consider measuring the length by reflecting light from a mirror fixed to one end of the rod with a light source at the other end of the rod. An observer in the frame S in which the rod is at rest, sends out a light pulse from one end, say A and the light is reflected from the mirror at other end, say B and travels back to the A where it is detected. This is shown in Fig. 12.4. The two events are: the sending of the light pulse from A at time $t = t_1 = 0$ from the point $x = 0$ and the arrival of the light pulse at A after reflection from B at time $t = t_3$ and $x = 0$. The space coordinates of both these events are the same and hence $\Delta x = 0$. The time difference between the two events is

$$\Delta t = t_3 - t_1 = t_3 = \frac{2L_0}{c}$$

Since c is known, one can measure Δt and determine the length L_0 in this frame in which the rod is at rest.

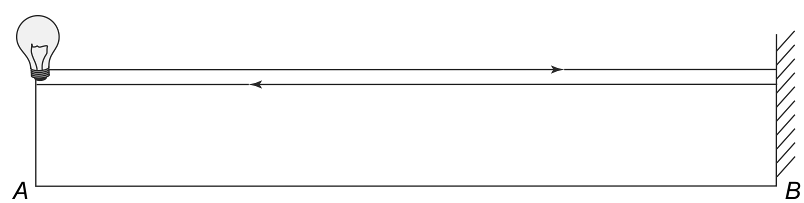


Fig. 12.4 *Measuring lengths using the time taken by a light pulse. Rays from a bulb at A travel to a mirror fixed at the other end at B and are reflected back. The total time taken by the light pulse for the round trip is* $\frac{2L_0}{c}$

Consider now the same measurement as done in the frame S' and let L be the length of the rod measured in S'. The light pulse once again starts at A with coordinates $x'_1 = 0, t'_1 = 0$. The rod is moving in the frame S' in the negative x direction with a speed v. Thus, at the time $t' = t'_2$ when the pulse strikes the mirror at the end B, the end B would have moved a distance vt'_2 from its original location. This motion will be in the -ve x direction. The distance therefore, travelled by the pulse would be less and given by

$$L - vt'_2$$

This is the distance travelled by the pulse in time t'_2 (since $t'_1 = 0$) with a velocity c. Therefore,

$$ct'_2 = L - vt'_2$$

or

$$t'_2 = \frac{L}{c+v} \tag{12.18}$$

After reflection, the pulse is now travelling towards A. It has started at time $t' = t'_2$. It reaches point A with the detector at time t'_3. However, in the time $t'_3 - t'_2$ the rod has travelled a distance $v(t'_3 - t'_2)$ to the left (i.e., the negative x direction). The pulse from B to A thus travels a distance of $L + v(t'_3 - t'_2)$ in a time $t_3' - t_2'$ with a velocity c. Thus,

$$t_3' - t_2' = \frac{L + v(t_3' - t_2')}{c}$$
$$= \frac{L}{c - v}$$

or (12.19)

$$t_3' = \frac{2L}{c(1 - \frac{v^2}{c^2})} \tag{12.20}$$

where we have used Eq. (12.18). This is the total time for the pulse to go from A to B and then back to A. In the frame S, this time was $t_3 = \frac{2L_0}{c}$. These are depicted in Fig. 12.5(a), (b) and (c).

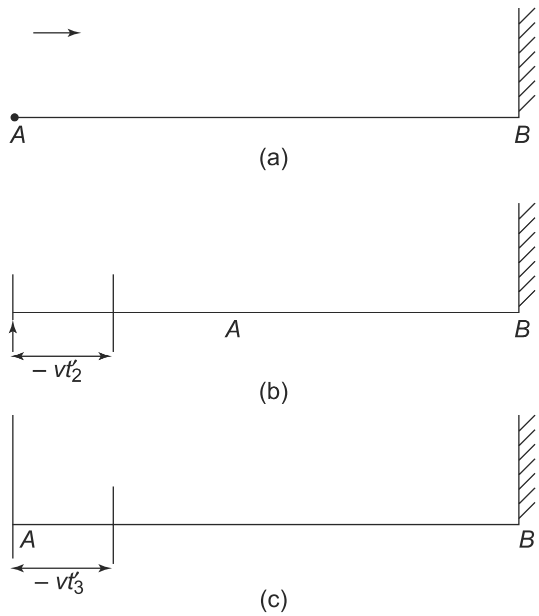


Fig. 12.5 *Measuring length in S' (a) The pulse starts from A at $x' = 0$ and $t' = t_1' = 0$ (b) Pulse arrives at B at $x' = x_2' = -vt'2$ with $t_2' = \frac{x_2'}{c}$ (c) Pulse arrives back at A at $x' = x_3' = -vt_3' - vt'2$ at time $t_3' = t_2' + \frac{L+v(t'3-t_2')}{c} - vt_3'$*

Since these observations were at the same value of x and so by time dilation

$$t_3' = \gamma t_3$$
$$= \gamma \frac{2L_0}{c} \tag{12.21}$$

Using Eqs. (12.21 and 12.19), we get

$$L = \sqrt{1 - \frac{v^2}{c^2}} L_0 \tag{12.22}$$

Although length contraction as such has not been directly verified experimentally but since it is related to time dilation as we saw above, it has been indirectly verified in experiments on time dilation.

EXAMPLE 12.4 A rod of length L is inclined at an angle θ' from the x-axis in an inertial frame of reference S' moving with velocity v in the x-direction w.r.t. another inertial frame S. What will be the angle θ of the rod from the axis in the frame S?

Solution

Let L_x and L_y be the components of the rod along x and y axes in S and $L_{x'}$ and $L_{y'}$ be the components of the rod along x' and y' axes in S'. Then, the angle subtended by the rod on x-axis is

$$\tan\theta = \frac{L_y}{L_x}$$

in S and

$$\tan\theta' = \frac{L_{y'}}{L_{x'}}$$

in S'. Now,

$$L_y = L_{y'}$$

and

$$L_x = L_{x'}\sqrt{1 - \frac{v^2}{c^2}}$$

as there is no length contraction along y direction. Therefore,

$$\tan\theta = \frac{\tan\theta'}{\sqrt{1 - \frac{v^2}{c^2}}}$$

Appearance of a Heavenly Body in Motion

The fact that length contraction applies to only the dimensions parallel to the direction of motion while leaving the directions perpendicular to it unaffected might lead one to believe that a spherical object in motion will appear as ellipsoidal. Interestingly, Einstein himself referred to this in his seminal 1905 paper which first proposed the Theory of Relativity. To quote him, “a rigid body, which measured in a state of rest has the form of a sphere, therefore, has in a state of motion—viewed from the stationary state—the form of an ellipsoid of revolution”.

It took more than 50 years after Einstein's paper for the correct analysis of the problem to be found. The analysis was in a paper by J. Terrel titled, 'Invisibility of Lorentz contraction', **Phys. Rev 116,1041 (1959)**. The basic point is that if the moving object is observed, say by a camera, i.e., the image is recorded on a photographic film or plate, then what is being recorded on the film are rays from the extremities of the object which have arrived simultaneously on the film. These rays are not emitted simultaneously though. Terrel showed that for a small spherical object, what will be recorded or imaged will be approximately spherical. A lucid presentation of this analysis can be found in an article by V. Wiesskopf in **Scientific American, 203, 74 (1960).**

12.3.4 Causality

Our observation of nature is based on certain intuitive concepts of space and time. For instance, the temporal ordering of a cause and effect sequence is a fundamental concept on which we base our understanding. Effects always follow causes in time and never the other way around. Thus, for instance, if we hear a gunshot at some time, we know that the cause of the sound of gunshot, namely the firing of the gun must have preceded our hearing of the sound. Or for instance, one is always born after one's parents and never the other way around. This ordering of cause and effect is ingrained in us and forms the basis of our understanding of nature.

However, relativity has told us that time itself is a relative concept. It is not the same in two inertial frames. A natural question then to ask is whether the temporal ordering of events, specifically the cause and effect sequence is also relative between two inertial frames, where one is moving with a velocity w.r.t. the other?

To see this, consider two events in S which form some cause and effect. Let us say that the cause or event 1 is, in the frame S, at the coordinates (x_1, t_1). We suppress the other two transverse directions in our discussion since they dont play any role. Similarly, let the effect, or event 2 be at (x_2, t_2) in S. Clearly, since these two events form a cause and effect sequence, the time interval, $\Delta t = t_2 - t_1 > 0$. However, these two events happened at two different space coordinates, x_1 and x_2. Thus, some signal must have been transmitted from one point to the other. The signal must have travelled from x_1 at time t_1 to reach x_2 at time t_2. The speed of the signal is thus,

$$|v_{\text{sig}}| = \frac{|x_2 - x_1|}{t_2 - t_1} \tag{12.23}$$

Now let us observe these same two events in another frame S' which is moving with a speed v along the positive x direction w.r.t S. The coordinates of the two events will be (x'_1, t'_1) and (x'_2, t'_2). These are given by

$$\begin{aligned}
x'_1 &= \gamma(x_1 - vt_1) \\
t'_1 &= \gamma\left(t_1 - \frac{vx_1}{c^2}\right) \\
x'_2 &= \gamma(x_2 - vt_2) \\
t'_2 &= \gamma\left(t_2 - \frac{vx_2}{c^2}\right)
\end{aligned} \tag{12.24}$$

Thus,

$$\begin{aligned}\Delta t' &\equiv t_2' - t_1' \\ &= \gamma\left[\Delta t - \frac{v(x_2 - x_1)}{c^2}\right] \\ &= \gamma\,\Delta t\left[1 - \frac{v v_{\text{sig}}}{c^2}\right] \qquad (12.25)\end{aligned}$$

We know that the upper limit of v is c. Thus, if $v_{\text{sig}} < c$ then $\Delta t' = t_2' - t_1' > 0$. This implies that in this case, the cause effect relationship is maintained in the moving frame too.

But what happens, if there are faster than light signals? In that case, $v_{\text{sig}} > c$. For such a case, we chose a relative velocity v between S and S' such that $c > v > \frac{c^2}{v_{sig}}$. We then will have $(t_2' - t_1') < 0$. That means the temporal sequence of events as observed in S' are reversed. Strange things may happen—you can be born before your parents were born as observed in S'. Or you can change history since the temporal ordering is reversed.

Fortunately, till date, signals travelling faster than light are only found in science fiction since no such signal exists in nature. But for the theory of relativity to be consistent, not only can particles not travel faster than light but even signals are restricted similarly.

12.3.5 Law of Addition of Velocities

When a particle in motion is observed from two different inertial frames, we can use the Lorentz transformation equations, (Eq. (12.8)) to obtain the relationship between the velocities of the particle observed by observers in the two frames.

Consider a particle moving along a trajectory $\vec{r} = \vec{r}(t)$ in the frame S. If S' is another frame moving with a velocity v along the x direction with respect to S, then the coordinates $\vec{r}'$ and time t' of the particle in S' are given by

$$\begin{aligned}x'(t') &= \gamma(x(t) - vt) \\ y'(t') &= y(t) \\ z'(t') &= z(t) \\ t' &= \gamma\left(t - \frac{vx(t)}{c^2}\right) \qquad (12.26)\end{aligned}$$

By definition, the velocity of the particle in S is $\vec{v} = \frac{d\vec{r}(t)}{dt}$ and that in S' is $\frac{d\vec{r}'(t')}{dt'}$. Using Eq. (12.26) we get

$$\begin{aligned}v_x' &= \frac{dx'(t')}{dt'} \\ &= \frac{\partial x'}{\partial t}\Big/\frac{\partial t}{\partial t'} \\ &= \gamma\left(\frac{dx}{dt} - v\right)\Big/\gamma\left(1 - \frac{v}{c^2}\frac{dx}{dt}\right) \\ &= \frac{v_x - v}{1 - \frac{vv_x}{c^2}} \qquad (12.27)\end{aligned}$$

$$
\begin{aligned}
v'_y &= \frac{dy'(t')}{dt'} \\
&= \frac{\partial y'}{\partial t} \Big/ \frac{\partial t}{\partial t'} \\
&= \frac{v_y}{\gamma(1 - \frac{vv_x}{c^2})}
\end{aligned}
\tag{12.28}
$$

$$
\begin{aligned}
v'_z &= \frac{dz'(t')}{dt'} \\
&= \frac{\partial z'}{\partial t} \Big/ \frac{\partial t}{\partial t'} \\
&= \frac{v_z}{\gamma\left(1 - \frac{vv_x}{c^2}\right)}
\end{aligned}
\tag{12.29}
$$

Equations (12.27)–(12.29) are the **relativistic law of composition of velocities**. They obviously are very different from their Galilean counterparts that we encountered in Eq. (12.6). We have considered the frame S' moving along the x-axis relative to S in deriving these relations. If S' moves relative to S in a direction given by the unit vector $\hat{n}$ with speed v, then of course the component of $\vec{v}'$ along $\vec{v}' \cdot \hat{n}$, would transform transform like v'_x with v_x in the rhs replaced by $\vec{v} \cdot \hat{n}$. Components of $\vec{v}'$ orthogonal to $\hat{n}$ would transform like Eq. (12.28) or (12.29).

The remarkable thing about the relativistic law for composition of velocities is that while the relative motion between the two frames is in the x-direction, all three components of velocities are affected. However, it is easy to see that in the limit of small values of v and v_x as compared to c such that $\frac{vv_x}{c^2} \ll 1$ and can be neglected, the relativistic laws (Eqs. (12.27)–(12.29)) go over to the corresponding Galilean laws.

The second noteworthy thing is that if the velocity of the object in S, v_x is c, then v'_x is also c since

$$
v'_x = \frac{c - v}{1 - \frac{v}{c}} = c \tag{12.30}
$$

In particular, this implies that a pulse of light which moves with a velocity c in one frame moves with the same velocity in all frames. This is obviously in conformity with the Second Postulate of Einstein that we encountered above.

The relativistic law of addition of velocities also implies that the magnitude of the velocity of any object cannot exceed c in any frame no matter how fast it might be moving towards or away from the object. Consider a particle moving with velocity $(\frac{-c}{2}, 0, 0)$ in frame S. Suppose S' is moving in the positive x direction w.r.t S with a velocity $\frac{9}{10}c$. Now according to Galilean relativity, the speed of the particle in S' would be $(-1.4c, 0, 0)$. However, applying Eq. (12.27), we get

$$
v'_x \simeq 0.99c \tag{12.31}
$$

Thus, we see that $|v'_x| < c$. Since the relativistic law of addition of velocities, Eqs. (12.27)–(12.29) differ substantially from the Galilean law of addition only when both v and v_x are comparable to c, a direct experimental verification of these are difficult to achieve.

EXAMPLE 12.5 An inertial system S' is moving with velocity v relative to another inertial system S in x direction. A ray of light is incident at an angle θ from vertical along xy plane in S so that its velocity is ($c_x = c\sin\theta, c_y = c\cos\theta, c_z = 0$). Find the components of velocity in the moving frame S' and hence, calculate the incident angle θ' of the ray from the vertical direction in S'. Also find the angle θ' for the special case when the ray is incident vertically downwards in S.

Solution

The incident angles of the light ray from vertical direction in S and S' are

$$\tan\theta = \frac{c_x}{c_y}$$

$$\tan\theta' = \frac{c'_x}{c'_y}$$

where

$$c'_x = \frac{c_x - v}{1 - \frac{vc_x}{c^2}}$$

$$c'_y = \frac{c_y}{\gamma(1 - \frac{vc_x}{c^2})}$$

Therefore,

$$\tan\theta' = \frac{c'_x}{c'_y} = \frac{c_x - v}{c_y}\gamma = \gamma\left(\tan\theta - \frac{v}{c\cos\theta}\right)$$

When the light ray incidents from vertical direction, $\theta = 0$ and $\tan\theta' = -\frac{v}{c}\gamma$. Approximately, this means $\theta' = -\frac{v}{c}$.

PROBLEM 12.4 Imagine an electron and a proton are travelling in a particle accelerator in the same direction. The speed of the electron is 0.8 c while that of the proton is 0.7 c both measured by the observer stationary in the laboratory. What is the speed of the proton as measured from the electron?

PROBLEM 12.5 In an inertial frame S, two spacecrafts travel in opposite directions along, straight, parallel trajectories which are separated by a distance d. The speed of each of the spacecraft is $\frac{c}{2}$. The observer in spacecraft 1 has a coordinate system which is parallel to that of S and the direction of motion is along the y axis. At some instant in the frame S, when the spacecraft are at the point of closest approach, the packet is ejected from Spacecraft 1 with a speed $\frac{3c}{4}$ as observed from S. At what angle must the pilot of spacecraft 1 aim the packet so that the packet is received by the pilot of spacecraft 2?

12.4 MOMENTUM AND ENERGY OF PARTICLES

The description and understanding of dynamics of particles in Newtonian mechanics involves not only velocities and positions, but also momentum of particles. The momentum is, in general, a function of coordinates and time. Thus, if we want to study the laws of motion of particles in different frames, we would need to know not only the transformation of coordinates and time, but also of momentum.

In Newtonian mechanics, the momentum of a particle of mass m moving with a velocity $\frac{d\vec{r}}{dt}$ in a frame S is defined as

$$\vec{p} = m\frac{d\vec{r}}{dt} \tag{12.32}$$

If one assumes that this definition is valid in relativity theory as well, then for the particle which was moving with velocity $\frac{d\vec{r}}{dt}$ in S, the momentum in frame S' will be

$$\vec{p}' = m\frac{d\vec{r}'}{dt'} \tag{12.33}$$

This is not quite right. Momentum has another important property in mechanics—in a system of particles interacting with each other via forces that satisfy Newton's third law, the total momentum of the system does not change with time. Unfortunately, momentum as defined by Eq. (12.33) does not lead to this conservation principle. Hence, one has to seek a different definition which conforms to conservation of momentum. Of course, momentum, in analogy with its Newtonian definition, must be a vector with three components and should vanish when the particle is at rest. Otherwise, we will not be able to get the Newtonian definition back in any limit. For a particle, the velocity $\vec{v}$ is the only such vector. Hence, we assume that the definition we are seeking has the form

$$\vec{P} = M(v)\vec{v} \tag{12.34}$$

In the above, we are using $\vec{P}$ to denote the relativistic momentum and $M(v)$ is a function of the magnitude of the velocity that we need to determine. In the Newtonian case, obviously $M(\vec{v})$ is simply the mass of the particle m. The function $M(v)$ is called the **relativistic mass**. We analyse a simple collision phenomenon in the two reference frames and determine the form of $M(v)$.

Consider two particles of equal mass m, approaching each other with equal and opposite velocities: Particle A with velocity $v_A = (v\cos\theta, v\sin\theta, 0)$ and Particle B with velocity $v_B = (-v\cos\theta, -v\sin\theta, 0)$. Their momenta, from Eq. (12.34), respectively are: $P_A = M(v)[v\cos\theta, v\sin\theta, 0]$ and $P_B = M(v)\ [-v\cos\theta, -v\sin\theta, 0]$ so that the total momentum $P = P_A + P_B$ is zero. We shall call the frame where these were the velocities and momentum as the **CM frame (short for Centre of Mass).** The two particles collide and after the collision they emerge not along the original directions but with their y-components of velocities reversed as in Fig. 12.6. The particles thus, emerge with the same speed as initially and their velocities after the collision in the CM frame, denoted by w_A and w_B are:

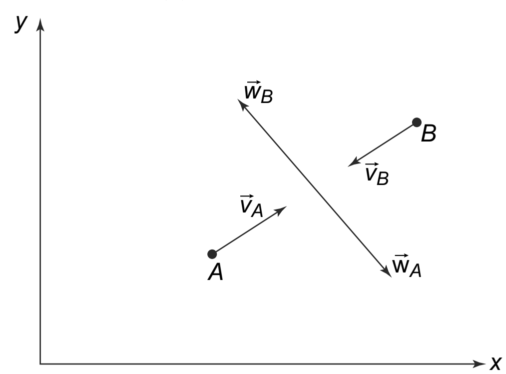


Fig. 12.6 *Collision of two particles*

$$\vec{w_A} = v(\cos\theta, -\sin\theta, 0)$$
$$\vec{w_B} = v(-\cos\theta, \sin\theta, 0) \tag{12.35}$$

Thus, irrespective of the form of $M(v)$, the total momentum is conserved, being zero both before and after the collision.

Consider now the same collision process viewed from another inertial frame which is moving with respect to the with a velocity $V = (v\cos\theta)$ along the x-direction. We will call this the L-frame (short for laboratory frame). The law of transformation of velocities that were obtained can now be used to work out the velocities before and after the collision as viewed in the L-frame. The velocity V of S' relative to S in Eqs. (12.27–12.29) for the present case is $v\cos(\theta)$. They are, using symbols v'_A, v'_B for the velocities before and w'_A, w'_B after the collision in the L-frame

$$\vec{v'_A} = 0\hat{i} + (\frac{v\sin\theta}{\gamma\left(1-\frac{v^2\cos^2\theta}{c^2}\right)}\hat{j} + 0\hat{k} \tag{12.36}$$

$$\vec{v'_B} = \frac{-2v\cos\theta}{1+\frac{v^2\cos^2\theta}{c^2}}\hat{i} + \left(-\frac{v\sin\theta}{\gamma(1+\frac{v^2\cos^2\theta}{c^2})}\right)\hat{j} + 0\hat{k} \tag{12.37}$$

After the collision, in the Centre of Mass frame, the x components remain the same while the y components reverse in sign. Hence, the velocities in the L frame after the collision are

$$\vec{w'_A} = 0\hat{i} + \left(\frac{-v\sin\theta}{\gamma(1-\frac{v^2\cos^2\theta}{c^2})}\right)\hat{j} + 0\hat{k} \tag{12.38}$$

and

$$\vec{w'_B} = \frac{-2v\cos\theta}{1+\frac{v^2\cos^2\theta}{c^2}}\hat{i} + \left(\frac{v\sin\theta}{\gamma(1+\frac{v^2\cos^2\theta}{c^2})}\right)\hat{j} + 0\hat{k} \tag{12.39}$$

where γ is now in the above equations,

$$\gamma(V) = \frac{1}{\sqrt{1-\frac{V^2}{c^2}}}$$

Consider now the total momentum $\vec{p}$ before and after the collision, using the Newtonian definition of momentum.

Before the collision,

$$\vec{p}_i = m\left[\frac{-2v\cos\theta}{1+\frac{v^2\cos^2\theta}{c^2}}\hat{i} + \left[\frac{v\sin\theta}{\gamma(1-\frac{v^2\cos^2\theta}{c^2})} - \frac{v\sin\theta}{\gamma(1+\frac{v^2\cos^2\theta}{c^2})}\right]\hat{j} + 0\hat{k}\right] \tag{12.40}$$

After the collision, the total momentum is

$$\vec{p}_f = m\left[\frac{-2v\cos\theta}{1+\frac{v^2\cos^2\theta}{c^2}}\hat{i} + \left[\frac{v\sin\theta}{\gamma(1+\frac{v^2\cos^2\theta}{c^2})} - \frac{v\sin\theta}{\gamma(1-\frac{v^2\cos^2\theta}{c^2})}\right]\hat{j} + 0\hat{k}\right] \tag{12.41}$$

Thus, we see that the x and z components of the total momentum do not change while the y component is different. This is unlike the Centre of Mass frame where none of the components are different. Hence, we can conclude that in relativity, the usual definition of momentum is unacceptable and we need to find another definition (i.e., find the function $M(v)$) such that momentum is conserved.

The problem can be traced to the definition of momentum in Newtonian mechanics. It involves differentiation with respect to time, a variable which in Newtonian physics is absolute and hence doesn't transform, but in relativity, depends on the frame of reference. This is what led to the fairly complicated law of addition of velocities as we saw above. We could try to find a variable which doesn't transform under Lorentz transformations between two inertial frames. Furthermore, this variable should in the limit of small velocities, reduce to the usual time variable since we know that at small velocities, Newtonian conception of nature of time is valid and gives us an accurate description. If we can find such a variable and then differentiate with respect to this new variable, the problem might be solved.

Such a variable indeed exists. Consider a particle which in an infinitesimal time interval dt travels a distance $d\vec{r}$ in frame S. Now in frame S' (which as above, is moving with a velocity V in the +ve x direction w.r.t S and has its axes parallel to those of S where we have used V to distinguish the velocity of the frame from that of the particle), the corresponding time and space intervals would be dt' and $d\vec{r}'$. These are related to the quantities in S by the Lorentz transformations

$$\begin{aligned} dx' &= \gamma(x - Vdt) \\ dy' &= dy \\ dz' &= dz \\ dt' &= \gamma(dt - \frac{Vdx}{c^2}) \end{aligned} \tag{12.42}$$

It is easy to see that while dt transforms as given above, the quantity $(d\tau)^2$ defined by

$$(c^2 d\tau^2) = c^2 dt^2 - dr^2 \tag{12.43}$$

is the same in S and in S' where it is

$$(c^2 d\tau)^2 = c^2 dt'^2 - dr'^2 \tag{12.44}$$

Thus, we have a quantity,

$$d\tau = dt\left(1 - \frac{1}{c^2}\left(\frac{d\vec{r}}{dt}\right)^2\right)^{1/2} = dt\left(1 - \frac{v^2}{c^2}\right)^{1/2}$$

which is an invariant and also goes to dt as $v \to 0$. Here $\vec{v} = \frac{d\vec{r}}{dt}$ is the velocity of a particle since we have used V as the relative velocity of the frames. **This quantity τ is called the proper time**. It is essentially the time that an observer will measure in a frame in which the particle is at rest.

We can now try to define a momentum using this invariant quantity. Let us consider $\frac{d\vec{r}}{d\tau}$. This is easily seen to be

$$\begin{aligned} m\frac{d\vec{r}}{d\tau} &= \frac{m}{\sqrt{1 - \frac{v^2}{c^2}}}\frac{d\vec{r}}{dt} \\ &= \frac{m}{\sqrt{1 - \frac{v^2}{c^2}}}\vec{v} \end{aligned} \tag{12.45}$$

This choice of momentum thus corresponds, in Eq. (12.34) of taking

$$\boxed{M(v) = \frac{m}{\sqrt{1 - \frac{v^2}{c^2}}}} \tag{12.46}$$

With this definition of momentum, we can work out the transformation of the momentum between two inertial frames. Let $\vec{P}$ and $\vec{P}'$ be the momenta of a particle in frame S and S' respectively. S and S' are related by a Lorentz transformation. Then the components of $\vec{P}'$ are

$$\begin{aligned}
P'_x &= m\frac{dx'}{d\tau} \\
&= m\gamma\left[\frac{dx}{d\tau} - V\frac{dt}{d\tau}\right] \\
&= \gamma P_x - \frac{\gamma m V}{\sqrt{1-\frac{v^2}{c^2}}} \\
&= \gamma P_x - \frac{\gamma m V}{\sqrt{1-\frac{v^2}{c^2}}}
\end{aligned} \tag{12.47}$$

and for the transverse components

$$P'_y = m\frac{dy'}{d\tau} = m\frac{dy}{d\tau} = P_y \tag{12.48}$$

$$P'_z = m\frac{dz'}{d\tau} = m\frac{dz}{d\tau} = P_z \tag{12.49}$$

With the new definitions of the momentum and the transformation properties, we can now explicitly check the momentum conservation in the collision process we considered above.

In Frame S:

Momentum of particle A before the collision

$$\vec{P}_A = m\frac{1}{\sqrt{1-\frac{v^2}{c^2}}}v_A = m\frac{1}{\sqrt{1-\frac{v^2}{c^2}}}(v\cos\theta, v\sin\theta, 0) \tag{12.50}$$

Momentum of particle B before the collision

$$\vec{P}_B = m\frac{1}{\sqrt{1-\frac{v^2}{c^2}}}v_B = m\frac{1}{\sqrt{1-\frac{v^2}{c^2}}}(-v\cos\theta, -v\sin\theta, 0) \tag{12.51}$$

Momentum of particle A after the collision

$$\vec{P}_A = m\frac{1}{\sqrt{1-\frac{v^2}{c^2}}}w_A = m\frac{1}{\sqrt{1-\frac{v^2}{c^2}}}(v\cos\theta, -v\sin\theta, 0) \tag{12.52}$$

Momentum of particle B after the collision

$$\vec{P}_B = m\frac{1}{\sqrt{1-\frac{v^2}{c^2}}}w_B = m\frac{1}{\sqrt{1-\frac{v^2}{c^2}}}(-v\cos\theta, v\sin\theta, 0) \tag{12.53}$$

Thus, the total momentum is conserved, being zero both before and after the collision.

The transformation laws for momentum, Eqs. (12.47)–(12.49) between inertial frames S and S' leave the y and z components of the momentum unchanged while changing the x component to $(\gamma(P_x - \frac{mV}{\sqrt{1-v^2/c^2}}))$ for every particle. Thus, the total momentum of the two particles, which was zero both before and after the collision in S, will be $(-\frac{2\gamma m V}{\sqrt{1-v^2/c^2}})$ in S' both before and after the collision. Total momentum thus is conserved with the relativistic definition of momentum: $\vec{p} = \frac{m\vec{v}}{\sqrt{1-v^2/c^2}}$.

What about the energy of the particle? The Newtonian definition of energy is related to the definition of momentum since it involves the work done which is related to the force which, by Newton's laws, is related to the rate of change of momentum. Thus, we expect that with the change of definition of momentum, there would be a change in the definition of energy also in relativity.

This is indeed the case. Recall that the work done in moving a particle through a distance $\vec{dr}$ against a force $\vec{F}$ is

$$\begin{aligned} dW &= \vec{F} \cdot \vec{dr} \\ &= \frac{d\vec{p}}{dt} \cdot \vec{dr} \\ &= \left[\frac{d\vec{p}}{dt} \cdot \vec{v}\right] dt \end{aligned} \tag{12.54}$$

Thus, if a particle is moved from a point A at a time t_A when its velocity was $\vec{v}_A$ to a point B at a time t_B when its velocity is $\vec{v}_B$, the total work done is

$$W_{AB} = \int_{t_A}^{t_B} dW \tag{12.55}$$

If we use the Newtonian definition of momentum, namely $\vec{p} = m\vec{v}$, then we get

$$\begin{aligned} W_{AB}^N &= \frac{m}{2} \int_{t_A}^{t_B} \frac{dv^2}{dt} dt \\ &= \frac{1}{2} m v_B^2 - \frac{1}{2} m v_A^2 \end{aligned} \tag{12.56}$$

We term the quantity $\frac{1}{2}mv^2$ as the kinetic energy of the particle. Then the work done in moving a particle from A to B is simply the change in the kinetic energy of the particle.

Let us now repeat this exercise with the new definition of momentum. Now, the force would be

$$\vec{F} = \frac{d\vec{P}}{dt} = \frac{d\left(\frac{m\vec{v}}{\sqrt{1-v^2/c^2}}\right)}{dt}$$

Then the incremental work done is

$$\begin{aligned} dW &= \frac{d\vec{P}}{dt} \cdot \vec{dr} \\ &= \left[\frac{d\frac{m\vec{v}}{\sqrt{1-v^2/c^2}}}{dt}\right] \cdot \vec{dr} \end{aligned}$$

$$= \left[\vec{v} \cdot \frac{d \frac{m\vec{v}}{\sqrt{1-v^2/c^2}}}{dt} \right] dt$$

$$= \left[\frac{d\left(\vec{v} \cdot \frac{m\vec{v}}{\sqrt{1-v^2/c^2}} \right)}{dt} \right] dt - \left[\frac{m\vec{v}}{\sqrt{1-v^2/c^2}} \right] \cdot d\vec{v} \tag{12.57}$$

and the total work done in moving the particle from A to B is

$$\boxed{\begin{aligned} W_{AB} &= \int_{t_A}^{t_B} dW \\ &= m \left[\frac{v_B^2}{\sqrt{1-\frac{v_B^2}{c^2}}} - \frac{v_A^2}{\sqrt{1-\frac{v_A^2}{c^2}}} \right] + mc^2 \left[\sqrt{1-\frac{v_B^2}{c^2}} - \sqrt{1-\frac{v_A^2}{c^2}} \right] \\ &= \frac{mc^2}{\sqrt{1-\frac{v_B^2}{c^2}}} - \frac{mc^2}{\sqrt{1-\frac{v_A^2}{c^2}}} \end{aligned}} \tag{12.58}$$

Since W_{AB} is the difference in the energy of the particle at B and A, we can define the energy as

$$E_{\mathrm{rel}} = \frac{mc^2}{\sqrt{1-\frac{v^2}{c^2}}} \tag{12.59}$$

This expression for the relativistic energy of the particle is very different from the Newtonian expression $E_N = \frac{mv^2}{2}$. For instance, when the particle is at rest, i.e., $v = 0$, the Newtonian energy vanishes while the relativistic energy is mc^2! However, for small velocities, i.e., $v \ll c$, we get

$$E_{\mathrm{rel}} = mc^2 + \frac{1}{2}mv^2 + \quad \text{terms of order} \frac{v^4}{c^4} \text{ or higher} \tag{12.60}$$

We can recognise, in this non-relativistic limit that the relativistic energy is the Newtonian energy plus a constant term mc^2.

We expect relativistic theories to go over smoothly to non-relativistic results for small velocities just like the Lorentz transformation equations go to Galilean transformation equations for small velocities. The constant mc^2, called the **rest energy** since it is the value of E_{rel} when $v = 0$, hence has to be reconciled with our ideas about relativistic and Newtonian laws of mechanics. We note however, that if when particles are interacting no new particles are created or none existing destroyed, the constant ($m_i c^2$) for the i^{th} particle in the system of particles interacting with each other simply adds to the total energy of the system of particles a constant term given by $\Sigma_i m_i c^2$. As far as the conservation of energy principle is concerned, addition of a constant to the energy expression makes no difference to the principle. Hence, we have that under the circumstances of no creation or destruction of particles, we can discard the rest energy and thus go over smoothly to the usual Newtonian expression for energy $\frac{1}{2}mv^2$ for velocities small compared to c.

It turns out however, that when particles interact with each other at very high energies, particles can be created or destroyed. Historically, the production of an electron (e^-) and its antiparticle, a positron (e^+) from radiation (γ^*) in interaction with matter was observed a long time ago.

$$\gamma^* \to e^- + e^+, \tag{12.61}$$

provided the energy of γ^* was higher than ($2m_e c^2$) where m_e is the mass of the electron or positron, the masses being the same since the positron is the antiparticle of the electron. The rest energy thus, is not an irrelevant constant always, but is as real as any other forms of energy.

Note that with the new expression for the relativistic energy of the particle, it is clear that if we have a particle with a non-zero mass m, then its velocity can never exceed c. As the speed approaches c, the energy keeps increasing and the velocity never reaches c. (Recently, in 2011, an experiment has been reported where the neutrinos were observed travelling at a speed faster than light. If confirmed, this will totally unsettle the theory of relativity which has been unchallenged for more than a century. For the latest summary of the results, see, e.g., the article entitled 'Neutrino experiment replicates faster than light finding' by **E.S. Reich, Nature News, November 18, 2011**.)

However, for particles with zero mass, the situation is very different. When $m = 0$ the energy and momentum becomes indeterminate

$$\vec{P} = \frac{m\vec{v}}{\sqrt{1 - \frac{v^2}{c^2}}}$$

$$E_{\text{rel}} = \frac{mc^2}{\sqrt{1 - \frac{v^2}{c^2}}}$$

Energy and momentum are, in general, related by

$$\boxed{E_{\text{rel}}^2 = P^2c^2 + m^2c^4} \tag{12.62}$$

so that for $m = 0$ we have

$$E = Pc \tag{12.63}$$

Amongst particles that exist freely, only photons satisfy Eq. (12.63). Till recently, neutrinos also were thought to have zero mass. Experiments of late have shown that some if not all the known neutrinos have small but non-zero mass. (See, e.g., a nice review of these results in **C. Waltham, 'Teaching Neutrino Oscillations', American Journal of Physics, 72, 743(2004)**).

EXAMPLE 12.6 A neutral Kaon (a kind of subatomic particle) of mass m_K is at rest and decays into a pair of pion and anti-pion, each with a mass m_π. Find the kinetic energy of each of the pions.

Solution

The decay process is

$$K^0 = \pi^+ + \pi^-$$

Momentum and energy are conserved in this decay. The initial momentum of the Kaon is zero since it is at rest. Thus, the final momentum of the pion and anti-pion pair would also be zero. The particles therefore, have the same magnitude of momentum in opposite directions. Thus,

$$\vec{p}_{\pi^+} + \vec{p}_{\pi^-} = \vec{p}_{K^0} = 0$$

Also because the masses of the pion and antipion are the same, and their momentum equal in magnitude, their kinetic energies are equal. Thus,

$$K_{\pi^+} = K_{\pi^-} = K$$

The initial energy of the Kaon before the decay is only its rest energy and hence,

$$E = m_{K^0}c^2$$

After the decay the energy is the rest energy of the pions as well as their kinetic energy.

$$E = E_{\pi^+} + E_{\pi^-} = \left(m_{\pi^+}c^2 + K_{\pi^+}\right) + \left(m_{\pi^-}c^2 + K_{\pi^-}\right)$$

Energy conservation gives us

$$m_{K^0}c^2 = \left(m_{\pi^+}c^2 + K_{\pi^+}\right) + \left(m_{\pi^-}c^2 + K_{\pi^-}\right)$$

Or

$$m_{K^0}c^2 - m_{\pi^+}c^2 - m_{\pi^-}c^2 = 2K$$

which gives us

$$K = \frac{1}{2}m_{K^0}c^2 - m_{\pi^+}c^2 - m_{\pi^-}c^2$$

Putting in the numbers, $m_{K^0} = 497.65$ MeV/c^2, $m_{\pi^+} = m_{\pi^-} = 139.57$ MeV/c^2, we get $K = 109.26$ MeV.

PROBLEM 12.6 Compute both classically and relativistically the momentum of two electrons, one moving with a speed of 10^7 m/s and one moving at 10^4 m/s. What is the percentage difference in the relativistic and classical values for both the electrons?

PROBLEM 12.7 A particle of mass m with momentum mc hits a target of mass $2\sqrt{2}m$ and forms a single particle in a completely inelastic collision. Calculate the speed with which the particle of mass m is moving, the mass of the new particle formed in the collision and its speed.

12.5 FOUR VECTORS

We have already seen that the Lorentz transformations describe the transformation of the coordinates of a point between two inertial frames. Recall the Lorentz transformations between two frames, moving with respect to each other with a speed v along the x axis are given by (Eq. (12.21))

$$\begin{aligned} x' &= \gamma(x - vt) \\ y' &= y \\ z' &= z \\ t' &= \gamma\left(t - \frac{vx}{c^2}\right) \end{aligned} \tag{12.64}$$

where γ now is

$$\gamma = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}$$

We had discussed the definition of a vector in the chapter on mathematical preliminaries. We saw there that a vector in space is defined by its transformation under rotations. Since any point in three dimensions is defined by a set of three quantities (the coordinates x_1, x_2, x_3), the basic transformation law is defined as how the coordinates of a point, namely the three quantities, (x_1, x_2, x_2) transform under rotations. We also saw that any set of three quantities which transform under rotations in the same way as the coordinates x_1, x_2, x_3, is then called a vector in three dimensions. (Mathematical Preliminaries, Equations (50)–(53)).

In a completely analogous way, we attempt the same for Lorentz transformations. We know how the coordinates of a point transform under Lorentz transformations (Eq. (12.64). Note that these are now four quantities, instead of three, in the case of ordinary vectors. We call any set of four quantities, which under a Lorentz transformation transform like Eq. (12.64) as **4-vectors**. For this purpose, it is convenient to introduce a new notation, as

$$\begin{aligned} x_0 &= ct \\ x_1 &= x \\ x_2 &= y \\ x_3 &= z \end{aligned} \tag{12.65}$$

Thus the coordinates of the coordinate 4-vector are (x_0, x_1, x_2, x_3).

With this notation, the Lorentz transformations become

$$\begin{aligned} x'_0 &= \gamma(x_0 - \beta x_1) \\ x'_1 &= \gamma(x_1 - \beta x_0) \\ x'_2 &= x_2 \\ x'_3 &= x_3 \end{aligned} \tag{12.66}$$

with $\beta = V/c$.

Thus any set of four quantities (A_0, A_1, A_2, A_3) which transform under Lorentz transformations as Eq. (12.66) is a 4-vector. We write a four vector as $\tilde{A}$

The Second Postulate of Relativity, namely the invariance of the speed of light in all inertial frames, we saw was built into the Lorentz transformations. In our new notation, this implies,

$$x_0^2 - x_1^2 - x_2^2 - x_3^2 = x_0'^2 - x_1'^2 - x_2'^2 - x_3'^2 \tag{12.67}$$

For any arbitrary four vector, $\tilde{A}$ with components (A_0, A_1, A_2, A_3), a similar situation will hold and we will have

$$A_0^2 - A_1^2 - A_2^2 - A_3^2 = A_0'^2 - A_1'^2 - A_2'^2 - A_3'^2 \tag{12.68}$$

or, since

$$|\vec{A}|^2 = A_1^2 + A_2^2 + A_3^2$$

we get for any 4-vector

$$A_0^2 - |\vec{A}|^2 = A_0'^2 - |\vec{A}'|^2 \tag{12.69}$$

It can also be shown easily that for any two 4-vectors, $\tilde{A}$ and $\tilde{B}$, the four dimensional 'scalar product', defined as

$$\tilde{A} \cdot \tilde{B} \equiv A_0 B_0 - \vec{A} \cdot \vec{B}$$

is invariant.

As an example of 4-vectors, we notice that the relativistic energy and momentum also form a 4-vector $\tilde{P}$ which has components $(\gamma mc, \gamma m\vec{v})$. That is, the 'time' component or the zeroth component is related to the relativistic energy $\frac{E_{\text{rel}}}{c}$ (Eq. (12.59) and the 'space' components are related to the relativistic momentum $\vec{p} = \gamma m\vec{v}$. Once again, we note that the scalar product of this four vector with itself will be an invariant quantity, that is

$$\boxed{E_{\text{rel}}^2 - |\vec{p}|^2 c^2 = m^2 c^4} \tag{12.70}$$

This is an extremely useful result since we can now find the energy E_{rel} if we know the momentum p or vice versa without finding out the velocity. It also is very useful in determining the energies and momenta of particles in collision problems with particles. This is because, since the magnitude of the energy momentum 4-vector, or indeed any scalar product of two energy-momentum 4-vectors is an invariant, we can compute the same in any convenient frame and then find out the results in any other frame. We illustrate this with a few examples of collision and decay processes of elementary particles.

EXAMPLE 12.7 A particle of mass M and 4-momentum P, decays into two particles of masses m_1 and m_2. Use the conservation of energy and momentum to show that the total energy of the first particle in the rest frame of the decaying particle is

$$E_1 = \frac{(M^2 + m_1^2 - m_2^2)c^2}{2M}$$

Solution

The initial four momentum of the decaying particle is given by $\tilde{P}$. But the particle is at rest and so the components of four-momentum are

$$\tilde{P} = \left(\frac{E}{c}, 0\right)$$

The final state consists of two particles of masses m_1 and m_2. Let their four-momenta be $\tilde{p}_1$ and $\tilde{p}_2$ respectively. Then the total 4-momenta after the decay will be

$$(\tilde{p}_1 + \tilde{p}_2)$$

where

$$\tilde{p}_1 = \left(\frac{E_1}{c}, \vec{p}_1\right)$$
$$\tilde{p}_2 = \left(\frac{E_2}{c}, \vec{p}_2\right)$$

Conservation of energy-momentum gives us

$$\tilde{P} = (\tilde{p}_1 + \tilde{p}_2)$$

or

$$\tilde{p}_2 = \tilde{P} - \tilde{p}_1$$

Squaring both sides gives us

$$\begin{aligned}\tilde{p}_2^2 &= (\tilde{P} - \tilde{p}_1)^2\\ m_2^2c^2 &= \tilde{P}^2 + \tilde{p}_1^2 - 2\tilde{p}_1 \cdot \tilde{P}\\ m_2^2c^2 &= M^2c^2 + m_1^2c^2 - 2E_1M\end{aligned}$$

where we have used the fact that

$$\tilde{p}_2^2 = m_2^2c^2$$

and, since the decaying particle is at rest and so has only rest energy

$$\tilde{P} = (Mc, 0)$$

Thus,

$$2E_1M = M^2c^2 + m_1^2c^2 - m_2^2c^2$$

The result follows.

EXAMPLE 12.8 The lambda particle (Λ) is a neutral particle of mass M that decays into a nucleon of mass m_1 and a pi-meson with mass m_2. Using conservation of energy and momentum, show that if the opening angle between the two daughter particles is θ, then the mass of the decaying particle is given by

$$M^2 = m_1^2 + m_2^2 + 2E_1E_2 - 2p_1p_2\cos\theta$$

Solution

Initial four-momentum (in the rest frame of the lambda particle)

$$\tilde{P} = (\frac{E}{c}, 0) = (Mc, 0)$$

The square of this four-momentum will obviously be

$$\tilde{P}^2 = M^2c^2$$

Now consider the final state with two particles. Let their four-momenta in laboratory frame be $\tilde{p}_1$ and $\tilde{p}_2$. Thus, the final four-momentum (in the lab frame)

$$(\tilde{p}_1 + \tilde{p}_2)$$

The square of this final four-momentum will be

$$(\tilde{p}_1 + \tilde{p}_2)^2 = m_1^2c^2 + m_2^2c^2 + 2\frac{E_1E_2}{c^2} - 2p_1p_2\cos\theta$$

where θ is the angle between $\vec{p}_1$ and $\vec{p}_2$. Now we use the fact that the square of the 4-vector is an invariant quantity. Hence, the square of the initial four-momentum of the lambda particle is M^2c^2. Equating the squares of the initial and final four-momenta, we get

$$\tilde{P}^2 = (\tilde{p}_1 + \tilde{p}_2)^2$$
$$M^2c^2 = m_1^2c^2 + m_2^2c^2 + 2\frac{E_1E_2}{c^2} - 2p_1p_2\cos\theta$$

Remember, we are equating the squares of the 4-momenta which are invariant. Obviously, in the rest frame of the lambda particle, the opening angle will be 180° since the decay products will be going back to back by conservation of 3-momentum.

Our next example is that of a collision between a photon and an electron. This is the well known Compton effect which was first observed by Arthur Compton and provided a remarkable verification of the particle-like nature of light in quantum theory.

EXAMPLE 12.9 Compton Effect: A photon with frequency ν hits a free electron (at rest). Find the relation between angle of photon deflection and frequency of deflected photon.

Solution

Compton Effect is the scattering of a photon from free electrons in a conductor. Let the initial four-momentum of the electron be $\tilde{Q}_i$. Since it is at rest,

$$\tilde{Q}_i = (m_ec, 0)$$

Let the final four-momentum of the electron after being hit by the photon be $\tilde{Q}_f$. Now, since the four-momentum squared is an invariant, we have

$$\tilde{Q}_i^2 = \tilde{Q}_f^2 = m_e^2c^2$$

Let $\tilde{P}_i$ and $\tilde{P}_f$ be the initial and final four-momenta of the photon. Recalling that a photon has zero mass, we have

$$\tilde{P}_i = (p_i, p_i\hat{n}_i)$$

and

$$\tilde{P}_f = (p_f, p_f\hat{n}_f)$$

where $\hat{n}_i$ and $\hat{n}_f$ are the unit vectors in the direction of the initial and final photon 3-momentum respectively and we have used Eq. (12.63) for the relationship between the energy and momentum of the photon. Here p_i and p_f are the magnitude of the initial and final momenta of the photon.

For the photon, we know that

$$\tilde{P}_i^2 = \tilde{P}_f^2 = E^2 - p^2c^2 = 0$$

Using conservation of energy momentum, we get

$$\begin{aligned}
\tilde{P}_i + \tilde{Q}_i &= \tilde{P}_f + \tilde{Q}_f \\
\tilde{Q}_f &= \tilde{Q}_i + \tilde{P}_i - \tilde{P}_f \\
\tilde{Q}_f^2 &= \tilde{P}_i^2 + \tilde{P}_f^2 + \tilde{Q}_i^2 - 2\tilde{P}_i \cdot \tilde{P}_f + 2\tilde{Q}_i \cdot (\tilde{P}_i - \tilde{P}_f) \\
\tilde{P}_i \cdot \tilde{P}_f &= \tilde{Q}_i \cdot (\tilde{P}_i - \tilde{P}_f) \\
p_i p_f - p_i p_f \cos\theta &= m_e c(p_i - p_f) \\
2\sin^2\left(\frac{\theta}{2}\right) &= m_e c\left(\frac{1}{p_f} - \frac{1}{p_i}\right) \\
2\sin^2\left(\frac{\theta}{2}\right) &= m_e c^2\left(\frac{1}{h\nu_f} - \frac{1}{h\nu_i}\right)
\end{aligned}$$

where we have used the fact that the energy of the photon of frequency μ is $h\nu$ and the energy and momentum are related by $E = pc$.

PROBLEM 12.8 An important process in the Sun is

$$p + p + e^- \rightarrow d + \nu$$

Assume that in the lab frame, the two protons have the same energy and impact angle, hit the electron which is at rest. Further assume that the neutrino and the electron are massless and the mass of the deuteron is $2m_P$. Calculate the energy of the neutrino in the rest frame of the deuteron.
[*Hint: Use the fact that the quantity* $E^2 - p^2c^2$ *is invariant*]

PROBLEM 12.9 An elementary particle with kinetic energy $2m_0c^2$ hits another similar particle at rest. Find the combined kinetic energy of the two particles and their individual momenta in the centre of mass frame.

PROBLEM 12.10 A particle at rest of rest mass m_1 disintegrates into three particles of rest masses m_2, m_3, m_4. Find the maximum total energy that the particle of mass m_2 can possess in this kind of process.

PROBLEM 12.11 A particle of rest mass m travelling at speed v in the x-direction, decays into two massless neutrinos, moving in the positive and negative x-direction relative to the original particle. What are their energies? What are the neutrino energies and directions if the neutrinos are emitted in the positive and negative z-direction relative to the original particle (i.e., perpendicular to the direction of motion, in the particle's rest frame).

12.6 TRANSFORMATION OF ELECTRIC AND MAGNETIC FIELDS

In our discussion above, we have considered the transformation of coordinates and other dynamical variables like velocity, energy and momentum under Lorentz transformations. That is, we discussed how these variables change when we go from one inertial frame to another. However, for considerations of electrodynamics, we also need to find out how the fundamental entities, namely electric and magnetic fields transform between frames under a Lorentz transformation. We proceed to do this in this section.

The starting point for deriving the transformation properties of the electric and magnetic fields are the dynamical equations obeyed by them, namely the Maxwell's equations. In a frame S, in vacuum with no charges and currents, these read,

$$\begin{aligned}
\vec{\nabla}\cdot\vec{E} &= 0\\
\vec{\nabla}\times\vec{E} &= -\frac{\partial\vec{B}}{\partial t}\\
\vec{\nabla}\cdot\vec{B} &= 0\\
\vec{\nabla}\times\vec{B} &= \mu_0\varepsilon_0\frac{\partial\vec{E}}{\partial t} = \frac{1}{c^2}\frac{\partial\vec{E}}{\partial t}
\end{aligned}\tag{12.71}$$

Now consider an inertial frame S' moving with a speed v along the $+x$ direction with its axes parallel to those of S and coinciding with it at $t = t' = 0$. Then the coordinates of the two frames are related by Lorentz transformations

$$\begin{aligned}
x' &= \gamma(x - vt)\\
y' &= y\\
z' &= z\\
t' &= \gamma(t - \frac{vx}{c^2})
\end{aligned}\tag{12.72}$$

where γ now is

$$\gamma = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}$$

By the postulates of Einstein's theory, Maxwell's equations are also valid in S'. Thus, if $\vec{E}'$ and $\vec{B}'$ denote the fields in the frame S', then two of the Maxwell's equations read

$$\begin{aligned}
\vec{\nabla}'\times\vec{E}' &= -\frac{\partial\vec{B}'}{\partial t'}\\
\vec{\nabla}'\times\vec{B}' &= \mu_0\varepsilon_0\frac{\partial\vec{E}'}{\partial t'}
\end{aligned}\tag{12.73}$$

Consider the x component of the Curl $\vec{E}$ equation in Eq. (12.71).

$$\frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} = -\frac{\partial B_x}{\partial t}\tag{12.74}$$

Now $\vec{E}, \vec{B}$ are obviously functions of (x, y, z, t). But from the Lorentz transformations, Eq. (12.72), they can equally be considered as functions of (x', y', z', t'). We can do this by transforming the derivatives as

$$\frac{\partial}{\partial y} = \frac{\partial}{\partial y'}$$

$$\frac{\partial}{\partial z} = \frac{\partial}{\partial z'}$$

$$\frac{\partial}{\partial t} = \left(\frac{\partial t'}{\partial t}\right)\frac{\partial}{\partial t'} + \left(\frac{\partial x'}{\partial t}\right)\frac{\partial}{\partial x'} = \gamma\frac{\partial}{\partial t'} + \gamma(-v)\frac{\partial}{\partial x'}$$

Therefore, we get from Eq. (12.74),

$$\frac{\partial E_z}{\partial y'} - \frac{\partial E_y}{\partial z'} = -\gamma\frac{\partial B_x}{\partial t'} + \gamma v\frac{\partial B_x}{\partial x'} \tag{12.75}$$

But from the divergence of $\vec{B}$ equation, we have

$$\vec{\nabla}\cdot\vec{B} = 0$$

$$\frac{\partial B_x}{\partial x} + \frac{\partial B_y}{\partial y} + \frac{\partial B_z}{\partial z} = 0$$

$$\left(\frac{\partial t'}{\partial x}\right)\frac{\partial B_x}{\partial t'} + \left(\frac{\partial x'}{\partial x}\right)\frac{\partial B_x}{\partial x'} = -\frac{\partial B_y}{\partial y'} - \frac{\partial B_z}{\partial z'}$$

$$\frac{\partial B_x}{\partial x'} = -\frac{-\left[\frac{\partial B_y}{\partial y'} + \frac{\partial B_z}{\partial z'}\right] - \left(-\frac{\gamma v}{c^2}\right)\frac{\partial B_x}{\partial t'}}{\gamma} \tag{12.76}$$

We can substitute this in Eq. (12.75) to get after rearranging the terms

$$-\frac{\partial}{\partial t'}\left[\gamma B_x - \frac{\gamma v^2}{c^2}B_x\right] = \frac{\partial}{\partial y'}\left[E_z + vB_y\right] - \frac{\partial}{\partial z'}\left[E_y - vB_z\right]$$

or

$$-\frac{\partial B_x}{\partial t'} = \gamma\frac{\partial}{\partial y'}\left(E_z + vB_y\right) - \gamma\frac{\partial}{\partial z'}\left(E_y - vB_z\right) \tag{12.77}$$

If we write the x component of the curl of $\vec{E}$ equation in Eq. (12.73), we get

$$-\frac{\partial B'_x}{\partial t'} = \frac{\partial E'_z}{\partial y'} - \frac{\partial E'_y}{\partial z'} \tag{12.78}$$

Now comparing Eqs. (12.78) and (12.77), we get

$$\begin{aligned} B'_x &= B_x \\ E'_y &= \gamma(E_y - vB_z) \\ E'_z &= \gamma(E_z + vB_y) \end{aligned} \tag{12.79}$$

In an analogous fashion, we can use the Curl of $\vec{B}$ to find the transformation of the other components as

$$\begin{aligned} E'_x &= E_x \\ B'_y &= \gamma(B_y + \frac{v}{c^2}E_z) \\ B'_z &= \gamma(B_z - \frac{v}{c^2}E_y) \end{aligned} \tag{12.80}$$

Equations (12.79) and (12.80) are the complete transformation equations for the electric and magnetic fields under a Lorentz transformation with the relative velocity between two inertial frames along the $+x$ axis.

For the general case of the frames moving with an arbitrary velocity $\vec{v}$, the transformations equations are more complicated but can be written in a succinct form as

$$\boxed{\begin{aligned} \vec{E}' &= \gamma\left[\vec{E} + \vec{\beta}\times\vec{B}\right] - \frac{\gamma^2}{\gamma+1}\vec{\beta}(\vec{\beta}\cdot\vec{E}) \\ \vec{B}' &= \gamma\left[\vec{B} - \vec{\beta}\times\vec{E}\right] - \frac{\gamma^2}{\gamma+1}\vec{\beta}(\vec{\beta}\cdot\vec{B}) \end{aligned}} \tag{12.81}$$

where, as usual $\vec{\beta} = \frac{\vec{v}}{c}$ and $\gamma = \frac{1}{\sqrt{1-\beta^2}}$.

Given these transformation properties, a few features are noteworthy.

1. Invariants

From the transformation equations given above, it is clear that some quantities do not change under a Lorentz transformation or they are invariant. These are

$$\vec{E}\cdot\vec{B}$$

and

$$\vec{B}^2 - \frac{1}{c^2}\vec{E}^2$$

It is easy to check that

$$\begin{aligned} \vec{E}'\cdot\vec{B}' &= \vec{E}\cdot\vec{B} \\ \vec{B}'^2 - \frac{1}{c^2}\vec{E}'^2 &= \vec{B}^2 - \frac{1}{c^2}\vec{E}^2 \end{aligned} \tag{12.82}$$

These lead to some interesting results. Consider a frame in which the electric and magnetic fields are orthogonal. Then, $\vec{E}\cdot\vec{B} = 0$ in this frame. But since this is an invariant quantity, it is zero in all frames related to the original frame by Lorentz transformations! In particular, since $\vec{E}$ and $\vec{B}$ are orthogonal in an electromagnetic wave, they continue to be orthogonal in all frames in which the EM wave is observed.

Now consider a frame in which $\vec{B} = 0$. Then the invariant quantity $\vec{B}^2 - \frac{1}{c^2}\vec{E}^2 < 0$ in this frame. But since it is invariant, it must be negative in all frames. What this means is that there is no frame in which these fields are pure magnetic, i.e., in which $\vec{E}$ is zero since that would imply that the invariant quantity $\vec{B}^2 - \frac{1}{c^2}\vec{E}^2 > 0$ which is not possible since it was negative in one frame.

2. Magnetic field due to a charge in motion

Consider a frame S in which there is a charge e at rest at the origin. Then the stationary charge obviously only has an electric field in S which is given by

$$\begin{aligned} \vec{E} &= \frac{e}{4\pi\varepsilon_0}\frac{\hat{r}}{r^2} \\ \vec{B} &= 0 \end{aligned} \tag{12.83}$$

or

$$E_x = \frac{e}{4\pi\varepsilon_0}\frac{x}{r^3}$$
$$E_y = \frac{e}{4\pi\varepsilon_0}\frac{y}{r^3}$$
$$E_z = \frac{e}{4\pi\varepsilon_0}\frac{z}{r^3}$$
$$\vec{B} = 0 \tag{12.84}$$

For an observer in S' moving once again with a speed v along the $+x$ axis, the charge would appear to be in motion with a velocity v along the $-x$ direction. The fields in this frame would then be, from the transformation properties of the fields given in Eqs. (12.79) and (12.80),

$$B'_x = 0$$
$$B'_y = \frac{\gamma v}{c^2}E_z$$
$$B'_z = \frac{-\gamma v}{c^2}E_y \tag{12.85}$$

Or putting it together, we have

$$\vec{B}' = \frac{1}{c^2}(\vec{v}_e \times \vec{E}) \tag{12.86}$$

where $\vec{v}_e = -\vec{v}$ is the velocity of the charged particle in S'. Equation (12.86) is the magnetic field due to a charged particle in motion as we saw in our discussion on magnetic fields.

EXAMPLE 12.10 In frame K, the electric and magnetic fields are given by

$$\vec{E} = 2\hat{k}$$
$$\vec{B} = \hat{j} + \hat{k}$$

Find a frame K' such that in K', $\vec{E}$ and $\vec{B}$ are parallel to each other.

Solution

We are given in frame K, the electric and magnetic fields

$$\vec{E} = 2\hat{k}$$
$$\vec{B} = \hat{j} + \hat{k}$$

We also know that $\vec{E}\cdot\vec{B}$ and $B^2 - \frac{E^2}{c^2}$ are invariants. In this frame, $\vec{E}\cdot\vec{B} = 2$ and $B^2 - \frac{E^2}{c^2} = 2 - \frac{4}{c^2}$. We need a frame such that $\vec{E}'$ and $\vec{B}'$ are parallel in that frame. Choose the velocity $\vec{v}$ of frame K' along the x-axis, i.e., $\vec{\beta} = \beta\hat{i}$ where $\vec{\beta} = \frac{\vec{v}}{c}$. Then in K',

$$E'_1 = 0$$
$$E'_2 = -\gamma\beta$$
$$E'_3 = 2\gamma + \gamma\beta$$

$$B_1' = 0$$
$$B_2' = \gamma + 2\beta\gamma$$
$$B_3' = \gamma$$

If $\vec{E}'$ and $\vec{B}'$ are parallel in some frame, then $\vec{E}' \times \vec{B}' = 0$. Hence,

$$\begin{aligned} E_2'B_3' - E_3'B_2' &= 0 \\ -(\gamma\beta)(\gamma) - (2\gamma + \gamma\beta)(\gamma + 2\beta\gamma) &= 0 \\ -\gamma^2\beta - 2\gamma^2 - 4\beta\gamma^2 - \gamma^2\beta - 2\beta^2\gamma^2 &= 0 \\ \beta^2 + 3\beta + 1 &= 0 \\ \beta &= \frac{-3 \pm \sqrt{5}}{2} \end{aligned}$$

Now since $|\beta| \leq 1$, we get $\beta = \frac{-3+\sqrt{5}}{2}$.

EXAMPLE 12.11 A particle of charge q and rest mass m is released with zero initial velocity in a region with an electric field $\vec{E}$ in the y direction and magnetic field $\vec{B}$ in the z direction. Find the conditions necessary for the existence of a frame in which $\vec{E} = 0$ and those for which $\vec{B} = 0$.

Solution

We have the in frame S, the laboratory frame, $\vec{E} = E\hat{j}$ and $\vec{B} = B\hat{k}$. In any frame S' moving with a velocity $\vec{v} = v\hat{i}$ w.r.t S, the transformed fields are

$$E_x' = E_x = 0$$
$$E_y' = \gamma(E - vB)$$
$$E_z' = 0$$

and

$$B_x' = B_x = 0$$
$$B_y' = 0$$
$$B_z' = \gamma\left(B - \frac{v}{c^2}E\right)$$

For $E' = 0$, we must have therefore, $E - vB = 0$ or $v = \frac{E}{B}$. But since $v \leq c$, this gives us $E \leq cB$.

For $B' = 0$, we must have $B - \frac{v}{c^2}E = 0$ or $v = \frac{c^2B}{E} \leq c$. Thus, we get $cB \leq E$.

EXAMPLE 12.12 A particle of mass m and charge e moves in the laboratory in a crossed, static, uniform, electric and magnetic fields. $\vec{E}$ parallel to the x-axis and $\vec{B}$ parallel to the y-axis. For $|\vec{E}| < |\vec{B}|$, make the necessary Lorentz transformation to obtain the parametric equations for the trajectory of the particle

Solution

Let the frame in question be K. Then in this frame $\vec{E} = E\hat{x}$ and $\vec{B} = B\hat{y}$. Let $|\vec{E}| < |\vec{B}|$ in this frame and therefore, since $B^2 - \frac{E^2}{c^2}$ is invariant we must have $B^2 > \frac{E^2}{c^2}$ in all frames. We can therefore, make a transformation to a frame in which $\vec{E}' = 0$. Let this frame be K' moving with a velocity $\vec{u}$ w.r.t. K. Now let

$$\vec{u} = \frac{c\vec{E} \times \vec{B}}{B^2} = \frac{cEB\hat{z}}{B^2} = \frac{cE}{B}\hat{z}$$

Then the fields in K' namely $\vec{E}'$ and $\vec{B}'$ are given by (with $\vec{\beta} = \frac{\vec{u}}{c}$)

$$\vec{E}' = \gamma\left[\vec{E} + \vec{\beta} \times \vec{B}\right] - \frac{\gamma^2}{\gamma+1}\vec{\beta}(\vec{\beta} \cdot \vec{E})$$

and

$$\vec{B}' = \gamma\left[\vec{B} - \vec{\beta} \times \vec{E}\right] - \frac{\gamma^2}{\gamma+1}\vec{\beta}(\vec{\beta} \cdot \vec{B})$$

Therefore,

$$\begin{aligned} E'_x &= \gamma\left[E_x + (\vec{u}/c \times \vec{B})_x\right] - \frac{\gamma^2}{\gamma+1}\left(\frac{\vec{u}}{c}\right)_x (\vec{\beta} \cdot \vec{E}) \\ &= \gamma\left[E_x + (1/c)\frac{cEB}{B}(\hat{z} \times \hat{y})_x\right] \\ &= \gamma\left[E_x - E_x\right] \\ &= 0 \end{aligned}$$

Similarly, it is easily seen that $E'_y = E'_z = 0$.
We also compute the magnetic field in the primed frame.

$$\begin{aligned} B'_x &= \gamma\left[B_x - 1/c(\vec{u} \times \vec{E})_x\right] - \frac{\gamma^2}{\gamma+1}\frac{u_x}{c}(\vec{\beta} \cdot \vec{E}) \\ &= -\frac{\gamma}{c}\frac{cEE}{B}(\hat{z} \times \hat{x})_x - \frac{\gamma^2}{\gamma+1} \\ &= 0 \end{aligned}$$

since $u_x = 0$ and $(\hat{z} \times \hat{x})_x = 0$.

Similarly, we can show that $B'_z = 0$. For B'_y,

$$\begin{aligned} B'_y &= \gamma\left[B_y - 1/c(\vec{u} \times \vec{E})_y\right] - \frac{\gamma^2}{\gamma+1}\frac{u_y}{c}(\vec{\beta} \cdot \vec{E}) \\ &= \gamma B_y - \frac{\gamma cEE}{cB}(\hat{z} \times \hat{x})_y \\ &= \gamma\left[B_y - \frac{E^2}{B}\right] \\ &= (B^2 - E^2)^{1/2} \end{aligned}$$

since $\gamma = \frac{B}{(B^2-E^2)^{1/2}}$.

Thus, we have

$$\vec{B}' = (B^2 - E^2)^{1/2}\hat{y}' = \frac{\vec{B}}{\gamma}$$

and

$$\vec{E}' = 0$$

in frame K'. We now consider the motion of the particle in frame K'. The equations of motion are

$$\frac{d\vec{p}}{dt} = \frac{q}{c}(\vec{v} \times \vec{B})$$

and

$$\frac{dE}{dt} = 0$$

where E is the energy and we have omitted the primes for simplicity but it has to be remembered that this is obviously the equation of motion in the primed frame in which there is only a B field.

Since the energy of the particle is constant in a pure, static magnetic field, the magnitude of the velocity is also constant and so is γ_v. Thus,

$$\vec{p} = m\vec{v}\gamma_v$$

or

$$\frac{d\vec{v}}{dt} = \frac{q}{mc\gamma_v}(\vec{v} \times \vec{B}) = \vec{v} \times \vec{\omega}_B$$

where $\vec{\omega}_B = \frac{q\vec{B}}{mc\gamma} = \frac{q\vec{B}c}{E}$ is a constant.

The solutions to the equations of motion are given by

$$\vec{v}(t) = \vec{v}_{||}\hat{\varepsilon}_3 + \omega_B a(\hat{\varepsilon}_1 - i\hat{\varepsilon}_2)e^{-i\omega_B t}$$

where $v_{||}$ is the constant initial velocity along the $\vec{B}$ direction and a is the gyration radius. The gyration radius is given by

$$a = \frac{cp_\perp}{eB}$$

and $p_\perp$ is the transverse momentum of the particle.

In our case, $\vec{B}$ is along the $\hat{y}$ direction and hence,

$$\vec{v}(t) = v_{||}\hat{y} + \omega_B a(\hat{z} - i\hat{x})e^{-i\omega_B t}$$

In component form, we get

$$v_x(t) = -\omega_B a \sin\omega_B t$$
$$v_y(t) = v_{||}$$
$$v_z(t) = \omega_B a \cos\omega_B t$$

Solving for the displacement, we get

$$\vec{x}(t) = \vec{X}_0(t) + v_{||}t\hat{y} + ia(\hat{z} - i\hat{x})e^{-i\omega_B t}$$

which, in component form, gives us (restoring the primes now to remind ourselves that this is the primed frame),

$$x'(t') = x'_0 + a\cos\omega_B t'$$
$$y'(t') = v_{||}t' + y'_0$$
$$z'(t') = z'_0 + a\sin\omega_B t'$$

We can now use the inverse Lorentz transformations to get the trajectory of the particle in the unprimed frame.

PROBLEM 12.12 An infinitely long conducting wire or radius r carries a constant current I and charge density zero as seen by a stationary observer. The current is caused by a stream of electrons of uniform density moving relativistically at a velocity U. A second observer travels with a speed v with respect to the wire. Calculate the electromagnetic field observed by the moving observer.

PROBLEM 12.13 The uniformly distributed charge per unit length in an infinite ion beam of constant circular cross section is λ. Calculate the force on a single ion of charge Q located at a distance r from the axis assuming that the beam radius is greater than R. Assume that all the ions move with the same velocity v.

PROBLEM 12.14 In an inertial frame S, the position r_Q of a point charge Q changes as $\vec{r}_Q = vt\hat{k}$. The particle crosses the origin at time $t = 0$. In moving frame S' where the charge is at rest at the origin, the electric field is given by Coulomb's Law, that is

$$\vec{E}' = \frac{Q\vec{r}'}{4\pi\varepsilon_0 r'^3}$$

Show that in frame S, at time $t = 0$, the electric field is given by

$$\vec{E} = \frac{Q}{4\pi\varepsilon_0}\frac{1-\beta^2}{(1-\beta^2\sin^2\theta)^{3/2}}\frac{\vec{r}}{r^3}$$

where θ is the polar angle.

SUMMARY

- An inertial frame is defined as those in which Newton's laws are valid. All frames which move with a uniform velocity with respect to an inertial frame are inertial. One can use the notion of a frame which is in uniform motion with respect to the fixed stars to define an inertial frame.
- Galilean transformations give us the relationship between the space and time coordinates of two inertial frames. Time is an absolute entity in Galilean physics.

- Electromagnetic waves were believed to be propagated in a medium called aether. Michleson and Morley attempted to observe the change in the velocity of light due to the motion in this medium and failed to do so.
- Einstein proposed a Theory of Relativity which was based on two postulates—that the laws of physics remain the same in all inertial frames and that the speed of light is the same in all inertial frames.
- Galilean transformations are not consistent with the second postulate of Einstein. The correct transformation equations, which are consistent are called Lorentz transformations.
- Time becomes a relative entity in Einstein's theory. Simultaneity, which was absolute in Galilean physics becomes relative.
- The consequences of Lorentz transformations include time dilation and length contraction. Causality will be violated if there are signals which travel faster than the speed of light.
- The law of addition of velocities which follows from Lorentz transformations ensures that the magnitude of any velocity cannot exceed the speed of light.
- Energy and momentum are redefined in Einstein's theory. The new definition of energy includes a rest energy which is the energy of an object even if it is at rest.
- Electric and magnetic fields also transform under Lorentz transformations. These transformation equations lead to certain invariants which are combinations of electric and magnetic fields.

CONCEPTUAL QUESTIONS

1. A particle of rest mass m_0 travels at a speed of $v = 0.2c$. How fast must the particle move for its momentum to be twice its original momentum?
   a. 0.4c
   b. 0.38c
   c. 099c
   d. None of the above

2. A particle moves with a speed of less than $0.5c$. If its speed is increased by twice, by what factor does its momentum increase?
   a. Less than 2
   b. More than 2
   c. Twice

3. Determine the difference in kinetic energy of an electron travelling at 0.99 c and one at $0.9999c$. Use Newtonian mechanics first and then relativity.

4. A proton is moving with a momentum of 3 GeV/c. The rest mass of the proton is 938 MeV/c^2. The velocity of the proton relative to the observer is
   a. 0.31c
   b. 0.33c

c. 0.91c
d. 0.95c
e. 3.1c

5. Which of the following quantities is an invariant, i.e., the same in all Lorentz frames?
a. Δt
b. Δx
c. Velocity $\vec{v}$
d. $c^2(\Delta t)^2 - (\Delta x)^2$

6. A clock in a spaceship measures a time interval of 1.0 seconds. If the spaceship is moving with a velocity of 0.86 c w.r.t. the observer on earth, the time period observed by the observer on earth is
a. 0.86 seconds
b. 1.00 seconds
c. 1.25 seconds
d. 1.77 seconds
e. 1.96 seconds

7. A star is 4.22 light years away. (A light year is the distance travelled by light in 1 year). A spaceship can travel at a speed .9 c. If you were in the spaceship, how long would it take you, in your frame, to travel to the star?
a. 2.04 years
b. 2.29 years
c. 3.42 years
d. 3.80 years
e. 4.22 years

8. An electron has a speed 0.99 c. What are its total energy, kinetic energy and momentum? Given that the rest mass of the electron is 0.511 MeV/c^2.

9. A spaceship approaching earth at 0.9 c fires a missile towards earth with a speed of 0.5 c relative to the spaceship. What is the speed of the missile as observed from the earth?

10. How much work is required to accelerate a proton from rest to a speed of 0.997 c?

PROBLEMS

1. An observer at rest in an inertial frame S' moves along the x direction past an inertial frame S at a speed of $v = 0.6c$. A ball is dropped in S and this according to the observer takes 1.5 seconds to fall to the ground. How long does the ball take to hit the ground for an observer in S?

2. The half life of pions at rest is 1.8×10^{-9} seconds. A beam of pions has a speed of 0.9 c w.r.t the laboratory frame. How long does it take for half the pions in the beam to decay as observed in the laboratory frame? What is the distance travelled by the pions in this time?

3. A space craft travels at a speed of 2×10^4 m/s. How much longer is a day on earth as opposed to the day observed by an observer on the spacecraft?

4. A meter rod is moving parallel to itself w.r.t. an observer. To the observer, the rod appear to be only 50 cm long. How fast is the rod moving w.r.t the observer? How long will it take to pass the observer?

5. In an inertial frame S, three events, 1, 2, 3 occur at coordinates (x, y, z) which are given by (5 m, 7 m, 4 m), (9 m, 5 m, 4 m) and (2 m, 2 m, 2 m) respectively at times $t_1 = 0$, $t_2 = (\frac{2\text{m}}{c})$ sec and $t_3 = (\frac{5\text{m}}{c})$ sec. Determine if there is another inertial frame S' moving with respect to S in the $x-$ direction such that events 1 and 2 are simultaneous. If so, in S' how much later will event 3 occur as compared with the first two events.

6. An observer J is stationed in a spacecraft which is an inertial frame. She sees two other spacecrafts: P, moving with a speed $0.99c$ relative to her and Q, moving with the same speed but at right angles to P. As seen by an observer in P, what would be the speed of Q?

7. An electron moves with a speed of 1.8×10^8 m/s. Find its total energy and its kinetic energy?

8. An electron has a kinetic energy of 0.3 MeV. Find its speed according to Newtonian and relativistic mechanics.

9. Show that a photon (which has zero rest mass and speed = c) howsoever energetic, cannot decay into an electron and a positron (positron is a positively charged particle with the same mass as the electron).

10. A proton of rest mass m collides with another proton. In this process, four particles of mass m are created and all of them have the same speed. Show that the kinetic energy of the incoming proton in this process is 6 mc^2?

11. Find the total energy of a neutron with rest mass = 940 MeV/c^2 whose momentum is 1000 MeV/c.

12. In a high energy experiment, a pi-meson of rest mass $m_\pi = 140\ \text{MeV/c}^2$ and energy 500 MeV is shot at a proton of rest mass $m_P = 930\ \text{MeV/c}^2$ which is at rest. The aim of this experiment was to produce a Δ particle which has a rest mass of $m_\Delta = 1230\ \text{MeV/c}^2$ through the reaction

$$\pi + p \to \Delta + \pi$$

Estimate whether the reaction could take place.

13. A neutral Kaon (a kind of subatomic particle) of mass m_K is at rest and decays into a pair of pion and anti-pion, each with a mass M_π. Find the kinetic energy of each of the pions.

14. A particle of mass m with momentum mc is fired at a stationary target of mass $2\sqrt{2}m$. The two particles then stick together and form a single entity. What is the speed of the incoming particle before the collision and the single entity formed after the collision?

15. The electron gun in a cathode ray tube produces electrons essentially with zero kinetic energy. The electrons are accelerated through a potential difference of 5, 000 V. Calculate the kinetic energy acquired by the electrons, their speed and momentum both using Newtonian mechanics and relativity.

16. In a certain reference frame, a static, uniform, electric field E_0 is parallel to the x axis, and a static, uniform magnetic field $B_0 = 2E_0$ lies in the $x - y$ plane, making an angle θ with the x axis. Determine the relative velocity of a reference frame in which the electric and magnetic fields are parallel. What are the fields in that frame for $\theta \ll 1$?

17. In a certain inertial frame, both the electric and magnetic fields are uniform and make an angle of 85° with each other. The electric field has has a magnitude $|\vec{E}| = \sqrt{5}c|\vec{B}|$. The same fields when observed in a second inertial frame are such that the magnitude of the $\vec{E}$ is $3c|\vec{B}|$. What will be

a. The relative magnitudes of the electric and magnetic in this frame, i.e., relative to their respective values in the first frame and

b. The angle between the two fields in the second frame.

18. A wire is uncharged in the frame S and carries a current $I = 10$ Amps which results from a line of positive charges moving to the right (along the x-axis) with speed $V = 0.8c$ and a line of negative charges at rest. What is the net electric charge density of the wire (in C/m) in a frame S' which is moving along the x-axis with velocity $V = 0.8c$. What is the current (in Amps) carried by the wire in the S' frame?

A
System of Units in Electrodynamics

Although the SI system of units has gained widespread acceptance, there still exist in literature, treatment of electrodynamics in other system of units. The dictionary for translating quantities and equations between the systems are given below.

System of Units	Coulomb's Law	(B, H) & (E, D) Relation	Maxwell's Eqns. in Vacuum	Lorentz Force
Gaussian	$\vec{F} = \frac{Q_1 Q_2 \vec{R}}{R^3}$	$\vec{D} = \vec{E} + 4\pi\vec{P}$ $\vec{H} = \vec{B} - 4\pi\vec{M}$	$\vec{\nabla}\cdot\vec{D} = 4\pi\rho$ $\vec{\nabla}\cdot\vec{B} = 0$ $\vec{\nabla}\times\vec{E} = -\frac{1}{c}\frac{\partial\vec{B}}{\partial t}$ $\vec{\nabla}\times\vec{H} =$ $(\frac{4\pi}{c})\vec{j} + \frac{1}{c}\frac{\partial\vec{D}}{\partial t}$	$\frac{Q}{c}(\vec{E} + \vec{v}\times\vec{B})$
esu	$\vec{F} = \frac{Q_1 Q_2 \vec{R}}{R^3}$	$\vec{D} = \vec{E} + 4\pi\vec{P}$ $\vec{H} = c^2\vec{B} - 4\pi\vec{M}$	$\vec{\nabla}\cdot\vec{D} = 4\pi\rho$ $\vec{\nabla}\cdot\vec{B} = 0$ $\vec{\nabla}\times\vec{E} = -\frac{\partial\vec{B}}{\partial t}$ $\vec{\nabla}\times\vec{H} = 4\pi\vec{j} + \frac{\partial\vec{D}}{\partial t}$	$Q(\vec{E} + \vec{v}\times\vec{B})$
emu	$\vec{F} = c^2\frac{Q_1 Q_2 \vec{R}}{R^3}$	$\vec{D} = \frac{\vec{E}}{c^2} + 4\pi\vec{P}$ $\vec{H} = \vec{B} - 4\pi\vec{M}$	$\vec{\nabla}\cdot\vec{D} = 4\pi\rho$ $\vec{\nabla}\cdot\vec{B} = 0$ $\vec{\nabla}\times\vec{E} = -\frac{\partial\vec{B}}{\partial t}$ $\vec{\nabla}\times\vec{H} = 4\pi\vec{j} + \frac{\partial\vec{D}}{\partial t}$	$Q(\vec{E} + \vec{v}\times\vec{B})$
SI	$\vec{F} = \frac{Q_1 Q_2 \vec{R}}{4\pi\epsilon_0 R^3}$	$\vec{D} = \epsilon_0\vec{E} + \vec{P}$ $\vec{H} = \frac{\vec{B}}{\mu_0} - \vec{M}$	$\vec{\nabla}\cdot\vec{D} = \rho$ $\vec{\nabla}\cdot\vec{B} = 0$ $\vec{\nabla}\times\vec{E} = -\frac{\partial\vec{B}}{\partial t}$ $\vec{\nabla}\times\vec{H} = \vec{j} + \frac{\partial\vec{D}}{\partial t}$	$Q(\vec{E} + \vec{v}\times\vec{B})$

Values of Constants in SI Units in this Book

$\epsilon_0 = 8.85 \times 10^{-12}\ \mathrm{C^2N^{-1}m^{-2}}$
$\mu_0 = 4\pi \times 10^{-7}\ \mathrm{NA^{-2}}$
$c = 2.998 \times 10^8\ \mathrm{ms^{-1}}$
Electron mass = 9.11×10^{-31} kg.
Electron charge = 1.60×10^{-19} C

Index