Electromagnetism

Book 1

An introduction to Maxwell’s equations

John Bolton
SMT359 Course Team

Course Team Chair
Stuart Freake

Academic Editors
John Bolton, Nicholas Braithwaite, Stuart Freake, Tom Smith

Authors
John Bolton, Nicholas Braithwaite, Stuart Freake, Bob Lambourne, Tom Smith, Mike Thorpe

Consultants
Derek Capper, Andrew Coates, Andrew Conway, Alan Durrant, Allister Forrest, David Grimes, Ian Halliday, Craig McFarlane, Robin Preston, Gillian Stansfield, Steve Swithenby, Stan Zochowski

Course Manager
Michael Watkins

Course Team Assistant
Tracey Woodcraft

LTS, Project Manager
Rafael Hidalgo

Editors
Peter Twomey, Alison Cadle, Rebecca Graham

TeX Specialist
Jonathan Fine

Graphic Design Advisors
Mandy Anton, Sarah Hofton

Graphic Artists
Roger Courthold, Steve Best, Sarah Hack

Picture Researchers/Copyrights
Lydia Eaton, Martin Keeling

Software Designers
Fiona Thomson, Will Rawes

Video Producers
Owen Horn, Martin Chiverton

External Assessor
Don Davis (University College, London)

This publication forms part of an Open University course SMT359 Electromagnetism. The complete list of texts which make up this course can be found at the back. Details of this and other Open University courses can be obtained from the Student Registration and Enquiry Service, The Open University, PO Box 197, Milton Keynes, MK7 6BJ, United Kingdom: tel. +44 (0)870 333 4340, email general-enquiries@open.ac.uk

Alternatively, you may visit the Open University website at http://www.open.ac.uk where you can learn more about the wide range of courses and packs offered at all levels by The Open University.

To purchase a selection of Open University course materials visit http://www.ouw.co.uk or contact Open University Worldwide, Michael Young Building, Walton Hall, Milton Keynes MK7 6AA, United Kingdom for a brochure. tel. +44 (0)1908 858785; fax +44 (0)1908 858787; email ouwenq@open.ac.uk

The Open University
Walton Hall, Milton Keynes
MK7 6AA

First published 2006. Copyright © 2006. The Open University

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, transmitted or utilised in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without written permission from the publisher or a licence from the Copyright Licensing Agency Ltd. Details of such licences (for reprographic reproduction) may be obtained from the Copyright Licensing Agency Ltd of 90 Tottenham Court Road, London W1T 4LP

Open University course materials may also be made available in electronic formats for use by students of the University. All rights, including copyright and related rights and database rights, in electronic course materials and their contents are owned by or licensed to The Open University, or otherwise used by The Open University as permitted by applicable law.

In using electronic course materials and their contents you agree that your use will be solely for the purposes of following an Open University course of study or otherwise as licensed by The Open University or its assigns.

Except as permitted above you undertake not to copy, store in any medium (including electronic storage or use in a website), distribute, transmit or retransmit, broadcast, modify or show in public such electronic materials in whole or in part without the prior written consent of The Open University or in accordance with the Copyright, Designs and Patents Act 1988.

Edited and designed by The Open University.

Typeset at The Open University.

Printed and bound in the United Kingdom at the University Press, Cambridge.

ISBN 0 7492 6985 5
# Contents

## AN INTRODUCTION TO MAXWELL’S EQUATIONS

### Introduction

- The founders of electromagnetism
- Why study electromagnetism?
- How to study this book

### Part I: The physics of electromagnetism

#### Chapter 1 Electric forces and fields

- 1.1 Electric charge
- 1.2 Electromagnetic forces
- 1.3 Electric and magnetic forces
- 1.4 Coulomb’s law
  - 1.4.1 Adding electrostatic forces
  - 1.4.2 Evidence for Coulomb’s law
  - 1.4.3 Limitations of Coulomb’s law
- 1.5 Electric fields
  - 1.5.1 Arrow maps and field lines
  - 1.5.2 Symmetry arguments
  - 1.5.3 Typical electric field values

#### Chapter 2 Gauss’s law

- 2.1 Charge density and electric flux
- 2.2 The road to Gauss’s law
- 2.3 Putting Gauss’s law to use
  - 2.3.1 Spherical symmetry
  - 2.3.2 Cylindrical symmetry
  - 2.3.3 Planar symmetry
- 2.4 The differential version of Gauss’s law

#### Chapter 3 Magnetic forces and fields

- 3.1 Current and current density
- 3.2 The Biot–Savart force law
- 3.3 Magnetic fields due to steady currents
  - 3.3.1 Splitting the Biot–Savart force law
  - 3.3.2 Visualizing magnetic fields
## Contents

3.3.3 Typical magnetic field values 77  
3.4 The Lorentz force law 79  
3.5 Afterword on Newton’s third law 84

**Chapter 4 Ampère’s law** 87  
4.1 The no-monopole law 87  
4.2 Ampère’s law 91  
4.3 Putting Ampère’s law to use 97  
  4.3.1 Cylinders and tubes 97  
  4.3.2 The reflection of electric and magnetic fields 100  
  4.3.3 Solenoids and toroidal solenoids 106  
  4.3.4 A hidden consequence of Ampère’s law 109  
4.4 The differential version of Ampère’s law 110

**Chapter 5 Electrostatic potential** 114  
Introduction 114  
5.1 Electrostatic fields are conservative 116  
  5.1.1 Shielding in a cavity 120  
5.2 The electrostatic potential 122  
  5.2.1 Introducing the electrostatic potential 122  
  5.2.2 The conservation of energy 124  
  5.2.3 Recovering the electrostatic field 125  
5.3 Adding electrostatic potentials 126  
5.4 Potential on conductors and capacitors 131  
  5.4.1 Capacitance 133  
  5.4.2 Energy in fields 135

**Chapter 6 Electromagnetic induction** 138  
Introduction 138  
6.1 Current flow in circuits 140  
  6.1.1 Non-conservative electric fields 140  
  6.1.2 Voltage drop and emf 141  
  6.1.3 Ohm’s law 145  
6.2 Induction in a stationary circuit 146  
  6.2.1 Faraday’s law 146  
  6.2.2 The sign in Faraday’s law 151  
  6.2.3 Electromagnetic induction in action 152
Contents

8.5.1 Volume integrals 216
8.5.2 Surface integrals 220
8.6 The divergence of a vector field 223
  8.6.1 The additivity of flux 223
  8.6.2 The divergence theorem 224
  8.6.3 Divergence as a derivative 225
  8.6.4 Divergence in other coordinate systems 227
8.7 Line integrals 228
8.8 The curl of a vector field 231
  8.8.1 A convention for perimeters of open surfaces 232
  8.8.2 The additivity of circulation 232
  8.8.3 The curl theorem 233
  8.8.4 Curl as a derivative 235
  8.8.5 Curl in other coordinate systems 237
8.9 Conservative vector fields 238
  8.9.1 Zero circulation and path-independence 238
  8.9.2 The scalar potential field 240
  8.9.3 The gradient of a scalar field 241
  8.9.4 More about the gradient 243
  8.9.5 Irrotational vector fields 245
  8.9.6 Integrating fields — a summary 247
8.10 Further topics in vector calculus 248
  8.10.1 Del notation 248
  8.10.2 Vector calculus identities 249
  8.10.3 The Laplacian operator 250

Solutions 251

Acknowledgements 272

Index 273
Introduction

Founders of electromagnetism

Michael Faraday came from a poor family; at times his weekly diet consisted of no more than a loaf of bread. He was apprenticed to a bookbinder and became aware of science by reading books and attending public lectures at the Royal Institution. He secured his first job in science by taking lecture notes, binding them and presenting them to the lecturer, Humphry Davy, who took him on as an assistant. But Faraday was endlessly inquisitive and soon explored on his own initiative. By the early 1830s, he had discovered electromagnetic induction, built the first transformer, established the principle of the electric motor and produced the first continuous electrical generator. In order to account for his observations, Faraday introduced the concept of a field.

James Clerk Maxwell produced a unified theory of the electromagnetic field and used it to show that light is a type of electromagnetic wave. This prediction dates from the early 1860s when Maxwell was at King’s College, London. Shortly afterwards Maxwell decided to retire to his family estate in Galloway in order to concentrate on research, unhindered by other duties. He was lured out of retirement in 1871, when he became the first professor of experimental physics in the Cavendish Laboratory, Cambridge. Given Maxwell’s present status as one of the greatest of all physicists, it is astonishing to learn that he was the third choice for this job. Incidentally, Clerk Maxwell (without a hyphen) is a surname; Maxwell’s father, John Clerk, simply appended ‘Maxwell’ to his own name in order to smooth a legal transaction.
Why study electromagnetism?

Electromagnetism was slow to get going. About a century after Newton published the law of universal gravitation, Coulomb discovered the law of electrostatic force, but it took another century before all the laws of electromagnetism were established. In retrospect, this slow development of electromagnetism is not surprising. Newton provided a mechanical world-view which worked marvellously well in the context of gravity, but which struggled to explain the more subtle effects of electricity and magnetism. It turned out that a key idea was missing — that of a field. Through long experience in the laboratory, Michael Faraday convinced himself that magnetic fields form part of the fabric of the world, every bit as real as particles. He explained the results of his experiments in terms of fields, drawing diagrams to help him visualize them. But the mathematics needed to describe fields was a significant hurdle, which Faraday was ill-equipped to surmount.

The full power of Faraday’s field concept was revealed by James Clerk Maxwell, who deliberately chose to read Faraday’s Experimental Researches in Electricity before applying any mathematics to the subject. In 1864, Maxwell distilled all the known properties of electric and magnetic fields into a set of four equations, known as Maxwell’s equations. These equations led to one of the greatest discoveries in science — the realization that light is an electromagnetic wave. The Earth is bathed in light from the Sun. The propagation of this light, the blueness of the sky, the sparkling of ocean spray, the colours of rainbows and butterfly wings, can all be explained by Maxwell’s equations. It is not every day that a whole branch of physics is merged with another, yet this is what Maxwell achieved — optics became a branch of electromagnetism. Before long, it became clear that electromagnetic waves also exist beyond the narrow band detected by our eyes. Radio waves and X-rays are other types of electromagnetic wave, important in engineering, medicine and astronomy. No wonder that Maxwell’s publication of the laws of electromagnetism has been described as the most significant event of the nineteenth century. These laws are the main subject of this book.

There are many good reasons for studying electromagnetism, and therefore Maxwell’s equations, in depth. Your motivation might come from a desire to understand fundamental physics, applications in science, applications in technology and medicine, you might be interested in seeing how mathematics is used to explain physical effects. Let me say a few words about each of these motives.

1 Fundamental physics Electromagnetism is one of four fundamental forces of Nature — the others being gravity, the strong nuclear force and the weak nuclear force. Of these forces, electromagnetism is the one that physicists understand best. Nuclear forces are complicated and gravity is hard to reconcile with quantum mechanics, but physicists are very confident that they know how electromagnetism works. Part of the importance of electromagnetism stems from the fact that it is a theory of fields. Electromagnetism was the first, and remains the most familiar, theory of fields. Most fundamental physics is about fields of various kinds, so electromagnetism has provided a sort of template from which other, more elaborate, theories have grown.

Combining electromagnetism with a quantum theory of fields, Richard Feynman and others developed quantum electrodynamics. This is the most precisely
confirmed physical theory ever devised; measurement and theory agree to at least 16 significant figures. The success of quantum electrodynamics has inspired the search for analogous theories of the other three forces, with the ultimate goal of obtaining a theory of everything — a super-unified theory of all four fundamental forces. Such ambitions lie beyond the scope of this book, and we will not use quantum theory at all, but it is interesting to note that classical electromagnetism has nourished all these developments. Electromagnetism has led to other profound ideas too. For example, Einstein was unconvinced by the explanations of electromagnetic induction found in the textbooks of his day. Special relativity was the fruit of this scepticism. Electromagnetism continues to suggest concepts that might be useful in other areas of physics. For example, astronomers are currently trying to detect a ‘gravimagnetic’ force. This is a modification of gravity that might arise when bodies move very rapidly. It can be thought of as a gravitational analogue of the magnetic force.

2 Applications in science The forces that bind atoms together in molecules and solids are electromagnetic in origin. This means that other sciences raise questions that require deep understanding of electromagnetism. In biology, for example, the membrane of a resting nerve cell has a negatively-charged inner surface and a positively-charged outer surface. Nerve impulses consist of localized reversals of this polarity, which sweep along the cell at speeds of up to 100 m s\(^{-1}\) (Figure 3). As you read this page, electromagnetic waves enter your eye. Even the transparent outer coating of the eye, the cornea, is a wonder of Nature. It is constructed from fibres of collagen, the same material that forms tendons, yet it is almost perfectly transparent. To understand the origin of this transparency, we need to know how electromagnetic waves propagate through media containing fibres, a topic that will be discussed in Book 3 of this series. Light focused on the back of the eye stimulates nerve impulses which propagate to the visual cortex of the brain, and in a complicated and poorly understood way, stimulate other areas of the brain. All of this activity involves electrical signals, subject to the laws of electromagnetism.

**Figure 3** A nerve impulse is a reversal of polarity across the membrane of a nerve cell. It sweeps along the cell, accompanied by a blip in the voltage difference across the membrane and sustained by flows of sodium ions (Na\(^+\)) and potassium ions (K\(^+\)) through voltage-sensitive channels in the membrane.

Step out of the house and look upwards. With clear skies you should see blue sky or stars, depending on the time of day. The blueness of the sky is due to the scattering of sunlight in the atmosphere. The stars reveal their presence by emitting electromagnetic waves, which propagate across vast distances of almost empty space. Of course, the whole subject of astronomy has progressed largely thanks to our ability to detect and analyze electromagnetic waves of various
wavelengths. In polar latitudes you may see the aurora; in England you are more likely to see a rainbow (Figure 4). All these phenomena are explained by the laws of electromagnetism.

![Image of aurora borealis and rainbow](image)

(Figure 4) (a) The aurora borealis and (b) a rainbow.

3 Applications in technology and medicine Modern society relies on a mastery of electromagnetic forces, currents, fields and waves. Think, for example, of electric lighting and heating, vacuum cleaners, car ignition systems, radar, mobile phones, televisions, satellite communication, body scanners, computers and the internet. Many sources of power are available to us — coal, gas, nuclear, solar, tidal or wind. In each case, there is a technology that produces electricity from the power source and distributes it across countries or even continents. It is easy to take these things for granted, but most of us would feel terribly deprived if electricity, and all its applications, were suddenly taken away from us. We are creatures of an electrical and electronic age.

Progress shows no sign of slackening. For a few years, at least, computers will continue to improve, with faster processors, more memory and better data storage. It seems likely that electrically-driven vehicles will eventually displace our petrol-fuelled cars. Magnetically-levitated trains may become commonplace, especially if superconductors can be made to work at room temperature (Figure 5). Photonic circuits, similar to electronic circuits, but based on the propagation of electromagnetic waves rather than electrons, may be used in household devices. It is probably unwise to gaze too closely into the crystal ball, but there are certainly more surprises and delights to come, with future generations of inventors and designers continuing to exploit fundamental electromagnetic concepts and laws.

4 Using mathematics Lastly, you may be interested in the mathematics that underlies electromagnetism. Maxwell’s equations can be described in various mathematical ways. This course uses the language of vector calculus, which means that you will use the ideas of divergence, curl and gradient and integrate simple functions over volumes and surfaces and along curves. If you are coming to the subject from a background in applied mathematics, you may be pleased that electromagnetism provides a concrete setting in which to practise these mathematical skills. This will be useful if you ever study other subjects, such as fluid mechanics, which are also based on vector calculus.
How to study this book

This book is divided into two parts.

Part I: The physics of electromagnetism introduces Maxwell’s equations in the simple context of charges and currents immersed in empty space.

Part II: The mathematics of electromagnetism consists of a Mathematical Toolkit which reviews the areas of mathematics that are needed to interpret and understand Maxwell’s equations. It emphasizes the practical business of how to use vector calculus. One advantage of separating the mathematics from the physics in this way is it gives you some flexibility in studying the material, depending on your background and preferences.

• If you are confident in your mathematical preparation, or if you need some physical motivation to spark your interest in mathematics, you will prefer to begin with Part I. Following this plan, the mathematics will appear as the servant of the physics, and the motivation for each mathematical concept should be clear. At various points throughout the text you will be advised to study selected mathematical topics from Part II, and you are expected to take these detours when they appear. Watch out for the ▶ flag, which indicates a recommended detour into the Mathematical Toolkit MT).

• If you feel that mathematics is likely to be a major obstacle, or if you are happy to study mathematics without strong physical motivation, you may prefer to read Part II before starting Part I. In this way, you can concentrate on mastering important mathematical techniques before embarking on physical discussions.

The relative amounts of time you spend on physics and mathematics will depend on your background, but a typical split for this book might be 70% physics, 30% mathematics. If you plan to read the book in 11 weeks, starting with all the mathematics, you will be on schedule if you start the physics chapters at the beginning of week 4. If, like me, you prefer to study physics and mathematics together, Table 1 suggests a reasonable way of allocating your time. The study times given here allow for lengthy detours into MT. You are strongly advised to treat these detours as essential parts of your study and be prepared to spend sufficient time on them. To make precise mathematical statements, to solve problems and to pass exams, you must be familiar with the material in Part II. If you treat Part II as an optional appendix, and choose not to read it, you may only achieve an uncertain and superficial understanding of electromagnetism.

We will refer to Section 8.X in the Mathematical Toolkit as MT 8.X.

Table 1  A study plan.

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Time/weeks</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.25</td>
</tr>
<tr>
<td>2</td>
<td>2.5</td>
</tr>
<tr>
<td>3</td>
<td>1.5</td>
</tr>
<tr>
<td>4</td>
<td>1.75</td>
</tr>
<tr>
<td>5</td>
<td>1.75</td>
</tr>
<tr>
<td>6</td>
<td>1.25</td>
</tr>
<tr>
<td>7</td>
<td>1.0</td>
</tr>
</tbody>
</table>
Part I: The physics of electromagnetism

Chapter 1  Electric forces and fields

Chapter 1 uses vector notation, unit vectors and vector addition. MT 8.1 discusses these topics. Be prepared to spend 25% of your study time for this chapter on this mathematics.

1.1  Electric charge

Knowledge of electricity can be traced back to a prehistoric forest. The trees died and their resin hardened to form golden-brown pebbles. The material of these pebbles is now called amber, but the ancient Greeks called it ‘elektron’. The Greeks noticed that amber, rubbed with fur, has the ability to attract dust. This is the first recorded observation of an electric force. You can observe a similar effect by combing your hair vigorously with a plastic comb. The comb will become electrically charged and attract tiny scraps of paper. Under favourable conditions, you will see that it deflects a fine jet of water running from a tap (Figure 1.1).

More usefully, electric forces are used to guide ink particles to appropriate areas of paper in printers and photocopiers and to separate fragments of DNA in forensic science laboratories.

To interpret electric forces we need the concept of electric charge. Nowadays, we simply treat charge as a physical property that certain elementary particles possess. Electrons and protons are both charged. We shall not delve any deeper than this because charge is taken to be a primitive concept — one that is so fundamental that it cannot be explained in simpler terms. We can, however, describe its properties.

Charge is the property that allows particles to exert and experience electromagnetic forces. It comes in two types — positive and negative. For example, a proton, one of the particles in the nucleus of an atom, is positively charged, while an electron is negatively charged. Charges of the same sign repel one another while charges of opposite sign attract one another. Thus, two protons repel one another, two electrons repel one another, and an electron is attracted to a proton. The attraction between an electron and a proton is strong enough for the two particles to stay bound together, forming an atom of hydrogen. Such an atom is uncharged and is said to be electrically neutral.

Electromagnetic forces decrease with increasing separation. There is practically no attraction between an electron on the Moon and a proton on the Earth, but there is a much stronger attraction between electrons and protons in the same atom. At a given separation, it is interesting to ask how the strength of the electromagnetic force compares with that of gravity. The answer is that there is simply no contest. Electromagnetic forces are much stronger than gravitational forces. In a hydrogen atom, the electron is both electromagnetically and gravitationally attracted to the proton, but the electromagnetic attraction is $2 \times 10^{39}$ times greater than the gravitational attraction. Electromagnetic forces can also be compared with two other forces known to physicists — the strong and weak nuclear forces. At
extremely short separations, the strong and weak nuclear forces are larger than the electromagnetic force, but these forces only act over very short ranges. At separations of more than $10^{-12}$ metres, the electromagnetic force is much stronger than either of the nuclear forces. Atoms and molecules are at least 100 times larger than this, so the electromagnetic force is the only force of any significance in chemistry and biochemistry. It is the only force needed to explain the melting point of ice, the hardness of diamond or the thoughts running through your brain.

If you are impressed by the vastness of the ratio $2 \times 10^{30}$, you might wonder why gravity is noticeable at all. Paradoxically, the answer lies in the strong attraction between opposite charges, which ensures that bulk matter is normally uncharged. For example, a positively-charged nucleus tends to pull negatively-charged electrons into its vicinity, forming a neutral atom. Large aggregates of matter tend to be made up of neutral atoms and so are themselves electrically neutral. That is why the motion of planets and stars is governed by gravity; the electromagnetic forces cancel out because planets and stars are practically uncharged.

To observe charge in detail, we need to disturb the usual neutrality of matter. This is what happens in a battery for example, or in an electron microscope which creates a stream of electrons. Fortunately, the properties of charge turn out to be remarkably simple. They can be summarized as follows:

**Charge is a scalar.** A scalar quantity is one that is described by a single real number together with an appropriate unit of measurement. In the case of charge, the sign of the number indicates whether the charge is positive or negative, and its magnitude (in a given system of units) tells us how much charge is present.

**Charge is additive.** The total charge within a given region is the sum of all the charges in that region. The sum is an algebraic one, with due account taken of the signs of the charges, so a system of two charges of equal magnitudes and opposite signs has a total charge of zero.

**Charge is conserved.** The total charge of the Universe remains constant in time. Actually, it is possible to make an even stronger statement: charge is conserved *locally*. This means that the total charge in any region of space remains constant, unless charged particles flow across the boundary of the region. If a positive charge were created at one point in space and a compensating negative charge were simultaneously created at a different point, the total charge of the Universe would remain constant. However, the local conservation of charge would be violated in the regions around both points, so such a process cannot happen.

Charge conservation goes beyond the idea that every particle carries a fixed charge. In modern physics, particles can be destroyed or created. It is possible for two high-energy particles to collide and annihilate one another, giving rise to completely new particles. Figure 1.2 shows tracks made in a bubble chamber following the collision of two neutral particles. These particles annihilate one another, producing a positively-charged positron and a negatively-charged electron. The newly-created particles have charges of equal magnitude but opposite sign, so the initial value of the charge (zero) is maintained locally at the site of the collision.

**Charge is invariant.** The value of a particle’s charge is agreed on by all observers. It does not depend on the observer’s choice of coordinate system or state of motion.

\[ 1.1 \text{ Electric charge} \]

\[ \text{Figure 1.2} \quad \text{A newly created electron–positron pair.} \]
Table 1.1 Charges of some elementary particles.

<table>
<thead>
<tr>
<th>Particle</th>
<th>Charge</th>
</tr>
</thead>
<tbody>
<tr>
<td>electron</td>
<td>$-e$</td>
</tr>
<tr>
<td>muon</td>
<td>$-e$</td>
</tr>
<tr>
<td>tauon</td>
<td>$-e$</td>
</tr>
<tr>
<td>proton</td>
<td>$+e$</td>
</tr>
<tr>
<td>neutron</td>
<td>0</td>
</tr>
<tr>
<td>neutrino</td>
<td>0</td>
</tr>
<tr>
<td>photon</td>
<td>0</td>
</tr>
<tr>
<td>W$^+$ boson</td>
<td>$+e$</td>
</tr>
<tr>
<td>W$^-$ boson</td>
<td>$-e$</td>
</tr>
<tr>
<td>up quark</td>
<td>$+2e/3$</td>
</tr>
<tr>
<td>charm quark</td>
<td>$+2e/3$</td>
</tr>
<tr>
<td>top quark</td>
<td>$+2e/3$</td>
</tr>
<tr>
<td>down quark</td>
<td>$-e/3$</td>
</tr>
<tr>
<td>strange quark</td>
<td>$-e/3$</td>
</tr>
<tr>
<td>bottom quark</td>
<td>$-e/3$</td>
</tr>
</tbody>
</table>

**Charge is quantized.** Charge comes in discrete lumps. Table 1.1 shows the charges of some elementary particles expressed in terms of the charge of a proton, which is given the symbol $e$. The charge on an electron is $-e$ and, so far as we know, all isolated particles have charges that are integer multiples of $e$. Quarks have charges that are integer multiples of $e/3$, but quarks have never been observed as isolated particles. They always occur as combinations with total charge $-e$, 0 or $+e$. For example, a proton consists of three quarks of charges $2e/3$, $2e/3$ and $-e/3$. By the additivity of charge, its total charge is $e$.

It is worth emphasizing that all these properties of charge are believed to be exactly true. I say this because strange things can happen in modern physics. In relativity, for example, mass is not additive, not conserved and not invariant. By contrast, charge is believed to be strictly additive, conserved and invariant. It is a scalar quantity, and it comes in quantized lumps. So things could hardly be more straightforward. Electromagnetism has its intricacies, but charge is not one of them. Charge is a simple concept and is always easy to deal with.

**Exercise 1.1** Why does a comb become positively charged when you run it through your hair? Is your explanation consistent with the conservation of charge? Why does a charged comb attract neutral scraps of paper?

### 1.2 Electromagnetic forces

Charged particles exert electromagnetic forces on one another. Before taking a closer look at these forces, let’s briefly recall why force is an important concept in classical physics. The main reason is Newton’s second law:

$$ F = \frac{dp}{dt} $$

which describes how a force $F$ influences the momentum $p$ of a particle. In this book, of course, we are concerned with electromagnetic forces.

By a particle, I mean a scrap of matter which, at any instant, can be thought of as occupying a single point in space. An electron is usually thought of as a particle, for example. However, electrons have a property called spin which produces small magnetic effects. I therefore introduce the concept of a **point charge**. By definition, a point charge is a charged particle with absolutely no internal structure, internal motion or spin. The concept of a point charge will help us make precise definitions and statements without worrying about spin or magnetism. This is really a legal nicety, part of the small print. For most purposes, the distinction between charged particles and point charges is unimportant. I will sometimes use the word ‘charge’ as a shorthand for ‘point charge’.

One of the surprising things about electromagnetic forces is that they depend on the velocities of the particles involved. For example, Figure 1.3 shows three different situations involving two negative point charges. In case (a) particle B moves uniformly along the dashed line while particle A is at rest. In case (b) both particles are held permanently at rest. In case (c) both particles move uniformly at the same velocity, perpendicular to their line of separation. These three situations are distinguished only by the motion of the particles. Just as you would expect, the charges repel one another. But you might be surprised to learn that the
magnitude of the force on particle A is not the same in these three cases. It is
greatest in case (a), smaller in case (b) and smaller still in case (c). In terms of
symbols we can write

\[ |F_A^{(a)}| > |F_A^{(b)}| > |F_A^{(c)}|, \]

where the subscript reminds us that we are discussing the force on particle A and
the superscript refers to the case under consideration.

![Diagram](image)

**Figure 1.3** The electromagnetic force on particle A due to particle B in three
different cases.

Admittedly, the variation in \( F_A \) is usually tiny. For it to be at all significant, the
two particles must move at speeds close to that of light. However, it is not
uncommon for charges to move at such high speeds. The electrons in a heavy
atom travel at an appreciable fraction of the speed of light, and so do the electrons
in a powerful electron microscope (Figure 1.4). In an electron microscope the
reduced repulsion illustrated in Figure 1.3c has a practical consequence — the
sideways spread of the electron beam is smaller than would be predicted on the
basis of Figure 1.3b.

If many particles are involved, the variation in the force may be evident at lower
speeds. Suppose that two neutral wires, A and B, are placed side-by-side. If there
are no currents flowing through the wires there is no electromagnetic force
between them. This is not surprising because the wires are neutral. In more detail,
some electrons in the wires become detached from their atoms, leaving positive
ions behind. Each metal wire can be therefore be pictured as a lattice of positive
ions immersed in a sea of mobile electrons. The electrons in wire A are repelled
by the electrons in wire B, but are equally attracted towards the ions in wire B.
The ions in wire A are repelled by the ions in wire B, but are equally attracted
towards the electrons in wire B. All the repulsions and attractions cancel out,
leaving no net force between the wires.

Now, suppose that steady parallel currents flow through the wires. In this case, the
wires are observed to attract one another (Figure 1.5). This is not an exotic
curiosity — it is an effect that is easily observed without any special equipment.
Why does it happen? It is important to realize that an electric current in a wire is
caused by a steady flow of electrons along the wire, while the positively-charged
ions in the wire remain fixed. You can picture the flow of electrons as being rather like the flow of water through a pipe. The motion of the electrons disrupts the precise balance between attractive and repulsive forces mentioned above. For example, Figure 1.3c suggests that the repulsion between parallel streams of moving electrons will be less than had they been at rest. The reduction in repulsion is matched by a compensating reduction in attraction between the electrons and the ions so, when all the forces are added together, the net effect is that the two current-carrying wires attract one another. One interesting feature of this attraction is that the electrons drift along the wire slowly, typically at a few tenths of a millimetre per second. This is 12 orders of magnitude smaller than the speed of light, so the net force on each electron will be tiny. However, a copper wire one millimetre in diameter has around $10^{21}$ mobile electrons in each centimetre of its length. This scales up the force and produces a noticeable effect.

### 1.3 Electric and magnetic forces

To make further progress we need to introduce force laws. These laws are usually phrased in terms of electric forces and magnetic forces, so our first task is to distinguish between these two types of electromagnetic force. Naturally enough, the electromagnetic force between two stationary point charges is called an electric force, but what is a magnetic force?

No doubt you have observed the effects of magnetic forces acting on ordinary magnets — a compass needle aligning in a South-North direction or a fridge-magnet sticking to the door of a fridge. At first sight, these magnetic forces appear to have nothing to do with the forces between charged particles, but there is actually a very deep connection. It turns out that a current-carrying coil behaves exactly like a magnet. The coil, too, aligns in a South-North direction and is attracted to a fridge door. In the 1820s, André-Marie Ampère (Figure 1.6) used this analogy to suggest that ordinary magnets owe their special properties to microscopic currents circulating within their volume. From this point of view, the essential feature of a magnetic force is that it acts on electric currents. Following this insight, we see that the force between parallel currents in neutral wires should be classified as a magnetic force. An electric current is just a flow of electric charge, so we can also say that magnetic forces act on charges that are in motion. This motivates the following definition.

**The distinction between electric and magnetic forces**

The electromagnetic force on a stationary point charge is defined to be an **electric force**. A stationary point charge experiences no magnetic force.

The electromagnetic force on a moving point charge may have both electric and magnetic contributions. The electric force is defined to be the same as for a stationary point charge at the same position. The **magnetic force** is the additional electromagnetic force that occurs because the charge is moving, rather than at rest.

This definition effectively splits electromagnetic forces into electric and magnetic contributions. The electric force is felt by all point charges, whether they are
moving or not. The magnetic force is felt only by a moving point charge. A stationary point charge experiences no magnetic force. This is a convention, rather than a deep fact about Nature, but it is a very important convention which permeates the whole subject. Let’s see how it works in simple cases.

- In Figures 1.3a and b, particle A is stationary so it cannot experience a magnetic force. It experiences only an electric force. The electromagnetic repulsion is observed to be stronger in Figure 1.3a than in Figure 1.3b. This fact is interpreted by saying that particle A experiences a stronger repulsive electric force in Figure 1.3a than in Figure 1.3b.

- In Figure 1.3c, particle A is moving so it can experience both electric and magnetic forces. The electric force is independent of the motion of particle A, and is therefore identical to the enhanced electric repulsion of Figure 1.3a. Nevertheless, the electromagnetic repulsion is observed to be weaker in Figure 1.3c than in Figure 1.3b. This fact is interpreted by saying that particle A experiences an attractive magnetic force in Figure 1.3c which more than compensates for the enhanced electric repulsion.

One other point is worth noting. By definition, a stationary point charge experiences no magnetic force. But who should judge whether a charge is stationary or not? If you are in a jet plane and I am in an armchair, we are likely to disagree about such matters. Albert Einstein (Figure 1.7) was the first to realize that different observers are entitled to make their own judgements. If a point charge is stationary relative to you, then you must say that it experiences no magnetic force in your reference frame. But, if the same charge is moving relative to me, I can say that it experiences a magnetic force in my reference frame. We are both right, although we have different viewpoints! The separation of electromagnetic forces into electric and magnetic contributions depends on the choice of reference frame. Strictly speaking we should specify our choice of reference frame at the outset, but this is seldom done. Usually, we focus on phenomena observed in a laboratory and implicitly assume that our descriptions refer to a reference frame that is stationary in the laboratory.

Exercise 1.2  Do electric forces depend on the motion of the charges that feel them?

Exercise 1.3  Do electric forces depend on the motion of the charges that exert them?

Exercise 1.4  A stationary electron can experience a magnetic force if it is near a strong magnet. Does this invalidate our classification of electric and magnetic forces?

1.4 Coulomb’s law

For the rest of this chapter we restrict attention to the forces between point charges that are at rest. Our previous discussion makes it clear that this is a special case, but it is an important one, and a good place to start. Because the charges are stationary, they cannot experience magnetic forces. Moreover, the electric forces exerted by the charges are not modified by their motion. The electric forces between stationary charges are called electrostatic forces and the electrostatic
The law of force between two stationary point charges is called **Coulomb’s law**. This law can be stated as follows:

### Coulomb’s law

The electrostatic force between two stationary point charges acts along their line of separation; it is repulsive for charges of the same sign and attractive for charges of opposite sign. The magnitude of the force is proportional to the product of the charges and is inversely proportional to the square of the distance between them.

Because of its dependence on distance, Coulomb’s law is said to be an *inverse square* law of force. Our first task is to express this law in a suitable mathematical form. We might, for example, write

\[ F_{\text{repulsive}} = k_{\text{elec}} \frac{q_1 q_2}{r_{12}^2} \]

where \( q_1 \) and \( q_2 \) are the values of the two point charges, \( r_{12} \) is the distance between them and \( k_{\text{elec}} \) is a positive constant of proportionality. The quantity on the left-hand side is the repulsive force experienced by a given charge in the pair. If \( q_1 \) and \( q_2 \) have the same sign, the repulsive force is positive, indicating a genuine repulsion away from the other charge. If \( q_1 \) and \( q_2 \) have opposite signs, the repulsive force is negative, which is interpreted as an attraction towards the other charge.

This equation tells no lies, but is almost useless for systems containing more than two charges. The trouble is that the direction of the force is not represented by symbols in the equation, but by the adjectives ‘repulsive’ or ‘attractive’. But, if three charges are not in a straight line, the force experienced by one of them is a combination of forces from the other two. These forces act in different directions, in a way that the above equation cannot hope to capture. To obtain a satisfactory representation of Coulomb’s law it is *essential* to use vectors, which brings us to the first mathematical detour of this book. Remember that I am assuming that you have taken such a detour *before* continuing with the rest of the text.

▶ Read MT 8.1 now if you have not already done so.

Vectors help us to define directions in space. Suppose that point charges \( q_1 \) and \( q_2 \) are at positions \( r_1 \) and \( r_2 \) (Figure 1.8). Then the displacement vector of charge 1 from charge 2 is

\[ r_{12} = r_1 - r_2. \]

Note that I have written \( r_{12} \) with its indices in the same order as those in \( r_1 - r_2 \). This is a convenient notation, but it means that the first index marks the destination of the displacement while the second index marks its start.

The displacement vector \( r_{12} \) has both magnitude and direction. Its magnitude is

\[ r_{12} = |r_1 - r_2|, \]
which is the distance between the charges. Its direction is given by the unit vector

\[ \mathbf{r}_{12} = \frac{\mathbf{r}_1 - \mathbf{r}_2}{|\mathbf{r}_1 - \mathbf{r}_2|}, \]

which is a vector of magnitude 1 (with no units) pointing in the direction shown in Figure 1.8. This unit vector is useful because it is aligned with the direction of the electrostatic force. It allows us to express Coulomb’s law as a vector equation:

\[ \mathbf{F}_{12} = k_{\text{elec}} \frac{q_1 q_2}{r_{12}^2} \mathbf{r}_{12}. \quad (1.1) \]

The left-hand side of this equation is the electrostatic force on charge 1 due to charge 2. This is represented by the force vector, \( \mathbf{F}_{12} \), where the first index indicates the particle experiencing the force and the second index indicates the particle responsible for the force. In this notation, the force on charge 2 due to charge 1 is written as \( \mathbf{F}_{21} \). The order of indices matters here because these two forces are not the same — they point in opposite directions.

The right-hand side of Equation 1.1 is the product of the scalar factor \( k_{\text{elec}} q_1 q_2 / r_{12}^2 \) and the unit vector \( \mathbf{r}_{12} \). The scalar factor ensures that the force is proportional to the product of the charges and is inversely proportional to the square of their separation. The unit vector ensures that the force points in the appropriate direction. To see how this works, consider two charges of the same sign. Since \( k_{\text{elec}} \) is positive, the unit vector is multiplied by a positive quantity, and the force on charge 1 points in the direction of \( +\mathbf{r}_{12} \), a repulsion directly away from charge 2. If the charges have opposite signs the unit vector is multiplied by a negative quantity, and the force on charge 1 points in the direction of \( -\mathbf{r}_{12} \), an attraction directly towards charge 2. Both predictions are correct. If you are ever in doubt about the order of the indices in Coulomb’s law, you should go through an analysis like this to check that everything is consistent with the rule that like charges repel one another.

It is conventional to write the proportionality constant \( k_{\text{elec}} \) as \( 1/4\pi \varepsilon_0 \), where \( \varepsilon_0 \) is rather grandly called the permittivity of free space. The reason for including a factor \( 1/4\pi \) at this stage is that it leads to simplifications elsewhere in the subject, especially in Maxwell’s equations, as you will see later in this book. We therefore choose to write Coulomb’s law in the standard form:

\[ \mathbf{F}_{12} = \frac{1}{4\pi \varepsilon_0} \frac{q_1 q_2}{r_{12}^2} \mathbf{r}_{12}. \quad (1.2) \]

Throughout this course we will use SI units, which means that length will be measured in metres (m), mass in kilograms (kg), time in seconds (s), force in newtons (N) and charge in coulombs (C). In SI units, the proportionality constant in Coulomb’s law has the value

\[ \frac{1}{4\pi \varepsilon_0} = 8.99 \times 10^9 \text{ N m}^2 \text{ C}^{-2}. \]

This means that the electrostatic force between two particles, each carrying a charge of one coulomb and separated by a distance of one metre, is \( 8.99 \times 10^9 \text{ N} \). The large magnitude of this force tells us that one coulomb is a very large charge in the context of electrostatics. The charges on everyday objects are often measured in microcoulombs (\( 1\mu\text{C} = 1 \times 10^{-6} \text{ C} \)) and the charge on an electron is only \( -1.60 \times 10^{-19} \text{ C} \).
There is another way of writing Coulomb’s law which is useful for some purposes. Using the definition of the unit vector \( \hat{r}_{12} \), we write

\[
F_{12} = \frac{1}{4\pi \varepsilon_0} \frac{q_1 q_2}{|\mathbf{r}_1 - \mathbf{r}_2|^3} (\mathbf{r}_1 - \mathbf{r}_2).
\]  

(1.3)

The extra factor of distance in the denominator (\(|\mathbf{r}_1 - \mathbf{r}_2|^3\) rather than \(|\mathbf{r}_1 - \mathbf{r}_2|^2\)) is compensated by the factor \((\mathbf{r}_1 - \mathbf{r}_2)\) in the numerator, so the inverse-square nature of Coulomb’s law is preserved, as it must be. The main advantage of Equation 1.3 is that it avoids the need to deal with unit vectors, and this can speed up some calculations, as you will see.

**Exercise 1.5** Suppose that one gram of pure electrons is separated from another gram of pure electrons by 1.5 \(\times 10^{11}\) m (the distance between the Earth and the Sun). Estimate the magnitude of the electrostatic force between these two concentrations of charge.

**Exercise 1.6** Is Equation 1.2 consistent with Newton’s third law which states that action and reaction are equal in magnitude and opposite in direction?  

### 1.4.1 Adding electrostatic forces

So far, we have considered two point charges, labelled 1 and 2. This situation is unusual. Normally many charges are present, and each charge exerts an electrostatic force on each of the others. Fortunately, the extension to many particles is straightforward. The total electrostatic force on a given particle can be found from the following principles.

- The total electrostatic force on a charge is the vector sum of the electrostatic forces it experiences due to all other charges.
- The electrostatic force on a given charge due to another charge is given by Coulomb’s law. This depends only on the two particles under consideration and is completely unaffected by the presence of other charges.

These principles express the law of **addition of force** in the context of electrostatics. They tell us that the total electrostatic force on charge \( i \) is given by the vector sum

\[
F_i = \sum_{j \neq i} F_{ij},
\]

where \( F_{ij} \) is the electrostatic force on particle \( i \) due to particle \( j \) and the sum runs over all the particles \( j \) that exert an appreciable electrostatic force on particle \( i \). Of course, there is no term with \( j = i \) because particle \( i \) cannot exert a force on itself. Since each individual electrostatic force obeys Coulomb’s law, we conclude that

\[
F_i = \frac{1}{4\pi \varepsilon_0} \sum_{j \neq i} \frac{q_i q_j}{|\mathbf{r}_i - \mathbf{r}_j|^3} (\mathbf{r}_i - \mathbf{r}_j).
\]

(1.4)

For any given arrangement of charges, the formal summation can be expanded to give an explicit formula for the electrostatic force on particle \( i \). However, this formula can only be evaluated if particle \( i \) and all the other relevant particles \( j \)
have known charges and positions. We will assume, for the moment, that this is the case. Note that we are dealing with a sum of vectors. In general, different force contributions will point in different directions, so adding or subtracting force magnitudes will not be good enough. That is why it is essential to express Coulomb’s law in vector form.

Problems based on Coulomb’s law can be tackled in two main ways. If the arrangement of charges is two-dimensional and highly symmetric, a method based on angles and geometry may be used.

**Worked Example 1.1**

Three particles, each of charge \( q \), are held stationary at the corners of an equilateral triangle with sides of length \( d \). Find the magnitude of the electrostatic force experienced by one of these particles due to the other two.

**Solution**

Figure 1.9 shows the arrangement of charges and a suitable choice of Cartesian axes.

Let’s calculate the total electrostatic force on particle 1. This particle experiences forces \( \mathbf{F}_{12} \) and \( \mathbf{F}_{13} \) due to particles 2 and 3. By symmetry, the \( x \)-components of these forces cancel one another. Using Coulomb’s law and some trigonometry to find the \( y \)-components of the forces, we see that the total force on particle 1 points along the \( y \)-axis and has magnitude

\[
F_1 = \frac{1}{4\pi\varepsilon_0} \frac{q^2}{d^2} \cos 30^\circ + \frac{1}{4\pi\varepsilon_0} \frac{q^2}{d^2} \cos 30^\circ = \frac{\sqrt{3}}{4\pi\varepsilon_0} \frac{q^2}{d^2}.
\]

By symmetry, particles 2 and 3 each experience forces of the same magnitude, pointing directly away from the centre of the triangle.
Chapter 1  Electric forces and fields

Exercise 1.7  Three identical point charges, \( q \), are stationary at the corners of an equilateral triangle with length of side \( d \). A point charge \( Q \) is placed exactly in the centre of the equilateral triangle. What value of \( Q \) ensures that all four charges experience zero electrostatic force?

If the particles do not all lie in one plane and there is a lack of symmetry, calculations based on geometry or trigonometry become cumbersome. Fortunately, we can always represent all the vectors in component form, and use the rules of vector algebra to combine them according to the recipe given in Equation 1.4. The following example illustrates this technique.

Essential skill
Using the vector form of Coulomb’s law

Worked Example 1.2
Particles 1, 2 and 3, of charges \( q_1 = 90 \mu \text{C}, q_2 = 30 \mu \text{C} \) and \( q_3 = 20 \mu \text{C} \), are stationary at positions \( r_1 = (4e_x + 8e_y + 5e_z) \text{ m} \), \( r_2 = (4e_x - 4e_y) \text{ m} \) and \( r_3 = (1e_x + 4e_y + 5e_z) \text{ m} \). What is the total electrostatic force on particle 1 due to particles 2 and 3?

Solution
The displacement vectors of particle 1 from particles 2 and 3 are

\[
\begin{align*}
\mathbf{r}_1 - \mathbf{r}_2 &= [(4 - 4)e_x + (8 + 4)e_y + (5 - 0)e_z] \text{ m} = (12e_y + 5e_z) \text{ m}, \\
\mathbf{r}_1 - \mathbf{r}_3 &= [(4 - 1)e_x + (8 - 4)e_y + (5 - 5)e_z] \text{ m} = (3e_x + 4e_y) \text{ m},
\end{align*}
\]

and the corresponding distances are

\[
\begin{align*}
|\mathbf{r}_{12}| &= \sqrt{12^2 + 5^2} = 13 \text{ m}, \\
|\mathbf{r}_{13}| &= \sqrt{3^2 + 4^2} = 5 \text{ m}.
\end{align*}
\]

Consequently,

\[
\begin{align*}
\mathbf{F}_{12} &= \frac{1}{4\pi \varepsilon_0} \frac{q_1 q_2}{|\mathbf{r}_1 - \mathbf{r}_2|^3} (\mathbf{r}_1 - \mathbf{r}_2) \\
&= 8.99 \times 10^9 \text{ N m}^2 \text{ C}^{-2} \times \frac{9 \times 10^{-5} \text{ C} \times 3 \times 10^{-5} \text{ C}}{13^3 \text{ m}^3} \times (12e_y + 5e_z) \text{ m} \\
&= (0.133e_y + 0.055e_z) \text{ N},
\end{align*}
\]

\[
\begin{align*}
\mathbf{F}_{13} &= \frac{1}{4\pi \varepsilon_0} \frac{q_1 q_3}{|\mathbf{r}_1 - \mathbf{r}_3|^3} (\mathbf{r}_1 - \mathbf{r}_3) \\
&= 8.99 \times 10^9 \text{ N m}^2 \text{ C}^{-2} \times \frac{9 \times 10^{-5} \text{ C} \times 2 \times 10^{-5} \text{ C}}{5^3 \text{ m}^3} \times (3e_x + 4e_y) \text{ m} \\
&= (0.388e_x + 0.518e_y) \text{ N}.
\end{align*}
\]

Adding these two contributions together, we conclude that

\[
\mathbf{F}_1 = (0.39e_x + 0.65e_y + 0.055e_z) \text{ N},
\]

to two significant figures. Calculations like this are always straightforward, endangered only by lapses in concentration.

Exercise 1.8  Particle 1, of charge \( 90 \mu \text{C} \), is stationary at a point with Cartesian coordinates \( (3, 2, -1) \text{ m} \). Particle 2, of charge \( -30 \mu \text{C} \), is stationary at a point
with Cartesian coordinates \( (2, 4, 1) \) m. What is the electrostatic force on particle 1 due to particle 2?

**Exercise 1.9** Two charges, \(-16q\) and \(3q\), where \(q\) is positive, are stationary at points \((2a, 0, 0)\) and \((0, a, 0)\). Find the electrostatic force on a charge \(q\) placed at the origin \((0, 0, 0)\). What is the magnitude of this force and what is its direction (specified by a unit vector)?

## 1.4.2 Evidence for Coulomb’s law

Coulomb’s law is believed to be a fundamental law of Nature, but is this belief well-founded? Instead of simply accepting Coulomb’s law, we had better examine the evidence. Let us assume that the magnitude of the electrostatic force decreases as \(1/r^n\), where \(n\) is a constant. Various laboratory experiments have measured the value of \(n\). All have found that \(n = 2\), to within the accuracies \(\Delta n\) shown in Table 1.2. The first two measurements pre-date Coulomb’s work, but were unpublished and forgotten for many years. It is the recent results that are truly impressive. We can be very confident indeed that electrostatic forces obey an inverse square law on the everyday scale of these laboratory experiments.

As well as testing Coulomb’s law on a laboratory scale, we should also test it at very short distances and at very large distances. Evidence at short distances comes from scattering experiments. A famous experiment of this type was carried out by Rutherford who fired alpha particles at metal foils and observed that a few of them bounced back in the direction from which they had come. Rutherford guessed that the atoms in the foil must contain tiny massive nuclei. Assuming that the positively-charged alpha particles and the positively-charged nuclei repel one another according to Coulomb’s law, he was able to explain the angular distribution of scattered alpha particles. This is celebrated as the discovery of the atomic nucleus, but it also provided evidence that Coulomb’s law works on sub-atomic length scales. Similar experiments involving electron–electron scattering confirm that Coulomb’s law is accurate down to length scales of order \(10^{-12}\) m.

Perhaps the most interesting challenge to Coulomb’s law occurs at large distances. Here the available evidence is based on quantum field theory, which interprets electric forces in terms of an exchange of photons between charged particles. Quantum field theory shows that an inverse square law is just what is expected in a three-dimensional space, provided that the photon has no mass. The validity of Coulomb’s law is therefore closely linked to the massless nature of the photon. Current physical theories assume that photons are massless, but it is fair to ask what would happen if the photon had a small mass. In this case, it is believed that Coulomb’s law would be replaced by

\[
F_{12} = \frac{1}{4\pi\varepsilon_0} \frac{q_1 q_2}{r_{12}^2} \left( 1 + \frac{r_{12}}{a} \right) e^{-r_{12}/a} \mathbf{r}_{12},
\]

where \(a = h/(2\pi mc)\) has the units of length, \(h\) is Planck’s constant, \(m\) is the photon mass and \(c\) is the speed of light. For \(r_{12} \ll a\), this force is essentially the same as the Coulomb force, but it becomes significantly smaller than the Coulomb force when the separation of the particles becomes comparable to, or greater than, \(a\).

### Table 1.2 Testing Coulomb’s law in the laboratory.

<table>
<thead>
<tr>
<th>Date</th>
<th>Physicist</th>
<th>(\Delta n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1769</td>
<td>Robison</td>
<td>±0.06</td>
</tr>
<tr>
<td>1773</td>
<td>Cavendish</td>
<td>±0.03</td>
</tr>
<tr>
<td>1785</td>
<td>Coulomb</td>
<td>±0.1</td>
</tr>
<tr>
<td>1873</td>
<td>Maxwell</td>
<td>±10^{-5}</td>
</tr>
<tr>
<td>1936</td>
<td>Plimpton</td>
<td>±10^{-9}</td>
</tr>
<tr>
<td>1970</td>
<td>Bartlett</td>
<td>±10^{-13}</td>
</tr>
<tr>
<td>1971</td>
<td>Williams</td>
<td>±10^{-16}</td>
</tr>
</tbody>
</table>
A non-zero mass for a photon would have measurable consequences. For example, it would imply that different colours of light travel at different speeds, even in a vacuum. These consequences can be searched for in laboratory experiments and astronomical observations. Current data show that $m < 10^{-54} \text{kg}$, corresponding to $a > 3.5 \times 10^{11} \text{m}$, which suggests that Coulomb’s law remains valid over vast distances, comparable to the diameter of the Earth’s orbit around the Sun.

1.4.3 Limitations of Coulomb’s law

Coulomb’s law can be used provided that:

- the charges are at rest;
- the locations of all the relevant charges are known.

Let’s examine these requirements in more detail, beginning with the need for the charges to be at rest. In fact, motion of the charge that experiences the force is unimportant. Remember that any additional force that a charge experiences as a result of being in motion, rather than being at rest, is classified as a magnetic force. This means that the electric force on a given charge is unaffected by its own motion. However, the electric force does depend on the motion of other charges as illustrated in Figure 1.3. Coulomb’s law gives the electric force in the special case where the source charges are permanently at rest.

At first sight, this appears to exclude Coulomb’s law from many interesting phenomena. However, it is important to keep a sense of proportion. The corrections to Coulomb’s law caused by the motion of charges turn out to be of order $\frac{1}{2} \frac{v^2}{c^2}$, where $v$ is the speed of the charges and $c$ is the speed of light ($3.00 \times 10^8 \text{m s}^{-1}$). So, even if the charges are moving at $100 \text{km s}^{-1}$, Coulomb’s law is accurate to better than one part in $10^7$. For particles moving much slower than this, the accuracy is far greater. It is therefore reasonable to use Coulomb’s law beyond its narrow domain of exactness. For all practical purposes, you can use Coulomb’s law to calculate the electric forces between non-relativistic particles — that is, particles moving much more slowly than light.

The second restriction seems harmless and self-evident; to use Coulomb’s law we need to know the locations of all the interacting charges. In fact, this requirement is far more troublesome than the speed restriction. To see why, suppose that two positive charges, A and B, are immersed in a conducting medium such as copper or silver (Figure 1.10). Electrons are free to flow through the conductor and are naturally drawn towards the two positive charges, forming clouds of negative charge around them. This phenomenon is called screening. Now charge A is repelled by charge B and attracted by the electron cloud around cloud B. The total force felt by two positive charges is therefore less than predicted by Equation 1.2. This is not because Coulomb’s law is wrong, but because the situation inherently involves more than two charges. In principle we can use Equation 1.4 to find the total force on a positive charge, but unless we know how the electrons accumulate around the positive charges, there is not enough information to use this equation. Screening can be highly effective. Each conductor has a characteristic screening length, typically of order $10^{-9} \text{m}$. If two stationary charges in a conductor are separated by a distance that is much larger than the screening length, each experiences essentially no net force.
Something similar, if less dramatic, occurs in non-conducting media such as water or alcohol. Electrons are not free to migrate in such materials, but the introduction of foreign charges causes small displacements of charge, either through distortions of electron clouds or re-orientations of molecules. This phenomenon is called polarization. These small displacements of charge, varying with the positions of molecules, produce a non-uniform charge distribution in the medium which leads to additional forces on the foreign charges. The net effect is again a reduction in the total force on each foreign charge. The effects of polarization need not be small. The net force between two stationary charges, separated by a macroscopic distance in pure water, is about $1/80$ of the force between the same charges in empty space.

In this book we restrict attention to situations where the effects of screening are simple and the effects of polarization are negligible. Polarization can certainly be ignored if the charges are surrounded by a vacuum and are far from other materials. It is also reasonable to ignore the polarization of air. The force between two charges in air is practically the same as in a vacuum, to within 0.05%. In general, it is possible to neglect the polarization in gases, but is unreasonable to do so in liquids or solids. The next book in this series will explain what to do when polarization really matters.

### 1.5 Electric fields

Think of two stationary charges, $Q$ and $q$, interacting via Coulomb’s law in empty space. Charge $q$ seems to be aware of charge $Q$ and it is natural to suppose that information about one charge is conveyed to the other, but Coulomb’s law does not describe this flow of information at all. It simply asserts that the electric force is determined by the two charges and their locations. If Coulomb’s law were true for all charges, whether stationary or not, displacing a charge on Earth would have a small but instantaneous effect on charges on Mars. This feature is called instantaneous action at a distance. Faraday found this feature implausible. He believed that the ultimate laws of physics must be local in nature, providing relationships between quantities in each small region of space. Searching for a deeper level of understanding, he devised the concept of a field.

We now think of the space around a charge $Q$ as being subtly modified by the presence of the charge. We say that $Q$ produces an electric field in its surroundings. The electric field does not require the presence of a medium, and exists even in empty space. It varies from point to point, and decreases as we move away from $Q$. Now place a second charge $q$ somewhere in the region. We assume that $q$ responds to the electric field at its immediate location. Note that the
Electric forces and fields

Charge is the source of electric field.

Charge \( q \) does not need to know anything about the charge \( Q \). It just responds to the electric field that it experiences. This electric field happens to be due to the charge \( Q \), but that is irrelevant. If any other set of charges produced the same electric field at \( q \), the response would be exactly the same. In this description, the interaction between charges is split into two steps. First, charge \( Q \) produces an electric field in its surrounding space. Then charge \( q \) responds locally to the electric field that it encounters.

These ideas can be made precise. In general, a field is a physical quantity which, at each instant, has definite values throughout a region of space. To define the value of the electric field at a given point, we place a charge at the point and measure the electric force on it. The value of the electric field is the electric force per unit charge. In terms of symbols, the electric field \( \mathbf{E}(\mathbf{r}) \) at a point \( \mathbf{r} \) is defined by

\[
\mathbf{E}(\mathbf{r}) = \frac{\mathbf{F}}{q},
\]

where \( \mathbf{F} \) is the electric force on a charge \( q \) placed at \( \mathbf{r} \). This definition applies at all points in space, and the electric field is the function of position specified by Equation 1.5. Because its values are vectors, the electric field is a vector field. Our notation is very concise. In a Cartesian coordinate system it expands to

\[
\mathbf{E}(\mathbf{r}) = E_x(x, y, z)\mathbf{e}_x + E_y(x, y, z)\mathbf{e}_y + E_z(x, y, z)\mathbf{e}_z,
\]

where \( E_x, E_y \) and \( E_z \) are the Cartesian components of the electric field and \( x, y \) and \( z \) are the Cartesian coordinates of the point. The magnitude \( E \) of the electric field is called the electric field strength and is given in Cartesian coordinates by

\[
E = |\mathbf{E}| = \sqrt{E_x^2 + E_y^2 + E_z^2}.
\]

The charge \( q \) in Equation 1.5 is sometimes called a test charge because it tests the value of the electric field. Measurements of electric fields can be tricky because the test charge may distort the charge distribution whose field we wish to measure. This difficulty is usually avoided by taking the test charge to be small enough to create a negligible disturbance. However, the quantization of charge limits our ability to select an arbitrarily small test charge, so some disturbance may be unavoidable. This is only a practical difficulty, not a theoretical one. When we calculate an electric field from theory (using Coulomb’s law or ultimately Maxwell’s equations) we simply imagine that the source charges have fixed positions. Such is the power of thought!

**Exercise 1.10** What is the SI unit of electric field?

**Exercise 1.11** An isolated point charge \( q_0 \) is stationary at a point \( \mathbf{r}_0 \). Show that the electric field due to this charge is

\[
\mathbf{E}(\mathbf{r}) = \frac{1}{4\pi \varepsilon_0} \frac{q_0}{|\mathbf{r} - \mathbf{r}_0|^3} (\mathbf{r} - \mathbf{r}_0).
\]

Describe the nature of this electric field in words.

We can use Coulomb’s law to find the electric field due to any arrangement of stationary charges. Suppose that charges \( q_1, q_2, \ldots, q_n \) are stationary at points
r_1, r_2, \ldots, r_n$. If we introduce a test charge \( q \) at point \( r \), not coincident with any of the stationary charges, Equation 1.4 shows that the electric force on this test charge is

\[
F = \frac{1}{4\pi \varepsilon_0} \sum_j \frac{qq_j}{|r - r_j|^3}(r - r_j).
\]

Using Equation 1.5 we conclude that

\[
E(r) = \frac{1}{4\pi \varepsilon_0} \sum_j \frac{q_j}{|r - r_j|^3}(r - r_j),
\]  

where the sum is over all the charges \( q_1, q_2, \ldots, q_n \). Notice that all traces of the test charge, \( q \), have disappeared from Equation 1.7. The electric field is created by its sources, the stationary charges \( q_1, q_2, \ldots, q_n \), and Equation 1.7 tells us how the charges and positions of these sources determine the electric field. The field is defined at any point \( r \) that is not coincident with one of the source charges; at the source charges themselves, it is undefined.

Notice, too, that Equation 1.7 is a sum of terms similar to the right-hand side of Equation 1.6. In other words, the electric field due to an arrangement of charges is the vector sum of the electric fields due to the individual charges in the arrangement. This is an example of a general principle, valid for all electric fields. The principle of superposition states that, when there is more than one source of electric field, the total electric field at any point is the vector sum of the electric fields contributed by each of the sources.

You might be tempted to think that Equation 1.7 provides an alternative definition for the electric field. Please resist this temptation. Equation 1.7 is based on Coulomb’s law, so it applies only in electrostatic situations. By contrast, Equation 1.5 defines the electric field under all conditions, whether electrostatic or not. It is the definition of the electric field.

If we know the electric field at a given point \( r \), we can find the force acting on any charge \( q \) placed at that point:

\[
F = qE(r).
\]  

As stressed earlier, this is a local description. If the electric field is known at the position of a charge, the electric force on the charge can be found without further enquiry — without knowing what is happening far away. However, we still need to use Equation 1.7 to find the electric field. Being based on Coulomb’s law, this is certainly not a local description. So our attempt to avoid action at a distance is only partially successful. The language of fields seems appropriate, but our equations still bear the non-local stamp of Coulomb’s law. Nevertheless we have made a start and it is clear what should be done next. We need to find out more about the electric field in different situations. If we can understand how the values of the field in one region are related to the values in a neighbouring region we can hope to obtain a truly local theory. This will be the task of the next chapter.

**Exercise 1.12** How long does it take a proton to accelerate from rest to a speed of \( 1.0 \times 10^7 \) m s\(^{-1} \) in a constant electric field of magnitude \( 100 \) N C\(^{-1} \)?
1.5.1 Arrow maps and field lines

It is useful to have a way of visualizing electric fields. The electric field at a given point can be represented by an arrow whose tail is at the point in question. The direction of the arrow is the direction of the electric field, and the length of the arrow is proportional to the magnitude of the field. The electric field throughout a region of space is then represented by an arrow map, which is a collection of arrows displayed at a selection of points in the region. Figure 1.11a is an arrow map for the electric field generated by an isolated positive charge, and Figure 1.11b is the corresponding arrow map for an isolated negative charge.

![Arrow maps for (a) an isolated positive charge and (b) an isolated negative charge.](image)

**Figure 1.11** Arrow maps for (a) an isolated positive charge and (b) an isolated negative charge. The lengths of the arrows decrease with distance according to Coulomb’s inverse square law.

An alternative way of representing fields is often used. We draw continuous lines in such a way that the direction of each line is the same as the direction of the electric field at each point along its path. These lines are known as electric field lines and a collection of field lines is called a field line pattern. Figures 1.12a and 1.12b show the field line patterns generated by an isolated positive charge and by an isolated negative charge. The field lines tend to be closer together in regions where the field is strong and further apart in regions where the field is weak, but quantitative information about the strength of the field is better represented using an arrow map.

Both arrow maps and field line patterns are restricted to the two dimensions of a flat sheet of paper. In reality the field lines of Figure 1.12 occupy three-dimensional space, pointing radially outwards like the spines of a spherical hedgehog, rather than like the spokes of a wheel. In general, you will need to use some imagination to visualize how field line patterns sketched on paper extend into three dimensions.
Using the law of superposition, we can work out the electric fields produced by more complicated arrangements of charge, such as those shown in Figure 1.13. The field shown in Figure 1.13b is especially important. A stationary pair of oppositely-charged particles, separated by a short distance, is called an electric dipole and the field that it produces is called a dipolar electric field. Fields like this are produced by simple molecules such as hydrogen chloride, where the centre of the distribution of negatively-charged electrons does not coincide with the centre of the distribution of positively-charged nuclei.
Some simple, but important points can be noted about these patterns:

- Electric field lines radiate outwards from positive charges and converge inwards towards negative charges. This captures the idea that charges are sources of electric field.
- Close enough to any point charge the electric field is similar to that for an isolated charge. Other charges have little influence in this region because they are much further away.
- Electric field lines cannot cross, except at points where the electric field vanishes. If two field lines did cross, the electric field would have two different values at the crossing point. This would give two different predictions for the acceleration of a charged particle at the crossing point, which is unreasonable.
- A situation of a given symmetry can be described by a pattern of electric field lines with the same symmetry. For example, a spherically-symmetric charge distribution generates an electric field with a spherically-symmetric pattern of electric field lines.

### 1.5.2 Symmetry arguments

When you first see an argument based on symmetry, you might wonder whether it is watertight. It might seem like intuition, guesswork or even cheating. I want to dispel such doubts. Symmetry arguments can be made completely rigorous. What is more, they are very useful in electromagnetism. Vector fields can be complicated and it helps enormously if some of this complexity can be removed at the outset.

The basic argument goes as follows. Suppose that an arrangement of charge is rotated (for example). Then the field pattern produced by these charges must be similarly rotated. This is because the laws of physics do not single out any special directions in space, so the orientation of a field pattern is completely determined by the orientation of its sources. Now, suppose that the rotated arrangement of charge is *indistinguishable* from the original arrangement. Then the new field pattern must be indistinguishable from the original field pattern. This is because a definite source must produce a definite field. We are therefore led to the following principle:

**Symmetry principle**

- Any operation that leaves the sources of an electromagnetic field unchanged also leaves the field unchanged. The field inherits the symmetry of its sources.

One important special case is a static, spherically-symmetric distribution of charge. A charge distribution has **spherical symmetry** if it is unchanged by any rotation about any axis through its centre. We can use this symmetry to show that a spherically-symmetric stationary charge distribution produces a spherically-symmetric electric field. This means that:

- at any given point P the field is radial, pointing directly towards, or directly away from, the centre of the charge distribution;
• the magnitude of the field is the same at all points that are the same distance from the centre of the charge distribution.

These properties can be proved by eliminating the alternatives. Suppose, for example, that the electric field at the given point P points in some non-radial direction, as shown by the solid arrow in Figure 1.14. If we rotate the charge distribution about an axis that passes through the centre of the distribution and through the point P, the field also rotates, from the solid arrow to the dashed arrow in Figure 1.14. However, the charge distribution (being spherically symmetric) is unaffected by the rotation, so the electric field cannot change. It follows that the proposed non-radial field is impossible: the field must be radial. Similarly, let's consider the electric fields at two points, P and Q, the same distance from the centre of the spherical charge distribution. Suppose that these fields have different magnitudes, as shown in Figure 1.15a. Then we can rotate the charge distribution and the fields about an axis that passes through the centre of the charge distribution and a point midway between P and Q. A rotation of 180° produces the situation shown in Figure 1.15b. Again, the charge distribution is unaffected by the rotation, so the electric field cannot change. It follows that the magnitudes of the electric fields at P and Q must be the same.

Figure 1.14  Ruling out the possibility of a non-radial electric field at P for a spherically-symmetric charge distribution.

Figure 1.15  Ruling out the possibility of an electric field with different magnitudes at P and Q for a spherically symmetric charge distribution (plan view). Situation (a) is before a 180° rotation and situation (b) is after.
We are also interested in cylindrically-symmetric distributions of charge. First, note that an object is said to have **axial symmetry** if it is unchanged by any rotation about a fixed axis (the axis of symmetry). For example, a pencil with a circular cross-section has axial symmetry (Figure 1.16a). However, a pencil does not have the full symmetry of a cylinder because its sharpened end is different from its blunt end. An object is said to have **cylindrical symmetry** if, in addition to axial symmetry, it is also unchanged by a $180^\circ$ rotation about any axis that passes through the midpoint, perpendicular to the axis of symmetry. Appropriately enough, a cylinder has cylindrical symmetry (Figure 1.16b). Now, consider a stationary cylinder which is infinitely-long and uniformly-charged. This distribution of charge has cylindrical symmetry. It also has **translational symmetry** because it is unchanged by any displacement along the long axis of the cylinder. These symmetries imply that:

- At any point P, not on the central axis of symmetry, the electric field is radial (pointing directly towards, or directly away from, the axis of symmetry).
- The magnitude of the electric field is the same at all points that are the same distance from the axis of symmetry.

These properties can be verified using arguments similar to those given for a sphere. In brief, the charge distribution is unchanged by a $180^\circ$ rotation about an axis that passes through the given point P and meets the axis of symmetry at right angles. This implies that the electric field has no component along the axis of symmetry, and has only a radial component in the plane perpendicular to the axis of symmetry. The charge distribution is also unchanged by translations along the axis of symmetry and rotations around the axis of symmetry. This implies that the magnitude of the electric field does not vary as we move parallel to the cylinder, or around its axis, provided we stay a fixed distance from the axis.

Until now, we have considered distributions of charge that are stationary, but Figure 1.17 shows the electric field of a positive charge moving at a high steady velocity $v$. This field is **not** spherically symmetric, which is not surprising because the direction of motion of the charge singles out one direction in space from all the others. From the symmetry of the situation, one would expect the field of a uniformly-moving charge to be axially symmetric. However, the field shown in Figure 1.17 has more symmetry than this: it is **cylindrically symmetric** because it is unchanged by a $180^\circ$ rotation about an axis passing through the charge, perpendicular to its line of motion. An alternative way of expressing this is to say that the electric field in Figure 1.17 is unchanged by reversing the velocity of the charge, which corresponds to reversing the direction of flow of time. This turns out to be a consequence of time-reversal symmetry — a deep symmetry which applies throughout electromagnetism. Taking charge to be invariant, the principle of **time-reversal symmetry** asserts that all electromagnetic forces are unchanged by a reversal in the direction of flow of time. This principle ensures that the electric field in Figure 1.17 is unchanged by time-reversal because the electric field is the force per unit test charge, and neither the force nor the test charge changes.

All these conclusions are general consequences of symmetry, and do not rely on Coulomb’s law. This is just as well in the last case because Coulomb’s law does not apply to rapidly-moving charges. I have spelt out the details to show that symmetry arguments are respectable, but with experience you can be much more
concise. Provided a question does not explicitly ask for a symmetry argument, you may simply use a suitable form for the field and briefly indicate which type of symmetry is being assumed (e.g. spherical symmetry or axial + time-reversal symmetry).

**Exercise 1.13** Use symmetry to discuss the direction of the electric field near a stationary uniformly-charged plane sheet.

**Exercise 1.14** Describe the direction of the electric field on the central axis of a short cylinder carrying a uniform positive charge. Sketch a set of arrows indicating roughly how you would expect the electric field to vary in direction and magnitude along the central axis of the cylinder.

### 1.5.3 Typical electric field values

It is worth noting some typical values of electric fields encountered in various circumstances (Table 1.3). Some of the fields listed in this table are rapidly oscillating and cannot be discussed in the context of electrostatics. Nevertheless, a broad feeling for typical values may help you spot gross errors in calculations and appreciate the range of fields that are needed for different purposes.

**Table 1.3** Some typical electric fields.

<table>
<thead>
<tr>
<th>Context</th>
<th>$E / \text{NC}^{-1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>strong TV or radio signal</td>
<td>0.01</td>
</tr>
<tr>
<td>inside copper wire of diameter 1 mm, carrying 1 A</td>
<td>0.02</td>
</tr>
<tr>
<td>time-averaged field near transmitting mobile phone</td>
<td>40</td>
</tr>
<tr>
<td>safety guideline at radio frequencies</td>
<td>60</td>
</tr>
<tr>
<td>30 cm from a hair dryer</td>
<td>80</td>
</tr>
<tr>
<td>average static field at Earth’s surface</td>
<td>100</td>
</tr>
<tr>
<td>on an electric blanket</td>
<td>$2 \times 10^3$</td>
</tr>
<tr>
<td>30 m from a 220 kV power line</td>
<td>$3 \times 10^3$</td>
</tr>
<tr>
<td>in DNA fingerprinting procedure</td>
<td>$5 \times 10^3$</td>
</tr>
<tr>
<td>static field at Earth’s surface below a thundercloud</td>
<td>$5 \times 10^3$</td>
</tr>
<tr>
<td>safety guideline at mains frequency</td>
<td>$5 \times 10^4$</td>
</tr>
<tr>
<td>just outside a charged photocopier drum</td>
<td>$1 \times 10^5$</td>
</tr>
<tr>
<td>breakdown field in dry air</td>
<td>$3 \times 10^6$</td>
</tr>
<tr>
<td>breakdown field in PVC insulating tape</td>
<td>$2 \times 10^7$</td>
</tr>
<tr>
<td>in a powerful particle accelerator</td>
<td>$3 \times 10^7$</td>
</tr>
<tr>
<td>in a hydrogen atom</td>
<td>$5 \times 10^{11}$</td>
</tr>
<tr>
<td>in the most intensely focused laser beams</td>
<td>$6 \times 10^{13}$</td>
</tr>
<tr>
<td>quantum electrodynamic critical field</td>
<td>$1.3 \times 10^{18}$</td>
</tr>
</tbody>
</table>

Several items in Table 1.3 refer to the electric fields around electrical equipment. It is reassuring to note that these fields are generally within safety guidelines. When assessing any risk, the frequency of the electric field is an important factor, for example, radio-frequency fields are far more hazardous than low-frequency or static fields. This explains why the safety of mobile phones has been investigated in depth. The issues are complex because mobile phones reduce their signals to
the minimum needed to communicate with the network and do not transmit at all unless the user is speaking. Also, to increase the number of simultaneous users on the network, each phone transmits in brief sub-millisecond bursts interspersed by much longer intervals of radio silence. The result is that, while the peak electric field may exceed the radio-frequency guideline, the field averaged over 1 s does not.

Table 1.3 also gives values for breakdown fields. Any insulator becomes a conductor in a sufficiently high electric field. The minimum field needed to cause this transformation is called the **breakdown field** of the insulator. The breakdown field of dry air is about \(3 \times 10^6 \text{ N C}^{-1}\), but that of moist air is only about \(1 \times 10^6 \text{ N C}^{-1}\). Air always contains a few electrons and ionized molecules. Breakdown of air occurs when the electric field accelerates these charged particles rapidly enough, so that their collisions with other molecules produce more electrons and ionized molecules. An avalanche of charged particles then cuts a conducting path through the air and sparks fly (Figure 1.18). If you have ever felt the snap of static electricity, you must have been briefly exposed to an electric field of at least a million newtons per coulomb.

Everyone has heard of the Earth’s magnetic field, but the **Earth’s electric field** is less well known. Nevertheless, the Earth does have an electric field which, on average, points vertically downwards and has a magnitude of about 100 N C\(^{-1}\). This field exists because the planet’s surface carries a negative charge of \(-5 \times 10^9 \text{ C}\) while the upper atmosphere carries a compensating positive charge. The atmosphere is not a perfect insulator so the Earth’s electric field drives a small current downwards. This would neutralize the Earth’s negative charge and remove the Earth’s electric field within minutes were it not for the effects of lightning (Figure 1.19). Below a thundercloud the electric field points vertically upwards and is roughly 50 times stronger than normal — even higher in places, enough to make your hair stand on end. Lightning conducts currents upwards, from the ground to the cloud. Over the entire Earth, around 40 000 thunderstorms per day keep our planet negatively charged, maintaining a relatively constant downward electric field.
Finally, the last row of the table takes us beyond classical electromagnetism, into the domain of quantum electrodynamics. In rough terms, quantum electrodynamics tells us that electrons (of charge $-e$) and positrons (of charge $e$) are continuously created and destroyed in a vacuum. These particles have a rather shadowy existence — they come and go very rapidly and do not have the masses of real electrons and positrons. For this reason, they are called virtual particles. To conserve charge, the electrons and positrons appear and disappear in pairs. The virtual pairs exert forces on other charges so the vacuum behaves rather like a tenuous insulating medium. This leads to a number of barely measurable effects. For example, the repulsive forces felt by two closely-spaced electrons deviate very slightly from Coulomb’s law at separations below $10^{-12}$ m. Such effects are ignored throughout this course. However, something much more spectacular is predicted to occur in the presence of an enormous electric field. If a static electric field exceeds a critical value throughout a region whose linear dimensions are much larger than $10^{-12}$ m, the electron and positron in a virtual pair should be able to gain enough energy from the field to transform into a real electron and a real positron. These charged particles separate rapidly in the field and the insulating nature of the vacuum is predicted to break down at the quantum electrodynamic critical field, $1.3 \times 10^{18}$ N C$^{-1}$, just as the insulating nature of moist air breaks down at $10^{6}$ N C$^{-1}$ in a lightning flash. Nobody has observed this effect because of the enormity of the required electric field, but recent advances in lasers raise hopes for a definitive test.

By real electrons, we mean of course the stable particles found in atoms.

**Summary of Chapter 1**

**Section 1.1** Electric charge is the property that allows particles to exert and experience electromagnetic forces. Electric charge is a scalar quantity which is additive, quantized, locally conserved and invariant.

**Section 1.2** Electromagnetic forces are velocity-dependent. This is important for particles moving at speeds comparable to that of light and is also significant in neutral systems containing a large number of slowly-moving particles (e.g. current-carrying wires).

**Section 1.3** It is customary to split electromagnetic forces into electric and magnetic contributions. A stationary point charge experiences only the electric force. A moving charge experiences the same electric force as a stationary charge; any additional electromagnetic force that it experiences by virtue of being in motion, rather than being at rest, is classified as a magnetic force.

**Section 1.4** Electric forces between stationary charges are called electrostatic forces. The electrostatic force between two charges is given by Coulomb’s law:

$$\mathbf{F}_{12} = \frac{1}{4\pi\varepsilon_0} \frac{q_1 q_2}{r_{12}^2} \mathbf{r}_{12},$$

where $\mathbf{F}_{12}$ is the force on particle 1 due to particle 2 and $\mathbf{r}_{12}$ is a unit vector pointing towards particle 1 from particle 2. The electrostatic force due to a number of stationary sources is found by vector addition, using Coulomb’s law and the law of addition of force.

Coulomb’s law has been experimentally tested over a wide range of length scales. It works well enough for slowly-moving particles and for charges in gaseous...
media but it is not valid for very rapidly moving particles and its implications can be obscured by the effects of screening or polarization in liquid or solid media.

Section 1.5  The electric field $\mathbf{E}(\mathbf{r})$ is a vector field defined throughout a region of space. Its spatial variation can be visualized using an arrow map or a field line pattern. At any given point the value of the electric field is given by

$$\mathbf{E}(\mathbf{r}) = \frac{\mathbf{F}}{q},$$

where $\mathbf{F}$ is the force that would be experienced by a charge $q$ placed at the point $\mathbf{r}$. Electric fields obey the principle of superposition: the electric field due to a set of sources is the vector sum of the individual electric fields due to each source. An electric field inherits the symmetry of its sources: any operation that leaves the sources unchanged also leaves the electric field unchanged.

**Achievements from Chapter 1**

*After studying this chapter you should be able to:*

1.1  Explain the meaning of the newly defined (emboldened) terms and symbols, and use them appropriately.

1.2  Distinguish between electric, electrostatic and magnetic forces.

1.3  State Coulomb’s law in vector form and use it to find the total electrostatic force due to a small number of point charges.

1.4  Define the electric field.

1.5  State the principle of superposition for electric fields and use it to find the total electric field due to a small number of point charges.

1.6  Use symmetry principles to deduce some properties of electric fields.

*After studying MT 8.1 you should also be able to:*

1.7  Use vector notation consistently.

1.8  Carry out basic calculations in vector algebra involving components, magnitudes, unit vectors, multiplication of vectors by scalars and vector addition.
Chapter 2  Gauss’s law

Introduction

This chapter has a simple aim. It takes Coulomb’s law and expresses it in the language of vector calculus — that is, in terms of volume integrals, surface integrals and partial derivatives. There are many reasons to make this transition. Some problems that are difficult to solve using Coulomb’s law become much easier when expressed in terms of vector calculus. More importantly, vector calculus is the natural language for fields. It allows us to construct a truly local theory of electromagnetism, expressed entirely in terms of relationships between physical quantities at individual points in space, and avoiding the troublesome idea of action at a distance. Finally, there is a marvellous gift from the gods. You will remember that Coulomb’s law is an electrostatic result — it only covers situations in which the source charges are at rest. Using vector calculus, we will derive a consequence of Coulomb’s law, known as Gauss’s law. But we get more than we bargain for. Gauss’s law turns out to be true in all situations, whether the charges are stationary or not. So vector calculus gives us a way of escaping the shackles of electrostatics. This is the path that Maxwell took. Although his notation was more cumbersome than ours, he had the basic concepts of vector calculus and set out to express all the known laws of electromagnetism (and any additional laws that he might discover) in terms of these concepts.

This chapter uses the mathematical concepts of fields, partial differentiation, volume integration, surface integration and divergence, which are covered in MT 8.3–8.6. Some of this material may be revision. Even so, you should be prepared to spend at least as much time on the mathematics as on the physics, doubling your study-time on this chapter.

2.1 Charge density and electric flux

Before developing Gauss’s law, it is useful to get some preliminaries out of the way. One of the tasks we need to perform is that of finding the total charge within a given region. In principle, this is easy. Electric charge resides on particles such as electrons or protons. For macroscopic purposes, these particles can be treated as point-like objects, which are either inside or outside the region. According to the law of addition of charge, the total charge within a given region is then given by

\[ Q = \sum_i q_i, \]  

(2.1)

where \( q_i \) is the charge on particle \( i \), and the sum extends over all the particles within the region. Although adding charges is a perfectly well-defined task, it is not always a very enviable one as huge numbers of particles might be involved. For most purposes it is better to treat charge as if it were spread out continuously through space, like a fluid. A continuous distribution of charge is characterized by a charge density, which is the charge per unit volume.

▶ Read MT 8.3–8.5 now. This will be a lengthy detour, but is all essential material.
The **charge density** \( \rho(\mathbf{r}) \) at a point \( \mathbf{r} \) is defined by taking a small volume element centred on the point, adding up the charge \( \Delta Q \) within the volume element, and then dividing by the volume \( \Delta V \) of the element. That is,

\[
\rho(\mathbf{r}) = \frac{\Delta Q}{\Delta V}.
\]

For small volumes, the value of \( \Delta Q/\Delta V \) will be almost independent of the shape and size of the volume \( \Delta V \). However, the volume element must not be too small. If we are interested in the charge distribution throughout a battery, for example, we would not choose a volume element on the scale of an atomic nucleus. Such a volume element would give a large charge density if it contained a nucleus and a small charge density if it did not. The charge density would then vary wildly and rapidly in space (and also in time, if the nuclei move). Such detail is usually unnecessary and unhelpful. We assume that the volume elements are large enough to smooth out such variations, but still small enough to characterize the charge density in any small part of the system. With an appropriate choice, the charge density is a smoothly-varying function, except possibly at material boundaries where it may change abruptly.

Given a smoothly-varying charge density, we can find the total charge within a region \( V \) by integrating the charge density over the volume. The sum in Equation 2.1 can then be replaced by the volume integral

\[
Q = \int_V \rho(\mathbf{r}) \, dV, \tag{2.2}
\]

where \( V \) is the volume of interest. In some cases, integrating a charge density is more fundamental than adding point charges. For example, a single electron in an atom behaves like a continuous cloud of negative charge. To find the total charge within a small region of an atom we therefore rely on Equation 2.2 rather than Equation 2.1.

---

**Essential skill**

Exploiting symmetry to evaluate volume integrals.

**Worked Example 2.1**

A spherically-symmetric charge distribution has charge density

\[
\rho(\mathbf{r}) = Ar \quad \text{for} \quad r \leq R,
\]

where \( r \) is the distance from the origin and \( A \) is a constant. For \( r > R \) the charge density is zero. What is the total charge of this charge distribution?

**Solution**

Because the charge distribution is spherically symmetric, we subdivide it into a set of thin spherical shells, rather like the layers of an onion (Figure 2.1). The surface area of a sphere of radius \( r \) is \( 4\pi r^2 \), so a thin spherical shell with inner radius \( r \) and outer radius \( r + \delta r \) has thickness \( \delta r \) and volume \( 4\pi r^2 \delta r \). The charge density in this shell is \( Ar \), so the shell contributes a charge

\[
\delta Q = Ar \times 4\pi r^2 \delta r = 4\pi Ar^3 \delta r.
\]
The total charge is obtained by integrating over all the shells from \( r = 0 \) to \( r = R \), giving

\[
Q = \int_0^R 4\pi Ar^3 \, dr = 4\pi A \left[ \frac{r^4}{4} \right]_0^R = \pi AR^4.
\]

The other concept we need is that of the **electric flux** over a surface. Flux is discussed in MT 8.5.2; here, we just summarize the main ideas. The simplest surface to consider is a plane element. We consider a plane element located at \( r \), with area \( \Delta S \) and unit normal \( \hat{n} \) (Figure 2.2).

There are two possible unit normals to choose from, so our specification of a plane element involves the selection of one of these. The element is taken to be so small that the electric field is constant all over it. Then we define

\[
\text{electric flux over element} = E_n \, \Delta S,
\]

where \( E_n \) is the normal component of the electric field on the element (that is, the electric field in the direction of the unit normal, \( \hat{n} \)). Writing \( E_n = E \cdot \hat{n} \), we have

\[
\text{electric flux over element} = (E \cdot \hat{n}) \, \Delta S = E \cdot \Delta S,
\]

where \( \Delta S = \hat{n} \, \Delta S \) is the **oriented area** of the plane element — a vector whose magnitude is the area of the element and whose direction is that of the unit normal of the element. We are generally interested in extended surfaces. To find the electric flux over an extended surface, we divide the surface into many tiny patches, each of which can be approximated by a plane element. The unit normals of neighbouring elements are chosen to be almost parallel (rather than almost antiparallel). Then the total flux over the surface is approximated by the sum of the fluxes over all the surface elements. In the limit of vanishingly small patches this approximation becomes exact and the sum is replaced by a surface integral:

\[
\text{electric flux over an extended surface} S = \int_S E \cdot dS.
\]
One special case is very important. If the normal component of the electric field is constant all over a surface, the electric flux is just the normal component of the electric field on the surface times the surface area of the surface.

This chapter will concentrate on closed surfaces. A **closed surface** is one that forms a complete barrier between its interior and exterior regions. It is conventional to take all the unit normals on a closed surface to point outwards into the exterior region. This means that the electric flux over a closed surface is an *outward* flux. It is positive for a closed surface containing an isolated positive charge and negative for a closed surface containing an isolated negative charge.

**Exercise 2.1**  
An electric field is constant throughout a region of space which contains a cube. Show that the electric flux over the surface of the cube is equal to zero.

### 2.2 The road to Gauss’s law

Coulomb’s law tells us that the electric field of a stationary charge is radially directed, spherically symmetrical and falls off as the inverse square of the distance from the charge. We can express this behaviour in a striking geometric way.

Consider a sphere of radius $R$, centred on a point charge $q$ that is stationary at the origin. No other charges are anywhere near the sphere. We will calculate the electric flux produced by $q$ over the surface of the sphere. The sphere is a *closed* surface and we adopt the standard convention of taking its unit normals to point *outwards* into the exterior space. Then, at any point on the surface of the sphere, the normal component of the field is $q/4\pi\varepsilon_0 R^2$, a result which follows from Coulomb’s law.

Because the normal component of the field is constant over the surface of the sphere, the required surface integral is obtained by multiplying the normal component by the surface area $4\pi R^2$ of the sphere. That is,

$$\int_{\text{sphere}} E \cdot dS = \frac{q}{4\pi\varepsilon_0 R^2} \times 4\pi R^2 = \frac{q}{\varepsilon_0}.$$

This is positive for $q > 0$ and negative for $q < 0$, as expected from Figure 2.3.

**Figure 2.3**  
(a) Positive flux due to a positive point charge;  
(b) negative flux due to a negative point charge.
Although the electric flux depends on the charge at the centre of the sphere it *does not depend on the radius of the sphere*. It is easy to see why. The electric field obeys an inverse square law, decreasing as $1/R^2$, while the surface area of the sphere grows as $R^2$. These two factors cancel out in Equation 2.3, leaving the electric flux independent of the radius of the sphere.

Equation 2.3 is the simplest example of a very powerful result. **Gauss’s law** states that the electric flux over *any* closed surface $S$ is equal to the total charge enclosed by the surface, divided by $\varepsilon_0$. Moreover, any charge outside the surface makes no contribution to the electric flux over the surface, so

$$\text{electric flux over } S = \int_S \mathbf{E} \cdot d\mathbf{S} = \frac{Q}{\varepsilon_0}, \quad (2.4)$$

where $Q$ is the total charge *enclosed* by $S$. You have seen that this is true for a spherical surface centred on an isolated stationary charge, but Gauss’s law impressively extends this result to all closed surfaces and all distributions of charge. Gauss’s law is the major subject of this chapter. I shall now explain why it is true.

For the sake of intellectual honesty you should follow the derivation through, but bear in mind that the final result is much more important than the supporting argument. The proof ends on page 45.

**Start of proof**

Suppose that an isolated point charge $q$ is surrounded by an arbitrary closed surface $S$. To begin with, let’s assume that $S$ is convex, which means that it bulges outwards like a rugby ball so that, viewed from the outside, there are no

---

**Figure 2.4** Carl Friedrich Gauss discovered Gauss’s law in 1835 but did not publish it, probably because he regarded it as only one step towards a complete theory of electromagnetism, which he hoped to develop. Gauss’s discovery was published posthumously in 1867 but by this time, it had been rediscovered by Lord Kelvin and identified as a fundamental law of electromagnetism by Maxwell.

**Figure 2.5** Projecting a mesh from an inner sphere to an outer surface $S$. 

---
hollows. We imagine a small sphere, centred on the charge and contained entirely within $S$. This sphere is covered by a fine mesh which is projected from the position of the charge onto the outer surface (Figure 2.5). What is interesting is that the flux over any patch on the sphere is equal to the flux over the corresponding projected patch on the surface $S$.

To see why, let’s initially look at the simplest case, in which the outer surface is a sphere. Consider a projected patch $\Delta S_{\text{sph}}$ that is part of the surface of a large sphere centred on the charge (Figure 2.6). Such a patch is an enlargement of a patch $\Delta S_0$ on the inner sphere, similar in shape but with each of its linear dimensions enlarged by a factor $R/R_0$, where $R$ and $R_0$ are the distances of the outer and inner patches from the charge. It follows that the area of the outer patch is greater than that of the inner patch by a factor $(R/R_0)^2$. The unit normals of both patches point radially away from the charge, so the normal component of the electric field over each patch is equal to the radial component of the field. Because the field obeys an inverse square law, this is smaller over the outer patch by a factor $(R_0/R)^2$. The two factors, $(R/R_0)^2$ and $(R_0/R)^2$ cancel out, confirming that the electric fluxes over the two patches are the same.

![Figure 2.6](image)

**Figure 2.6** A patch $\Delta S_0$ on an inner sphere is projected onto a patch $\Delta S_{\text{sph}}$ on an outer sphere. Both spheres are centred on the charge $q$.

Now let’s look at the general case, in which the outer surface is *not* a sphere (Figure 2.7a). A patch $\Delta S_0$ on the inner sphere is projected onto a patch $\Delta S$ of the outer surface. This patch is not perpendicular to the radial direction from the charge, and it is not an enlargement of $\Delta S_0$, but is stretched more in some directions than others. To see whether this affects our conclusion, it is helpful to imagine a sphere, centred on the charge, whose surface passes through $\Delta S$. We then project $\Delta S_0$ onto this spherical surface, producing a patch $\Delta S_{\text{sph}}$. The two patches, $\Delta S$ and $\Delta S_{\text{sph}}$, are both projected from $\Delta S_0$ and are both the same distance from the charge, but they are inclined at different angles: $\Delta S_{\text{sph}}$ is perpendicular to the radial direction but $\Delta S$ is not. Bear in mind that these patches have been drawn large enough to see, but that our analysis will assume that they are arbitrarily small. This allows us to approximate each patch by a plane element, with negligible variation of electric field over the element, and for the dotted projection lines to be effectively parallel in the vicinity of the element. These approximations are illustrated in the enlarged view shown in Figure 2.7b.
We now compare the electric fluxes over $\Delta S$ and $\Delta S_{\text{sph}}$. First note that the normal component of the electric field over $\Delta S_{\text{sph}}$ is $E_r$, the radial component of the electric field due to $q$, while the normal component of the electric field over $\Delta S$ is

$$E_n = E_r \cos \alpha,$$

where $\alpha$ is the angle between the normals of $\Delta S$ and $\Delta S_{\text{sph}}$. Because the surface $S$ bulges outwards, $\alpha$ is an acute angle, so $E_n$ has the same sign as $E_r$ but is smaller. On the other hand, the area of $\Delta S$ is greater than that of $\Delta S_{\text{sph}}$. Elementary trigonometry in Figure 2.7b shows that

$$\Delta S = \frac{\Delta S_{\text{sph}}}{\cos \alpha}.$$

The flux over a plane element is the product of the normal component of the field times the area of the element so, when we compare the two fluxes, the factors of
\[ \cos \alpha \text{ cancel out. That is,} \]

\[
\text{flux over } \Delta S = E_n \Delta S \\
= E_r \cos \alpha \frac{\Delta S_{\text{sph}}}{\cos \alpha} \\
= E_r \Delta S_{\text{sph}} \\
= \text{flux over } \Delta S_{\text{sph}}.
\]

Our previous argument showed that the flux over \( \Delta S_{\text{sph}} \) is the same as that over \( \Delta S_0 \), so the flux over \( \Delta S \) must also be the same as that over \( \Delta S_0 \). Because the surface \( S \) bulges outwards, each of its patches has a partner in \( S_0 \), and vice versa. We therefore conclude that the total flux over \( S \) is the same as the total flux over \( S_0 \), and this, we already know, is \( q/\varepsilon_0 \).

The main part of the proof is now complete, but there are some loose ends to tidy up. So far, we have restricted attention to convex closed surfaces but Figure 2.8 shows a closed surface \( S \) of a more complex shape which contains an isolated point charge \( q \). We construct a sphere, \( S_0 \), centred on the charge and entirely inside \( S \). The sphere is again covered by arbitrarily small patches which are projected onto \( S \). In this case, however, one patch on \( S_0 \) may project onto several patches on \( S \). This requires a modification of our argument.

Consider all the patches that are produced when a given patch \( \Delta S_0 \) of the inner mesh is projected onto \( S \). For reasons described earlier, the magnitude of the flux is the same over all these patches. However, we need to think carefully about the sign of the flux in each case. In the case of a spherical or rugby-ball shaped surface all the patches contribute fluxes of the same sign (positive for \( q > 0 \) and negative for \( q < 0 \)), but this is not true for the surface in Figure 2.8. The electric field lines can enter or leave this surface, depending on the orientation of a given patch. Because the unit normals of a closed surface always point outwards, the flux is positive in regions where the field lines leave \( S \) and negative in regions where they enter it. Consequently, as we step radially outwards from the charge,
the fluxes contributed by successive projected patches cancel out in pairs. The outward journey away from the charge towards infinity must always involve an odd number of crossings of the surface, so \( \Delta S_0 \) projects onto an odd number of patches on \( S \). Because of the cancellations, the total flux contributed by all these patches reduces to the flux over the last patch in the set. Our argument then goes through as before: the flux over this last patch is equal to the flux over \( \Delta S_0 \). So, summing over all the patches on \( S \), the total flux over \( S \) is equal to the total flux over \( S_0 \), namely \( q/\varepsilon_0 \).

We should also consider the flux produced by an isolated point charge \( q \) outside a closed surface, \( S \) (Figure 2.9). To do this we again project outwards from the charge onto \( S \). Our previous argument goes through almost as before. The only important difference is that, as we step radially outwards from the charge towards infinity, we cross \( S \) an even number of times, so projecting in a given direction produces an even number of patches on \( S \). The flux contributions from these patches again cancel out in pairs so, summing over all the patches on \( S \), we obtain a total flux of zero. Charges outside a closed surface do not contribute to the flux over that surface.

Finally, all the different cases we have considered can be drawn together, using the principle of superposition and the law of additivity of charge. Suppose we have a number of point charges \( q_1, q_2, \ldots, q_n \), producing electric fields \( \mathbf{E}_1(\mathbf{r}), \mathbf{E}_2(\mathbf{r}), \ldots, \mathbf{E}_n(\mathbf{r}) \). Then the total electric field at any point is given by the vector sum \( \mathbf{E}_1(\mathbf{r}) + \mathbf{E}_2(\mathbf{r}) + \ldots + \mathbf{E}_n(\mathbf{r}) \). It follows that the total electric flux over a closed surface is the sum of the individual electric fluxes due to the charges \( q_1, q_2, \ldots, q_n \). We know that each charge outside the surface makes no contribution to the electric flux whereas each charge inside the surface contributes a flux equal to the value of the charge divided by \( \varepsilon_0 \). The total flux over the closed surface is therefore the sum of all the charges inside the surface, divided by \( \varepsilon_0 \). By the additivity of charge, this is just the total enclosed charge, \( Q \), divided by \( \varepsilon_0 \), which is the general statement of Gauss’s law, established now for all closed surfaces and all stationary distributions of charge.
Chapter 2  Gauss’s law

The scope and status of Gauss’s law

Gauss’s law is a gem, combining simplicity with impressive generality. The proof outlined above was based on three physical assumptions — Coulomb’s law, the law of addition of charge and the principle of superposition of electric fields. These assumptions were combined with some elementary geometry in three dimensions and Gauss’s law emerged as a result.

Taking the geometry, addition of charge and superposition of fields on trust, Gauss’s law is sometimes said to provide an alternative expression of Coulomb’s law. However, this is a very loose description. In fact, our proof of Gauss’s law did not use every aspect of Coulomb’s law. It only assumed that:

- In any given direction, the electric field of a point charge is radial and dies off according to an inverse square law.
- The electric flux over the surface of a sphere centred on an isolated point charge $q$ is $q/\varepsilon_0$.

These assumptions certainly follow from Coulomb’s law, but they can also be true in situations where Coulomb’s law does not apply. For example, the electric field of a uniformly-moving charge (shown in Figure 1.17) does not have the spherical symmetry required by Coulomb’s law. Even so, it turns out that this field obeys both the above assumptions and so does obey Gauss’s law.

At this point we make a bold leap of faith. We assert that Gauss’s law is true under all circumstances — for charges that are moving uniformly or non-uniformly, as well as for particles that are at rest. For stationary particles, we have shown that Gauss’s law follows from Coulomb’s law, but the final step of asserting that Gauss’s law remains valid for moving particles is taken to be a basic fact of Nature. No attempt is made to justify this fact using deeper knowledge because the universality of Gauss’s law is itself regarded as a fundamental truth. Maxwell was the first person to make this leap and to pursue its consequences, and Gauss’s law is the first of Maxwell’s four celebrated equations of electromagnetism. The ultimate justification of Gauss’s law in all its generality comes from experiment — not direct experiments that probe this particular law, but from the triumphant predictions of the whole of Maxwell’s theory.

Gauss’s law is the part of Coulomb’s law that is universally valid — it is true for charges in motion, as well as for charges at rest. This generality is made possible by the fact that Gauss’s law is less specific than Coulomb’s law; it gives the surface integral of the electric field over a closed surface, not the value of the field at any point in space. The surface integral turns out to be independent of the motion of charges, even though the electric field itself depends on this motion (compare Figures 1.12 and 1.17, for example).

Any information that is left out of Gauss’s law is contained in the rest of Maxwell’s equations, which you will meet later in this book. In favourable circumstances, however, the missing information can be deduced from the symmetry of the situation. For example, suppose we have a point charge $q$ at rest at the origin. Because charge is a scalar quantity with no directional character, a stationary point charge singles out no special direction in space. This implies that the electric field of a stationary point charge is spherically symmetrical, pointing away from, or towards, the charge, with a magnitude that depends only on the distance from the charge. If we now suppose that the charge is at the centre of a
sphere of radius \( R \), the electric flux over the surface of this sphere is \( E_r \times 4\pi R^2 \), where \( E_r \) is the radial component of the field on the spherical surface. Using Gauss’s law we conclude that
\[
E_r \times 4\pi R^2 = \frac{q}{\varepsilon_0},
\]
so
\[
E_r = \frac{q}{4\pi \varepsilon_0 R^2},
\]
and
\[
E_r = \frac{q}{4\pi \varepsilon_0 R^2} \, e_r.
\]
which is essentially Coulomb’s law, expressed in terms of the electric field. So Gauss’s law, supplemented by spherical symmetry, leads to Coulomb’s law. The assumption of a spherically-symmetric field is obvious for a stationary charge, but not for one that is moving, which explains why Coulomb’s law is restricted to the static case.

**Exercise 2.2** An isolated point charge is placed at the centre of a sphere. Is the total electric flux over the closed surface of the sphere changed by: (a) moving the charge off-centre inside the sphere; (b) moving the charge just outside the sphere; (c) splitting the charge into two fragments, both of which remain inside the sphere; (d) allowing the charge to oscillate to and fro within the sphere; (e) adding an extra charge just inside the sphere; (f) adding an extra charge just outside the sphere; (g) increasing the radius of the sphere or (h) deforming the sphere slightly?

**Exercise 2.3** If the photon had a non-zero mass, would Gauss’s law be exactly true? *(Hint: see Section 1.4.2.)*

### 2.3 Putting Gauss’s law to use

At the risk of repetition, here is a definitive statement of Gauss’s law:

**Gauss’s law**

The electric flux over any closed surface \( S \) is equal to the total charge \( Q \) enclosed by the surface, divided by \( \varepsilon_0 \). That is,
\[
\int_S \mathbf{E} \cdot d\mathbf{S} = \frac{Q}{\varepsilon_0}. \tag{2.5}
\]
This law is true for all closed surfaces, no matter what their shape, and for all distributions of charge, whether they are stationary or not. In particular, any charges outside the surface make no contribution to the electric flux over the surface. Gauss’s law cannot be extended to open surfaces because the concept of the total enclosed charge only makes sense for a surface that is closed.

The total charge inside the closed surface can be expressed as a volume integral of the charge density, so Gauss’s law can also be written as
\[
\int_S \mathbf{E} \cdot d\mathbf{S} = \frac{1}{\varepsilon_0} \int_V \rho(r) \, dV, \tag{2.6}
\]
where \( S \) is any closed surface and \( V \) is the region inside this surface. Because this statement involves surface and volume integrals, it is called the integral version of Gauss’s law.
As emphasized earlier, knowledge of the electric flux does not usually reveal the value of the electric field at individual points in space. To have any chance of finding the electric field from Gauss’s law, the situation must be highly symmetric. In practice, Gauss’s law is most valuable when the sources of the electric field have spherical symmetry, cylindrical symmetry or planar symmetry. Ideally, the direction of the electric field should be perpendicular to, or parallel to, the chosen closed surface and the magnitude of the electric field should be constant over the surface. We are free to choose whichever surface we like, but the choice had better be made wisely, with symmetry in mind. A choice of surface made for a particular application of Gauss’s law is called a Gaussian surface.

### 2.3.1 Spherical symmetry

**Essential skill**

Applying Gauss’s law in cases of spherical symmetry.

**Worked Example 2.2**

A static uniform distribution of total charge $Q$ occupies a sphere of radius $R$, centred on the origin. Find the electric field at all points (a) outside and (b) inside this distribution of charge.

**Solution**

The static uniform charge distribution has spherical symmetry, so the electric field must also have spherical symmetry. At each point it is directed towards, or away from, the centre of the sphere and has a radial component that depends only on the distance from the centre of the sphere. So

$$
\mathbf{E}(\mathbf{r}) = E_r(\mathbf{r}) \mathbf{e}_r,
$$

where $E_r(\mathbf{r})$ is the radial component at radius $r$ and $\mathbf{e}_r$ is the radial unit vector at the point $\mathbf{r}$. (Symmetry arguments like this are legitimate, safe and invaluable, as explained in Chapter 1.)

(a) Consider the field outside the sphere of charge. We exploit the spherical symmetry by choosing a spherical Gaussian surface, centred on the origin, with radius $r > R$. This closed surface contains the whole charge distribution, and therefore encloses charge $Q$. The electric field due to the charge distribution is perpendicular to the surface and has a constant normal component, $E_r(r)$, on this surface. Applying Gauss’s law, we obtain

$$
E_r(r) \times 4\pi r^2 = \frac{Q}{\varepsilon_0}
$$

so

$$
E_r(r) = \frac{Q}{4\pi \varepsilon_0 r^2} \quad \text{and} \quad \mathbf{E}(\mathbf{r}) = \frac{Q}{4\pi \varepsilon_0 r^2} \mathbf{e}_r \quad \text{(for } r \geq R).\n$$

Viewed from the outside, the sphere behaves as if all its charge were concentrated at its centre. Looking back at the derivation, it is easy to see that this result does not rely on a uniform charge density; it is true for any spherically-symmetric distribution of charge.

(b) Inside the sphere we choose a spherical Gaussian surface, centred on the origin, with $r < R$. This closed surface does not contain the whole charge.
Putting Gauss’s law to use

2.3

$Q$, but encloses a charge

$$Q_{\text{enc}} = \int_0^r \rho \times 4\pi s^2 \, ds,$$

where I have written the variable of integration as $s$ to avoid any confusion with the upper limit, $r$. The symbol used for a variable of integration has no physical significance, so this is a legitimate step.

In our case the charge density is uniform so

$$\rho = \frac{\text{charge}}{\text{volume}} = \frac{Q}{4\pi R^3/3}$$

and the enclosed charge is

$$Q_{\text{enc}} = \frac{Q}{4\pi R^3/3} \int_0^r 4\pi s^2 \, ds = Q \times \frac{r^3}{R^3}.$$  

This is just the total charge $Q$, multiplied by the ratio of the volumes of the Gaussian sphere and the whole sphere — just as you would expect for a uniform charge density. Finally, Gauss’s law gives

$$E_r(r) \times 4\pi r^2 = \frac{Q}{\varepsilon_0} \times \frac{r^3}{R^3}$$

so

$$E_r(r) = \frac{Q \, r}{4\pi \varepsilon_0 R^3} \quad \text{and} \quad E(r) = \frac{Q \, r}{4\pi \varepsilon_0 R^3} \, \mathbf{e}_r \quad (\text{for } r \leq R).$$

The question has now been fully answered, but it is always worth checking your answers. Even the most reliable computers have error-detecting codes built into their software and you should also develop the routine of checking that your answers are reasonable. In the present case, a number of points can be confirmed:

- The units are correct. This is obvious outside the sphere, where the answer is the electric field of a point charge. It is also true inside the sphere because $r/R^3$ has the same units as $1/R^2$.
- The electric field is zero at the origin. A non-zero field at the origin would be inexplicable, given the spherical symmetry of the charge distribution.
- The interior field becomes equal to the exterior field at $r = R$.
  A discontinuity would imply an infinite charge density at the surface of the sphere, which is physically unreasonable.

Of course, these checks do not guarantee that our answer is right, but they certainly help to boost our confidence in it. Figure 2.10 is a graph of $E_r$ versus $r$ for all $r$.

![Figure 2.10](image)  

A graph of $E_r$ versus $r$ for use in Worked Example 2.2.
Exercise 2.4 A static spherically-symmetric distribution of charge is centred on the origin and vanishes outside a sphere of radius $R$. Inside the sphere, the charge density at distance $r$ from the origin is

$$\rho(r) = Ar,$$

where $A$ is a constant. Find the electric field at all points (a) outside and (b) inside the distribution of charge.

Exercise 2.5 If the electric field in air becomes too great, the air undergoes breakdown, becoming a good conductor and allowing charge to leak away from objects. Show that the magnitude of static charge carried by a spherical hailstone of radius 3.0 mm cannot exceed one nanocoulomb ($1 \text{nC} = 10^{-9} \text{C}$). You may assume that the charge is distributed in a spherically symmetric way in the hailstone and that the breakdown field of damp air is $1.0 \times 10^6 \text{N C}^{-1}$.

Exercise 2.6 In fine weather there is a weak electric field pointing vertically downwards at the Earth’s surface. A typical value, averaged over the surface of the Earth, is $100 \text{N C}^{-1}$. Estimate the total charge on planet Earth (that is the total charge within its solid or liquid surface). You may take the Earth to be a perfect sphere of radius $6.6 \times 10^6 \text{m}$, with a spherically-symmetric charge distribution.

Our next application of Gauss’s law forms the basis of a famous experiment. An isolated spherical shell of radius $R$ is centred on the origin. The shell is constructed from a conducting material such as copper and given a net charge. Charge flows easily in the copper so the mutual repulsion of like charges causes the charge to spread out over the shell. Very soon, a state of equilibrium is reached in which the charge is spread out uniformly and the shell provides a static, spherically-symmetric distribution of charge. The spherical cavity inside the shell is empty and contains no charge. Under these circumstances we can show that there is no electric field inside the cavity.

We choose a spherical Gaussian surface (of radius $r < R$) inside the spherical cavity, with the same centre as the shell (Figure 2.11). There is no flux over this Gaussian surface because there is no charge inside it. We then make use of symmetry. By spherical symmetry, the magnitude of the electric field is the same at all points on the Gaussian surface. Spherical symmetry also shows that, if there is an electric field, it must be radial. No other direction is compatible with the spherical symmetry of the charge distribution. This leaves only one way of explaining the zero flux — the electric field must be zero at all points on the Gaussian surface. This argument works for all spherical Gaussian surfaces with $0 < r < R$. The point at the centre of the cavity is exempt from the argument, but spherical symmetry guarantees that the field vanishes here as well. So the electric field is predicted to be zero throughout the cavity. In 1773, Henry Cavendish tested this prediction by direct experiment and later physicists, including Maxwell himself, refined the sensitivity of this measurement. The electric field inside a hollow spherical conductor has always been below the limits of detection, providing good evidence for the validity of Gauss’s law and for the underlying inverse square law of electrostatic force.
2.3.2 Cylindrical symmetry

We now turn to situations with cylindrical symmetry. Consider first a long cylinder of radius $R$ and length $L$ with a charge $Q$ that is spread out uniformly throughout the cylinder. The charge per unit length is $\lambda = Q/L$. Close to the axis of the cylinder and far from its ends, the electric field is well-approximated by the field of an infinitely long cylinder with the same radius and the same charge density. The approximation of taking a cylinder to be infinitely long avoids the need to discuss end effects — the modifications in the field that occur near the ends of the cylinder. We will therefore restrict attention to infinitely-long cylinders. In the context of problems in electromagnetism, you may take the description of a cylinder as being long as a coded invitation to regard it as being infinitely long and hence to ignore the end effects. The electric field around an infinitely-long cylinder is cylindrically symmetric around the axis of the cylinder and does not vary along the axis of the cylinder. As shown in Figure 2.12, the field is perpendicular to the axis of the cylinder. Moreover, the radial component of the field depends only on the distance from the axis. We therefore have

$$\mathbf{E}(r) = E_r(r) \mathbf{e}_r,$$

where $E_r(r)$ is the radial component of the electric field at distance $r$ from the axis of the cylinder and $\mathbf{e}_r$ is the unit radial vector at point $r$. Note that these quantities refer to cylindrical coordinates, not spherical coordinates. So $r$ is the distance from the central axis; it is not the distance from the origin. And $\mathbf{e}_r$ is a unit vector which is perpendicular to the central axis and points directly away from it; it does not point away from the origin. There is never any ambiguity about whether we are dealing with cylinders or spheres, so the intended meaning of our symbols, and of words like radius or radial, will be clear from the context.

Figure 2.12  The electric field $\mathbf{E}$ in a plane perpendicular to a uniformly-charged infinite cylinder.
**Essential skill**

Applying Gauss’s law in cases of cylindrical symmetry.

---

**Worked Example 2.3**

Find the electric field outside a stationary uniformly-charged infinite cylinder with a charge per unit length of \( \lambda \).

**Solution**

This situation is cylindrically symmetric around the axis of the cylinder so the appropriate choice of Gaussian surface is a cylinder of radius \( r > R \) and length \( \Delta l \), with the same axis as the charged cylinder (Figure 2.13). The Gaussian surface must be closed, so it includes the two end-faces.

![Figure 2.13 Gaussian surface for Worked Example 2.3.](image)

The electric field is parallel to the end-faces of the cylinder so they contribute no flux. The curved surface of the cylinder has area \( 2\pi r \Delta l \). The electric field is perpendicular to this surface and has a constant normal component over it, so the flux contributed by the curved surface is \( E_r(r) \times 2\pi r \Delta l \). The total charge enclosed by the Gaussian surface is \( \lambda \Delta l \) so Gauss’s law gives

\[
E_r(r) \times 2\pi r \Delta l = \frac{\lambda \Delta l}{\epsilon_0}.
\]

Hence

\[
E_r(r) = \frac{\lambda}{2\pi \epsilon_0 r} \quad \text{and} \quad \mathbf{E}(r) = \frac{\lambda}{2\pi \epsilon_0 r} \mathbf{e}_r.
\]

(2.7)

One way of checking this answer is to show that its dimensions (or units) are correct. This is easily done since \( \lambda \) is a charge per unit length, so the expressions in Equation 2.7 have the same dimensions as \( Q/4\pi \epsilon_0 r^2 \), which can be recognized as the electric field of a point charge.

**Exercise 2.7** Would Equation 2.7 still be valid if the cylinder were moving at constant velocity in the direction of its own axis?
Exercise 2.8  
(a) Is the electric field outside a uniformly-charged infinite cylinder the same as the field of a uniform infinite line of charge, with the same charge per unit length, lying along the central axis of the cylinder?  
(b) Is a similar statement true for the electric field inside a uniformly-charged infinite cylinder?  

2.3.3 Planar symmetry  

Finally we discuss situations with planar symmetry. Consider a plane of area $A$ with a charge $Q$ spread uniformly over its surface. We use the term areal charge density to describe the charge per unit area. This areal charge density has the constant value $\sigma = Q/A$ all over the plane. For simplicity, we ignore any modifications in the field that occur near the edges of the plane. This is achieved by imagining the plane to be infinite in extent, without any edges.  

The electric field of a uniform infinite plane of charge has considerable symmetry. It is perpendicular to the plane, does not vary in any direction parallel to the plane and is reversed by a reflection in the plane. These symmetry properties allow us to write  

$$ E(r) = E_n \, e_n, $$  

where $E_n$ depends only on the distance of the point $r$ from the plane and $e_n$ is a unit normal to the plane, conventionally chosen to point away from the plane, towards the point $r$.  

**Worked Example 2.4**  

Find the electric field near a uniformly-charged infinite insulating plate with areal charge density $\sigma$.  

**Solution**  

An appropriate Gaussian surface is shown in Figure 2.14. This is a squat cylinder (sometimes called a pillbox) with end-faces of area $\Delta A$. The axis of the cylinder is perpendicular to the plate and straddles it symmetrically, so half of it is in front of the plate (solid lines) and half is behind (dashed lines).  

The electric field is parallel to the curved surface of the cylinder so this surface makes no contribution to the flux. However, the electric field is perpendicular to the flat ends of the cylinder and has a constant normal component $E_n$ over each end-face, so the total flux contributed by the two end-faces is $2E_n \, \Delta A$. The total charge enclosed by the cylinder is $\sigma \, \Delta A$ so Gauss’s law gives  

$$ 2E_n \, \Delta A = \frac{\sigma \, \Delta A}{\varepsilon_0}, $$  

and  

$$ E(r) = \frac{\sigma}{2\varepsilon_0} \, e_n, \quad (2.8) $$  

where $e_n$ is a unit vector pointing perpendicularly away from the plate towards $r$. This correctly gives a field pointing away from the plate for $\sigma > 0$ and a field pointing towards the plate for $\sigma < 0$.  

**Essential skill**  

Applying Gauss’s law in cases of planar symmetry.  

**Figure 2.14** A Gaussian surface for a plate of charge.
It is interesting to note that Equation 2.8 does not depend on the distance from the plate, even though such a dependence would be consistent with planar symmetry. The lack of dependence on distance is a consequence of Gauss’s law or, equivalently, the inverse square law of force. It is only exactly true for an infinite plate of charge. Nevertheless, for a finite square plate of charge, the field is almost independent of distance provided that we keep far from the edges of the plate and the distance from the plate is small compared to the plate’s lateral dimensions.

Next, we consider an isolated charged conducting plate. If the plate is large enough in its lateral dimensions, the charge distributes itself almost uniformly in the plane of the plate with areal charge density \( \sigma \). However, the charge is not distributed uniformly through the thickness of the plate. Mutual repulsion causes charge separation and produces two similar sheets of charge on opposite surfaces of the plate. In equilibrium, each of the two charge sheets has the same areal charge density, \( \sigma/2 \). This is usually called a surface charge density because it is associated with a given surface. Of course, there is a distinction between the surface charge density \( \sigma/2 \) and the areal charge density of the plate, \( \sigma \). The electric fields due to the two charge sheets cancel out inside the conducting plate. This is not surprising. In equilibrium, there can be no electric field inside a conductor. This is because any electric field in a conductor would drive a current, and there are no currents in a state of true equilibrium.

There are two alternative ways of applying Gauss’s law in this situation. First, we can use a pillbox that straddles the whole plate. The calculation then repeats that given in Worked Example 2.4. Alternatively, we can use the Gaussian surface shown in Figure 2.15 — a squat cylinder with end-faces of area \( \Delta A \), one of which is inside the plate. The flux over this end-face vanishes because there is no electric field inside the plate, while the flux over the external end-face is \( E_n \Delta A \). The pillbox contains only one of the two surface charge sheets, so it encloses a charge \( \sigma/2 \times \Delta A \). Thus, Gauss’s law gives

\[
E_n \Delta A = \frac{\sigma \Delta A}{2\varepsilon_0},
\]

which rearranges to give the same field as before.

Finally, let’s consider a case of practical importance — a capacitor. A capacitor is a device used to store electrical energy by keeping positive and negative charges separated. Capacitors are used in defibrillators that save the lives of heart-attack victims and in circuits that tune radios and televisions. The membrane of a nerve cell also acts like a capacitor whose properties affect the speed of transmission of nerve impulses. To take the simplest possible case, we will consider an empty parallel plate capacitor (Figure 2.16). This consists of a pair of parallel conducting plates, each of area \( A \), separated by a narrow gap which is empty. The plates carry opposite charges, \( +Q \) and \( -Q \), and their areal charge densities are \( \sigma = +Q/A \) and \( -\sigma = -Q/A \). We wish to know the electric field inside and just outside the capacitor.

If we keep away from the edges of the plates and are either in the narrow gap between the plates, or outside the gap but close to the plates, any edge effects can be neglected. This means that the capacitor can be modelled as having infinite plates with uniform areal charge densities \( \sigma \) and \( -\sigma \). In this situation, the electric
field can be found using Equation 2.8, together with the principle of superposition. The electric field of an infinite plane of charge does not decrease with distance. Hence the two infinite planes of charge produce fields of the same magnitude. Outside the capacitor, these fields have opposite directions and cancel out. So the field outside the capacitor vanishes. In the gap between the plates, the fields have the same direction and add together. So the field in the gap is

\[ E = \frac{\sigma}{\varepsilon_0} e_n, \]

where \( e_n \) is a unit normal, pointing from the plate with charge \( +Q \) towards the plate with charge \( -Q \). The field in the gap is therefore uniform and perpendicular to the plates.

It is instructive to take a fresh look at this problem, deriving the results more directly from Gauss’s law. Remember that, in equilibrium, there is no electric field inside either plate. Inside plate 1, the field due to charges on the surface of plate 1 must cancel the field due to plate 2. To achieve this cancellation, all the charge on plate 1 migrates to its inner surface, leaving no charge on its outer surface. The same happens on plate 2, so the charge accumulates on the inner surfaces of the plates, as shown in Figure 2.17. The distribution of charge on each plate is totally unlike that on an isolated charged plate (Figure 2.15) because the two plates in a capacitor are not isolated, but interact strongly with one another. To apply Gauss’s law in this situation, we can choose the cylindrical pillbox Gaussian surfaces of Figure 2.18. Each of these pillboxes has an end-face inside the plate, where the electric field vanishes. The total flux over the surface of each pillbox is therefore \( E_n \Delta A \), where \( E_n \) is the normal component of the electric field over the external end-face and \( \Delta A \) is the cross-sectional area of the cylinder. Pillbox (a) contains no charge, so Gauss’s law shows that \( E_n = 0 \) outside the capacitor. Pillbox (b) contains charge \( \sigma \Delta A \) so Gauss’s law gives

\[ E_n = \frac{\sigma}{\varepsilon_0}, \]

between the plates, in agreement with results obtained earlier.

**Exercise 2.9** An infinite parallel plate capacitor has uniformly-charged plates. Plate 1 has areal charge density \( \sigma_1 \) and plate 2 has areal charge density \( \sigma_2 \). (Normally a capacitor has oppositely-charged plates so \( \sigma_1 + \sigma_2 = 0 \), but this need not be the case.) What is the electric field in the gap between the plates in general?

### 2.4 The differential version of Gauss’s law

Gauss’s law, as described so far, involves integrals; the electric flux is a surface integral and the total charge enclosed by a surface is often calculated as a volume integral. This form of Gauss’s law is called the integral version of Gauss’s law. In this section you will see that Gauss’s law can be re-expressed using derivatives rather than integrals, giving a differential version of Gauss’s law.
You are familiar with the idea that mass is additive. If we divide an object into many parts, the mass of the object is the sum of the masses of its parts. This allows us to express the mass of the object as a volume integral of its density (the mass per unit volume), taken over the volume of the object.

Now, following MT 8.6, something very similar can be done for electric flux. Recall that the flux of any vector field is additive. In particular, electric flux is additive. So, if we divide a volume into many parts, the electric flux over the surface of the volume is the sum of the electric fluxes over the surfaces of its parts. This allows us to express the electric flux over the surface of a region as the volume integral of the electric flux per unit volume, taken over the volume of the region. The electric flux per unit volume is called the divergence of the electric field and given the symbol \( \text{div } \mathbf{E} \). This is a scalar field defined at each point in space. The electric flux over any closed surface \( S \) can therefore be expressed as a volume integral of \( \text{div } \mathbf{E} \). We write

\[
\int_S \mathbf{E} \cdot d\mathbf{S} = \int_V \text{div } \mathbf{E} \, dV, \tag{Eqn 8.34}
\]

where \( V \) is the volume enclosed by \( S \). This is the content of the divergence theorem, as applied to the electric field.

To obtain the differential version of Gauss’s law, we combine the divergence theorem with the integral version of Gauss’s law (Equation 2.6) to give

\[
\int_V \text{div } \mathbf{E} \, dV = \int_V \frac{\rho(\mathbf{r})}{\varepsilon_0} \, dV, \tag{2.9}
\]

where both integrals extend over the same volume \( V \) with boundary \( S \). The key step is then to notice that Equation 2.9 applies to any volume whatsoever. In particular, it applies to volumes that are arbitrarily small. Under these circumstances, the only way to satisfy Equation 2.9 is to insist that the integrands are identical on both sides. We therefore conclude that

\[
\text{div } \mathbf{E} = \frac{\rho(\mathbf{r})}{\varepsilon_0}. \tag{2.10}
\]

Finally, as you saw in the MT 8.6.3, the divergence of a field can be expressed in terms of partial derivatives of the field. In Cartesian coordinates, there is a simple and symmetrical expression for the divergence of \( \mathbf{E} \). It is

\[
\text{div } \mathbf{E} = \frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z}.
\]

So we conclude that

\[
\frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z} = \frac{\rho(\mathbf{r})}{\varepsilon_0}. \tag{2.11}
\]

Equations 2.10 and 2.11 both express the differential version of Gauss’s law. Just like the integral version, the differential version of Gauss’s law is a
fundamental law of electromagnetism, valid for all charge distributions, whether they are stationary or moving. Both versions of Gauss’s law are cornerstones of Maxwell’s theory of electromagnetism.

The differential version of Gauss’s law is a major step forward. It is a local description relating the charge density at a given point and time to the divergence of the electric field at the same point and time. There is no mention of distant charges so the concept of action at a distance is avoided. This is in the true spirit of a field theory and is a great advance on Coulomb’s law. However, Gauss’s law does not by itself give a complete description of the electromagnetic field. For example, we might expect that jiggling a charge in one region will cause disturbances in the field that propagate outwards, just as jiggling a stick in water causes water waves to spread out over a pond. Such effects are not described by Gauss’s law because it does not include derivatives with respect to time. Gauss himself thought this was a defect, so he failed to publish his law. However, Maxwell guessed (correctly) that there is nothing wrong with Gauss’s law, but Gauss’s law is only part of a complete theory of electromagnetism, which contains other equations as well. These other equations are contained in the rest of Maxwell’s theory, which will be developed in the rest of this book.

### Worked Example 2.5

A non-conducting slab lies between $z = -d/2$ and $z = +d/2$, and extends to infinity in the $x$- and $y$-directions. Throughout its volume this slab has a uniform charge density, $\rho_0$. The slab is isolated from all other influences, and the electric field for $z > 0$ is the reverse of the field for $z < 0$. Use the differential version of Gauss’s law to find the electric field inside the slab.

#### Solution

The planar symmetry means that the electric field of the slab only has a $z$-component, and this depends only on $z$. So

$$ E = E_z(z) \, e_z. $$

According to Gauss’s law

$$ \frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z} = \frac{\rho_0}{\varepsilon_0}, $$

inside the slab. The partial derivatives $\partial E_x/\partial x$ and $\partial E_y/\partial y$ are equal to zero because the field has no $x$- or $y$-components. The partial derivative $\partial E_z/\partial z$ can be written as an ordinary derivative because $E_z$ depends only on $z$. Hence Gauss’s law becomes

$$ \frac{dE_z}{dz} = \frac{\rho_0}{\varepsilon_0}. $$

Integrating both sides with respect to $z$ gives

$$ E_z(z) = \frac{\rho_0 z}{\varepsilon_0} + C, $$

where $C$ is an arbitrary constant of integration.

The charge distribution, and hence the electric field, has a reflection symmetry in the plane $z = 0$ so that $E_z(z) = -E_z(-z)$. This implies that

$$ \frac{\rho_0 z}{\varepsilon_0} + C = - \left( \frac{-\rho_0 z}{\varepsilon_0} + C \right). $$
Hence \( C = 0 \), and
\[
E_z(z) = \frac{p_0 z}{\varepsilon_0}.
\]

It is interesting to check the answer using the integral version of Gauss’s law. We choose the cylindrical Gaussian surface shown in Figure 2.19, of cross-sectional area \( \Delta A \) and with end-faces located at \( z \) and \( -z \), where \( 0 < z < d/2 \). The integral version of Gauss’s law then gives
\[
(E_z(z) - E_z(-z)) \Delta A = \frac{\rho_0 \times 2z \Delta A}{\varepsilon_0}.
\]

Then, since \( E_z(z) = -E_z(-z) \), we recover our previous result.

It is not always appropriate to work in Cartesian coordinates. If the electric field has spherical or cylindrical symmetry, it is much more sensible to use spherical or cylindrical coordinates. Expressions for divergence in these coordinate systems (too long to remember) are listed inside the back cover of the book, but simplifications can often be made.

If an electric field \( \mathbf{E}(r) = E_r(r) \mathbf{e}_r \) is radial with respect to the origin, with \( E_r(r) \) depending only on the distance \( r \) from the origin, the divergence simplifies to
\[
\text{div } \mathbf{E} = \frac{1}{r^2} \frac{d}{dr} (r^2 E_r) \quad \text{in spherical coordinates.} \tag{2.12}
\]

If an electric field \( \mathbf{E}(r) = E_r(r) \mathbf{e}_r \) is radial with respect to the \( z \)-axis, with \( E_r(r) \) depending only on the distance \( r \) from the \( z \)-axis, the divergence simplifies to
\[
\text{div } \mathbf{E} = \frac{1}{r} \frac{d}{dr} (r E_r) \quad \text{in cylindrical coordinates.} \tag{2.13}
\]

**Exercise 2.10** Use the differential version of Gauss’s law to find the electric field inside a spherical distribution of charge with charge density \( \rho(r) = Ar \), where \( A \) is a constant.

The above successes are special cases. Symmetry arguments allowed us to convert Equation 2.10 (a partial differential equation in three variables) into an ordinary differential equation in one variable. Symmetry was also used to fix the value of the constant of integration. However, a lack of symmetry generally blocks any such simplifications. This difficulty is related to the existence of **divergence-free fields** — that is, vector fields whose divergence vanishes everywhere. An example of such a field is
\[
\mathbf{G}(r) = G_x(y,z) \mathbf{e}_x + G_y(x,z) \mathbf{e}_y + G_z(x,y) \mathbf{e}_z,
\]
which obviously satisfies \( \text{div } \mathbf{G} = 0 \), because the partial derivatives \( \partial G_x/\partial x \), \( \partial G_y/\partial y \) and \( \partial G_z/\partial z \) all vanish. Suppose that we find a vector field \( \mathbf{E}(r) \) that satisfies Gauss’s law,
\[
\text{div } \mathbf{E} = \frac{\rho(r)}{\varepsilon_0},
\]
for a given charge density. Then
\[
\text{div}(\mathbf{E} + \mathbf{G}) = \text{div } \mathbf{E} + \text{div } \mathbf{G} = \frac{\rho(r)}{\varepsilon_0} + 0 = \frac{\rho(r)}{\varepsilon_0},
\]
so \( \mathbf{E}(r) + \mathbf{G}(r) \) is another vector field that satisfies Gauss’s law for this charge density. Which solution is correct? By itself, Gauss’s law cannot tell us. We need further information — either supplied by symmetry or by other principles. Remember, there are more Maxwell equations to come in later chapters!

Although it is not always possible to use Gauss’s law to deduce an electric field from a known charge density, it is always possible to do the reverse. If we know the electric field throughout a region, we can find the charge density in that region. According to Equation 2.10, the charge density is found by taking the divergence of the electric field and multiplying by \( \varepsilon_0 \).

**Exercise 2.11** Throughout a spherical region centred on the origin, the electric field is

\[
\mathbf{E} = A(x^3 \mathbf{e}_x + y^3 \mathbf{e}_y + z^3 \mathbf{e}_z),
\]

where \( A \) is a positive constant. What is the charge density throughout this region? Does the region contain only positive charges, only negative charges or a mixture of the two?

A software package on the DVD allows you to explore the integral and differential versions of Gauss’s law. This package is best studied some time after completing this chapter. The DVD also contains a video of a tutorial which uses Gauss’s law to solve a typical problem.

**Summary of Chapter 2**

**Section 2.1** Charge density is the charge per unit volume. The total charge in a region is the volume integral of the charge density over the region. Electric flux is the surface integral of the electric field over a given surface.

**Section 2.2** The integral version of Gauss’s law states that the electric flux over a closed surface \( S \) is equal to the total charge enclosed by the surface, divided by \( \varepsilon_0 \). That is,

\[
\int_S \mathbf{E} \cdot d\mathbf{S} = \frac{Q}{\varepsilon_0} = \frac{1}{\varepsilon_0} \int_V \rho(\mathbf{r}) \, dV,
\]

where \( Q \) is the charge enclosed by \( S \) and \( V \) is the volume enclosed by \( S \). This law applies to all distributions of charge (whether stationary or moving) and to all closed surfaces (no matter what their shape). It is unaffected by the presence of charges outside the closed surface. For the special case of stationary charges, Gauss’s law can be derived from Coulomb’s law, the principle of superposition and the additivity of charge.

**Section 2.3** To apply Gauss’s law, we exploit the symmetry of the charges to constrain the possible form of the electric field, and choose a suitable closed surface (a Gaussian surface). Ideally, the field has a constant normal component over this surface, or over easily identified faces of the surface. In cases of spherical, cylindrical or planar symmetry it is possible to deduce the electric field from Gauss’s law.
Section 2.4  The divergence of the electric field is the electric flux per unit volume. This is a scalar field, represented in Cartesian coordinates by

\[ \text{div } \mathbf{E} = \frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z}. \]

The divergence theorem tells us that the surface integral of a vector field over a closed surface \( S \) is the volume integral of the divergence of the field over the region \( V \) inside \( S \). That is,

\[ \int_S \mathbf{E} \cdot d\mathbf{S} = \int_V \text{div } \mathbf{E} \, dV. \]

Using this theorem, together with the integral version of Gauss’s law, we obtain the differential version of Gauss’s law:

\[ \text{div } \mathbf{E} = \frac{\rho(r)}{\varepsilon_0}. \]

This applies to all distributions of charge (whether stationary or moving). In highly symmetrical situations it leads to a differential equation which can be solved for the electric field.

Achievements from Chapter 2

After studying this chapter you should be able to do the following:

2.1  Explain the meaning of the newly defined (emboldened) terms and symbols, and use them appropriately.

2.2  Find the total charge in a volume by integrating a charge density.

2.3  State the integral version of Gauss’s law and use it in simple cases involving spherical, cylindrical and planar symmetry.

2.4  State the differential version of Gauss’s law and explain how it follows from the integral version. Use the differential version of Gauss’s law in simple cases.

2.5  Recognize that Gauss’s law is one of Maxwell’s equations, with universal validity.

After studying MT 8.3 — 8.5 you should also be able to:

2.6  Recognize the distinction between closed and open surfaces.

2.7  Select and use appropriate coordinate systems for situations with planar, cylindrical or spherical symmetry.

2.8  Evaluate simple volume and surface integrals.

2.9  Use the divergence theorem to link volume integrals and surface integrals.

2.10  Express divergence in terms of partial derivatives and evaluate the divergence of a vector field.