# Basic Books in Science

## Book 10

## More Physics: electric charges and fields - electromagnetism
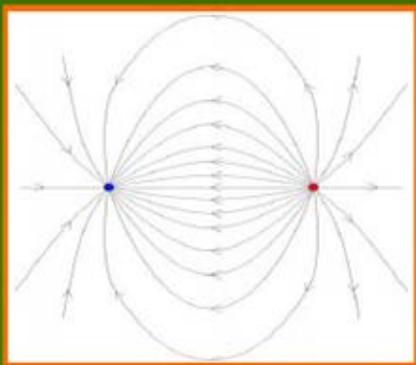


## Roy McWeeny

# BASIC BOOKS IN SCIENCE

– a Series of books that start *at the beginning*

# Book 10

# More Physics: electric charges and fields – electromagnetism

**Roy McWeeny**

Professore Emerito di Chimica Teorica, Università di Pisa (Italy)

Books in the Series are available from:

**<www.paricenter.com>** (see 'Basic Books in Science')
**<www.learndev.org>** (see 'For the Love of Science')

(Last updated 29 July 2009)

# BASIC BOOKS IN SCIENCE

## Acknowledgements

**BASIC BOOKS IN SCIENCE**

**About this book**

This book, like the others in the Series[1], is written in simple English – the language most widely used in science and technology. It builds on the foundations laid in Books 1-4, which covered many parts of Mathematics and Physics, and on some parts of Book 5, which refers to the structure and properties of matter in general.

The present book takes up the story from Book 4, which introduced the science of Mechanics and showed how it could account for the motion of objects ranging from particles to planets. Things move under the action of applied **forces** (like 'pushes' and 'pulls'). But some forces, like the one due to 'gravity', arise from the **attraction** between massive objects. You can *feel* the force due to gravity, which makes a body fall to the ground even when nothing is touching it, but you can't *see* it. We say there is a **field** of force at any point in space and it is proportional to the **mass** $m$ of the body on which it acts.

Besides having a mass $m$, a body may carry an **electric charge** $q$; and another kind of force may act on the body when it is placed in an **electric field**. If you double the

---

[1]The aims of the Series are described elsewhere (see Book 1)

quantity of charge $q$ the force is also doubled; so the electric field felt by a charge is proportional to the charge. The electric field is so similar to the gravitational field that you'll be able to take over from Book 4 many things you know about already. In Book 5 you learnt that electric charge is carried by **elementary particles** such as the **electron** and the **proton**; and that their charges are of two kinds, *negative* and *positive* respectively. So you're already well on the way to understanding electricity.

Much of this book is about fields. Electricity plays an important part in everyday life: think of electric lights, power stations where electricity is generated in **dynamos**, and **electric motors**, which use electricity to drive the machines in our factories; think of radios and the telephone, through which we can communicate with people all over the world – and even beyond, in space! When you speak into the telephone, the sound of your voice sets up vibrations and these are used to make **electrons** move along wires; the **currents** produced can be used to send **radio signals** through space; and these can be picked up by a **radio receiver** millions of miles away and converted back into sounds! By the end of this book you'll already be starting to understand all this.

# Looking ahead –

In Book 4, when you started on Physics, we said "Physics is a big subject and you'll need more than one book". Here is the next one! Book 4 was mainly about **particles**, the ways they move when *forces* act on them, and how the same 'laws of motion' still hold good for objects built up from particles – however big.

In Book 10 we'll be more interested in the **fields** that govern the interactions between particles. The only field we knew about in Book 4 was that 'due to gravity', which acts on the mass of a particle and makes it fall, but now we'll be thinking about the fields which act on particles that carry an **electric charge**.

- Chapter 1 explains how two kinds of electric charge, positive and negative, can be defined in terms of the *forces* between particles that carry charge. The force law ($F = q_1q_2/r^2$) is so similar to that for masses that you can still use most of what you already know, provided you put in the $+$ or $-$ signs and take a positive force as a *repulsion*, tending to increase the separation ($r$) of the particles, and a negative one as an *attraction*, tending to decrease the separation. It's as simple as that! And it all links up with Book 5, where you first met the **electron** and the **proton** – the basic negative and positive charges that are found in Nature.

  In this chapter the charges are at rest: we're talking

about **electrostatics**. You'll learn about machines for making 'static' electricity and about how it can be stored in **condensers** and used to make **sparks**. And you'll learn how all of electrostatics can be understood in terms of a single basic principle – Gauss's law.

- In Chapter 2 you'll be thinking about charges in motion, which produce **electric curents**. The currents are carried through **conductors** – materials which allow charges, usually electrons, to flow through them. Other materials, which don't allow charge to pass, are called **insulators**; and in between there are **resistances**, which allow *some* charge to pass but are not very good at it!

  Long thin **copper wires** are good conductors and are used in all electrical gadgets. You'll study some simple **electric circuits**, in which 'circuit elements' like condensers and resistances are connected by wires so that current can flow between them.

  Finally, we'll make a connection with Chemistry (Book 5). Batteries make a plentiful supply of electricity from certain **chemical reactions** which depend on the transfer of electrons between the atoms taking part.

- Chapter 3 again starts from discoveries made over 3000 years ago! – that certain heavy stones could attract small pieces of iron, and that the iron itself could be given the same property by stroking the stone

with it. The property is called **magnetism** and a piece of iron with this property is a **magnet**. When a magnet is free to move, one end of it always points the same way – towards the 'North pole' of the axis around which the Earth turns; so magnets were first used in steering ships at sea.

From such simple beginnings the whole theory of the **magnetic field** was built up over the centuries. The magnetic field ($\mathsf{B}$), like the electric field ($\mathsf{E}$), is a vector; but unlike the electric field it acts only on a **moving charge**, pushing it away from the direction of motion. As a moving charge is an electric *current*, magnetism and electricity are closely related: you'll discover how an electric current can produce a magnetic field; and how a magnetic field can exert a force on a current-carrying wire.

- You're now ready to put all these ideas together and look at **Maxwell's equations** – one of those great breakthroughs that happen perhaps only once in a hundred years. In Chapter 4 you'll find how the fields $\mathsf{E}$ and $\mathsf{B}$, and the electric charges and currents that produce them, can all be related through just *four* equations, set up by Maxwell about 130 years ago. This is difficult stuff, but if you've got so far you're ready to make the last step into **electromagnetism**.

You're going to need many of the ideas you've picked up about fields, that vary from point to point in space (and in time!), and the differential equations that they

obey. You don't have to *solve* them at this point; it's enough just to understand what they mean. Then, in the rest of the book, you'll find they tell you how to build electric power stations, electric motors, radio transmitters, and many other things that can be found ready-made nowadays – thanks to the work of all the other people who've made the rest of the journey from the *principles* to the practice.

- Coming up, in Chapter 5, you'll see how just one of Maxwell's equations (which can be written on the back of a bus ticket!) says all we need to know about dynamos and electric motors. As usual, we start from a simple 'model' in which we imagine a loop of wire spinning in a magnetic field. Faraday discovered that the changing **flux** of the *magnetic* field, passing through a wire loop, could produce an *electric* field that moves charge round the wire. Maxwell set up the equation that describes this effect; and this leads to two types of dynamo (usually called "generators" nowadays). One type gives *direct current* – always flowing in the same direction – while the other gives *alternating current* – flowing alternately backwards and forwards. The DC and AC types differ only through a simple change in the generator.

The electric motor is like a generator running in reverse: instead of using mechanical power to turn the loop and generate an electric current it *takes in* electric current and generates power! It's as easy as that.

Power, as you know already from Book 4, is the rate of doing work. The generator works by turning mechanical energy into electrical energy; and you'll find that the principle of **energy conservation** still holds good: you never *make* energy – you just push it around, from one form to another! By the end of Chapter 5 you'll understand how electric power can be **distributed** – being carried into thousands of homes and factories.

- Chapter 6 is about sending electric and magnetic effects through *space*, without using wires. Fields can go through space in the form of **waves**, so we have to start by thinking about waves in general. All kinds of 'wave motion' are very similar, involving ideas like **wavelength**, **frequency**, and the **velocity** with which they travel. And they are described by the same equations – so when you've studied one you've done them all! You'll find examples of setting up and solving the **wave equation** e.g. for waves on a vibrating string, which produce sound waves.

Then we come back to Maxwell's equations and show that any disturbance of the **electromagnetic field** travels out through space (is "propagated") with the enormous speed of almost 300 million metres per second! And it turns out that electromagnetic waves are of many different kinds, some with wavelengths of several kilometres (in radio) and others going down to 'atomic' dimensions – with ordinary 'daylight' coming

in between!

- In Chapter 7 we specialize on radio waves and how they can be produced by electric circuits. We look at the way different circuit elements behave when a voltage is applied to them: for a resistance $R$ the applied voltage $V$ and the current $I$ passing through it are related by $V = RI$; but for a condenser, with capacity $C$, and a coil with 'inductance' $L$, things are different. Instead of $R$ we need to use an **impedance** $Z$, which turns out to be a **complex number** – which you first learnt about in Book 1. This is the beginning of AC circuit theory, which is the 'backbone' of radio. With simple **oscillator circuits** you can make electrical oscillations of any frequency you choose! This allows you to send out radio waves of that frequency from a **transmitter** and to 'pick them up' with a **reciver**, which can be adjusted (or 'tuned') to select only that frequency.

# CONTENTS

# Chapter 1

# Charges at rest: Electrostatics

## 1.1 Electric charges and fields

**Electric charges**

About 3000 years ago people began to notice that certain
materials had 'magical' properties: if a piece of amber was
rubbed with a woollen cloth it would attract small objects,
such as bits of straw, and pick them up; and a rod made
of glass and rubbed with a silk cloth would have a similar
property. Much later, the rubbed objects were described as
"charged with electricity" – thought to be an invisible fluid
of which there were two kinds, '**negative**' for the amber,
'**positive**' for the glass.

Nowadays we know (see Section 1.2 of Book 5) that matter of all kinds is built up from **atoms** and that the atoms contain **elementary particles**: **electrons**, which carry a unit of *negative* charge, and **protons**, which carry a unit of *positive* charge. The electrons are very light and move easily, while the protons are nearly 2000 times heavier and stay inside the **nuclei** of the atoms. There's another kind of heavy particle, with almost exactly the same mass as the proton but with zero charge: it's called the **neutron**. The two kinds of heavy particle (protons and neutrons) are **nucleons**: they have almost exactly the same mass and together they account for nearly all the mass of the atom.

The **atomic number** is the number of *protons* in the nucleus and is usually denoted by $Z$: so if we write $-e$ for the charge of the electron the charge on the nucleus will be $+Ze$. The whole atom, nucleus plus electrons, is normally uncharged, the positive charge on the nucleus being exactly cancelled by the charge $-Ze$ of $Z$ electrons *outside* the nucleus.

There are more than ninety different kinds of atom, one for each of the **chemical elements**: they are usually listed, along with their names and main properties, in a **Periodic Table**. The first two atoms in the Table are

Hydrogen(Atomic number $Z = 1$, atomic mass 1.0 $m_u$),

Helium (Atomic number $Z = 2$, atomic mass 4.0 $m_u$),

where $m_u$ is the **atomic mass unit**. After that, the atoms are listed in rows, the values of $Z$ increasing along each row in unit steps. The first row starts with Lithium (atomic

number $Z = 3$, atomic mass 7.0 m$_u$) and goes up to Neon with $Z = 10$. It's now clear why the atomic masses, which go in the order 1,4,7,... can be so much 'out of step' with the atomic numbers $Z = 1, 2, 3, ....$ The mass of the nucleus depends simply on the number of heavy particles (both types of '**nucleon**') which it contains. For example, the nucleus of the He atom must contain 4 nucleons, two of them protons, the other two being neutrons; and similarly the Li nucleus ($Z = 3$) must contain 3 protons, the remaining 4 being neutrons.
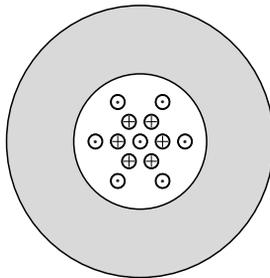


**Figure 1.** Constituents of a Carbon atom (schematic)

Figure 1 will remind you of how things stand in the case of an atom of carbon: the spots shown as $\oplus$ indicate the protons, while $\odot$ stands for a neutron. The 6 electrons outside the nucleus give carbon its *chemical* properties, which you began to study in Book 5. The 7 neutrons contribute only to the *mass* of the atom.

What about the much lighter particles, the electrons? It used to be thought that they moved around the nucleus in

orbits, like the planets around the sun (studied in Book 4); but now it's known that they don't follow definite paths. Think of them instead as buzzing around the nucleus like a swarm of bees: the number of electrons is definite and is equal to $Z$; but you can never know exactly where they are!

Figure 1 may help you to get a picture of what's inside an atom, but remember this picture is very schematic: the light gray ball just indicates the region in which the electrons are most likely to be found. The diameter of the carbon atom is around 1 Å unit, 1 ten-millionth of a centimeter – so it's no surprise that you can't see atoms. And the diameters of their *nuclei* are roughly ten thousand times smaller still! So the white area in Fig.1 really gives no idea of the size of nucleus; it's there only to indicate the *contents* of the nucleus. The spots shown as $\oplus$ represent the protons, while those shown as $\odot$ stand for the neutrons, but again you mustn't think of them being placed at particular positions to form some kind of structure: like the electrons, they're always on the move and there's no way of knowing exactly where they are. In Chemistry all that doesn't matter: what counts is the atomic number $Z$, the number of protons in Fig.1, which determines the 'chemistry of carbon'.

Now we're back in Physics and starting to think in more detail about the effects of electric charge. The first thing to remember is the enormous numbers of particles involved: 1 **mole** (the unit of quantity used in Chemistry) contains about $6 \times 10^{23}$ atoms. Let's talk about the metal copper, which is used a lot in electrical gadgets because it allows

electrons to flow through it very easily. It is a good **conductor** of electricity, whereas many things (like glass or silk or dry wood) don't have that property and are called **insulators**. (You've probably seen **power lines**, which carry electricity from one place to another, hanging from glass insulators so that the electricity doesn't leak away.)

The next thing we need to remember, from Book 5, is that the forces between particles which carry an electric charge are very much bigger than those that arise from **gravitational attraction**. The *gravitational* force $F_{\mathrm{grav}}$ between two *masses*, $m_1$ and $m_2$, is proportional to the product of the masses, but inversely proportional to the square of the distance between them. The force $F_{\mathrm{elec}}$ between two particles carrying *electric charges*, $q_1$ and $q_2$, is given by a law of exactly the same form. So we can write

$$F_{\mathrm{grav}} \propto \frac{m_1 m_2}{r^2} \quad (a), \qquad F_{\mathrm{elec}} \propto \frac{q_1 q_2}{r^2} \quad (b). \qquad (1.1)$$

But whereas masses are always *positive* quantities, and the force $F_{\mathrm{grav}}$ is an attraction, the charges can be either positive or negative: if both have the same sign $F_{\mathrm{elec}}$ is a repulsion (tending to increase the separation $r$); while if $q_1$ and $q_2$ are of opposite sign $F_{\mathrm{elec}}$ is an attraction (tending to reduce the separation $r$). In short, **like charges repel one another, unlike charges attract.**

As we said above, there is normally an almost exact balance between the negative and positive charges in the atoms of matter in any form. What people discovered long ago was that the balance could sometimes be disturbed by rubbing

some kinds of material with a different material. For example, some of the atoms may lose an electron and the numbers of negative and positive charges then get out of balance: in that case the material is left with an excess of positive charge. But the important thing is that the disturbance is *extremely small*; perhaps only one atom in every 500 million loses an electron, but that's enough to have an enormous effect! To find how enormous is easy. Let's work it out as an Example:

---

**Example 1.** Forces between charged bodies

There are about $6 \times 10^{23}$ atoms in the mole of copper, so if one in 500 million of them loses an electron the mole of copper will get an excess positive charge of $\approx 1.2 \times 10^{16} e$, where the magnitude of the charge on an electron is $e \approx 1.6 \times 10^{-19}$ C (C being the **unit of electric charge**, which is called the "Coulomb"). So the mole of copper will, in that case, get a charge of about $(1.92 \times 10^{-2}$ C.

Now let's get the force between two such charged bits of copper, two small disks 1cm apart. The experimentally defined proportionality constant in the force law given in (1.1), which we come back to later, has the approximate value $9 \times 10^9$N m$^2$ C$^{-2}$ (if you've forgotten about **physical dimensions** and SI units you'd better look back at Books 4 and 5) and if we take the charges as $q_1 = q_2 = 1.92 \times 10^{-2}$ C the result will be

$$F_{\text{elec}} \approx (9 \times 10^9 \text{N m}^2 \text{ C}^{-2}) \times (1.92 \times 10^{-2}\text{C})^2 \times (10^{-2}\text{m})^{-2}.$$

If you do the arithmetic (Do it! – thinking carefully about the units and noting that the answer, which is a force, comes out

correctly in Newtons, the SI unit of force), you'll find

$$F_{\text{elec}} \approx 33.18 \times 10^3 \text{N}.$$

To put this in units of "kilogram weight", which you can 'feel', remember (Book 4) that 1 kg wt = 9.81 N. So the force between the two copper disks, 1 cm apart, is the force you'd have to apply to lift an object weighing about $3.38 \times 10^3$ kg, or nearly 4 metric tons! And that's all because you took away – in your imagination – an electron from one copper atom in every 5 hundred million. So a little bit of electric charge goes a long way; moving those charges around is enough to drive heavy machinery, as you'll discover in later chapters.

---

One last thing: you may be wondering if the force between two particles depends on their masses, as well as any charges they may carry. The answer is simply that the gravitational force is so small (by a factor of $10^{-20}$ or more), compared with the electric force, that in Book 10 you can just forget about it!

**Electric fields**

In Book 4 you studied bodies moving under the action of applied forces, including the force due to gravity; and learnt that a force $F$, acting on a particle of mass $m$, made it move with an **acceleration** $a$ such that $F = ma$. You also learnt that force and acceleration are **vector** quantities, depending on direction in space, and that the vector equation $\mathsf{F} = m\mathsf{a}$ (special symbols are used for quantities of this kind) means the same as

7

three 'ordinary' equations, one for each **component**:

$$\mathsf{F} = m\mathsf{a} \quad \text{means} \quad F_x = ma_x, \; F_y = ma_y, \; F_z = ma_z,$$
$$(1.2)$$

where $F_x, a_x$ etc. are the components of the vectors $\mathsf{F}, \mathsf{a}$ in the x-direction. Remember that the mass is a property of the particle, not depending on how you choose the x-, y-, and z-axes: it is a **scalar** quantity and the three equations relating the components in (1.2) are scalar equations. We also used **calculus** notation for rates of change, so that the x-component equation in (1.2) can be written as

$$F_x = ma_x = m\frac{\mathrm{d}v_x}{\mathrm{d}t} = m\frac{\mathrm{d}^2 x}{\mathrm{d}t^2}, \qquad (1.3)$$

with similar equations for the y- and z-components. (Remember that acceleration is the rate of change of velocity, with respect to time, and that $v_x$ is already the rate of change of distance $x$; so $a_x$ is the *second* derivative of $x$ with respect to time.)

If you're not yet comfortable with vectors and calculus, you'd better go back to Books 2, 3, and 4, where everything you need has been explained and used.

From now on we'll be neglecting gravity and the only forces we need will be those arising from **electric charges** on the particles. The interaction between particles carrying charges $q_1, q_2$ will thus result in the force $F_{\mathrm{elec}}$ in (1.1) and ths subscript "elec" will be dropped as the forces we talk about will *all* be electric in origin.

If a particle with electric charge $q$ is put at some point in space and feels a force $F$, proportional to the charge, we say that $F$ measures the electric field $\mathsf{E}$ at that point. Once the *unit* of charge has been agreed on, the unit of **field intensity** – also called **field strength** – is thus defined experimentally: it is the force acting on unit charge placed at the field point. The physical dimensions of $\mathsf{E}$ are also defined as those of (force/electric charge): $[\mathsf{E}] = \mathrm{MLT}^{-2}\mathrm{Q}^{-1}$

From the experimental fact that two charged particles exert a force on each other, according to (1.1), if follows that any charged particle *produces* an electric field, besides being able to 'feel' one. Thus, the force law in (1.1) may be written (using $k$ for the proportionality constant)

$$F = k\frac{q_1 q_2}{r_{12}^2}$$

and gives the force between two particles, at Points 1 and 2, at a distance $r_{12}$, when they carry charges $q_1$ and $q_2$. This means that the particle at Point 1 feels a force of magnitude

$$F_1 = q_1 \times \left(k\frac{q_2}{r_{12}^2}\right),$$

due to the presence of the second particle. But from the definition of electric field intensity as the force *per unit charge* this tells us that the field at Point 1 is $E_1 = F_1/q_1$. The proportionality constant, which we've just called '$k'$', is very fundamental and is always denoted by $(1/4\pi\epsilon_0)$ ("one over four pi epsilon zero") – more about that later. For now, we'll write our result in the form:

Electric field at Point 1, due to presence of a charge $q_2$ at Point 2, is given by

$$E_1 = \left(\frac{1}{4\pi\epsilon_0}\right)\left(\frac{q_2}{r_{12}^2}\right),$$

where $r_{12}$ is the separation of Points 1 and 2.

We mustn't forget that here the field is a *vector*, pointing from 1 to 2 – but which way round? The unit 'test charge' at Point 1 is taken positive; so if $q_2$ is negative the force, and the field, will point in the direction $1 \rightarrow 2$. That's easy to remember, so we won't use vector notation for fields when it's not needed. as here. This is a key equation, so we'll put it in a box:

---

Field intensity $E$ at distance $r$
from point charge $q$ is given by
$$E = (1/4\pi\epsilon_0)(q/r^2),$$

---

(1.4)

where the labels (1,2) are left out, as $q$ is any charge and $r$ is its distance from it the field point.

One place where we *do* need to use vector notation is when we think about *combining* fields. We know (see Chapter 1 of Book 4) that forces are combined by using the laws of **vector addition**; and since electric fields are measured by the forces acting on a test charge we expect that when two

fields act on a charge they will have a **resultant** which is simply their vector sum.

This is exactly what is found by experiment. It is called a **superposition principle** and in its general form states that:

> The resultant field $\mathsf{E}$ at any given point in space, when any number of fields $\mathsf{E}_1, \mathsf{E}_2, \mathsf{E}_3$, ... act at the same time, is simply their **vector sum**:
> $\mathsf{E} = \mathsf{E}_1 + \mathsf{E}_2 + \mathsf{E}_3 + $ ...

$$(1.5)$$

Now let's look at a simple example to show how you can use the principle to get the field, at any point in space, due to to any given set of electric charges.

---

**Example 2.** Calculation of a field vector $\mathsf{E}$

Calculate the field vector $\mathsf{E}$ at one corner of a square, with sides of length $L = 10$ cm, when charges of $1 \times 10^{-10}$ C are placed at the three other corners. (Use the known value of the fundamental constant $(1/4\pi\epsilon_0) \approx 9 \times 10^9$ N m$^2$ C$^{-2}$)

Take the corner where you want to get the field as origin of coordinates and the coordinates of the three other corners as $(10, 0)$, $(10\sqrt{2}, \ 10\sqrt{2})$, $(0, 10)$, all in units of 1 cm. Then the field $(\mathsf{E}_1)$ due to the first charge will have magnitude $(1/4\pi\epsilon_0) \times (1 \times 10^{-10}$ C$)/(0.1 \, \text{m})^2$ and will point along the x-axis towards

the origin; that ($\mathsf{E}_2$) due to the second charge will $1\times10^{-10}$ C have magnitude $(1/4\pi\epsilon_0) \times (1 \times 10^{-10}\,\text{C})/(0.1\,\sqrt{2}\,\text{m})^2$ and will be directed along the diagonal towards the origin; and $\mathsf{E}_3$ will be like $\mathsf{E}_1$, but lying along the y-axis. (Draw a picture to make this clear and get the field contributions one at a time: the first one is $(9 \times 10^9\,\text{N m}^2\,\text{C}^{-2}) \times (1 \times 10^{-10}\,\text{C}) \times (0.1\,\text{m})^{-2}) = 90\; rmN\;\text{C}^{-1}$, directed towards the origin along the x-axis. Note that the unit of field strength is 1 Newton per Coulomb and that the numerical value comes out large owing to the enormous value of $(1/4\pi\epsilon_0)$, as in Example 1.

Get the other two field vectors and then the vector sum of all three, as in (). (If you've forgotten how to work with vectors go back to Section 1.5 of Book 4.)

---

## 1.2 Some properties of the electric field

In Chapter 2 of Book 4, we first met the ideas of **work** and **energy** and found how they could be defined and used. We now need to extend these ideas to particles carrying an electrric charge and moving in an electric field (or, more exactly, an **electrostatic field** as we first take the case of a field produced by charges *at rest*).

Remember that when a force $F$ is applied to a particle and moves it through a distance $d$, in the direction of the force, mechanical **work** is done – *by* the force, *on* the particle. The force does work $W = Fd$. When the

force and the displacement are *not* in the same direction it's convenient to use vector methods, as in Book 4 (Chapter 3). In a very small displacement, represented by a **displacement vector** ds, the work done by the force is

$$w = \mathsf{F} \cdot \mathsf{ds} = F_x \mathrm{d}x + F_y \mathrm{d}y + F_z \mathrm{d}z, \qquad (1.6)$$

where the second form expresses the **scalar product** $\mathsf{F} \cdot \mathsf{ds}$ in terms of the x- y- and z-components $(F_x, F_y, F_z)$ and $(\mathrm{d}x, \mathrm{d}y, \mathrm{d}z)$ of the vectors $\mathsf{F}$ and $\mathsf{ds}$. In general, ds points along the path followd by the particle – it is a **path element** – while the force vector may be in any direcction.

When the force $\mathsf{F}$ was due to gravity, acting on the particle *mass*, it was found useful to define a **potential energy** function $V(x, y, z)$ for every point P $(x, y, z)$ in space, such that the *difference* in PE between two points A and B measures the work you have to do in carrying the particle slowly from A to B

$$V_B - V_A = W(A \to B). \qquad (1.7)$$

The work you have to do gives the *increase* in PE, just as if you're climbing a hill; and you must carry the particle slowly because if you let go it will accelerate under the force due to the field and all the energy arising from the field will be 'lost' as kinetic energy. The important

thing about the gravitational field was that the work $W(A \to B)$ didn't depend at all on the *path* followed in going from A to B. Fields of this kind are called **conservative**: they allow us to define **path-independent** functions, with a unique value for every choice of the variables $x, y, z$. Now, we're thinking of a test particle with electric charge $q_0$, in an electric field, but everything goes in exactly the same way. We define an **electrostatic potential** function, call it $\phi(x, y, z)$, as the potential energy per unit charge of $q_0$ at the field point; so $\phi(x, y, z) = V(x, y, z)/q_0$ and we can say

$$V_B - V_A = q_0\phi_B - q_0\phi_A = W(A \to B), \qquad (1.8)$$

where $W(A \to B)$ is the work you must do in carrying the charge slowly between the two field points. If Point A is taken as a 'reference point', at which $\phi_A$ is defined to be *zero*, then $\phi_B$ will have a unique value for *any* point in the field. Usually the electrostatic potential is chosen to be zero at a point, with coordinates $(x_0, y_0, z_0)$, an infinite distance from all sources of electric fields: $\phi(x_0, y_0, z_0) = 0$. As an Example let's calculate the electric potential due to a point charge $q$:

---

**Example 3.** Electric potential due to a point charge $q$

Start from the experimental law, **Coulomb's law**, for the force between two particles, at a distance $r$ and carrying electric charges

14

$q$, the 'given' charge, and $q_0$, the 'test' charge:

$$\mathsf{F} = \left(\frac{1}{4\pi\epsilon_0}\right)\left(\frac{q_0 q}{r^2}\right).$$

Here $q_0$ is put at the point where we want to find the potential due to the given charge $q$ which is producing the field.

To move the test charge slowly we must apply a force $-\mathsf{F}$, to hold it in equilibrium against the field. The work we have to do as we go along the path from A to B, will then be

$$W(A \rightarrow B) = -\left(\frac{q_0 q}{4\pi\epsilon_0}\right)\int_{r_A}^{r_B}\left(\frac{1}{r^2}\right)\mathrm{d}r,$$

where the definite integral (see Book 3) goes from $r = r_A$ (start of the path) to $r = r_B$ (end of the path). Since $\int(1/r^2)\mathrm{d}r = -1/r$ it follows easily that

$$W(A \rightarrow B) = \left(\frac{q_0 q}{4\pi\epsilon_0}\right)\left(\frac{1}{r_B} - \frac{1}{r_A}\right).$$

On putting this result in (1.8) and dividing both sides of the equation by $q_0$ it follows that

$$\phi_B - \phi_A = \left(\frac{q}{4\pi\epsilon_0}\right)\left(\frac{1}{r_B} - \frac{1}{r_A}\right).$$

As for the gravitational potential energy, we choose a 'reference point' at which to define the potential as zero. In the case of the *electric field* of a point charge $q$, let's first imagine the charge being taken away to an infinite distance without moving $q_0$, and put $V_A \rightarrow 0$ when $r_A \rightarrow \infty$ in the last equation. The result is

$$\phi_B = \left(\frac{q}{4\pi\epsilon_0}\right)\left(\frac{1}{r_B}\right).$$

This is true for any field point $r_B$ at which the test charge is put, so the label B is not needed:

---

Example 3 has led us to another key result:

'

> Electric potential $\phi$ at any point distant $r$ from point charge $q$, is
> $\phi(r) = (1/4\pi\epsilon_0)(q/r)$

$$(1.9)$$

This result gives a simple way of getting the potential at any point in space due to any distribution of electric charges. Thus, the electric potential due to a collection of charges $q_1, q_2, \ldots$ at points $\mathsf{r}_1, \mathsf{r}_2, \ldots$, respectively, is

$$\phi = \phi(\mathsf{r}_1) + \phi(\mathsf{r}_2) + \ldots = \left(\frac{1}{4\pi\epsilon_0}\right)\left(\frac{q_1}{r_1} + \frac{q_2}{r_2} + \ldots\right),$$
$$(1.10)$$

where $r_i$ is the *length* of the vector $\mathsf{r}_i$ pointing from the field point (where $\phi$ is evaluated) to the charge $q_i$. It is clearly much easier to get the potential at a point than to get the field intensity $\mathsf{E}$, given in (1.1), because $\mathsf{E}$ is a sum of *vectors* (perhaps involving complicated geometry) whereas $\phi$ is just a sum of scalar quantities (which

depend only on distances). At this point, however, you should think about *units* – numbers by themselves are not enough!

---

**A note on units**

The unit of charge, first used in Section 1.1, was called the Coulomb and the electric field intensity, being force per unit charge, was therefore measured in "Newtons per Coulomb" i.e. unit field $= 1$ N C$^{-1}$. More usually, however, the field intensity is measured using the *potential* $\phi$: from (1.8) this has the dimensions of work per unit charge, or force×distance per unit charge, and is therefore measured in N m C$^{-1}$. The SI unit of potential is called the "**Volt**", so in summary.

$$\text{Unit of potential} = 1 \text{ V} = 1 \text{ N m C}^{-1}$$
$$\text{Unit of field intensity} = 1 \text{ N C}^{-1} = 1 \text{ V m}^{-1}$$

---

When the field **E** itself is required, it can always be obtained from $\phi$ by differentiating. Thus, suppose unit charge moves a distance d$x$ along the x-axis, so the field does work $E_x$d$x$ on it. This is the *loss* of potential for doing work, so we call it $-$d$\phi$ and write (in the limit of infinitely small changes) $E_x = -$d$\phi$/d$x$. In words "the x-component of the electric field at any point is given by the rate of decrease of potential $\phi$ in the x-direction".

The next Example shows how this can be done:

---

**Example 4** Potential and field due to several charges

Use (1.10) to get the electric potential $\phi$ at the origin in Example 3 due to charges $q = 1 \times 10^{-10}$ C at each of the three other corners of the square. What are the units in which $\phi$ is given?

Repeat the calculation to get the potential at points 1 cm away from the origin in the x- and y-directions, in turn. Then make rough estimates of the field components $E_x$ and $E_y$ at the origin.

---

For a point displaced from the origin by a small vector $\mathsf{dr}$, a similar result holds good for each of the three components, provided the derivatives are replaced by *partial* derivatives. Thus, with $\phi = \phi(x, y, z)$, the field at the origin has components

$$E_x = -\frac{\partial \phi}{\partial x}, \quad E_y == \frac{\partial \phi}{\partial y}, \quad E_z = -\frac{\partial \phi}{\partial z}. \qquad (1.11)$$

The operation of getting the vector components $(E_x, E_y,$ and $E_z)$, from a scalar function $\phi(x, y, z)$ of the three spatial coordinates, is called taking the **gradient** of the function. And the result (1.12) is often written in the short form

$$\mathsf{E} = -\mathrm{grad}\ \phi. \qquad (1.12)$$

To *visualize* the electric field – to get a picture of it – we can either work in terms of the vector $\mathsf{E}$ and the

18

way it varies in space; or in terms of the scalar potential function $\phi(x, y, z)$.

## 1.3 Visualizing the electric field

The Examples above illustrate two direct methods of calculating the electric field from a number of point charges. To get a picture of the field let's start from the simplest example – a single point charge. The potential $\phi$ is in this case a function of one variable alone, the distance $r$ of the field point from the charge $q$ at the origin, and has the same value at all points on a spherical surface of radius $r$. Such surfaces are called **equipotential surfaces**: they are indicated below as circles centred on the positive charge $q$ (shown as $\oplus$).
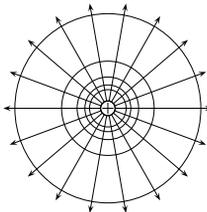


**Figure 2.** Electric field around a point charge
– showing equipotentials (spheres)
– and lines of force (radii)

Note that the equipotentials become further apart as $r$ increases and have not been shown near to the charge as they become so close together. The potential has a **singularity**(see Book 3) at any point charge: $\phi \rightarrow \infty$ as $r \rightarrow 0$. The radial lines, shown as arrows pointing out from the charge, are always perpendicular (or 'normal') to the equipotential surfaces. They are called "lines of force" because, as they follow the direction of the field vector $\mathsf{E}$, they follow the force that the field would exert on a test charge.

These properties are general and apply to the fields produced by *any* distribution of electric charges. The lines of force (shown by the arrows) come *out* from a positive charge, but go *into* a negative charge. They were very much used in the early days of electrostatics, by Michael Faraday and others, though nowadays they are used only in getting a general picture of the field. They do, however, lead us to define an important quantity called the **flux** of the vector field.

In Fig.2 the equipotential surfaces, indicated in the Figure by circles, are in fact *spheres* and the magnitude of the field intensity $E = |\mathsf{E}|$ has the same value at all points on the surface, being a function of $r$ alone. The *vector* $\mathsf{E}$ is normal to the surface and if you think of the field as something 'coming out' from the charge $q$ then lines of force, crossing the surface, show you the way it is 'flowing' at that point. When the distance from the

charge is small the lines of force are close together and $E$ is large; further from the charge the lines are far apart and $E$ is small.

Let's *define* the flux generally with the help of Figure 3.
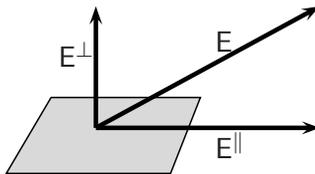


**Figure 3.** Definition of flux
$\mathsf{E}$ = field vector
$\mathsf{E}^{\perp}$ = normal component
$\mathsf{E}^{\parallel}$ = parallel component

The **flux** of $E$ across a small bit of surface, of area $\mathrm{d}S$, is (area)$\times$ (component of $\mathsf{E}$ normal to surface). With the notation used in Fig.3, this will be $E^{\perp}\mathrm{d}S$, where $E^{\perp}$ is the perpendicular component of $\mathsf{E}$.

We can also think of charge as a **creator of flux** and rewrite the 'boxed' statement (1.4) as

> Flux created by point charge $q$ (field at $r$ times area of spherical surface at $r$) is $q/\epsilon_0$.

$$(1.13)$$

– which looks nicer with no $4\pi$ and $r^2$ to remember.

Now take another Example, a bit more complicated than a single point charge, using a system of *two* point charges, one negative but the other positive: this is called an **electric dipole**.

---

**Example 5.** Electric field outside a dipole

For two charges of equal but opposite sign the electric potential at any point can be found easily from (1.10). Let's also use Cartesian coordinates (Book 2), putting the two charges on the z-axis at a small distance $\pm\delta$ above and below the xy-plane, as in Fig.4.
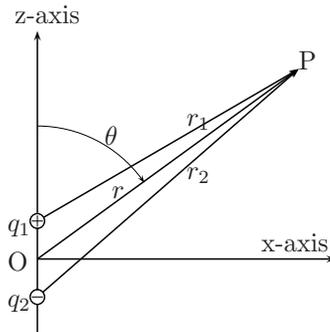


**Figure 4.** For calculating dipole field

The potential at a field point P, with coordinates $(x, y, z)$, will then follow as

$$\phi(x, y, z) = \left(\frac{1}{4\pi\epsilon_0}\right)\left(\frac{q}{r_1} - \frac{q}{r_2}\right),$$

where P lies in the xz-plane and the two charges are $q_1 = q$ and $q_2 = -q$.

The squared distances are $r_1^2 = x^2 + (z-\delta)^2$ and $r_2^2 = x^2 + (z+\delta)^2$; and thus, neglecting terms in $\delta^2$,

$$r_1^2 = x^2 + (z^2 - 2z\delta) = r^2 - 2z\delta = r^2[1 - (2z\delta/r^2)],$$
$$r_2^2 = x^2 + (z^2 + 2z\delta) = r^2 + 2z\delta = r^2[1 + (2z\delta/r^2)].$$

The two terms in the $\phi$-equation may now be expanded, using the binomial theorem (Book 3) to write $(1/r_1) = (r_1^2)^{-\frac{1}{2}}$ as an expansion in powers of the small quantity $2z\delta/r^2$ and similarly for $(1/r_2)$.

If you do that (and you should do it!) you'll find the two terms become

$$\frac{-q}{r}\left(1 + \frac{z\delta}{r^2}\right) + \frac{q}{r}\left(1 - \frac{z\delta}{r^2}\right)$$

and their sum leads to a final expression for the potential $\phi$ at point P in Fig.4:

$$\phi(x, y, z) = \left(\frac{1}{4\pi\epsilon_0}\right)\frac{q}{r}\frac{2z\delta}{r^2} = \left(\frac{1}{4\pi\epsilon_0}\right)\frac{d\cos\theta}{r^2},$$

where $d = 2\delta \times q = $ (charge-separation)×(charge-magnitude) is called the **electric moment** of the dipole and $\theta$ is the angle shown in Figure 4.

---

The result given in Example 5 can be put in vector form (see Book2, Section 5.2) by defining an **electric dipole vector d**, of magnitude $d$ and pointing from $-q$ to $+q$. The result then becomes

$$\phi(\mathsf{r}) = \left(\frac{1}{4\pi\epsilon_0}\right)\frac{\mathsf{d}\cdot\mathsf{r}}{r^3}, \qquad (1.14)$$

where $\phi(\mathsf{r})$ just means that $\phi$ depends on the three components $x, y, z$ of the position vector $\mathsf{r}$ of Point P.

To get the electric field intensity $\mathsf{E}$, which is a vector, we must get the three components $E_x, E_y, E_z$. These are given in (1.11) and involve partial differentiation (see Section 6.1, Book 3). The results are

$$E_x = \left(\frac{d}{4\pi\epsilon_0}\right)\frac{3zx}{r^5}, \quad E_y = \left(\frac{d}{4\pi\epsilon_0}\right)\frac{3zy}{r^5},$$
$$E_z = \left(\frac{d}{4\pi\epsilon_0}\right)\frac{3z^2 - r^2}{r^5}. \tag{1.15}$$

By putting trial values of the coordinates of P into these formulas you can find the field direction at any point and (with a lot of patience!) work your way along any line of force to see where it goes. Just do it for a few points and then study the picture that follows (Fig.5).
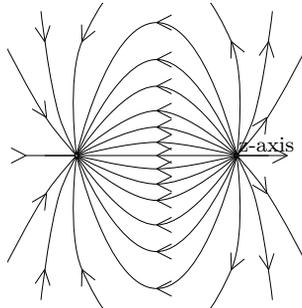


**Figure 5.** Electric field around a dipole
– showing some lines of force

## 1.4  A general principle.  Some applications

There's another way of finding the field due to a distribution of charge: it's based on a very general principle, which you can understand by looking at Fig.2 for the field arising from a single point charge. The **flux** of the field $\mathsf{E}$, through a small element of surface $\mathrm{d}S$, was defined in Fig.3 as (normal component of field)$\times$ (area of surface) i.e. $E^{\perp} \times \mathrm{d}S$. Now look at any equipotential surface, of radius $r$, in Fig.2: the *total* flux through it will be $E \times S = E \times (4\pi r^2)$, because the vector $\mathsf{E}$ is already normal to the surface and the *whole* area (which you get by summing all the little bits $\mathrm{d}S$) is simply that of a sphere. On putting in the value of E, which follows from (1.4), the total flux of $\mathsf{E}$ across any spherical surface in Fig.2 is found as

$$\text{Total flux} = \left(\frac{1}{4\pi\epsilon_0}\right)\frac{q}{r^2} \times 4\pi r^2 = \frac{q}{\epsilon_0}. \qquad (1.16)$$

This is a very remarkable result: whichever surface you choose, the total flux of field intensity through it is the same – it's just the amount of charge it encloses, divided by the universal constant $\epsilon_0$. But the result is much more general than you could ever imagine: it's still true even if you could chop the charge into pieces and put the bits *anywhere* within the surface, not just at the centre!

And the surfaces you choose don't have to be spheres, or even equipotentials, they can be of any shape whatever as long as they are *closed*. We can't yet be sure that this is always true, but the idea is so beautiful that we'll write it down as a general principle – calling it "Gauss's law" – and then start using it to show how well it works:

> The total flux of electric field intensity out of any **closed surface** is equal to the total charge it encloses, divided by $\epsilon_0$.

(1.17)

Note that the numerical value of $1/\epsilon_0$ in the formula depends on how the units are chosen: the physical dimensions of $(1/\epsilon_0)$ are, from (1.1), [force×length$^2$ /charge$^2$], so in SI units it will be measured in Newton metre$^2$ per Coulomb$^2$ i.e. N m$^2$ C$^{-2}$.

To use the principle (1.17) to find the field at some point you must look for a suitable 'box' (called a "Gaussian box") to enclose the charges you have. Then you can often find a simple argument that will give you the field at the surface of the box. The following examples involve the fields in and around **conducting materials**, such as copper and other metals. The first thing to note is that since charges move freely in a conductor the field

intensity $E$ must be zero at all points inside it – for otherwise the charges could never come to rest. The fields *outside* metal objects, however, are very important: we look at three examples.

---

**Example 6.** Field at surface of a charged conductor

Suppose we want the field intensity $E$ just outside a charged conductor, whose surface is plane if we look at a small enough part of it. In this case we can imagine a Gaussian box with two plane faces, both parallel to the surface, one inside the conductor and the other just outside, as in Fig.6. To get the field at a 'field point', marked by the black dot on the surface, we argue as follows.
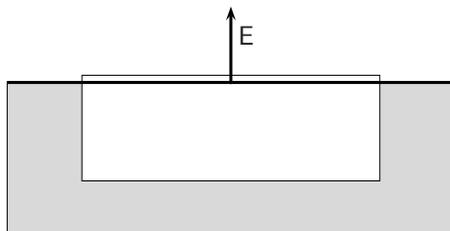


**Figure 6.**
Gaussian box (shown in white)
enclosing piece of surface
$E$ = field vector normal to surface

The field $E$ just outside the conductor must be *normal* to the surface, because any component parallel to it would keep surface

27

charges in constant motion – whereas we are considering what happens when they have come to rest. The fields on the other surfaces of the Gaussian box (shown in white) must also be zero, as they are *inside* the conducting material We can then use (1.17) by considering a piece of surface of area $S$, forming the top of the box. The total flux of $\mathsf{E}$ out of the box will be $E \times S$, from the top side – and nothing else. The electric charge on the piece of surface that lies just inside the box will be $q = S\times$ (amount of charge per unit area); and the quantity in parentheses is the **surface density of charge**, usually denoted by $\sigma$ ( the Greek letter "sigma"). From (1.17) it follows that $E \times S = S \times \sigma/\epsilon_0$ and, dividing by the area $S$, that $E = \sigma/\epsilon_0$. This result is general, because any small enough piece of surface at the point where we want to know the field is 'flat'.

---

From Example 6 we conclude that **the electric field at the surface of a conducting material is equal to the surface charge density divided by** $\epsilon_0$:

$$E = \sigma/\epsilon_0. \qquad (1.18)$$

This is true for a conductor of any size and shape, as long as the bit of surface we're looking at is more or less flat (no sharp points sticking out from it! – something we'll come back to later).

---

**Example 7.** Field outside a charged wire

**Wires** are very important in all electrical gadgets: they are long strings of conducting material (usually copper, often with a plastic

cover) for carrying electric charge from one place to another. We often want to know how strong the field E is near a wire, and again we can get it by using a Gaussian box. In Fig.7 the thick black line represents part of a wire, of length $l$, which we can imagine surrounded by a box in the form of a cylindrical tube of radius $r$. Let's ask for the field E at distance $r$ from the wire i.e. at the surface of the Gaussian box. In Example 6 the important thing was the *surface* density of charge; but here it's going to be the *line* density, which we'll call $\lambda$ (Greek "lambda") – the charge per unit length.
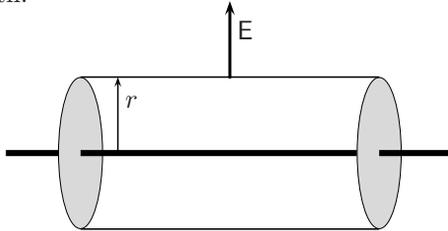


**Figure 7.**
Gaussian box (tube) around wire
E = field vector normal to wire
(box ends shown in grey)

If we take a small piece of wire, of length $l$, it will contain charge $q = l\lambda$. As in Example 6 this is the total charge inside the Gaussian box; and this is what produces the total flux out of the box. To get the flux we first note that the field E, at *all* points on the wire (which look exactly the same), must be perpendicular to the wire and therefore the flux from the box ends will be zero. The only non-zero flux will thus be from the curved surface of the tube: it will be (field $E^{\perp} = E$) × (surface area of tube). And, as you'll remember from Book 2, the area will be (length) × ($2\pi r$). Putting it all together, (1.17) gives for the total outward

flux $E \times 2\pi rl = l\lambda/\epsilon_0$. Thus, re-arranging the equation, we find $E = (\lambda/2\pi\epsilon_0)(1/r)$.

---

The result from Example 7 is that the field intensity at distance $r$ from a conducting wire, carrying charge $\lambda$ per unit length, is given by

$$E = \left(\frac{\lambda}{2\pi\epsilon_0}\right)\left(\frac{1}{r}\right) \qquad (1.19)$$

– so the field goes down slowly (like $1/r$) as you go away from the wire.

As a last example let's take a charged spherical surface with nothing inside it – a *hollow* metal sphere.

---

**Example 8.** Field around a closed spherical conductor

Suppose the conductor carries a charge $Q$ and has radius $R$. The surface density of charge is then $Q/4\pi R^2$ and according to (1.18) the field intensity at the surface is $E = (1/4\pi\epsilon_0)(Q/R^2)$. This agrees with the formula (1.4), which gives

$$E = F/q_0 = (1/4\pi\epsilon_0)(q/r^2)$$

for the field of a *point* charge $q$ at distance $r$, provided we put all the charge $Q$ at a single point and use the point-charge formula with $r = R$, the radius of the big sphere. In other words, the charged sphere (however big!) behaves like a point charge at all points *outside* i.e. with $r \geq R$.

But what about points *inside* the sphere in Example 8? (We said it was hollow.) Is there a charge at any point on the *inner* surface? And is there a field E at any point inside the surface – in the empty part? To answer these questions it's best to use the idea of electric potential (see Section 1.2), as follows in Example 9.

**Example 9.** Field *inside* a closed spherical conductor

When we draw a "line of force" it connects points between which there is a difference of potential; and E at any point on the line shows the rate of decrease of the potential as you go along the line. On the surface of a conductor there is no parallel component $E^\parallel$ because the charges are in equilibrium (we're talking about *electrostatics*; and this is the same as saying the conducting surface is an equipotential – $\phi$ has the same value at all points on it, $\phi_0$ say.

If there *was* a positive charge at some point on the inner surface of the hollow conductor, then a line of force would start from that point (look at Fig.6 for example). But where would it end? Following the direction of E, the potential $\phi$ would continually decrease: the line of force would be 'looking for' a negative charge to connect with! But there *are* no charges in the empty space; and if the line reaches another point on the inner surface it will be at potential $\phi = \phi_0$. There will be zero potential difference between starting point and end point; and no charge will flow if you connect them with a wire. It was therefore wrong to imagine a surface charge at any point on the inner surface of a hollow conductor: there can be no surface charge and the whole inner

31

region must be at the constant potential $\phi = \phi_0$. Moreover, the same argument applies for a hollow conductor of *any* shape, not just for a sphere.

---

From Example 9 it follows that the electric field intensity must be zero at all points inside a hollow conductor. You can sit inside a metal box, charged to a potential of ten million volts, without getting an electric shock! And it works the other way round: any electrical experiments you may do will have no effects outside the box.

# 1.5 Making electricity – and storing it

It seems a long way from principles to practice. Rubbing pieces of amber or some kinds of plastic with a woollen cloth isn't a very good way of generating electricity but that's how it all started; and we have to go step by step.

The first step was to invent simple devices for detecting electric charges, moving them around, and storing them.

### The electroscope

The **electroscope** detects electric charge, either positive or negative, by using the principle (Section 1.1) that charges of the same sign repel one another.
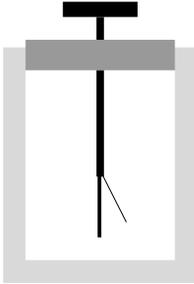
**Figure 8.**
Simple electroscope
– glass jar shown in light grey,
insulating lid in dark grey,
conducting material in black.

To see if an object is charged you touch the electroscope with it (i.e. the metal plate at the top). The charge flows freely into the conducting material, which ends with a small plate and (lying against it) a thin metal 'leaf' (usually 'gold-leaf'). Both the fixed plate and the leaf are then charged, either with negative or positive electricity, and repel each other. The leaf then sticks out at an angle, as shown in Fig.6, the small weight of the leaf being overcome by the repulsion: the bigger the angle, the bigger the charge. (When such a device is used for *measuring* the charge it's called an 'electrometer'.)

**The electrophorus**

This was one of the first devices for carrying electricity from one place to another ('phoros' being the Greek word for 'carrier'). Figure 9 shows how it works.
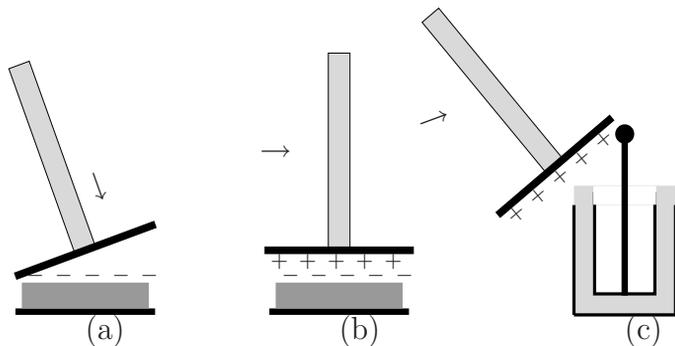
**Figure 9. Using the Electrophorus**
(a) Put metal plate on −ve charged insulator
(b) Touch top surface. Carry +ve charge away
(c) Put charge into store (Leyden jar) – see text.

You start by rubbing the insulating material (shown in dark grey) to give it an electric charge: if it's amber or any other 'resinous' material you'll use a woollen cloth and the charge on the surface will be *negative* (there will be extra electrons there). Then, in (a), you put the metal plate of the electrophorus, which has a glass (insulating) handle, down on top of the charged insulator. The plate won't be smooth enough to touch at every point on the surface, so over most of it there will be a small gap between the metal and the surface charge on the insulator.

What happens? The negative charge repels the electrons in the metal and drives them away from the surface,

which then gets a *positive* charge. This, shown in (b) as $+++++$ , is what you want to carry away and store somewhere. But you don't want it to get mixed up with the negative charge that was driven away; and that's why you have to touch the top surface of the metal plate to let the electrons escape through your body - which is a conductor - and finally through the ground you're standing on into the Earth, which is an *infinite* conductor and can absorb any amount of charge. This is called **"earthing"** the top surface of the electrophorus. Then, when you take it away, the plate will be left with the positive charge. But there's not much of it and where can you store it?

In the last picture (c) you can see the answer to the question. The glass jar, shown in light grey, has a metal lining (black) and also stands in a metal container. This a charge **storage device** or **condenser**; this particular one being a 'Leyden jar', named after the place in Holland where it was invented. If you touch the central metal rod with the charged plate of the electrophorus, the positive charge it carries will run freely into the Leyden jar, accumulating on the inner lining. Notice also that the charge you started with, in (a), is still there; it hasn't been used up in any way. In fact, it's charge of the *opposite* sign (in this example positive) that has been produced, carried away, and stored in the condenser. The positive charge on the plate of the electrophorus

was *'induced'* by the nearby negative charge, which attracts it: and we say it was put there by **induction** – nothing like '*con*duction', which depends on *contact.*

You can go through the steps (a), (b), and (c) many times (until finally the original charge 'leaks away', because no material is a *perfect* insulator). In this way, the charge in the condenser can be gradually built up; and the electric potential $\phi$, defined in Section 1.2 in terms of work done in moving a test charge, will also build up. Eventually, the condenser will hold a charge $Q$, say, on the inner surface and $-Q$ on the outer surface (the whole system being electrically 'neutral'. The outer surface in (c) is usually 'earthed' and its potential $\phi$ is taken as a convenient 'zero'. The potential of the inner surface may then reach thousands of volts and if you replace the electrophorus in Fig.9(c) by an earthed wire you will get a bright blue **spark** – a tiny flash of lightning – between it and the central conductor.

**Other electrostatic devices**

It's quite tiring to build up a lot of charge by using the electrophorus, because you may have to go through the (abc)-cycle hundreds of times and whenever you move electric charge through a field you have to do work (remember Section 1.2). So people had to invent ways of doing the job automatically. Two machines for doing it are (1) the Wimshurst machine, and (2) the van de Graaff generator, both named after their inventors.

36

*The Wimshurst machine*

Here the insulating material of the electrophorus is re-
placed by a circular disk, with metal plates stuck onto
its surface so that by rotating the disk they can be put
in contact with 'brushes' (made of thin metal wires).
These charge collectors can pick up or put down electric
charge as the plates pass them, as shown (very schemat-
ically) in Figure 10.
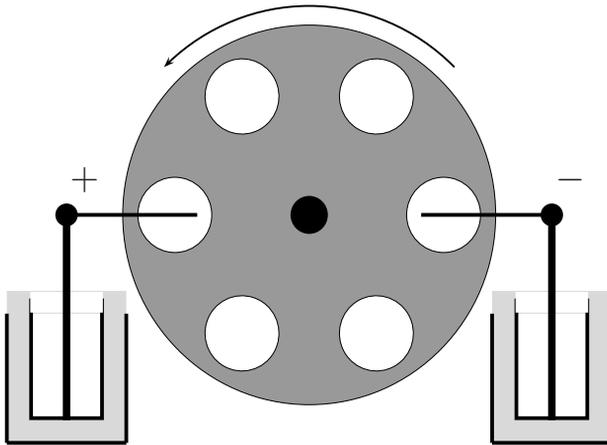


**Figure 10.** Wimshurst Machine (schematic)
Insulating disk (dark grey), attached metal plates (white)
Leyden Jars as in Fig.9 (cut open to show interior)
Curved arrow shows rotation, from $-$ to $+$ (see text).

Each metal plate, sitting on the electrified disk, is like

a small electrophorus and as it passes the right-hand collector it loses the electrons that were driven onto its top surface. The induced *positive* charge remains as the plate moves onwards, to the left. The same thing happens to the next plate as it comes along. When the charged plates reach the left-hand collector the brush picks up the positive charge from each plate in turn. The charges which are collected are then led away, through the conducting rods, onto the inner surfaces of the Leyden Jars; and as the plates continue on their way they again become positively charged, underneath, by induction (the big insulating disk still having its original negative charge). What a clever invention! No wonder Volta called his electrophorus " a perpetual electricity carrier". After a while the right-hand Leyden Jar stores a lot of negative electricity, the other one holding all the positive charge. Usually, a Wimshurst Machine has a *pair* of disks, rotating in opposite directions, and picks up twice as much charge in every rotation. And if you try to connect the positive and negative 'terminals', which may have a potential difference of several thousand volts, all the electricity you've produced will **discharge** in a spark.

*The van de Graaff Generator*

This is a more modern way of generating high voltages, still using the same idea of transporting charge (either $+$ or $-$) into a metal container, where it is to be stored.

It was invented only in the 1930s and is still in use for high-voltage experiments.

Instead of a Leyden Jar, the electricity is stored in a single large metal sphere, with an opening in the bottom just big enough to let charge be carried in. Think of the plate of an electrophorus being replaced by a metal disk or ball, a large number of them being carried on a continuous belt of insulating material (e.g. strong nylon fabric). Suppose a small metal ball, carrying positive charge (some electrons having being removed), is taken into the large conducting sphere and is touched by a wire connected to the inside of the sphere. Will it give up its charge or not? That will depend on the **potential difference** between the two ends of the wire: the charge will go only in the direction of decreasing potential. But in the following Example you'll be able to show that the electric potential *always decreases* in going from the small sphere to the large one; so the electric charge will always go that way, however much positive charge has already accumulated on the big sphere!

---

**Example 10.** Which way will the charge go?

From Example 3 we know that the field outside a spherical conductor, carrying a total charge $Q$, is the same as if $Q$ were a *point* charge at the centre of the sphere. This result followed from (1.17), with a Gaussian surface just outside the charged sphere,

and didn't depend on how $Q$ was made up – as long as was the total charge enclosed. Now suppose the big sphere (of radius $R$, say) has a small sphere (radius $r$) *inside it*, carrying its own charge $q$.

The total charge to be used in getting the field ouside the big sphere is then $Q + q$ and this means the potential $\phi$ at any point on the surface will be

$$\phi_R = \frac{1}{4\pi\epsilon_0}\frac{Q+q}{R} = \frac{1}{4\pi\epsilon_0}\frac{Q}{R} + \frac{1}{4\pi\epsilon_0}\frac{q}{R},$$

where the first term comes from the charge $Q$ on the big sphere itself, while the second comes from the charge $q$ carried by the small sphere.

Example 3 also showed that the potential from $Q$ was constant inside the hollow sphere, with the surface value

$$\frac{1}{4\pi\epsilon_0}\frac{Q}{R}.$$

The potential at the surface of the small sphere, however, will be

$$\frac{1}{4\pi\epsilon_0}\frac{q}{r},$$

due to its own charge $q$, *plus*

$$\frac{1}{4\pi\epsilon_0}\frac{Q}{R},$$

coming from $Q$ on the large sphere. The total potential at the surface of the small sphere will be the sum of the two terms:

$$\phi_r = \frac{1}{4\pi\epsilon_0}\frac{q}{r} + \frac{1}{4\pi\epsilon_0}\frac{Q}{R}.$$

On comparing this with the expression for $\phi_R$, the potential difference comes out as

$$V = \phi_r - \phi_R = \frac{q}{4\pi\epsilon_0}\left(\frac{1}{r} - \frac{1}{R}\right).$$

Assuming $q$ positive, $V$ must also be positive (since $R > r$), which means that charge will always flow 'downhill' from the small sphere to the large – *however large Q may be.*

From this result, it follows that the charge on the big sphere of a van de Graaff machine, and its electric potential $\phi$, will both go on increasing as long as the belt is being driven. Of course there is a limit: as the potential increases the *field* outside the big sphere gets bigger and bigger and it gets hard to carry more positive charge towards it. And even before that, when the potential reaches millions of volts, there will be a **discharge** in the form of a flash of lightening between the sphere and the ground, or the nearest object at a lower potential – and all the work you've done in building up the charge will be wasted.

*Condensers in general*

In the machines we've just looked at the electricity generated has been stored, either in Leyden jars or in a large metal sphere: these storage devices are both simple forms of *condenser*, of which there are many other kinds. One which shows up the general principles, which apply to all of them, is the **parallel plate condenser**.

Figure 11 indicates how it works.
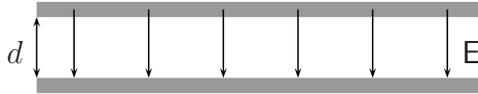


**Figure 11.** Parallel-plate condenser
Top plate, area $A$, surface charge $Q$
Bottom plate, surface charge $-Q$
$d$ = separation, $\mathsf{E}$ = typical field vector

If the top conducting plate. of area $A$, is given a positive charge $Q$ and the bottom plate a charge $-Q$, the charges will spread out uniformly until (when equilibrium is reached) they sit on the *inner* surfaces of the two plates. The lines of force between them connect the positive and negative charges and the field intensity $\mathsf{E}$ will have a magnitude $E$ given by equation (1.18), $E = \sigma/\epsilon_0$, where $\sigma$ is the surface charge per unit area.

Let $\phi_1$ and $\phi_2$ be the potentials of the upper and lower plates: then (see Section 1.2) the PD $\phi_1 - \phi_2$ measures the work done when unit positive charge moves through distance $d$ in the direction of the field: $\phi_1 - \phi_2 = Ed$. Putting charge in is called "charging the condenser" and it's useful to know how much charge has to be put in to get a voltage difference $V$ between the two plates. The next Example shows how you can find this.

**Example 11.** Capacity of a parallel-plate condenser

Suppose charge $Q$ has been put on the top plate in Fig.11. The surface charge density is $\sigma = Q/A$ for a plate of area A, all the charge being on the inner surface only. From Example 6 the field E between the plates has magnitude $E = \sigma/\epsilon_0 = Q/\epsilon_0 A$ and the PD will thus be

$$\phi_1 - \phi_2 = Ed = \frac{Qd}{\epsilon_0 A} = V,$$

where the PD, which is measured in Volts, is also indicated by $V$ and called the **voltage difference** between the plates. The amount of charge on either plate, $Q$. is thus directly proportional to the voltage difference: $Q = CV$.

The proportionality constant $C$ is called the **capacity** of the condenser. For a parallel-plate condenser, the previous equation shows that $C = \epsilon_0 A/d$. This is an approximate result because we haven't worried about 'edge effects', small corrections that are needed when the plates are not infinitely large and the lines of force near the edge are not straight but bent.

---

The important conclusion from Example 11 is that: for any electric storage device that works by separating the charges $\pm Q$ and setting up a PD $(V)$ between them the magnitude of the charge is directly proportional to the PD between the separated charges:

$$Q = CV \tag{1.20}$$

This is true for condensers of all kinds and gives us a general definition of capacity. But Example 11, for a

parallel-plate device is also of practical interest because it shows how you can *change* the capacity. For a parallel-plate condenser we found that $C$ has the value

$$C = \epsilon_0 A/d \qquad (1.21)$$

where $A$ is the area of overlap of the two plates and $d$ is their distance of separation. Thus, you can increase $C$ (i) by increasing the overlap-area of the plates, or (ii) by bringing them closer together (decreasing the separation $d$, or (iii) by putting a different insulating material between the plates, to change $\epsilon_0$ (the "permittivity of free space") to $\epsilon$, the permittivity of the material. More about this later; but condensers are very important, being found in nearly all electrical gadgets, and it's very useful to be able to vary their capacity for holding charge. Note that the unit of capacity will be, from (1.18), 1 coulomb per volt; this unit is called the 'farad' (after Faraday), so 1 farad = 1 C V$^{-1}$.

*Energy of a charged condenser*

Nothing much has been said so far about the amount of *energy* stored in a condenser. But the change in potential energy $V_B - V_A$, of a charge moved from point A to point B, was used in (1.8) to define the electric potential in terms of the work done. In the next Example, we'll use this this procedure to find how much work has to be done to charge a parallel-plate condenser by moving an infinitesimal amount $\mathrm{d}q$ from Plate 1 to Plate 2 and then

repeating this (over and over again) until the condenser has $+Q$ on Plate 2 and $-Q$ on Plate 1.

---

**Example 12.** Building up the energy in a condenser

The work done in moving $dq$ from Plate 1 to Plate 2 is $dq\,\phi_2 - dq\,\phi_1 = dq\,V$ where $V$ is the PD between the plates *at that time* – when charge has already built up to $q$, say.

The total work done in building up the final charge on each plate then follows (go back to Book 3 if you've forgotten about integration!) as

$$W = \int dw = \int_0^Q (q/C)dq = |\tfrac{1}{2}q^2/C|_0^Q = \tfrac{1}{2}Q^2/C$$

– and that is the electrical energy of the charged condenser. Since, from (1.20), the total charge is $Q = CV$, this can be written $W = \tfrac{1}{2}CV^2$.

Now where is that energy actually stored? The condenser looks exactly the same as it did before we started putting charge into it: it is only the *electric field* that has changed from zero to $\mathsf{E}$ at all points in space. So perhaps the energy is stored in the *field*? and, if so, how much is stored in an element $dV$ of volume? For the condenser in Fig.11, the field is uniform throughout the space between the plates and negligible outside: so the volume in which the field is non-zero will be (area of each plate)×(separation), $A \times d$. If $W$ is divided by the volume $A\,d$ it will give the field energy per unit volume i.e. the *energy density* of the field, usually denoted by $u$

To complete the calculation we need only substitute for the capacity $C$, given in (1.19), and divide by $A\,d$. The result (check it!) is $u = \tfrac{1}{2}\epsilon_0 E^2$.

The important results from Example 12 are the expressions for the energy of a charged condenser $(W)$, which is the work you must do to charge it, and the **energy density formula**, expressing this energy in terms of the *electric field* you've created. In summary

$$W = \tfrac{1}{2}CV^2, \tag{1.22}$$

for the work done in charging a condenser of capacity $C$, and

$$u = \tfrac{1}{2}\epsilon_0 E^2, \tag{1.23}$$

for the energy density.

Of course the 'proofs' are valid (and only roughly so) for a very special case, the parallel-plate condenser, but – remarkably! – the results are widely valid. In particular, the energy density formula (3.7) applies to the electrostatic field arising from any distribution of electric charge.

## 1.6 Differential form of Gauss's law

**A note to the reader.** The next two sections bring in some new and difficult ideas. Skip them on first reading and come back to them when you're ready.

The **flux** of a vector field across a surface was defined with the help of Fig.3 and some of its remarkable properties were noted just after equation (1.13). We used the idea in formulating Gauss's law in a very general way in (1.17) and found how easily it could give expressions for the field around conductors like plates and wires. But the law in that form can be applied only in very special cases, where the *whole system* has a simple shape. Later we're going to need a *differential* form of the law, which will let us talk about fields that change rapidly from point to point. The differential calculus (Book 3) has been used so far only in showing how a vector field, like $\mathsf{E}$, could be derived from a scalar field $\phi$ by means of the **gradient** operation, according to equation (1.12). Now we want to know about the variation of a *vector* field as we go from point to point; and again we'll need the calculus.

Let's start by talking about the flux of the field out of a very small 'box', a volume element $\mathrm{d}V = \mathrm{d}x\mathrm{d}y\mathrm{d}z$, like the ones you've used in Book 3. Think of the box as a cube, with six faces which we can call A, B, C, D, E, F, all with the same surface area $S$ – just to make things simple. Take the first pair (A and B, perpendicular to the x-axis) on which the field vectors are $\mathsf{E}_A$ and $\mathsf{E}_B$ as indicated in Figure 12.
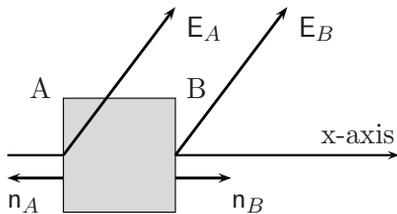
**Figure 12.** Volume element and field vectors
Faces A and B are perpendicular to x-axis,
Corresponding fields are $\mathsf{E}_A, \mathsf{E}_B$
and outward normals are $\mathsf{n}_A, \mathsf{n}_B$.

The important thing to keep in mind is the *direction* of
the vectors. In this example the field hasn't changed in
going from Face A to Face B, so the vectors are equal
(same length, same direction). But the flux out of a vol-
ume is defined relative to the *outward-pointing* normal;
so for Face B it is $S \times E_B^{\perp}$ (with the notation of Fig.3)
and is counted positive because the perpendicular pro-
jection points along the *positive* x-axis. On the other
hand, for Face A, the flux is counted negative because
the vector $\mathsf{E}_A$ points *into* the cube while the outward
normal is along the *negative* x-axis.

In vector language, which you first used in Sections 5.4 and
6.3 of Book 2, the projection of a vector $\mathsf{E}$ along the direction
of the unit vector $\mathsf{n}$ is the **scalar product** $\mathsf{E} \cdot \mathsf{n}$.); it is the

magnitude $E$ of the vector times the cosine of the angle it makes with $\mathsf{n}$. If you reverse the direction of the normal you must change the sign of the projection. So if you use the scalar-product definition the signs are taken care of automatically!

The flux of $\mathsf{E}$ out of the cube, through the faces A and B, can now be calculated as follows, using $\mathsf{n}$ for the unit vector along the positive x-axis:

Outward flux of $\mathsf{E}_A$ at Face A $= S\mathsf{E}_A \cdot \mathsf{n}_A = -S\mathsf{E}_A \cdot \mathsf{n}$,

outward flux of $\mathsf{E}_B$ at Face B $= S\mathsf{E}_B \cdot \mathsf{n}_B = S\mathsf{E}_B\mathsf{n}$.

When the two field vectors are equal. they make equal but opposite contributions to the flux out of the volume element – if no field is created (or destroyed) inside it, then what goes out must equal what comes in! All this is true for each pair of opposite faces and by adding the contribtions we can get the total flux of field intensity out of the volume element $\mathrm{d}V$. And, if no flux is created or destroyed inside it, this total outward flux must be zero.

But what if there *is* some charge in the volume element? We started in Section 1.1 from the idea that the field was *created* by 'something' that we called 'electric charge'; and that there were two kinds of charge, negative (meaning rich in electrons) and positive (short of electrons). The field was defined in terms of forces (which we knew how to measure) and this led, in Section 1.1, to the defi-

nition of the unit of charge – the Coulomb. In Coulomb's law (1.4) we were thinking of charge as a generator of the field, but after getting the idea of **flux** (Figure 3) we found another interpretation:

> Electric charge at any point in space creates an outward flux at the point – which can be positive, for positive charge; negative, for negative charge; or zero, if there is no charge there. The flux created by a point charge $q$ is $q/\epsilon_0$.

Now let's go back to Fig.12 and suppose the volume element contains charge $q$, which will create an outward flux $q/\epsilon_0$ if it's a positive charge or destroy the same amount if it's negative. Everything is beginning to fall into place! All we have to do now is to express these ideas in mathematical form. The x-axis is perpendicular to both Face A and Face B of the cube, so $\mathsf{E}_A \cdot \mathsf{n}$ can be called instead $E_x^A$ – the x-component of the vector $\mathsf{E}_A$ (where the label 'A' has been moved upstairs to keep it out of the way); and in general the field will depend on *position*, so if we suppose the bottom left-hand corner of the cube in Fig.12 is at point (x,y,z) then the field vectors $\mathsf{E}^A$ and $\mathsf{E}^B$ will correspond to $\mathsf{E}(x,y,z)$ and $\mathsf{E}(x+\mathrm{d}x,y,z)$, respectively. Note that $x=0$ at points on Face A, $x = x+\mathrm{d}x$ at points on Face B, and the x-components are thus $\mathsf{E}_x(x,y,z)$ and $\mathsf{E}_x(x+\mathrm{d}x,y,z)$. Here $\mathrm{d}x$ is the

infinitesimal change in $x$ in going from Face A to Face B; it is a **differential** (See Section 2.3 of Book 3). The corresponding increase in the x-component of the field is (rate of change)$\times \mathrm{d}x$ or

$$\mathrm{d}E_x = \frac{\partial E_x}{\partial x}\mathrm{d}x,$$

where the derivative on the right is a *partial* derivative because only $x$ has been changed, the other variables being kept fixed. (Section 6.1 of Book 3 will remind you of such things.)

There will be a similar increase in the component $E_y$ in going from Face C to Face D, by changing $y$ alone, both faces perpendicular to the y-axis,

$$\mathrm{d}E_y = \frac{\partial E_y}{\partial y}\mathrm{d}y.$$

The same is true for $E_z$ as you go from Face E to Face F, both perpendicular to the z-axis. Now multiply all field components by the areas out of which they point and add together the results for all three directions (noting that in general these areas will be $\mathrm{d}y\mathrm{d}z$, $\mathrm{d}z\mathrm{d}x$, and $\mathrm{d}x\mathrm{d}y$). The result will be (check it!)

$$\text{Total flux out of } \mathrm{d}V = \left(\frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z}\right)\mathrm{d}x\mathrm{d}y\mathrm{d}z.$$
$$(1.24)$$

This formula for the flux of the field out of a volume element $dV = dxdydz$, at any point in 3-dimensional space where there is no charge, is very important. It arises whenever we talk about vector fields and the way they can change. The quantity it defines is given a special name: it is the **divergence** of the field. Here it measures how quickly the lines of force move apart (i.e. 'diverge') as they go out from any point in space.

Just as the **gradient** of a scalar field $\phi(x, y, z)$ was defined in (1.13) in terms of its first derivatives, the divergence of a vector field (with three components) is defined in terms of first derivatives of its three components: we write (for short)

$$\text{div } \mathsf{E} = \left( \frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z} \right). \qquad (1.25)$$

The total outward flux from $dV$, given in (1.24), appears in the form (flux *density*)$\times$(volume). We now see that the statement div $\mathsf{E} = 0$ at some point P is a neat mathematical way of saying that there is zero outward flux per unit volume at that point.

If, on the other hand, there *is* some charge ($q$ say) inside $dV$ we can write $q = \rho dV$, where the letter $\rho$ is the usual name for the charge *per unit volume* and is called the **charge density** at the point where $dV$ is located. We started by talking only about *point* charges, but now we're thinking of the charge as being 'smeared out' – and

as long as $\rho = \rho(x, y, z)$ is concentrated inside the box $\mathrm{d}x\mathrm{d}y\mathrm{d}z$ it won't make any difference. However, (1.24) will now give, using the quantity div $\mathsf{E}$ defined in (1.25)

$$\text{Total flux out of } \mathrm{d}V = \text{div } \mathsf{E}\,\mathrm{d}V = \rho\mathrm{d}V/\epsilon_0.$$

This applies even when electric charge is spread out all over space, with a density $\rho$ which is any function of position. The volume elements may be cancelled and the final equation, which applies when charge is present, becomes

$$\text{div } \mathsf{E} = \rho/\epsilon_0. \qquad (1.26)$$

This is the *differential form* of Gauss's law: it is a **partial differential equation**. (See also Chapter 6 of Book 3.)

We already guessed the general form of the law in (1.19), which referred not to a *point* in space but rather to a volume of any size enclosing any number of charges. How are the two ideas connected? We're now going to start from charges and fields *at a point* and build up to whole systems in which the charges may be spread out in any way. This will give us a *proof* of the general form – the 'integral' form.

# 1.7 Integral form of Gauss's law

The key step in finding the law is to add a second box to the one used in Fig.12; then we can continue and build up the whole system out of tiny boxes. In Figure 13, we look at the first two volume elements – showing them slightly separated so as to see what's going on.
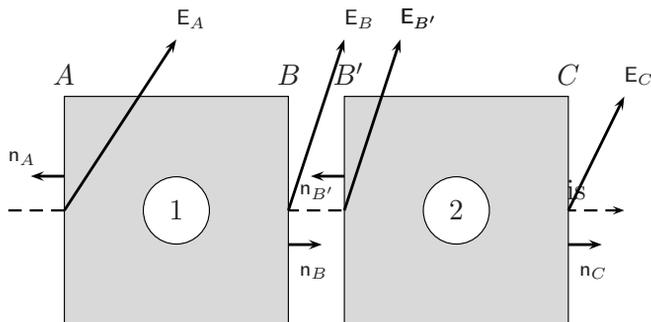


**Figure 13.** For proof of Gauss's theorem – showing two adjacent volume elements

---

**Example 13.** Building up the boxes

Think of the flux coming out of each box in Fig.13 (they have been labelled 1 and 2) and for the moment let's suppose the boxes are 'empty', containing no charge. As in Fig.12, the faces of the cubes can be named A, B, C, etc. but we first consider only the faces

perpendicular to the x-axis: we'll call them A and B, for Box 1, B′ and C, for Box 2. The surfaces B and B′ are not really separate: when the boxes are put together there's only *one* and it's called the **interface** between the two. The field intensity at each face is shown as a vector coming out from its centre and you'll note that $\mathsf{E}_B$ and $\mathsf{E}_{B'}$ are shown equal, in both length and direction, because they are the same vector. Now let's list the flux contributions for the four surfaces, using $\mathsf{n}$ for a unit vector along the positive x-axis, just as we did for Fig.12.

The flux of $\mathsf{E}$ out of the first cube, through the faces A and B, can now be calculated as follows, again using $\mathsf{n}$ for the unit vector along the positive x-axis:

Outward flux of $\mathsf{E}_A$ at Face A $= S\mathsf{E}_A \cdot \mathsf{n}_A = -S\mathsf{E}_A \cdot \mathsf{n}$,

outward flux of $\mathsf{E}_B$ at Face B $= S\mathsf{E}_B \cdot \mathsf{n}_B = S\mathsf{E}_B \cdot \mathsf{n}$.

The flux of $\mathsf{E}$ out of the second cube, through the faces B′ and C, are:

Outward flux of $\mathsf{E}'_B$ at Face B′ $= S\mathsf{E}_{B'} \cdot \mathsf{n}_{B'} = -S\mathsf{E}_{B'} \cdot \mathsf{n}$,

outward flux of $\mathsf{E}_C$ at Face C $= S\mathsf{E}_C \cdot \mathsf{n}_C = S\mathsf{E}_C\mathsf{n}$.

By adding all four fluxes, we get the x-part of the total flux out of both volume elements taken together. But the contributions from B and B′, which is really a single *interface* when the cubes are put together, is clearly zero: $S\mathsf{E}_B \cdot \mathsf{n} - S\mathsf{E}_{B'} \cdot \mathsf{n} = 0$, because $\mathsf{E}_B = \mathsf{E}_{B'}$, the two vectors being identical. The result is that:
Total outward flux in x-direction = flux from face A + flux from face C

---

The conclusion from Example 13 can be stated generally as follows:

The total flux of E out of two adjacent volume elements, put together, can be calculated as either (i) the sum of the two separate fluxes, or (ii) the total flux from both together – **but not counting their common interface**.

And if this is true for *two* volume elements it must be true for any number, as we add them one at a time and in any way we please. Whenever two faces come together, forming one *interface*, we can simply forget about their flux contributions – which will cancel out.

By putting together a large number of very small volume elements, we can build up a good approximation to a solid object of any shape; and as the elements become *infinitesimal* the approximation becomes more and more perfect. This is the method of the **Integral Calculus** (Book 3, Section 3.2). The important thing is that, in getting the flux out of the whole object, we have to think only of volume elements *at the surface*. The total outward flux will then be the sum of contributions from all elements of volume and this becomes, in the limit, a **volume integral**. The presence of charge at any point in space also contributes to the flux, as in the displayed statement just before (1.26), giving a volume integral of

the charge density. The final result is then

$$\text{Total flux out of } V = \int_V \text{div } \mathbf{E} \, dV = \int_V \rho dV/\epsilon_0,$$
$$(1.27)$$

where $\int_V$ simply means "integral over the whole volume". This is a new mathematical statement of Gauss's law in the form original form (1.17), the integral on the right-hand side giving the total charge contained within the volume, divided by $\epsilon_0$.

It only remains to express the outward flux as a **surface integral**. The flux arises from the exposed faces of all the volume elements at the surface; but as the boxes become infinitely small they will, in the limit, approach a smooth surface. The flux out of a **surface element** $d\mathbf{S} = dS \, \mathbf{n}$ (where $\mathbf{n}$ is the outward-pointing unit vector normal to the surface) is just $\mathbf{E} \cdot d\mathbf{S}$: the total flux is thus the surface integral

$$\text{Total flux of } \mathbf{E} \text{ out of volume } V = \int_S \mathbf{E} \cdot d\mathbf{S}. \quad (1.28)$$

On putting this alternative expression for the flux into (1.27) we get a very fundamental result, called **Gauss's Theorem**:

$$\int_S \mathbf{E} \cdot d\mathbf{S} = \int_V \text{div } \mathbf{E} \, dV. \quad (1.29)$$

This is usually proved in books on Mathematics: here it comes out naturally, along with its meaning, in talking

about fields in Physics. Even if we didn't have it we'd need to invent it!

What the theorem does is to relate a surface integral of the field vector $\mathsf{E}$ to a volume integral of another quantity, div $\mathsf{E}$, which can be derived from the field: in this way a *2-dimensional* integral is expressed as a *3-dimensional* integral of a density, which is a scalar quantity defined at all points within the volume enclosed by the surface. The vector and its divergence are both important quantities in Physics and it's often useful to be able to pass from one to the other.

There's a similar theorem which relates the **line integral** of a field vector, around a closed circuit, to the integral of another quantity taken over any surface bounded by the circuit (giving a 1-dimensional integral in terms of a 2-dimensional one). You might be guessing that this has something to do with the idea of a **path integral**, which we met in talking about the electric potential – and you would be right.

If you read Section 1.2 again, you'll remember that the electric potential $\phi(x, y, z)$ at any point, with coordinates $x, y, z$, was defined by a path integral: the work you have to do in taking unit positive charge from Point A to Point B is

$$W(A \rightarrow B) = \int_A^B \mathsf{E} \cdot \mathsf{ds}$$

where the tiny vector ds represents an element of displacement along the path and the scalar product $\mathsf{E} \cdot \mathsf{ds}$ is the work done in that small step. The integral gives you the total *change* in potential, $\Delta\phi(A \rightarrow B)$. The fundamental property of this quantity is that it doesn't depend on which path you take from A to B and this means that the potential is a unique function of position $x, y, z$. So if you come back to the starting point (putting B = A) there will be zero change. This property of the field is often expressed by writing the work equation (above) as

$$\oint \mathsf{E} \cdot \mathsf{ds} = 0, \qquad\qquad (1.30)$$

where the symbol $\oint$ simply means the integral "taken round any closed circuit". Equation (1.30) is what ensures that the scalar quantity $\phi(x, , y, z)$ can be defined and that it is related to the field $\mathsf{E}$ by (1.14), namely $\mathsf{E} = -\mathrm{grad}\ \phi$.

To end this long chapter, let's squeeze the whole of Electrostatics (everyhing to do with charges at rest and the fields they produce) into one simple-looking statement:

The electric field $\mathsf{E}(x, y, z)$ at any point $(x, y, z)$ in space, due to a static distribution of charge, of density $\rho(x, y, z)$ per unit volume, can be obtained by solving the single equation $\operatorname{div} E = \rho/\epsilon_0$.

$$(1.31)$$

Of course this is not easy to solve (it's a partial differential equation in three variables), but it represents an enormous step forward in our understanding.

# Chapter 2

# Charges in motion: Electric currents

## 2.1 Flow of charge in a conductor

In the last chapter, we were dealing mainly with charges in equilibrium i.e. when they had come to rest. The charges may be positive or negative, like those on the two plates of a condenser (Fig.11), and an electrically neutral body will have zero net charge (equal amounts of positive and negative charge). But if the charges are allowed to mix (e.g. by putting a conductor between the two plates of a condenser) they will reach a new

equilibrium: **current flow** will take place until finally there is no longer any electric field at any point in the conductor. In this first section we want to understand what is happening.

Suppose then we have a **continuous medium** which conducts electricity and is exactly the same at all points. During the time that currents are flowing, the charge density $\rho$ will be changing as charge moves from one place to another: it may be positive or negative at any point (more $+$ charge than $-$ charge, or the other way round) and it will depend on time $t$ as well as position $x, y, z$. As an equation

$$\rho = \rho(x, y, z, t). \tag{2.1}$$

The electric field components $E_x, E_y, E_z$ and the potential $\phi$ will all be functions of the same four variables.

At any instant $t$, every little bit of charge $q$ (or $\rho dV$ if it is spread out continuously) will feel the electric field $\mathsf{E}$ which tries to move it. Suppose it moves with velocity $\mathsf{v}$, a vector pointing in the direction of the field: then $q\mathsf{v}$ gives the velocity with which the charge is being moved – and this is an **electric field**.

When the charge is spread out with density $\rho$ per unit volume, a **current density vector** may be defined as

$$\mathsf{j} = \rho\mathsf{v}. \tag{2.2}$$

This is a current *density* because it is "per unit volume": the amount of charge being moved (in volume element d$V$) is $\rho$d$V$.

In a continuous conductor, charge may be flowing in any direction and we may want to know how much crosses any bit of surface, of area d$S$, per second. This will be called the **flux** of current density across the surface element. And flux is something we know about already from Chapter 1, Figure 3. Here, instead of the electric field $\mathsf{E}$, we'll be thinking of charge moving with velocity $\mathsf{v}$ across d$S$ as in the Figure below.



(a)                    (b)
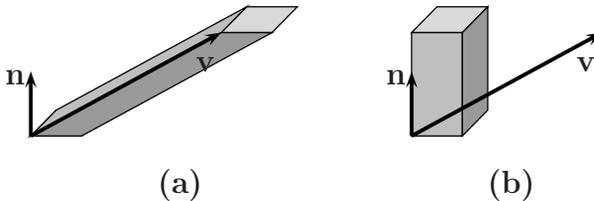
**Figure 14. The flux vector**

In Fig.14(a) the charge coming out of d$S$ in 1 second, will have travelled a distance $v$ along the vector $\mathsf{v}$ and will occupy the long thin 'box' shown, closed at its bottom end by the surface element d$S$ and at its top end by a parallel piece of surface (shown in pale grey). The *amount* of charge in that box will be (volume of box)× (charge per unit volume); but what is the volume of such

an odd shape?

This is something you did long ago in Book 2. If you imagine the long thin box cut into thin slices, parallel to the base, it could be pushed over to one side as in Fig.14(b), where the unit vector $\mathsf{n}$ is normal to the surface element. It then stands on the same base and has the same vertical height (the same number of slices, of the same thickness) as before you pushed it. The volume is thus *invariant* – $\mathrm{d}V$ =(area of base)×(vertical height)= $\mathrm{d}S \times \mathsf{v} \cdot \mathsf{n}$ – and the amount of charge it holds is $\rho\mathrm{d}V$. This will be the charge passing through $\mathrm{d}S$ in 1 second; and this is the flux density,

$$\text{Flux of } \mathsf{j} \text{ across } \mathrm{d}S = \mathsf{j} \cdot \mathsf{n}\mathrm{d}S = \mathsf{j} \cdot \mathrm{d}\mathsf{S}, \qquad (2.3)$$

where $\mathrm{d}\mathsf{S}$ in the last form is called (Book 2, Section 6.3) the **vector area** of the surface element. ($\mathrm{d}\mathsf{S}$ combines the magnitude of the area with the direction $\mathsf{n}$ of the normal, $\mathrm{d}\mathsf{S} = \mathsf{n}\mathrm{d}S$.)

For a continuous surface of any kind, not just a tiny element $\mathrm{d}S$, the **total current** through the surface may be found by summing the flux out of every small element into which it may be divided. The result is written

$$I = \int_S \mathsf{j} \cdot \mathsf{n}\mathrm{d}S = \int_S \mathsf{j} \cdot \mathrm{d}\mathsf{S}, \qquad (2.4)$$

where $\int_S$ is the notation for a **surface integral**, as used already in (1.29)for total flux of the *field* $\mathsf{E}$.

For a *closed* surface containing no charges (which generate the field), the surface integral of the field intensity E must be zero. And in the same way the total outward flux of the current density j will be zero whenever no current is *put in* (or taken away) at points inside the surface. In other words

$$\int_S \mathsf{j} \cdot \mathsf{dS} \quad \text{(no source or sink of current within S)}. \quad (2.5)$$

Here 'source' means a point where current *enters* the system, 'sink' being a point where it *exits*: such points usually correspond to **electrodes** or **terminals**. More about that later. First we need to know how currents are related to the electric fields that produce them.

## 2.2 Conductivity

We know from Chapter 1 that the field intensity E can be related to the electric potential $\phi$: the field vector at any point measures the rate of *decrease* of $\phi$ and it is often convenient to introduce the potential difference (in Volts), so that $E$ will be in "Volts per metre". For a given PD, the rate of current flow is found experimentally to be very variable: it is high for a good conductor, like copper, but very low for many non-metals, such as dry wood or silk. Moreover, it is usually proportional, in

good approximation, to the PD. This suggests a general relationship of the form

$$\mathsf{j} = \sigma \mathsf{E}. \tag{2.6}$$

The proportionality constant $\sigma$ depends only on the material (and to some extent on the temperature), but has nothing to do with the shape of the conductor: it is the **conductivity** of the material. The experimental result (2.6) is called **Ohm's law**.

One special kind of conductor is a long thin **wire**, used for carrying electricity from one place to another. The total current through the wire is then given by (2.4). If you have a current supply (e.g. a battery, like the one in your torch or portable radio) you can send current into a conductor of any size and shape, through a terminal, and take it out through another. What happens inside can then be determined by the equations we already have. Of course this can be complicated for conductors of any shape – but we shalln't need to do it here! We'll only deal with simple **electric circuits** in which wires are used to connect things like condensers; and we'll do that in a later section. However, it's nice to know that the basic equations can very easily be extended to allow for currents flowing into and out of a continuous medium. (You can skip the next bit at this point if you like!)

**Example 1.** Conduction in a continuous medium

Suppose the conductor, of volume $V$ is enclosed within a surface $S$ and contains charge, smeared out with a density $\rho$. The total charge it holds is then $\int_V \rho \, dV$ and this will change with time at a rate $\int_V (\partial \rho / \partial t) dV$; and this rate of *decrease* must be equal to the amount of charge flowing out across the surface $S$. As an equation,

$$-\int_V (\partial \rho / \partial t) dV = \int_S \mathsf{j} \cdot dS.$$

This doesn't look very promising! But remember Gauss's theorem (1.29), which relates the surface integral of a vector field to the volume integral of the *divergence* of the field. If we put $\mathsf{j}$ in place of $\mathsf{E}$, the theorem gives

$$\int_S \mathsf{j} \cdot dS = \int_V \operatorname{div} \mathsf{j} \, dV$$

and putting this result in the previous equation leads to

$$\int_V \left( \operatorname{div} \mathsf{j} - \frac{\partial \rho}{\partial t} \right) dV = 0.$$

Now this must be true for *any volume whatever*; and that can be so only if the coefficient of $dV$ is zero for *all* volume elements i.e. at all points in space. *The integrand must therefore be zero everywhere.*

---

The final result from Example 1 is thus

$$\left( \operatorname{div} \mathsf{j} - \frac{\partial \rho}{\partial t} \right) = 0. \tag{2.7}$$

Let's repeat: this equation simply says that "the rate of decrease of charge, within any volume $V$ enclosed by a surface $S$ is equal to the rate at which charge flows out across the surface" – which is just common sense! Of course, in a **steady state**, where the charge density is everywhere independent of the time, $\partial\rho/\partial t = 0$. In that case, (2.7) reduces to the simpler equation $\operatorname{div} \mathsf{j} = 0$. which is much easier to solve. In later sections we'll usually be dealing with steady states in which there may be potentials ($\phi$), fields ($\mathsf{E}$), and currents ($\mathsf{j}$) at any point in space – but they don't depend on time.

## 2.3   Some simple circuits

In Chapter 1 it was noted that one specially simple and useful kind of conductor was a long thin **wire**, which could carry charge from one place to another. In this section we start to look at **electric circuits** in which things like condensers and other **circuit elements** are connected together by wires. You can think of a wire as a long cylinder of conducting material (usually metal) with a circular cross-section. It can be bent into any shape you wish, as in Figure 15 below, and the total current going into or out of it at its ends
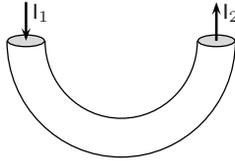
**Figure 15.** Wire, bent into semi-circle

$I_1$ = current vector in,    $I_2$ = current vector out

can be calculated as a surface integral of the flux density
$j$. The nice thing about this current $I$ is that $I_1$ and $I_2$
are the same, even when the wire is bent, because of the
'conservation equation' (2.7): this says that what goes
in equals what comes out – even when the currents point
in different directions (as in Fig.15) they must have the
same magnitude. (The flux density $j$ is zero over the
sides of the wire, because that's where the metal stops,
so the only contributions to the outward flux are those
from the ends, giving $I_2 - I_1 = 0$ in a steady state.)

Now that we know about wires we can go back to the ex-
ample considered at the beginning of Section 2.1 and ask
what happens if we connect the two plates of a charged
condenser. When we're thinking of wires, carrying cur-
rent from one place to another, it's useful to work with
the total current $I$ and to ask how it depends on the

69

potential difference between the two ends of the wire. Instead of (2.6), which defines the *conductivity* of a material under the influence of an electric field, we measure the **resistance** $R$ of a conducting wire by writing

$$I = V/R, \tag{2.8}$$

where $V$ is the PD, usually measured in volts, between the ends (or **terminals**) of the wire, and $I$ is the current it produces. The current is thus *inversely* proportional to the resistance: the bigger the resistance, the smaller the rate of charge flow. This equation is also often referred to as "Ohm's law", being simply another way of putting (2.6). The unit of $R$ is 1 Volt per Coulomb and is called the "Ohm": its symbol is the Greek capital letter $\Omega$, pronounced "omega". (And an easy way of remembering (2.8) is by saying to yourself "Volts equal Amps times Ohms".)

Let's draw a **circuit diagram** to show what we're talking about. Figure 16 shows the two plates of a charged condenser, on the left, which can be connected by a wire, through a **switch**. The switch is shown 'open' so as to make a break in the conducting wire, but by 'closing' it the current can pass from one side to the other. After the switch the wire can be as long as you like, so as to offer a bigger resistance, and is shown as a 'jimpy' line. After that, going clockwise, the wire finally connects with the other plate of the condenser. So when

the switch is open the plates are not connected and the charge on them has nowhere to go; but on closing the switch the circuit between the top plate and the bottom plate is completed and current can flow. When the charges $+Q$ and $-Q$ meet they make a total charge zero: the condenser is then fully discharged and there will be zero potential difference between the plates.
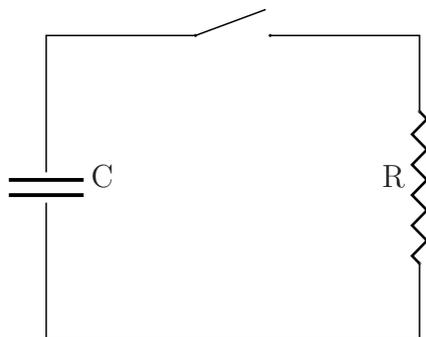


**Figure 16.** Discharge of condenser C
Condenser plates (left) are connected,
when switch (top) is closed, through
resistance R ('saw-tooth' line on right)

But what is happening in the short time it takes to discharge the condenser? To find out we have to solve a simple **differential equation**, of the kind we first met in Book 3 (Section 6.2). This will be our first example of how to deal with an electric circuit.

---

**Example 2.** Discharge of a condenser

The *rate* at which the charge $Q$ 'decays' is, in calculus notation, $-(\mathrm{d}Q/\mathrm{d}t)$ and this rate of outflow of charge is the current $I$ in the wire connecting the plates when the switch is closed:

$$I = -\frac{\mathrm{d}Q}{\mathrm{d}t}.$$

But we know from (2.8) that $I$ is related to the PD (also called the "voltage") across the plates of the condenser by $I = V/R$ and that $V = Q/C$ in terms of the capacity $C$; so $I = Q/(RC)$. On putting this last expression for $I$ into the previous equation we find

$$\frac{\mathrm{d}Q}{\mathrm{d}t} = -\frac{Q}{RC}$$

and this is the differential equation we need: $Q$ varies with time, as the condenser discharges, but $R$ and $C$ are constants. The solution of this equation is a standard result (Section 3.4 of Book 3). It is $Q = A \exp\left(-\frac{1}{RC}\,t\right)$, where $A$ is any constant. To fix the value of $A$ let's use $Q_0$ for the value of $Q$ when the condenser is fully charged at $t = 0$. The last equation then says $Q_0 = A$, telling us what the constant means.

---

The conclusion from Example 2 is that when the plates of a charged condenser are connected through a resistance $R$ the initial charge $Q_0$ (at $t = 0$) falls exponentially to zero as $t \to \infty$:

$$Q = Q_0 \exp\left(-\frac{1}{RC}\,t\right). \qquad (2.9)$$

72

The charge $Q$ goes down quite slowly if the capacity $C$ and resistance $R$ are big, so that $1/(RC)$ is small; but very fast if they are small, so that the coefficient of $t$ is large.

Before looking at other examples, we need to add another **circuit element**, a **battery**, in order to *generate* an electric current. So let's replace the condenser in Fig.16 by a battery, as in Fig.17, where it stands in the same position, without changing the rest of the circuit except that we'll include a lamp in it – just to show when a current is flowing.
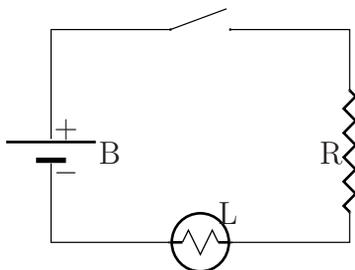


**Figure 17.**
Circuit with battery B and lamp L
When switch is closed, current goes from
$+$ to $-$, first through resistance R, then
through lamp filament (see text).

73

---

**Example 3.** Circuit with a battery and a lamp.

You'll find how batteries work in Section 2.4, but for now it's
enough to know that they produce a **voltage** between their two
terminals, indicated by the + on the upper plate in Fig.17 (higher
potential) and the − on the other plate (lower potential). The
charges on the two plates don't just disappear when the switch is
closed: they give a steady electric current that runs around the
circuit until the switch is opened again – or until the battery is
'run down'. And if you put a small lamp in the circuit, as shown
in the Figure, it lights up.

The light can be switched on or off whenever you like: it's not
like the condenser discharge, which can happen only once. And
instead of just giving a spark the battery is doing useful work.

---

Electric circuits can become very complicated if many
different elements are connected together (think of your
portable radio) but they can all be understood by us-
ing two simple principles, which we already know about
from earlier sections. Both apply to properties of the
electric field $\mathsf{E}$, and the potential $\phi$, inside a conductor;
so they are important also for wires.

The first thing to remember is that the wires are there
only to connect the circuit elements: if they were 'per-
fect' they wouldn't hold any charge in themselves and
they wouldn't offer any resistance to a current passing

through them – they would be perfect conductors. We can only imagine them, so we call them **"ideal wires"**. Thin wires of copper are pretty good approximations.

The field inside a wire, that makes electric charge move, is often called the **electromotive force**, or "emf" for short, and is denoted by $\mathcal{E}$ to make it different from the field vector $\mathsf{E}$ at any *point* in the conducting material. It is measured in volts (not volts *per unit distance*) and the voltage between the terminals of a battery is the emf it can produce. If unit positive charge passes through a circuit element, the work done on it is just the PD between its terminals i.e. the 'voltage drop' across the element. If it passes through any number of circuit elements, getting back finally to the starting point, then the total work done must be zero (its potential is the same as at the start). To include a battery as one of the circuit elements you only need include its emf $\mathcal{E}$ as one of the voltage drops, getting the signs right by noting the direction of the current flow (from positive to negative potential). This is one of **"Kirchoff's laws"**: the sum of the voltage drops, going round any closed circuit must be zero.

Of course, if two or more wires join at a point, the currents flowing into that point must add up to zero – for otherwise charge at that point would go on building up forever!

These conclusions apply to any **network** in which circuit elements are connected by ideal wires. They are usually stated as follows, where the word "node" is used to mean any connected set of terminals:

Kirchoff's first law:
The sum of the currents flowing into any "node" $n$ in a network of elements connected by ideal wires must be zero: $\sum_n I_n = 0$.

Kirchoff's second law:
The sum of the potential drops $V_N$ across the circuit elements $N$ in any closed circuit within a network must be zero: $\sum_N V_N = 0$.

The next two Examples show how easy it is to apply these two rules.

---

**Example 4.** Resistances in series.

This example shows what happens if you connect two resistances 'in series' – one after the other, as in Fig.18 where the current goes through $R_1$ and then on through $R_2$.
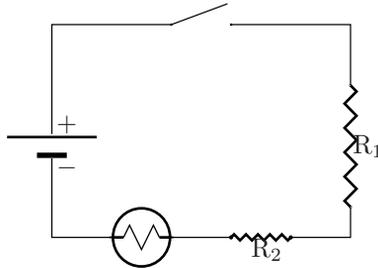
**Figure 18.** Resistances in series.
Current $I$ passes first through $R_1$,
then through $R_2$, and finally through
the lamp.

When the switch is closed, the (positive) current flows from the top plate of the battery towards points at lower potential: call it $I$ when it enters the resistance $R_1$. By Kirchoff's first law the current leaving $R_1$ will also be $I$ – what goes in must equal what comes out – and by (2.8) the potential drop between entering and leaving will be $V_1 = IR_1$. The current $I$ continues on through $R_2$ and when it comes out there will be another voltage drop, again from (2.8), $V_2 = IR_2$. Then there will be a voltage drop, $V_L$ say, when the current goes through the lamp and gets back to the bottom terminal of the battery. The total potential change on reaching that point will thus be $-V_1 - V_2 - V_L$, with the minus signs because the potential is *dropping*. But then, to complete the circuit, there will be a *positive* change in potential – namely the emf of the battery, which keeps the top plate at potential $\mathcal{E}$ above that of the lower plate. So the total change in coming back to the starting point will be

$$\mathcal{E} - V_1 - V_2 - V_L = \mathcal{E} - V_L - I(R_1 + R_2)$$

and by Kirchoff's second law this must be zero.

---

The important thing to note from Example 4 is that the potential drop across the two resistances *in series* is just the same as a single term $IR$: in other words you can replace the two resistances, connected that way, by a single resistance $R = R_1 + R_2$. And if you connect any number of resistances in series they will clearly give you a similar result,

$$R = R_1 + R_2 + R_3 + \ ... \ = \sum_i R_i. \qquad (2.10)$$

Next look at Figure 19, which shows the same circuit, but with the resistances connected 'in parallel'.
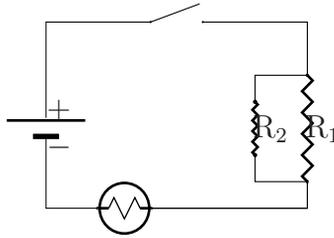


**Figure 19.** Resistances in parallel. Current $I$ passes through R$_1$ and R$_2$, connected 'in parallel', before going through the lamp.

**Example 5**. Resistances in parallel.

This example shows what happens if you connect two resistances 'side-by-side'. Here all the circuit elements are just the same as in Fig.18 and we can go ahead, using Kichoff's laws in exactly the same way. In this case, however, there is an important **node** where the top terminals of $R_1$ and $R_2$ are both connected by wires to the wire coming from the switch. At the point where the three wires meet, the current coming in is $I$ while the currents going out are $I_1$ (into resistance $R_1$) and $I_2$ (into resistance $R_2$). So Kirchoff's first law tells us that $I - I_1 - I_2 = 0$: the sum of the currents through the two resistances is therefore $I_1 + I_2 = I$ (the current coming from the battery, when the switch is closed).

To apply Kirchoff's second law, let's use $V$ for the voltage drop between the node at the top of the two resistances and that at the bottom. We can then say, using (2.8), that $V = I_1 R_1$, $V = I_2 R_2$. But since we just found that $I_1 + I_2 = I$ we can also say

$$\frac{V}{R_1} + \frac{V}{R_2} = I.$$

Now think of the two resistances in parallel as a single circuit element. If you divide all terms in the last equation by $V$ you find

$$\frac{1}{R_1} + \frac{1}{R_2} = \frac{I}{V} = \frac{1}{R},$$

where, according to (2.8), $R$ is the value of a *single* resistance equivalent to $R_1$ and $R_2$ connected in parallel. If we go round the whole circuit, as we did in Example 4444 for resistances in series, the second law will give

$$\mathcal{E} - V_L - I \left( \frac{1}{R_1} + \frac{1}{R_2} \right).$$

Here $V_L$ is again the resistance offered by the lamp, but the last term is now $-IR$ where $1/R$ is the sum of the *reciprocals* of the parallel-connected resistances.

---

The result in Example 6 extends to any number of resistances connected in parallel:

$$\frac{1}{R} = \left( \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \ ... \ \right) = \sum_i \frac{1}{R_i}. \qquad (2.11)$$

The above Examples are enough to give you a start: now draw a few other circuits (fairly simple!), with circuit elements connected together in various ways, to get new **networks**. By applying Kirchoff's two laws you can always get enough equations to find all the currents and voltage drops in all parts of a network – no matter how complicated!

# 2.4 Electricity, Chemistry, and Batteries

Just over 200 years ago it was found that electricity could be produced in a new way, completely different from the one used in making 'static electricity' (Chapter 1). Alessandro Volta, an Italian, noticed that two plates of different metals, separated by absorbent mate-

rial soaked in saltwater, came to different electric potentials. This combination of two plates with a salt solution between them came to be known as a **"voltaic cell"**.

The potential difference (PD) produced is very small (around 1 volt), compared with the thousands of volts coming from electrostatic generators. But by piling up a large number of plates, alternating the two metals and separating every pair with a layer of salty material, you get a **voltaic pile** or **battery** in which the PD between the top and bottom plates (of different metals) can easily reach 100 V or more: this is enough to feel if you touch one plate with each hand (especially if your two hands are wet to give good contact); and if you connect them with a wire you'll get a spark. The electricity is exactly the same 'stuff' as you get by rubbing a piece of amber with a woollen cloth; but now it comes from a **chemical reaction** (see Book 5) between the metal plates and the salt solution. The study of the relationship between electricity and chemistry is called **electrochemistry**. (You may need to look back at some parts of Book 5 before going on with this section.)

Remember (Section 1.1 of Book 5) that some kinds of solid matter, like common salt (sodium chloride, NaCl) consist of **ions** (atoms that have lost or gained one or more electrons, becoming +ve or −ve ions respectively). In the solid state the ions form a regular 'lattice' as in Fig.9(a) of Book 5 and are held together by electric

forces. But if the salt is dissolved in water the ions are 'loosened' and can go away and get mixed up with the **molecules** ($H_2O$) of the liquid, giving an **aqueous solution** of the salt.

The positive and negative ions in the solution can be moved apart by putting two conductors into the solution and keeping them at different electric potentials – supposing you have some way of doing that. The two conductors are then called **electrodes**: the one at higher potential, with positive charge, is the **anode**, while the one at lower potential is the **cathode**. This process of separation is called **electrolysis** and the solution between the electrodes is said to be the **electrolyte**.

In the example of sodium chloride, the positive ions $Na^+$ go towards the electrode of lower potential, while the negative ions $Cl^-$ go towards that of higher potential. The positive ions are 'neutralized' by regaining the lost electron ($Na^+ \rightarrow Na$) from the negatively charged cathode, while the $Cl^-$ ions lose their extra electron to the positive electrode: the neutral atoms, Na and Cl, finally form sodium (metal) and chlorine (gas) at their respective electrodes.

Keeping this picture in mind, let's think about the voltaic cell with two electrodes, zinc (chemical symbol Zn) and copper (Cu), metals which are both easy to find. The metal atom Zn loses *two* electrons very easily to form the ion $Zn^{2+}$, while the atom Cu can *take on* 2 elec-

trons to form $Cu^{2-}$. These processes can be written as 'chemical reactions'

$$\text{Zn}\rightarrow \text{Zn}^{2+} +2e^- \uparrow, \quad \text{Cu} +2e^- \downarrow \rightarrow Cu^{2-}$$

where $\rightarrow$ means "goes to", $\uparrow$ means "goes out", and $\downarrow$ means "comes in".

Whereas electrolysis takes place only when an electric current is *put into* the system, the voltaic cell works the other way round, as Volta discovered: the voltaic cell lets a chemical reaction 'go by itself', delivering electricity to the electrodes – from which it can be *taken out* of the system through wires. In other words, **Electrolysis needs an *input* of electrical energy. The voltaic cell gives an *output* of electrical energy.**

What we want to know next is How much? – and can we get a quantitative understanding of how chemical energy is changed into electrical energy?

First, remember that the 'chemical' unit of quantity is the **"mole"** (short form 'mol'), which means that 1 mole of zinc is $L$ zinc atoms ($L$ being Avogadro's number, $6.024{\times}10^{23}$), and similarly for 1 mole of $Zn^{2+}$ ions. In the same way, we can talk about 1 mole of *electrons*; and since each electron carries an electric charge $e \approx -1.6 \times 10^{-19}$ C we can say

One mole of electrons carries an electric charge of
$(6.024 \times 10^{23}) \times (1.6 \times 10^{-19})$ C $\approx 96,500$ coulombs

The charge carried by 1 mole of electrons is so large and important that it is given a special name:

$$1 \text{ faraday} = 96,500 \text{ C}, \qquad (2.12)$$

after Faraday the scientist who laid the foundations of electrochemistry.

All this makes it clear that by using the power of chemical reactions we can expect to produce quantities of electricity very much greater then those we were dealing with in Chapter 1. The charge carried by an electrophorus (Section 1.2 of Chapter 1) may be only a very small fraction of a coulomb! In the rest of this section we'll be talking about a very simple form of voltaic cell, based on Volta's original invention. It uses electrodes of Zinc and Copper, instead of Zinc and Silver; and the electrodes dip into a solution formed from water and a little "hydrochloric acid" (so we use a weak – 'dilute' – acid instead of common salt). All you need know about the acid is that the molecules of HCl are then almost completely 'dissociated' into the ions $H^+$ and $Cl^-$. The whole cell is indicated in the Figure below, along with a lamp connected between the two electrodes:
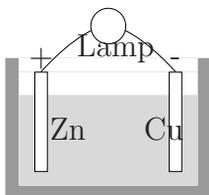
**Figure 20.**
A simple voltaic cell
– glass jar shown in dark grey,
– zinc anode (+ve) on left
– copper cathode (−ve) on right
– electrolyte shown light grey

How does it work? The lamp, shown as an open circle at the top, lights up as soon as it is connected to the two electrodes, because the anode, which is short of electrons, can take them in from the cathode, which is electron-rich. As the electrons pass through the lamp, on their way to the anode, which needs them, they heat a fine wire **filament** inside the glass bulb until it becomes 'white-hot'. And that's what makes the light.

The anode is short of electrons because the zinc surface is attacked by the acid: the surface atoms lose electrons according to $Zn \rightarrow Zn^{2+} + 2e^-$, so the metal becomes electrically positive and the electrons can move into solution, being picked up by anything that can carry them.

Back in the solution, there are $H^+$ ions which are repelled by the positive anode (helping them on their way, through the solution, towards the cathode. But there are also $Cl^-$ ions which are *attracted* by the anode and stay close to it until they meet up with the zinc posi-

tive ions who are looking for positive partners. When that happens there's another reaction: $Zn^{2+} + 2Cl^- \rightarrow ZnCl_2$, a neutral molecule called zinc chloride.

But what about the $H^+$ ions, journeying towards the electron-rich cathode? When they finally get there they find a negative surface charge of electrons waiting for them. The reaction that follows is $H^+ + e^- \rightarrow H$, which gives a neutral hydrogen atom – a new-born species called "nascent hydrogen". It doesn't stay that way for long because two of them get together immediately, forming the neutral hydrogen molecule $H_2$ which comes out of solution in the form of gas bubbles.

That's the end of the story – a sort of love story on an atomic scale, where everyone finds the right partner in the end! The driving force is chemistry: neutral species come together, have a short life as ions, generate electricity, and are partly used up in the process – some of the zinc electrode being 'eaten away' by the acid and converted into the salt, zinc chloride.

*Some other kinds of battery*

Nowadays there are many kinds of battery, often made from dry materials (like the small 'dry batteries' used in calculators and portable radios) so they can be carried around without spilling. Some (like the ones for driving an electric motor, to start a car or to operate an electric drill – which both need a lot of power) may also be *re-*

*chargeable.* A rechargeable battery, or **accumulator**, needs an *input* of current to charge it, by changing its chemical condition, but then releases its charge when connected to an external circuit – providing power as it discharges. Many small cars and trucks are now driven in this way, which is very much cleaner than by burning oil and making the air in our cities almost unbreathable.

But, basically, all such batteries work on the principles explained in this Section. It remains only to ask *how much* power electricity can give us – and where does it all go to when it's done its job?

## 2.5 Power from electric currents

In Section 1.2, you got used to the idea of **electric potential**, which has since been used many times. If a point charge $q$ is taken from Point A to Point B, by doing work $W(A \rightarrow B)$, then the change of potential energy is as given in (1.8):

$$V_B - V_A = W(A \rightarrow B) = q\phi_B - q\phi_A. \qquad (2.13)$$

(Don't get mixed up with the letter 'V', used here for potential energy, with its use to mean 'voltage' in talking about electric circuits! or used for the Volt as the unit of potential. The meaning will always be clear if you really think about what you're doing.)

If the charge is positive and $\phi_B > \phi_A$, then work $W(A \to B)$ is also positive – meaning the particle has gone from Point A to a Point B with higher potential energy: you have done work *on the charge* to get it there. But if $\phi_B < \phi_A$, then the work term is negative and the charge has gone to a position of lower potential energy: negative work done on the charge means that the charge was able to do work on its surroundings. Remember also that positive charge always flows in the direction of *decreasing* potential i.e. *'downhill'*.

Now look at a typical circuit, Fig.17, where a current $I$ is flowing from the top terminal of the battery round the circuit in a clockwise direction (as long as the switch is closed), finally reaching the negative terminal. The electric potential $\phi$ is going down at each circuit element: when unit positive charge goes into the resistance $R$ at terminal A, and comes out at terminal B, its potential energy will change by $\phi_B - \phi_A$ and this will be a voltage *drop*.

Since $I$ is the amount of charge passing through the resistance *per second*, the work done per unit time by the moving charge will be $V \times I$. But we know from (2.8) that $V = I \times R$ and it therefore follows that

> Work done per unit time when current
> $I$ flows through a resistance $R$ is
>   Work done per second $= RI^2$.
> This is called the **power** consumed.

$$(2.14)$$

The unit of power is called the **watt**, as in Book 4 – where we were dealing with mechanical systems. Here, where electricity is involved, the watt is most naturally defined as 1 unit of current (1 amp) flowing between points at unit potential difference (1 volt).

You can do the same with other circuit elements, each being defined in terms of the voltage drop between its terminals and the current flowing through it. Thus, if the lamp bulb in Fig.17 is rated as 100 W and is operating on an electricity supply at 230 V, the current through the filament will be given by $I \times 230 = 100$ which gives $I = 0.43$ A.

**But what does electric power really mean?**

In Chapter 1 we talked about machines that generated static electricity at potentials of thousands, or millions, of volts. Surely they give us a lot of power? But in fact they don't. Power depends on both the potential difference between two terminals (the 'potential drop') and the electric charge that's dropping through that PD

*in every second.* Let's make a comparison between the Wimshurst machine (Section 1.5), supposing it gives an output at 100,000 volts; and the humble battery in your electric torch, which gives an output at little more than 1 volt from each single cell.

---

## Example 7. Power from a Wimshurst machine

The Wimshurst machine can easily produce 100,000 volts, as you can see from the sparks if gives – which may be several centimetres long! The electricity is taken off the disks and stored in condensers (Leyden jars), so we need to know how much charge they can hold. It's hard to calculate the capacity $C$ of a jar, so let's approximate it by using instead a parallel-plate condenser (Figure 11), with two square plates 20 cm $\times$ 20 cm separated by 0.5 cm.

From equation (1.21), the capacity would be $C = \epsilon_0 \times$ (Area divided by separation); but that's for plates separated only by *air*, with $\epsilon_0 \approx 9 \times 10^{-12}$ F m$^{-1}$.

Now a Leyden jar has *glass* between the metal conductors, not air – which would 'break down' and allow sparks to pass and discharge the condenser completely. With glass insulation between the plates, the constant $\epsilon_0$ must be multiplied by about 5; and the formula for $C$ then gives

Capacity $C \approx 5 \times (9 \times 10^{-12}$ (F/m) $\times$ 400 cm$^2$/ 0.5 cm $\approx 3.6 \times 10^{-10}$ F.

(Check the calculation for yourself, noting that F is the usual abbreviation for the Farad and that the constant $\epsilon_{\text{glass}} \approx 5\epsilon_0$ is an experimentally measured constant.)

Suppose now that the Leyden jars are charged to a potential of 100,000 V and have a joint capacity of the order $2C \approx 10^{-9}$ F.

How much charge will they hold? From the formula (1.22) we get $Q = CV \approx 10^{-9} \times 10^5$ C – where we remember the unit F V$^{-1}$ is the *Coulomb* (C).

In summary, our 100,000 V generator has given us only about $10^{-4}$ C of electric charge!

---

Now let's compare this result with what you can get from a voltaic cell. Go back to the circuit in Fig.17, using a single cell of the kind studied in Section 2.4 as the battery. It produces a PD of roughly 1 V and we'll suppose it gives an electric current of about 0.1 A, passing through the resistance (R) and the lamp (L): the two things together, connected in series, would then have a combined resistance of $R = (1.0/0.1)$ V A$^{-1}$ = $10\,\Omega$. We can then argue as follows:

---

**Example 8.** Power from a single Voltaic cell

When unit charge (1 C) falls through a PD of 1 V the energy it loses is 1 J, the unit of **energy**. The joule was first defined in Book 4, in terms of 'mechanical' quantities like force and work; but here it's expressed, still in the same terms, using Coulomb's law for the force between electric charges. The **power**, as noted above, is the *rate* at which work is being done. When $I$ units of charge fall through voltage $V$ *per second* the current is $I$ and the rate of working is $I\,V$ J s$^{-1}$ (joules per second). With the values taken above, the power used is $P = IV = 0.1 \times 1.0$ J s$^{-1}$.

---

The conclusion from Examples 7 and 8 is clear: the Voltaic cell is sending round the circuit in Fig.17, in every second, a charge of 0.1 coulomb – *an amount a thousand times greater than that used by the high voltage generator in producing a single spark!* So don't be misled by sparks alone! What counts, in using electricity, is the useful *work* it can do for you; and that usually depends on the *wattage* you can get from a device, which is a *product* of amps and volts. And remember that the old 'mechanical' unit of working was 1 'horse power' – the work per second that a very large horse was expected to give you! In the units we use nowadays that's about 746 W.

**And where does the electricity go to when it's done its job?**

This is the last question we have to think about. Remember that in Physics one of the great principles is that of **energy conservation** (look back at Section 2.4 in Book 4). When energy seems to be lost it doesn't just disappear: it simply changes from one form to another. In the Wimshurst machine, the *mechanical work* you have to do in turning the handle reappears as electric charge, which is stored at a certain potential – giving you *electrical* energy. In the Voltaic cell, the *chemical* energy, released in the reactions taking place, reappears as the electric charge going round the circuit. Sometimes the energy produced is put to good use, like lighting a

lamp; sometimes it doesn't do anything very useful, like making a spark; but **it is never lost**. Very often it reappears as **heat**, which you learnt about in Book 5.

Look again at Example 8, the Voltaic cell driving electric current ($I$) through circuit elements with a combined resistance $R$. The calculation of the power used led to an expression for the power $P$, which gave the work done per second as $10^{-1}$ W. This reminds us that the Joule was first used in Book 5 (Section 3.1) as a unit of *thermal* energy, being related to the older unit, the **calorie**, by 1 calorie $\approx$ 4.2 J. As Joules are used up in an electric circuit it is therefore not surprising that they are being transformed into **heat**. Some of the power used may be spent usefully (e.g. in cooking or in heating the room with an 'electric fire' – which is a long high-resistance wire wound round a bar of heat-proof insulating material); but part of it is always wasted, because even an almost 'ideal wire' has a small but non-zero resistance.

What form does the thermal energy take? As you'll remember from Book 5, matter of all kinds is composed of atoms and molecules which can store energy in the form of *vibrations* or other forms of internal motion. If you dig even deeper the atoms and their nuclei are themselves built up out of the **fundamental particles**, some of which (e.g. electrons and protons) carry their own electric charges – which keep their special values, whatever may happen to the matter they compose. You

can think of many other questions to ask – why only those very special values of charges and masses? – and why are these particular groups of interacting particles only allowed to take certain special 'quantized' energy values? There's no end to it. But these are problems for present-day Physics, for which you'll need a few more books!

# Chapter 3

# Magnetism

## 3.1 Another bit of history

Roughly 2500 years ago both the Greeks and the Chinese discovered that certain heavy stones would attract pieces of iron; and that this property could be passed on to the iron itself by stroking the stone with it. The property is what we now call **'magnetism'** and when it is given to a piece of iron the iron is said to be 'magnetized' – it becomes a **'magnet'**.

If a small iron needle is magnetized in this way the attracting power seems to be strongest near the two ends – at points which were called 'poles'. And if the needle hangs from a thread, attached to its midpoint, it

has the mysterious property of always pointing in the same direction. The end which points to the North is called the 'north-seeking pole', while the other one is the 'south-seeking pole': they are the 'N-pole' and 'S-pole', respectively.

As you will know, this device – with the magnetic needle free to turn about a vertical axis – is a **compass**. And compasses have been in use for many hundreds of years as a means of finding your direction (e.g. in navigation, when your ship is in mid-ocean with no land in sight). It seems that the Earth itself is an enormous magnet and that all magnets, wherever they are, line up in the same way.

Other strange things are that, if a magnetized wire is chopped into small pieces every piece becomes a separate magnet, with its poles keeping the same direction; and that similar poles (both N or both S) *repel* each other, while opposite poles *attract*, always with an 'inverse-square' law – like that between electric charges.

People went on making and using magnets for centuries, without understanding what magnetism was, or where it came from. So magnetism developed as a purely *experimental* science, until the idea of a **field** as a means of transmitting force between different objects came to be accepted. At first, the N and S poles were thought of as some kind of 'magnetic charge' (like the + and − charges in electrostatics). This idea was supported

by the fact that the force law was very similar to that for electric charges: like poles *repelled* each other, with a force inversely proportional to the square of the distance between them, while unlike poles *attracted* each other in just the same way.

But then, just less than two hundred years ago, Oersted (in Denmark) noticed that a current-carrying wire was moved by the presence of a magnet. In other words, the moving electric charges in a conducting wire must be 'pushed' by a **magnetic field**: they experienced a force *perpendicular* to the direction of the current and this pushed the wire to one side, in the direction of the magnetic field. There followed many other discoveries, by Ampère, Faraday, and others which proved beyond all doubt that electriciy and magnetism were closely related. The strength of a magnetic field could thus be *measured* by the force it exerted on a moving charge: if a charge $q$ moves with velocity $\mathsf{v}$, through a magnetic field $\mathsf{B}$, the force acting on it will be

$$\mathsf{F}_{\mathrm{mag}} = q\mathsf{v} \times \mathsf{B}. \tag{3.1}$$

The cross ($\times$) here means the **vector product** of $q\mathsf{v}$ and the field vector $\mathsf{B}$, which gives the third vector $\mathsf{F}$; so the field will have physical dimensions

$$[opB] = [\text{force}][\text{charge}]^{-1}[\text{velocity}]^{-1}.$$

The three vectors are related as shown in Figure 21.
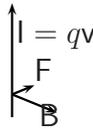
**Figure 21.** Force on a current-carrying wire
Applied field $B$ is taken perpendicular to wire.
Resultant force $F$ is perpendicular to both $I$, $B$.

You met vector products in Section 5.4 of Book 2: $F$ is perpendicular to the plane containing $v$ and $B$. Here $I = qv$ is the current vector in a wire and all three vectors are mutually perpendicular. The magnitude of $B$ will be measureable in units of Newton Amp$^{-1}$. (Notice that various names are used for the electric and magnetic fields: $E$ is often called the "electric field intensity" or "electric field strength", while $B$ is also called the "magnetic flux density" – for reasons that will be clear later. Usually, however, we'll just call them the "electric field" and the "magnetic field".)

With the discoveries of the early 17th century, everything seemed to be falling into place: but there was one big mystery – nobody ever found a single magnetic pole! They always came in *pairs*, one N and one S, no matter how small you made the bits of magnetized iron wire when you chopped it into pieces!

## 3.2 The magnetic field

How was it possible to make sense of all the experimental observations on magnets and the way they behaved? It seemed clear that the interactions between magnets were due to *fields* between them, and that the fields must be similar to those we met in electrostatics (as they lead to force laws of the same mathematical form).

So in this Section let's start looking for **field equations** which now include the magnetic field $\mathsf{B}$, but are not very different from the ones we've already met. Most of electrostatics comes out of Gauss's law: that the flux of the electric field $\mathsf{E}$ out of any closed surface is proportional to the total amount of charge it encloses. So let's take the basic property of the magnetic field to be that the flux of $\mathsf{B}$ out of any closed surface enclosing any distribution of 'magnetic material' must be *zero*. This simply says that any 'poles' could only be found in N-S pairs and would give only equal but opposite flux contributions. As in Chapter 1, for the electric field, there will be two mathematical forms of this statement: one 'integral' and one 'differential'. The first form applies to any finite volume, bounded by a closed surface. This means simply that all the points considered lie *inside* the surface. For a magnetic field the corresponding statement reads:

> The total flux of magnetic field intensity $\mathsf{B}$ out of any **closed surface** is equal to zero.

$$(3.2)$$

The second form applies when the volume 'shrinks' to zero: at any *point* in space

$$\operatorname{div} \mathsf{B} = 0, \qquad (3.3)$$

where the 'divergence' of a vector field is defined in (1.27). The zero arises, in each case, because there is no 'magnetic charge' to *create* the field $\mathsf{B}$. Instead, it is found experimentally that $\mathsf{B}$ arises whenever there are *electric currents*. The key experiment, first made by Ampère, demonstrates the production of a magnetic field by an electric current (see Figure 22).
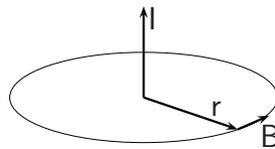


**Figure 22.** Field $\mathsf{B}$ near current-carrying wire. Vectors $\mathsf{r}, \mathsf{B}$ lie in plane perpendicular to $\mathsf{I}$. Lines of force are circles with wire as axis – like the one shown.

In Section 2.3 we used the idea of 'lines of force' in visualizing the *electric* field: every line of force followed an electric field vector from a it ended. The points on any line could be found by experiment (noting the force acting on a small 'test charge' at each point), or by calculation (provided the charges that produce the field, and their positions, are known). But the *magnetic* field is easier: a simple experiment lets you actually 'see' the field!

You stick the current-carrying wire in Fig.22 through a stiff white card and sprinkle the card with 'iron filings' (the small bits of metal that fall off when you file an iron bar to make it smooth); and then you switch on the current and gently shake the card. What will you see? Each small piece of iron is a tiny magnet, which lines up in the field like a compass needle in the field of the Earth. The magnets set themselves 'head-to-tail' (N pole to S pole) to show you a line of force for the field B! You'll see a whole 'family' of lines marked out in this way.

The vertical wire in Fig.22 gives you *circles*, all with their centres on the wire. They have no ends, showing there are no free poles (div B = 0). The lines of force always *circulate* around the wire in horizontal planes, to which the wire is normal; and the field vector B – wherever you look – always points along a tangent to one of the circles.

To describe all this mathematically we'll go along the same lines as we did for the *electric* field in Chapter 1, noting carefully the differences. The *magnetic* field is produced by electric currents, described by the **current density**, j, instead of the scalar density of charge $\rho$. And the main property of B is that it *circulates* around the currents that produce the field (think of the current-carrying wire and the field lines that go round it). So in general we need to think about the **circulation** of B around any closed path encircling the currents that produce it. This should make you remember the idea of a **path integral**, touched on at the end of Section 1.7.

Let's make a 'shot in the dark', replacing (1.17) by a similar sort of principle that we suspect might describe the *magnetic* field. Using similar language we'll suppose that

---

The circulation of magnetic field intensity B, round any **closed circuit**, is equal to the total current passing through the circuit (i.e. the flux), divided by $\bar{\mu}_0$.

---

$$(3.4)$$

Here "passing through the circuit" means passing through the area which has the circuit as its boundary – no matter what shape it has; and the proportionality constant

($\bar{\mu}_0$) has been chosen simply to make the equation look more like (1.17). (The 'bar' has been put on the Greek letter 'mu' only because 'mu' without the bar is the standard name for something else – called the 'magnetic permeability'.)

Now we'll look at this principle more closely to see if it works.

## 3.3   Circulation and the flux of current

In Section 1.2 we were able to define the electric potential $\phi(x, y, z)$ at any point in space $P(x, y, z)$ by studying the work done $W(A \to B)$ in taking unit charge along any path leading from A to B. This was our first example of a path integral. For a tiny path element ds the work done was $\mathsf{E} \cdot \mathsf{ds}$, while for the whole path we were able to write

$$W(A \to B) = \int \mathsf{E} \cdot \mathsf{ds}.$$

For a *closed* path, the integral was zero (potential change zero on returning to the starting point) and this allowed

us to define another basic property of the electric field:

Circulation of $\mathsf{E}$ round any closed path $= \oint \mathsf{E} \cdot \mathsf{ds} = 0$.

$$(3.5)$$

The circulation of $\mathsf{B}$ in (3.4) is defined in exactly the same way, but it is *not* zero when there is a flow of electric current across the area bounded by the closed path: instead

Circulation of $\mathsf{B}$ round any closed path $= \oint \mathsf{B} \cdot \mathsf{ds}$

$$= \frac{1}{\mu_0} \times \text{current through area bounded by path.}$$

$$(3.6)$$

And since the total current $I$ across any surface is given as an integral of the normal component of the *flux vector* $\mathsf{j}$, as in (2.7), this leads to a very neat mathematical form of (3.6):

$$\oint \mathsf{B} \cdot \mathsf{ds} = \frac{1}{\mu_0} \int_S \mathsf{j} \cdot \mathsf{dS}. \qquad (3.7)$$

The relationship (3.6) between field $\mathsf{B}$ and the density vector $\mathsf{j}$ of the electric currents which produce it corresponds closely to that between $\mathsf{E}$ and the density $\rho$ of the charge distribution which produces it, given in (1.20). This is called 'Ampère's law, after its discoverer; and it is just as important as Gauss's law in electrostatics. The Examples that follow will show how it can be

used. First, however, we'd better be sure we get the **units** right!

**A note on units** (If you get confused by these things, jump straight to the end of the small print – you can come back to it when you've worked through some of the Examples

How do we *measure* B? If you go back to (3.1), where the magnetic field was first mentioned, you'll be able to find its **physical dimensions**. (The idea has been used quite a lot in Book 4, and in Section 1.1 of the present book; so go there if you need to.)

In an equation $F = q\,v\,B$ ("force equals charge times velocity times $B$"), each factor has certain 'dimensions': thus, velocity is distance per unit time and we write $[v] = L/T = L\,T^{-1}$ and say "the dimensions of velocity are distance over time or $L\,T^{-1}$"; similarly acceleration $a$ is an increase of velocity per unit time and we write $[a] = L\,T^{-2}$; and force $f$ is mass times acceleration, so we write $[f] = M\,L\,T^{-2}$. The basic quantities M,L,T are measured in the **primary units**, kilogrammes, metres, and seconds (kg, m, s): these allow us to assign dimensions to all 'mechanical' quantities and to fix the units in which to measure them. Thus, force will be measured in units of $\mathrm{kg\,m\,s^{-2}}$.

Now we are including electric and magnetic effects and have already introduced one more basic quantity, the electric charge Q, calling the unit of charge the "Coulomb" C. We also found units of electric field intensity and electric potential in Section 1.3, namely 1 N $\mathrm{C^{-1}}$ ("1 newton per coulomb") and 1 N m $\mathrm{C^{-1}}$ ("1 newton-metre per coulomb", respectively. But what about the 'magnetic' quantities?

Returning to (3.1), the quantities on the two sides of the equation must have the same dimensions (otherwise they wouldn't stay

equal if the units were changed) and this means [F]= [q][v][B]=
$Q\,L\,T^{-1}$ [B]. But we know [F]= $M\,L\,T^{-2}$. So if we suppose the
dimensions of B are [B]=$M^r L^s T^t Q^u$ then the only way to make
the exponents match is to choose $1 + r = 1$, $1 + s = 1$, $-1 + t = -2$, $1 + u = 0$. In other words, the dimensions of the magnetic
field are [B]=$M^0\,L^0\,T^{-1}\,Q^{-1}$

Instead of using always the primary units it is often more conve-
nient to use **derived units**, like force or energy; any combination
of units will do – as long as the dimensions are equivalent. For
example, from (3.1) it follows also that the dimensions of B are
those of force/(charge×velocity) and this can be measured in units
of newtons/(coulombs×m/s). Since 1 C = 1 amp-second this can
be re-arranged (do it!) to give

  1 unit of field intensity (B) = 1 newton per amp-metre = 1 T.

This is the unit most commonly used: it is called the "**tesla**" in
honour of Nikola Tesla (1856-1941) and is the last new unit you're
going to need for a long time!

# 3.4   Magnetic field near a current-carrying wire

Let's start from Ampère's first experiment (see Figure
22), in which the electric current $I$ is simply the surface
integral of the flux density j over a cross-section of the
wire. (The idea used already in Section 2.3, Figure 15)

Here the path integral of the field vector B is most nat-
urally made by simply following B round a circle of, say,

radius $r$ with its centre on the wire. The field is directed along the tangent to the circle and will have the same magnitude at all points 'by symmetry' (all points looking just the same as you go round the circle); and the length of the path is $2\pi r$, as you'll recall from Book 2.

The field magnitude $B$ at distance $r$ from the wire therefore gives a path integral $B \times 2\pi r$ (supposing (3.6) is correct!) and this should be equated to $I/\bar{\mu}_0$. Thus, the magnetic field is directed along the tangent to any circle, at distance $r$ from the wire and in a plane normal to the wire; and it has magnitude

$$B = \left( \frac{1}{4\pi\bar{\mu}_0} \right) \frac{2I}{r} \qquad (3.8)$$

which you can compare with that for the *electric* field $E$ at distance $r$ from a charged wire, with charge/unit length $\lambda$ as given in (1.19).

There's another way of getting this result, which we'll need later. It depends on knowing the field d$B$ due to a little bit d$x$ of the current-carrying wire, pointing along the x-axis; and then summing the contributions from all such bits in the whole wire by doing an **integration** (see Book 3). The formula we need will give the contribution to the field $B$, at P, that comes from the current in the element d$x$ of the wire. This tiny contribution is given,

in vector language, by

$$dB = \left(\frac{1}{4\pi\bar{\mu}_0}\right)\frac{I d\mathsf{x} \times \mathsf{r}}{r^3} = \left(\frac{1}{4\pi\bar{\mu}_0}\right)\frac{I dx}{r^2}\sin\theta \; \hat{\mathsf{n}}. \quad (3.9)$$

Here the field dB is normal to the plane containing the wire and the vector r, which points from the element dx to the field point P – as in Figure 23.

Note that r is not *perpendicular* to the wire, in general, but makes an angle $\theta$ with it; and the 'cross' in (3.9) means the **vector product** of dx and r – a new vector pointing along the unit vector n̂ normal to the plane. (Look back at Section 5.4 of Book 2 if you need to be reminded!)
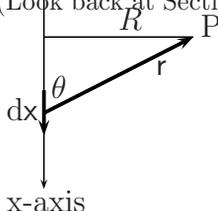


**Figure 23.** Alternative calculation of the field
Vertical wire (x-axis) carries current $I$.
Field B at Point P is perpendicular to plane and is
obtained by integration over $x$ in interval $(-\infty, +\infty)$.

The result (3.9) is called the **Biot-Savart law**, after the people who first used it. You needn't try to prove it: we'll just show how it leads to the right answer (3.8) for an infinitely long straight wire. The details are a bit difficult, but they're given in the next Example – to be kept for later.

---

**Example 1.** Field near a current-carrying wire

Look at Figure 23. Here the field contribution dB is normal to the plane containing the wire and the vector r, which points from the wire to P – as indicated in the Figure. Remember that r is not *perpendicular* to the wire, but makes an angle $\theta$ with it. However, $x$ can be measured from the foot of the perpendicular from P to the wire and by changing its value we can move the element dx along the whole wire.

To do this we first note that all contributions dB are in the same direction (along the upward normal n̂, so we only need their *magnitudes* d$B$ and can write for their infinite sum

$$B = \frac{I}{4\pi\bar{\mu}_0} \int_{x\to-\infty}^{x\to+\infty} \frac{\sin\theta}{r^2} dx$$

Also, from Fig.23, $r^2 = R^2 + x^2$. Since $R$ is a constant, for all elements, and $\sin\theta = x/r$ the integral becomes

$$B = \frac{I}{4\pi\bar{\mu}_0} \int_{x\to-\infty}^{x\to+\infty} \frac{x}{(R^2 + x^2)^{3/2}} dx.$$

This is quite a tricky thing to evaluate, but when it's done the result agrees perfectly with (3.8) where $r$ (not $R$) now means the perpendicular distance of P from the wire. The point of doing all this is to show that every little bit of the current path makes 'its own' contribution to the field at any point P, according to the Biot-Savart law (3.9). This law is very general and applies to *any* current-carrying wire, not just a straight one (and the current needn't even be inside a wire!).

109

---

As a second example, we'll take a wire in the form of a circular ring and calculate the field at any point on the axis (see Figure 24).
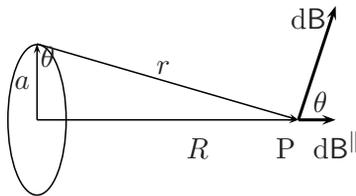


**Figure 24.** Field at Point P on axis of a current-carrying ring of radius $a$: $B^{\parallel}$ at Point P is perpendicular to plane of the ring and points along the axis.

---

**Example 2.** Field near a current-carrying ring

For an element of the ring of length $ds$, carrying the current $I$, the Biot-Savart law shows it will give a field contribution $dB$ at Point P as shown in the Figure. This will be at some angle $\theta$ to the axis, but will have components $dB^{\perp}$ and $dB^{\parallel}$ perpendicular and parallel to the axis.

All elements on the circular ring will give contributions of the same magnitude 'by symmetry' but, on summing to get the resultant field, the components perpendicular to the axis will cancel out to give zero. The components pointing along the axis, however, will all have the same magnitude $dB \cos\theta = dB(a/r)$ and may simply

110

be added together to get the total field $B$.

The axial field due to an element d$s$ will be

$$\mathrm{d}B = \frac{I}{4\pi\bar{\mu}_0} \times \frac{a}{(a^2 + x^2)^{3/2}} \times \mathrm{d}s$$

– from (3.9), with a factor $\cos\theta = a/\sqrt{a^2 + x^2}$ included to give the axial component. Since all elements d$s$ on the ring give identical contributions, summation gives only a factor $2\pi a$.

The end result for the axial field at distance $x$ from the ring centre will thus be $B = \frac{1}{2\bar{\mu}_0} \times \frac{Ia^2}{(a^2+x^2)^{3/2}}$. Later we shall find how useful this important result can be.

---

Example 2 shows that the B-field at any point on the axis of a circular current-carrying ring will point along the axis and have a magnitude

$$B = \frac{1}{2\bar{\mu}_0} \times \frac{Ia^2}{(a^2 + x^2)^{3/2}}, \qquad (3.10)$$

$x$ being the distance of the point from the ring centre. Later we shall find how useful this important result can be.

# 3.5   Magnetic field along the axis of a long solenoid

If you wind a long wire round the outside of a tube of insulating material, as indicated below in Figure 25, you

get a **solenoid**.
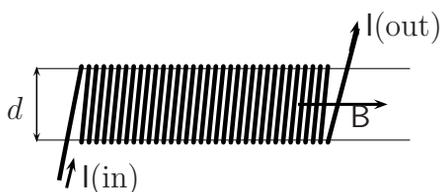


**Figure 25.** A long solenoid
$N$ turns of wire, tightly wound
round insulating tube of diameter $d$.
Length of winding $= L$.

Note that the wire itself must also be covered with in-
sulating material (e.g. silk or enamel) so that current
can't jump from one turn of wire to the next – causing
a 'short circuit'. Solenoids are very important, as we'll
see later. In the next Example we calculate the B-field
produced inside when a current passes through the wire.

In order to use Ampère's law let's look at what's going
on *inside* the solenoid, taking away the wires to get the
picture below.
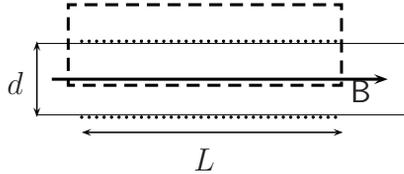
**Figure 26** Solenoid: cross-section to show path for calculating field B along the axis

$N$ turns of wire, tightly wound round insulating tube of diameter $d$. Length of winding $= L$.

---

**Example 3.** Field inside a long solenoid

Inside the solenoid the lines of force must all go along parallel to the axis (the field has zero divergence) and we can calculate it easily by choosing a path like the one shown as a broken line. This goes along the axis of the tube, where the field is B say, then perpendicular to the axis where B has zero component along the path, and then back *outside* the solenoid – where the field is very small and can be neglected. So the complete path integral will have the value $B \times L$ and this gives you the left side of (3.6). This path encircles $N$ wires, all carrying the same current $I$, so the flux integral on the right will be simply $N \times I/\bar{\mu}_0$. When you equate the two things you find at once $B = \frac{nI}{\bar{\mu}_0}$ $(n = N/L)$.

---

The conclusion from Example 3 is that the B-field inside a long solenoid (length $L$) points along the axis and has magnitude

$$B = \frac{nI}{\bar{\mu}_0} \quad (n = N/L), \qquad (3.11)$$

where $n = N/L$ is the number of turns per unit length. Solenoids are very useful things: simply by switching on an electric current you can generate a strong magnetic field inside the tube, so the solenoid becomes a magnet until you switch the current off and the field disappears. A solenoid of this kind becomes an **electromagnet**. If the air inside it is replaced by an iron 'core' (e.g. a bundle of 'soft' iron wires), which has a different value of the constant $\bar{\mu}_0$, you can make a powerful magnet. Electromagnets are widely used in industry. Remember something similar happened in electrostatics: by changing the material between the plates of a condenser from air (with permittivity $\epsilon \approx \epsilon_0$) to glass, with a much higher $\epsilon$, the capacity could be increased enormously. The iron core in an electromagnet has a similar effect in concentrating the field and making it much stronger. And if the core is properly treated it may even keep its magnetization when the current is switched off – becoming a **'permanent magnet'**.

## 3.6 Force between two current-carrying wires

So far we've been using both coulombs (C) and amps (A) in talking about electric charges, but the amp was defined as the unit of current which would carry unit charge when it flowed for unit time (i.e. 1 A s = 1 C). In practice, now that we know about magnetic effects, it's more convenient to express coulombs in terms of amps: 1 C = 1 A-s; and current is most easily measured in terms of its 'mechanical' effects. For this purpose we can combine the result (3.8), from Example 1, with the basic force law (3.1) which gives the force acting on a moving charge $q$ in a magnetic field of intensity B.

Let's think of a current-carrying wire as a cylinder of conducting material containing bits of flowing charge: call them 'charge particles', it doesn't matter what they are, so we'll just suppose they all carry charge $q$ and that there are $N$ of them per unit volume of the wire. In time $t$ the amount of charge flowing across the area $A$ (cross-section of the wire) will be $Q = Nq(A \times vt)$ ($v$ being the magnitude of the velocity vector). But $Nqv$ is the magnitude of the current density vector j and $Q$ can thus be re-written as $Q = j \times t \times A$; or, since $jA$ is the total current in the wire, $Q = It$.

Now (3.1) tells us that the total transverse force on the

charge in volume $V = LA$ (length×area) of the wire will be $\mathsf{F} = LA(\mathsf{j} \times \mathsf{B}) = L\mathsf{I} \times \mathsf{B}$. In other words,

> The force $F$ per unit length on an ideal wire carrying current $I$, in a transverse field $B$, is given by $F = IB$. The vectors $\mathsf{F}, \mathsf{I}$, and $\mathsf{B}$ are all mutually perpendicular.

(3.12)

Finally, we come back to the *definition* of the unit of current, the amp, and calculate the force between two current-carrying wires, as in the next Example.

---

**Example 4.** Force between current-carrying wires

We imagine two ideal parallel wires, carrying curents $I_1$ and $I_2$. The force per unit length on Wire 2, due to current $I_1$ in Wire 1, will be – using (3.8) and (3.10) –

$$F_2 = I_2 B_2 = I_2 \frac{1}{4\pi\bar{\mu}_0} \times \frac{2I_1}{r_{12}},$$

where $r_{12}$ is the distance between the parallel wires, Wire 1 and Wire 2. And if you do things the other way round, to find the force on Wire 1, due to current $I_2$ in Wire 2, you get the same result (check it!) – the forces are equal but may be in opposite directions (think about the directions of the three vectors in the vector product! will the wires attract or repel each other?).

From the result in Example 4 we could say that two straight parallel wires, each carrying a current of 1 Amp, would attract or repel each other with a force of 1 newton per metre. This would mean taking the value of $2/(4\pi\bar{\mu}_0)$ in Example 4 to be 1 N $A^{-2}$. In the SI system, however, the standard definition of the amp is chosen to make $1/(4\pi\bar{\mu}_0)$ equal to $1\times10^{-7}$ N $A^{-2}$. This means simply that all 'magnetic' units are fixed by international convention through the choice

$$\frac{\mu_0}{4\pi} = \frac{1}{4\pi\bar{\mu}_0} = 10^{-7}\text{NA}^{-2}. \qquad (3.13)$$

Here the constant $\bar{\mu}_0$, used so far, has been eliminated by using the internationally accepted $\mu_0 = 1/\bar{\mu}_0$, which is called the **magnetic permeability of free space**. (Note, however, that sometimes we'll still use $\bar{\mu}_0$ when it helps to make things clearer.)

Why did it take so long for SI units to be agreed on by (almost!) all the world's scientists? In this case it was because so many mistakes were made in talking about electric and magnetic effects – with two systems growing up independently, each with its own concepts and sets of units. When people realized that electricity and magnetism were closely related it was not easy to change old names and habits – so we still use some of the ideas and names from long ago, and perhaps they will stay with

us for another generation or two.

## 3.7 How magnetism differs from electrostatics

Before ending this Chapter we should compare the laws that give $\mathsf{E}$ (in terms of charges) and $\mathsf{B}$ (in terms of currents). First think of a static charge $q$ at the origin of coordinates: the field it generates at point $\mathrm{P}(x, y, z)$ with position vector $\mathsf{r}$ is given by

$$\mathsf{E} = \frac{1}{4\pi\epsilon_0}\frac{q\mathsf{r}}{r^3}$$

– written in vector notation. (Note that this is not an inverse-*cube* law, because $\mathsf{r} = r\hat{\mathsf{r}}$, $\hat{\mathsf{r}}$ being a *unit* vector in the direction of $\mathsf{r}$.) The *magnitude* of the field is thus given by the old familiar inverse-square law

$$E = |\mathsf{E}| = \frac{1}{4\pi\epsilon_0}\frac{q}{r^2} \qquad (3.14)$$

– as written long ago in (1.4).

In the next Example we'll get the $\mathsf{B}$-field produced by a circulating current, in which charge $q$ moves around the origin in a circular orbit of radius $a$, giving a current $\mathsf{I} = q\mathsf{v}$ in each element of path d$\mathsf{s}$.

**Example 5.** Field due to charge moving in a circular orbit

We ask for the magnetic field B at a point P on the axis of the orbit and at a distance $x$ from its centre (as in Fig.24). In this case, using (3.10), the magnitude of the axial field due to the orbital current will be

$$B = \frac{1}{2\bar{\mu}_0} \times \frac{Ia^2}{(a^2 + x^2)^{3/2}}$$

(with the notation used in Fig.24) but for an orbit of 'atomic' dimensions $a$ can safely be neglected in the denominator. On using $r$ instead of $x$ for the axial distance and noting that $\pi a^2 = A$, the *area* of the orbit, this result for the field magnitude $B$ becomes (check it!)

$$B = \frac{1}{4\pi\bar{\mu}_0} \times \frac{2IA}{r^3},$$

which should be compared with (3.14).

---

The final result from Example 5 is that the magnetic field due to a point charge moving in a tiny circular orbit has magnitude

$$B = \frac{1}{4\pi\bar{\mu}_0} \times \frac{2IA}{r^3}, \qquad (3.15)$$

where $I$ is the current due to the moving charge. The field is directed along the axis of the orbit and (3.15) gives its magnitude at a distance $r$ from it. Note that the

magnetic field $B$ falls off as the *cube* of the distance from the object producing it (the tiny circulating current) and the proportionality factor is $1/4\pi\bar{\mu}_0$ instead of $1/4\pi\epsilon_0$ for the field $E$ produced by an electric charge.

To interpret this result, look back at Section 1.3 where we studied the field created by an **electric dipole** consisting of two charges ($+$ and $-$) a short distance apart. The electric field at points along the axis of the dipole (the z-axis in Fig 4) was found to have z-component

$$E_z = \frac{d}{4\pi\epsilon_0} \times \frac{3z^2 - r^2}{r^5}$$

where $d$ is the dipole 'strength' or **electric moment**, the magnitude of each charge times their distance apart. For any point P on the z-axis, where $x = y = 0, z = r$, the axial electric field is therefore

$$E = \frac{1}{4\pi\epsilon_0} \times \frac{2d}{r^3}$$

and shows an inverse-cube dependence on distance from the dipole.

We can make (3.15) look similar by writing it as

$$B = \frac{1}{4\pi\bar{\mu}_0} \times \frac{2d_{\mathrm{mag}}}{r^3}, \tag{3.16}$$

where

$$d_{\mathrm{mag}} = IA \tag{3.17}$$

120

is the **magnetic dipole moment** arising from electric current $I$ circulating around the perimeter of an area $A$.

The key quantity in producing a magnetic field is in fact a magnetic dipole; and the components of the field B produced follow from the same formulas we found in Section 1.3, provided the constant $\epsilon_0$ for the electric permittivity of free space is changed to its magnetic counterpart $\bar{\mu}_0$ and the electric dipole moment $d$ is changed to the *magnetic* moment $d_{\mathrm{mag}}$. Current $I$, circulating round the boundary of an area $A$, is equivalent to a **magnetic shell** in which *imaginary* 'North poles' and 'South poles' lie close together on the two sides of the surface and every dipole makes its own contribution to the field B at any point P. You'll never need to use this picture but it helps us to understand why separate 'magnetic charges', N and S, are never found; and in Chapter 4 we'll see how all our ideas about electric and magnetic fields can be unified in a single 'framework' called **electromagnetism**.

**Energy density in a magnetic field**

You may be wondering why, in comparing electric and magnetic effects, nothing has been said about *energy density* – to compare it with that given in equation () for an *electric field*. The reason is only that this is not something you can do easily, even for a simple example. But it can be done quite generally; and here we'll just note that there is a very similar expression for the energy

density in a magnetic field $\mathsf{B}$. Instead of () the density is found to be

$$u_{\text{mag}} = \tfrac{1}{2}\bar{\mu}_0 B^2, \tag{3.18}$$

where again $\bar{\mu}_0$, the *reciprocal* of the conventional magnetic permeability of free space, is used to show up the perfect similarity of the two expressions.

# Chapter 4

# Getting it all together: Maxwell's equations

## 4.1 More on differential operators and functions

The idea of a differential **operator** was first used in Book 3, where the rate of change of some quantity $y = f(x)$, depending on a single variable $x$, was written as $y'(x) = f'(x) = dy/dx$. The process of finding $f'(x)$ from $f(x)$ was regarded as applying to it the 'operator' $\mathsf{D} = d/dx$: written in full,

$$f'(x) = \mathsf{D}f(x) = \frac{d}{dx}f(x) \qquad (4.1)$$

and the symbol $\mathsf{D}$ stood for the whole process of increasing the variable $x$ by a tiny amount $\mathrm{d}x$, finding the corresponding increase in $f(x)$, dividing it by $\mathrm{d}x$, and getting the *limit* of this ratio as $\mathrm{d}x \to 0$ –which is called the **first derivative** $f'(x)$ of the function $f(x)$. So $\mathsf{D}$ stands for quite a lot!

Later in Book 3 the idea was extended to functions of *several* independent variables, such as the coordinates $x, y, z$ of a point P in 3-dimensional space, and *partial* differential operators were defined in a similar way but with only one variable at a time being changed. Thus

$$\mathsf{D}_x f(x, y, z) = \frac{\partial}{\partial x} f(x, y, z) \qquad (4.2)$$

would indicate the rate of change of $f(x, y, z)$ with respect to variation of $x$ alone, $y$ and $z$ being left unchanged. In Section 6.1 (of Book 3) many properties of partial differential operators were found; and they appeared again and again in Book 4 (– in Physics) and Book 5 (– in Chemistry). Here, in Book 10, it's clear that they give us the language we need in talking about electric and magnetic fields. For example, the electric potential $\phi = \phi(x, y, z)$ is a function of position in 3-space, depending on the coordinates $x, y, z$ of a **field point** P, and the components of the fields $\mathsf{E}$ and $\mathsf{B}$ are all functions of $x, y, z$. Moreover, when the fields vary with *time*, all such quantities will be functions of *four*

variables:

$$E_x = E_x(x, y, z, t), \quad E_y = E_y(x, y, z, t), \ \dots . \quad (4.3)$$

and similarly for $B_x$, $B_y$, $B_z$. (Remember that a *quantity* and the *function* which describes it mathematically are often given the same symbol (e.g. $E_x$) – something you should be used to by now!)

For a field with three components (apart from the time, which need not be shown when the field it not varying) there are three partial differential operators

$$D_x = \frac{\partial}{\partial x}, \quad D_y = \frac{\partial}{\partial y}, \quad D_z = \frac{\partial}{\partial z} \quad (4.4)$$

and, as we saw in Section 6.1 of Book 3, these 'first-order' operators can be combined to give 'higher-order' operators. To do this, they are allowed to work on the function one after the other. Thus, for any function $f(x, y)$ we can define second-order operators by, for example,

$$D_x D_y f = \frac{\partial}{\partial x} \frac{\partial f}{\partial y}.$$

This defines the **product** of the two operators $\mathsf{D}_x$ and $\mathsf{D}_y$ and has the same meaning for *any* function $f(x, y)$; so the $f$ may be left out and we simply write

$$D_x D_y = \frac{\partial}{\partial x} \frac{\partial}{\partial y}. \quad (4.5)$$

– where the operators are left 'hungry' for something to work on. For all 'well-behaved' functions, the order in which $D_x$ and $D_y$ are applied is unimportant: $D_y D_x = D_x D_y$ and we say the operators **commute**.

In Section 1.2, the **gradient** of the electric potential $\phi = \phi(x, y, z)$ was defined as the vector with three components, given in (1.11): the vector is the electric field $E$ and the *operation* of finding its components is described in (1.12) by writing $E = -\text{grad } \phi$. The gradient is a **vector operator**, producing a vector field from the scalar function $\phi$ (which depends only on *position* of the field point).

And in Section 1.6 the **divergence** of a vector field was defined according to (1.25), namely

$$\text{div } E = \left( \frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z} \right). \qquad (4.6)$$

This operation works the other way round, producing a *scalar* from a vector field.

The physical meaning of the divergence rests on the way it was first defined, in (1.24), in terms of the **flux** of $E$ out of a volume element $dV = dxdydz$. For a *finite* volume $V$, it was shown in Section 1.7 that the total flux of $E$ out of $V$ could be expressed as the *integral* of its divergence over the whole volume. And this led to an 'integral form' of Gauss's law that the total outward flux was proportional to the amount of electric charge within

the volume. Towards the end of Section 1.7 we obtained Gauss's Theorem (1.29), that the integral of the flux over a closed *surface* was equal to the *volume* integral of the divergence, over the volume enclosed by the surface. And we guessed that there might be a similar theorem relating an integral of 'something' along a closed path (i.e. a *line* integral) to a *surface* integral of 'something else' over a surface bounded by the path. In Section 1.7 the 'something' was $\mathsf{E} \cdot \mathsf{ds}$, the work done in moving unit positive charge along an element of path $\mathsf{ds}$ in field $\mathsf{E}$. Now, we want to find the 'something else': it will turn out to be another vector, defining a new vector field related to $\mathsf{E}$ and called curl $\mathsf{E}$ – the 'curl' of $\mathsf{E}$.

**The 'curl' of a vector field**

In Chapter 1 we studied the work done in moving a particle through a field of force: the work done by the force $\mathsf{F}$ in displacing the particle through a small element of path $\mathsf{ds}$ is

$$w = \mathsf{F} \cdot \mathsf{ds} = F_x \mathrm{d}x + F_y \mathrm{d}y + F_z \mathrm{d}z$$

and the total work done in moving the particle from Point A to Point B is the line integral $\int_s \mathsf{F} \cdot \mathsf{ds}$ – obtained by summing the work done in every elementary step $\mathsf{ds}$ in going from A to B. For a *closed* curve, where B is the same as A (you've come back to the starting point) the work done, which gives the total change of potential energy (a function of position only), must be zero. The

result is then written

$$\text{Circulation of } \mathsf{F} \text{ round closed path } = \oint \mathsf{F} \cdot \mathsf{ds}. \quad (4.7)$$

Now let's think of the circulation round the *perimeter* of a small, rectangular area of surface, d$S$, lying in the xy-plane, as in Figure 27.
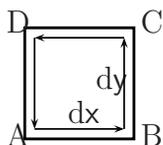


**Figure 27.** For calculation of the curl. Rectangular surface element is ABCD, with corners at A(x,y), B(x+dx,y), C(x+dx,y+dy), D(x,y+dy). Surface area is d$S$ = d$x$d$y$. Arrows show direction of circulation.

The following Example shows how to calculate the circulation around a surface element d$S$ = d$x$d$y$.

---

**Example 1.** Circulation of a field-vector $\mathsf{F}$ around d$S$

The field components $F_x$ and $F_y$ on the four sides of d$S$ in Fig.27 are $F_x^{AB}$, $F_y^{AB}$, for AB, $F_x^{BC}$, $F_y^{BC}$, etc. and the circulation round the surface element will be

$$F_x^{AB}\mathrm{d}x + F_y^{BC}\mathrm{d}y - F_x^{CD}\mathrm{d}x - F_y^{DA}\mathrm{d}y,$$

where the minus signs arise for steps in the negative directions. This is

$$(F_y^{BC} - F_y^{DA})\mathrm{d}y - (F_x^{CD} - F_x^{AB})\mathrm{d}x.$$

But the y-component of $\mathsf{F}$ increases by $(\partial F_y/\partial x)\mathrm{d}x$ when $x$ increases by $\mathrm{d}x$ (going from side AD to side BC); so

$$(F_y^{BC} - F_y^{DA}) \approx (\partial F_y/\partial x)\mathrm{d}x.$$

And the x-component of $\mathsf{F}$ increases by $(\partial F_x/\partial y)\mathrm{d}y$ when $y$ increases by $\mathrm{d}y$ (going from side AB to side DC); so

$$(F_x^{CD} - F_x^{AB}) \approx (\partial F_x/\partial y)\mathrm{d}y.$$

When these two terms are put into the expression for the total circulation the result is

$$\text{Circulation of } \mathsf{F} \text{ round ABCD } = \left(\frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y}\right)\mathrm{d}x\mathrm{d}y.$$

The expression obtained in this way becomes exact in the usual limit where $\mathrm{d}x$ and $\mathrm{d}y$ tend to zero and all derivatives are evaluated at the point $\mathrm{A}(x,y)$. )

---

For a surface element in the xy-plane the circulation corresponds to a right-handed 'screw' motion around the upward normal to the plane – in this case the z-axis. When the x-, y-, and z-axes are defined in this way the coordinate system is said to be right-handed. The circulation is then counted positive and represented by a *vector* pointing along the normal. Example 1 shows that

this vector has magnitude

$$\text{Circulation of } \mathsf{F} \text{ round ABCD} = \left( \frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y} \right) \mathrm{d}x\mathrm{d}y.$$
(4.8)

Now let's define the more general vector, called curl $\mathsf{F}$, with components

$$
\begin{aligned}
(\text{curl } \mathsf{F})_x &= \left( \frac{\partial F_z}{\partial y} - \frac{\partial F_y}{\partial z} \right), \\
(\text{curl } \mathsf{F})_y &= \left( \frac{\partial F_x}{\partial z} - \frac{\partial F_z}{\partial x} \right), \\
(\text{curl } \mathsf{F})_z &= \left( \frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y} \right).
\end{aligned}
$$
(4.9)

As you can see, the z-component is the one we found first in (4.8), where it multiplied the element of area $\mathrm{d}x\mathrm{d}y$ in the xy-plane. The others can be written down easily by noting that the labels $x, y, z$ in the first term follow the 'cyclic order' $x\,y\,z \;\rightarrow\; y\,z\,x, \rightarrow\; z\,x\,y$ as you go from one component to the next. Thus, the first component in (4.9) starts with

$$(\text{curl } \mathsf{F})_\mathbf{x} = \mathsf{D}_\mathbf{y} F_\mathbf{z} \,...$$

and you can fill in the rest using the rule for cyclic order (close the book and do it yourself!)

Now (4.8) is the circulation round an element of area $\mathrm{d}S = \mathrm{d}x\mathrm{d}y$ in the xy-plane: as a *vector* element of area it

is $d\mathsf{S} = dS\mathsf{n}$, where the unit vector $\mathsf{n}$ points along the z-axis. In this case the vector $d\mathsf{S}$ has only one component; and so has $\operatorname{curl}\mathsf{F}$ – in both cases the z-component. And the elementary circulation in (4.8) is now seen to be a **scalar product** of the two (single-component) vectors, $\operatorname{curl}\mathsf{F}$ and $d\mathsf{S}$.

In words,

> The circulation of a vector field $\mathsf{F}$ round the perimeter of an element of area $d\mathsf{S}$ is equal to the scalar product $(\operatorname{curl}\mathsf{F}) \cdot d\mathsf{S}$.

This is a remarkable result because, although derived for a very special case – a surface element lying flat in the xy-plane – it must be true *even if the whole picture is rotated*, because we know (Section 6.1 of Book 2) that a scalar product is **invariant** against such a change.

An even more amazing result follows when we *integrate* the last one! The circulation round the perimeter of the element $dS$ is $\oint \mathsf{F} \cdot d\mathsf{s}$, where $d\mathsf{s}$ is an element of *path* as you go round the curve and back to the starting point; and for this tiny circuit we've shown that

$$\oint \mathsf{F} \cdot d\mathsf{s} = \operatorname{curl}\mathsf{F} \cdot d\mathsf{S}.$$

(Note that $d\mathsf{s}$ is a *path* element, while $d\mathsf{S}$ is a *surface* element!)

When we want to deal with a whole surface, of any size and shape, and the closed circuit that forms its edge, we can go ahead much as we did in proving Gauss's Theorem in Section 1.7 (where we divided the whole volume into elements $dV = dxdydz$). But now we divide the whole *surface* into tiny rectangular pieces $dS$, where every piece may share one or more sides with other pieces as in the Figure below.
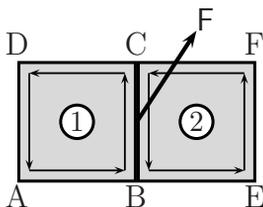


**Figure 28.** For Stokes's Theorem.
Adjacent surface elements, ABCD and BEFC
Side BC common to both elements (1 and 2)
Circulation direction shown by arrows.
Field F on BC same for both elements.

Note that on any shared side, like BC, there is a common field vector F but the contributions to the circulation round Element 1 and Element 2 *from that side* are of opposite sign: they cancel in taking the sum and only elements at the *outer edge* of the 'double-element' contribute to the surface integral.

Any surface can be divided in this way into a great number of tiny rectangular elements, provided you take enough of them, and all the 'internal' elements that share sides with their neighbours can be ignored because their circulations cancel: only elements at the *boundary* of the surface need be taken into account – and these give a 'jimpy' line from which the actual line integral can be approximated as accurately as you wish. The final conclusion then is that

$$\oint_s \mathsf{F} \cdot \mathrm{d}\mathsf{s} = \int_S \mathrm{curl}\,\mathsf{F} \cdot \mathrm{d}\mathsf{S}. \qquad (4.10)$$

This result is called Stokes's Theorem, which was mentioned at the end of Section 1.7 after getting Gauss's Theorem. If you compare Figure 28 with Figure 13 you'll see how similar these two important results are: the **flux** of a vector field out of any volume, bounded by a closed surface, is equal to the volume integral of the **divergence** of the field; and the **circulation** of a vector field round any closed curve is equal to the surface integral of the **curl** of the field, taken over any surface bounded by the closed curve. We haven't really *proved* either of these theorems – that's stuff for professional mathematicians – but we can say we've understood what they're about and what they allow us to do. As you begin to use them in talking about electric and magnetic fields you'll find how far we can go with little more than the

basic ideas developed in the first three Chapters of this Book.

In fact, armed with the three operations symbolized by 'grad', 'div', and 'curl', we're all set to come to grips with Maxwell's famous equations – which underpin so much of the science and technology of the present day.

## 4.2   A reminder of the equations for E and B

Let's start by writing down, in vector language, the things we know already. The field E is created by electric charge and the **outward flux** of E from any closed volume containing a total amount of charge $Q$ is given by Gauss's law (1.17):

$$\text{Flux of } \mathsf{E} = \int_S \mathsf{E} \cdot \mathsf{dS} = Q/\epsilon_0. \qquad (4.11)$$

When the charge is spread out continuously, with density $\rho$ per unit volume, $Q = \int_V \rho dV$ – a volume integral. Gauss's law then becomes

$$\int_S \mathsf{E} \cdot \mathsf{dS} = (1/\epsilon_0) \int_V \rho dV. \qquad (4.12)$$

This is the 'integral form' of the law, applying to volumes of any size and shape. But after introducing the

idea of **divergence** in Section 1.6, which refers to the field at a single point in space, we were able to express the flux integral on the left as a volume integral of div $\mathsf{E}$, obtaining (1.26), which is equivalent to (1.24). The 'differential form' of Gauss's law applies at all *points* in space:

$$\operatorname{div} \mathsf{E} = \rho/\epsilon_0. \qquad (4.13)$$

This simple-looking partial differential equation can be solved to give the electric field throughout the whole of empty space, created by any distribution of static charges.

In Chapter 1 (Section 1.2) the idea of **electric potential** was introduced by thinking about the work done when unit positive charge moves from one point to another: the work done $(w)$ defines the *change* of potential $(\phi)$ as a *path integral* of the force acting on the charge; and for a *closed* path the change was zero. Since the force acting is represented by the vector $q\mathsf{E}$, this means that

$$\oint \mathsf{E} \cdot \mathsf{ds} = 0$$

and this is the reason why it is *possible* to define a potential function $\phi(x, y, z)$, depending only on position in space. And now that we know about the 'curl' operator this integral statement can be expressed in a differential form much more like (4.13). Thus, from Stokes's

theorem (4.10) it follows that

$$\int_S \operatorname{curl} \mathsf{E} \cdot \mathrm{d}\mathsf{S} = 0 \qquad (4.14)$$

for any surface element, however small, on which the field is evaluated. This means

$$\operatorname{curl} \mathsf{E} = 0 \qquad (4.15)$$

at all points in empty space.

The two equations (4.13) and (4.15) – that the divergence of $\mathsf{E}$ at any point is equal to the density of charge at the point, divided by $\epsilon_0$, while the curl of $\mathsf{E}$ is zero – are the first two of **Maxwell's equations** for stationary fields in empty space. There are two more, which determine the *magnetic* field created by electric *currents*. They have been given in Chapter 3, in the 'integral form' which involves the *flux* of $\mathsf{B}$ out of the volume bounded by any closed surface (see (3.2)) and the *circulation* of $\mathsf{B}$ round any closed path through which there is a flux of electric current (see (3.7) – which we called Ampère's law.

The zero flux of div $\mathsf{B}$ led at once to the differential form div $\mathsf{B} = 0$, which means there are no single 'magnetic poles' (only N-S pairs). But a differential form of (3.7) was not given. Now that the curl operator has been defined, Stokes's theorem in (4.3) can be used to replace

the line integral $\oint_s \mathsf{B} \cdot d\mathsf{s}$ by a *surface* integral of curl $\mathsf{B}$; and the rest follows on shrinking the surface element to zero, as in the derivation of (4.15). This results (check it all for yourself!) in the differential form, which applies at any point in space.

$$\text{curl } \mathsf{B} = \mathsf{j}/\bar{\mu}_0, \qquad (4.16)$$

where $\bar{\mu}_0$ is the analogue of $\epsilon_0$ in electrostatics.

Let's collect all the key equations, which contain all of electrostatics and magnetostatics, together in one nice little box. On introducing $\mu_0 = 1/\bar{\mu}_0$, which is the **magnetic permeability of free space**, we obtain the standard forms:

$$\text{(a) div } \mathsf{E} = \rho/\epsilon_0 \quad \text{(c) div } \mathsf{B} = 0$$

$$\text{(b) curl } \mathsf{E} = 0 \quad \text{(d) curl } \mathsf{B} = \mu_0 \mathsf{j}$$

$$(4.17)$$

That's all we need in order to calculate the $\mathsf{E}$ and $\mathsf{B}$ fields arising from any given distribution of charges and steady currents – throughout the whole of empty space!

They contain Gauss's law, Ampère's law, and everything we've done so far, including the applications we've made to particular systems like charged spheres and plates. current-carrying wires, and so on.

By thinking only of *empty* space we must leave out the charges and currents *themselves* by supposing they are non-zero only within very small volumes (e.g. points and lines). This may seem strange; but the 'empty space' we want to talk about is normally *vast* compared with the objects that create the fields.

Something looks odd, however; because the fields $E$ and $B$ seem to be completely independent. Equations (4.17a) and (4.17b) allow us to find $E$ at all points in space, given the distribution of electric charge, *without knowing anything about* $B$; while $B$ can be found from (4.17c) and (4.17d), for any given set of currents, without reference to $E$. In fact, many simple experiments tell us that when the fields are *changing* all this is no longer true. In particular, if the flux of $B$ through the area bounded by a closed circuit changes with time, then the circulation of $E$ round the circuit is *not zero* – and this means that curl $E \neq 0$, contradicting (4.17b).

## 4.3 And if E and B vary in time? – Maxwell's complete equations

The first thing we need to change is (4.17b) – because Faraday discovered long ago that changing the magnetic field passing through a closed loop of wire produced an *electric* field in the wire. The field set up in this way moved the charges in the wire, making an electric current; and if the magnetic field was changed fast enough the current set up was big enough to be measured. Faraday's experiment showed that the circulation of E round the closed circuit was *not* always zero, but depended on the rate of change of any *magnetic flux* passing through the area bounded by the wire.

We have to be careful here, because Faraday used *wires* in his work – whereas now we're talking about fields E and B in empty space, and our 'circuits' may be simply *imaginary* paths. Remember Section 1.3, where the current $I$ was the current *density* j integrated over the cross-section of the wire; and instead of E we used $\mathcal{E}$ for the **electromotive force (emf)**, measured in volts, that pushed it along the wire. To make things easier we'll usually express Faraday's results in terms of the fields themselves.

Finding how to improve (4.17b), to take account of a changing magnetic field, will be our "Step 1". But then equation (4.17d) will look odd, because it doesn't show

any dependence of the B-field on how the *electric* field is changing. To put that right we'll have to take a second step – "Step 2". We'll take these steps in the next two Examples

---

**Example 2.** Faraday's experiment on changing magnetic flux

Let's assume that the E produced in Faraday's experiment is directly proportional to the rate of change of the B-flux through the circuit. The new rule will then be

$$\left( \begin{array}{c} \text{circulation of } \mathsf{E} \\ \text{round circuit} \end{array} \right) = \left( \begin{array}{c} \text{rate of increase of } \mathsf{B} - \text{flux} \\ \text{through area bounded by circuit} \end{array} \right),$$

– with a $\pm$ sign to add, as usual, depending on directions of the vectors. Experiment shows that the minus sign applies: the directions of E, ds, and B are then related by the usual 'corkscrew rule'. How can this statement be put in mathematical form?

The E-field due to changing flux is what propels charge along the wire and gives an **induced current**. The above result can be written in symbols, using the definitions of circulation and flux in (4.7) and (1.28), as

$$\oint \mathsf{E} \cdot \mathsf{ds} = -\frac{\mathrm{d}}{\mathrm{d}t} \int_S \mathsf{B} \cdot \mathsf{dS},$$

which is the 'integral form' of the relationship, the term on the right replacing the zero in (4.15). The corresponding 'differential form' is obtained by looking at a single infinitesimal element dS and going from the integrals to their point values:

$$\mathrm{curl}\, \mathsf{E} = -\frac{\partial}{\partial t}\mathsf{B}$$

and this will be the equation that replaces (4.17b) when the field $\mathsf{B}$ is varying with time. Note that the *partial* derivative $(\partial \mathsf{B}/\partial t)$ has been used here, because $\mathsf{B}$ is a function of four variables $x, y, z, t$, but at any particular point in space only the time is changing – the other variables being held fixed.

The conclusion from Example 2 is that, when the magnetic field is changing with time, the equation curl $\mathsf{E} = 0$ in (4.17) should be replaced by

$$\text{curl } \mathsf{E} = -\frac{\partial}{\partial t}\mathsf{B}. \qquad (4.18)$$

You can make a simple experiment, like the one Faraday did, to show that an electric current can be set up in this way. Take a loop of copper wire, or even better wind it into a solenoid (see Fig.26) with several loops, and connect the ends to a small lamp (like one from an electric torch). If you can get a 'bar magnet', plunge it quickly in and out of the solenoid – and the lamp will glow as the current heats its filament. Here, instead of using an electric current to produce a magnetic field (as in Section 4.4) you're doing the reverse, using the field from a rapidly moving permanent magnet to produce an electric field. This is the principle used in the **dynamo**, which runs the lights on your bicycle, or the enormous **generators** found at the power stations in our cities: both do the same job, using *mechanical* energy to produce *electrical* energy.

Much more about this later! – but first let's take the next step, by asking how the circulation of B round a closed circuit might depend on the rate of change of the E-flux through it. One of Maxwell's great discoveries was to find this missing term.

---

**Example 3.** The missing term in curl B = 0.

To show why there must be a missing term in (4.17d) it's enough to look at one simple experiment – the charging of a parallel-plate condenser, like the one shown in Fig.11, where current $I$ is flowing onto the top plate and building up a charge $Q$. Current $I$ flows *away* from the bottom plate, building up *negative* charge on it.

The situation is as shown in Figure 29 below:
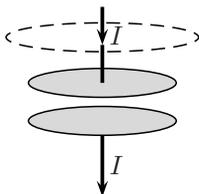


**Figure 29.** For Maxwell's new term.
Top circuit (broken line) encircles the wire,
carrying current $I$ to top plate of condenser.
Upper plate holds positive charge, current $I$
leaves from lower (negative) plate.

When the path used in getting the line integral of E is the circle indicated with a broken line in Fig.29, the integral form of

142

Ampère's law in (3.7) becomes

$$\oint \mathsf{B} \cdot d\mathsf{s} = \frac{1}{\mu_0} \int_S \mathsf{j} \cdot d\mathsf{S} = \frac{1}{\mu_0} I.$$

Now in terms of the current density j the total current $I$ in the wire is simply the *flux* of j over the cross-section of the wire. This gives a first form (call it **A**) of the current

**A:** $I =$ flux of j over cross-section of wire.

But what if the plane surface, whose edge is indicated by the broken line in Fig.29, is moved down *between* the plates of the condenser? The current passing through it is then zero – there's only a *field* between the plates; and the field B can't depend on where the *imaginary* surface is drawn! The missing term in (4.17d), on the right-hand side of the equation, must clearly 'take over' at points where the current density j vanishes – i.e.where there is a 'gap' in the conducting path.

To find its form we try to express the total current $I$ in terms of the *field* in the gap between the plates, noting that $I$ gives the rate at which charge $Q$ builds up on the top plate in Fig.29. Thus, $I = dQ/dt$; and we know from (1.20) that the field outside a plate of area $A$ with surface charge density $\sigma = Q/A$ is given by $E = \sigma/\epsilon_0$. It follows that there must be an alternative form (call it **B**) of the current in the wire, namely

$$\textbf{B:} \ I = \epsilon_0 \frac{d}{dt}(EA) = \epsilon_0 \frac{d}{dt} \left( \begin{array}{c} \text{flux of E across surface} \\ \text{between condenser plates} \end{array} \right).$$

Forms **A** and **B** are simply alternatives: the first one applies when a '*real*' current passes through the surface you have in mind; and

143

the second one when there is only a field, the current being zero. Both give $I$ as the flux of a current density across a surface; in the First form, the current j is in a conducting medium (the wire) and the surface is the cross-section of the wire; but in the Second form the 'real' current is replaced by

$$j_D = \epsilon_0 \frac{\partial}{\partial t} \mathsf{E}$$

and the surface is *imaginary*, being in the 'gap' between the condenser plates.

Maxwell called this term the **displacement current**: it replaces the real current when charge is *displaced*, building up on the conducting plates and producing the *field* betwen them.

To be on the safe side, we should allow for both alternatives in Example 3 by using $\mathsf{J} = \mathsf{j} + \mathsf{j}_D$ in place of j. This would mean that (4.17d) should be replaced by the more general equation

$$\text{curl } \mathsf{B} = \frac{1}{\bar{\mu}_0} \mathsf{J},$$

where $\mathsf{J}$ now includes the displacement current:

$$\mathsf{J} = \mathsf{j} + \left( \epsilon_0 \frac{\partial \mathsf{E}}{\partial t} \right). \tag{4.19}$$

We can now collect the equations, as Maxwell left them 130 years ago, for fields in empty space. They provide a solid foundation for the science of Electromagnetism and

have been verified by countless applications to systems of many kinds. Their final forms are:

$$\text{(a) div } \mathsf{E} = \rho/\epsilon_0 \qquad \text{(c) div } \mathsf{B} = 0$$

$$\text{(b) curl } \mathsf{E} = -\partial\mathsf{B}/\partial t \quad \text{(d) curl } \mathsf{B} = \mu_0\mathsf{J}$$

$$(4.20)$$

– with the total current $\mathsf{J}$ defined in (4.19).

In the remaining chapters of this book you'll begin to understand just how important these results are. Like Newton's laws, Maxwell's equations are here to stay: they'll still be serving the needs of humanity, long after today's politicians, dictators, generals and warlords are dead and forgotten.

Before going to Chapter 5 we'll put on record one important result that comes directly from the free-space form (4.20) of Maxwell's equations. Now that time variation of the fields is included we can ask if equations like (3.7), for the energy density in an electrostatic field, and (3.18), for the density in a magnetic field, remain

true. The proof is too difficult to give here, but the result looks very simple: the energy stored in the electromagnetic field can be calculated from the single **energy density formula**

$$u = u_{\text{elec}} + u_{\text{mag}} = \tfrac{1}{2}\epsilon_0 |\mathsf{E}|^2 + \tfrac{1}{2}\bar{\mu}_0 |\mathsf{B}|^2 \qquad (4.21)$$

– just as if the energies of the $\mathsf{E}$-field and the $\mathsf{B}$-field were simply additive and spread out through all space with an amount $u$ per unit volume. And this is true even when the fields are varying with time!

# Chapter 5

# Dynamos, motors and electric power

## 5.1   Where can we go from the principles? – and how?

You've all seen **dynamos**, like the one on your bike that gives enough electric power to light the lamps, so you can see where you're going at night. They change *mechanical* energy, supplied by pedalling, into *electrical* energy.

You may also have seen **electric motors**, often used in driving the pumps that bring water from deep wells,

or smaller ones (usually running on batteries) in electric drills for boring holes in wood or stone. In this case *electrical* energy is changed into *mechanical* energy: the motor does the exact opposite of what the dynamo does.

Both depend on the same basic principles, discovered experimentally by Faraday and others and neatly packed away in Maxwell's equations. In this chapter we want to show how just one equation

$$\text{curl } \mathsf{E} = -\partial \mathsf{B}/\partial t, \tag{5.1}$$

set up in Chapter 4 as equation (4.17b), can tell us everything we need to know about dynamos and motors of all kinds!

To do this, we'll start from a very simple 'model' in which a loop of wire is rotating in a magnetic field. But first let's remember that (5.1) is the *differential* expression of Faraday's discovery that changing the magnetic flux through a wire loop could set up an electric current in it. The 'integral form' of this discovery was used in Example 2 of Chapter 4: we'll write it out again as

$$\mathcal{E} = \oint \mathsf{E} \cdot d\mathsf{s} = -\frac{d}{dt} \int_S \mathsf{B} \cdot d\mathsf{S}. \tag{5.2}$$

Here $\mathcal{E}$ represents the emf set up in any closed circuit, while the term on the right represents the rate of change of $\mathsf{B}$-flux through that circuit – which in Faraday's experiment really *was* a loop of wire. (Notice that $d/dt$

appears in (5.2), instead of $\partial/\partial t$, because the thing being differentiated is a function of one variable only – the others being 'used up' in doing the integration.)

In many cases equation (5.2) gives the easiest way of getting a result – so it's the one we'll use. Figure 30 shows (very schematically) a rectangular loop turning around its long axis (the broken vertical line). The magnetic field is indicated by the magnetic intensity vector B, directed at right angles to the axis of rotation and (in this drawing) perpendicular to the plane of the loop. The area of the loop will be $S = ab$, where $a, b$ are the lengths of its short and long sides, repectively.
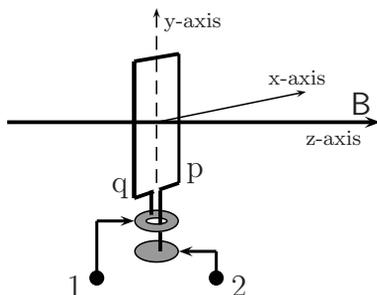


**Figure 30.** Principle of the dynamo

The rotation axis in the Figure is perpendicular to the field and 'brushes', shown as arrows, connect the loop to the terminals (shown as large black dots).

149

The picture shows other things also: the bottom side of the wire loop is cut in the middle and vertical wires go down from each side of the cut, to give a way of getting current into it or out of it. The left-hand wire ends on a metal disk, with a hole in it to let the right-hand wire pass through on its way down to the lower disk. The 'brushes' (often short sticks of carbon, which is a conductor) rub on the two disks and let electricity pass to the **terminals** of the dynamo. The 'beginning' and 'end' of the loop, going round it in the positive (anticlockwise) direction, are marked by the 'p' and 'q'.

(Remember that whenever two conductors are at a different potential and come into contact a current will flow between themi. So if you don't want that to happen the conductors must be **insulated** from each other: that's why the right-hand wire is shown passing through a hole in the top disk, not touching it; and if the disks are mounted on the same axle they must be carefully insulated from it (e.g. by plastic 'sleeves'). But these are small details and are important only if you want to *make* a dynamo.)

---

### Example 1. Simple form of a dynamo

How does the thing shown in Fig.30 actually work? In the position shown, with the plane of the loop perpendicular to B, the flux through it has its maximum value $SB$ (area×normal component

of field), the normal making an angle $\theta = 0°$ with the field B. But for any other value of the angle $\theta$ the flux through the loop. which we'll call $\Phi_{\text{mag}}$, will be $SB \cos \theta$ and the (negative) *rate of change* of flux as the loop rotates – which appears in (5.2), will be (Do the differentiation yourself, going back to Chapter 3 of Book 3 if you need to)

$$\mathcal{E} = -\frac{\mathrm{d}\Phi_{\text{mag}}}{\mathrm{d}t} = -\frac{\mathrm{d}}{\mathrm{d}t} SB \cos \omega t = SB\omega \sin \omega t,$$

where $\omega = \mathrm{d}\theta/\mathrm{d}t$ is the rate of rotation or **angular velocity** of the rotating loop. Note that when $\theta$ ($= \omega t$) is an integer $n \times \pi$ the plane of the loop has its normal along the field direction and the flux through it has its greatest or least (positive or negative) value. In such cases, the equation shows that $\mathcal{E} = 0$.

Notice that the emf generated at any instant $t$ depends also on $\omega$ and that $\mathcal{E}$ now arises from the field induced in an actual wire – rather than in empty space.

As already noted, the circuit in Fig.30 has been cut on its bottom side, to allow charge to pass into and out of the wire loop; but the gap can be as small as we wish and the line integral of E, on going round the circuit, will still give the difference of potential between the two sides of the cut i.e. between the two ends of the wire loop. Also, since E points round the loop in the anticlockwise direction (the 'right-hand screw rule' corresponding to B as shown in the Figure), positive charge will be urged in the direction p→q: so Terminal '1' will be at the *higher* potential – positive relative to Terminal '2'.

---

The conclusion from Example 1 is that the the emf pro-

duced by the rotating loop is

$$\mathcal{E} = SB\omega \sin \omega t. \qquad (5.3)$$

This means that when the magnetic flux through the loop is changing a PD will be induced between its ends: the 'dynamo' is behaving like a battery (Section 2.4) in setting up a PD (or 'voltage') between its two terminals. The PD is generated, in the present case, by the changing magnetic flux instead of by chemical reactions. But the action of the dynamo (or 'generator') is similar to that of a voltaic battery in the sense that it creates a *discontinuity of potential* at the point where it is inserted into a circuit. In Section 2.4 the battery was symbolized by its positive and negative plates, as in Fig.17. If the battery is replaced by a generator the corresponding circuit will be shown as in Figure 31:
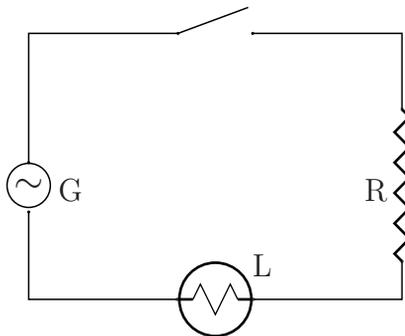


**Figure 31.** Circuit with generator G and lamp L

When the switch in Fig.31 is closed, current goes first through resistance R and then through the lamp filament.

The PD set up by a battery is what pushes electric charge round a circuit. Section 2.4 was where we first called it the **electromotive force**, 'emf' for short, and denoted it by $\mathcal{E}$ to remind us that it's not the same as the field E (Can you explain why?).

The PD set up between the terminals of a dynamo has the same effect and that's why we use the same symbol in (5.2). The final result is given in (5.3) $\mathcal{E} = SB\omega \sin \omega t$. That's the form of the emf produced by any kind of dynamo in which a circuit is rotating with anglar velocity $\omega$, with its axis perpendicular to a magnetic field B. So it's a pretty important result – think of the enormous dynamos, nowadays called **generators** (the word we'll use from now on for all kinds of dynamo) at the power stations which supply electricity to whole cities. Of course, it's a long way from the rotating loop of wire to a real generator, but the principle is the same in each case. Let's look at some of the steps along the way.

---

**Example 2.** From loops to coils

First it's clear that a single loop of wire won't have much flux going through it, even if the field B is big, because the *area* enclosed by

the circuit will be small. But remember the solenoid in Section 3.5, which looks like $N$ circular loops of wire, each one joined to the next so that current can flow from one to the other. In a generator we can put $N$ loops side-by-side to make a **coil** by winding $N$ turns of a continuous wire round a short cylinder of area $A$. The flux through the whole circuit will then be $NA \times B$ and when it rotates in a field $B$, directed along the axis, every loop will produce its own contribution to the emf in the circuit – namely $\mathcal{E}$ given in (5.3). (Why is that? – think of the line integral of $\mathsf{E}$ along the whole circuit.) So, by using a coil of $N$ turns instead of a single loop of wire, the voltage produced across the terminals is multiplied by $N$.

---

Using $V$ for the voltage produced by the generator with an $N$-turn coil rotating in the field, Example 2 tells us that

$$V = V_0 \sin \omega t \qquad (V_0 = NSB\omega). \qquad (5.4)$$

The next important thing to notice is that the output voltage $V$ is *not steady*, as it would be for a battery, but wiggles up and down between limits $+V_0$ and $-V_0$: $V = 0$ when $\omega t = n\pi$ ($n$ being any integer), but then

$$V = V_0 \text{ when } \omega t = (n - \tfrac{1}{2})\pi \ (n \text{ odd}),$$
$$V = -V_0 \text{ when } \omega t = (n + \tfrac{1}{2})\pi \ (n \text{ even}).$$

The variation of voltage, as the coil rotates, is shown in Figure 32:
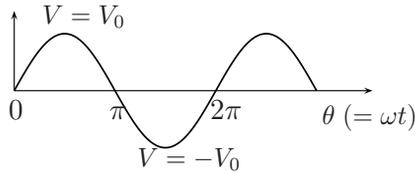
**Figure 32**. Output voltage $V$ from an AC generator.

That's why the symbol $\sim$ (in a circle) has been used to stand for the generator in the circuit diagram of Fig.31 – because it produces an **alternating current** in the circuit to which it is connected. Most of the generators in our big power stations are **AC Generators**; but you can also get **direct current** (which always flows in one direction, from $+$ to $-$ as with a battery) by reversing the direction of current flow every time the coil has turned through $180°$ (i.e. $\pi$ radians) – as you'll see in the next Section. That way, you can get a DC Generator. Both AC and DC generators depend on the same principles, so we'll deal with them both together.

First, however, let's note that the speed of rotation is also important because $V_0$ in (5.4) is directly proportional to the angular velocity $\omega$: if you pedal your bike twice as fast the lights get twice as bright!

## 5.2 AC and DC generators

**A real AC generator.**   In a real generator, the loop of wire in Fig.30 is replaced by a coil consisting of many turns of wire (all insulated from each other) wound around a slab of insulating material to make the **armature**. This is shown below, in Figure 33, lying in a vertical plane with an axle through the middle for making it spin (indicated by the big black dot).
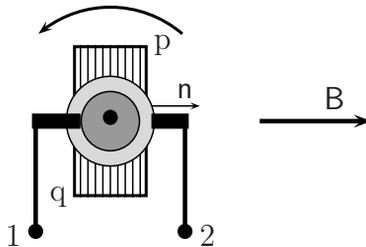


**Figure 33.** Diagram of AC generator (see text)

The magnetic field B is perpendicular to the axle, so when the armature spins the flux through the coil varies rapidly and creates an emf in the wire. The wires from the two ends of the coil (marked 'p' and 'q' as in Fig.30) are led to two metal disks, mounted on the same axle but carefully insulated from it. (The connections to the disks are not shown in Fig.33) A **brush** (shown in black)

rides on each disk. The brushes are connected to the output terminals (just as they were in Fig.30) so that Terminal 1 will be at the voltage corresponding to 'q' and Terminal 2 will be at the same voltage as 'p'.

In Fig.33 the axle of the armature is horizontal (indicated by the large black dot) and is perpendicular to field **B**. It is shown here with the normal **n** pointing along the field and it spins in the direction shown by the curved arrow. Brushes ride on the two separate disks, as in Fig.31, taking current to terminals 1 and 2.

---

**Example 3.** The output from an AC generator

What happens as the armature turns? In the vertical position shown, the normal **n** to the plane of the coil makes an angle $\theta = 0$ with the field **B**. The flux through the coil will be at a maximum, but its rate of change will be zero and the induced emf will thus also be zero by (5.4).

Let's start the rotation from this position. As $\theta$ begins to increase the flux will decrease and the emf $(V)$, given by its *rate* of decrease, will go on increasing until it reaches a maximum when $\theta = \pi/2$, according to (5.4). This corresponds to the first maximum in the graph of Fig.32. As rotation continues the voltage starts to fall, becoming zero at $\theta = \pi$ where the curve in Fig.32 crosses the x-axis. After that, $V$ goes *negative* and reaches its lowest negative value $(V = -V_0)$ when $\theta = 3\pi/2$.

The armature has then reached the position shown in Fig.33, except that 'p' and 'q' *have changed places*: the flux *has changed sign* – the normal **n** then pointing in the opposite direction to the

B-vector. At that point, the induced emf takes its largest *negative value*, corresponding to the first *minimum* in Fig.32. As $\theta = \omega t$ goes on increasing the **sinusoidal** ('sine-wave') form of $V$ continues indefinitely. The voltage passes through a whole **cycle** of values, all in the range $V_0$ to $-V_0$, whenever $\theta$ increases by $2\pi$; and this pattern of values is repeated over and over again as long as the armature goes on turning. The time taken for completing a cycle is clearly $\tau = 2\pi/\omega$, because when $t \rightarrow t + \tau$ the angle $\theta$ $(= \omega t)$ increases by $\omega \times \tau$ $(= 2\pi)$. The time for a complete up-down 'oscillation' is called its **period**; and the number of such oscillations per second (namely $1/\tau$) is called its **frequency**. These numbers are important for any kind of periodic change. You may remember meeting them in Section 6.2 of Book 4, in talking about the vibrations that make musical sounds.

Here, the alternating emf produces an alternating current in any circuit connected to the output terminals; and it goes on for as long as the armature is turning – unlike the battery which runs down when the chemical reaction is completed. Usually, the electricity that comes from the power station has a frequency of around 50 or 60 cycles/second and is supplied at between 100 and 250 volts.

---

The very simple AC generator described above can be improved in many ways. For example, by winding the armature coil on a material such as iron, with a magnetic permeability much bigger than $\mu_0$ (for free space), it becomes much more effective in concentrating the magnetic flux (see Section 3.5). And by increasing the strength of the static field B, not so far considered, it fol-

158

lows from (5.4) that the output emf can also be greatly increased. In most big generators, the field B is produced by an *electromagnet* – not by a small permanent magnet of the kind used in the dynamo on your bike. But these are matters for a whole new field of applied science – "Electrical Engineering". Here we'll have to be content with the basic Physics.

**Real DC generators.** Some way back we said that *direct* current (DC) could also be obtained from a spinning coil in a magnetic field, provided we could change the direction of flow (and that means reversing the voltage) after every half turn. To see why, look again at Fig.32 which shows the emf or 'voltage difference' between the ends of the coil as it rotates. Suppose we start from $\theta = 0$. where the flux through the coil is a maximum. As $\theta$ increases from 0 to $\pi/2$ the voltage across the terminals goes from 0 to $V_0$, as (5.4) shows. At that point the *voltage* is a maximum; but then it falls to zero as $\theta$ goes from $\pi/2$ to $\pi$, where n, the normal to the plane of the coil, points in the direction *opposite* to the field B. The flux and its rate of change then start to grow, but their values carry a *negative* sign. The first negative hump in Fig.32 is at $\theta = 3\pi/2$ and the voltage stays negative until $\theta$ reaches the value $2\pi$. After that, as you can see from the graph, the voltage changes sign whenever $\theta$ increases by an integer times $\pi$ – which means the current will change direction. You may have

wanted DC (e.g. for charging a battery, where getting the direction wrong would ruin the battery), but you can only get AC from the generator in Fig.33 – unless you can think of some way of changing the *output* voltage so that it will always be positive.

In fact, you can do that very easily, without having to start again at the beginning and build a completely different kind of generator. Just look at the next Figure, which shows *almost* the same generator, but with the armature in two different positions. In Fig.34a the armature stands at 10° to the right of the vertical position shown in Fig.31, while in Fig.34b it is 10° to the left.
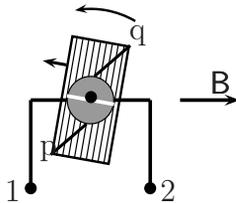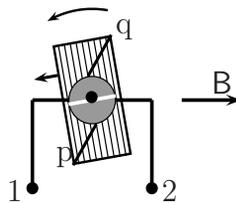


**Figure 34a**                    **Figure 34b**

These positions correspond to points just to the left and just to the right of the first crossing point of the graph in Fig.32 (at $\theta = \pi$), where the armature has rotated to $\theta = 180° - 10°$ and $\theta = 180° + 10°$, respectively.

Now it's clear why we said the picture showed *almost* the same generator. In fact it's not *exactly* the same,

because the brushes in Fig.31 rested on two separate disks, while here there is only *one* disk (called a **commutator**) but it's divided into two halves (insulated from each other at the 'white line').

Such a simple change makes an enormous difference. Instead of leading the wires from the coil ends to different disks (shown in dark-grey and light-grey in Fig.31) they are taken to the two halves of the commutator. The wires coming from the coil ends, at 'p' and 'q', are both shown as bold black lines and carry current to the brushes. In Fig.34a, the end marked with a 'p' represents the *beginning* of the coil (where you start winding the wire anticlockwise around the direction of the normal) and 'q' marks the end where the wire comes out (just as in Fig.30). In that case 'q' will again be at a higher potential than 'p' and you'll see that with the same setup the emf produced points in the direction q→p). The voltage difference $V = V_q - V_p$ during this part of the rotation is shown in the first 'hump' of the curve in Fig.31, where the top of the hump comes for rotation angle $\theta = \pi/2$. Up to that point, $V$ is always positive; but then it begins to go down – and when $\theta$ passes $\pi$ it goes *negative* which we don't want! But the commutator is waiting for it. When $\theta$ is between 0 and $\pi$ (Fig.34a) 'p' is connected to Terminal 1 (as you can see if you follow the conducting path, through the commutator) while 'q' is connected to Terminal 2; so

1 is negative and 2 is positive. But when $\theta$ is just a little bit bigger than $\pi$ **the commutator becomes a switch**: the conducting path from 'p' then goes through the other half of the commutator, below the white line, and connects the lower potential output (from 'p') to Terminal 2 – where it *should* go! In the same way, Terminal 1 is connected to 'q', which is now the higher potential output from the coil. Whenever $V = V_p - V_q$ goes negative, the commutator switches the connections to the two output terminals, so that Terminal 1 is always negative and Terminal 2 positive. What a clever trick! The output voltage never changes sign. Example 3 shows that the output voltage from a DC generator is always in the same direction, never changing sign (see Fig.35). And since the output voltage is always positive the current through any external circuit will flow from Terminal 2 (the higher potential) down to Terminal 1 : it will be Direct Current, even though it goes up and down.
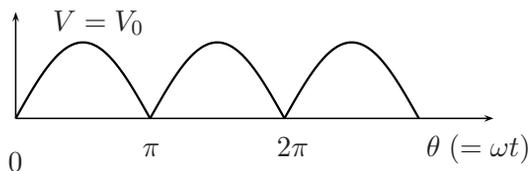


**Figure 35**. Output voltage $V$ from a DC generator.

162

There are many ways of improving a DC generator, smoothing the up-down 'ripple' and increasing the power of the output. But the important thing is that we now know how DC current can be produced without limit – we can throw away the chemical cells, except when we need a small and portable supply of electric power.

## 5.3 Electric motors

Before going on to talk about electric motors, which are used throughout our homes, in doing *mechanical* work on a small scale, and in industry for large scale applications – like driving heavy machinery – we should stop for a minute and look again at anything that may have troubled you. If you want to construct an enormous generator for a power station it's not enough to make a giant armature, by winding thousands of turns of copper wire round a block of iron with an axle on which it can spin; you have to supply enough power to get it spinning in a *uniform magnetic field* B. But how can you get the field when the armature is really big, weighing perhaps half a ton? Small bars of magnetized iron, like the ones in the dynamo on your bike, can't do that kind of job!

Instead, you have to build an **electromagnet** big enough to hold the whole rotating armature; and to supply it with enough current to produce a strong B-field over the

whole volume that contains it. That sounds like a tall order! But you already know about solenoids (Section 3.5); and you know that by winding many turns of wire round an iron 'core' instead of an empty cylinder you can easily produce a very strong magnetic field. (Remember that $\mu_{\text{iron}}$ has a value thousands of times bigger than $\mu_0$ for empty space.) All you have to do now is cut the solenoid into two halves, putting the spinning armature between them as in the Figure below, and supply it with DC current. That way you can get an approximately uniform B-field over the whole armature.

Figure 36 shows a spinning armature between two **field coils** which produce the strong magnetic field: the current supply comes from Terminals 3 and 4. The same picture can be used for both the generator and the electric motor.
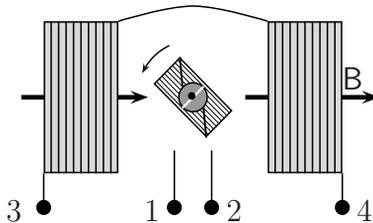


**Figure 36.** Motor with electromagnets

The big difference is that for the generator you must

put *mechanical* energy into it, to make the armature spin and send out *electrical* energy from Terminals 1 and 2; while for the motor you must do the opposite – putting electrical energy into it, through Terminals 1 and 2, to make the armature spin and drive any machinery connected to it (e.g. through pulleys or gear wheels mounted on the same axle). In that way electrical energy will be converted into mechanical energy.

In the Figure the armature is shown in the centre, with field coils on both sides of it. Brushes etc. (not shown) carry current to it from the input terminals 1 and 2. Input to the field coils is from 3 and 4. Running the device as a generator, current goes out, but running it as a motor the current goes in.

Now we're talking about energy you'll be wondering how much energy it takes to keep that strong magnetic field in place. The surprising answer is *"very little"!* That's because the field is not doing any work and is therefore not using any energy. The little bit of energy used is really *wasted* energy, just enough to overcome the resistance offered by the field coils, and in doing so it is turned into heat. We look at the details in the next Example.

---

**Example 4.** The Joule heat

In Section 2.3 the resistance ($R$) of a circuit was defined in terms

of the voltage $(V)$ applied to it and the current $(I)$ forced through it, by the equation

$$V = IR.$$

But $I$ is the amount of charge passing per second; and when unit charge goes from potential $V_1$ to potential $V_2$ the work done on it is the PD $V = V_2 - V_1$. So the work done per second on the flowing charge – the *rate* of working – is given by

$$\mathrm{d}W/\mathrm{d}t = IV = I^2R,$$

where the value of $R$ depends only on the nature of the circuit. The work done gives the charge carriers (electrons, ions, or whatever they may be) **kinetic energy**, which is **dissipated** in collisions with the atoms of the conducting material (in this case the wire); and re-appears as heat – the so-called "Joule heat".

---

What Example 4 tells us is that, when current $I$ passes round any circuit with total resistance $R$, the rate of production of heat is given by

$$\mathrm{d}W/\mathrm{d}t = IV = I^2R. \tag{5.5}$$

The **Joule heat** produced in this way may keep you warm but isn't much good for anything else, like running machinery, and in that sense is 'wasted'. Nevertheless (5.5) is an important equation: it gives you the **wattage**, the rate of electric power production of your generator in watts (joules per second); and if you want to turn all that energy into heat by sending it through a

166

high-resistance material it gives you the heat output of your **radiator**. A common domestic radiator will have a power of a few **kilowatt**, while a light bulb may be marked "100 watt".

On the other hand, the electricity produced at the power station is going to be used in many other ways. In particular we may want to turn it back into *mechanical* energy, which is what the electric motor is supposed to do. How can we put all that electric power to use in doing mechanical work. An electric motor depends on just the same principles as a generator, but you can't always make one simply by running a generator in reverse – putting a voltage across the output terminas and expecting the armature to start turning. To find out more about the 'interconversion' of electrical and mechanical energy let's go back to the 'model' generator in Fig.30.

---

**Example 5.** Energy interconversion

When the wire loop in Fig.30 is rotated in the field B charges are set in motion by the forces arising from the emf induced in the circuit. But, at bottom, a charge moves because a *force* acts on it; and the basic equation needed is the one first used in *defining* the magnetic field, namely (3.1). That's where we should start from and we've already found (Section 3.6) how it leads to an expression for the force per unit length on an ideal wire in a magnetic field. Here we'll use the same 'model' in which a current carrying wire is pictured as a collection of moving charges, $N$ of them per unit

167

length of wire, each carrying charge $q$ and having (on average) velocity $v$ along the wire. The work done per second on the $N$ moving charges, the *rate* of working, must be

$$dW/dt \;\; = \;\; N \oint v \times (\text{force acting}) ds$$

$$= \;\; N \oint vq \times (F/q) ds = Nvq \oint E ds,$$

where $ds$ is an element of path around the circuit and $Nvq$ is the charge flow per second, which is the current $I$. The path integral of the electric field intensity $(E = F/q)$ is the emf in the circuit $(\mathcal{E})$ and it follows that the rate of producing electrical energy is

$$(\mathbf{A}): \quad dW/dt = I\mathcal{E}, \quad \left( \mathcal{E} = \oint E ds \right)$$

$\mathcal{E}$ being simply another name for the generator voltage $V$ in (5.5).

Where does this energy come from? Of course it can only come from the *rotational motion* of the loop of wire, which is the only source of *mechanical* energy. We'd like to express this rate of working in terms of quantities like forces and torques, which we used in mechanics (Book 4).
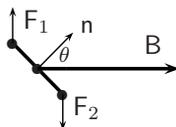


**Figure 37.** Simple dynamo (Fig.30), seen from above.

To do so, look down on the simple dynamo, indicated in Fig.30, from the top. You'll see only the top side of the loop of wire, shown in Fig.37 as a thick black line with a central dot marking the axis of rotation and outer dots marking the tops of the side-wires.

(To understand what follows, you may need to turn back to Book 4 to be reminded of 'torques' and 'moments' etc.; and to Section 3.2 of Book 2 for something on 'angles'.)

The normal ($\mathsf{n}$) to the plane of the loop is shown at an angle of 45° to the field $\mathsf{B}$. The outer dots mark the tops of the vertical side-wires of the loop: they each have length $b$ and carry current $I$. And we already know from (3.12) that each of them will therefore feel a transverse force of magnitude $F = IbB$.

The mechanical 'turning force' or **torque** acting on the loop is simply the total **moment** of the forces around the axis of rotation. As you can see from Fig.37, both $\mathsf{F}_1$ and $\mathsf{F}_2$ apply the same torque $(a/2) \sin \theta \times F$ around the upward axis in the clockwise sense. (Work this out for yourself!)

The total torque on the loop is thus given by

$$\tau = IabB \sin \theta = ISB \sin \omega t$$

– since the whole angle turned through (at constant angular velocity $\omega$) is $\theta = \omega t$ at time $t$.

Now the work done in turning the loop through a tiny angle $\mathrm{d}\theta$ is $\mathrm{d}W = \tau \mathrm{d}\theta$; and if this is done in the short time $\mathrm{d}t$, the *rate* of working will be

$$(\mathbf{B}): \qquad \frac{\mathrm{d}W}{\mathrm{d}t} = ISB \sin \omega t \frac{\mathrm{d}\theta}{\mathrm{d}t} = ISB\omega \sin \omega t.$$

This equation tells us the rate at which mechanical work must be done to keep the armature of the generator spinning and producing electricity; while $(\mathbf{A})$ (above) told us the electric power it

169

could put out into any circuit connected to the output terminals. Both equations are about the rate of 'doing work' or 'supplying energy'. The two kinds of energy seem very different, but they are measured in the same units (the joule) and the *rate* at which one kind is converted into the other is also measured in the same units (1 joule/second is 1 watt). So we can easily compare them: from equation (5.3) we get the emf generated in the spinning armature; and on putting this into **(A)** we find that the rate of sending electrical energy into any circuit connected to the output terminals must be

$$\frac{\mathrm{d}W}{\mathrm{d}t} = ISB\sin\omega t\frac{\mathrm{d}\theta}{\mathrm{d}t} = ISB\omega\sin\omega t.$$

This is exactly the same expression as **(B)!**

---

The conclusion from Example 5 is simply that

> The rate of production of electrical energy by an ideal generator is exactly equal to the rate at which mechanical energy must be supplied in turning the armature

It doesn't matter how the *mechanical* work is being done – by bullocks turning a wooden wheel, or water coming down a pipe from a lake high up in the hills and driving a 'turbine' – and it doesn't matter how the *electrical* energy produced is being used. The generator simply changes one form into another – **without changing the *amount* of energy converted.** This is just another example of the principle of **Conservation of Energy**, which we meet so often in Physics (see Books 4

and 5). The energy we get out of a system is the same as the energy we put in, provided none of it is lost in the form of un-usable energy like the Joule heat mentioned above. If such losses were negligible a dynamo would be 'ideal': its efficiency would be 100%.

When we're thinking of motors instead of generators it's only the *direction* of energy flow that changes. With an electric motor, electrical energy *goes in* and mechanical energy *comes out*. But it's not always possible to reverse the direction: if you try to run an AC generator as a motor, by putting an alternating voltage across the output terminals (1 and 2 in Fig.36), it just won't go. As long as the voltage is positive, the torque on the armature is in the same direction and it begins to rotate: but when the voltage falls to zero (the first crossing point in Fig.32) there is no longer any torque and the moving armature slows down; and when the voltage goes negative the torque direction *reverses* and tries to send the armature back. The heavy armature can't keep up with the alternating current and eventually stops: the current has no work to do except producing Joule heat – so much of it that the armature gets hot and its insulated windings may even catch fire! So you always have to keep in mind the basic physics of what is happening.

With a DC generator, that doesn't happen because the commutator ensures that the current you put in, and therefore the torque, is always in the same direction.

The armature then steadily builds up speed and can safely drive machinery. Fortunately we know, by now, how to produce both AC and DC power. And it's time to turn to the last Section of this Chapter – on electric power and how to move it from one place to another.

## 5.4   Electric power – and how we can move it around

Suppose a huge generator has been built, big enough to produce electric power for a whole city, but it's housed in a power station somewhere up in the mountains and is driven by water from a large dam. That's **hydro-electric power** – and it can't be carried around! The power it produces is needed mainly in the cities, for heating and lighting and for running the machines in factories. The big problem is how to bring electric power to those homes and factories; and to the many far-away farms where it will also be needed.

The first thing you will think of is the use of wires. But that will bring many problems which all have to be solved: first you have to set up a **'national grid'** of wires, extending over the whole country, keeping them safely out of the way of people and animals. Then you have to make sure that the system will be efficient;

and that means you have to think about the Joule heat wasted in miles of curent carrying copper wires.

We know that the rate of working (i.e. the power) transmitted through a wire carrying current $I$, resulting from an emf $\mathcal{E}$ (measured in volts and often denoted by $V$) will be

$$P = (\mathrm{d}W/\mathrm{d}t) = IV. \qquad (5.6)$$

(For example, a 1kW electric heater connected in a circuit with emf 230 volts, will take a current $I$ $(= P/V)$ of $(1000/230)$ amps.)

On the other hand a wire carrying current $I$ in a circuit with emf $V$ will lose energy, as Joule heat, at the rate

$$\mathrm{d}W/\mathrm{d}t = IV = I^2 R, \qquad (5.7)$$

where $R$ is the resistance of the wire and is determined by its thickness and the kind of conducting material (e.g. copper is better than steel). To keep the heat loss small, $R$ should be as small as possible: but there is a limit set by the cost per metre of the wire, which depends on how strong and thick it must be to do its job. Anyway, for given $R$, it is clearly best to keep the current $I$ down to an absolute minimum.

Suppose now we want the wire to transmit power $P$ of, say, 1000 kW. In that case the current will be $I = P/V$ – **inversely proportional to the voltage at which it is transmitted**. For whatever power is needed we

shall have to choose the highest possible emf – even if it's many thousands of volts! That's why you see the wires of the national grid trailing across the countryside, hanging from tall pylons at a great height to keep them out of harm's way: they may be carrying current at 250,000 volts!

This brings us to the last problem in this chapter: if the generator at the power station produces electricity at only, say, 500 volts how can we change the voltage of the supply by a factor of 500? This can be done easily by using a 'transformer', but only for AC current; and that's why all large scale power transmission calls for AC generators.

### Inductances and Transformers

We've come a long way from Faraday's discovery that a changing magnetic flux through a coil of wire can produce an emf in it, and Maxwell's neat expression of it in equation (5.2). But what if the changing flux is produced by an emf, and the circulating current it produces, in *another* coil of wire? The changing flux *induces* an emf and we now need to study this property of **inductance** in more detail.

Suppose we have two coils and we put an AC current $I_1$ through Coil 1: being an alternating current means it will produce a *changing* magnetic field all around it and therefore a changing flux through Coil 2. This will give

rise to an emf and thus, when the circuit is closed, an **induced current** $I_2$.

Suppose Coil 1 is a solenoid, like the one pictured in Figure 25 of Section 3.5, consisting of $N_1$ turns of wire wound round a cylinder of length $L$ and cross-sectional area $A$. Then suppose we wind $N_2$ turns of a second wire on top of it. This particular arrangement is a very simple and basic form of **transformer**, so called because it gives us a way of transforming electric current from one voltage to another – just what we need to do! How does it work?
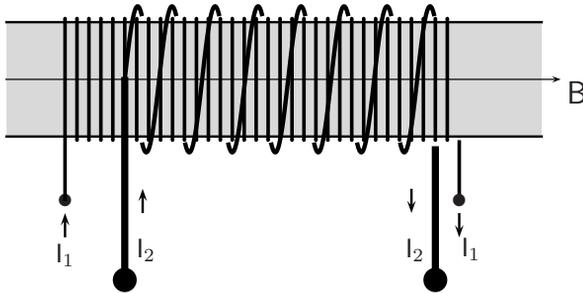


**Figure 38.** A simple transformer (see text)

In the device pictured above, $N_1$ turns of wire are wound round a cylinder, to make a 'primary coil'. On top of this, $N_2$ turns of thicker wire are wound, to make a 'secondary coil'. When an emf $\mathcal{E}_1$, is applied to the primary

(e.g. from an AC generator) it sets up a current $I_1$. But an emf $\mathcal{E}_2$ is also *induced* in the secondary, by the changing magnetic flux passing through it, and appears as a voltage between the 'output terminals' (shown as large black dots). If you connect an external circuit to the terminals a current $I_2$ will flow – but at a voltage corresponding to $\mathcal{E}_2$ instead of the input voltage $V_1 = \mathcal{E}_1$.

We want to know how this thing works, but we have to do a bit more work first. We know from Section 3.5 that the B-field along the axis of the solenoid, formed when a current $I_1$ passes through the primary coil (Coil 1) will be

$$\mathsf{B}_1 = \mu_0(N_1/L)I_1. \qquad (5.8)$$

Here $\mu_0$ (the magnetic permeability of free space) can easily be changed by using, for example, an iron 'core' instead of an empty cylinder.

To find the induced emf in the secondary coil (Coil 2), we start from

$$\mathcal{E}_2 = -\frac{\mathrm{d}}{\mathrm{d}t}\int_{S_2} \mathsf{B}_1 \cdot \mathrm{d}\mathsf{S}_2,$$

which is the 'integral form' of the Maxwell equation (4.18). Here quantities relating to Coil 2 carry a subscript '2' and

$$\mathcal{E}_2 = \oint \mathsf{E}_2 \cdot \mathrm{d}\mathsf{s}_2$$

is the induced emf in Coil 2. (Note that $B_1$ is the field produced by Coil 1, and that's why it carries the subscript '1'.) On substituting for $B_1$ from (5.8), the induced emf in the second coil becomes

$$\mathcal{E}_2 = \mu_0(N_1 N_2)(S/L)\frac{dI_1}{dt} \qquad (5.9)$$

– since the surface integral which gives the flux of $B_1$ through Coil 2 is just (area, $S$)×(normal component, $B_1$)×(number of turns, $N_2$). Notice that only the rate of change of the *current* appears, the other quantities being independent of $t$.

This formula will lead us to everything we need! It is usually written

$$\mathcal{E}_2 = \mathcal{M}_{21}\frac{dI_1}{dt} = \mathcal{M}_{21}\dot{I}_1 \qquad (5.10)$$

where the constant $\mathcal{M}_{21}$ is called the **mutual inductance** of the two coils and the $I_1$ with a dot over it is short for "the rate of change of $I_1$".

The mutual inductance depends only on the 'geometry' of the setup $(S, L)$, together with the numbers of turns in the coils, and the magnetic permeability of the 'core' (e.g. $\mu_0$ for an empty cylinder or $\mu_{\text{iron}}$ – which may be thousands of times bigger – for one with an iron 'core'). For the transformer shown in Fig.38, the mutual inductance of the two coils is

$$\mathcal{M}_{21} = \mu_0(N_1 N_2)(S/L), \qquad (5.11)$$

– as you will see if you compare (5.10) with (5.9) – but the subscripts are usually dropped because (although it's not easy to prove) $\mathcal{M}_{12} = \mathcal{M}_{21}$ if the coils change places. (This is true for all kinds of coil – not just the setup shown in Fig.38.)

Even when there is only *one* coil, the changing flux due to changes in the current passing through it will also induce an emf: this is called a 'back-emf' and arises from the **self-inductance** of the coil. The back-emf set up is again proportional to the rate of change of current going through the coil and is given by a formula similar to (5.10) except that the proportionality coefficient is *negative*, meaning that the emf is in the opposite direction to the one producing the field. The self-inductance is usually denoted by $-\mathcal{L}$, so as not to mix it up with $\mathcal{M}$ for mutual inductance. For simple coils like the ones we're talking about here the self-inductance follows from a formula very similar to (5.11) – but less easy to prove:

$$\mathcal{L} = \mu_0(N^2)(S/L), \qquad (5.12)$$

– just as if you had two coils and put $N_1 = N_2 = N$. When there really are two coils of the same kind, their self- and mutual-inductances will be

$$\begin{aligned} \mathcal{L}_1 &= \mu_0(N_1^2)(S/L), \quad \mathcal{L}_2 == \mu_0(N_2^2)(S/L), \\ \mathcal{M} &= \mu_0(N_1 N_2)(S/L). \end{aligned} \qquad (5.13)$$

Before going on we should think for a minute about units.

**A note on units**

From (5.11), remembering that the emf is measured in volts (V) and the current in amps (A), the inductance will have the same dimensions as $V/(\mathrm{d}I/\mathrm{d}t)$ and will be expressed in volts/amps per sec. If the current in one circuit is changing at the rate of 1 A s$^{-1}$ and induces an emf of 1 V in another circuit, the two circuits will have 1 unit of mutual inductance. This unit is called the 'henry', after the American scientist Henry who discovered the effect at about the same time as Faraday; so the unit is

Unit of inductance is the henry: 1 H = 1 V A$^{-1}$ s.

Now, at last, we're ready to understand the transformer in Fig.38 – which gives us a good Example on solving the equations.

---

**Example 6.** A simple transformer (Fig.38)

If you connect the primary coil (Coil 1) to a source of voltage ($V_1$), the emf $\mathcal{E}_1 = V_1$ and this will set up a curent $I_1$ the coil. If now the current is made to vary (e.g. by rapidly 'making and breaking' the circuit, or simply by using AC current – which goes up and down anyway) what will happen? An emf will be induced in Coil 2 and, according to (5.10,) this will be proportional to $\mathcal{M}_{21}$ and to the rate at which the primary current is changing – because the changing flux of B through Coil 1 is also felt by Coil 2.

Let's write down the total emf driving electric charge round each

179

of the two circuits, primary and secondary. In the primary, we have $\mathcal{E}_1$ from the generator providing a voltage $V_1$; but then there is a back-emf $-\mathcal{L}_1\dot{I}_1$ due to the changing current; and also another $-\mathcal{M}\dot{I}_2$ due to the changing current *in the secondary*. (Note that the inductance is *mutual* as it arises from interaction between the two coils.)

What do the two emfs do? They drive current around the two circuits, the input circuit which includes the generator and the output circuit which includes anything connected to the ouput terminals (usually called the 'load'). And each emf is used to overcome the *resistance* of the conducting circuit to the current being pushed through it. Putting all this together, and remembering Ohm's law (2.9), we get two equations:

$$
\begin{aligned}
(A): \quad \mathcal{E}_1 - \mathcal{L}_1\dot{I}_1 - \mathcal{M}\dot{I}_2 &= R_1 I_1 \\
(B): \quad \mathcal{E}_2 - \mathcal{L}_2\dot{I}_2 - \mathcal{M}\dot{I}_1 &= R_2 I_2.
\end{aligned}
$$

These look like very nasty differential equations, worse than the ones you came across in Books 3 and 3A. The two of them are even coupled together! But we're going to find that, for the simple transformer, a solution comes out very easily.

What we want to find is the output voltage $V_2$ measuring the total emf in the secondary coil, but the first term ($\mathcal{E}$ in the second equation is usually missing – the generator belongs only to the primary circuit, giving the *input*. So we put that term to zero. And next we can make an approximation, by neglecting the resistance $R_1$ of the primary coil: this is reasonable because we always try to use good conducting wires anyway, to keep the wasted Joule heat to a minimum. If we do this, putting $R_1$ and $\mathcal{E}_2$ to zero in the last two equations, we are left with

$$
\begin{aligned}
(C): \quad \mathcal{E}_1 - \mathcal{L}_1\dot{I}_1 - \mathcal{M}\dot{I}_2 &= 0 \\
(D): \quad -\mathcal{L}_2\dot{I}_2 - \mathcal{M}\dot{I}_1 &= R_2 I_2.
\end{aligned}
$$

According to Ohm's law $V_2 = R_2 I_2$, so that's the term we must focus on in the second equation. We don't know anything about the two terms on the left of the $=$ sign except that $\dot{I}_1$ and $\dot{I}_2$ are both found in the first equation: so let's work on that one. The steps are as follows: Multiply (C) by $\mathcal{L}_2$ and you'll get

$$\mathcal{E}_1 \mathcal{L}_2 = \mathcal{L}_2 \mathcal{L}_1 \dot{I}_1 + \mathcal{L}_2 \mathcal{M} \dot{I}_2.$$

But (5.13) shows that $\mathcal{L}_2 \mathcal{L}_1 = \mathcal{M}^2$, so the last equation can be rewritten as $\mathcal{E}_1 \mathcal{L}_2 = \mathcal{M}(\mathcal{M}\dot{I}_1 + \mathcal{L}_2 \dot{I}_2)$ and the thing in brackets, apart from a negative sign, is just what you have in (D). So you can get rid of the nasty time derivatives, replacing $(\mathcal{M}\dot{I}_1 + \mathcal{L}_2 \dot{I}_2)$ by $-R_2 I_2$. We're getting somewhere! At this point we have

$$\mathcal{E}_1 \mathcal{L}_2 = -\mathcal{M} R_2 I_2$$

and $R_2 I_2$ is what we're looking for – the output voltage in the secondary circuit.

All we have to do now is put in the values of the inductances $\mathcal{L}_2$ and $\mathcal{M}$, given in (5.13), to get

$$V_2 = R_2 I_2 = -\mathcal{E}_1 \left( \frac{\mathcal{L}_2}{\mathcal{M}} \right) = -\mathcal{E}_1 \frac{N_2}{N_1}.$$

Of course, $V_1$ is just another name for the emf $\mathcal{E}_1$ supplied by the generator in Circuit 1, and the final result is therefore (writing $\mathcal{E}_1 = V_1$ and dropping a minus sign, which only tells you that $V_1$ and $V_2$ are in opposite directions)

$$N_1 V_2 = N_2 V_1.$$

What a beautifully simple result after all that work!

What Example 6 has shown us is that for any simple transformer, where heat losses are neglected, the output and input voltages stand in the same ratio as the numbers of turns of wire in Coil 2 (the secondary) and Coil 1 (the primary):

$$\frac{V_2}{V_1} = \frac{N_2}{N_1}. \tag{5.14}$$

If you want to reduce the voltage of the power supplied to your home (probably about 250 V) to a safe level, so it won't kill you when you try to use it, you'd better be sure there are at least ten times as many turns in the primary (input) coil as in the secondary. That way you'll get an output at not more than 25 V. But remember you're not reducing the *power*. As we noted in talking about the Joule heat, the power supplied by current $I$ flowing between points at a difference of potential $V$ is $P = VI$ and is measured in watts: it's the rate of doing work – transfering energy – and in an ideal transformer (which is what we're always talking about) the small amount of wasted Joule heat is neglected. In that case the principle of Energy Conservation says that what comes out (of the secondary) equals what goes in (from the primary): $V_2 I_2 = V_1 I_1$. So if the input voltage is divided by 10, the input current will be *multiplied* by 10: $(I_2/I_1) = (V_1/V_2)$.

Now it's clear why the wire of the secondary coil in Fig.38 is shown much thicker than that of the primary;

it has to carry a current ten times as heavy and that means that the Joule heat produced will be a *hundred* times as great! The coil would over-heat and the wire might even melt! Sometimes, in fact, this is useful: in **electric welding** you have to melt two pieces of metal where they come into contact and this is done by passing a heavy current through them. In that case you are using a **step-down** transformer (stepping down the voltage to increase the current).

In other cases, you may want to step *up* the voltage. This is what has to be done at the power station, to transform the voltage produced at the generator to the very high voltage used in long distance power trnamission – possibly 250,000 V. But however you may need to change the voltage you can do it using transformers. There will be many technical problems of course, keeping things cool when they carry heavy currents and keeping wires well insulated when they are at a high voltage; but as long as you know the principles the difficulties can always be overcome.

# Chapter 6

# Waves that travel through empty space

## 6.1 What is a wave?

Most people think first of waves on the sea, or of the smaller waves set up when you drop a stone in a pond; but then there are sound waves, where there is nothing to see, but a noise may come to your ears when someone calls you from a distance. In both cases 'something' (let's call it a 'disturbance') travels from one point to another at a certain speed, which determines how long it takes to get there. In the first example, the waves travel through water and it's an up-and-down motion

that travels; in the second, the waves travel though the air and although you can't see anything happening your ears can *feel* the arrival of the noise as a kind of 'vibration' of your eardrums. In both cases the waves are carried by a 'medium' (water or air) and the disturbance is passed on directly from one bit of the medium to the next bit. But in this chapter we'll be thinking of disturbances you can neither see nor hear; they travel through empty space in which there's no 'stuff' of any kind to carry them. They are carried by **fields** and what they carry is not an up-and-down motion of stuff, but rather a variation of **field strengths** as you go from one point to the next.

What do all these kinds of wave have in common? They all involve a 'disturbance' in which something we can measure, call it $\phi$, varies with time – it is a function of time: $\phi = \phi(t)$; and it also depends on the point in space, with coordinates $x, y, z$ say, where it is measured. The function we're talking about may be the height of a water wave above the water level on a still day; or the pressure of the air at point $x, y, z$ in a sound wave. But whatever it may be it will increase or decrease (that's the 'disturbance') as time passes; so it is correct to say

$$\phi = \phi(x, y, z, t) \tag{6.1}$$

– which is general enough to show all the main properties of wave motion.

Although this is a function of four variables, three for space and one for time, we can go a long way by looking at the simplest possible case of a 1-dimensional system, where the wave travels along the x-axis and no other coordinates need be shown: $\phi = \phi(x, t)$. At any given moment (e.g. $t = 0$) the wave will have a certain 'shape'

$$\phi(x, 0) = f(x) \qquad (6.2)$$

which shows how $\phi$ varies with $x$ alone. at the initial time $t = 0$. We don't know what mathematical form the function will take: that's what we want to find in any particular problem.

Next we'll suppose that the wave moves along the x-axis with *constant speed c*, without changing its shape. So 'photographs' of the wave, taken at $t = 0$ (the initial time) and at time t, might look as in Figure 39.
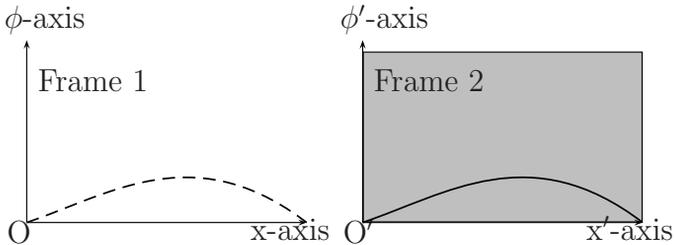


**Figure 39.** Travelling wave profile.

The graph on the left shows the *shape* of the wave, called the **wave profile**, at $t = 0$ when it sets off, while the

one on the right shows the same profile pushed along the positive x-axis through a distance $ct$ – the distance it has travelled in time $t$.

Each wave profile has been put in its own 'picture frame'. The profile shown in Frame 1 shows what the wave looked like when it set off at $t = 0$. The one shown in Frame 2 shows the wave profile at time $t$. It is exactly the same function of $x'$, we are supposing, where the distance $x'$ is measured from the origin O$'$ at the bottom left-hand corner of Frame 2; so $\phi = f(x')$. But $x'$ is related to $x$ measured from the origin O, fixed in space, by $x' = x - ct$ as you can see from Fig.39. It follows that for the moving wave

$$\phi = f(x - ct), \qquad (6.3)$$

which is referred to the fixed origin and now includes the time dependence.

Equation 6.3 gives the standard description of a wave moving with constant velocity $c$ along the positive x-axis, without change of shape. The equation for a similar wave moving in the *negative* direction is obtained simply by changing the sign of the velocity and is thus

$$\phi = f(x + ct). \qquad (6.4)$$

In Section 5.1 of the last chapter, we already found an example of 'wavelike' behaviour in the voltage variation

from an AC generator: the voltage was a function of time alone, with the form $V = V_0 \sin \omega t$, $\omega$ being a constant that determines the frequency with which $V$ goes up and down. This wavelike variation is shown in Fig.32 and a function with a cosine instead of the sine has a similar form, but starting at the maximum $V = V_0$ instead of at $V = 0$. Let's play with a function of the cosine form

$$\phi = A \cos ax, \tag{6.5}$$

where $A$ and $a$ are simply constants, and pretend that it is a wave profile for some disturbance moving along the positive x-axis. This will be very different from the function in Fig.39, which was *localized* in a small region of space; for the sine and cosine functions go on forever, in both directions. But it has the same property of looking exactly the same when you shift it through a certain distance along the x-axis. Both $\phi$ and the slope $d\phi/dx$ repeat their values when the argument of the function $(ax)$ increases in value by any integer multiple (positive or negative) of $2\pi$ – for $ax$ plays the part of an *angle* and the 'circular functions' (like sine and cosine) are unchanged under any complete revolution. Thus, whenever $x$ changes to $x + \Delta x$ such that $a\Delta x = n \times 2\pi$ the function (6.4) will appear to be unchanged: the smallest non-zero value of $\Delta x$ with this property follows when $n = 1$. It is called the **wavelength** and is usually denoted by the letter 'lambda': $\lambda = 2\pi/a$. The wave function (6.5) can

therefore be written in terms of wavelength as

$$\phi = A\cos(2\pi x/\lambda). \qquad (6.6)$$

The constant $A$ is called the **amplitude** of the wave: the value of $\phi$ 'oscillates' between $\pm A$. One complete 'up-down' oscillation extends over one wavelength, but $\phi$ itself continues indefinitely in both directions.

(Note that in talking about 'wavelength' we're now thinking of only a small part of the whole wave profile – the part in which $\phi$ runs through all its possible values just once, as $x$ increases by $\lambda$ – before being repeated over and over again. Such functions are said to be **periodic** and you met them long ago in Section 1.3 of Book 3. If you want to make it clear that you're talking about the whole function, with all its repetitions, you can call it a **'wave train'**.)

There is still no *time variable* in equation (6.6): this $\phi$ gives only the *shape* of a wave, not saying whether or not it is moving. The general form of a wave moving along the positive x-axis at velocity $c$, without change of shape, is given in (6.3), so to get a 'travelling wave' of this shape we only need to replace $x$ by $x - ct$. The following Example shows what happens when you use the cosine function (6.6).

---

**Example 1.** Making the wave move

When $x$ is replaced by $x - ct$ we get a new wave function

$$\phi = A\cos(2\pi/\lambda)(x - ct).$$

Now it's clear how $\phi$ will vary with *time*. If you look at one point (fixing $x$) but wait for time to pass, letting $t \to t + \Delta t$, the

argument of the cosine function will change by $-(2\pi c/\lambda)\Delta t$; and whenever this is an integral multiple of $2\pi$ (negative or positive makes no difference) the form of the wave is exactly repeated. The smallest value of $\Delta t$ is called the **period** $T$ of the wave: $T = \lambda/c$. So the wave function may be written in other forms:

$$\phi = A\cos 2\pi\left(\frac{x}{\lambda} - \frac{t}{T}\right) = A\cos 2\pi(kx - \nu t).$$

The reciprocals of the period and the wavelength also have a clear meaning and are given special names: they are

**frequency** $\nu = 1/T,$ **wave number** $k = 1/\lambda.$

Thus $\nu$ is the number of oscillations per unit time, while $k$ is the number of waves per unit distance in the direction of propagation. The 'argument' of the cosine function (call it $\theta$) is the **phase** of the wave. Two wave functions, $\phi_1$ and $\phi_2$, with the same value of $\theta$ are said to be 'in phase': if $\theta_1 \neq \theta_2$ they show a 'phase difference'. The phase of the functions in the form given above is

$$\theta = 2\pi\left(\frac{x}{\lambda} - \frac{t}{T}\right) = 2\pi(kx - \nu t).$$

---

Important results from Example 1, for a wave of cosine form travelling along the x-axis with velocity $c$, are collected below:

$$\text{Simplest form :} \quad \phi = A\cos(2\pi/\lambda)(x - ct), \quad (6.7)$$

but other forms are clearly

$$\phi = A\cos 2\pi\left(\frac{x}{\lambda} - \frac{t}{T}\right) = A\cos 2\pi(kx - \nu t) \quad (6.8)$$

The phase of the wave function in the forms given above (the argument of the cosine factor) is

$$\theta = 2\pi \left( \frac{x}{\lambda} - \frac{t}{T} \right) = 2\pi(kx - \nu t).$$

– showing that the wave function is periodic in both space and time. What this means will be clear if you study Figure 40 below:
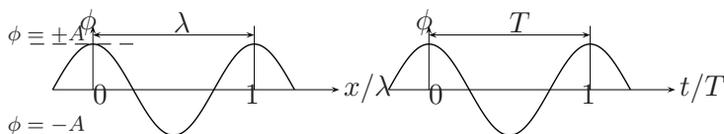


**Figure 40**. Periodic variation of a wave, in space (left) and in time (right).

The picture on the left shows how $\phi$ varies as $x$ changes, at any given time, while that on the right shows how it varies with time at any given point.

It is important to note that $\nu \times \lambda$ is the distance travelled by the wave in unit time. This is the **velocity of propagation**, being the number of waves going out every second times the length of each one:

$$\nu \times \lambda = \frac{\nu}{k} = c. \qquad (6.9)$$

This relationship is true for any kind of wave propagation without change of shape.

## 6.2 The wave equation

Now we know how to *describe* certain simple kinds of wave motion it's time to start thinking about where the mathematical description comes from. Usually the description comes from the solution of a **wave equation**. Every kind of wave motion will have its own equation, one for water waves, one for sound waves, one for waves on a vibrating string, and so on; but the same equation will usually apply to many different kinds of wave – only the physical meaning of the symbols changing. So you only have to study one problem and find how to solve it and you can use the solutions over and over again.

We're now going to look for a wave equation that could lead to a solution like those we've been looking at it the last section. This is really very easy because we're going 'backwards' from a function that represents, from the way it was set up, a disturbance travelling along the positive x-axis with velocity c and without change of shape: we're starting from what must be a *solution*, $\phi$, of the differential equation we want. And since it's going to involve *rates of change* of the wave function with respect to the only variables, $x$ and $t$, we can simply try differentiating $\phi$ and look for a relationship between

$$\phi, \quad \partial\phi/\partial x, \quad \partial\phi/\partial t, \quad \partial^2\phi/\partial x^2, \quad ...\text{etc.}$$

which must always be satisfied. And that will be the

equation we want! To see how easy it is, let's do it.

___

**Example 2.** Setting up a wave equation

We can take $A = 1$ (why?) and start from the form (6.8) with $1/\lambda = k$ and $1/T = \nu$. So we start with

$$\phi = \cos 2\pi(kx - \nu t)$$

and play around with partial derivatives (as we have two independent variables). You find easily (look back at Book 3 if you need help)

$$\frac{\partial \phi}{\partial x} = -\sin 2\pi(kx - \nu t) \times 2\pi k$$

$$\frac{\partial^2 \phi}{\partial x^2} = -\cos 2\pi(kx - \nu t) \times (2\pi k)^2 = -(2\pi k)^2 \phi$$

$$\frac{\partial \phi}{\partial t} = -\sin 2\pi(kx - \nu t) \times (-2\pi\nu)$$

$$\frac{\partial^2 \phi}{\partial t^2} = -\cos 2\pi(kx - \nu t) \times (-2\pi\nu)^2 = -(2\pi\nu)^2 \phi$$

and that's all you'll need.

It follows at once that

$$\frac{1}{(2\pi k)^2} \frac{\partial^2 \phi}{\partial x^2} = \frac{1}{(2\pi\nu)^2} \frac{\partial^2 \phi}{\partial t^2} = -\phi$$

and therefore that

$$\frac{\partial^2 \phi}{\partial x^2} = \frac{k^2}{\nu^2} \frac{\partial^2 \phi}{\partial t^2}.$$

That's an equation relating the second derivatives of $\phi$ with respect to the two independent variables $x$ and $t$: it's the wave equation we want.

Now from (6.9) it is known that $k/\nu = 1/c$ and Example 2 therefore tells us that the standard **wave equation**

$$\frac{\partial^2 \phi}{\partial x^2} = \frac{1}{c^2}\frac{\partial^2 \phi}{\partial t^2} \qquad (6.10)$$

has, among its possible solutions, functions of the kind we've been looking for. In fact (6.10) is the standard partial differential equation for propagation of any kind of disturbance along the x direction, with velocity $c$ and without change of shape!

Of course we live in a 3-dimensional world and a 1-dimensional equation may seem too special to be useful: sound waves, for example, spread out in all directions. But you already know enough to be able to guess the more general equation.

**Waves in 3 dimensions**

In earlier chapters all equations for the electric and magnetic fields have been written for 3-space, where any point is labelled by its x-,y- and z-coordinates and any vector by its corresponding components. We also know that space is **isotropic**, so that (roughly speaking) 'what's good for one direction is good for any other'. For example, the equation for a wave travelling along the positive y-axis can be obtained from (6.10) just by changing $x$ to $y$.

To get a 3-dimensional form of the wave equation, we need to replace the differential operator $(\partial^2\phi/\partial x^2)$ by an 'equivalent' operator for working on a function of all the variables $\phi(x, y, z, t)$. We've already come across such operators in Section 1.4 of this book, namely the gradient, the divergence and the curl.

Suppose now that $\phi$ stands for the electric potential at some point $(x, y, z)$ in space. We know from earlier Sections that $-\text{grad }\phi = \mathsf{E}$ defines a **vector field** with components

$$E_x = -\frac{\partial\phi}{\partial x}, \quad E_y = -\frac{\partial\phi}{\partial y}, \quad E_z = -\frac{\partial\phi}{\partial z}.$$

We know also that div $\mathsf{E}$ is obtained from $\mathsf{E}$ by differentiating each component with respect to 'its own' coordinate and adding the results:

$$\text{div } \mathsf{E} = \frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z}.$$

Like the potential $\phi$, this is a scalar function of position. If now we start from $\phi$ and operate first with 'grad' and *then* with 'div' we'll find the result of a compound operation 'div grad' (also called "the Laplacian"):

$$\text{div grad } \phi = \left(\frac{\partial^2\phi}{\partial x^2} + \frac{\partial^2\phi}{\partial y^2} + \frac{\partial^2\phi}{\partial z^2}\right). \qquad (6.11)$$

This operator is often denoted by $\nabla \cdot \nabla$, or $\nabla^2$, and called 'del-squared', but take care not to think of it as 'del' or 'div' applied twice: it is simply the operator defined in (6.11) and it gives us the correct generalization of the single term $\partial^2\phi/\partial x^2$ when we want to study waves in *three* dimensions (in '3-space') instead of just one.

To see how this can be, let's try replacing (6.10) by what seems might work when $\phi$ depends on all three spatial variables. If we suppose the right equation is

$$\nabla^2 \phi = \left( \frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} + \frac{\partial^2 \phi}{\partial z^2} \right) = \frac{1}{c^2} \left( \frac{\partial^2 \phi}{\partial t^2} \right), \quad (6.12)$$

then it's easy to check that the one-dimensional solutions we already know will also satisfy this new equation. (You may need reminding of the 3-dimensional geometry you studied in Book 2, but you can skip the details if you want to keep moving.)

The wave defined in (6.7) travels along the x-axis with velocity $c$, but now we're thinking of 3-space and on the x-axis the y- and z-coordinates are always zero. If, however, we move off the x-axis, letting $y$ and $z$ take any values we wish, but keeping $x$ constant, the points we're thinking of will all lie on the same **plane**. This plane defines a **wave-front**: it's where the wave-front setting out from the plane with $x = 0$ will have got to at time $t$. The whole travelling wave will be a procession of wave-fronts, all parallel to each other and perpendicular

to the x-axis – along which they move with velocity $c$.

With this picture in mind let's go back to (6.7) and see what happens when $\phi = \phi(x, y, z, t)$ and we look for a wave-front moving along the x-axis.

---

**Example 3.** Going into three dimensions

All the differential operators are now *partial*, so $\mathsf{D}_y = \partial/\partial y$ and $\mathsf{D}_z = \partial/\partial z$ will just destroy $\phi$ when they act at points on a wave-front, because $\phi$ is *constant* for all values of $y$ and $z$ on that surface. The second and third terms in (6.12) will thus disappear, leaving us with

$$\nabla^2 \phi = \frac{\partial^2 \phi}{\partial x^2} = \frac{1}{c^2} \frac{\partial^2 \phi}{\partial t^2}.$$

But this looks like the one-dimensional equation we set up in (6.10) – the $y$ and $z$ variables are just 'sleeping' and are not even shown! Moreover, the solution given in (6.8) is now seen to satisfy also the 3-space equation (6.12).

In just the same way you can show (do it!) that the same equation (6.12) has three 'families' of plane-wave solutions, all with wavenumber $k$ and frequency $\nu$. They are:

(a): $\phi = A \cos 2\pi(kx - \nu t)$

– wave-fronts propagating along the x-direction, $\phi$ independent of $y, z$,

(b): $\phi = A \cos 2\pi(ky - \nu t)$

– wave-fronts propagating along the y-direction, $\phi$ independent of $z, x$,

(c): $\phi = A \cos 2\pi(kz - \nu t)$

– wave-fronts propagating along the z-direction, $\phi$ independent of

$x, y$.

These are not the only solutions of (6.12). Think of sound waves, for example, where the sound from an explosion at the origin O spreads out in *spherical* waves, so that the noise arrives at all points at the same distance $r = ct$ at the same time $t$: in that case all points $(x, y, z)$ for which $x^2 + y^2 + z^2 = (ct)^2$ will lie on a spherical wave-front of radius $ct$. But even the plane-wave solutions do not all belong to the families indicated above. Plane wave-fronts travelling along a *rotated* x-axis, with direction cosines $l, m, n$ (see Section 5.4 of Book 2), would also be solutions – all directions in space being just as acceptable. In that case the distance travelled in the direction of the *rotated* x-axis is $x' = lx + my + nz$ and if $x$ in (6.8) is replaced by $x'$ we'll find an exactly similar plane wave (same frequency, wavelength, speed, and shape) of the form

$$\phi = A \cos 2\pi[k(lx + my + nz) - \nu t],$$

where (by definition, see Section 5.4 of Book 2) the direction cosines take any values, provided $l^2 + m^2 + n^2 = 1$.

---

The last result in Example 3 shows that

$$\phi = A \cos 2\pi[k(lx + my + nz) - \nu t] \qquad (6.13)$$

describes a plane wave travelling along an axis with direction cosines $l, m, n$, all wavefronts being perpendicular to the axis.

It's not easy to see the connection between this result and the three families listed above, but it's already clear that the *general* solution of (6.12) must contain a large

number of *particular* solutions, with various values of any 'parameters' such as $k$, $\nu$, $l$, $m$, $n$. You may remember that in Book 3 (Section 6.2 Example 3 on 'making music') we set up a wave equation for waves on a vibrating string; and we even found a useful way of getting solutions by 'separating the variables'. Here we can start off along the same lines, exploring other kinds of solution.

## 6.3   Separable solutions

In Section 6.2 of Book 3 we studied the partial differential equation

$$\frac{\partial^2 y}{\partial x^2} = \frac{1}{c^2}\frac{\partial^2 y}{\partial t^2},$$

where $y$ was the displacement at point $x$ of a stretched string, pulled aside and then left to vibrate. That equation was the same as (6.10) except that the 'disturbance' was called $y$, while in this chapter we're calling it $\phi$, It's the *equation* that matters, not what the symbols may stand for! So let's start again with the general wave equation (6.12) and use what we learnt in Book 3 about 'separating the variables'.

Forget about plane-wave solutions! Instead we can do as we did in Book 3, looking for solutions of an altogether

different kind, trying

$$\phi(x, y, z, t) = X(x)Y(y)Z(z)T(t) \qquad (6.14)$$

– a product of four factors, one for each of the independent variables. Here we'll suppose $X$ is a function of $x$ only, not depending in any way on the values of $y, z, t$, and so on. Only the last factor will depend on the time $t$ at which we look at the disturbance $\phi$. It may not be possible to find solutions of this particular form – but it's worth a try! In Section 6.2 of Book 3 the idea worked well: for 1-dimensional waves on a string it was possible to satisfy the wave equation with a function of the form $F(x)T(t)$, provided each factor was chosen as the solution of a simple differential equation in one variable.

Remember how the argument went? We supposed that the displacement $y$ of the vibrating string, at a point distant $x$ from one end and at time $t$, could perhaps be written as $y(x, t) = F(x)T(t)$. By substituting this product in the wave equation $(\partial^2 y/\partial t^2) = (1/c^2)(\partial^2 y/\partial t^2)$, and then dividing both sides by the same product, it was possible to rewrite the equation in the **separated form**

$$\frac{1}{F}\left(\frac{\mathrm{d}^2 F}{\mathrm{d}x^2}\right) = \frac{1}{c^2 T}\left(\frac{\mathrm{d}^2 T}{\mathrm{d}t^2}\right). \qquad (6.15)$$

The two sides of the equation are entirely independent: the left-hand side involves only the variable $x$, the right-hand side involves only $t$. (Note also that *partial* derivatives are no longer needed because $F$ and $T$ are each

functions of only *one* variable.) The two sides can only be equal for all values of $x$ and $t$ if both are equal to the same *constant*, $C$ say. The equation that the function $T(t)$ must satisfy is thus

$$\frac{1}{T}\left(\frac{d^2 T}{dt^2}\right) = Cc^2$$

and you already found the general solution of this equation in Section 6.2 of Book 3 (Example 2). When the constant $C$ is given a negative value (call it $-\omega^2$) the solution is oscillatory and is in general

$$T(t) = A\sin\omega ct + B\cos\omega ct, \qquad (6.16)$$

where $A$ and $B$ are arbitrary constants.

This showed us that solutions of the separated form $y(x,t) = F(x)T(t)$ could indeed be found, with a 'shape' factor $F(x)$ and a time factor $T(t)$ which simply moved the shape up and down with a frequency $\nu = \omega/2\pi$. (You'll remember that $\omega$ is often called the 'angular frequency' because whenever the 'angle' $\omega t$, in the sine and cosine functions, increases by $2\pi$ the time factor has gone through a whole cycle of possible values and starts repeating: $t = 2\pi/\omega$ is the period of a complete oscillation and its reciprocal is the number of cycles per unit time i.e. the frequency $\nu$.)

Now let's move on to the 3-dimensional case in which the disturbance (6.14), namely $\phi = X(x)Y(y)Z(z)T(t)$,

must satisfy the equation (6.12). The next Example will show how the four factors in the product can be found.

---

**Example 4.** Separating the equation in three dimensions

We're going to put the product $XYZT$ in place of $\phi$ on both sides of the equation $\nabla^2\phi = (1/c^2)\phi$ and then try to choose the four factors so that the equation is satisfied.

By putting that product into $\nabla^2\phi$ on the left-hand side and noting that each term in the operator works on only one of the four factors, leaving the others unchanged, you'll find

$$\nabla^2\phi = YZT\left(\frac{\mathrm{d}^2X}{\mathrm{d}x^2}\right) + XZT\left(\frac{\mathrm{d}^2Y}{\mathrm{d}y^2}\right) + XYT\left(\frac{\mathrm{d}^2Z}{\mathrm{d}z^2}\right)$$

and this must be equal to the term on the right-hand side, namely

$$\frac{1}{c^2}\frac{\partial^2\phi}{\partial t^2} = \frac{1}{c^2}XYZ\left(\frac{\mathrm{d}^2T}{\mathrm{d}t^2}\right).$$

Now set the two things equal and divide everything by the product $XYZT$. This will give you (do it yourself!) the 'separated equation' in which each term depends only on one variable:

$$(\mathrm{A}):\quad \frac{1}{X}\left(\frac{\mathrm{d}^2X}{\mathrm{d}x^2}\right) + \frac{1}{Y}\left(\frac{\mathrm{d}^2Y}{\mathrm{d}y^2}\right) + \frac{1}{Z}\left(\frac{\mathrm{d}^2Z}{\mathrm{d}z^2}\right) = \frac{1}{c^2T}\left(\frac{\mathrm{d}^2T}{\mathrm{d}t^2}\right).$$

The term on the right-hand side must be a constant, as in the 1-dimensional case above where we called it $C$ and were led to the simple differential equation

$$\frac{1}{c^2T}\left(\frac{\mathrm{d}^2T}{\mathrm{d}t^2}\right) = C,\ \text{or}\ \left(\frac{\mathrm{d}^2T}{\mathrm{d}t^2}\right) = -\omega^2c^2T.$$

Here $C$ has again been chosen *negative* $(-\omega^2)$, as we're interested only in *oscillatory* solutions, and the solution is again (6.16).

The terms on the left in (A) must also be constants, each one being completely independent of the others, but their sum must be equal to the constant on the right (that's what the equation tells us!). As long as we're looking for oscillating (i.e. wave-like) solutions, these constants must also be negative numbers; so we write them as $-\alpha^2, -\beta^2, -\gamma^2$ and require that $\alpha^2 + \beta^2 + \gamma^2 = c^2\omega^2$. When these conditions are satisfied the product

$$\phi(x, y, z, t) = X(x)Y(y)Z(z)T(t)$$

will certainly be a solution of the wave equation (6.12).

---

We're now ready to put down a rather general solution of the wave equation in three dimensions. It can be written in the form

$$
\phi(x, y, z, t) = \binom{\sin}{\cos}(\alpha x)\binom{\sin}{\cos}(\beta y)
$$
$$
\times \binom{\sin}{\cos}(\gamma z)\binom{\sin}{\cos}(\omega c t) \quad (6.17)
$$

This means that any product of the four factors indicated, using either the sine or the cosine in any of the four, will satisfy the wave equation (6.12). For example, taking once again the case of a vibrating string, stretched along the x-axis with its ends at $x = 0$ and $x = L$ it is easy to get back the particular solution we found in Book 3.

**Example 5.** Waves on a stretched string

In this case we choose the product (two factors only, one for $x$ and one for $t$): $\phi(x,t) = \sin \alpha x \sin \omega ct$. But although this is a particular solution it still isn't particular enough! Where does the length of the string come in? and how do we know the value of the frequency factor $\omega$. In fact, to complete the solution we have to satisfy any **boundary conditions** that there may be. Here the ends of the vibrating string are fixed (remember it is stretched!) and this means $\phi = 0$ at both $x = 0$ and $x = L$. The first condition rules out a cosine solution, which would give $\phi = 1$ at $x = 0$, while the second requires that $\sin \omega cL = 0$, which means that $\omega cL = n\pi$ where $n$ is any *integer*.

The particular solutions that satisfy the boundary conditions are thus (remembering the condition $\alpha^2 = c^2 \omega^2$)

$$\phi(x,t) = A \sin(n\pi x/L) \sin(n\pi ct/L), \quad (n \text{ integral}).$$

There are many solutions, corresponding to different choices of the integer $n$, all of the same kind – with $\phi = 0$ at the two ends of the string. They are called "standing waves". You should sketch their forms for a few values of $n$ (e.g. $n = 1, 2, 3$) and calculate the vibration *frequencies*, which correpond to different musical notes.

---

You may be wondering why nothing was said in Example 5 about details like the material of the string, or how heavy it is, or how tightly it is stretched. That's because they don't matter: all that we need to know in

talking about vibrations and wave motion of any kind is the velocity of propagation $c$ – the other things are important only in determining the value of $c$. That's Physics, the rest is just Mathematics. Now it's time to turn back to Physics and the way electric and magnetic fields propagate through space.

## 6.4 Electromagnetic waves

In Chapter 5 you learnt how electricity could be generated in great quantities and carried over great distances through metal conductors in the form of wires; but you also know, from everyday experience with radio and television, that even without wires it's possible to send electric and magnetic effects from one place to another. How is it possible? You can perhaps imagine that electric charges and currents, rushing up and down in wires, generate rapidly changing *fields* and that these 'disturbances' can travel through space like other kinds of wave motion. In this Section we want to prove, from Maxwell's equations, that this is all possible and that even the speed with which an **electromagnetic wave** travels can be predicted from the equations.

We'll be talking about the difficult part, where the space is really empty – no conductors that can hold charge, no wires that can carry current. So we can leave out the

charge density $\rho$ and the current density $j$ and start from Maxwell's equations in the form:

$$
\begin{array}{ll}
\text{(a) div } \mathsf{E} = 0 & \text{(c) div } \mathsf{B} = 0 \\[2em]
\text{(b) curl } \mathsf{E} = -\partial \mathsf{B}/\partial t & \text{(d) curl } \mathsf{B} = \mu_0 \epsilon_0 \left( \partial \mathsf{E}/\partial t \right)
\end{array}
$$

(6.18)

These look quite a bit simpler than the complete Maxwell equations (4.20). But, simple as they may be, they will hold good from here to the stars! It's only where you start the waves off (at a **transmitter**) and where you 'pick them up' (at a **receiver**) that you'll need to worry about wires and currents.

What do we want to do with the equations? We're not trying to solve them, to find the fields at all points in space: all we want is a single quantity – perhaps one component of the electric field – and to set up an equation that will describe how it travels through space. That will be a wave equation, like (6.12) and should contain something that stands in place of $c$ – a velocity of propagation. But to focus on one quantity in (6.20) and separate it from the others is not easy: we have to learn how to play with the differential operators.

We know that the curl operation takes us from one vector field to another; from $\mathsf{E}$, say, to $\operatorname{curl}\mathsf{E}$, and the components of the vectors are related by

$$
\begin{aligned}
(\operatorname{curl}\mathsf{E})_x &= \frac{\partial}{\partial y}E_z - \frac{\partial}{\partial z}E_y, \\
(\operatorname{curl}\mathsf{E})_y &= \frac{\partial}{\partial z}E_x - \frac{\partial}{\partial x}E_z, \\
(\operatorname{curl}\mathsf{E})_z &= \frac{\partial}{\partial x}E_y - \frac{\partial}{\partial y}E_x.
\end{aligned}
$$

Now let's start from (6.18b) and try to get rid of the term in $\mathsf{B}$, so as to concentrate on the E-field. To do that we can use the curl operator again, getting

$$
(\operatorname{curl}\operatorname{curl}\mathsf{E}) = -\frac{\partial}{\partial t}\operatorname{curl}\mathsf{B} \tag{6.19}
$$

and then using the last equation in (6.18) to replace $\operatorname{curl}\mathsf{B}$ by a term depending only on $\mathsf{E}$. Thus

$$
(\operatorname{curl}\operatorname{curl}\mathsf{E}) = -\epsilon_0\mu_0\,(\partial^2\mathsf{E}/\partial t^2) \tag{6.20}
$$

and we're halfway there! – with a differential equation involving only one field and a second derivative with respect to the time $t$, just as we had in the general wave equation (6.12).

The second half of the story is more difficult. We'll take it bit by bit in the next Example, which shows how you

can handle equations like these. Skip it on first reading
and come back to it when you feel like trying it!

---

**Example 6.** Whatever does "curl curl" mean?

The left-hand side of (6.20) needs a bit of work, but you can do
it if you try hard. Break it down one component at a time, using
the curl formula, given a few lines earlier, to get the curl of the
vector field $\mathsf{v} = \operatorname{curl}\mathsf{E}$.

Thus, for the x-component,

$$
\begin{aligned}
(\operatorname{curl}\operatorname{curl}\mathsf{E})_x &= \frac{\partial}{\partial y}v_z - \frac{\partial}{\partial z}v_y \\
&= \frac{\partial}{\partial y}\left(\frac{\partial}{\partial x}E_y - \frac{\partial}{\partial y}E_x\right) - \frac{\partial}{\partial z}\left(\frac{\partial}{\partial z}E_x - \frac{\partial}{\partial x}E_z\right).
\end{aligned}
$$

This doesn't look very promising, but it does contain two terms of
the 'del-squared' operator you've already met: they are $(\partial^2/\partial y^2)$
and $\partial^2/\partial z^2$), both with a $-$ sign and working on the field com-
ponent $E_x$. This suggests that you re-group the terms in the last
equation, writing it in the form

$$
\left(-\frac{\partial^2 E_x}{\partial y^2} - \frac{\partial^2 E_x}{\partial z^2}\right) + \left(\frac{\partial^2 E_y}{\partial y \partial x} + \frac{\partial^2 E_z}{\partial z \partial x}\right)
$$

and then start jiggling things around until they look nicer. (You
have to do this a lot in difficult mathematics – when it 'looks nice'
you've probably got it right!)

So let's add a term $-(\partial^2 E_x/\partial x^2)$ in the first parentheses (.....)
and $+(\partial^2 E_x/\partial x^2)$ in the second – so you're adding zero to the
whole thing. The first parenthesis then contains $-\nabla^2 E_x$, while

the second becomes

$$\left( \frac{\partial^2 E_x}{\partial x^2} + \frac{\partial^2 E_y}{\partial y \partial x} + \frac{\partial^2 E_z}{\partial z \partial x} \right).$$

That looks quite nasty; but you may remember from Book 3 that, as long as they work on 'well-behaved' functions, the *order* in which $(\partial/\partial x), (\partial/\partial y)$, etc. are applied doesn't matter (the operators *commute*. The field components are normally of that kind, varying smoothly throughout space, and in that case the last expression can be re-written as

$$\frac{\partial}{\partial x} \left( \frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z} \right).$$

Check this out carefully to be sure it's true! and then you'll see that it's just $\partial/\partial x$ of the *divergence* of the electric field $\mathsf{E}$.

Now you can go back to equation (6.20) and see how far we've come. We were trying to find a wave equation and got stuck with the left-hand side, which was $\operatorname{curl} \operatorname{curl} \mathsf{E}$. But now we know that

$$(\mathrm{A}): \quad (\operatorname{curl} \operatorname{curl} \mathsf{E})_x = \frac{\partial}{\partial x} (\operatorname{div} \mathsf{E}) - \nabla^2 E_x$$

and there will be a similar equation for each of the other components. And if an equation is satisfied by all the vector components it must be satisfied by the vectors themselves, which means that

$$(\mathrm{B}): \quad (\operatorname{curl} \operatorname{curl} \mathsf{E}) = \operatorname{grad} \operatorname{div} \mathsf{E} - \nabla^2 \mathsf{E}$$

– since the derivatives in the first term of (A) are the components of the **gradient** operator, working on the scalar field $(\operatorname{div} \mathsf{E})$

---

The conclusion from Example 6 is an important and general result, applying to any vector field that arises as

209

the gradient of a scalar potential:

$$\operatorname{curl}\operatorname{curl}\mathsf{E} = \operatorname{grad}\operatorname{div}\mathsf{E} - \nabla^2\mathsf{E} \qquad (6.21)$$

That's the end of the difficult bit! We'll use the result straight away to get the final wave equation we've been looking for.

Going back to (6.20) we can now use (6.21) to evaluate the 'curl-curl' term; and we can also use the fact that in free space, which is what we're talking about, $\operatorname{div}\mathsf{E}$ is everywhere zero.

The final result is thus

$$\nabla^2\,\mathsf{E} = \epsilon_0\mu_0\,\frac{\partial^2\mathsf{E}}{\partial t^2}, \qquad (6.22)$$

which means that any component (call it $\phi$) of the electric field $\mathsf{E}$ must satisfy the standard wave equation

$$\nabla^2\,\phi = \frac{1}{c^2}\,\frac{\partial^2\phi}{\partial t^2},$$

with the velocity of propagation $c$ given by

$$c^2 = \frac{1}{\epsilon_0\mu_0}. \qquad (6.23)$$

The permittivity and permeability of free space, $\epsilon_0$ and $\mu_0$, were defined long ago (Sections 1.1 and 3.6) in terms

of experiments in the laboratory (forces on charged particles and between current-carrying wires). They have *experimentally determined* values:

$$\epsilon_0 = 8.8542 \times 10^{-12} \text{ farad per metre}$$

$$\mu_0 = 4\pi \times 10^{-7} \text{ henry per metre}$$

and when you put these into (6.23) you come out with the incredibly large value

$$c = 2.9979 \times !0^8 \text{ metre per second}$$

– very nearly 300 million metres every second! That is also the speed with which light travels (also determined by experiments in the laboratory). But we were talking about *electric and magnetic* fields: now it seems clear that light, whether it comes from the sun or from a white-hot filament in an electric light bulb, is simply an electromagnetic disturbance! As you probably know, it's also the speed with which radio waves travel: all these different kinds of wave travel at exactly the same speed (in empty space), the differences between them being their frequency $\nu$ – or their wavelength $\lambda$, for we know that the two are related by $\nu\,\lambda = c$. These travelling waves are all forms of **electromagnetic radiation** but your eye is only sensitive to radiation whose wavelength falls within a very small range: space is full of radiation of all kinds, with wavelengths going from almost zero to amost infinity and we feel hardly any of it! The entire range is called the **electromagnetic spectrum**.

(So many things are going on in the universe and the most amazing thing of all is that they are all related and all conform to a few basic rules!)

# 6.5 The electromagnetic spectrum

Maxwell's equations have shown that any disturbance of the electromagnetic fields $\mathsf{E}$ and $\mathsf{B}$ will be propagated, even through empty space (containing no electric charges or currents), with the enormous speed $c = 1/\sqrt{\epsilon_0\mu_0}$ in the form of **travelling waves**. Now we have to find out more about the nature of the waves. Then, in Chapter 7, we'll turn to radio waves in particular and ask how they can be produced by *transmitters*, sent out into space, and 'picked up' by *receivers*.

What does an electromagnetic wave look like? Nothing at all, of course, because we can't see it; we only know it's there because the electromagnetic field, at any point in space, is described by giving the components of the two vectors, $\mathsf{E}$ and $\mathsf{B}$, and we can picture *in our minds* the little arrows *representing* the fields at any point. The arrows indicate the forces acting on a moving charge at that point: the $\mathsf{E}$-arrow shows the force arising from the *presence* of other charges, while the $\mathsf{B}$-arrow shows the

force due to their *motion* (i.e. to currents). Some E-arrows are shown in the next Figure. But to get a more complete picture we must do a bit more work.

In the last Section we studied one field component $E_x$ and found that it satisfied a wave equation (6.25) with particular solutions describing **plane waves** travelling along the x-axis. We need a bit more detail: what about the other components $(E_y, E_z)$ and the components $B_x, B_y, B_z$ of the *magnetic* field? From the components we can work out which way the E- and B-vectors point and how big they are.

For a plane wave we can take the direction of propagation as the z-axis, so the wavefront will be a plane, with its **normal** n pointing in the z-direction. The x- and y-axes will then be perpendicular to n. Note that the E-vectors are *transverse* to the direction of propagation (shown by the arrow n on the wavefront at $z = 0$). The three axes are related by the 'right-hand screw rule' (turning the x-axis towards the y-axis would move a corkscrew along the z-axis).
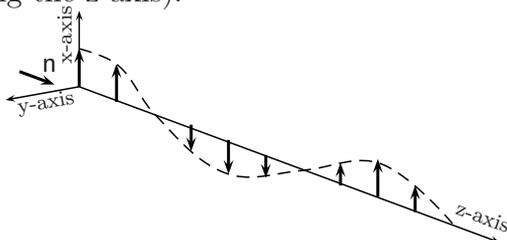


**Figure 41.** Illustrating plane-polarized light

In fact the three vectors $\mathsf{E}, \mathsf{B}, \mathsf{n}$ form a 'right-handed system', with $\mathsf{E}$ pointing along the x-direction – that of **'polarization'** – and $\mathsf{B}$ along the y-direction. Figure 41 shows how you can try to imagine a plane wave travelling in the z-direction! In the Figure, the wave travel along the z-axis, with $\mathsf{E}$ always in the x-direction; and the vector $\mathsf{E}$ is the same at all points in any plane of given $z$ (i.e. on a wave-front). All this follows from Maxwell's equations. We know that in free space div $E = 0$; and in Section 6.2 we found plane-wave solutions with $\mathsf{E}$ perpendicular to the propagation direction. On a wave-front, $\mathsf{E}$ has components $E_x, E_y$ with *constant* values at all points, so

$$\frac{\partial E_x}{\partial x}, \quad \frac{\partial E_y}{\partial x}, \quad \frac{\partial E_x}{\partial y}, \quad \frac{\partial E_y}{\partial y}$$

must all be zero. The first two terms in div $\mathsf{E}$ therefore vanish – and the third must also be zero: $(\partial E_z/\partial z) = 0$. In words, the $\mathsf{E}$-vector can have no component in the propagation direction; it must be transverse.

What about the $\mathsf{B}$-field? Again div $\mathsf{B} = 0$ and a similar argument applies. It shows that $\mathsf{B}$ also has zero component in the propagation direction, but by using the equation curl $\mathsf{B} = \mu_0 \epsilon_0 (\partial \mathsf{E}/\partial t)$ it gives you much more: it shows that when $\mathsf{E}$ at any point on a wave-front is in the x-direction $\mathsf{B}$ points in the y-direction and that as they vary with time, as well as distance along the

z-axis, they stay in phase. That's getting a bit compli-
cated! (Just think about it and try to put these ideas
into a diagram like Fig.41: that's about the best pic-
ture you can get of a plane wave propagating along the
z-axis. If you want the details, try the next Example by
yourself.)

---

**Example 7.** Field vectors for a plane-polarized wave

For a plane wave travelling in the z-direction, polarized in the
x-direction, you can take $E_x = E_0 \cos 2\pi(kz - \nu t)$ and use

$$\text{curl}\, \mathsf{B} = \frac{1}{c^2}\frac{\partial E}{\partial t}$$

to relate the $\mathsf{B}$ and $\mathsf{E}$ vectors.

*Hints* The curl equation for the $\mathsf{B}$-field becomes $-\partial B_y/\partial z =$

$(1/c^2)(\partial E_x/\partial t)$. To obtain $B_y$ in terms of $E_x$ you have to do
one differentiation with respect to $t$ and one integration with re-
spect to $z$. To make things easier, use the phase $\theta = 2\pi(kz - \nu t)$
as a new variable. (You may need to look at Book 3.)

---

### Different kinds of electromagnetic waves

What's the difference between one kind of electromag-
netic wave and another? The main differences arise from
the frequency $\nu$ or the wavelength $\lambda$: you can use either
one in describing the wave, as $\lambda\mu = c$ (the velocity of

215

light, $\approx 3 \times 10^8$ m s$^{-1}$). One complete oscillation per second is called '1 Hertz' (1 Hz) and the observable part of the spectrum extends from about $10^5$ Hz up to more than $10^{20}$ Hz. This corresponds to wavelengths running from about 3 km down to far less than 'atomic' dimensions.

To make a 'picture' of the whole spectrum one usually works in powers of 10, adopting a **logarithmic scale**. Figure 42 shows the frequency spectrum in this form, labelling different ranges with their conventional names.
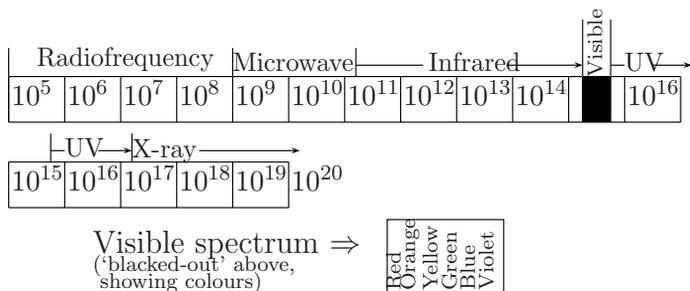


**Figure 42.** The electromagnetic spectrum
(frequency range from $10^5$ Hz to $10^{20}$ Hz)

Starting from the long-wave end ($\nu = 10^5$ Hz, $\lambda \approx 3$ km), the Radiofrequency range (used in commercial radio broadcasting) is followed by the Microwave range (used in microwave ovens), which goes from about $10^9$ to $10^{11}$ Hz. There it enters the Infrared (IR) range,

corresponding to heat radiation from fires or white-hot lamp filaments. This is followed by the Visible range, the only part of the spectrum your eye can detect, marked in Fig.42 as the narrow band labelled "visible" – 'blacked-out' to show how small it is compared with the whole spectrum. The visible band is shown below in magnified form, and contains all the colours you can see, going from red (on the left) to violet (on the right). The mixture of all colours is "white light", which is ordinary 'daylight'. Beyond violet, as the frequency increases, comes the Ultra-Violet (UV) range; like the Infra-Red (IR), radiation in this range is invisible, but both IR and UV are absorbed by your skin, IR being felt as heat, UV producing 'sunburn' and (if you get too much of it) serious damage. Above frequencies of about $10^{17}$ Hz comes the X-Ray range, of radiation that can pass through your body and show (on a photographic plate) shadows of your internal organs. At still higher frequencies there are Gamma($\gamma$)-rays, produced by 'radioactive' chemical elements like radium, and finally Cosmic rays, which arrive from other parts of the Universe, when stars explode and 'die'. When we go outside and are not protected by clouds, tents, buildings and other materials that can absorb radiation, we get it all! Lucky for us that most of all this radiation is very weak by the time it reaches us.

**Radiation and energy flow**

You'll be wondering what happens to all that radiation

which the Universe is so full of. It's only just now that we've started talking about radiation being *absorbed*; but that's because we've been thinking only of empty space, with no atoms or molecules or matter of any kind. In order to include things like that in our model of the Universe, we have to think about the **interaction** between radiation and matter. That's a very difficult business, but usually we can manage quite well with simple approximations. The first rule is to deal separately with the things that we know about: *radiation* in empty – or *nearly* empty – space (which is what Maxwell's equations are about); and *particles* of matter, from which atoms and molecules can be built up (as we've seen in Book 5); and finally think about the ways they can interact, using the Physics we know from Book 4.

Radiation is a form of **energy**. In Chapter 1 we saw how the energy of the electric field could be 'divided out' over all space, with so much per unit volume at any given point: that amount is the **energy density of the electric field**. And equation (1.25) gave us a simple formula for getting this amount: $u_{\text{elec}} = \frac{1}{2}\epsilon_0 E^2$. We 'guessed' this formula, on the basis of a simple example – a parallel plate condenser holding electric charge; but it turns out to be true for any kind of field arising from static charges.

Of course in radiation the field includes a *magnetic* part B; and both fields vary with time. But (although we

didn't derive it in Chapter 3) we noted a very similar formula for the energy density of the magnetic field due to a distribution of steady currents, namely

$$u_{\mathrm{mag}} = \tfrac{1}{2}\bar{\mu}_0 B^2, \qquad (6.24)$$

where $\bar{\mu}_0$, defined in equation (3.4) of Ch.3, is the reciprocal of the conventional magnetic permeability $\mu_0$ and corresponds more closely to $\epsilon_0$ in electrostatics.

But one of the most amazing things to come out of Maxwell's equations is an expression for the energy density of electromagnetic radiation, which is universally true and looks just like the sum of the two terms for static electric and magnetic fields respectively:

$$u = \tfrac{1}{2}\epsilon_0 E^2 + \tfrac{1}{2}\bar{\mu}_0 B^2. \qquad (6.25)$$

A second result follows if you calculate the rate at which field energy comes out of any volume throughout which the $\mathsf{E}$ and $\mathsf{B}$ vectors are defined. You may remember that this kind of **flux** follows (remember Gauss's theorem back in Chapter 1) by taking the *divergence* of the scalar quantity $u$. This is difficult stuff, but the result is very simple. The energy flow is represented by a vector $\mathcal{S}$, called the **Poynting vector**, whose normal component $\mathsf{S} \cdot \mathsf{n}$ gives the field energy flowing across unit area, per unit time, in the direction of $\mathsf{n}$. The Poynting vector is defined by

$$\mathcal{S} = \bar{\mu}_0 \mathsf{E} \times \mathsf{B}. \qquad (6.26)$$

Let's take one example to show how things work out.

---

**Example 8.** Energy flow in a light wave

For the plane wave indicated in Fig.41, with $E_x = E_0 \cos 2\pi(kz - \nu t)$, the B and E vectors are perpendicular and their magnitudes are related by $B = (1/c)E$. Thus

$$\mathcal{S} = (\epsilon_0 c^2)(E^2/c) = \epsilon_0 c E^2.$$

The field is rapidly changing however, while the value just given is *instantaneous* – with $E_x$ going up and down as in Fig.40 – and the average value of $E_x^2$ over a complete cycle is just half the amplitude $E_0$ squared (check it by doing the integration). Finally, then,

$$\langle \mathcal{S} \rangle_{\text{av}} = \tfrac{1}{2}\epsilon_0 c E_0^2$$

and this average rate of energy flow is the same at all points in the beam of light.

---

To end this chapter let's come back to the question of where does all that radiation go? It seems it should go on forever, in all directions, once it's been started off. But, as we guessed earlier, that's only because we've been thinking of what happens in *empty space*; and in empty space there's nothing to stop it.

Here on Earth there are plenty of material objects that can interact with radiation and take in some of its energy. It's not possible to fully understand what happens until you've studied more Physics. but everyday experi-

ence tells you a lot about the **absorption** of radiation. If you're in a darkened room and let a beam of sunlight come in through a small hole you'll feel the energy flow (the Poynting vector!) when it falls on your hand. The light that doesn't get into the room is absorbed by the shutters – and they get hot from the energy, especially that in the infra-red part of the spectrum, when they take it in. Where does that energy go? If you look back at Book 5 you'll remember that **heat** can be taken in by atoms and molecules and can be stored as the mechanical energy associated with vibrational and rotational motions. You also learnt that much of that energy is 'lost' forever in the sense that we can never get it back in a form that can do useful work. We say the energy has been 'dissipated' or 'degraded'. These ideas are fundamental to a large part of Science; they are built into the principles of Thermodynamics and led us in Book 5 to the concept of **entropy**; and they show us that all natural changes take place in such a way that the entropy of the world inreases; they make us think about time and the way the world is 'running down' like a clock!

In real space, even outside the Earth's atmosphere, there are countless millions of atoms and molecules that can interact with any radiation that comes their way. So there's little danger of the Universe getting overcrowded with all the radiation we're producing. In the next, and

last, chapter we'll start thinking about *how* electromagnetic radiation can be produced and used, in giving us the means of communication through **radio and television**.

# Chapter 7

# How to make electromagnetic waves – radio communication

## 7.1 Electric circuits: more circuit elements

Maxwell's equations have shown that any disturbance of the electromagnetic fields $\mathsf{E}$ and $\mathsf{B}$ will be propagated, even through empty space (containing no electric charges or currents), with the enormous speed $c = 1/\sqrt{\epsilon_0\mu_0}$ in the form of **travelling waves**. In this last chapter we'll be asking how such waves can be set up;

and we'll be thinking mostly of those included in the radio-frequency range (Fig.42), which are produced by using electric circuits.

As you've probably guessed, the waves can be created by rapidly moving electric charges in a special form of conductor called an **antenna**; often this consists of a vertical metal bar or tube, held up high in an open space (e.g. on a hill top) with no buildings or other nearby obstacles that would absorb the radiation coming from it. The antenna must of course be insulated from everything except the wires that supply it with electric power by sending charge into it. When charge moves from one end of the antenna to the other an electric dipole is created and when the charge goes backwards and forwards you get an **oscillating dipole**. This creates an electromagnetic field, oscillating with the same frequency as the dipole, and this is the field that propagates outward with the speed of light $c$. The only thing we have to do now is produce the oscillating dipole. In Section 1.3 we studied some simple circuits, containing **circuit elements** such as resistances (R), capacities (C) (often called by their older name "condensers"), and inductances (L), connected by wires: these are 'passive' elements, since they only allow current to pass through them. Other circuit elements, like batteries and generators, may actually *produce* currents by setting up a potential difference between their terminals. This dis-

continuity of potential (or voltage, $V$) arises from the emf which urges charge from one terminal towards the other: in Chapter 2 the PD was produced by a battery and was constant, giving a steady **direct current** (DC); but in Chapter 5 it was produced by a generator and varied like a sine or cosine function (see Fig.32), giving an **alternating current** (AC). We'll suppose the generator produces an output voltage

$$V = V_0 \cos \omega t, \qquad (7.1)$$

which will be applied to various circuits.

The passive circuit elements may behave in different ways, depending on whether the current passing through them is DC or AC. Now we're going to deal with oscillating currents, so we need to know their AC behavior. Let's take them one at a time, starting with resistance R, and take the case of 'slow' variation (valid for all the frequencies we need to deal with).

**Resistance (R).**

When the PD across a resistance is constant, with value $V$, the relationship between current and voltage is given by Ohm's law as $V = RI$. For alternating currents the same law holds good, provided $V$ and $I$ are corresponding *instantaneous* values of voltage and current.

In the AC-case, however, the instantaneous values of $V$ and $I$ are not very interesting because they're always

varying between extreme values $\pm V_0$ and $\pm I_0$, respectively, often with a very high frequency. What matters, usually, is an *average* value, taken over a whole cycle of possible values. For example, the power dissipated as 'Joule heat' when a generator sends current through a resistance is given in (5.5) as

$$P = \mathrm{d}W/\mathrm{d}t = VI = I^2 R \qquad (7.2)$$

and, on substituting the alternating values of voltage and current, this becomes

$$P = (V_0 \cos \omega t)(I_0 \cos \omega t) = (V_0^2/R) \cos^2 \omega t.$$

The time average of this quantity, indicated as $\langle P \rangle$ is then

$$\langle P \rangle = (V_0^2/R)\langle \cos^2 \omega t \rangle.$$

To evaluate the mean value of $\cos^2 \omega t$ we remember the identity $\cos 2\theta = 2 \cos^2 \theta - 1$ (Read again Chapter 4 of Book 2.) and write $\cos^2 \omega t = \frac{1}{2}(\cos 2\theta + 1)$. The average of the cosine term over a whole cycle of values is zero (look at its graph), while that of the constant term is just $\frac{1}{2}$. Thus,

$$\langle P \rangle = \frac{V_0^2}{R}\frac{1}{2} = \frac{V_{\mathrm{rms}}^2}{R}, \qquad (V_{\mathrm{rms}} = V_0/\sqrt{2}) \qquad (7.3)$$

where, instead of the *maximum* voltage $V_0$, we've introduced the **root-mean-square** value $V_{\mathrm{rms}}$. In working

with AC circuits the output voltage from a generator is usually taken to mean $V_{\text{rms}}$, so if you say "a 100 volt AC generator" you mean one in which the voltage oscillates between $\pm 100\sqrt{2}$ volts i.e. between $\pm 141.4$ volts.

## Capacity (C).

We've already done the work in earlier Sections! In Section 4.2 we introduced the idea of "displacement current" by studying the charging and discharging of a condenser when the applied voltage across its plates was changing. When the condenser is connected to an AC generator, as in Figure 43, there is a displacement current arising from alternately charging and discharging the condenser – the current changing direction when the voltage reverses. That doesn't happen when you use a battery instead of the AC generator: in that case the constant PD across the plates produces a current for a very short time, until the condenser is fully charged; but then it stops when the voltage comes up to that of the battery.
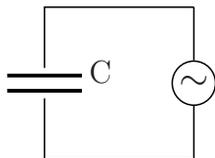


**Figure 43.** AC circuit with condenser C

The condenser plates, on the left in Fig.43, are con-

nected to an AC generator. With *direct* current, instead, the steady state is one in which there is *no current flow*: the condenser acts like an insulator. Back in Section 2.3 we discovered that on connecting the plates of a charged condenser with a wire a current *did* flow, but only for a very short time – until the charge had disappeared and the condenser was fully *discharged*. In fact, at time $t$ after making the connection, the charge left was only

$$Q = Q_0 \exp\left(-\frac{1}{RC}t\right),\qquad(7.4)$$

where $Q_0$ is the charge at $t = 0$, $R$ is the resistance of the connecting wire, and $C$ the capacity of the condenser. If you really want to discharge the condenser fast you make $R$ small (e.g. use a short piece of thick copper wire – after disconnecting the battery, or you'll discharge that too!); the fraction in front of the $t$ is then very big and the exponential falls rapidly to zero.

Now, however, we want to know what happens when a condenser is connected to a source of *alternating* voltage? This will produce an alternating current (AC) as the condenser alternately charges and discharges. And if the condenser is going to be used in an AC circuit we need to know how voltage and current will be related. For a simple resistance, $V$ and $I$ are related by Ohm's law $V = IR$ and this remains so even when both voltage and current depend on time – as long as we use

their *instantaneous values*. But this simple proportionality may not apply to other circuit elements. In the two Examples that follow we'll suppose a condenser, or a coil, is connected to an AC generator (see Chapter 5) which produces a voltage (7.1), namely $V = V_0 \cos \omega t$. This goes up and down between $V_0$ and $-V_0$ and has the same form as in Fig.40, of Chapter 6 (second picture). Here $V$ is plotted on the y-axis and $\omega t$ along the x-axis. (Remember that $\omega t$ is called the **phase** of the variation and $V$ runs through all possible values between $\pm V_0$ as the phase goes from 0 to $2\pi$.)

---

**Example 1.** Capacity (C) in an AC circuit

If the charge on the positive plate of a condenser is $Q$, at time $t$, the current flowing into it will be $I = \mathrm{d}Q/\mathrm{d}t$. On the other hand we know from (1.20) that the charge is related to the voltage between the plates by $Q = CV$, the constant $C$ being the capacity of the condenser. Thus

$$I = \frac{\mathrm{d}Q}{\mathrm{d}t} = C \left( \frac{\mathrm{d}V}{\mathrm{d}t} \right).$$

This is a **differential equation** whose solution leads to both $I$ and $V$ as functions of the independent variable $t$. When you first make the connection ("switch on") $I$ jumps up and down a bit, but things soon settle down into a **steady state** – which is what we're looking for. The steady-state solution is easy to find: differentiating the generator voltage $V = V_0 \cos \omega t$ gives (look back at Book 3 if you need to) $\mathrm{d}V/\mathrm{d}t = -\omega V_0 \sin \omega t$ and the

equation for $I$ is therefore satisfied by

$$I = -\omega V_0 C \sin \omega t.$$

The current going in thus varies with the same angular frequency as the voltage; but $I$ is not directly proportional to $V$, as it would be with a simple resistance $R$ in place of the capacity.

To understand this result it's enough to look back at the properties of the 'circular' or 'trigonometric' functions you first met in Book 2: at the end of Chapter 4 you found that for any two angles, $A$ and $B$, the sine and cosine of their sum could be expressed in terms of the sines and cosines of the separate angles. For example,

$$\cos(A + B) = \cos A \cos B - \sin A \sin B.$$

You also know that $\cos \pi/2 = 0, \sin \pi/2 = 1$ and that that's all you need: think of $\omega t$ as the angle $A$ and take $B = \pi/2$. The formula then gives

$$\cos(\omega t + \pi/2) = -\sin \omega t$$

and, on putting this result in the current equation we just found, it becomes instead $I = \omega V_0 C \cos(\omega t + \pi/2)$.

---

The conclusion from Example 1, namely

$$I = \omega V_0 C \cos(\omega t + \pi/2), \qquad (7.5)$$

can thus be put in words as

> The current passing through a capacity $C$, in
> an AC circuit, has the same cosine-variation

as the voltage across its plates, except for a
**phase shift** in which $\omega t$ is changed to $\omega t + \pi/2$.

The time variation shows up nicely in a 'vector diagram'
where $V$ is shown as a vector, of length $V_0$, rotating
around the origin with angular velocity $\omega$, so that at
time $t$ it makes an angle $\theta_V = \omega t$ with the x-axis. The
current $I$ is then represented by a vector of length $I_0$,
making an angle $\theta_I = \omega t + \pi/2$ with the x-axis. The
actual values of the voltage and current, at time $t$, are
given as the *projections* of the two vectors on the x-axis.
The situation is shown in Figure 44:
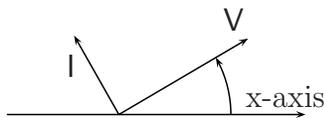


**Figure 44.**
Vector diagram for AC circuit in Fig.43

The vectors V and I rotate anti- clockwise with angular
frequency $\omega$.

Note that the phase of the current *leads* that of the volt-
age by $\pi/2$, so that when the voltage reaches its maxi-
mum value $V_0$ the current (which is putting charge into

231

the condenser) has already fallen to zero – the condenser being fully charged at that instant. Also, in terms of maximum values, $I_0 = \omega C V_0$, whereas for a circuit containing only a resistance Ohm's law gives $I_0 = V_0/R$. Thus, $(1/\omega C)$ plays the part of an 'effective resistance'. At the same time, current and voltage are no longer **in phase**.

**Inductance (L)**

Let's next look at the situation when the condenser of capacity $C$ is replaced by a coil of inductance $L$. In this case the emf of the generator will immediately (at $t = 0$) begin to force current round the circuit, the coil being a continuous wire connecting the output terminals. The new circuit is shown in Figure 45 below:
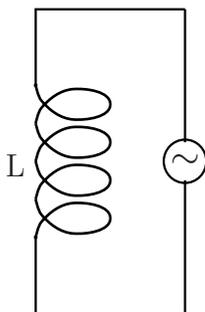


**Figure 45.**
AC circuit with coil L
Coil (left) connected
to generator (right).

In this case we know from Section 5.4 that when the current through a coil is changing it sets up a "back-emf", which is proportional to the rate of change, $dI/dt$, and

*opposes* the emf $V$ applied to the terminals. The proportionality constant was called the "self-inductance" of the coil and was denoted by $\mathcal{L}$: we'll now just use $L$ and write the voltage change between the terminals of the coil as $-L(\mathrm{d}I/\mathrm{d}t)$. Since the sum of the voltage changes in going round any closed circuit is zero (Kirchoff's first law in Section 2.3) the circuit diagram in Fig.45 shows that

$$V - L\frac{\mathrm{d}I}{\mathrm{d}t} = 0. \tag{7.6}$$

---

**Example 2.** Inductance (L) in an AC circuit.

When the generator applies the voltage (7.1) substitution in (7.6) leads to the differential equation

$$L\frac{\mathrm{d}I}{\mathrm{d}t} = V_0 \cos \omega t,$$

which can be integated immediately (Chapter 3 of Book 3). For any function $y = \sin ax$, we know the 'standard form' $(\mathrm{d}y/\mathrm{d}x) = a\cos ax$ and this tells us that $y = \int a\cos ax\,\mathrm{d}x$. But the result above then says that $\int (V_0/L)\cos \omega t\,\mathrm{d}t = I$; so on changing the names of the variables $(y \to I,\ x \to t)$ we find $I = (V_0/L\omega)\sin \omega t$. Going ahead just as in Example 1 (do it!), then leads finally to $I = (V_0/L\omega)\cos(\omega t - \pi/2)$.

---

The result from Example 2 is thus

$$I = \frac{V_0}{L\omega}\cos(\omega t - \pi/2) \tag{7.7}$$

and the relationship between the applied AC voltage
(7.1) and the current through the coil, $I$, is then clear
from Figure 46:
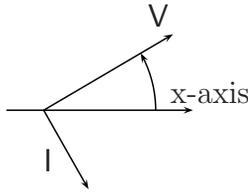


**Figure 46.**
Vector diagram for AC circuit in Fig.45

The vectors V and I in Fig.46 rotate anti-clockwise with
angular frequency $\omega$.

Maximum values of current and voltage are then related
by $I_0 = V_0/\omega L$, as if the inductance were an 'effective
resistance' of magnitude $\omega L$. And again there is a phase
difference of $\pi/2$, though this time the current *lags* be-
hind the voltage by $\pi/2$. When the voltage across the
coil has its maximum value, the V-vector – of length $V_0$
– pointing along the +ve x-axis, the current through it
is only just starting to grow (the I-vector pointing ver-
tically down, with zero projection on the x-axis).

## 7.2 A cool way of doing it: complex numbers

For more complicated circuits there'll be more and more equations to set up and solve – but don't get into a sweat, thinking about all that trigonometry! There's an easy way of avoiding it.

In Figures 44 and 46, the voltage and current were represented as *vectors* rotating anti-clockwise around the origin in a vector diagram with angular velocity $\omega$ radians per second; but the actual voltage and current at any instant were measured by the *projections* of V and I on the x-axis. This may have reminded you of **complex numbers**, which you first met long ago in Books 1 and 2. (Look again at Section 5.2 in Book 1 and at Chapter 4 in Book 2. Just to remind you, the following note will help to get you through.)

*A note on complex numbers*

---

A 'complex' number is a combination of 'ordinary' (real) numbers and a 'new' number, called the **imaginary unit** and denoted by $i$, with the property $i^2 = -1$. When $x, y$ belong to the field of real numbers the combination $x + iy$ belongs to the field of complex numbers. Usually the letter $z$ is reserved for a complex number and when $z = x + iy$ $x$ and $y$ are called the "real and imaginary parts" of $z$. If $x$ and $y$ are measured along the usual x- and y-axes, then $z$ may be pictured as a vector in an **Argand diagram**,

with $x$ and $y$ as the horizontal and vertical *components*. This way of representing a complex number in a diagram is described as **cartesian** (as in plane geometry). Figure 47 shows the representation of $z$ as a vector of length $r$, with components $x$ (horizontal axis) and $y$ (vertical axis). The length of the vector is called the **modulus** of $z$, denoted by $|z|$.

In Chapter 4 of Book 2, another way of describing the complex number was also used: instead of giving the cartesian components of the vector $(x, y)$ the same object is described by giving its length $(r)$ and its inclination $(\theta)$ to the x-axis. This is described as the **polar** form of $z$ (corresponding to polar coordinates in plane geometry). The $(r, \theta)$ form is also clear from Fig.47, along with the relations

$$x = r\cos\theta, \quad y = r\sin\theta, \quad |z|^2 = x^2 + y^2.$$
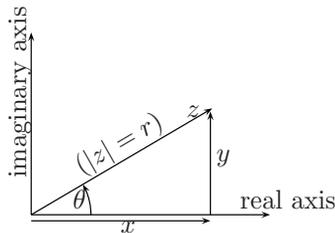


**Figure 47.** Argand diagram for $z = x + iy$

In the Figure $z$ is represented as a vector of length $|z| = r$ The real part $(x)$ is the projection of $z$ on the x-axis, while the imaginary part $(y)$ is that on the vertical axis.

To show how things work out, let's look again at what we did in the last Section; and then go on to study circuits that can produce *oscillating* currents and voltages.

Look again at Fig.47 and suppose the angle $\theta$ is set equal to $\omega t$: in that way you can put in a *time factor* which will make the complex number $z$ (or rather the vector representing it) rotate as time passes. But in Fig.44, for example, we repesented an alternating voltage by a vector $\mathsf{V} = \mathsf{V}_0 \cos \omega t$, such that the actual voltage at any instant was the *projection* of the vector on the x-axis. Now it's clear that we could equally well represent the voltage by a complex number $V = V_0 e^{i\omega t}$ and use its *real part* to represent the voltage we actually measure between the terminals of the generator. The nice thing about complex numbers is that they obey the same basic rules as real numbers (that's how they were invented!) and can therefore be handled using the normal rules of mathematical analysis (see for example Book 3).

Suppose we now take

$$V = V_0 e^{i\omega t}, \quad I = I_0 e^{i\omega t}, \quad (7.8)$$

as the (complex) quantities standing for voltages and currents at any points in an AC circuit, $V_0$ and $I_0$ being independent of time. The circuit elements we've studied already are resistances (R), capacities (C), and inductances (L). Let's look at them again:

## Resistance (R)

When current $I$ passes through a resistance $R$ the voltage (PD) set up between its terminals is $V = RI$, where $R$ is a simple proportionality constant. This is not changed in any way if the current is alternating, provided *instantaneous* values are used; so if time factors are included it makes no difference and we can continue to use the complex quantities in (7.8). As time passes these rotate with the same factor $e^{i\omega t}$ and thus always point in the same direction: $V = RI$.

## Capacity (C)

When the circuit element is a condenser of capacity $C$ the current is related to the building up of charge on the plates; $I = \mathrm{d}Q/\mathrm{d}t$ and we know that the charge at time $t$ is proportional to the voltage, with proportionality constant $C$, so the basic equation is $I = C\mathrm{d}V/\mathrm{d}t$.

Now we take account of the time variation by using the complex quantities in (7.8). No more sines and cosines, we just do the differentiation in the last equation, getting

$$I = CV_0(i\omega)e^{i\omega t} = i\omega CV.$$

This looks a bit like Ohm's law, but instead of $V = RI$ we have

$$V = \left(\frac{1}{i\omega C}\right) I = \left(\frac{-i}{\omega C}\right) I$$

(after muliplying top and bottom by $i$, with $i^2 = -1$).

The factor in parentheses, which plays the part of some kind of 'effective resistance', is the **complex impedance** of the condenser: we denote it by

$$Z_C = -i(1/\omega C).\qquad\qquad (7.9)$$

But what does a complex impedance mean? Before thinking about that let's use the complex representation of current and voltage in (7.8) to deal with a coil of inductance $L$.

**Inductance (L)**

Remember, when the output from an AC generator is applied to a coil, as in the circuit shown in Fig.45, the changing magnetic flux results in a 'back emf' which opposes the applied voltage $V$. By Kirchoff's first law, the sum of the PDs round the closed circuit must be zero and this gives the basic equation $V = L(\mathrm{d}I/\mathrm{d}t)$. When voltage and current are oscillating as, in (7.8), it follows that

$$V = L\frac{\mathrm{d}}{\mathrm{d}t}I_0 e^{i\omega t} = LI_0(i\omega)e^{i\omega t} = i(\omega L)I$$

Thus, $V = (i\omega L) = Z_L I$, where

$$Z_L = i\omega L\qquad\qquad (7.10)$$

is the complex impedance of an inductance $L$. Now we can come back to the question of what it all means.

## Interpretation of a complex impedance

Start from the example we've just looked at, namely the inductance of a coil. The complex impedance has the property that, when current $I$ passes, it produces a voltage $V = Z_L I = i\omega L I$ between the terminals of the coil (Fig.45). The real factor $(\omega L)$ simply multiplies the modulus of $I$, leaving its direction unchanged – but what does the factor $i$ do?

If you multiply any complex number $z = x + iy$ by $i$ it changes into $z' = ix - y$; and if you draw the vectors representing $z$ and $z'$ you'll find (check it!) that $z'$ looks like $z$ *rotated by* $+90°$. In the same way, multiplying $z$ by $-i$ rotates it by $-90°$. Figure 48 shows the effect of multiplying the $z$ in Fig.47 by $i$.
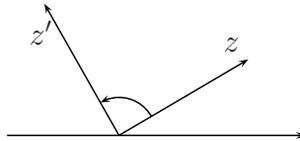


**Figure 48.** Effect of muliplying $z$ by $i$
Any $z$ is rotated anticlockwise by $90°$
without change of modulus ('length').

Having dealt with resistance, capacity, and inductance, we're ready to go ahead and study the way any kind of circuit will work when it's connected to a source of AC power – *however these three kinds of circuit elements*

*may be connected together!* But before going on you may want to be sure that using complex numbers leads to the same results as before – summarized in Figs.44 and 46.

In defining the complex impedance $Z$ we've supposed that the voltage across a circuit element is related to the current passing through it by $V = ZI$, so that $Z$ will look like some kind of 'effective' resistance, replacing the $R$ in Ohm's law $V = RI$. But we started this Chapter by thinking of $V$ (the generator voltage) as the independent variable, taking $I$ as the dependent quantity and writing $I = \text{constant} \times V$.

Now that we're using the impedance $Z$, the current/voltage relationship would be written as $I = Z^{-1}V$. The reciprocal of the impedance is called the **admittance** and sometimes it's convenient to put things this way round – as long as you don't get mixed up! The following Example shows how you can get back the results in Figs.44 and 46.

---

**Example 3.** Using complex impedance – capacity and inductance

To write $I$ in terms of $V$, so that $I = Z^{-1}V$, we note that:

for capacity $C$ $\quad Z_C{}^{-1} = i\omega C,$ $\qquad$ for inductance $L$ $\quad Z_L{}^{-1} = \dfrac{-i}{\omega L}.$

The effect of $Z_C{}^{-1}$ on a voltage vector $V$ is thus to multiply its

length $V_0$ by $\omega C$ and then to rotate it by 90°, as shown in Fig.44 – giving the current $I$ in (7.5). The current *leads* the voltage by $\frac{1}{2}\pi$.

On the other hand, the effect of $Z_L{}^{-1}$ on a voltage vector $V$ is to multiply its length $V_0$ by $(1/\omega L$ and then to rotate it by $-90°$, as shown in Fig.46 – giving the current $I$ in (7.7). The current *lags* behind the voltage by $\frac{1}{2}\pi$.

In short, representing the current and voltage vectors by complex numbers gives results in perfect agreement with those found in Section 7.1 – only the *language* is a bit different.

---

Now we know how to deal with simple circuit elements, like capacity and inductance, using complex numbers. But suppose we have a box, with two terminals, and something inside it that we don't know about: all we can say is that it's an impedance which can be represented by a complex number $Z = X + iY$. This simply means that when a voltage $V$ is applied across the terminals a current $I$ will flow and that the voltage-current relationship will be

$$V = ZI, \quad (Z = X + iY, \ X \text{ and } Y \text{ real}). \qquad (7.11)$$

In an AC circuit, the current and voltage will contain a complex factor $\exp i\omega t$ and $ZI$ will thus be a product of two complex numbers, $Z$ usually being expressed in Cartesian form, but $I$ in the polar form $I = I_0 \exp i\omega t$. Figure 47 shows how such products can be handled: the

simplest way is to put both factors in *polar* form, for if

$$z_1 = r_1 \exp i\theta_1, \quad z_2 = r_2 \exp i\theta_2,$$
$$\text{then } z_1 z_2 = r_1 r_2 \exp i(\theta_1 + \theta_2). \qquad (7.12)$$

In other words, to multiply two complex numbers you must multiply their moduli (or 'lengths'), $r_1 = |z_1|$ and $r_2 = |z_2|$, and then give the result a 'phase factor' $\exp i\theta$ formed by *adding* their phases $(\theta = \theta_1 + \theta_2)$.

To get the moduli and the phases from the Cartesian forms is easy if you look at Fig.47 and remember your geometry:

$$r^2 = |z|^2 = x^2 + y^2, \qquad \tan\theta = y/x. \qquad (7.13)$$

In later Sections there will be Examples to show how these methods can be used.

# 7.3 Connecting circuit elements together

In Section 2.3 we first met some simple circuits in which various circuit elements were connected together by ideal wires, in such a way that on connnecting the ends of the circuit to a DC voltage source (such as a battery) current could flow through all the elements. In the Examples we studied, the only 'passive' elements were resistances and condensers; and the currents were usually

taken to be steady – not changing with time. A complicated **network** could be set up by connecting many elements in many diferent ways, but two kinds of connection were of special importance. For instance, resistances $R_1, R_2, R_3, ...$) could be connected **in series**, where the same *current* $I$ passes through each one in turn; or **in parallel**, where the different elements are 'side-by-side' but with the same *voltage* drop $V$ between their ends. The two cases are indicated in the Figure below:



(a) Resistances in series     (b) Resistances in parallel
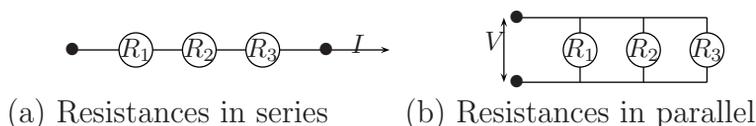
**Figure 49**

The results we found for the two cases (a) and (b) were that

(a) resistances connected in series are equivalent to a single resistance

$$R = R_1 + R_2 + R_3 + ...;$$

(b) resistances connected in parallel are equivalent to a single one with $R$ given by

$$(1/R) = (1/R_1) + (1/R_2) + (1/R_3) + ....$$

244

What we want to show next is that, on replacing the DC power supply (usually a battery) by an AC generator and the resistances by impedances $Z_1, Z_2, Z_3, ...$ of any kind, Kirchoff's laws still apply and allow us to calculate the currents and voltages *throughout the network.* The networks illustrating series and parallel connections will then be as in the Figure below. (Note that the "impedances" we're talking about are usually *complex*, so we won't always say so.)



(a) Impedances in series        (b) Impedances in parallel
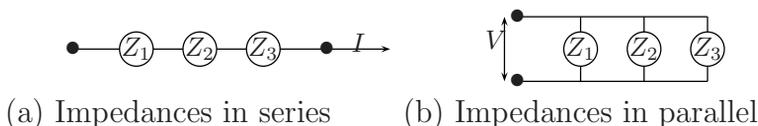
**Figure 50**

In (a) the circuit elements are separated by **nodes**, where they are connected together, and the current going into any node must be the same as that going out (otherwise charge at that point would build up continually without limit); that's often called Kirchoff's second law. The same current $I$ must therefore go through every circuit element, giving $V_1 = Z_1 I$, $V_2 = Z_2 I$, and so on. And the total voltage drop in following the current round the circuit must equal the voltage $V$ applied across the terminals. That means

$$V = V_1 + V_2 + \ ... \ = Z_1 I + Z_2 I + \ ... \ = (Z_1 + Z_2 + \ ...)I.$$

245

But the impedance $(Z)$ of the whole circuit is defined so as to make $ZI = V$. It follows that

$$Z = Z_1 + Z_2 + \; ... \quad \text{(impedances in series)} \qquad (7.14)$$

– the impedances must be added together, just as if they were simple resistances, but now they are complex numbers!

In (b), every circuit element feels the same applied voltage $V$; but the current passing through it is related to $V$ by $V = Z_1 I_1$, for the first element, by $V = Z_2 I_2$ for the second, and so on. The total current going into the circuit at the top node in Fig.50(b) (and coming out at the bottom) is then

$$I = I_1 + I_2 + \; ... \; = \frac{V}{Z_1} + \frac{V}{Z_2} + \; ... \; = V\left(\frac{1}{Z_1} + \frac{1}{Z_2} + \; ... \right).$$

But the impedance $Z$ of the whole circuit is defined by $V = IZ$; so $I$ on the left-hand side of the sum can be replaced by $I = V/Z$. The last equation then tells us (cancelling the common factor $V$) that

$$\frac{1}{Z} = \left(\frac{1}{Z_1} + \frac{1}{Z_2} + \; ... \right) \quad \text{(impedances in parallel)}$$

$$(7.15)$$

– again just as if the impedances were simple resistances.

As a first example of connecting impedances let's think about an AC generator, forcing current $I$ through a resistance $(R)$, a condenser (of capacity $C$) and a coil (of

inductance $L$), all connected in series. The whole circuit will be as shown below (Figure 51).
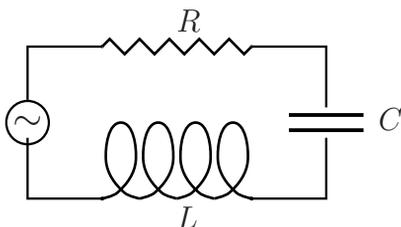


**Figure 51.**
AC circuit (R-C-L)
(generator on left)

The impedances are all in series; so they will be equivalent to a single impedance $Z = Z_R + Z_C + Z_L$, with $Z_R = R$ (the usual *real* resistance), $Z_C = -i/\omega C$ by (7.9) and $Z_L = i\omega L$ by (7.16). The impedance of the whole circuit is thus

$$Z = R + i\left(\omega L - \frac{1}{\omega C}\right), \qquad (7.16)$$

which has the form $Z = X + iY$, where $X$ and $Y$ are the real quantities

$$X = R, \qquad Y = \left(\omega L - \frac{1}{\omega C}\right).$$

And now we have everything we need to know!

The applied voltage from the generator is $V = ZI$ and all quantities are complex; but remember that $I$ is of the form $I_0 e^{i\theta}$ and only the real part of $I$ has any physical meaning, the phase factor just making $I$ go round and round in our vector diagram, as its real part varies between $\pm I_0$.

Now what does $Z$ do when it multiplies $I$? The answer is given in (7.12): you multiply the modulus of $I$, namely $|I| = I_0$, by the modulus of $Z$; and you change the phase angle $\theta = \omega t$, contained in $I$, to $\theta + \phi$ where $\phi$ is the phase of $Z$.

When $Z$ has the form given above in (7.16), as $X + iY$, it's easy to find both $|Z|$ and $\phi$ (look at Fig.47 if you need reminding how): here $|Z|^2 = X^2 + Y^2$ (Pythagoras!) and $\tan\phi = Y/X$. Thus

$$|Z|^2 = R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2, \qquad (7.17)$$

$$\tan\phi = \frac{1}{R}\left(\omega L - \frac{1}{\omega C}\right). \qquad (7.18)$$

It's easy to visualize our conclusions by drawing a vector diagram to show the relationship $V = ZI$.
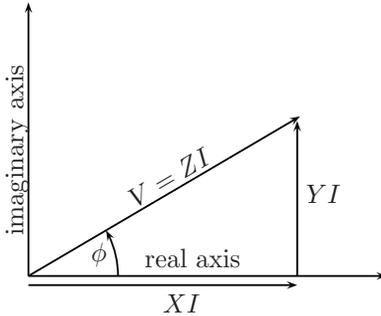
**Figure 52.** Vector diagram for $V = ZI$
V is for R-L-C circuit (Fig. 51)
$V = ZI$ = voltage vector
$XI$ = projection on real-axis,
$YI$ = that on vertical axis.

In the Figure, I points along the x-axis, corresponding
to $t = 0$ and phase $\theta = \omega t$ also zero; its modulus is the
maximum value $I_0$ of the current as time passes and the
vectors rotate. The vector V *leads* the current vector I
by the phase angle $\phi$, whose tangent is given in (7.18);
and the magnitude of the voltage vector is $|Z|I_0$ where
$|Z|$ is given in (7.17); finally $RI$ represents the voltage
drop across the resistance. Note that the exact form of
the vector diagram depends on the actual values of $R$,
$C$ and $L$: Fig.52 corresponds to the case $\omega L > (1/\omega C)$,

which makes the phase angle $\phi$ positive. If $L$ is reduced enough the sign of the phase may be changed and the voltage vector will then *lag* behind the current vector.

Even this fairly simple circuit, with only three elements connected in series, is enough to show the main principles on which all forms of radio communication depend. More of that in the next Section!

## 7.4 Oscillator circuits

The circuit indicated in Fig.51 describes an electric **oscillator** with remarkable properties. The current from the generator oscillates with angular frequency $\omega$, going through a complete cycle of values when $\omega t$ increases by $2\pi$ i.e. when $t$ increases by $2\pi/\omega$. We remember that $T = 2\pi/\omega$ is the **period** of the oscillation and $\nu = (1/T) = (\omega/2\pi)$ is its **frequency**, the number of oscillations per unit time. The generator 'drives' the circuit at this frequency.

The first thing to note is that the oscillator has a very special angular frequency $\omega = \omega_0$ which makes

$$\left( \omega L - \frac{1}{\omega C} \right)$$

vanish. This is called the **natural frequency** of the oscillator and has nothing to do with the frequency of

the alternating current that is driving it. Clearly

$$\omega_0 = \sqrt{\left(\frac{1}{LC}\right)}. \qquad (7.19)$$

(The frequency, as usually defined in cycles per second (i.e. in herz), is then $\nu_0 = \omega_0/2\pi$.)

If the AC generator supplies power at this frequency, (7.17) shows that the circuit behaves like a pure resistance $R$ – as if the coil and condenser were not there! The impedance $|Z|$ has its smallest possible value $(R)$ and the phase difference $\phi$ between current and voltage is zero – the $I$ and $V$ vectors always point in the same direction and rotate at the same speed.

The next Example shows how much power is used in keeping things going.

---

**Example 4.** Power used in driving an oscillator.

According to Fig.52 the voltage vector, in general, leads the current by the angle $\phi$ defined in (7.18). Taking the instantaneous value of the current as the real part of $I_0 e^{i\omega t}$, namely $I_0 \cos \omega t$, that of the voltage will be $V_0 e^{i(\omega t+\phi)}$, namely $V_0 \cos(\omega t + \phi)$.

The rate at which work is being done, at the instant $t$, will therefore be $\mathrm{d}W/\mathrm{d}t = VI = V_0 \cos(\omega t+\phi) \times I_0 \cos \omega t$. But this quantity is oscillating with frequency $\omega$. What we need is the *average* value, taken over a whole period, and this rate of working will be the

power consumed:

$$P = \langle \mathrm{d}W/\mathrm{d}t \rangle = V_0 I_0 \langle \cos(\omega t + \phi) \cos \omega t \rangle.$$

To get the average value (indicated as usual by the 'angle brackets') we have to integrate as time goes from 0 to $T$ and then divide by the whole time interval $T$. (See if you can do this for yourself, using Book 3 if you need to and remembering the 'sum-rule' for $\cos A + B$ – as used in Example 1.) You should find the simple result

$$P = \tfrac{1}{2} V_0 I_0 \cos \phi$$

and you can evaluate this for any values of resistance, capacitance and inductance.

---

We said that the oscillator had some remarkable properties: the first one comes straight out of Example 4, which gave the result

Power consumed by a simple oscillator $= \tfrac{1}{2} V_0 I_0 \cos \phi$.
(7.20)

Here $V_0$ and $I_0$ are the *amplitudes* of the input voltage and current, while $\phi$ is their *phase difference* which follows from

$$\tan \phi = \frac{1}{R} \left( \omega L - \frac{1}{\omega C} \right).$$

It's now clear that if the circuit is oscillating at its *natural frequency* $\omega_0$, defined in (7.19), the generator will have to supply very little power. As the generator frequency $\omega$ approaches $\omega_0$, the voltage and current 'come

into phase' ($\phi \rightarrow 0$) and the impedance $Z$ in (7.17) goes to its minimum value $R$. This is called the case of **resonance** – when the frequency of the applied voltage exactly matches the natural frequency of the circuit.

In fact $R$ is often just the resistance of the wire connecting it to the circuit elements $L$ and $C$. For an 'ideal' wire, $R$ would be zero and once the circuit was oscillating it would go on by itself forever! If the generator is still applying a voltage of amplitude $V_0$ the current (and the power absorbed) will 'blow up' as $\omega \rightarrow \omega_0$ – just as if the generator were short-circuited by connecting its terminals directly. Nobody would want to destroy an expensive generator in that way!

On the other hand, a large inductance is often used to *control* the current taken from a power supply, since it acts by providing a back-emf to oppose the applied voltage; unlike an ordinary resistance, which simply wastes energy by producing Joule heat.

Next we'll do a simple experiment (in our minds) by taking the generator away at the moment when the condenser is fully charged, with voltage $V_0$ across its plates, and allowing it to discharge through the coil. First we'll put $R = 0$, for the case of 'ideal' wires:

**Example 5.** Discharge of a condenser through a coil.

We now think of the circuit in Figure 51, but without the generator (just replacing it by a connecting wire) and suppose the wires have very small resistance, so we can put $R = 0$. Remember that $V_C$ across the condenser is related to the current $I$ by $I = C(dV_C/dt)$, while $V_L$ between the teminals of the coil depends on its inductance $L$ and is related to $I$ by $V_L = L(dI/dt)$.

With neglect of resistance $R$ and with no generator voltage ($V = 0$), Kirchoff's first law for the circuit becomes $V_C + V_L = 0$. If this must be zero, at all times, then its rate of change must also be zero:

$$\frac{dV_C}{dt} + \frac{dV_L}{dt} = 0.$$

But, from above, the first term is $I/C$, while the second term will be

$$\frac{dV_L}{dt} = L\frac{d^2 I}{dt^2},$$

where the current $I$ is now differentiated *twice* with respect to the time $t$. On setting the sum of the two terms equal to zero, we get a **differential equation** which must be satisfied by $I$:

$$\frac{I}{C} + L\frac{d^2 I}{dt^2} = 0.$$

Perhaps you remmber meeting this equation before, in Chapter 6 of Book 3. There, in Example 2, the dependent variable was called $y$ and denoted the displacement of a swinging weight from it equilibrium position. The equation describing the swinging 'pendulum' was written $\mathsf{L}y = \mathsf{D}^2 y - \omega^2 y = 0$, where $\mathsf{L} = \mathsf{D}^2 - \omega^2$ is a **linear operator** and $\mathsf{D}^2$ is short for "differentiate twice with respect to time". The pendulum is in fact a *mechanical* oscillator with natural frequency $\omega/2\pi$. If you replace the displacement $y$ by the current $I$, and the positive constant $\omega^2$ by $1/LC$, what do you

254

get? You'll find exactly the differential equation for the *electrical* oscillator, as given above. Note that $\omega$ has the value $\omega_0$ found in (7.19): it is indeed the frequency with which the circuit oscillates when left to itself, with no generator driving it (provided of course the resistance of the connecting wires is negligible).

---

The last Example has shown us that when a charged condenser, of capacity $C$, is discharged by connecting its plates by a coil of inductance $L$, the current doesn't go smoothly down to zero. Instead $I$ satisfies the differential equation

$$\frac{I}{C} + L\frac{\mathrm{d}^2 I}{\mathrm{d}t^2} = 0 \qquad (7.21)$$

which has an oscillating solution $I \propto \exp \pm i\omega_0 t$, where $\omega_0$ is the natural frequency given in (7.19). For a circuit with ideal wires $(R = 0)$ the oscillation would continue forever. The next Example will show what happens when $R \neq 0$. In that case Kirchoff's first law becomes $V_C + V_L + V_R = 0$, where the extra term (voltage drop across the resistance) is simply $IR$. The rate of change of the sum must also be zero and this gives, instead of (7.21) the new differential equation

$$\frac{I}{C} + L\frac{\mathrm{d}^2 I}{\mathrm{d}t^2} + R\frac{\mathrm{d}I}{\mathrm{d}t} = 0 \qquad (7.22)$$

Let's see what difference the new term makes:

---

**Example 6.** Discharge through a coil including resistance

The differential equation we must solve is now (rearranging the terms to make it look nice)

$$L\frac{\mathrm{d}^2 I}{\mathrm{d}t^2} + R\frac{\mathrm{d}I}{\mathrm{d}t} + \frac{I}{C} = 0.$$

A standard way of solving an equation of this type (you don't need to go into the details) is to make a 'trial', choosing something you think might fit, containing one or more parameters that you can adjust. The 'something' must be easy to differentiate, so you can try it out easily. In this case a good 'trial function' might be $I = Ae^{\omega t}$, which is pretty general if the parameter $\omega$ is allowed to take complex values. Substituting this function in the above equation and doing the differentiations (do them!) you'll find

$$L\omega^2\, I + R\omega\, I + (1/C)\, I = 0.$$

The thing on the left is *quadratic* in the unknown $\omega$ and we know from Book 1 Section 5.3 that it will only be zero for *two* possible values. called the "roots" of the equation. If you look back at Book 1 you'll find that the standard quadratic equation is usually written

$$ax^2 + bx + c = 0, \qquad \text{with roots} \quad x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

So we must use $\omega$ instead of $x$ as the quantity we're looking for and put $a = L,\ b = R,\ c = 1/C$ as the known constants. When you do that you'll find the two solutions (taking the upper or lower signs in the $\pm$)

$$\omega = \frac{-R \pm \sqrt{R^2 - 4L/C}}{2L}.$$

The result looks nicer if you do a bit of algebra after putting $\omega_0^2 = 1/LC$ and $k^2 = R^2/4L^2$, where $\omega_0$ being the natural frequency of the circuit with $R = 0$. The two solutions then become (check it all carefully!)

$$\omega = -k \pm \sqrt{k^2 - \omega_0^2} = -k \mp i\sqrt{\omega_0^2 - k^2}.$$

The sum of the two is also a solution and gives finally (you'll need to use $\exp \pm i\theta = \cos\theta \pm i\sin\theta$)

$$I = I_0\, e^{-kt} \cos\omega_0' t.$$

---

The beautifully simple result from Example 6 shows that when the small resistance of the connecting wires is included (i) the current still oscillates in time (the factor $\cos\omega_0' t$), though the frequency is slightly less than the natural frequency $\omega_0$, and (ii) the amplitude $I_0$ is replaced by $I_0' = I_0 e^{-kt}$ and therefore goes down slowly with time. In fact

$$I = I_0\, e^{-kt} \cos\omega_0' t, \tag{7.23}$$

where $k = R/2L$ and $\omega_0'^2 = \omega_0{}^2 - k^2$. The behaviour of the current, before it dies out, is indicated in the Figure below, which decribes a 'damped' oscillation.
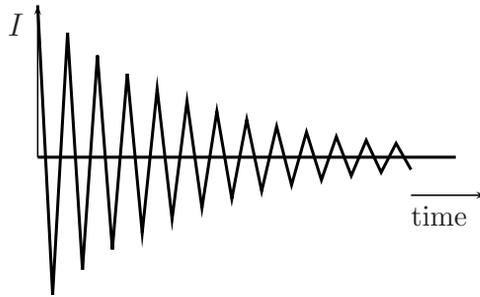
**Figure 53.** Decaying pulse from oscillator.
– schematic representation of
$$I = I_0 \, e^{-kt} \cos \omega_0' t.$$

# 7.5   Transmitting and receiving the waves

We started Chapter 7 by thinking about what is possible *today*, when radio signals can be sent out from **transmitters** and 'picked up' by observers in satellites, using **receivers**. But a hundred years ago all this would have seemed unbelievable: this branch of Applied Science, Radio Communication, was just at its beginning and it was a great achievement to send signals even a few hundred metres. To connect all this with Maxwell's equations has been part of a great adventure, and we're only just at the beginning. So let's go back 100 years and look at the first steps.

At the beginning of the chapter we were wanting to produce oscillating charges in a conductor called an antenna, in the hope that the resulting *oscillating dipole* would produce an oscillating electromagnetic field. We know from Maxwell's equations that it *should* and that the waves set up in this way would travel outwards from the dipole with the speed of light. And we guessed that we'd need a source of **high voltage**, oscillating at high frequency. Now, having learnt how to produce electric oscillations in a circuit containing a capacity ($C$) and an inductance ($L$), at any frequency we wish, we must fill in the details. Why, for example, will we need high

voltages? Because the charges $\pm Q$ at the ends of an antenna must be moved backwards and forwards at high frequency by an emf (measured in volts); that needs a lot of work and the bigger the oscillating voltage the faster it can be done.

The theory of radiation from an oscillating dipole was worked out by Hertz about 120 years ago, as an early application of Maxwell's equations. Hertz showed that wavefronts (Chapter 6) would move outwards from the dipole in all directions *except* along the axis of the dipole; so by keeping the dipole vertical waves would go out in all directions along the surface of the earth. The wavelength would depend on the length of the dipole and in the early days of radio it was expected that the longer the wavelength the further it would reach. Long waves were best produced by using very long vertical antennas, or "aerials", in the form of 'masts' ("antenna" is in fact the Greek word for a ship's mast); but times have changed and nowadays tall masts are not often seen.

Here we'll look at the earliest examples of radio transmitters and receivers.

**Early radio transmitters**

The earliest attempts to send out signals using electromagnetic waves (before the days of power stations and AC generators) depended on the use of batteries to produce the high voltages needed. We know that high volt-

ages can be produced easily, using the principle of the **transformer** (see Section 5.4), but only from an *alternating current* (AC) source. If you have only a battery, which is a direct current (DC) source, then you have to make the input current vary rapidly – because the transformer action depends on the *rate of change* of the magnetic flux produced by the current in the primary coil. In the **induction coil** this rapid variation is produced by an **interruptor** which 'makes and breaks' the connection to the primary coil; at one moment the current is flowing in, but at the next moment the connection is broken and the input current is zero. The next Figure shows how this can be done.
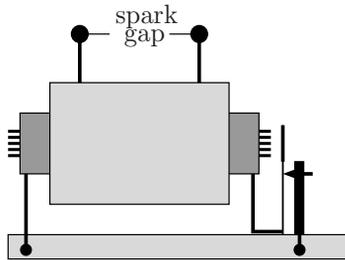


**Figure 54.** Induction coil (schematic)
Primary coil (dark grey) wound over core
Secondary (light grey) surrounds both.
Input terminals and wires at bottom,
connecting path shown in black (see text).

Current is supplied to the primary coil from the input

terminals, passing round the circuit shown in black. On the right of Fig.54 you see a pillar, with a pointed screw going through it and pressing on the vertical spring of the **interruptor** (or 'vibrator'). From there the current enters the primary coil, which is a solonoid, and produces magnetic flux along its axis. The core (a bundle of iron wires) then becomes a powerful electromagnet and attracts the iron disk at the top of the vibrator spring, pulling it towards the core. That breaks the circuit at the point where the screw touches the spring; the current falls to zero, switching off the electromagnet, and the spring goes back to the upright position; as soon as it touches the screw contact is restored, the magnet is switched on, and everything begins again. At the same time a bright blue spark jumps across the 'spark gap' shown at the top of Fig.54. Where does it come from?

Remember the transformer you studied in Section 5.4, which changed the input voltage $V_1$ applied to a primary coil into an output voltage $V_2$ of the emf induced in the secondary coil? The ratio of the two voltages came out to be the same as the ratio of the numbers of turns in the two coils: $V_2 = (N_2/N_1)V_1$. That result didn't depend on *how* the primary current was changing or on whether the voltage was being stepped up or stepped down. So if we now have 10,000 times as many turns in the secondary coil as in the primary, and jiggle the input voltage up and down in any way we please, a 6 V

car battery will give us a 60,000 V spark! We don't need a power station, with its big AC generators, if we only need to send radio signals – which is just as well, because the earliest transmitters were used for signalling between ships at sea and batteries are much more portable then power stations!

Now we know how to get the high voltages we can return to the problem of making the high frequency oscillations and launching them into space from some kind of antenna. The simplest form of transmitter, in use about 100 years ago, depended on an induction coil to supply the power and is indicated in the next Figure.
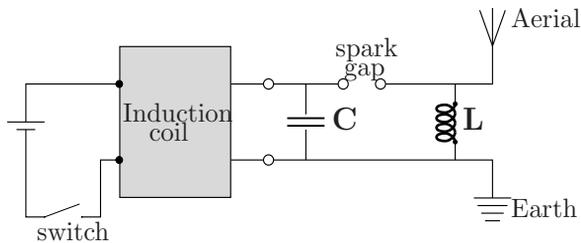


**Figure 55.** Circuit diagram for a spark-gap transmitter

The circuit included a 'tuned' oscillator, whose frequency could be changed by adjusting the values of the capacity $C$ and the inductance $L$ in a circuit of the kind shown in Fig.51. The oscillating high voltage was applied to an

263

antenna, usually in the form of a vertical 'aerial' with the lower end embedded in the earth (i.e. 'earthed' – fixing the zero of potential). Figure 55 shows a typical transmitter circuit of the day.

How does it work? The induction coil is in the grey box, with input from the battery (on the left) through a switch. When it's switched on it produces a high voltage across the output terminals. Between them is a condenser which charges up until the PD across the spark-gap is so big that a spark strikes (like a flash of lightening) between the two metal balls. The spark completes the circuit connecting the condenser and the coil; and this starts a high-voltage oscillation exactly as in Example 6 of Section 7.4 (the small resistance of the wires is not shown – but it's always there). The result is a 'pulse' in which current and voltage vary rapidly; and this is followed by another pulse as soon as the condenser has had time to re-charge and discharge again through the coil. This series of voltage oscillations across the coil feeds the Aerial and Earth sides of the dipole, shown on the right with the usual symbols. The 'switch' indicated in Fig.55 is really a **Morse key**, which lets the operator send out coded messages by switching the induction coil on and off. (Morse was the name of the inventor.) When the coil is switched on it charges the condenser, which discharges and produces the voltage oscillations that go to the antenna. When the pulse has decayed it

is followed by another – and so on, for as long as the switch is closed. A train of voltage pulses obtained in this way is indicated in Figure 56.
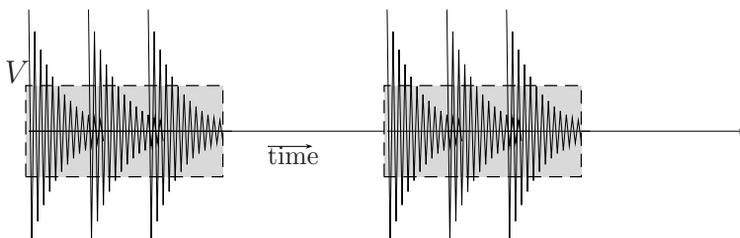


**Figure 56.** Series of pulses
– schematic representation of
antenna voltage $V = V_0 \, e^{-kt} \cos \omega'_0 t$.

Fig.56 shows three pulses in a shaded box, extending over a short time; then an interval when the switch is opened and the voltage falls to zero; and then another boxful of pulses when the switch is closed. In this way signals can be sent out from the antenna. The next thing to do is to find a way of receiving them – many miles away.

**Early radio receivers**

The signals that come from a transmitter arrive at the point where you want to receive them in the form of pulses of radio waves. If you look back at Fig.42 you'll see that such waves have oscillation frequencies going

from about $10^5$ Hz to $10^8$ Hz. On the other hand, the frequencies of *sound* waves are very much lower, being of the order of only a few hundred cycles per second ($10^2$ Hz). So there are two problems in changing the radio waves into something you can *hear*: you have to detect the incoming signal, using it to set up electrical oscillations in the receiver, at the same frequency; and then you have to use the electric circuit to drive some mechanical device at a frequency perhaps a thousand times lower.

The high frequency oscillations in each pulse of radiation (see Fig.56) are those of the **carrier wave**, but the *amplitude* of the wave may vary – as it does in Fig.53 where it falls slowly to zero. In that case the amplitude is said to be *modulated* and the modulation can be as slow as we please. This tells us how we can change the frequency. Instead of looking at the frequency of the oscillations within each decaying pulse of the carrier wave (Fig.53) we could use the frequency with which 'boxfuls' of pulses come out from the oscillator – which could be thousands of times smaller. Or we could use the frequency with which any modulated bits of the carrier wave come out. If we want to transmit the low-frequency sounds of speech all we have to do is impress on the carrier wave the pattern of the noises we make. That can be done by speaking into a **microphone**, which changes mechanical vibrations into electrical oscillations.

Let's look first at the circuit diagram for the simplest kind of radio receiver, the **crystal set** – so called because it used the special property of a crystal called **galena**. The diagram is shown in Figure 57, starting with the antenna on the left which picks up the waves from the transmitter.
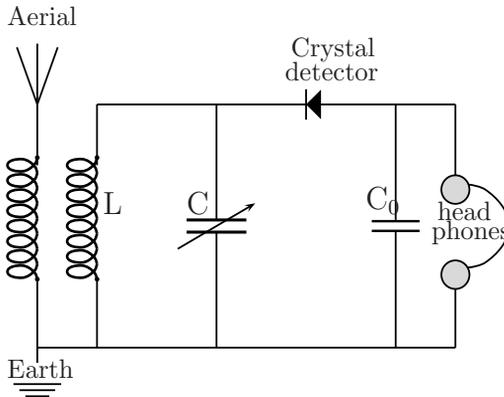


**Figure 57.** Circuit diagram for an early form of radio receiver, the 'crystal set'.

What happens in this circuit? The electromagnetic waves induce tiny electric currents in the aerial wire. These pass through a coil, on their way down to Earth, and this is 'inductively coupled' to a second coil (you'll remember about 'mutual inductance from Chapter 5) and

267

this is part of the basic **oscillator circuit** which also contains a condenser (marked C). This is a special kind of condenser, whose capacity can be varied (that's indicated in the diagram by putting an 'arrow' through it). You learnt about simple condensers in Section 1.5 of Chapter 1; those with variable capacity usually work by changing the overlap area of their plates when you turn a knob. That means you can change the **resonance frequency** of the circuit until it matches the frequency of the signal coming from the aerial. When you've done that, the tiny oscillating current coming in sets up a much bigger one in the circuit – big enough to measure!

That brings us to the second part of the problem: we've 'trapped' a high-frequency carrier wave coming from the aerial, but we still have to make it do something useful – like making a noise that would correspond to the lower-frequency signals impressed on the carrier wave at the transmitter. And that's where the **crystal detector** comes in! On the right of Fig.57 you'll see the symbol for a pair of headphones, one for each ear with a headband to hold them in place. Each headphone has a very thin soft-iron disk, called a "diaphragm", which can vibrate like the membrane of a drum when you touch it with a drum-stick. Underneath the diaphragm, but very close to it, is an **electromagnet** in the form of a letter 'U' with a current-carrying coil round one of the 'legs', as

shown in Fig.58 below.
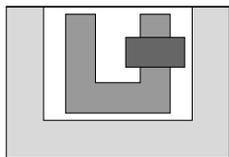
**Figure 58.**

Section through a headphone

A thin iron diaphragm is attracted
by a magnet, energized by a varying
current in the coil (dark grey)
(connecting wires not shown)

The magnetic field goes up and down as the electric current varies. It attracts the diaphragm, in proportion to the field strength, and produces a mechanical vibration – making sound. That can only happen if the variation is *slow* (low frequency) and always pulls the diaphragm in the same direction. (Remember the problem we met in Section 5.3 of making an AC generator run as a motor – it just wouldn't go when the impulse was rapidly changing direction.)

Here, the current in the oscillator circuit is *modulated* – at a low frequency compared with that of the carrier wave (see Fig.56). But it still *alternates*, changing sign within every square pulse. To make sure we don't have the 'AC-motor' problem we have to **rectify** the current so it won't change direction. That is done by the **crystal detector** in Fig.57, which only allows current to flow in the direction of the 'point'. The galena crystal, with a

269

pointed bronze wire pressing on it, is in fact an 'ancester' of the **transistor**, the device used nowadays in nearly all electronic gadgets like radios and computers. The incoming signal shows a voltage alternation like that in Fig.56, but when the voltage wants to send current in the 'wrong' direction it just won't go! The result is that only the positive part of the modulated carrier wave in Fig.56 gets through to the headphones – on the right in Fig.57 – and that's just what is needed to produce the low frequency mechanical vibrations of **sound**.

There's only one small thing left that may still seem mysterious: what is the second condenser ($C_0$) doing there? connected across the input to the headphones in Fig.57. You'll find the answer if you go back to Section 7.1 where you learnt about **complex impedance**. Remember this is a kind of 'effective' resistance. In Example 3 you found that the current passing through a condenser was given by $I = V/Z_C$, where $V$ is the voltage across its terminals, and that $Z_C = -i/\omega C$. Thus, $I = i\omega C \times V$.

What does it mean? The magnitude of the voltage that drives the headphones, making sound, is multiplied by $C_0\omega$ to give you the magnitude of the current going through the condenser (the factor $i$ only telling you that $V$ and $I$ are 90° out of phase). So low-frequency voltage variation ($\omega \to 0$) sends very little current through $C_0$: it all goes to the headphones – as it should. But any

high-frequency variation (that may have slipped through the crystal detector) is 'short-circuited' by the impedance, $C_0\omega$ being large. The condenser $C_0$ simply acts as a **filter**, taking out any unwanted (high-frequency) components that may be left in the incoming signal. Without that condenser the receiver will still work, but it will give very poor, noisy reception.

# Looking back –

When you started this book you knew only about the Physics you learnt in Book 4, mainly 'Classical Mechanics', which is concerned with the principles governing the motion of objects that possess **mass**. Now you know that was only a small part of Physics. Objects may possess **electric charge**, as well as mass, and when they do they behave differently: the principles and 'laws' need to be extended.

- In Chapter 1 you learnt about charges and the **fields** they produce *when they are at rest* (relative to the 'observer'). All this is part of **Electrostatics**. You can visualize the field using pictures like the one on the front cover of this book, which show 'lines of force' indicating the *direction* of the force acting on a unit charge at any point in space. You learnt how to calculate the field, using a single principle – **'Gauss's law'**; and how to build machines for making and storing **static electricity**. Finally, you found two mathematical ways of expressing Gauss's law (a bit more mathematics that you'll need again and again).

- When charges are in motion, in a **conducting material**, you had to learn (in Chapter 2) about the **currents** they set up – the **flux** of charge and

how to describe it. This led you to the study of **electric circuits** in which the currents flow through metal wires. You learnt to distinguish between current ($I$) and **voltage** ($V$), which drives it along the circuit. In electrostatics, the voltage (which is a difference of electric potential – a 'PD') may be very high, while the current is usually very small. On the other hand, in the study of 'current electricity', it's the other way round: low voltage moving a lot of charge and giving a high current. You learnt about **batteries**, which can produce heavy currents – from chemical reactions – but at a low voltage. This made you think about **electric power** and how to define it.

- Chapter 3 started (again!) from a bit of history – the experimental discovery of **magnetism** and the invention of the **compass**. You found that the magnetic field was very similar to the electric field – except that there was nothing corresponding to a 'magnetic charge'. Instead there were **magnetic 'poles'**, which always came in *pairs* – a 'North pole', where a line of force started, and a 'South pole', where it ended. The **magnetic field** was generated by a *current* – in other words a *moving* electric charge, not from a static charge.

Instead of Gauss's law, relating the flux of elec-

tric field E out of any closed *surface* to the total charge enclosed by the surface, you learnt that the **circulation** of magnetic field vector B around any closed *circuit* is proportional to the total current passing through it. (This is a difficult idea to grasp; but if you've studied the Examples of fields produced by current-carrying wires you'll begin to understand it.) The rest of Chapter 3 got you ready for the next big jump!

- In Chapter 4, the most mathematical in the whole of Book 10, you really reached the heart of electromagnetic theory. The four equations formulated by Maxwell about 130 years ago 'pull together' into just a few lines all the concepts and principles built up so far. You'll have to look back at them (and the way they were set up!) many times before you can say you completely understand them. But they give you a solid 'platform' on which to build everything else you'll need in applying the principles. The rest of the book contains only *applications* – no really new ideas!

- The first application is enormously important in Applied Science, because it showed how just one of Maxwell's equations could be used in designing and building **generators** big enough to supply all the electric power needed in our cities;

and the **electric motors** for turning that power back into mechanical power, for driving machinery in homes and factories. You found there were two main types of output from a generator, **direct current** (DC), always flowing in the same direction; and **alternating curent** (AC), alternately changing direction. In studying the transmission of power over large distances you found that AC had great advantages: by using very high voltages less power was wasted as heat; and the voltage could easily be changed (stepped up or down) in **transformers** (again depending on Maxwell's equations!)

- A second, and equally important, application is to the transmission of energy through space – without wires! Chapter 6 started by formulating the general ideas of **wave motion**, introducing **frequency**, **wavelength**, **amplitude**, and so on. You spent a lot of time on types of **wave equation**, whose solutions gave various kinds of wave propagation. Then finally you came to **electromagnetic** waves, which go through *empty* space, not needing any kind of 'medium' – like air or water – to carry them. They arise from disturbance of the field at any point and you were able to see how they followed from Maxwell's equations – with a predicted velocity $c$, the velocity of light, which

could be determined *in the laboratory*. You were able to follow up with a quick 'tour' of the whole **spectrum**, extending from long-wavelength radiation down to wavelengths of less than 'atomic' dimensions; verifying that all types of radiation carried *energy*.

- As a final application you studied the first successful attempts at **radio communication**. You had to prepare for this with a bit more work on AC circuits, containing **circuit elements** such as condensers and coils – whose behaviour depends on the **frequency** of the incoming or outgoing waves. It was probably a surprise that the **complex numbers** you learnt about years ago in Book 1 now came back into the picture! But that's always happening in Science –you can't keep things in separate boxes, the most unexpected things often turn out to be closely related. In this case the capacities and inductances of condensers and coils turned out to be similar to resistances (defined first for DC circuits), except that for AC circuits they have complex values! – they are **complex impedances**.

Now, after all that hard work, you've come close to the finishing line! You've looked at the circuit diagrams for a **radio transmitter** and a **radio**

**receiver** and discovered you can work your way round each circuit and see just how it works!

Before closing this book and putting it aside for a while, you may be worried that there are many questions that we haven't even asked. We've nearly always been talking about electric and magnetic fields in *free space* – in a perfect vacuum with no particles in it except the ones that we're studying. But the space we live in is not like that: usually it's filled with air or water or other kinds of material; and in some cases we've thought about materials such as glass and iron, noting only that Maxwell's equations and everything else we've done still seem to work, provided the fundamental constants $\epsilon_0$ and $\mu_0$ are replaced by $\epsilon$ and $\mu$ for whatever material we're dealing with. How can that be true?

The fact is that, as far as we can detect, the free-space equations are *always true* (that's why we studied them in such detail); but as soon as we start filling space with *real materials* we have to accept that materials are made up from **atoms** and **molecules**, and that they in turn are made up from **electrons** and **atomic nuclei** – and so it goes on!

This brings us to another big and growing area of Science – the **Science of Materials**, which will need other books in the Series. And before we can enter that field, we'll need to study a whole new area of fundamental

277

Physics –**quantum physics**. We'll start on this in Book 11, where we begin to look at matter on an atomic scale and find that the 'classical physics' developed so far must be *extended*. Nothing you've learnt so far, however, has to be thrown away: it still holds good *as far as it goes* and forms the starting point for going on into new areas. And in these new areas we'll have to create new 'models' to represent the atoms and molecules that 'space' is full of: quantum physics will give us the tools for understanding how such systems respond to the fields that we've studied in Book 10. Only then will it be possible to fully understand the properties of matter in its many forms.

INDEX – to follow (24 May 2009)

# Index

282