

Classical Electromagnetism IB

Prof Oleg Brandt et al.

September 29, 2021

Dedicated to all who continue to play after their childhood years...

Contents

1	Introduction	1
1.1	General thoughts	1
1.2	Resources	2
1.3	Literature	2
1.4	Electromagnetism	3
2	Electrostatic fields	5
2.1	Electrostatic Force	5
2.2	Electrostatic Potential	7
2.3	Vector gradient of a scalar – grad	9
2.4	Spatial derivatives of an electric field	11
2.5	The curl of a vector field and Stokes's theorem	13
2.6	Electric Monopoles	16
2.7	Dipoles	17
2.8	Force on a dipole in a uniform field	20
2.9	Potential energy of a dipole in a uniform field	20
2.10	Force on a dipole in a non-uniform field	21
2.11	Multipole expansions	23
2.12	Electric flux	25
2.13	The divergence of a field – div	26
2.14	Example – charge conservation	27
2.15	The divergence theorem for electric fields	28
2.16	Maxwell's equation for electrostatic fields	30
2.17	Applications of Gauss's law	30
2.18	Laplace's and Poisson's Equations	32
2.19	An introduction to Green's functions	33
2.20	Boundary conditions and uniqueness	34

2.21 Conducting sphere in a uniform electric field	37
2.22 Method of images	39
2.23 Image for conducting cylinder and line of charge	41
2.24 Capacitance	43
2.25 Capacitance of two long parallel cylinders	44
2.26 Electrostatic energy	45
2.27 Energy stored in a capacitor	47
2.28 Energy stored in an electric field	48
2.29 Forces and virtual work	50
2.30 Force on a charged conductor	53
3 Electrostatic fields in dielectrics	55
3.1 Introduction	55
3.2 Isotropic dielectrics	55
3.3 Polarisation charge density	58
3.4 Divergence theorem and polarisation	60
3.5 Gauss's law for dielectric materials	60
3.6 Summary of key points	62
3.7 A word of warning	63
3.8 Use of \mathbf{D} and \mathbf{E} in electrostatic problems	63
3.9 Inhomogeneous dielectrics, boundary conditions	64
3.10 Behaviour of field lines at boundaries	66
3.11 Boundary-value problems with dielectrics	68
3.12 Energy density in dielectrics	71

4 Magnetostatic fields	75
4.1 Introduction	75
4.2 Magnetostatic force	75
4.3 The force between two current elements	77
4.4 The force between currents	77
4.5 The field on the axis of a current loop	78
4.6 The field on the axis of a short solenoid	78
4.7 Gauss's Law for magnetic fields	79
4.8 Loops and magnetic dipoles	80
4.9 Potential energy of a magnetic dipole in a uniform field .	84
4.10 Force on a magnetic dipole in a non-uniform field	85
4.11 Magnetic scalar potential	86
4.12 Magnetic scalar potential of a current loop	88
4.13 Ampère's law	90
4.14 The magnetic field of a long wire	92
4.15 The force between two parallel wires	93
4.16 Magnetic field of a long solenoid	94
4.17 The magnetic scalar potential revisited	94
4.18 Magnetic vector potential	97
4.19 Maxwell's equations for magnetostatic fields	99
4.20 Resistance and Conductivity	100
5 Magnetostatic fields in magnetic materials	101
5.1 Introduction	101
5.2 Magnetisation currents	101
5.3 Surface magnetisation currents	105
5.4 Magnetic field strength	106
5.5 Summary of key points	109
5.6 Inhomogeneous magnetic materials	109
5.7 Comparing the properties of fields	111
5.8 Boundary-value problems	112
5.9 Electromagnets	115

6 Electromagnetic induction	119
6.1 Introduction	119
6.2 Faraday's law	119
6.3 Faraday's law—an alternative perspective	121
6.4 Self-inductance	123
6.5 Self-inductance of a long solenoid	124
6.6 Self-inductance of coaxial cylinders	125
6.7 Self-inductance of a pair of wires	126
6.8 Self-inductance and voltage	128
6.9 Energy stored in an inductor	129
6.10 Mutual inductance	130
6.11 Combining self- and mutual inductance	132
6.12 The relationship between self- and mutual inductance	133
6.13 The ideal transformer	135
6.14 Transformer circuits	136
6.15 Energy flow in resonant circuits	138
6.16 Magnetic energy	139
7 Maxwell's Equations and Waves	143
7.1 Completing Maxwell's equations—the displacement current	143
7.2 Free-space waves	144
7.3 Plane waves in isotropic media	147
7.4 Frequency of a plane wave	149
7.5 General plane waves	150
7.6 Energy flow	153
7.7 Poynting vector	155
7.8 Radiation Pressure and Momentum	155
7.9 Complex power	156
7.10 Reflection and transmission at interfaces	159
7.11 Reflection and transmission coefficients	161
7.12 Waves in plasmas	163
7.13 Waves in conducting media	168
7.14 The skin effect	171
7.15 Metals <i>vs</i> plasmas	174

8 Guided waves	175
8.1 Introduction	175
8.2 Parallel wires	175
8.3 Parallel wires	178
8.4 Coaxial cylinders	179
8.5 Strip transmission line	180
8.6 Power flow on transmission lines	182
8.7 Reflections on transmission lines	183
8.8 Input impedances of transmission lines	184
8.9 Waveguides	186
8.10 Plane-wave pairs	187
8.11 Conducting plates	188
8.12 Transverse electric modes	190
8.13 Rectangular waveguides	192
8.14 The General TE_{mn} mode	193
8.15 The Waveguide Equation	194
9 Conclusion and appendix	195
9.1 Conclusion	195
9.2 Useful results—non-examinable	196
9.3 The generalised divergence theorem	197
9.4 The generalised Stokes Theorem	197
9.5 Couple on Current Loop	198

Part IB Physics B — Electromagnetism

20 Lectures

Michaelmas term, 2021

Prof Oleg Brandt

I would like to gratefully acknowledge the kind support of previous lecturers of this course, in particular Prof Chris Ford.

1. Introduction

1.1 General thoughts

- The concepts of this course are organically built on one another. Hence, it is very important to give steady attention to the lecture course and try to follow continuously.
- It is highly advisable to put aside at least 20 minutes and ideally more in the afternoon/evening of the day of the lecture, to go through the material once again and identify the relevant concepts, think about them, and incorporate them.
- Being able to discuss complex scientific questions with others is an integral part of undergraduate training. It is highly beneficial to discuss the contents of the lecture with other fellow students. In times of social distancing this can happen in the context of supervisions and through ad-hoc video/audio calls with other students enrolled in the course.
- Make full benefit of the handouts that are given our regularly
 - have a very quick glance at the relevant pages for the upcoming lecture *just before* the lecture. These 5 minutes are a small investment compared to the immense benefit of knowing the general outline of where the lecture is heading!
- Whenever a new type of problem/question comes up, it is advisable to pause for a few seconds to reflect on it and to think what one's intuition says before reading on. Good scientific intuition is a central part of being a physicist, and it can be trained!
- Whenever you find a subject difficult to grasp, consider reading up on it in literature, where it may be presented from a slightly different perspective that will make a difference. Remember, physics is about understanding concepts and not learning by heart.

1.2 Resources

- 20 lectures + 20 slides — Uploaded on TiS (link below)
- Handout lecture script + vector calculus — Given out at start of term.
- 2 Problem sheets — Please report any errors.
- Worked examples — Given out around lecture 15 and early next term.
- Web page — For copies of the handouts, and links to web sites containing electromagnetism demonstrations and other useful information:
 - See the course web page on the TiS (Handouts, slides, problem sheets, etc):
<https://www-teach.phy.cam.ac.uk/students/courses/electromagnetism/82>
 - See the course web page on VLE / Moodle (lecture capture + demonstrations):
<https://www.vle.cam.ac.uk/course/view.php?id=89161§ionid=601461>

1.3 Literature

Duffin, W.J.: Electricity and Magnetism (4th Edition, W J Duffin Publishing 2001). Start here if you find electromagnetism a mystery.

Grant, I.S. & Phillips, W.R.: Electromagnetism (2nd edn, Wiley 1990). This treatment is at about the level of the course. It is easier to read than Bleaney & Bleaney, but does not go so far.

Halliday D., Resnick R., and Krane K.S.: Physics (Volume 2) (5th edn, Wiley 2002). This is fairly basic but is good for background and has interesting examples and pictures; it may also be useful for optics and quantum mechanics.

Bleaney, B.I. & Bleaney, B.: Electricity and Magnetism: OUP (3rd edn. OUP, two volumes). This is a classic text that will see you through Part IB and Part II.

David J. Griffiths: Introduction to Electrodynamics (3rd edn, Pearson 1998). This book treats the subject in a rigorous manner, and typically small print aspects are discussed, too. Consult this textbook to better understand concepts.

Richard P. Feynman, Robert B. Leighton, & Matthew Sands: Feynman Lectures on Physics, Vol. II (Basic Books; New Millennium ed. edition 2011). This book is a treasure chest with lots of interesting and exciting insights into the subject. Consult this textbook to better understand concepts.

1.4 Electromagnetism

During this course we shall study the relationships between charge, current, electric fields, and magnetic fields, and explore how they behave in free space and in materials.

Amazingly, the whole of electromagnetism is described by just four equations, called Maxwell's equations, and we shall study their structure and application.

Key concepts to understand are scalar and vector fields:

A **scalar field** is a region of space where some scalar quantity is specified at every point. Usually we shall assume that the scalar fields of interest and their first derivatives are continuous.

A **vector field** is a region of space where three scalar quantities, a vector, are specified at every point. We shall assume that the vector fields of interest and their first derivatives are continuous.

Higher order fields, called **tensor fields**, also exist, where typically nine quantities are specified at every point. The full power of Maxwell's equations is unleashed when tensor formalism is used, but here we shall only refer to tensors in passing.

The whole of classical electromagnetism can be described by just four equations. **Maxwell's equations** can be written in vector differential form:

$$\left\{ \begin{array}{ll} M1 & \boldsymbol{\nabla} \cdot \mathbf{D} = \rho_{\text{free}} \\ M2 & \boldsymbol{\nabla} \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \\ M3 & \boldsymbol{\nabla} \cdot \mathbf{B} = 0 \\ M4 & \boldsymbol{\nabla} \times \mathbf{H} = \mathbf{J}_{\text{free}} + \frac{\partial \mathbf{D}}{\partial t} \end{array} \right. \quad \begin{array}{l} \text{Gauss' theorem} \\ \text{for electric field} \\ \text{Faraday's law of} \\ \text{electromagnetic} \\ \text{induction} \\ \text{Gauss' theorem} \\ \text{for magnetic field} \\ \text{Ampère's law and} \\ \text{displacement current} \end{array} \quad (1.1)$$

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \quad \text{Lorentz force law} \quad (1.2)$$

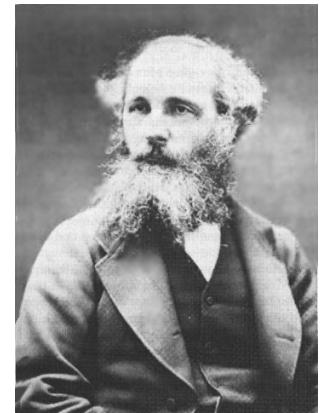
The whole course focuses on understanding and using these equations.

Maxwell's equations relate vector and scalar fields, where ρ is the charge density— C m^{-3} (coulombs/metre³)

\mathbf{D} is the electric flux density— C m^{-2} (coulombs/metre²)

\mathbf{B} is the magnetic flux density— Wb m^{-2} (webers/metre²) or (tesla)

\mathbf{E} is the electric field strength— V m^{-1} (volts/metre)



James Clerk Maxwell
(1831–1879)

\mathbf{H} is the magnetic field strength— A m^{-1} (amps/metre)
 \mathbf{J} is the electric current density— A m^{-2} (amps/metre²)

Maxwell's equations can also be written in integral form:

$$\begin{aligned} \text{M1} \Rightarrow \quad & \oint_{\partial V} d\mathbf{S} \cdot \mathbf{D} = \int_V dV \rho & (1.3) \\ \text{M3} \Rightarrow \quad & \oint_{\partial V} d\mathbf{S} \cdot \mathbf{B} = 0 \\ \text{M2} \Rightarrow \quad & \oint_{\partial S} dl \cdot \mathbf{E} = -\frac{\partial}{\partial t} \int_S d\mathbf{S} \cdot \mathbf{B} \\ \text{M4} \Rightarrow \quad & \oint_{\partial S} dl \cdot \mathbf{H} = \int_S d\mathbf{S} \cdot \mathbf{J} + \frac{\partial}{\partial t} \int_S d\mathbf{S} \cdot \mathbf{D} \end{aligned}$$

These equations underpin all classical electromagnetic phenomena, from the physics of electromagnetic wave propagation, scattering, radiative heat transfer, optics, non-linear optics, plasma physics, early-Universe physics, to the operation of optical fibres, radio antennas, cables and printed circuit boards. They also form the starting point for the quantisation of the electromagnetic field, and have a key role in quantum physics, in particular quantum optics. They have also been the starting point for some of the most important theoretical and experimental discoveries in physics. There is no area of physics that is left untouched by these equations.

Because Maxwell's equations include such a wealth of phenomena, it is not surprising that their manipulation and use is complex, but as we shall see, the structures that emerge are very elegant and can be understood relatively easily.

To understand Maxwell's equations and their use, it is vitally important to acquire a mental picture of their geometrical form. In fact, the most important requirement when following this course is to ensure that the geometrical form of every construction, problem, and solution is understood.

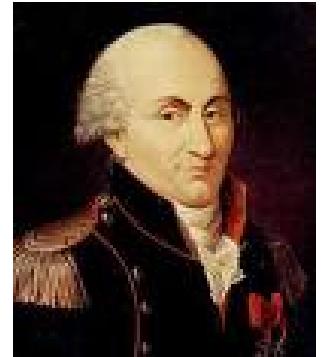
There is no doubt that Maxwell's equations are amongst the most profound equations in physics, and that electromagnetism is an essential tool for every practising physicist.

2. Electrostatic fields

2.1 Electrostatic Force

Coulomb showed experimentally that the magnitude of the force between two charged particles is inversely proportional to the square of the distance between them, i.e., $|\mathbf{F}| \propto 1/r^2$. The inverse square-law has been confirmed to high precision on many occasions. It is known that, if $|\mathbf{F}| \propto 1/r^n$, $|n - 2| < 3 \times 10^{-16}!$

The force is linearly proportional to the charge of each particle; in particular this implies that same charges repel each other, while opposite charges attract. The force is directed along the line joining them, see Fig. 2.1.



Charles-Augustin de Coulomb (1736–1806)

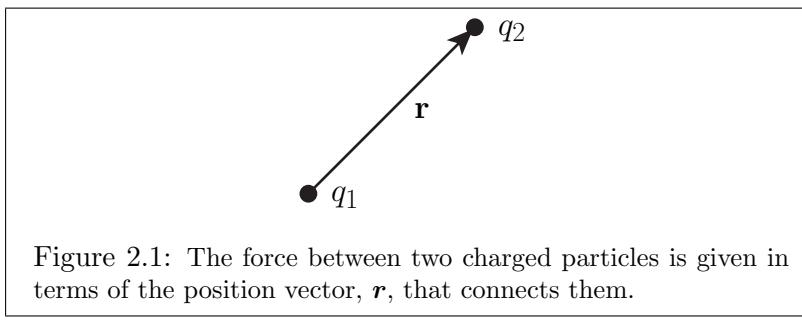


Figure 2.1: The force between two charged particles is given in terms of the position vector, \mathbf{r} , that connects them.

If the position vector of particle 2 with respect to particle 1 is \mathbf{r} , then the force experienced by particle 2 due to particle 1 is given by

$$\boxed{\mathbf{F}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \frac{q_1 q_2}{r^3} \mathbf{r} = \frac{1}{4\pi\epsilon_0} \frac{q_1 q_2}{r^2} \hat{\mathbf{r}}, \quad \text{Electrostatic force}} \quad (2.1)$$

from which we find

$$|\mathbf{F}(\mathbf{r})| = \frac{1}{4\pi\epsilon_0} \frac{q_1 q_2}{r^2}. \quad (2.2)$$

The constant of proportionality, ϵ_0 , is called permittivity of free space and depends on the system of units. In SI units, charge is measured in coulombs, and $\epsilon_0 = 8.85 \times 10^{-12} \text{ F m}^{-1}$ ($\text{F} = \text{C/V}$).

Coulomb's law is valid over all measurable length scales, from 10^{-16} m , subatomic scales, upwards. We shall assume, correctly, that it holds for moving charge as well as for static charge.

The convention for negative and positive charges is arbitrary, and goes back to Benjamin Franklin (1750), who defined positive charge as the one that remains on a glass rod after rubbing it with silk tissue, while negative charge resides on the silk.

In analogy to the gravitational field and its effect on a probe mass, in electrostatics it is practical to introduce the concept of electric field through its effect on a probe charge.

The **electric field** $\mathbf{E}(\mathbf{r})$ is the force per unit charge that would be exerted on an infinitesimally small positive test charge placed at \mathbf{r} .

$\mathbf{E}(\mathbf{r})$ is a vector that may vary in magnitude and direction as a function of position. It may be non-zero everywhere (say for an isolated source charge in empty space) or it may be zero in some places (where the effect of the source charge is screened by an intervening conductor). $\mathbf{E}(\mathbf{r})$ is an example of a **vector field**, because it is necessary to specify three scalar quantities E_i , $i = 1, 2, 3$ at every point \mathbf{r} in order to describe its form (i.e., it is represented by a vector).

It is necessary to use a infinitesimally small test charge in order not to disturb the positions of the other charges, which might move significantly if the test charge is too large. If the charges are fixed, then we can use a test charge of *any* size.

Obviously, a static charge cannot exert a force on itself, and therefore in calculations it is essential to ensure that a charge's own electric field is not included when calculating the electric field that the charge experiences. Care is required.

The force on a particle with charge q in an electric field $\mathbf{E}(\mathbf{r})$ is given by

$$\mathbf{F}(\mathbf{r}) = q\mathbf{E}(\mathbf{r}). \quad \text{Electrostatic field and force} \quad (2.3)$$

It follows from (2.1) that the electric field associated with a single point charge q is

$$\mathbf{E}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \frac{q}{r^2} \hat{\mathbf{r}}. \quad \text{Electric field} \quad (2.4)$$

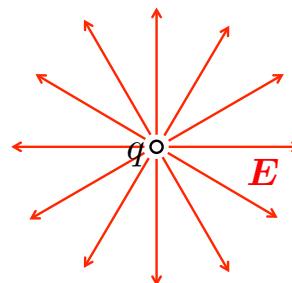


Figure 2.2: Electric field of a single positive point charge.

According to the **superposition principle**, if a system is comprised of a number of charged particles, then the electric fields associated with each of them can be added vectorially – electrostatics is linear (unless a non-linear dielectric medium is present).

Clearly, the positions of the source charges must be known, and so care is required because the addition of an extra charge may cause the positions of the existing charges to move.

2.2 Electrostatic Potential

To move a charge from \mathbf{r} to $\mathbf{r} + d\mathbf{r}$ in an electric field, work must be done:

$$dW = -d\mathbf{r} \cdot \mathbf{F}(\mathbf{r}). \quad (2.5)$$

The minus sign is needed because energy must be supplied to the system in order to move a charge *against* the direction of the force. As usual, we must distinguish between the force \mathbf{F} exerted *on* a test charge and the equal, opposite force $-\mathbf{F}$ that has to be provided in order to move a test charge against the direction of the field.

The **electric potential difference** between two points is the energy that must be supplied, per unit charge, in order to move a small test charge from one point to another. For points separated by an infinitesimally small distance $d\mathbf{r}$, the electric potential difference, dV , is

$$dV = -d\mathbf{r} \cdot \mathbf{E}(\mathbf{r}). \quad (2.6)$$

The above arguments have been made in the context of potential differences, which is the correct physical interpretation. Sometimes it is convenient to think about the potential at a point as being the potential difference between the point and a point at infinity, where the influence of the source charges is negligible. In other words, one often hears about the electric potential as being the energy needed to bring up unit test charge from infinity to a point. In reality, only potential differences are meaningful, and the point of zero potential, the reference point, is chosen arbitrarily.

In order to calculate the total electrical potential difference, V_{BA} , between two widely separated points, A and B , it is necessary to add up all of the infinitesimal differential potentials when moving along a specified path:

$$V_{BA} = - \int_A^B d\mathbf{r} \cdot \mathbf{E}(\mathbf{r}). \quad (2.7)$$

So the potential difference is found by evaluating a line integral. It is clear that the units of $\mathbf{E}(\mathbf{r})$ are volts per metre, V m^{-1} . Notice that the potential difference is *minus* the line integral of \mathbf{E} , cf. (2.5).

Once the reference point of zero potential has been defined, it is possible to calculate the electric potential at every other point throughout some region by performing line integrals.

The potential $V(\mathbf{r})$ constitutes a **scalar field**, because it is only necessary to specify a single real number at every point in order to describe the physical property of interest.

In order to be meaningful, the potential difference between two points can only depend on the positions of the start and end points,

and not on the path taken when the line integral is evaluated. Can we show that this is true?

Consider moving a test charge around a closed loop in a potential field. More specifically, consider moving it from A to B , and then back again from B to A , as illustrated below:

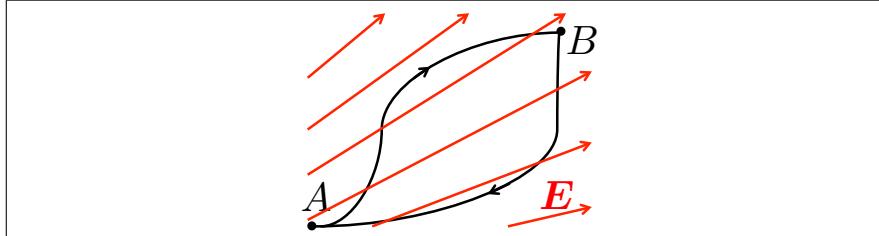


Figure 2.3: Moving a test charge along a closed loop between A and B .

The total potential difference around this closed loop is

$$V_{ABA} = -\underbrace{\int_A^B d\mathbf{r} \cdot \mathbf{E}(\mathbf{r})}_{\text{path 1}} - \underbrace{\int_B^A d\mathbf{r} \cdot \mathbf{E}(\mathbf{r})}_{\text{path 2}}. \quad (2.8)$$

Trivially, we have

$$V_{ABA} = -\underbrace{\int_A^B d\mathbf{r} \cdot \mathbf{E}(\mathbf{r})}_{\text{path 1}} + \underbrace{\int_A^B d\mathbf{r} \cdot \mathbf{E}(\mathbf{r})}_{\text{path 2}}, \quad (2.9)$$

but the potential difference between a point and itself is zero, because energy is not required in order to leave the test charge at the same place. Therefore $V_{ABA} = 0$, and

$$\underbrace{\int_A^B d\mathbf{r} \cdot \mathbf{E}(\mathbf{r})}_{\text{path 1}} = \underbrace{\int_A^B d\mathbf{r} \cdot \mathbf{E}(\mathbf{r})}_{\text{path 2}}. \quad (2.10)$$

The path of integration does not matter – only the positions of the end points matter, q.e.d.

If this were not true, energy would be lost or delivered to the system, but if the system is conservative with no external sources or sinks of energy, then the potential differences around a closed loop must sum to zero. The potential is thus a property of the space in which the test charge moves, rather than being a property of the way in which it moves.

The situation is entirely analogous to the way in which the gravitational potential differences must sum to zero when we walk around on a hillside on a path that takes us back to the start.

2.3 Vector gradient of a scalar – grad

Up to now, we have considered how to calculate the potential field, $V(\mathbf{r})$, if the electric field, $\mathbf{E}(\mathbf{r})$, is known everywhere throughout the region of interest. We might wonder if it is possible to calculate the vector electric field $\mathbf{E}(\mathbf{r})$ if the scalar electrical potential $V(\mathbf{r})$ is known.

The procedure is straightforward. Note that

$$dV = \frac{\partial V}{\partial x} dx + \frac{\partial V}{\partial y} dy + \frac{\partial V}{\partial z} dz, \quad (2.11)$$

which can also be written

$$dV = \left(\hat{\mathbf{x}} \frac{\partial V}{\partial x} + \hat{\mathbf{y}} \frac{\partial V}{\partial y} + \hat{\mathbf{z}} \frac{\partial V}{\partial z} \right) \cdot (\hat{\mathbf{x}} dx + \hat{\mathbf{y}} dy + \hat{\mathbf{z}} dz), \quad (2.12)$$

where $\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}$ are Cartesian unit vectors.

The first bracket on the RHS is the vector differential operator ∇ operating on V , defined through

$$\nabla V \equiv \left(\hat{\mathbf{x}} \frac{\partial}{\partial x} + \hat{\mathbf{y}} \frac{\partial}{\partial y} + \hat{\mathbf{z}} \frac{\partial}{\partial z} \right) V(\mathbf{r}). \quad \nabla \text{ ("nabla"-operator)}$$

(2.13)

∇ is an operator that operates on a scalar field, in this case $V(\mathbf{r})$, to produce a vector field, in this case $\nabla(\mathbf{r})$. \Rightarrow The vector components of ∇ give the rate of change of the scalar field in each of the coordinate directions.

Using (2.12) and (2.13),

$$dV = \nabla V \cdot d\mathbf{r}, \quad (2.14)$$

which on comparing with (2.6) leads to

$$\mathbf{E}(\mathbf{r}) = -\nabla V(\mathbf{r}).$$

(2.15)

$\nabla V(\mathbf{r})$ is called the **gradient** of $V(\mathbf{r})$ ("grad V "), and it demonstrates one of three possible uses of the vector differential operator ∇ . In this usage, ∇ generates a vector field from a scalar field.

$\nabla V(\mathbf{r})$ is a vector that has both *direction* and *magnitude* at every point in space. What do these tell us?

We know that dV is the scalar product between $\nabla V(\mathbf{r})$ and the differential displacement $d\mathbf{r}$. \Rightarrow The magnitude of dV is thus at a maximum when $d\mathbf{r}$ points in the same direction as $\nabla V(\mathbf{r})$, $\Rightarrow \nabla V(\mathbf{r})$ points in the direction in which the potential increases most rapidly with position, or \Rightarrow equivalently in the opposite direction

to the electric field, which tells us the direction of the force that would act on a test particle.

The magnitude of $\nabla V(\mathbf{r})$, i.e., $|\nabla V(\mathbf{r})|$ tells us the maximum rate of change of the potential with position.

The potential function is only defined to within a constant offset, because a (spatially invariant) constant can be added to $V(\mathbf{r})$ without changing the electric field that would be calculated. This ambiguity confirms the notion that any point can be chosen as the reference point of zero potential. Setting the undefined constant to some specific value is known as choosing the **gauge**, a concept and terminology that has great importance when time-varying fields are considered.

Note that, while $V(\mathbf{r})$ is a useful physical concept, it is *not* a physical observable unlike $\mathbf{E}(\mathbf{r})$ that can be measured. This is evident from the arbitrariness in the definition of $V(\mathbf{r})$ from its gauge.

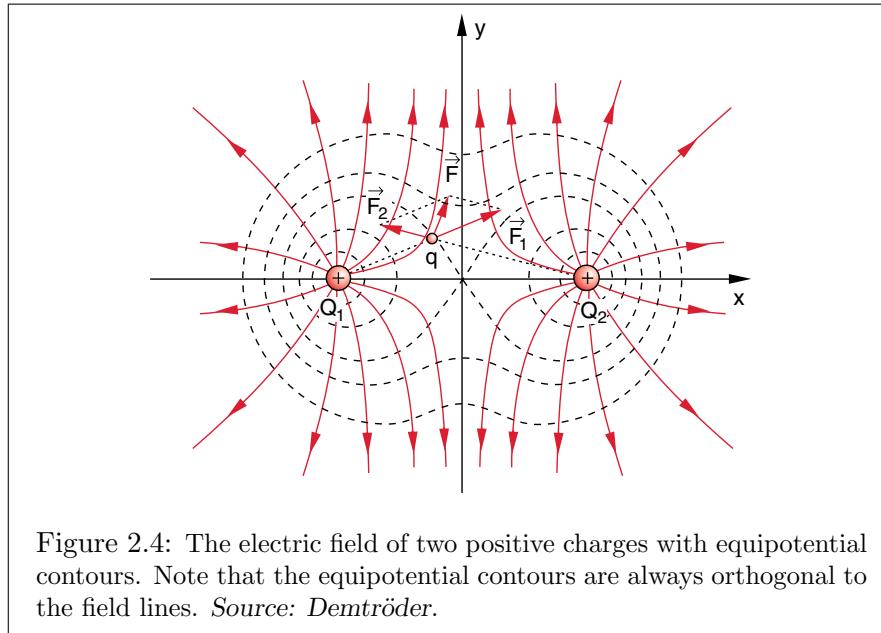


Figure 2.4: The electric field of two positive charges with equipotential contours. Note that the equipotential contours are always orthogonal to the field lines. Source: Demtröder.

For any collection of charge we can draw diagrams showing the electric field lines, because at each point we have the total field by summing vectorially the fields associated with each of the individual charges according to the superposition principle. We can then integrate from point to point, building up a map of the electric potential. Equivalently, we can draw surfaces of constant potential, and use the vector differential operator to calculate the field at every point. This is illustrated for the respective cases of two same and opposite charges in Figs. 2.4 and 2.5

If we move along an equipotential surface, the potential does not change, which means that there can be no component of $\mathbf{E}(\mathbf{r})$ in that direction.

In other words, at every point the field vector points in the direction of the maximum rate of change of the potential, and therefore the field lines are orthogonal to the equipotential surfaces.

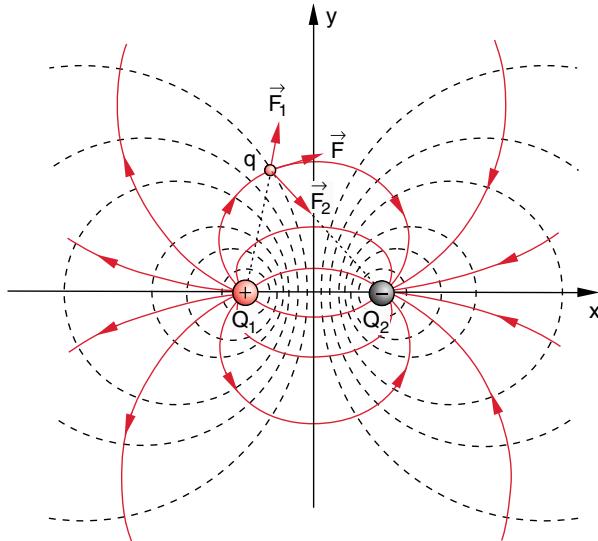


Figure 2.5: The electric field of opposite charges with equipotential contours. Note that the equipotential contours are always orthogonal to the field lines. Source: Demtröder.

Intuitively, just as for contours on a map, it is clear that there will be regions where equipotential surfaces are packed closely together, and other regions where they are spaced widely apart. We can understand the meaning of these differences in terms of the rate of change of potential, which is the electric field.

Equally, there will be regions where electric field lines are crowded closely together (representing high electric field), and other regions where they are spaced wide apart (low electric field). Care is needed because the density of field lines just depends on how many are drawn (whereas for equipotentials each can be labelled with the potential)! We will study this problem later.

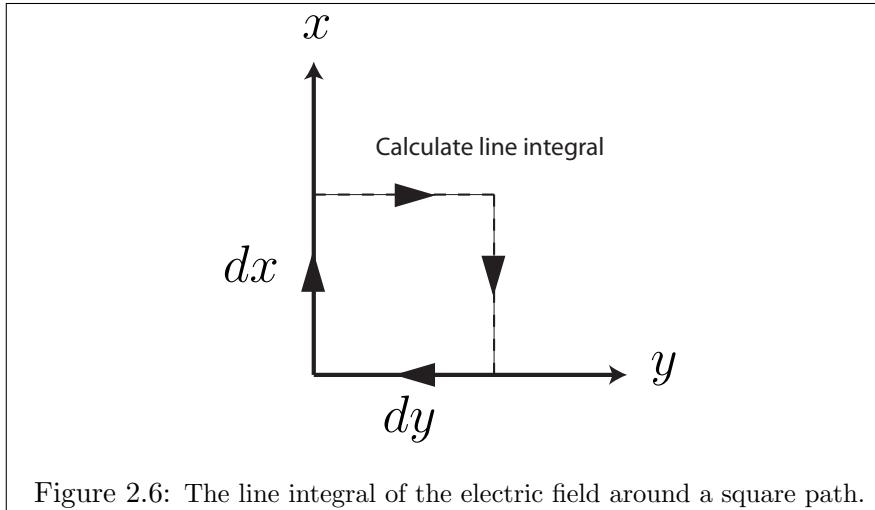
2.4 Spatial derivatives of an electric field

We can now derive important relationships between the spatial derivatives of the three components of the electric field at a point. Be careful: we are now concerned with the derivatives of an electric (vector) field, not a potential (scalar) field.

We know that the potential differences around a closed loop must sum to zero. Take the loop to be an infinitesimal rectangle in the $x - y$ plane of a Cartesian coordinate system, and assume that the side lengths are dx and dy (see Fig. 2.6).

Summing the potential differences along the sides gives

$$\begin{aligned} E_x dx + \left(E_y + \frac{\partial E_y}{\partial x} dx \right) dy \\ - \left(E_x + \frac{\partial E_x}{\partial y} dy \right) dx - E_y dy = 0, \end{aligned} \tag{2.16}$$



or, after multiplying out,

$$\frac{\partial E_y}{\partial x} dx dy = \frac{\partial E_x}{\partial y} dy dx, \quad (2.17)$$

but because this must hold for any dx and dy , we find

$$\frac{\partial E_y}{\partial x} = \frac{\partial E_x}{\partial y}. \quad (2.18)$$

The same procedure can be carried out in the $x-z$ and $y-z$ planes, giving a set of three relationships:

$$\begin{aligned} \frac{\partial E_y}{\partial x} &= \frac{\partial E_x}{\partial y} \\ \frac{\partial E_z}{\partial x} &= \frac{\partial E_x}{\partial z} \\ \frac{\partial E_y}{\partial z} &= \frac{\partial E_z}{\partial y}. \end{aligned} \quad (2.19)$$

Thus the spatial derivatives, at a point, of the components of the electric field cannot be chosen arbitrarily, but are linked through the above equations. Note that, while the specific relationships in (2.19) hold in electrostatics only, the components of \mathbf{E} are *always* linked. Equations (2.19) can be derived in a much more elegant way, but first we need to introduce the concept of the curl of a vector field, see next Section 2.5

Thought for the day 1: Consider what is meant by the operation $\nabla \mathbf{E}$, and why it constitutes a tensor field. Be careful, \mathbf{E} is a vector, not a scalar. How can the equations listed in (2.19) be described in terms of a single property of the tensor $\nabla \mathbf{E}$?

2.5 The curl of a vector field and Stokes's theorem

The **curl** of a vector field $\mathbf{E}(\mathbf{r})$ is defined according to

$$\begin{aligned}\operatorname{curl} \mathbf{E} = \nabla \times \mathbf{E} &\equiv \hat{\mathbf{x}} \left(\frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} \right) \\ &+ \hat{\mathbf{y}} \left(\frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} \right) \\ &+ \hat{\mathbf{z}} \left(\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} \right),\end{aligned}\quad (2.20)$$

which can be remembered symbolically through

$$\nabla \times \mathbf{E} = \begin{vmatrix} \hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ E_x & E_y & E_z \end{vmatrix}, \quad (2.21)$$

where the vertical lines indicate that the elements of the ‘matrix’ should be manipulated as if calculating the determinant.

The curl of a vector field thus generates another vector field; in a loose sense it can be regarded as taking the vector cross product between the vector differential operator ∇ and the vector field of interest.

To get a physical sense of what the curl operator means, we must derive Stokes’s theorem.

Let us calculate the line integral of a vector field $\mathbf{A}(\mathbf{r})$ around an infinitesimally small rectangle in the $x - y$ plane, as in Fig. 2.6 of Sect. 2.4, but for a general vector field $\mathbf{A}(\mathbf{r})$, for which we cannot assume that the integral evaluates to zero.

Thus the line integral becomes, in general, the quantity

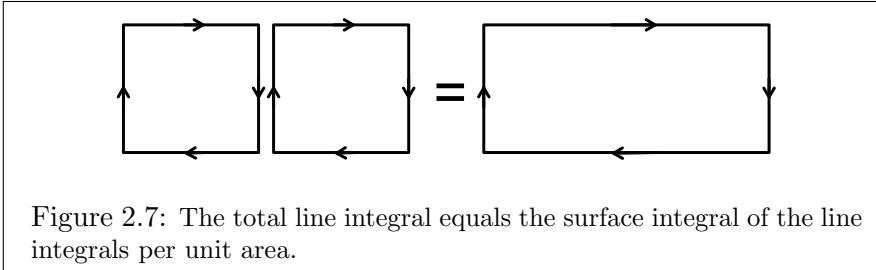
$$\frac{\partial A_y}{\partial x} dx dy - \frac{\partial A_x}{\partial y} dx dy \quad (2.22)$$

or

$$\left(\frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y} \right) dx dy. \quad (2.23)$$

The quantity in brackets can be regarded as the line integral around an infinitesimally small rectangle, per unit area. The notion of a line integral, around a closed path, per unit area is not as strange as it may first seem.

Consider placing two elemental areas together, such that they share a common edge. Because the line integrals on the common edge are evaluated in different directions, the line integral around the outside of the combined areas is the sum of the line integrals around the



two areas considered separately. Thus we can calculate the total line integral, by means of a surface integral of the line integrals per unit area, as illustrated in Fig. 2.7.

The real beauty of the curl operator is that it is not limited to a plane. First, we can derive expressions for the line integrals per unit area in each of the $y - z$, $x - z$, and $x - y$ planes, giving

$$\begin{aligned} & \left(\frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z} \right) dy dz, \\ & \left(\frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x} \right) dz dx, \\ & \left(\frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y} \right) dx dy, \end{aligned} \quad (2.24)$$

which can be written

$$\begin{aligned} & \left(\frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z} \right) \hat{x} \cdot d\mathbf{S}_1, \\ & \left(\frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x} \right) \hat{y} \cdot d\mathbf{S}_2, \\ & \left(\frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y} \right) \hat{z} \cdot d\mathbf{S}_3. \end{aligned} \quad (2.25)$$

$d\mathbf{S}_1$, $d\mathbf{S}_2$, $d\mathbf{S}_3$ are vectors representing the elemental areas of three faces. In the usual way, the direction of $d\mathbf{S}$ orientates the surface, whereas its magnitude gives the area.

Imagine now taking *three* faces of a box, where each face corresponds to one of the principal planes, as demonstrated in Fig. 2.8. Because the edges of the faces are connected, we can calculate the line integral around the free edge of the total structure by adding together the above three (scalar) contributions, or equivalently calculating the surface integral over the three faces. Thus,

$$\sum_i (\nabla \times \mathbf{A}) \cdot d\mathbf{S}_i = \sum_i (\nabla \times \mathbf{A})_i d\mathbf{S}_i = (\nabla \times \mathbf{A}) \cdot d\mathbf{S}, \quad i \in \{1, 2, 3\}. \quad (2.26)$$

Complex surfaces can be built up by adding more and more infinitesimally small rectangles. Thus, in the differential limit, the

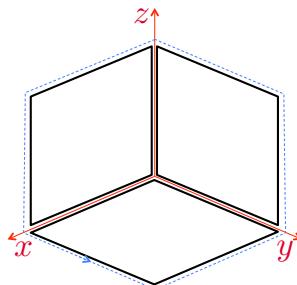


Figure 2.8: Three faces of a box, where each face corresponds to one of the principal planes.

notion of a line integral per unit area can therefore be applied to surfaces of any shape.

Although we have developed an intuitive idea through considering small rectangles, the principle is very general. If we evaluate the curl of a vector field at a point, and take the dot product with respect to some unit vector, the result is the line integral per unit area of the vector field around an infinitesimal closed loop that is perpendicular to the unit vector.

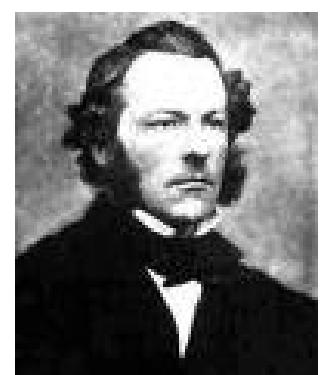
By calculating the surface integral of the curl of a vector field over some surface that has a free boundary, one calculates the line integral around the free edge of the surface. It can be seen that if the integral is carried out over the whole of a closed surface, the result must sum to zero. For example, if the surface is a sphere, the line integrals around the edges of two hemispheres, which form the equator of the full sphere, must be equal in magnitude and opposite in sign, and therefore add to zero. For a closed surface there are no free edges; or, equivalently, all of the edges of sub-areas are connected!

It is a mistake to imagine that the curl of a vector field says something about the apparent curvature of the field lines. This interpretation is unhelpful and misleading. The curl actually tells us about the line integral around a small area at a *point*. In this sense, it is a property of the field at a point.

The overall concept is called **Stokes's theorem**, and is applicable to any vector field, say \mathbf{A} . Stokes's theorem is usually written in the form

$$\oint_{\partial S} d\mathbf{l} \cdot \mathbf{A}(\mathbf{r}) = \int_S d\mathbf{S} \cdot \nabla \times \mathbf{A}(\mathbf{r}). \quad \text{Stokes's theorem} \quad (2.27)$$

where the line integral is taken around the free edge ∂S of the surface S used in the surface integral, see Fig. 2.9. Notice the use of a circle on the integral sign to denote that the path along



Sir George Gabriel Stokes (1819–1903)

which the line integral is evaluated is closed. Stokes's theorem is used frequently to convert forwards and backwards between line and surface integrals.

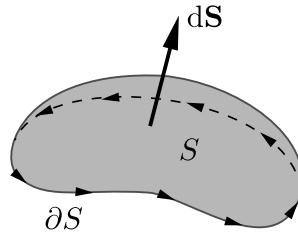


Figure 2.9: Illustration of a surface S with a boundary ∂S .

To finish let us briefly understand the consequences of Stokes's theorem for electrostatic fields by coming back to Sect. 2.4.

We certainly know that the line integral of the *electric field* around any closed loop is zero, as discussed previously in Sect. 2.4, and therefore the line integral around the free edge of any surface is zero. Hence, the left hand side of (2.27) is zero, and therefore because the surface integral must be zero for all surfaces we conclude

$$\nabla \times \mathbf{E}(\mathbf{r}) = \mathbf{0} \quad (2.28)$$

for any \mathbf{r} .

Be careful, the curl of a vector field is a vector field, and so the object on the RHS of (2.28) is the zero vector, not the zero scalar. Also note that, while the expression in (2.28) holds in electrostatics that we are concerned with here, it is not generally valid in electrodynamics.

Taking a step back, (2.28) can be expanded out using (2.20) to yield (2.19). Thus, we simply recover the system of three equations that connects the spatial derivatives of an electric field at a point. Eq. (2.28) is of substantial importance in electrostatics and should be remembered!

2.6 Electric Monopoles

An electric **monopole** comprises a single point charge q . Accordingly, the potential at any distance r' from the charge is given by

$$V(r') = - \int_{\infty}^{r'} dr \frac{q}{4\pi\epsilon_0 r^2} = \frac{q}{4\pi\epsilon_0 r'}, \quad (2.29)$$

where the reference position for zero potential is conveniently chosen at infinity.

It is now possible to calculate the electric field from the potential. In spherical polar coordinates, the vector differential operator ∇ is given by

$$\nabla \equiv \left(\hat{r} \frac{\partial}{\partial r} + \hat{\theta} \frac{1}{r} \frac{\partial}{\partial \theta} + \hat{\phi} \frac{1}{r \sin \theta} \frac{\partial}{\partial \phi} \right). \quad (2.30)$$

Because the potential does not depend on θ or ϕ , the last two terms of ∇V evaluate to zero, and so

$$\mathbf{E}(\mathbf{r}) = -\nabla V(\mathbf{r}) = -\hat{r} \frac{dV(r)}{dr} = \hat{r} \frac{q}{4\pi\epsilon_0 r^2}, \quad (2.31)$$

as expected.

The field lines are radial, and therefore the equipotential surfaces are spherical, cf. Fig. 2.10. It is seen that the charge can be regarded as the ‘source’ of the field lines. Indeed, field lines always begin and end on charge.

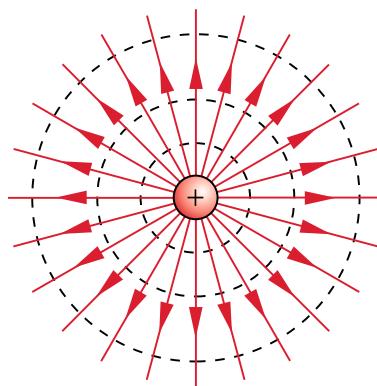


Figure 2.10: The electric field of an electric monopole q with equipotential contours. Note that the field lines are radial and that the equipotential contours are always orthogonal to the field lines. Source: Demtröder.

2.7 Dipoles

One of the most important arrangements of charge is called a **dipole**. A dipole comprises two point charges of opposite sign that are placed in close proximity to each other. Overall, the system has no *net* charge, but because the charges are separated, albeit by some small amount, an electric field exists around them. In the following, we want to understand what is the form of this field?

Consider the geometry shown in Fig. 2.11, and assume that the two charges are separated by a distance a .

To find the total electric field at any point, we merely add the electric fields associated with the individual charges according to the superposition principle:

$$\mathbf{E}(\mathbf{r}) = \mathbf{E}_1(\mathbf{r}) + \mathbf{E}_2(\mathbf{r}). \quad (2.32)$$

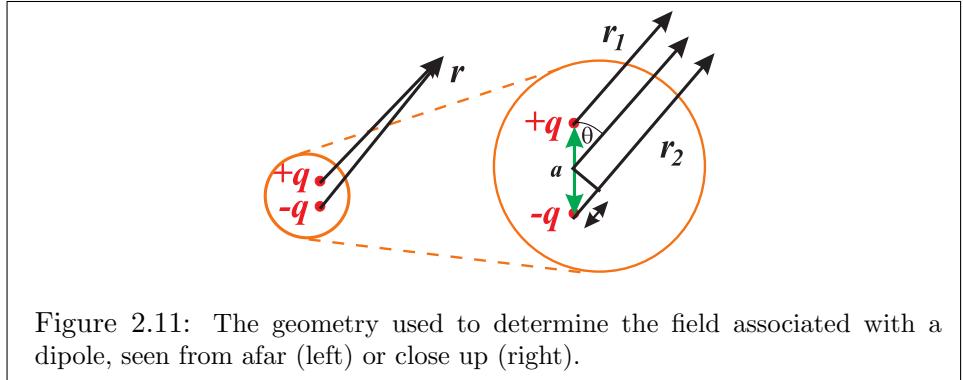


Figure 2.11: The geometry used to determine the field associated with a dipole, seen from afar (left) or close up (right).

This means that the potentials add also. In other words the potential difference between two points is the sum of the potential differences that would exist as a result of the charges acting separately.

Thought for the day 2: Always be cautious of loose statements like this one. We know that the forces will add if several charges are collected together, because that is an experimental fact, but we have not proven that potentials add. Prove this mathematically. Does it matter if the individual potential fields have different reference positions?

We therefore have

$$V(\mathbf{r}) = V_1(\mathbf{r}) + V_2(\mathbf{r}). \quad (2.33)$$

At large distances from the dipole, where “large” means $r = |\mathbf{r}| \gg a$,

$$\begin{aligned} r_1 &\approx r - \frac{a}{2} \cos \theta \\ r_2 &\approx r + \frac{a}{2} \cos \theta. \end{aligned} \quad (2.34)$$

The total potential is

$$V(r_1, r_2) = \frac{q}{4\pi\epsilon_0 r_1} - \frac{q}{4\pi\epsilon_0 r_2}, \quad (2.35)$$

and therefore

$$\begin{aligned} V(r, \theta) &= \frac{q}{4\pi\epsilon_0 r} \\ &\left[\left(1 - \frac{a}{2r} \cos \theta\right)^{-1} - \left(1 + \frac{a}{2r} \cos \theta\right)^{-1} \right], \end{aligned} \quad (2.36)$$

which becomes, after working out the term in square brackets and remembering $r \gg a$,

$$V(r, \theta) = \frac{1}{4\pi\epsilon_0} \cdot \frac{qa \cos \theta}{r^2}. \quad (2.37)$$

The potential at large distances compared with the separation, $|r| \gg a$, depends only on the product qa . It does not depend on the absolute value of q or a separately.

The **electric dipole moment** is defined as the product

$$\mathbf{p} \equiv qa, \quad \text{Electric dipole moment} \quad (2.38)$$

where \mathbf{a} is defined as the vector connecting the negative to the positive charge. In other words $\hat{\mathbf{p}}$ is defined as pointing to the positive charge.

We write for $p \equiv |\mathbf{p}|$

$$V(r, \theta) = \frac{1}{4\pi\epsilon_0} \cdot \frac{p \cos \theta}{r^2}. \quad (2.39)$$

Although we do not know the field close to the charges, if the charges are separated by some tiny amount then we have a simple functional form for the field at large distances. It should be noticed the potential falls as $1/r^2$, as opposed to the $1/r$ dependence for a monopole.

Now calculate the electric field, which is best done using spherical polar coordinates:

$$\mathbf{E}(\mathbf{r}) = -\nabla V(\mathbf{r}) = -\left[\hat{\mathbf{r}} \frac{\partial V(r, \theta)}{\partial r} + \hat{\theta} \frac{1}{r} \frac{\partial V(r, \theta)}{\partial \theta} \right], \quad (2.40)$$

which evaluates to

$$\mathbf{E}(r, \theta) = \frac{1}{4\pi\epsilon_0} \frac{p}{r^3} \left(\hat{\mathbf{r}} 2 \cos \theta + \hat{\theta} \sin \theta \right). \quad (2.41)$$

Notice that in the limit where the separation between the charges is reduced to zero, while keeping the charge constant, the field tends to zero, as would be expected. Also, for $\theta = 0$ the field is radial, whereas for $\theta = 90^\circ$ the field is azimuthal (cf. Fig. 2.5).

If the distance between the two charges is modulated sinusoidally at an angular rate ω , then

$$\mathbf{E}(r, \theta, t) = \frac{1}{4\pi\epsilon_0} \frac{p_0 \sin(\omega t)}{r^3} \left(\hat{\mathbf{r}} 2 \cos \theta + \hat{\theta} \sin \theta \right), \quad (2.42)$$

which shows that the electric field also varies sinusoidally in time.

Using the dipole definition from (2.38), the potential field generated by a dipole is

$$V(r, \theta) = \frac{1}{4\pi\epsilon_0} \frac{1}{r^2} \mathbf{p} \cdot \hat{\mathbf{r}}. \quad \text{Dipole potential} \quad (2.43)$$

2.8 Force on a dipole in a uniform field

We have already considered the force that is exerted on a charge, or monopole, when it is placed in an electric field, but what is the nature of the force that is exerted on a dipole when it is placed in a field?

Consider the setup in Fig. 2.12. Initially, assume that the field is uniform. Also, assume that the charges are held rigidly (at an invariant distance) with respect to each other. The situation where dipoles are *induced*, because charges are pulled apart by an applied field, will be considered later.

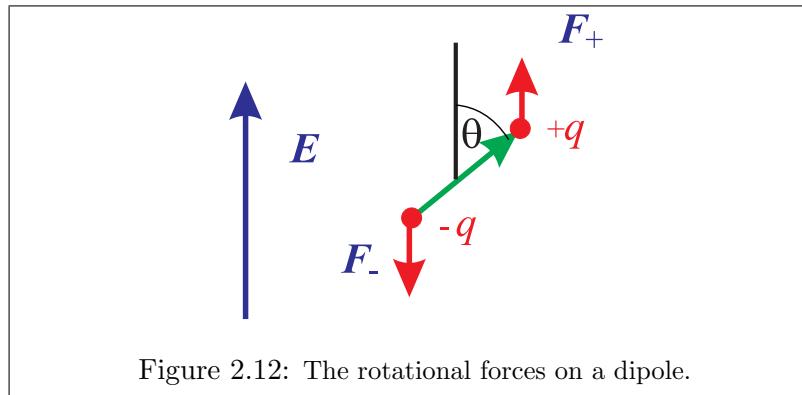


Figure 2.12: The rotational forces on a dipole.

Because the field is uniform, the individual charges experience the same force in opposite directions, and because they are held rigidly with respect to each other, there is no net translational force. They do, however, experience a *couple* that can lead to rotation.

If a system comprises two equal, but oppositely directed, parallel forces, then the product of the perpendicular distance between them and the magnitude of one of the rotational forces is called the **couple**. The axis of the couple is a vector that is perpendicular to the plane in which the two forces act. If the couple is denoted by \mathbf{G} then for the dipole

$$|\mathbf{G}| = q|\mathbf{E}|a \sin \theta = p|\mathbf{E}| \sin \theta, \quad (2.44)$$

where θ is the angle between the direction of \mathbf{E} and the direction of \mathbf{p} . In vector notation, we have

$$\mathbf{G} = \mathbf{p} \times \mathbf{E}. \quad \text{Couple of a dipole} \quad (2.45)$$

The couple becomes zero when the dipole is aligned with the field, showing the tendency of the dipole to align itself with the field.

2.9 Potential energy of a dipole in a uniform field

Previously, we considered the potential field created by a dipole; in other words the energy required to move a test charge around in

the vicinity of a dipole. We can now consider the dipole to be the test charge, and consider what happens as we move it around in a uniform field.

Because there is no net force on the dipole, displacement does not require any work to be done; rotation, however, does require work to be done.

Suppose that we apply an external couple to the dipole in order to increase the angle θ between the dipole \mathbf{p} and the field \mathbf{E} by an amount $d\theta$. The work done when increasing the angle of rotation by $d\theta$ such that the charge moves against the direction of the force is

$$dW = |\mathbf{G}(\theta)|d\theta, \quad (2.46)$$

which can be integrated to give the change in potential after rotation through some total angle:

$$U_{\theta\theta_0} = \int_{\theta_0}^{\theta} d\theta' |\mathbf{G}(\theta')| \stackrel{(2.44)}{=} \int_{\theta_0}^{\theta} d\theta' p|\mathbf{E}| \sin \theta' \quad (2.47)$$

$$= -p|\mathbf{E}|(\cos \theta - \cos \theta_0). \quad (2.48)$$

It is clear that the potential will be at a minimum when the dipole is fully aligned with the field, which occurs at $\theta = 0$, and at a maximum when the dipole is aligned with the field in the opposite direction, $\theta = \pi$.

Let us take the reference angle for the potential to be $\theta_0 = \pi/2$ so $\cos \theta_0 = 0$, and then the potential difference takes positive and negative values with respect to this reference. Thus we have simply

$$U = -\mathbf{p} \cdot \mathbf{E}, \quad (2.49)$$

as the **electrostatic potential associated with the dipole's orientation** (see Fig. 2.13).

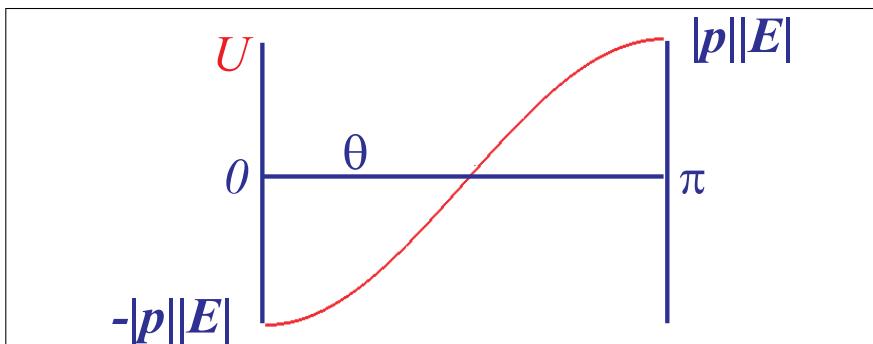
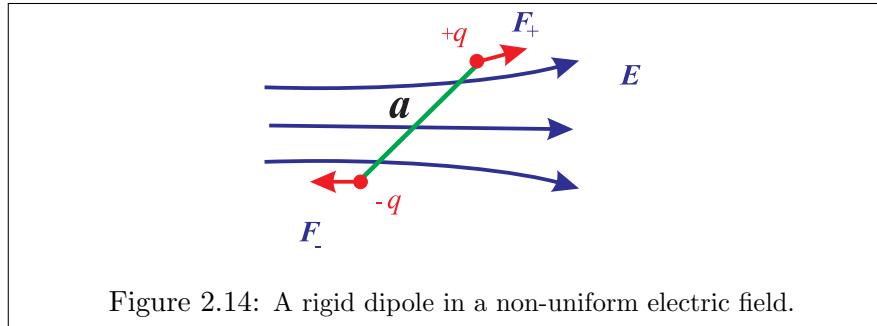


Figure 2.13: The potential of a dipole as it is rotated in a uniform field.

2.10 Force on a dipole in a non-uniform field

Now consider the slightly more challenging problem of finding the translational force that acts on a rigid dipole when it is placed in a non-uniform electric field (see Fig. 2.14).



The x -directed electric field at the position of the positive charge E_x^+ can be written to first order of a Taylor expansion as

$$E_x^+ = E_x^- + a_x \frac{\partial E_x}{\partial x} + a_y \frac{\partial E_x}{\partial y} + a_z \frac{\partial E_x}{\partial z}, \quad (2.50)$$

where E_x^- is the x -directed field at the position of the negative charge, and a_x , a_y , and a_z are the components of a in the principal directions.

The total translational force acting in the x direction is therefore

$$F_x = q (E_x^+ - E_x^-), \quad (2.51)$$

or

$$F_x = q \left(a_x \frac{\partial E_x}{\partial x} + a_y \frac{\partial E_x}{\partial y} + a_z \frac{\partial E_x}{\partial z} \right). \quad (2.52)$$

Similar expressions exist for the y and z directions, which can be collected together and expressed in terms of the components of the electric dipole moment:

$$\begin{aligned} F_x &= p_x \frac{\partial E_x}{\partial x} + p_y \frac{\partial E_x}{\partial y} + p_z \frac{\partial E_x}{\partial z} \\ F_y &= p_x \frac{\partial E_y}{\partial x} + p_y \frac{\partial E_y}{\partial y} + p_z \frac{\partial E_y}{\partial z} \\ F_z &= p_x \frac{\partial E_z}{\partial x} + p_y \frac{\partial E_z}{\partial y} + p_z \frac{\partial E_z}{\partial z}. \end{aligned} \quad (2.53)$$

This can be written more conveniently in

$$\mathbf{F} = (\mathbf{p} \cdot \nabla) \mathbf{E}, \quad (2.54)$$

or, in suffix notation, and remembering the summation convention for repeated suffices,

$$F_i = p_j \frac{\partial E_i}{\partial x_j}. \quad (2.55)$$

Using $\nabla \times \mathbf{E} = 0$ (2.19), (2.53) can be written

$$\begin{aligned} F_x &= p_x \frac{\partial E_x}{\partial x} + p_y \frac{\partial E_y}{\partial x} + p_z \frac{\partial E_z}{\partial x} \\ F_y &= p_x \frac{\partial E_x}{\partial y} + p_y \frac{\partial E_y}{\partial y} + p_z \frac{\partial E_z}{\partial y} \\ F_z &= p_x \frac{\partial E_x}{\partial z} + p_y \frac{\partial E_y}{\partial z} + p_z \frac{\partial E_z}{\partial z}, \end{aligned} \quad (2.56)$$

Or in suffix notation,

$$F_i = p_j \frac{\partial E_j}{\partial x_i}. \quad (2.57)$$

If the components of the dipole moment are constant ($\mathbf{p} = \mathbf{const}$, *fixed* dipole), they can be taken inside the derivatives, and so we have

$$\mathbf{F}(\mathbf{r}) = \nabla [\mathbf{p} \cdot \mathbf{E}(\mathbf{r})]. \quad (2.58)$$

Interestingly (2.58) can be written, using (2.49), in terms of the potential energy associated with rotation

$$\mathbf{F}(\mathbf{r}) = -\nabla U(\mathbf{r}). \quad (2.59)$$

Eq. (2.59) should not be confused with $\mathbf{E}(\mathbf{r}) = -\nabla V(\mathbf{r})$ (2.15), but it does have similarities in that $\mathbf{E}(\mathbf{r})$ is the force on unit charge, and it is given by minus the gradient of the electrostatic potential, whereas $\mathbf{F}(\mathbf{r})$ is the force on a dipole, even though its overall charge is neutral, and $U(\mathbf{r})$ is the rotational potential energy.

In other words, for a *fixed* dipole, the difference in potential is caused by the alignment between the field and the dipole changing, and the force is given by minus the gradient of this potential. Clearly, the dipole will tend to move to some position, where the field and dipole are aligned. This approximation ignores the ability of the dipole to rotate.

Our expression for the translational force on a dipole in a non-uniform field should be accompanied by a word of warning because it assumes that the orientation of the dipole is fixed. In reality, if the dipole were free to move, it would not only move in a translational way, but it would also rotate, and this rotation should be taken into account when calculating the force at the new point. Moreover, the dipole must be a *rigid* dipole, rather than an *induced* dipole.

2.11 Multipole expansions

Our analysis of the fields associated with monopoles and dipoles used the addition of scalar potentials to get the total potential, and then the gradient operator to get the electric field. It is often more convenient to add scalar fields, rather than vector fields, particularly when a large number of charges is involved.

Suppose that we have a large number of charges collected together in some finite region of space. Far from the region it seems reasonable that the electric field will correspond to that of a point charge equal to the net charge—the sum of the charges present. In this case the potential will decay as $1/r$ and the field as $1/r^2$.

Often, however, the region will have no net charge, but it may have separated charge that takes the form of a dipole. In this case, the potential will decay as $1/r^2$, and the field as $1/r^3$. The functional form of the field will therefore reveal the existence of dipole-like structure.

In reality, of course, the charge may be distributed in more complicated ways. Consider, for example, two dipoles pointing in opposite directions, and slightly displaced with respect to each other. This charge distribution is called a **quadrupole**, and it produces a quadrupole field.

It can be shown that the potential of a quadrupole decays as $1/r^3$, whereas the field decays as $1/r^4$.

	Potential	Electric Field
Monopole	$1/r$	$1/r^2$
Dipole	$1/r^2$	$1/r^3$
Quadrupole	$1/r^3$	$1/r^4$

The functional dependencies of the potential and field for different charge distributions.

This series can be continued to more and more complicated potentials and field distributions.

For any general distribution of charges, monopole, dipole, quadrupole and higher-order fields may all be present *simultaneously*.

Suppose that we have some general continuous charge distribution $\rho(\mathbf{r})$ occupying some finite region of space.

- The first, or monopole, moment (zeroth order) is then given by integrating over the volume V enclosing all the charge (or entire space)

$$Q = \int dV \rho(\mathbf{r}), \quad (2.60)$$

which is the net charge associated with the distribution.

- The second, or dipole moment (first order), is given by

$$\mathbf{p} = \int dV \mathbf{r} \rho(\mathbf{r}), \quad (2.61)$$

which is a vector quantity. Notice that it is not a function of position; it is a single vector quantity.

- The third, or quadrupole, moment (second order) is a tensor, which we shall not consider here because several definitions based on different normalisations are possible.

The potential functions associated with the multipole moments have special significance because they form a complete set of functions in terms of which any potential can be expanded. Even if the charge distribution is not known, but its potential distribution is known, then we can expand the potential in terms of a weighted linear combination of this special infinite set of functions. These special functions are called spherical harmonics, and you will learn more about them in the Mathematics and Quantum Mechanics courses.

2.12 Electric flux

We have studied the fields produced by certain, simple charge distributions. Usually, however, a charged body has vast numbers of individual charges associated with it, and it is not possible to consider the fields associated with each of the charges individually. We must therefore investigate models for determining the electrostatic fields associated with continuous distributions of charge. The notion of a continuous charge distribution, $\rho(\mathbf{r})$, was already touched on in the previous section.

An intuitive idea of the electric field strength is conveyed through electric field lines. We know that the packing of electric field lines seems to vary from position to position. In some regions fields lines are packed more tightly than others, which intuitively conveys a relatively higher electric field strength. However, this observation has no mathematical substance because the number of field lines is determined solely by how many we choose to draw!

It is reasonable, however, to consider the quantity

$$\int_S d\mathbf{S} \cdot \mathbf{E}(\mathbf{r}), \quad \text{Electric flux} \quad (2.62)$$

which exactly the concept we are looking for, as it is clearly intimately related to the ‘number of field lines’ passing normally through the surface S . If the vector field in this equation were to correspond to electric current density, then the integral of the normal component of the vector current density would give the total amount of current flowing through the surface.

Once the electric field lines are established it is inevitable that we can define and calculate a quantity that we might choose to call the **electric flux**. Unlike current, it would be a mistake to relate it to the flow of anything.

Ultimately, we are interested measurable quantities, e.g., in the forces acting on systems of charged particles, and the energy transferred between systems of charged particles. **The electric flux enters merely as a (useful) mathematical concept that is an inevitable consequence of having established the notion of field lines.**

2.13 The divergence of a field – div

We can nevertheless explore the properties of electric flux, and establish whether it provides a useful tool for modelling the behaviour of charged systems. To make physical sense of the flux, we must introduce the third and final use of the vector differential operator ∇ .

The **divergence** of a vector field, $\mathbf{E}(\mathbf{r})$, is defined by

$$\nabla \cdot \mathbf{E}(\mathbf{r}) = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right) \cdot (E_x, E_y, E_z), \quad (2.63)$$

which becomes

$$\nabla \cdot \mathbf{E}(\mathbf{r}) = \frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z}. \quad \text{Divergence}$$

(2.64)

The divergence of a vector field generates a scalar field.

Consider a small volume, in a Cartesian coordinate system, having side lengths dx , dy , and dz . The total net “flow of field lines”, i.e., $d\mathbf{S} \cdot \mathbf{E}(\mathbf{r})$, through the opposite sides of the volume in the x -direction is

$$\left(E_x + \frac{\partial E_x}{\partial x} dx \right) dy dz - E_x dy dz, \quad (2.65)$$

which gives

$$\frac{\partial E_x}{\partial x} dx dy dz. \quad (2.66)$$

The total net outflow from all 6 sides is

$$\left(\frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z} \right) dx dy dz. \quad (2.67)$$

The divergence of a vector field at a point gives the net outflow from the point per unit volume.

The concept of net outflow per unit volume may seem a little strange, but consider two small cubic volumes placed next to each other. Because the outflows from each of the areas on the touching sides are in opposite directions, and therefore sum to zero, the total net outflow is the sum of the two divergences from each of the volumes. Thus we can find the total net outflow through the surface of some complicated macroscopic shape by evaluating a volume integral (see Fig. 2.15).

We are then left with the **divergence theorem**,

$$\oint_{S=\partial V} d\mathbf{S} \cdot \mathbf{E}(\mathbf{r}) = \int_V dV \nabla \cdot \mathbf{E}(\mathbf{r}). \quad \text{Divergence theorem}$$

(2.68)

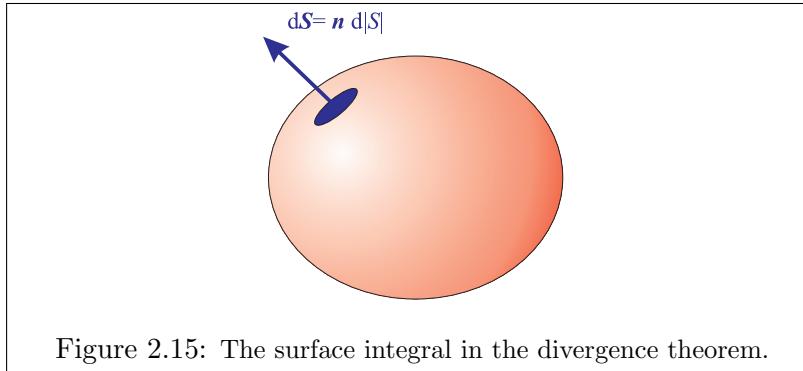


Figure 2.15: The surface integral in the divergence theorem.

The divergence theorem is applicable to any vector field, and it relates the surface integral of the normal component of the vector field, over some closed surface, to the volume integral of the divergence of the vector field throughout the volume. It is clear that the divergence of a vector field is intimately related to the existence of sources, because it indicates regions of space where more field lines travel out through a surface than enter the surface. For example, a spherical surface around a point charge. Notice the use of the circle on the integral sign of (2.68) to indicate that the surface is closed.

2.14 Example – charge conservation

Here is an important example of the use of the divergence theorem.

Consider the arbitrary volume shown (Fig. 2.16). At any point on the surface let the current density be \mathbf{J} (the current per unit area perpendicular to the direction of current flow). Let an infinitesimal element of area on the surface be $d\mathbf{S}$. Then the current through a surface element is $d\mathbf{S} \cdot \mathbf{J}$.

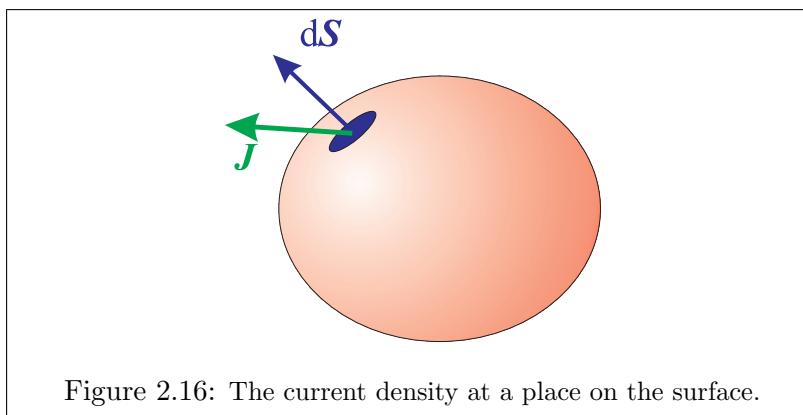


Figure 2.16: The current density at a place on the surface.

The total charge enclosed is $\int dV \rho$, where ρ is the charge density per unit volume. Conservation of charge implies that the total current

$$\begin{aligned} I &= \oint d\mathbf{S} \cdot \mathbf{J} = -\frac{\partial}{\partial t} \int dV \rho \\ &= \int dV \nabla \cdot \mathbf{J}. \end{aligned} \quad (2.69)$$

The result is true for *any* volume, and so we arrive at the **continuity equation**:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{J} = 0 \quad \text{Continuity equation} \quad (2.70)$$

This is a very important formula and well worth remembering!

2.15 The divergence theorem for electric fields

The discussion in Sect. 2.13 has been in the context of any vector field; we can now explore how the divergence theorem can be applied specifically to electrostatic fields.

Consider a single point charge, and integrate the normal component of the electric field over a spherical surface that has the charge at its centre:

$$\oint_S d\mathbf{S} \cdot \mathbf{E}(\mathbf{r}) = \oint_S d\mathbf{S} \cdot \hat{\mathbf{r}} \frac{q}{4\pi\epsilon_0 r^2} = 4\pi r^2 \frac{q}{4\pi\epsilon_0 r^2} = \frac{q}{\epsilon_0}. \quad (2.71)$$

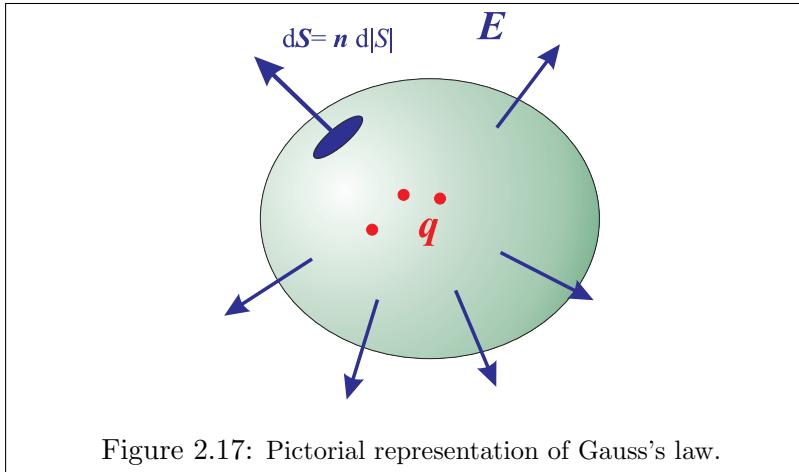
The integral of the electric field over the surface is related to the charge enclosed. This expression turns out to be true regardless of where the charge is located within the surface, or indeed the shape of the surface. It is certainly clear in the above case that the size of the sphere does not change the value of the integral. These facts indicate that charge can be regarded as the ‘source’ of electric field lines.

If a number of charges are present within some volume then it follows straightforwardly that

$$\oint_S d\mathbf{S} \cdot \mathbf{E}(\mathbf{r}) = \sum_i \frac{q_i}{\epsilon_0}, \quad \text{Gauss's Law for discrete charges,} \quad (2.72)$$

again indicating that electric charge is the source of electric flux (see Fig. 2.17).

Thought for the day 3: How can we prove that (2.71) is true regardless of the shape of the surface, and the position of the charge? Have a go. Place a point charge inside an arbitrarily shaped surface. Calculate the normal component of the electric field at some point, surrounded by some infinitesimally small region that forms the base of a cone whose vertex is at the position of the charge. Project the base of the cone onto the surface, and, using the notion of solid angle, integrate the result to find the answer.



Now assume that the charge is distributed in some continuous manner according to the charge density distribution $\rho(\mathbf{r})$, which of course is a scalar field. The units of $\rho(\mathbf{r})$ are C m^{-3} .

We can write the equivalent of (2.72) as:

$$\oint_{S=\partial V} d\mathbf{S} \cdot \mathbf{E}(\mathbf{r}) = \frac{1}{\epsilon_0} \int_V dV \rho(\mathbf{r}) \quad \text{Gauss's Law (integral form)} \quad (2.73)$$



Carl Friedrich Gauss
(1777–1855)

From the divergence theorem we get

$$\int_V dV \nabla \cdot \mathbf{E}(\mathbf{r}) = \oint_S d\mathbf{S} \cdot \mathbf{E}(\mathbf{r}) = \frac{1}{\epsilon_0} \int_V dV \rho(\mathbf{r}), \quad (2.74)$$

but because this must be true for all possible volumes

$$\nabla \cdot \mathbf{E}(\mathbf{r}) = \frac{\rho(\mathbf{r})}{\epsilon_0} \quad \text{Gauss's Law (differential form)} \quad (2.75)$$

Eqs. (2.73) and (2.75) are equivalent expressions of **Gauss's law**.

As an aside, it is common practice to introduce a new quantity called the **electric displacement** or **electric flux density** of free space:

$$\mathbf{D}(\mathbf{r}) = \epsilon_0 \mathbf{E}(\mathbf{r}). \quad \text{Electric displacement} \quad (2.76)$$

It will be seen later that, with some minor modification, $\mathbf{D}(\mathbf{r})$ is an especially useful quantity when insulating materials are introduced into electrostatic systems.

We can write down Gauss's law in terms of the electric displacement equivalently to Eq. (2.73) and (2.75) that are given in terms of the electric field:

$$\begin{aligned} \oint_{S=\partial V} d\mathbf{S} \cdot \mathbf{E}(\mathbf{r}) = \frac{1}{\epsilon_0} \int_V dV \rho(\mathbf{r}) &\Leftrightarrow \oint_{S=\partial V} d\mathbf{S} \cdot \mathbf{D}(\mathbf{r}) = \int_V dV \rho(\mathbf{r}), \\ \nabla \cdot \mathbf{E}(\mathbf{r}) = \frac{\rho(\mathbf{r})}{\epsilon_0} &\Leftrightarrow \boxed{\nabla \cdot \mathbf{D}(\mathbf{r}) = \rho(\mathbf{r})}. \end{aligned} \quad (2.77)$$

2.16 Maxwell's equation for electrostatic fields

We have now established the two most important equations in electrostatics, which will eventually become two of Maxwell's equations when we allow charge to move with time.

In summary, we have

$$\boxed{\nabla \cdot \mathbf{E}(\mathbf{r}) = \frac{\rho(\mathbf{r})}{\epsilon_0}. \quad \text{Maxwell's equation 1}} \quad (2.78)$$

and

$$\boxed{\nabla \times \mathbf{E}(\mathbf{r}) = \mathbf{0} \quad \text{Maxwell's equation 2 in electrostatics}} \quad (2.79)$$

Remember these.

Note that Maxwell's equation 2 in (2.79) is valid in electrostatics only.

2.17 Applications of Gauss's law

Let us use Gauss's law to derive the field distributions of certain simple geometries. The geometries are simple in that it is straightforward to identify surfaces of integration where the field lines are everywhere normal, thus simplifying evaluation.

- **A uniform sheet of charge:**

Suppose that we have an infinitely large, uniform sheet of charge. Intuition would suggest that the field lines are everywhere normal to the sheet – consider adding up the fields from a sheet of point charges. Also, if this were not the case, charges in the surface would move until there was no component of the force in the surface.

Now enclose a surface with a rectangular box. By Gauss's law

$$2A|\mathbf{E}(\mathbf{r})| = \frac{\sigma A}{\epsilon_0}, \quad (2.80)$$

where σ is the charge per unit *area*. The field is uniform, and

$$|\mathbf{E}| = \frac{\sigma}{2\epsilon_0}. \quad (2.81)$$

We conclude

$$\boxed{\mathbf{E} = \frac{\sigma}{2\epsilon_0} \hat{\mathbf{n}}.} \quad (2.82)$$

The field does not depend on the distance to the surface, as expected.

Thought for the day 4: How does this expression compare with the field between two parallel sheets of charge of opposite sign? Again, consider adding vectorially the field lines associated with individual charges, and the result is obvious.

- **A uniform line of charge:**

In the case of a uniform line of charge, applying Gauss's law to a surface of coaxial shape with length L gives

$$L2\pi r |\mathbf{E}(r)| = \frac{\lambda L}{\epsilon_0}, \quad (2.83)$$

where λ is the charge per unit length, or

$$|\mathbf{E}(r)| = \frac{\lambda}{2\pi\epsilon_0 r}. \quad (2.84)$$

Finally,

$$\mathbf{E}(r) = \frac{\lambda}{2\pi\epsilon_0 r} \hat{\mathbf{r}}. \quad (2.85)$$

Remember that Gauss's law requires the surface of integration to be closed, but because field lines do not cross the ends of the cylinder of integration, only the curved surface contributes.

The potential difference between a point at radius r_0 and one at radius r can now be found through

$$V(r) = - \int \mathbf{E} \cdot d\mathbf{l} = - \int_{r_0}^r \frac{\lambda}{2\pi\epsilon_0 r'} dr', \quad (2.86)$$

giving

$$V(r) = - \frac{\lambda}{2\pi\epsilon_0} [\ln(r) - \ln(r_0)], \quad (2.87)$$

i.e.,

$$V(r) = - \frac{\lambda}{2\pi\epsilon_0} \ln(r/r_0). \quad (2.88)$$

- **A coaxial cable:**

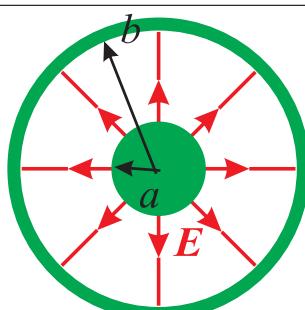


Figure 2.18: A coaxial cable.

Now consider a coax cable, which consists of two coaxial cylinders as shown in Fig. 2.18, where the inner and outer radii of the intervening region are a and b , respectively, and the voltage on the central conductor relative to the outer conductor is V . Then the capacitance per unit length is

$$C \equiv \lambda/V = \frac{2\pi\epsilon_0}{\ln(b/a)}. \quad (2.89)$$

This equation will be useful later when considering a coax cable used as a transmission line.

- **A point charge:**

Although we have already considered point charges, it is instructive to go backwards and to derive the field using Gauss's law. In the case of a point charge, the field lines are radial, and integrating over a spherical surface and using Gauss's law, we have

$$4\pi r^2 |\mathbf{E}(r)| = \frac{q}{\epsilon_0}, \quad (2.90)$$

giving

$$|\mathbf{E}(r)| = \frac{q}{4\pi\epsilon_0 r^2}, \quad (2.91)$$

as expected. Therefore

$$\boxed{\mathbf{E}(r) = \frac{q}{4\pi\epsilon_0 r^2} \hat{\mathbf{r}}.} \quad (2.92)$$

Easy!

2.18 Laplace's and Poisson's Equations

Up to now, we have concentrated on the potential at a point, albeit at an arbitrary one, but we might wonder about what limits the smoothness of the potential function; in other words, how is the potential at one point related to the potential at a neighbouring point? Clearly the answer will depend on the underlying charge distribution. To address this problem we need to find a differential equation for the potential.

The solution is straightforward. Using vector differential calculus

$$\begin{aligned} \nabla \cdot \mathbf{E}(\mathbf{r}) &= \nabla \cdot (-\nabla V) \\ &= -\left(\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2}\right) \\ &= -\nabla^2 V(\mathbf{r}), \end{aligned} \quad (2.93)$$

which follows straightforwardly by substituting the definition of grad into the definition of div.

Thus we get **Poisson's equation** from the first Maxwell's equation (2.78):

$$\nabla^2 V(\mathbf{r}) = \frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2} = -\frac{\rho(\mathbf{r})}{\epsilon_0}. \quad \text{Poisson's equation}$$
(2.94)

We have a differential equation, which relates the ‘curvature’ of the potential function to the local charge density. If the charge distribution is known, and the boundary conditions on the potential are known, the full potential function can be constructed by solving (2.94).

A special case arises when no charge is present, in which case Laplace's equation holds, which is simply

$$\nabla^2 V(\mathbf{r}) = \frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2} = 0. \quad \text{Laplace's equation}$$
(2.95)

When the source term is set to zero in this way, the equation is called a *homogeneous* differential equation.

Poisson's and, where appropriate, Laplace's equations are of fundamental importance; indeed, every electrostatics problem has a solution that is a solution of Poisson's equation.



Siméon-Denis Poisson
(1781–1840)



Pierre-Simon de
Laplace (1749–1827)

Thought for the day 5: Explain why there cannot be any electric field inside an empty, perfectly conducting metallic box. A container that has this effect is called a ‘Faraday cage’; can you give a practical example of a Faraday cage?

2.19 An introduction to Green's functions

Numerous elegant techniques have been developed over the years for solving (2.94) and (2.95). The most important formulation of the solution, which is easy to prove, is

$$V(\mathbf{r}) = \frac{1}{\epsilon_0} \int G(\mathbf{r}, \mathbf{r}') \rho(\mathbf{r}') d^3 r', \quad (2.96)$$

where $d^3 r'$ is simply a different way of writing $dV' = dx' dy' dz'$. $G(\mathbf{r}, \mathbf{r}')$ is called the **Green's function** of the differential equation.

The Green's function of the Poisson equation is the solution of the equation

$$\nabla^2 G(\mathbf{r}, \mathbf{r}') = -\delta(\mathbf{r} - \mathbf{r}') \quad (2.97)$$

that satisfies ‘homogeneous’ boundary conditions (i.e., some linear combination of G and its first derivative is *zero* at each point on the boundary). $\delta(\mathbf{r} - \mathbf{r}')$ is the Dirac delta function, and it represents

a point charge located at \mathbf{r}' , contained within an infinitely small region.

Once the Green's function is known for the system of interest, defined in terms of the boundary conditions, the potential function can be found for any distribution of charge within the system.

Thought for the day 6: Show that, once the Green's function has been found by solving (2.97) for particular homogeneous boundary conditions, the potential field for any charge distribution with those same boundary conditions is given by (2.96). [You can then cope with inhomogeneous b.c.'s by adding the complementary function satisfying those b.c.'s.]

Equation 2.96 shows that the potential at a point is equal to the sum of the potentials that would exist at the point as a consequence of the charge density distribution being divided into infinitesimally small volumes that contain the appropriate infinitesimal amounts of charge. For example, for free space,

$$V(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d^3 r', \quad (2.98)$$

which can be compared with (2.96) to give

$$G(\mathbf{r}, \mathbf{r}') = \frac{1}{4\pi|\mathbf{r} - \mathbf{r}'|}. \quad (2.99)$$

Eq. the right boundary conditions for this problem (free space), as it goes to zero as $|\mathbf{r}| \rightarrow \infty$. In general, it is clear that, since the boundary conditions must be satisfied by $G(\mathbf{r}, \mathbf{r}')$ for any fixed \mathbf{r}' , (2.96) states that the overall potential is a weighted linear combination of these Green's functions.

2.20 Boundary conditions and uniqueness

In order to find a solution to Poisson's equation, we need to know the charge distribution and certain information about the potential over the boundary that encloses the region of interest, but what constitutes a complete set of information for Poisson's equation?

In the case of a one-dimensional equation we might know the potential at each end of the appropriate interval, the potential and the derivative of the potential at one end, or the potential at one end and the derivative at the other.

What happens in the case of a closed three-dimensional surface?

Aside: define terminology (valid for any differential equation):

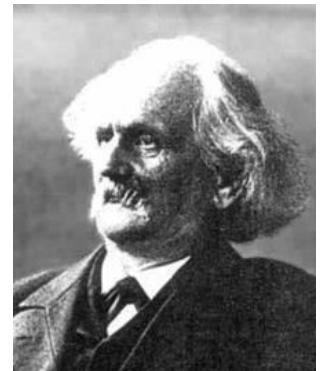
- If the quantity of interest is specified over the boundary, the boundary condition is called **Dirichlet**.
- If the normal derivative of the quantity of interest is specified over the boundary, the boundary condition is called **Neumann**.
- Mixed boundary conditions, where at any place on the boundary either the function or its normal derivative is specified, are called **Cauchy**.

In the case of Poisson's equation:



Peter Gustav Lejeune
Dirichlet (1805–1859)

- The boundary conditions are well-posed if *either* Dirichlet *or* Neumann boundary conditions are specified over the whole surface surrounding the region. Part or all of the surface may be at infinity.
- On a closed surface, the problem is overdetermined if *both* Dirichlet *and* Neumann conditions are specified.
- If Neumann conditions are specified, the problem is undetermined to within an additive constant, which, as we have seen, is one of the features of an electrostatic potential, where only potential differences are of significance.



Carl Gottfried Neumann (1832–1925)

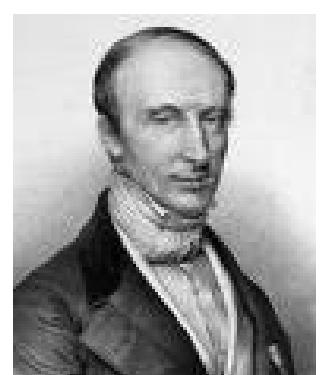
In the case of electrostatics where the system is often unbounded, we require the solution to tend to zero, or at least some finite value, at infinity. In the case of a perfect conductor, where charge spreads itself out to keep the electric field normal to the surface, the potential must be constant over the surface, which provides important information about the boundary. The fact that there are boundary conditions guarantees that the solution is unique (perhaps to within an additive constant).

Suppose that we find a function that satisfies Poisson's equation, and that is consistent with the known, fixed charge distribution; is the solution unique, or is it merely one of a number of possible solutions?

For the boundary conditions specified above, the solution is unique, and therefore it does not matter how it is found. The uniqueness of the solution can be proved as follows.

Suppose that we claim to have two solutions of Poisson's equation. Let us call these two solutions $V(\mathbf{r})$ and $U(\mathbf{r})$. Then

$$\begin{aligned}\nabla^2 V(\mathbf{r}) &= -\frac{\rho(\mathbf{r})}{\epsilon_0} \\ \nabla^2 U(\mathbf{r}) &= -\frac{\rho(\mathbf{r})}{\epsilon_0}.\end{aligned}\tag{2.100}$$



Augustin-Louis Cauchy (1789–1857)

It follows that $\Phi(\mathbf{r}) = V(\mathbf{r}) - U(\mathbf{r})$ is a solution of Laplace's equation, because subtracting the two equations in (2.100) gives

$$\text{From (2.100): } \nabla^2 \Phi(\mathbf{r}) \equiv \nabla^2 V(\mathbf{r}) - \nabla^2 U(\mathbf{r}) = 0 \quad (2.101)$$

throughout the region.

For the proof, first consider the case where Dirichlet boundary conditions are specified, $\Phi(\mathbf{r}) = 0$ on the boundary. These follow because the boundary conditions and the charge distributions are the same for both 'solutions'.

We can now use the vector differential identity

$$\nabla \cdot (\Phi \nabla \Phi) = |\nabla \Phi|^2 + \Phi \nabla^2 \Phi, \quad (2.102)$$

which can be verified by substitution. Both sides of this equation are scalar fields, and so we can integrate over the total volume to give

$$\int dV \nabla \cdot (\Phi \nabla \Phi) = \int dV |\nabla \Phi|^2 + \int dV \Phi \nabla^2 \Phi \quad (2.103)$$

$$= \int_S d\mathbf{S} \cdot (\Phi \nabla \Phi) \quad (2.104)$$

by the divergence theorem.

We know, however, that $\nabla^2 \Phi(\mathbf{r}) = 0$ throughout the region, giving

$$\int dV |\nabla \Phi|^2 = \int_S d\mathbf{S} \cdot (\Phi \nabla \Phi). \quad (2.105)$$

If the Dirichlet boundary conditions are satisfied by both solutions, $\Phi(\mathbf{r}) = 0$ on the boundary, and

$$\int dV |\nabla \Phi|^2 = 0. \quad (2.106)$$

However, because this is the volume integral of a positive quantity, we must have

$$|\nabla \Phi(\mathbf{r})| = 0 \quad (2.107)$$

at all points.

Finally, because the gradient of $\Phi(\mathbf{r})$ is everywhere zero, and $\Phi(\mathbf{r})$ is zero on the boundary, $\Phi(\mathbf{r})$ must be zero throughout the whole region. This proves $V(\mathbf{r}) = U(\mathbf{r})$ everywhere, i.e., the two solutions are therefore identical, q.e.d.

If Neumann boundary conditions had been specified, then according to (2.105) the normal component of $\nabla \Phi$ would be zero on the boundary, and we would have arrived at the same conclusion. In this case, however, the relationship between the solutions is only constrained to within a constant factor k : in other words, $V(\mathbf{r}) - U(\mathbf{r}) = k$, because only the normal derivative of Φ was specified to be zero.

It can also be appreciated that mixed boundary conditions could be given, as long as *either* the gradient *or* the potential is specified at each point. In this case the solution is unique, and referenced to the given potentials.

2.21 Conducting sphere in a uniform electric field

We can build up quite complicated field distributions by adding together the fields associated with different charge distributions; one interesting example comprises a dipole in, and aligned to, a uniform electric field.

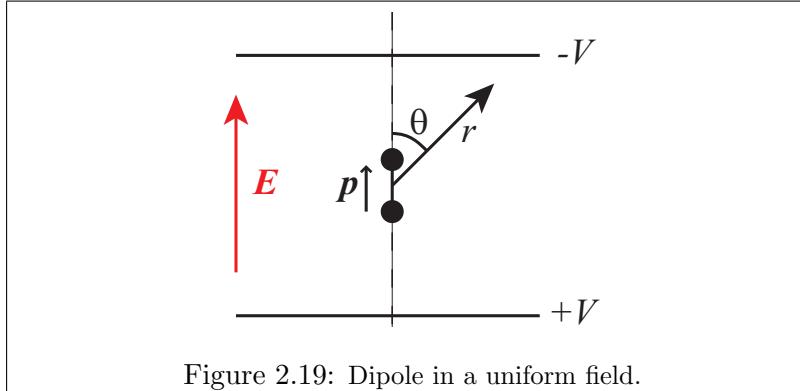


Figure 2.19: Dipole in a uniform field.

Suppose that we place a tiny dipole between two parallel, large sheets of charge; or equivalently between the plates of a large, parallel-plate capacitor (see Fig. 2.19). Assume that a , the size of the dipole, is very much smaller than the distance between the plates. Without the dipole, the potential associated with the parallel plates is uniform, and is given by

$$V_{\text{capacitor}}(z) \stackrel{\text{(Euclidean coord.)}}{=} -E_0 z \stackrel{\text{(polar coord.)}}{=} -E_0 r \cos \theta, \quad (2.108)$$

where $z = 0$, the centre of the coordinate system, is taken to be centred on the dipole. The field is zero outside the plates, and a long distance away from the sides of the plates.

With the dipole within the plates, and with the plates a long way from the dipole, the total potential is given by the sum of the two potentials:

$$\begin{aligned} V(\mathbf{r}) &= V_{\text{capacitor}}(\mathbf{r}) + V_{\text{dipole}}(\mathbf{r}) \\ &\stackrel{(2.39)}{=} -E_0 r \cos \theta + \frac{1}{4\pi\epsilon_0} \frac{p \cos \theta}{r^2} = \left(-E_0 r + \frac{p}{4\pi\epsilon_0 r^2} \right) \cos \theta. \end{aligned} \quad (2.109)$$

Since both contributions are proportional to $\cos \theta$, the potential is zero for

$$r = \left(\frac{p}{4\pi\epsilon_0 E_0} \right)^{1/3}, \quad (2.110)$$

which is a spherical surface ($r = \text{const}$, no θ dependence). Thus, there is a spherical surface in the system over which the potential is constant. We could, therefore, place a spherical conductor on this surface without changing the form of the field.

It is possible to reverse the argument and say that if we place a spherical conductor of radius s between two parallel plates, then

we can model the field by superposing a dipole field, having dipole moment

$$p = \underbrace{4\pi\epsilon_0 s^3}_{\alpha} E_0, \quad (2.111)$$

and a uniform field.

It is as if the applied field has induced a dipole moment, where the constant of proportionality, α , called the **polarisability**, is given by

$$\alpha = 4\pi\epsilon_0 s^3. \quad (2.112)$$

Note that the dipole moment is actually a vector quantity that points in the same direction as the field, and therefore we can write, more completely,

$$\mathbf{p} = \alpha \mathbf{E}_0. \quad \text{Dipole moment def'd through polarisability}$$

(2.113)

Although the applied field has indeed induced this dipole moment, it is not actually created by a small displacement of charge at the origin, but is caused by the separation of charge on the surface of the spherical conductor.

Later we shall study *induced* dipoles in more detail. In general, in certain materials, the induced dipole moment does not have to be in the same direction as the applied field, and then it is necessary to define a more complicated object, called the polarisability *tensor*. Indeed it is easy to imagine a system where charge is able to separate in one direction, but not in the orthogonal direction, and therefore if a field is applied at 45° , the field and the induced dipole will be in different directions.

We can now calculate the electric field. The potential in (2.109) can be written as

$$V(\mathbf{r}) = E_0 \cos \theta \left(\frac{s^3}{r^2} - r \right). \quad (2.114)$$

Note that the field dependence on r is determined by the radius of the sphere s that in turn defines the strength of the dipole p through (2.110).

Calculate the electric field now:

$$E_r(\mathbf{r}) = -\frac{\partial V(\mathbf{r})}{\partial r} = E_0 \cos \theta \left(\frac{2s^3}{r^3} + 1 \right),$$

(2.115)

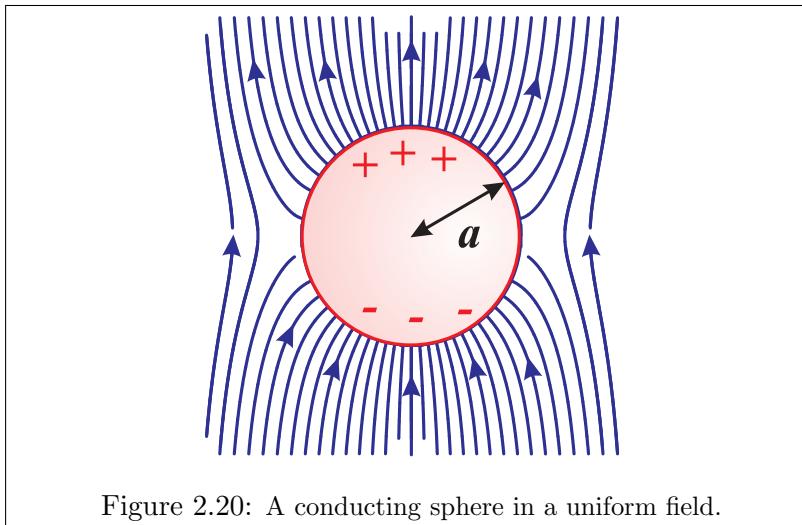
and

$$E_\theta(\mathbf{r}) = -\frac{1}{r} \frac{\partial V(\mathbf{r})}{\partial \theta} = E_0 \sin \theta \left(\frac{s^3}{r^3} - 1 \right).$$

(2.116)

It follows that at $r = s$, $E_\theta = 0$ which confirms that the field lines are orthogonal to the surface of the sphere, and $E_r = 3E_0 \cos \theta$ (see Fig. 2.20). We can also calculate the surface charge σ through $\sigma = \epsilon_0 E_r$, giving

$$\sigma = 3\epsilon_0 E_0 \cos \theta. \quad (2.117)$$



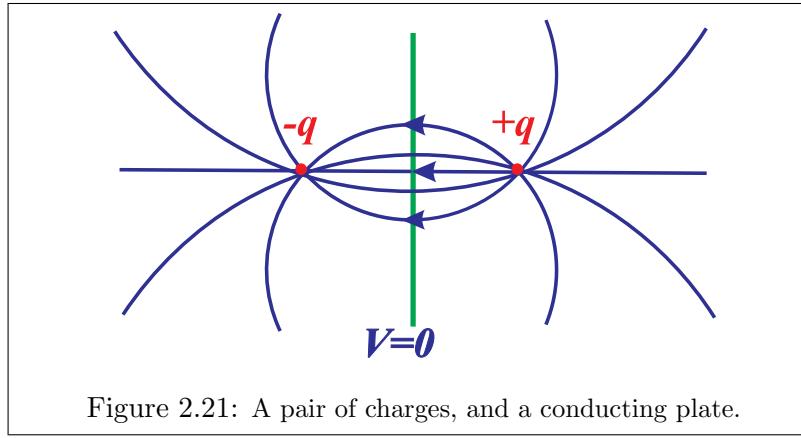
Thought for the day 7: Although charge, flux density, electric field, and potential are easy to describe mathematically, it is always beneficial to plot them as a function of position. For a conducting sphere in a parallel field, plot the key electrostatic quantities along a line through the centre of the system, $(-\infty < r < \infty)$, for $\theta = 0$. Show the separate contributions from the plates and the induced dipole, and for the combination.

2.22 Method of images

In the previous section we superposed two charge distributions to create a total potential that could be used to determine the total electric field. It turned out that there was a spherical surface in the system over which the potential was constant, and therefore it was possible to place a conducting sphere at that surface without changing the form of the field. According to the uniqueness theorem the field outside the sphere is uniquely determined because the boundary conditions are specified. If we remove the dipole, the field lines must remain unchanged, and we have the solution to the problem of placing a conducting sphere in a large-scale uniform field. This general approach can be used to great effect to study the properties of systems that have simple geometrical forms.

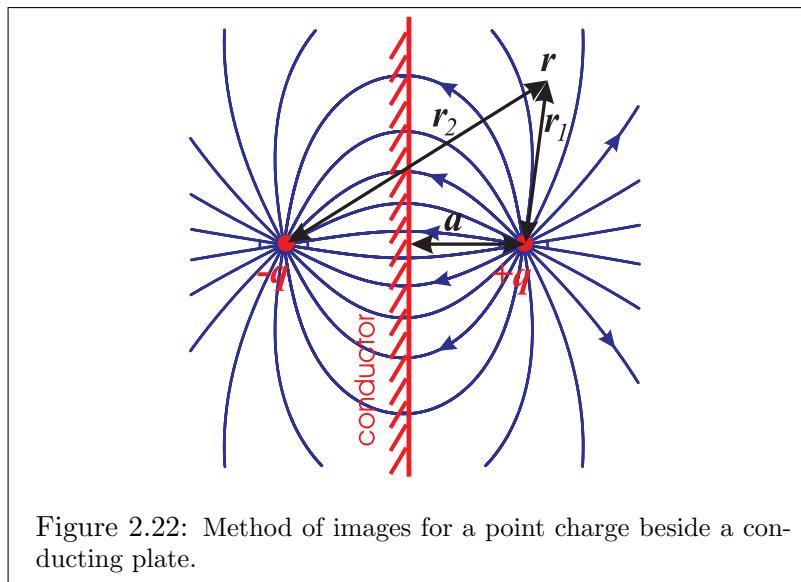
Consider a pair of charges (see Fig. 2.21). Half-way between them is a plane of symmetry, over which $V = 0$. We could place a conducting plate at the same position without changing the form of the field. If the plate is held at zero potential, then we could remove the negative charge without changing the form of the field to the right of the plate, or we could remove the positive charge without changing the form of the field to the left of the plate.

It is possible to reverse the argument. Suppose that we have a system comprising a single charge beside a conducting plate that



is held at zero potential. The conductor could then be replaced by an *image* charge that ensures that the potential remains constant over the surface even when the conductor is removed.

The technique of replacing conducting surfaces with charges is called the **method of images**, and it is a powerful method that was used extensively before the invention of computers. The method works well when we know the charge distributions that give rise to the equipotential surfaces that match problems of interest. Libraries of such distributions have been assembled over the years.



Before describing a number of general examples, it is beneficial to describe the procedure by analysing the charge beside a ground plane in more detail. As shown in Fig. 2.22, we are interested in finding the field lines when a charge is placed in proximity to a plane conductor, say q at $x = a$. It is possible to maintain the plane equipotential surface by placing an image charge $-q$ at $x = -a$, and removing the conductor. The total potential becomes

$$V(\mathbf{r}) = \frac{q}{4\pi\epsilon_0} \left(\frac{1}{r_1} - \frac{1}{r_2} \right), \quad (2.118)$$

where

$$\begin{aligned} r_1 &= \sqrt{(x-a)^2 + y^2 + z^2} \\ r_2 &= \sqrt{(x+a)^2 + y^2 + z^2}. \end{aligned} \quad (2.119)$$

On the surface of the plate we have $x = 0$, giving $r_1 = r_2$, and therefore $V = 0$. This is true for all y and z , and therefore a plane equipotential surface exists at $x = 0$.

The field in the x direction at $x = 0$ can be found easily through

$$E_x = -\frac{\partial V}{\partial x} \Big|_{x=0} = -\frac{q}{2\pi\epsilon_0} \frac{a}{[a^2 + y^2 + z^2]^{3/2}}. \quad (2.120)$$

It is also easy to verify that at $x = 0$, $E_y = 0$ and $E_z = 0$. Finally, using $E = E_\perp = \sigma/\epsilon_0$, the surface charge at $x = 0$ becomes

$$\sigma = \frac{q}{2\pi} \frac{a}{[a^2 + y^2 + z^2]^{3/2}}, \quad (2.121)$$

which has a maximum on-axis value of

$$\sigma = \frac{q}{2\pi a^2}. \quad (2.122)$$

Thought for the day 8: Various systems can be built up using planes. Consider how you would calculate the field when a charge is placed near the intersection of two perpendicular conducting planes - it is important not to forget images of images! What would happen if the angle between the planes were not so simple?

2.23 Image for conducting cylinder and line of charge

Consider the equipotential surfaces of two line charges placed side by side (see Fig. 2.23). We know that the total potential associated with a single line charge of λ in units of coulombs per metre, for example) is

$$V(\mathbf{r}) = -\frac{\lambda}{2\pi\epsilon_0} \ln \frac{r}{r_0}, \quad (2.123)$$

and therefore the total potential associated with two equal but opposite line charges is

$$V(\mathbf{r}) = \frac{\lambda}{2\pi\epsilon_0} \ln \frac{r_2}{r_1}. \quad (2.124)$$

It is clear that the equipotential surfaces associated with a single line of charge are cylinders, but the same is true of *all* the equipotential surfaces, $r_2/r_1 = \text{constant}$, associated with two parallel lines of charge. This can be shown as follows (see Fig. 2.24).

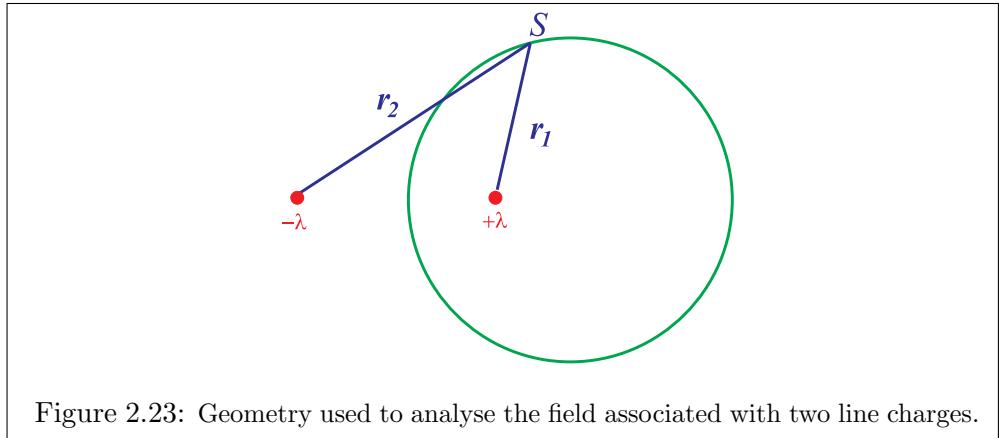


Figure 2.23: Geometry used to analyse the field associated with two line charges.

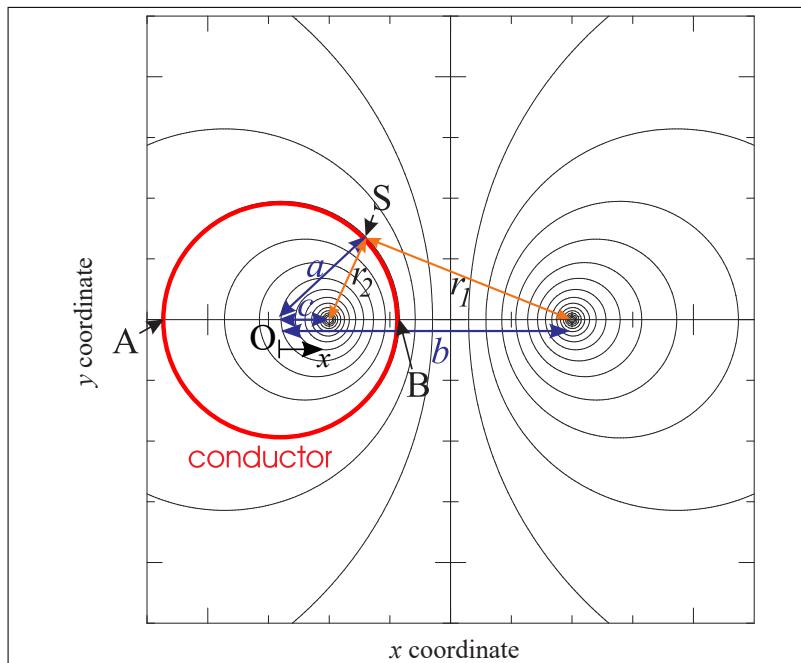


Figure 2.24: The geometry used to analyse the field associated with two line charges.

If the coordinates of the observation point, with respect to some general Cartesian system having its centre O on the x -axis, are x_s and y_s , then the squares of the distances to the two line charges are

$$\begin{aligned} r_2^2 &= (x_s - c)^2 + y_s^2 \\ r_1^2 &= (x_s - b)^2 + y_s^2, \end{aligned} \quad (2.125)$$

where b and c are the positions of the line charges: see Fig. 2.24.

For an equipotential surface the ratio r_2/r_1 must be constant, say $r_2^2/r_1^2 = k$ or $r_2^2 = kr_1^2$, giving

$$(x_s - c)^2 + y_s^2 = k(x_s - b)^2 + ky_s^2. \quad (2.126)$$

Equation (2.126) can be multiplied out to show that when $kb = c$, the surface is a cylinder, centred at the origin of the coordinate

system, and with a radius of

$$r^2 = \frac{kb^2 - c^2}{1 - k}. \quad (2.127)$$

Therefore if the equipotential surface represents a cylinder of radius a , we have

$$a^2 = \frac{kb^2 - c^2}{1 - k}, \quad (2.128)$$

as $kb = c$, giving

$$a^2 = bc. \quad (2.129)$$

Let us suppose that we wish to find the centre of the equipotential cylinder having radius a if we only know the distance between the line charges, say $d = b - c$. Then we have

$$\begin{aligned} bc &= a^2 \\ b - c &= d, \end{aligned} \quad (2.130)$$

which, through a quadratic equation, can be used to find c or b , and the problem is solved. In fact, the two solutions to the quadratic give the two cylindrical equipotential surfaces of radius a associated with the system.

The actual potential of the surface is then

$$V(\mathbf{r}) = \frac{\lambda}{2\pi\epsilon_0} \ln \left[\sqrt{\frac{c}{b}} \right] = \frac{\lambda}{2\pi\epsilon_0} \ln \left[\frac{a}{b} \right]. \quad (2.131)$$

In the limit where the centre of the cylinder is a long way from the line charges, we have $c/b \rightarrow 1$ and $a \rightarrow \infty$. This gives an equipotential surface of zero over a plane midway between the two line charges. In terms of images, if we have a line of charge at a distance b from the centre of a perfect cylindrical conductor having radius a , then we can replace the cylinder with an equal but opposite charge placed at distance $c = a^2/b$ from the centre. Very similar maths can be used to describe a charge near a conducting sphere rather than a cylinder, although the image charge is no longer equal and opposite to the real charge.

2.24 Capacitance

The **capacitance**, C , of a two-surface metallic structure is defined as the amount of net positive charge that resides on the high-potential surface divided by the potential difference between them:

$$C \equiv \frac{Q}{V} \quad \text{Capacitance} \quad (2.132)$$

The capacitance is a measure of the charge on each surface that is caused by a given potential difference between the two conductors. As we saw previously, the potential difference between two points is independent of the path of integration, and because the metallic surfaces are equipotential surfaces, it does not matter which point is taken on each electrode. Usually, for mathematical simplicity, the path integral is evaluated along a field line.

2.25 Capacitance of two long parallel cylinders

Suppose that we have two parallel cylindrical conductors separated by a distance $2D$. It is clear, from the previous analysis and Fig. 2.24, that there is an image system where the conductors can be replaced by lines of charge. For simplicity assume that both cylinders have radius a .

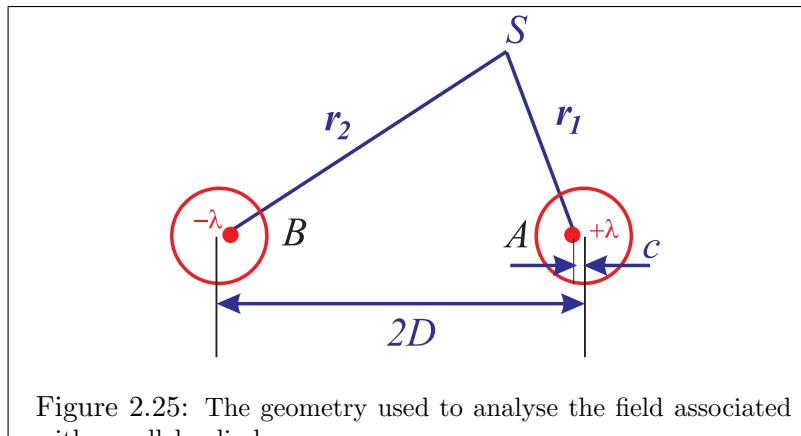


Figure 2.25: The geometry used to analyse the field associated with parallel cylinders.

Now we know from Sect. 2.23, (2.129) that $c = a^2/b$, or

$$a^2 = \underbrace{(2D - c)}_b c, \quad (2.133)$$

giving

$$c^2 - 2Dc + a^2 = 0. \quad (2.134)$$

Taking the *smaller* of the two roots solving this quadratic equation (the other corresponds to the other image charge)

$$c = D - (D^2 - a^2)^{1/2}. \quad (2.135)$$

Consequently, we know where the lines of image charge should be placed.

It is now possible to calculate the potentials of each of the surfaces. We know from Sect. 2.23 that

$$V(\mathbf{r}) = \frac{\lambda}{2\pi\epsilon_0} \ln \frac{r_2}{r_1}. \quad (2.136)$$

It is easiest to calculate the potentials of the cylinders on the line joining the two centres. At A

$$\frac{r_2}{r_1} = \frac{(2D - c) - a}{a - c} = \frac{b - a}{a - c} = \frac{a^2/c - a}{a - c} = \frac{a}{c}, \quad (2.137)$$

and similarly at B ,

$$\frac{r_1}{r_2} = \frac{a}{c} \Rightarrow \frac{r_2}{r_1} = \frac{c}{a}, \quad (2.138)$$

and we conclude

$$V_A = -V_B. \quad (2.139)$$

Finally we have $V = V_A - V_B$ or, using (2.135):

$$V = 2V_A = \frac{\lambda}{\pi\epsilon_0} \ln \frac{a}{D - (D^2 - a^2)^{1/2}}. \quad (2.140)$$

The capacitance per unit length becomes

$$C = \frac{\lambda}{V} = \pi\epsilon_0 \left[\ln \frac{a}{D - (D^2 - a^2)^{1/2}} \right]^{-1}. \quad (2.141)$$

In the limit where $D \gg a$

$$D - (D^2 - a^2)^{1/2} = D - D(1 - a^2/D^2)^{1/2} \approx \frac{a^2}{2D}, \quad (2.142)$$

and

$$C \approx \frac{\pi\epsilon_0}{\ln 2D/a}. \quad (2.143)$$

In this limit the image line charges move close to the centres of the cylinders. Clearly, the capacitance falls as the cylinders are moved apart.

2.26 Electrostatic energy

Although we have discussed concepts such as electric field, electric flux, and capacitance, it is important to remember that the primary principle is that energy is transferred into and out of physical systems as charge is moved. A critical question, therefore, is how much energy is required in order to assemble a complete system of charges? There are various ways in which this calculation can be performed.

Consider assembling a system of charges by bringing up one charge at a time from infinity (see Fig. 2.26).

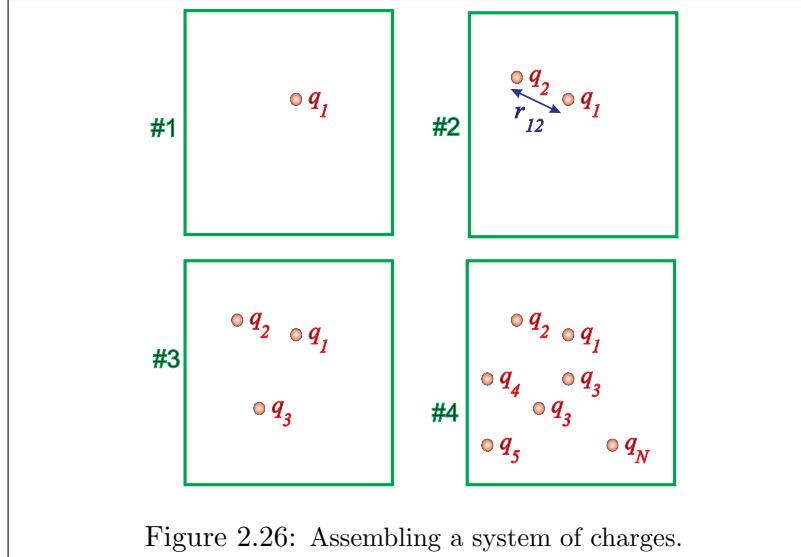


Figure 2.26: Assembling a system of charges.

Each time a charge is added, energy must be supplied in the following steps:

$$\begin{aligned} U_1 &= 0, \\ U_2 &= \frac{q_2 q_1}{4\pi\epsilon_0 r_{12}}, \\ U_3 &= U_2 + \frac{q_3}{4\pi\epsilon_0} \left(\frac{q_1}{r_{13}} + \frac{q_2}{r_{23}} \right), \end{aligned} \quad (2.144)$$

giving

$$U_N = \sum_{j=1}^N \sum_{i < j} \frac{q_j q_i}{4\pi\epsilon_0 r_{ij}}. \quad (2.145)$$

The second sum can, however, be modified according to

$$U_N = \frac{1}{2} \sum_{j=1}^N \sum_{i \neq j} \frac{q_j q_i}{4\pi\epsilon_0 r_{ij}}, \quad (2.146)$$

and therefore

$$U_N = \frac{1}{2} \sum_{j=1}^N q_j \sum_{i \neq j} \frac{q_i}{4\pi\epsilon_0 r_{ij}} = \frac{1}{2} \sum_{j=1}^N q_j V_j, \quad (2.147)$$

where V_j is the potential that exists at position j when all of the charges are present apart from the j th charge itself.

We conclude that the energy needed to assemble a system of charges can be written

$$U_N = \frac{1}{2} \sum_{j=1}^N q_j V_j. \quad (2.148)$$

An alternative way of deriving the same result is to imagine that we slowly increase the charges from 0 to q_j (even though it probably

didn't happen that way...). The charges are αq_j ($0 < \alpha < 1$). The potentials αV_j are proportional to the charges. αV_j is the potential (the energy per unit charge) when $q_j d\alpha$ is added, so the work done in changing from $\alpha \rightarrow \alpha + d\alpha$ is

$$\sum_{j=1}^N (\alpha V_j) (q_j d\alpha)$$

Thus the total work done is

$$U_N = \int_0^1 d\alpha \alpha \sum_j q_j V_j = [\frac{1}{2} \alpha^2]_0^1 \sum_j q_j V_j = \frac{1}{2} \sum_j q_j V_j.$$

This is identical to (2.148), as expected.

It is straightforward to extend the calculation to a continuous distribution of charge. If ρ is the charge density, then we can break this into a collection of charges, each being $\rho(\mathbf{r}) dV$ in an infinitesimal volume element dV at position \mathbf{r} . Then the **energy to produce a charge density $\rho(\mathbf{r})$** becomes

$$U = \frac{1}{2} \int d^3\mathbf{r} \rho(\mathbf{r}) V(\mathbf{r}), \quad (2.149)$$

where $V(\mathbf{r})$ is the potential at \mathbf{r} in the absence of the infinitesimal charge that is ultimately placed at \mathbf{r} . The key point is that charge cannot feel its own force!

2.27 Energy stored in a capacitor

Consider the following simple example comprising a parallel-plate capacitor (see Fig. 2.27).

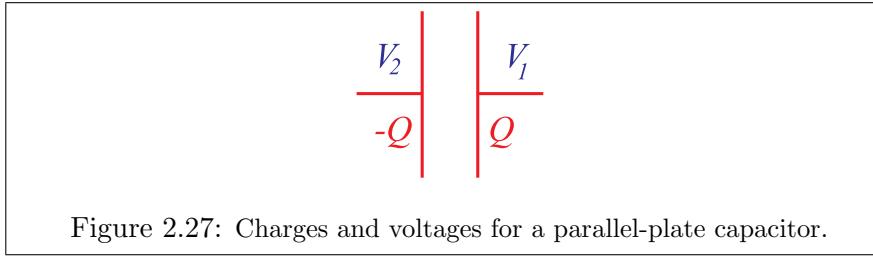


Figure 2.27: Charges and voltages for a parallel-plate capacitor.

The total charge of the final system on one side of the capacitor is $+Q$, whereas the total charge on the other side is $-Q$. Suppose that we divide this charge into N elemental parts dQ_j . Obviously, $Q = \sum_{j=1}^N dQ_j$.

To implement (2.148), for each j we must use the potential V_j across the capacitor when *all* elemental charges *except* dQ_j are in position. In this case, the potential is simply

$$V_j = \frac{(Q - dQ_j)}{C}. \quad (2.150)$$

Eq. (2.148) then becomes

$$U_N = \frac{1}{2} \sum_{j=1}^N dQ_j \frac{(Q - dQ_j)}{C}. \quad (2.151)$$

Assuming that each elemental part has the same charge, and remembering that Q is a fixed quantity (because it comes from the expression for the potential from all charges except dQ_j),

$$U = \frac{1}{2} N \frac{dQ Q}{C} - \frac{1}{2} N \frac{dQ^2}{C}. \quad (2.152)$$

However, we can use smaller and smaller elemental parts, increasing N accordingly with $Q = NdQ$, so the second term tends to zero and

$$U = \frac{1}{2} \frac{Q^2}{C}. \quad (2.153)$$

We conclude that the energy stored in a capacitor is given by

$$U = \frac{1}{2} QV = \frac{1}{2} CV^2. \quad (2.154)$$

We could just have used $i = 1, 2, V_i = 0, V$ in Eq. (2.148)...

2.28 Energy stored in an electric field

We have seen how it is possible to calculate the energy needed to assemble a system of charges. The energy needed is the same as the energy stored because it is assumed that no losses are present. Is it possible to calculate the energy stored solely from knowledge of the electric field? In this way we might consider that the energy is stored in the field itself.

Consider the case of a parallel-plate capacitor (see Fig. 2.28).

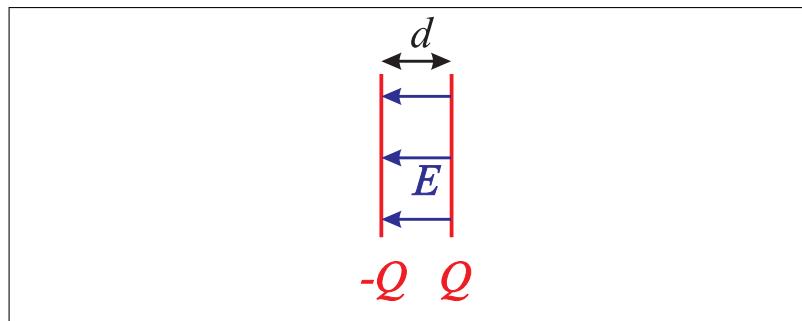


Figure 2.28: Charge and electric field for a parallel-plate capacitor.

From Sect. 2.17, we know that

$$|E| = \frac{\sigma}{\epsilon_0} = \frac{Q}{A\epsilon_0} \quad (2.155)$$

and

$$V = |\mathbf{E}|d, \quad (2.156)$$

and therefore

$$U = \frac{1}{2}QV = \underbrace{\frac{1}{2}\epsilon_0|\mathbf{E}|^2}_{\text{energy density}} \cdot \underbrace{Ad}_{\text{volume}}. \quad (2.157)$$

The term in the right-hand braces is the total volume and so the term in the left-hand braces can be considered to be the energy density.

This result is completely general, and therefore the energy density is given by

$U_E(\mathbf{r}) = \frac{1}{2}\epsilon_0 \mathbf{E}(\mathbf{r}) ^2$	Energy density of the electric field
---	---

(2.158)

or considering the electric displacement through (2.76),

$U_E(\mathbf{r}) = \frac{1}{2}\mathbf{D}(\mathbf{r}) \cdot \mathbf{E}(\mathbf{r})$	Energy density of the electric field
--	---

(2.159)

an equation of considerable importance, because it applies even when dielectric materials are present.

Then, the *total* energy stored becomes

$U = \int d^3\mathbf{r} U_E(\mathbf{r}).$	
---	--

(2.160)

The total energy can thus be regarded as being stored in the field itself.

We now have two expressions for the energy stored in an electrostatic system:

$$\text{Eq. (2.149)} : U = \frac{1}{2} \int d^3\mathbf{r} \rho(\mathbf{r})V(\mathbf{r}), \quad (2.161)$$

$$\text{Eq. (2.158)} : U = \frac{1}{2}\epsilon_0 \int d^3\mathbf{r} |\mathbf{E}|^2(\mathbf{r}). \quad (2.162)$$

Can these be derived for a general field without having to refer to a capacitor?

We have

$$U = \frac{1}{2} \int d^3\mathbf{r} \rho(\mathbf{r})V(\mathbf{r}) \quad (2.163)$$

Gauss's law

$$= \frac{1}{2}\epsilon_0 \int d^3\mathbf{r} [\nabla \cdot \mathbf{E}(\mathbf{r})] V(\mathbf{r}),$$

where Gauss's law has been used, but we also have the vector identity

$$\nabla \cdot (c\mathbf{A}) = c\nabla \cdot \mathbf{A} + \mathbf{A} \cdot \nabla c, \quad (2.164)$$

where c is any scalar field, and \mathbf{A} is any vector field, and therefore

$$U = \frac{1}{2}\epsilon_0 \int d^3\mathbf{r} \{ \nabla \cdot [V(\mathbf{r})\mathbf{E}(\mathbf{r})] - \mathbf{E}(\mathbf{r}) \cdot \nabla V(\mathbf{r}) \}. \quad (2.165)$$

Look at each of the terms in turn. The first term can be replaced, through the divergence theorem, by an integration over the surface containing the field,

$$\frac{1}{2}\epsilon_0 \int d^3\mathbf{r} \nabla \cdot [V(\mathbf{r})\mathbf{E}(\mathbf{r})] = \frac{1}{2}\epsilon_0 \int d\mathbf{S} \cdot [V(\mathbf{r})\mathbf{E}(\mathbf{r})]. \quad (2.166)$$

The surface must, however, be taken at infinity, if it is to include all of the field, but for a localised charge distribution $V(\mathbf{r})$ falls as $1/r$ and $\mathbf{E}(\mathbf{r})$ falls as $1/r^2$ (multipole contributions fall even faster), whereas the area increases as r^2 , and therefore the integral tends to zero for large r .

Only the second term remains, and

$$\begin{aligned} U &= -\frac{1}{2}\epsilon_0 \int d^3\mathbf{r} \mathbf{E}(\mathbf{r}) \cdot \nabla V(\mathbf{r}) \\ U &= \frac{1}{2}\epsilon_0 \int d^3\mathbf{r} \mathbf{E}(\mathbf{r}) \cdot \mathbf{E}(\mathbf{r}) \\ &= \frac{1}{2}\epsilon_0 \int d^3\mathbf{r} |\mathbf{E}|^2(\mathbf{r}). \end{aligned} \quad (2.167)$$

Quod erat demonstrandum.

2.29 Forces and virtual work

It is possible to learn a great deal about the forces acting on a complicated system of charges through a technique known as **virtual work**. The basic idea is to identify, and equate, the energy that flows into and out of a system when its physical form is changed in some way. It is essential to ensure that any work done by the external mechanical system is correctly taken into account, as well as any electrical energy that may be supplied or dissipated by an external circuit. It is also essential to ensure that the signs on the various terms are correct.

Let us suppose that x quantifies some dimension of a system, say the distance between the plates of a capacitor. Then the work that must be done on the system in order to change x by some small amount dx must be either stored in the electrostatic field or dissipated in some way. Mathematically,

$$F dx = \left. \frac{\partial U_s}{\partial x} \right|_{\text{const}} \cdot dx + \left. \frac{\partial U_d}{\partial x} \right|_{\text{const}} \cdot dx, \quad (2.168)$$

where U_s is the stored electrostatic energy and U_d the dissipated energy, and the vertical bars are included to emphasise that certain

aspects of the system will be held constant, usually the charge, or the potential. When evaluating the partial derivatives it is important to ensure that any quantities that vary with x are correctly taken into account.

Consider first the force acting between a pair of charged, electrically isolated, parallel plates (see Fig. 2.29).

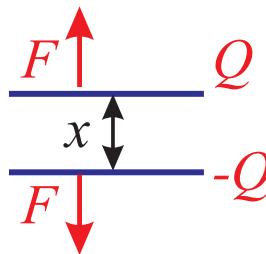


Figure 2.29: Force between charged parallel plates.

The plates are attracted to each other, and therefore work must be done to increase their separation by dx . There is no mechanism for dissipating energy, and, in this case, the charge is constant. According to Gauss's law, constant charge means a constant electric field, and a constant electric field means that the potential will increase as the separation increases.

Energy stored is given according to (2.167) by

$$U_s = \frac{1}{2}\epsilon_0|\mathbf{E}|^2Ax, \quad (2.169)$$

but for a parallel-plate capacitor

$$|\mathbf{E}| = \frac{Q}{A\epsilon_0}, \quad (2.170)$$

and therefore

$$U_s = \frac{1}{2\epsilon_0} \frac{Q^2}{A}x, \quad (2.171)$$

giving

$$\left. \frac{\partial U_s}{\partial x} \right|_Q dx = \frac{1}{2\epsilon_0} \frac{Q^2}{A} dx. \quad (2.172)$$

The change in stored energy is a positive quantity, showing that all of the external work done on the system goes into the stored energy.

Finally, equating the virtual work,

$$Fdx = \frac{1}{2\epsilon_0} \frac{Q^2}{A} dx, \quad (2.173)$$

and therefore

$F = \frac{Q^2}{2\epsilon_0 A}.$

(2.174)

Remembering that Q is the total charge, if the charge per unit area is constant, the force scales with area, as might be expected on intuitive grounds.

Now consider a more complicated case where the potential between the plates is held constant by some external power supply (see Fig. 2.30).

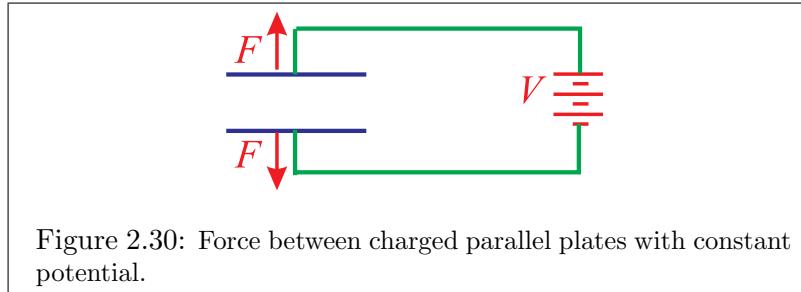


Figure 2.30: Force between charged parallel plates with constant potential.

In this case, as the separation is increased, the electric field must fall because the potential difference is held constant.

The electric field can only fall if the charge on the plates falls. The charge on the plates falls through the process of positive charge moving through the battery from the positive terminal to the negative terminal. But as charge moves from a high potential to a low potential it gives up energy, which must be returned to, or dissipated in, the battery.

Consider how this can be formulated mathematically.

Energy stored in the capacitor is given by

$$U_s = \frac{1}{2}\epsilon_0|\mathbf{E}|^2Ax. \quad (2.175)$$

Note that the electric field is no longer constant as x increases. Instead,

$$V = |\mathbf{E}|x \stackrel{\text{(power supply)}}{=} \text{const}, \quad (2.176)$$

and therefore

$$U_s = \frac{1}{2}\epsilon_0\frac{V^2}{x^2}Ax = \frac{1}{2}\epsilon_0\frac{V^2}{x}A. \quad (2.177)$$

It follows that

$$\left. \frac{\partial U_s}{\partial x} \right|_{V=\text{const}} dx = -\frac{1}{2}\epsilon_0\frac{V^2}{x^2}A dx \quad (2.178)$$

Notice that this is a negative quantity showing that the stored energy actually decreases with increasing separation!

Now consider the energy lost to the power supply. The charge on the positive plate of the capacitor is given by

$$Q = |\mathbf{E}|\epsilon_0 A = \frac{V}{x}\epsilon_0 A, \quad (2.179)$$

and therefore the change in charge is

$$dQ = \frac{\partial Q}{\partial x} \Big|_{V=\text{const}} dx = -\frac{V}{x^2} \epsilon_0 A dx. \quad (2.180)$$

This decrease in charge on the plates dissipates energy in the power supply, and therefore

$$\frac{\partial U_d}{\partial x} \Big|_{V=\text{const}} dx = -dQ V = \epsilon_0 \frac{V^2}{x^2} A dx. \quad (2.181)$$

Finally, equating the work done to the change in stored energy and the energy lost gives

$$F dx = \frac{\partial U_s}{\partial x} \Big|_{V=\text{const}} dx + \frac{\partial U_d}{\partial x} \Big|_{V=\text{const}} dx \quad (2.182)$$

$$= -\frac{1}{2} \epsilon_0 \frac{V^2}{x^2} A dx + \epsilon_0 \frac{V^2}{x^2} A dx, \quad (2.183)$$

$$(2.184)$$

which can be tidied up to reveal

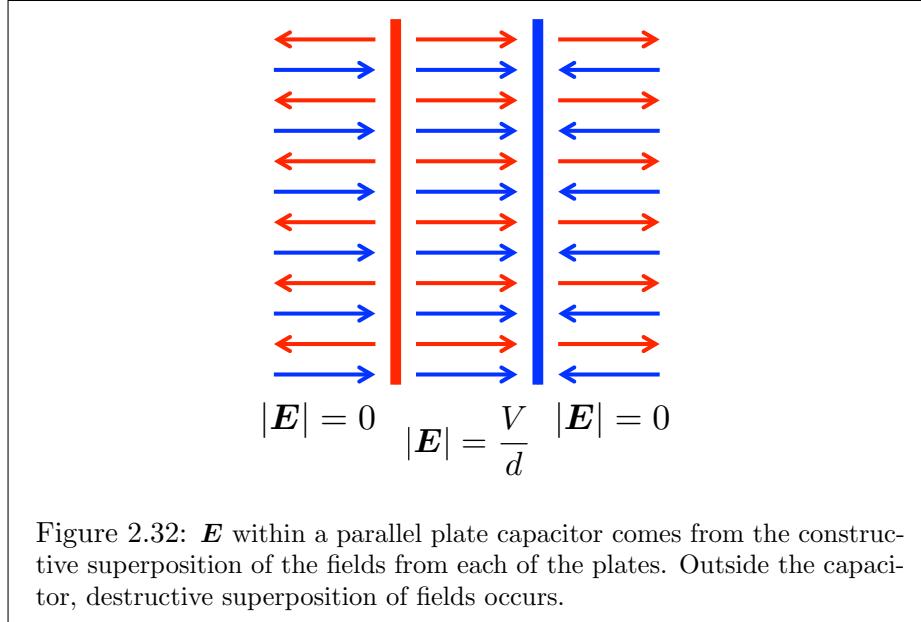
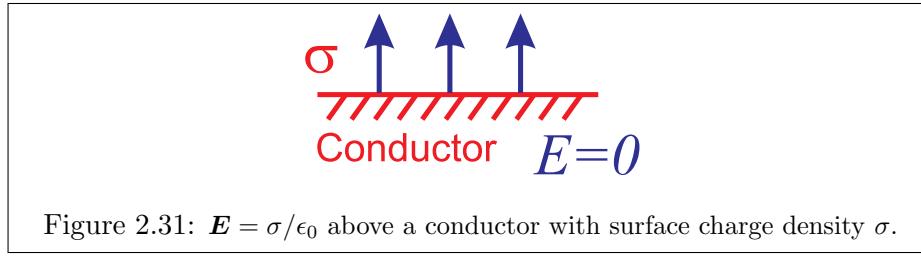
$$F = \frac{1}{2} \epsilon_0 V^2 \frac{A}{x^2}, \quad (2.185)$$

which is the force on the plates of a parallel-plate capacitor whose potential difference is held fixed. It should also be noted that for a fixed potential, the force scales with area. Eq. (2.185) can be rearranged to show that it is identical to the expression for the force between the two electrically isolated plates, even though, in the last case charge flows through the battery in order to maintain a constant potential difference.

2.30 Force on a charged conductor

An interesting question is whether we could have determined the force acting between the plates of a parallel-plate capacitor in a more direct way, without the use of virtual work. Ideally, one should just be able to think of the charge on one plate of the capacitor as sitting within the field produced by the charge on the other plate, but how does this work, as all the field lines begin and end on the surface charge? The way forward can be appreciated if the field lines associated with the individual *sheets* of charge are drawn individually rather than their sum.

The field lines associated with the positive plate travel normally outwards in both directions; the field lines associated with the sheet of negative charge travel normally inwards in both directions. When these combine there is twice the field between the plates, but no field outside the plates (see Figures 2.31 and 2.32).



One sheet of charge can now be seen to reside within the field produced by the other sheet of charge.

Previously we saw, using Gauss's law, that the field due to a sheet of charge is given by

$$|\mathbf{E}| = \frac{\sigma}{2\epsilon_0} = \frac{Q}{2\epsilon_0 A}, \quad (2.186)$$

where σ is the charge per unit area. The force therefore pulls the plates towards each other. It has magnitude

$$F = |\mathbf{E}|Q = \frac{Q^2}{2\epsilon_0 A}, \quad (2.187)$$

which, using (2.179), equates to (2.185).

It can be seen above that the field corresponds to that of a sheet of charge, rather than the field that exists within the parallel plates of the capacitor: there is a factor of 2 difference. It would be incorrect to use the field internal to the capacitor, because this field is produced by both sheets of charge, whereas we only require the field produced by one sheet in order to calculate the force acting on the other. This is an example of where it is easy to inadvertently, and mistakenly, include the field produced by a charge itself when calculating the force on it—be careful.

3. Electrostatic fields in dielectric materials

3.1 Introduction

Up to this point we have been concerned with the behaviour of static electric fields in free space, although we did allow perfectly conducting boundaries. We would now like to understand the behaviour of static electric fields in the presence of insulating materials, such as crystals and plastics. We will find that, with some small modification, the techniques that have been described previously can be applied directly even when dielectric bodies are present.

3.2 Isotropic dielectrics

Faraday found that when an insulator is placed between the plates of a capacitor, held at a constant potential difference, the charge on the plates **increases** (alternatively, for constant charges, the potential **decreases**, as demonstrated in the lecture). In fact, if the insulator completely fills the space between the plates, the charge increases by a factor ϵ , which is called the **relative dielectric constant** or **relative permittivity** of the material. This is sometimes written ϵ_r or k instead of ϵ . The dielectric constant of most materials is of order 1 to 10, but it can be as high as 1000.

We are interested in insulating materials, and therefore, by definition, there is no free charge in the material itself. All of the charge is bound, and we shall assume that, overall, the material is neutral. Hence, it is crucially important to distinguish between **bound** charge and **free** charge.

Despite the absence of any free charges in an insulator, an applied electric field can cause positive and negative bound charge to separate, such that a dipole moment is induced. Microscopically, the electrons and the nucleus are displaced in different directions. For moderate fields in most materials, the induced dipole moment in a unit cell, or atom, is linearly proportional to the applied field.

An isotropic material is one for which the magnitude of the induced dipole moment does not depend on the orientation of the field with respect to the material. Most materials are anisotropic at some level; for example, sapphire.

In a region where the electric field is uniform, and the material homogeneous, charge **only appears on the external surfaces**. Since we are considering insulators, this effect is not due to free charge being displaced, but is caused by the cancellation of the separated charge internal to the material. A model to understand this effect conceptually is discussed in the following.

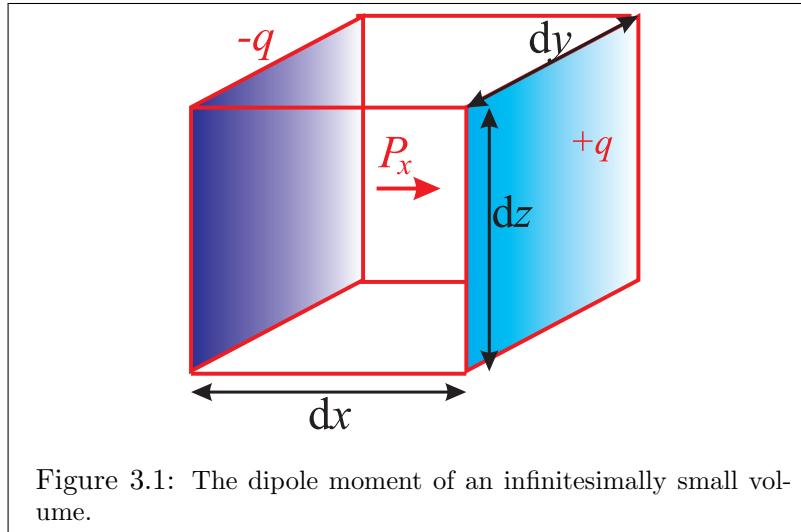


Figure 3.1: The dipole moment of an infinitesimally small volume.

Consider some infinitesimally small volume $d^3\mathbf{r} \equiv dV$, as shown in Fig. 3.1. The application of an external electric field causes positive charge $+q$ to be displaced towards one face, and negative charge $-q$ towards the opposite face. As a consequence there is an excess of charge at the two faces, even though, overall, the volume is neutral.

If a field is applied in the x -direction, the dipole moment is given by

$$p_x = qdx = \sigma_x dxdydz, \quad (3.1)$$

where σ_x is the surface charge density on the \hat{i} -directed surface. We could also have applied a field in the y and z directions:

$$\begin{aligned} p_x &= \sigma_x dxdydz, \\ p_y &= \sigma_y dxdydz, \\ p_z &= \sigma_z dxdydz. \end{aligned} \quad (3.2)$$

It is more usual to express a dipole moment as the **dipole moment per unit volume**, \mathbf{P} . The concept of \mathbf{P} is meaningful because if we, say, place two volume elements side by side, then the charge on the combined surfaces is twice that of the single surface, but the displacement is the same; if we place two volume elements end to end, the charges on the internal surface cancel, the charges on the other surfaces remain the same, but now the separation has doubled. Therefore, it does not matter *how* we combine the small volumes—the dipole moment increases proportionally to the volume.

For each of x , y and z , the dipole moments per unit volume become

$$\begin{aligned} P_x &= \sigma_x, \\ P_y &= \sigma_y, \\ P_z &= \sigma_z. \end{aligned} \quad (3.3)$$

When a material is *anisotropic*, the polarisation density \mathbf{P} depends on the direction of the inducing field, i.e., for a given $|\mathbf{E}|$, for each of x , y , and z , $\sigma_x = \sigma_y = \sigma_z$ is not fulfilled.

The expressions in Eq. (3.3) can now be combined into a single vector quantity, giving the vector dipole moment per unit volume \mathbf{P} :

$$\mathbf{P} = (P_x, P_y, P_z). \quad (3.4)$$

For example, say that the dipole moment associated with a single atom is \mathbf{p} , and that there are N atoms per unit volume, then in some small region the dipole moment per unit volume will be

$$\mathbf{P} = n\mathbf{p}, \quad (3.5)$$

where it is assumed that the medium is sufficiently diffuse that dipoles do not ‘interact’.

If \mathbf{P} is known, the polarisation (bound) charge density at the surface, with normal $\hat{\mathbf{n}}$, is given by

$$\sigma = |\mathbf{P}_\perp| = \mathbf{P} \cdot \hat{\mathbf{n}}. \quad (3.6)$$

Now consider a macroscopically large volume of material, with a uniform external field, as shown in Fig. 3.2.

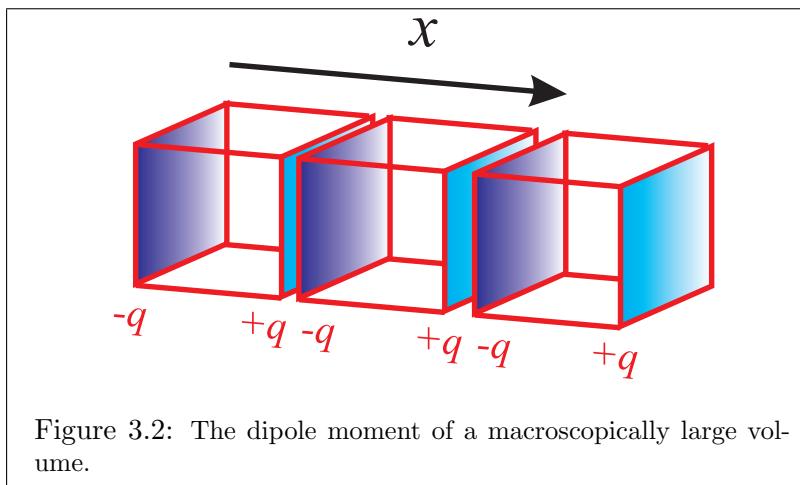


Figure 3.2: The dipole moment of a macroscopically large volume.

In this case, the charges on the internal surfaces cancel, and charge is only left on the external surfaces. We are thus left with the idea that when an electric field is applied to an insulating material, bound charge separates, such that equal and opposite charge appears on opposite faces.

When an insulating material is placed in a capacitor, a given internal electric field requires a given amount of charge on the plates, but we must compensate for the polarisation charge on the surfaces of the insulator, which tends to cancel the original charge. This must occur because the potential difference is fixed, and of course the separation between the plates is fixed, and therefore the electric field must be fixed ($E = V/d$). In other words, the free charge on the plates increases to offset the separated bound charge, which is consistent with Faraday’s observation.

A self-consistent solution is reached when the total charge on the positive plate Q is equal to the free charge that is required to establish the field, \mathbf{E} , plus the additional free charge needed to offset the induced bound charge:

$$Q = \underbrace{\epsilon_0 |\mathbf{E}| A}_{\text{field w/o dielectric}} + \underbrace{\epsilon_0 \chi |\mathbf{E}| A}_{\text{offset induced bound charge}}, \quad (3.7)$$

where χ is the constant of proportionality that gives the surface polarisation charge for a given electric field.

Factorising,

$$Q = \epsilon_0 (1 + \chi) |\mathbf{E}| A. \quad (3.8)$$

By definition

$$C = \frac{Q}{V} = \frac{\epsilon_0 (1 + \chi) A}{d}. \quad (3.9)$$

Finally,

$$\epsilon = \frac{C_{\text{with dielectric}}}{C_{\text{w/o dielectric}}} = (1 + \chi), \quad (3.10)$$

We thus arrive at an important result: the relationship between the **relative permittivity** ϵ and the **susceptibility** χ :

$$\epsilon = (1 + \chi) \quad \text{Relative permittivity as a function of susceptibility}$$

(3.11)

Of course we also have the capacitance in a capacitor filled with a dielectric:

$$C = \frac{Q}{V} = \frac{\epsilon_0 \epsilon A}{d}.$$

(3.12)

The key point is that the charge Q always refers to the *free charge* on the plates—it does not include the polarisation (bound) charge on the surface of the dielectric. The presence of the bound charge is handled “automatically” through the introduction of the relative dielectric constant. Appreciating this point is central to understanding the classical description of the electrostatic behaviour of materials.

3.3 Polarisation charge density

It is now possible to use the basic concepts of polarisation, susceptibility, and relative dielectric constant in more sophisticated ways.

Consider what happens if the electric field is not uniform, which means that the polarisation is no longer uniform either, as shown in

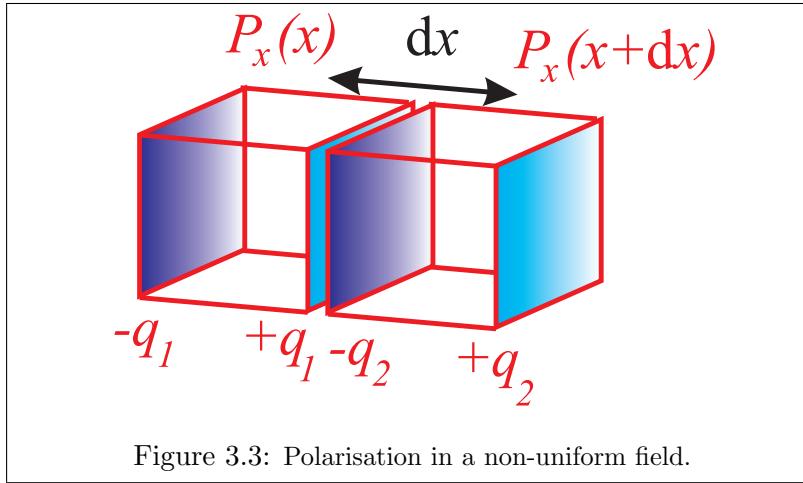


Fig. 3.3. In this case the separated, bound charge no longer cancels inside the material. The net internal, bound charge gives rise to a **polarisation** charge density ρ_p . The two internal surfaces (with $x = \text{constant}$), which in reality correspond to the same surface, therefore have surface charge density $\sigma = \mathbf{P} \cdot \hat{\mathbf{n}}$ and hence total charges

$$\begin{aligned} q_1 &= P_x(x) dy dz, \\ -q_2 &= -P_x(x + dx) dy dz. \end{aligned} \quad (3.13)$$

Thus, the total charge on the internal surface due to the spatial dependence of P_x is

$$\begin{aligned} q &= q_1 - q_2 \\ &= [P_x(x) - P_x(x + dx)] dy dz \\ &= -\frac{\partial P_x}{\partial x} dx dy dz, \end{aligned} \quad (3.14)$$

and therefore the charge per unit volume due to the spatially varying P_x is

$$\rho_{p,x} = -\frac{\partial P_x}{\partial x}. \quad (3.15)$$

Similar charges will appear if P_y or P_z changes with position, and therefore the total accumulated polarisation charge density is

$$\rho_p = - \left[\frac{\partial P_x}{\partial x} + \frac{\partial P_y}{\partial y} + \frac{\partial P_z}{\partial z} \right], \quad (3.16)$$

and it follows for the **polarisation charge density**:

$$\boxed{\rho_p = -\nabla \cdot \mathbf{P}(\mathbf{r}), \quad \text{Polarisation charge density}} \quad (3.17)$$

where the position dependence of the polarisation is shown explicitly.

The divergence of the polarisation per unit volume at a point gives the net polarisation charge density at that point. It is clear that bound charge is the source of $(-\mathbf{P})$, in the same way that free charge is the source of \mathbf{D} .

3.4 Divergence theorem and polarisation

Presented with (3.17) one is immediately inclined to apply the divergence theorem to some finite, closed volume. In other words,

$$\int_V d^3r \rho_p = - \int_V d^3r \nabla \cdot \mathbf{P} = - \oint_S d\mathbf{S} \cdot \mathbf{P}. \quad (3.18)$$

The situation is shown in Fig. 3.4.

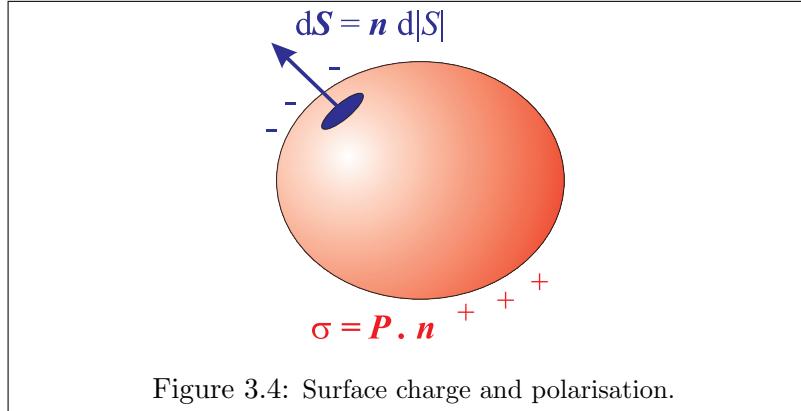


Figure 3.4: Surface charge and polarisation.

Previously, we saw that the dot product between the polarisation vector and the surface normal gives the surface charge density at the position where the dot product is evaluated ($\sigma = \mathbf{P} \cdot \hat{\mathbf{n}}$). The term on the RHS of (3.18) gives the total surface charge, but this must be equal and opposite to the internal bound charge from which it was separated. If there is no internal charge, because the electric field is everywhere uniform, then the surface integral of the normal component of the polarisation evaluates to zero.

3.5 Gauss's law for dielectric materials

In general, a volume of space may contain free charge ρ_f as well as bound charge ρ_p : Fig. 3.5. For this volume, it is possible to apply Gauss's law, but now both of these charges must be taken into consideration.

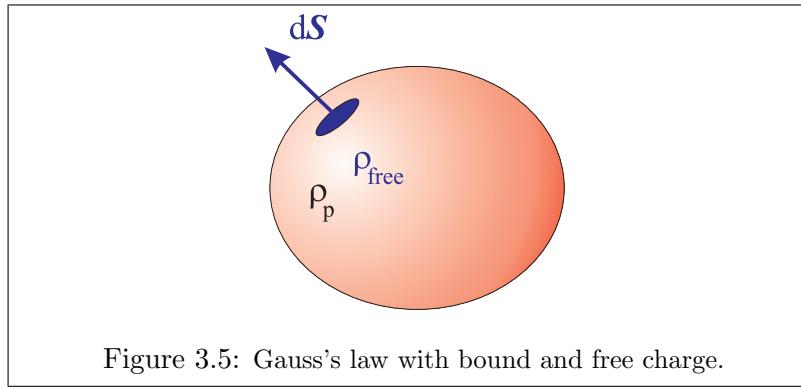


Figure 3.5: Gauss's law with bound and free charge.

According to Gauss's law we have

$$\int d\mathbf{S} \cdot \mathbf{E} = \frac{1}{\epsilon_0} \int d^3\mathbf{r} [\rho_f + \rho_p]. \quad (3.19)$$

Applying the divergence theorem,

$$\begin{aligned} \int d^3\mathbf{r} \nabla \cdot \mathbf{E} &= \frac{1}{\epsilon_0} \int d^3\mathbf{r} [\rho_f + \rho_p] \\ &= \frac{1}{\epsilon_0} \int d^3\mathbf{r} [\rho_f - \nabla \cdot \mathbf{P}]. \end{aligned} \quad (3.20)$$

Rearranging we get

$$\int d^3\mathbf{r} \nabla \cdot [\epsilon_0 \mathbf{E} + \mathbf{P}] = \int d^3\mathbf{r} \rho_f. \quad (3.21)$$

Finally, because this expression must hold for all volumes,

$$\boxed{\nabla \cdot [\epsilon_0 \mathbf{E} + \mathbf{P}] = \rho_f.} \quad (3.22)$$

We can now define the **electric displacement** \mathbf{D} properly:

$$\boxed{\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P}. \quad \text{Electric displacement}} \quad (3.23)$$

We can thus conclude that the source of \mathbf{D} are *free* charges and define **Gauss's law for dielectrics**:

$$\boxed{\nabla \cdot \mathbf{D} = \rho_f. \quad \text{Gauss's law for dielectrics}} \quad (3.24)$$

Previously we assumed that the induced polarisation is proportional to the electric field,

$$\mathbf{P} = \epsilon_0 \chi \mathbf{E}, \quad (3.25)$$

and therefore, using $\epsilon = (1 + \chi)$, we can define the **Electric displacement in terms of the electric field**:

$$\boxed{\mathbf{D}(\mathbf{r}) = \epsilon_0 \epsilon \mathbf{E}(\mathbf{r}) \quad \text{Electric displacement and electric field}} \quad (3.26)$$

where position dependence has been referred to explicitly.

With the useful notation introduced above, we can now rewrite Eq. (3.22):

$$\nabla \cdot [\epsilon_0 \epsilon \mathbf{E}] = \rho_f. \quad (3.27)$$

This equation shows that (if ϵ is constant)

$$\nabla \cdot [\epsilon_0 \mathbf{E}] = \frac{\rho_f}{\epsilon}, \quad (3.28)$$

which indicates that the total charge enclosed is effectively reduced as a consequence of the induced bound surface charge.

Equation (3.27) is a restatement of Gauss's law, but where the permittivity of free space ϵ_0 has been replaced by the permittivity of free space multiplied by the relative dielectric constant ϵ .

As described previously, the quantity $\mathbf{D}(\mathbf{r})$, which is a vector field, is called the **electric displacement** or the **electric flux density**. The notion of a flux density comes from our early lectures where we integrated the normal component of the electric field over a surface to create the concept of **electric flux**.

The key point is that the charge, $\rho_f(\mathbf{r})$, in (3.24) relates to free charge, as it did before, but now the role of bound charge is included, indirectly, through the use of the ‘relative permittivity’. Whenever the relative permittivity is used in electrostatic field calculations, any reference to charge corresponds to free charge only: it does not include bound charge. It is remarkable that the effects of polarisation-charge separation can be taken care of solely by multiplying ϵ_0 by a simple multiplicative factor (for linear dielectrics).

Thought for the day 9: Consider a parallel-plate capacitor having a surface charge of ρ_f on one plate. Imagine that a plane-parallel dielectric is placed midway between the plates in a such a way that gaps are left between the surface of the dielectric and the plates on each side. Draw a diagram showing clearly the following quantities as a function of position along a perpendicular line crossing from one plate to the other: polarisation $|\mathbf{P}|$; free charge density ρ_f ; bound charge density ρ_p ; electric field $|\mathbf{E}|$; and electric displacement $|\mathbf{D}|$.

3.6 Summary of key points

It is worthwhile summarising a number of key points relating to the behaviour of electrostatic fields in insulating materials:

- The constant of proportionality between the polarisation and the electric field is called the *susceptibility*, χ , giving $\mathbf{P}(\mathbf{r}) = \epsilon_0 \chi \mathbf{E}(\mathbf{r})$.
- The relationship between the susceptibility and the relative dielectric constant, ϵ , is $\epsilon = 1 + \chi$.
- The susceptibility is the parameter of choice when describing the detailed behaviour of materials, whereas the relative dielectric constant is preferred when describing the overall behaviour of electrostatic systems.
- For a homogeneous dielectric, we can simply replace ϵ_0 with $\epsilon_0 \epsilon$, remembering that any reference to charge relates to free charge only. The effect of bound charge is contained in the effective dielectric constant.
- In particular, the electric displacement $\mathbf{D}(\mathbf{r}) = \epsilon_0 \mathbf{E}(\mathbf{r})$ becomes $\mathbf{D}(\mathbf{r}) = \epsilon_0 \epsilon \mathbf{E}(\mathbf{r})$.

- In most materials $\mathbf{P}(\mathbf{r})$ is in the same direction as $\mathbf{E}(\mathbf{r})$, and the material is called **isotropic**. Depending on the microscopic behaviour of the atoms and molecules, however, E_x may generate P_x , P_y , and P_z , and then the material is then called anisotropic. In this case the susceptibility takes the form of a matrix, or tensor.
- Usually, $\mathbf{P}(\mathbf{r})$ is **linearly** proportional to $\mathbf{E}(\mathbf{r})$, certainly for low field strengths, but this need not be the case, giving rise to numerous, important non-linear optical phenomena.

3.7 A word of warning

The above formalism is used extensively in the study of electrostatic systems, and you are likely to go through life without ever worrying about anything different. A word of warning is, however, appropriate.

We have assumed that at each point in the material, the local electric field only induces a dipole. It does not induce a current, because no free charge is present, and neither does it induce higher-order poles, such as a quadrupole term. Remember that any charge distributions can be described by a hierarchy of terms, and so in principle any of these could be present. Thus, we have, implicitly, extracted the lowest-order term from a series of possibilities. Usually, this series converges so rapidly that it is sufficient to consider only the dipole term, but in certain materials this assumption can be violated.

In fact, the electric field induces a dipole, the gradient of the field induces a quadrupole, and the second derivative of the field induces an octupole. In almost all circumstances, the first term is dominant, and so this complication can be ignored. Indeed, this assumption is hidden in the majority of electromagnetic calculations.

3.8 Use of \mathbf{D} and \mathbf{E} in electrostatic problems

Usually, electrostatic problems come in two flavours, which are best approached in slightly different ways:

- (i) a collection of conducting surfaces is established and the potential differences between the surfaces are known, and remain fixed;
- (ii) a collection of conducting surfaces is established, and the free charge on the surfaces is known, and remains fixed.

In the case of (i), and a homogeneous dielectric, the constant-potential surfaces determine the electric field \mathbf{E} throughout the

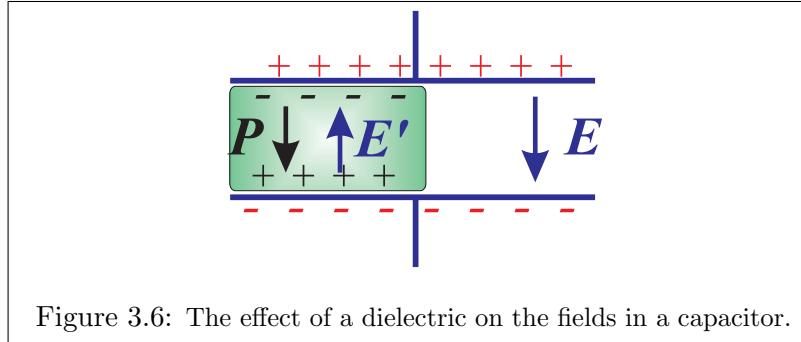


Figure 3.6: The effect of a dielectric on the fields in a capacitor.

region, through Poisson's equation. For example, in the case of a parallel-plate capacitor, Fig. 3.6, $|\mathbf{E}| = V/d$, regardless of the presence of any dielectric. We then have $\mathbf{D} = \epsilon_0\epsilon\mathbf{E}$, and then Gauss's law gives the surface charge, $\sigma = |\mathbf{D}| = \epsilon_0\epsilon V/d$. As the potential is held constant, and dielectric material is introduced, the free charge on the conducting surfaces increases to offset the induced bound charge, and to hold the potential constant.

In the case of (ii), a fixed distribution of free charge on the conductors determines \mathbf{D} through Gauss's law $\nabla \cdot \mathbf{D}(\mathbf{r}) = \rho_f(\mathbf{r})$. For example, in the case of a parallel-plate capacitor $|\mathbf{D}| = \sigma$. The electric field is then given by $\mathbf{E} = \mathbf{D}/\epsilon_0\epsilon$, which in turn determines the potential difference, $V = |\mathbf{E}|d$. Therefore, $|\mathbf{E}| = \sigma/\epsilon_0\epsilon$, and $V = \sigma d/\epsilon_0\epsilon$. As the surface charge is held constant and dielectric material introduced, the voltage falls because the total electric field between the plates is smaller than would be expected on the basis of free charge alone.

Another way of thinking about the relationship between \mathbf{D} and \mathbf{E} is that free charge is the source of \mathbf{D} , such that the field lines associated with \mathbf{D} can only start and end on free charge, whereas bound and free charges are the source of \mathbf{E} , and the field lines associated with \mathbf{E} can begin and end on polarisation charge and free charge. This model explains why, if the free charge is known, it is necessary to calculate \mathbf{D} first, whereas if the potential difference is known, it is necessary to calculate \mathbf{E} first. In this sense \mathbf{E} is more fundamental than \mathbf{D} . Certainly, any test charge within a material will experience a force due to the total field \mathbf{E} , not merely due to the field that is associated with the free charge on the conductors.

3.9 Inhomogeneous dielectrics and boundary conditions

We must now consider the case when ϵ is not homogeneous throughout the system of interest. First of all consider the relationships between the various field quantities on either side of a discontinuous change in ϵ . These relationships are called the **boundary conditions**.

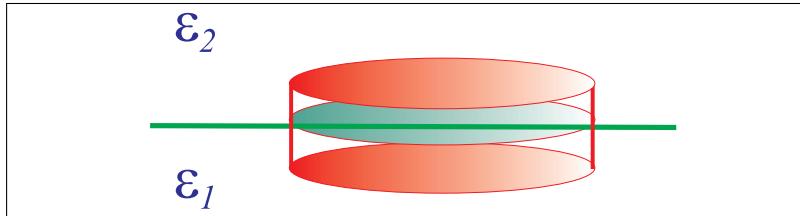


Figure 3.7: The boundary between dissimilar dielectric materials.

To approach the problem of a boundary between two dissimilar materials, set up a “pillbox” that cuts across the surface, as shown in Fig. 3.7, and apply Gauss’s law:

$$\int \mathbf{D}(\mathbf{r}) \cdot d\mathbf{S} = \int \rho_f(\mathbf{r}) d^3 r = 0, \quad (3.29)$$

where the second equality follows because there is no free surface charge at the boundary — unless there are surface imperfections.

Now shrink the parallel faces of the pillbox down, such that they are separated by an infinitesimally small distance. Also, the parallel faces of the pillbox are small enough that D is essentially constant over the area, and $d\mathbf{S}$ points in opposite directions on the two surfaces. If the normal component of \mathbf{D} in region 1 is called $\mathbf{D}_{1\perp}$ and the normal component of \mathbf{D} in region 2 is $\mathbf{D}_{2\perp}$, then

$$A\mathbf{D}_{2\perp} - A\mathbf{D}_{1\perp} = \mathbf{0}, \quad (3.30)$$

where A is the area, from which it follows that

$$\mathbf{D}_{2\perp} = \mathbf{D}_{1\perp}. \quad (3.31)$$

We conclude that the normal component of \mathbf{D} must be continuous across the boundary. This is independent of whether the relative permittivity changes or not. Only free charge is the source of \mathbf{D} . In turn, this shows that the normal component of the electric field must be discontinuous across the boundary, because $\mathbf{D} = \epsilon_0 \epsilon \mathbf{E}$, which is self-consistent, because electric field ends on bound charge as well as free charge.

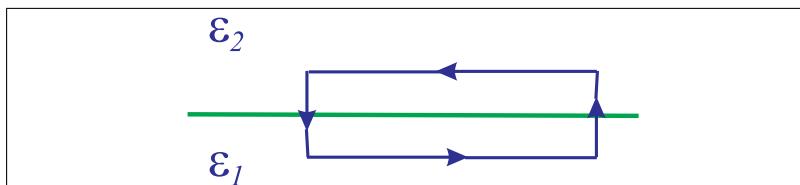


Figure 3.8: The boundary between dissimilar dielectric materials.

Now consider the boundary condition for the electric field. Place a closed loop around the boundary, as shown in Fig. 3.8, and apply Stokes’s theorem,

$$\int \mathbf{E}(\mathbf{r}) \cdot d\mathbf{l} = 0. \quad (3.32)$$

If the sides of the loop are infinitesimally small, and the \mathbf{E} field is constant along the sides, then we have

$$\mathbf{E}_{2\parallel}L - \mathbf{E}_{1\parallel}L = \mathbf{0}, \quad (3.33)$$

from which it follows

$$\mathbf{E}_{1\parallel} = \mathbf{E}_{2\parallel}. \quad (3.34)$$

We conclude that the parallel component of \mathbf{E} must be continuous across the boundary, which is independent of whether the relative permittivity changes or not. The parallel component of \mathbf{D} must be discontinuous across the boundary because $\mathbf{E} = \mathbf{D}/\epsilon_0\epsilon$.

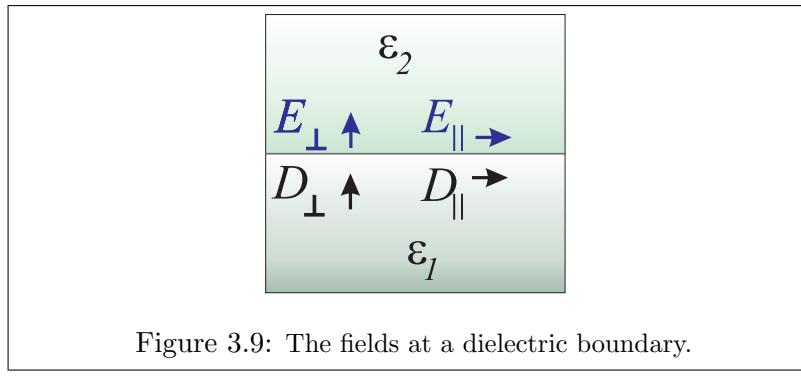


Figure 3.9: The fields at a dielectric boundary.

In conclusion, across a dielectric boundary, using Fig. 3.9,

- The normal component of \mathbf{D} is continuous (\mathbf{D}_\perp continuous).
- The normal component of \mathbf{E} is discontinuous.
- The parallel component of \mathbf{E} is continuous (\mathbf{E}_\parallel continuous).
- The parallel component of \mathbf{D} is discontinuous.

3.10 The behaviour of field lines at dielectric boundaries

Consider how the boundary conditions on \mathbf{D} and \mathbf{E} affect the form of the field lines.

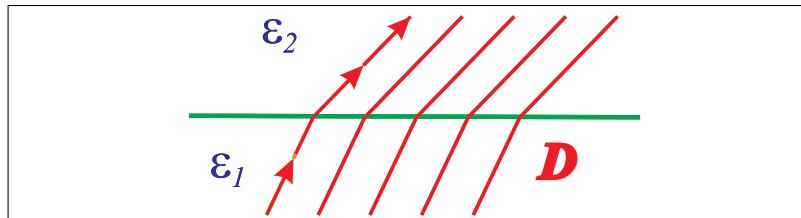


Figure 3.10: The fields at a dielectric boundary.

For $\epsilon_2 > \epsilon_1$, it is obvious that the \mathbf{D} field lines change direction as shown in Fig. 3.10, because $\mathbf{D}_{2\perp} = \mathbf{D}_{1\perp}$, and $\mathbf{D}_{2\parallel} > \mathbf{D}_{1\parallel}$.

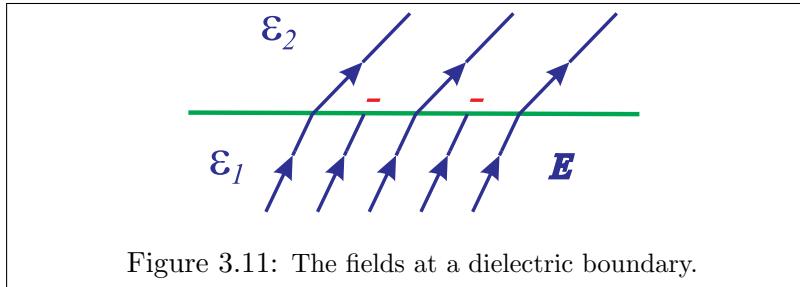


Figure 3.11: The fields at a dielectric boundary.

The \mathbf{E} field has the same form, as shown in Fig. 3.11, because $E_{2\parallel} = E_{1\parallel}$, and $E_{2\perp} < E_{1\perp}$. In fact \mathbf{D} and \mathbf{E} point in the same direction for every point in space.

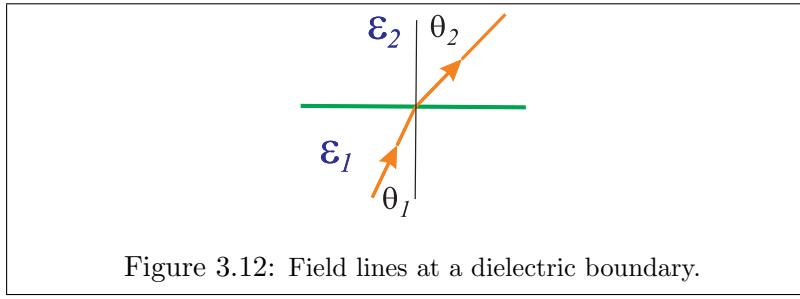


Figure 3.12: Field lines at a dielectric boundary.

The change in the direction of the field lines can be quantified as follows. For the \mathbf{D} field, from D_{\perp} being constant across a boundary, using Fig. 3.12, we have

$$D_1 \cos \theta_1 = D_2 \cos \theta_2, \quad (3.35)$$

whereas for the \mathbf{E} field, from E_{\parallel} being constant across a boundary:

$$E_1 \sin \theta_1 = E_2 \sin \theta_2. \quad (3.36)$$

Also $\mathbf{D} = \epsilon_0 \epsilon \mathbf{E}$, so (3.35) becomes

$$\epsilon_0 \epsilon_1 E_1 \cos \theta_1 = \epsilon_0 \epsilon_2 E_2 \cos \theta_2. \quad (3.37)$$

Dividing by eq. 3.36, we get for the change in the direction of the field lines at dielectric interfaces:

$$\epsilon_1 \cot \theta_1 = \epsilon_2 \cot \theta_2 \quad (3.38)$$

or

$$\frac{\cot \theta_2}{\cot \theta_1} = \frac{\epsilon_1}{\epsilon_2}. \quad (3.39)$$

We now have quantitatively derived the change in the direction of the field lines as a function of the relative permittivities of two materials across an interface. It is this change in direction that makes the analysis of general systems of dielectric bodies quite complicated. Of course, if the field lines are perpendicular to the surface, such as a parallel-plate capacitor with two dielectrics, or coaxial conductors with a coaxial dielectric filler, then the situation is straightforward to analyse. In fact, this is a key feature of all systems that can be analysed easily.

3.11 Boundary-value problems with dielectrics

Once dielectric bodies having arbitrary shapes are combined with conducting surfaces having arbitrary shapes, electrostatic problems become complicated to solve, and the modern approach would be to set up a numerical model based on, say, Poisson's equation. Historically, other methods of solution had to be found, and this resulted in a rich structure of mathematical techniques. Indeed significant advances were made in mathematics itself because of the need to understand how to solve the integral and differential equations associated with electrostatic problems.

In the remainder of this Section, we look at a number of examples of increasing complexity.

- **Long thin rod parallel to a uniform field**

Consider placing a long thin rod parallel to a uniform electric field of value \mathbf{E}_0 : Fig. 3.13.

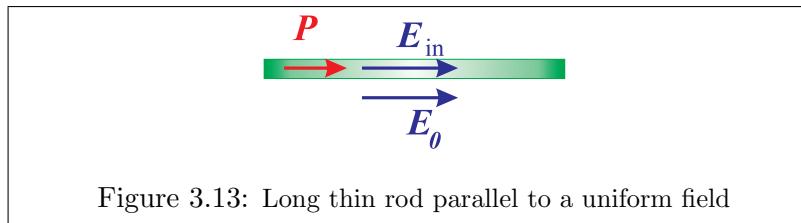


Figure 3.13: Long thin rod parallel to a uniform field

Because the tangential component of the electric field must be continuous across the dielectric boundary, the internal electric field must be the same as the external electric field:

$$\mathbf{E}_{\text{in}} = \mathbf{E}_0. \quad (3.40)$$

We can therefore find the polarisation

$$\mathbf{P} = \epsilon_0 \mathbf{E}_0 \chi. \quad (3.41)$$

- **Thin slab perpendicular to a uniform field**

Now consider a thin slab perpendicular to the field, as shown in Fig. 3.14.

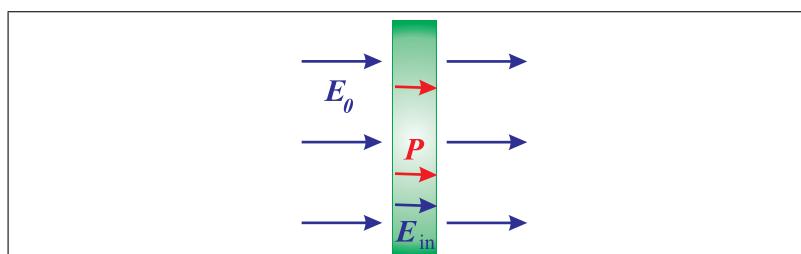


Figure 3.14: Thin slab perpendicular to a uniform field.

Because the slab is very large, we expect the internal electric field to be uniform and parallel to the external electric

field. This is justified because, by symmetry, there can be no components in other directions. Since \mathbf{D} and \mathbf{E} are parallel, as shown in Section 3.10, also \mathbf{D} is perpendicular to the boundary, i.e., $\mathbf{D} = \mathbf{D}_\perp$. Since \mathbf{D}_\perp is continuous across the boundary, as we have seen in Section 3.9, it follows that

$$\epsilon_0 \epsilon \mathbf{E}_{\text{in}} = \epsilon_0 \mathbf{E}_0, \quad (3.42)$$

giving

$$\mathbf{E}_{\text{in}} = \frac{\mathbf{E}_0}{\epsilon} = \frac{\mathbf{E}_0}{1 + \chi}, \quad (3.43)$$

and finally

$$\mathbf{P} = \epsilon_0 \mathbf{E}_0 \frac{\chi}{1 + \chi}. \quad (3.44)$$

These two examples serve to demonstrate that the polarisation and electric field inside a dielectric body are dependent on the shape of the body—as one would expect. What is not so obvious is that for many simple shapes, such as a cylinder or sphere, the relationship always takes the form

$$\mathbf{P} = \epsilon_0 \mathbf{E}_0 \frac{\chi}{1 + n\chi}, \quad (3.45)$$

where $0 < n < 1$, and for a cylinder $n = 1/2$ and for a sphere $n = 1/3$, as we will see.

• Dielectric sphere in a uniform field

Consider a dielectric sphere with radius a in a uniform electric field, as shown in Fig. 3.15.

Our Ansatz is to find the solution of Poisson's equation $\nabla^2 V = -(\rho_f + \rho_p)/\epsilon_0$ that satisfies the boundary conditions. Because there is no free charge in this problem — although formally free charge would be needed to create the parallel field — Poisson's equation reduces to $\nabla^2 V = -\rho_p/\epsilon_0$.

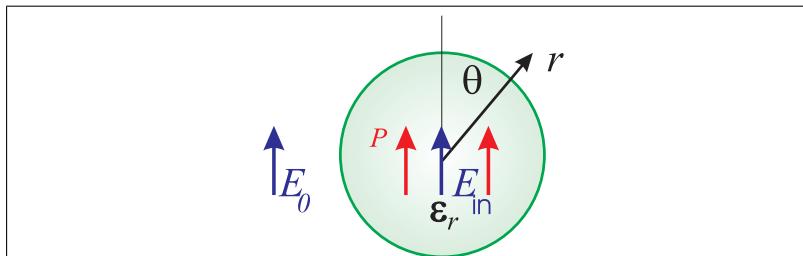


Figure 3.15: Geometry of a dielectric *sphere* in a uniform field.

Let us guess that the internal field is uniform, and that the external field is the original uniform field plus a dipole field generated by the surface polarisation charge on the sphere, which is similar to the case of the metallic sphere considered previously (cf. Section 2.21).

Using spherical polar coordinates, the potentials then become

$$\begin{aligned} V_{\text{in}} &= -E_{\text{in}} r \cos \theta = E_{\text{in}} z \\ V_0 &= -E_0 r \cos \theta + \frac{\kappa \cos \theta}{r^2}, \end{aligned} \quad (3.46)$$

where κ is some constant of proportionality, whose value is to be found.

We require that the parallel component E_{\parallel} be continuous across the boundary, where

$$E_{\parallel} = E_{\theta} = -\frac{1}{r} \frac{\partial V}{\partial \theta} \Big|_{r=a}. \quad (3.47)$$

Therefore

$$-E_{\text{in}} \sin \theta = -E_0 \sin \theta + \frac{\kappa \sin \theta}{r^3} \Big|_{r=a}, \quad (3.48)$$

and

$$E_{\text{in}} = E_0 - \frac{\kappa}{a^3}, \quad (3.49)$$

which can also be derived by requiring the potential to be continuous.

We also require that the normal component D_{\perp} be continuous, where

$$D_{\perp} = -\epsilon_0 \epsilon \frac{\partial V}{\partial r} \Big|_{r=a}. \quad (3.50)$$

Now,

$$\begin{aligned} D_{\perp \text{in}} &= \epsilon_0 \epsilon E_{\text{in}} \cos \theta \\ D_{\perp 0} &= \epsilon_0 E_0 \cos \theta + \epsilon_0 \frac{2\kappa \cos \theta}{a^3} \end{aligned} \quad (3.51)$$

where we have used the fact that $\epsilon = 1$ outside the dielectric. Therefore, using $D_{\perp \text{in}} = D_{\perp 0}$ and Eq. (3.49)

$$\epsilon E_{\text{in}} \stackrel{D_{\perp \text{in}} = D_{\perp 0}}{=} E_0 + \frac{2\kappa}{a^3} \stackrel{(3.49)}{=} \epsilon \left(E_0 - \frac{\kappa}{a^3} \right). \quad (3.52)$$

Hence we find

$$\kappa = \left(\frac{\epsilon - 1}{\epsilon + 2} \right) a^3 E_0, \quad (3.53)$$

and therefore can conclude considering Eq. (3.49) that the internal field is uniform:

$$\mathbf{E}_{\text{in}} = \frac{3}{\epsilon + 2} \mathbf{E}_0, \quad (3.54)$$

and finally that the polarisation is uniform, too:

$$\begin{aligned} \mathbf{P} &= \epsilon_0 E_{\text{in}} \chi \\ &= \frac{3\chi}{2 + \epsilon} \epsilon_0 \mathbf{E}_0 \\ &= \frac{\chi}{1 + \chi/3} \epsilon_0 \mathbf{E}_0. \end{aligned} \quad (3.55)$$

In summary, we find

$$\mathbf{P} = \frac{\chi}{1 + \chi/3} \epsilon_0 \mathbf{E}_0, \quad (3.56)$$

confirming our original guess about the relationship between polarisation and external field. Since it satisfies Poisson's equation and the boundary conditions, by the uniqueness theorem it is **the** solution, cf. Sect. 2.20. The \mathbf{D} field and the equipotentials are shown in Fig. 3.16.

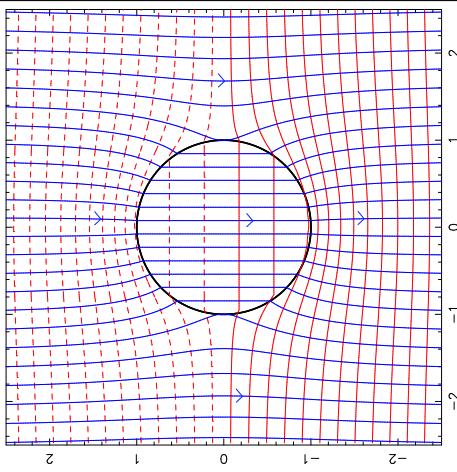


Figure 3.16: Dielectric sphere in a uniform field. Lines of \mathbf{D} are roughly horizontal (blue, dark), equipotentials are roughly vertical (red, light). In this example, $\epsilon = 4$.

• Other dielectric bodies

Finding the electrostatic fields associated with other dielectric bodies is more complicated, and requires more sophisticated solutions of Poisson's equation. As representative examples, the solutions for oblate and prolate spheroids are shown in Fig. 3.17 and Fig. 3.18, respectively. Once the electric field is known, many other quantities follow, such as the couple exerted on a spheroid.

We shall not continue this topic further, but it is easy to gain an impression of the strength of the tools and the richness of the structures described.

3.12 Energy density in dielectrics

The only remaining issue is how to calculate the energy stored in an electric field when dielectric bodies are present.

We have shown that the energy can be calculated from knowledge of the electric field alone when only free charge and conducting surfaces are present, but we have to be careful when dielectrics are introduced. The problem is that previously we calculated the energy

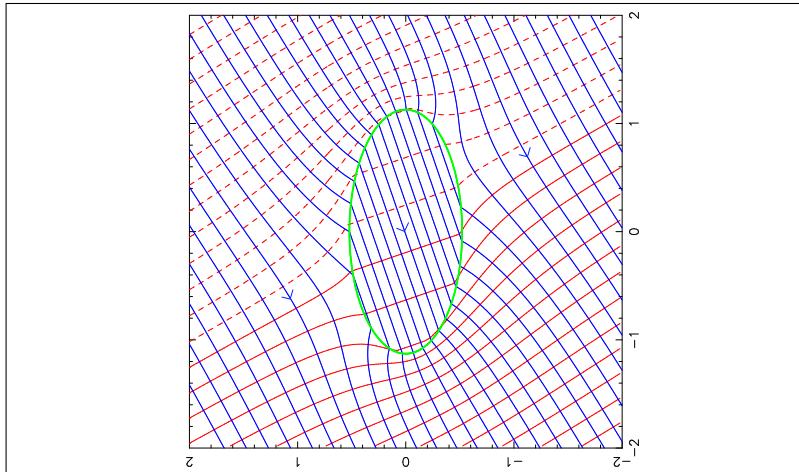


Figure 3.17: Oblate dielectric spheroid in a uniform field. Lines of \mathbf{D} slope down to the right.

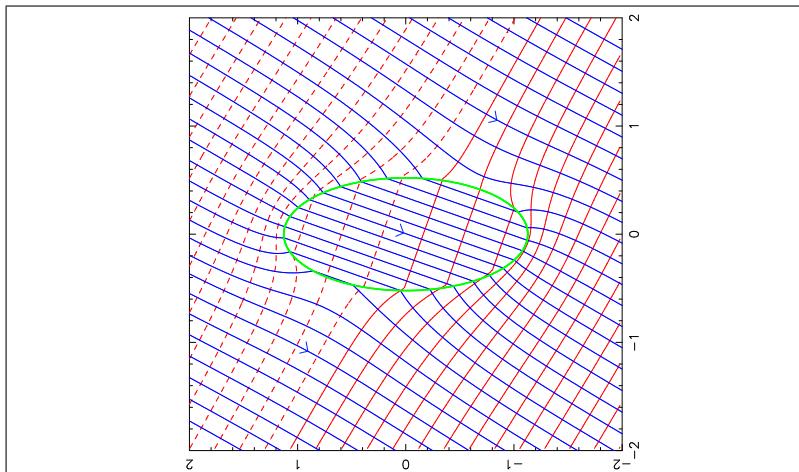


Figure 3.18: Prolate dielectric spheroid in a uniform field. Lines of \mathbf{D} slope down to the right.

required to bring up one charge at a time from infinity, and then summed all such contributions in order to find the energy needed to build a complete system. When dielectric bodies are present, however, not only is energy stored in the free charge that is moved, but energy is also stored in the bound charge that is separated. Rather than trying to reproduce the previous argument with induced charge, bringing up complete blocks of charge and placing them at the appropriate places one point at a time, which would be tedious, it is better to look at the other approach shown earlier, building up the complete charge distribution gradually, using a factor α going from 0 to 1.

We found in Eq. (2.149) that

$$U = \frac{1}{2} \int d^3r \rho(\mathbf{r}) V(\mathbf{r}). \quad (3.57)$$

In that derivation nothing was said that contradicts the argument when bound charge is present, and hence we can apply a similar

argumentation here. Bound (polarisation) charge is not explicitly included in $\rho(\mathbf{r})$, but rather gradually induced as the free charge is assembled.

Thought for the day 10: Although equations such as this one are easy to write down and manipulate, it is important to keep an eye on their physical meaning. For example, consider what the integrals mean in terms of how the charge is assembled, and how polarisation charge is induced, and understand why the potential is linear. This could be done in terms of a simple system comprising two free charges of opposite sign, with an induced dipole midway between them.

We can now proceed efficiently, in almost exactly the same way as before in Chapter ?? for $\epsilon = 1$. We have

$$\begin{aligned} U &= \frac{1}{2} \int d^3\mathbf{r} \rho(\mathbf{r}) V(\mathbf{r}) \\ &= \frac{1}{2} \int d^3\mathbf{r} \nabla \cdot \mathbf{D}(\mathbf{r}) V(\mathbf{r}). \end{aligned} \quad (3.58)$$

where Gauss's law has been used, but we also can integrate by parts using the vector identity

$$\nabla \cdot (c\mathbf{A}) = c\nabla \cdot \mathbf{A} + \mathbf{A} \cdot \nabla c, \quad (3.59)$$

where c is any scalar field, and \mathbf{A} is any vector field, and therefore

$$U = \frac{1}{2} \int d^3\mathbf{r} \{ \nabla \cdot [V(\mathbf{r})\mathbf{D}(\mathbf{r})] - \mathbf{D}(\mathbf{r}) \cdot \nabla V(\mathbf{r}) \}. \quad (3.60)$$

Now look at each of the two RHS terms in turn.

The *first term* can be replaced, through the divergence theorem, by an integration over the surface containing the field,

$$\frac{1}{2} \int d^3\mathbf{r} \nabla \cdot [V(\mathbf{r})\mathbf{D}(\mathbf{r})] = \frac{1}{2} \int d\mathbf{S} \cdot \mathbf{D}(\mathbf{r}) V(\mathbf{r}). \quad (3.61)$$

The integration surface must, however, be taken at infinity, if it is to include all of the field. For a localised charge $V(\mathbf{r})$ falls as $1/r$ and $\mathbf{D}(\mathbf{r})$ falls as $1/r^2$, whereas the area increases as r^2 , and therefore the integral on the RHS of Eq. (3.61) tends to zero for $r \rightarrow \infty$.

Only the *second term* remains, and hence the energy density in dielectrics is

$$\begin{aligned} U &= -\frac{1}{2} \int d^3\mathbf{r} \mathbf{D}(\mathbf{r}) \cdot \nabla V(\mathbf{r}) \\ U &= \frac{1}{2} \int d^3\mathbf{r} \mathbf{D}(\mathbf{r}) \cdot \mathbf{E}(\mathbf{r}). \end{aligned} \quad (3.62)$$

The argumentation above is similar to the dielectric-free version. In fact, for free space, where $\mathbf{D}(\mathbf{r}) = \epsilon_0 \mathbf{E}(\mathbf{r})$ it gives exactly the same result.

We conclude that the **energy stored in an electrostatic field** is given by the general equation

$$U = \frac{1}{2} \int d^3r \mathbf{D}(\mathbf{r}) \cdot \mathbf{E}(\mathbf{r}). \quad \begin{array}{l} \text{Energy stored in} \\ \text{an electrostatic field} \end{array} \quad (3.63)$$

This relation is generally applicable, in free space or when insulating dielectric materials are present.

4. Magnetostatic fields

4.1 Introduction

We now turn our attention to *magnetostatics*: the study of static magnetic fields. In many ways, the mathematical structures used to study magnetic fields are similar to those used to study electric fields.

In the same way that static charge gives rise to electric fields, electrical current, or equivalently moving charge, gives rise to magnetic fields.

This association implies that there is a close relationship between electric and magnetic fields, which is the case.

For example, if a stationary observer sees charge move past, does the observer detect a time-varying electric field or a time-varying magnetic field? This is an important aspect of the theory of special relativity.

In time-invariant (static) systems, we *can* treat the electric and magnetic fields as separate.

Because we are concerned with the behaviour of static magnetic fields, we shall assume that all electrical currents are time-invariant.

We shall also assume that there is no net charge anywhere.

Thought for the day 11: How is it possible to have a current, but no overall free electrical charge? After all, a current corresponds to moving charge.

4.2 Magnetostatic force

Define a **current element** as a filament of wire of infinitesimal length carrying current I . Let the infinitesimal vector corresponding to this current element be denoted by dl .

The direction of the vector obviously corresponds to the direction of current flow.

The magnetic field at a point tells us about the force, \mathbf{F} , that will be exerted on a small test current filament that is inserted at the point. Unlike electrical charge, however, a current element has an orientation, and its orientation with respect to the magnetic field will be important in determining the force.



Jean-Baptiste Biot
(1774–1862)

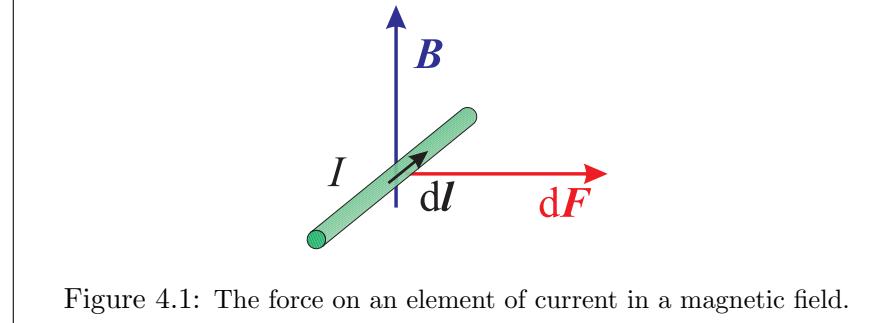


Figure 4.1: The force on an element of current in a magnetic field.

The force on a current element is given by

$$d\mathbf{F} = I dl \times \mathbf{B} \quad \text{Force on current element due to magnetic field.} \quad (4.1)$$

This is the **definition** of the magnetic field \mathbf{B} .

N.B. This formula is consistent with the Lorentz force on a moving charge: put $I dl = \left(\frac{dq}{dt}\right) \mathbf{v} dt = dq \mathbf{v} \Rightarrow d\mathbf{F} = dq \mathbf{v} \times \mathbf{B}$.

Conversely, the magnetic field produced by a current element dl is given by the **Biot-Savart law**:



Felix Savart (1791–1841)

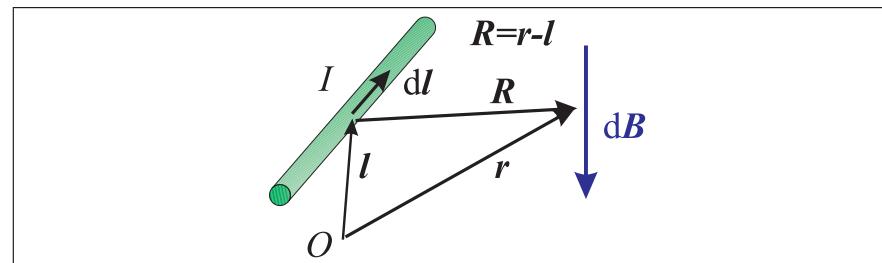


Figure 4.2: The magnetic field produced by an element of current. Note the direction of \mathbf{R} .

$$d\mathbf{B} = \frac{\mu_0 I}{4\pi R^3} dl \times \mathbf{R} = \frac{\mu_0 I}{4\pi R^2} dl \times \hat{\mathbf{R}} \quad \text{Biot-Savart law} \quad (4.2)$$

Here, μ_0 is a constant of proportionality, called the **permeability** of free space.

The magnetic field decreases as the square of the distance from the element.

Clearly, the magnetic field lines ‘circulate’ around the current element.

The Biot-Savart law is a particularly powerful tool for analysing magnetostatic systems. It is most useful when the field does not have any particular symmetry.

In fact, for any current distribution, we can write

$$\boxed{\mathbf{B} = \frac{\mu_0}{4\pi} \int d\mathbf{l} \times \mathbf{R} \frac{I}{R^3}}, \quad (4.3)$$

where the current appears under the integral, because, in general, it too can be a function of position.

4.3 The force between two current elements

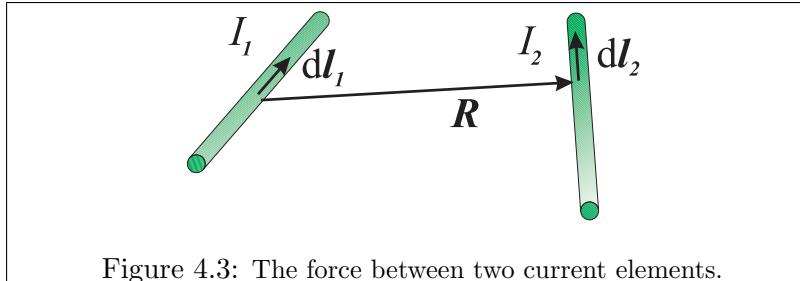


Figure 4.3: The force between two current elements.

Combining equations (4.2)

$$d\mathbf{B}_2 = \frac{\mu_0 I_1}{4\pi R^3} d\mathbf{l}_1 \times \mathbf{R} \quad (4.4)$$

and (4.1)

$$d\mathbf{F}_2 = I_2 d\mathbf{l}_2 \times d\mathbf{B}_2, \quad (4.5)$$

we obtain

$$d\mathbf{F}_2 = \frac{I_1 I_2 \mu_0}{4\pi R^3} d\mathbf{l}_2 \times (d\mathbf{l}_1 \times \mathbf{R}). \quad (4.6)$$

Be careful, the vector triple product is not associative: the order in which the cross products are evaluated is important.

The force is greatest when the current elements are aligned, and is attractive when the currents are flowing **in the same** direction.

4.4 The force between currents

The force between two parallel wires was used in the preliminary definition of SI units (valid up to 20 May 2019) to define the unit of electrical current.

The ampere is the current that, when flowing in each of two parallel straight conductors, of infinite length and negligible cross section, placed 1 m apart in a vacuum, produces a force of 2×10^{-7} N m⁻¹.

This force can be calculated using the Biot-Savart law for the force between current elements (see problem sheet), or using Ampère's law (see later). The chosen value **defines** μ_0 to be

$$\boxed{\mu_0 = 4\pi \times 10^{-7} \text{ H m}^{-1}.} \quad (4.7)$$

4.5 The field on the axis of a current loop

Consider how the Biot-Savart law can be used to calculate the magnetic field on the axis of a loop of wire carrying current I .

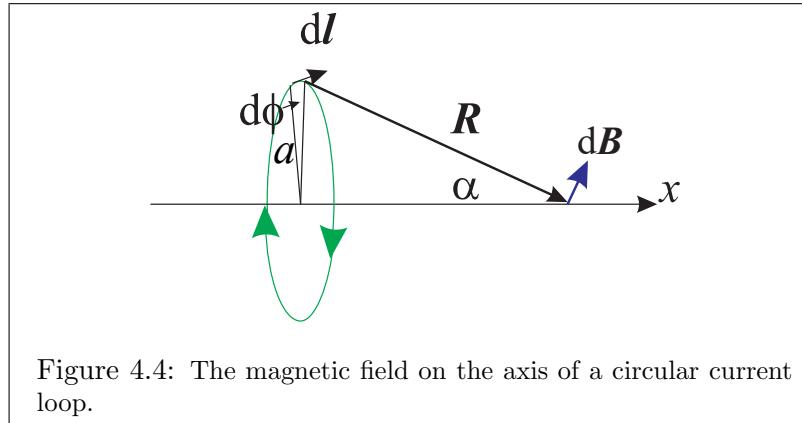


Figure 4.4: The magnetic field on the axis of a circular current loop.

Consider the loop to be made up of many individual line elements. The Biot-Savart law allows us to calculate the field associated with any one of these elements at any position:

$$d\mathbf{B} = \frac{\mu_0 I}{4\pi R^3} dl \times \mathbf{R}. \quad (4.8)$$

After adding together the fields associated with each of the current elements, only the x component remains, and therefore

$$\begin{aligned} B_x &= a \sin \alpha \frac{\mu_0 I}{4\pi R^2} \int_0^{2\pi} d\phi \\ &= a \sin \alpha \frac{\mu_0 I}{2R^2}. \end{aligned} \quad (4.9)$$

Or, using $R^2 = a^2 + x^2$ and the definition of $\sin \alpha$:

$$B_x = \frac{\mu_0 I a^2}{2(a^2 + x^2)^{3/2}}. \quad (4.10)$$

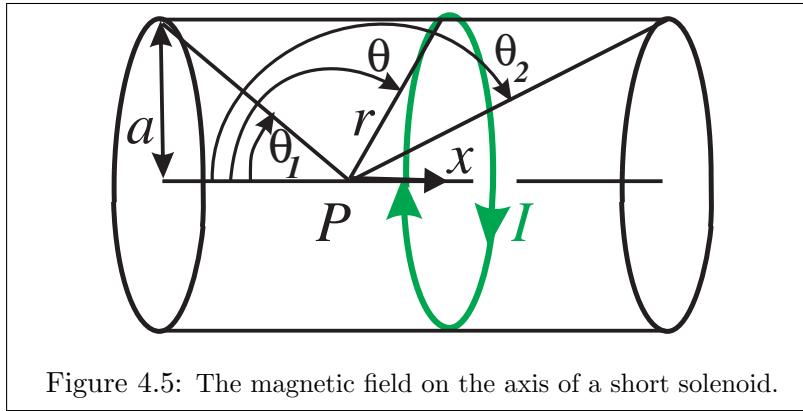
4.6 The field on the axis of a short solenoid

The field on the axis of a solenoid can be found by merely adding together the fields associated with a line of closely packed loops.

The field at P due to the loop at x becomes, considering Eq. (4.9):

$$dB_P = a \sin \theta \frac{\mu_0 I}{2r^2} n dx, \quad (4.11)$$

where n is the number of turns per unit length along x , and $\theta \equiv \alpha$ from Sect. 4.5.



We also have

$$x = -a \cot \theta \quad \text{and} \quad r = a / \sin \theta, \quad (4.12)$$

and therefore,

$$dx = \frac{a}{\sin^2 \theta} d\theta = \frac{r^2}{a} d\theta. \quad (4.13)$$

It follows that

$$dB_P = \frac{\mu_0 I n}{2} \sin \theta d\theta, \quad (4.14)$$

and

$$\begin{aligned} B_P &= \frac{\mu_0 I n}{2} \int_{\theta_1}^{\theta_2} \sin \theta d\theta \\ &= \frac{\mu_0 I n}{2} (\cos \theta_1 - \cos \theta_2). \end{aligned} \quad (4.15)$$

In the middle of a long solenoid, $\theta_1 \rightarrow 0$ and $\theta_2 \rightarrow \pi$, and therefore

$$B = \mu_0 n I. \quad (4.16)$$

The field is proportional to the number of turns per unit length and the current.

4.7 Gauss's Law for magnetic fields

In general, a magnetic field will change direction and strength as we move around in some region of space. In the same way as we defined electric flux for electrostatic problems, we can define magnetic flux, Φ , for magnetostatic problems. In fact, once a vector field has been established, the mathematical definition of **magnetic flux** over a surface is immediate:

$$\Phi = \int_S d\mathbf{S} \cdot \mathbf{B}(\mathbf{r}). \quad \text{Magnetic flux} \quad (4.17)$$

Notice that the flux is a scalar quantity, and in a crude sense is a measure of the total amount of magnetic field passing normally through S . The magnetic flux is only useful, however, if it proves to be valuable when analysing physical problems. We shall understand its physical usefulness later. Because we integrate $\mathbf{B}(\mathbf{r})$ to get a flux Φ , $\mathbf{B}(\mathbf{r})$ is often called the **magnetic flux density**.

Magnetic field lines have a number of characteristic features:

- Magnetic field lines wrap around electrical currents, and form closed loops.
- There are no point sources of magnetic field, in the sense that magnetic fields do not appear to radiate from points. There is no evidence that magnetic monopoles exist anywhere in the Universe.
- There is no experimental evidence that magnetic monopoles exist
- Because magnetic field lines always wrap around on themselves, the total magnetic flux passing through a closed surface must be zero. The same amount of flux must enter a closed surface as leaves it.

Formulating the last point mathematically, we have

$$\oint_S d\mathbf{S} \cdot \mathbf{B}(\mathbf{r}) = 0. \quad (4.18)$$

Using the divergence theorem

$$\oint_S d\mathbf{S} \cdot \mathbf{B}(\mathbf{r}) = \int_V dV \nabla \cdot \mathbf{B}(\mathbf{r}) = 0, \quad (4.19)$$

but because this must be true for all closed surfaces

$$\boxed{\nabla \cdot \mathbf{B}(\mathbf{r}) = 0} \quad \text{Maxwell's equation 3} \quad (4.20)$$

everywhere. **This is one of Maxwell's equations.** It should be contrasted with the equivalent expression for an electric field:

$$\boxed{\nabla \cdot \mathbf{D}(\mathbf{r}) = \rho(\mathbf{r}).} \quad \text{Maxwell's equation 1}$$

The observation that the divergence of a magnetic field is everywhere zero is an experimental fact of considerable importance.

4.8 Loops and magnetic dipoles

There are no magnetic monopoles; in fact the simplest physically realisable sources are magnetic dipoles. Notice that our infinitesimal current element cannot exist physically because there is nowhere for the current to circulate. A small loop of current, however, produces, at large distances, a magnetic dipole field, and this can be regarded as the basic building block of all magnetic fields. In fact we may think of permanent magnets as being made up of large numbers of magnetic dipoles.

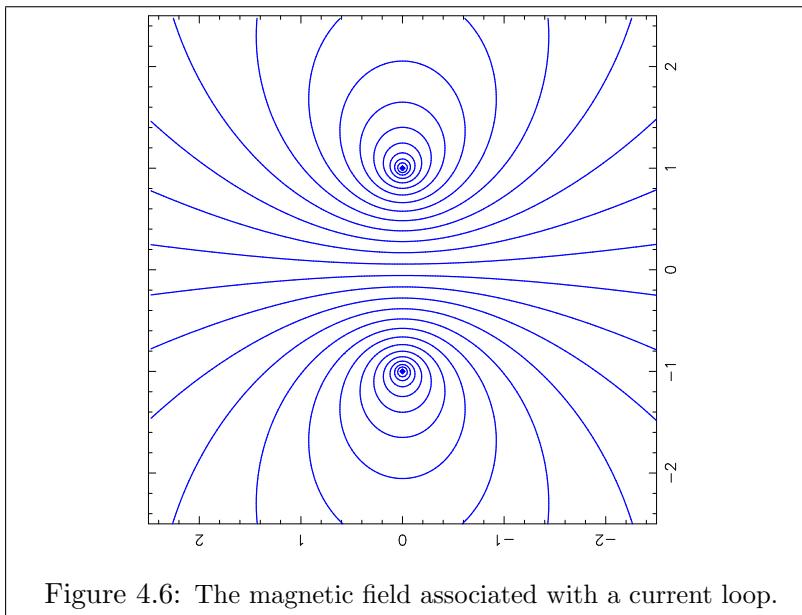


Figure 4.6: The magnetic field associated with a current loop.

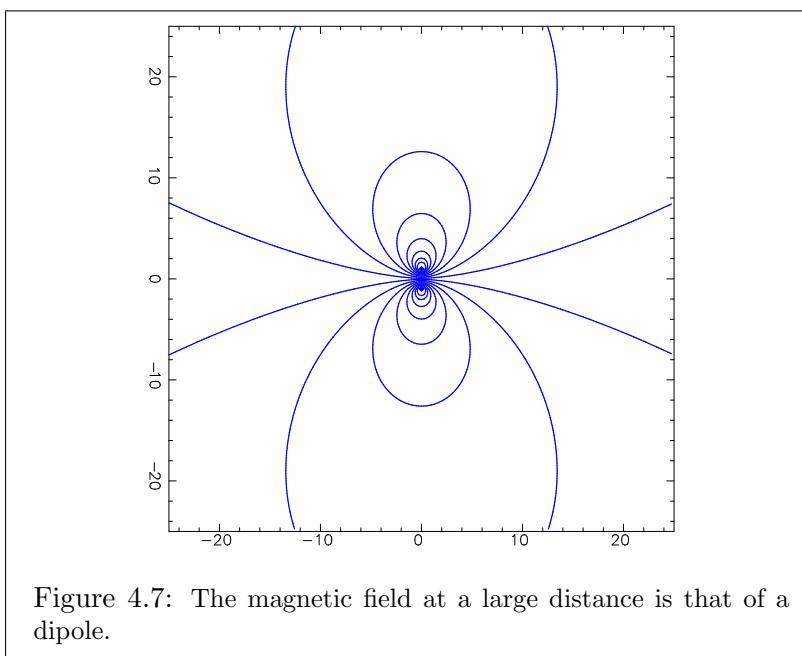


Figure 4.7: The magnetic field at a large distance is that of a dipole.

Consider an elementary current loop in a uniform magnetic field.

What force does the loop experience?

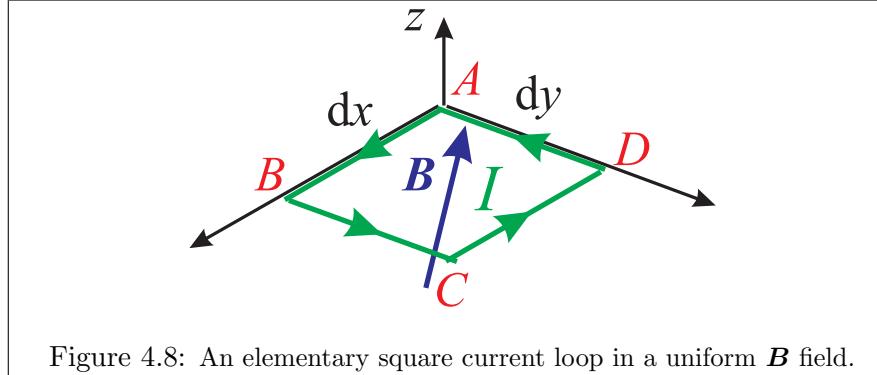


Figure 4.8: An elementary square current loop in a uniform \mathbf{B} field.

There is no net translation force on the loop because the forces on opposite sides act in opposite directions. The loop does experience a **couple**. Remember from Section 2.8 that when a system comprises two equal, but oppositely directed forces, the product of the perpendicular distance between them and the magnitude of one of the forces gives the magnitude of the couple. Consider each of the components of \mathbf{B} in turn:

- The x -directed component of \mathbf{B} does not produce a force on AB or CD;
- The force on DA is in the positive- z direction, and has magnitude $IdyB_x$;
- The force on BC is in the negative- z direction and has the same magnitude.

The couple is therefore in the y direction, and

$$dG_y = IdxdyB_x. \quad (4.21)$$

Likewise:

- B_y does not produce a force on BC or DA;
- The force on AB is in the positive- z direction, and has magnitude $IdxB_y$;
- The force on CD is in the negative- z direction and has the same magnitude.

The couple is therefore in the $-x$ direction, and

$$dG_x = -IdxdyB_y. \quad (4.22)$$

The z -directed component of \mathbf{B} does not produce a translational force or a couple, but attempts to expand the loop, which is rigid.

Thus,

$$d\mathbf{G} = (-IdxdyB_y, IdxdyB_x, 0). \quad (4.23)$$

Define the **infinitesimal magnetic dipole moment**, $d\mathbf{m}$, by

$$d\mathbf{m} = Id\mathbf{S}, \quad (4.24)$$

which is a vector perpendicular to the plane of the loop. So $Id\mathbf{S} = (0, 0, Idxdy)$ for this loop. Then the **infinitesimal couple on a magnetic dipole moment due to a magnetic field** is

$$d\mathbf{G} = d\mathbf{m} \times \mathbf{B}, \quad (4.25)$$

which can be verified by substitution.

Let us now suppose that we have some arbitrarily shaped loop. What couple does it experience in a uniform field?

We can imagine the loop to be built up from a large number of infinitesimally small current loops. If all of these are held rigidly with respect to each other, the total couple, \mathbf{G} , is just the vector sum of the individual couples associated with the elemental dipoles:

$$\mathbf{G} = I \int_S d\mathbf{S} \times \mathbf{B}, \quad (4.26)$$

Since I is a constant factor, it simply scales the magnitude of the elemental dipoles.

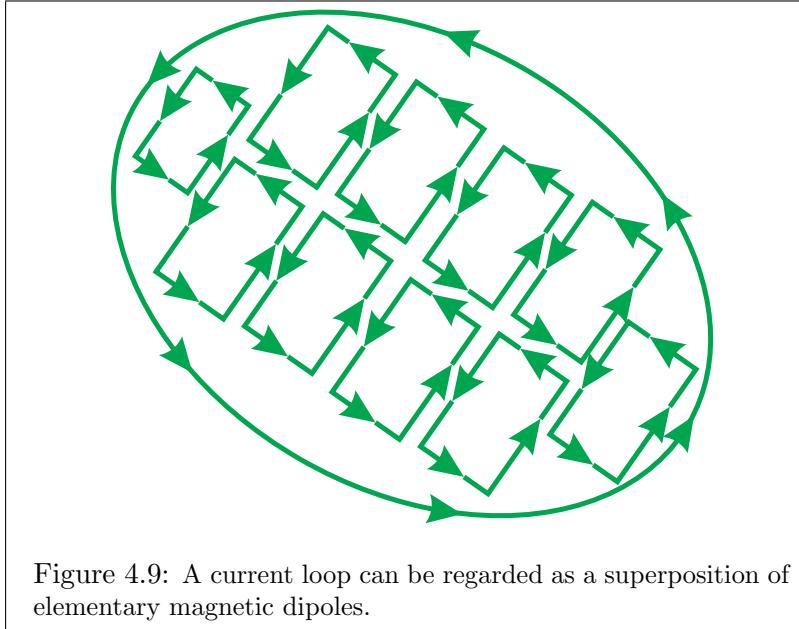


Figure 4.9: A current loop can be regarded as a superposition of elementary magnetic dipoles.

Then we can write for the **couple on a magnetic dipole moment due to a magnetic field**:

$\mathbf{G} = \mathbf{m} \times \mathbf{B}$,	Couple on a magnetic dipole moment
---	------------------------------------

(4.27)

where the **magnetic dipole moment** is given as

$$\mathbf{m} = I \int_S d\mathbf{S}. \quad \text{Magnetic dipole moment} \quad (4.28)$$

In other words, the dipole moment of the complete loop can be found by integrating the surface normal over any surface, S , that has the loop as its boundary.

It can also be appreciated that the current I can be identified with the current that flows around the loop, which follows because the currents associated with the individual dipoles cancel on any point in the interior of the surface. Notice that the loop does not have to lie in a plane; the surface used to find the dipole can have any form.

4.9 Potential energy of a magnetic dipole in a uniform field

As we have seen, there is no translational force on a magnetic dipole in a uniform magnetic field; there is, however, a couple and if we rotate the dipole against the field, energy must be supplied, and therefore stored in the form of potential energy.

Suppose that we apply an external couple to the dipole in order to increase the angle θ with respect to the direction of the field. The work done is then

$$dW = |\mathbf{G}| d\theta', \quad (4.29)$$

which can be integrated to give the change in potential energy after rotation through some total angle:

$$U_{\theta_B \theta_A} = \int_{\theta_A}^{\theta_B} d\theta' |\mathbf{G}| = \int_{\theta_A}^{\theta_B} d\theta' |\mathbf{m}| |\mathbf{B}| \sin \theta'. \quad (4.30)$$

It is clear that the potential energy will be at a minimum when the dipole is aligned with the field, which occurs at $\theta = 0$, and at a maximum when the dipole is aligned with the field in the opposite direction, $\theta = \pi$.

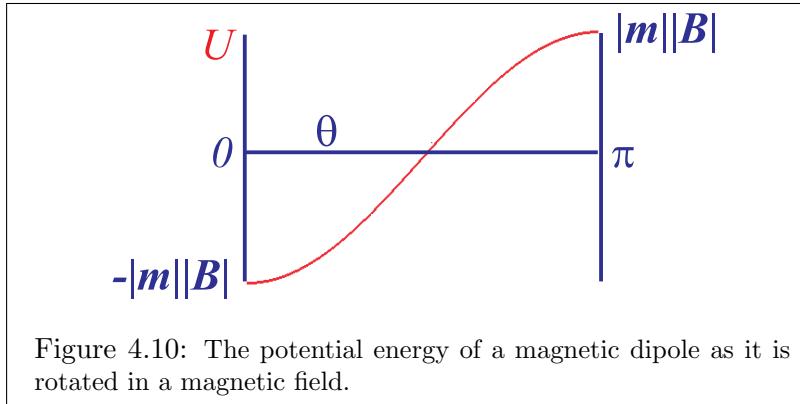
Take the reference angle for the potential to be $\theta_A = \pi/2$. The potential can then take on positive or negative values. Therefore

$$U = \int_{\pi/2}^{\theta} d\theta' |\mathbf{m}| |\mathbf{B}| \sin \theta' = -|\mathbf{m}| |\mathbf{B}| \cos \theta. \quad (4.31)$$

We conclude for the **potential of a magnetic dipole in \mathbf{B}** :

$$U = -\mathbf{m} \cdot \mathbf{B} \quad \text{Potential of a magnetic dipole,} \quad (4.32)$$

which is identical to the form of the electric-dipole case.



In the case of a macroscopic current loop, we can assemble this from a set of microscopic loops (dipoles) $d\mathbf{m} = Id\mathbf{S}$, each in a uniform field \mathbf{B} , so the total energy becomes

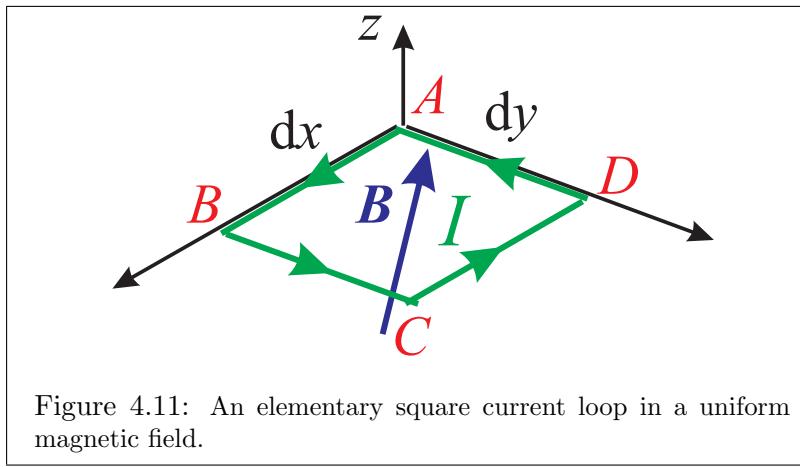
$$U = - \int d\mathbf{m} \cdot \mathbf{B} = - \int_S Id\mathbf{S} \cdot \mathbf{B} = -I\Phi, \quad (4.33)$$

where Φ is the magnetic flux passing normally through the loop.

4.10 Force on a magnetic dipole in a non-uniform field

When a magnetic dipole is placed in a non-uniform field, it experiences a translational force in addition to a couple. The translational force can be found as follows.

Consider the elementary current loop introduced earlier.



The B_x component of the magnetic field induces a translational force in the z direction of

$$B_x Idy - (B_x + \frac{\partial B_x}{\partial x} dx) Idy = -\frac{\partial B_x}{\partial x} Idxdy; \quad (4.34)$$

likewise the B_y component introduces a translational force of

$$B_y Idx - (B_y + \frac{\partial B_y}{\partial y} dy) Idx = -\frac{\partial B_y}{\partial y} Idxdy. \quad (4.35)$$

Combining,

$$F_z = -\left(\frac{\partial B_x}{\partial x} + \frac{\partial B_y}{\partial y}\right)Idxdy = \frac{\partial B_z}{\partial z}Idxdy, \quad (4.36)$$

where the second equality follows because $\nabla \cdot \mathbf{B}(\mathbf{r}) = 0$ everywhere.

Likewise, for this same loop we also have

$$F_x = \frac{\partial B_z}{\partial x}Idxdy, \quad (4.37)$$

and

$$F_y = \frac{\partial B_z}{\partial y}Idxdy. \quad (4.38)$$

Now consider connecting together rigidly three elemental current loops in the three principal planes, such that the overall dipole moment can point in any direction. Adding all of the contributions, we get

$$\begin{aligned} F_x &= m_x \frac{\partial B_x}{\partial x} + m_y \frac{\partial B_y}{\partial x} + m_z \frac{\partial B_z}{\partial x} \\ F_y &= m_x \frac{\partial B_x}{\partial y} + m_y \frac{\partial B_y}{\partial y} + m_z \frac{\partial B_z}{\partial y} \\ F_z &= m_x \frac{\partial B_x}{\partial z} + m_y \frac{\partial B_y}{\partial z} + m_z \frac{\partial B_z}{\partial z}, \end{aligned} \quad (4.39)$$

where the appropriate notation has been inserted for the dipole moments, or, more simply,

$F_i = m_j \frac{\partial B_j}{\partial x_i}, \quad \text{Force on a magnetic dipole in a non-uniform } \mathbf{B}$

(4.40)

using suffix notation and the Einstein summation convention.

If the dipole moments are constant in space, they can be taken inside the spatial derivatives, and we have

$\mathbf{F}(\mathbf{r}) = \nabla [\mathbf{m} \cdot \mathbf{B}(\mathbf{r})].$

(4.41)

The form is identical to that for the force on an electric dipole in a non-uniform electric field. If the dipole cannot rotate, it will attempt to move to a position where the dipole and field are aligned.

4.11 Magnetic scalar potential

In our work on electrostatics, we saw that it is convenient to generate the electric field from the gradient of a scalar field. Can we imagine some landscape where the vector field of interest, in this case $\mathbf{B}(\mathbf{r})$, always points downhill with respect to some potential surface with equipotential contours? Of course, we must imagine

this surface in three dimensions. In other words, we would like to write

$\mathbf{H}(\mathbf{r}) = -\nabla\phi_m(\mathbf{r}),$	Magnetic field strength from magnetic scalar potential
---	---

(4.42)

with

$$\mathbf{B}(\mathbf{r}) = \mu_0 \mathbf{H}(\mathbf{r}), \quad (4.43)$$

such that one obtains for the magnetic field, also known as magnetic flux density:

$\mathbf{B}(\mathbf{r}) = -\mu_0 \nabla\phi_m(\mathbf{r}).$

(4.44)

At this stage, $\mathbf{B}(\mathbf{r})$ and $\mathbf{H}(\mathbf{r})$, are related simply through the permeability of free space, μ_0 . Later, when magnetic materials are introduced they will be seen to have different physical interpretations, but here we can simply regard one as being a scaled version of the other.

We might wonder if every vector field can be represented in terms of a potential surface, or whether there are certain fields that do not permit such a description. The answer is clear: consider representing a field as a potential function, and then calculating the curl of the result:

$$\nabla \times [-\mu_0 \nabla\phi_m(\mathbf{r})] = \mathbf{0}, \quad (4.45)$$

where the equality follows because the curl of the gradient of any scalar field always evaluates to zero. Thus by representing a field in this way, we have forced its curl to be zero. In the case of static electric fields, the curl of \mathbf{E} is always zero, as we saw, but we do not yet know whether the curl of a magnetic field is always zero.

Thought for the day 12: Consider a plane surface, with a two-dimensional vector field defined on the surface; why, in some small region where the curl is not zero, is it not possible to describe the field in terms of the gradient of a scalar field? Hint: consider a geometrical construction that represents the situation. Does Escher's staircase explain the difficulty?

We shall see that the curl of a magnetostatic field is zero except where there are currents. Thus for current-free, or equivalently source-free, regions

$$\nabla \times \mathbf{B}(\mathbf{r}) = \mathbf{0}, \quad (4.46)$$

and we can derive the magnetic field from some scalar field.

This approach is reasonable because usually we are interested in fields in regions that are away from the currents that generate them. This restriction is not too different from the case of electrostatic potentials. Care is needed, however, because in the case of electrostatic fields, we introduced the potential function in terms of the

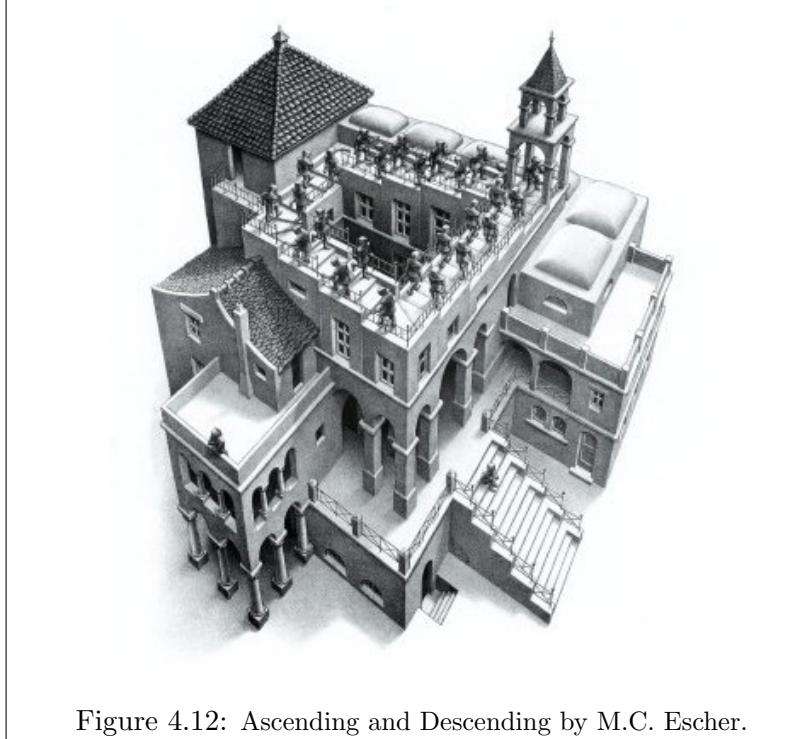


Figure 4.12: Ascending and Descending by M.C. Escher.

work that must be done to bring up charge from infinity. Here, we have introduced the magnetic scalar potential solely as the shape of the landscape that generates the magnetic field through its gradient; we have not attributed any particular physical meaning to the magnetic scalar potential.

4.12 Magnetic scalar potential of a current loop

Can we derive the functional form of the magnetic scalar potential associated with a current loop? After all, if we cannot calculate potential functions easily, they are not going to be much use.

First, it is convenient to introduce a mathematical tool relating to solid angles.

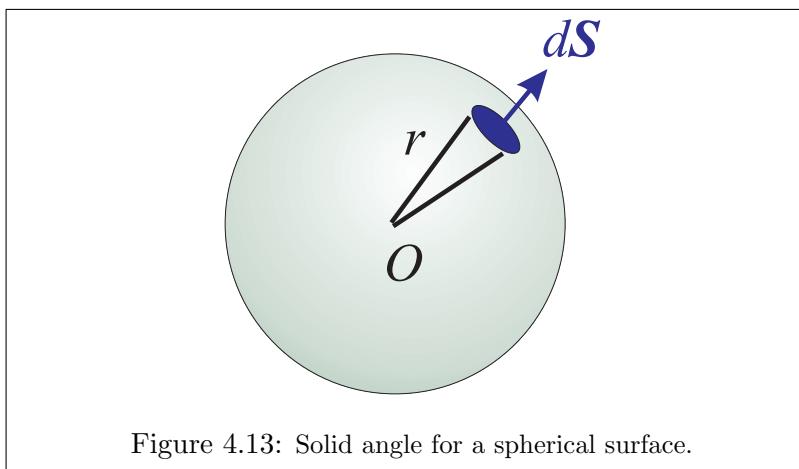


Figure 4.13: Solid angle for a spherical surface.

Solid angle $d\Omega$ subtended by an element of surface area $d\mathbf{S}$ on a sphere of radius r , centred on O (see Fig. 4.13), is defined by

$$d\Omega = \frac{|d\mathbf{S}|}{r^2}. \quad (4.47)$$

The solid angle is measured in **steradians**, and a whole sphere subtends 4π steradians. More generally, if the centre of the coordinate system is not at the centre of the sphere, or if the surface is not spherical, then the solid angle subtended by some elemental area on the surface is given by

$$d\Omega = \frac{|d\mathbf{S}|}{r^2} \cos \theta = \frac{d\mathbf{S} \cdot \mathbf{r}}{r^3}, \quad (4.48)$$

where θ is the angle shown in Fig. 4.14.

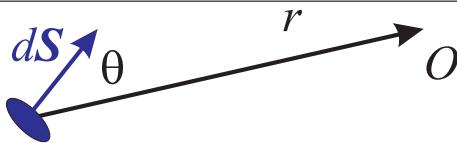


Figure 4.14: Solid angle for a non-spherical surface.

Note the sign convention: the surface normal would usually point *outwards* from a closed surface. For solid angles, the “radius vector” \mathbf{r} to the centre of the coordinate system points *away* from the surface, rather than towards it, and the surface normal points *inwards* also.

Now consider how to calculate the magnetic scalar potential associated with a current loop $I d\mathbf{S}$. The associated infinitesimal dipole moment is

$$d\mathbf{m} = I d\mathbf{S}. \quad (4.49)$$

The Biot-Savart law shows that the magnetic field has the same dipole form as the electric field from an electric dipole, as shown in Figs. 4.6 and 4.7. [For simplicity, we only proved this on the axis of the loop.] In each case the gradient of the scalar potential generates a dipole field, so we can see that the magnetic scalar potential for a small current loop with dipole moment \mathbf{m} should have the same functional form as the electric potential for an electric dipole \mathbf{p} .

Thus, to match the expression for an electric dipole,

$$\phi_m(r, \theta) = \frac{|d\mathbf{m}| \cos \theta}{4\pi r^2}. \quad (4.50)$$

It follows that

$$\begin{aligned} \phi_m &= \frac{d\mathbf{m} \cdot \mathbf{r}}{4\pi r^3} \\ &= \frac{I d\mathbf{S} \cdot \mathbf{r}}{4\pi r^3} \\ &= \frac{I d\Omega}{4\pi}. \end{aligned} \quad (4.51)$$

By inspection, it can be seen that $d\Omega$ is the solid angle of the current loop at \mathbf{r} , the point of observation. This same result must apply for larger loops, as can be appreciated by summing the contributions from a collection of small loops.

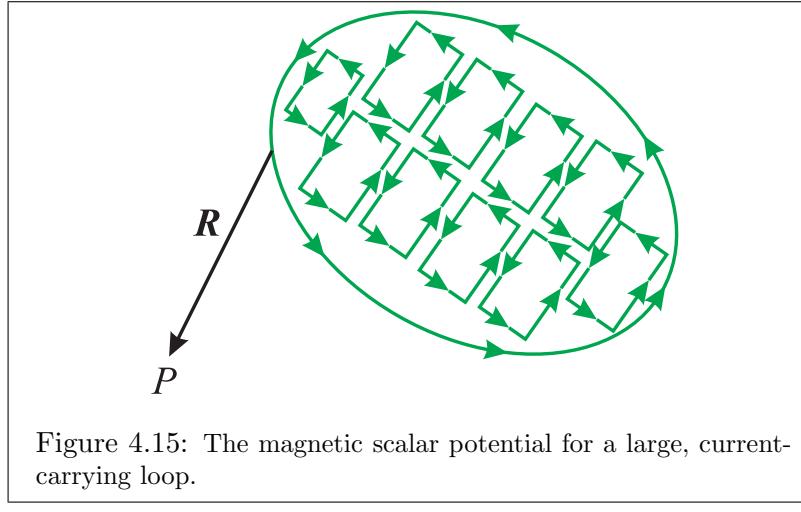


Figure 4.15: The magnetic scalar potential for a large, current-carrying loop.

We conclude that, for a macroscopic loop, the magnetic scalar potential at some observation point is proportional to the solid angle subtended by the loop when viewed from the observation point:

$$\phi_m = \frac{I\Omega}{4\pi}. \quad \text{Magnetic scalar potential for a macroscopic loop} \quad (4.52)$$

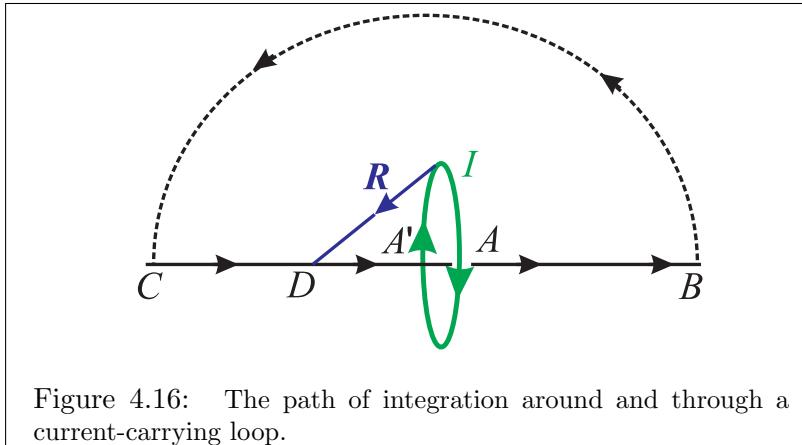
4.13 Ampère's law

We shall now uncover one of the most important laws in electromagnetism, namely Ampère's law:

$$\oint d\mathbf{l} \cdot \mathbf{B} = \mu_0 I \quad \text{Ampère's law.} \quad (4.53)$$

Consider the arrangement shown in Fig. 4.16, and start by looking at the solid angle subtended by a current loop as a function of position around a closed path that intersects the loop. Notice that the direction of the associated dipole moment is $C \rightarrow B$.

- Just to the right of the loop, at A, on the axis, $\Omega = +2\pi$ (half of the full solid angle).
- Moving outwards towards B (near ∞), Ω tends to zero.
- Travelling around the upper semicircle to C (near $-\infty$), Ω continues to be approximately zero.
- At D (intermediate distance), Ω takes on some negative value, because of the relative directions of the dipole moment and the position vector.



- Just to the left of the loop, at A' , $\Omega = -2\pi$. Notice that $\Omega_{A'} \neq \Omega_A$.

We therefore have

$$\phi_m(A') = -\frac{I}{2}, \quad (4.54)$$

and

$$\phi_m(A) = \frac{I}{2}. \quad (4.55)$$

It can be seen that at the centre of the loop, the magnetic scalar potential is not single-valued: if we approach it from the right, we get a positive quantity, whereas if we approach it from the left, we get a negative quantity. In fact

$$\int_A^{A'} d\phi_m = -I, \quad (4.56)$$

which can be written using the definition of ϕ_m in Eq. (4.44) as

$$\int_A^{A'} d\phi_m = \int_A^{A'} dl \cdot \nabla \phi_m \stackrel{(4.44)}{=} -\frac{1}{\mu_0} \oint dl \cdot \mathbf{B}. \quad (4.57)$$

From this we obtain **Ampère's law**:

$$\oint dl \cdot \mathbf{B} = \mu_0 I. \quad \text{Ampère's law} \quad (4.58)$$

The line integral of \mathbf{B} around any closed path is equal to μ_0 times the current flowing through the path.

We can also write Ampère's law more succinctly for the magnetic field strength

$$\oint dl \cdot \mathbf{H} = I, \quad (4.59)$$



André-Marie Ampère
(1775–1836)

Ampère's law can also be expressed in terms of the **electrical current density**:

$$\oint_{\partial S} dl \cdot \mathbf{H} = \int dS \cdot \mathbf{J}, \quad \begin{array}{l} \text{Ampère's law} \\ \text{for current density } \mathbf{J} \end{array} \quad (4.60)$$

where we have used

$$I = \int dS \cdot \mathbf{J}. \quad (4.61)$$

It is now possible to use Stokes' theorem to express Ampère's law in terms of vector differential operators. We have

$$\oint_{\partial S} dl \cdot \mathbf{H} = \int dS \cdot \nabla \times \mathbf{H} = \int dS \cdot \mathbf{J}, \quad (4.62)$$

and since this must be true for all closed loops, we arrive at **Maxwell's equation 4**:

$$\nabla \times \mathbf{H} = \mathbf{J} \quad \text{Maxwell's equation 4 (simplified).} \quad (4.63)$$

Although we have not proven Ampère's law—in fact we have only considered one geometry—we have uncovered a law that is fundamentally true for all closed paths in space. The applications are considerable.

4.14 The magnetic field of a long wire

Let us look at an example of how Ampère's law can be used. Consider the magnetic field associated with a long straight wire carrying constant current, I .

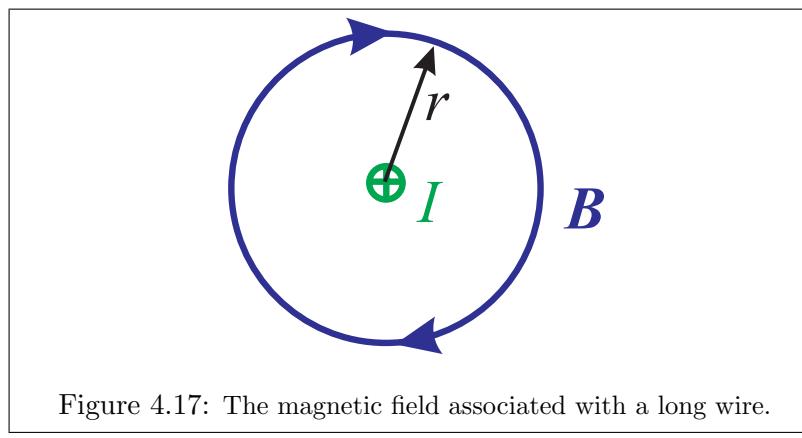


Figure 4.17: The magnetic field associated with a long wire.

To find the magnetic field \mathbf{B} at a distance r , we can simply apply Ampère's law on a circular path that has the wire at its centre:

$$\oint dl \cdot \mathbf{B} = 2\pi r |\mathbf{B}(r)| = \mu_0 I, \quad (4.64)$$

where we have used the fact that, by symmetry, $|\mathbf{B}|$ is only a function of r , and the field always points in a direction tangential to the path of integration, i.e., $|\mathbf{B}| = B_\phi$.

We conclude that the **magnetic field for an infinitely long wire** is

$$B(r) = \frac{\mu_0 I}{2\pi r}, \quad (4.65)$$

which has been trivial to derive. This simple example shows the utility of Ampère's law.

4.15 The force between two parallel wires

Consider the force between two infinite, parallel wires.

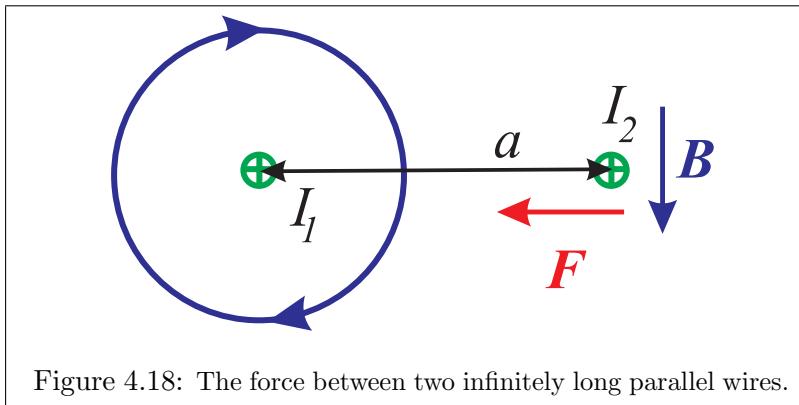


Figure 4.18: The force between two infinitely long parallel wires.

We know that the total force acting on a rigid wire carrying current I_2 is given by

$$\mathbf{F} = \int I_2 d\mathbf{l} \times \mathbf{B}, \quad (4.66)$$

where we have simply summed the contributions from individual elements.

We now know, however, that the field associated with a long wire is given by

$$B(r) = \frac{\mu_0 I_1}{2\pi a}, \quad (4.67)$$

and therefore the **force between two infinite, parallel wires per unit length** becomes

$$F = \frac{\mu_0 I_1 I_2}{2\pi a}. \quad (4.68)$$

Notice that when the currents flow in the same direction, the force is attractive.

This same result could have been derived using the Biot-Savart law, but the use of Ampère's law is particularly efficient. Generally, Ampère's law is particularly useful when the system has some symmetry that allows the direction of the field lines to be known prior to analysis.

4.16 Magnetic field of a long solenoid

In the case of a long solenoid, we can use Ampère's law, taking the path of integration along a line inside the solenoid, radially outwards to a point outside the solenoid, along the surface of the outside, and then radially inwards back to the starting point.

To apply Ampère's law, we need to identify and exploit the symmetry properties of the setup. Because the solenoid is infinitely long, the field along the outside is negligibly small, the field tangential to the radial parts of the path is zero, and we are only left with the field inside the solenoid.

The total current passing normally through the surface defined by the path of integration is nI , where n is the number of turns per unit length, and therefore

$$B = \mu_0 n I. \quad (4.69)$$

Ampère's law also shows that the field is constant over the whole of the cross-section of the solenoid. Again, we could have derived this same result using the Biot-Savart law, but Ampère's law is much more efficient.

4.17 The magnetic scalar potential revisited

Now consider the magnetic scalar potential associated with a long wire.

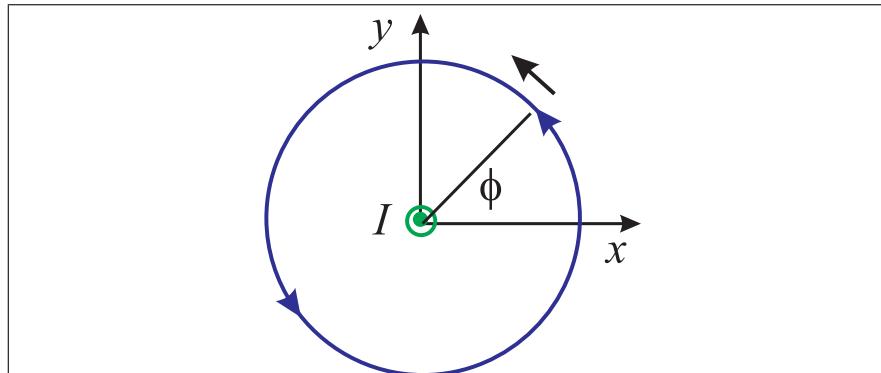


Figure 4.19: The magnetic scalar potential associated with a long wire.

We know from (4.65) that for a long wire

$$\mathbf{B} = \frac{\mu_0 I}{2\pi r} \hat{\phi}, \quad (4.70)$$

but we wish to describe the field in terms of the gradient of a potential, ϕ_m :

$$\mathbf{B} = -\mu_0 \nabla \phi_m. \quad (4.71)$$

In cylindrical polar coordinates, the $\hat{\phi}$ component of the gradient of a scalar field is given by

$$(\nabla)_\phi = \frac{1}{r} \frac{\partial}{\partial \phi}, \quad (4.72)$$

and therefore

$$\phi_m = -\frac{I\phi}{2\pi}. \quad (4.73)$$

This result shows that the magnetic scalar potential changes linearly with angle as we move around a current-carrying wire. The geometrical interpretation is straightforward, because the magnetic scalar potential can be regarded as a potential surface, and the field lines must always point downhill at any point on the surface. It is also obvious that the scalar potential must be multi-valued, e.g., along $\phi = 0$ which is the same line as $\phi = 2\pi$. This reveals a hazard of trying to describe magnetic fields in terms of potentials.

The case of parallel current-carrying wires is similar.

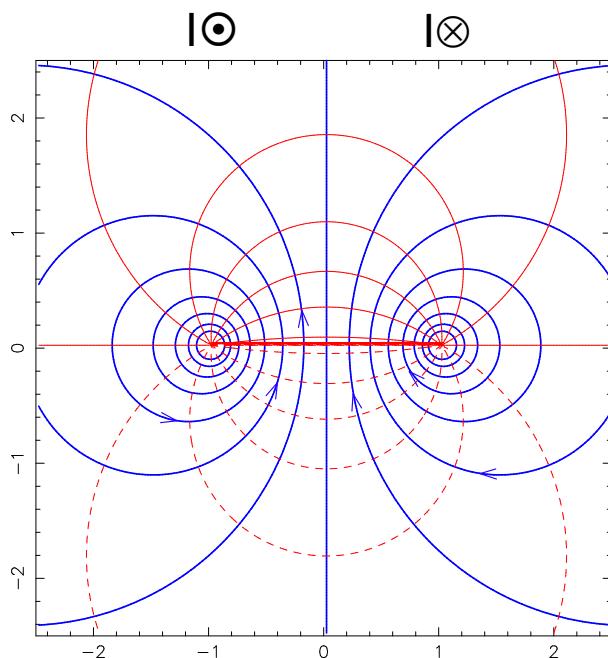
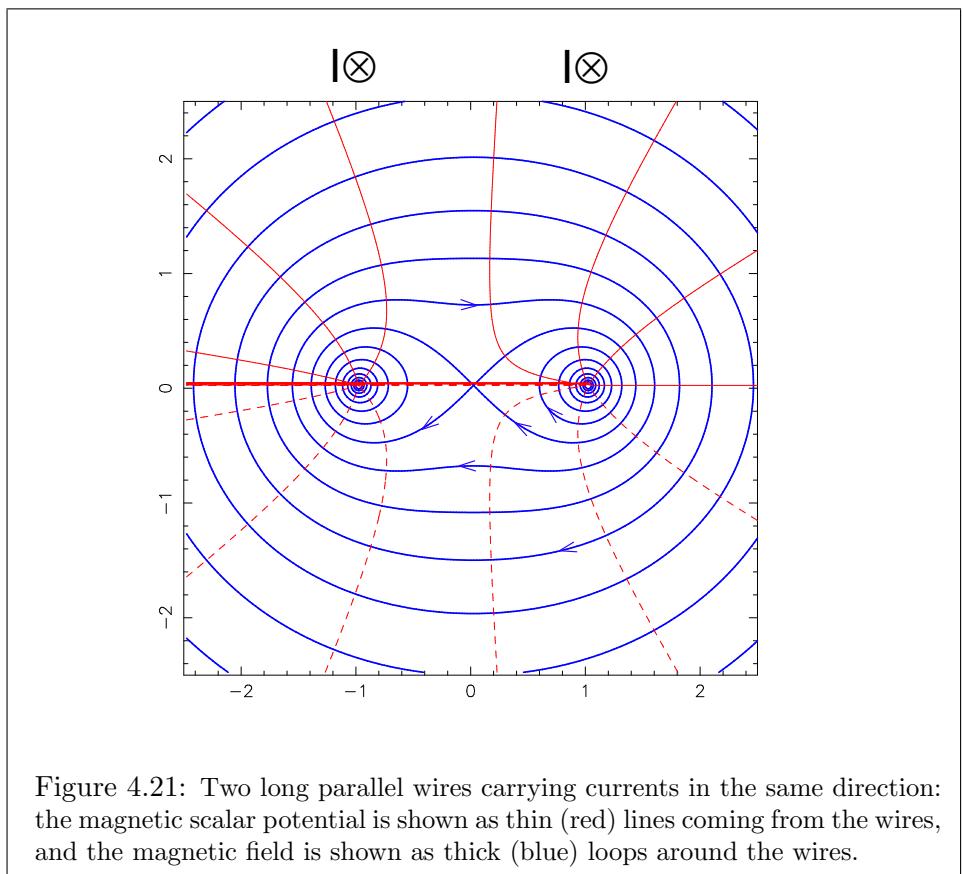


Figure 4.20: Two long parallel wires carrying currents in opposite directions: the magnetic scalar potential is shown as thin (red) lines coming from the wires, and the magnetic field is shown as thick (blue) loops around the wires.

In the case where the wires carry currents in opposite directions, shown in Fig. 4.20, there is a discontinuity along the line joining the two wires. In the case where the wires carry currents in the same direction, shown in Fig. 4.21, there is a discontinuity along the x -axis to the left of the left hand side wire and the line joining the right hand side wire and the middle point between the two wires. Thus a discontinuity only occurs along a closed path when



it encloses a current. This is entirely consistent with the notion of a potential landscape, and whether the potential, which defines a vector field, can return to the same value after having moved around a closed path.

4.18 Magnetic vector potential

Previously, we introduced the notion of a magnetic scalar potential, by requiring that the magnetic field can be derived from the gradient of a scalar field. This approach followed our work on the electric potential. In both cases, because the gradient of a constant is zero, the potential is undefined to within a constant; in other words the zero point can be chosen anywhere, because it is only the gradient that is of interest. One complication arises because the magnetic potential ϕ_m is multi-valued if there are currents.

We also know, however, that the curl of a vector field is a vector field, and therefore we might wonder if we can define some vector field such that its curl gives the magnetic field of interest:

$$\mathbf{B}(\mathbf{r}) = \nabla \times \mathbf{A}(\mathbf{r}). \quad \text{Magnetic vector potential} \quad (4.74)$$

$\mathbf{A}(\mathbf{r})$ is called the **magnetic vector potential**, and it turns out to be physically more important than the magnetic scalar potential. If we take the divergence of $\mathbf{B}(\mathbf{r})$, defined in this way, then because $\nabla \cdot (\nabla \times \mathbf{A}) = 0$ for any vector field \mathbf{A} , then $\nabla \cdot \mathbf{B}(\mathbf{r}) = 0$, which is always true for magnetic fields. Thus the magnetic vector potential does seem to be a reasonable concept.

There is, however, an ambiguity when choosing a vector field, $\mathbf{A}(\mathbf{r})$, whose curl is equal to some predefined vector field, $\mathbf{B}(\mathbf{r})$. Imagine that we choose some $\mathbf{A}(\mathbf{r})$ to represent $\mathbf{B}(\mathbf{r})$, then

$$\nabla \times \mathbf{A} = \begin{vmatrix} \hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ A_x & A_y & A_z \end{vmatrix}, \quad (4.75)$$

but we can add to $\mathbf{A}(\mathbf{r})$ any vector, $\mathbf{K}(\mathbf{r})$, having the form

$$\mathbf{K}(\mathbf{r}) = \hat{\mathbf{x}} K_x(x) + \hat{\mathbf{y}} K_y(y) + \hat{\mathbf{z}} K_z(z), \quad (4.76)$$

without changing the result, which can be seen by inspection. The magnetic vector potential is therefore undefined to within the addition of a vector whose x -component is a function of x only, whose y component is a function of y only, and whose z component is a function of z only.

The divergence of the modified vector potential then becomes

$$\begin{aligned} \nabla \cdot (\mathbf{A} + \mathbf{K}) &= \frac{\partial A_x}{\partial x} + \frac{\partial A_y}{\partial y} + \frac{\partial A_z}{\partial z} \\ &+ \frac{\partial K_x}{\partial x} + \frac{\partial K_y}{\partial y} + \frac{\partial K_z}{\partial z}, \end{aligned} \quad (4.77)$$

which *does* depend on \mathbf{K} . In fact we can change the divergence to anything we like while still retaining the same curl. Thus the requirement that the curl of $\mathbf{A}(\mathbf{r})$ equals $\mathbf{B}(\mathbf{r})$ does not constrain the divergence of $\mathbf{A}(\mathbf{r})$.

When representing a vector field through the curl of some other vector field, the process of setting the divergence to some chosen value is called **choosing the gauge**. In the case of the magnetic vector potential, one common choice is

$$\nabla \cdot \mathbf{A}(\mathbf{r}) = 0 \quad (4.78)$$

everywhere. This is chosen for convenience but many other choices are possible.

We might now wonder about the relationship between current sources and the magnetic vector potential. The analysis is straightforward:

$$\begin{aligned} \mu_0 \mathbf{J} &= \nabla \times \mathbf{B} \\ &= \nabla \times (\nabla \times \mathbf{A}) \\ &= \nabla(\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A} \\ &= -\nabla^2 \mathbf{A}. \end{aligned} \quad (4.79)$$

In conclusion, we found the **Poisson equation for the magnetic vector potential**:

$-\nabla^2 \mathbf{A} = \mu_0 \mathbf{J},$

Poisson equation for magnetic vector potential \mathbf{A}

(4.80)

which can be written out in terms of the components:

$$\begin{aligned} -\nabla^2 A_x &= \mu_0 J_x \\ -\nabla^2 A_y &= \mu_0 J_y \\ -\nabla^2 A_z &= \mu_0 J_z. \end{aligned} \quad (4.81)$$

Equations (4.81) compare directly with the electrostatic case

$$-\nabla^2 V = \frac{\rho}{\epsilon_0}, \quad (4.82)$$

and in fact Poisson's equation can be used for finding the components of the vector potential for any given current distribution. Also, by comparing with the electrostatic case given in Eq. (2.98) of Section 2.19, we can write down the magnetic vector potential $\mathbf{A}(\mathbf{r})$ directly by simply applying Green's function from Eq. (2.99) to $\mathbf{J}(\mathbf{r})$:

$\mathbf{A}(\mathbf{r}) = \frac{\mu_0}{4\pi} \int \frac{\mathbf{J}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d^3 r',$

**Magnetic vector potential
from the current distribution**

(4.83)

which calculates the magnetic vector potential at a point by summing over all of the current sources that contribute to the potential.

The similarity between the functional form of the components of the magnetic vector potential and current, and the functional form of the electric scalar potential and charge, is of fundamental physical importance. In special relativity the components of the magnetic vector potential and the electric scalar potential are combined into a single object called a **4-vector**. Indeed it explains the close relationship between electric and magnetic fields: static charge is the source of electric field, whereas moving charge is the source of magnetic field. Whether you see an electric field or a magnetic field just depends your frame of reference. An observer moving at a constant velocity sees something different from what a stationary observer sees. Likewise, ρ and \mathbf{J} are related: when ρ changes, we get a current \mathbf{J} . (ρ, \mathbf{J}) also forms a relativistic 4-vector.

4.19 Maxwell's equations for magnetostatic fields

In this part of the course, we have derived two of the most important equations in magnetostatics:

$$\nabla \cdot \mathbf{B} = 0, \quad \text{Maxwell's equation 2} \quad (4.84)$$

and

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J}. \quad \text{Maxwell's equation 4 in electrostatics} \quad (4.85)$$

These are two of Maxwell's equations. It is important to remember these, as they form the basis for solving essentially all problems in time-invariant magnetic systems.

Calculating the current from the magnetic field is straightforward, but usually we wish to calculate the magnetic field from the current. We have seen that the magnetic field associated with some current distribution can be found indirectly by solving

$$-\nabla^2 \mathbf{A} = \mu_0 \mathbf{J}, \quad (4.86)$$

for the magnetic vector potential, and then calculating the field through

$$\mathbf{B} = \nabla \times \mathbf{A}. \quad (4.87)$$

However, for simple geometries, the integral laws of Biot-Savart and Ampère are far more convenient, or the magnetic scalar potential may be useful.

4.20 Resistance and Conductivity — Reminder

Ohm's law, $V = IR$, is very familiar, but it can also be expressed in terms of the electric field \mathbf{E} and the current density \mathbf{J} , as shown below. A resistor of resistance R obeys Ohm's law:

$$R = V/I \quad (4.88)$$

where V is the “electromotive force” (e.m.f.), and R is the “resistance”.

Consider a volume of material, with faces of area S a distance ℓ apart, as shown in Fig. 4.22. Applying V across the faces of the

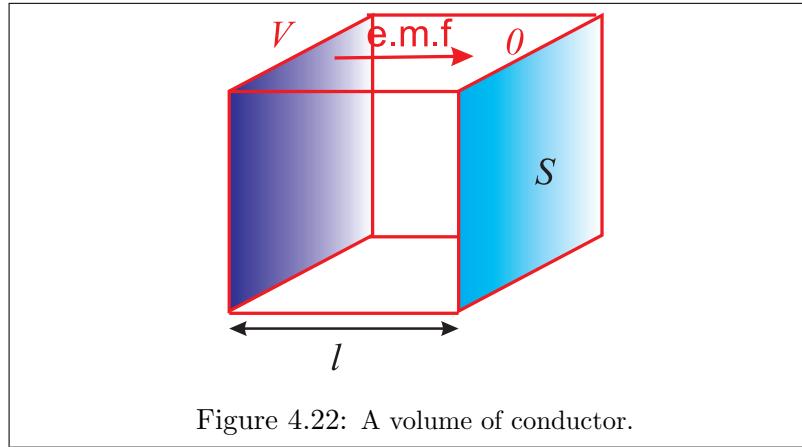


Figure 4.22: A volume of conductor.

material, we expect the following dependencies:

- Current $\propto V$;
- Current $\propto S$;
- Current $\propto 1/\ell$.

In combination:

$$I = \frac{VS\sigma}{\ell}; \quad \Rightarrow R = \frac{\ell}{S\sigma} \quad (4.89)$$

The proportionality constant σ in the above equation is the *conductivity* of the material ($\sigma = 1/\rho$ [$\Omega^{-1}\text{m}^{-1}$], where ρ is the resistivity).

Using $V/\ell = |\mathbf{E}|$, and $J = I/S$ in the direction of \mathbf{E} , find the constitutive relation for **Ohm's law**:

$$\mathbf{J} = \sigma \mathbf{E} \quad \text{Ohm's law.}$$

(4.90)

5. Magnetostatic fields in magnetic materials

5.1 Introduction

To this point we have not mentioned magnetic materials, and yet the behaviour of a magnetic system will clearly change if a magnetisable material is introduced. As in the case of electrically polarisable materials, where induced electric dipoles change the behaviour of electric field lines, magnetically polarisable materials change the behaviour of magnetic field lines. There are, essentially, three different kinds of magnetic material:

- diamagnetic
- paramagnetic
- ferromagnetic

Their different properties stem from the precise nature of the magnetic dipole moments of the atoms and molecules that make up the materials.

There are two main contributions to the magnetic dipole moment of an atom, or molecule: (i) the orbital motions of electrons correspond to current loops and thus form dipoles; (ii) the electrons themselves have intrinsic magnetic dipole moments arising from their quantum-mechanical spin. The sum of these two contributions determines the overall magnetic properties of most materials:

- Closed current loops react **diamagnetically**, which means that they oppose any applied external magnetic field, where ‘oppose’ means that the net field is reduced, as a consequence of the dipole field combining with the external field.
- Some materials react **paramagnetically**, which means that permanent dipole moments of the material align themselves with the external field, leading to an enhancement of the field.
- Some materials behave **ferromagnetically**, which means that they behave in a way that is similar to paramagnetic materials, but the effect is considerably enhanced.

Regardless of the precise physical origin of the magnetic properties of materials, we require a general way of describing the macroscopic behaviour of materials when they are exposed to an external magnetic field.

5.2 Magnetisation currents

The key concept is that magnetic materials acquire a net magnetic dipole moment when placed in a field.

Assume, for simplicity, that this net dipole moment occurs as a result of microscopic circulating currents that align when a field is applied. If all of the dipoles are arranged randomly, as shown in Fig. 5.1, there is no net circulating current, and there is no net dipole moment. If the dipoles are aligned, as in Fig. 5.2, there is a net circulating current, and a net dipole moment. How can we describe how these individual dipoles add together to create an overall magnetic field?

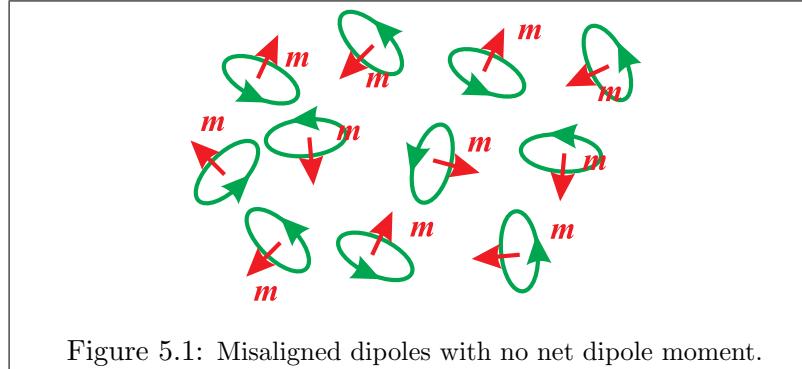


Figure 5.1: Misaligned dipoles with no net dipole moment.

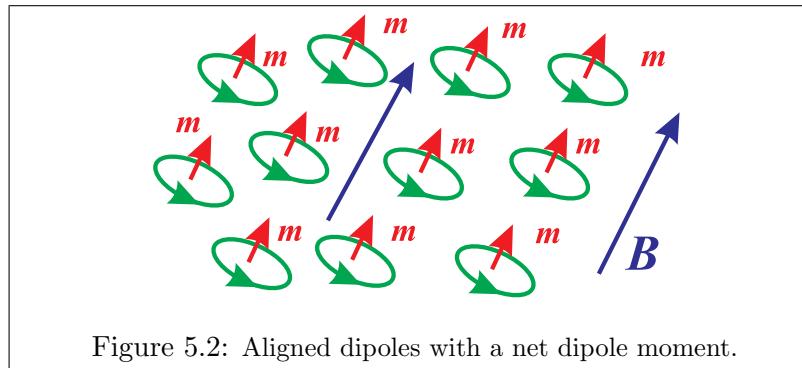


Figure 5.2: Aligned dipoles with a net dipole moment.

Consider the situation where four loops of wire are connected together with a common edge, which points in the direction of \hat{y} (see Fig. 5.3). The four loops all contribute to the current I_y shown.

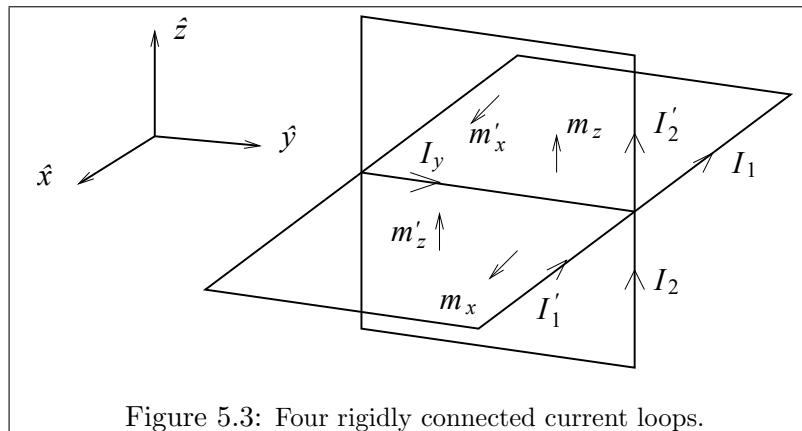


Figure 5.3: Four rigidly connected current loops.

The dipole moment due to I_1 is

$$m_z = I_1 dx dy, \quad (5.1)$$

and the dipole moment due to I'_1 is

$$m'_z = I'_1 dx dy. \quad (5.2)$$

We know, however, that

$$m'_z = m_z + \frac{\partial m_z}{\partial x} dx, \quad (5.3)$$

and therefore

$$\begin{aligned} m_z - m'_z &= -\frac{\partial m_z}{\partial x} dx \\ &= (I_1 - I'_1) dx dy \\ &= I_y^1 dx dy, \end{aligned} \quad (5.4)$$

where I_y^1 is the net current in the central path due to I_1 and I'_1 .

Hence we find for I_y^1

$$I_y^1 dy = -\frac{\partial m_z}{\partial x}. \quad (5.5)$$

Similarly, for the other pair of loops, the dipole moment due to I_2 is

$$m_x = I_2 dy dz, \quad (5.6)$$

and the dipole moment due to I'_2 is

$$m'_x = I'_2 dy dz. \quad (5.7)$$

We know that

$$m'_x = m_x + \frac{\partial m_x}{\partial z} dz, \quad (5.8)$$

and therefore

$$\begin{aligned} m_x - m'_x &= -\frac{\partial m_x}{\partial z} dz \\ &= (I_2 - I'_2) dy dz \\ &= -I_y^2 dy dz, \end{aligned} \quad (5.9)$$

where I_y^2 is the net current in the central path due to I_2 and I'_2 .

Hence we find for I_y^2 :

$$I_y^2 dy = \frac{\partial m_x}{\partial z}. \quad (5.10)$$

The **total current** on the central path due to all four loops I_y is

$$I_y = I_y^1 + I_y^2 = \frac{1}{dy} \left[\frac{\partial m_x}{\partial z} - \frac{\partial m_z}{\partial x} \right], \quad (5.11)$$

which we can convert to the current density

$$J_y = \frac{1}{dxdydz} \left[\frac{\partial m_x}{\partial z} - \frac{\partial m_z}{\partial x} \right]. \quad (5.12)$$

We introduce a new quantity called the **magnetisation**, denoted by \mathbf{M} , which is the magnetic dipole moment per unit volume:

$$\mathbf{M} = \mathbf{m} \frac{1}{dxdydz}. \quad (5.13)$$

The magnetic dipole moment per unit volume, \mathbf{M} , which is a vector field, is very similar to the electric dipole moment per unit volume \mathbf{P} , introduced previously in the context of electrostatic systems.

The total dipole moment of an object, say \mathbf{m}_{tot} , can be found by integrating the dipole moment per unit volume over the object:

$$\mathbf{m}_{\text{tot}} = \int d^3r \mathbf{M}. \quad (5.14)$$

In other words, if an object is broken down into infinitesimally small elements, the dipole moment of the complete system is found by adding together the contributions from the elements.

If the individual dipole moments point in random directions, the vector sum over the whole object leads to a zero net dipole moment. If they are aligned, even partially, a net dipole moment results.

Thought for the day 13: It may seem curious that dipole moments can be integrated in this way, but remember that they are intimately related to couples, which can be added vectorially. Consider two couples, one comprising two oppositely directed forces, of magnitude F_1 , separated by a distance dx , and the other comprising two oppositely directed forces, of magnitude F_2 , separated by a distance dx . Calculate the resulting couple when these two systems are connected rigidly end to end, and show that the result can be derived simply by adding the couples numerically. How does this lead to a dipole moment per unit length, and in three dimensions to a couple per unit volume?

Returning to the original problem, Eq. (5.12) can be written

$$J_y = \left[\frac{\partial M_x}{\partial z} - \frac{\partial M_z}{\partial x} \right]. \quad (5.15)$$

The same procedure can be carried out for two other sets of loops having common paths in the \hat{x} and \hat{y} directions, resulting in a total of three equations:

$$\begin{aligned} J_x &= \left[\frac{\partial M_z}{\partial y} - \frac{\partial M_y}{\partial z} \right] \\ J_y &= \left[\frac{\partial M_x}{\partial z} - \frac{\partial M_z}{\partial x} \right] \\ J_z &= \left[\frac{\partial M_y}{\partial x} - \frac{\partial M_x}{\partial y} \right], \end{aligned} \quad (5.16)$$

which is finally written

$$\mathbf{J}_m = \nabla \times \mathbf{M}. \quad \text{Magnetisation current density} \quad (5.17)$$

We have introduced a subscript on the current to show that it represents the intrinsic magnetic behaviour of the individual atoms and molecules. It must be distinguished from, and is in addition to, the field produced by the movement of free charge. In fact it can be a fictitious current introduced merely to represent a magnetic dipole moment that has physical origins other than current. Again, this model is analogous to the electrostatic case, where we distinguished between externally applied free charge, and the separation of bound (polarisation) charge, which was associated entirely with the materials used in the system.

It is clear that if the magnetisation is uniform, the curl, which is calculated through spatial derivatives, is zero, and the magnetisation current is zero. This is equivalent to the current on neighbouring loops summing to zero on the common path.

5.3 Surface magnetisation currents

If the magnetisation is uniform within some object, then the magnetisation currents must reside on the surface: conversely, for example, the current in the wires of a solenoid can be regarded as a ‘surface’ current that produces a uniform magnetic field.

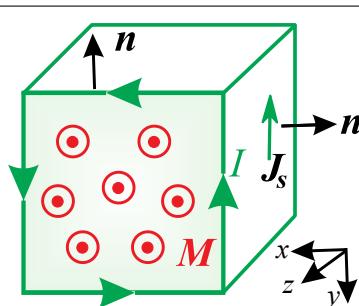


Figure 5.4: Sheet current density \mathbf{J}_s arising from the magnetisation \mathbf{M} .

To show this, consider a uniformly magnetised block (see Fig. 5.4). All the little current loops corresponding to the magnetic dipoles cancel each other out inside, leaving a sheet of current flowing around the surface. We call this a surface current density \mathbf{J}_s [Am^{-1}], like the current density but with one dimension shrunk to zero. So it is rather like surface charge σ flowing in the surface. The current flowing in the x -direction in a surface in the xz plane scales only with the extent in the z -direction.

Let the block have sides of length dx , dy and dz , and magnetisation M_z in the z -direction. Then the current I in the surface whose cross-section is shown is given by $I = |\mathbf{J}_s|dz$ and $M_z dx dy dz = I dx dy$. So $M_z = |\mathbf{J}_s|$, and \mathbf{J}_s is perpendicular to \mathbf{M} and to the unit normal to the surface \mathbf{n} .

The **surface current density** is therefore

$$\mathbf{J}_s = \mathbf{M} \times \mathbf{n} \quad \text{Surface current density} \quad (5.18)$$

As an aside reminder, the surface current density is introduced merely to represent a magnetic dipole moment that may have physical origins other than current.

Thought for the day 14: Consider the boundary between a uniformly magnetised material and free space. Use Stokes' theorem to show that a magnetisation current must flow on the surface as a consequence of the parallel component of the magnetisation changing from some constant value to zero as the surface is crossed. Can you prove that the magnetisation current has no normal component? The normal component of \mathbf{M} does not need a surface current.

5.4 Magnetic field strength

The total current in a material comprises the true conduction current, which is associated with the bulk movement of free charge, and the magnetisation current, which represents the intrinsic internal magnetic behaviour of the atoms and molecules.

Ampère's law can be invoked by simply taking into account this extra current:

$$\nabla \times \mathbf{B} = \mu_0 [\mathbf{J}_{\text{free}} + \mathbf{J}_m], \quad (5.19)$$

where we explicitly distinguish between free current \mathbf{J}_{free} and magnetisation current \mathbf{J}_m . Remember that \mathbf{B} simply gives the force on some small test current placed in the system, so it is perfectly reasonable, and necessary, to add on the magnetisation currents when calculating \mathbf{B} .

Simple manipulation using Eq. 5.17 gives

$$\nabla \times [\mathbf{B} - \mu_0 \mathbf{M}] = \mu_0 \mathbf{J}_{\text{free}}. \quad (5.20)$$

Define the **magnetic field strength** \mathbf{H} , according to

$$\mu_0 \mathbf{H} = \mathbf{B} - \mu_0 \mathbf{M}, \quad \text{Magnetic field strength in magnetic materials} \quad (5.21)$$

i.e.,

$$\mathbf{B} = \mu_0 (\mathbf{H} + \mathbf{M}), \quad \text{Magnetic field in magnetic materials} \quad (5.22)$$

such that

$$\nabla \times \mathbf{H} = \mathbf{J}_{\text{free}}. \quad (5.23)$$

The magnetic field strength has the advantage that it is related directly to the conduction current flowing at a point, regardless of the magnetic properties of the material in which the current is flowing. Of course if we wish to know the magnetic field \mathbf{B} , we would need to know \mathbf{M} .

Using Stokes' theorem, we have

$$\begin{aligned} \int d\mathbf{S} \cdot \nabla \times \mathbf{H}(\mathbf{r}) &= \oint_{\partial S} dl \cdot \mathbf{H}(\mathbf{r}) \\ &= \int d\mathbf{S} \cdot \mathbf{J}_{\text{free}}(\mathbf{r}) \\ &= I, \end{aligned} \quad (5.24)$$

where specific reference to position has been included for clarity. Therefore

$$\oint dl \cdot \mathbf{H}(\mathbf{r}) = I. \quad (5.25)$$

From now on we will always assume that $\mathbf{J}(\mathbf{r})$, without the subscript, corresponds to conduction (free) current.

Equations (5.23) and (5.25) constitute Ampère's law when magnetic materials are present. They are entirely consistent with our previous definition of Ampère's law, when no magnetisable material was present. Therefore, if we calculate the line integral of the magnetic field strength \mathbf{H} around any closed path, the result is equal to the current flowing through any surface that has the path as its bounding edge. It can be seen that the magnetic field strength is intimately related to the flow of conduction current. When written in this way, Ampère's law implicitly accounts for the properties of the materials making up the system.

For the purpose of calculating forces, we require \mathbf{B} , which in turn requires us to know \mathbf{M} , as appreciated by inspection of (5.21). In reality, as external current is applied and increased, \mathbf{H} increases, and \mathbf{M} , the dipole moment per unit volume, increases, as the state of magnetisation of the material changes.

For many materials, at least in the small-field limit, \mathbf{M} is linearly proportional to \mathbf{H} , and we define

$$\mathbf{M} = \chi_m \mathbf{H}, \quad (5.26)$$

where χ_m is a constant of proportionality called the **magnetic susceptibility**.

Thus, Eq. (5.21) becomes

$$\mu_0 \mathbf{H} = \mathbf{B} - \mu_0 \mathbf{M} \quad \Rightarrow \quad \mu_0 \mathbf{H} = \mathbf{B} - \mu_0 \chi_m \mathbf{H}, \quad (5.27)$$

which on rearranging produces

$$\mathbf{B} = \mu_0 \underbrace{(1 + \chi_m)}_{\equiv \mu} \mathbf{H}. \quad (5.28)$$

Now we can rewrite the relationship between the **magnetic field** and the **magnetic field strength**:

$$\boxed{\mathbf{B} = \mu_0 \mu \mathbf{H}}, \quad (5.29)$$

where the **relative permeability** μ of the material is given by

$$\boxed{\mu = 1 + \chi_m, \quad \text{Relative permeability}} \quad (5.30)$$

Note that μ is sometimes written as μ_r .

Also note that μ implicitly accounts for the change in the magnetic state of the material that results from a magnetic field \mathbf{H} being applied. It is common practice to use μ when describing the behaviour of magnetic fields, and χ_m when describing the detailed properties of magnetic materials.

For most insulators, which are non-magnetic, $\mu \approx 1$, whereas for magnetic materials, μ can take on very large values. The magnetic susceptibility can take on both positive and negative values, which was not true of the electric susceptibility. In fact, materials are defined in the following way:

- **Paramagnetic** materials have a positive susceptibility.
- **Diamagnetic** materials have a negative susceptibility.
- **Ferromagnetic** materials have an exceedingly high positive susceptibility.

Ferromagnetic materials are also often non-linear for relatively low field strengths, meaning that a $B - H$ curve needs to be used to describe the properties of a material, rather than just using a single constant of proportionality. This non-linearity gives rise to saturation in transformers, where the waveform of the current at the output does not follow the waveform of the current at the input. In fact, most magnetic materials are highly hysteretic, which means that a graph of B against H describes a loop as the applied magnetic field is increased, decreased, and reversed in sign.

A further complication is that certain materials may be associated with a magnetic field even in the absence of external excitation. In this case, the material is said to be **permanently magnetised**. The material has a constant dipole moment per unit volume, sometimes denoted \mathbf{M}_0 , which is not at all related to the applied field \mathbf{H} .

Finally, \mathbf{B} and \mathbf{H} do not have to lie in the same direction, in which case it is necessary to relate all of the components of \mathbf{B} to all of the components of \mathbf{H} through a matrix, or tensor. It is clear that the magnetic properties of materials can be quite complicated...

5.5 Summary of key points

It is useful to summarise certain key points pertaining to magnetic materials:

- The constant of proportionality between the magnetic dipole moment per unit volume and the magnetic field strength is called the magnetic susceptibility, where $\mathbf{M}(\mathbf{r}) = \chi_m \mathbf{H}(\mathbf{r})$.
The susceptibility can take on positive or negative values depending on the nature of the material (and it is not constant for ferromagnets).
- The relationship between the susceptibility and the relative permeability, μ , is $\mu = 1 + \chi_m$.
- For an isotropic, homogeneous material we can simply replace μ_0 with $\mu_0\mu$. Any reference to current refers to free current; magnetisation current is included indirectly through μ .
- Often the magnetisation is in the direction of the applied field, and linearly proportional to it, but neither of these needs to be true, leading to a wealth of physical phenomena.

5.6 Inhomogeneous magnetic materials and boundary conditions

Magnetic systems rarely have homogeneous magnetic properties, and therefore we are led to looking at situations where the magnetic properties change, possibly abruptly, at some boundary. As in the case of electrostatic problems, we must consider the magnetic field \mathbf{B} and the magnetic field strength \mathbf{H} separately.

Set up a ‘pillbox’ that cuts across the boundary between two different magnetic materials, as shown in Fig. 5.5. We know that,

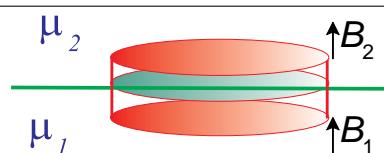


Figure 5.5: A pillbox on the boundary between dissimilar magnetic materials.

according to Maxwell’s equation 3 for the magnetic field \mathbf{B} ,

$$\oint_S d\mathbf{S} \cdot \mathbf{B}(\mathbf{r}) = 0. \quad (5.31)$$

Shrink the parallel faces of the pillbox down so that they are separated by an infinitesimally small distance. Also, make the parallel

faces small enough so that $\mathbf{B}(\mathbf{r})$ is essentially constant over the surfaces. If the normal component of \mathbf{B} in region 1 is called $\mathbf{B}_{1\perp}$, and the normal component of \mathbf{B} in region 2 is called $\mathbf{B}_{2\perp}$, then

$$A\mathbf{B}_{2\perp} - A\mathbf{B}_{1\perp} = \mathbf{0}, \quad (5.32)$$

where A is the area, from which it follows that

$$\boxed{\mathbf{B}_{2\perp} = \mathbf{B}_{1\perp}.} \quad (5.33)$$

The normal component of \mathbf{B} is continuous across a boundary, regardless of whether the relative permeability changes or not. The normal component of \mathbf{H} must change discontinuously across a boundary because $\mathbf{B} = \mu\mu_0\mathbf{H}$.

Now set up an imaginary loop that cuts across the boundary between two dissimilar magnetic materials, as shown in Fig. 5.6. Be-

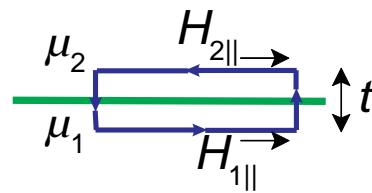


Figure 5.6: A loop on the boundary between dissimilar magnetic materials.

cause there is no conduction (free) current on the surface

$$\oint d\mathbf{l} \cdot \mathbf{H}(\mathbf{r}) = 0. \quad (5.34)$$

If the sides of the loop, having length L , are brought infinitely close together,

$$L\mathbf{H}_{2\parallel} - L\mathbf{H}_{1\parallel} = \mathbf{0}, \quad (5.35)$$

from which it follows that

$$\boxed{\mathbf{H}_{2\parallel} = \mathbf{H}_{1\parallel}.} \quad (5.36)$$

The tangential component of the magnetic field strength \mathbf{H} is continuous across a boundary, regardless of whether the relative permeability changes or not. The tangential component of \mathbf{B} must change discontinuously across a boundary because $\mathbf{B} = \mu\mu_0\mathbf{H}$.

In conclusion, across a magnetic boundary,

- The normal component of \mathbf{B} is continuous (\mathbf{B}_\perp continuous).
- The normal component of \mathbf{H} is discontinuous.
- The parallel component of \mathbf{H} is continuous (\mathbf{H}_\parallel continuous).
- The parallel component of \mathbf{B} is discontinuous.

These observations are comparable to the boundary conditions for electric fields.

5.7 Comparing the properties of magnetic fields and electric fields

It seems that magnetic fields have much in common with electric fields, and it is worthwhile comparing the similarities, and highlighting the differences:

- The magnetic field (magnetic flux density) \mathbf{B} is the fundamental quantity in magnetostatics, in the sense that it gives directly the force on a test current.

The electric field \mathbf{E} is the fundamental quantity in electrostatics because it gives the force on a test charge.

- The magnetic field strength \mathbf{H} is defined for the convenience of including magnetic materials, and it is related to \mathbf{B} through $\mathbf{B} = \mu\mu_0\mathbf{H}$.

The electric displacement (electric flux density) \mathbf{D} is defined for the purpose of including dielectric materials, and it is related to the electric field \mathbf{E} through $\mathbf{D} = \epsilon\epsilon_0\mathbf{E}$.

- In terms of the boundary conditions, \mathbf{B} and \mathbf{D} are alike, and \mathbf{H} and \mathbf{E} are alike.
- Because the normal components of \mathbf{B} and \mathbf{D} are continuous across a boundary, they can be imagined, for conceptual purposes, to have the properties of fluxes, which must be conserved.

A graphical overview of the behaviour of the fields at material boundaries is given in Fig. 5.7.

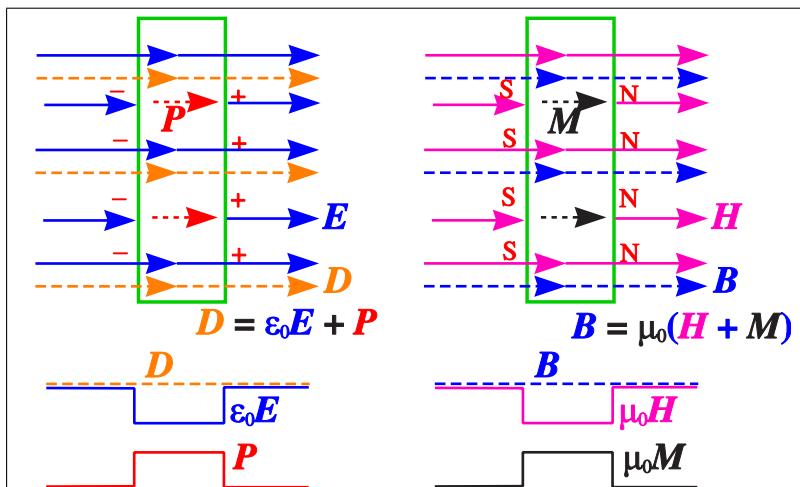


Figure 5.7: Comparison of the boundary conditions for magnetic and electric fields perpendicular to a slab.

In the case of electric fields, we saw that free charge is the source of \mathbf{D} , such that field lines associated with \mathbf{D} can only start and

end on free charge, whereas field lines associated with \mathbf{E} can start and end on either free charge or bound charge: Fig. 5.7. In the case of magnetic fields, the movement of free charge, in the sense of electrical current, is the source of \mathbf{H} , whereas both electrical current and magnetisation current are sources of \mathbf{B} . There seem to be magnetisation currents on the outer surface of a magnetic material to which the external magnetic field lines are pinned. We can also see how the direction of magnetic field lines changes at a magnetic boundary in the same way that the direction of electric field lines changes at a dielectric boundary.

5.8 Boundary-value problems with magnetic materials

How do we calculate the form of the magnetic field lines when electric currents and magnetic bodies are present? The modern approach would be to carry out numerical solutions on a computer, but systems with certain symmetries can be solved relatively easily. Few examples will be considered in the following, with increasing level of complexity.

- **Long thin rod parallel to uniform field**

Consider a long thin rod of magnetic material in a uniform field, as shown in Fig. 5.8. Because the tangential component

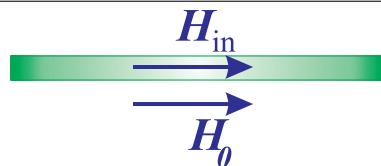


Figure 5.8: Long thin rod in a parallel magnetic field.

of \mathbf{H} must be continuous across the boundary, the internal magnetic field must be the same as the external field:

$$\mathbf{H}_{\text{in}} = \mathbf{H}_0, \quad (5.37)$$

from which it follows that

$$\mathbf{B}_{\text{in}} = \mu \mathbf{B}_0. \quad (5.38)$$

For paramagnetic and ferromagnetic materials ($\mu > 1$), the number of flux lines per unit area is larger on the inside than on the outside; the field is concentrated in the material.

For diamagnetic materials ($\mu < 1$), the number of flux lines per unit area is smaller on the inside than on the outside. Flux appears to be expelled.

- **Thin slab perpendicular to uniform field**

Consider a slab of magnetic material perpendicular to a uniform field (much thinner than its width so that field lines remain parallel), as shown in Fig. 5.9. Because the perpendic-

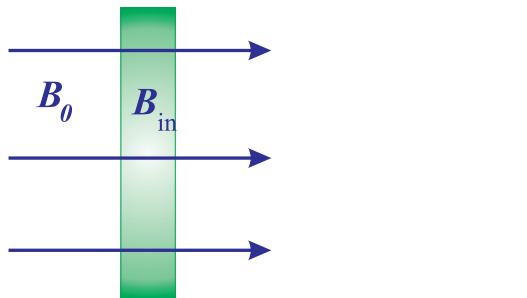


Figure 5.9: Thin slab perpendicular to a uniform magnetic field.

ular component of \mathbf{B} must be continuous across the boundary, the internal magnetic field must be the same as the external magnetic field:

$$\mathbf{B}_{\text{in}} = \mathbf{B}_0, \quad (5.39)$$

which is essentially the conservation of flux across the boundary. It follows that

$$\mathbf{H}_{\text{in}} = \frac{1}{\mu} \mathbf{H}_0. \quad (5.40)$$

For paramagnetic and ferromagnetic materials the magnetic field strength inside the material is less than that outside the material

For diamagnetic materials, the field strength internally is greater than that externally

- **Magnetisable sphere in a uniform field**

Consider a magnetisable (either diamagnetic or paramagnetic) sphere in a uniform external field, as shown in Fig. 5.10. This

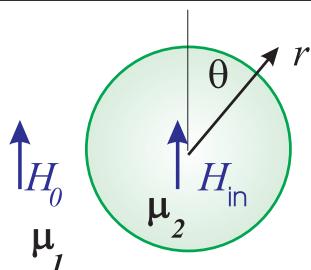


Figure 5.10: Magnetisable sphere in a uniform magnetic field.

problem can be solved in an almost identical way to that of the dielectric sphere. We shall use the magnetic scalar potential $\phi_m(\mathbf{r})$, which is related to the magnetic field through

$$\mathbf{H}(\mathbf{r}) = -\nabla \phi_m(\mathbf{r}).$$

Notice that because \mathbf{H} is finite at the boundary, the normal derivative of ϕ_m is finite, and therefore ϕ_m must be continuous.

As in the case of a dielectric sphere, let us assume that the internal field is uniform, and the external field is the externally applied uniform field plus a dipole field generated by the magnetisation current. The potential becomes

$$\begin{aligned}\phi_m(r) &= -H_{\text{in}}r \cos \theta && \text{for } r < a \\ \phi_m(r) &= -H_0r \cos \theta + \frac{A \cos \theta}{r^2} && \text{for } r > a,\end{aligned}\quad (5.41)$$

where A is some constant of proportionality whose value is to be found.

Because the potential is continuous across the boundary [or using H_{\parallel} continuous]

$$H_{\text{in}}a \cos \theta = H_0a \cos \theta - \frac{A \cos \theta}{a^2}, \quad (5.42)$$

or

$$H_{\text{in}} = H_0 - \frac{A}{a^3}. \quad (5.43)$$

We also require that the normal component of \mathbf{B} be continuous, where

$$B_{\perp} = B_r = -\mu\mu_0 \frac{\partial \phi_m}{\partial r}, \quad (5.44)$$

at $r = a$, which gives, using Eq. (5.41)

$$\mu\mu_0 H_{\text{in}} = \mu_0 \left(H_0 + \frac{2A}{a^3} \right). \quad (5.45)$$

Solving for A gives

$$A = \frac{\mu - 1}{\mu + 2} H_0 a^3, \quad (5.46)$$

from which it follows that

$$H_{\text{in}} = \frac{3}{\mu + 2} H_0. \quad (5.47)$$

This result is the same as that for the dielectric sphere with $\epsilon \rightarrow \mu$.

• Uniformly magnetised cylinder

Consider a cylinder that is uniformly magnetised along its length. In fact, this is a reasonable model for a bar magnet. In this case, magnetisation currents flow around the surface of the cylinder, in much the same way as the current in the wires of a solenoid. As a consequence, a cylindrical bar magnet has the same \mathbf{B} -field as a short solenoid (see Fig. 5.11).

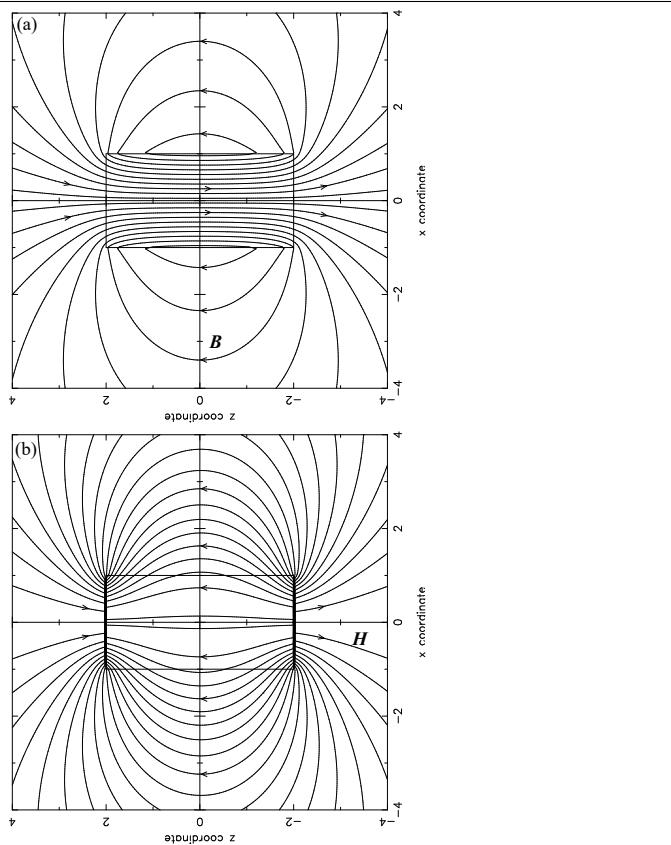


Figure 5.11: (a) The \mathbf{B} -field of a magnetised cylinder. (b) The \mathbf{H} -field of a magnetised cylinder. [Note that there are too few lines plotted to the left and right outside the magnet, where the density should increase as it does for \mathbf{B} .]

5.9 Electromagnets

Using the concepts described, we can model the behaviour of electromagnets.

Consider the arrangement where a coil of wire is wound around a toroidal core of magnetisable material. Assume initially that the toroid is continuous, without a gap. Assume that there are N turns

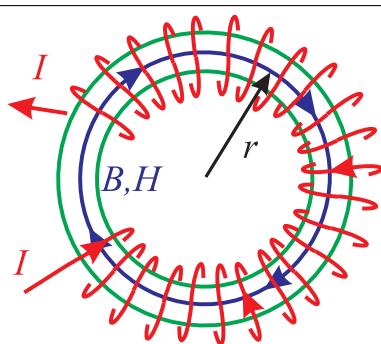


Figure 5.12: A coil of wire on a toroidal core of magnetisable material.

of wire in total, each carrying a current I . Also assume that the core is made of a material having a high relative permeability μ .

Because the system has a simple form, we can apply Ampère's law around a closed loop, inside the material, of radius r , as shown in Fig. 5.12. Ampère's law states

$$\oint dl \cdot \mathbf{H}(\mathbf{r}) = \int_S d\mathbf{S} \cdot \mathbf{J}(\mathbf{r}). \quad (5.48)$$

Notice that the current flowing perpendicular to the surface having the boundary defined by the path of integration is NI . The only thing that matters is the current flowing through the surface, as Ampère's law is always true. Also, the magnetic field always lies in a direction parallel to the path of integration, thus

$$\oint dl \cdot \mathbf{H}(\mathbf{r}) = 2\pi r H_{\text{in}} = NI, \quad (5.49)$$

from which we get

$$H_{\text{in}} = \frac{NI}{2\pi r}, \quad (5.50)$$

and

$$B_{\text{in}} = \frac{\mu\mu_0 NI}{2\pi r}. \quad (5.51)$$

In this analysis we used our intuition about the direction of the field for the purpose of solving the problem. Ampère's law is very powerful for simple geometries, and this technique should be remembered.

Now consider the case where a small air gap is opened up in the toroid, as shown in Fig. 5.13. The air gap, if short enough, can be considered to be a slab of magnetically dissimilar material that is perpendicular to the field. We know, therefore, that the normal

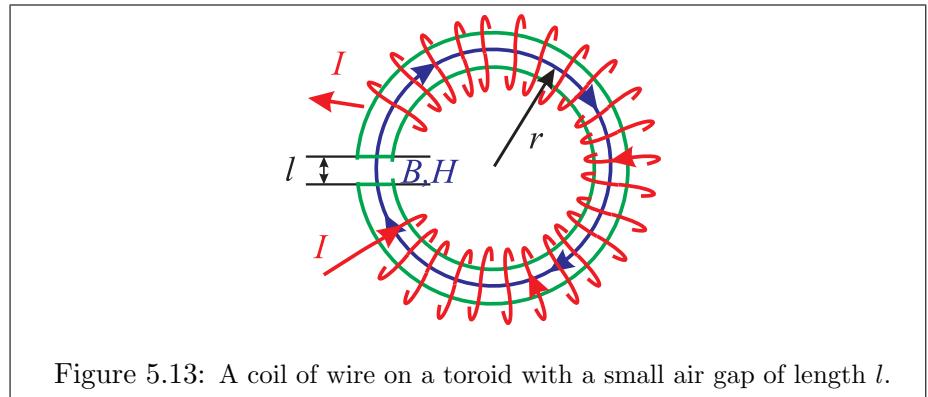


Figure 5.13: A coil of wire on a toroid with a small air gap of length l .

component of \mathbf{B} is continuous across the gap, giving

$$B_{\text{gap}} = B_{\text{in}}. \quad (5.52)$$

It follows that

$$\mu_0 H_{\text{gap}} = \mu_0 \mu H_{\text{in}}. \quad (5.53)$$

Ampère's law can again be applied to the whole loop:

$$\oint dl \cdot \mathbf{H}(\mathbf{r}) = (2\pi r - l)H_{\text{in}} + lH_{\text{gap}} = NI. \quad (5.54)$$

Of course, this assumes that the cross-sectional width of the toroid is much smaller than the radius of the loop, so that the field across the toroid can be considered uniform. Substituting and rearranging, we have

$$H_{\text{gap}} = \frac{\mu NI}{2\pi r + (\mu - 1)l}, \quad (5.55)$$

giving

$$B_{\text{gap}} = \mu_0 H_{\text{gap}} = \frac{\mu_0 \mu NI}{2\pi r + (\mu - 1)l}. \quad (5.56)$$

For typical electromagnets $\mu \gg 1$ and $\mu l \gg 2\pi r$, giving

$$B_{\text{gap}} = \frac{\mu_0 NI}{l}. \quad (5.57)$$

Consequently, we know the magnetic field in the gap.

Notice that in this system, with a highly magnetisable material, it is as if all of the ‘magnetic path’ is in the gap; in other words it is the gap that makes the biggest contribution to the line integral in Ampère’s law. This occurs because the flux density B is conserved around the loop, therefore,

$$H_{\text{gap}} = \mu H_{\text{in}}, \quad (5.58)$$

and so $H_{\text{gap}} \gg H_{\text{in}}$. The magnetic field in the toroid can, for the purpose of using Ampère’s law, be ignored in this limit.

Finally, it is also straightforward to calculate the magnetic flux, Φ , flowing across the gap, for which we have

$$\Phi = \int_S d\mathbf{S} \cdot \mathbf{B}(\mathbf{r}). \quad (5.59)$$

In the case of the toroid, the flux density is uniform, the field is everywhere perpendicular to the surface of interest, and we have

$$\Phi = \frac{A\mu_0 NI}{l}, \quad (5.60)$$

where A is the cross-sectional area.

It should be appreciated that in all calculations of this kind, the central issue is to have a clear understanding of where the field lines are pointing, the directions in which the currents are flowing, and how these relate to the surfaces and paths over which the various integrals are evaluated.

Electromagnetism is a highly ‘geometric’ subject!

6. Electromagnetic induction

6.1 Introduction

We have considered the behaviour of electrostatic and magnetostatic fields in some detail, and now we turn to the case where fields change with time. It would be naive to assume that it is only necessary to make all of the field quantities functions of time, and that the various laws remain the same—perhaps mother nature did not build the Universe in this way! What happens as we increase the rate of change? Do electric and magnetic systems continue to respond quasi-statically? We certainly need to worry about current because static charge gives rise to electric fields, whereas moving charge gives rise to magnetic fields, but what is the difference? Is not current simply moving charge?

6.2 Faraday's law

Michael Faraday observed that when the magnetic flux in a circuit changes, an electromotive force is induced, and the magnitude of the e.m.f. is proportional to the rate of change of flux.

Mathematically, **Faraday's law** is written

$$\mathcal{E} = -\frac{d}{dt} \int_S d\mathbf{S} \cdot \mathbf{B}(\mathbf{r}) = -\frac{d\Phi}{dt} \quad \text{Faraday's law} \quad (6.1)$$

$$i.e., \oint dl \cdot \mathbf{E}(\mathbf{r}) = -\frac{d}{dt} \int_S d\mathbf{S} \cdot \mathbf{B}(\mathbf{r}) \quad (6.2)$$

The line integral of the electric field around a closed loop is equal to the negative time rate of change of the magnetic flux that passes normally through any surface that has the loop as its boundary.

It is vitally important to appreciate that the e.m.f. is defined as the line integral of the electric field, which is not the same as voltage. In electromagnetic induction, e.m.f. can be defined around a closed loop as the electromagnetic work that would be transferred to a unit of charge if it travelled once around that loop. While the charge travels around the loop, it can simultaneously lose the energy via resistance into thermal energy.

The potential difference between two points is the *negative* of the line integral of the electric field, because work is done when moving charge against a force.

Faraday deduced his law after performing a number of experiments in 1831–2:



Michael Faraday (1791–1867)

- If the current in a coil changes, a current is induced in a neighbouring coil, because the amount of the first coil's magnetic flux that **links** (i.e., crosses) the second coil changes.
- If a coil moves *relative* to a source of flux (whatever its origin), such that the flux linked changes, then a current is induced in the coil.
- If part of a conducting circuit moves, and therefore cuts magnetic flux, then a current is induced in the circuit.
- The induced current is proportional to the conductance of the wire, showing that a changing flux gives rise to a definite **electromotive force** (“–voltage”), rather than a fixed current.

In 1832, Joseph Henry was the first to discover **self-inductance**, whereby a change in the current flowing in a coil of wire induces an e.m.f. across its own terminals. In fact, this is true for any circuit, including a short length of wire.

Thought for the day 15: In electric circuit theory, a constant-current source is a source of electrical energy that must have, under all circumstances, a fixed amount of current flowing through it. A constant-voltage source is a source of electrical energy that must have, under all circumstances, a fixed voltage across it. Why is it clear, even theoretically, that perfect current sources and perfect voltage sources cannot physically exist, even in principle? Real sources of electrical energy always have internal resistance. How does this overcome this fundamental problem? How does the internal resistance help us when we come to connect current sources in series, and voltage sources in parallel?

Faraday was vague about the direction of the induced e.m.f., and it was Emil Lenz in 1834 who first made a clear statement: *the induced e.m.f. is always in a such a direction as to promote a current flow that creates a magnetic field that opposes the change in flux.*

We can show that these observations lead to Faraday's law, namely that a change in the magnetic flux coupling into a circuit leads to a voltage, regardless of whether it is the magnetic field that is changing, or whether it is the circuit that is moving, or a combination of the two. Faraday's law can also be written in differential form, by assuming that the area over which the integral is defined is rigid. Then the time derivative can be calculated before the integral to give

$$\oint \mathrm{d}\mathbf{l} \cdot \mathbf{E}(\mathbf{r}) = - \int_S \mathrm{d}\mathbf{S} \cdot \frac{\partial \mathbf{B}(\mathbf{r})}{\partial t}. \quad (6.3)$$

Applying Stokes' theorem to the left-hand side gives

$$\oint \mathrm{d}\mathbf{l} \cdot \mathbf{E}(\mathbf{r}) = \int_S \mathrm{d}\mathbf{S} \cdot \nabla \times \mathbf{E}, \quad (6.4)$$

so

$$\int_S d\mathbf{S} \cdot \nabla \times \mathbf{E} = - \int_S d\mathbf{S} \cdot \frac{\partial \mathbf{B}(\mathbf{r})}{\partial t}, \quad (6.5)$$

but because this must be true for all surfaces,

$\nabla \times \mathbf{E} = - \frac{\partial \mathbf{B}(\mathbf{r})}{\partial t}.$

Maxwell's equation 2

(6.6)

This is one of the most important equations in electromagnetism, and forms one of Maxwell's equations. It is usually simpler to remember it as

$\nabla \times \mathbf{E} = - \dot{\mathbf{B}}.$

(6.7)

When the magnetic flux does not change with time,

$\nabla \times \mathbf{E} = 0,$

(6.8)

as we have already found for electrostatics

6.3 Faraday's law—an alternative perspective

Consider an elemental path in free space, or in a material of some kind, as shown in Fig. 6.1.

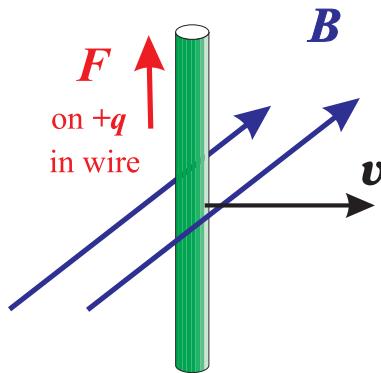


Figure 6.1: An elemental path in space, which could be a short length of wire.

Assume that the path is moving at velocity \mathbf{v} in a region where a static magnetic field is present. Any charge q on the path experiences a Lorentz force \mathbf{F} , where

$$\mathbf{F} = q(\mathbf{v} \times \mathbf{B}). \quad (6.9)$$

In the rest frame of the charge, this force appears to be due to an electric field

$$\mathbf{E} = \mathbf{F}/q = \mathbf{v} \times \mathbf{B} \quad (6.10)$$

The contribution $d\mathcal{E}$ to the e.m.f., for an element of length dl , is

$$d\mathcal{E} \equiv \mathbf{E} \cdot dl = (\mathbf{v} \times \mathbf{B}) \cdot dl \quad (6.11)$$

An electrical potential difference has been created merely by moving charge at a constant velocity through a magnetic field.

Now consider a complete loop moving through a constant, but not necessarily uniform, magnetic field at velocity \mathbf{v} , as shown in Fig. 6.2.

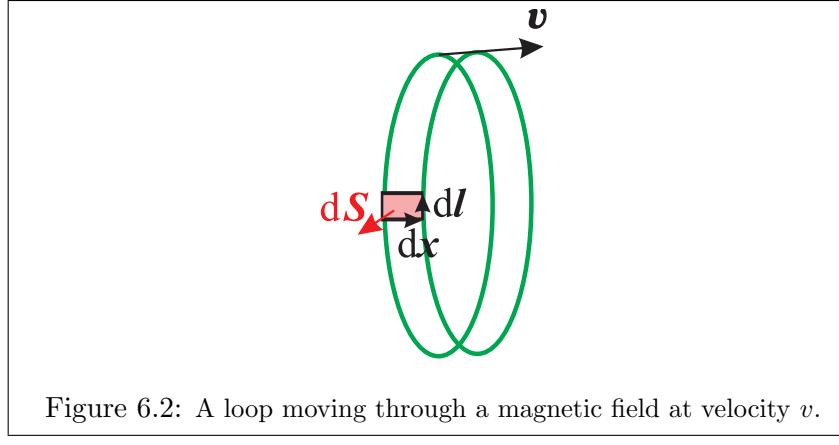


Figure 6.2: A loop moving through a magnetic field at velocity v .

The total e.m.f. around the circuit is given by

$$\mathcal{E} = \oint d\mathcal{E} = \oint (\mathbf{v} \times \mathbf{B}) \cdot d\mathbf{l} = - \oint (\mathbf{B} \times \mathbf{v}) \cdot d\mathbf{l} = - \oint \mathbf{B} \cdot (\mathbf{v} \times d\mathbf{l}), \quad (6.12)$$

where the order of \mathbf{B} and \mathbf{v} has been changed and the brackets in the scalar triple product have been swapped around.

Now, from Fig. 6.2 we see that

$$\mathbf{v} \times d\mathbf{l} = \frac{dx}{dt} \times d\mathbf{l} = \frac{d\mathbf{S}}{dt}, \quad (6.13)$$

and so

$$\mathcal{E} = - \int_{\text{strip}} \frac{d\mathbf{S}}{dt} \cdot \mathbf{B}, \quad (6.14)$$

where the integration is carried out over the surface of the strip swept out by the path as it moves.

Because the \mathbf{B} field is static, pull out d/dt out of the integral

$$\mathcal{E} = - \frac{1}{dt} \int_{\text{strip}} d\mathbf{S} \cdot \mathbf{B} = - \frac{d}{dt} \int_{\text{loop}} d\mathbf{S} \cdot \mathbf{B}, \quad (6.15)$$

since the integral over the loop after time dt is the old integral plus the integral over the strip around the edge.

Thus

$$\mathcal{E} = - \frac{d}{dt} \int_S d\mathbf{S} \cdot \mathbf{B} = - \frac{d\Phi}{dt}, \quad (6.16)$$

where Φ is the flux passing normally through the surface S of the loop itself. This is Faraday's law, as formulated in Eq. (6.1).

Above, Faraday's law was derived for the case of a rigid circuit moving in a static magnetic field. However, Faraday showed that it was

only the relative motion of the circuit and the source of magnetic field that mattered. Thus the law also applies to the case of a stationary circuit where the flux linked changes with time because the source is moving. Note, however, that we cannot prove the latter result using the derivation above! However, we cannot tell how a magnetic field was produced just by measuring it at a point. Thus Faraday's law should hold when the flux varies for whatever reason, such as a varying current in another coil, and not just when the source is moving. For this, let us consider a Gedankenexperiment: when measuring the potential difference between the terminals of a loop of wire, we cannot tell whether the linked flux is changing because the loop is moving and the magnetic field is constant, or because the loop is stationary and the field is changing.

In fact, Faraday's law is applicable when the field is static and the circuit is changing, or when the circuit is fixed, and the field is changing, or any combination of the two. This approach to Faraday's law again suggests a very close relationship between \mathbf{E} and \mathbf{B} fields, a relationship that is understood using Einstein's special theory of relativity.

6.4 Self-inductance

Often we talk about the flux *linked* into a circuit, because the flux may actually be constant, and it is the circuit that is moving. The **self-inductance** is the flux that is linked back into a circuit as a consequence of unit current flowing in the same circuit.

It is vitally important to be able to find the flux linked for a variety of cases, as this will give insight into what is meant by 'linked flux'. For example, in Fig. 6.3 the current in the loop produces a

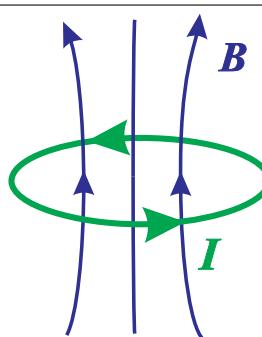


Figure 6.3: The flux linked into a current loop.

magnetic field, but the magnetic flux density \mathbf{B} can be integrated over the loop to yield the total flux, Φ , passing through the loop. This field passes normally through the loop and constitutes the flux that is linked back into the loop. In fact, if the current changes, the magnetic field changes, the flux linked changes, and a voltage is generated across the terminals of the loop. It is always beneficial

to think in this two-step way when solving problems of this kind: the current produces a magnetic field, and magnetic flux links back into the circuit.

The **self-inductance** is then defined as

$$L \equiv \frac{\Phi}{I}, \quad \text{Self-inductance} \quad (6.17)$$

where Φ is the linked flux. The (SI) unit of inductance is the henry, where $1\text{H} \equiv 1\text{Wb A}^{-1}$, and $1\text{ Wb} \equiv 1\text{ T m}^2$. The weber, written Wb , is the unit of magnetic flux. Clearly, different geometries will have different inductances.

6.5 Self-inductance of a long solenoid

Consider how to calculate the self-inductance of a long solenoid, of length l and cross-sectional area S , as shown in Fig. 6.4.

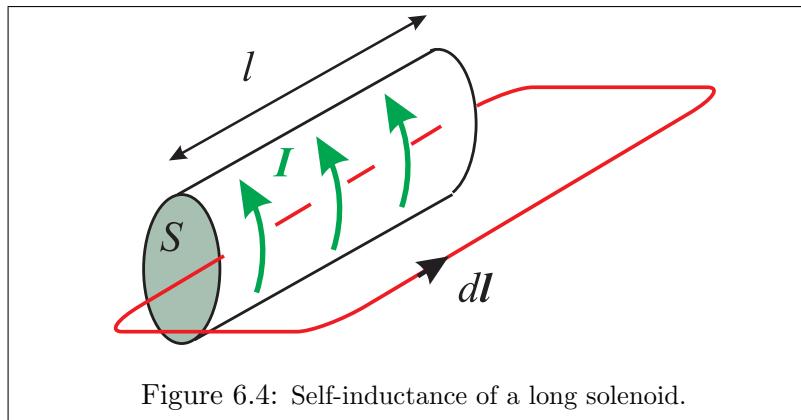


Figure 6.4: Self-inductance of a long solenoid.

Assume that the solenoid is long enough that end effects can be ignored. Let there be n turns per unit length, and assume that \mathbf{B} is zero outside the coil, which is a good approximation for a very long coil.

Inside the solenoid, $|\mathbf{B}| = B_{\text{in}}$. As we showed in section 4.16, if Ampère's law is applied to the path shown,

$$\oint dl \cdot \mathbf{H} = \int_S \mathbf{J} \cdot d\mathbf{S} = NI = nlI \quad (6.18)$$

$$= H_{\text{in}}l \quad (6.19)$$

and therefore

$$H_{\text{in}} = nI. \quad (6.20)$$

Notice that we can move the path of integration around inside the solenoid without changing the result, showing that H_{in} is uniform across the inside of the coil.

Assuming that the solenoid has an air core

$$B_{\text{in}} = \mu_0 n I. \quad (6.21)$$

Thus the magnetic flux density is proportional to the number of turns per unit length.

Care is now required in calculating the flux linked. The total flux passing through any *single* loop is

$$\Phi_{\text{single}} = S B_{\text{in}}, \quad (6.22)$$

because the flux is normal to *each* coil. We require, however, the flux linked back into the *complete* circuit, and therefore we must take into account the fact that the flux passes through each of the coils. The total flux linked becomes

$$\Phi = nlS B_{\text{in}}, \quad (6.23)$$

giving

$$\Phi = n^2 l S \mu_0 I. \quad (6.24)$$

Thus, the flux linked is proportional to the square of the number of turns, one factor of n comes from the fact that the more turns there are the higher the field, and the second comes from the fact that the more turns there are the higher the voltage generated for a given flux.

The **self-inductance** becomes

$L = \frac{\Phi}{I} = n^2 l S \mu_0,$

Self-inductance of long solenoid

(6.25)

which can also be used to give the self-inductance per unit length.

6.6 Self-inductance of coaxial cylinders

Let us consider the self-inductance of coaxial cylinders, as shown in Fig. 6.5. This problem turns out to be of central importance when calculating the electromagnetic behaviour of coaxial cables, such as those used with television aerials.

Assume that the cylinders have inner and outer radii a and b respectively. Consider a length l of the system, and assume that the cylinders are connected together electrically at the far end so that a current passed down the inner returns along the outer. Imagine that ultimately one needs to know the voltage generated between the cylinders, at the input.

There is no magnetic field outside the outer cylinder, because there is no net current flowing through a path of integration having a constant radius outside the outer: the two internal currents are

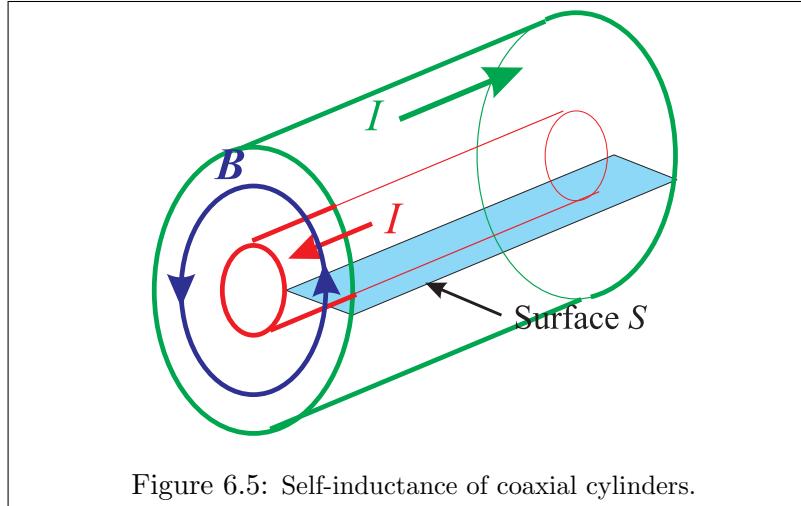


Figure 6.5: Self-inductance of coaxial cylinders.

equal and flow in opposite directions. Consider a circular coaxial path of integration of radius r , where $a < r < b$ (for example following the \mathbf{B} -field in blue in Fig. 6.5). In this case, Ampère's law,

$$\oint \mathrm{d}\mathbf{l} \cdot \mathbf{H} = I, \quad (6.26)$$

gives according to Eq. (4.64)

$$B(r) = \frac{\mu_0 I}{2\pi r}, \quad (6.27)$$

because the field always points in the direction of the path of integration, and has constant magnitude for a constant radius.

Now that the magnetic field is known, the next step is to calculate the magnetic flux in order to apply Faraday's law. The surface S used for evaluating Faraday's law, and hence the path of integration defined by the boundary ∂S , are shown in Fig. 6.5. The magnetic flux passing normally through S is

$$\Phi = l \int_a^b \mathrm{d}r B(r), \quad (6.28)$$

or

$$\Phi = \frac{\mu_0 I l}{2\pi} \ln \left(\frac{b}{a} \right), \quad (6.29)$$

from which it follows that the **self-inductance** is

$L = \frac{\Phi}{I} = \frac{\mu_0 l}{2\pi} \ln \left(\frac{b}{a} \right).$
Self-inductance of coaxial cable

(6.30)

Note that, as expected, L scales with the length.

6.7 Self-inductance of a pair of wires

Consider the parallel wires shown in Fig. 6.6. This type of geometry is often used in applications such as telephone signal cables

or digital data-transmission cables. The calculation proceeds in a similar way to that for coaxial cylinders.

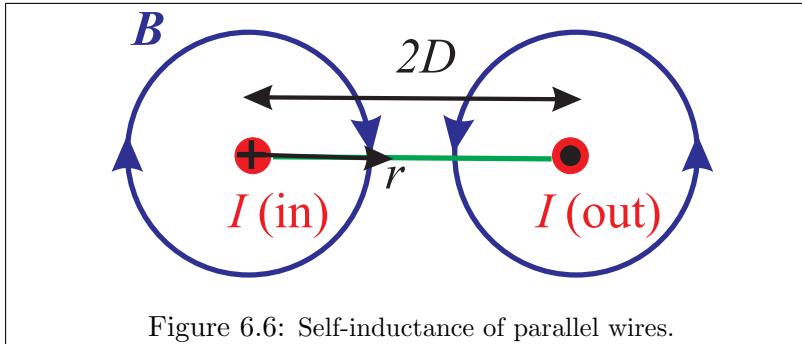


Figure 6.6: Self-inductance of parallel wires.

Assume that the system has length l , and that the individual wires have radius a , and are separated by distance $2D$. We shall also assume that $a \ll D$. The magnetic field due to one wire is

$$B(r) = \frac{\mu_0 I}{2\pi r}. \quad (6.31)$$

If one end of the cable is shorted, then the total flux passing through the loop that is formed is

$$\Phi_1 = \frac{\mu_0 Il}{2\pi} \int_a^{2D-a} dr \frac{1}{r}, \quad (6.32)$$

or, since $2D - a \approx 2D$,

$$\Phi_1 \stackrel{2D-a \approx 2D}{\approx} \frac{\mu_0 Il}{2\pi} \ln \left(\frac{2D}{a} \right). \quad (6.33)$$

The flux in (6.33) results from just one wire. However, the returning wire also carries current I , and hence produces the same amount of flux, and in the same direction. Thus, the flux due to both wires becomes

$$\Phi \approx \frac{\mu_0 Il}{\pi} \ln \left(\frac{2D}{a} \right), \quad (6.34)$$

and therefore the **self-inductance** is

$$L = \frac{\Phi}{I} = \frac{\mu_0 l}{\pi} \ln \left(\frac{2D}{a} \right). \quad \text{Self-inductance of two parallel wires}$$

(6.35)

Notice that we have assumed that there is no flux within the wires themselves. As we shall see later, flux does penetrate into a conductor that has a finite conductivity, but the penetration depth becomes smaller as the frequency increases. At high frequencies, all of the current can be assumed to flow in a thin shell on the surface, which justifies our assumption that the internal flux can be neglected, but strictly speaking it should be included. At low frequencies, this self-inductance of each wire can also be shown to be small, $\sim 0.05 \mu\text{H m}^{-1}$. As an example, in nanotechnology, where wires can be very tiny, the internal flux does play an important role in the operation of a conductor.

6.8 Self-inductance and voltage

Self-inductance provides a way of relating the voltage that appears across a circuit to the rate of change of current that flows through the same circuit.

Consider a loop of wire, where a source of current is connected across a small gap that is opened up in the wire.

Faraday's law can be used to calculate the voltage that appears across the gap:

$$\mathcal{E} = -\frac{d\Phi}{dt} \Rightarrow \oint_{\partial S} dl \cdot \mathbf{E} = -\frac{d}{dt} \int_S dS \cdot \mathbf{B}(\mathbf{r}), \quad (6.36)$$

where the path of integration follows the wire and defines the boundary of the surface S .

If there is no current flowing, or if the conductor is perfect, there is no electric field in the conductor, and so the only part of the path that can contribute to the integral is within the gap, where there can be a non-zero electric field. The electric potential difference is related to the field across the gap by

$$V_{\text{gap}} = - \int_{\text{gap}} dl \cdot \mathbf{E} = +\frac{d\Phi}{dt}, \quad (6.37)$$

It is vitally important to get the signs correct, and to appreciate which end of the gap corresponds to the high-potential side and which to the low-potential side, when, say, the current is increasing.

It then follows, trivially, that

$V_{\text{gap}} = L \frac{dI}{dt},$

(6.38)

which is one of the main equations in circuit theory.

Thought for the day 16: In derivations of this kind it is easy to lose sight of the sign conventions being used. In the above derivation, be very careful to examine the direction in which the field across the gap is pointing, and what this means for the voltage. Consider a circuit comprising a voltage source, a resistor, and an inductor connected in series. It is often said that as the source voltage is changed, the voltage that appears across the terminals of the inductor tries to oppose the change in current. In what direction is the voltage across the inductor when the current is (a) increasing and (b) decreasing, and how can this be interpreted as attempting to oppose the change in current? Why is it easy to get an electric shock from a benign-looking circuit?

Eq. 6.38 can be compared with the equivalent expression for a capacitor, for which we have

$$Q = CV. \quad (6.39)$$

If a source of current is connected across a capacitor, positive charge moves from one plate to the other at a rate I (charge per unit time), and therefore

$$I = \frac{dQ}{dt} = C \frac{dV}{dt}, \quad (6.40)$$

i.e.,

$I = C \frac{dV}{dt}.$

(6.41)

Equations 6.38 and 6.41 are very similar, apart from the fact that the roles of voltage and current are interchanged; indeed, this similarity makes them easy to remember.

6.9 Energy stored in an inductor

Consider an $L - R$ circuit and a voltage source, as shown in Fig. 6.7.

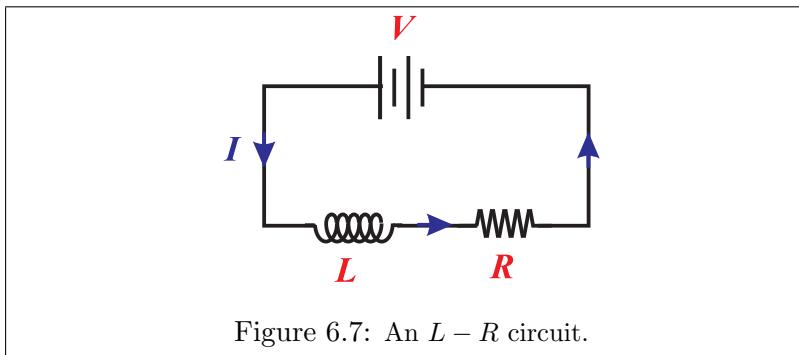


Figure 6.7: An $L - R$ circuit.

If there is no changing magnetic field threading the circuit, as distinct from the coils of the inductor, by virtue of Faraday's law the voltages around the circuit must sum to zero:

$$V = IR + L \frac{dI}{dt}. \quad (6.42)$$

[As an aside, note that if flux did thread the circuit, we could treat it as a stray inductance, in a perfect circuit.] The rate at which the voltage source delivers energy to the circuit is VI (note that positive charge leaves the positive side of the voltage source at a higher potential than enters the negative side), and therefore

$$VI = I^2 R + LI \frac{dI}{dt} = I^2 R + \frac{d}{dt} \left(\frac{1}{2} LI^2 \right), \quad (6.43)$$

where the chain rule has been used. In other words, the rate at which energy is delivered by the source is equal to the rate at which

energy is dissipated in the resistor plus the rate at which energy is stored in the magnetic field of the inductor.

The energy stored in the inductor is therefore given by

$$U_L = \frac{1}{2}LI^2. \quad (6.44)$$

It can also be shown, by a similar argument, that the energy stored in a capacitor is given by

$$U_C = \frac{1}{2}CV^2. \quad (6.45)$$

Inductors can be regarded as circuit elements that represent the storage of energy in a magnetic field; capacitors represent the storage of energy in an electric field; and resistors are components that represent the loss of energy from a circuit.

In an equivalent circuit, the components do not even have to represent actual inductors, capacitors, or resistors, but can represent any physical mechanism that behaves in an appropriate way. For example:

- A resistance is measured when looking into the terminals of a loudspeaker, because acoustic waves transfer energy out of the system.
- Inductance appears in the equivalent circuit of a superconductor because energy is stored in the kinetic, undamped motion of superconducting paired electrons.

You will see equivalent circuit components appearing in numerous areas of physics.

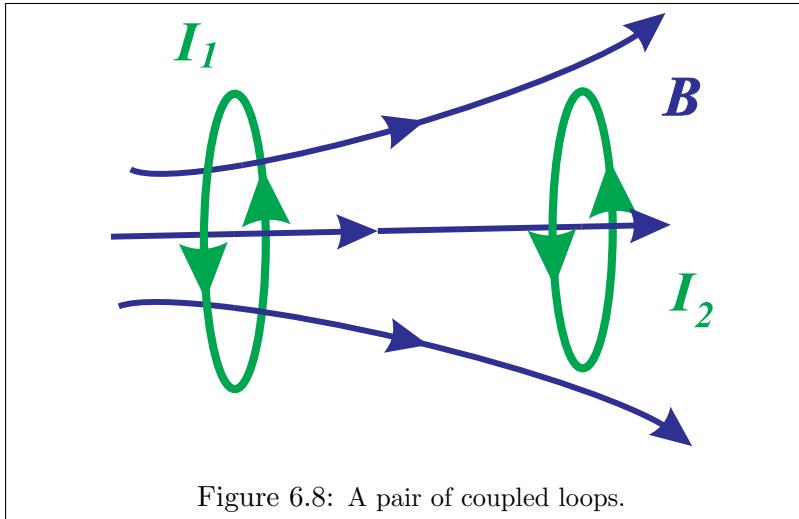
6.10 Mutual inductance

We have all of the main components of an electrical circuit; in fact the whole of electric circuit theory seems to be based on the electromagnetic field equations we have discussed. There is a situation, however, where an additional concept is required.

Consider what happens when two coils are placed in close proximity to each other, as shown in Fig. 6.8.

The current I_1 in loop 1 produces a flux Φ_2 that links into loop 2. This motivates the definition of the **mutual inductance**, M_{21} , according to

$$M_{21} = \frac{\Phi_2}{I_1}, \quad \text{Mutual inductance} \quad (6.46)$$



where subscripts have been used to show that flux is linked into loop 2 as a consequence of current flowing in loop 1.

It is instructive to show that mutual inductance is symmetric, i.e., that $M_{12} = M_{21}$, which we will do in the following.

First, suppose that $I_1 = I_2 = 0$, and that I_1 is gradually turned on by placing a current source in a small gap in the wire. This increase in current will lead to a voltage across the current source:

$$V_{11} = L_1 \frac{dI_1}{dt}, \quad (6.47)$$

where the subscript indicates that the voltage is induced in loop 1 by current 1. The energy stored in the magnetic field of the inductor is thus given by

$$U_1 = \frac{1}{2} L_1 I_1^2. \quad (6.48)$$

Now suppose that I_2 is turned on keeping I_1 fixed at its final value. The additional energy stored in the system, above that already stored, is

$$U_2 = \frac{1}{2} L_2 I_2^2. \quad (6.49)$$

As the second current is increased, however, it induces a voltage across the terminals of the first coil, where a constant current is now flowing. This voltage is given by

$$V_{12} = M_{12} \frac{dI_2}{dt}, \quad (6.50)$$

and therefore energy flows into the magnetic field at a rate

$$V_{12} I_1 = M_{12} \frac{dI_2}{dt} I_1 \stackrel{I_1=\text{const}}{=} \frac{d}{dt} M_{12} I_1 I_2. \quad (6.51)$$

The *final total energy* stored in the system when both currents are at their final values is therefore

$$U = \frac{1}{2}L_1I_1^2 + \frac{1}{2}L_2I_2^2 + M_{12}I_1I_2. \quad (6.52)$$

The final total energy stored must be the same regardless of the order in which the current sources were turned on, meaning that the indices can be swapped, and we find that

$$M_{12} = M_{21}. \quad (6.53)$$

In fact, for this reason, it is common practice not to include the subscripts at all, and simply to refer to the mutual inductance as M .

6.11 Combining self- and mutual inductance

In this section, we shall consider how self- and mutual inductances are combined in circuit analysis. Before doing so, we shall define the sign convention for the mutual inductance: M is positive if the currents produce fluxes that are in the same direction, as indicated in Fig. 6.9.

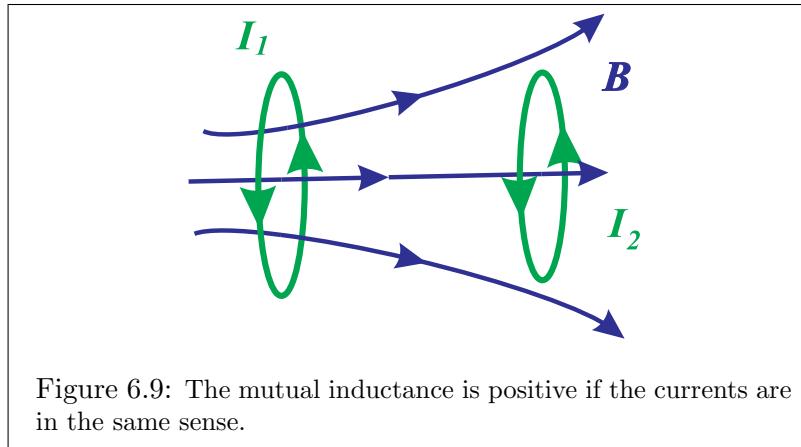


Figure 6.9: The mutual inductance is positive if the currents are in the same sense.

The total fluxes linked into loops 1 and 2 are

$$\begin{aligned}\Phi_1 &= L_1I_1 + MI_2 \\ \Phi_2 &= MI_1 + L_2I_2,\end{aligned} \quad (6.54)$$

which are simply the superpositions of the fluxes associated with the currents individually. It follows that

$$\begin{aligned}V_1 &= L_1 \frac{dI_1}{dt} + M \frac{dI_2}{dt} \\ V_2 &= M \frac{dI_1}{dt} + L_2 \frac{dI_2}{dt},\end{aligned} \quad (6.55)$$

which is a set of coupled differential equations that relates the currents at the terminals to the voltages at the terminals. The equivalent circuit corresponding to these equations is shown in Fig. 6.10.

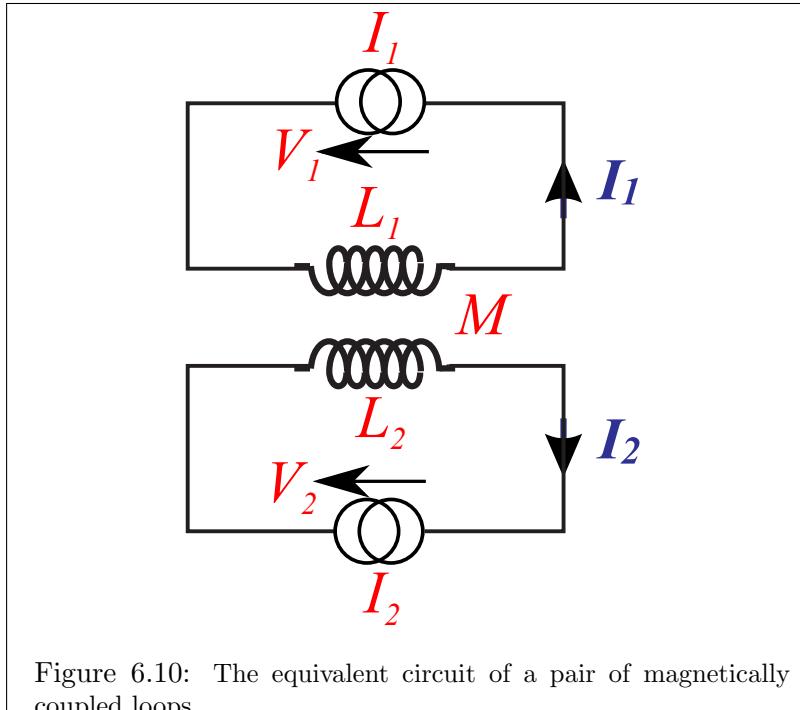


Figure 6.10: The equivalent circuit of a pair of magnetically coupled loops.

It is instructive to calculate an expression for the total energy stored from a circuit-theoretic perspective. We know that the total rate at which energy is fed into the magnetic field of the system is given by

$$W = V_1 I_1 + V_2 I_2, \quad (6.56)$$

Thus, substituting from Eq. (6.55),

$$W = I_1 L_1 \frac{dI_1}{dt} + I_1 M \frac{dI_2}{dt} + I_2 L_2 \frac{dI_2}{dt} + I_2 M \frac{dI_1}{dt}, \quad (6.57)$$

i.e.,

$$W = \frac{d}{dt} \left(\frac{1}{2} I_1^2 L_1 \right) + \frac{d}{dt} \left(\frac{1}{2} I_2^2 L_2 \right) + \frac{d}{dt} (I_1 I_2 M). \quad (6.58)$$

The total energy stored is therefore

$$U = \frac{1}{2} I_1^2 L_1 + \frac{1}{2} I_2^2 L_2 + I_1 I_2 M, \quad (6.59)$$

which is the same as our previous expression (6.52).

6.12 The relationship between self- and mutual inductance

It is not apparent from our discussion of self- and mutual inductance that there is a close relationship between the two. Indeed, there is one, as we shall see in the following.

We know that the total energy stored is given by

$$U = \frac{1}{2}I_1^2L_1 + I_1I_2M + \frac{1}{2}I_2^2L_2, \quad (6.60)$$

which, after completing the square, can be written

$$U = \frac{1}{2}L_1\left(I_1 + \frac{M}{L_1}I_2\right)^2 + \frac{1}{2}\left(L_2 - \frac{M^2}{L_1}\right)I_2^2. \quad (6.61)$$

The total energy must be positive, and so there is a restriction on the value of M . U must be positive even when the first bracket is zero (remember that the currents can have opposite signs). It follows that

$$L_1L_2 \geq M^2. \quad (6.62)$$

In fact, a **coupling coefficient k** , $0 \leq k \leq 1$ is usually defined such that

$$M = k(L_1L_2)^{1/2}, \quad \boxed{\text{Coupling coefficient } k} \quad (6.63)$$

The coupling coefficient quantifies the degree to which two coils having certain self-inductances are coupled together. $k = 0$ indicates no coupling whatsoever, whereas $k = 1$ indicates perfect coupling.

As an example, consider the doubly-wound solenoid shown in Fig. 6.11. The numbers of turns per unit length are n_1 and n_2 respectively. A and l are the same for both coils.

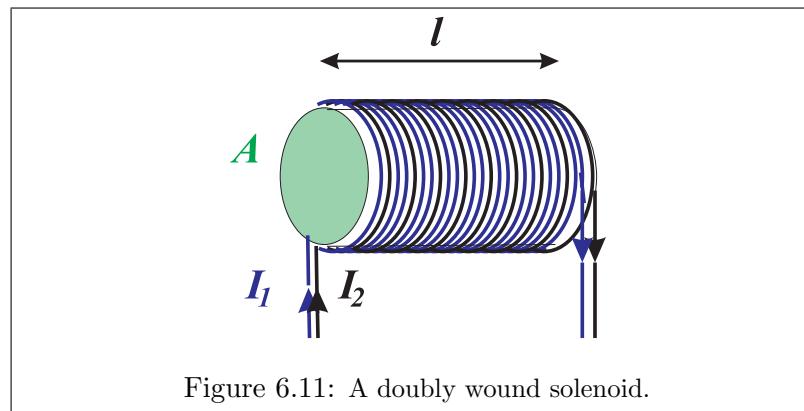


Figure 6.11: A doubly wound solenoid.

We know that

$$B_1 = \mu_0 n_1 I_1, \quad (6.64)$$

and the flux linking coil 1 becomes

$$\Phi_1 = B_1 A n_1 l, \quad (6.65)$$

$$\Rightarrow L_1 = \mu_0 n_1^2 A l. \quad (6.66)$$

The flux linking coil 2 becomes

$$\Phi_2 = B_1 A n_2 l, \quad (6.67)$$

$$\Rightarrow M = \mu_0 n_1 n_2 A l. \quad (6.68)$$

Similarly,

$$L_2 = \mu_0 n_2^2 A l. \quad (6.69)$$

Considering the above, we see that

$$M = (L_1 L_2)^{1/2}. \quad (6.70)$$

In this case, the same flux is enclosed by both coils, and the coils are perfectly coupled, so $k = 1$. For perfectly coupled circuits, it is only necessary to know the self-inductances in order to calculate the mutual inductance.

6.13 The ideal transformer

Now we turn to the case of the ideal transformer, where two coils are coupled by means of a core of ferromagnetic material. We shall assume that all of the flux couples to both coils, such that the coupling coefficient is unity. In this case, it is more convenient to refer to the total number of turns N , rather than the number of turns per unit length, n . In the case of ideal transformers we assume that the wires have no resistance, and that the core is linear and does not exhibit hysteresis.

If the coils are perfectly coupled, the same flux Φ passes through both coils, so the linked fluxes, Φ_1 and Φ_2 , become

$$\Phi_1 = N_1 \Phi, \quad \Phi_2 = N_2 \Phi, \quad (6.71)$$

and

$$\begin{aligned} V_1 &= \frac{d\Phi_1}{dt} = N_1 \frac{d\Phi}{dt} \\ V_2 &= \frac{d\Phi_2}{dt} = N_2 \frac{d\Phi}{dt}, \end{aligned} \quad (6.72)$$

giving the **ratio of voltages for an ideal transformer**:

$$\frac{V_2}{V_1} = \frac{N_2}{N_1}.$$

Ratio of voltages for an ideal transformer

$$(6.73)$$

The applications of transformers are considerable, ranging from power supplies that transform the domestic socket voltage of 230 V down to typically 5-12 V that are used in most electronic devices to gigantic metre-sized, oil-cooled devices that transform the voltage up to 400 kV in power lines to reduce losses.

The self-inductances of the individual coils are

$$\begin{aligned} L_1 &= \mu \mu_0 \left(\frac{N_1}{l_1} \right)^2 A l_1 \\ L_2 &= \mu \mu_0 \left(\frac{N_2}{l_2} \right)^2 A l_2 \end{aligned} \quad (6.74)$$

and therefore

$$\frac{L_1}{L_2} = \left(\frac{N_1}{N_2} \right)^2 \frac{l_2}{l_1}. \quad (6.75)$$

Thought for the day 17: The above analysis only applies to ideal transformers; real transformers use wire having finite conductivity, and not all of the flux that passes through one coil couples to the other coil. Show how four additional components can be added to the circuit of a perfect transformer to represent these imperfections. What other losses might be present?

6.14 Transformer circuits

Fig. 6.12 shows a simple circuit containing a perfect transformer.

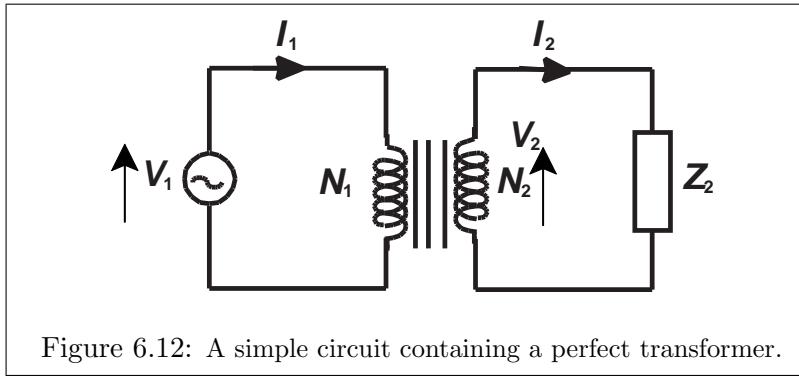


Figure 6.12: A simple circuit containing a perfect transformer.

In particular, a sinusoidally varying voltage source is connected to the primary of a perfect transformer that has an impedance of Z_2 connected to its secondary. We might wonder how to calculate the current, I_1 , that flows in the primary.

Assume a sign convention: when the primary and secondary currents both flow into the top of the transformer, their resultant fluxes add. Applying this convention to our case shown in Fig. 6.12, the two fluxes subtract because the secondary current is shown flowing in the opposite direction.

[Aside: these are just sign *conventions*: we do not know the actual directions until the problem has been solved!]

The total flux for circuit 1, Φ_1 , is simply

$$\Phi_1 = L_1 I_1 - M I_2, \quad (6.76)$$

which leads to

$$\begin{aligned} V_1 &= L_1 \frac{dI_1}{dt} - M \frac{dI_2}{dt} \\ V_2 &= -L_2 \frac{dI_2}{dt} + M \frac{dI_1}{dt}. \end{aligned} \quad (6.77)$$

It is common practice to assume that all of the voltages and currents in a circuit can be written in the form of complex exponentials:

$$\begin{aligned} V(t) &= V e^{j\omega t} \\ I(t) &= I e^{j\omega t}. \end{aligned} \quad (6.78)$$

The actual current has to be found by taking the real part of these complex expressions. It is convenient to use complex analysis here because it provides a powerful way of manipulating trigonometric functions using vectors in the complex plane.

Using a complex representation, we note

$$\begin{aligned} V_1 &= j\omega L_1 I_1 - j\omega M I_2 \\ V_2 &= -j\omega L_2 I_2 + j\omega M I_1. \end{aligned} \quad (6.79)$$

Moreover, we also know that the output voltage is constrained by Z_2 :

$$V_2 = Z_2 I_2, \quad (6.80)$$

and therefore the second equation becomes

$$j\omega M I_1 = (Z_2 + j\omega L_2) I_2. \quad (6.81)$$

Substituting I_2 into (6.79) gives:

$$\begin{aligned} V_1 &= j\omega L_1 I_1 - j\omega M I_2 \\ &= j\omega L_1 I_1 - \frac{(j\omega M)^2 I_1}{Z_2 + j\omega L_2}, \end{aligned} \quad (6.82)$$

and therefore

$$Z_1 \equiv \frac{V_1}{I_1} = j\omega L_1 - \frac{(j\omega M)^2}{Z_2 + j\omega L_2}. \quad (6.83)$$

We know from Eq. (6.70) that

$$M^2 = L_1 L_2, \quad (6.84)$$

and from Eq. (6.75)

$$\frac{L_1}{L_2} = \left(\frac{N_1}{N_2} \right)^2, \quad (6.85)$$

if we assume that the coils are of equal length.

Finally, after combining the above results, we have

$$Z_1 = \frac{j\omega L_1 Z_2 (N_1/N_2)^2}{j\omega L_1 + Z_2 (N_1/N_2)^2}. \quad (6.86)$$

Thus the input impedance comprises $j\omega L_1$ in parallel with $Z_2(N_1/N_2)^2$.

For typical electric devices and experimental setups

$$\omega L_1 \gg Z_2 (N_1/N_2)^2, \quad (6.87)$$

and therefore we can write for the **load impedance ratio across the transformer**:

$Z_1 \approx Z_2 (N_1/N_2)^2. \quad \text{Load impedance ratio across the transformer}$

(6.88)

In other words, the load impedance is transformed across the transformer so as to appear multiplied by the square of the turns ratio. In fact, one of the principal uses of transformers is to impedance-match loads to sources.

6.15 Energy flow in resonant circuits

In the following, we shall study the flow of energy in resonant circuits. For this, consider a circuit comprising a resistor, inductor, and capacitor connected in series. This circuit is connected to a current source having the form

$$I(t) = I_0 \cos \omega t, \quad (6.89)$$

where we have reverted to the ordinary time-domain representation. The voltages across the resistor, inductor, and capacitor are

$$\begin{aligned} V_R(t) &= I(t)R &= I_0 R \cos \omega t \\ V_L(t) &= L \frac{dI(t)}{dt} &= -I_0 \omega L \sin \omega t \\ V_C(t) &= \frac{1}{C} \int dt I(t) &= I_0 \frac{1}{\omega C} \sin \omega t \end{aligned}$$

The instantaneous powers flowing into each of the components are then

$$\begin{aligned} P_R(t) &= I(t)V_R(t) &= I_0^2 R \cos^2 \omega t \\ P_L(t) &= I(t)V_L(t) &= -I_0^2 \omega L \frac{1}{2} \sin 2\omega t \\ P_C(t) &= I(t)V_C(t) &= I_0^2 \frac{1}{\omega C} \frac{1}{2} \sin 2\omega t \end{aligned}$$

We see that at times when energy is flowing into the magnetic field of the inductor, it is flowing out of the electric field of the capacitor; conversely, when energy is flowing out of the magnetic field of the inductor, it is flowing into the electric field of the capacitor. Energy always flows into the resistor.

The total rate at which energy is stored is given by

$$P_S(t) = P_C(t) + P_L(t) = \frac{I_0^2}{2} \left(\frac{1}{\omega C} - \omega L \right) \sin 2\omega t. \quad (6.90)$$

If $1/\omega C = \omega L$, $P_S(t) = 0$ at all times. The rate at which energy flows in and out of the capacitor is precisely the same as the rate at which energy flows in and out of the inductor, but they have opposite signs. No additional energy needs to be supplied at any part of the cycle. This condition is called **resonance**: energy sloshes backwards and forwards between the inductor and capacitor. The only energy that needs to be supplied is that required to sustain the losses, and the circuit looks, overall, like a resistor of value R . The **resonant frequency**, ω_0 , is given by

$$\omega_0 = \frac{1}{\sqrt{LC}}. \quad \text{Resonant frequency}$$

(6.91)

For $\omega > \omega_0$, $1/\omega C < \omega L$, and the external current source must lend energy to the circuit in such a way that the overall circuit looks inductive, because not enough energy was stored in the capacitor.

In other words, electrical circuit theory is merely a way of describing how energy moves in and out of magnetic and electric fields, and in

the case of resistors is lost permanently to other forms. We shall see that all other classical electromagnetic phenomena can be described in terms of the sloshing backwards and forwards of energy between electric and magnetic fields, and that equivalent circuits exist for physical arrangements that have no obvious electrical components. For example, an electromagnetic wave propagating down a metallic tube can be described by an electrical circuit!

Thought for the day 18: Draw, as a function of time, a graph of the current, and the instantaneous power flow into each of the resistor, capacitor, and inductor, and show that the circuit behaves as described.

6.16 Magnetic energy

In the case of electrostatic fields, we saw that it is possible to imagine that energy is stored in the field itself when charges are assembled. In the case of a magnetic field, it is possible to consider energy as being stored in the magnetic field itself as current loops are assembled.

To calculate the energy that is stored in the magnetic field, we note that the energy stored in an inductor is given by

$$U = \frac{1}{2}LI^2. \quad (6.92)$$

We also know that

$$L = \frac{\Phi}{I}, \quad (6.93)$$

and therefore, for *one* circuit,

$$U = \frac{1}{2}\Phi I. \quad (6.94)$$

Eq. (6.94) expresses the energy stored in terms of the total flux linked and the current.

Now consider *two* circuits between which flux is shared. In Section 6.10 we saw that

$$U = \frac{1}{2}L_1I_1^2 + \frac{1}{2}L_2I_2^2 + I_1I_2M, \quad (6.95)$$

which can be written

$$U = \frac{1}{2}(L_1I_1 + MI_2)I_1 + \frac{1}{2}(L_2I_2 + MI_1)I_2, \quad (6.96)$$

or

$$U = \frac{1}{2}\Phi_1I_1 + \frac{1}{2}\Phi_2I_2. \quad (6.97)$$

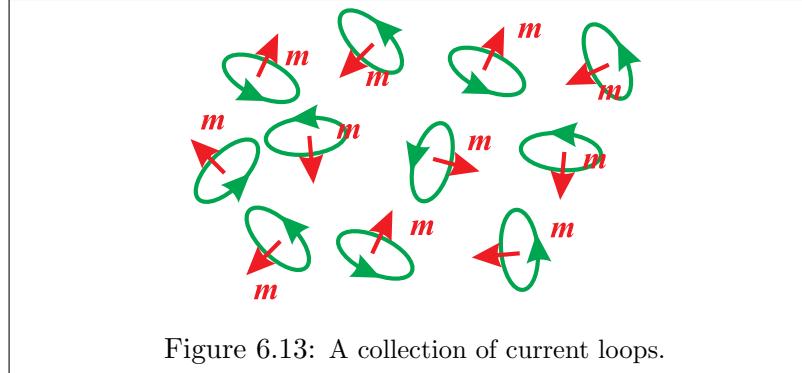


Figure 6.13: A collection of current loops.

Φ_1 and Φ_2 are the fluxes that thread loops 1 and 2, respectively.

This argument can be extended, and therefore, for a collection of current loops, as shown in Fig. 6.13,

$$U = \sum_{i \in \text{loops}} \frac{1}{2} \Phi_i I_i, \quad (6.98)$$

where I_i is the current flowing in loop i , and Φ_i is the total flux threading loop i .

Eq. (6.98) does not, however, represent the total energy needed to assemble a collection of current loops, because if the loops are moved apart such that the coupling coefficient M tends to zero, U does not tend to zero, because it includes the energy stored in the individual loops.

For our example two-loop system

$$U \rightarrow \frac{1}{2} L_1 I_1^2 + \frac{1}{2} L_2 I_2^2, \quad (6.99)$$

as $M \rightarrow 0$. In fact, the first two terms in Eq. (6.95) can be regarded as the **self-energies** of the two loops, whereas the third term is the **interaction energy**.

If Φ_i in Eq. (6.98) is instead considered to be the flux passing through loop i as a consequence of the currents in the other $N - 1$ loops, then Eq. (6.98) is the energy needed to assemble the complete system. We shall stay with the notion that U is the total energy stored in the system, rather than the energy needed to assemble the system from existing current loops.

We can now make progress rapidly. Remember that by virtue of Stokes' theorem

$$\Phi = \int d\mathbf{S} \cdot \mathbf{B} = \int d\mathbf{S} \cdot (\nabla \times \mathbf{A}) = \oint dl \cdot \mathbf{A}, \quad (6.100)$$

where the line integral follows the path of the current loop, and \mathbf{A} is the magnetic vector potential.

The total energy associated with a collection of current loops becomes, using Eq. (6.98):

$$U = \sum_{i \in \text{loops}} \frac{1}{2} \left(\oint_i I_i d\mathbf{l} \cdot \mathbf{A} \right), \quad (6.101)$$

where different paths must be taken for each of the current loops present.

The currents can be written in terms of current densities to yield

$$U = \sum_{i \in \text{loops}} \frac{1}{2} \int_i d^3\mathbf{r} \mathbf{J} \cdot \mathbf{A}. \quad (6.102)$$

The discrete sum over different loops can be combined with the volume integrals over specific loops to give

$$U = \frac{1}{2} \int d^3\mathbf{r} \mathbf{J} \cdot \mathbf{A}, \quad \text{Total magnetic energy} \quad (6.103)$$

which is already a powerful expression because it gives the total energy stored in terms of the scalar product between the current density \mathbf{J} and the magnetic vector potential \mathbf{A} . Eq. (6.103) is often used as the starting point for quantising the electromagnetic field. When used in the context of practical problems, Eq. (6.103) is usually expressed in a different form.

In the context of *magnetostatic* problems

$$\nabla \times \mathbf{H} = \mathbf{J}, \quad (6.104)$$

and therefore

$$U = \frac{1}{2} \int d^3\mathbf{r} \mathbf{J} \cdot \mathbf{A} = \frac{1}{2} \int d^3\mathbf{r} \mathbf{A} \cdot (\nabla \times \mathbf{H}). \quad (6.105)$$

One can use the vector calculus identity

$$\nabla \cdot (\mathbf{A} \times \mathbf{H}) = \mathbf{H} \cdot (\nabla \times \mathbf{A}) - \mathbf{A} \cdot (\nabla \times \mathbf{H}), \quad (6.106)$$

and therefore

$$\begin{aligned} U &= \frac{1}{2} \int d^3\mathbf{r} \mathbf{H} \cdot (\nabla \times \mathbf{A}) \\ &\quad - \frac{1}{2} \int d^3\mathbf{r} \nabla \cdot (\mathbf{A} \times \mathbf{H}). \end{aligned} \quad (6.107)$$

Applying the divergence theorem on the second term

$$\begin{aligned} U &= \frac{1}{2} \int d^3\mathbf{r} \mathbf{H} \cdot (\nabla \times \mathbf{A}) \\ &\quad - \frac{1}{2} \oint d\mathbf{S} \cdot (\mathbf{A} \times \mathbf{H}), \end{aligned} \quad (6.108)$$

or

$$\begin{aligned} U &= \frac{1}{2} \int d^3r \mathbf{H} \cdot \mathbf{B} \\ &- \frac{1}{2} \oint d\mathbf{S} \cdot (\mathbf{A} \times \mathbf{H}). \end{aligned} \quad (6.109)$$

As the surface over which the integral is evaluated increases in size, its area increases as R^2 , whereas \mathbf{A} falls as R^{-1} and \mathbf{H} falls as R^{-2} , and therefore the second term on the RHS tends to zero as $R \rightarrow \infty$.

The **stored magnetic energy** becomes

$$U_m = \frac{1}{2} \int d^3r \mathbf{B}(\mathbf{r}) \cdot \mathbf{H}(\mathbf{r}), \quad \text{Stored magnetic energy}$$

(6.110)

which implies that

$$\text{Magnetic energy density} = \frac{1}{2} \mathbf{B}(\mathbf{r}) \cdot \mathbf{H}(\mathbf{r}).$$

(6.111)

In these equations, the position dependence of the field quantities has now been shown explicitly. The expression for the magnetic energy density holds generally, and is of considerable importance.

Eq. (6.110) should be compared with the equivalent expression for the energy stored in the electric field of an electrostatic system, which we derived earlier:

$$U_e = \frac{1}{2} \int d^3r \mathbf{D}(\mathbf{r}) \cdot \mathbf{E}(\mathbf{r}). \quad (6.112)$$

The similarity between the two is striking.

7. Maxwell's Equations and Waves

7.1 Completing Maxwell's equations—the displacement current

We have studied the behaviour of electrostatic and magnetostatic fields, and have considered how time-varying magnetic fields lead to voltages. We have discussed a number of equations that seem to solve adequately all of the physical problems encountered. In particular, we have derived a set of equations that describe electrical circuit theory. The most important equations are

$$\left. \begin{array}{lcl} \nabla \cdot \mathbf{D} & = & \rho \\ \nabla \times \mathbf{E} & = & -\dot{\mathbf{B}} \\ \nabla \cdot \mathbf{B} & = & 0 \\ \nabla \times \mathbf{H} & = & \mathbf{J} \end{array} \right\} \text{(incomplete!)}$$

and we might be content having reached this point, but there is a big problem:

This set of field equations is not consistent with the conservation of charge!

To demonstrate the problem, remember that in Chapter 2 we showed that for charge to be conserved,

$$\oint d\mathbf{S} \cdot \mathbf{J} + \int d^3\mathbf{r} \frac{\partial \rho}{\partial t} = 0, \quad (7.1)$$

or, using the divergence theorem,

$$\int d^3\mathbf{r} \nabla \cdot \mathbf{J} + \int d^3\mathbf{r} \frac{\partial \rho}{\partial t} = 0. \quad (7.2)$$

Because this must be true for all volumes, we arrive at the **continuity equation**:

$\nabla \cdot \mathbf{J} + \frac{\partial \rho}{\partial t} = 0.$ **Continuity equation**

(7.3)

In the steady state, the charge density does not change anywhere with time, and so

$$\nabla \cdot \mathbf{J} = 0, \quad (7.4)$$

showing that currents form closed circulating loops. Using the last of the field equations $\nabla \times \mathbf{H} = \mathbf{J}$ to evaluate the continuity equation gives

$$\nabla \cdot \mathbf{J} = \nabla \cdot (\nabla \times \mathbf{H}) = 0, \quad (7.5)$$

because the divergence of the curl of any vector field is zero, but this means that the continuity equation cannot be satisfied for anything other than the steady state—magnetostatics.

Oh dear, the whole of the scheme seems to have fallen apart just when we were doing so well! All of the best physics comes out of unforeseen problems, so what is going on?

Maxwell suggested that the last of the field equations should be modified by adding an additional term $\partial\mathbf{D}/\partial t$,

$$\boxed{\nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t}, \quad \text{Maxwell's equation 4}} \quad (7.6)$$

which he called the **displacement current**. The continuity equation then becomes

$$\begin{aligned} \nabla \cdot \left(\nabla \times \mathbf{H} - \frac{\partial \mathbf{D}}{\partial t} \right) + \frac{\partial \rho}{\partial t} &= 0 \\ \nabla \cdot \left(-\frac{\partial \mathbf{D}}{\partial t} \right) + \frac{\partial \rho}{\partial t} &= 0 \\ -\frac{\partial \nabla \cdot \mathbf{D}}{\partial t} + \frac{\partial \rho}{\partial t} &= 0, \end{aligned} \quad (7.7)$$

which is consistent with the first of the field equations, and so the equations are self-consistent even when charge is accumulating or decreasing within some region.

The addition of the displacement current term to Eq. (7.6) seems rather simple, but was one of Maxwell's greatest contributions, and led to a revolution in physics. Indeed, we have finally arrived at the complete set of **Maxwell's equations**:

$$\boxed{\begin{aligned} \nabla \cdot \mathbf{D} &= \rho \\ \nabla \times \mathbf{E} &= -\dot{\mathbf{B}} \\ \nabla \cdot \mathbf{B} &= 0 \\ \nabla \times \mathbf{H} &= \mathbf{J} + \dot{\mathbf{D}} \end{aligned} \quad \text{Maxwell's equations (complete)}}$$

Maxwell's equations are self-consistent, and describe *every possible* physical phenomenon in classical electromagnetism.

7.2 Free-space waves

We can now unleash the full power of Maxwell's equations by predicting the existence of travelling waves. The additional term, $\partial\mathbf{D}/\partial t$, which was guessed at by Maxwell, is central to this process.

In free space, we have

$$\begin{aligned} \rho &= 0 \\ \mathbf{D} &= \epsilon_0 \mathbf{E} \\ \mathbf{J} &= \mathbf{0} \\ \mathbf{B} &= \mu_0 \mathbf{H}, \end{aligned} \quad (7.8)$$

because there is no free charge, and there are no conduction currents. Maxwell's equations then become

$$\boxed{\begin{aligned}\nabla \cdot \mathbf{E} &= 0 \\ \nabla \times \mathbf{E} &= -\mu_0 \dot{\mathbf{H}} \\ \nabla \cdot \mathbf{H} &= 0 \\ \nabla \times \mathbf{H} &= \epsilon_0 \dot{\mathbf{E}}.\end{aligned}}$$

Taking the curl of the second equation, $\nabla \times \mathbf{E} = -\mu_0 \dot{\mathbf{H}}$, gives

$$\nabla \times (\nabla \times \mathbf{E}) = -\mu_0 \frac{\partial \nabla \times \mathbf{H}}{\partial t}. \quad (7.9)$$

We expand the triple product and use $\nabla \cdot \mathbf{E} = 0$:

$$\nabla \times (\nabla \times \mathbf{E}) = \nabla (\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E} = -\nabla^2 \mathbf{E}. \quad (7.10)$$

Thus

$$\nabla^2 \mathbf{E} = \mu_0 \frac{\partial \nabla \times \mathbf{H}}{\partial t}. \quad (7.11)$$

Finally, using the last of Maxwell's equations in free-space, $\nabla \times \mathbf{H} = \epsilon_0 \dot{\mathbf{E}}$, yields the **wave equation in free space**

$$\boxed{\nabla^2 \mathbf{E} = \mu_0 \epsilon_0 \frac{\partial^2 \mathbf{E}}{\partial t^2}. \quad \text{Wave equation in free space}} \quad (7.12)$$

By an identical sequence of arguments, we can also show that

$$\boxed{\nabla^2 \mathbf{H} = \mu_0 \epsilon_0 \frac{\partial^2 \mathbf{H}}{\partial t^2}, \quad \text{Wave equation in free space}} \quad (7.13)$$

How can we understand these equations as representing *waves*? Before addressing this question, we shall familiarise ourselves with the notation. We are used to $\nabla^2 \phi$ where ϕ is a scalar, but here we have $\nabla^2 \mathbf{E}$ where \mathbf{E} is a vector. We can, however, just think of ∇^2 as acting on each component of \mathbf{E} . Remember that ∇^2 is a scalar differential operator, that acts on a scalar field to generate a scalar field. The operator ∇^2 therefore acts on a vector field to produce a vector field. In other words, in Cartesian coordinates $\nabla^2 \mathbf{E}$ takes the form

$$\nabla^2 \mathbf{E} = (\nabla^2 E_x, \nabla^2 E_y, \nabla^2 E_z). \quad (7.14)$$

We can now appreciate that each component of these equations, for each component of \mathbf{E} and \mathbf{H} , call it $\psi(\mathbf{r}, t)$, takes the form

$$\nabla^2 \psi(\mathbf{r}, t) = \frac{1}{c^2} \frac{\partial^2 \psi(\mathbf{r}, t)}{\partial t^2} \quad (7.15)$$

for three-dimensional problems, and for one-dimensional problems

$$\frac{\partial^2 \psi(x, t)}{\partial x^2} = \frac{1}{c^2} \frac{\partial^2 \psi(x, t)}{\partial t^2}. \quad (7.16)$$

These are forms of the well-known wave equation. Pure plane-wave solutions in three-dimensional space are of the form

$$\psi = \psi_0 e^{i(\mathbf{k}\cdot\mathbf{r} - \omega t)} \quad (7.17)$$

with oscillation frequency $\omega = 2\pi f$ and wave vector $|\mathbf{k}| = \frac{2\pi}{\lambda}$. Similarly, in one dimension,

$$\psi = \psi_0 e^{i(kx - \omega t)}. \quad (7.18)$$

Because, certainly in the one-dimensional case, the components of the vector \mathbf{E} and \mathbf{H} fields take the form of waves as shown in Eqs. (7.12) and (7.13), and because we know Eq. (7.15) that c corresponds to the speed of the wave, we can conclude for the speed of light in free space:

$c = \frac{1}{\sqrt{\epsilon_0 \mu_0}}, \quad \text{Speed of light in free space}$

(7.19)

which is $299,792,458 \text{ m s}^{-1}$. Eq. (7.19) is a beautiful unification of three very different quantities, two of which were originally introduced to describe the behaviour of electrostatic and magnetostatic fields.

The constant c is a profoundly important quantity; Einstein realised that it is the same for all observers, whatever their own constant speed of motion. In 1905, Einstein used Maxwell's equations, and the constancy of c , to formulate his special theory of relativity, and to discover that $E = mc^2$.

In the SI system, the metre is now defined in terms of c and the second. This *defines* the constants for the permittivity and permeability of free space:

$$\begin{aligned} \mu_0 &\equiv 4\pi \times 10^{-7} \text{ H m}^{-1} \text{ (or in N A}^{-2}\text{)} \\ \epsilon_0 &\equiv (c^2 \mu_0)^{-1} = 8.85 \times 10^{-12} \text{ F m}^{-1}. \end{aligned}$$

Thought for the day 19: Examine the second and the fourth of Maxwell's equations for the case of free space. There is a striking asymmetry between these two equations: one of them has a minus sign, while the other one does not. Which physical implications does this relative minus sign between the second and the fourth Maxwell's equation have? By contrast, how would our world look like if there were no relative minus sign?

7.3 Plane waves in isotropic media

The wave equations, derived through Maxwell's free-space field equations, contain a wealth of physical phenomena in their own right. Let us look for a **plane-wave** solution to the wave equation for a field propagating in the z -direction. By *plane-wave solution*, we mean that the field, at a fixed point in time, does not change as we move on the plane transverse to the propagation direction (here: in the x and y directions), but that it does change as we move in the propagation direction (here: z).

By definition, we certainly know that

$$\begin{aligned}\frac{\partial\psi(\mathbf{r},t)}{\partial x} &= 0 \\ \frac{\partial\psi(\mathbf{r},t)}{\partial y} &= 0.\end{aligned}\quad (7.20)$$

And therefore

$$\nabla \times \mathbf{H} = \begin{vmatrix} \hat{\mathbf{i}} & \hat{\mathbf{j}} & \hat{\mathbf{k}} \\ 0 & 0 & \frac{\partial}{\partial z} \\ H_x & H_y & H_z \end{vmatrix} = \left(-\frac{\partial H_y}{\partial z}, \frac{\partial H_x}{\partial z}, 0 \right). \quad (7.21)$$

Considering Maxwell's equation $\nabla \times \mathbf{H} = \epsilon_0 \dot{\mathbf{E}}$ yields

$$\begin{aligned}-\frac{\partial H_y}{\partial z} &= \epsilon_0 \frac{\partial E_x}{\partial t} \quad \text{Transverse polarisation P1} \\ \frac{\partial H_x}{\partial z} &= \epsilon_0 \frac{\partial E_y}{\partial t} \quad \text{Transverse polarisation P2} \\ 0 &= \epsilon_0 \frac{\partial E_z}{\partial t} \quad \text{Longitudinal}\end{aligned}\quad (7.22)$$

Likewise:

$$\nabla \times \mathbf{E} = \begin{vmatrix} \hat{\mathbf{i}} & \hat{\mathbf{j}} & \hat{\mathbf{k}} \\ 0 & 0 & \frac{\partial}{\partial z} \\ E_x & E_y & E_z \end{vmatrix} = \left(-\frac{\partial E_y}{\partial z}, \frac{\partial E_x}{\partial z}, 0 \right), \quad (7.23)$$

and then, considering Maxwell's equation $\nabla \times \mathbf{E} = -\mu_0 \dot{\mathbf{H}}$ results in

$$\begin{aligned}-\frac{\partial E_y}{\partial z} &= -\mu_0 \frac{\partial H_x}{\partial t} \quad \text{Transverse polarisation P2} \\ \frac{\partial E_x}{\partial z} &= -\mu_0 \frac{\partial H_y}{\partial t} \quad \text{Transverse polarisation P1} \\ 0 &= -\mu_0 \frac{\partial H_z}{\partial t} \quad \text{Longitudinal}\end{aligned}\quad (7.24)$$

We see that any longitudinal fields are *static* fields, and so will be ignored, and any *time-varying* fields are transverse. Thus, plane

electromagnetic waves in free space are transverse waves, as they only have field components that are *transverse* to the direction of propagation.

Notice that the transverse solutions come in pairs: E_x is linked to H_y , and E_y is linked to H_x . In fact, each pair can exist independently of the other, giving rise to two distinct solutions, which are known as linear **polarisations**. These can be grouped for convenience: the x -polarised waves

$$\begin{aligned}\frac{\partial H_y}{\partial z} &= -\epsilon_0 \frac{\partial E_x}{\partial t} \\ \frac{\partial E_x}{\partial z} &= -\mu_0 \frac{\partial H_y}{\partial t}\end{aligned}\tag{7.25}$$

and the y -polarised waves

$$\begin{aligned}\frac{\partial H_x}{\partial z} &= \epsilon_0 \frac{\partial E_y}{\partial t} \\ \frac{\partial E_y}{\partial z} &= \mu_0 \frac{\partial H_x}{\partial t}.\end{aligned}\tag{7.26}$$

Notice that the polarisation describes which component of the **E** -field is present. Eq. (7.25) can be solved by differentiating each equation a second time (w.r.t. t and z , respectively), and eliminating the common term (using the fact that $\partial^2/\partial p\partial q \equiv \partial^2/\partial q\partial p$), to give

$$\frac{\partial^2 E_x}{\partial z^2} - \epsilon_0 \mu_0 \frac{\partial^2 E_x}{\partial t^2} = 0,\tag{7.27}$$

which is the wave equation.

In conclusion, the E_x component of the field propagates in the z -direction in the form of a wave travelling at the speed of light. The H_y component does the same. Similar equations can be derived from Eq. 7.26 for the orthogonal polarisation.

If the medium through which the wave is travelling is linear, isotropic, homogeneous, and non-conducting, then similar solutions are found, but now the speed becomes

$$v = \frac{1}{\sqrt{\epsilon\epsilon_0\mu\mu_0}} = \frac{c}{\sqrt{\epsilon\mu}} = \frac{c}{n},\tag{7.28}$$

where the refractive index n is given by

$$n = \sqrt{\epsilon\mu}.\tag{7.29}$$

In non-magnetic dielectrics, where the relative permeability is unity, the refractive index is simply the square root of the relative dielectric constant.

7.4 Frequency of a plane wave

Let us suppose that the field is generated by means of a sinusoidally oscillating current, and that, because the medium is linear, all of the field components have the form

$$E_x = E_0 \exp[i(kz - \omega t)], \quad (7.30)$$

which is a travelling sinusoidal wave. Strictly, we mean

$$E_x = \operatorname{Re} \{E_0 \exp[i(kz - \omega t)]\}, \quad (7.31)$$

where $\operatorname{Re}()$ means the real part; this only matters when multiplying two such quantities together—**you must take real parts before multiplying**, unless you use the expression for the complex power (see Eq. 7.82 later in this chapter)!

Substituting the assumed form from Eq. (7.30) into the wave equation (7.27) and using $\omega = 2\pi f$ and $k = 2\pi/\lambda$ for frequency f and wavelength λ , gives

$$\frac{\omega}{k} = \frac{1}{\sqrt{\epsilon_0 \mu_0}} = c = \frac{2\pi f \lambda}{2\pi} = f\lambda, \quad (7.32)$$

which is a well-known expression. The parameter k is called the **wave number** (or **propagation constant**), and gives, for a fixed point in time, the rate at which the phase changes with position. Fourier analysis tells us that any propagating pulse, or time-varying waveform, can be built up from a set of plane-wave components of this kind.

Eq. (7.30) can also be used with Eq. (7.25) to find the relationships between the individual field components. We find

$$\frac{E_x}{H_y} = \frac{k}{\epsilon_0 \omega} = \frac{1}{\epsilon_0 c} = \sqrt{\frac{\mu_0}{\epsilon_0}}, \quad (7.33)$$

or equivalently

$$\frac{E_x}{H_y} = \frac{\mu_0 \omega}{k} = \mu_0 c = \sqrt{\frac{\mu_0}{\epsilon_0}}. \quad (7.34)$$

Thus, there is a definite relationship between the two field quantities. Firstly, they are in phase, and therefore propagate locked in perfect phase with respect to each other. Secondly, their magnitudes are related by

$$Z_0 = \sqrt{\frac{\mu_0}{\epsilon_0}} = \frac{E_x}{H_y}, \quad (7.35)$$

which is called the impedance of free space, because it is a quantity that has the dimensions of impedance, i.e., V A^{-1} : E has the units of V m^{-1} and H has the units of A m^{-1} .

Using the values of the ϵ_0 and μ_0 constants we find for the **characteristic impedance of free space**:

$$Z_0 \equiv \sqrt{\frac{\mu_0}{\epsilon_0}} = 377 \Omega. \quad \text{Impedance of free space} \quad (7.36)$$

It must be appreciated that the characteristic impedance of free space does not dissipate power; its origin can be traced to the fact that an electromagnetic wave can either supply power or carry power away.

For a linear, isotropic, homogeneous, and non-conducting, medium

$$Z = \sqrt{\frac{\mu\mu_0}{\epsilon\epsilon_0}} = 377\sqrt{\frac{\mu}{\epsilon}} \Omega. \quad (7.37)$$

It is also apparent that

$$E_x = \mu_0 c H_y = c B_y. \quad (7.38)$$

In other words, the \mathbf{E} and \mathbf{H} fields are related through the speed of light, showing that they are fundamentally connected, as was later explained by Einstein. Similarly, for a wave with electric field E_x , travelling in the negative z -direction,

$$Z = -\sqrt{\frac{\mu\mu_0}{\epsilon\epsilon_0}}. \quad (7.39)$$

It is possible to calculate an identical set of expressions for the y -polarised wave, giving

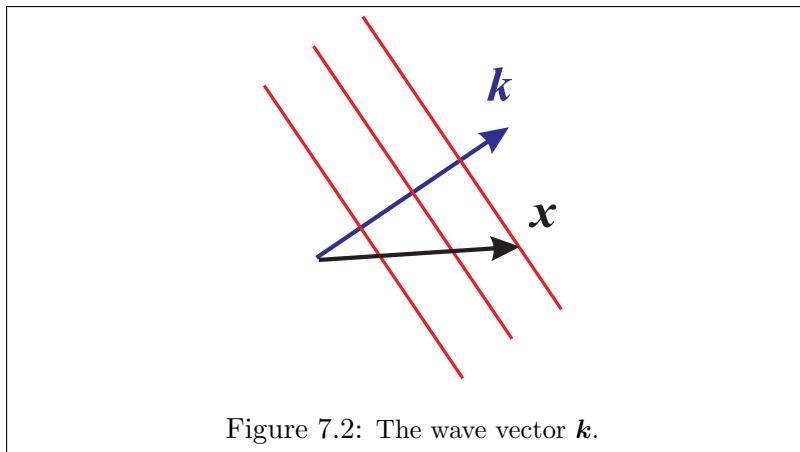
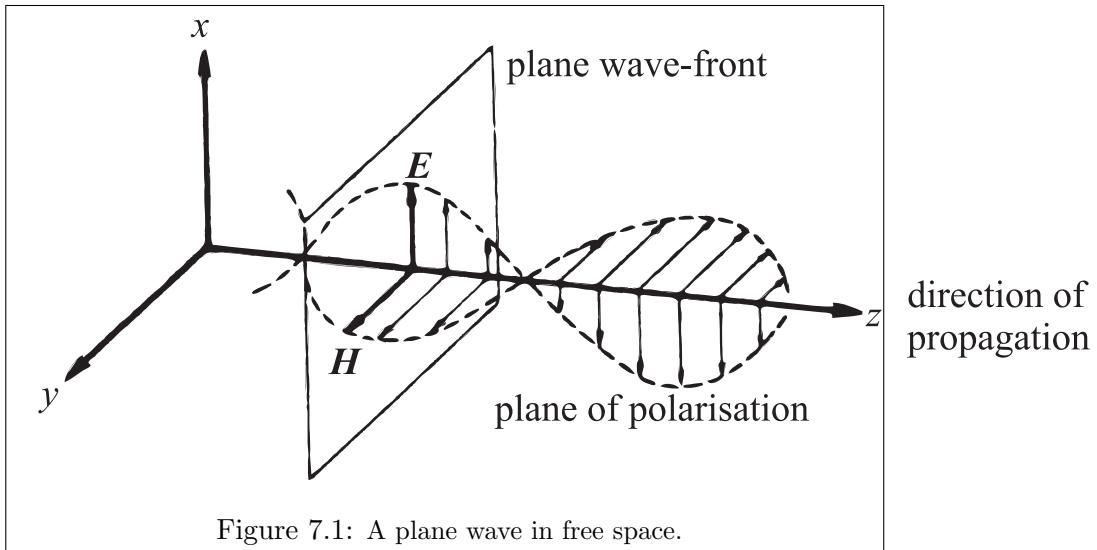
$$\frac{E_y}{H_x} = -\frac{\mu_0 \omega}{k} = -\mu_0 c = -\sqrt{\frac{\mu_0}{\epsilon_0}}. \quad (7.40)$$

The negative sign appears because the \mathbf{H} field actually points in the negative- x direction. Thus, the \mathbf{E} field, \mathbf{H} field, and the direction of propagation form a right-handed coordinate system, cf. Fig. 7.1. In free space, this is generally true for any propagation direction, as will be seen in the next section. Each linear polarisation has the same form, but they are rotated with respect to each other.

7.5 General plane waves

We have already studied the key features of plane electromagnetic waves, but we chose the direction of propagation to be the z -direction. Can we handle any general direction of propagation in a mathematically elegant way? Assume that a plane wave is travelling in the direction of \mathbf{k} , where \mathbf{k} is called the **wave vector**. In this case, the phases of the various field components have the functional form

$$(\mathbf{k} \cdot \mathbf{x} - \omega t), \quad (7.41)$$



where \mathbf{x} is any position vector, as can be appreciated from Fig. 7.2.

It follows that the fields themselves can be written

$$\begin{aligned}\mathbf{E}(\mathbf{x}, t) &= \mathbf{E}_0 \exp[i(\mathbf{k} \cdot \mathbf{x} - \omega t)] \\ \mathbf{H}(\mathbf{x}, t) &= \mathbf{H}_0 \exp[i(\mathbf{k} \cdot \mathbf{x} - \omega t)].\end{aligned}\quad (7.42)$$

Remember that we really mean the real parts of these quantities, which is vital if combining such quantities unless using the complex power! In this case, it is straightforward to show that

$$\begin{aligned}\nabla \cdot \mathbf{E} &= i\mathbf{k} \cdot \mathbf{E} \\ \nabla \times \mathbf{E} &= i\mathbf{k} \times \mathbf{E}.\end{aligned}\quad (7.43)$$

Thought for the day 20: The equalities listed in (7.43) are exceedingly useful. Prove, using suffix notation, that they are true.

Using the expressions in Eq. (7.43), Maxwell's equations can be

written as

$$\begin{aligned} M1 : \quad \mathbf{k} \cdot \mathbf{E}_0 &= 0 \\ M3 : \quad \mathbf{k} \cdot \mathbf{H}_0 &= 0 \\ M2 : \quad \mathbf{k} \times \mathbf{E}_0 &= \omega\mu_0 \mathbf{H}_0 \\ M4 : \quad \mathbf{k} \times \mathbf{H}_0 &= -\omega\epsilon_0 \mathbf{E}_0. \end{aligned} \quad (7.44)$$

The first two of these equations show that the electric and magnetic field vectors are always transverse to the direction of propagation. The second two show that \mathbf{E}_0 , \mathbf{H}_0 and \mathbf{k} form a right-handed system of vectors, as shown in Fig. 7.3.

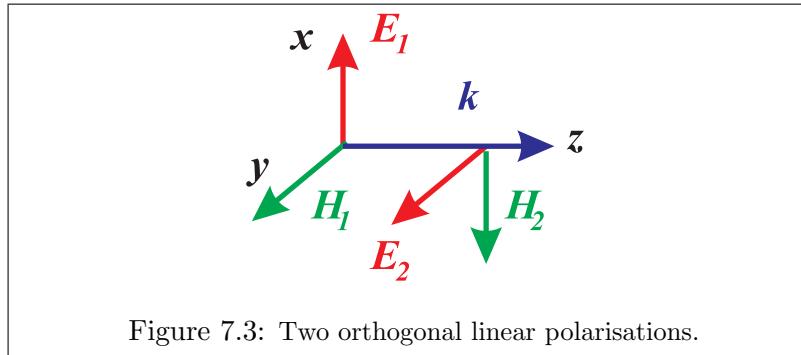


Figure 7.3: Two orthogonal linear polarisations.

We also have

$$\begin{aligned} -\omega\epsilon_0 \mathbf{E}_0 &= \mathbf{k} \times \mathbf{H}_0 \\ &= \frac{1}{\omega\mu_0} \mathbf{k} \times (\mathbf{k} \times \mathbf{E}_0) \\ &= -\frac{1}{\omega\mu_0} k^2 \mathbf{E}_0, \end{aligned} \quad (7.45)$$

where in the last line the vector triple product and $\mathbf{k} \perp \mathbf{E}_0$ has been used. It then follows that

$$\frac{\omega^2}{k^2} = \frac{1}{\epsilon_0\mu_0} = c^2, \quad (7.46)$$

as we found previously.

Finally,

$$\begin{aligned} \mathbf{H}_0 &= \frac{k}{\omega\mu_0} \hat{\mathbf{k}} \times \mathbf{E}_0 \\ &= \frac{1}{c\mu_0} \hat{\mathbf{k}} \times \mathbf{E}_0 \\ &= \frac{1}{Z_0} \hat{\mathbf{k}} \times \mathbf{E}_0, \end{aligned} \quad (7.47)$$

where $\hat{\mathbf{k}}$ is the unit vector pointing in the direction of propagation. This last expression highlights the TEM ('transverse electromagnetic') nature of the wave, and the role of the characteristic impedance of free space.

The plane-wave form

$$\mathbf{E}(\mathbf{x}, t) = \mathbf{E}_0 \exp [i(\mathbf{k} \cdot \mathbf{x} - \omega t)] \quad (7.48)$$

is of considerable importance. It can be written as

$$\mathbf{E}(\mathbf{x}, t) = \mathbf{E}_0 e^{ik_x x} e^{ik_y y} e^{ik_z z} e^{-i\omega t}. \quad (7.49)$$

In fact, we can describe any field as a weighted linear combination of plane waves, where for each temporal frequency, a continuous distribution of plane waves, travelling in different directions, is required.

Formally, the four-dimensional (3 spatial and 1 temporal) Fourier transform of a field,

$$\begin{aligned} \mathbf{E}(\mathbf{x}, t) = & \\ & \int \int \int \int dk_x dk_y dk_z d\omega \mathbf{A}_s(\mathbf{k}, \omega) e^{ik_x x} e^{ik_y y} e^{ik_z z} e^{-i\omega t}, \end{aligned} \quad (7.50)$$

represents a free-space propagating field in terms of a superposition of plane waves. However, note that the components of \mathbf{k} , i.e., k_x , k_y , and k_z , are not independent of each other. \mathbf{A}_s is called the spectral function. The plane-wave decomposition from Eq. (7.50) is the starting point of quantum optics, because each plane wave can be regarded as a simple-harmonic oscillator, which must be quantised. Indeed, it is central to the whole notion of photons in modes, and is used in the derivation of Planck's law for the black body radiation.

7.6 Energy flow

We have studied how energy can propagate in the form of plane electromagnetic waves; we might wonder if there are any general theorems relating to the flow of energy at a point regardless of the precise large-scale form of the field.

Work is done by forces when the point of application moves in the direction of the force \mathbf{F} . The force on a charge q is $\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B})$. Only the electric field does work ($\mathbf{v} \perp \mathbf{v} \times \mathbf{B}$). In moving the charge a distance $d\mathbf{l}$, the work done *by* the field *on* the charge is

$$dW = \mathbf{F} \cdot d\mathbf{l} = q\mathbf{E} \cdot d\mathbf{l}. \quad (7.51)$$

So the rate at which the field does work is $q\mathbf{E} \cdot \mathbf{v}$. For *distributed* charges, the rate at which the field does work is

$$\int d^3r \rho \mathbf{E} \cdot \mathbf{v} = \int d^3r \mathbf{E} \cdot \rho \mathbf{v} = \int d^3r \mathbf{E} \cdot \mathbf{J}. \quad (7.52)$$

This power is dissipated by scattering of these moving charges in the medium (as it has resistance), so the rate at which currents dissipate energy per unit volume is

$$\mathbf{E} \cdot \mathbf{J}. \quad (7.53)$$

To calculate the total flow of energy, we start with the important vector identity

$$\nabla \cdot (\mathbf{E} \times \mathbf{H}) = \mathbf{H} \cdot (\nabla \times \mathbf{E}) - \mathbf{E} \cdot (\nabla \times \mathbf{H}). \quad (7.54)$$

We also have, from Maxwell's equations,

$$\begin{aligned} \nabla \times \mathbf{E} &= -\mu_0 \frac{\partial \mathbf{H}}{\partial t} \\ \nabla \times \mathbf{H} &= \mathbf{J} + \epsilon_0 \frac{\partial \mathbf{E}}{\partial t}, \end{aligned} \quad (7.55)$$

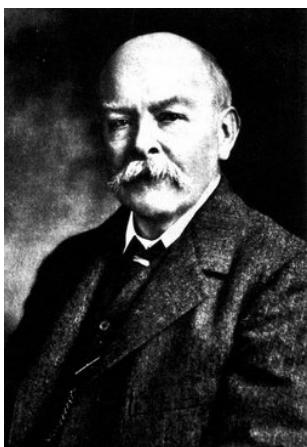
and therefore

$$\nabla \cdot (\mathbf{E} \times \mathbf{H}) = -\mu_0 \mathbf{H} \cdot \frac{\partial \mathbf{H}}{\partial t} - \epsilon_0 \mathbf{E} \cdot \frac{\partial \mathbf{E}}{\partial t} - \mathbf{E} \cdot \mathbf{J}. \quad (7.56)$$

Taking the volume integral of both sides, and using the divergence theorem,

$$\begin{aligned} & - \oint_S d\mathbf{S} \cdot (\mathbf{E} \times \mathbf{H}) \\ &= \int_V d^3 \mathbf{r} \left(\mu_0 \mathbf{H} \cdot \frac{\partial \mathbf{H}}{\partial t} + \epsilon_0 \mathbf{E} \cdot \frac{\partial \mathbf{E}}{\partial t} + \mathbf{E} \cdot \mathbf{J} \right) \\ &= \int_V d^3 \mathbf{r} \left(\frac{\partial}{\partial t} \left[\frac{1}{2} \mu_0 \mathbf{H} \cdot \mathbf{H} \right] + \frac{\partial}{\partial t} \left[\frac{1}{2} \epsilon_0 \mathbf{E} \cdot \mathbf{E} \right] + \mathbf{E} \cdot \mathbf{J} \right). \end{aligned} \quad (7.57)$$

On the right-hand side, the first term in the integral corresponds to the rate of increase of the energy stored per unit volume in the magnetic field, the second to the rate of increase of the energy stored per unit volume in the electric field, and the third to the rate of energy dissipation per unit volume. Comparing the first and the second term, we notice an important fact: for an electromagnetic wave, the energy stored in the electric field is the *same* as that stored in the magnetic field, since $|H| = \sqrt{(\epsilon_0/\mu_0)|E|}$.



John Henry Poynting
(1852–1914)

The whole of the right-hand side of Eq. (7.57) gives the rate at which energy is flowing across the surface into the volume V . Remembering that the surface integral on the left-hand side is evaluated with respect to the outward normal, the integrand on the left-hand side must give the rate at which energy is flowing across the surface per unit area. The power flow is therefore given by the **Poynting vector**

$$\mathbf{N} = \mathbf{E} \times \mathbf{H}. \quad \text{Poynting vector} \quad (7.58)$$

The magnitude of \mathbf{N} gives the power flow per unit area, and the direction of \mathbf{N} gives the direction in which the power is flowing. The Poynting vector applies at a point—it constitutes a vector field—and can be used even when magnetic and dielectric materials are present. It provides an exceedingly powerful tool for investigating energy flow in electromagnetic systems.

7.7 Poynting vector

The Poynting vector, which we derived in the previous section,

$$\mathbf{N}(\mathbf{r}) = \mathbf{E}(\mathbf{r}) \times \mathbf{H}(\mathbf{r}), \quad (7.59)$$

is a vector field that gives the direction, and energy flux, of an electromagnetic field. Notice that \mathbf{N} is not linear in the field, and therefore the superposition principle does not apply to Poynting fields.

Consider a plane electromagnetic wave travelling in the z direction; say

$$\begin{aligned}\mathbf{E} &= (E_x, 0, 0) \\ \mathbf{H} &= (0, H_y, 0),\end{aligned}\quad (7.60)$$

giving

$$\mathbf{N} = (0, 0, E_x H_y) = (0, 0, \frac{E_x^2}{Z_0}), \quad (7.61)$$

as expected.

If the material is non-conducting, but is magnetic or a dielectric, then

$$\mathbf{N} = (0, 0, E_x H_y) = (0, 0, \frac{E_x^2}{Z}), \quad (7.62)$$

where Z is simply the impedance of the medium from Eq. (7.37).

7.8 Radiation Pressure and Momentum

Radiation carries momentum as well as energy. We could calculate this classically, using the effect of the wave's electric field on electrons in a conductor, and the force exerted on the resulting current by the wave's magnetic field. However, it is easier to consider photons of energy E and use Einstein's relation $E^2 = p^2 c^2 + m_0^2 c^4$ (with rest mass $m_0 = 0$ for photons), i.e.,

$$E = pc. \quad (7.63)$$

Thus, dividing by the volume per photon (we don't care what this is since we are interested in densities), the energy density U is related to the radiation momentum density \mathbf{g} by

$$U = |\mathbf{g}|c. \quad (7.64)$$

Consider now the radiation incident normally on area A of a surface in time dt , as shown in Fig. 7.4. In terms of the Poynting vector \mathbf{N} , the energy moving through the surface in this time is

$$dE = \int_A d\mathbf{S} \cdot \mathbf{N} dt = |\mathbf{N}| A dt. \quad (7.65)$$

At the start of the time interval, this radiation occupied the volume element shown in Fig. 7.4, which has length $c dt$ and cross-sectional area A . So the energy density is given by

$$U = \frac{dE}{A c dt} = \frac{|\mathbf{N}|}{c}. \quad (7.66)$$

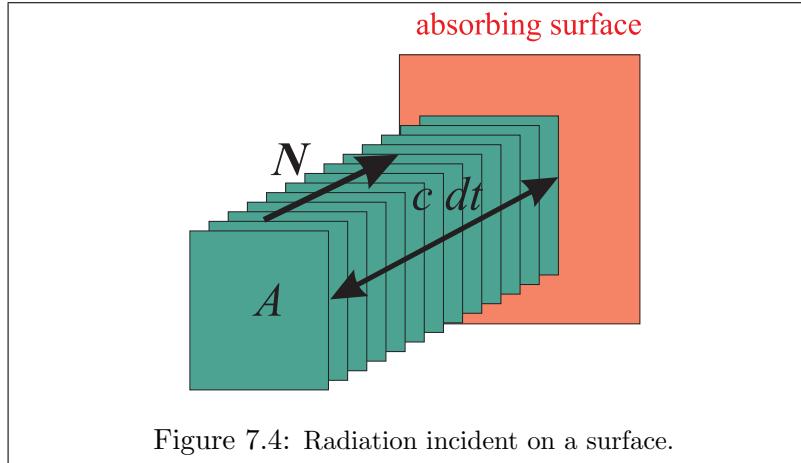


Figure 7.4: Radiation incident on a surface.

Thus, using Eq. (7.64), $U = |\mathbf{g}|c$, and noting that \mathbf{N} and \mathbf{g} will point in the same direction,

$$\mathbf{g} = \frac{\mathbf{N}}{c^2}. \quad (7.67)$$

The total momentum in the volume element shown in Fig. 7.4 is $d\mathbf{p} = \mathbf{g} A c dt$. If the radiation is completely absorbed by the surface, all the momentum is absorbed in time dt .

The radiation pressure \mathbf{R} on the surface is the rate of change of momentum per unit area:

$$\mathbf{R} = \frac{1}{A} \frac{d\mathbf{p}}{dt} = \frac{\mathbf{g} A c dt}{A dt} = \mathbf{g} c. \quad (7.68)$$

Thus, the **radiation pressure** of a wave in free space is

$$\mathbf{R} = \frac{\mathbf{N}}{c} . \quad \text{Radiation pressure} \quad (7.69)$$

If, instead, the surface *reflects* the radiation, then the pressure is *doubled*.

7.9 Complex power

In the following, we aim to find and motivate the expression for complex power. For this, consider the arrangement where some electrical component has a voltage

$$V = V_0 \cos(\omega t - \phi), \quad (7.70)$$

across it, and a current

$$I = I_0 \cos \omega t, \quad (7.71)$$

flowing through it.

The voltage can also be written

$$V = V_0 [\cos \omega t \cos \phi + \sin \omega t \sin \phi], \quad (7.72)$$

and then the instantaneous power becomes

$$\begin{aligned} P &= VI \\ &= V_0 I_0 \left[\cos \phi \cos^2 \omega t + \frac{1}{2} \sin \phi \sin 2\omega t \right]. \end{aligned} \quad (7.73)$$

This power is the sum of two terms:

- $V_0 I_0 \cos \phi \cos^2 \omega t$, which has a time-averaged value of

$$P_r = \frac{1}{2} V_0 I_0 \cos \phi, \quad (7.74)$$

and describes the average rate at which energy is transferred to the component, and

- $V_0 I_0 \frac{1}{2} \sin \phi \sin 2\omega t$, which has a time-averaged value of zero, and a peak value of

$$P_i = V_0 I_0 \frac{1}{2} \sin \phi, \quad (7.75)$$

and describes the maximum rate at which energy sloshes in and out of the component.

Now consider representing the current and voltage in terms of vectors in the complex plane:

$$\begin{aligned} I &= I_0 \exp [i\omega t] \\ V &= V_0 \exp [i(\omega t - \phi)], \end{aligned} \quad (7.76)$$

where we have used the same symbols as before for the current and voltage, even though they are now complex quantities. We are now in a position to define the **complex power**

$P = \frac{1}{2} IV^*$

Complex power

(7.77)

i.e., in our case

$$P = \frac{1}{2} I_0 V_0 \exp (i\phi), \quad (7.78)$$

which has a real part

$$P_r = \frac{1}{2} V_0 I_0 \cos \phi, \quad (7.79)$$

and an imaginary part

$$P_i = \frac{1}{2}V_0I_0 \sin \phi. \quad (7.80)$$

The real part P_r of the complex power gives the average rate at which energy flows into the component, and the imaginary part P_i gives the maximum rate at which energy sloshes in and out of the component. *Power flow* is thus represented as a vector in the complex plane, which is distinct from the complex representation of the current and voltage.

As one might imagine from the physical interpretation, if two systems are connected together, the complex power flow associated with the combination is the vector sum of the complex powers associated with the individual components, regardless of whether the systems are connected in parallel or in series.

Engineers call $\frac{1}{2}V_0I_0$ the **apparent power** because it has no obvious physical meaning, $\frac{1}{2}V_0I_0 \cos \phi$ the **real power**, because it gives the rate at which energy is lost, and $\frac{1}{2}V_0I_0 \sin \phi$ the **reactive power**, because it gives the maximum rate at which energy sloshes around. Domestic electricity meters are particularly ingenious because they only charge you for the energy you use, not for the energy you borrow, say to store in the magnetic field of a washing machine's motor. (Solar panels on homes feed significant amounts of energy into the national grid, though this does not make the meter run backwards!)

In the context of the Poynting vector, when the fields are real time-varying quantities,

$$|\mathbf{N}(\mathbf{r})| = |\mathbf{E}(\mathbf{r}) \times \mathbf{H}(\mathbf{r})| \quad (7.81)$$

gives the *instantaneous power flow per unit area* at a given point. When the fields are represented as complex vectors,

$$\boxed{\left| \frac{1}{2} \operatorname{Re} [\mathbf{E}(\mathbf{r}) \times \mathbf{H}(\mathbf{r})^*] \right|} \quad (7.82)$$

gives the time-averaged power flow per unit area at a point, where the modulus sign denotes the magnitude of the enclosed vector (in a spatial sense), and

$$\boxed{\left| \frac{1}{2} \operatorname{Im} [\mathbf{E}(\mathbf{r}) \times \mathbf{H}(\mathbf{r})^*] \right|} \quad (7.83)$$

gives the maximum rate at which energy is sloshing backwards and forwards per unit area at a point.

In the case of plane waves, and TEM transmission lines (see later), the magnetic and electric fields are in phase, the complex power is a real quantity, and energy only flows in one direction; it does not slosh backwards and forwards. There are many electromagnetic fields, however, for which this is not the case, and energy sloshes backwards and forwards across a surface.

7.10 Reflection and transmission at interfaces

Generally, it is quite complicated to calculate the form of the travelling electromagnetic waves that exist when dielectric or magnetic bodies are present, and sophisticated analysis techniques have to be used. An important problem that can be analysed, however, is when a plane wave is incident on the boundary between two dielectric materials that have different refractive indices.

Consider the situation shown in Fig. 7.5. A plane wave having wave vector \mathbf{k}_i is incident on a plane dielectric boundary. A transmitted

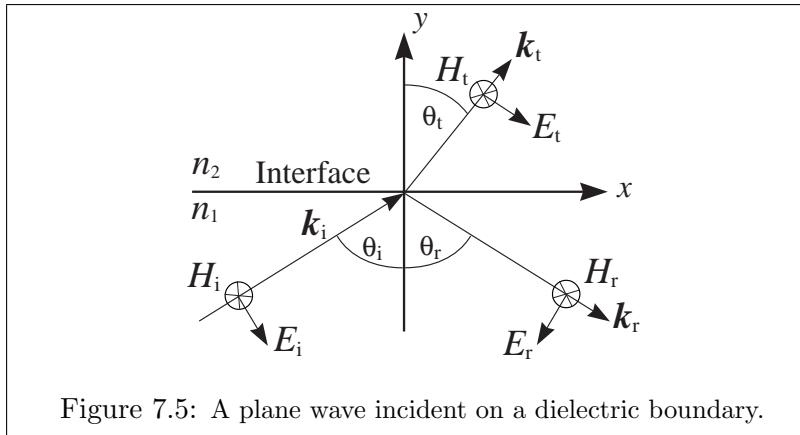


Figure 7.5: A plane wave incident on a dielectric boundary.

wave, \mathbf{k}_t , and a reflected wave, \mathbf{k}_r , are produced, and we cannot make any assumptions about their directions of travel. The angles of incidence, reflection and transmission have been marked as θ_i , θ_r , and θ_t respectively.

In this example, we have chosen the incident wave to be polarised in the **plane of incidence** (the plane containing both the incident wavevector and the normal to the interface). Without loss of generality, choose coordinate system such that z is perpendicular to the plane of incidence, so there is no z dependence for any of the fields, and so we are justified in drawing the diagram as shown.

The incident, reflected, and transmitted plane waves have the functional forms

$$\begin{aligned} \mathbf{E}_i &= \mathbf{E}_{i0} \exp [i(\mathbf{k}_i \cdot \mathbf{r} - \omega_i t)] \\ \mathbf{E}_r &= \mathbf{E}_{r0} \exp [i(\mathbf{k}_r \cdot \mathbf{r} - \omega_r t)] \\ \mathbf{E}_t &= \mathbf{E}_{t0} \exp [i(\mathbf{k}_t \cdot \mathbf{r} - \omega_t t)], \end{aligned} \quad (7.84)$$

where we have not even assumed that the transmitted and reflected waves have the same temporal frequencies as the incident field.

In the following, we want to draw up relationships between ω , \mathbf{k} and θ of the incident, reflected, and transmitted waves. For each wave (using $k \equiv |\mathbf{k}|$)

$$k_x = k \sin \theta, \quad (7.85)$$

and, remembering that on the x -axis, where $y = 0$, the component of the electric field *parallel* to the surface $|\mathbf{E}_{\parallel}| = E_x$ must be continuous across the boundary:

$$E_{i,x} + E_{r,x} = E_{t,x} \quad (7.86)$$

i.e.,

$$\begin{aligned} & E_{i0} \exp [i(k_i x \sin \theta_i - \omega_i t)] \cos \theta_i \\ - & E_{r0} \exp [i(k_r x \sin \theta_r - \omega_r t)] \cos \theta_r \\ = & E_{t0} \exp [i(k_t x \sin \theta_t - \omega_t t)] \cos \theta_t. \end{aligned} \quad (7.87)$$

Eq. (7.87) must be true for all x and t , and therefore

$$\omega_i = \omega_r = \omega_t, \quad (7.88)$$

and

$$k_i x \sin \theta_i = k_r x \sin \theta_r = k_t x \sin \theta_t. \quad (7.89)$$

The above procedure is known as *phase matching*. In other words, the fields must have the same temporal, and x -directed spatial, frequencies.

We also know that

$$k = \frac{n\omega}{c}, \quad (7.90)$$

where n is the refractive index in that region, and therefore

$$\begin{aligned} k_1 &\equiv k_i = k_r \\ k_2 &\equiv k_t, \end{aligned} \quad (7.91)$$

with

$$\frac{k_1}{k_2} = \frac{n_1}{n_2}. \quad (7.92)$$

We conclude that, because of Eqs. (7.91) and (7.89),

$$\theta_i = \theta_r, \quad (7.93)$$

and we find **Snell's law**, $n \sin \theta = \text{constant}$:

$$\frac{\sin \theta_t}{\sin \theta_i} = \frac{k_1}{k_2} = \frac{n_1}{n_2}, \quad \text{Snell's law}$$

(7.94)

The same analysis can be applied to a linearly polarised wave with the electric field perpendicular to the plane of incidence—there are just no $\cos \theta$ factors in Eq. (7.87). The same phase-matching requirements apply, and a similar set of expressions is derived. In particular, Snell's law still holds, as would be expected.

7.11 Reflection and transmission coefficients

We have established the relationship between the angles of incidence and reflection, but we have said nothing about the *amount* of power that is reflected. This will be discussed in this Section.

Because the arguments of the exponentials in Eq. 7.87 are always equal,

$$\begin{aligned} E_{i0} \cos \theta_i - E_{r0} \cos \theta_r &= E_{t0} \cos \theta_t \\ \Rightarrow (E_{i0} - E_{r0}) \cos \theta_i &= E_{t0} \cos \theta_t, \end{aligned} \quad (7.95)$$

where the last line follows because the angles of incidence and reflection are equal. If we match the parallel component of the \mathbf{H} field in an identical way, we find

$$\begin{aligned} H_{i0} + H_{r0} &= H_{t0} \\ \Rightarrow n_1 E_{i0} + n_1 E_{r0} &\stackrel{E/H = Z_0/n}{=} n_2 E_{t0}, \end{aligned} \quad (7.96)$$

where the second line follows because we always have $\frac{E}{H} = \frac{Z_0}{n}$. It follows that

$$(E_{i0} - E_{r0}) \cos \theta_i = \frac{n_1}{n_2} (E_{i0} + E_{r0}) \cos \theta_t, \quad (7.97)$$

which can be rearranged to give

$$\frac{E_{r0}}{E_{i0}} = \frac{(n_2/n_1) \cos \theta_i - \cos \theta_t}{(n_2/n_1) \cos \theta_i + \cos \theta_t} = \frac{\tan(\theta_i - \theta_t)}{\tan(\theta_i + \theta_t)}, \quad (7.98)$$

where Snell's law has been used for the last step.

We conclude that the reflection coefficient for a parallel-polarised plane wave is

$$r_{\parallel} = \frac{E_{r0}}{E_{i0}} = \frac{\tan(\theta_i - \theta_t)}{\tan(\theta_i + \theta_t)}. \quad (7.99)$$

In a similar way, it can be shown that the transmission coefficient is given by

$$t_{\parallel} = \frac{E_{t0}}{E_{i0}} = \frac{2 \cos \theta_i}{(n_2/n_1) \cos \theta_i + \cos \theta_t}. \quad (7.100)$$

An entirely equivalent procedure can be carried out when the plane of polarisation is perpendicular to the plane of incidence, giving a reflection coefficient of

$$r_{\perp} = \frac{E_{r0}}{E_{i0}} = -\frac{\sin(\theta_i - \theta_t)}{\sin(\theta_i + \theta_t)}, \quad (7.101)$$

and a transmission coefficient of

$$t_{\perp} = \frac{E_{t0}}{E_{i0}} = \frac{2 \cos \theta_i}{\cos \theta_i + (n_2/n_1) \cos \theta_t}. \quad (7.102)$$

These four equations are known as **Fresnel's relations**.

To illustrate the usefulness of the coefficients above, consider an example where $n_1 = 1$ and $n_2 = n$. Then at normal incidence, $\theta_i = \theta_r = \theta_t = 0$, so, using small-angle formulae as $\theta_i \rightarrow 0$ in the first part of Eq. (7.98) or in Eq. (7.99), the *power* reflection coefficient becomes

$$R_{\parallel} = |r_{\parallel}|^2 = \left(\frac{n-1}{n+1} \right)^2 = |r_{\perp}|^2 = R_{\perp}. \quad (7.103)$$

In the case of glass, $n = 1.5$, and so $R = 4\%$.

As the angle of incidence and the polarisation are changed, the behaviour becomes quite complicated. The Fresnel equations have been used to produce the plots shown in Fig. 7.6. Notice that the

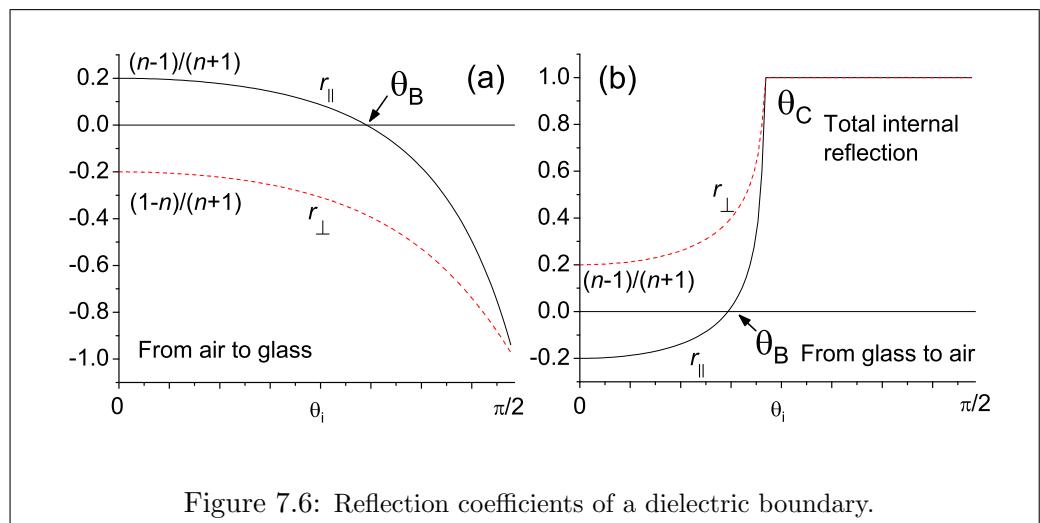


Figure 7.6: Reflection coefficients of a dielectric boundary.

reflection coefficients can take on negative values, showing that the fields can be 180° out of phase, i.e., $E_{r0} = -E_{i0}$.

An analysis of Fig. 7.6 shows that there is a particular angle of incidence where $r_{\parallel} = 0$. This angle is called the **Brewster angle**, and is given by

$$\tan \theta_B = \frac{n_2}{n_1} \quad \text{Brewster angle} \quad (7.104)$$

where the wave is incident in the material with index n_1 and transmitted into the material with index n_2 . If going from free space, the Brewster angle is simply

$$\tan \theta_B = n_2$$

The Brewster angle depends on whether the wave is travelling from air into glass or glass into air. Brewster-angle windows are often used to eliminate reflections from optical windows, as shown in Fig. 7.7, or to produce linearly polarised light from unpolarised light. Hence, they are often found in experimental optical setups.



Figure 7.7: Photo of a window viewed near the Brewster angle, with a camera polariser filter aligned (left) vertically (\mathbf{E} perpendicular to the plane of incidence), and (right) horizontally (\mathbf{E} parallel to the plane of incidence, eliminating the reflection) (from *Wikipedia*).

For incidence from within the material, $n_1 > n_2$, so there is some critical angle θ_c such that when $\theta_i > \theta_c$ a problem occurs:

$$\sin \theta_t = \frac{n_1}{n_2} \sin \theta_i > 1. \quad (7.105)$$

There is no (real) solution to this equation, and hence all of the incident power is reflected, regardless of the polarisation of the wave! The **critical angle of total internal reflection** is given by

$$\boxed{\sin \theta_c = \frac{n_2}{n_1}. \quad \text{Critical angle}} \quad (7.106)$$

Once the critical angle of incidence is exceeded, an **evanescent** wave is produced that decays exponentially from the surface and travels along the surface. In Cambridge practicals, this is studied using a wax block and microwaves. Critical angles and polarisation properties are now being used extensively in the field of surface-plasmon physics.

Thought for the day 21: Fresnel's equations predict that there is some critical angle beyond which all power is reflected. Can you understand the physics of why this occurs? Concentrate on what is required of the wave vector of the transmitted wave in order for propagation to be possible. Can you give an example of where total internal reflection is seen on a daily basis?

7.12 Waves in plasmas

Up to this point we have considered the case where the medium in which the wave is travelling does not have conduction currents, but there are numerous examples of where conduction currents are present. We will start with plasmas, and then look at metals.

A plasma is a region of space where free electrons and their parent ions are present. The mass of an electron is much less than that of

an ion, and so they are more mobile than ions. In this analysis, we shall ignore the movement of ions.

We know that the equation of motion for an electron in an electromagnetic field is

$$m_e \frac{d^2 \mathbf{r}}{dt^2} = -e (\mathbf{E} + \mathbf{v} \times \mathbf{B}), \quad (7.107)$$

where \mathbf{r} is the electron's position, and \mathbf{v} its velocity.

Suppose that we illuminate a free electron with a plane wave, then

$$\frac{E_x}{H_y} = \sqrt{\frac{\mu_0}{\epsilon_0}}, \quad (7.108)$$

i.e.,

$$\frac{E_x}{B_y} = c. \quad (7.109)$$

The cross-product term in Eq. (7.107) will have its greatest magnitude when \mathbf{v} and \mathbf{H} are at right angles, but when $|\mathbf{v}| \ll c$, it can be ignored, compared to the \mathbf{E} -field term. Let us suppose that the temporal variation of the wave is represented as a vector in the complex plane:

$$\mathbf{E} = \mathbf{E}_0 \exp [i(kz - \omega t)]. \quad (7.110)$$

It then follows that

$$\mathbf{r} = \frac{e}{m_e \omega^2} \mathbf{E}_0 \exp [i(kz - \omega t)], \quad (7.111)$$

which can be verified by substitution into Eq. (7.107).

Equation (7.111) indicates that the position of the electron oscillates with an amplitude that is inversely proportional to the mass and the squared frequency of the incident radiation. The inertial mass of the electron "regulates" the amplitude of the oscillation as the frequency increases.

In a plasma, we have to take into account the ions, in the sense that as the electrons move, they separate from the ions, inducing dipoles. The dipole moment of a single separating pair is

$$\mathbf{p} = -e\mathbf{r} = -\frac{e^2}{m_e \omega^2} \mathbf{E}, \quad (7.112)$$

and the dipole moment per unit volume is

$$\mathbf{P} = N\mathbf{p} = -\frac{Ne^2}{m_e \omega^2} \mathbf{E}, \quad (7.113)$$

where N is the number of electrons per unit volume, and it is assumed that they are non-interacting.

Recall that χ is defined through

$$\mathbf{P} = \epsilon_0 \chi \mathbf{E}, \quad (7.114)$$

and therefore

$$\epsilon = 1 + \chi = 1 - \frac{Ne^2}{\epsilon_0 m_e \omega^2}, \quad (7.115)$$

the **relative permittivity in plasma** is usually written as

$$\epsilon = 1 - \frac{\omega_p^2}{\omega^2}, \quad \text{Relative permittivity in plasma} \quad (7.116)$$

where the **plasma frequency** is defined as

$$\omega_p^2 = \frac{Ne^2}{\epsilon_0 m_e}. \quad \text{Plasma frequency} \quad (7.117)$$

The plasma frequency ω_p characterises the electromagnetic properties of a plasma. The plasma frequency can be expressed in the convenient form:

$$f_p = \frac{\omega_p}{2\pi} = \frac{1}{2\pi} \sqrt{\frac{Ne^2}{\epsilon_0 m_e}} \approx 9\sqrt{N/m^{-3}} \text{ Hz.} \quad (7.118)$$

Taking the ionosphere as a concrete example, $N \approx 10^{12} \text{ m}^{-3}$, and therefore $f_p \approx 10 \text{ MHz}$. Thus, at frequencies below 10 MHz, electromagnetic waves are reflected off the ionosphere, enabling low-frequency communications, as indicated in Fig. 7.8.

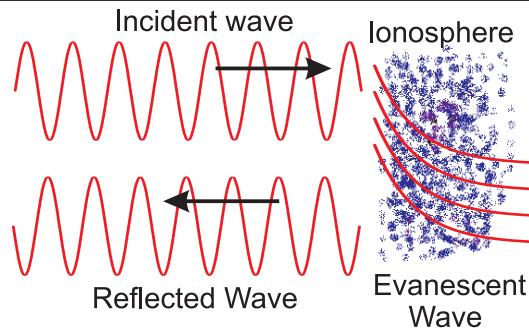


Figure 7.8: Reflection from the ionosphere, below the plasma frequency.

The relative permittivity of a plasma ϵ is shown as a function of oscillation frequency in Fig. 7.9. The refractive index becomes a function of the oscillation frequency, too:

$$n = \sqrt{\epsilon} = \left(1 - \frac{\omega_p^2}{\omega^2}\right)^{1/2}. \quad (7.119)$$

The refractive index is imaginary for frequencies below the plasma frequency, and we expect an evanescent wave.

Let us study the case where $\omega < \omega_p$ in more detail. Since the refractive index is imaginary, it is convenient to define

$$n = i\beta, \quad (7.120)$$

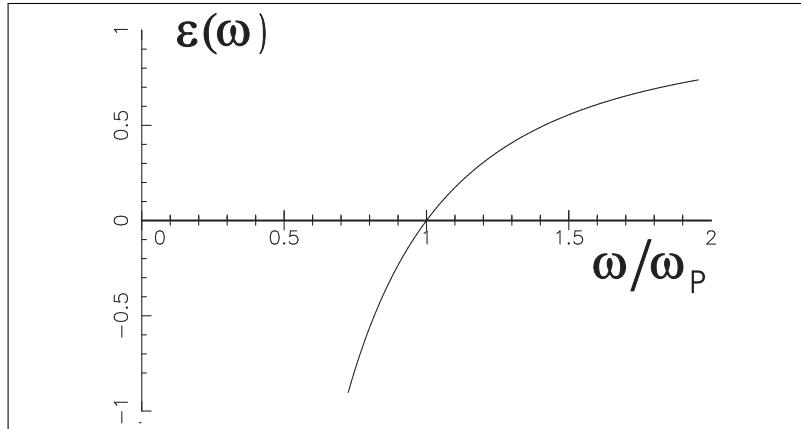


Figure 7.9: The effective dielectric constant of a plasma.

where

$$\beta = \sqrt{|\epsilon|}. \quad (7.121)$$

Then

$$k = \frac{n\omega}{c} = \frac{i\beta\omega}{c}, \quad (7.122)$$

which can be substituted into Eq. (7.110) to give

$$\mathbf{E} = \mathbf{E}_0 \exp \left[-\frac{\omega\beta}{c} z \right] \exp [-i\omega t]. \quad (7.123)$$

The resulting \mathbf{E} field no longer takes the form of a travelling wave, but decays with z , as shown in Fig. 7.10.

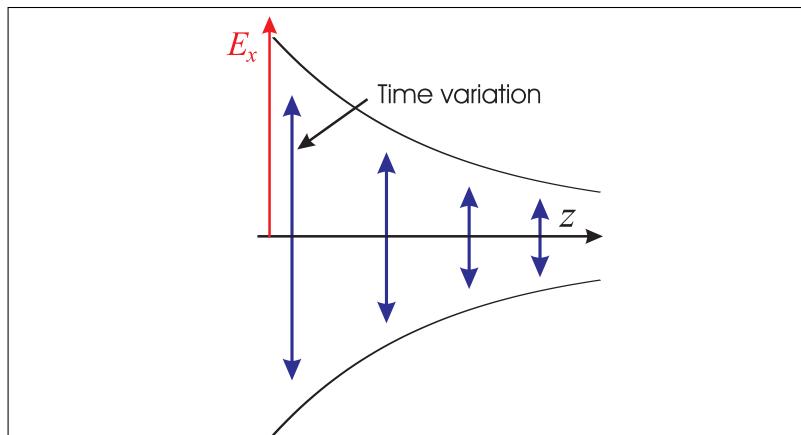


Figure 7.10: The electric field of an electromagnetic wave decays in a plasma below cut-off.

The expression in Eq. (7.123) can also be written explicitly in terms of ω_p using Eq. (7.119) for $\omega < \omega_p$,

$$\mathbf{E} = \mathbf{E}_0 \exp \left[-k_0 z \left| 1 - \frac{\omega_p^2}{\omega^2} \right|^{1/2} \right] \exp [-i\omega t], \quad (7.124)$$

where $k_0 = \omega/c$. The lower the frequency, the shorter the distance over which the field decays. Notice that the wave is *not* a decaying propagating wave—it does not propagate at all. A non-propagating, decaying wave of this kind is called an **evanescent wave**.

It is also useful to calculate the magnetic field, which is given by

$$H_y = \sqrt{\frac{\epsilon\epsilon_0}{\mu_0}} E_x = \frac{i\beta}{Z_0} E_x. \quad (7.125)$$

The electric and magnetic fields are $\pi/2$ out of phase, as shown in Fig. 7.11. In other words, the surface impedance of the plasma is reactive.

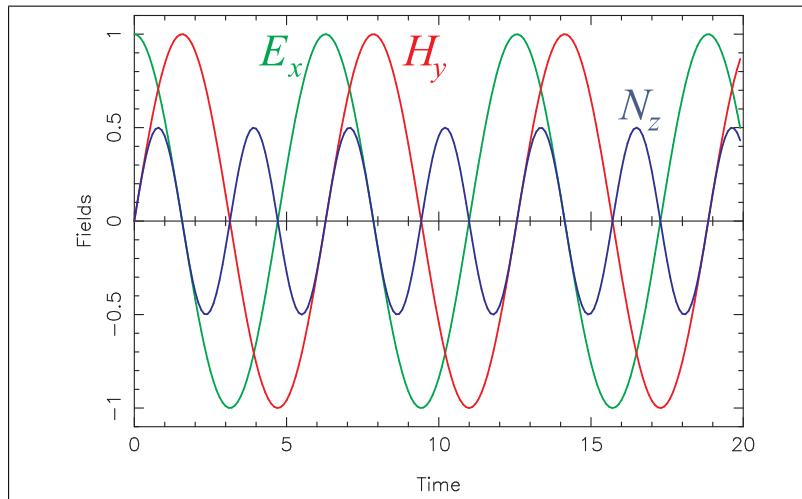


Figure 7.11: The time variation of the electric and magnetic fields in a plasma below cut-off. The sign of the Poynting vector alternates, indicating the sloshing of energy.

The average and instantaneous magnitudes of the Poynting vector become respectively

$$\left| \frac{1}{2} \operatorname{Re} [\mathbf{E}(\mathbf{r}) \times \mathbf{H}(\mathbf{r})^*] \right| = 0 \quad \text{and} \quad \left| \frac{1}{2} \operatorname{Im} [\mathbf{E}(\mathbf{r}) \times \mathbf{H}(\mathbf{r})^*] \right| = \frac{\beta}{2Z_0} E_x^2. \quad (7.126)$$

Thus, there is no net transfer of energy within the plasma, but energy sloshes backwards and forwards in the z direction at any point. This is effectively caused by the energy being stored in the motion of the oscillating electrons. All of the energy in the original wave that is incident on the plasma must, on average, be reflected.

Finally, we can look at propagation velocities. When the frequency of the incident radiation is above the plasma frequency, the phase velocity becomes

$$v = \frac{\omega}{k} = \frac{c}{\sqrt{1 - \omega_p^2/\omega^2}}. \quad (7.127)$$

Given that v explicitly depends on ω , the waves are **dispersive**, i.e., different frequencies travel with different velocities. As a result,

the amplitude envelope of a wave packet will in general change as the packet propagates. Notice that the phase velocity is greater than the speed of light.

Thought for the day 22: Plot the dispersion relation of a plasma, Eq. 7.127, and comment on the salient features.

It is straightforward to show that the group velocity is given by

$$v_g = \frac{d\omega}{dk} = c\sqrt{1 - \omega_p^2/\omega^2}, \quad (7.128)$$

which is less than c , as expected. It is interesting to observe that the product of the phase and group velocities is c^2 . This general behaviour is also seen in many other structures, such as waveguides.

7.13 Waves in conducting media

In a plasma, we have assumed the density of electrons (and their parent atoms) is low enough that electrons do not collide with each other during one or more cycles of the incoming EM radiation. In a metal, electrons are separated from their parent atoms as in a plasma, but their density is much higher, and so scattering becomes significant. We will now investigate what happens if we attempt to pass an electromagnetic wave through a sheet of metal.

Remember that for all materials we have the constitutive relations

$$\begin{aligned} \mathbf{D} &= \epsilon\epsilon_0\mathbf{E} \\ \mathbf{B} &= \mu\mu_0\mathbf{H} \\ \mathbf{J} &= \sigma\mathbf{E}. \quad [\text{Ohm's law}] \end{aligned} \quad (7.129)$$

Using the last of Maxwell's equations, we find

$$\begin{aligned} \nabla \times \mathbf{H} &= \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} \\ &= \sigma\mathbf{E} + \epsilon\epsilon_0 \frac{\partial \mathbf{E}}{\partial t}. \end{aligned} \quad (7.130)$$

If we now assume solutions of the form

$$\begin{aligned} \mathbf{E} &= \mathbf{E}_0 \exp[-i\omega t] \\ \mathbf{H} &= \mathbf{H}_0 \exp[-i\omega t] \end{aligned} \quad (7.131)$$

we find

$$\begin{aligned} \nabla \times \mathbf{H} &= (\sigma - i\omega\epsilon\epsilon_0)\mathbf{E} \\ &= -i\omega\epsilon_0 \left(\epsilon + \frac{i\sigma}{\omega\epsilon_0} \right) \mathbf{E}. \end{aligned} \quad (7.132)$$

An insulating material with dielectric constant ϵ' would give

$$\nabla \times \mathbf{H} = -i\omega\epsilon'\epsilon_0\mathbf{E}, \quad (7.133)$$

and, therefore, for a conducting material, we can define an **effective dielectric constant**:

$$\epsilon' \equiv \left(\epsilon + \frac{i\sigma}{\omega\epsilon_0} \right). \quad \text{Effective dielectric constant} \quad (7.134)$$

The effects of the current are contained within the effective dielectric constant ϵ' . In fact, the effective dielectric constant is a complex quantity. It can now be seen that a material can be considered to behave as a dielectric if the real part of the complex dielectric constant dominates, whereas it behaves as a metal if the imaginary part of the complex dielectric constant dominates. For example, in the case of copper, $\sigma = 5 \times 10^7 \Omega^{-1} \text{ m}^{-1}$, operating up to optical frequencies ($\nu = 10^{15} \text{ Hz}$), we find that the real part of the complex dielectric constant is negligible (about 5000 times smaller than the imaginary part). For metals, we therefore have

$$\epsilon' \approx \frac{i\sigma}{\omega\epsilon_0}. \quad (7.135)$$

Usually, when the dielectric constant appears in the solution of a field problem it enters by way of its square root, but now the effective dielectric constant is a complex quantity. The refractive index is therefore also a complex quantity. For a metal,

$$\begin{aligned} n &= \sqrt{\epsilon'\mu} \\ &= \sqrt{\frac{i\sigma\mu}{\omega\epsilon_0}} \\ &= \pm \frac{(1+i)}{\sqrt{2}} \sqrt{\frac{\sigma\mu}{\omega\epsilon_0}}, \end{aligned} \quad (7.136)$$

as can be appreciated by inspecting Fig. 7.12.

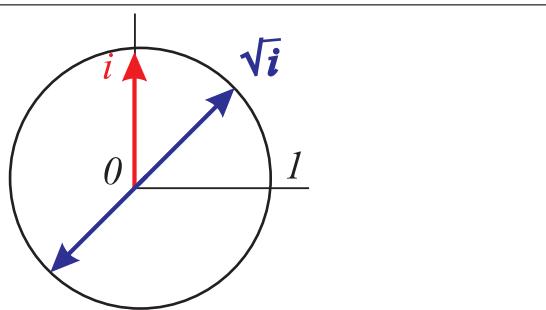


Figure 7.12: The square roots of i .

We have already shown that Maxwell's equations can be solved to give plane-wave solutions of the form

$$\begin{aligned} \mathbf{E} &= \mathbf{E}_0 \exp [i(kz - \omega t)] \\ \mathbf{H} &= \mathbf{H}_0 \exp [i(kz - \omega t)], \end{aligned} \quad (7.137)$$

where

$$k = \frac{\omega}{c/n}. \quad (7.138)$$

In the case of a metal,

$$k = \frac{\omega}{c} (1 + i) \sqrt{\frac{\sigma\mu}{2\omega\epsilon_0}} = (1 + i) \sqrt{\frac{\sigma\omega\mu_0\mu}{2}} = \frac{1 + i}{\delta}, \quad (7.139)$$

where the **skin depth** of the material is defined as

$\delta \equiv \sqrt{\frac{2}{\sigma\omega\mu_0\mu}}$

Skin depth

(7.140)

Substituting the wave number k into Eq. (7.137) gives

$$\mathbf{E} = \mathbf{E}_0 \exp\left[-\frac{z}{\delta}\right] \exp\left[i\left(\frac{z}{\delta} - \omega t\right)\right]. \quad (7.141)$$

This new expression has a number of features: firstly, it represents a travelling wave with ‘wave number’ $1/\delta$, and secondly, the amplitude of the wave decays on propagation. Notice that if we were to use the other solution for \sqrt{i} we would have a solution whose amplitude increases indefinitely with propagation, which is not physical. In fact, the amplitude of the field has decayed by e^{-1} after a distance δ . The attenuation in z is very severe—the wave decays by a factor of $e^{-2\pi} \approx 1/535$ in every wavelength ($\lambda/\delta = 2\pi$). This is shown in Fig. 7.13.

The skin depth in metals is usually very small, and so electromagnetic waves decay very rapidly once they enter a highly conductive material. For example, in the case of copper, $\sigma = 5 \times 10^7 \Omega^{-1}m^{-1}$, at 100 MHz the skin depth is nearly $7 \mu m$, which is a tiny fraction of the free-space wavelength, 3 m. Even though the attenuation in a metal is very strong, if the metal is thin enough, some transmission is possible. In other words, electromagnetic waves can penetrate into metals, but they are attenuated severely.

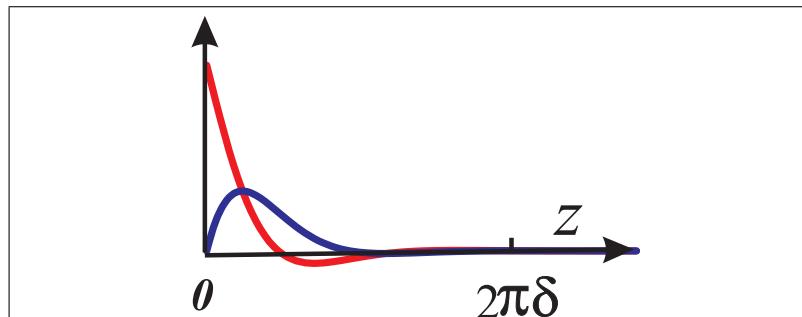


Figure 7.13: The propagation of a field into a good conductor. The lighter, red curve corresponds to time $t = 0$, and the darker, blue curve to $\omega t = \pi/2$.

It is also worth looking at the relationship between the \mathbf{E} and the \mathbf{H} fields. If the wave is polarised in the x -direction, then

$$H_y = \sqrt{\frac{\epsilon_0\epsilon}{\mu_0\mu}} E_x = \sqrt{\frac{i\sigma}{\omega\mu\mu_0}} E_x = \sqrt{\frac{\sigma}{2\omega\mu\mu_0}} (1 + i) E_x. \quad (7.142)$$

A new feature has emerged: the \mathbf{E} and \mathbf{H} fields are 45° out of phase. Fig. 7.14 shows the fields inside the surface of a good conductor as

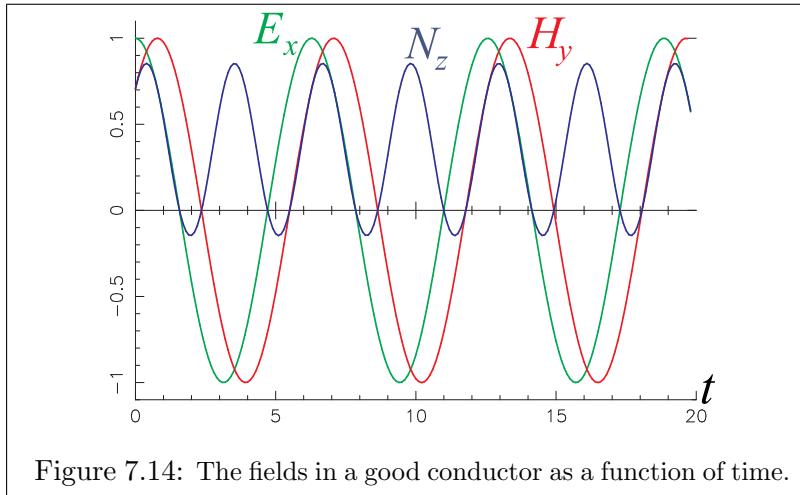


Figure 7.14: The fields in a good conductor as a function of time.

a function of time. Notice that H_y lags behind E_x .

Figure 7.14 also shows the instantaneous value of the Poynting vector ($E_x H_y$), which points in the z direction. The Poynting vector displays a number of features: firstly, it takes on positive and negative values, indicating that power can be travelling backwards at particular times in the cycle; secondly, on average it travels forwards, showing that power must be dissipated in the conductor. Power dissipation is caused by ohmic loss within the material.

7.14 The skin effect

The partial propagation of a field into a good conductor has particular significance for the flow of current in a wire, which we will address in this Section.

Consider a wire carrying a current I that oscillates at frequency ω . Power flows along the wire, and so the largest component of the Poynting vector points in the same direction as the wire. We also know, however, that because of the current flow on the surface, and the finite conductivity of the metal, there is a non-zero component of \mathbf{E} on the surface. When the Poynting vector is calculated using this \mathbf{E} field and the \mathbf{H} field associated with the current flow, the result points perpendicularly into the surface of the wire, showing that energy flows *into* the surface.

Approximating the surface of the wire as a plane surface as shown in Fig. 7.15, with $z \approx a - r$, where $a \gg \delta$ is the radius of the wire, δ is the skin depth, and the x -axis is along the wire (parallel to the current), we know that

$$E_x = E_0 \exp\left[-\frac{z}{\delta}\right] \exp\left[i\left(\frac{z}{\delta} - \omega t\right)\right], \quad (7.143)$$



Figure 7.15: High-frequency currents flow on the surface of a wire.

and therefore, using Ohm's law $\mathbf{J} = \sigma \mathbf{E}$,

$$J_x = J_0 \exp\left[-\frac{z}{\delta}\right] \exp\left[i\left(\frac{z}{\delta} - \omega t\right)\right]. \quad (7.144)$$

In other words, the amplitude of the current decays away from the

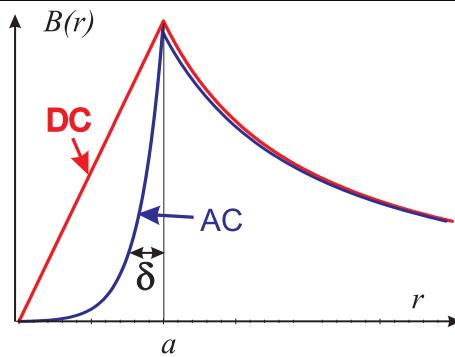


Figure 7.16: The magnetic field inside and outside a current-carrying wire.

surface of the wire. Consequently, the resulting magnetic field B decays rapidly too, as shown in Fig. 7.16. The oscillating currents are confined to the surface of the wire, and the skin depth gets smaller as the frequency increases. The electromagnetic behaviour of a (cylindrical) wire can be analysed rigorously using cylindrical basis functions, and essentially the same result is found.

With the above toolset it is now possible to calculate the resistance of a wire at high frequencies; as the frequency increases, the resistance increases due to the current being confined to a smaller and smaller region. When the skin depth is much smaller than a , we can approximate the circular case by 'unwrapping' the shell in which the current flows. The total current in the wire is given by

$$I = \int dS J_x \approx 2\pi a \int dz J_x(z). \quad (7.145)$$

Taking $z = 0$ on the surface, and letting the other limit tend to ∞ for simplicity since $\delta \ll a$, we get

$$I \approx 2\pi a J_0 \exp[-i\omega t] \int_0^\infty dz \exp\left[\frac{z}{\delta}(-1 + i)\right], \quad (7.146)$$

which evaluates to

$$I \approx 2\pi a J_0 \exp[-i\omega t] \frac{\delta}{1-i} \frac{1+i}{1+i}, \quad (7.147)$$

or

$$I \approx \pi a J_0 \delta (1+i) \exp[-i\omega t]. \quad (7.148)$$

The actual physical current can be derived from its complex representation through

$$\begin{aligned} I(t) = \operatorname{Re}[I] &= \operatorname{Re}[\pi a J_0 \delta (1+i) \exp[-i\omega t]] \quad (7.149) \\ &= \pi a J_0 \delta (\cos \omega t + i \sin \omega t), \end{aligned}$$

which has a mean square value of

$$\langle I(t)^2 \rangle = (\pi a J_0 \delta)^2. \quad (7.150)$$

The actual physical current density can also be calculated from its complex representation to give

$$J(t) = \operatorname{Re}[J_x] = J_0 \exp\left[-\frac{z}{\delta}\right] \cos\left(\frac{z}{\delta} - \omega t\right), \quad (7.151)$$

which has a mean square value of

$$\langle J(t)^2 \rangle = \frac{J_0^2}{2} \exp\left[-\frac{2z}{\delta}\right]. \quad (7.152)$$

The power dissipated per unit length of the wire in an elemental annulus of the cross-section, of infinitesimally small area $dA = 2\pi adz$, is

$$dP = \frac{I^2 R}{L} = J^2 dA^2 \left(\frac{L}{\sigma dA} \right) \frac{1}{L} = \frac{J^2 dA}{\sigma}. \quad (7.153)$$

The power dissipated per unit length of the wire in the whole cross-section of the wire is therefore

$$P = \frac{J_0^2}{2\sigma} 2\pi a \int_0^\infty dz \exp\left[-\frac{2z}{\delta}\right] = \frac{J_0^2 \pi a \delta}{2\sigma}. \quad (7.154)$$

Finally, we can define an effective resistance by requiring that the actual dissipated power is given when the total current and the effective resistance are used to calculate the power. The effective resistance, R , per unit length is therefore

$$R = \frac{P}{\langle I(t)^2 \rangle} = \frac{1}{2\pi a \delta \sigma}. \quad (7.155)$$

In other words, **from the point of view of power dissipation, the resistance per unit length is simply the resistance that is calculated when the current is assumed to flow uniformly in a thin shell of thickness δ** . Thus ordinary wires become very lossy at short wavelengths, and other ways of guiding electromagnetic energy must be found.

7.15 Metals vs plasmas

The propagation of electromagnetic waves in metals is not dissimilar to the propagation of electromagnetic waves in plasmas. However, there are some differences:

- In metals, the electrons tend to be *scattered* on time-scales less than one oscillation cycle, leading to power dissipation.
- In plasmas, all of the electrons move together, and the movement is undamped. Below the plasma frequency, power is reflected from the plasma, rather than being absorbed.

At high frequencies, when the electrons move only a short distance in each cycle, collective motion occurs, and the plasma-frequency concept becomes appropriate, even for a metal.

It can be seen that, in general terms, the surface impedance of a metal will have a real part, corresponding to power absorption, and an imaginary part, corresponding to the energy stored in the motion of the electrons.

Superconductors have a particularly high kinetic inductance term due to undamped motion of Cooper pairs. The behaviour of superconducting transmission lines is modified significantly by this kinetic-inductance effect.

8. Guided waves

8.1 Introduction

We have seen that wires become exceedingly lossy when used at high frequencies.

If the length of the wire or the dimension of the circuit d is comparable to or larger than a given wavelength, the assumption that the voltage and current are constant along the wire breaks down. This is illustrated in units of d/λ in Fig. 8.1. Fortunately, there are many other ways of guiding electromagnetic fields. The treatment of transition lines ($\lambda \approx d$) and wave guides ($\lambda \ll d$) is the subject of this Chapter.

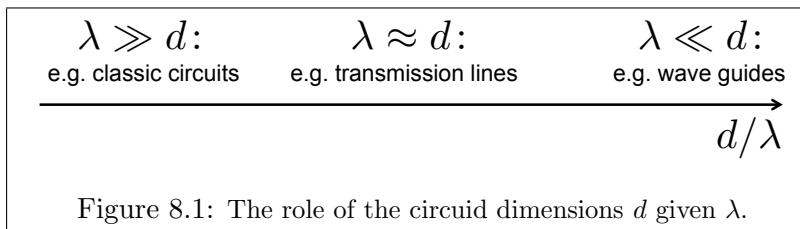


Figure 8.1: The role of the circuital dimensions d given λ .

It should be stressed that the transition between classical circuits, transmission lines, and wave guides is not a function of d only, but also λ . A familiar example are CPUs in smart phones that typically run at a frequency of about 1 GHz, corresponding to λ of about 1 m. Such a CPU can be treated as a classical circuit since d is about 1 cm $\ll \lambda \approx 1$ m. Yet, a similar frequency of 2.45 GHz ($\lambda \approx 0.1$ m) is used in microwave ovens that require wave guides. While surprising at first, this can be understood considering that d for microwave ovens is typically a couple of orders of magnitude larger than for CPUs!

We start by considering how wires guide electromagnetic waves, and then extend the concepts to more complex setups.

8.2 Parallel wires

Guiding structures made up of pairs of wires are particularly important, both conceptually and practically. Consider the arrangement shown in Fig. 8.2.

If the wire is much longer than a wavelength, we cannot assume that the voltage and current are constant along the wire. In general, V and I will be functions of position, as well as time. We know that the line has some inductance L per unit length, and some capacitance C per unit length. Hence, certain transmission lines can be modelled as a series connection of distributed inductors, and

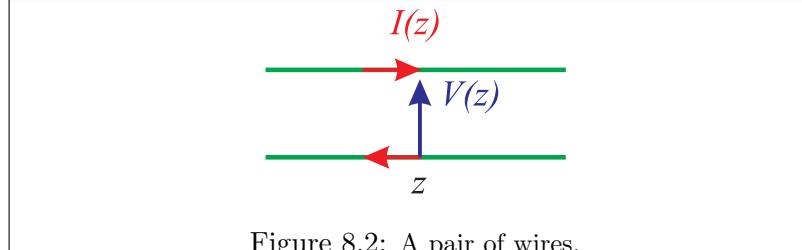


Figure 8.2: A pair of wires.

a parallel connection of distributed capacitors. These distributed inductances and capacitances imply that the currents and voltages change with z .

To approach the problem, we first consider the voltage drop across a small length dz of the line that will be given as

$$dV = V_2 - V_1 = -(Ldz) \frac{\partial I}{\partial t}, \quad (8.1)$$

and the current change along this small length of line will be

$$dI = I_2 - I_1 = -(Cdz) \frac{\partial V}{\partial t}. \quad (8.2)$$

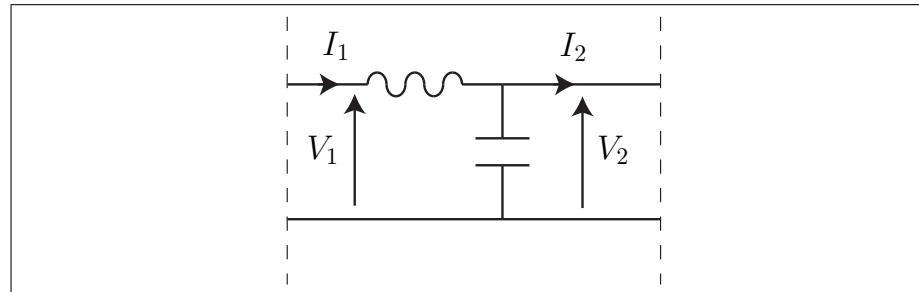


Figure 8.3: Equivalent circuit of a small length of a transverse electromagnetic (TEM) transmission line.

As dz tends to zero, the line takes the form of distributed inductance and capacitance, as anticipated. The equivalent circuit of a small length of a transmission line that supports a transverse electromagnetic (TEM) wave is shown in Fig. 8.3.

Hence, we can rewrite the above equation differentially as

$$\begin{aligned} \frac{\partial V}{\partial z} dz &= -(Ldz) \frac{\partial I}{\partial t} \\ \frac{\partial I}{\partial z} dz &= -(Cdz) \frac{\partial V}{\partial t}, \end{aligned} \quad (8.3)$$

i.e., dropping the differential,

$$\frac{\partial V}{\partial z} = -L \frac{\partial I}{\partial t} \quad (8.4)$$

$$\frac{\partial I}{\partial z} = -C \frac{\partial V}{\partial t}. \quad (8.5)$$

These equations are valid for any transmission line that supports a transverse electromagnetic wave.

Differentiating again and combining, we obtain

$$\begin{aligned}\frac{\partial^2 V}{\partial z^2} &= LC \frac{\partial^2 V}{\partial t^2} \\ \frac{\partial^2 I}{\partial z^2} &= LC \frac{\partial^2 I}{\partial t^2}.\end{aligned}\quad (8.6)$$

As the above wave equation indicates, we obtained voltage and current waves that travel along the transmission line at velocity

$$v = \pm \frac{1}{\sqrt{LC}}. \quad (8.7)$$

We can look for solutions of the form

$$\begin{aligned}V &= V_0 \exp[-i(kz - \omega t)] \\ I &= I_0 \exp[-i(kz - \omega t)],\end{aligned}\quad (8.8)$$

where we have used the engineering convention for the Fourier transform. These solutions can be substituted into Eq. (8.4) to give

$$kV = \omega LI,$$

i.e.,

$$\frac{V}{I} = \frac{\omega L}{k}. \quad (8.9)$$

Considering the definition of the wave velocity,

$$\frac{\omega}{k} = v = \frac{1}{\sqrt{LC}}, \quad (8.10)$$

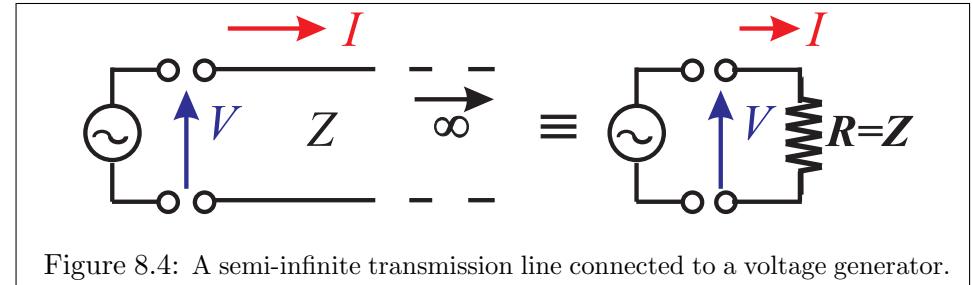
we obtain the **characteristic impedance** of the transmission line

$\frac{V}{I} = Z = \sqrt{\frac{L}{C}}, \quad \text{Characteristic impedance}$

(8.11)

The characteristic impedance of the transmission line is conceptually similar to the characteristic impedance of free space. As for a plane wave, the characteristic impedance of the wave travelling in the negative- z direction is $-Z$. It can be seen that the characteristic impedance is a real number, and so the voltage and current are in phase; in fact they propagate locked in phase with respect to each other.

Now consider connecting a semi-infinite transmission line to a voltage generator, as shown in Fig. 8.4. Because the impedance $Z = V/I$ is a real number, the transmission line can be swapped for a resistor without changing the electrical behaviour at the terminals. In fact, we can connect a resistor having the value of the



characteristic impedance, Z , to the end of a transmission line without changing the behaviour at the input terminals. This is referred to as *impedance-matching*. In physics and engineering, a load is often impedance-matched to a given transmission line to avoid undesired wave reflection.

Note that we have just adopted a new concept in this Chapter: wires should be regarded as guiding electromagnetic energy, rather than simply conducting current.

Thought for the day 23: Practical transmission lines have resistive loss in their conductors, and even resistive loss due to leakage in their dielectric. Show that a resistive loss in the wires leads to a wave that decays as it propagates. Why is it energy-inefficient to lay power transmission lines underground?

8.3 Parallel wires

A pair of parallel wires represents a good example to study guided waves, as it simple, supports a transverse electromagnetic wave, and can be described using a simple transmission-line model.

Consider two parallel wires, each of radius a , separated by a distance $2D$, where $2D \gg a$, as shown in Fig. 8.5. The electric and

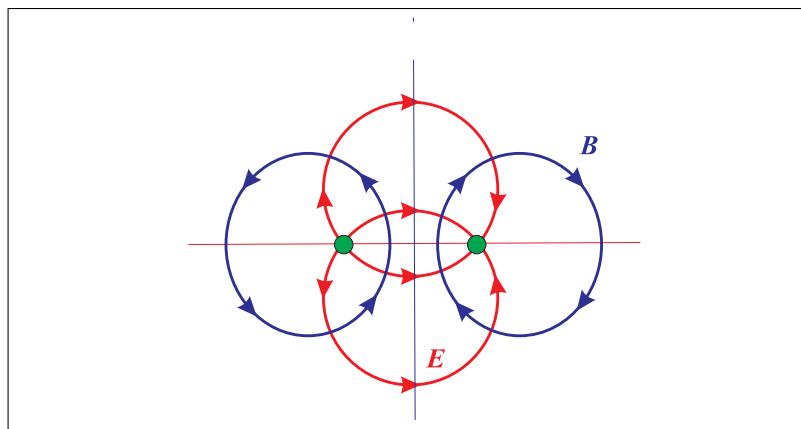


Figure 8.5: A two-wire transmission line.

magnetic field lines are everywhere transverse, and orthogonal to each other. The Poynting vector $\mathbf{N} = \mathbf{E} \times \mathbf{H}$ always points in the same direction as the wire, which implies that power flows along the wire.

We have already calculated that

$$C = \frac{\pi\epsilon_0}{\ln(2D/a)} \quad \text{and} \quad L \stackrel{(6.35)}{=} \frac{\mu_0}{\pi} \ln(2D/a), \quad (8.12)$$

from which it follows that

$$v = \frac{1}{\sqrt{LC}} = \frac{1}{\sqrt{\epsilon_0\mu_0}} = c \quad (8.13)$$

and

$$Z = \sqrt{\frac{L}{C}} = Z_0 \frac{\ln(2D/a)}{\pi}. \quad (8.14)$$

If the wires are embedded in a dielectric with relative permittivity ϵ and refractive index $n = \sqrt{\epsilon}$, the above expressions have to be modified as follows:

$$v = \frac{1}{\sqrt{LC}} = \frac{c}{\sqrt{\epsilon}} = \frac{c}{n}, \quad (8.15)$$

and

$$Z' = \sqrt{\frac{L}{C}} = \frac{Z_0}{\sqrt{\epsilon}} \frac{\ln(2D/a)}{\pi} = \frac{Z}{n}. \quad (8.16)$$

Also,

$$k = k_0 n, \quad (8.17)$$

or

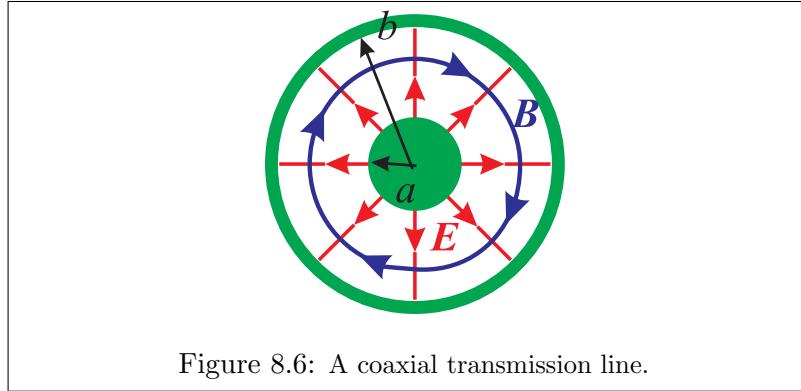
$$\lambda = \frac{\lambda_0}{n}. \quad (8.18)$$

As anticipated, we found that the wave vector is larger and the wavelength smaller than they would be in an air-spaced cable.

Note that if the wires are not completely embedded in a dielectric, but are, say, routed over the surface of a dielectric half-space, then the transmission cannot support a TEM wave. Usually, however, the wave behaves in a TEM-like manner as long as the frequency is low enough. Often, such pairs of wires are twisted together to prevent external magnetic field coupling noise into the circuit, but as long as the twist is long compared to the wavelength, the line behaves as if it were a simple parallel pair.

8.4 Coaxial cylinders

Now consider the coaxial cylinders shown in Fig. 8.6, where the inner and outer radii of the intervening dielectric are a and b , respectively. Typically, coaxial cables are filled with a dielectric. Hence,



we assume that the medium between the cylinders has permittivity ϵ . The electric field is radial, the magnetic field is azimuthal, and the Poynting vector points in the direction of the transmission line. Again, it is clear that this system supports a transverse electromagnetic wave.

We know from Eqns. (2.89) and (6.30) that

$$C = \frac{2\pi\epsilon\epsilon_0}{\ln(b/a)} \quad \text{and} \quad L = \frac{\mu_0}{2\pi} \ln(b/a), \quad (8.19)$$

and therefore

$$v = \frac{1}{\sqrt{LC}} = \frac{c}{\sqrt{\epsilon}} = \frac{c}{n} \quad (8.20)$$

and

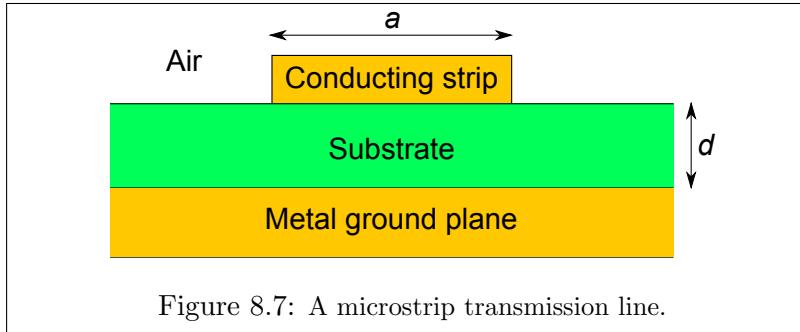
$$Z = \sqrt{\frac{L}{C}} = \frac{Z_0}{n} \frac{\ln(b/a)}{2\pi}. \quad (8.21)$$

For example, a polyethylene insulator has $\epsilon = 2.3$, and therefore $Z = 50\Omega$ for $b/a = 3.5$, which is a typical impedance. In fact most commercial coaxial cables are manufactured with $Z = 50\Omega$ or $Z = 75\Omega$.

If a coaxial cable is only part filled with dielectric, it will still support a TEM wave if the dielectric forms concentric cylinders; any azimuthal structure in the dielectric will, however, prevent a pure TEM wave from propagating.

8.5 Strip transmission line

The strip transmission line is an important example to study here, particularly as printed circuit boards (PCB) are being made to operate at higher and higher frequencies. Personal computers run at frequencies in the GHz range, with harmonics at higher frequencies, which, for the typical dimensions of a mainboard, is well into the region where radio-frequency (RF) engineering must be used. This is less of an issue on CPUs, which often can be approximately treated as classical circuits due to their small dimensions.



Consider the arrangement where two flat strips of metal are used to guide a wave, as shown in Fig. 8.7. If their separation, d , is small compared to the width of the narrower strip, a , edge effects can be ignored, and

$$C = \frac{\epsilon\epsilon_0 a}{d} \quad \text{and} \quad L = \frac{\mu_0 d}{a}, \quad (8.22)$$

from which we find

$$v = \frac{c}{n}, \quad (8.23)$$

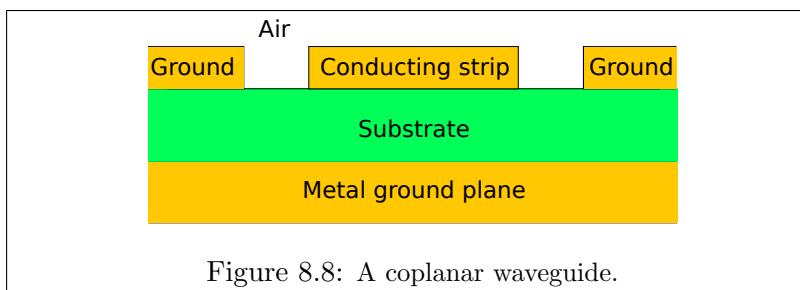
and

$$Z = \frac{Z_0}{n} \frac{d}{a}. \quad (8.24)$$

Interestingly, the dimensions of the transmission line appear in the expression for the impedance. Consequently, the impedance can be changed merely by adjusting the geometry for a given dielectric.

In the case of real printed circuit boards, one track comprises a continuous ground plane, and the other track a sheet conductor; fringing factors then appear in the equations for the impedance and guide wavelength, which we have neglected here. This type of transmission line is called a **microstrip line**.

Strictly speaking, a microstrip line does not support a TEM wave because the conductors are not completely embedded in a homogeneous dielectric, but are separated by a dielectric half space. A TEM analysis is quite satisfactory, however, as long as the frequency is not too high, and the width of the main conductor as well as the thickness of the dielectric substrate, are small compared with the wavelength. Another type of stripline, with less leakage of field lines to other conductors (crosstalk) is a coplanar waveguide, which is sketched in Fig. 8.8.



8.6 Power flow on transmission lines

It is interesting to use a technique similar to that used for introducing the Poynting vector, to study power flow on transmission lines.

Consider a transmission line, where the voltages and currents are labelled as shown in Fig. 8.9. The energy stored, U , on a transmis-

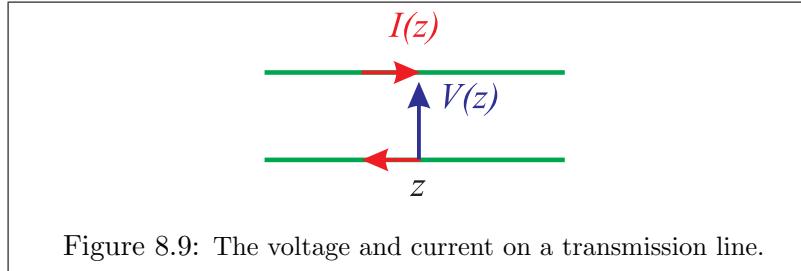


Figure 8.9: The voltage and current on a transmission line.

sion line between points $z = a$ and $z = b$ is given by:

$$U = \int_a^b dz \left(\frac{1}{2} LI^2 + \frac{1}{2} CV^2 \right), \quad (8.25)$$

which is the energy stored in the electric and magnetic fields of the *distributed* inductance and capacitance.

The rate of change of stored energy is

$$\frac{dU}{dt} = \int_a^b dz \left(LI \frac{\partial I}{\partial t} + CV \frac{\partial V}{\partial t} \right), \quad (8.26)$$

which, by virtue of the transmission-line equations (8.4) and (8.5), becomes

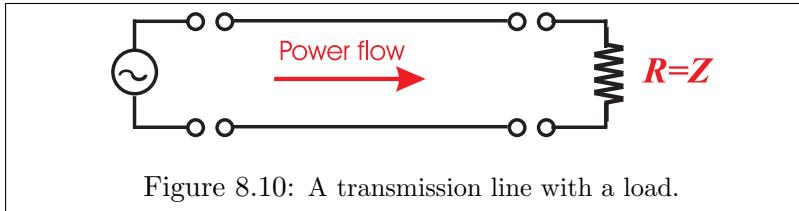
$$\frac{dU}{dt} = - \int_a^b dz \left(I \frac{\partial V}{\partial z} + V \frac{\partial I}{\partial z} \right). \quad (8.27)$$

The integral evaluates to

$$\frac{dU}{dt} = [IV]_a - [IV]_b. \quad (8.28)$$

The rate of change of the stored energy is equal to the rate at which energy flows in, at point $z = a$, minus the rate at which energy flows out, at $z = b$. As expected, the rate at which energy flows along the transmission line at a point is simply VI .

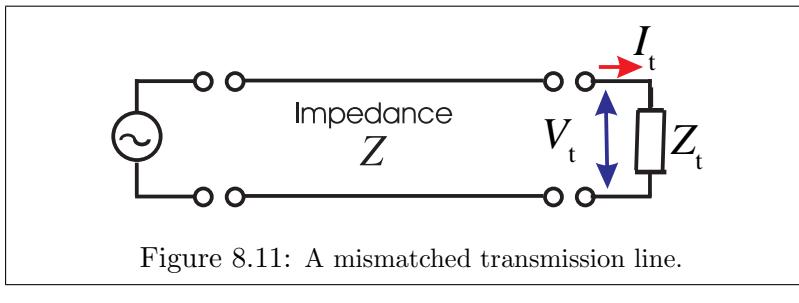
Now consider connecting a load equal to the characteristic impedance, $R = Z$, to the end of a transmission line, as shown in Fig. 8.10. The voltage source creates a travelling wave that carries energy away from the source. In our case, the load absorbs all of the power that is incident on it. This follows because a load with $R = Z$ is connected to our transmission line, i.e., the power flowing along the line is given by $P = VI$ with $V = ZI = RI$ at the end of the transmission line and across the load.



If, as above, a transmission line is terminated with a load equal to the characteristic impedance, no power will be reflected. In this case, the load is said to be **matched** to the line. By contrast, if we terminate a line with a resistance other than Z , some of the power will have to be reflected, because the incident power cannot be the same as the absorbed power. Power transmission lines setups are constructed such as to avoid this undesired behaviour.

8.7 Reflections on transmission lines

In the following, we want to study reflections on transmission lines. For this, consider a line terminated with a resistance other than Z , as shown in Fig. 8.11.



Even though $Z_t \neq Z$, we must have

$$\frac{V_t}{I_t} = Z_t. \quad (8.29)$$

at the terminal load. The forward-travelling incident waves have the form

$$\begin{aligned} V_i &= V_1 \exp[-i(kz - \omega t)] \\ I_i &= I_1 \exp[-i(kz - \omega t)], \end{aligned} \quad (8.30)$$

but, due to the reflection, backward-travelling reflected waves must also exist:

$$\begin{aligned} V_r &= V_2 \exp[-i(-kz - \omega t)] \\ I_r &= I_2 \exp[-i(-kz - \omega t)]. \end{aligned} \quad (8.31)$$

At the boundary between the transmission line and the load we must have

$$V_t = V_i + V_r = \{V_1 e^{-ikz} + V_2 e^{ikz}\} e^{i\omega t}, \quad (8.32)$$

and

$$I_t = I_i + I_r = \left\{ \frac{V_1}{Z} e^{-ikz} - \frac{V_2}{Z} e^{ikz} \right\} e^{i\omega t}, \quad (8.33)$$

where the load is located at z . To simplify the notation, let us place the origin of the coordinate system, $z = 0$, at the load, so then

$$\frac{V_t}{I_t} = \frac{V_1 + V_2}{V_1/Z - V_2/Z} = Z_t. \quad (8.34)$$

The **voltage reflection coefficient** becomes

$$r \equiv \frac{V_2}{V_1} = \frac{Z_t - Z}{Z_t + Z}. \quad \text{Voltage reflection coefficient}$$

(8.35)

The voltage reflection coefficient r becomes zero when $Z_t = Z$, as anticipated in the previous Section.

The actual voltage across the load, which can be regarded as the **transmitted voltage**, is

$$V_t = V_1 + V_2 = tV_1, \quad (8.36)$$

where t is the voltage transmission coefficient. We therefore have the **voltage transmission coefficient**

$$t \equiv \frac{V_t}{V_1} = 1 + r = \frac{2Z_t}{Z_t + Z}. \quad \text{Voltage transmission coefficient}$$

(8.37)

Thought for the day 24: Show that the incident power is equal to the transmitted power plus the reflected power; express your answer in terms of the reflection and transmission coefficients.

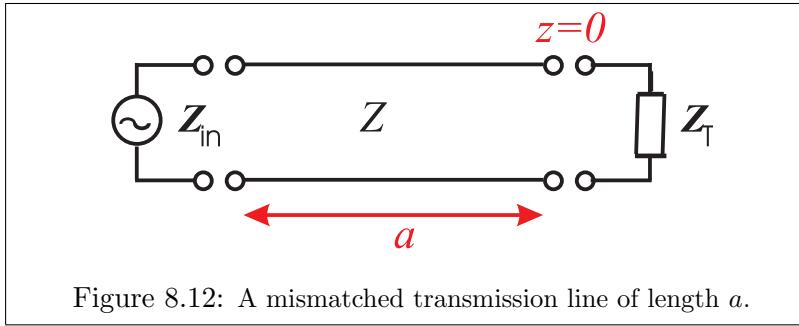
8.8 Input impedances of transmission lines

Having established that there can be reflections on transmission lines, in this Section we shall study their effects seen at the input of the transmission line.

Consider the arrangement shown in Fig. 8.12, where a length of transmission line is terminated by an impedance Z_t , where $Z_t \neq Z$. Assume that the length of the line is a . Assume that the impedance of the *source* is equal to Z , so that it absorbs all reflections.

By the definition of the voltage reflection coefficient r , we know that

$$\begin{aligned} V_i &= V_1 \exp[-i(kz - \omega t)] \\ V_r &= rV_1 \exp[-i(-kz - \omega t)], \end{aligned} \quad (8.38)$$



and

$$\begin{aligned} I_i &= \frac{V_i}{Z} \\ I_r &= -\frac{V_r}{Z}. \end{aligned} \quad (8.39)$$

We further can write

$$Z_{in} = \left. \frac{V_i + V_r}{I_i + I_r} \right|_{z=-a} = \frac{e^{ika} + re^{-ika}}{e^{ika} - re^{-ika}} Z. \quad (8.40)$$

Now, using the definition of the voltage reflection coefficient

$$r = \frac{Z_t - Z}{Z_t + Z}, \quad (8.41)$$

we can show that

$$\boxed{\frac{Z_{in}}{Z} = \frac{Z_t \cos ka + iZ \sin ka}{Z \cos ka + iZ_t \sin ka}}. \quad (8.42)$$

We arrived at an interesting result: the impedance that would be measured in an experiment depends on the position of the measurement, a , and the nature of the load, Z_t .

In the following, we shall consider certain special cases:

Short-circuited line

In the case of a short-circuited line, i.e., $Z_t \rightarrow 0$,

$$\boxed{\frac{Z_{in}}{Z} = i \tan ka.} \quad (8.43)$$

Notice that the input impedance is always imaginary because the load cannot absorb power, and all of the power must be reflected.

It can be seen that for $0 < ka < \pi/2$, the impedance is *inductive*, and for $\pi/2 < ka < \pi$, it is *capacitive*. In fact, shorted transmission lines are often used in high-frequency engineering to synthesise impedances, and therefore to tune circuits, amplifiers, and antennas.

Open-circuited line

In the case of an open-circuited line, i.e., $Z_t \rightarrow \infty$,

$$\boxed{\frac{Z_{in}}{Z} = -i \cot ka.} \quad (8.44)$$

It can be seen that for $0 < ka < \pi/2$, the impedance is *capacitive*, while for $\pi/2 < ka < \pi$ it is *inductive*.

Quarter-wavelength line

A transmission line that is a quarter of a wavelength long, so that $ka = \pi/2$, as shown in Fig. 8.13, has particular significance. In this

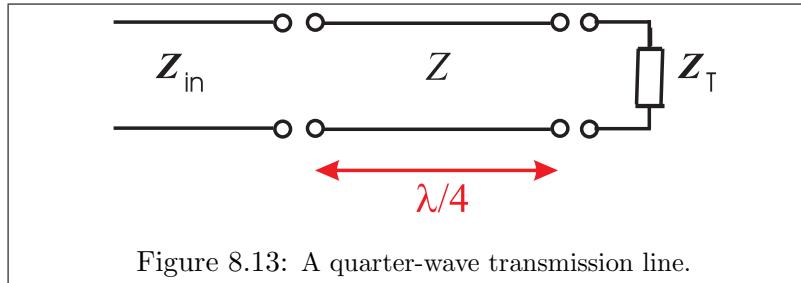


Figure 8.13: A quarter-wave transmission line.

case, $\cos ka = 0$, and

$$\frac{Z_{in}}{Z} = \frac{Z}{Z_t}, \quad (8.45)$$

i.e.,

$$Z_{in} = \frac{Z^2}{Z_t}. \quad (8.46)$$

Let us assume that we have some load that we would like to connect to a transmission line, but that the load and transmission line have different impedances. In this case we can use a *quarter-wavelength line* to match one to the other, ensuring that there are no reflections.

Care is required when impedance matching using quarter-wavelength lines, because the matching will only be effective at one particular frequency: as the frequency is changed, the line will no longer be one quarter of a wavelength long, and the match will deteriorate, introducing reflections on the line. It turns out that multiple lengths of transmission line having different characteristic impedances can be connected together to achieve broadband matching. In fact, this kind of matching works in the same way as antireflection coatings on optical components (see question 42). Again, these techniques are used extensively at GHz (microwave) and THz frequencies.

8.9 Waveguides

Transmission lines of the kinds just considered are used extensively at high frequencies. In fact, coaxial cable can be purchased for frequencies up to tens of GHz, and superconducting striplines are used in astronomical detectors for frequencies of up to 1 THz. Each of these structures has the advantage that it supports a transverse electromagnetic wave, or at least a wave that is very close to being TEM for the frequency of operation. However, as we have seen in

Chapter 7, the losses increase dramatically at high frequencies, due to the skin effect in the conductors.

Hence, the question arises if we can guide waves in ways that do not require a TEM wave on a two-conductor transmission line? After all, we can guide optical beams by the use of mirrors. Indeed, electromagnetic waves can propagate *without* the need for a second conductor, down hollow metal tubes of various cross-sections (for microwave frequencies) or along optical fibres (for optical wavelengths). These are usually what we mean by the term *waveguides*.

In fact, the transmission lines we have just considered are themselves waveguides too, but the voltages and currents in them are simple enough that we can analyse the system using them instead of the electric and magnetic fields. We always have this choice of whether to consider I and V everywhere in the surfaces or to calculate \mathbf{E} and \mathbf{H} everywhere around the conductors.

In the following Sections, we shall be concerned with the question how we can calculate the modes that can propagate on waveguides. Obviously, these are not plane waves, but we will find that, for certain structures, the modes are closely related to plane waves.

8.10 Plane-wave pairs

To attack the problem, consider a preliminary setup that consists of two plane waves, with wave vectors \mathbf{k}_1 and \mathbf{k}_2 , travelling at angles $\pm\theta$ to the z axis, as sketched in Fig. 8.14. Using the coordinate

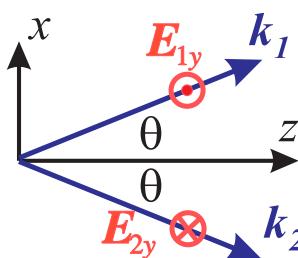


Figure 8.14: A pair of y -polarised plane waves.

system defined in Fig. 8.14, we can write

$$\begin{aligned}\mathbf{k}_1 &= (k \sin \theta, 0, k \cos \theta) \\ \mathbf{k}_2 &= (-k \sin \theta, 0, k \cos \theta).\end{aligned}\quad (8.47)$$

It follows that

$$\begin{aligned}\mathbf{k}_1 \cdot \mathbf{r} &= kx \sin \theta + kz \cos \theta \\ \mathbf{k}_2 \cdot \mathbf{r} &= -kx \sin \theta + kz \cos \theta.\end{aligned}\quad (8.48)$$

A composite wave that comprises a linear combination of these two plane waves has the functional form

$$E_y = E_0 \{ e^{i(\mathbf{k}_1 \cdot \mathbf{r})} - e^{i(\mathbf{k}_2 \cdot \mathbf{r})} \} e^{-i\omega t}, \quad (8.49)$$

where, without loss of generality, y polarisation has been assumed for simplicity in Fig. 8.14. Similarly, a minus sign has been chosen for the amplitudes, resulting in a simpler algebra.

By substitution of Eq. (8.48) we find that

$$\begin{aligned} E_y &= E_0 \exp[i(kz \cos \theta - \omega t)] \\ &\quad \{ \exp[i(kx \sin \theta)] - \exp[i(-kx \sin \theta)] \} \quad (8.50) \\ &= E_0 \exp[i(kz \cos \theta - \omega t)] 2i \sin(kx \sin \theta). \end{aligned}$$

A close inspection reveals a remarkable feature of the result: there are **planes of constant x** for which the E_y component is zero, at all times t ! The m^{th} such plane occurs at

$$kx \sin \theta = m\pi, \quad (8.51)$$

where m is an integer of either sign. In the above expression, k and θ are assumed to be given, characterising an experimental setup. In other words, we have found a **standing wave in the x direction**, and a **propagating wave in the z direction**.

[As an aside, note that if we had combined the \mathbf{E} fields with a + sign instead of a – sign, we would instead have $\cos(kx \sin \theta)$, which is not zero at $x = 0$. However, it has zeroes at $kx \sin \theta = m\pi/2$ for odd-integer m , which would bring us to the same physical conclusion, but would be far less convenient algebraically.]

The phase velocity in the z direction is

$$v_{\text{ph}} = \frac{\omega}{k_g}, \quad (8.52)$$

where the effective wave vector along the guide is

$$k_g \equiv k \cos \theta. \quad (8.53)$$

Combining the two expressions, we find

$$v_{\text{ph}} = \frac{\omega}{k \cos \theta} = \frac{c}{\cos \theta}, \quad (8.54)$$

Since $\cos \theta \leq 1$, we have found that the phase velocity v_{ph} is greater than the speed of light, a phenomenon that is seen in many optical systems. One can show that the group velocity $v_g < c$.

8.11 Conducting plates

In the previous example, we could have placed conducting plates at the planes for which E_y —the tangential \mathbf{E} field—is zero, and the

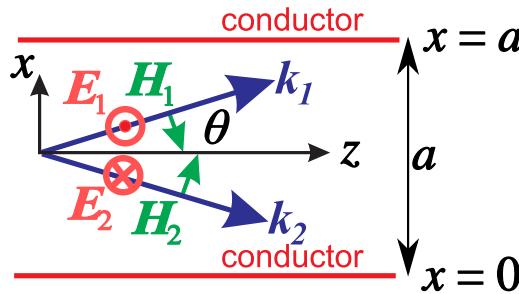


Figure 8.15: A pair of y -polarised plane waves propagating between two parallel plates.

solution to the problem would not have changed. Thus a field can propagate between two conducting parallel plates, and it will take the form of a combined pair of plane waves. The situation is shown in Fig. 8.15.

We know that the field is zero for

$$kx \sin \theta = m\pi, \quad (8.55)$$

so we choose to put the lower plate at $x = 0$. If the total distance between the plates is a , the other plate is at $x = a$, and so we require

$$ka \sin \theta = m\pi, \quad (8.56)$$

for some integer $m > 0$. It follows that

$$k \sin \theta = k_x = \frac{m\pi}{a}, \quad (8.57)$$

which establishes a particular set of angles for a given separation a . In other words, for a given free-space wavelength, only plane waves travelling at very specific angles can propagate.

Now, using $\cos^2 \theta + \sin^2 \theta = 1$, and remembering that $k_g = k \cos \theta$, we find that

$$k_g^2 = k^2 - \frac{m^2 \pi^2}{a^2}. \quad (8.58)$$

In fact, this could be found more easily through

$$k^2 = k_x^2 + k_z^2, \quad (8.59)$$

where $k_x = \frac{m\pi}{a}$ is determined by the separation between the plates, and $k_z = k_g$.

This solution has been constructed in a rather artificial way, and we should be careful to consider the boundary conditions not only of the \mathbf{E} field, but also of the \mathbf{H} field. For each plane wave, we know that \mathbf{H} is perpendicular to \mathbf{E} and the direction of propagation, either \mathbf{k}_1 or \mathbf{k}_2 . It follows that the \mathbf{H} field has both x and z

components:

$$\begin{aligned}\mathbf{H}_1 &= H_0(-\cos \theta, 0, \sin \theta) \\ &\quad \exp \left[i \frac{m\pi}{a} x \right] \exp [i(kz \cos \theta - \omega t)] \quad (8.60) \\ \mathbf{H}_2 &= H_0(-\cos \theta, 0, \sin \theta) \\ &\quad \exp \left[-i \frac{m\pi}{a} x \right] \exp [i(kz \cos \theta - \omega t)].\end{aligned}$$

Adding the two components, we find

$$\begin{aligned}\mathbf{H} &= 2H_0 \exp [i(kz \cos \theta - \omega t)] \\ &\quad \left(-i \cos \theta \sin \left(\frac{m\pi}{a} x \right), 0, \sin \theta \cos \left(\frac{m\pi}{a} x \right) \right).\end{aligned} \quad (8.61)$$

In other words, when $x = 0$ or a , i.e., at the conducting plates, the x component of the \mathbf{H} field becomes zero, as does the y component of the \mathbf{E} field: $H_x = 0, E_y = 0$. Because $\mathbf{H}_\perp = 0$, the normal component of the magnetic field \mathbf{B}_\perp is zero at the conducting plates. We know that the normal component of the \mathbf{B} field across a boundary must be continuous, and since there is no magnetic field inside the conductor, there cannot be any normal component of magnetic field outside the conductor. So this boundary condition is automatically satisfied.

There *can* be a parallel component of the \mathbf{H} field at the surface of a perfect conductor, but since there is no field inside, there must, according to Ampère's Law, be a thin sheet of current at the surface. We have found a remarkable result: For our travelling waves in our preliminary waveguide, current flows near the surfaces of the conducting plates, within the skin depth, and over this distance the parallel component of the \mathbf{H} field goes to zero.

8.12 Transverse electric modes

Having constructed a preliminary waveguide of two parallel plates in the previous section, we shall study its properties in more detail in the following. The field has the functional form

$$E_y = E_0 \sin \left(\frac{m\pi}{a} x \right) \exp [i(k_g z - \omega t)], \quad (8.62)$$

where m is any positive integer. The first two modes have the form shown in Fig. 8.16.

It can now be seen that whereas our original two-strip transmission line supported a TEM wave, the modes that we have now discovered are not TEM, despite the fact that the individual plane waves from which they are constructed are TEM. The transition from TEM to non-TEM behaviour occurs when the spacing of the plates becomes greater than half a free-space wavelength, which follows because we found in Eq. (8.57)

$$ka \sin \theta = m\pi, \quad (8.63)$$

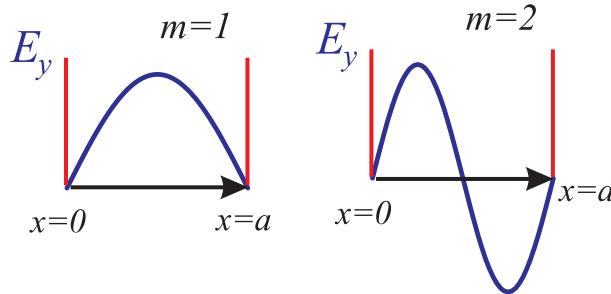


Figure 8.16: The two lowest-order modes for a parallel-plate waveguide.

for $m \in \mathbb{Z}^+$, i.e.,

$$\sin \theta = \frac{m\pi}{ka} = \frac{m\pi\lambda}{2\pi a} = \frac{m\lambda}{2a}. \quad (8.64)$$

Therefore, in order for our setup to act as a waveguide, it must fulfil the condition

$$\frac{m\lambda}{2a} < 1. \quad (8.65)$$

For a given wavelength, the above condition is most likely to be satisfied for the lowest-order mode, $m = 1$, and for this mode to propagate

$$a > \lambda/2. \quad (8.66)$$

If the separation between the parallel plates is fixed, then, as the frequency increases, and the wavelength falls, the lowest-order mode will start to propagate inside the waveguide when the distance between the plates is greater than half a wavelength. Sometimes we refer to the critical wavelength, λ_c , where

$$\boxed{\lambda_c = 2a.} \quad (8.67)$$

As an example, a microstrip line will behave as a TEM transmission line as long as the lines' cross-sectional dimensions are small compared to the given wavelength, which is typical of RF circuit boards. However, as the frequency increases, parallel-plate waveguide modes will start to propagate, and the behaviour changes considerably. In fact, this is one of the primary reasons why low-pass transmission-line filters start to leak power at high frequencies. The only solution is to make the line's geometry very small, which becomes a challenge above about 1 THz, where $\lambda/2 = 150 \mu\text{m}$.

This same reasoning applies to all TEM transmission-line systems, and again explains why coaxial lines become smaller and smaller as the frequency of operation is increased.

8.13 Rectangular waveguides

So far, our preliminary waveguide consists of two parallel plates. In the following, we shall construct a realistic waveguide by introducing additional plates. Because the \mathbf{E} field is in the y direction, it is possible to place plates in planes of constant y , while ensuring that there is no \mathbf{E} field parallel to the surfaces of the new plates. We also know that the \mathbf{H} field is always in the $x - z$ plane, $H_y = 0$, and therefore the two new plates can be placed anywhere, while still ensuring that the normal component of the \mathbf{H} (and \mathbf{B}) field is zero. We choose to place them at $y = 0$ and $y = b$. The guiding structure now consists of a perfectly conducting rectangular tube of dimensions $a \times b$.

The lowest-order mode shown is designated the TE_{10} mode using the TE_{mn} notation, where $m = 1$ indicates that there is one half-wavelength of the \mathbf{E} field oscillation in the x -direction, and $n = 0$ implies no variation in the y -direction. TE refers to **Transverse Electric**, which indicates that the \mathbf{E} field is everywhere transverse to the direction of propagation of the total field in the waveguide. By contrast, the \mathbf{H} field has x and z components, i.e., it is not transverse. Fig. 8.17 shows the \mathbf{E} (dotted line) and \mathbf{H} (solid line) fields of the TE_{10} mode in a rectangular waveguide.

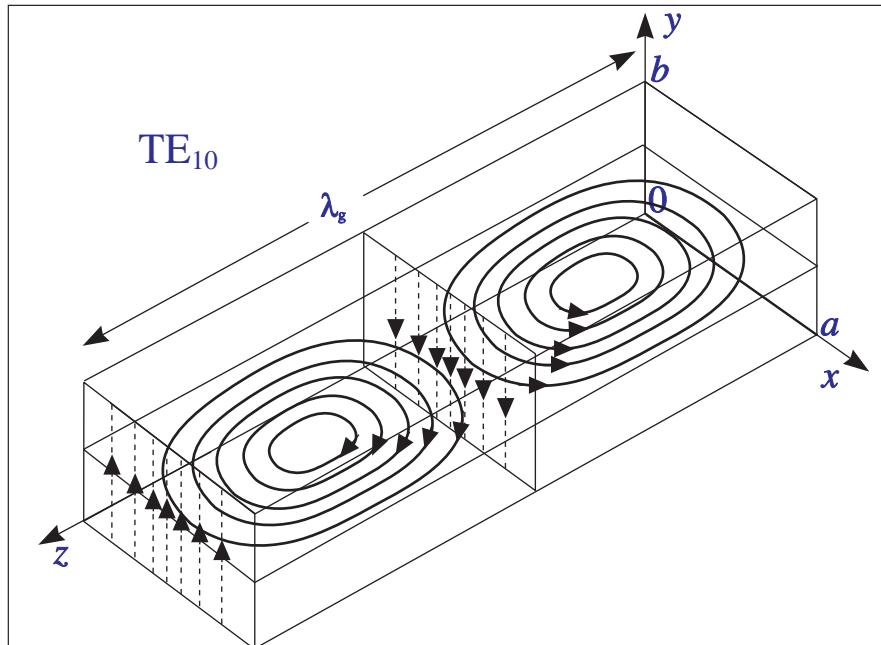


Figure 8.17: The TE_{10} mode in a rectangular waveguide: \mathbf{E} (dotted line) and \mathbf{H} (solid line) fields.

It is important to point out that the \mathbf{H} field is non-zero at the surface of the top and bottom conductor plates, and gives rise to the current shown in Fig. 8.18. Also note that the currents flow parallel to the z axis along the *centres* of the top and bottom surfaces, i.e., at $\frac{a}{2}$. In fact, thin slots can be cut along the centres of the

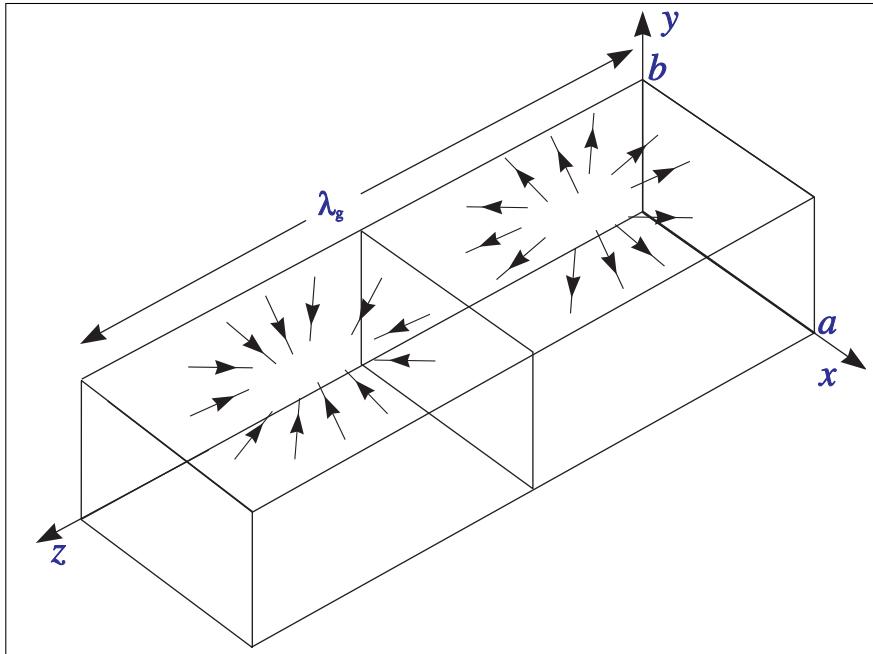


Figure 8.18: The currents associated with the TE_{10} mode of a rectangular waveguide.

walls without affecting the waveguide's ability to guide the TE_{10} mode. When making cuts in the walls of metallic waveguides, say to introduce components, it is important to ask whether a cut prevents the flow of current; if it does, it will interfere with the nature of the propagating mode, and may prevent propagation altogether.

Keep in mind that the original TEM transmission-line mode cannot propagate in a realistic waveguide. This is because we have now connected together the two strips that originally formed the top and bottom guiding plates, and hence there can be no voltage difference between them.

8.14 The General TE_{mn} mode

We can allow \mathbf{E} to vary with both x and y across the waveguide, whilst remaining transverse. The general solution is

$$\begin{aligned} E_x &= A_0 k_y \cos(k_x x) \sin(k_y y) \cos(k_z z - \omega t) \\ E_y &= -A_0 k_x \sin(k_x x) \cos(k_y y) \cos(k_z z - \omega t) \\ E_z &= 0 \end{aligned} \quad (8.68)$$

where

$$(k_x, k_y, k_z) \equiv \left(\frac{m\pi}{a}, \frac{n\pi}{b}, k_g \right). \quad (8.69)$$

Note that m and n must be integers in order to fulfil the boundary conditions. The mode in Eq. (8.69) is known as the TE_{mn} mode, where m controls the variation in the x direction, and n controls the variation with y .

We require integers $m, n \geq 0$, but we cannot have $m = n = 0$ at the same time.

Note that, even though k_x and k_y are now quantised, $|\mathbf{k}^2|$ must still satisfy

$$\frac{\omega^2}{c^2} = k_x^2 + k_y^2 + k_g^2. \quad (8.70)$$

8.15 The Waveguide Equation

From Eq. (8.70) above, to satisfy the wave equation we must have

$$k_z^2 = k_g^2 = \frac{\omega^2}{c^2} - \frac{m^2\pi^2}{a^2} - \frac{n^2\pi^2}{b^2}. \quad (8.71)$$

Thus, for a given guide and mode (m, n) , $k_x^2 + k_y^2 \equiv k_c^2$ is fixed. The k_c parameter is the ‘critical’ k . Conventionally, k_0 is used instead of k , with $k_0 \equiv \omega/c$ ensuring that $k_0^2 = k_x^2 + k_y^2 + k_z^2$. Using the k_0 notation, the **waveguide equation** can be written neatly as

$$k_g^2 = k_0^2 - k_c^2. \quad \text{Waveguide equation} \quad (8.72)$$

For propagating modes, k_g^2 must be positive, otherwise there would be an exponential decay. This gives a cut-off frequency

$$\nu_c = \frac{ck_c}{2\pi} = c \left(\frac{m^2}{4a^2} + \frac{n^2}{4b^2} \right)^{1/2}, \quad (8.73)$$

where $\omega_c = ck_c$ has been used. Below the cut-off frequency for a given guide and given mode, waves will *not* propagate and we have evanescent waves.

9. Conclusion and non-examinable appendix

9.1 Conclusion

This course has established Maxwell's equations from experimentally deduced laws. The equations are completely general, and are consistent with Special Relativity. We predicted from these equations, based on low-frequency results, that electric and magnetic fields could propagate as electromagnetic radiation, at any frequency, even e.g., 10^{20} Hz, at the speed of light, c , which is a constant in a vacuum. We have also found how to guide such waves, down transmission lines and waveguides.

We have also worked out the properties of circuit components such as capacitors, inductors and transformers, and have shown how these depend on the surrounding materials, based on how the individual atoms in those materials behave.

In the Quantum Mechanics course you have learned that Electromagnetism is at the basis of quantisation of black body radiation, the cradle of quantum mechanics. Next year, you will learn more about electromagnetism in the Optics and Electrodynamics course. Beyond that, one can study quantum field theory in the Quantum Electrodynamics course, to see how interaction probabilities can be calculated from fundamental principles, and how photons come into the picture.

One current direction of research is into the use of single photons in secure communications. Each photon may have one of two polarisation directions (e.g. left or right circular polarisation), and quantum-mechanically may be in a mixture (superposition) of the two states, and 'entangled' with another such photon. Then one photon may be sent somewhere, and the presence of any eavesdropper detected because it would change the states of the photons. The generation of such photons often requires great control of the cavity in which the photon is emitted. For example, the presence of the cavity can control the frequency at which the photon is emitted.

Another hot topic is that of **negative** refractive index materials—'metamaterials'. These are artificial media (consisting of arrays of inductors *etc.*) or perhaps even some crystals. Rigorous calculations using Maxwell's equations predict, for example, perfect lenses (revealing sub-wavelength details), or 'invisibility cloaks'...

Moreover, the great degree of control of quantum field theoretical calculations of quantum electrodynamical processes predestine it for as a perfect testbed of the Standard Model of Elementary Particle Physics, and hence allow for indirect searches for new physics beyond the Standard Model through precision measurements of key observables.

The above representative examples clearly demonstrate that electromagnetism is an exciting and still developing field, whilst at the same time underpinning many other disciplines, and being technologically vital to the modern world.

9.2 Useful results—non-examinable

The rest of this chapter contains some useful or interesting results that are *not* examinable.

9.3 The generalised divergence theorem

The following results may seem very mathematical, but if you are not put off completely, they can be very powerful. They enable results to be proved without recourse to particular loops, since all the work has been done for us already in proving the divergence theorem and Stokes' theorem! They will not come up in the exam, but are worth keeping in the back of your mind in case you need them in the future.

- The divergence theorem

$$\oint d\mathbf{S} \cdot \mathbf{A} = \int d\tau \nabla \cdot \mathbf{A} \quad (9.1)$$

is much more general than you think!

- The generalised form is

$$\oint d\mathbf{S}(\bullet) = \int d\tau \nabla(\bullet) \quad (9.2)$$

The argument \bullet can be anything...

$$\text{i.e., } \oint dS_i f = \int d\tau \frac{\partial}{\partial x_i} f \quad (9.3)$$

where f is any quantity (with or without indices). This is an operator identity.

- Two new useful theorems: for a scalar field Φ and a vector \mathbf{A}

$$\oint d\mathbf{S} \Phi = \int d\tau \nabla \Phi ; \quad \oint d\mathbf{S} \times \mathbf{A} = \int d\tau \nabla \times \mathbf{A} \quad (9.4)$$

9.4 The generalised Stokes Theorem

- Stokes' theorem

$$\oint d\mathbf{l} \cdot \mathbf{A} = \int d\mathbf{S} \cdot \nabla \times \mathbf{A} = \int dS_i \epsilon_{ijk} \frac{\partial}{\partial x_j} A_k \quad (9.5)$$

may be written

$$\oint d\mathbf{l} \cdot \mathbf{A} = \int (d\mathbf{S} \times \nabla) \cdot \mathbf{A} \quad (9.6)$$

- The generalised form is

$$\oint d\mathbf{l}(\bullet) = \int (d\mathbf{S} \times \nabla)(\bullet) \quad (9.7)$$

$$\text{i.e., } \oint dl_i f = \int dS_j \epsilon_{ijk} \frac{\partial}{\partial x_k} f \quad (9.8)$$

where f is any quantity (with or without indices).

- Here are two useful new theorems:

$$\oint d\mathbf{l} \Phi = \int d\mathbf{S} \times \nabla \Phi ; \quad \oint d\mathbf{l} \times \mathbf{A} = \int (d\mathbf{S} \times \nabla) \times \mathbf{A} \quad (9.9)$$

- Remember the brackets!

- The ∇ operates on **everything** that follows it.

9.5 Couple on Current Loop

This example of the use of the generalised Stokes theorem proves the formula for the couple on a current loop of arbitrary shape, without us having to think about the cancellation of many small loops as we did earlier. The number of suffices is daunting, but just follow the strict rules of summing repeated suffices, and it all comes out without much thought.

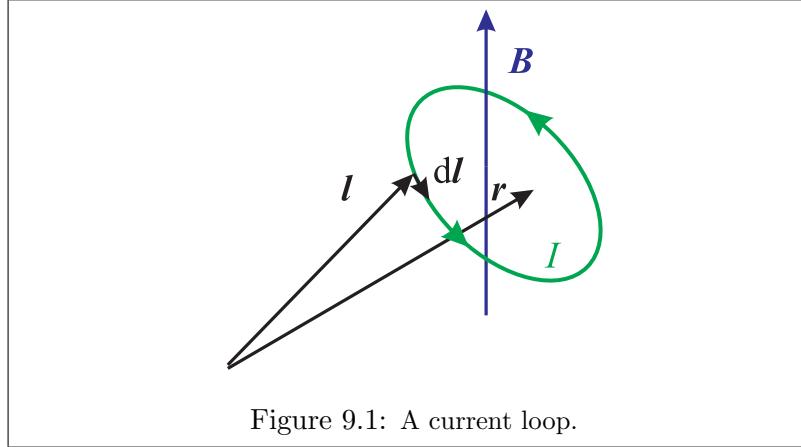


Figure 9.1: A current loop.

- Uniform magnetic field.
 - \mathbf{l} is on the loop; \mathbf{r} is a general point.
 - Force: $\mathbf{F} = I \oint d\mathbf{l} \times \mathbf{B}$ i.e. $F_i = I \oint \epsilon_{ilm} dl_l B_m$.
 - Couple: $\mathbf{G} = I \oint \mathbf{l} \times (d\mathbf{l} \times \mathbf{B})$ i.e.
- $$G_i = I \oint \epsilon_{ijk} l_j \epsilon_{klm} dl_l B_m. \quad (9.10)$$

- Generalised Stokes: $\oint dl_l f = \int dS_p \epsilon_{lpq} \frac{\partial}{\partial x_q} f$ with $f = \epsilon_{ijk} l_j \epsilon_{klm} B_m$.

The fact that the suffix l is also in the above integral, and so is repeated (and hence summed over) is nothing to worry about. It just means we will sum three such integrals.

$$\text{So } G_i = I \int dS_p \epsilon_{lpq} \frac{\partial}{\partial x_q} [\epsilon_{ijk} l_j \epsilon_{klm} B_m].$$

- Not quite as bad as it looks...

$\mathbf{l} = \mathbf{r}$, and $\epsilon_{ijk} \epsilon_{klm} = \delta_{il} \delta_{jm} - \delta_{im} \delta_{jl}$, so

$$G_i = I \int dS_p \epsilon_{lpq} (\delta_{il} \delta_{jm} - \delta_{im} \delta_{jl}) \frac{\partial}{\partial x_q} (x_j B_m)$$

- \mathbf{B} is constant, and $\frac{\partial x_j}{\partial x_q} \equiv \delta_{jq}$, so $G_i = I \int dS_p (\epsilon_{ipq} B_j - \epsilon_{jpq} B_i) \delta_{qj}$

$$\Rightarrow G_i = I \int dS_p (\epsilon_{ipq} B_q - \epsilon_{jpj} B_i)$$

But $\epsilon_{jpj} \equiv 0$ (two indices the same)

$$\bullet \Rightarrow \mathbf{G} = I \int d\mathbf{S} \times \mathbf{B} = \mathbf{m} \times \mathbf{B}$$