ELECTRICITY AND MAGNETISM

CHAPTER 1
ELECTRIC FIELDS

1.1 *Introduction*

This is the first in a series of chapters on electricity and magnetism. Much of it will be aimed at an introductory level suitable for first or second year students, or perhaps some parts may also be useful at high school level. Occasionally, as I feel inclined, I shall go a little bit further than an introductory level, though the text will not be enough for anyone pursuing electricity and magnetism in a third or fourth year honours class. On the other hand, students embarking on such advanced classes will be well advised to know and understand the contents of these more elementary notes before they begin.

The subject of electromagnetism is an amalgamation of what were originally studies of three apparently entirely unrelated phenomena, namely *electrostatic* phenomena of the type demonstrated with pieces of amber, pith balls, and ancient devices such as Leyden jars and Wimshurst machines; magnetism, and the phenomena associated with lodestones, compass needles and Earth's magnetic field; and current electricity – the sort of electricity generated by chemical cells such as Daniel and Leclanché cells. These must have seemed at one time to be entirely different phenomena. It wasn't until 1820 that Oersted discovered (during the course of a university lecture, so the story goes) that an electric current is surrounded by a magnetic field, which could deflect a compass needle. The several phenomena relating the apparently separate phenomena were discovered during the nineteenth century by scientists whose names are immortalized in many of the units used in electromagnetism – Ampère, Ohm, Henry, and, especially, Faraday. The basic phenomena and the connections between the three disciplines were ultimately described by Maxwell towards the end of the nineteenth century in four famous equations. This is not a history book, and I am not qualified to write one, but I strongly commend to anyone interested in the history of physics to learn about the history of the growth of our understanding of electromagnetic phenomena, from Gilbert's description of terrestrial magnetism in the reign of Queen Elizabeth I, through Oersted's discovery mentioned above, up to the culmination of Maxwell's equations.

This set of notes will be concerned primarily with a description of electricity and magnetism as natural phenomena, and it will be treated from the point of view of a "pure" scientist. It will *not* deal with the countless electrical devices that we use in our everyday life – how they work, how they are designed and how they are constructed. These matters are for electrical and electronics engineers. So, you might ask, if your primary interest in electricity is to understand how machines, instruments and electrical equipment work, is there any point in studying electricity from the very "academic" and abstract approach that will be used in these notes, completely divorced as they appear to be from the world of practical reality? The answer is that electrical engineers *more than anybody* must understand the basic scientific principles before they even begin to apply them to the design of practical appliances. So – do not even think of electrical engineering until you have a thorough understanding of the basic scientific principles of the subject.

This chapter deals with the basic phenomena, definitions and equations concerning *electric fields.*

1.2 *Triboelectric Effect*

In an introductory course, the basic phenomena of electrostatics are often demonstrated with "pith balls" and with a "gold-leaf electroscope". A *pith ball* used to be a small, light wad of pith extracted from the twig of an elder bush, suspended by a silk thread. Today, it is more likely to be either a ping-pong ball, or a ball of styrofoam, suspended by a nylon thread – but, for want of a better word, I'll still call it a pith ball. I'll describe the gold-leaf electroscope a little later.

It was long ago noticed that if a sample of amber (fossilized pine sap) is rubbed with cloth, the amber became endowed with certain apparently wonderful properties. For example, the amber would be able to attract small particles of fluff to itself. The effect is called the *triboelectric effect.* [Greek τρίβος (rubbing) + ήλεκτρον (amber)] The amber, after having been rubbed with cloth, is said to bear an *electric charge*, and space in the vicinity of the charged amber within which the amber can exert its attractive properties is called an *electric field.*

Amber is by no means the best material to demonstrate triboelectricity. Modern plastics (such as a comb rubbed through the hair) become easily charged with electricity (provided that the plastic, the cloth or the hair, and the atmosphere, are *dry*). Glass rubbed with silk also carries an electric charge – but, as we shall see in the next section, the charge on glass rubbed with silk seems to be not quite the same as the charge on plastic rubbed with cloth.

1.3 *Experiments with Pith Balls*

A pith ball hangs vertically by a thread. A plastic rod is charged by rubbing with cloth. The charged rod is *brought close to the pith ball without touching it.* It is observed that the charged rod weakly *attracts* the pith ball. This may be surprising – and you are right to be surprised, for the pith ball carries no charge. For the time being we are going to put this observation to the back of our minds, and we shall defer an explanation to a later chapter. Until then it will remain a small but insistent little puzzle.

We now *touch* the pith ball with the charged plastic rod. Immediately, some of the magical property (i.e. some of the electric charge) of the rod is transferred to the pith ball, and we observe that thereafter the ball is strongly *repelled* from the rod. We conclude that two electric charges repel each other. Let us refer to the pith ball that we have just charged as Ball A.

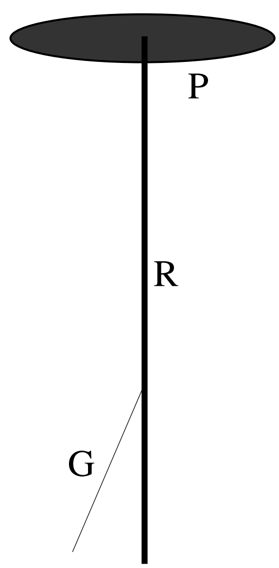
Now let's do exactly the same experiment with the glass rod that has been rubbed with silk. We bring the charged glass rod close to an uncharged Ball B. It initially attracts it weakly – but we'll have to wait until Chapter 2 for an explanation of this unexpected behaviour. However, as soon as we *touch* Ball B with the glass rod, some charge is transferred to the ball, and the rod thereafter repels it. So far, no obvious difference between the properties of the plastic and glass rods.

But... now bring the *glass* rod close to *Ball A*, and we see that Ball A is strongly *attracted*. And if we bring the *plastic* rod close to *Ball B*, it, too, is strongly *attracted.* Furthermore, Balls A and B *attract each other.*

We conclude that there are *two kinds* of electric charge, with exactly opposite properties. We arbitrarily call the kind of charge on the glass rod and on Ball B *positive* and the charge on the plastic rod and Ball A *negative*. We observe, then, that *like* charges (i.e. those of the *same* sign) *repel* each other, and *unlike* charges (i.e. those of *opposite* sign) *attract* each other.

1.4 *Experiments with a Gold-leaf Electroscope*

FIGURE I.1



A gold-leaf electroscope has a vertical rod R attached to a flat metal plate P. Gold is a malleable metal which can be hammered into extremely thin and light sheets. A light gold leaf G is attached to the lower end of the rod.

If the electroscope is positively charged by touching the plate with a positively charged glass rod, G will be repelled from R, because both now carry a positive charge.

You can now experiment as follows. Bring a positively charged glass rod close to P. The leaf G diverges further from R. We now know that this is because the metal (of which P, R and G are all composed) contains *electrons*, which are negatively charged particles that can move about more or less freely inside the metal. The approach of the positively charged glass rod to P attracts electrons towards P, thus increasing the excess positive charge on G and the bottom end of R. G therefore moves away from R.

If on the other hand you were to approach P with a negatively charged plastic rod, electrons would be repelled from P down towards the bottom of the rod, thus reducing the excess positive charge there. G therefore approaches R.

Now try another experiment. Start with the electroscope uncharged, with the gold leaf hanging limply down. (This can be achieved by touching P briefly with your finger.) Approach P with a negatively charged plastic rod, but don't touch. The gold leaf diverges from R. Now, briefly touch P with a finger of your free hand. Negatively charged electrons run down through your body to ground (or earth). Don't worry – you won't feel a thing. The gold leaf collapses, though by this time the electroscope bears a positive charge, because it has lost some electrons through your body. Now remove the plastic rod. The gold leaf diverges again. By means of the negatively charged plastic rod and some deft work with your finger, you have induced a positive charge on the electroscope. You can verify this by approaching P alternately with a plastic (negative) or glass (positive) rod, and watch what happens to the gold leaf.

1.5 *Coulomb's Law*

If you are interested in the history of physics, it is well worth reading about the important experiments of Charles Coulomb in 1785. In these experiments he had a small fixed metal sphere which he could charge with electricity, and a second metal sphere attached to a vane suspended from a fine torsion thread. The two spheres were charged and, because of the repulsive force between them, the vane twisted round at the end of the torsion thread. By this means he was able to measure precisely the small forces between the charges, and to determine how the force varied with the amount of charge and the distance between them.

From these experiments resulted what is now known as *Coulomb's Law*. Two electric charges of like sign repel each other with a force that is proportional to the product of their charges and inversely proportional to the square of the distance between them:

$$F \propto \frac{Q_1 Q_2}{r^2}. \qquad 1.5.1$$

Here Q_1 and Q_2 are the two charges and r is the distance between them.

We could in principle use any symbol we like for the constant of proportionality, but in standard SI (Système International) practice, the constant of proportionality is written as $\frac{1}{4\pi\epsilon}$, so that Coulomb's Law takes the form

$$F = \frac{1}{4\pi\varepsilon}\frac{Q_1 Q_2}{r^2}. \qquad 1.5.2$$

Here ε is called the *permittivity* of the medium in which the charges are situated, and it varies from medium to medium. The permittivity of a *vacuum* (or of "free space") is given the symbol ε_0. Media other than a vacuum have permittivities a little greater than ε_0. The permittivity of air is very little different from that of free space, and, unless specified otherwise, I shall assume that all experiments described in this chapter are done either in free space or in air, so that I shall write Coulomb's Law as

$$F = \frac{1}{4\pi\varepsilon_0}\frac{Q_1 Q_2}{r^2}. \qquad 1.5.3$$

You may wonder – why the factor 4π? In fact it is very convenient to define the permittivity in this manner, with 4π in the denominator, because, as we shall see, it will ensure that all formulas that describe situations of spherical symmetry will include a 4π, formulas that describe situations of cylindrical symmetry will include 2π, and no π will appear in formulas involving uniform fields. Some writers (particularly those who favour cgs units) prefer to incorporate the 4π into the definition of the permittivity, so that Coulomb's law appears in the form $F = Q_1 Q_2 /(\varepsilon_0 r^2)$, though it is standard SI practice to define the permittivity as in equation 1.5.3. The permittivity defined by equation 1.5.3 is known as the "rationalized" definition of the permittivity, and it results in much simpler formulas throughout electromagnetic theory than the "unrationalized" definition.

The SI unit of charge is the *coulomb*, C. Unfortunately at this stage I cannot give you an exact definition of the coulomb, although, if a current of 1 amp flows for a second, the amount of electric charge that has flowed is 1 coulomb. This may at first seem to be very clear, until you reflect that we have not yet defined what is meant by an amp, and that, I'm afraid, will have to come in a much later chapter.

Until then, I can give you some small indications. For example, the charge on an electron is about -1.6022×10^{-19} C, and the charge on a proton is about $+1.6022 \times 10^{-19}$ C. That is to say, a collection of 6.24×10^{18} protons, if you could somehow bundle them all together and stop them from flying apart, amounts to a charge of 1 C. A *mole* of protons (i.e. 6.022×10^{23} protons) which would have a mass of about one gram, would have a charge of 9.65×10^4 C, which is also called a *faraday* (which is *not at all* the same thing as a *farad*).

[The current definition of the coulomb and the amp, which will be given in Chapter 6, requires some knowledge of electromagnetism. However, it is likely that, in 2015, the coulomb will be redefined in such a manner that the magnitude of the charge on a single electron is exactly 1.60217×10^{-19} C.]

The charges involved in our experiments with pith balls, glass rods and gold-leaf electroscopes are very small in terms of coulombs, and are typically of the order of nanocoulombs.

The permittivity of free space has the approximate value

$$\varepsilon_0 = 8.8542 \times 10^{-12} \text{ C}^2 \text{ N}^{-1} \text{ m}^{-2}.$$

Later on, when we know what is meant by a "farad", we shall use the units F m^{-1} to describe permittivity – but that will have to wait until section 5.2.

You may well ask how the permittivity of free space is measured. A brief answer might be "by carrying out experiments similar to those of Coulomb". However – and this is rather a long story, which I shall not describe here – it turns out that since we today define the metre by *defining* the speed of light, c, to be exactly $2.997\,925\,58 \times 10^8 \text{ m s}^{-1}$, the permittivity of free space has a *defined value*, given, in SI units, by

$$4\pi\varepsilon_0 = \frac{10^7}{c^2}.$$

It is therefore not necessary to *measure* ε_0 any more than it is necessary to *measure* c. But that, as I say, is a long story.

[But if, as is likely, the new definition of the coulomb, referred to on the previous page, becomes official in 2015, ε_0 will no longer have an exact defined value, but its measured value will be approximately $8.8542 \times 10^{-12} \text{ C}^2 \text{ N}^{-1} \text{ m}^{-2}$. Many teaching laboratories run an undergraduate experiment in which students measure the charge on a capacitor of known physical dimensions and a measured potential difference between the plates, and this enables the measured value of ε_0 to be calculated.]

From the point of view of *dimensional analysis*, electric charge cannot be expressed in terms of M, L and T, but it has a dimension, Q, of its own. (This assertion is challenged by some, but this is not the place to discuss the reasons. I may add a chapter, eventually, discussing this point much later on.) We say that the dimensions of electric charge are Q.

Exercise: Show that the dimensions of permittivity are

$$[\varepsilon_0] = \text{M}^{-1}\,\text{L}^{-3}\,\text{T}^2\,\text{Q}^2.$$

I shall strongly advise the reader to work out and make a note of the dimensions of every new electric or magnetic quantity as it is introduced.

Exercise: Calculate the magnitude of the force between two point charges of 1 C each (that's an enormous charge!) 1 m apart *in vacuo*.

The answer, of course, is $1/(4\pi\varepsilon_0)$, and that, as we have just seen, is $c^2/10^7 = 9 \times 10^9$ N, which is equal to the weight of a mass of 9.2×10^5 tonnes or nearly a million tonnes.

Exercise: Calculate the ratio of the electrostatic to the gravitational force between two electrons. The numbers you will need are: $Q = 1.60 \times 10^{-19}$ C, $m = 9.11 \times 10^{-31}$ kg, $\varepsilon_0 = 8.85 \times 10^{-12} \text{ N m}^2 \text{ C}^{-2}$, $G = 6.67 \times 10^{-11} \text{ N m}^2 \text{ kg}^{-2}$.

The answer, which is independent of their distance apart, since both forces fall off inversely as the square of the distance, is $Q^2/(4\pi\epsilon_0 Gm^2)$, (and you should verify that this is dimensionless), and this comes to 4.2×10^{42}. This is the basis of the oft-heard statement that electrical forces are 10^{42} times as strong as gravitational forces – but such a statement out of context is rather meaningless. For example, the gravitational force between Earth and Moon is much more than the electrostatic force (if any) between them, and cosmologists could make a good case for saying that the strongest forces in the Universe are gravitational.

The ratio of the permittivity of an insulating substance to the permittivity of free space is its *relative permittivity*, also called its *dielectric constant.* The dielectric constants of many commonly-encountered insulating substances are of order "a few". That is, somewhere between 2 and 10. Pure water has a dielectric constant of about 80, which is quite high (but bear in mind that most water is far from pure and is not an insulator.) Some special substances, known as *ferroelectric* substances, such as strontium titanate $SrTiO_3$, have dielectric constants of a few hundred.

1.6 *Electric Field* ***E***

The region around a charged body within which it can exert its electrostatic influence may be called an *electric field*. In principle, it extends to infinity, but in practice it falls off more or less rapidly with distance. We can define the *intensity* or *strength E* of an electric field as follows. Suppose that we place a small test charge q in an electric field. This charge will then experience a force. The ratio of the force to the charge is called the *intensity of the electric field*, or, more usually, simply the *electric field*. Thus I have used the words "electric field" to mean either the region of space around a charged body, or, quantitatively, to mean its intensity. Usually it is clear from the context which is meant, but, if you wish, you may elect to use the longer phrase "intensity of the electric field" if you want to remove all doubt. The field and the force are in the same direction, and the electric field is a vector quantity, so the definition of the electric field can be written as

$$\mathbf{F} = Q\mathbf{E}\,. \qquad 1.6.1$$

The SI units of electric field are newtons per coulomb, or N C^{-1}. A little later, however, we shall come across a unit called a *volt*, and shall learn that an alternative (and more usual) unit for electric field is volts per metre, or V m^{-1}. The dimensions are $MLT^{-2}Q^{-1}$.

You may have noticed that I supposed that we place a "small" test charge in the field, and you may have wondered why it had to be small, and how small. The problem is that, if we place a large charge in an electric field, this will change the configuration of the electric field and hence frustrate our efforts to measure it accurately. So – it has to be sufficiently small so as not to change the configuration of the field that we are trying to measure. How small is that? Well, it will have to mean infinitesimally small. I hope that is clear! (It is a bit like that pesky particle of negligible mass m that keeps appearing in mechanics problems!)

We now need to calculate the intensity of an electric field in the vicinity of various shapes and sizes of charged bodes, such as rods, discs, spheres, and so on.

1.6.1 *Field of a point charge*

It follows from equation 1.5.3 and the definition of electric field intensity that the electric field at a distance r from a point charge Q is of magnitude

$$E = \frac{Q}{4\pi\epsilon_0 r^2}. \qquad 1.6.2$$

This can be written in vector form:

$$\mathbf{E} = \frac{Q}{4\pi\epsilon_0 r^2}\hat{\mathbf{r}} = \frac{Q}{4\pi\epsilon_0 r^3}\mathbf{r}. \qquad 1.6.3$$

Here $\hat{\mathbf{r}}$ is a unit vector in the radial direction, and $\mathbf{r}$ is a vector of length r in the radial direction.

1.6.2 *Spherical Charge Distributions*

I shall not here give calculus derivations of the expressions for electric fields resulting from spherical charge distributions, since they are identical with the derivations for the gravitational fields of spherical mass distributions in the Classical Mechanics "book" of these physics notes, provided that you replace mass by charge and G by $-1/(4\pi\epsilon_0)$. See Chapter 5, subsections 5.4.8 and 5.4.9 of Celestial Mechanics. Also, we shall see later that they can be derived more easily from Gauss's law than by calculus. I shall, however, give the *results* here.

At a distance r from the centre of a *hollow spherical shell* of radius a bearing a charge Q, the electric field is *zero* at any point *inside* the sphere (i.e. for $r < a$). For a point *outside* the sphere (i.e. $r > a$) the field intensity is

$$E = \frac{Q}{4\pi\epsilon_0 r^2}. \qquad 1.6.4$$

This is the same as if all the charge were concentrated at a point at the centre of the sphere.

If you have a *spherically-symmetric distribution of charge Q contained within a spherical volume of radius a,* this can be considered as a collection of nested hollow spheres. It follows that at a point *outside* a spherically-symmetric distribution of charge, the field at a distance r from the centre is again

$$E = \frac{Q}{4\pi\epsilon_0 r^2}. \qquad 1.6.5$$

That is, it is the same as if all the charge were concentrated at the centre. However, at a point *inside* the sphere, the charge beyond the distance r from the centre contributes zero to the electric field; the electric field at a distance r from the centre is therefore just

$$E = \frac{Q_r}{4\pi\epsilon_0 r^2}. \qquad 1.6.6$$

Here Q_r is the charge within a radius r. If the charge is uniformly distributed throughout the sphere, this is related to the total charge by $Q_r = \left(\frac{r}{a}\right)^3 Q$, where Q is the total charge. Therefore, for a uniform spherical charge distribution the field inside the sphere is

$$E = \frac{Qr}{4\pi\epsilon_0 a^3}. \qquad 1.6.7$$

That is to say, it increases linearly from centre to the surface, where it reaches a value of $\frac{Q}{4\pi\epsilon_0 a^2}$, whereafter it decreases according to equation 1.6.5.

It is not difficult to imagine some electric charge distributed (uniformly or otherwise) throughout a finite spherical volume, but, because like charges repel each other, it may not be easy to realize this idealized situation in practice. In particular, if a *metal* sphere is charged, since charge can flow freely through a metal, the self-repulsion of charges will result in all the charge residing on the *surface* of the sphere, which then behaves as a hollow spherical charge distribution with zero electric field within.

1.6.3 *A Long, Charged Rod*

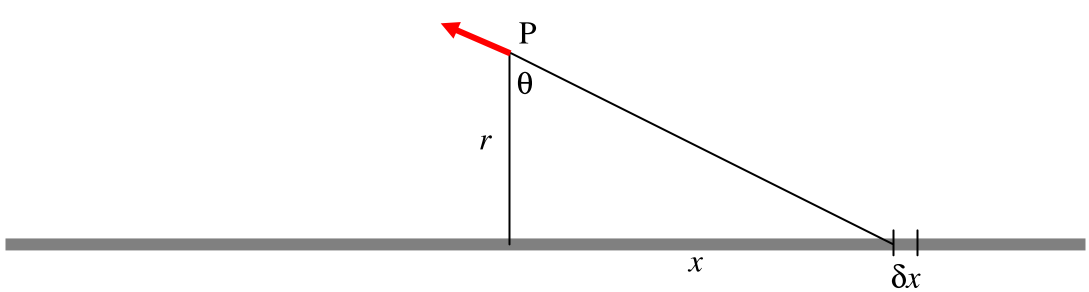


FIGURE I.1

A long rod bears a charge of λ coulombs per metre of its length. What is the strength of the electric field at a point P at a distance r from the rod?

Consider an element δx of the rod at a distance $(r^2 + x^2)^{1/2}$ from the rod. It bears a charge $\lambda\,\delta x$. The contribution to the electric field at P from this element is $\frac{1}{4\pi\epsilon_0}\cdot\frac{\lambda\,\delta x}{r^2 + x^2}$ in the direction shown. The radial component of this is $\frac{1}{4\pi\epsilon_0}\cdot\frac{\lambda\,\delta x}{r^2 + x^2}\cos\theta$. But $x = r\tan\theta$, $\delta x = r\sec^2\theta\,\delta\theta$ and $r^2 + x^2 = r^2\sec^2\theta$. Therefore the radial component of the field from the element δx is $\frac{\lambda}{4\pi\epsilon_0 r}\cos\theta\,\delta\theta$. To find the radial component of the field from the entire rod, we integrate along the length of the rod. If the rod is infinitely long (or if its length is much greater than r), we integrate from $\theta = -\pi/2$ to $+\pi/2$, or, what amounts to the same thing, from 0 to $\pi/2$, and double it. Thus the radial component of the field is

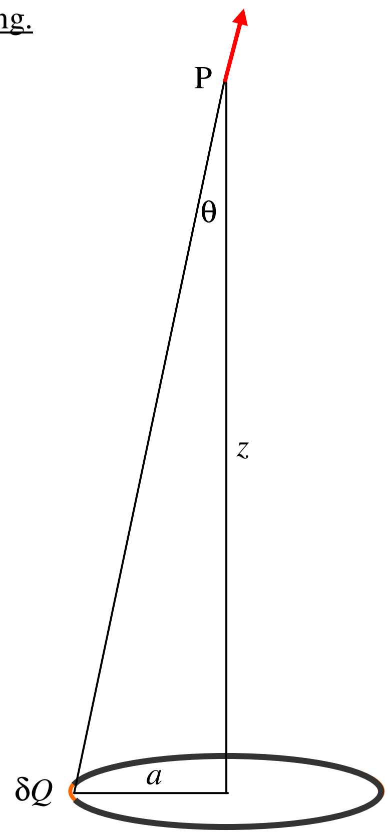
$$E = \frac{2\lambda}{4\pi\epsilon_0 r}\int_0^{\pi/2}\cos\theta\;\delta\theta = \frac{\lambda}{2\pi\epsilon_0 r}. \qquad 1\,6.8$$

The component of the field parallel to the rod, by considerations of symmetry, is zero, so equation 1.6.8 gives the total field at a distance r from the rod, and it is directed radially away from the rod.

Notice that equation 1.6.4 for a spherical charge distribution has $4\pi r^2$ in the denominator, while equation 1.6.8, dealing with a problem of cylindrical symmetry, has $2\pi r$.

1.6.4 *Field on the Axis of and in the Plane of a Charged Ring*

<u>Field on the axis of a charged ring.</u>



Ring, radius a, charge Q. Field at P from element of charge $\delta Q = \dfrac{\delta Q}{4\pi\epsilon_0(a^2 + z^2)}$.

Vertical component of this $= \dfrac{\delta Q\cos\theta}{4\pi\epsilon_0(a^2 + z^2)} = \dfrac{\delta Q z}{4\pi\epsilon_0(a^2 + z^2)^{3/2}}$.

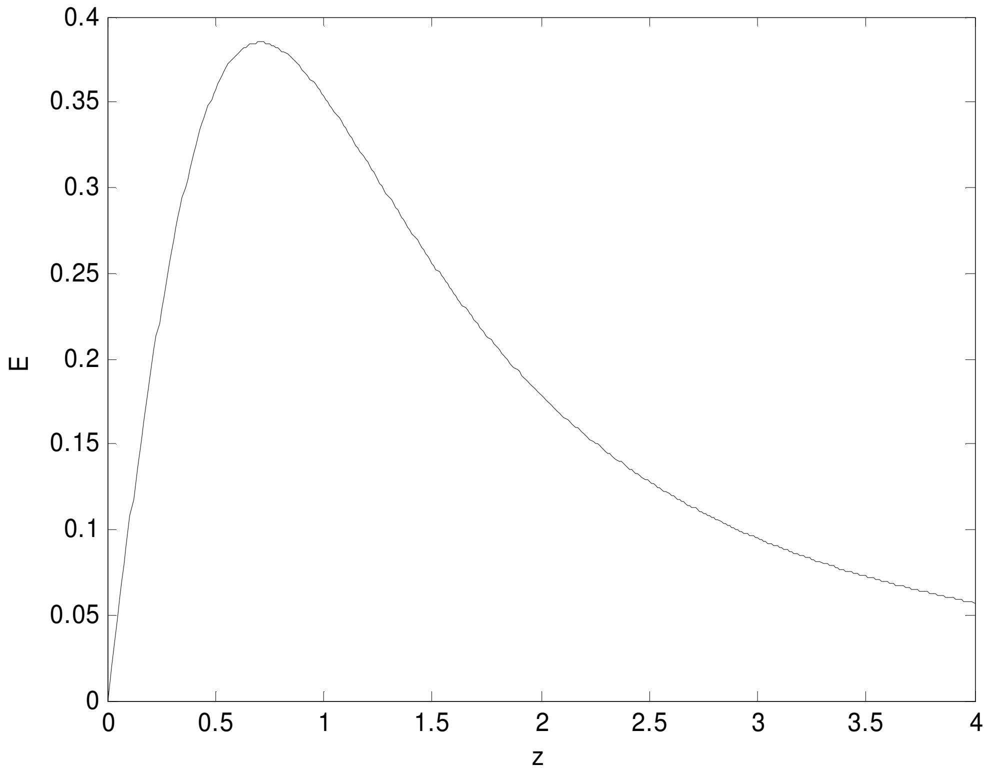
Integrate for entire ring:

Field $\underline{\underline{E = \frac{Q}{4\pi\varepsilon_0}\frac{z}{(a^2 + z^2)^{3/2}}}}$.

In terms of dimensionless variables:

$$\underline{\underline{E = \frac{z}{(1 + z^2)^{3/2}}}} ,$$

where E is in units of $\frac{Q}{4\pi\varepsilon_0 a^2}$, and z is in units of a.



From calculus, we find that this reaches a maximum value of $\frac{2\sqrt{3}}{9} = \underline{\underline{0.3849}}$ at $z = 1/\sqrt{2} = \underline{\underline{0.7071}}$.

It reaches half of its maximum value where $\frac{z}{(1 + z^2)^{3/2}} = \frac{\sqrt{3}}{9}$.

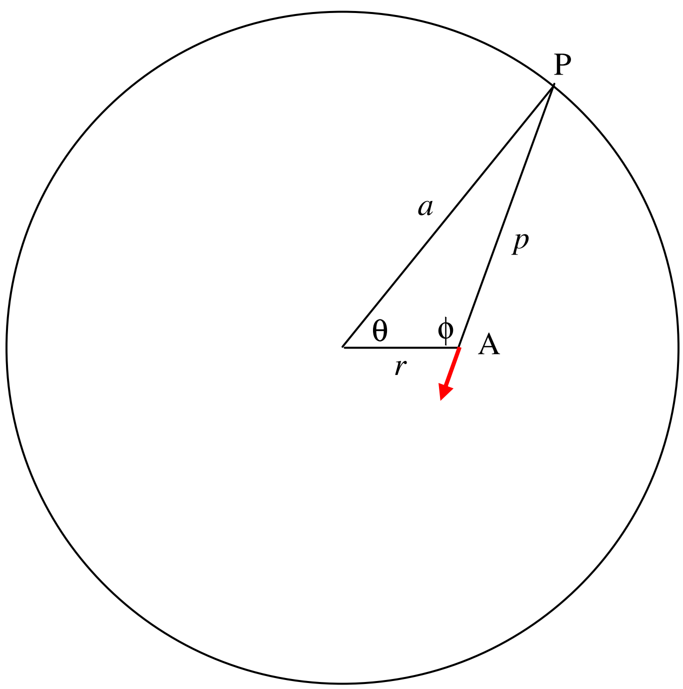
That is, $3 - 72Z + 9Z^2 + 3Z^3 = 0$, where $Z = z^2$.

The two positive solution are $Z = 0.041889$ and 3.596267.

That is, $\underline{\underline{z = 0.2047 \text{ and } 1.8964.}}$

Field in the plane of a charged ring.

We suppose that we have a ring of radius a bearing a charge Q. We shall try to find the field at a point in the plane of the ring and at a distance r $(0 \le r < a)$ from the centre of the ring.



Consider an element $\delta\theta$ of the ring at P. The charge on it is $\frac{Q\delta\theta}{2\pi}$. The field at A due this element of charge is

$$\frac{1}{4\pi\epsilon_0}\cdot\frac{Q\delta\theta}{2\pi}\cdot\frac{1}{a^2+r^2-2ar\cos\theta} = \frac{Q}{4\pi\epsilon_0 . 2\pi a^2}\cdot\frac{\delta\theta}{b-c\cos\theta},$$

where $b = 1 + r^2/a^2$ and $c = 2r/a$. The component of this toward the centre is

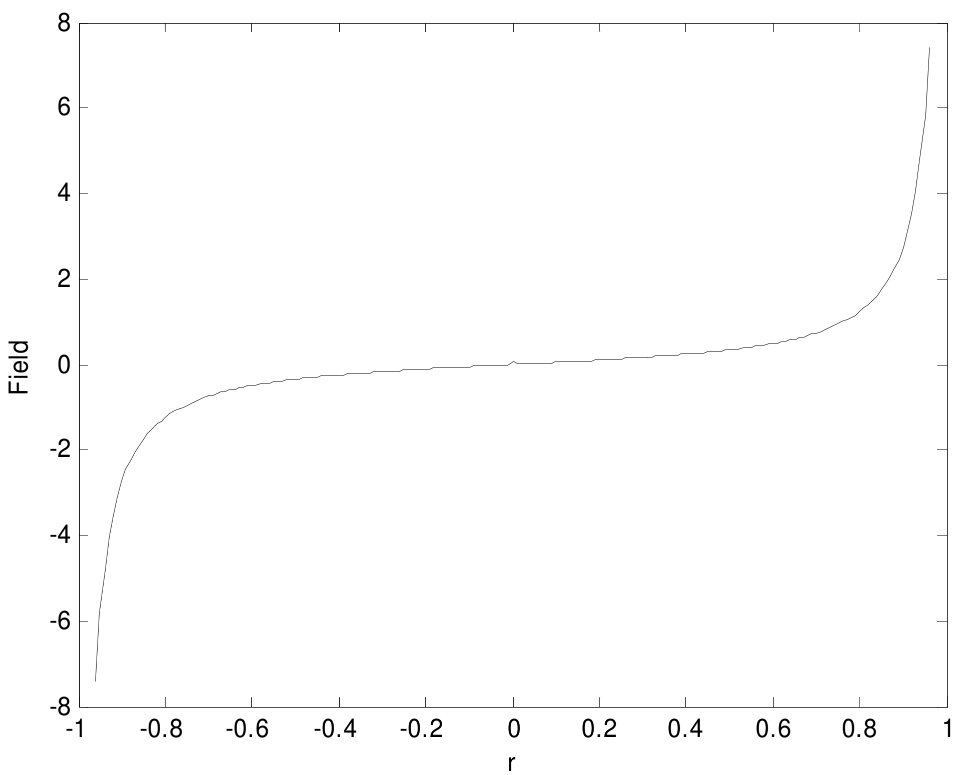
$$-\frac{Q}{4\pi\varepsilon_0 . 2\pi a^2} \cdot \frac{\cos\phi\,\delta\theta}{b - c\cos\theta} \,.$$

To find the field at A due to the entire ring, we must express ϕ in terms of θ, r and a, and integrate with respect to θ from 0 to 2π (or from 0 to π and double it). The necessary relations are

$$p^2 = a^2 + r^2 - 2ar\cos\theta,$$

$$\cos\phi = \frac{r^2 + p^2 - a^2}{2rp} \,.$$

The result of the numerical integration is shown below, in which the field is expressed in units of $Q/(4\pi\varepsilon_0 a^2)$ and r is in units of a.



1.6.5 *Field on the Axis of a Uniformly Charged Disc*

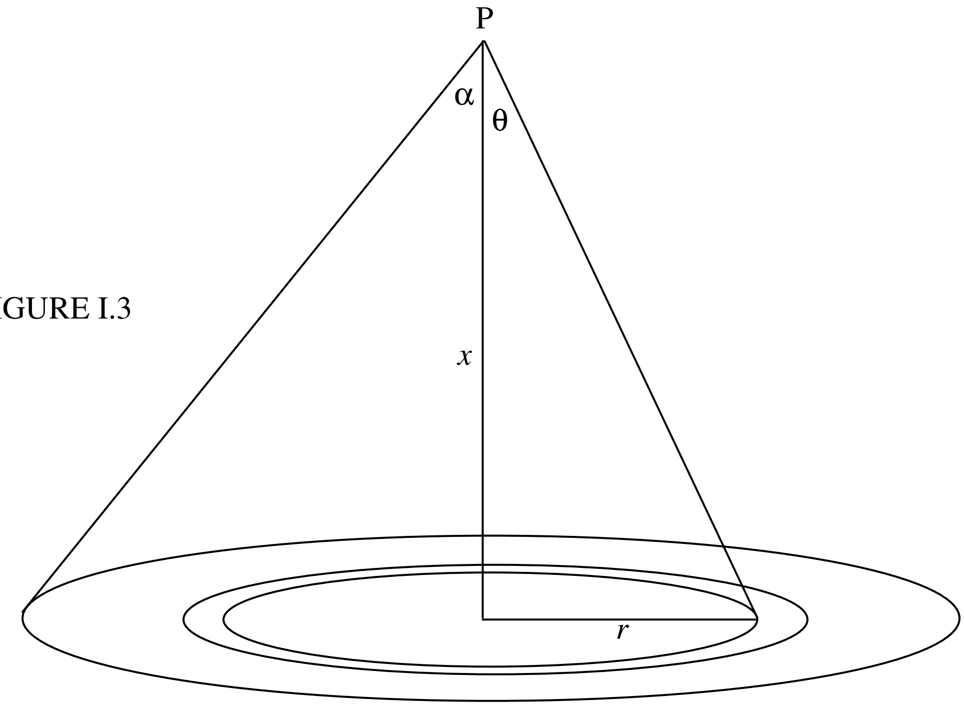


FIGURE I.3

We suppose that we have a circular disc of radius a bearing a surface charge density of σ coulombs per square metre, so that the total charge is $Q = \pi a^2\sigma$. We wish to calculate the field strength at a point P on the axis of the disc, at a distance x from the centre of the disc.

Consider an elemental annulus of the disc, or radii r and $r + \delta r$. Its area is $2\pi r\delta r$ and so it carries a charge $2\pi\sigma r\delta r$. Using the result of subsection1.6.4, we see that the field at P from this charge is

$$\frac{2\pi\sigma\, r\,\delta r}{4\pi\epsilon_0}\cdot\frac{x}{(r^2 + x^2)^{3/2}} = \frac{\sigma\, x}{2\epsilon_0}\cdot\frac{r\,\delta r}{(r^2 + x^2)^{3/2}}\,.$$

But $r = x\tan\theta$, $\delta r = x\sec^2\theta\,\delta\theta$ and $(r^2 + x^2)^{1/2} = x\sec\theta$. Thus the field from the elemental annulus can be written

$$\frac{\sigma}{2\epsilon_0}\sin\theta\,\delta\theta.$$

The field from the entire disc is found by integrating this from $\theta = 0$ to $\theta = \alpha$ to obtain

$$E = \frac{\sigma}{2\epsilon_0}(1 - \cos\alpha) = \frac{\sigma}{2\epsilon_0}\left(1 - \frac{x}{(a^2 + x^2)^{1/2}}\right). \qquad 1.6.11$$

This falls off monotonically from $\sigma/(2\epsilon_0)$ just above the disc to zero at infinity.

1.6.6 *Field of a Uniformly Charged Infinite Plane Sheet*

All we have to do is to put $\alpha = \pi/2$ in equation 1.6.10 to obtain

$$E = \frac{\sigma}{2\epsilon_0}. \qquad 1.6.12$$

This is independent of the distance of P from the infinite charged sheet. The electric field lines are uniform parallel lines extending to infinity.

Summary

Point charge Q: $\quad E = \dfrac{Q}{4\pi\epsilon_0 r^2}.$

Hollow Spherical Shell: $\quad E$ = zero inside the shell,

$E = \dfrac{Q}{4\pi\epsilon_0 r^2}$ outside the shell.

Infinite charged rod: $\quad E = \dfrac{\lambda}{2\pi\epsilon_0 r}.$

Infinite plane sheet: $\quad E = \dfrac{\sigma}{2\epsilon_0}.$

1.7 *Electric Field* ***D***

We have been assuming that all "experiments" described have been carried out in a vacuum or (which is almost the same thing) in air. But what if the point charge, the infinite rod and the infinite charged sheet of section 1.6 are all immersed in some medium whose permittivity is not ϵ_0, but is instead ϵ? In that case, the formulas for the field become

$$E = \frac{Q}{4\pi\epsilon r^2} \quad , \quad \frac{\lambda}{2\pi\epsilon\, r} \quad , \quad \frac{\sigma}{2\epsilon} .$$

There is an ϵ in the denominator of each of these expressions. When dealing with media with a permittivity other than ϵ_0 it is often convenient to describe the electric field by another vector, **D**, defined simply by

$$\mathbf{D} = \epsilon\mathbf{E} \qquad 1.7.1$$

In that case the above formulas for the field become just

$$D = \frac{Q}{4\pi r^2} \quad , \quad \frac{\lambda}{2\pi\, r} \quad , \quad \frac{\sigma}{2} .$$

The dimensions of D are Q L^{-2}, and the SI units are C m^{-2}.

This may seem to be rather trivial, but it does turn out to be more important than it may seem at the moment.

Equation 1.7.1 would seem to imply that the electric field vectors **E** and **D** are just vectors in the same direction, differing in magnitude only by the scalar quantity ϵ. This is indeed the case *in vacuo* or in any isotropic medium – but it is more complicated in an anisotropic medium such as, for example, an orthorhombic crystal. This is a crystal shaped like a rectangular parallelepiped. If such a crystal is placed in an electric field, the magnitude of the permittivity depends on whether the field is applied in the x- , the y- or the z-direction. For a given magnitude of E, the resulting magnitude of D will be different in these three situations. And, if the field **E** is not applied parallel to one of the crystallographic axes, the resulting vector **D** will not be parallel to **E**. The permittivity in equation 1.7.1 is a *tensor* with nine components, and, when applied to **E** it changes its direction as well as its magnitude.

However, we shan't dwell on that just yet, and, unless specified otherwise, we shall always assume that we are dealing with a vacuum (in which case $\mathbf{D} = \epsilon_0\mathbf{E}$) or an isotropic medium (in which case $\mathbf{D} = \varepsilon\mathbf{E}$). In either case the permittivity is a scalar quantity and $\mathbf{D}$ and $\mathbf{E}$ are in the same direction.

1.8 *Flux*

The product of electric field intensity and area is the *flux* Φ_E. Whereas E is an *intensive* quantity, Φ_E is an *extensive* quantity. It dimensions are $ML^3T^{-2}Q^{-1}$ and its SI units are N m^2 C^{-1}, although later on, after we have met the unit called the *volt*, we shall prefer to express Φ_E in V m.

With increasing degrees of sophistication, flux may be defined mathematically as:

$$\Phi_E = EA$$

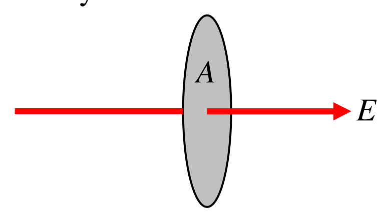


FIGURE I.4

$$\Phi_E = EA\cos\theta = \mathbf{E} \cdot \mathbf{A}$$

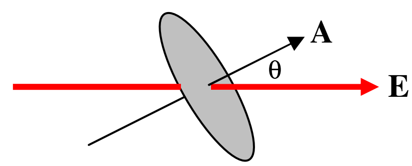


FIGURE I.5

Note that $\mathbf{E}$ is a vector, but Φ_E is a scalar.

$$\Phi_E = \iint \mathbf{E} \cdot d\mathbf{A}$$

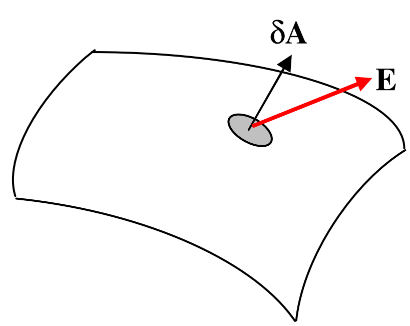


FIGURE I.6

We can also define a *D*-flux by $\Phi_D = \iint \mathbf{D} \cdot d\mathbf{A}$. The dimensions of Φ_D are just Q and the SI units are coulombs (C).

An example is in order:

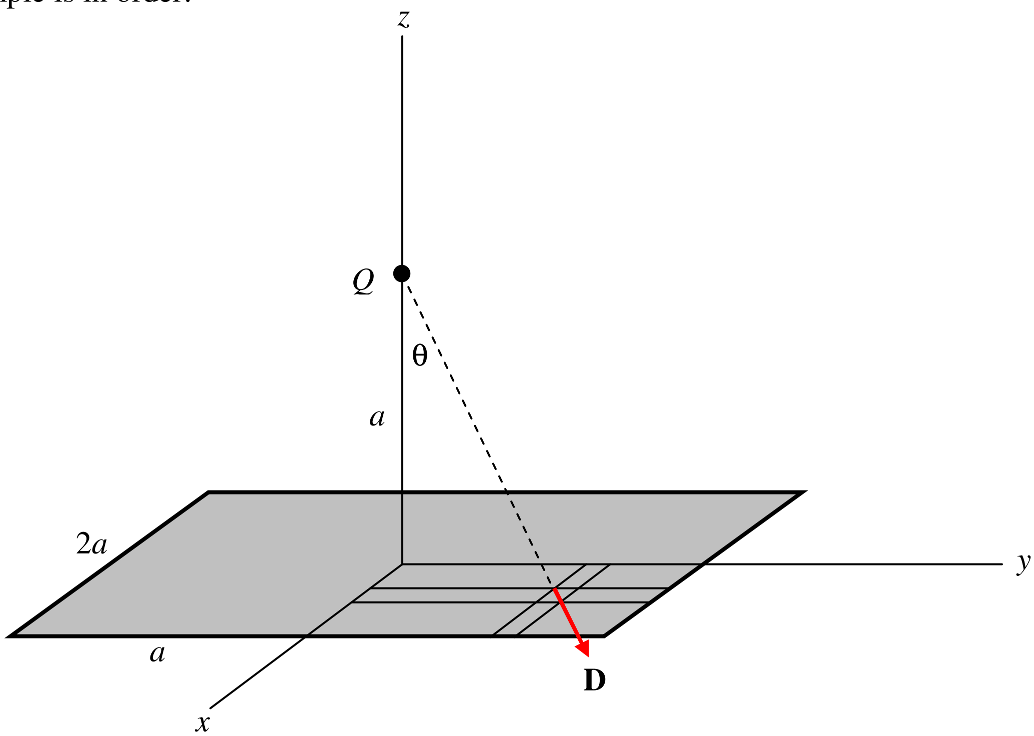


FIGURE I.7

Consider a square of side 2*a* in the *xy*-plane as shown. Suppose there is a positive charge *Q* at a height *a* on the *z*-axis. Calculate the total *D*-flux, Φ_D through the area.

Consider an elemental area *dxdy* at (x, y , 0). Its distance from *Q* is $(a^2 + x^2 + y^2)^{1/2}$, so the magnitude of the *D*-field there is $\frac{Q}{4\pi}\cdot\frac{1}{a^2 + x^2 + y^2}$. The scalar product of this with the area is $\frac{Q}{4\pi}\cdot\frac{1}{a^2 + x^2 + y^2}\cdot\cos\theta\, dxdy$, and $\cos\theta = \frac{a}{(a^2 + x^2 + y^2)^{1/2}}$. The surface integral of **D** over the whole area is

$$\iint D\bullet dA = \frac{Qa}{\pi}\int_0^a\int_0^a \frac{dxdy}{(a^2 + x^2 + y^2)^{3/2}}\,. \qquad 1.8.1$$

Now all we have to do is the nice and easy integral. Let $x = \sqrt{a^2 + y^2}\tan\psi$, and the inner integral $\int_0^a \frac{dx}{(a^2 + x^2 + y^2)^{3/2}}$ reduces, after some modest algebra, to $\frac{a}{(a^2 + y^2)\sqrt{2a^2 + y^2}}$. Thus we now have

$$\iint \mathbf{D} \cdot d\mathbf{A} = \frac{Qa^2}{\pi}\int_0^a \frac{dy}{(a^2 + y^2)\sqrt{2a^2 + y^2}}. \qquad 1.8.2$$

With the further substitution $a^2 + y^2 = a^2 \sec\omega$, this reduces, after more careful algebra, to

$$\iint \mathbf{D} \cdot d\mathbf{A} = \frac{Q}{6}. \qquad 1.8.3$$

Two additional examples of calculating surface integrals may be found in Chapter 5, section 5.6, of the Celestial Mechanics section of these notes. These deal with gravitational fields, but they are essentially the same as the electrostatic case; just substitute Q for m and $-1/(4\pi\epsilon)$ for G.

I urge readers actually to go through the pain and the algebra and the trigonometry of these three examples in order that they may appreciate all the more, in the next section, the power of Gauss's theorem.

1.9 *Gauss's Theorem*

A point charge Q is at the centre of a sphere of radius r. Calculate the D-flux through the sphere. Easy. The magnitude of D at a distance a is $Q/(4\pi r^2)$ and the surface area of the sphere is $4\pi r^2$. Therefore the flux is just Q. Notice that this is independent of r; if you double r, the area is four times as great, but D is only a quarter of what it was, so the total flux remains the same. You will probably agree that if the charge is surrounded by a shape such as shown in figure I.8, which is made up of portions of spheres of different radii, the D-flux through the surface is still just Q. And you can distort the surface as much as you like, or you may consider any surface to be made up of an infinite number of infinitesimal spherical caps, and you can put the charge anywhere you like inside the surface, or indeed you can put as many charges inside as you like – you haven't changed the total normal component of the flux, which is still just Q. This is Gauss's theorem, which is a consequence of the *inverse square* nature of Coulomb's law.

The total normal component of the D-flux through any closed surface is equal to the charge enclosed by that surface.

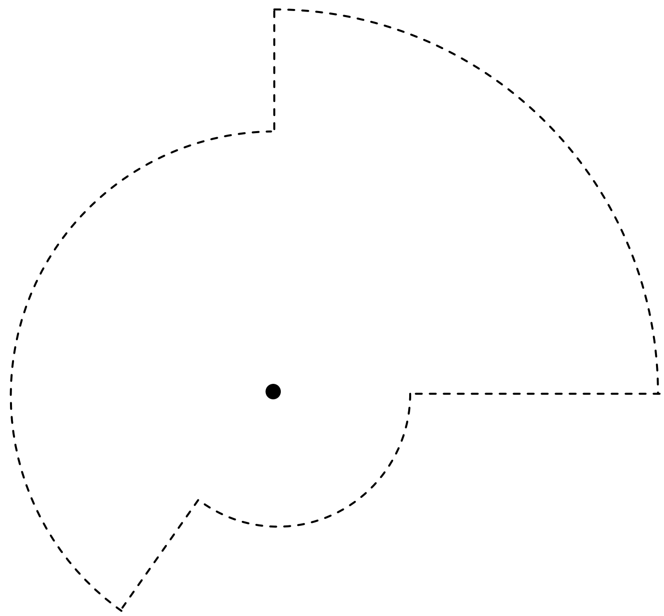


FIGURE I.8

Examples

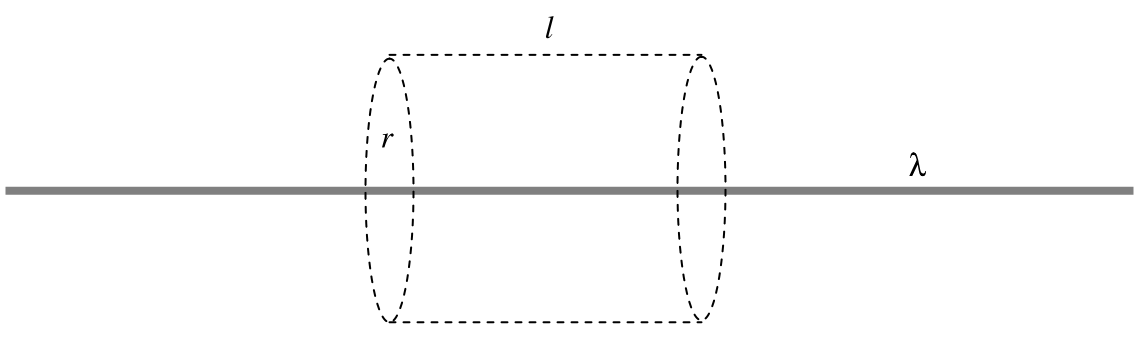


FIGURE I.9

A long rod carries a charge of λ per unit length. Construct around it a cylindrical surface of radius r and length l. The charge enclosed is $l\lambda$, and the field is directed radially outwards, passing only through the curved surface of the cylinder. The D-flux through the cylinder is $l\lambda$ and the area of the curved surface is $2\pi rl$, so $D = l\lambda/(2\pi rl)$ and hence $E = \lambda/(2\pi\varepsilon r)$.

A flat plate carries a charge of σ per unit area. Construct around it a cylindrical surface of cross-sectional area A. The charge enclosed by the cylinder is $A\sigma$, so this is the D-flux through the cylinder. It all goes through the two ends of the cylinder, which have a total area $2A$, and therefore $D = \sigma/2$ and $E = \sigma/(2\varepsilon)$.

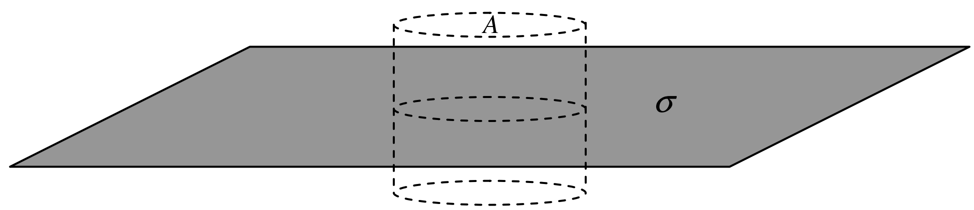


FIGURE I.10

A hollow spherical shell of radius a carries a charge Q. Construct two gaussian spherical surfaces, one of radius less than a and the other of radius $r > a$. The smaller of these two surfaces has no charge inside it; therefore the flux through it is zero, and so E is zero. The charge through the larger sphere is Q and is area is $4\pi r^2$. Therefore $D = Q/(4\pi r^2)$ and $E = Q/(4\pi\varepsilon r^2)$. (It is worth going to Chapter 5 of Celestial Mechanics, subsection 5.4.8, to go through the calculus derivation, so that you can appreciate Gauss's theorem all the more.)

A point charge Q is in the middle of a cylinder of radius a and length $2l$. Calculate the flux through the cylinder.

An infinite rod is charged with λ coulombs per unit length. It passes centrally through a spherical surface of radius a. Calculate the flux through the spherical surface.

These problems are done by calculus in section 5.6 of Celestial Mechanics, and furnish good examples of how to do surface integrals, and I recommend that you work through them. However, it is obvious from Gauss's theorem that the answers are just Q and $2a\lambda$ respectively.

A point charge Q is in the middle of a cube of side $2a$. The flux through the cube is, by Gauss's theorem, Q, and the flux through one face is $Q/6$. I hope you enjoyed doing this by calculus in section 1.8.

CHAPTER 2
ELECTROSTATIC POTENTIAL

2.1 *Introduction*

Imagine that some region of space, such as the room you are sitting in, is permeated by an electric field. (Perhaps there are all sorts of electrically charged bodies outside the room.) If you place a small positive test charge somewhere in the room, it will experience a force $\mathbf{F} = Q\mathbf{E}$. If you try to move the charge from point A to point B against the direction of the electric field, you will have to do work. If work is required to move a positive charge from point A to point B, there is said to be an electrical *potential difference* between A and B, with point A being at the lower potential. If one joule of work is required to move one coulomb of charge from A to B, the potential difference between A and B is one *volt* (V).

The dimensions of potential difference are $\mathrm{ML^2T^{-2}Q^{-1}}$.

All we have done so far is to define the potential *difference* between *two points.* We cannot define "the" potential at *a point* unless we arbitrarily assign some reference point as having a defined potential. It is not always necessary to do this, since we are often interested only in the potential differences between point, but in many circumstances it is customary to define the potential to be *zero* at an *infinite distance* from any charges of interest. We can then say what "the" potential is at some nearby point. Potential and potential difference are scalar quantities.

Suppose we have an electric field E in the positive x-direction (towards the right). This means that potential is decreasing to the right. You would have to do work to move a positive test charge Q to the left, so that potential is increasing towards the left. The force on Q is QE, so the work you would have to do to move it a distance dx to the right is $-QE\ dx$, but by definition this is also equal to $Q\ dV$, where dV is the potential difference between x and $x + dx$.

Therefore
$$E = -\frac{dV}{dx}\,. \qquad 2.1.1$$

In a more general three-dimensional situation, this is written

$$\mathbf{E} = -\mathbf{grad}\,V = -\nabla V = -\left(\mathbf{i}\frac{\partial V}{\partial x} + \mathbf{j}\frac{\partial V}{\partial x} + \mathbf{k}\frac{\partial V}{\partial x}\right). \qquad 2.1.2$$

We see that, as an alternative to expressing electric field strength in newtons per coulomb, we can equally well express it in volts per metre ($\mathrm{V\ m^{-1}}$).

The inverse of equation 2.1.1 is, of course,

$$V = -\int E\,dx + \text{constant}. \qquad 2.1.3$$

2.2 *Potential Near Various Charged Bodies*

2.2.1 *Point Charge*

Let us arbitrarily assign the value zero to the potential at an infinite distance from a point charge Q. "The" potential at a distance r from this charge is then the work required to move a unit positive charge from infinity to a distance r.

At a distance x from the charge, the field strength is $\frac{Q}{4\pi\epsilon_0 x^2}\cdot$ The work required to move a unit charge from x to $x + \delta x$ is $-\frac{Q\,\delta x}{4\pi\epsilon_0 x^2}$. The work required to move unit charge from r to infinity is $-\frac{Q}{4\pi\epsilon_0}\int_r^\infty \frac{dx}{x^2} = -\frac{Q}{4\pi\epsilon_0 r}\cdot$ The work required to move unit charge from infinity to r is minus this.

Therefore
$$V = +\frac{Q}{4\pi\epsilon_0 r}\cdot \qquad 2.2.1$$

The *mutual potential energy* of two charges Q_1 and Q_2 separated by a distance r is the work required to bring them to this distance apart from an original infinite separation. This is

$$\text{P.E.} = +\frac{Q_1 Q_2}{4\pi\epsilon_0 r}\cdot \qquad 2.2.2$$

Before proceeding, a little review is in order.

Field at a distance r from a charge Q:

$$E = \frac{Q}{4\pi\epsilon_0 r^2}, \qquad \text{N C}^{-1} \ \text{ or } \ \text{V m}^{-1}$$

or, in vector form,
$$\mathbf{E} = \frac{Q}{4\pi\epsilon_0 r^2}\hat{\mathbf{r}} = \frac{Q}{4\pi\epsilon_0 r^3}\mathbf{r}. \qquad \text{N C}^{-1} \ \text{ or } \ \text{V m}^{-1}$$

Force between two charges, Q_1 and Q_2:

$$F = \frac{Q_1 Q_2}{4\pi\varepsilon r^2}. \qquad \text{N}$$

Potential at a distance r from a charge Q:

$$V = \frac{Q}{4\pi\varepsilon_0 r}. \qquad \text{V}$$

Mutual potential energy between two charges:

$$\text{P.E.} = \frac{Q_1 Q_2}{4\pi\varepsilon_0 r}. \qquad \text{J}$$

We couldn't possibly go wrong with any of these, could we?

2.2.2 *Spherical Charge Distributions*

Outside any spherically-symmetric charge distribution, the field is the same as if all the charge were concentrated at a point in the centre, and so, then, is the potential. Thus

$$V = \frac{Q}{4\pi\varepsilon_0 r}. \qquad 2.2.3$$

Inside a *hollow spherical shell* of radius a and carrying a charge Q the field is zero, and therefore the potential is uniform throughout the interior, and equal to the potential on the surface, which is

$$V = \frac{Q}{4\pi\varepsilon_0 a}. \qquad 2.2.4$$

A solid sphere of radius a bearing a charge Q that is uniformly distributed throughout the sphere is easier to imagine than to achieve in practice, but, for all we know, a proton might be like this (it might be – but it isn't!), so let's calculate the field at a point P inside the sphere at a distance r $(< a)$ from the centre. See figure II.1

We can do this in two parts. First the potential from the part of the sphere "below" P. If the charge is uniformly distributed throughout the sphere, this is just $\frac{Q_r}{4\pi\varepsilon_0 r}$. Here Q_r is the charge contained within radius r, which, if the charge is uniformly distributed throughout the sphere, is $Q(r^3/a^3)$. Thus, that part of the potential is $\dfrac{Qr^2}{4\pi\epsilon_0 a^3}\cdot$

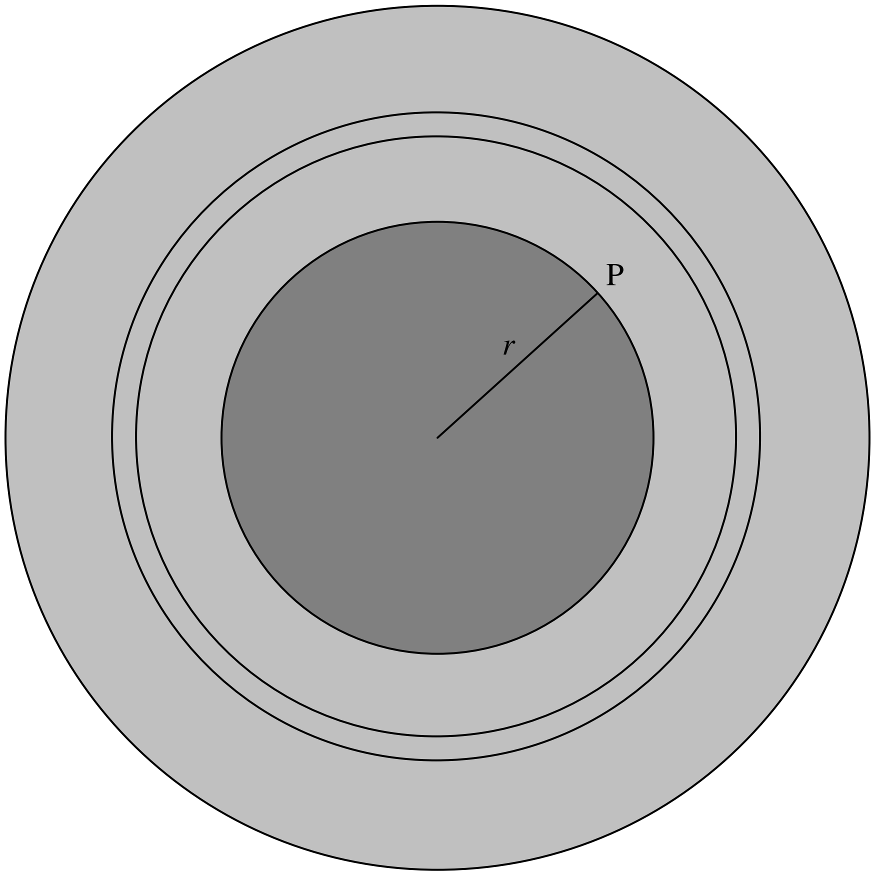


FIGURE II.1

Next, we calculate the contribution to the potential from the charge "above" P. Consider an elemental shell of radii x , $x + \delta x$. The charge held by it is $\delta Q = \dfrac{4\pi x^2\,\delta x}{\frac{4}{3}\pi a^3}\times Q = \dfrac{3Qx^2\,\delta x}{a^3}\cdot$ The contribution to the potential at P from the charge in this elemental shell is $\dfrac{\delta Q}{4\pi\epsilon_0 x} = \dfrac{3Q\,x\,\delta x}{4\pi\epsilon_0 a^3}\cdot$ The contribution to the potential from all the charge "above" P is $\dfrac{3Q}{4\pi\epsilon_0 a^3}\displaystyle\int_r^a x\,dx = \dfrac{3Q(a^2 - r^2)}{4\pi\epsilon_0 .2a^3}\cdot$ Adding together the two parts of the potential, we obtain

$$V = \frac{Q}{8\pi\epsilon_0 a^3}(3a^2 - r^2). \qquad 2.2.5$$

2.2.3 *Long Charged Rod*

The field at a distance r from a long charged rod carrying a charge λ coulombs per metre is $\frac{\lambda}{2\pi\epsilon_0 r}$. Therefore the potential difference between two points at distances a and b from the rod $(a < b)$ is

$$V_b - V_a = -\frac{\lambda}{2\pi\epsilon_0}\int_a^b \frac{dr}{r}.$$

$$\therefore \qquad V_a - V_b = \frac{\lambda}{2\pi\epsilon_0}\ln(b/a). \qquad 2.2.6$$

2.2.4 *Large Plane Charged Sheet*

The field at a distance r from a large charged sheet carrying a charge σ coulombs per square metre is $\frac{\sigma}{2\epsilon_0}$. Therefore the potential difference between two points at distances a and b from the sheet $(a < b)$ is

$$V_a - V_b = \frac{\sigma}{2\epsilon_0}(b - a). \qquad 2.2.7$$

2.2.5 *Potential on the Axis of a Charged Ring*

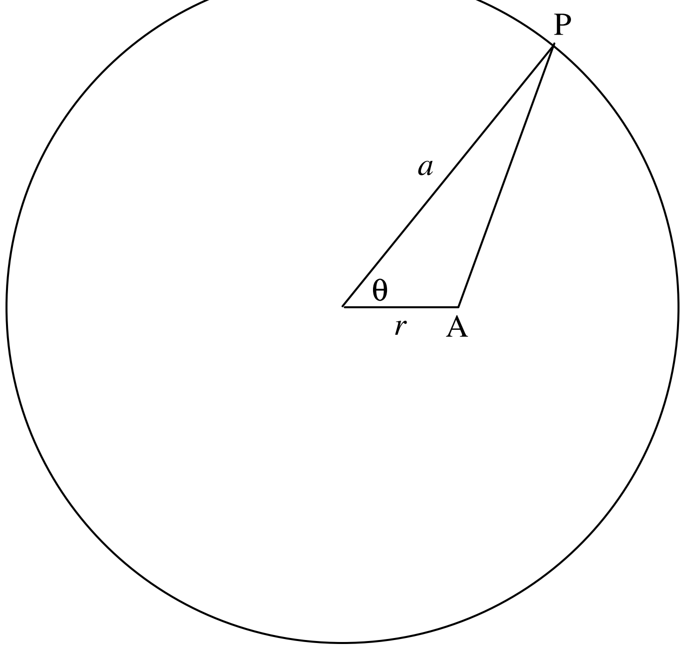
The field on the axis of a charged ring is given in section 1.6.4. The reader is invited to show that the potential on the axis of the ring is

$$V = \frac{Q}{4\pi\epsilon_0(a^2 + x^2)^{1/2}}. \qquad 2.2.8$$

You can do this either by integrating the expression for the field or just by thinking about it for a few seconds and realizing that potential is a scalar quantity.

2.2.6 *Potential in the Plane of a Charged Ring*

We suppose that we have a ring of radius a bearing a charge Q. We shall try to find the potential at a point in the plane of the ring and at a distance r $(0 \le r < a)$ from the centre of the ring.



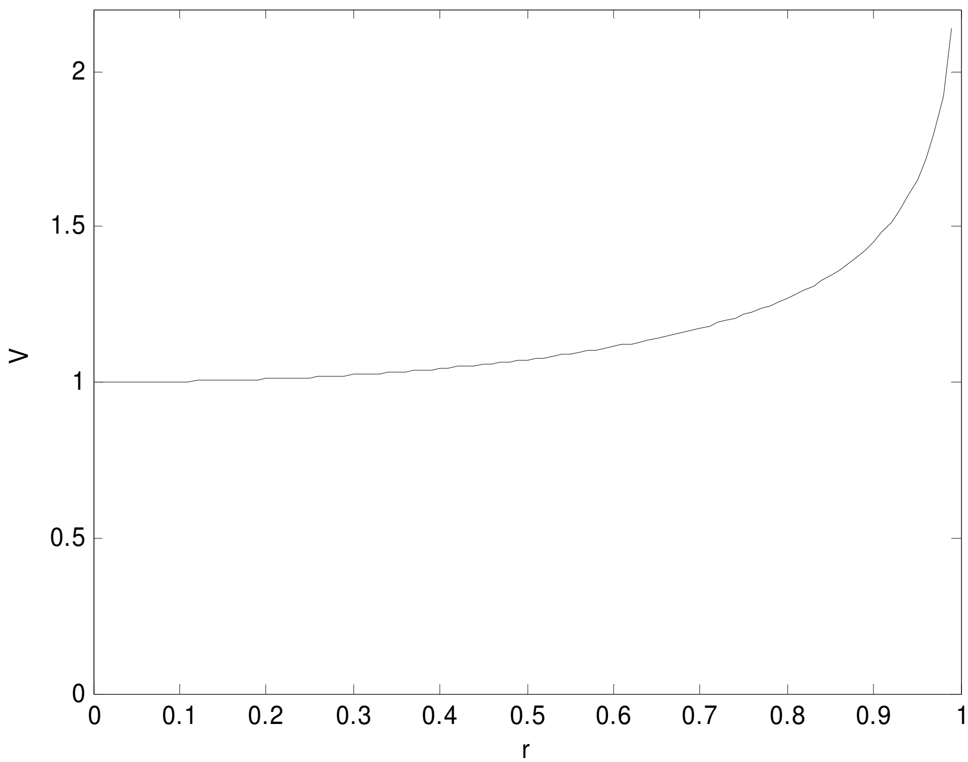
Consider an element $\delta\theta$ of the ring at P. The charge on it is $\dfrac{Q\delta\theta}{2\pi}$. The potential at A due this element of charge is

$$\frac{1}{4\pi\epsilon_0}\cdot\frac{Q\delta\theta}{2\pi}\cdot\frac{1}{\sqrt{a^2 + r^2 - 2ar\cos\theta}} = \frac{Q}{4\pi\epsilon_0 . 2\pi a}\cdot\frac{\delta\theta}{\sqrt{b - c\cos\theta}}, \qquad 2.2.9$$

where $b = 1 + r^2/a^2$ and $c = 2r/a$. The potential due to the charge on the entire ring is

$$V = \frac{Q}{4\pi\epsilon_0 . \pi a}\int_0^\pi \frac{d\theta}{\sqrt{b - c\cos\theta}}\cdot \qquad 2.2.10$$

I can't immediately see an analytical solution to this integral, so I integrated it numerically from $r = 0$ to $r = 0.99$ in steps of 0.01, with the result shown in the following graph, in which r is in units of a, and V is in units of $\dfrac{Q}{4\pi\epsilon_0 a}$.



The field is equal to the gradient of this and is directed towards the centre of the ring. It looks as though a small positive charge would be in stable equilibrium at the centre of the ring, and this would be so if the charge were constrained to remain in the plane of the ring. But, without such a constraint, the charge would be pushed away from the ring if it strayed at all above or below the plane of the ring.

Some computational notes.

Any reader who has tried to reproduce these results will have discovered that rather a lot of heavy computation is required. Since there is no simple analytical expression for the integration, each of the 100 points from which the graph was computed entailed a numerical integration of the expression for the potential. I found that Simpson's Rule did not give very satisfactory results, mainly because of the steep rise in the function at large r, so I used Gaussian quadrature, which proved much more satisfactory.

Can we avoid the numerical integration? One possibility is to express the integrand in equation 2.2.10 as a power series in $\cos\theta$, and then integrate term by term.

Thus $\sqrt{b - c\cos\theta} = \sqrt{b}.\sqrt{1 - e\cos\theta}$, where $e = \frac{c}{b} = \frac{2(r/a)}{(r/a)^2 + 1}\cdot$ And then

$$\sqrt{1 - e\cos\theta} = 1 + \tfrac{1}{2}e\cos\theta + \tfrac{3}{8}e^2\cos^2\theta + \tfrac{5}{16}e^3\cos^3\theta + \tfrac{35}{128}e^4\cos^4\theta$$
$$+ \tfrac{231}{1024}e^5\cos^5\theta + \tfrac{63}{256}e^6\cos^6\theta + \tfrac{231}{1024}e^7\cos^7\theta + \tfrac{715}{32768}e^8\cos^8\theta + \ldots \qquad 2.2.11$$

We can then integrate this term by term, using $\int_0^\pi \cos^n\theta d\theta = \dfrac{(n-1)!!\pi}{n!!}$ if n is even, and obviously zero if n is odd.

We finally get:

$$V = \frac{Q}{4\pi\epsilon_0 a}(1 + \tfrac{3}{16}e^2 + \tfrac{105}{1024}e^4 + \tfrac{1155}{16384}e^6 + \tfrac{25025}{4194304}e^8 \ldots). \qquad 2.2.12$$

For computational purposes, this is most efficiently rendered as

$$V = \frac{Q}{4\pi\epsilon_0 a}(1 + e^2(\tfrac{3}{16} + e^4(\tfrac{105}{1024} + e^6(\tfrac{1155}{16384} + \tfrac{25025}{4194304}e^8)))). \qquad 2.2.14$$

I shall refer to this as Series I. It turns out that it is not a very efficient series, as it converges very slowly. This is because e is not a small fraction, and is always greater than r/a. Thus for $r/a = \frac{1}{2}$, $e = 0.8$.

We can do much better if we can obtain a power series in r/a. Consider the expression $\dfrac{1}{\sqrt{a^2 + r^2 - 2ar\cos\theta}} = \dfrac{1}{a\sqrt{1 + (r/a)^2 - 2(r/a)\cos\theta}}$, which occurs in equation 2.2.9. This expression, and others very similar to it, occur quite frequently in various physical situations. It can be expanded by the binomial theorem to give a power series in r/a. (Admittedly, it is a trinomial expression, but do it in stages). The result is

$$(1 + (r/a)^2 - 2(r/a)\cos\theta)^{-1/2} = P_0(\cos\theta) + P_1(\cos\theta)(\tfrac{r}{a}) + P_2(\cos\theta)(\tfrac{r}{a})^2 + P_3(\cos\theta)(\tfrac{r}{a})^3 + \ldots \qquad 2.2.15$$

where the coefficients of the powers of $(\frac{r}{a})$ are polynomials in cos θ, which have been extensively tabulated in many places, and are called *Legendre polynomials*. See, for example my notes on Celestial Mechanics, http://orca.phys.uvic.ca/~tatum/celmechs.html Sections 1.1.4 and 5.11. Each term in the Legendre polynomials can then be integrated term by term, and the resulting series, after a bit of work, is

$$V = \frac{Q}{4\pi\epsilon_0 a}(1 + \tfrac{1}{4}(\tfrac{r}{a})^2 + \tfrac{9}{64}(\tfrac{r}{a})^4 + \tfrac{25}{256}(\tfrac{r}{a})^6 + \tfrac{1225}{16384}(\tfrac{r}{a})^8 \ldots). \qquad 2.2.16$$

Since this is a series in $(\frac{r}{a})$ rather than is e, it converges much faster than equation 2.2.13. I shall refer to it as series II. Of course, for computational purposes it should be written with nested parentheses, as we did for series I in equation 2.2.14.

Here is a table of the results using four methods. The first column gives the value of r/a. The next four columns give the values of V, in units of $\frac{Q}{4\pi\epsilon_0 a}$, calculated by four methods. Column 2, integration by Gaussian quadrature. Column 3, integration by Simpson's Rule. Column 4, approximation by Series I. Column 5, approximation by series II. In each case I have given the number of digits that I believe to be reliable. It is seen that Gaussian quadrature gives by far the best results. Series I is not very good at all, while Series II is almost as good as Simpson's Rule.

0.0	1.000 000 000	1.000 000 000	1.000 000 000	1.000 000 000
0.1	1.002 514 161	1.002 514	1.000 514 161	1.002 514
0.2	1.010 231 448	1.010 23	1.010 231 4	1.010 2
0.3	1.023 715 546	1.023 7	1.023 72	1.02
0.4	1.044 056 341	1.0441	1.044	1.04
0.5	1.073 182 007	1 .732	1.073	
0.6	1.114 564 487	1.115	1.11	
0.7	1.175 005	1.175	1.17	
0.8	1.270 25	1.270	1.3	
0.9	1.451 8	1.452	1.4	

Of course any of these methods is completed almost instantaneously on a modern computer, so one may wonder if it is worthwhile spending much time seeking the most efficient solution. That will depend on whether one wants to do the calculation just once, or whether one wants to do similar calculations millions of times.

2.2.7 *Potential on the Axis of a Charged Disc*

The field on the axis of a charged disc is given in section 1.6.5. The reader is invited to show that the potential on the axis of the disc is

$$V = \frac{2Q}{4\pi\epsilon_0 a^2}[(a^2 + x^2)^{1/2} - x]. \qquad 2.2.9$$

2.3 *Electron-volts*

The *electron-volt* is a unit of *energy* or *work*. An electron-volt (eV) is the work required to move an electron through a potential difference of one volt. Alternatively, an electron-volt is equal to the kinetic energy acquired by an electron when it is accelerated through a potential difference of one volt. Since the magnitude of the charge of an electron is about 1.602×10^{-19} C, it follows that an electron-volt is about 1.602×10^{-19} J. Note also that, because the charge on an electron is negative, it requires work to move an electron from a point of high potential to a point of low potential.

Exercise. If an electron is accelerated through a potential difference of a million volts, its kinetic energy is, of course, 1 MeV. At what speed is it then moving?

First attempt. $\frac{1}{2}m\upsilon^2 = eV.$

(Here eV, written in italics, is not intended to mean the unit electron-volt, but e is the magnitude of the electron charge, and V is the potential difference (10^6 volts) through which it is accelerated.) Thus $\upsilon = \sqrt{2eV/m}$. With $m = 9.109 \times 10^{-31}$ kg, this comes to $\upsilon = 5.9 \times 10^8$ m s^{-1}. Oops! That looks awfully fast! We'd better do it properly this time.

Second attempt. $(\gamma - 1)mc^2 = eV.$

Some readers will know exactly what we are doing here, without explanation. Others may be completely mystified. For the latter, the difficulty is that the speed that we had calculated was even greater than the speed of light. To do this properly we have to use the formulas of special relativity. See, for example, Chapter 15 of the Classical Mechanics section of these notes.

At any rate, this results in $\gamma = 2.958$, whence $\beta = 0.9411$ and $\upsilon = 2.82 \times 10^8$ m s^{-1}.

2.4 *A Point Charge and an Infinite Conducting Plane*

An infinite plane metal plate is in the xy-plane. A point charge $+Q$ is placed on the z-axis at a height h above the plate. Consequently, electrons will be attracted to the part of the plate immediately below the charge, so that the plate will carry a negative charge density σ which is greatest at the origin and which falls off with distance ρ from the origin. Can we determine $\sigma(\rho)$? See figure II.2

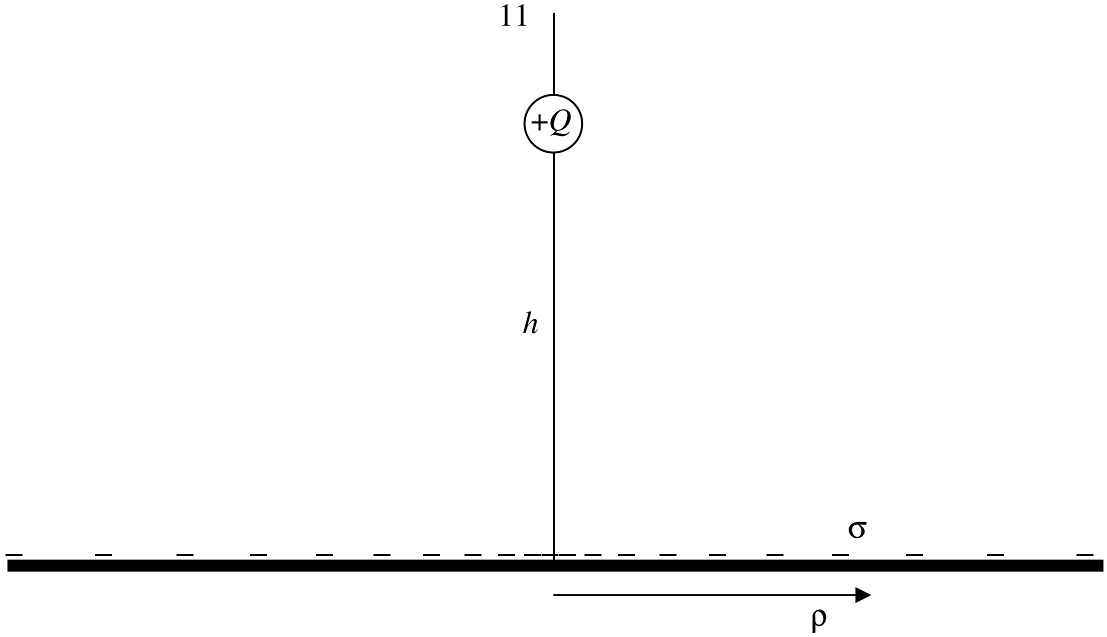


FIGURE II.2

First, note that the metal surface, being a conductor, is an *equipotential* surface, as is any metal surface. The potential is uniform anywhere on the surface. Now suppose that, instead of the metal surface, we had (in addition to the charge $+Q$ at a height h above the xy-plane), a second point charge, $-Q$, at a distance h below the xy-plane. The potential in the xy-plane would, by symmetry, be uniform everywhere. That is to say that the potential in the xy-plane is the same as it was in the case of the single point charge and the metal plate, and indeed the potential at any point above the plane is the same in both cases. For the purpose of calculating the potential, we can replace the metal plate by an *image* of the point charge. It is easy to calculate the potential at a point $(z\ ,\ \rho)$. If we suppose that the permittivity above the plate is ε_0, the potential at $(z\ ,\ \rho)$ is

$$V \;=\; \frac{Q}{4\pi\varepsilon_0}\left(\frac{1}{[\rho^2 + (h-z)^2]^{1/2}} - \frac{1}{[\rho^2 + (h+z)^2]^{1/2}}\right). \qquad 2.4.1$$

The field strength E in the xy-plane is $-\partial V/\partial z$ evaluated at $z = 0$, and this is

$$E \;=\; -\frac{2Q}{4\pi\varepsilon_0}\cdot\frac{h}{(\rho^2 + h^2)^{3/2}}\ . \qquad 2.4.2$$

The D-field is ε_0 times this, and since all the lines of force are above the metal plate, Gauss's theorem provides that the charge density is $\sigma = D$, and hence the charge density is

$$\sigma \;=\; -\frac{Q}{2\pi}\cdot\frac{h}{(\rho^2 + h^2)^{3/2}}\ . \qquad 2.4.3$$

This can also be written $$\sigma = -\frac{Q}{2\pi}\cdot\frac{h}{\xi^3}, \qquad 2.4.4$$

where $\xi^2 = \rho^2 + h^2$, with obvious geometric interpretation.

Exercise: How much charge is there on the surface of the plate within an annulus bounded by radii ρ and $\rho + d\rho$? Integrate this from zero to infinity to show that the total charge induced on the plate is $-Q$.

2.5 *A Point Charge and a Conducting Sphere*

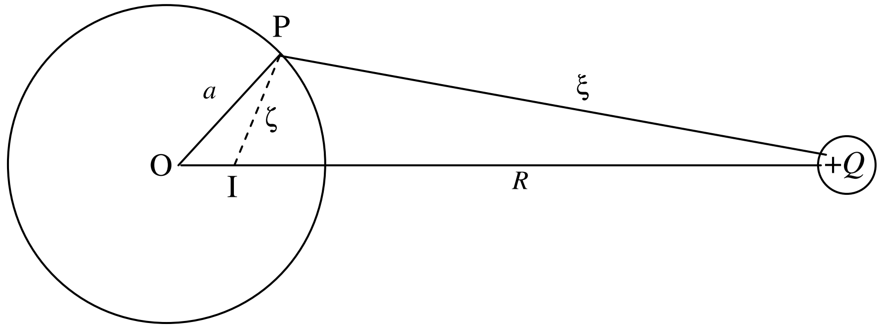


FIGURE II.3

A point charge $+Q$ is at a distance R from a metal sphere of radius a. We are going to try to calculate the surface charge density induced on the surface of the sphere, as a function of position on the surface. We shall bear in mind that the surface of the sphere is an equipotential surface, and we shall take the potential on the surface to be zero.

Let us first construct a point I such that the triangles OPI and PQO are similar, with the lengths shown in figure II.3. The length OI is a^2/R. Then $R/\xi = a/\zeta$, or

$$\frac{1}{\xi} - \frac{a/R}{\zeta} = 0. \qquad 2.5.1$$

This relation between the variables ξ and ζ is in effect the equation to the sphere expressed in these variables.

Now suppose that, instead of the metal sphere, we had (in addition to the charge $+Q$ at a distance R from O), a second point charge $-(a/R)Q$ at I. The locus of points where the potential is zero is where

$$\frac{Q}{4\pi\epsilon_0}\left(\frac{1}{\xi} - \frac{a/R}{\zeta}\right) = 0. \qquad 2.5.2$$

That is, the surface of our sphere. Thus, for purposes of calculating the potential, we can replace the metal sphere by an *image* of Q at I, this image carrying a charge of $-(a/R)Q$.

Let us take the line OQ as the z-axis of a coordinate system. Let X be some point such that OX = r and the angle XOQ = θ. The potential at P from a charge $+Q$ at Q and a charge $-(a/R)Q$ at I is (see figure II.4)

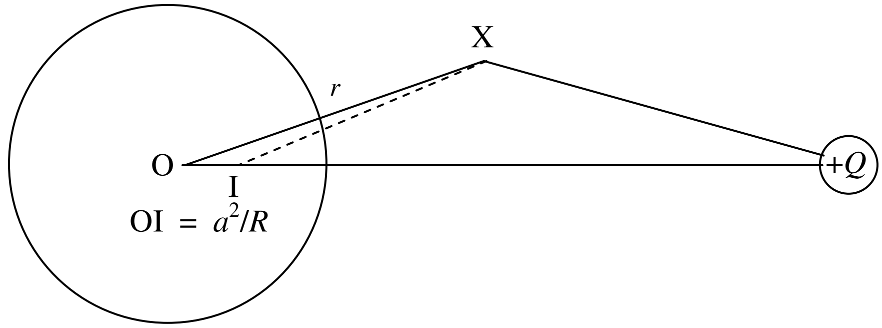


FIGURE II.4

$$V = \frac{Q}{4\pi\epsilon_0}\left(\frac{1}{(r^2+R^2-2rR\cos\theta)^{1/2}} - \frac{a/R}{(r^2+a^4/R^2-2a^2r\cos\theta/R)^{1/2}}\right). \qquad 2.5.2$$

The E field on the surface of the sphere is $-\partial V/\partial r$ evaluated at $r = a$. The D field is ϵ_0 times this, and the surface charge density is equal to D. After some patience and algebra, we obtain, for a point X on the surface of the sphere

$$\sigma = -\frac{Q}{4\pi}\cdot\frac{R^2-a^2}{a}\cdot\frac{1}{(\mathrm{XQ})^3}. \qquad 2.5.3$$

2.6 *Two Semicylindrical Electrodes*

This section requires that the reader should be familiar with functions of a complex variable and conformal transformations. For readers not familiar with these, this section can be skipped without prejudice to understanding following chapters. For readers who are familiar, this is a nice example of conformal transformations to solve a physical problem.

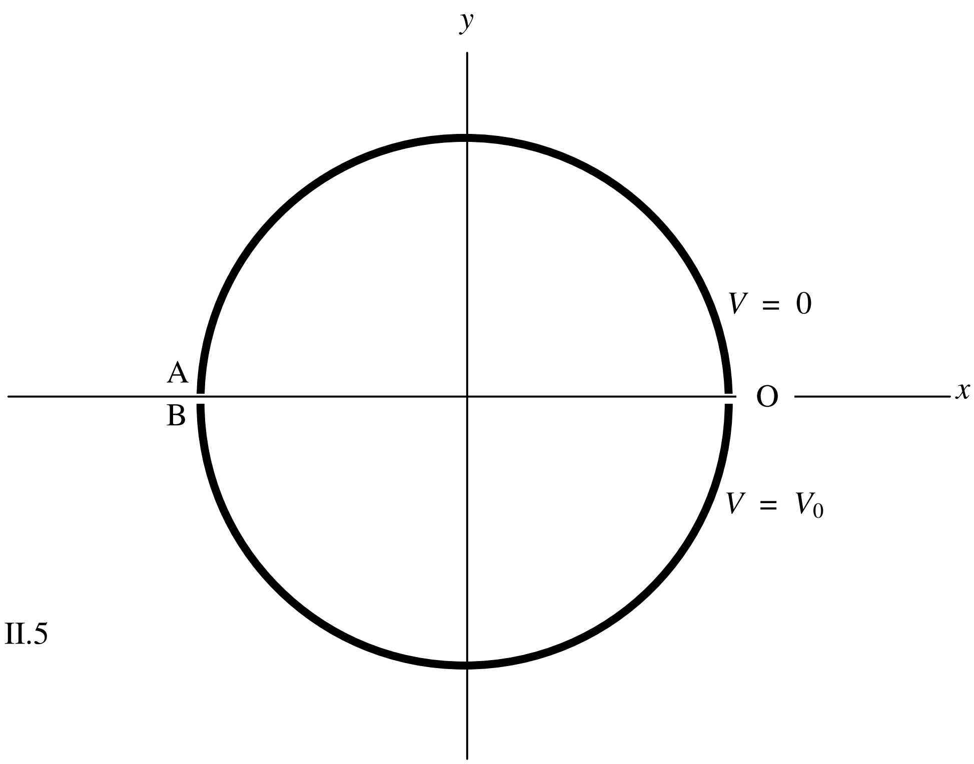


FIGURE II.5

We have two semicylindrical electrodes as shown in figure II.5. The potential of the upper one is 0 and the potential of the lower one is V_0. We'll suppose the radius of the curcle is 1; or, what amounts to the same thing, we'll express coordinates x and y in units of the radius. Let us represent the position of any point whose coordinates are (x , y) by a complex number $z = x + iy$.

Now let $w = u + iv$ be a complex number related to z by $w = i\left(\dfrac{1 - z}{1 + z}\right)$; that is, $z = \dfrac{1 + iw}{1 - iw}$. Substitute $w = u + iv$ and $z = x + iy$ in each of these equations, and equate real and imaginary parts, to obtain

$$u = \frac{2y}{(1+x)^2 + y^2}; \qquad \upsilon = \frac{1 - x^2 - y^2}{(1+x)^2 + y^2}; \qquad 2.6.1$$

$$x = \frac{1 - u^2 - \upsilon^2}{u^2 + (1+\upsilon)^2}; \qquad y = \frac{2u}{u^2 + (1+\upsilon)^2}. \qquad 2.6.2$$

In that case, the upper semicircle ($V = 0$) in the xy-plane maps on to the positive u-axis in the $u\upsilon$-plane, and the lower semicircle ($V = V_0$) in the xy-plane maps on to the negative u-axis in the $u\upsilon$-plane. (Figure II.6.) Points inside the circle bounded by the electrodes in the xy-plane map on to points above the u-axis in the $u\upsilon$-plane.

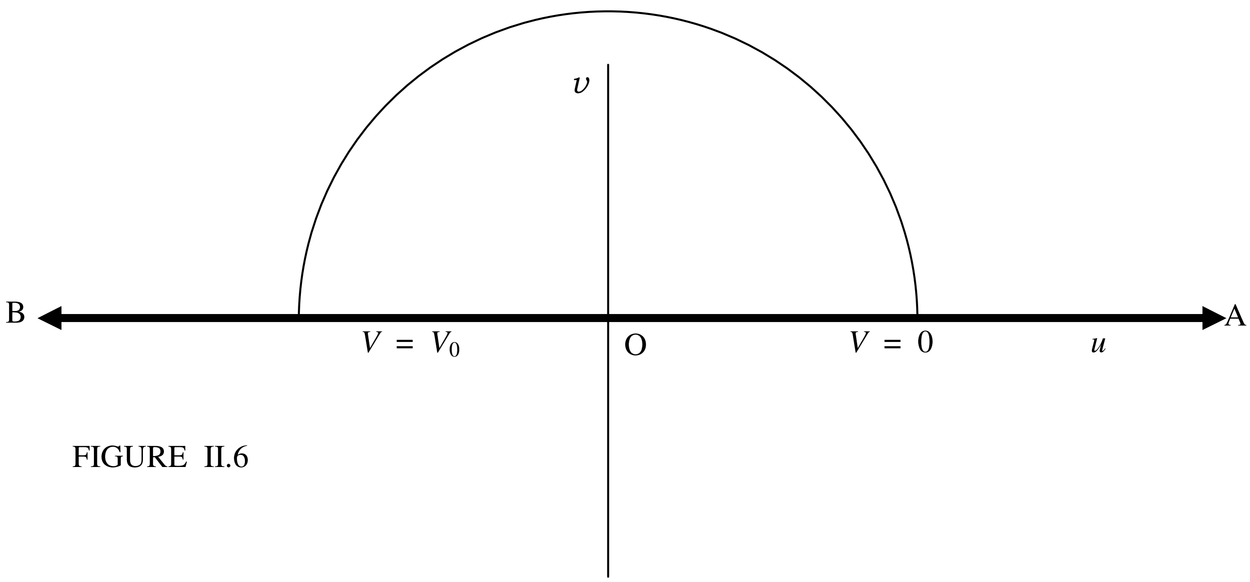


FIGURE II.6

In the $u\upsilon$-plane, the lines of force are semicircles, such as the one shown. The potential goes from 0 at one end of the semicircle to V_0 at the other, and so equation to the semicircular line of force is

$$\frac{V}{V_0} = \frac{\arg w}{\pi} \qquad 2.6.3$$

or

$$V = \frac{V_0}{\pi}\tan^{-1}(\upsilon / u). \qquad 2.6.4$$

The equipotentials (V = constant) are straight lines in the $u\upsilon$-plane of the form

$$\upsilon = fu. \qquad 2.6.5$$

(You would prefer me to use the symbol m for the slope of the equipotentials, but in a moment you will be glad that I chose the symbol f.)

If we now transform back to the xy-plane, we see that the equation to the lines of force is

$$V = \frac{V_0}{\pi} \tan^{-1}\left(\frac{1 - x^2 - y^2}{2y}\right), \qquad 2.6.6$$

and the equation to the equipotentials is

$$1 - x^2 - y^2 = 2fy, \qquad 2.6.7$$

or

$$x^2 + y^2 + 2fy - 1 = 0. \qquad 2.6.8$$

Now aren't you glad that I chose f ? Those who are handy with conic sections (see Chapter 2 of Celestial Mechanics) will understand that the equipotentials in the xy-plane are circles of radii $\sqrt{f^2 + 1}$, whose centres are at $(0, \pm f)$, and which all pass through the points $(\pm 1, 0)$. They are drawn as blue lines in figure II.7. The lines of force are the orthogonal trajectories to these, and are of the form

$$x^2 + y^2 + 2gy + 1 = 0. \qquad 2.6.9$$

These are circles of radii $\sqrt{g^2 - 1}$ and have their centres at $(0, \pm g)$. They are shown as dashed red lines in figure II.7.

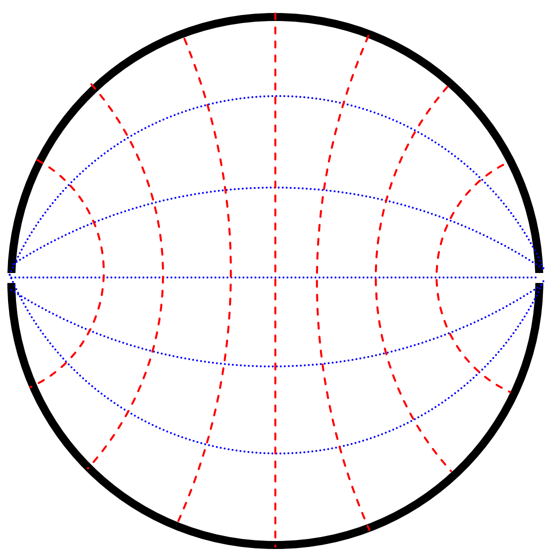


FIGURE II.7

CHAPTER 3
DIPOLE AND QUADRUPOLE MOMENTS

3.1 *Introduction*

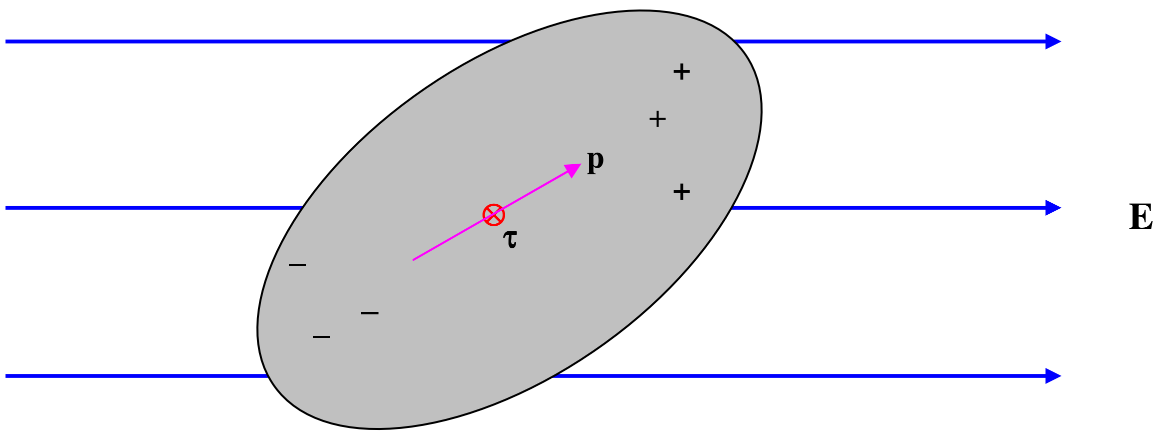


FIGURE III.1

Consider a body which is on the whole electrically neutral, but in which there is a separation of charge such that there is more positive charge at one end and more negative charge at the other. Such a body is an *electric dipole*.

Provided that the body as a whole is electrically neutral, it will experience no *force* if it is placed in a uniform external electric field, but it will (unless very fortuitously oriented) experience a *torque*. The magnitude of the torque depends on its orientation with respect to the field, and there will be two (opposite) directions in which the torque is a *maximum*.

The maximum torque that the dipole experiences when placed in an external electric field is its *dipole moment*. This is a vector quantity, and the torque is a maximum when the dipole moment is at right angles to the electric field. At a general angle, the torque $\boldsymbol{\tau}$, the dipole moment $\mathbf{p}$ and the electric field $\mathbf{E}$ are related by

$$\boldsymbol{\tau} = \mathbf{p} \times \mathbf{E}. \qquad 3.1.1$$

The SI units of dipole moment can be expressed as N m (V/m)$^{-1}$. However, work out the dimensions of p and you will find that its dimensions are Q L. Therefore it is simpler to express the dipole monent in SI units as coulomb metre, or C m.

Other units that may be encountered for expressing dipole moment are cgs esu, debye, and atomic unit. I have also heard the dipole moment of thunderclouds expressed in kilometre coulombs. A cgs esu is a centimetre-gram-second electrostatic unit. I shall describe the cgs esu system in a later chapter; suffice it here to say that a cgs esu of dipole moment is about 3.336×10^{-12} C m, and a debye (D) is 10^{-18} cgs esu. An atomic unit of electric dipole moment is $a_0 e$, where a_0 is the radius of the first Bohr orbit for hydrogen and e is the magnitude of the electronic charge. An atomic unit of dipole moment is about 8.478×10^{-29} C m.

I remark in passing that I have heard, distressingly often, some such remark as "The molecule has a dipole". Since this sentence is not English, I do not know what it is intended to mean. It would be English to say that a molecule is a dipole or that it has a dipole moment.

3.2 *Mathematical Definition of Dipole Moment*

In the introductory section 3.1 we gave a *physical* definition of dipole moment. I am now about to give a *mathematical* definition.

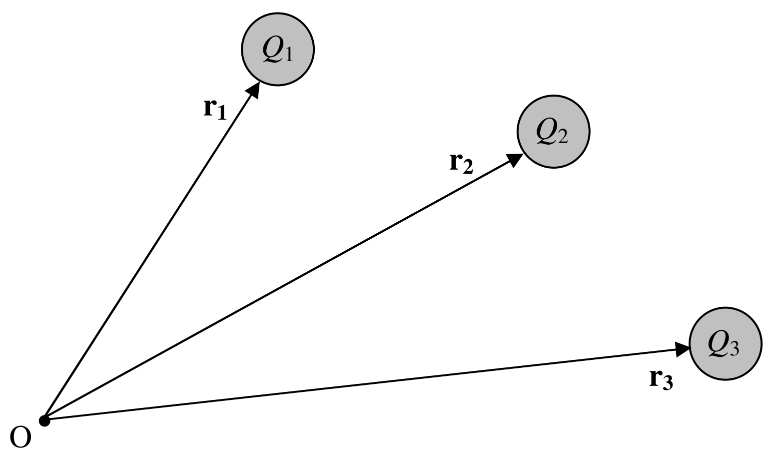


FIGURE III.2

Consider a set of charges Q_1, Q_2, Q_3 ... whose position vectors with respect to a point O are $\mathbf{r_1}$, $\mathbf{r_2}$, $\mathbf{r_3}$ ... with respect to some point O. The vector sum

$$\mathbf{p} = \sum Q_i \mathbf{r_i}$$

is the dipole moment of the system of charges with respect to the point O. You can see immediately that the SI unit has to be C m.

Exercise. Convince yourself that if the system as a whole is electrically neutral, so that there is as much positive charge as negative charge, the dipole moment so defined is independent of the position of the point O. One can then talk of "the dipole moment of the system" without adding the rider "with respect to the point O".

Exercise. Convince yourself that if any electrically neutral system is placed in an external electric field **E**, it will experience a torque given by $\boldsymbol{\tau} = \mathbf{p} \times \mathbf{E}$, and so the two definitions of dipole moment – the physical and the mathematical – are equivalent.

Exercise. While thinking about these two, also convince yourself (from mathematics or from physics) that the moment of a simple dipole consisting of two charges, $+Q$ and $-Q$ separated by a distance l is Ql. We have already noted that C m is an acceptable SI unit for dipole moment.

3.3 *Oscillation of a Dipole in an Electric Field*

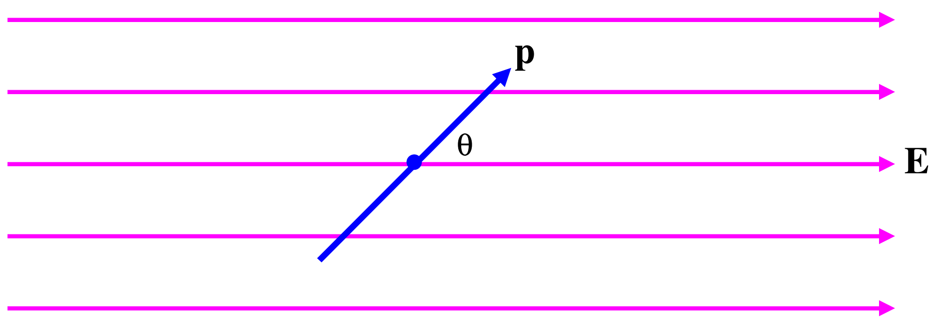


FIGURE III.3

Consider a dipole oscillating in an electric field (figure III.3). When it is at an angle θ to the field, the magnitude of the restoring torque on it is $pE \sin \theta$, and therefore its equation of motion is $I\ddot{\theta} = -pE\sin\theta$, where I is its rotational inertia. For small angles, this is approximately $I\ddot{\theta} = -pE\theta$, and so the period of small oscillations is

$$P = 2\pi\sqrt{\frac{I}{pE}}\,. \tag{3.3.1}$$

Would you expect the period to be long if the rotational inertia were large? Would you expect the vibrations to be rapid if p and E were large? Is the above expression dimensionally correct?

3.4 *Potential Energy of a Dipole in an Electric Field*

Refer again to figure III.3. There is a torque on the dipole of magnitude $pE \sin \theta$. In order to increase θ by $\delta\theta$ you would have to do an amount of work $pE \sin \theta \, \delta\theta$. The amount of work you would have to do to increase the angle between **p** and **E** from 0 to θ would be the integral of this from 0 to θ, which is $pE(1 - \cos \theta)$, and this is the potential energy of the dipole, provided one takes the potential energy to be zero when **p** and **E** are parallel. In many applications, writers find it convenient to take the potential energy (P.E.) to be zero when **p** and **E** perpendicular. In that case, the potential energy is

$$\text{P.E.} = -pE\cos\theta = -\mathbf{p}\bullet\mathbf{E}. \qquad 3.4.1$$

This is negative when θ is acute and positive when θ is obtuse. You should verify that the product of p and E does have the dimensions of energy.

3.5 *Force on a Dipole in an Inhomogeneous Electric Field*

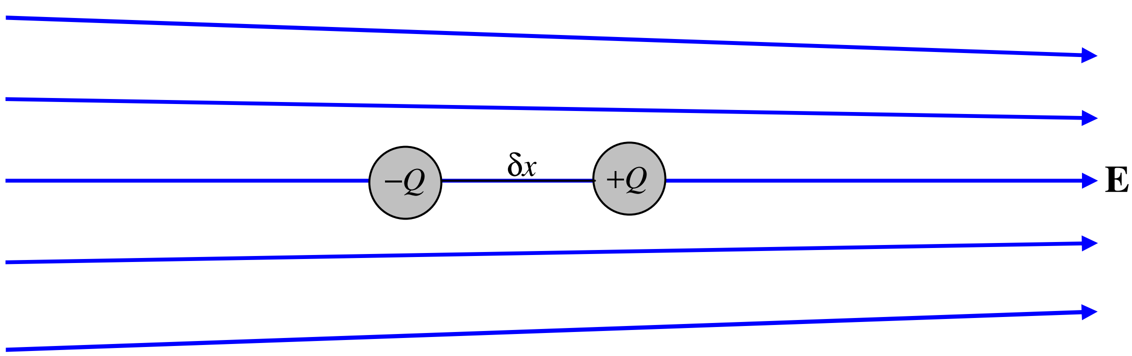


FIGURE III.4

Consider a simple dipole consisting of two charges $+Q$ and $-Q$ separated by a distance δx, so that its dipole moment is $p = Q\,\delta x$. Imagine that it is situated in an inhomogeneous electrical field as shown in figure III.4. We have already noted that a dipole in a *homogeneous* field experiences no net force, but we can see that it *does* experience a net force in an *inhomogeneous* field. Let the field at $-Q$ be E and the field at $+Q$ be $E + \delta E$. The force on $-Q$ is QE to the left, and the force on $+Q$ is $Q(E + \delta E)$ to the right. Thus there is a net force to the right of $Q\,\delta E$, or:

$$\text{Force} = p\frac{dE}{dx}. \qquad 3.5.1$$

Equation 3.5.1 describes the situation where the dipole, the electric field and the gradient are all parallel to the x-axis. In a more general situation, all three of these are in different directions. Recall that electric field is minus potential gradient. Potential is a scalar function, whereas electric field is a vector function with three component, of which the x-component, for example is $E_x = -\frac{\partial V}{\partial x}$. Field *gradient* is a symmetric *tensor* having nine components (of which, however, only six are distinct), such as $\frac{\partial^2 V}{\partial x^2}$, $\frac{\partial^2 V}{\partial y \partial z}$, etc. Thus in general equation 3.5.1 would have to be written as

$$\begin{pmatrix} E_x \\ E_y \\ E_z \end{pmatrix} = -\begin{pmatrix} V_{xx} & V_{xy} & V_{xz} \\ V_{xy} & V_{yy} & V_{yz} \\ V_{xz} & V_{yz} & V_{zz} \end{pmatrix} \begin{pmatrix} p_x \\ p_y \\ p_z \end{pmatrix}, \qquad 3.5.2$$

in which the double subscripts in the potential gradient tensor denote the second partial derivatives.

3.6 *Induced Dipoles and Polarizability*

We noted in section 1.3 that a charged rod will attract an *uncharged* pith ball, and at that time we left this as a little unsolved mystery. What happens is that the rod *induces a dipole moment* in the uncharged pith ball, and the pith ball, which now has a dipole moment, is attracted in the *inhomogeneous field* surrounding the charged rod.

How may a dipole moment be induced in an uncharged body? Well, if the uncharged body is metallic (as in the gold leaf electroscope), it is quite easy. In a metal, there are numerous free electrons, not attached to any particular atoms, and they are free to wander about inside the metal. If a metal is placed in an electric field, the free electrons are attracted to one end of the metal, leaving an excess of positive charge at the other end. Thus a dipole moment is induced.

What about a nonmetal, which doesn't have free electrons unattached to atoms? It may be that the individual molecules in the material have permanent dipole moments. In that case, the imposition of an external electric field will exert a torque on the molecules, and will cause all their dipole moments to line up in the same direction, and thus the bulk material will acquire a dipole moment. The water molecule, for example, has a permanent dipole moment, and these dipoles will align in an external field. This is why pure water has such a large dielectric constant.

But what if the molecules do not have a permanent dipole moment, or what if they do, but they cannot easily rotate (as may well be the case in a solid material)? The bulk material can still become polarized, because a dipole moment is induced in the individual molecules, the electrons inside the molecule tending to be pushed towards one end of the molecule. Or a molecule such as CH_4, which is symmetrical in the absence of an external electric field, may become distorted from its symmetrical shape when placed in an electric field, and thereby acquire a dipole moment.

Thus, one way or another, the imposition of an electric field may induce a dipole moment in most materials, whether they are conductors of electricity or not, or whether or not their molecules have permanent dipole moments.

If two molecules approach each other in a gas, the electrons in one molecule repel the electrons in the other, so that each molecule induces a dipole moment in the other. The two molecules then attract each other, because each dipolar molecule finds itself in the inhomogeneous electric field of the other. This is the origin of the van der Waals forces.

Some bodies (I am thinking about individual molecules in particular, but this is not necessary) are more easily polarized that others by the imposition of an external field. The ratio of the induced dipole moment to the applied field is called the *polarizability* α of the molecule (or whatever body we have in mind). Thus

$$\mathbf{p} = \alpha \mathbf{E}. \qquad 3.6.1$$

The SI unit for α is C m (V m^{-1})$^{-1}$ and the dimensions are $\mathrm{M^{-1}T^2Q^2}$.

This brief account, and the general appearance of equation 3.6.1, suggests that **p** and **E** are in the same direction – but this is so only if the electrical properties of the molecule are isotropic. Perhaps most molecules – and, especially, long organic molecules – have *anisotropic polarizability.* Thus a molecule may be easy to polarize with a field in the *x*-direction, and much less easy in the *y*- or *z*-directions. Thus, in equation 3.6.1, the polarizability is really a symmetric *tensor*, **p** and **E** are not in general parallel, and the equation, written out in full, is

$$\begin{pmatrix} p_x \\ p_y \\ p_z \end{pmatrix} = \begin{pmatrix} \alpha_{xx} & \alpha_{xy} & \alpha_{xz} \\ \alpha_{xy} & \alpha_{yy} & \alpha_{yz} \\ \alpha_{xz} & \alpha_{yz} & \alpha_{zz} \end{pmatrix} \begin{pmatrix} E_x \\ E_y \\ E_z \end{pmatrix}. \qquad 3.6.2$$

(Unlike in equation 3.5.2, the double subscripts are not intended to indicate second partial derivatives; rather they are just the components of the polarizability tensor.) As in several analogous situations in various branches of physics (see, for example, section 2.17 of Classical Mechanics and the inertia tensor) there are three mutually orthogonal directions (the eigenvectors of the polarizability tensor) for which **p** and **E** will be parallel.

3.7 *The Simple Dipole*

As you may expect from the title of this section, this will be the most difficult and complicated section of this chapter so far. Our aim will be to calculate the field and potential surrounding a simple dipole.

A simple dipole is a system consisting of two charges, $+Q$ and $-Q$, separated by a distance $2L$. The dipole moment of this system is just $p = 2QL$. We'll suppose that the dipole lies along the x-axis, with the negative charge at $x = -L$ and the positive charge at $x = +L$. See figure III.5.

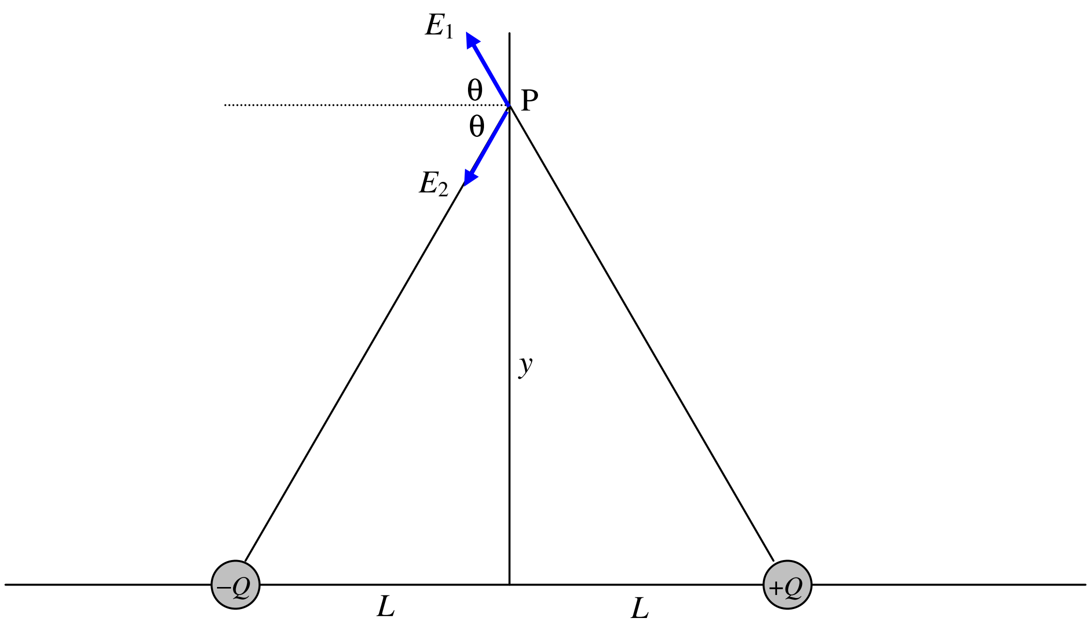


FIGURE III.5

Let us first calculate the electric field at a point P at a distance y along the y-axis. It will be agreed, I think, that it is directed towards the left and is equal to

$$E_1 \cos\theta \ + \ E_2 \cos\theta, \text{ where } E_1 \ = \ E_2 \ = \ \frac{Q}{4\pi\epsilon_0(L^2 + y^2)} \ \text{ and } \ \cos\theta = \frac{L}{(L^2 + y^2)^{1/2}} \ .$$

Therefore $$E \ = \ \frac{2QL}{4\pi\epsilon_0(L^2 + y^2)^{3/2}} \ = \ \frac{p}{4\pi\epsilon_0(L^2 + y^2)^{3/2}} \ . \qquad 3.7.1$$

For large y this becomes

$$E = \frac{p}{4\pi\varepsilon_0 y^3}\,. \qquad 3.7.2$$

That is, the field falls off as the cube of the distance.

To find the field on the x-axis, refer to figure III.6.

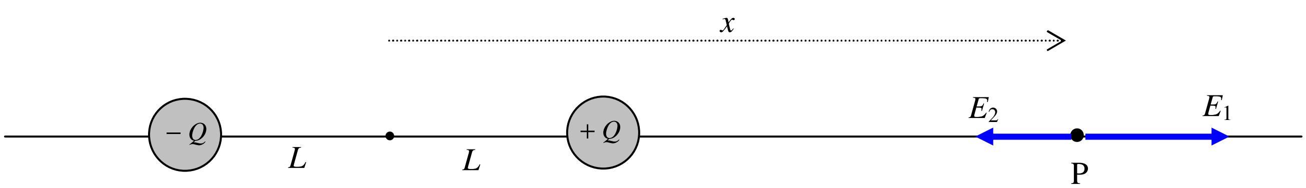


FIGURE III.6

It will be agreed, I think, that the field is directed towards the right and is equal to

$$E = E_1 - E_2 = \frac{Q}{4\pi\varepsilon_0}\left(\frac{1}{(x-L)^2} - \frac{1}{(x+L)^2}\right). \qquad 3.7.3$$

This can be written $\frac{Q}{4\pi\varepsilon_0 x^2}\left(\frac{1}{(1-L/x)^2} - \frac{1}{(1+L/x)^2}\right)$, and on expansion of this by the binomial theorem, neglecting terms of order $(L/x)^2$ and smaller, we see that at large x the field is

$$E = \frac{2p}{4\pi\varepsilon_0 x^3}\,. \qquad 3.7.4$$

Now for the field at a point P that is neither on the axis (x-axis) nor the equator (y-axis) of the dipole. See figure III.7.

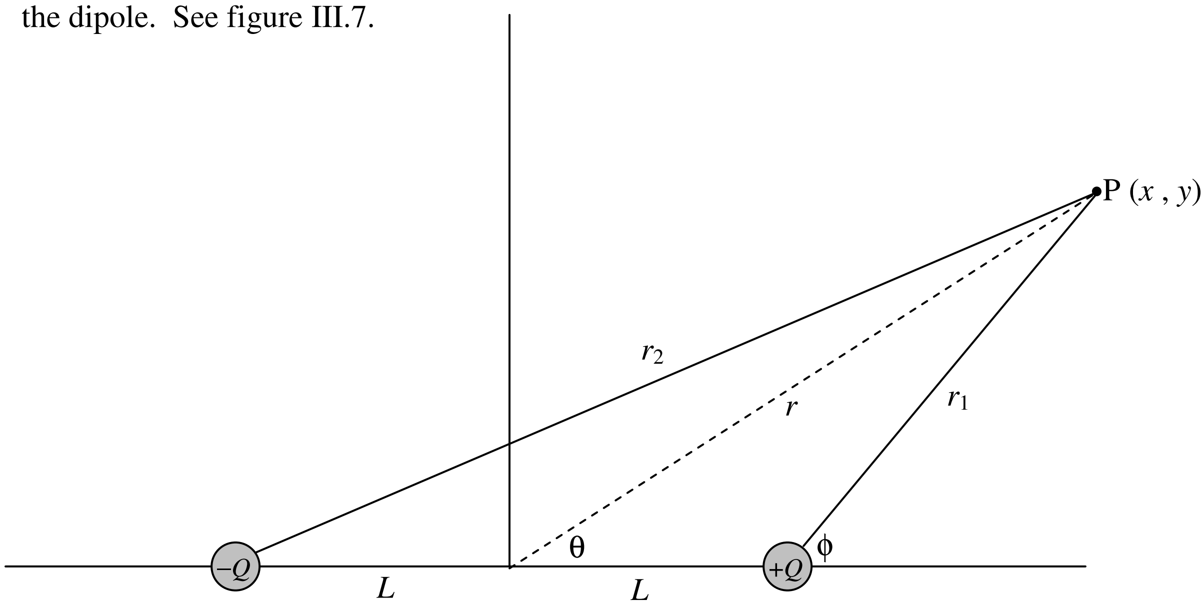


FIGURE III.7

It will probably be agreed that it would not be particularly difficult to write down expressions for the contributions to the field at P from each of the two charges in turn. The difficult part then begins; the two contributions to the field are in different and awkward directions, and adding them vectorially is going to be a bit of a headache.

It is much easier to calculate the *potential* at P, since the two contributions to the potential can be added as scalars. Then we can find the *x*- and *y*-components of the field by calculating $\partial V / \partial x$ and $\partial V / \partial y$.

Thus $$V = \frac{Q}{4\pi\epsilon_0}\left(\frac{1}{\{(x-L)^2 + y^2\}^{1/2}} - \frac{1}{\{(x+L)^2 + y^2\}^{1/2}}\right). \qquad 3.7.5$$

To start with I am going to investigate the potential and the field at a *large distance* from the dipole – though I shall return later to the near vicinity of it.

At *large distances* from a small dipole (see figure III.8), we can write $r^2 = x^2 + y^2$,

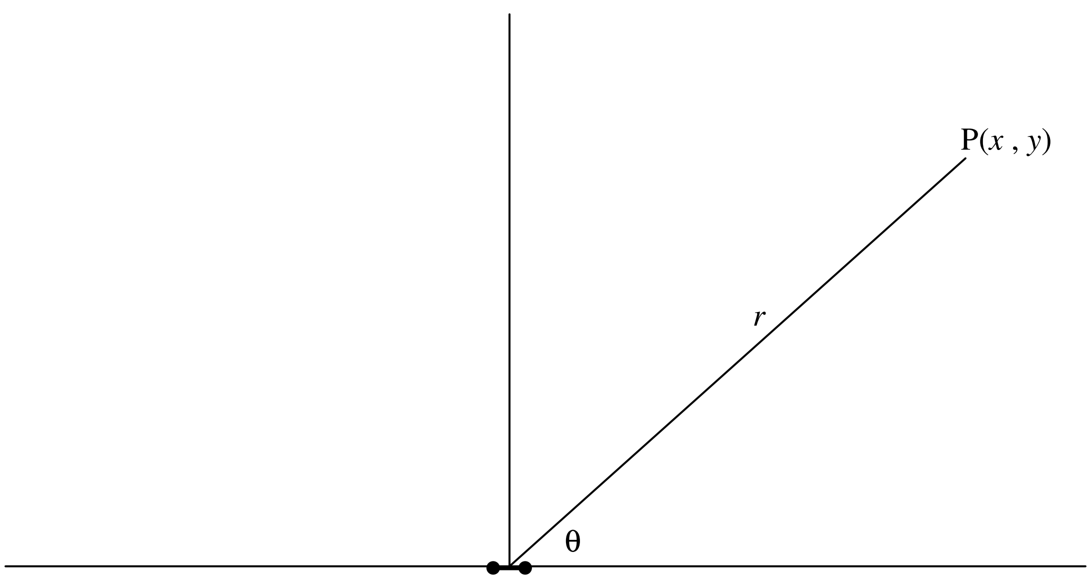


FIGURE III.8

and, with $L^2 << r^2$, the expression 3.7.5 for the potential at P becomes

$$V = \frac{Q}{4\pi\epsilon_0}\left(\frac{1}{(r^2 - 2Lx)^{1/2}} - \frac{1}{(r^2 + 2Lx)^{1/2}}\right) = \frac{Q}{4\pi\epsilon_0 r}\left((1 - 2Lx/r^2)^{-1/2} - (1 + 2Lx/r^2)^{-1/2}\right).$$

When this is expanded by the binomial theorem we find, to order L/r , that the potential can be written in any of the follwing equivalent ways:

$$V = \frac{2QLx}{4\pi\epsilon_0 r^3} = \frac{px}{4\pi\epsilon_0 r^3} = \frac{p\cos\theta}{4\pi\epsilon_0 r^2} = \frac{\mathbf{p \cdot r}}{4\pi\epsilon_0 r^3}. \qquad 3.7.6$$

Thus the equipotentials are of the form

$$r^2 = c\cos\theta, \qquad 3.7.7$$

where

$$c = \frac{p}{4\pi\epsilon_0 V}. \qquad 3.7.8$$

Now, bearing in mind that $r^2 = x^2 + y^2$, we can differentiate $V = \frac{px}{4\pi\epsilon_0 r^3}$ with respect to x and y to find the x- and y-components of the field.

Thus we find that

$$E_x = \frac{p}{4\pi\epsilon_0}\left(\frac{3x^2 - r^2}{r^5}\right) \quad \text{and} \quad E_y = \frac{pxy}{4\pi\epsilon_0 r^5}. \qquad 3.7.9\text{a,b}$$

We can also use polar coordinates find the radial and transverse components from $E_r = -\frac{\partial V}{\partial r}$ and $E_\theta = -\frac{1}{r}\frac{\partial V}{\partial \theta}$ together with $V = \frac{p\cos\theta}{4\pi\epsilon_0 r^2}$ to obtain

$$E_r = \frac{2p\cos\theta}{4\pi\epsilon_0 r^3} \quad \text{and} \quad E_\theta = \frac{p\sin\theta}{4\pi\epsilon_0 r^3}. \qquad 3.7.10\text{a,b}$$

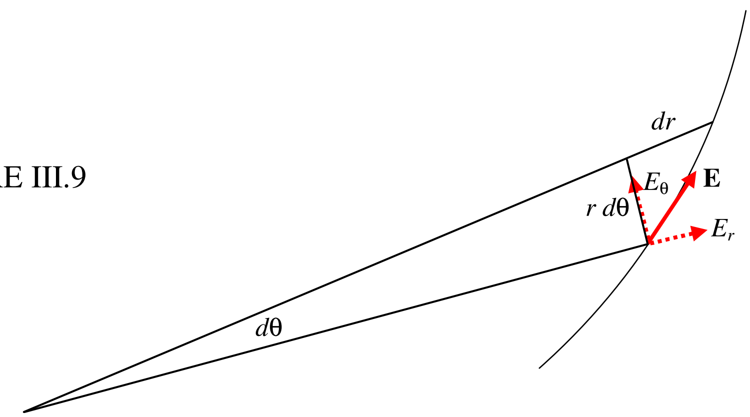
For those who enjoy vector calculus, we can also say $\mathbf{E} = -\frac{1}{4\pi\epsilon_0}\boldsymbol{\nabla}\left(\frac{\mathbf{p \cdot r}}{r^3}\right)$, from which, after a little algebra and quite a lot of vector calculus, we find

$$\mathbf{E} = \frac{1}{4\pi\epsilon_0}\left(\frac{3(\mathbf{p \cdot r})\mathbf{r}}{r^5} - \frac{\mathbf{p}}{r^3}\right). \qquad 3.7.11$$

This equation contains all the information that we are likely to want, but I expect most readers will prefer the more explicit rectangular and polar forms of equations 3.7.9 and 3.7.10.

Equation 3.7.7 gives the equation to the equipotentials. The equation to the lines of force can be found as follows. Referring to figure III.9, we see that the differential equation to the lines of force is

FIGURE III.9



$$r\frac{d\theta}{dr} = \frac{E_\theta}{E_r} = \frac{\sin\theta}{2\cos\theta} = \tfrac{1}{2}\tan\theta, \qquad 3.7.12$$

which, upon integration, becomes

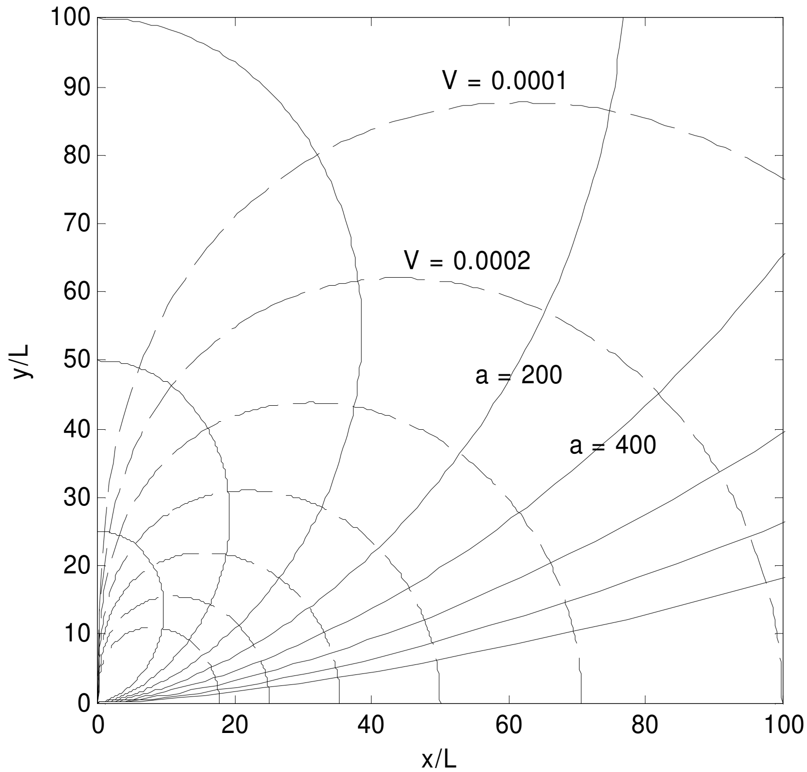
$$r = a\sin^2\theta. \qquad 3.7.13$$

Note that the equations $r^2 = c\cos\theta$ (for the equipotentials) and $r = a\sin^2\theta$ (for the lines of force) are orthogonal trajectories, and either can be derived from the other. Thus, given that the differential equation to the lines of force is $r\dfrac{d\theta}{dr} = \tfrac{1}{2}\tan\theta$ with solution $r = a\sin^2\theta$, the differential equation to the orthogonal trajectories (i.e. the equipotentials) is $-\dfrac{1}{r}\dfrac{dr}{d\theta} = \tfrac{1}{2}\tan\theta$, with solution $r^2 = c\cos\theta$.

In figure III.10, there is supposed to be a tiny dipole situated at the origin. The unit of length is L, half the length of the dipole. I have drawn eight electric field lines (continuous), corresponding to a = 25, 50, 100, 200, 400, 800, 1600, 3200. If r is expressed in units of L, and if V is expressed in units of $\dfrac{Q}{4\pi\epsilon_0 L}$, the equations 3.7.7 and

3.7.8 for the equipotentials can be written $r = \sqrt{\frac{2\cos\theta}{V}}$, and I have drawn seven equipotentials (dashed) for V = 0.0001, 0.0002, 0.0004, 0.0008, 0.0016, 0.0032, 0.0064. It will be noticed from equation 3.7.9a, and is also evident from figure III.10, that E_x is zero for $\theta = 54^\circ\ 44'$.

FIGURE III.10



At the end of this chapter I append a (geophysical) exercise in the geometry of the field at a large distance from a small dipole.

Equipotentials near to the dipole

These, then, are the field lines and equipotentials at a *large distance* from the dipole. We arrived at these equations and graphs by expanding equation 3.7.5 binomially, and neglecting terms of higher order than L/r. We now look *near to* the dipole, where we cannot make such an approximation. Refer to figure III.7.

We can write 3.7.5 as

$$V(x,y) = \frac{Q}{4\pi\epsilon_0}\left(\frac{1}{r_1} - \frac{1}{r_2}\right), \qquad 3.7.14$$

where $r_1^2 = (x-L)^2 + y^2$ and $r_2^2 = (x+L)^2 + y^2$. If, as before, we express distances in terms of L and V in units of $\frac{Q}{4\pi\epsilon_0 L}$, the expression for the potential becomes

$$V(x,y) = \frac{1}{r_1} - \frac{1}{r_2}, \qquad 3.7.15$$

where $r_1^2 = (x+1)^2 + y^2$ and $r_2^2 = (x-1)^2 + y^2$.

One way to plot the equipotentials would be to calculate V for a whole grid of (x , y) values and then use a contour plotting routine to draw the equipotentials. My computing skills are not up to this, so I'm going to see if we can find some way of plotting the equipotentials directly.

I present two methods. In the first method I use equation 3.7.15 and endeavour to manipulate it so that I can calculate y as a function of x and V. The second method was shown to me by J. Visvanathan of Chennai, India. We'll do both, and then compare them.

First Method.

To anticipate, we are going to need the following:

$$r_1^2 r_2^2 = (x^2 + y^2 + 1)^2 - 4x^2 = B^2 - A, \qquad 3.7.16$$

$$r_1^2 + r_2^2 = 2(x^2 + y^2 + 1) = 2B, \qquad 3.7.17$$

and $$r_1^4 + r_2^4 = 2[(x^2 + y^2 + 1)^2 + 4x^2] = 2(B^2 + A), \qquad 3.7.18$$

where $$A = 4x^2 \qquad 3.7.19$$

and $$B = x^2 + y^2 + 1. \qquad 3.7.20$$

Now equation 3.7.15 is $r_1 r_2 V = r_2 - r_1$. In order to extract y it is necessary to square this twice, so that r_1 and r_2 appear only as r_1^2 and r_2^2. After some algebra, we obtain

$$r_1^2 r_2^2[2 - V^4 r_1^2 r_2^2 + 2V^2(r_1^2 + r_2^2)] = r_1^4 + r_2^4. \qquad 3.7.21$$

Upon substitution of equations 3.7.16,17,18, for which we are well prepared, we find for the equation to the equipotentials an equation which, after some algebra, can be written as a quartic equation in B:

$$a_0 + a_1B + a_2B^2 + a_3B^3 + a_4B^4 = 0, \tag{3.7.22}$$

where

$$a_0 = A(4 + V^4A), \tag{3.7.23}$$

$$a_1 = 4V^2A, \tag{3.7.24}$$

$$a_2 = -2V^2A, \tag{3.7.25}$$

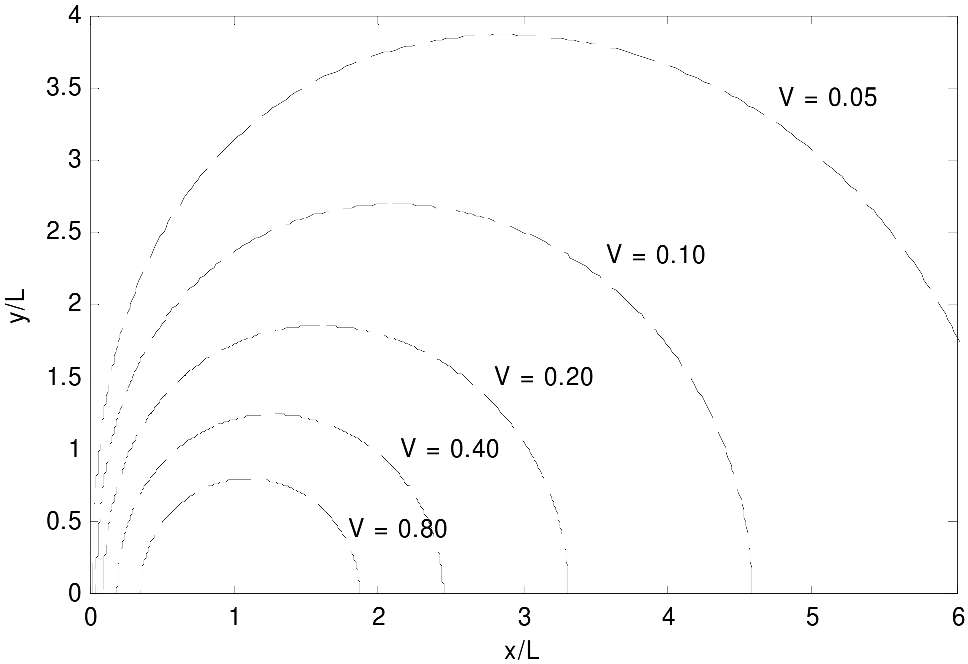
$$a_3 = -4V^2, \tag{3.7.26}$$

and

$$a_4 = V^4. \tag{3.7.27}$$

The algorithm will be as follows: For a given V and x, calculate the quartic coefficients from equations 3.7.23-27. Solve the quartic equation 3.7.22 for B. Calculate y from equation 3.7.20. My attempt to do this is shown in figure III.11. The dipole is supposed to have a negative charge at $(-1 , 0)$ and a positive charge at $(+1 , 0)$. The equipotentials are drawn for $V = 0.05, 0.10, 0.20, 0.40, 0.80$.

FIGURE III.11



Second method (J. Visvanathan).

In this method, we work in polar coordinates, but instead of using the coordinates (r, θ), in which the origin, or pole, of the polar coordinate system is at the centre of the dipole (see figure III.7), we use the coordinates (r_1, ϕ) with origin at the positive charge.

From the triangle, we see that

$$r_2^2 = r_1^2 + 4L^2 + 4Lr_1 \cos\phi. \qquad 3.7.28$$

For future reference we note that

$$\frac{\partial r_2}{\partial r_1} = \frac{r_1 + 2L\cos\phi}{r_2}. \qquad 3.7.29$$

Provided that distances are expressed in units of L, these equations become

$$r_2^2 = r_1^2 + 4r_1 \cos\phi + 4, \qquad 3.7.30$$

$$\frac{\partial r_2}{\partial r_1} = \frac{r_1 + 2\cos\phi}{r_2}. \qquad 3.7.31$$

If, in addition, electrical potential is expressed in units of $\dfrac{Q}{4\pi\epsilon_0 L}$, the potential at P is given, as before (equation 3.17.15), by

$$V(r_1, \phi) = \frac{1}{r_1} - \frac{1}{r_2}. \qquad 3.7.32$$

Recall that r_2 is given by equation 3.7.30, so that equation 3.7.32 is really an equation in just V, r_1 and ϕ.

In order to plot an equipotential, we fix some value of V; then we vary ϕ from 0 to π, and, for each value of ϕ we have to try to calculate r_1.This can be done by the Newton-Raphson process, in which we make a guess at r_1 and use the Newton-Raphson process to obtain a better guess, and continue until successive guesses converge. It is best if we can make a fairly good first guess, but the Newton-Raphson process will often converge very rapidly even for a poor first guess.

Thus we have to solve the following equation for r_1 for given values of V and ϕ,

$$f(r_1) = \frac{1}{r_1} - \frac{1}{r_2} - V = 0, \qquad 3.7.33$$

bearing in mind that r_2 is given by equation 3.7.31.

By differentiation with respect to r_1, we have

$$f'(r_1) = -\frac{1}{r_1^2} + \frac{1}{r_2^2}\frac{\partial r_2}{\partial r_1} = -\frac{1}{r_1^2} + \frac{r_1 + 2\cos\phi}{r_2^3}, \qquad 3.7.34$$

and we are all set to begin a Newton-Raphson iteration: $r_1 = r_1 - f/f'$. Having obained r_1, we can then obtain the (x, y) coordinates from $x = 1 + r_1\cos\phi$ and $y = r_1\sin\phi$.

I tried this method and I got exactly the same result as by the first method and as shown in figure III.11.

So which method do we prefer? Well, anyone who has worked through in detail the derivations of equations 3.7.16 -3.7.27, and has then tried to program them for a computer, will agree that the first method is very laborious and cumbersome. By comparison Visvanathan's method is much easier both to derive and to program. On the other hand, one small point in favour of the first method is that it involves no trigonometric functions, and so the numerical computation is potentially faster than the second method in which a trigonometric function is calculated at each iteration of the Newton-Raphson process. In truth, though, a modern computer will perform the calculation by either method apparently instantaneously, so that small advantage is hardly relevant.

So far, we have managed to draw the *equipotentials* near to the dipole. The *lines of force* are orthogonal to the equipotentials. After I tried several methods with only partial success, I am grateful to Dr Visvanathan who pointed out to me what ought to have been the "obvious" method, namely to use equation 3.7.12, which, in our $(r_1\ , \phi)$ coordinate system based on the positive charge, is $r_1 \dfrac{d\phi}{dr_1} = \dfrac{E_\phi}{E_{r_1}}$, just as we did for the large distance, small dipole, approximation. In this case, the potential is given by equations 3.7.30 and 3.7.32. (Recall that in these equations, distances are expressed in units of L and the potential in units of $\dfrac{Q}{4\pi\epsilon_0 L}$.) The radial and transverse components of the field are given by $E_{r_1} = -\dfrac{\partial V}{\partial r_1}$ and $E_\phi = -\dfrac{1}{r_1}\dfrac{\partial V}{\partial \phi}$, which result in

$$E_{r_1} = \frac{1}{r_1^2} - \frac{r_1 + 2\cos\phi}{r_2^3} \qquad 3.7.35$$

and

$$E_\phi = \frac{2\sin\phi}{r_2^3}. \qquad 3.7.36$$

Here, the field is expressed in units of $\dfrac{Q}{4\pi\epsilon_0 L^2}$, although that hardly matters, since we are interested only in the ratio. On applying $r_1 \dfrac{d\phi}{dr_1} = \dfrac{E_\phi}{E_{r_1}}$ to these field components we obtain the following differential equation to the lines of force:

$$d\phi = \frac{2r_1 \sin\phi}{(r_1^2 + 4 + 4r_1\cos\phi)^{3/2} - r_1^2(r_1 + 2\cos\phi)} dr_1. \qquad 3.7.37$$

Thus one can start with some initial ϕ_0 and small r_2 and increase r_1 successively by small increments, calculating a new ϕ each time. The results are shown in figure III.12, in which the equipotentials are drawn for the same values as in figure III.11, and the initial angles for the lines of force are 30º, 60º, 90º, 120º, 150º.

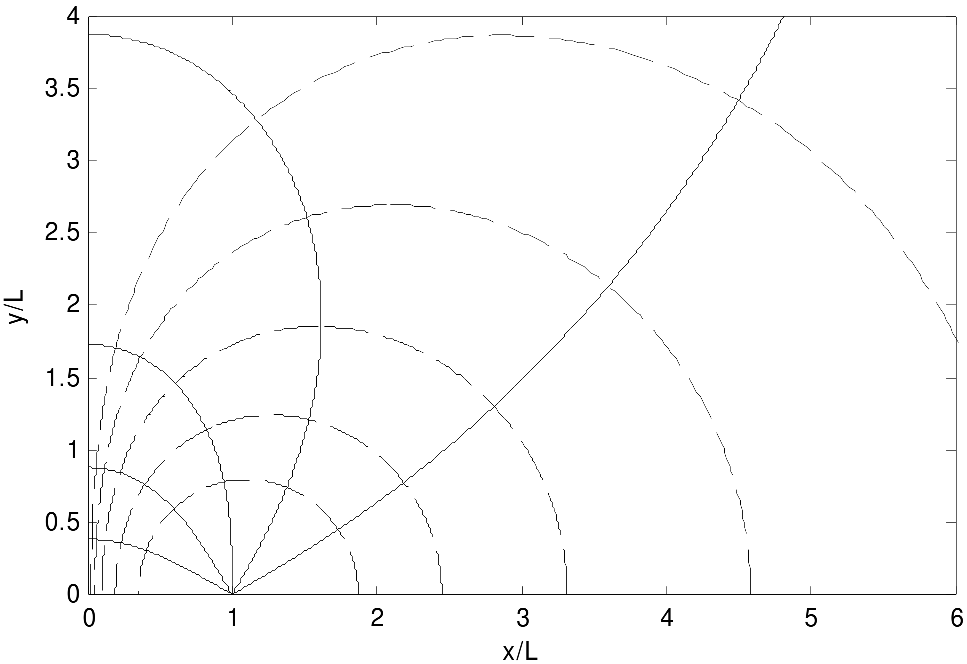


FIGURE III.12

3.8 *Quadrupole Moment*

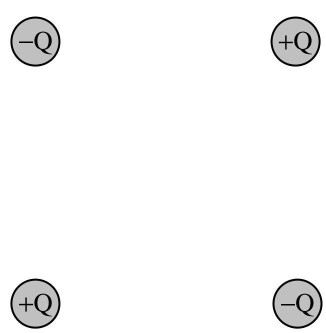


FIGURE III.13

Consider the system of charges shown in figure III.13. It has no net charge and no net dipole moment. Unlike a dipole, it will experience neither a net force nor a net torque in *any uniform* field. It may or may not experience a net force in an external nonuniform field. For example, if we think of the quadrupole as two dipoles, each dipole will experience a force proportional to the local field gradient in which it finds itself. If the field gradients at the location of each dipole are equal, the forces on each dipole will be equal but opposite, and there will net force on the quadrupole. If, however, the field gradients at the positions of the two dipoles are unequal, the forces on the two dipole will be unequal, and there will be a net force on the quadruople. Thus there will be a net force if there is a non-zero gradient of the field gradient. Stated another way, there will be no net force on the quadrupole if the mixed second partial derivatives of the field componets (the third derivatives of the potential!) are zero. Further, if the quadrupole is in a nonuniform field, increasing, say, to the right, the upper pair will experience a force to the right and the lower pair will experience a force to the left; thus the system will experience a net torque in an inhomogeneous field, though there will be not net force unless the field gradients on the two pairs are unequal.

The system possesses what is known as a *quadrupole moment*. While a single charge is a scalar quantity, and a dipole moment is a vector quantity, the quadrupole moment is a second order symmetric tensor.

The dipole moment of a system of charges is a vector with three components given by $p_x = \sum Q_i x_i$, $p_y = \sum Q_i y_i$, $p_z = \sum Q_i z_i$. The quadrupole moment $\mathbf{q}$ has nine components (of which six are distinct) defined by $q_{xx} = \sum Q_i x_i^2$, $q_{xy} = \sum Q_i x_i y_i$, etc., and its matrix representation is

$$\mathbf{q} = \begin{pmatrix} q_{xx} & q_{xy} & q_{xz} \\ q_{xy} & q_{yy} & q_{yz} \\ q_{xz} & q_{yz} & q_{zz} \end{pmatrix}. \qquad 3.8.1$$

For a continuous charge distribution with charge density ρ coulombs per square metre, the components will be given by $q_{xx} = \int \rho x^2 d\tau$, etc., where $d\tau$ is a volume element, given in rectangular coordinates by $dxdydz$ and in spherical coordinates by $r^2 \sin\theta \, dr d\theta d\phi$. The SI unit of quadrupole moment is C m^2, and the dimensions are L^{2}Q,

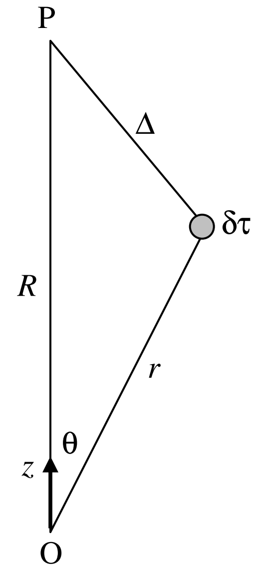
By suitable rotation of axes, in the usual way (see for example section 2.17 of Classical Mechanics), the matrix can be diagonalized, and the diagonal elements are then the eigenvalues of the quadrupole moment, and the trace of the matrix is unaltered by the rotation.

3.9 *Potential at a Large Distance from a Charged Body*

We wish to find the potential at a point P at a large distance R from a charged body, in terms of its total charge and its dipole, quadrupole, and possibly higher-order moments. There will be no loss of generality if we choose a set of axes such that P is on the z-axis.

We refer to figure III.14, and we consider a volume element $\delta\tau$ at a distance r from some origin. The point P is at a distance r from the origin and a distance Δ from $\delta\tau$. The potential at P from the charge in the element $\delta\tau$ is given by

FIGURE III.14



$$4\pi\epsilon_0 \delta V = \frac{\rho\,\delta\tau}{\Delta} = \frac{\rho}{R}\left(1 + \frac{r^2}{R^2} - \frac{2r}{R}\cos\theta\right)^{-1/2}\delta\tau\,, \qquad 3.9.1$$

and so the potential from the charge on the whole body is given by

$$4\pi\epsilon_0 V = \frac{1}{R}\int\rho\left(1 + \frac{r^2}{R^2} - \frac{2r}{R}\cos\theta\right)^{-1/2}\delta\tau. \qquad 3.9.2$$

On expanding the parentheses by the binomial theorem, we find, after a little trouble, that this becomes

$$4\pi\epsilon_0 V = \frac{1}{R}\int\rho\, d\tau + \frac{1}{R^2}\int\rho\, r\, P_1(\cos\theta)\, d\tau + \frac{1}{2!R^3}\int\rho\, r^2 P_2\,(\cos\theta)\, d\tau$$
$$+ \frac{1}{3!R^4}\int\rho\, r^3 P_3\,(\cos\theta)\, d\tau + \ldots\,, \qquad 3.9.3$$

where the polynomials P are the Legendre polynomials given by

$$P_1(\cos\theta) = \cos\theta, \qquad 3.9.4$$

$$P_2(\cos\theta) = \tfrac{1}{2}(3\cos^2\theta - 1), \qquad 3.9.5$$

and $$P_3(\cos\theta) = \tfrac{1}{2}(5\cos^3\theta - 3\cos\theta). \qquad 3.9.6$$

We see from the forms of these integrals and the definitions of the components of the dipole and quadrupole moments that this can now be written:

$$4\pi\epsilon_0 V = \frac{Q}{R} + \frac{p}{R^2} + \frac{1}{2R^3}(3q_{zz} - \mathrm{Tr}\,\mathbf{q}) + \ldots, \qquad 3.9.7$$

Here Tr $\mathbf{q}$ is the trace of the quadrupole moment matrix, or the (invariant) sum of its diagonal elements. Equation 3.9.7 can also be written

$$4\pi\epsilon_0 V = \frac{Q}{R} + \frac{p}{R^2} + \frac{1}{2R^3}[2q_{zz} - (q_{xx} + q_{yy})] + \ldots\ . \qquad 3.9.8$$

The quantity $2q_{zz} - (q_{xx} + q_{yy})$ of the diagonalized matrix is often referred to as "the" quadrupole moment. It is zero if all three diagonal components are zero or if $q_{zz} = \tfrac{1}{2}(q_{xx} + q_{yy})$. If the body has cylindrical symmetry about the z-axis, this becomes $2(q_{zz} - q_{xx})$.

Exercise.

Show that the potential at (r, θ) at a large distance from the linear quadrupole of figure III.15 is

$$V = \frac{QL^2(3\cos^2\theta - 1)}{4\pi\epsilon_0 r^3}.$$

(The gap in the dashed line is intended to indicate that r is very large compared with L.)

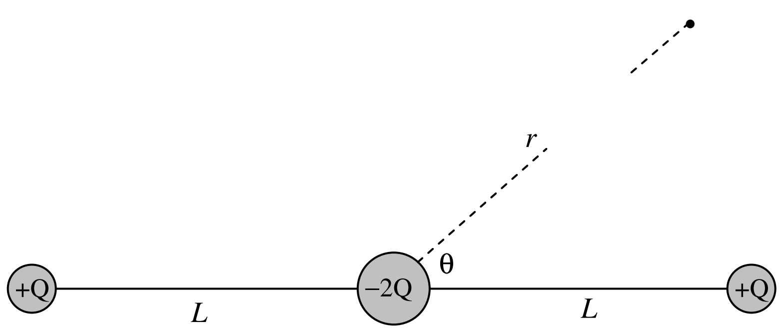


FIGURE III.15

The solution to this exercise is easy if you know about *Legendre polynomials*. See Section 1.14 of my notes on Celestial Mechanics. What you need to know is that the expansion of $(1-2ax+x^2)^{-1/2}$ can be written as a series of Legendre polynomials, namely $P_0(x) + xP_1(x) + x^2P_2(x) + \ldots$. You also need a (very small) table of Legendre polynamials, namely $P_0(x) = 1$, $P_1(x) = x$, $P_2(x) = \frac{1}{2}(3x^2-1)$. Given that, you should find the exercise very easy.

3.10 *A Geophysical Example*

Assume that planet Earth is spherical and that it has a little magnet or current loop at its centre. By "little" I mean small compared with the radius of the Earth. Suppose that, at a large distance from the magnet or current loop the geometry of the magnetic field is the same as that of an electric field at a large distance from a simple dipole. That is to say, the equation to the lines of force is $r = a\sin^2\theta$ (equation 3.7.13), and the differential equation to the lines of force is $\frac{dr}{d\theta} = \frac{2r}{\tan\theta}$ (equation 3.7.12).

Show that the angle of dip D at geomagnetic latitude L is given by

$$\tan D = 2\tan L. \qquad 3.10.1$$

The geometry is shown in figure III.16.

The result is a simple one, and there is probably a simpler way of getting it than the one I tried. Let me know (jtatum@uvic.ca) if you find a simpler way. In the meantime, here is my solution.

I am going to try to find the slope m_1 of the tangent to Earth (i.e. of the horizon) and the slope m_2 of the line of force. Then the angle D between them will be given by the equation (which I am hoping is well known from coordinate geometry!)

$$\tan D = \frac{m_1 - m_2}{1 + m_1m_2}. \qquad 3.10.2$$

The first is easy:

$$m_1 = \tan(90^\circ + \theta) = -\frac{1}{\tan\theta}. \qquad 3.10.3$$

For m_2 we want to find the slope of the line of force, whose equation is given in polar coordinates? So, how do you find the slope of a curve whose equation is given in polar coordinates? We can do it like this:

$$x = r\cos\theta, \qquad 3.10.4$$
$$y = r\sin\theta, \qquad 3.10.5$$
$$dx = \cos\theta dr - r\sin\theta d\theta, \qquad 3.10.6$$
$$dy = \sin\theta dr + r\cos\theta d\theta. \qquad 3.10.7$$

From these, we obtain

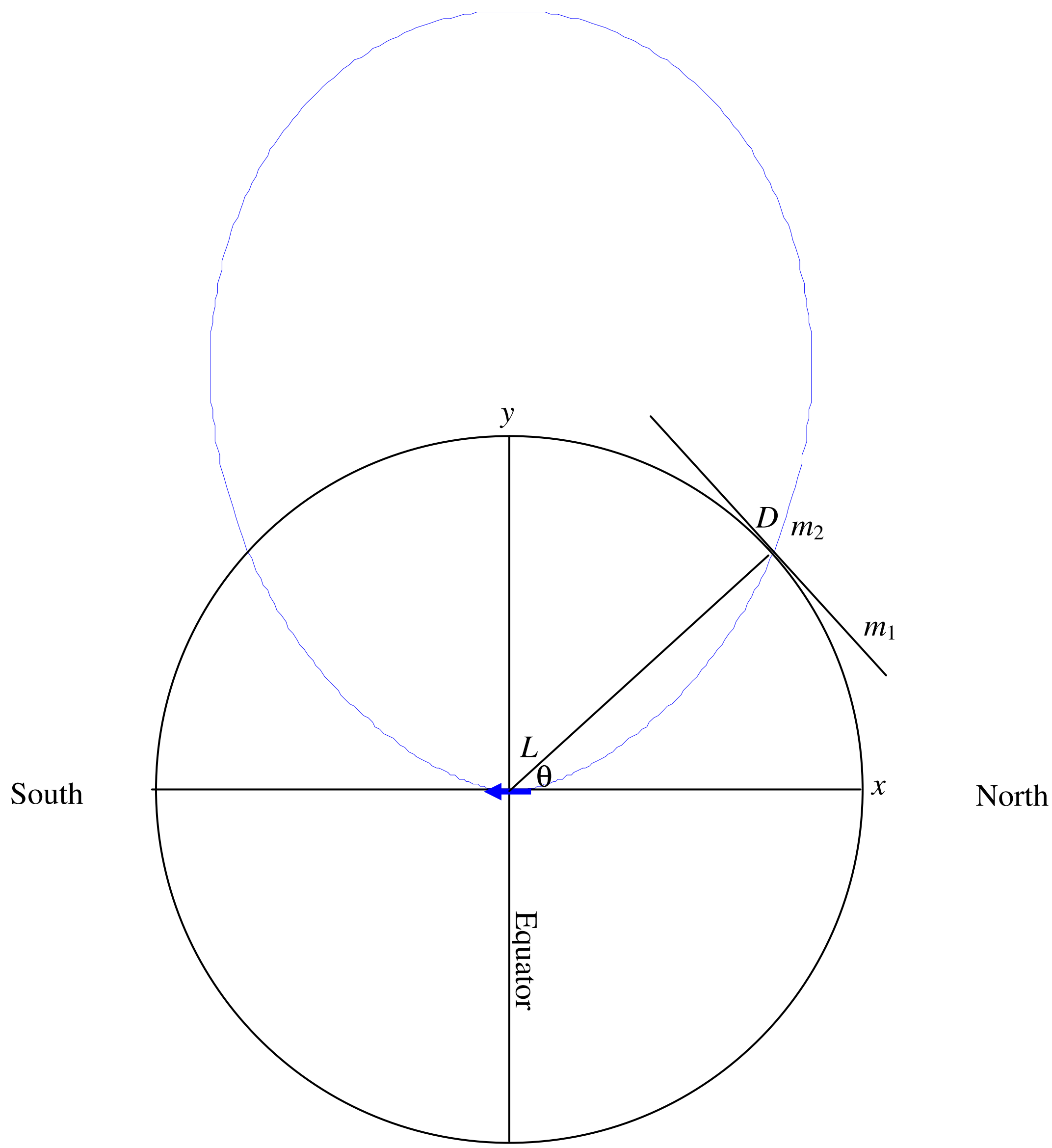
$$\frac{dy}{dx} = \frac{\sin\theta\frac{dr}{d\theta} + r\cos\theta}{\cos\theta\frac{dr}{d\theta} - r\sin\theta}. \qquad 3.10.8$$

In our particular case, we have $\frac{dr}{d\theta} = \frac{2r}{\tan\theta}$ (equation 3.7.12), so if we substitute this into equation 3.10.8 we soon obtain

$$m_2 = \frac{3\sin\theta\cos\theta}{3\cos^2\theta - 1}. \qquad 3.10.9$$

Now put equations 3.10.3 and 3.10.9 into equation 3.10.2, and, after a little algebra, we soon obtain

$$\underline{\underline{\tan D = \frac{2}{\tan\theta} = 2\tan L.}} \qquad 3.10.10$$



Here is another question. The magnetic field is generally given the symbol B. Show that the strength of the magnetic field $B(L)$ at geomagnetic latitude L is given by

$$B(L) = B(0)\sqrt{1 + 3\sin^2 L}, \qquad 3.10.11$$

where $B(0)$ is the strength of the field at the equator. This means that it is twice as strong at the magnetic poles as at the equator.

Start with equation 3.7.2, which gives the electric field at a distant point on the equator of an electric dipole. That equation was $E = \frac{p}{4\pi\epsilon_0 y^3}$. In this case we are dealing with a magnetic field and a magnetic diople, so we'll replace the electric field E with a magnetic field B. Also $p/(4\pi\epsilon_0)$ is a combination of electrical quantities, and since we are interested only in the geometry (i.e. on how B varies from equation to pole, let's just write $p/(4\pi\epsilon_0)$ as k. And we'll take the radius of Earth to be R, so that equation 3.7.2 gives for the magnetic field at the surface of Earth on the equator as

$$B(0) = \frac{k}{R^3}. \qquad 3.10.12$$

In a similar vein, equations 3.7.10a,b for the radial and transverse components of the field at geomagnetic latitude L (which is 90º – θ) become

$$B_r(L) = \frac{2k \sin L}{R^3} \quad \text{and} \quad B_\theta(L) = \frac{k \cos L}{R^3}. \qquad 3.10.13\text{a,b}$$

And since $B = \sqrt{B_r^2 + B_\theta^2}$, the result immediately follows.

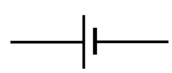
CHAPTER 4
BATTERIES, RESISTORS AND OHM'S LAW

4.1 *Introduction*

An electric *cell* consists of two different metals, or carbon and a metal, called the *poles*, immersed or dipped into a liquid or some sort of a wet, conducting paste, known as the *electrolyte*, and, because of some chemical reaction between the two poles and the electrolyte, there exists a small potential difference (typically of the order of one or two volts) between the poles. This potential difference is much smaller than the hundreds or thousands of volts that may be obtained in typical laboratory experiments in electrostatics, and the electric field between the poles is also correspondingly small.

Definition. The potential difference across the poles of a cell when no current is being taken from it is called the *electromotive force* (EMF) of the cell.

The circuit symbol for a cell is drawn thus:



The longer, thin line represents the positive pole and the shorter, thick line represents the negative pole.

Several cells connected together form a *battery* of cells. Thus in principle a single cell should strictly be called just that – a *cell* – and the word *battery* should be restricted to a battery of several cells. However, in practice, most people use the word battery to mean either literally a battery of several cells, or a single cell.

I shall not discuss in this chapter the detailed chemistry of why there exists such a potential difference, nor shall I discuss in detail the chemical processes that take place inside the several different varieties of cell. I shall just mention that in the cheaper types of flashlight battery (cell), the negative pole, made of zinc, is the outer casing of the cell, while the positive pole is a central carbon rod. The rather dirty mess that is the electrolyte is a mixture that is probably known only to the manufacturer, though it probably includes manganese oxide and ammonium chloride and perhaps such goo as flour or glue and goodness knows what else. Other types have a positive pole of nickelic hydroxide and a negative pole of cadmium metal in a potassium hydroxide electrolyte. A 12-volt car battery is typically a battery of 6 cells in series, in which the positive poles are lead oxide PbO_2, the negative poles are metallic lead and the electrolyte is sulphuric acid. In some batteries, after they are exhausted, the poles are irreversibly damaged and the battery has to be discarded. In others, such as the nickel-cadmium or lead-acid cells, the chemical reaction is reversible, and so the cells can be recharged. I have heard the word "accumulator" used for a rechargeable battery, particularly the lead-acid car battery, but I don't know how general that usage is.

Obviously the purpose of a battery is to extract a current from it. An *electrolytic cell* is quite the opposite. In an electrolytic cell, an electric current is forced into it from outside. This may be done in a laboratory, for example, to study the flow of electricity through an electrolyte, or in industrial processes such as electroplating. In an electrolytic cell, the current is forced into the cell by two *electrodes*, one of which (the *anode*) is maintained at a higher potential than the other (the *cathode*). The electrolyte contains positive ions (*cations*) and negative ions (*anions*), which can flow through the electrolyte. Naturally, the positive ions (cations) flow towards the negative electrode (the cathode) and the negative ions (the anions) flow towards the positive electrode (the anode).

The direction of flow of electricity in an electrolytic cell is the opposite from the flow when a battery is being used to power an external circuit, and the roles of the two poles or electrodes are reversed. Thus some writers will refer to the *positive* pole of a *battery* as its "cathode". It is not surprising therefore, that many a student (and, one might even guess, many a professor and textbook writer) has become confused over the words cathode and anode. The situation is not eased by referring to negatively charged electrons in a gaseous discharge tube as "cathode rays".

My recommendation would be: When referring to an electrolytic cell, use the word "electrodes"; when referring to a battery, use the word "poles". Avoid the use of the prefixes "cat" and "an" altogether. Thus, refer to the positive and negative electrodes of an electrolytic cell, the positive and negative poles of a battery, and the positive and negative ions of an electrolyte. In that way your meaning will always be clear and unambiguous to yourself and to your audience or your readers.

4.2 *Resistance and Ohm's Law*

When a potential difference is maintained across the electrodes in an electrolytic cell, a current flows through the electrolyte. This current is carried by positive ions moving from the positive electrode towards the negative electrode and also, simultaneously, by negative ions moving from the negative electrode towards the positive electrode. The conventional direction of the flow of electricity is the direction in which positive charges are moving. That is to say, electricity flows from the positive electrode towards the negative electrode. The positive ions, then, are moving in the same direction as the conventional direction of flow of electricity, and the negative ions are moving in the opposite direction.

When current flows in a *metal*, the current is carried exclusively by means of negatively charged electrons, and therefore the current is carried exclusively by means of particles that are moving in the opposite direction to the conventional flow of electricity. Thus "electricity" flows from a point of high potential to a point of lower potential; electrons move from a point of low potential to a point of higher potential.

When a potential difference V is applied across a resistor, the ratio of the potential difference across the resistor to the current I that flows through it is called the *resistance*, R, of the resistor. Thus

$$V = IR. \qquad 4.2.1$$

This equation, which defines resistance, appears at first glance to say that *the current through a resistor is proportional to the potential difference across it*, and this is *Ohm's Law.* Equation 4.2.1, however, implies a simple proportionality between V and I only if R is constant and independent of I or of V. In practice, when a current flows through a resistor, the resistor becomes hot, and its resistance increases – and then V and I are no longer linearly proportional to one another. Thus one would have to state Ohm's Law in the form that *the current through a resistor is proportional to the potential difference across it, provided that the temperature is held constant*. Even so, there are some substances (and various electronic devices) in which the resistance is not independent of the applied potential difference even at constant temperature. Thus it is better to regard equation 4.2.1 as a definition of resistance rather than as a fundamental law, while also accepting that it is a good description of the behaviour of most real substances under a wide variety of conditions as long as the temperature is held constant.

Definitions. If a current of one amp flows through a resistor when there is a potential difference of one volt across it, the resistance is one *ohm* (Ω). (Clear though this definition may appear, however, recall from chapter 1 that we have not yet defined exactly what we mean by an amp, nor a volt, so suddenly the meaning of "ohm" becomes a good deal less clear! I do promise a definition of "amp" in a later chapter – but in the meantime I crave your patience.)

The *dimensions* of resistance are $\dfrac{\mathrm{ML^2T^{-2}Q^{-1}}}{\mathrm{T^{-1}Q}} = \mathrm{ML^2T^{-1}Q^{-2}}$.

The reciprocal of resistance is *conductance*, G. Thus $I = GV$. It is common informal practice to express conductance in "mhos", a "mho" being an ohm^{-1}. The official SI unit of conductance, however, is the siemens (S), which is the same thing as a "mho", namely one A V^{-1}.

The *resistance* of a *resistor* is proportional to its length l and inversely proportional to its cross-sectional area A:

$$R = \frac{\rho l}{A}. \qquad 4.2.2$$

The constant of proportionality ρ is called the *resistivity* of the material of which the resistor is made. Its dimensions are $\mathrm{ML^3T^{-1}Q^{-2}}$, and its SI unit is ohm metre, or Ω m.

The reciprocal of resistivity is the *conductivity*, σ. Its dimensions are $M^{-1}L^{-3}TQ^2$, and its SI unit is siemens per metre, S m^{-1}.

For those who enjoy collecting obscure units, there is an amusing unit I once came across, namely the unit of surface resistivity. One is concerned with the resistance of a thin sheet of conducting material, such as, for example, a thin metallic film deposited on glass. The resistance of some rectangular area of this is proportional to the length l of the rectangle and inversely proportional to its width w:

$$R = \frac{\rho l}{w}.$$

The resistance, then, depends on the ratio l/w – i.e. on the shape of the rectangle, rather than on its size. Thus the resistance of a 2 mm × 3 mm rectangle is the same as that of a 2 m × 3 m rectangle, but quite different from that of a 3 mm × 2 mm rectangle. The surface resistivity is defined as the resistance of a rectangle of unit length and unit width (i.e. a square) – and it doesn't matter what the size of the square. Thus the units of surface resistivity are ohms per square. (End of sentence!)

As far as their resistivities are concerned, it is found that substances may be categorized as *metals*, *nonconductors* (insulators), and *semiconductors*. Metals have rather low resistivities, of the order of 10^{-8} Ω m. For example:

Silver:	1.6×10^{-8} Ω m
Copper:	1.7×10^{-8}
Aluminium:	2.8×10^{-8}
Tungsten:	5.5×10^{-8}
Iron:	10×10^{-8}

Nonconductors have resistivities typically of order 10^{14} to 10^{16} Ω m or more. That is, for most practical purposes and conditions they don't conduct any easily measurable electricity at all.

Semiconductors have intermediate resistivities, such as

Carbon:	1500×10^{-8} Ω m
Germanium:	4.5×10^{-1}
Silicon	$6.4 \times 10^{+2}$

There is another way, besides equation 4.2.1, that is commonly used to express Ohm's law. Refer to figure IV.1.

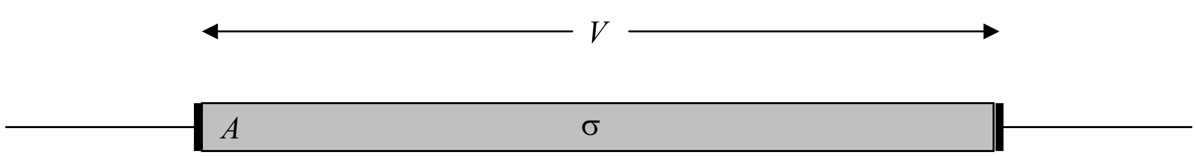


FIGURE IV.1

We have a metal rod of length l, cross-sectional area A, electrical conductivity σ, and so its resistance is $l/(\sigma A)$. We clamp it between two points which have a potential difference of V between them, and consequently the magnitude of the electric field in the metal is $E = V/l$. Equation 4.2.1 (Ohm's law) therefore becomes $El = Il/(\sigma A)$. Now introduce $J = I/l$ as the *current density* (amps per square metre). Them Ohm's law becomes $J = \sigma E$. This is usually written in vector form, since current and field are both vectors, so that Ohm's law is written

$$\mathbf{J} = \sigma\mathbf{E}. \qquad 4.2.3$$

4.3 *Resistance and Temperature*

It is found that the resistivities of metals generally increase with increasing temperature, while the resistivities of semiconductors generally decrease with increasing temperature.

It may be worth thinking a little about how electrons in a metal or semiconductor conduct electricity. In a solid metal, most of the electrons in an atom are used to form covalent bonds between adjacent atoms and hence to hold the solid together. But about one electron per atom is not tied up in this way, and these "conduction electrons" are more or less free to move around inside the metal much like the molecules in a gas. We can estimate roughly the speed at which the electrons are moving. Thus we recall the formula $\sqrt{3kT/m}$ for the root-mean-square speed of molecules in a gas, and maybe we can apply that to electrons in a metal just for a rough order of magnitude for their speed. Boltzmann's constant k is about 1.38×10^{-23} J K^{-1} and the mass of the electron, m, is about 9.11×10^{-31} kg. If we assume that the temperature is about 27°C or 300 K, the root mean square electron speed would be about 1.2×10^{5} m s^{-1}.

Now consider a current of 1 A flowing in a copper wire of diameter 1 mm – i.e. cross-sectional area 7.85×10^{-7} m^2. The density of copper is 8.9 g cm^{-3}, and its "atomic weight" (molar mass) is 63.5 g per mole, which means that there are 6.02×10^{23} (Avogadro's number) of atoms in 63.5 grams, or 8.44×10^{22} atoms per cm^3 or 8.44×10^{28} atoms per m^3. If we assume that there is one conduction electron per atom, then there are 8.44×10^{28} conduction electrons per m^3, or, in our wire of diameter 1 mm, 6.63×10^{22} conduction electrons per metre.

The speed at which the electrons are carrying the current of one amp is the current divided by the charge per unit length, and with the charge on a single electron being 1.60×10^{-19} C, we find that the speed at which the electrons are carrying the current is about 9.4×10^{-5} m s^{-1}.

Thus we have this picture of electrons moving in random directions at a speed of about 1.2×10^{5} m s^{-1} (the thermal motion) and, superimposed on that, a very slow drift speed of only 9.4×10^{-5} m s^{-1} for the electron current. If you were able to see the electrons, you would see them dashing hither and thither at very high speeds, but you wouldn't even notice the very slow drift in the direction of the current.

When you connect a long wire to a battery, however, the current (the slow electron drift) starts almost instantaneously along the entire length of the wire. If the electrons were in a complete vacuum, rather than in the interior of a metal, they would accelerate as long as they were in an electron field. The electrons inside the metal also accelerate, but they are repeatedly stopped in their tracks by collisions with the metal atoms – and then they start up again. If the temperature is increased, the vibrations of the atoms within the metal lattice increase, and this presumably somehow increases the resistance to the electron flow, or decreases the mean time or the mean path-length between collisions.

In a *semiconductor*, most of the electrons are required for valence bonding between the atoms – but there are a few (much fewer than one per atom) free, conduction electrons. As the temperature is increased, more electrons are shaken free from their valence duties, and they then take on the task of conducting electricity. Thus the conductivity of a semiconductor increases with increasing temperature.

The temperature coefficient of resistance, α, of a metal (or other substance) is the fractional increase in its resistivity per unit rise in temperature:

$$\alpha = \frac{1}{\rho}\frac{d\rho}{dT}. \qquad 4.3.1$$

In SI units it would be expressed in K^{-1}. However, in many practical applications the temperature coefficient is defined in relation to the change in resistance compared with the resistivity at a temperature of 20°C, and is given by the equation

$$\rho = \rho_{20}[1 + \alpha(t - 20)], \qquad 4.3.2$$

where t is the temperature in degrees Celsius.

Examples:

Silver:	$3.8 \times 10^{-3}\ \text{C}^{\circ\,-1}$
Copper:	3.9×10^{-3}
Aluminium:	3.9×10^{-3}
Tungsten:	4.5×10^{-3}
Iron:	5.0×10^{-3}
Carbon:	$-0.5 \times 10^{-3}\ \text{C}^{\circ\,-1}$
Germanium:	-48×10^{-3}
Silicon	-75×10^{-3}

Some metallic alloys with commercial names such as nichrome, manganin, constantan, eureka, etc., have fairly large resistivities and very low temperature coefficients.

As a matter of style, note that the *kelvin* is a unit of temperature, much a the *metre* is a unit of length. Thus, when discussing temperatures, there is no need to use the "degree" symbol with the kelvin. When you are talking about some other temperature scale, such as Celsius, one needs to say "20 degrees on the Celsius scale" – thus 20°C. But when one is talking about a temperature *interval* of so many Celsius degrees, this is written C°. I have adhered to this convention above.

The resistivity of platinum as a function of temperature is used as the basis of the *platinum resistance thermometer*, useful under conditions and temperatures where other types of thermometers may not be useful, and it is also used for defining a practical temperature scale at high temperatures. A *bolometer* is an instrument used for detecting and measuring infrared radiation. The radiation is focussed on a blackened platinum disc, which consequently rises in temperature. The temperature rise is measured by measuring the increase in resistance. A *thermistor* is a semiconducting device whose resistance is very sensitive to temperature, and it can be used for measuring or controlling temperature.

4.4 *Resistors in Series*

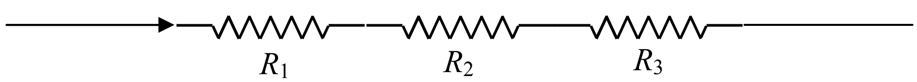


FIGURE IV.2

$$R = R_1 + R_2 + R_3. \qquad 4.4.1$$

The current is the same in each. The potential difference is greatest across the largest resistance.

4.5 *Conductors in Parallel*

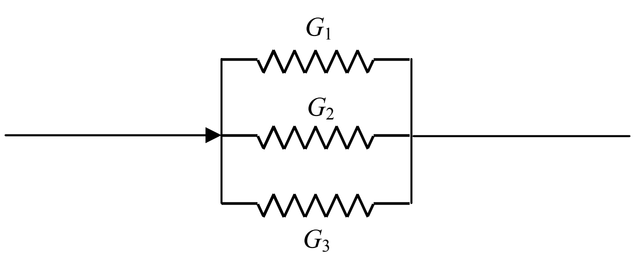


FIGURE IV.3

$$G = G_1 + G_2 + G_3. \qquad 4.5.1$$

That is to say $$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} \cdot \qquad 4.5.2$$

The potential difference is the same across each. The current is greatest through the largest conductance – i.e. through the smallest resistance.

4.6 *Dissipation of Energy*

When current flows through a resistor, electricity is falling through a potential difference. When a coulomb drops through a volt, it loses potential energy 1 joule. This energy is dissipated as heat. When a current of I coulombs per second falls through a potential difference of V volts, the rate of dissipation of energy is IV, which can also be written (by making use of Ohm's law) I^2R or V^2/R.

If two resistors are connected in series, the current is the same in each, and we see from the formula I^2R that more heat is generated in the larger resistance.

If two resistors are connected in parallel, the potential difference is the same across each, and we see from the formula V^2/R that more heat is generated in the small resistance.

4.7 *Electromotive Force and Internal Resistance*

The reader is reminded of the following definition from section 4.1:

Definition. The potential difference across the poles of a cell when no current is being taken from it is called the *electromotive force* (EMF) of the cell.

I shall use the symbol $\mathcal{E}$ for EMF.

Question. A 4 Ω resistance is connected across a cell of EMF 2 V. What current flows? The immediate answer is 0.5 A – but this is likely to be wrong. The reason is that a cell has a resistance of its own – its *internal resistance.* The internal resistance of a lead-acid cell is typically quite small, but most dry cells have an appreciable internal resistance. If the external resistance is R and the internal resistance is r, the total resistance of the circuit is $R + r$, so that the current that flows is $\mathcal{E}/(R + r)$.

Whenever a current is taken from a cell (or battery) the potential difference across its poles *drops* to a value less than its EMF. We can think of a cell as an EMF in series with an internal resistance:

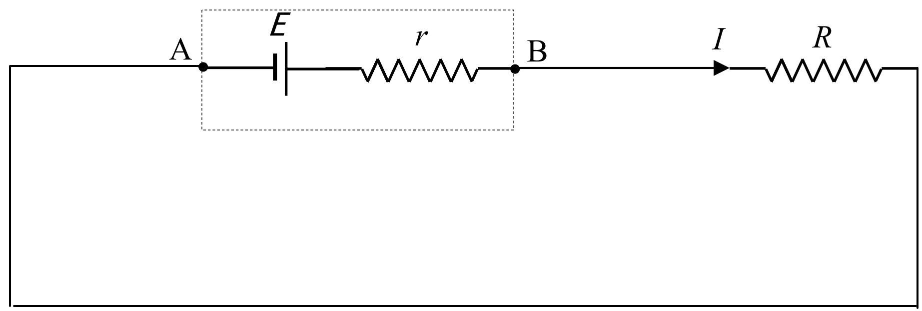


FIGURE IV.4

If we take the point A as having zero potential, we see that the potential of the point B will be $E - Ir$, and this, then, is the potential difference across the poles of the cell when a current I is being taken from it.

Exercise. Show that this can also be written as $\frac{ER}{R + r}$.

4.8 *Power Delivered to an External Resistance*

Question: How much heat will be generated in the external resistance R if $R = 0$?
Answer: None!

Question: How much heat will be generated in the external resistance R if $R = \infty$?
Answer: None!

Question: How much heat will be generated in the external resistance R if R is something?
Answer: Something!

This suggests that there will be some value of the external resistance for which the power delivered, and heat generated, will be a maximum, and this is indeed the case.

The rate at which power is delivered, and dissipated as heat, is

$$P = I^2 R = \frac{E^2 R}{(R + r)^2}. \qquad 4.8.1$$

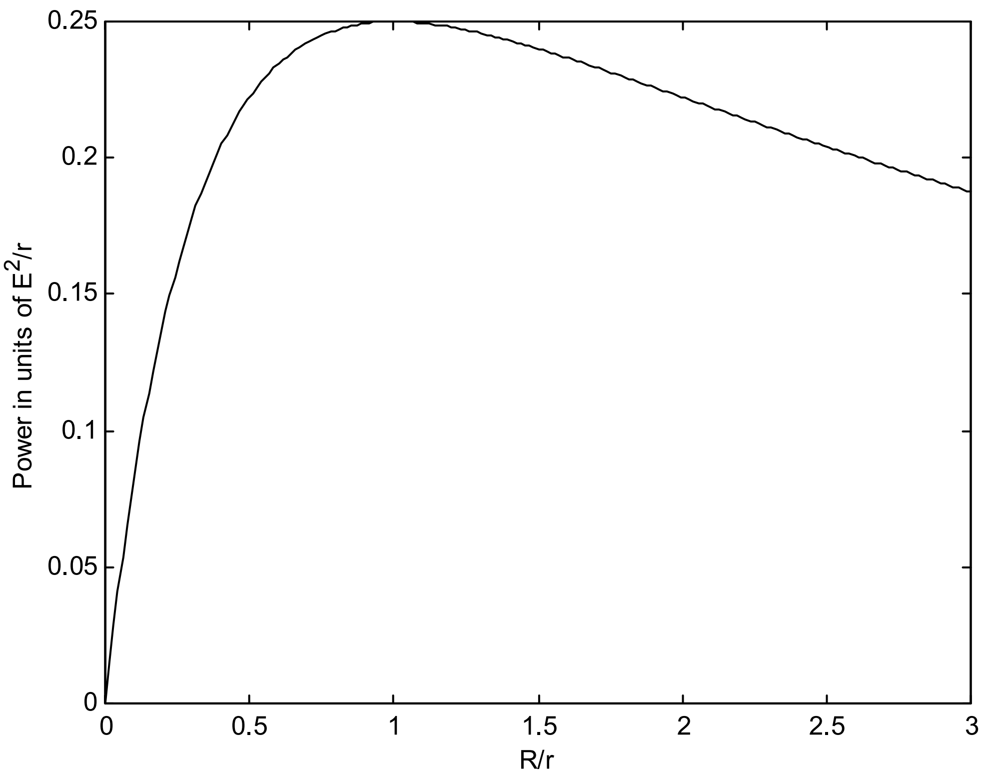
In figure IV.5 I have plotted the power (in units of E^2/R) versus R/r. Differentiation of the above expression (do it!) will show that the power delivered reaches a maximum of

$E^2/(4R)$ when $R = r$; that is, when the external resistance is "matched" to the internal resistance of the cell. This is but one example of many in physics and engineering in which maximum power is delivered to a load when the load is matched to the internal load of the power source.

Exercise. A 6V battery with an internal resistance of 0.5 Ω is connected to an external resistance. Heat is generated in the external resistance at a rate of 12 W. What is the value of the external resistance?

Answer. 1.87 Ω or 0.134 Ω.

FIGURE IV.5



4.9 *Potential Divider*

The circuit illustrated in figure IV.6 is a *potential divider*.

It may be used to supply a variable voltage to an external circuit. It is then called a *rheostat*.

Or it may be used to compare potential differences, in which case it is called a *potentiometer.* (In practice many people refer to such a device as a "pot", regardless of the use to which it is put.)

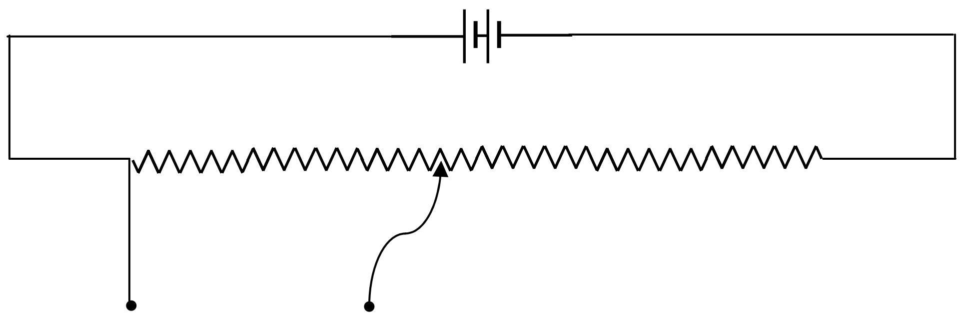


FIGURE IV.6

For example, in figure IV.7, a balance point (no current in the ammeter, A) is found when the potential drop down the length x of the resistance wire is equal to the EMF of the small cell. (Note that, since no current is being taken from the small cell, the potential difference across its poles is indeed the EMF.) One could compare the EMFs of two cells in this manner, one of which might be a "standard cell" whose EMF is known.

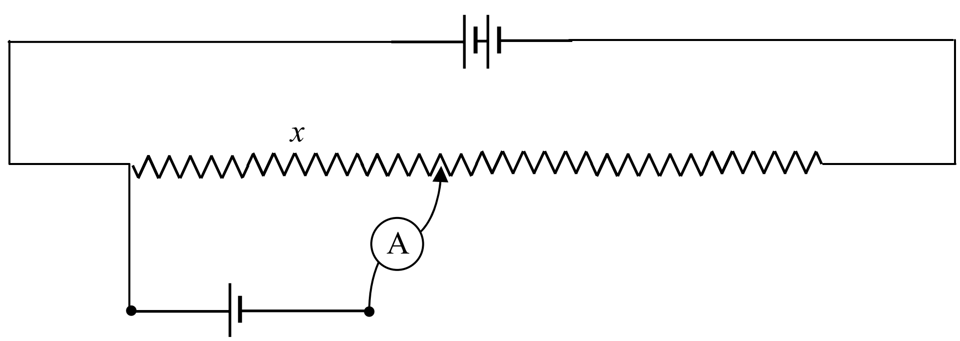


FIGURE IV.7

In figure IV.8, a current is flowing through a resistor (which is assumed to be in part of some external circuit, not drawn), and, assuming that the potential gradient down the potentiometer has been calibrated with a standard cell, the potentiometer is being used to measure the potential difference across the resistor. That is, the potentiometer is being used as a voltmeter.

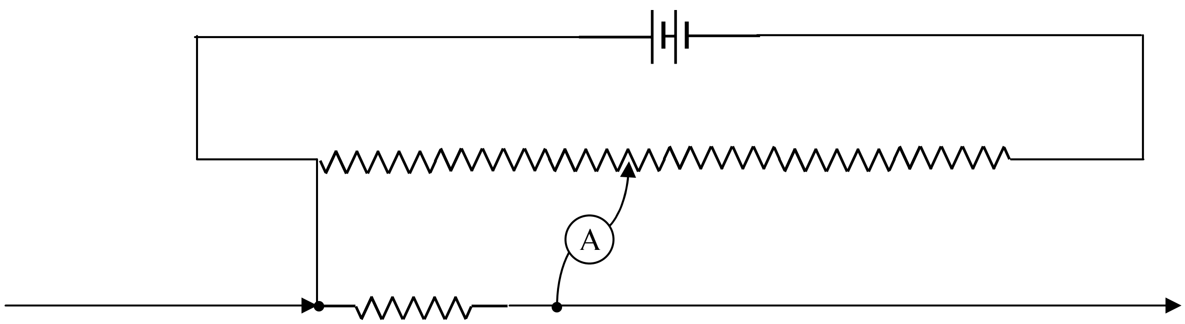


FIGURE IV.8

4.10 *Ammeters and Voltmeters*

For the purpose of this section it doesn't matter how an ammeter actually works. Suffice it to say that a current flows through the ammeter and a needle moves over a scale to indicate the current, or else the current is indicated as numbers in a digital display. In order to measure the current through some element of a circuit, the ammeter is placed, of course, in *series* with the element. Generally an ammeter has rather a low resistance.

An inexpensive *voltmeter* is really just an ammeter having rather a high resistance. If you want to measure the potential difference across some circuit element, you place the voltmeter, of course, *across* that element (i.e. in *parallel* with it). A small portion of the current through the element is diverted through the meter; the meter measures this current, and, from the known resistance of the meter, the potential difference can be calculated – though in practice nobody does any calculation – the scale is marked in volts. Placing a meter *across* a circuit element in fact slightly reduces the potential difference across the element – that is, it reduces the very thing you want to measure. But, because a voltmeter typically has a high resistance, this effect is small. There are, of course, modern (and more expensive) voltmeters of a quite different design, which take no current at all, and genuinely measure potential difference, but we are concerned in this section with the commonly-encountered ammeter-turned-voltmeter. It may be noticed that the potentiometer described in the previous section takes no current from the circuit element of interest, and is therefore a true voltmeter.

There are meters known as "multimeters" or "avometers" (for amps, volts and ohms), which can be used as ammeters or as voltmeters, and it is with these that this section is concerned.

A typical inexpensive ammeter gives a full scale deflection (FSD) when a current of 15 mA = 0.015 A flows through it. It can be adapted to measure higher currents by connecting a small resistance (known as a "shunt") *across* it.

Let's suppose, for example, that we have a meter that which shows a FSD when a current of 0.015 A flows through it, and that the resistance of the meter is 10 Ω. We would like to use the meter to measure currents as high as 0.15 A. What value of shunt resistance shall we put across the meter? Well, when the total current is 0.15 A, we want 0.015 A to flow through the meter (which then shows FSD) and the remainder, 0.135 A, is to flow through the shunt. With a current of 0.015 A flowing through the 10 Ω meter, the potential difference across it is 0.15 V. This is also the potential difference across the shunt, and, since the current through the shunt is 0.135 A, the resistance of the shunt must be 1.11 Ω.

We can also use the meter as a voltmeter. Suppose, for example, that we want to measure voltages (horrible word!) of up to 1.5 V. We place a large resistance R in *series* with the meter, and then place the meter-plus-series-resistance across the potential difference to measured. The total resistance of meter-plus-series-resistance is $(10 + R)$, and it will show a FSD when the current through it is 0.015 A. We want this to happen when the potential difference across it is 1.5 volts. This $1.5 = 0.015 \times (10 + R)$, and so $R = 90$ Ω.

4.11 *Wheatstone Bridge*

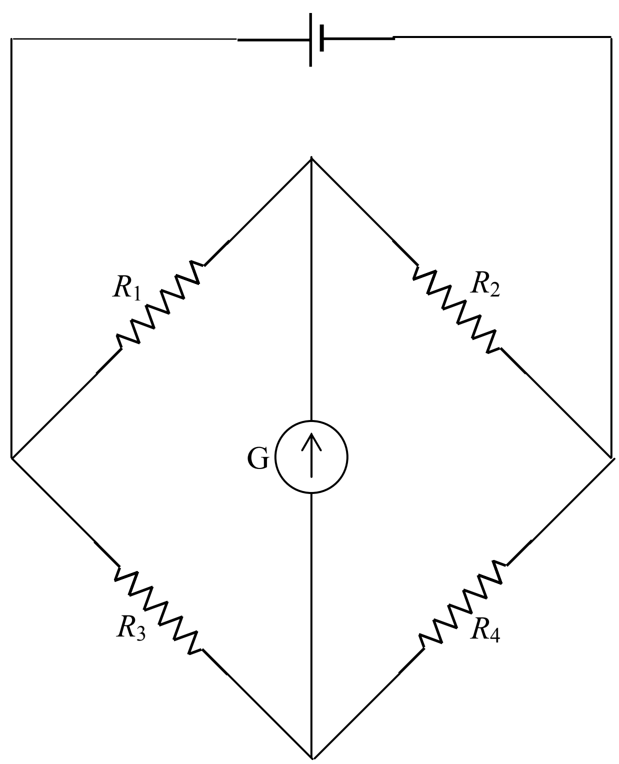


FIGURE IV.9

The Wheatstone bridge can be used to *compare* the value of two resistances – or, if the unknown resistance is compared with a resistance whose value is known, it can be used to *measure* an unknown resistance. R_1 and R_2 can be varied. R_3 is a standard resistance whose value is known. R_4 is the unknown resistance whose value is to be determined. G is a *galvanometer*. This is just a sensitive ammeter, in which the zero-current position has the needle in the middle of the scale; the needle may move one way or the other, depending on which way the current is flowing. The function of the galvanometer is not so much to *measure* current, but merely to *detect* whether or not a current is flowing in one direction of another. In use, the resistances R_1 and R_2 are varied until no current flows in the galvanometer. The bridge is then said to be "balanced" and $R_1/R_2 = R_3/R_4$, and hence the unknown resistance is given by $R_4 = R_1R_3/R_2$.

4.12 *Delta-Star Transform*

Consider the two circuits (each enclosed in a black box) of figure IV.10.

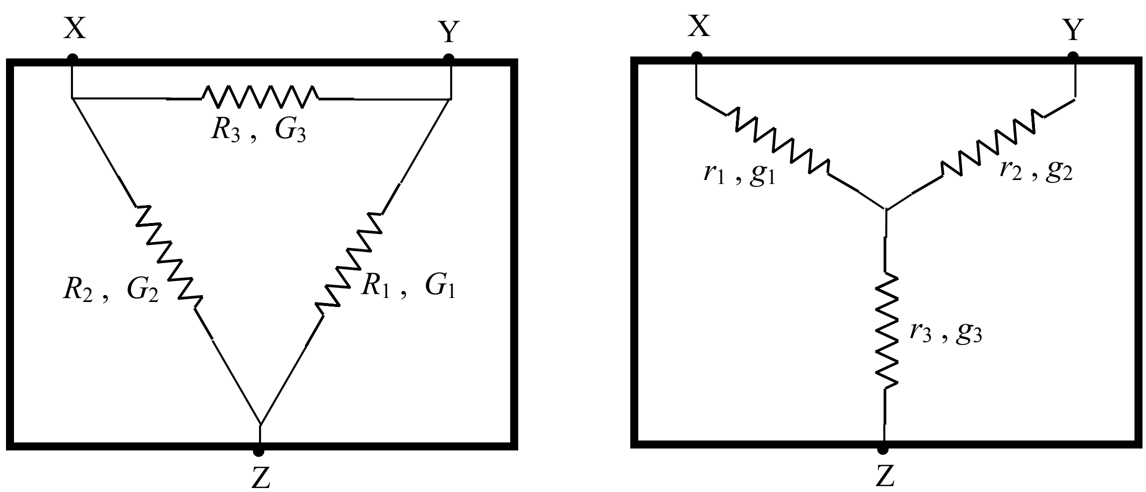


FIGURE IV.10

The configuration in the left hand box is called a "delta" (Δ) and the configuration in the right hand box is called a "star" or a "Y". I have marked against each resistor its resistance and its conductance, the conductance, of course, merely being the reciprocal of the resistance. I am going to suppose that the resistance between the terminals X and Y is the same for each box. In that case:

$$r_1 + r_2 = \frac{R_3(R_1 + R_2)}{R_1 + R_2 + R_3}. \qquad 4.12.1$$

We can get similar equations for the terminal pairs Y,Z and Z,X. Solving the three equations for r_1 , r_2 and r_3, we obtain

$$r_1 = \frac{R_2R_3}{R_1 + R_2 + R_3}\ , \tag{4.12.2}$$

$$r_2 = \frac{R_3R_1}{R_1 + R_2 + R_3} \tag{4.12.3}$$

and

$$r_3 = \frac{R_1R_2}{R_1 + R_2 + R_3}\ . \tag{4.12.4}$$

In terms of the conductances, these are

$$g_1 = \frac{G_2G_3 + G_3G_1 + G_1G_2}{G_1}\ , \tag{4.12.5}$$

$$g_2 = \frac{G_2G_3 + G_3G_1 + G_1G_2}{G_2} \tag{4.12.6}$$

and

$$g_3 = \frac{G_2G_3 + G_3G_1 + G_1G_2}{G_3}\ . \tag{4.12.7}$$

The converses of these equations are:

$$R_1 = \frac{r_2r_3 + r_3r_1 + r_1r_2}{r_1}\ , \tag{4.12.8}$$

$$R_2 = \frac{r_2r_3 + r_3r_1 + r_1r_2}{r_2}\ , \tag{4.12.9}$$

$$R_3 = \frac{r_2r_3 + r_3r_1 + r_1r_2}{r_3}\ , \tag{4.12.10}$$

$$G_1 = \frac{g_2g_3}{g_1 + g_2 + g_3}\ , \tag{4.12.11}$$

$$G_2 = \frac{g_3g_1}{g_1 + g_2 + g_3} \tag{4.12.12}$$

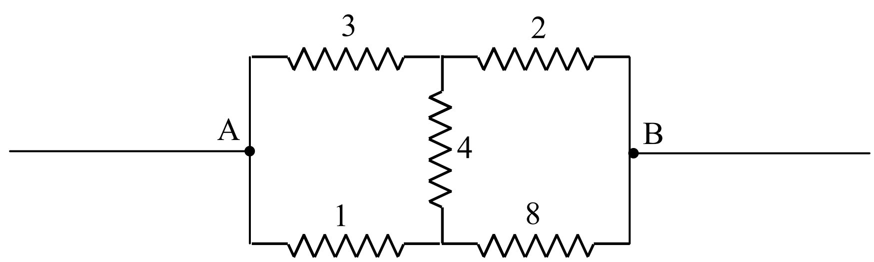
and

$$G_3 = \frac{g_1g_2}{g_1 + g_2 + g_3}\ . \tag{4.12.13}$$

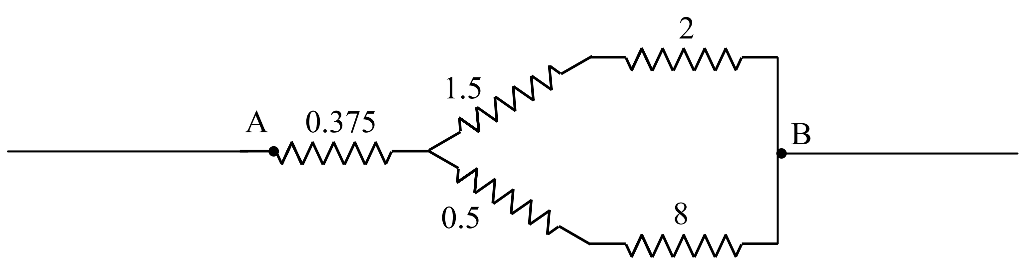
That means that, if the resistances and conductances in one box are related to the resistances and conductances in the other by these equations, then you would not be able to tell, if you had an ammeter, and a voltmeter and an ohmmeter, which circuit was in which box. The two boxes are indistinguishable from their electrical behaviour.

These equations are not easy to commit to memory unless you are using them every day, and they are sufficiently awkward that mistakes are likely when evaluating them numerically. Therefore, to make the formulas useful, you should programme your calculator or computer so that they will instantly convert between delta and star without your ever having to think about it. The next example shows the formulas in use. It will be heavy work unless you have programmed your computer in advance – but if you *have* done so, you will see how very useful the transformations are.

Example. Calculate the resistance between the points A and B in the figure below. The individual resistances are given in ohms.



At first, one doesn't know how to start. But notice that the 1, 3 and 4 ohm resistors are connected in delta and the circuit is therefore equivalent to



After that, it is easy, and you will soon find that the resistance between A and B is 2.85 Ω.

4.13 *Kirchhoff's Rules*

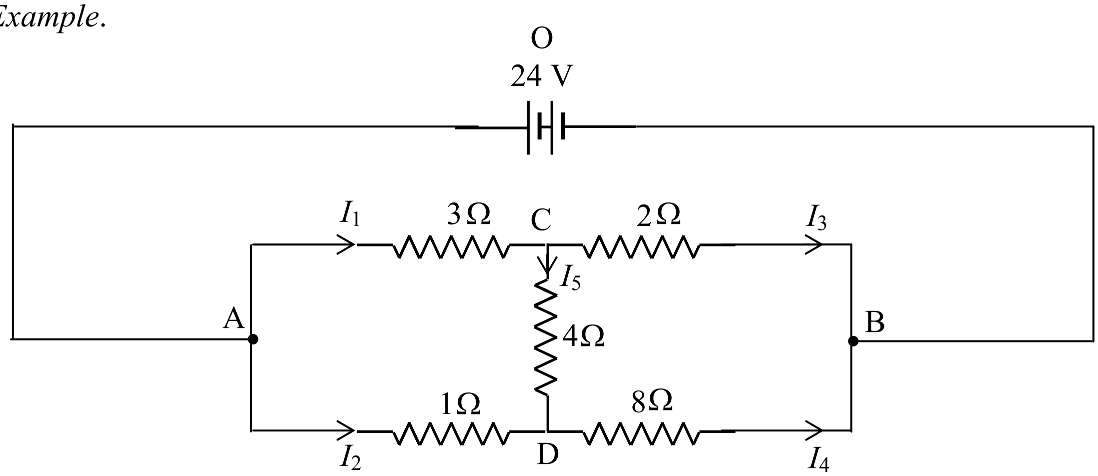
There are two h's in his name, and there is no *tch* sound in the middle. The pronunciation is approximately keerr–hhofe.

The rules themselves are simple and are self-evident. What has to be learned, however, is the art of using them.

K1: The net current going into any *point* in a circuit is zero; expressed otherwise, the sum of all the currents entering any point in a circuit is equal to the sum of all the currents leaving the point.

K2: The sum of all the EMFs and *IR* products in a *closed circuit* is zero. Expressed otherwise, as you move around a closed circuit, the potential will sometimes rise and sometimes fall as you encounter a battery or a resistance; but, when you come round again to the point where you started, there is no change in potential.

Example.



In the above circuit, the 24 V battery is assumed to have negligible internal resistance. Calculate the current in each of the resistors.

The art of applying Kirchhoff's rules is as follows.

1. Draw a large circuit diagram *in pencil*.

2. Count the number of independent resistors. (Two is series with nothing in between don't count as independent.) This tells you how many independent equations you can obtain, and how many unknowns you can solve for. In this case, there are five independent resistors; you can get five independent equations and you can solve for five unknowns.

3. Mark in the unknown currents. If you don't know the directions of some of them, don't spend time trying to think it out. Just make a wild guess. If you are wrong, you will merely get a negative answer for it. Those who have some physical insight might already guess (correctly) that I have marked I_5 in the wrong direction, but that doesn't matter.

4. Choose any closed circuit and apply K2. Go over that closed surface *in ink*. Repeat for several closed circuits until the entire diagram is inked over. When this happens, you cannot get any further independent equations using K2. If you try to do so, you will merely end up with another equation that is a linear combination of the ones you already have,

5. Make up the required number of equations with K1.

Let us apply these to the present problem. There are five resistors; we need five equations. Apply K2 to OACBO. Start at the negative pole of the battery and move counterclockwise around the circuit. When we move up to the positive pole, the potential has gone up by 24 V. When we move down a resistor in the direction of the current, the potential goes down. For the circuit OACBO, K2 results in

$$24 - 3I_1 - 2I_3 = 0.$$

Now do the same this with circuit OADBO:

$$24 - I_2 - 8I_4 = 0,$$

and with circuit ACDA:

$$3I_1 + 4I_5 - I_2 = 0.$$

If you have conscientiously inked over each circuit as you have done this, you will now find that the entire diagram is inked over. You cannot gain any further independent equations from K2. We need two more equations. Apply K1 to point C:

$$I_1 = I_3 + I_5,$$

and to point D: $\quad I_4 = I_2 + I_5.$

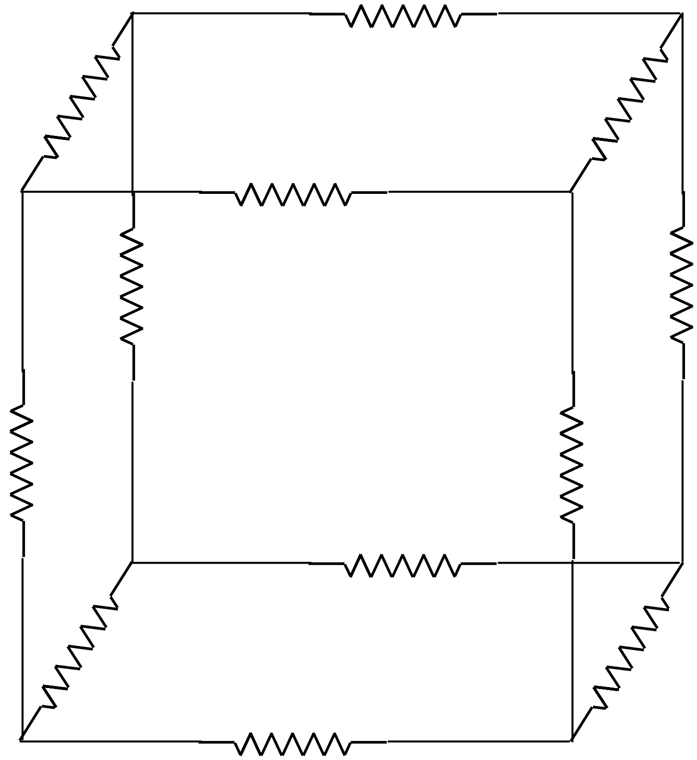
You now have five independent linear equations in five unknowns and you can solve them. (Methods for solving simultaneous linear equations are given in Chapter 1, Section 1.7 of Celestial Mechanics.) The solutions are:

$$I_1 = +4.029\text{A},\ I_2 = +4.380\text{A},\ I_3 = +5.956\text{A},\ I_4 = +2.453\text{A},\ I_5 = -1.927\text{A}.$$

4.14 *Tortures for the Brain*

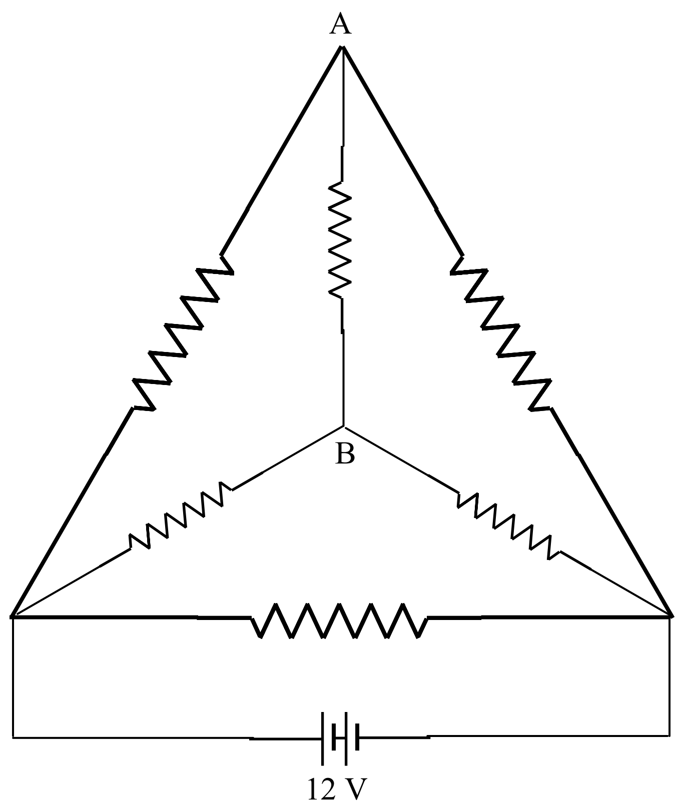
I don't know if any of the examples in this section have any practical applications, but they are excellent ways for torturing students, or for whiling away rainy Sunday afternoons.

4.14.1



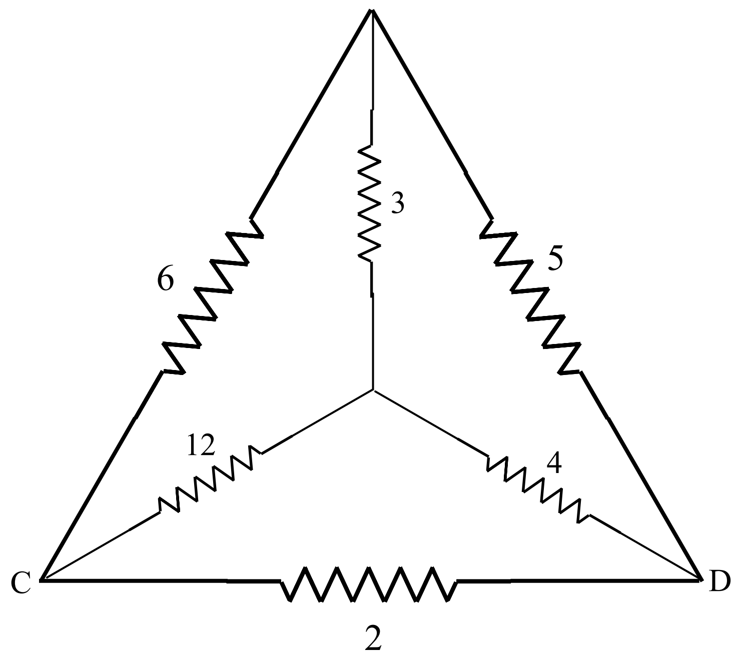
The drawing shows 12 resistances, each of value $r\ \Omega$, arranged along the edges of a cube. What is the resistance across opposite corners of the cube?

4.14.2



The drawing shows six resistors, each of resistance 1 Ω, arranged along the edges of a tetrahedron. A 12 V battery is connected across one of the resistors. Calculate the current between points A and B.

4.14.3

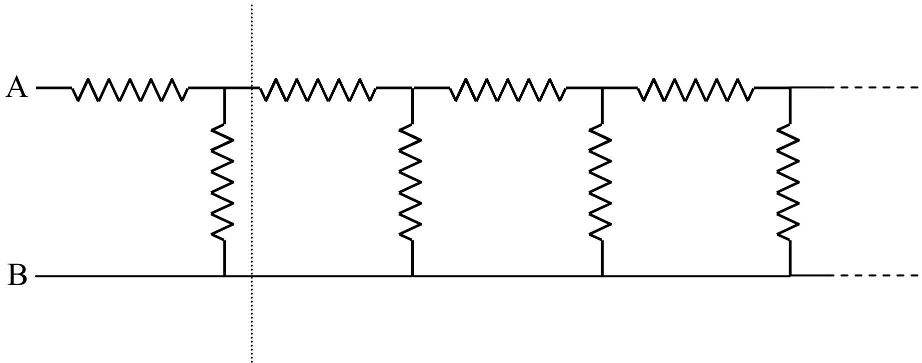


The figure shows six resistors, whose resistances in ohms are marked, arranged along the edges of a tetrahedron. Calculate the net resistance between C and D.

4.14.4 $R_1 = 8\ \Omega$ and $R_2 = 0.5\ \Omega$ are connected across a battery. The rate at which heat is generated is the same whether they are connected in series or in parallel. What is the internal resistance r of the battery?

4.14.5 $R_1 = 0.25\ \Omega$ and $R_2 = ?$ are connected across a battery whose internal resistance r is 0.5 Ω. The rate at which heat is generated is the same whether they are connected in series or in parallel. What is the value of R_2?

4.14.6



In the above circuit, each resistance is 1 ohm. What is the net resistance between A and B if the chain is of infinite length?

4.14.7 What is the resistance between A and B in question 4.14.6 if the chain is not of infinite length, but has n "links" – i.e. $2n$ resistors in all?

4.15 *Solutions, Answers or Hints to 4.14*

Hints for 4.14.1. Imagine a current of $6I$ going into the bottom left hand corner. Follow the current through the cube, writing down the current through each of the 12 resistors. Also write down the potential drop across each resistor, and hence the total potential drop across the cube. I make the answer for the effective resistance of the whole cube $\frac{5}{6}r$.

Solution for 4.14.2. By symmetry, the potentials of A and B are equal. Therefore there is no current between A and B.

Hint for 4.14.3. Replace the heavily-drawn delta with its corresponding star. After that it should be straightforward, although there is a little bit of calculation to do. I make the answer $1.52\ \Omega$.

Solution for 4.14. From equation 4.8.1, the rate at which heat is generated in a resistance R connected across a battery of EMF $\mathcal{E}$ and internal resistance r is $\frac{\mathcal{E}^2 R}{(R+r)^2}$.

If the resistors are connected in series, $R = R_1 + R_2$, while if they are connected in parallel, $R = \frac{R_1 R_2}{R_1 + R_2}$. If the heat generated is the same in either case, we must have

$$\frac{R_1 + R_2}{(R_1 + R_2 + r)^2} = \frac{\dfrac{R_1R_2}{R_1 + R_2}}{\left(\dfrac{R_1R_2}{R_1 + R_2} + r\right)^2}.$$

After some algebra, we obtain

$$r = \frac{R_1 + R_2 - \sqrt{R_1 R_2}}{\sqrt{\dfrac{R_1}{R_2}} + \sqrt{\dfrac{R_2}{R_1}} - 1}. \qquad 4.15.1$$

With $R_1 = 8\ \Omega$ and $R_2 = 0.5\ \Omega$, we obtain $\underline{r = 2.00\ \Omega\ .}$

Solution for 4.14.5.

In equation 4.15.1, let $\dfrac{r}{R_1} = a$ and $\sqrt{\dfrac{R_2}{R_1}} = x$. The equation 4.15.1 becomes

$$a = \frac{1 + x^2 - x}{(1/x) + x - 1}. \qquad 4.15.2$$

Upon rearrangement, this is

$$a - (a+1)x + (a+1)x^2 - x^3 = 0. \qquad 4.15.3$$

In our example, $a = \dfrac{r}{R_1} = \dfrac{0.50}{0.25} = 2$, so that equation 4.15.4 is

$$2 - 3x + 3x^2 - x^3 = 0,$$

or

$$(2 - x)(1 - x + x^2) = 0.$$

The only real root is $x = 2$. But $R_2 = R_1x^2 = 0.25x^2 = 1\Omega$.

<u>Solution to 4.14.6</u> Suppose that there are n links ($2n$ resistors) to the right of the dotted line, and that the effective resistance of these n links is R_n. Add one more link, to the left. The effective resistance of the $n + 1$ links is then

$$R_{n+1} = \frac{2R_n + 1}{R_n + 1}. \qquad 4.15.4$$

As $n \to \infty$, $R_{n+1} \to R_n \to R$. $\therefore$ $R = \frac{2R + 1}{R + 1}$, or $R^2 - R - 1 = 0$.

Whence, $R = \frac{1}{2}(\sqrt{5} + 1) = 1.618\ 033\ 989\ \Omega$.

<u>Solution to 4.14.7</u> By repeated application of equation 4.15.4, we find:

n	R_n	
1	2	
2	$\frac{5}{3}$	= 1.666 666 667
3	$\frac{13}{8}$	= 1.625 000 000
4	$\frac{34}{21}$	= 1.619 047 619
5	$\frac{89}{55}$	= 1.618 181 818
6	$\frac{233}{144}$	= 1.618 055 556
7	$\frac{610}{377}$	= 1.618 037 135
8	$\frac{1597}{987}$	= 1.618 034 448
9	$\frac{4181}{2584}$	= 1.618 034 056
10	$\frac{10946}{6765}$	= 1.618 033 999
11	$\frac{28657}{17711}$	= 1.618 033 990
12	$\frac{75025}{46368}$	= 1.618 033 989

Inspection shows that $R_n = \frac{F_{2n+1}}{F_{2n}}$, where F_m is the mth member of the Fibonacci sequence: 1 1 2 3 5 8 13 21 ......

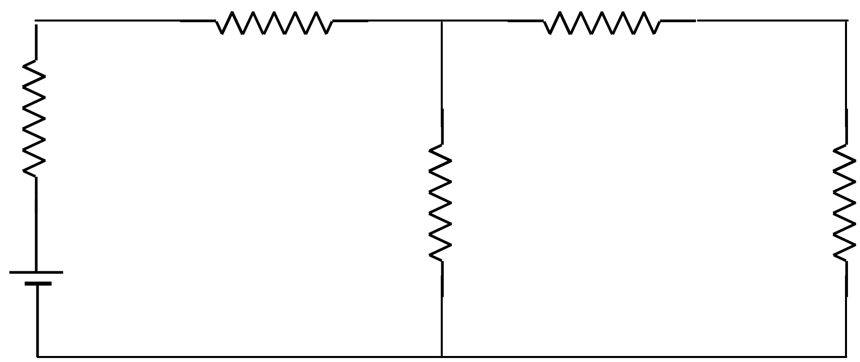
But, from the theory of Fibonacci sequences,

$$F_m = \frac{1}{\sqrt{5}}\left\{\left(\frac{1 + \sqrt{5}}{2}\right)^m - \left(\frac{1 - \sqrt{5}}{2}\right)^m\right\}.$$

Hence $$R_n = \frac{1}{2}\left(\frac{(1+\sqrt{5})^{2n+1} - (1-\sqrt{5})^{2n+1}}{(1+\sqrt{5})^{2n} - (1-\sqrt{5})^{2n}}\right)\ \Omega$$

4.16 *Attenuators*

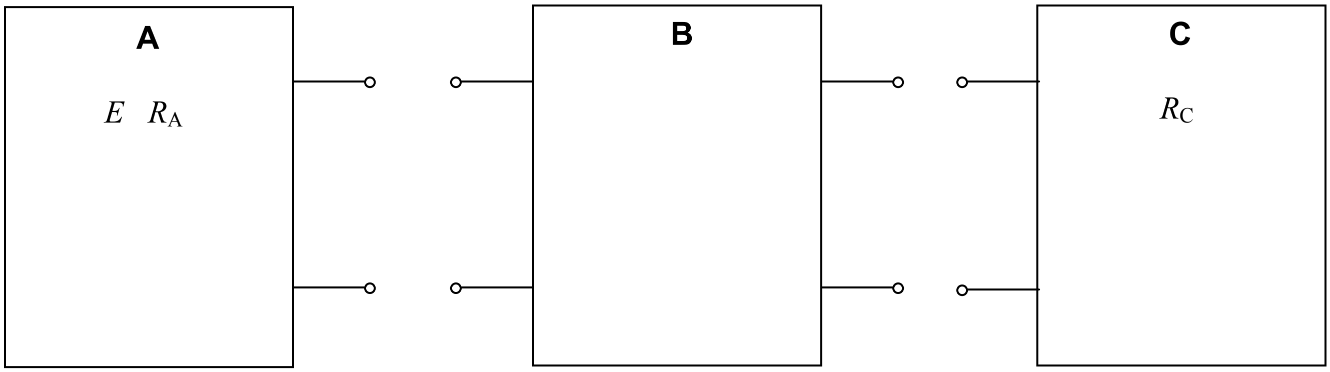
These are networks, usually of resistors, that serve the dual purpose of supplying more examples for students or for reducing the voltage, current or power from one circuit to another. An example of the former might be:



You might be told the values of the four left-hand resistances and of the EMF of the cell, and you are asked to find the current in the right hand resistor.

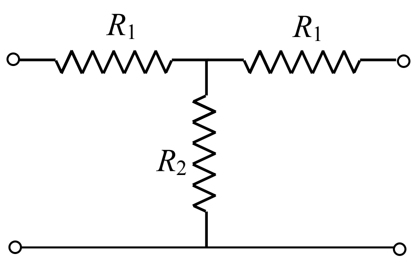
On the other hand, if the object is to design an attenuator, you might be told the values of the resistances at the two ends, and you are required to find the resistances of the three middle resistors such that the current in the rightmost resistor is half the current in the left hand resistor. The three intermediate resistances perform the function of an *attenuator.*

In the drawing below, **A** is some sort of a device, or electrical circuit, or, in the simplest case, just a battery, which has an electromotive force E and an internal resistance $R_{\rm A}$. **C** is some other device, whose internal resistance is $R_{\rm C}$. **B** is an attenuator, which is a collection of resistors which you want to design so that the current delivered to **C** is a certain fraction of the current flowing from **A**; or so that the voltage delivered across the terminals of **C** is a certain fraction of the voltage across the terminals of **A**; or perhaps again so that the power delivered to **C** is a certain fraction of the power generated by **A**. The circuit in the attenuator has to be designed so as to achieve one of these goals.

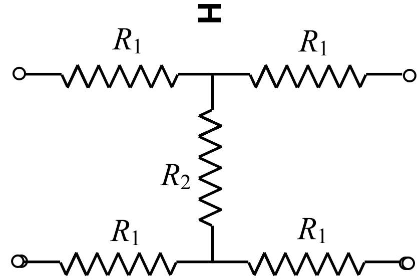


Four simple attenuators are known as **T**, **H,** Π or square, named after their shapes. In the drawing below, the **H** is on its side, like this: **H**

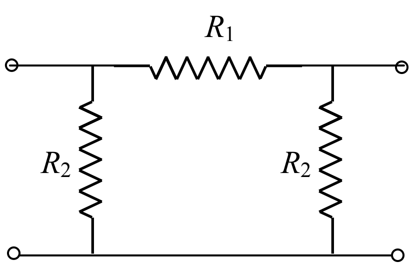
T



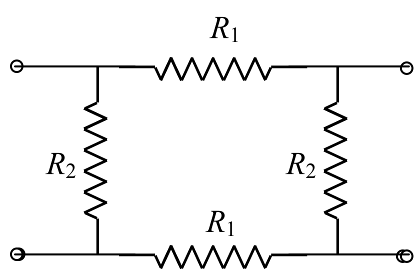
H



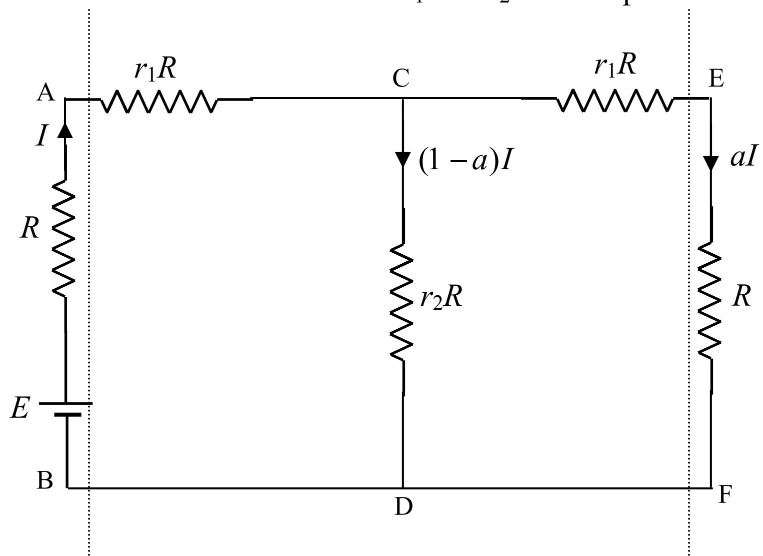
Π



□



Let us look at a **T** attenuator. We'll suppose that the **A** device has an electromotive force E and an internal resistance R, and indeed it can be represented by a cell in series with a resistor. And we'll suppose that the resistance of the **C** device is also R and it can be represented by a single resistor. We'll suppose that we want the current that flows into **C** to be a fraction a of the current flowing out of **A**, and the voltage to be supplied to **C** to be a fraction a of the potential difference across the output terminals of **A**. What must be the values of the resistances in the **T** attenuator? I'll call them r_1R and r_2R, so that we have to determine the dimensionless ratios r_1 and r_2. The equivalent circuit is:



The current leaving the battery is I, and we want the current entering the load at the right hand side to be aI. The current down the middle resistor is then necessarily, by Kirchhoff's first rule, $(1-a)I$. If we apply Kirchhoff's second rule to the outermost circuit, we obtain (after algebraic reduction)

$$E = (1+a)(1+r_1)IR. \tag{4.16.1}$$

We also want the potential difference across the load (i.e. across EF), which, by Ohm's law, is aIR, to be a times the potential difference across the source AB. AB are the terminals of the source. Recall that E is the EMF of the source, and R its internal resistance, so that, when a current I is being taken from the source, the potential difference across its terminals AB is $E - IR$, and we want aIR to be a times this. The fraction a can be called the *voltage reduction factor* of the attenuator. Thus we have

$$aIR = a(E - IR). \tag{4.16.2}$$

From these two equations we obtain

$$\underline{\underline{r_1 = \frac{1-a}{1+a}}}. \tag{4.16.3}$$

Application of Kirchhoff's second rule to the right hand circuit gives

$$r_2(1 - a)IR \;=\; a(1 + r_1)IR, \tag{4.16.4}$$

which, in combination with equation 4.16.3, yields

$$\underline{\underline{r_2 \;=\; \frac{2a}{1 - a^2}}}\,. \tag{4.16.5}$$

Thus, if source and load resistance are each equal to R, and we want a voltage reduction factor of $\frac{1}{2}$, we must choose r_1R to be $\frac{1}{3}R$, and r_2R to be $\frac{4}{3}R$.

You might like to try the same problem for the ⌶ , Π and square attenuators. I am not absolutely certain (I haven't checked them carefully), but I believe the answers are:

For ⌶ :

$$r_1 \;=\; \frac{1}{2}\left(\frac{1 - a}{1 + a}\right) \qquad\qquad r_2 \;=\; \frac{2a}{1 - a^2} \tag{4.16.6}$$

For Π :

$$r_1 \;=\; \frac{1 - a^2}{2a} \qquad\qquad r_2 \;=\; \frac{1 + a}{1 - a} \tag{4.16.7}$$

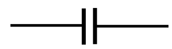
For square :

$$r_1 \;=\; \frac{1}{2}\left(\frac{1 - a^2}{2a}\right) \qquad\qquad r_2 \;=\; \frac{1 + a}{1 - a} \tag{4.16.7}$$

CHAPTER 5
CAPACITORS

5.1 *Introduction*

A capacitor consists of two metal plates separated by a nonconducting medium (known as the *dielectric medium* or simply the *dielectric*, or by a vacuum. It is represented by the electrical symbol

 .

Capacitors of one sort or another are included in almost any electronic device. Physically, there is a vast variety of shapes, sizes and construction, depending upon their particular application. This chapter, however, is not primarily concerned with real, practical capacitors and how they are made and what they are used for, although a brief section at the end of the chapter will discuss this. In addition to their practical uses in electronic circuits, capacitors are very useful to professors for torturing students during exams, and, more importantly, for helping students to understand the concepts of and the relationships between electric fields **E** and **D**, potential difference, permittivity, energy, and so on. The capacitors in this chapter are, for the most part, imaginary academic concepts useful largely for pedagogical purposes. Need the electronics technician or electronics engineer spend time on these academic capacitors, apparently so far removed from the real devices to be found in electronic equipment? The answer is surely and decidedly <u>*yes*</u> – more than anyone else, the practical technician or engineer must thoroughly understand the basic concepts of electricity before even starting with real electronic devices.

If a potential difference is maintained across the two plates of a capacitor (for example, by connecting the plates across the poles of a battery) a charge $+Q$ will be stored on one plate and $-Q$ on the other. The ratio of the charge stored on the plates to the potential difference V across them is called the *capacitance* C of the capacitor. Thus:

$$Q = CV. \tag{5.1.1}$$

If, when the potential difference is one volt, the charge stored is one coulomb, the capacitance is one *farad*, F. Thus, a farad is a coulomb per volt. It should be mentioned here that, in practical terms, a farad is a very large unit of capacitance, and most capacitors have capacitances of the order of microfarads, μF.

The dimensions of capacitance are $\dfrac{\mathrm{Q}}{\mathrm{ML^2T^{-2}Q^{-1}}} = \mathrm{M^{-1}L^{-2}T^2Q^2}$.

It might be remarked that, in older books, a capacitor was called a "condenser", and its capacitance was called its "capacity". Thus what we would now call the "capacitance of a capacitor" was formerly called the "capacity of a condenser".

In the highly idealized capacitors of this chapter, the linear dimensions of the plates (length and breadth, or diameter) are supposed to be very much larger than the separation between them. This in fact is nearly always the case in real capacitors, too, though perhaps not necessarily for the same reason. In real capacitors, the distance between the plates is small so that the capacitance is as large as possible. In the imaginary capacitors of this chapter, I want the separation to be small so that the electric field between the plates is uniform. Thus the capacitors I shall be discussing are mostly like figure V.1, where I have indicated, in blue, the electric field between the plates:

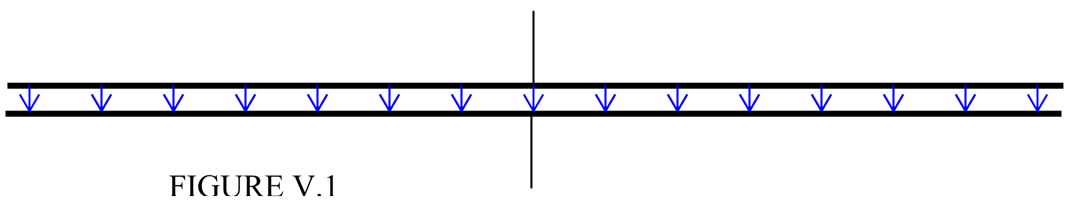
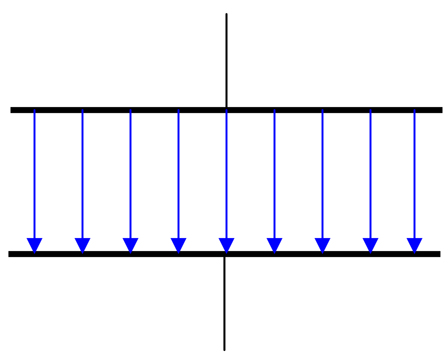


FIGURE V.1

However, I shall not always draw them like this, because it is rather difficult to see what is going on inside the capacitor. I shall usually much exaggerate the scale in one direction, so that my drawings will look more like this:

FIGURE V.2

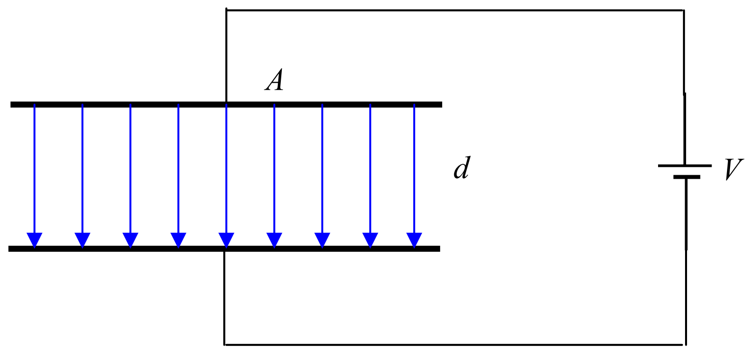


If the separation were really as large as this, the field would not be nearly as uniform as indicated; the electric field lines would greatly bulge outwards near the edges of the plates.

In the next few sections we are going to derive formulas for the capacitances of various capacitors of simple geometric shapes.

5.2 *Plane Parallel Capacitor*

FIGURE V.3



We have a capacitor whose plates are each of area A, separation d, and the medium between the plates has permittivity ε. It is connected to a battery of EMF V, so the potential difference across the plates is V. The electric field between the plates is $E = V/d$, and therefore $D = \varepsilon E/d$. The total D-flux arising from the positive plate is DA, and, by Gauss's law, this must equal Q, the charge on the plate.

Thus $Q = \varepsilon AV/d$, and therefore the capacitance is

$$C = \frac{\varepsilon A}{d}. \qquad 5.2.1$$

Verify that this is dimensionally correct, and note how the capacitance depends upon ε, A and d.

In Section 1.5 we gave the SI units of permittivity as $\mathrm{C^2\,N^{-1}\,m^{-2}}$. Equation 5.2.1 shows that a more convenient SI unit for permittivity is $\mathrm{F\,m^{-1}}$, or farads per metre.

Question: If the separation of the plates is not small, so that the electric field is not uniform, and the field lines bulge outwards at the edge, will the capacitance be less than or greater than $\varepsilon A/d$?

5.3 *Coaxial Cylindrical Capacitor*

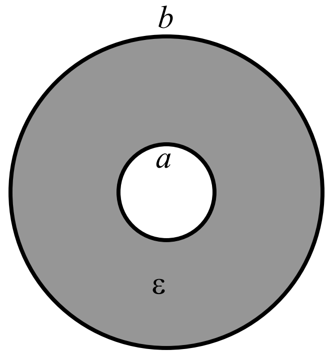


FIGURE V.4

The radii of the inner and outer cylinders are a and b, and the permittivity between them is ε. Suppose that the two cylinders are connected to a battery so that the potential difference between them is V, and the charge per unit length on the inner cylinder is $+\lambda$ C m^{-1}, and on the outer cylinder is $-\lambda$ C m^{-1}. We have seen (Subsection 2.2.3) that the potential difference between the cylinders under such circumstances is $\frac{\lambda}{2\pi\varepsilon}\ln(b/a)$. Therefore the capacitance per unit length, C', is

$$C' = \frac{2\pi\varepsilon}{\ln(b/a)}. \qquad 5.3.1$$

This is by no means solely of academic interest. The capacitance per unit length of coaxial cable ("coax") is an important property of the cable, and this is the formula used to calculate it.

5.4 *Concentric Spherical Capacitor*

Unlike the coaxial cylindrical capacitor, I don't know of any very obvious practical application, nor quite how you would construct one and connect the two spheres to a battery, but let's go ahead all the same. Figure V.4 will do just as well for this one.

The two spheres are of inner and outer radii a and b, with a potential difference V between them, with charges $+Q$ and $-Q$ on the inner and outer spheres respectively. The potential difference between the two spheres is then $\frac{Q}{4\pi\varepsilon}\left(\frac{1}{a}-\frac{1}{b}\right)$, and so the capacitance is

$$C = \frac{4\pi\varepsilon}{\frac{1}{a}-\frac{1}{b}}. \qquad 5.4.1$$

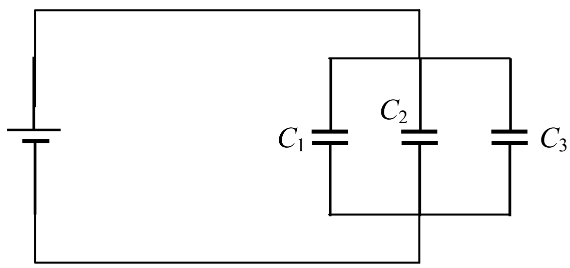
If $b \to \infty$, we obtain for the capacitance of an isolated sphere of radius a:

$$C = 4\pi\varepsilon a. \tag{5.4.2}$$

Exercise: Calculate the capacitance of planet Earth, of radius 6.371×10^3 km, suspended in free space. I make it 709 μF - which may be a bit smaller than you were expecting.

5.5 *Capacitors in Parallel*

FIGURE V.5



The potential difference is the same across each, and the total charge is the sum of the charges on the individual capacitor. Therefore:

$$C = C_1 + C_2 + C_3. \tag{5.5.1}$$

5.6 *Capacitors in Series*

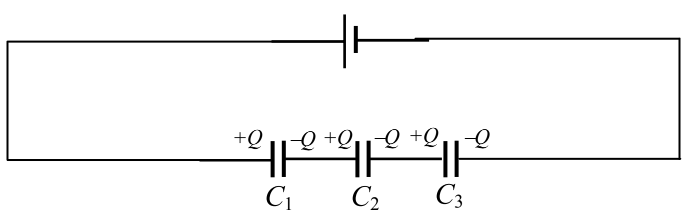


FIGURE V.6

The charge is the same on each, and the potential difference across the system is the sum of the potential differences across the individual capacitances. Hence

$$\frac{1}{C} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3}. \tag{5.6.1}$$

5.7 *Delta-Star Transform*

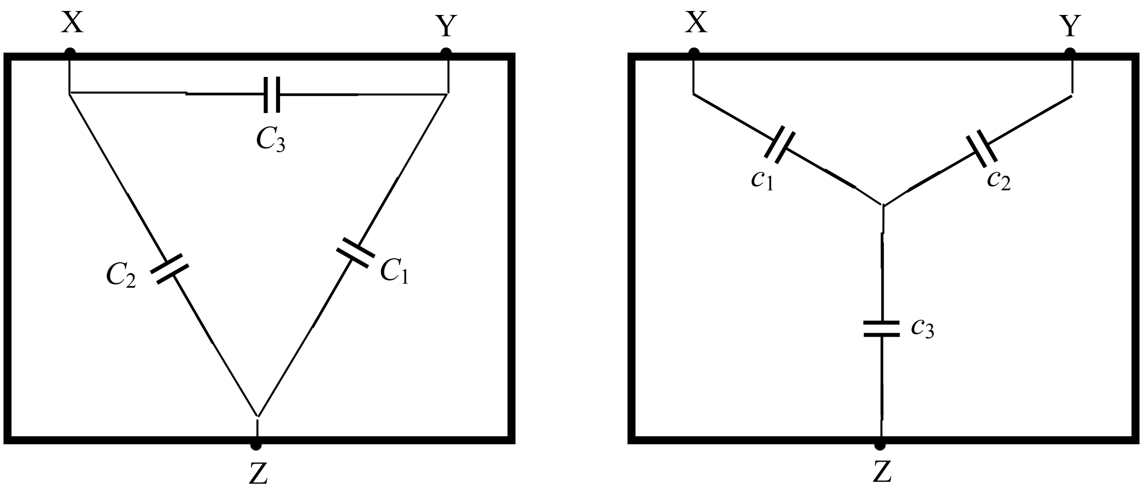


FIGURE V.7

As we did with resistors in Section 4.12, we can make a delta-star transform with capacitors. I leave it to the reader to show that the capacitance between any two terminals in the left hand box is the same as the capacitance between the corresponding two terminals in the right hand box provided that

$$c_1 = \frac{C_2C_3 + C_3C_1 + C_1C_2}{C_1}, \qquad 5.7.1$$

$$c_2 = \frac{C_2C_3 + C_3C_1 + C_1C_2}{C_2} \qquad 5.7.2$$

and

$$c_3 = \frac{C_2C_3 + C_3C_1 + C_1C_2}{C_3}. \qquad 5.7.3$$

The converse relations are

$$C_1 = \frac{c_2c_3}{c_1 + c_2 + c_3}, \qquad 5.7.4$$

$$C_2 = \frac{c_3c_1}{c_1 + c_2 + c_3} \qquad 5.7.5$$

and

$$C_3 = \frac{c_1c_2}{c_1 + c_2 + c_3}. \qquad 5.7.6$$

For example, just for fun, what is the capacitance between points A and B in figure V.8, in which I have marked the individual capacitances in microfarads?

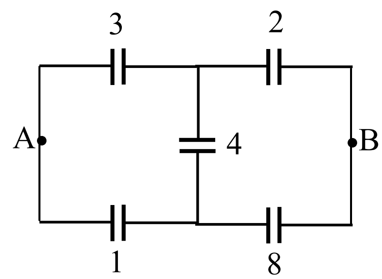


FIGURE V.8

The first three capacitors are connected in delta. Replace them by their equivalent star configuration. After that it should be straightforward. I make the answer 0.402 μF.

5.8 *Kirchhoff's Rules*

We can even adapt Kirchhoff's rules to deal with capacitors. Thus, connect a 24 V battery across the circuit of figure V.8 – see figure V.9

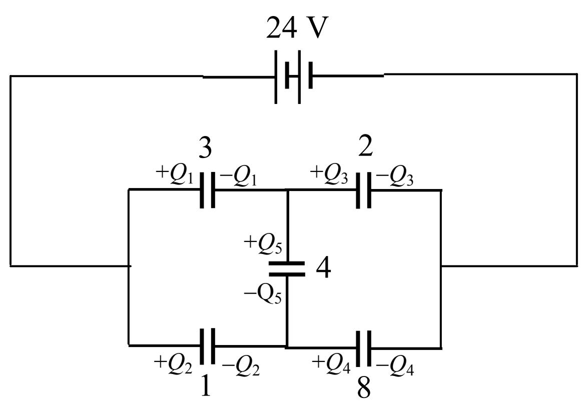


FIGURE V.9

Calculate the charge held in each capacitor. We can proceed in a manner very similar to how we did it in Chapter 4, applying the capacitance equivalent of Kirchhoff's second rule to three closed circuits, and then making up the five necessary equations by applying "Kirchhoff's first rule" to two points. Thus:

$$24 - \frac{Q_2}{3} - \frac{Q_3}{2} = 0\,, \qquad 5.8.1$$

$$24 - Q_2 - \frac{Q_4}{8} = 0\,, \qquad 5.8.2$$

$$\frac{Q_1}{3} - Q_2 + \frac{Q_5}{4} = 0, \qquad 5.8.3$$

$$Q_1 = Q_3 + Q_5 , \qquad 5.8.4$$

and

$$Q_4 = Q_2 + Q_5 . \qquad 5.8.5$$

I make the solutions

$$Q_1 = +41.35\,\mu\text{C},\ Q_2 = +19.01\,\mu\text{C},\ Q_3 = +20.44\,\mu\text{C},\ Q_4 = +39.92\,\mu\text{C},\ Q_5 = +20.91\,\mu\text{C}.$$

5.9 *Problem for a Rainy Day*

Another problem to while away a rainy Sunday afternoon would be to replace each of the resistors in the cube of subsection 4.14.1 with capacitors each of capacitance c. What is the total capacitance across opposite corners of the cube? I would start by supposing that the cube holds a net charge of $6Q$, and I would then ask myself what is the charge held in each of the individual capacitors. And I would then follow the potential drop from one corner of the cube to the opposite corner. I make the answer for the effective capacitance of the entire cube $1.2c$.

5.10 *Energy Stored in a Capacitor*

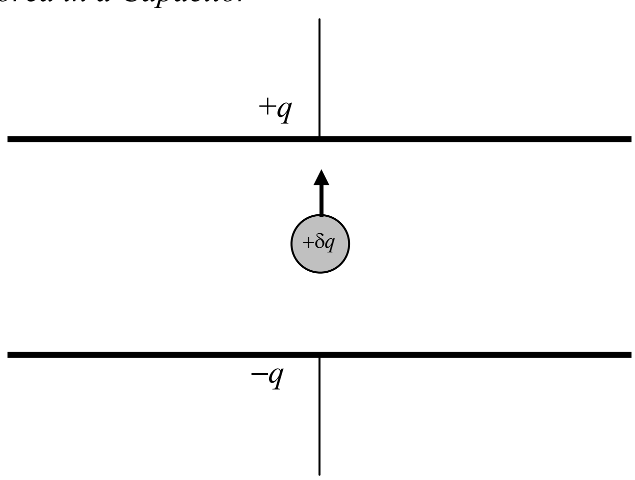


FIGURE V.10

Let us imagine (figure V.10) that we have a capacitor of capacitance C which, at some time, has a charge of $+q$ on one plate and a charge of $-q$ on the other plate. The potential difference across the plates is then q/C. Let us now take a charge of $+\delta q$ from the bottom plate (the negative one) and move it up to the top plate. We evidently have to do work to do this, in the amount of $\frac{q}{C}\delta q$. The total work required, then, starting with the plates completely uncharged until we have transferred a charge Q from one plate to the other is $\frac{1}{C}\int_0^Q q\,dq = Q^2/(2C)$. This is, then, the energy E stored in the capacitor, and, by application of $Q = CV$ it can also be written $\mathsf{E} = \frac{1}{2}QV$, or, more usually,

$$\mathsf{E} = \tfrac{1}{2}CV^2. \qquad 5.10.1$$

Verify that this has the correct dimensions for energy. Also, think about how many expressions for energy you know that are of the form $\frac{1}{2}ab^2$. There are more to come.

The symbol E is becoming rather over-worked. At present I am using the following:

E = magnitude of the electric field
$\mathbf{E}$ = electric field as a vector
E = electromotive force
E = energy

Sorry about that!

5.11 *Energy Stored in an Electric Field*

Recall that we are assuming that the separation between the plates is small compared with their linear dimensions and that therefore the electric field is uniform between the plates.

The capacitance is $C = \varepsilon A/d$, and the potential differnece between the plates is Ed, where E is the electric field and d is the distance between the plates. Thus the energy stored in the capacitor is $\frac{1}{2}\varepsilon E^2 Ad$. The volume of the dielectric (insulating) material between the plates is Ad, and therefore we find the following expression for the *energy stored per unit volume in a dielectric material in which there is an electric field*:

$$\tfrac{1}{2}\varepsilon E^2.$$

Verify that this has the correct dimensions for energy per unit volume.

If the space between the plates is a vacuum, we have the following expression for the enrgy stored ber unit volume in the electric field

$$\tfrac{1}{2}\varepsilon_0 E^2$$

- even though there is absolutely nothing other than energy in the space. Think about that!

I mentioned in Section 1.7 that in an *anisotropic medium* **D** and **E** are not parallel, the permittivity then being a tensor quantity. In that case the correct expression for the energy per unit volume in an electric field is $\frac{1}{2}\mathbf{D}\bullet\mathbf{E}$.

5.12 *Force Between the Plates of a Plane Parallel Plate Capacitor*

We imagine a capacitor with a charge $+Q$ on one plate and $-Q$ on the other, and initially the plates are almost, but not quite, touching. There is a force F between the plates. Now we gradually pull the plates apart (but the separation remains small enough that it is still small compared with the linear dimensions of the plates and we can maintain our approximation of a uniform field between the plates, and so the force remains F as we separate them). The work done in separating the plates from near zero to d is Fd, and this must then equal the energy stored in the capacitor, $\frac{1}{2}QV$. The electric field between the plates is $E = V/d$, so we find for the force between the plates

$$F = \tfrac{1}{2}QE. \qquad 5.12.1$$

We can now do an interesting imaginary experiment, just to see that we understand the various concepts. Let us imagine that we have a capacitor in which the plates are horizontal; the lower plate is fixed, while the upper plate is suspended above it from a spring of force constant k. We connect a battery across the plates, so the plates will attract each other. The upper plate will move down, but only so far, because the electrical attraction between the plates is countered by the tension in the spring. Calculate the equilibrium separation x between the plates as a function of the applied voltage V. (Horrid word! We don't say "metreage" for length, "kilogrammage" for mass or "secondage" for time – so why do we say "voltage" for potential difference and "acreage" for area? Ugh!) We should be able to use our invention as a voltmeter – it even has an infinite resistance!

Refer to figure V.11.

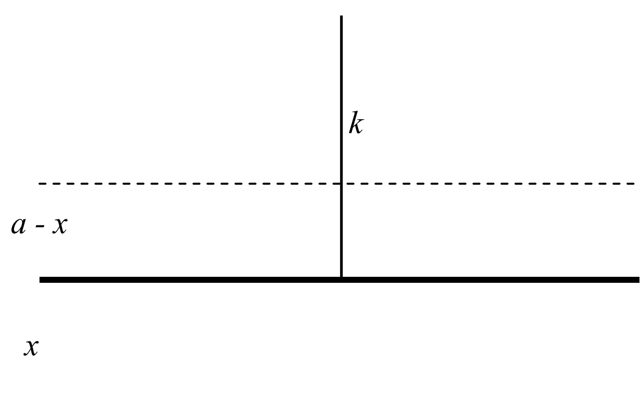


FIGURE V.11

We'll suppose that the separation when the potential difference is zero is a, and the separation when the potential difference is V is x, at which time the spring has been extended by a length $a - x$.

The electrical force between the plates is $\frac{1}{2}QE$. Now $Q = CV = \frac{\varepsilon_0 AV}{x}$ and $E = \frac{V}{x}$, so the force between the plates is $\frac{\varepsilon_0 AV^2}{2x^2}$. Here A is the area of each plate and it is assumed that the experiment is done in air, whose permittivity is very close to ε_0. The tension in the stretched spring is $k(a - x)$, so equating the two forces gives us

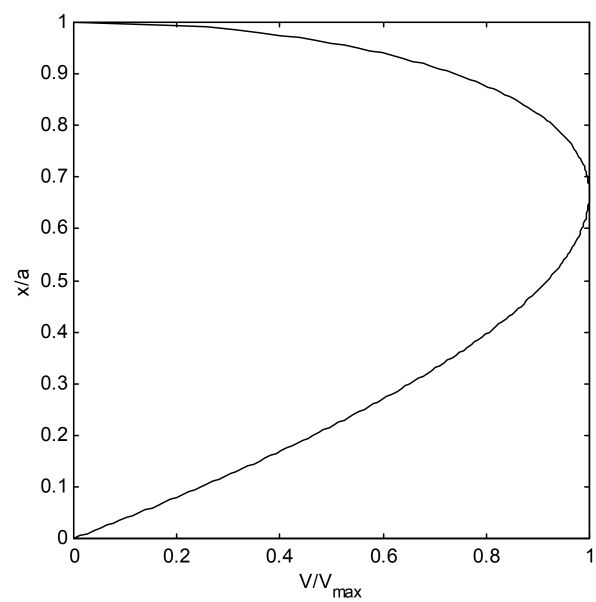
$$V^2 = \frac{2kx^2(a-x)}{\varepsilon_0 A}. \tag{5.12.2}$$

Calculus shows [do it! – just differentiate $x^2(1 - x)$] that V has a maximum value of $V_{\text{max}} = \sqrt{\frac{8ka^3}{27\varepsilon_0 A}}$ for a separation $x = \frac{2}{3}a$. If we express the potential difference in units of V_{max} and the separation in units of a, equation 5.12.2 becomes

$$V^2 = \frac{27x^2(1-x)}{4}. \tag{5.12.3}$$

In figure V.12 I have plotted the separation as a function of the potential difference.

FIGURE V.12



As expected, the potential difference is zero when the separation is 0 or 1 (and therefore you would expect it to go through a maximum for some intermediate separation).

We see that for $V < V_{\max}$ there are *two* equilibrium positions. For example, if $V = 0.8$, show that $x = 0.396\ 305$ or $0.876\ 617$. The question also arises – what happens if you apply across the plates a potential difference that is *greater than* $V_{\max}$?

Further insight can be obtained from energy considerations. The potential energy of the system is the work done in moving the upper plate from $x = a$ to $x = x$ while the potential difference is V:

$$E = \frac{\varepsilon_0 A V^2}{2a} - \frac{\varepsilon_0 A V^2}{2x} + \tfrac{1}{2}k(a-x)^2. \qquad 5.12.4$$

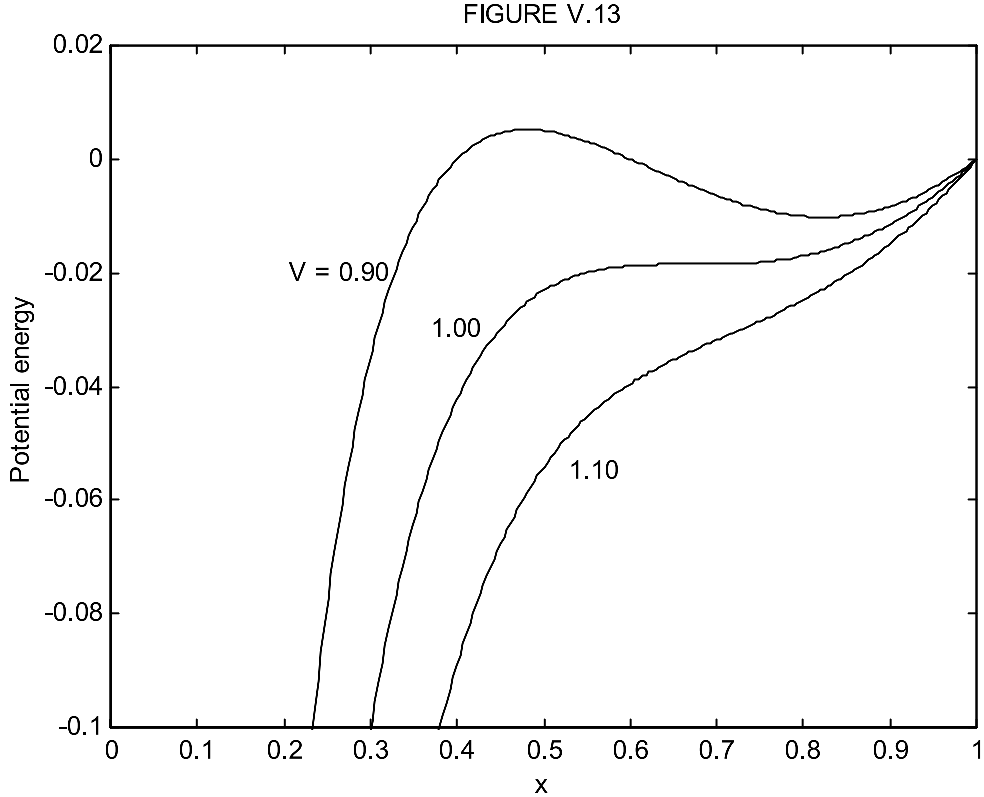
You may need to refer to Section 5.15 to be sure that we have got this right.

If we express V in units of $V_{\max}$, x in units of a and E in units of ka^2, this becomes

$$E = \tfrac{4}{27}V^2(1 - 1/x) + \tfrac{1}{2}(1 - x)^2. \qquad 5.12.5$$

In figure V.13 I have plotted the energy versus separation for three values of potential difference, 90% of $V_{\max}$, $V_{\max}$ and 110% of $V_{\max}$.

FIGURE V.13



We see that for $V < V_{\max}$, there are two equilibrium positions, of which the *lower* one (smaller x) is *unstable*, and we see exactly what will happen if the upper plate is displaced slightly upwards (larger x) from the unstable equilibrium position or if it is displaced slightly downwards (smaller x). The *upper* equilibrium position is stable.

If $V > V_{\max}$, there is no equilibrium position, and x goes down to zero – i.e. the plates clamp together.

5.13 *Sharing a Charge Between Two Capacitors*

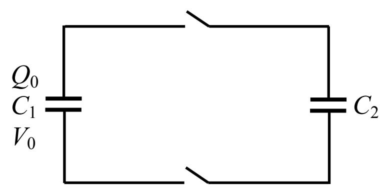


FIGURE V.14

We have two capacitors. C_2 is initially uncharged. Initially, C_1 bears a charge Q_0 and the potential difference across its plates is V_0, such that

$$Q_0 = C_1 V_0 , \qquad 5.13.1$$

and the energy of the system is

$$E_0 = \tfrac{1}{2} C_1 V_0^2 . \qquad 5.13.2$$

We now close the switches, so that the charge is shared between the two capacitors:

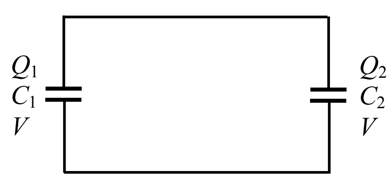


FIGURE V.15

The capacitors C_1 and C_2 now bear charges Q_1 and Q_2 such that $Q_0 = Q_1 + Q_2$ and

$$Q_1 = \frac{C_1}{C_1 + C_2} Q_0 \qquad \text{and} \qquad Q_2 = \frac{C_2}{C_1 + C_2} Q_0 . \qquad 5.13.3\text{a,b}$$

The potential difference across the plates of either capacitor is, of course, the same, so we can call it V without a subscript, and it is easily seen, by applying $Q = CV$ to either capacitor, that

$$V = \frac{C_1}{C_1 + C_2} V_0 . \qquad 5.13.4$$

We can now apply $E = \frac{1}{2}CV^2$ to each capacitor in turn to find the energy stored in each. We find for the energies stored in the two capacitors:

$$E_1 = \frac{C_1^3 V_0^2}{2(C_1 + C_2)^2} \quad \text{and} \quad E_2 = \frac{C_2 C_1^2 V_0^2}{2(C_1 + C_2)^2} . \qquad 5.13.5\text{a,b}$$

The total energy stored in the two capacitors is the sum of these, which is

$$E = \frac{C_1^2 V_0^2}{2(C_1 + C_2)} , \qquad 5.13.6$$

which can also be written

$$E = \frac{C_1}{C_1 + C_2} E_0 . \qquad 5.13.7$$

Surprise, surprise! The energy stored in the two capacitors is less than the energy that was originally stored in C_1. What has happened to the lost energy?

A perfectly reasonable and not incorrect answer is that it has been dissipated as heat in the connecting wires as current flowed from one capacitor to the other. However, it has been found in low temperature physics that if you immerse certain metals in liquid helium they lose *all* electrical resistance and they become *superconductive.* So, let us connect the capacitors with superconducting wires. Then there is no dissipation of energy as heat in the wires – so the question remains: where has the missing energy gone?

Well, perhaps the dielectric medium in the capacitors is heated? Again this seems like a perfectly reasonable and probably not entirely incorrect answer. However, my capacitors have a *vacuum* between the plates, and are connected by superconducting wires, so that no heat is generated either in the dielectric or in the wires. Where has that energy gone?

This will have to remain a mystery for the time being, and a topic for lunchtime conversation. In a later chapter I shall suggest another explanation.

5.14 *Mixed Dielectrics*

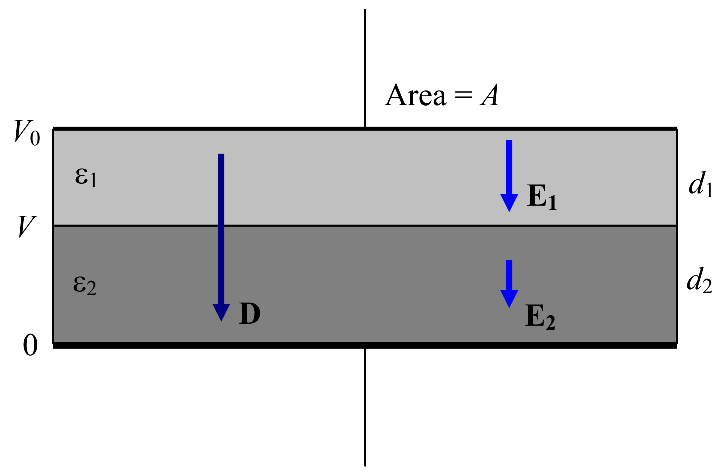
This section addresses the question: If there are two or more dielectric media between the plates of a capacitor, with different permittivities, are the electric fields in the two media different, or are they the same? The answer depends on

1. Whether by "electric field" you mean E or D;

2. The disposition of the media between the plates – i.e. whether the two dielectrics are in series or in parallel.

Let us first suppose that two media are in series (figure V.16).

FIGURE V.16



Our capacitor has two dielectrics in series, the first one of thickness d_1 and permittivity ε_1 and the second one of thickness d_2 and permittivity ε_2. As always, the thicknesses of the dielectrics are supposed to be small so that the fields within them are uniform. This is effectively two capacitors in series, of capacitances $\varepsilon_1 A/d_1$ and $\varepsilon_2 A/d_2$. The total capacitance is therefore

$$C = \frac{\varepsilon_1 \varepsilon_2 A}{\varepsilon_2 d_1 + \varepsilon_1 d_2}. \qquad 5.14.1$$

Let us imagine that the potential difference across the plates is V_0. Specifically, we'll suppose the potential of the lower plate is zero and the potential of the upper plate is V_0. The charge Q held by the capacitor (positive on one plate, negative on the other) is just given by $Q = CV_0$, and hence the surface charge density σ is CV_0/A. Gauss's law is that the total D-flux arising from a charge is equal to the charge, so that in this geometry $D = \sigma$, and this is not altered by the nature of the dielectric materials between the plates. Thus, in this capacitor, $D = CV_0/A = Q/A$ in both media. Thus D is continuous across the boundary.

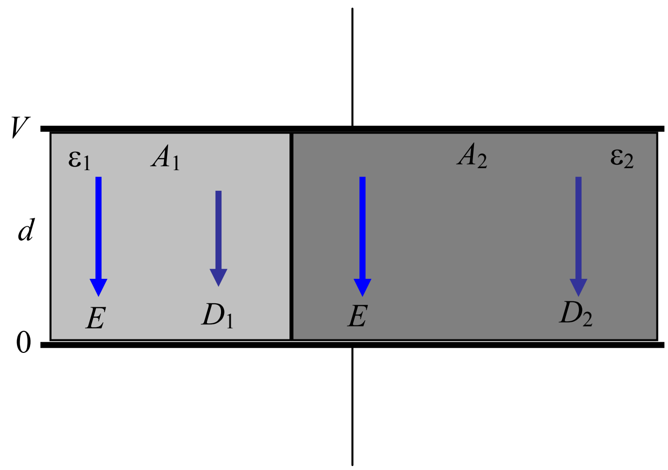
Then by application of $D = \varepsilon E$ to each of the media, we find that the E-fields in the two media are $E_1 = Q/(\varepsilon_1 A)$ and $E_2 = Q/(\varepsilon_2 A)$, the E-field (and hence the potential gradient) being larger in the medium with the smaller permittivity.

The potential V at the media boundary is given by $V/d_2 = E_2$. Combining this with our expression for E_2, and $Q = CV$ and equation 5.14.1, we find for the boundary potential:

$$V = \frac{\varepsilon_1 d_2}{\varepsilon_2 d_1 + \varepsilon_1 d_2} V_0 . \qquad 5.14.2$$

Let us now suppose that two media are in parallel (figure V.17).

FIGURE V.17



This time, we have two dielectrics, each of thickness d, but one has area A_1 and permittivity ε_1 while the other has area A_2 and permittivity ε_2. This is just two capacitors in parallel, and the total capacitance is

$$C = \frac{\varepsilon_1 A_1}{d} + \frac{\varepsilon_2 A_2}{d} . \qquad 5.14.3$$

The E-field is just the potential gradient, and this is independent of any medium between the plates, so that $E = V/d$. in each of the two dielectrics. After that, we have simply that $D_1 = \varepsilon_1 E$ and $D_2 = \varepsilon_2 E$. The charge density on the plates is given by Gauss's law as $\sigma = D$, so that, if $\varepsilon_1 < \varepsilon_2$, the charge density on the left hand portion of each plate is less than on the right hand portion – although the *potential* is the same throughout each plate. (The surface of a metal is always an equipotential surface.) The two different charge densities on each plate is a result of the different *polarizations* of the two dielectrics – something that will be more readily understood a little later in this chapter when we deal with media polarization.

We have established that:

1. The component of **D** perpendicular to a boundary is continuous;

2. The component of **E** parallel to a boundary is continuous.

In figure V.18 we are looking at the *D*-field and at the *E*-field as it crosses a boundary in which $\varepsilon_1 < \varepsilon_2$. Note that D_y and E_x are the same on either side of the boundary. This results in:

$$\frac{\tan\theta_1}{\tan\theta_2} = \frac{\varepsilon_1}{\varepsilon_2}. \qquad 5.14.4$$

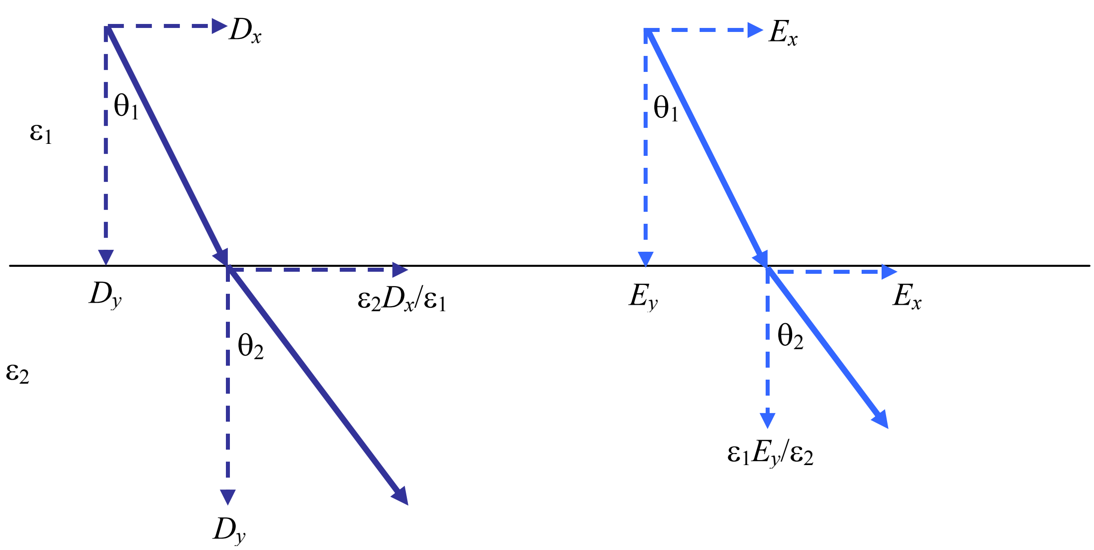


FIGURE V.18

5.15 *Changing the Distance Between the Plates of a Capacitor*

If you gradually increase the distance between the plates of a capacitor (although always keeping it sufficiently small so that the field is uniform) does the intensity of the field change or does it stay the same? If the former, does it increase or decrease?

The answer to these questions depends

1. on whether, by the field, you are referring to the *E*-field or the *D*-field;

2. on whether the plates are *isolated* or if they are *connected to the poles of a battery*.

We shall start by supposing that the plates are *isolated*.

In this case the charge on the plates is constant, and so is the charge density. Gauss's law requires that $D = \sigma$, so that D remains constant. And, since the permittivity hasn't changed, E also remains constant.

The potential difference across the plates is Ed, so, as you increase the plate separation, so the potential difference across the plates in increased. The capacitance decreases from $\varepsilon A/d_1$ to $\varepsilon A/d_2$ and the energy stored in the capacitor increases from $\frac{Ad_1\sigma^2}{2\varepsilon}$ to $\frac{Ad_1\sigma^2}{2\varepsilon}$. This energy derives from the work done in separating the plates.

Now let's suppose that the plates are *connected to a battery* of EMF V, with air or a vacuum between the plates. At first, the separation is d_1. The magnitudes of E and D are, respectively, V/d_1 and $\varepsilon_0 V/d_1$. When we have increased the separation to d_2, the potential difference across the plates has not changed; it is still the EMF V of the battery. The electric field, however, is now only $E = V/d_2$ and $D = \varepsilon_0 V/d_2$. But Gauss's law still dictates that $D = \sigma$, and therefore the charge density, and the total charge on the plates, is less than it was before. It has gone into the battery. In other words, in doing work by separating the plates we have recharged the battery. The energy stored in the capacitor was originally $\frac{\varepsilon_0 AV^2}{2d_1}$; it is now only $\frac{\varepsilon_0 AV^2}{2d_2}$. Thus the energy held in the capacitor has been reduced by $\frac{1}{2}\varepsilon_0 AV^2\left(\frac{1}{d_1} - \frac{1}{d_2}\right)$.

The charge originally held by the capacitor was $\frac{\varepsilon_0 AV}{d_1}$. After the plate separation has been increased to d_2 the charge held is $\frac{\varepsilon_0 AV}{d_2}$. The difference, $\varepsilon_0 AV\left(\frac{1}{d_1} - \frac{1}{d_2}\right)$, is the charge that has gone into the battery. The energy, or work, required to force this amount of charge into the battery against its EMF V is $\varepsilon_0 AV^2\left(\frac{1}{d_1} - \frac{1}{d_2}\right)$. Half of this came from the loss in energy held by the capacitor (see above). The other half presumably came from the mechanical work you did in separating the plates. Let's see if we can verify this.

When the plate separation is x, the force between the plates is $\frac{1}{2}QE$, which is $\frac{1}{2}\frac{\varepsilon_0 AV}{x}\cdot\frac{V}{x}$ or $\frac{\varepsilon_0 AV^2}{2x^2}$. The work required to increase x from d_1 to d_2 is $\frac{\varepsilon_0 AV^2}{2}\int_{d_1}^{d_2}\frac{dx}{x^2}$, which is indeed $\frac{1}{2}\varepsilon_0 AV^2\left(\frac{1}{d_1} - \frac{1}{d_2}\right)$. Thus this amount of mechanical work, plus an equal amount of energy from the capacitor, has gone into recharging the battery. Expressed otherwise, the work done in separating the plates equals the work required to charge the battery minus the decrease in energy stored by the capacitor.

Perhaps we have invented a battery charger (figure V.19)!

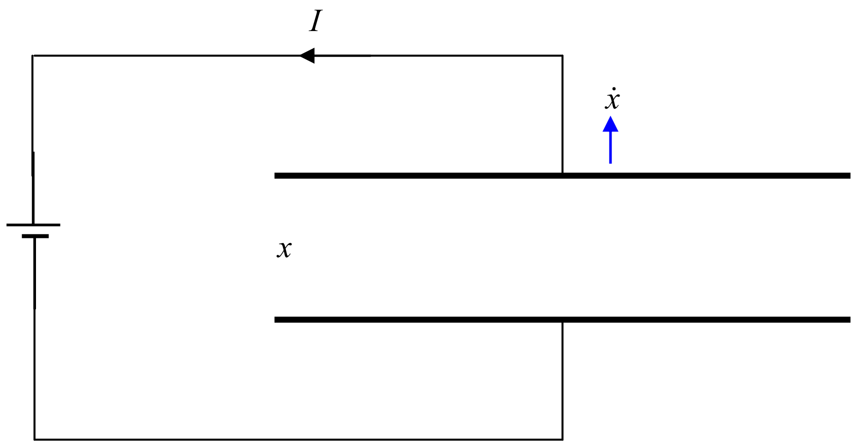


FIGURE V.19

When the plate separation is x, the charge stored in the capacitor is $Q = \frac{\epsilon_0 AV}{x}$. If x is increased at a rate $\dot{x}$, Q will increase at a rate $\dot{Q} = -\frac{\epsilon_0 AV\dot{x}}{x^2}$. That is, the capacitor will discharge (because $\dot{Q}$ is negative), and a current $I = \frac{\epsilon_0 AV\dot{x}}{x^2}$ will flow counterclockwise in the circuit. (Verify that this expression is dimensionally correct for current.)

5.16 *Inserting a Dielectric into a Capacitor*

Suppose you start with two plates separated by a vacuum or by air, with a potential difference across the plates, and you then insert a dielectric material of permittivity ϵ_0 between the plates. Does the intensity of the field change or does it stay the same? If the former, does it increase or decrease?

The answer to these questions depends

1. on whether, by the field, you are referring to the E-field or the D-field;

2. on whether the plates are *isolated* or if they are *connected to the poles of a battery*.

We shall start by supposing that the plates are *isolated*. See figure V.20.

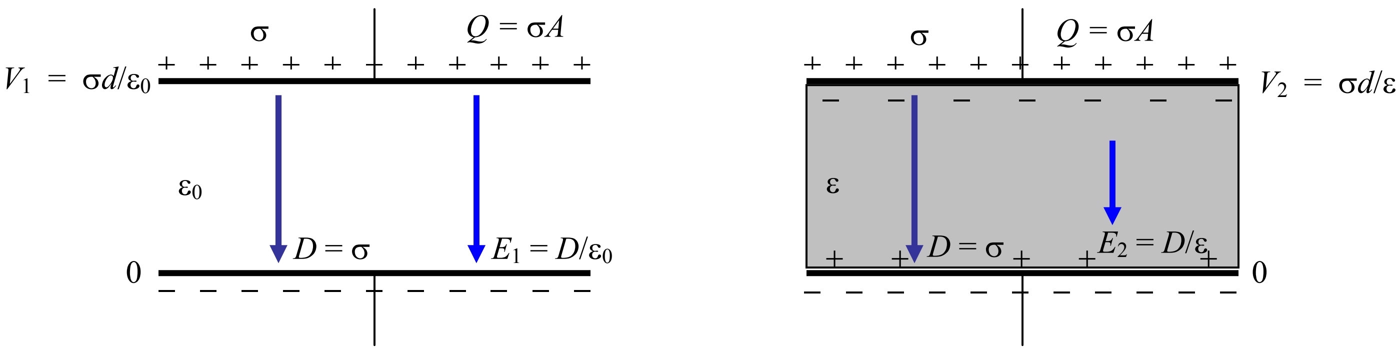


FIGURE V.20

Let Q be the charge on the plates, and σ the surface charge density. These are unaltered by the introduction of the dielectric. Gauss's law provides that $D = \sigma$, so this, too, is unaltered by the introduction of the dielectric. The electric field was, initially, $E_1 = D/\epsilon_0$. After introduction of the dielectric, it is a little less, namely $E_1 = D/\epsilon$.

Let us take the potential of the lower plate to be zero. Before introduction of the dielectric, the potential of the upper plate was $V_1 = \sigma d/\epsilon_0$. After introduction of the dielectric, it is a little less, namely $V_1 = \sigma d/\epsilon$.

Why is the electric field E less after introduction of the dielectric material? It is because the dielectric material becomes *polarized.* We saw in Section 3.6 how matter may become polarized. Either molecules with pre-existing dipole moments align themselves with the imposed electric field, or, if they have no permanent dipole moment or if they cannot rotate, a dipole moment can be induced in the individual molecules. In any case, the effect of the alignment of all these molecular dipoles is that there is a slight surplus of positive charge on the surface of the dielectric material next to the negative plate, and a slight surplus of negative charge on the surface of the dielectric material next to the positive plate. This produces an electric field opposite to the direction of the imposed field, and thus the total electric field is somewhat reduced.

Before introduction of the dielectric material, the energy stored in the capacitor was $\frac{1}{2}QV_1$. After introduction of the material, it is $\frac{1}{2}QV_2$, which is a little bit less. Thus it will require work to remove the material from between the plates. The empty capacitor will tend to suck the material in, just as the charged rod in Chapter 1 attracted an uncharged pith ball.

Now let us suppose that the plates are *connected to a battery*. (Figure V.21)

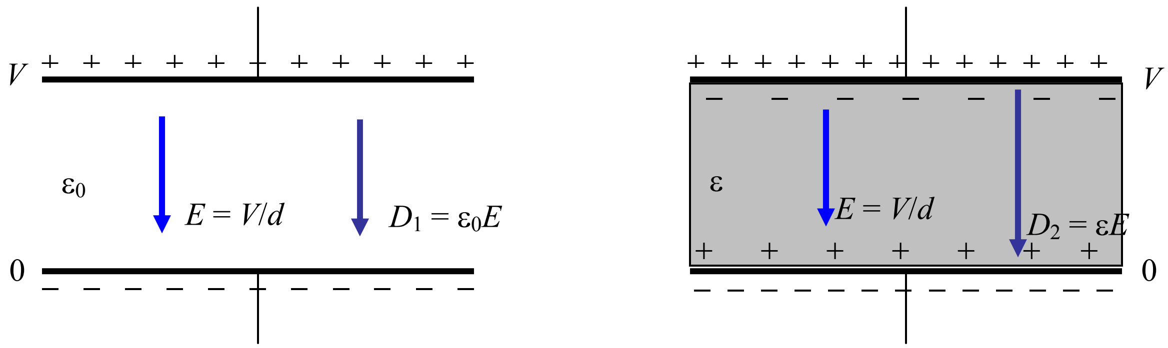


FIGURE V.21

This time the potential difference remains constant, and therefore so does the E-field, which is just V/d. But the D-field increases from $\epsilon_0 E$ to ϵE, and so, therefore, does the surface charge density on the plates. This extra charge comes from the battery.

The capacitance increases from $\frac{\epsilon_0 A}{d}$ to $\frac{\epsilon A}{d}$ and the charge stored on the plates increases from $Q_1 = \frac{\epsilon_0 AV}{d}$ to $Q_2 = \frac{\epsilon AV}{d}$. The energy stored in the capacitor increases from $\frac{1}{2}Q_1 V$ to $\frac{1}{2}Q_2 V$.

The energy supplied by the battery = the energy dumped into the capacitor + the energy required to suck the dielectric material into the capacitor:

$$(Q_2 - Q_1)V = \tfrac{1}{2}(Q_2 - Q_1)V + \tfrac{1}{2}(Q_2 - Q_1)V.$$

You would have to do work to remove the material from the capacitor; half of the work you do would be the mechanical work performed in pulling the material out; the other half would be used in charging the battery.

In Section 5.15 I invented one type of battery charger. I am now going to make my fortune by inventing another type of battery charger.

Example 1.

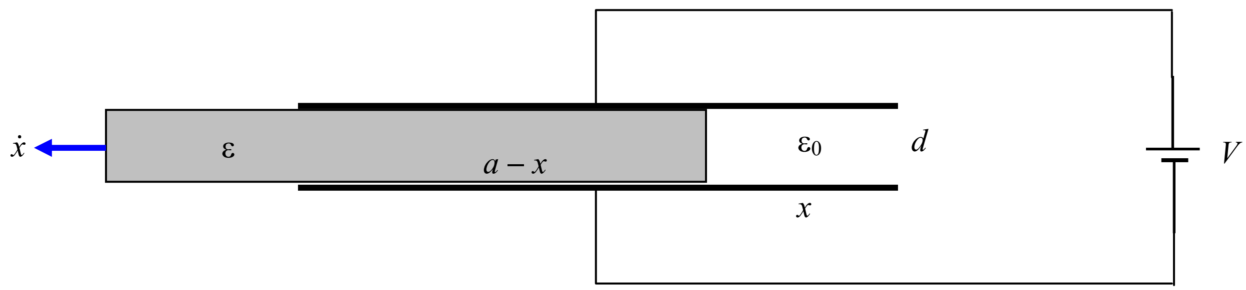


FIGURE V.22

A capacitor is formed of two square plates, each of dimensions $a \times a$, separation d, connected to a battery. There is a dielectric medium of permittivity ε between the plates. I pull the dielectric medium out at speed $\dot{x}$. Calculate the current in the circuit as the battery is recharged.

Solution.

When I have moved a distance x, the capacitance is

$$\frac{\varepsilon a(a-x)}{d} + \frac{\varepsilon_0 ax}{d} = \frac{\varepsilon a^2 - (\varepsilon - \varepsilon_0)ax}{d}.$$

The charge held by the capacitor is then

$$Q = \left[\frac{\varepsilon a^2 - (\varepsilon - \varepsilon_0)ax}{d}\right]V.$$

If the dielectric is moved out at speed $\dot{x}$, the charge held by the capacitor will increase at a rate

$$\dot{Q} = \frac{-(\varepsilon - \varepsilon_0)a\dot{x}V}{d}.$$

(That's negative, so Q decreases.) A current of this magnitude therefore flows clockwise around the circuit, into the battery. You should verify that the expression has the correct dimensions for current.

Example 2.

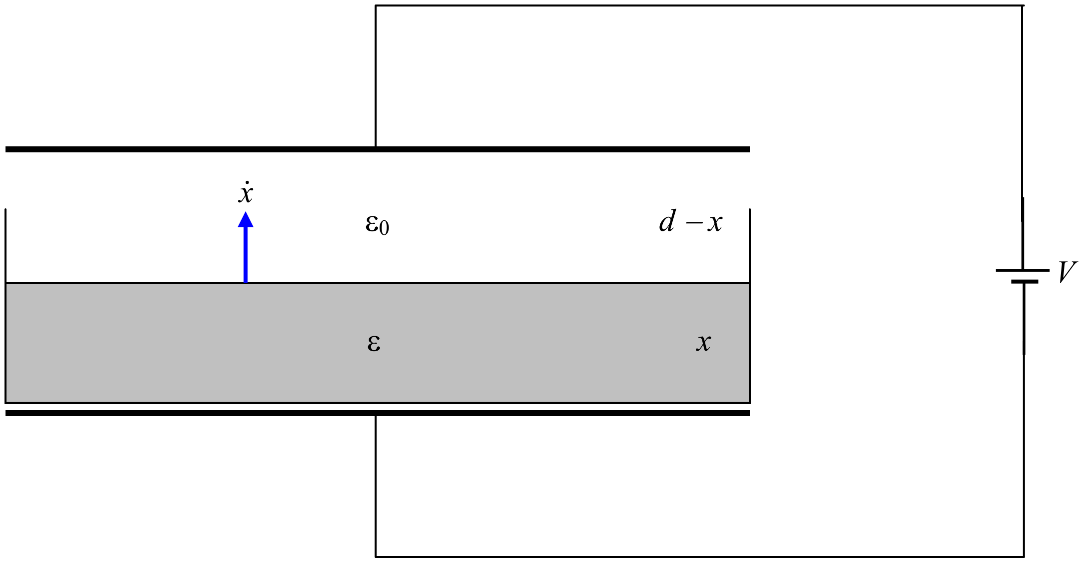


FIGURE V.23

A capacitor consists of two plates, each of area A, separated by a distance x, connected to a battery of EMF V. A cup rests on the lower plate. The cup is gradually filled with a nonconducting liquid of permittivity ε, the surface rising at a speed $\dot{x}$. Calculate the magnitude and direction of the current in the circuit.

It is easy to calculate that, when the liquid has a depth x, the capacitance of the capacitor is

$$C = \frac{\varepsilon\varepsilon_0 A}{\varepsilon d - (\varepsilon - \varepsilon_0)x}$$

and the charge held by the capacitor is then

$$Q = \frac{\varepsilon\varepsilon_0 AV}{\varepsilon d - (\varepsilon - \varepsilon_0)x} \, .$$

If x is increasing at a rate $\dot{x}$, the rate at which Q, the charge on the capacitor, is increasing is

$$\dot{Q} = \frac{\varepsilon\varepsilon_0(\varepsilon - \varepsilon_0)AV\dot{x}}{[\varepsilon d - (\varepsilon - \varepsilon_0)x]^2} \, .$$

A current of this magnitude therefore flows in the circuit counterclockwise, draining the battery. This current increases monotonically from zero to $\frac{\varepsilon(\varepsilon-\varepsilon_0)AV\dot{x}}{\varepsilon_0 d^2}$.

5.17 *Polarization and Susceptibility*

When an insulating material is placed in an electric field, it becomes *polarized*, either by rotation of molecules with pre-existing dipole moments or by induction of dipole moments in the individual molecules. Inside the material, D is then greater than $\varepsilon_0 E$. Indeed,

$$D = \varepsilon_0 E + P. \qquad 5.17.1$$

The excess, P, of D over $\varepsilon_0 E$ is called the *polarization* of the medium. It is dimensionally similar to, and expressed in the same units as, D; that is to say C m^{-2}. Another way of looking at the polarization of a medium is that it is the *dipole moment per unit volume.*

In vector form, the relation is

$$\mathbf{D} = \varepsilon_0 \mathbf{E} + \mathbf{P}. \qquad 5.17.2$$

If the medium is isotropic, all three vectors are parallel.

Some media are more susceptible to becoming polarized in a polarizing field than others, and the ratio of P to $\varepsilon_0 E$ is called the electric *susceptibility* χ_e of the medium:

$$P = \chi_e \varepsilon_0 E. \qquad 5.17.3$$

This implies that P is linearly proportional to E but only if χ_e is independent of E, which is by no means always the case, but is good for small polarizations.

When we combine equations 5.17.1 and 5.17.3 with $D = \varepsilon E$ and with $\varepsilon_r = \varepsilon/\varepsilon_0$, the *relative permittivity* or *dielectric constant*, we obtain

$$\chi_e = \varepsilon_r - 1. \qquad 5.17.4$$

5.18 *Discharging a Capacitor Through a Resistor*

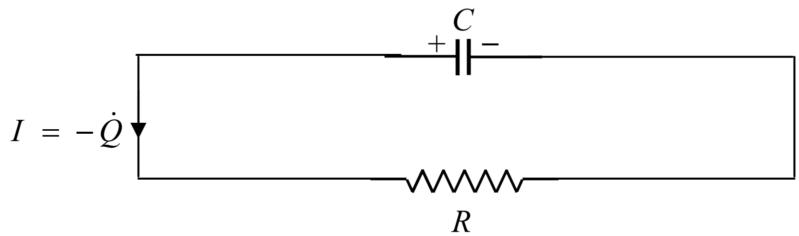


FIGURE V.24

What you have to be sure of in this section and the following section is to get the *signs* right. For example, if the charge held in the capacitor at some time is Q, then the symbol $\dot{Q}$, or dQ/dt, means the rate of increase of Q with respect to time. If the capacitor is discharging, $\dot{Q}$ is negative. Expressed otherwise, the symbol to be used for the rate at which a capacitor is *losing* charge is $-\dot{Q}$.

In figure V.24 a capacitor is discharging through a resistor, and the current as drawn is given by $I = -\dot{Q}$. The potential difference across the plates of the capacitor is Q/C, and the potential difference across the resistor is $IR = -\dot{Q}R$.

Thus: $$\frac{Q}{C} - IR = \frac{Q}{C} + \dot{Q}R = 0. \qquad 5.18.1$$

On separating the variables (Q and t) and integrating we obtain

$$\int_{Q_0}^{Q} \frac{dQ}{Q} = -\frac{1}{RC}\int_0^t dt, \qquad 5.18.2$$

where Q_0 is the charge in the capacitor at $t = 0$.

Hence $$Q = Q_0 e^{-t/(RC)}. \qquad 5.18.3$$

Here RC is the *time constant.* (Verify that it has the dimensions of time.) It is the time for the charge to be reduced to $1/e = 36.8\%$ of the initial charge. The half life of the charge is $RC\ln 2 = 0.6931RC$.

5.19 *Charging a Capacitor Through a Resistor*

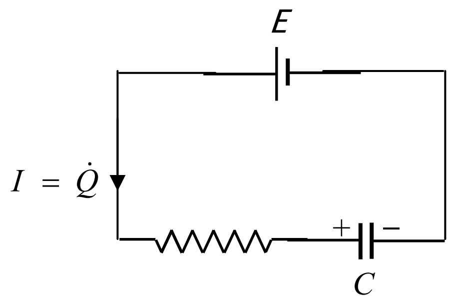


FIGURE V.25

This time, the charge on the capacitor is increasing, so the current, as drawn, is $+\dot{Q}$. Thus

$$E \;-\; \dot{Q}R \;-\; \frac{Q}{C} \;=\; 0. \qquad 5.19.1$$

Whence: $$\int_0^Q \frac{dQ}{EC - Q} \;=\; \frac{1}{RC}\int_0^t dt. \qquad 5.19.2$$

[Note: Don't be tempted to write this as $\int_0^Q \frac{dQ}{Q - EC} \;=\; -\frac{1}{RC}\int_0^t dt$. Remember that, at any finite t, Q is less than its asymptotic value EC, and you want to keep the denominator of the left hand integral positive.]

Upon integrating, we obtain

$$Q \;=\; EC\left(1 - e^{-t/(RC)}\right). \qquad 5.19.3$$

Thus the charge on the capacitor asymptotically approaches its final value EC, reaching 63% $(1 - e^{-1})$ of the final value in time RC and half of the final value in time $RC \ln 2 \;=\; 0.6931\,RC$.

The potential difference across the plates increases at the same rate. Potential difference cannot change instantaneously in any circuit containing capacitance.

Here's a way of making a neon lamp flash periodically.

In figure V.$25\frac{1}{2}$ (sorry about the fraction – I slipped the figure in as an afterthought!), the things that looks something like a happy face on the right is a discharge tube; the dot inside it indicates that it's not a complete vacuum inside, but it has a little bit of gas

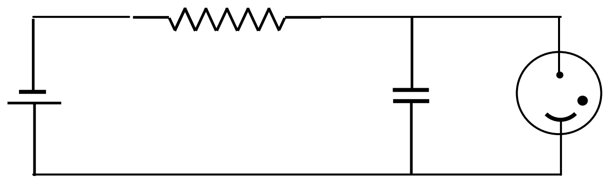


FIGURE V. $25\frac{1}{2}$

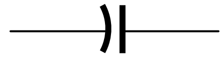
inside. It will discharge when the potential difference across the electrodes is higher than a certain threshold. When an electric field is applied across the tube, electrons and positive ions accelerate, but are soon slowed by collisions. But, if the field is sufficiently high, the electrons and ions will have enough energy on collision to ionize the atoms they collide with, so a cascading discharge will occur. The potential difference rises exponentially on an *RC* time-scale until it reaches the threshold value, and the neon tube suddenly discharges. Then it starts all over again.

5.20 *Real Capacitors*

Real capacitors can vary from huge metal plates suspended in oil to the tiny cylindrical components seen inside a radio. A great deal of information about them is available on the Web and from manufacturers' catalogues, and I only make the briefest remarks here.

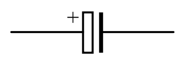
A typical inexpensive capacitor seen inside a radio is nothing much more than two strips of metal foil separated by a strip of plastic or even paper, rolled up into a cylinder much like a Swiss roll. Thus the separation of the "plates" is small, and the area of the plates is as much as can be conveniently rolled into a tiny radio component.

In most applications it doesn't matter which way round the capacitor is connected. However, with some capacitors it is intended that the outermost of the two metal strips be grounded ("earthed" in UK terminology), and the inner one is shielded by the outer one from stray electric fields. In that case the symbol used to represent the capacitor is



The curved line is the outer strip, and is the one that is intended to be grounded. It should be noted, however, that not everyone appears to be aware of this convention or adheres to it, and some people will use this symbol to denote *any* capacitor. Therefore care must be taken in reading the literature to be sure that you know what the writer intended, and, if you are describing a circuit yourself, you must make very clear the intended meaning of your symbols.

There is a type of capacitor known as an *electrolytic capacitor*. The two "plates" are strips of aluminium foil separated by a conducting paste, or electrolyte. One of the foils is covered by an extremely thin layer of aluminium oxide, which has been electrolytically deposited, and it is this layer than forms the dielectric medium, not the paste that separates the two foils. Because of the extreme thinness of the oxide layer, the capacitance is relatively high, although it may not be possible to control the actual thickness with great precision and consequently the actual value of the capacitance may not be known with great precision. It is very important that an electrolytic capacitor be corrected the right way round in a circuit, otherwise electrolysis will start to remove the oxide layer from one foil and deposit it on the other, thus greatly changing the capacitance. Also, when this happens, a current may pass through the electrolyte and heat it up so much that the capacitor may burst open with consequent danger to the eyes. The symbol used to indicate an electrolytic capacitor is:



The side indicated with the plus sign (which is often omitted from the symbol) is to be connected to the positive side of the circuit.

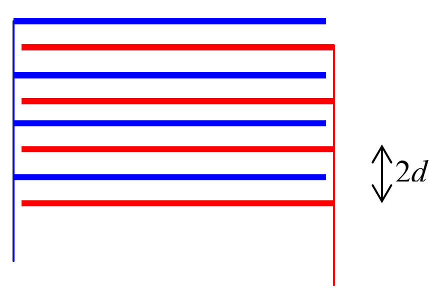
When you tune your radio, you will usually find that, as you turn the knob that changes the wavelength that you want to receive, you are changing the capacitance of a variable air-spaced capacitor just behind the knob. A variable capacitor can be represented by the symbol

Such a capacitor often consists of two sets of interleaved partiallyoverlapping plates, one set of which can be rotated with respect to the other, thus changing the overlap area and hence the capacitance.

Thinking about this suggests to me a couple of small problems for you to amuse yourself with.

Problem 1.

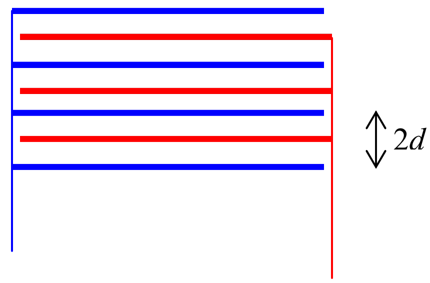
FIGURE V.26



A capacitor (figure V.26) is made from two sets of four plates. The area of each plate is A and the spacing between the plates in each set is $2d$. The two sets of plates are interleaved, so that the distance between the plates of one set and the plates of the other is d. What is the capacitance of the system?

Problem 2

FIGURE V.27



This is just like Problem 1, except that one set has four plates and the other has three. What is the capacitance now?

Answers on the next page.

Solutions. The answer to the first problem is $7\varepsilon_0 A/d$ and the answer to the second probolem is $6\varepsilon_0 A/d$ – but it isn't good enough just to assert that this is the case. We must give some reasons.

Let us suppose that the potential of the left-hand (blue) plates is zero and the potential of the right-hand (blue) plates is V. The electric field in each space is V/d and $D = \varepsilon_0 V/d$. The surface charge density on each plate, by Gauss's theorem, is therefore $2\varepsilon_0 V/d$ except for the two end plates, for which the charge density is just $\varepsilon_0 V/d$. The total charge held in the capacitor of Problem 1 is therefore $\varepsilon_0 AV/d + 3 \times 2\varepsilon_0 AV/d = 7\varepsilon_0 AV/d$, and the capacitance is therefore $7\varepsilon_0 A/d$. For Problem 2, the blue set has two end-plates and two middle-plates, so the charge held is $2 \times \varepsilon_0 AV/d + 2 \times 2\,\varepsilon_0 AV/d = 6\varepsilon_0 AV/d$. The red set has three middle- plates and no end-plates, so the charge held is $3 \times 2\varepsilon_0 AV/d = 6\varepsilon_0 AV/d$. The capacitance is therefore $6\varepsilon_0 A/d$.

5.21 *More on* **E**, **D, P**, *etc.*

I'll review a few things that we have already covered before going on.

The electric field **E** between the plates of a plane parallel capacitor is equal to the potential gradient – i.e. the potential difference between the plates divided by the distance between them.

The electric field **D** between the plates of a plane parallel capacitor is equal to the surface charge density on the plates.

Suppose at first there is nothing between the plates. If you now thrust an isotropic* dielectric material of relative permittivity ε_r between the plates, what happens? Answer: If the plates are *isolated* **D** remains the same while **E** (and hence the potential difference across the plates) is reduced by a factor ε_r. If on the other hand the plates are connected to a battery, the potential difference and hence **E** remains the same while **D** (and hence the charge density on the plates) increases by a factor ε_r.

*You will have noticed the word *isotropic* here. Refer to Section 1.7 for a brief mention of an anisotropic medium, and the concept of permittivity as a tensor quantity. I'm not concerned with this aspect here.

In either case, the block of dielectric material becomes *polarized*. It develops a charge density on the surfaces that adjoin the plates. The block of material develops a *dipole moment*, and the dipole moment divided by the volume of the material – i.e. the dipole moment per unit volume – is the *polarization* **P** of the material. **P** is also equal to

$\mathbf{D} - \varepsilon_0\mathbf{E}$ and, of course, to $\varepsilon\mathbf{E} - \varepsilon_0\mathbf{E}$. The ratio of the resulting polarization **P** to the polarizing field $\varepsilon_0\mathbf{E}$ is called the electric *susceptibility* χ of the medium. It will be worth spending a few moments convincing yourself from these definitions and concepts that $\varepsilon = \varepsilon_0(1 + \chi)$ and $\chi = \varepsilon_r - 1$, where ε_r is the dimensionless *relative permittivity* (or *dielectric constant*) $\varepsilon / \varepsilon_0$.

What is happening physically inside the medium when it becomes polarized? One possibility is that the individual molecules, if they are asymmetric molecules, may already possess a *permanent dipole moment*. The molecule carbon dioxide, which, in its ground state, is linear and symmetric, O=C=O, does not have a permanent dipole moment. Symmetric molecules such as CH_4, and single atoms such as He, do not have a permanent dipole moment. The water molecule has some elements of symmetry, but it is not linear, and it does have a permanent dipole moment, of about 6×10^{-30} C m, directed along the bisector of the HOH angle and away from the O atom. If the molecules have a permanent dipole moment and are free to rotate (as, for example, in a gas) they will tend to rotate in the direction of the applied field. (I'll discuss that phrase "tend to" in a moment.) Thus the material becomes polarized.

A molecule such as CH_4 is symmetric and has no permanent dipole moment, but, if it is placed in an external electric field, the molecule may become distorted from its perfect tetrahedral shape with neat 109º angles, because each pair of CH atoms has a dipole moment. Thus the molecule acquires an *induced dipole moment*, and the material as a whole becomes polarized. The ratio of the induced dipole moment **p** to the polarizing field **E** *polarizability* α of the molecule. Review Section 3.6 for more on this.

How about a single atom, such as Kr? Even that can acquire a dipole moment. Although there are no bonds to bend, under the influence of an electric field a preponderance of electrons will migrate to one side of the atom, and so the atom acquires a dipole moment. The same phenomenon applies, of course, to a molecule such as CH_4 in addition to the bond bending already mentioned.

Let us consider the situation of a dielectric material in which the molecules have a permanent dipole moment and are free (as in a gas, for example) to rotate. We'll suppose that, at least in a weak polarizing field, the permanent dipole moment is significantly larger than any induced dipole moment, so we'll neglect the latter. We have said that, under the influence of a polarizing field, the permanent dipole will *tend to* align themselves with the field. But they also have to contend with the constant jostling and collisions between molecules, which knock their dipole moments haywire, so they can't immediately all align exactly with the field. We might imagine that the material may become fairly strongly polarized if the temperature is fairly low, but only relatively weakly polarized at higher temperatures. Dare we even hope that we might be able to predict the variation of polarization P with temperature T? Let's have a go!

We recall (Section 3.4) that the potential energy U of a dipole, when it makes an angle θ with the electric field, is $U = -pE\cos\theta = -\mathbf{p}\bullet\mathbf{E}$. The energy of a dipole whose direction makes an angle of between θ and $\theta + d\theta$ with the field will be between U and $U + dU$, where $dU = pE\sin\theta d\theta$. What happens next requires familiarity with Boltzmann's equation for distribution of energies in a statistical mechanics. See for example my Stellar Atmospheres notes, Chapter 8, Section 8.4. The fraction of molecules having energies between U and $U + dU$ will be, following Boltzmann's equation,

$$\frac{e^{-U/(kT)}dU}{\int_{-pE}^{+pE} e^{-U/(kT)}dU}, \tag{5.21.1}$$

(Caution: Remember that here I'm using U for potential energy, and E for electric field.)

That is, the fraction of molecules making angles of between θ and $\theta + d\theta$ with the field is

$$\frac{pEe^{pE\cos\theta/(kT)}\sin\theta d\theta}{\int_0^\pi pEe^{pE\cos\theta/(kT)}\sin\theta d\theta} = \frac{e^{pE\cos\theta/(kT)}\sin\theta d\theta}{\int_0^\pi e^{pE\cos\theta/(kT)}\sin\theta d\theta}. \tag{5.21.2}$$

The component in the direction of **E** of the dipole moment of this fraction of the molecules is

$$\frac{pe^{pE\cos\theta/(kT)}\sin\theta\cos\theta d\theta}{\int_0^\pi e^{pE\cos\theta/(kT)}\sin\theta d\theta}, \tag{5.21.3}$$

so the component in the direction of **E** of the dipole moment all of the molecules is

$$\frac{p\int_0^\pi e^{-pE\cos\theta/(kT)}\sin\theta\cos\theta d\theta}{\int_0^\pi e^{-pE\cos\theta/(kT)}\sin\theta d\theta}, \tag{5.21.4}$$

and this expression represents the induced dipole moment in the direction of the field of the entire sample, which I'll call p_{s}. The *polarization* of the sample would be this divided by its volume.

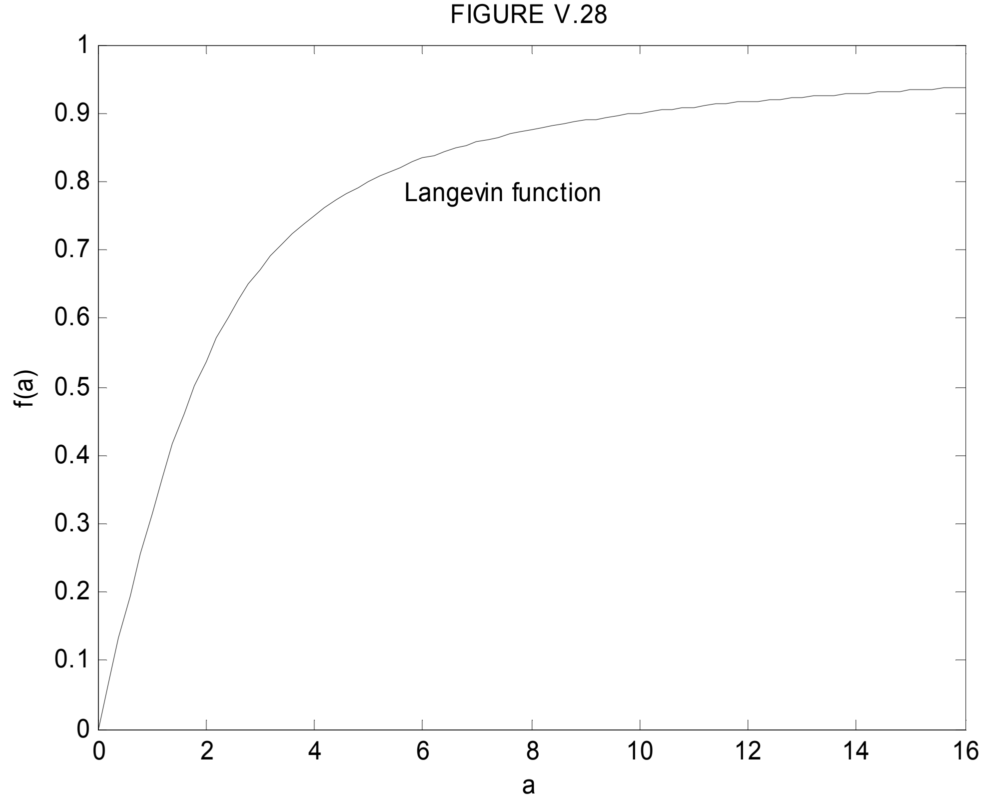
Let
$$x = \frac{pE}{kT}\cos\theta = a\cos\theta. \tag{5.21.5}$$

Then the expression for the dipole moment of the entire sample becomes (some care is needed):

$$p_s = p \times \frac{\int_{-a}^{+a} xe^x dx}{a\int_{-a}^{a} e^x dx} = p \times \left(\frac{e^a + e^{-a}}{e^a - e^{-a}} - \frac{1}{a} \right). \qquad 5.22.6$$

The expression in parentheses is called the *Langevin function*, and it was first derived in connection with the theory of paramagnetism. If your calculator or computer supports the hyperbolic coth function, it is most easily calculated as coth $a - 1/a$. If it does not support coth, calculate it as $\frac{1+b}{1-b} - \frac{1}{a}$, where $b = e^{-2a}$. In any case it is a rather interesting, even challenging, function. Let us call the expression in parentheses $f(a)$. What would the function look like it you were to plot $f(a)$ versus a? The derivative with respect to a is $\frac{1}{a^2} - \frac{4b}{(1-b)^2}$. It is easy to see that, as $a \to \infty$, the function approaches 1 and its derivative, or slope, approaches zero. But what are the function and its derivative (slope) at $a = 0$? You may find that a bit of a challenge. The answer is that, as $a \to 0$, the function approaches zero and its derivative approaches 1/3. (In fact, for small a, the Langevin function is approximately $\frac{a}{3(1-a)}$, and for very small a, it is $\frac{1}{3}a$.) Thus, for small a (i.e. hot temperatures) p_s approaches $p \times \frac{pE}{3kT}$ and no higher. The Langevin function looks like this:

FIGURE V.28

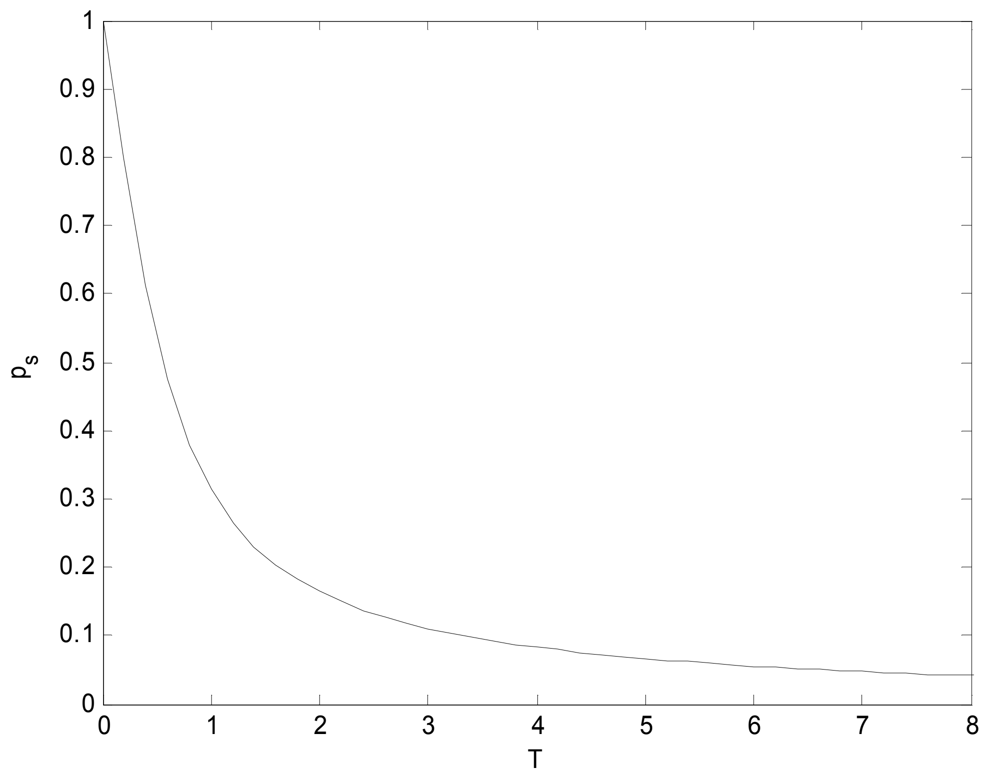


It may be more interesting to see directly how the sample dipole moment varies with temperature. If we express the sample dipole moment p_s in units of the molecular dipole moment p, and the temperature in units of pE/k, then equation 5.22.6 becomes

$$p_s = \frac{e^{1/T} + e^{-1/T}}{e^{1/T} - e^{-1/T}} - T = \coth(1/T) - T, \qquad 5.22.7$$

and that looks like this:

FIGURE V.29



The contribution to the polarization of a sample from the other two mechanisms – namely bond bending, and the pushing of electrons to one side, is independent of temperature. Thus, if we find that the polarization is temperature dependent, this tells us of the existence of a permanent dipole moment, as, for example, in methyl chloride CH_3Cl and H_2O. Indeed the temperature dependence of the polarization is part of the evidence that tells us that the water molecule is nonlinear. For small a (recall that $a = \frac{pE}{kT}$), the polarization of the material is $\frac{pE}{3kT}$, and so a graph of the polarization versus $1/T$ will be a straight line from which one can determine the dipole moent of the molecule – the greater the slope, the greater the dipole moment. One the other hand, if the polarization is temperature-independent, then the molecule is symmetric, such as methane CH_4 and

OCO. Indeed the independence of the polarization on temperature is part of the evidence that tells us that CO_2 is a linear molecule.

5.22 *Dielectric material in a alternating electric field.*

We have seen that, when we put a dielectric material in an electric field, it becomes polarized, and the **D** field is now $\epsilon\mathbf{E}$ instead of merely $\epsilon_0\mathbf{E}$. But how long does it take to become polarized? Does it happen instantaneously? In practice there is an enormous range in relaxation times. (We may define a relaxation time as the time taken for the material to reach a certain fraction – such as, perhaps $1 - e^{-1} = 63$ percent, or whatever fraction may be convenient in a particular context – of its final polarization.) The relaxation time may be practically isntantaneous, or it may be many hours.

As a consequence of the finite relaxation time, if we put a dielectric material in oscillating electric field $E = \hat{E}\cos\omega t$ (e.g. if light passes through a piece of glass), there will be a phase lag of D behind E. D will vary as $D = \hat{D}\cos(\omega t - \delta)$. Stated another way, if the E-field is $E = \hat{E}e^{i\omega t}$, the D-field will be $D = \hat{D}e^{i(\omega t - \delta)}$. Then $\frac{D}{E} = \frac{\hat{D}}{\hat{E}}e^{-i\delta} = \epsilon(\cos\delta - i\sin\delta)$. This can be written

$$D = \epsilon^* E, \qquad 5.22.1$$

where $\epsilon^* = \epsilon' - i\epsilon''$ and $\epsilon' = \epsilon\cos\delta$ and $\epsilon'' = \epsilon\sin\delta$.

The complex permittivity is just a way of expressing the phase difference between D and E. The magnitude, or modulus, of ϵ^* is ϵ, the ordinary permittivity in a static field.

Let us imagine that we have a dielectric material between the plates of a capacitor, and that an alternating potential difference is being applied across the plates. At some instant the charge density σ on the plates (which is equal to the D-field) is changing at a rate $\dot{\sigma}$, which is also equal to the rate of change $\dot{D}$ of the D-field), and the current in the circuit is $A\dot{D}$, where A is the area of each plate. The potential difference across the plates, on the other hand, is Ed, where d is the distance between the plates. The instantaneous rate of dissipation of energy in the material is $AdE\dot{D}$, or, let's say, the instantaneous rate of dissipation of energy per unit volume of the material is $E\dot{D}$.

Suppose $E = \hat{E}\cos\omega t$ and that $D = \hat{D}\cos(\omega t - \delta)$ so that

$$\dot{D} = -\hat{D}\omega\sin(\omega t - \delta) = -\hat{D}\omega(\sin\omega t\cos\delta - \cos\omega t\sin\delta).$$

The dissipation of energy, in unit volume, in a complete cycle (or period $2\pi/\omega$) is the integral, with respect to time, of $E\dot{D}$ from 0 to $2\pi/\omega$. That is,

$$\hat{E}\hat{D}\omega\int_0^{2\pi/\omega}\cos\omega t(\sin\omega t\cos\delta \;-\; \cos\omega t\sin\delta)dt.$$

The first integral is zero, so the dissipation of energy per unit volume per cycle is

$$\hat{E}\hat{D}\omega\sin\delta\int_0^{2\pi/\omega}\cos^2\omega t dt \;=\; \pi\hat{E}\hat{D}\omega\sin\delta.$$

Since the energy loss per cycle is proportional to $\sin\delta$, $\sin\delta$ is called the *loss factor*. (Sometimes the loss factor is given as $\tan\delta$, although this is an approximation only for small loss angles.)

CHAPTER 6
MAGNETIC EFFECT OF AN ELECTRIC CURRENT

6.1 *Introduction*

Most of us are familiar with the more obvious properties of magnets and compass needles. A magnet, often in the form of a short iron bar, will attract small pieces of iron such as nails and paper clips. Two magnets will either attract each other or repel each other, depending upon their orientation. If a bar magnet is placed on a sheet of paper and iron filings are scattered around the magnet, the iron filings arrange themselves in a manner that reminds us of the electric field lines surrounding an electric dipole. All in all, a bar magnet has some properties that are quite similar to those of an electric dipole. The region of space around a magnet within which it exerts its magic influence is called a *magnetic field*, and its geometry is rather similar to that of the electric field around an electric dipole – although its *nature* seems a little different, in that it interacts with iron filings and small bits of iron rather than with scraps of paper or pith-balls.

The resemblance of the magnetic field of a bar magnet to the electric field of an electric dipole was sometimes demonstrated in Victorian times by means of a Robison Ball-ended Magnet, which was a magnet shaped something like this:

FIGURE VI.1

The geometry of the magnetic field (demonstrated, for example, with iron filings) then *greatly* resembled the geometry of an electric dipole field. Indeed it looked as though a magnet had two *poles* (analogous to, but not the same as, electric charges), and that one of them acts as a *source* for magnetic field lines (i.e. field lines diverge from it), and the other acts as a *sink* (i.e. field lines converge to it). Rather than calling the poles "positive" and "negative", we somewhat arbitrarily call them "north" and "south" poles, the "north" pole being the source and the "south" pole the sink. By experimenting with two or more magnets, we find that like poles repel and unlike poles attract.

We also observe that a freely-suspended magnet (i.e. a compass needle) will orient itself so that one end points approximately north, and the other points approximately south, and it is these poles that are called the "north" and "south" poles of the magnet.

Since unlike poles attract, we deduce (or rather William Gilbert, in his 1600 book *De Magnete, Magneticisque Corporibus, et de Magno Magnete Tellure* deduced) that Earth itself acts as a giant magnet, with a south magnetic pole somewhere in the Arctic and a north magnetic pole in the Antarctic. The Arctic magnetic pole is at present in Bathurst Island in northern Canada and is usually marked in atlases as the "North Magnetic Pole", though magnetically it is a *sink*, rather than a source. The Antarctic magnetic pole is at present just offshore from Wilkes Land in the Antarctic continent. The Antarctic magnetic pole is a *source*, although it is usually marked in atlases as the "South Magnetic Pole". Some people have advocated calling the end of a compass needle that points north the "north-seeking pole", and the other end the "south-seeking pole. This has much to commend it, but usually, instead, we just call them the "north" and "south" poles. Unfortunately this means that the Earth's magnetic pole in the Arctic is really a south magnetic pole, and the pole in the Antarctic is a north magnetic pole.

The resemblance of the magnetic field of a bar magnet to a dipole field, and the *very* close resemblance of a "Robison Ball-ended Magnet" to a dipole, with a point source (the north pole) at one end and a point sink (the south pole) at the other, is, however, deceptive.

In truth a magnetic field has *no* sources and *no* sinks. This is even expressed as one of Maxwell's equations, div $\mathbf{B} = 0$, as being one of the defining characteristics of a magnetic field. The magnetic lines of force always form closed loops. *Inside* a bar magnet (even inside the connecting rod of a Robison magnet) the magnetic field lines are directed from the south pole to the north pole. If a magnet, even a Robison magnet, is cut in two, we do not isolate two separate poles. Instead each half of the magnet becomes a dipolar magnet itself.

All of this is very curious, and matters stood like this until Oersted made an outstanding discovery in 1820 (it is said while giving a university lecture in Copenhagen), which added what may have seemed like an additional complication, but which turned out to be in many ways a great simplification. He observed that, if an electric current is made to flow in a wire near to a freely suspended compass needle, the compass needle is deflected. Similarly, if a current flows in a wire that is free to move and is near to a fixed bar magnet, the wire experiences a force at right angles to the wire.

From this point on we understand that a magnetic field is something that is *primarily associated with an electric current.* All the phenomena associated with magnetized iron, nickel or cobalt, and "lodestones" and compass needles are somehow secondary to the fundamental phenomenon that an electric current is always surrounded by a magnetic field. Indeed, Ampère speculated that the magnetic field of a bar magnet may be caused by many circulating current loops within the iron. He was right! – the little current loops are today identified with electron spin.

If the direction of the magnetic field is taken to be the direction of the force on the north pole of a compass needle, Oersted's observation showed that the magnetic field around a current is in the form of concentric circles surrounding the current. Thus in figure VI.2, the current is assumed to be going away from you at right angles to the plane of your computer screen (or of the paper, if you have printed this page out), and the magnetic field lines are concentric circles around the current,

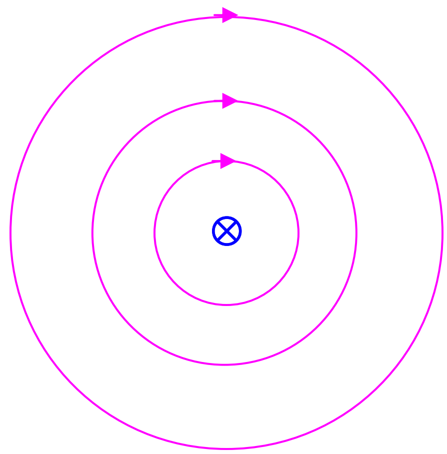


FIGURE VI.2

In the remainder of this chapter, we shall no longer be concerned with magnets, compass needles and lodestones. These may come in a later chapter. In the remainder of this chapter we shall be concerned with the magnetic field that surrounds an electric current.

6.2 *Definition of the Amp*

We have seen that an electric current is surrounded by a magnetic field; and also that, if a wire carrying a current is situated in an external magnetic field, it experiences a force at right angles to the current. It is therefore not surprising that two current-carrying wires exert forces upon each other.

More precisely, if there are two parallel wires each carrying a current in the same direction, the two wires will *attract* each other with a force that depends on the strength of the current in each, and the distance between the wires.

Definition. One *amp* (also called an ampère) is that steady current which, flowing in each of two parallel wires of negligible cross-section one metre apart *in vacuo*, gives rise to a force between them of 2×10^{-7} newtons per metre of their length.

At last! We now know what an amp is, and consequently we know what a coulomb, a volt and an ohm are. We have been left in a state of uncertainty until now. No longer!

But you may ask: Why the factor 2×10^{-7}? Why not define an amp in such a manner that the force is 1 N m^{-1}? This is a good question, and its answer is tied to the long and tortuous history of units in electromagnetism. I shall probably discuss this history, and the various "CGS" units, in a later chapter. In brief, it took a long time to understand that electrostatics, magnetism and current electricity were all aspects of the same basic phenomena, and different systems of units developed within each topic. In particular a so-called "practical" unit, the *amp* (defined in terms of the rate of deposition of silver from an electrolytic solution) became so entrenched that it was felt impractical to abandon it. Consequently when all the various systems of electromagnetic units became unified in the twentieth century (starting with proposals by Giorgi based on the metre, kilogram and second (MKS) as long ago as 1895) in the "Système International" (SI), it was determined that the fundamental unit of current should be identical with what had always been known as the ampère. (The factor 2, by the way, is not related to their being two wires in the definition.) The amp is the only SI unit in which any number other than "one" is incorporated into its definition, and the exception was forced by the desire to maintain the amp.

[A proposal to be considered (and probably passed) by the Conférence Générale des Poids et Mesures i n 2015 would re-define the coulomb in such a manner that the magnitude of the charge on a single electron is exactly 1.60217×10^{-19} C.]

One last point before leaving this section. In the opening paragraph I wrote that "It is therefore not surprising that two current-carrying wires exert forces upon each other." Yet when I first learned, as a student, of the mutual attraction of two parallel electric currents, I was very astonished indeed. The reason why this is astonishing is discussed in Chapter 15 (Special Relativity) of the Classical Mechanics section of these notes.

6.3 *Definition of the Magnetic Field*

We are going to define the magnitude and direction of the magnetic field entirely by reference to its effect upon an electric current, without reference to magnets or lodestones. We have already noted that, if an electric current flows in a wire in an externally-imposed magnetic field, it experiences a force at right angles to the wire.

I want you to imagine that there is a magnetic field in this room, originating, perhaps, from some source outside the room. This need not entail a great deal of imagination, for there already *is* such a magnetic field – namely, Earth's magnetic field. I'll tell you that the field within the room is uniform, but I shan't tell you anything about either its magnitude or its direction.

You have a straight wire and you can pass a current through it. You will note that there is a force on the wire. Perhaps we can define the direction of the field as being the direction of this force. But this won't do at all, because the force is always at right angles to the wire no matter what its orientation! We do notice, however, that the *magnitude* of the force depends on the orientation of the wire; and there is one *unique orientation* of the wire in which it experiences *no force at all*. Since this orientation is unique, we choose to *define the direction of the magnetic field* as being parallel to the wire when the orientation of the wire is such that it experiences no force.

This leaves a two-fold ambiguity since, even with the wire in its unique orientation, we can cause the current to flow in one direction or in the opposite direction. We still have to resolve this ambiguity. Have patience for a few more lines.

As we move our wire around in the magnetic field, from one orientation to another, we notice that, while the direction of the force on it is always at right angles to the wire, the *magnitude* of the force depends on the orientation of the wire, being zero (by definition) when it is parallel to the field and greatest when it is perpendicular to it.

Definition. The intensity B (also called the flux density, or field strength, or merely "field") of a magnetic field is equal to the maximum force exerted per unit length on unit current (this maximum force occurring when the current and field are at right angles to each other).

The dimensions of B are $\dfrac{\mathrm{MLT^{-2}}}{\mathrm{LQT^{-1}}} = \mathrm{MT^{-1}Q^{-1}}$.

Definition. If the maximum force per unit length on a current of 1 amp (this maximum force occurring, of course, when current and field are perpendicular) is 1 N m^{-1}, the intensity of the field is 1 *tesla* (T).

By definition, then, when the wire is parallel to the field, the force on it is zero; and, when it is perpendicular to the field, the force per unit length is IB newtons per metre.

It will be found that, when the angle between the current and the field is θ, the force per unit length, F', is

$$F' = IB\sin\theta. \tag{6.3.1}$$

In vector notation, we can write this as

$$\mathbf{F'} = \mathbf{I}\times\mathbf{B}, \tag{6.3.2}$$

where, in choosing to write $\mathbf{I}\times\mathbf{B}$ rather than $\mathbf{F'} = \mathbf{B}\times\mathbf{I}$, we have removed the two-fold ambiguity in our definition of the direction of **B**. Equation 6.3.2 expresses the "right-hand rule" for determining the relation between the directions of the current, field and force.

6.4 *The Biot-Savart Law*

Since we now know that a wire carrying an electric current is surrounded by a magnetic field, and we have also decided upon how we are going to define the intensity of a magnetic field, we want to ask if we can calculate the intensity of the magnetic field in the vicinity of various geometries of electrical conductor, such as a straight wire, or a plane coil, or a solenoid. When we were calculating the *electric* field in the vicinity of various geometries of *charged bodies*, we started from *Coulomb's Law*, which told us what the field was at a given distance from a point charge. Is there something similar in electromagnetism which tells us how the *magnetic* field varies with distance from an *electric current*? Indeed there is, and it is called the *Biot-Savart Law*.

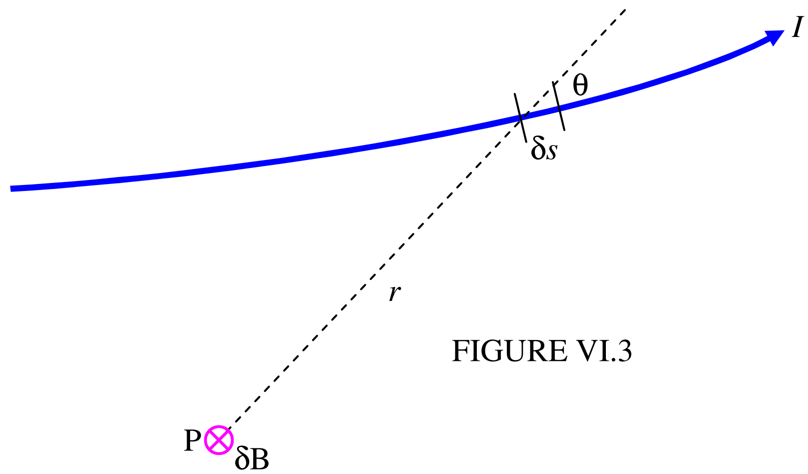


FIGURE VI.3

Figure VI.3 shows a portion of an electrical circuit carrying a current I. The Biot-Savart Law tells us what the contribution δB is at a point P from an elemental portion of the electrical circuit of length δs at a distance r from P, the angle between the current at δs and the radius vector from P to δs being θ. The Biot-Savart Law tells us that

$$\delta B \propto \frac{I\,\delta s \sin\theta}{r^2}. \qquad 6.4.1$$

This law will enable us, by integrating it around various electrical circuits, to calculate the total magnetic field at any point in the vicinity of the circuit.

But – can I prove the Biot-Savart Law, or is it just a bald statement from nowhere? The answer is neither. I cannot prove it, but nor is it merely a bald statement from nowhere. First of all, it is a not unreasonable guess to suppose that the field is proportional to I and to δs, and also inversely proportional to r^2, since δs, in the limit, approaches a point source. But you are still free to regard it, if you wish, as speculation, even if reasonable speculation. Physics is an experimental science, and to that extent you cannot "prove" anything in a mathematical sense; you can experiment and measure. The Biot-Savart law enables us to calculate what the magnetic field ought to be near a straight wire, near a plane circular current, inside a solenoid, and indeed near any geometry you can imagine. So far, after having used it to calculate the field near millions of conductors of a myriad shapes and sizes, the predicted field has always agreed with experimental measurement. Thus the Biot-Savart law is *likely* to be true – but you are perfectly correct in asserting that, no matter how many magnetic fields it has correctly predicted, there is always the chance that, some day, it will predict a field for some unusually-shaped circuit that disagrees with what is measured. All that is needed is *one* such example, and the law is disproved. You may, if you wish, try and discover, for a Ph.D. project, such a circuit; but I would not recommend that you spend your time on it!

There remains the question of what to write for the constant of proportionality. We are free to use any symbol we like, but, in modern notation, we symbol we use is $\frac{\mu_0}{4\pi}\cdot$ Why the factor 4π? The inclusion of 4π gives us what is called a "rationalized" definition, and it is introduced for the same reasons that we introduced a similar factor in the constant of proportionality for Coulomb's law, namely that it results in the appearance of 4π in spherically-symmetric geometries, 2π in cylindrically-symmetric geometries, and no π where the magnetic field is uniform. Not everyone uses this definition, and this will be discussed in a later chapter, but it is certainly the recommended one.

In any case, the Biot-Savart Law takes the form

$$\delta B = \frac{\mu_0}{4\pi}\frac{I\,\delta s\sin\theta}{r^2}. \qquad 6.4.2$$

The constant μ_0 is called the *permeability of free space*, "free space" meaning a vacuum. The subscript allows for the possibility that if we do an experiment in a medium other than a vacuum, the permeability may be different, and we can then use a different subscript, or none at all. In practice the permeability of air is very little different from that of a vacuum, and hence I shall normally use the symbol μ_0 for experiments performed in air, unless we are discussing measurement of very high precision.

From equation 6.4.2, we can see that the SI units of permeability are T m A^{-1} (tesla metres per amp). In a later chapter we shall come across another unit – the *henry* – for a quantity (*inductance*) that we have not yet described, and we shall see then that a more convenient unit for permeability is H m^{-1} (henrys per metre) – but we are getting ahead of ourselves.

What is the numerical value of μ_0? I shall reveal that in the next chapter.

Exercise. Show that the *dimensions* of permeability are MLQ^{-2}. This means that you may, if you wish, express permeability in units of kg m C^{-2} – although you may get some queer looks if you do.

Thought for the Day.

The sketch shows two current elements, each of length δs, the current being the same in each but in different directions. Is the force on one element from the other equal but opposite to the force on the other from the one? If not, is there something wrong with Newton's third law of motion? Discuss this over lunch.

6.5 *Magnetic Field Near a Long, Straight, Current-carrying Conductor*

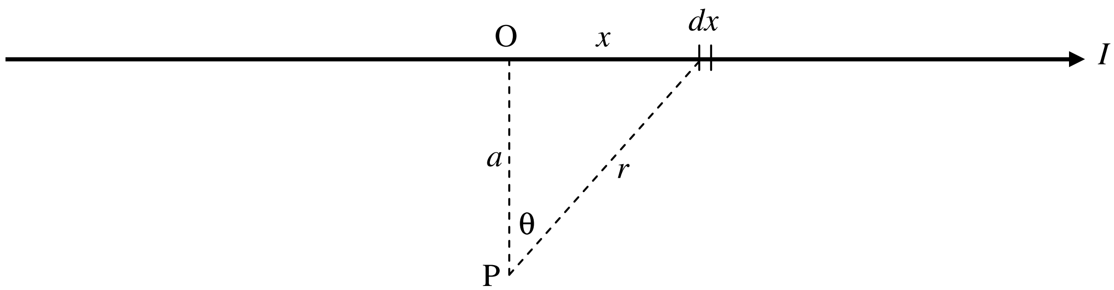


FIGURE VI.4

Consider a point P at a distance a from a conductor carrying a current I (figure VI.4). The contribution to the magnetic field at P from the elemental length dx is

$$dB = \frac{\mu}{4\pi} \cdot \frac{I\, dx \cos\theta}{r^2}. \qquad 6.5.1$$

(Look at the way I have drawn θ if you are worried about the cosine.)

Here I have omitted the subscript zero on the permeability to allow for the possibility that the wire is immersed in a medium in which the permeability is not the same as that of a vacuum. (The permeability of liquid oxygen, for example, is slightly greater than that of free space.) The direction of the field at P is into the plane of the "paper" (or of your computer screen).

We need to express this in terms of one variable, and we'll choose θ. We can see that $r = a\sec\theta$ and $x = a\tan\theta$, so that $dx = a\sec^2\theta\, d\theta$. Thus equation 6.5.1 becomes

$$dB = \frac{\mu I}{4\pi a} \sin\theta\, d\theta. \qquad 6.5.2$$

Upon integrating this from $-\pi/2$ to $+\pi/2$ (or from 0 to $\pi/2$ and then double it), we find that the field at P is

$$B = \frac{\mu I}{2\pi a}. \qquad 6.5.3$$

Note the 2π in this problem with cylindrical symmetry.

6.6 *Field on the Axis and in the Plane of a Plane Circular Current-carrying Coil*

I strongly recommend that you compare and contrast this derivation and the result with the treatment of the electric field on the axis of a charged ring in Section 1.6.4 of Chapter 1. Indeed I am copying the drawing from there and then modifying it as need be.

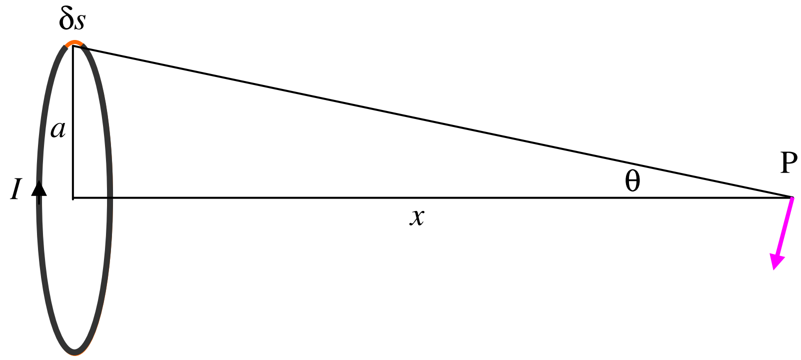


FIGURE VI.5

The contribution to the magnetic field at P from an element δs of the current is $\dfrac{\mu I\,\delta s}{4\pi(a^2+x^2)}$ in the direction shown by the coloured arrow. By symmetry, the total component of this from the entire coil perpendicular to the axis is zero, and the only component of interest is the component along the axis, which is $\dfrac{\mu I\,\delta s}{4\pi(a^2+x^2)}$ times $\sin\theta$.

The integral of δs around the whole coil is just the circumference of the coil, $2\pi a$, and if we write $\sin\theta = \dfrac{a}{(a^2+x^2)^{1/2}}$, we find that the field at P from the entire coil is

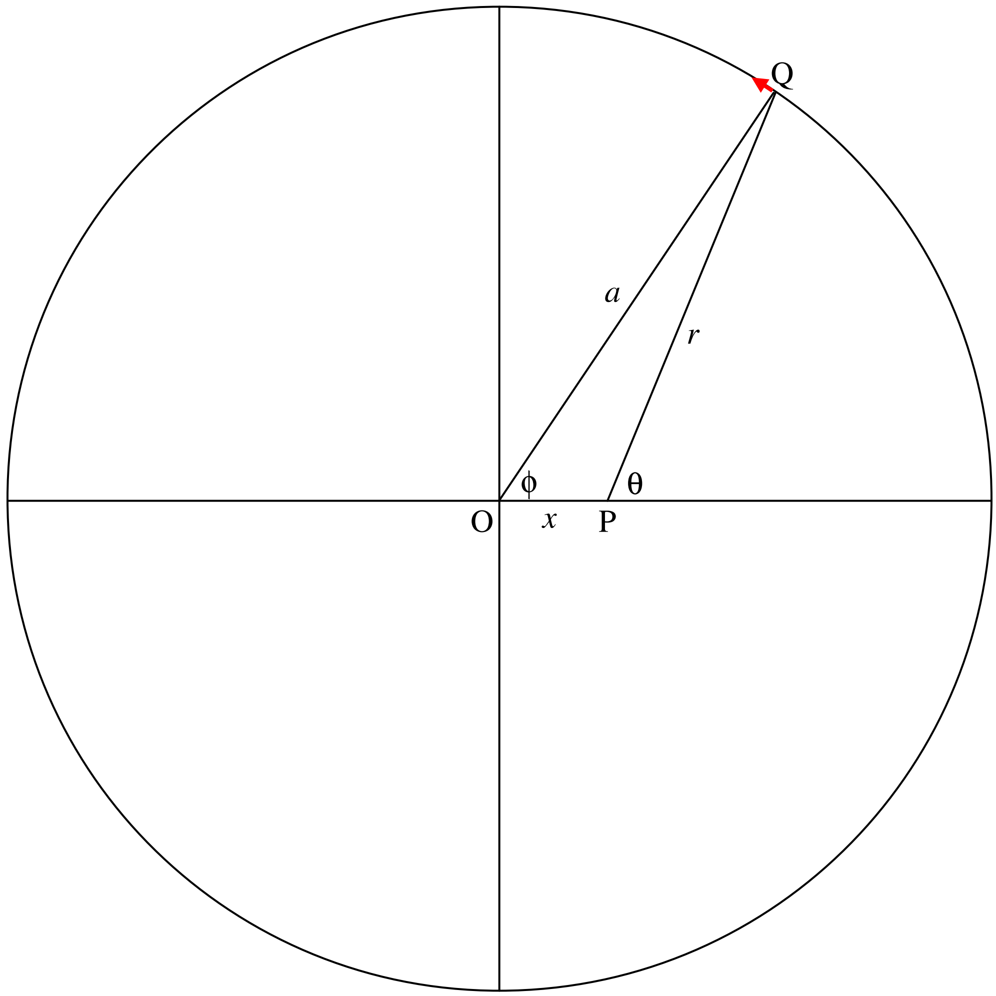
$$B = \frac{\mu I a^2}{2(a^2+x^2)^{3/2}}, \qquad 6.6.1$$

or N times this if there are N turns in the coil. At the centre of the coil the field is

$$B = \frac{\mu I}{2a}. \qquad 6.6.2$$

The field is greatest at the centre of the coil and it decreases monotonically to zero at infinity. The field is directed to the left in figure IV.5.

We can calculate the field in the plane of the ring as follows.



Consider an element of the wire at Q of length $ad\phi$. The angle between the current at Q and the line PQ is $90^\circ - (\theta - \phi)$. The contribution to the *B*-field at P from the current I this element is

$$\frac{\mu_0}{4\pi} \cdot \frac{Ia\cos(\theta - \phi)d\phi}{r^2} .$$

The field from the entire ring is therefore

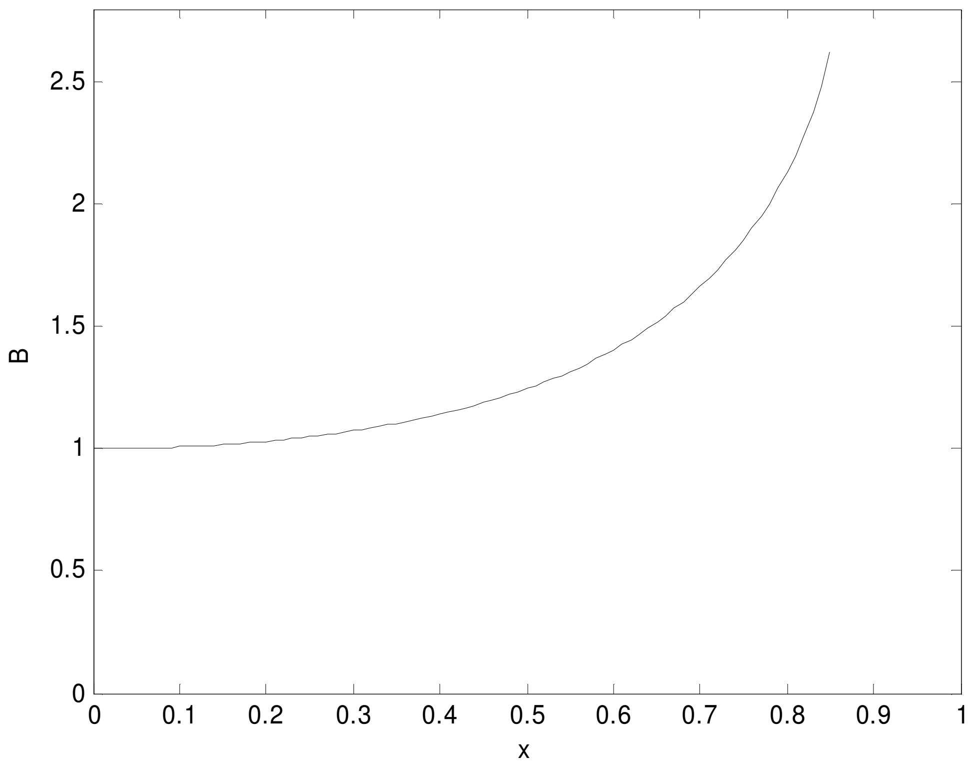
$$\frac{2\mu_0 Ia}{4\pi}\int_0^\pi \frac{\cos(\theta-\phi)d\phi}{r^2},$$

where $$r^2 = a^2 + x^2 - 2ax\cos\phi,$$

and $$\cos(\theta-\phi) = \frac{a^2 + r^2 - x^2}{2ar}.$$

This requires a numerical integration. The results are shown in the following graph, in which the abscissa, x, is the distance from the centre of the circle in units of its radius, and the ordinate, B, is the magnetic field in units of its value $\mu_0 I/(2a)$ at the centre. Further out than $x = 0.8$, the field increases rapidly.

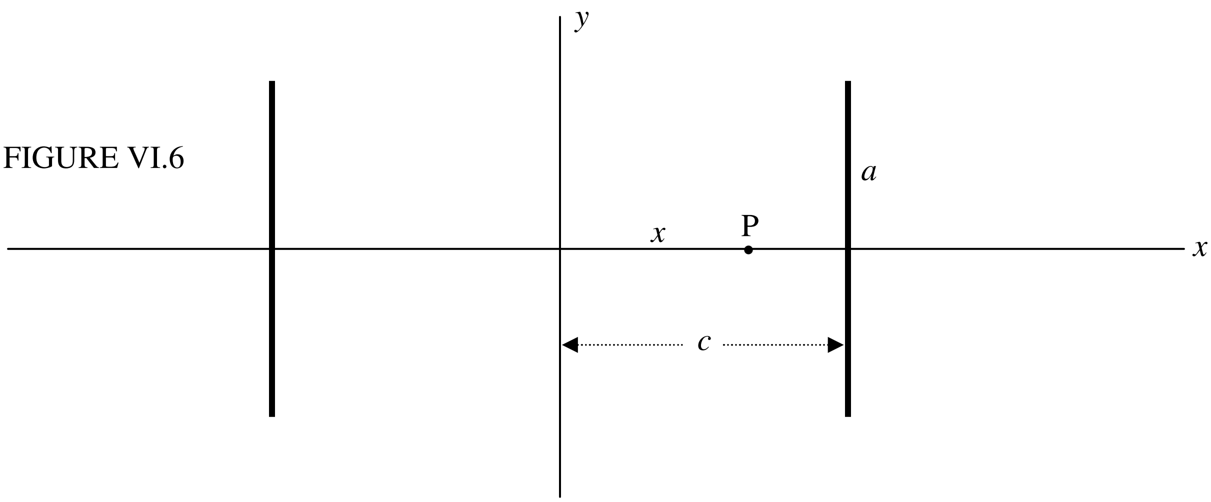
Magnetic field in the plane of a ring current.



6.7 *Helmholtz Coils*

Let us calculate the field at a point halfway between two identical parallel plane coils. If the separation between the coils is equal to the radius of one of the coils, the arrangement is known as "Helmholtz coils", and we shall see why they are of particular interest. To

FIGURE VI.6



begin with, however, we'll start with two coils, each of radius a, separated by a distance $2c$.

There are N turns in each coil, and each carries a current I.

The field at P is

$$B = \frac{\mu NIa^2}{2}\left(\frac{1}{[a^2 + (c-x)^2]^{3/2}} + \frac{1}{[a^2 + (c+x)^2]^{3/2}}\right). \tag{6.7.1}$$

At the origin ($x = 0$), the field is

$$B = \frac{\mu NIa^2}{(a^2 + c^2)^{3/2}}. \tag{6.7.2}$$

(What does this become if $c = 0$? Is this what you'd expect?)

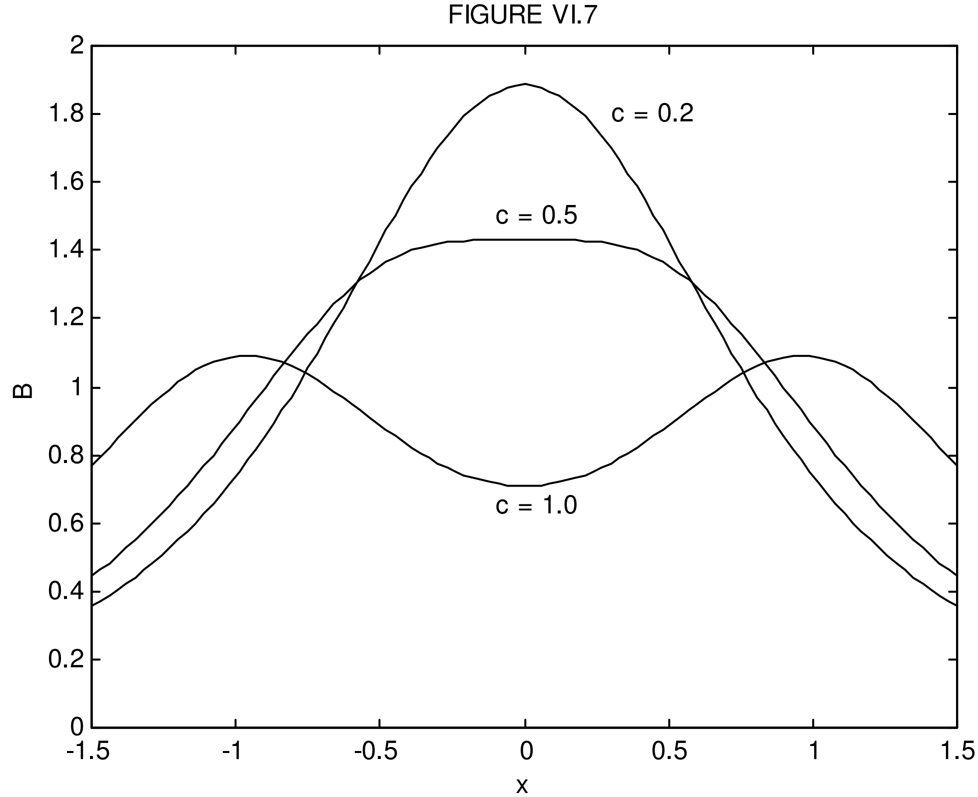
If we express B in units of $\mu NI/(2a)$ and c and x in units of a, equation 6.7.1 becomes

$$B = \frac{1}{[1 + (c - x)^2]^{3/2}} + \frac{1}{[1 + (c + x)^2]^{3/2}}. \tag{6.7.4}$$

Figure VI.7 shows the field as a function of x for three values of c. The coil separation is $2c$, and distances are in units of the coil radius a. Notice that when $c = 0.5$, which means that the coil separation is equal to the coil radius, the field is uniform over a large range, and this is the usefulness of the Helmholtz arrangement for providing a uniform field. If you are energetic, you could try differentiating equation 6.7.4 twice with respect to x and show that the second derivative is zero when $c = 0.5$.

For the Helmholtz arrangement the field at the origin is $\frac{8\sqrt{5}}{25}\cdot\frac{\mu NI}{a} = \frac{0.7155\mu NI}{a}$.

FIGURE VI.7



6.8 *Field on the Axis of a Long Solenoid*

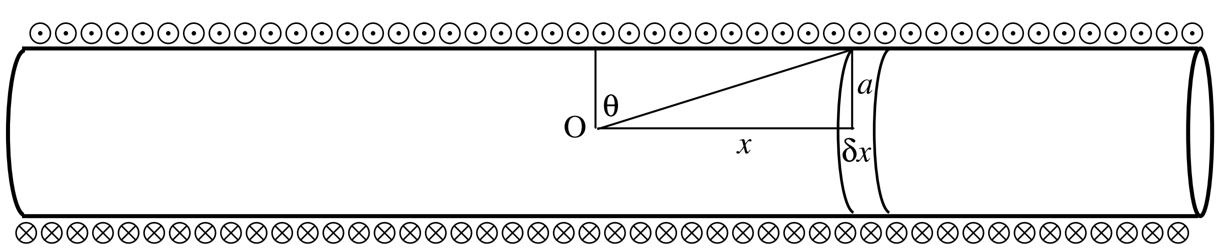


FIGURE VI.8

The solenoid, of radius a, is wound with n turns per unit length of a wire carrying a current in the direction indicated by the symbols ⊗ and ⊙. At a point O on the axis of the solenoid the contribution to the magnetic field arising from an elemental ring of width δx (hence having $n\,\delta x$ turns) at a distance x from O is

$$\delta B = \frac{\mu n \delta x I a^2}{2(a^2 + x^2)^{3/2}} = \frac{\mu n I}{2a} \cdot \frac{a^3\,\delta x}{(a^2 + x^2)^{3/2}}. \qquad 6.8.1$$

This field is directed towards the right.

Let us express this in terms of the angle θ.

We have $x = a\tan\theta$, $\delta x = a\sec^2\theta\,\delta\theta$, and $\dfrac{a^3}{(a^2 + x^2)^{3/2}} = \cos^3\theta$. Equation 6.8.1 becomes

$$\delta B = \tfrac{1}{2}\mu n I \cos\theta. \qquad 6.8.2$$

If the solenoid is of infinite length, to find the field from the entire infinite solenoid, we integrate from $\theta = \pi/2$ to 0 and double it. Thus

$$B = \mu n I \int_0^{\pi/2} \cos\theta\, d\theta. \qquad 6.8.3$$

Thus the field on the axis of the solenoid is

$$B = \mu n I. \qquad 6.8.4$$

This is the field on the *axis* of the solenoid. What happens if we move away from the axis? Is the field a little greater as we move away from the axis, or is it a little less? Is the field a maximum on the axis, or a minimum? Or does the field go through a maximum, or a minimum, somewhere between the axis and the circumference? We shall answer these questions in section 6.11.

6.9 *The Magnetic Field H*

If you look at the various formulas for the magnetic field B near various geometries of conductor, such as equations 6.5.3, 6.6.2, 6.7.1, 6.8.4, you will see that there is always a μ on the right hand side. It is often convenient to define a quantity $H = B/\mu$. Then these equations become just

$$H = \frac{I}{2\pi a}, \tag{6.9.1}$$

$$H = \frac{I}{2a}, \tag{6.9.2}$$

$$H = \frac{NIa^2}{2}\left(\frac{1}{[a^2 + (c-x)^2]^{3/2}} + \frac{1}{[a^2 + (c+x)^2]^{3/2}}\right), \tag{6.9.3}$$

$$H = nI. \tag{6.9.10}$$

It is easily seen from any of these equations that the SI units of H are A m^{-1}, or amps per metre, and the dimensions are QT^{-1}M^{-1}.

Of course the magnetic field, whether represented by the quantity B or by H, is a vector quantity, and the relation between the two representations can be written

$$\mathbf{B} = \mu\mathbf{H}. \tag{6.9.11}$$

In an isotropic medium **B** and **H** are parallel, but in an anisotropic medium they are not parallel (except in the directions of the eigenvectors of the permeability tensor), and permeability is a tensor. This was discussed in section 1.7.1 with respect to the equation $\mathbf{D} = \varepsilon\mathbf{E}$.

6.10 *Flux*

Recall from Section 1.8 that we defined two extensive scalar quantities $\Phi_{\rm E} = \iint \mathbf{E} \cdot d\mathbf{A}$ and $\Phi_{\rm D} = \iint \mathbf{D} \cdot d\mathbf{A}$, which I called the E-flux and the D-flux. In an entirely similar manner I can define the B-flux and H-flux of a magnetic field by

$$\Phi_{\rm B} = \iint \mathbf{B} \cdot d\mathbf{A} \tag{6.10.1}$$

and

$$\Phi_{\rm H} = \iint \mathbf{H} \cdot d\mathbf{A}. \tag{6.10.2}$$

The SI unit of $\Phi_{\rm B}$ is the tesla metre-squared, or T m^2, also called the *weber* Wb.

A summary of the SI units and dimensions of the four fields and fluxes might not come amiss here.

E	V m^{-1}	MLT^{-2}Q^{-1}
D	C m^{-2}	L^{-2}Q

B	T	$MT^{-1}Q^{-1}$
H	A m^{-1}	$L^{-1}T^{-1}Q$
Φ_E	V m	$ML^3T^{-2}Q^{-1}$
Φ_D	C	Q
Φ_B	Wb	$ML^2T^{-1}Q^{-1}$
Φ_H	A m	$LT^{-1}Q$

6.11 *Ampère's Theorem*

In Section 1.9 we introduced Gauss's theorem, which is that the total normal component of the *D*-flux through a closed surface is equal to the charge enclosed within that surface. Gauss's theorem is a consequence of Coulomb's law, in which the electric field from a point source falls off inversely as the square of the distance. We found that Gauss's theorem was surprisingly useful in that it enabled us almost immediately to write down expressions for the electric field in the vicinity of various shapes of charged bodies without going through a whole lot of calculus.

Is there perhaps a similar theorem concerned with the magnetic field around a current-carrying conductor that will enable us to calculate the magnetic field in its vicinity without going through a lot of calculus? There is indeed, and it is called *Ampère's Theorem*.

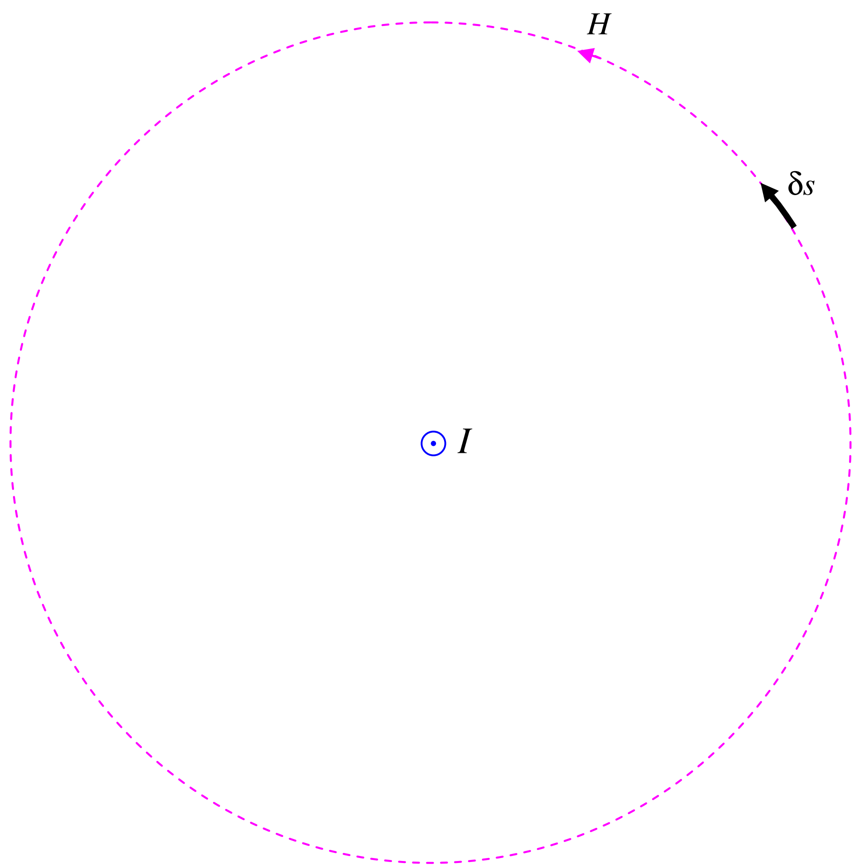


FIGURE VI.9

In figure VI.9 there is supposed to be a current I coming towards you in the middle of the circle. I have drawn one of the magnetic field lines – a dashed line of radius r. The strength of the field there is $H = I/(2\pi r)$. I have also drawn a small elemental length ds on the circumference of the circle. The line integral of the field around the circle is just H times the circumference of the circle. That is, the line integral of the field around the circle is just I. Note that this is independent of the radius of the circle. At greater distances from the current, the field falls off as $1/r$, but the circumference of the circle increases as r, so the product of the two (the line integral) is independent of r.

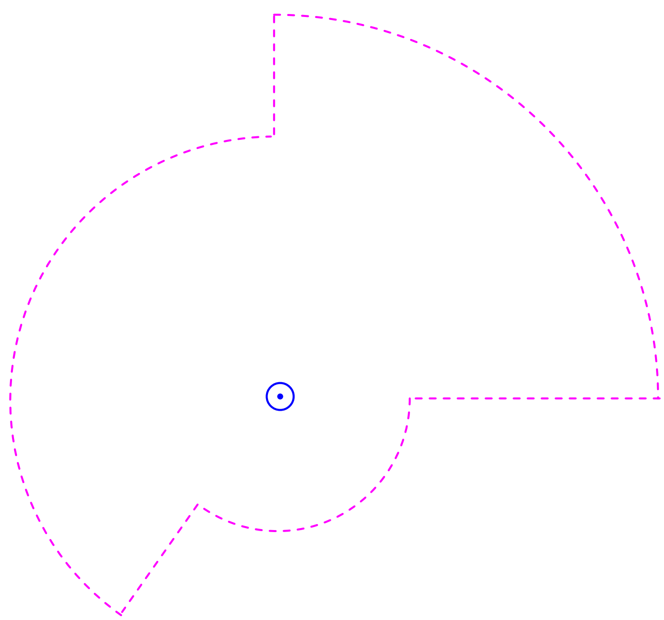


FIGURE VI.10

Consequently, if I calculate the line integral around a circuit such as the one shown in figure VI.10, it will still come to just I. Indeed it doesn't matter what the shape of the path is. The line integral is $\int \mathbf{H} \cdot \mathbf{ds}$. The field **H** at some point is perpendicular to the line joining the current to the point, and the vector **ds** is directed along the path of integration, and $\mathbf{H} \cdot \mathbf{ds}$ is equal to H times the component of **ds** along the direction of **H**, so that, regardless of the length and shape of the path of integration:

The line integral of the field **H** *around any closed path is equal to the current enclosed by that path.*

This is *Ampère's Theorem*.

So now let's do the infinite solenoid again. Let us calculate the line integral around the rectangular amperian path shown in figure VI.11. There is no contribution to the line integral along the vertical sides of the rectangle because these sides are perpendicular to the field, and there is no contribution from the top side of the rectangle, since the field there is zero (if the solenoid is infinite). The only contribution to the line integral is along the bottom side of the rectangle, and the line integral there is just Hl, where l is the length of the rectangle. If the number turns of wire per unit length along the solenoid is n, there will be nl turns enclosed by the rectangle, and hence the current enclosed by the rectangle is nlI, where I is the current in the wire. Therefore by Ampère's theorem, $Hl = nlI$, and so $H = nI$, which is what we deduced before rather more laboriously. Here H is the strength of the field at the position of the lower side of the rectangle; but we can place the rectangle at any height, so we see that the field is nI <u>anywhere inside the solenoid</u>. That is, the field inside an infinite solenoid is <u>uniform</u>.

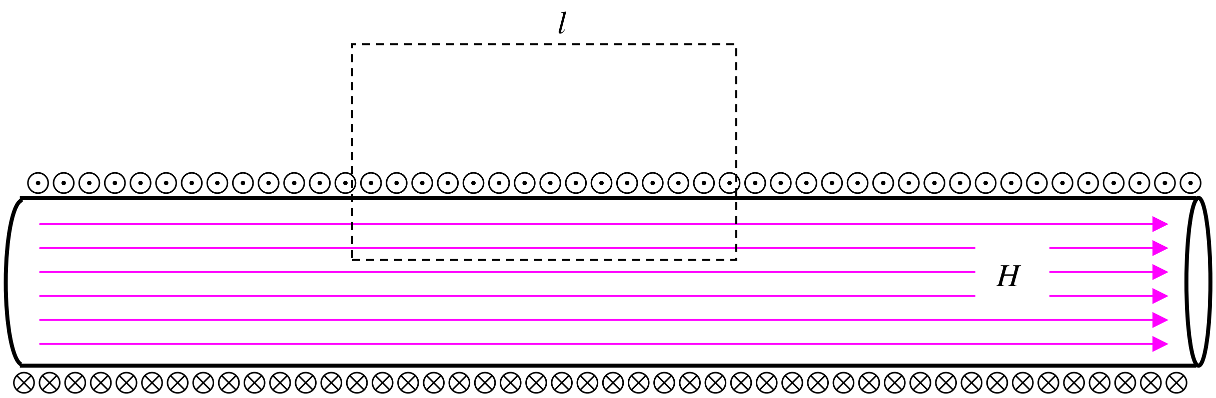
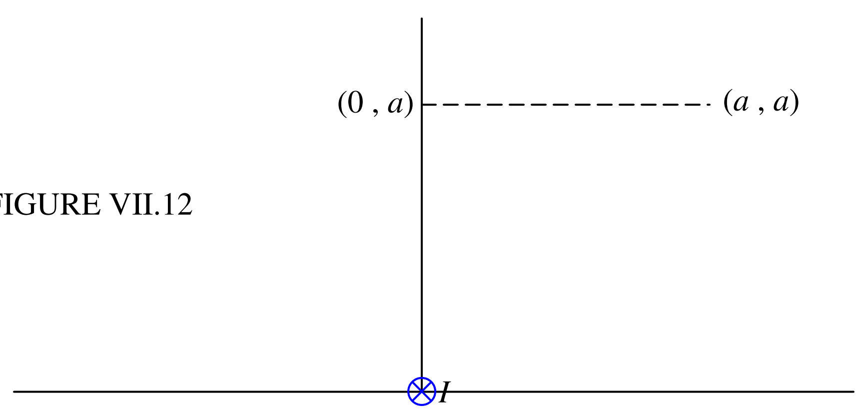


FIGURE VI.11

It is perhaps worth noting that Gauss's theorem is a consequence of the inverse square diminution of the electric field with distance from a point charge, and Ampère's theorem is a consequence of the inverse first power diminution of the magnetic field with distance from a line current.

Example.
Here is an example of the calculation of a line integral (figure VII.12)

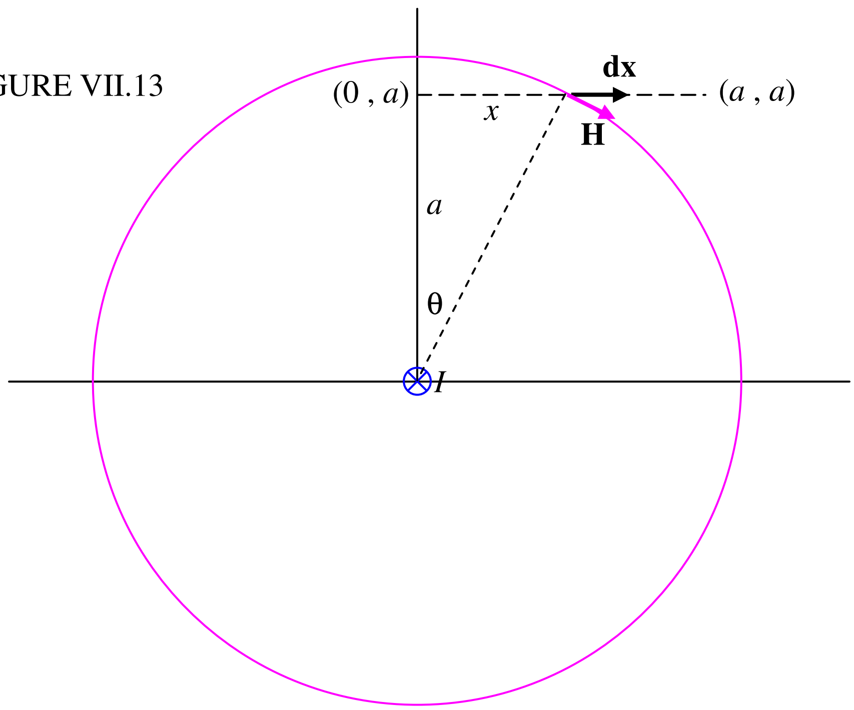
FIGURE VII.12



An electric current I flows into the plane of the paper at the origin of coordinates. Calculate the line integral of the magnetic field along the straight line joining the points $(0\ ,\ a)$ and $(a\ ,\ a)$.

In figure VII.13 I draw a (circular) line of force of the magnetic field **H**, and a vector **dx** where the line of force crosses the straight line of interest.

FIGURE VII.13



The line integral along the elemental length dx is $\mathbf{H}\ .\ \mathbf{dx} = H\,dx\cos\theta$. Here $H = \dfrac{I}{2\pi(a^2 + x^2)^{1/2}}$ and $\cos\theta = \dfrac{a}{(a^2 + x^2)^{1/2}}$, and so the line integral along **dx** is $\dfrac{aI\,dx}{2\pi(a^2 + x^2)}$. Integrate this from $x = 0$ to $x = a$ and you will find that the answer is $I/8$.

Figure VII.14 shows another method. The line integral around the square is, by Ampère's theorem, I, and so the line integral an eighth of the way round is $I/8$.

You will probably immediately feel that this second method is much the better and very "clever". I do not deny this, but it is still worthwhile to study carefully the process of line integration in the first method.

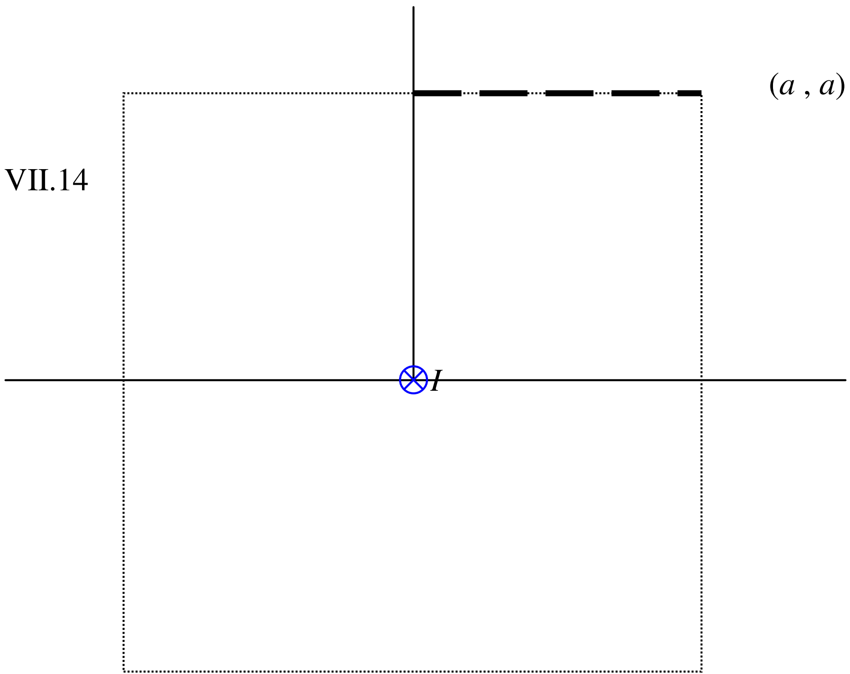
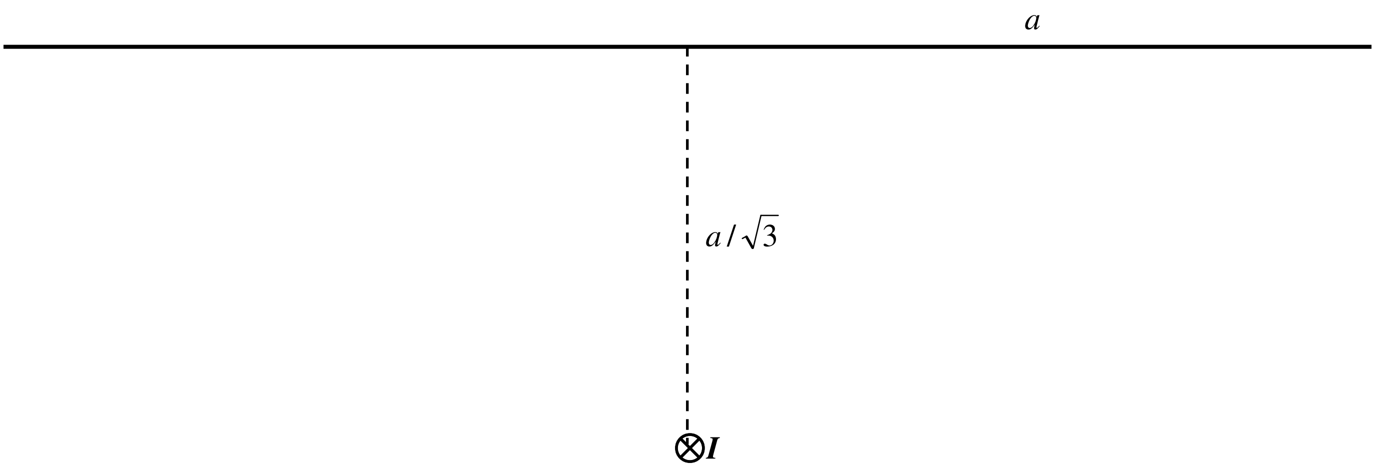


FIGURE VII.14

Another Example

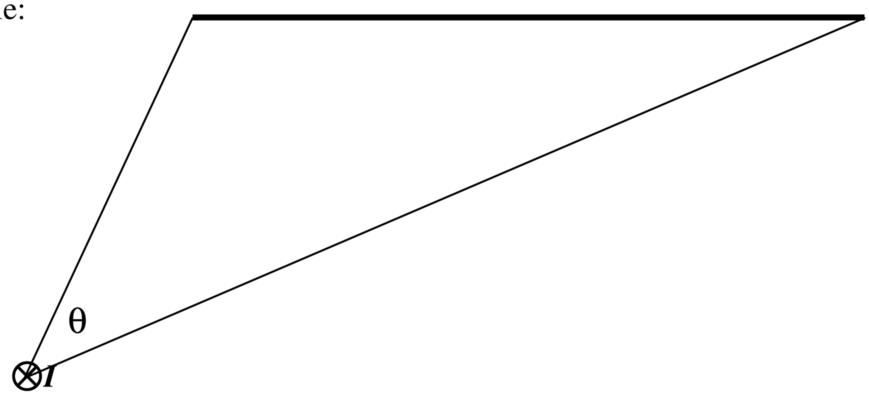


An electric current I flows into the plane of the paper. Calculate the line integral of the magnetic field along a straight line of length $2a$ whose mid-point is at a distance $a/\sqrt{3}$ from the current.

If you are not used to line integrals, I strongly urge you to do it by integration, as we did in the previous example. Some readers, however, will spot that the line is one side of an equilateral triangle, and so the line integral along the line is just $\frac{1}{3}I$.

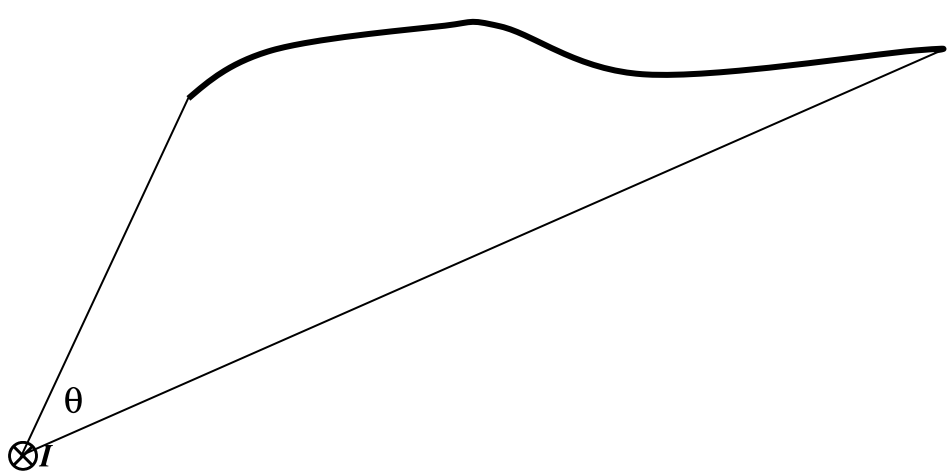
We can play this game with other polygons, of course, but it turns out to be even easier than that.

For example:



Show, by integration, that the line integral of the magnetic *H*-field along the thick line is just $\dfrac{\theta}{2\pi}$ times *I*.

After that it won't take long to convince yourself that the line integral along the thick line in the drawing below is also $\dfrac{\theta}{2\pi}$ times *I*.



Another Example

A straight cylindrical metal rod (or a wire for that matter) of radius a carries a current I. At a distance r from the axis, the magnetic field is clearly $I/(2\pi r)$ if $r > a$. But what is the magnetic field *inside* the rod at a distance r from the axis, $r < a$?

FIGURE VII.15

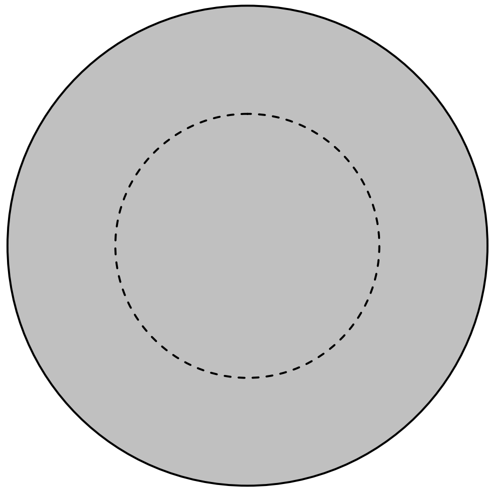


Figure VII.15 shows the cross-section of the rod, and I have drawn an amperian circle of radius r. If the field at the circumference of the circle is H, the line integral around the circle is $2\pi rH$. The current enclosed within the circle is Ir^2/a^2. These two are equal, and therefore $H = Ir/(2\pi a^2)$.

<u>More and More Examples</u>

In the above example, the current density was uniform. But now we can think of lots and lots of examples in which the current density is not uniform. For example, let us imagine that we have a long straight hollow cylindrical tube of radius a, perhaps a linear particle accelerator, and the current density J (amps per square metre) varies from the middle (axis) of the cylinder to its edge according to $J(r) = J_0(1 - r/a)$. The total current is, of course, $I = 2\pi\int_0^a J(r)rdr = \frac{1}{3}\pi a^2 J_0$, and the mean current density is $\bar{J} = \frac{1}{3}J_0$.

The question, however, is: what is the magnetic field H at a distance r from the axis? Further, show that the magnetic field at the edge (circumference) of the cylinder is $\frac{1}{6}J_0 a$, and that the field reaches a maximum value of $\frac{3}{16}J_0 a$ at $r = \frac{3}{4}a$.

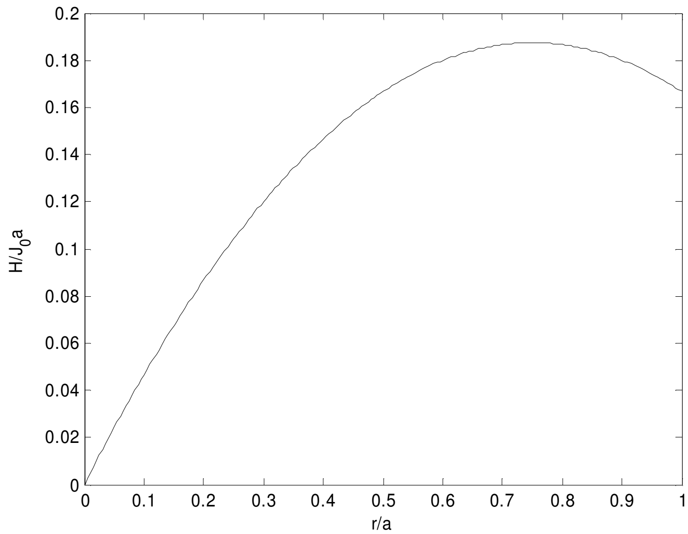
Well, the current enclosed within a distance r from the axis is

$$I = 2\pi\int_0^r J(x)xdx = \pi J_0 r^2(1 - \tfrac{2r}{3a}),$$

and this is equal to the line integral of the magnetic field around a circle of radius *r*, which is $2\pi rH$. Thus

$$H = \tfrac{1}{2} J_0 r\left(1 - \tfrac{2r}{3a}\right).$$

At the circumference of the cylinder, this comes to $\frac{1}{6}J_0 a$. Calculus shows that H reaches a maximum value of $\frac{3}{16}J_0 a$ at $r = \frac{3}{4}a$. The graph below shows $H/(J_0 a)$ as a function of *x*/*a*.



Having whetted our appetites, we can now try the same problem but with some other distributions of current density, such as

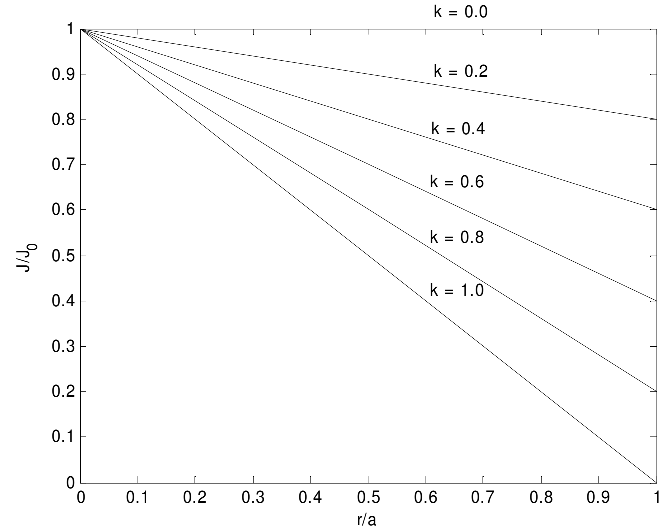
1	2	3	4
$J(r) = J_0\left(1 - \dfrac{kr}{a}\right)$,	$J(r) = J_0\left(1 - \dfrac{kr^2}{a^2}\right)$,	$J(r) = J_0\sqrt{1 - \dfrac{kr}{a}}$,	$J(r) = J_0\sqrt{1 - \dfrac{kr^2}{a^2}}$.

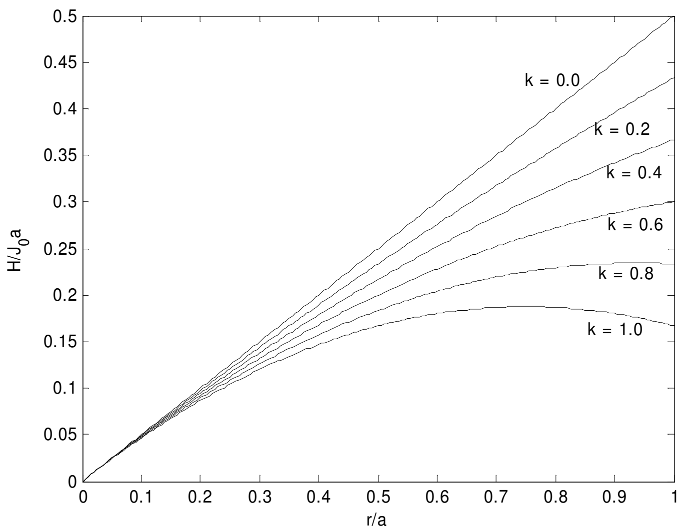
The mean current density is $\bar{J} = \dfrac{2}{a^2}\int_0^a rJ(r)dr$, and the total current is πa^2 times this.

The magnetic field is $H(r) = \dfrac{1}{r}\int_0^r xJ(x)dx$.

Here are the results:

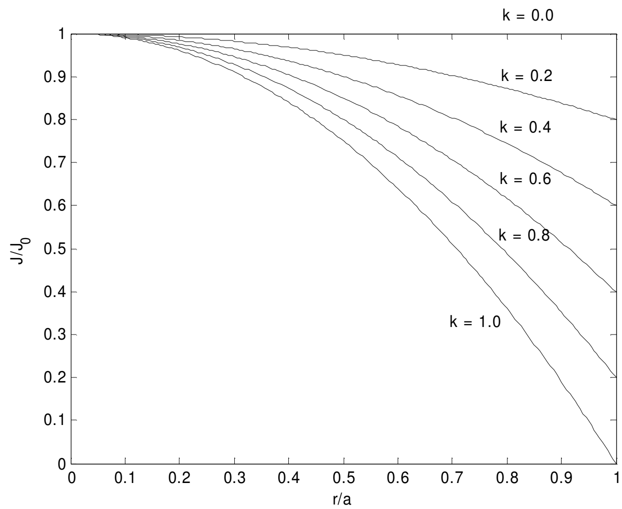
1. $$J = J_0\left(1 - \frac{kr}{a}\right), \quad \bar{J} = J_0\left(1 - \frac{2k}{3}\right), \quad H(r) = J_0\left(\frac{r}{2} - \frac{kr^2}{3a}\right).$$

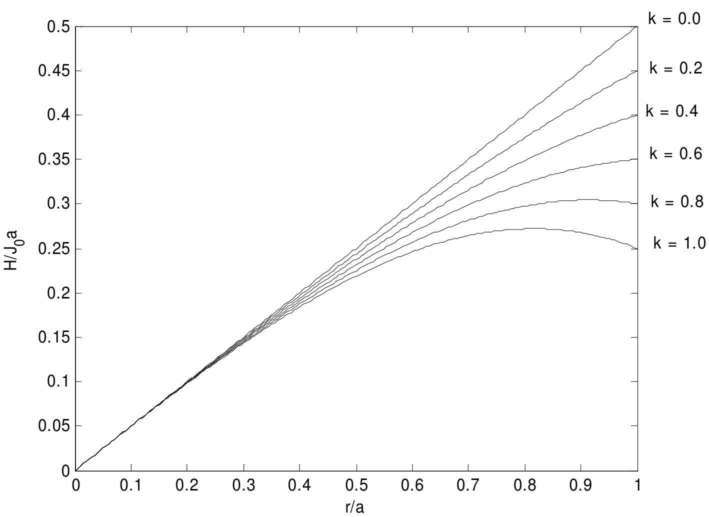




H reaches a maximum value of $\frac{3J_0a}{16k}$ at $r = \frac{3a}{4k}$, but this maximum occurs inside the cylinder only if $k > \frac{3}{4}$.

2. $J = J_0\left(1 - \dfrac{kr^2}{a^2}\right)$, $\bar{J} = J_0\left(1 - \tfrac{1}{2}k\right)$, $H(r) = J_0\left(\dfrac{r}{2} - \dfrac{kr^3}{4a^2}\right)$.



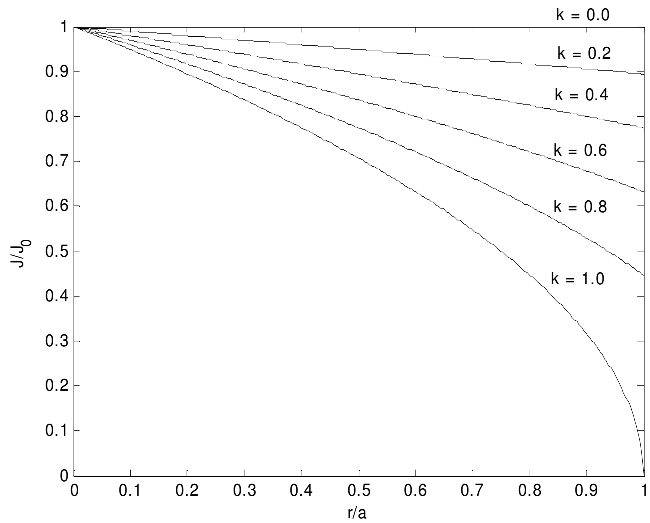


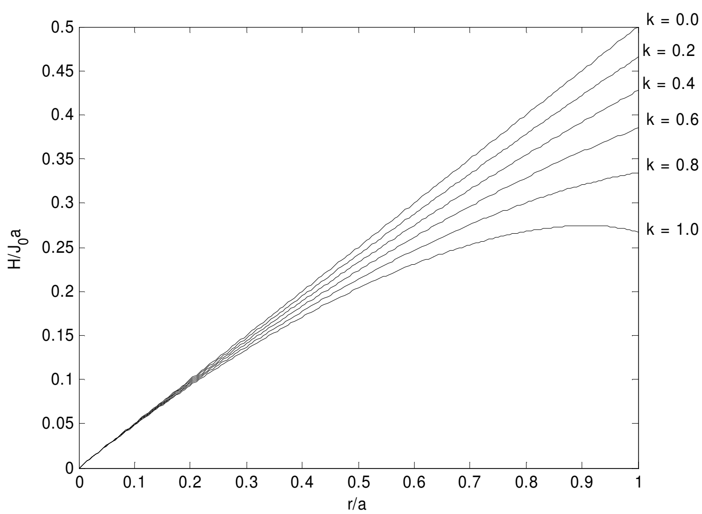
H reaches a maximum value of $\sqrt{\dfrac{2}{27k}}J_0 a$ at $\dfrac{r}{a} = \sqrt{\dfrac{2}{3k}}$, but this maximum occurs inside the cylinder only if $k > \frac{2}{3}$.

3.

$$J = J_0\sqrt{1 - \frac{kr}{a}}\,, \qquad \bar{J} = \frac{J_0}{15k^2}\left[8 - 20(1-k)^{3/2} + 12(1-k)^{5/2}\right],$$

$$H(r) = \frac{2J_0 a^2}{15k^2 r}\left[2 - 5\left(1 - \frac{kr}{a}\right)^{3/2} + 3\left(1 - \frac{kr}{a}\right)^{5/2}\right].$$

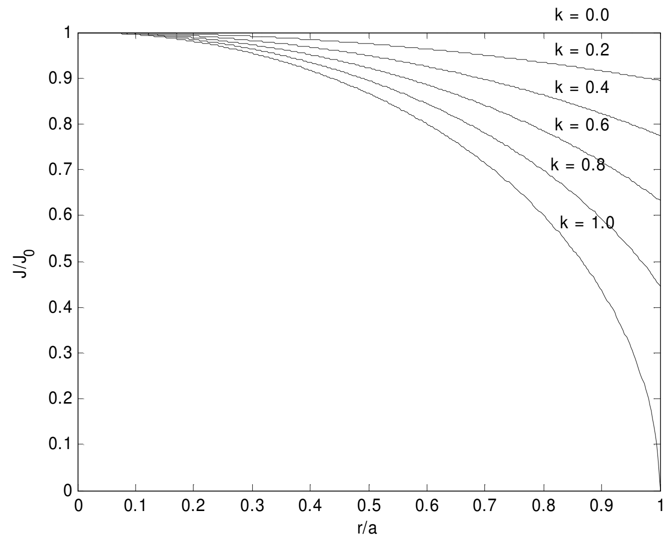


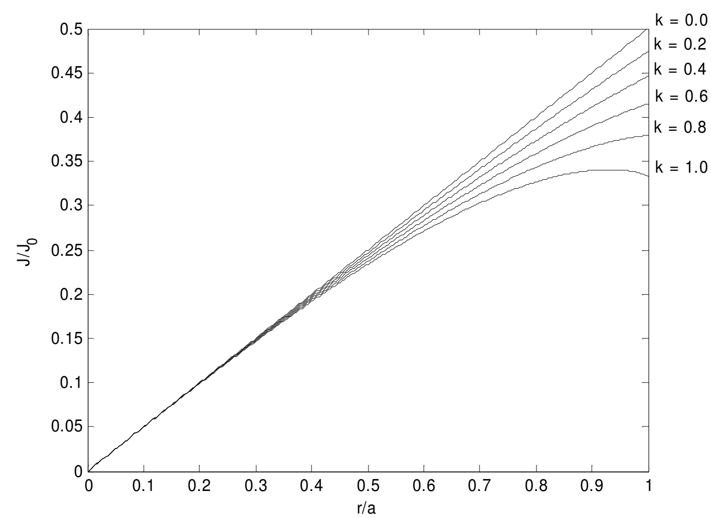


I have not calculated explicit formulas for the positions and values of the maxima. A maximum occurs inside the cylinder if $k > 0.908901$.

4. $$J = J_0\sqrt{1 - \frac{kr^2}{a^2}}\ , \qquad \bar{J} = \frac{2J_0}{3k}\left[1 - (1-k)^{3/2}\right],$$

$$H(r) = \frac{J_0 a^2}{3kr}\left[1 - \left(1 - \frac{kr^2}{a^2}\right)^{3/2}\right].$$





A maximum occurs inside the cylinder if $k > 0.866025$.

In all of these cases, the condition that there shall be no maximum H inside the cylinder – that is, between $r = 0$ and $r = a$ – is that $\frac{J(a)}{\bar{J}} > \frac{1}{2}$. I believe this to be true for *any* axially symmetric current density distribution, though I have not proved it. I expect that a fairly simple proof could be found by someone interested.

Additional current density distributions that readers might like to investigate are

$$J = \frac{J_0}{1 + x/a} \qquad J = \frac{J_0}{1 + x^2/a^2} \qquad J = \frac{J_0}{\sqrt{1 + x/a}}$$

$$J = \frac{J_0}{\sqrt{1 + x^2/a^2}} \qquad J = J_0 e^{-x/a} \qquad J = J_0 e^{-x^2/a^2}$$

CHAPTER 7
FORCE ON A CURRENT IN A MAGNETIC FIELD

7.1 *Introduction*

In Chapter 6 we showed that when an electric current is situated in an external magnetic field it experiences a force at right angles to both the current and the field. Indeed we used this to *define* both the *magnitude* and *direction* of the magnetic field. The magnetic field is defined in magnitude and direction such that the force per unit length $\mathbf{F'}$ on the current is given by

$$\mathbf{F'} = \mathbf{I} \times \mathbf{B}. \qquad 1.7.1$$

7.2 *Force Between Two Current-carrying Wires*

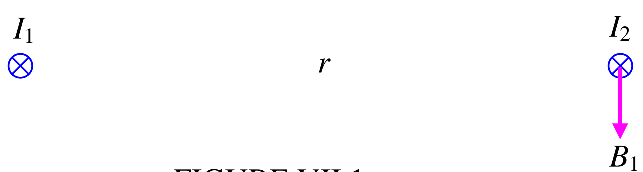


FIGURE VII.1

In figure VII.1, we have two parallel currents, I_1 and I_2, each directed away from you (i.e. into the plane of the paper) and a distance r apart. The current I_1 produces a magnetic field at I_2, directed downward as shown, and of magnitude $B = \mu I_1/(2\pi r)$, where μ is the permeability of the medium in which the two wires are immersed. Therefore, following equation 7.1.1, I_2 experiences a force per unit length towards the left $F' = \mu I_1 I_2/(2\pi r)$. You must also go through the same argument to show that the force per unit length on I_1 from the magnetic field produced by I_2 is of the same magnitude but directed towards the right, thus satisfying Newton's third law of motion.

Thus the force of attraction per unit length between two parallel currents a distance r apart is

$$F' = \frac{\mu I_1 I_2}{2\pi r}. \qquad 7.2.1$$

7.3 *The Permeability of Free Space*

If each of the currents in the arrangement of Section 7.2 is one amp, and if the distance r between to two wires is one metre, and if the experiment is performed in a vacuum, so that $\mu = \mu_0$, then the force per unit length between the two wires is $\mu_0/(2\pi)$ newtons per metre. But we have already (in Chapter 6) *defined the amp* in such a manner that this force is 2×10^{-7} N m^{-1}. Therefore it follows from our definition of the amp that the permeability of free space, by definition, has a value of exactly

$$\mu_0 = 4\pi \times 10^{-7} \text{ T m A}^{-1}, \qquad 7.3.1$$

or, as we shall learn to express it in a later chapter, $4\pi \times 10^{-7}$ henrys per metre, H m^{-1}.

[It was mentioned briefly in Chapters 1 and 6 that there is a proposal , likely to become official in 2015, to re-define the coulomb (and hence the amp) in such a manner that the magnitude of the charge on a single electron is exactly 1.60217×10^{-19} C. If this proposal is passed (as is likely), μ_0 will no longer have a defined value, but will have a measured value of approximately 12.5664×10^{-7} T m A^{-1}.]

7.4 *Magnetic Moment*

If a compass needle, or indeed any bar magnet, is placed in an external magnetic field, it experiences a *torque* – the one exception being if the needle is placed exactly along the direction of the field. The torque is greatest when the needle is oriented at right angles to the field.

Definition. The *magnetic moment* of a magnet is equal to the maximum torque it experiences when in unit magnetic field.

As already noted this maximum torque is experienced when the magnet is at right angles to the magnetic field. In SI units, "unit magnetic field" means, of course, one tesla, and the SI units of magnetic moment are N m T^{-1}, or newton metres per tesla. The reader should look up (or deduce) the dimension of magnetic field (teslas) and then show that the dimensions of magnetic moment are L^2 T^{-1}Q.

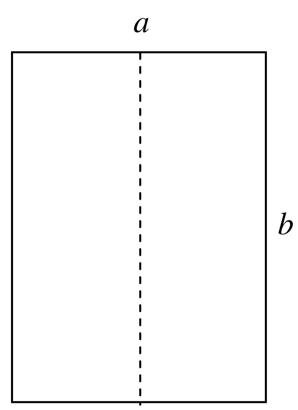
It is noted here that many different definitions of and units for magnetic moment are to be found in the literature, not all of which are correct or even have the correct dimensions. This will be discussed in a later chapter. In the meantime the definition we have given above is standard in the Système International.

7.5 *Magnetic moment of a Plane, Current-carrying Coil*

A plane, current carrying coil also experiences a torque in an external magnetic field, and its behaviour in a magnetic field is quite similar to that of a bar magnet or a compass needle. The torque is maximum when the *normal* to the coil is perpendicular to the magnetic field, and the magnetic moment is defined in exactly the same way, namely the maximum torque experienced in unit magnetic field.

Let us examine the behaviour of a rectangular coil of sides a and b, which is free to rotate about the dashed line shown in figure VII.2.

FIGURE VII.2



In figure VII.3 I am looking down the axis represented by the dashed line in figure VII.2, and we see the coil sideways on. A current I is flowing around the coil in the directions indicated by the symbols $\odot$ and $\otimes$. The normal to the coil makes an angle θ with respect to an external field $\mathbf{B}$.

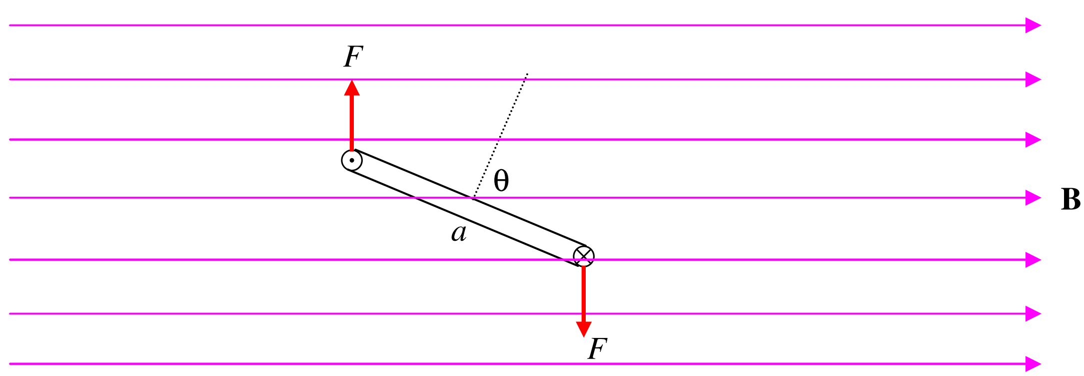


FIGURE VII.3

According to the Biot-Savart law there is a force F on each of the b-length arms of magnitude bIB, or, if there are N turns in the coil, $F = NbIB$. These two forces are opposite in direction and constitute a couple. The perpendicular distance between the two forces is $a \sin \theta$, so the torque τ on the coil is $NabIB \sin \theta$, or $\tau = NAIB \sin \theta$, where A is the area of the coil. This has its greatest value when $\theta = 90^\circ$, and so the magnetic moment of the coil is *NIA*. This shows that, in *SI* units, magnetic moment can equally well be expressed in units of A m^2, or ampere metre squared, which is dimensionally entirely equivalent to N m T^{-1}. Thus we have

$$\tau = p_{\mathrm{m}} B \sin\theta, \tag{7.5.1}$$

where, for a plane current-carrying coil, the magnetic moment is

$$p_m = NIA\,. \tag{7.5.2}$$

This can conveniently be written in vector form:

$$\boldsymbol{\tau} = \mathbf{p_m} \times \mathbf{B}, \tag{7.5.3}$$

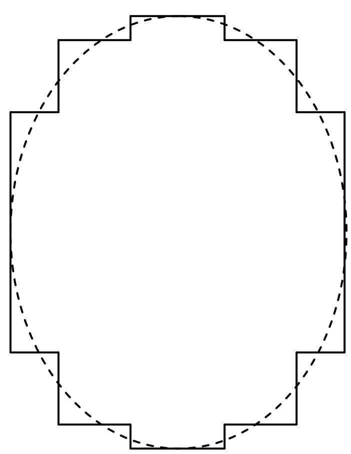
where, for a plane current-carrying coil,

$$\mathbf{p_m} = NI\mathbf{A}. \tag{7.5.4}$$

Here **A** is a vector normal to the plane of the coil, with the current flowing clockwise around it. The vector $\boldsymbol{\tau}$ is directed into the plane of the paper in figure VII.3

The formula *NIA* for the magnetic moment of a plane current-carrying coil is not restricted to rectangular coils, but holds equally for plane coils of any shape; for (see figure VII.4) any curve can be described in terms of an infinite number of infinitesimally small vertical and horizontal segments.

FIGURE VII.4



7.6 *Period of Oscillation of a Magnet or a Coil in an External Magnetic Field*

$$P = 2\pi\sqrt{\frac{I}{p_{\mathrm{m}}B}} \,. \qquad 7.6.1$$

For a derivation of this, see the derivation in Section 3.3 for the period of oscillation of an electric dipole in an electric field. Also, verify that the dimensions of the right hand side of equation 7.6.1 are T. In this equation, what does the symbol *I* stand for?

7.7 *Potential Energy of a Magnet or a Coil in a Magnetic Field*

$$E = \text{constant} - \mathbf{p_m} \bullet \mathbf{B} \,. \qquad 7.7.1$$

For a derivation of this, see the derivation in Section 3.4 for the potential energy of an electric dipole in an electric field. Also, verify that the dimensions of the right hand side of equation 7.7.1 are $\mathrm{ML^2T^{-2}}$ (energy).

7.8 *Moving-coil Ammeter*

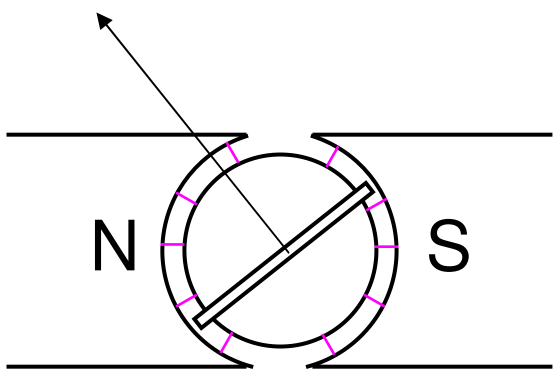


FIGURE VII.5

The current is led into the coil of *N* turns through a spiral spring of torsion constant *c*. The coil is between two poles of a specially-shaped magnet, and there is an iron cylinder inside the coil. This ensures that the magnetic field is everywhere parallel to the plane of the coil; that is, at right angles to its magnetic moment vector. This ensures that the deflection of the coil increases linearly with current, for there is no sin θ factor. When a current flows through the coil, the torque on it is *NABI*, and this in counteracted by the torque $c\theta$ of the holding springs. Thus the current and deflection are related by

$$NABI = c\theta. \qquad 7.8.1$$

7.9 *Magnetogyric Ratio*

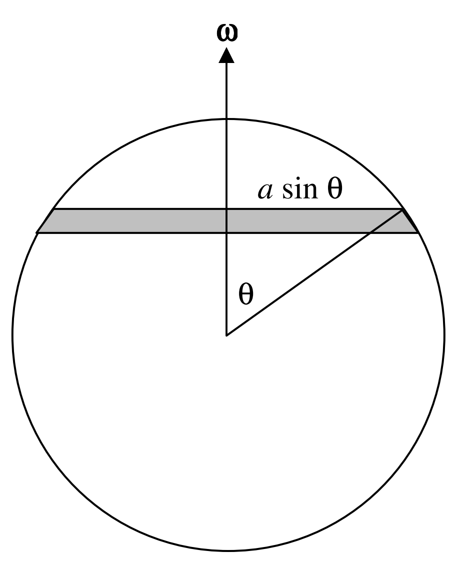
The magnetic moment and the angular momentum are both important properties of subatomic particles. Each of them, however, depends on the angular speed of rotation of the particle. The *ratio* of magnetic moment to angular momentum, on the other hand, is independent of the speed of rotation, and tells us something about how the mass and charge are distributed within the particle. Also, it can be measured with higher precision than either the magnetic moment or the angular momentum separately. This ratio is called the *magnetogyric ratio* (or, perversely and illogically, by some, the "gyromagnetic ratio"). You should be able to show that the dimensions of the magnetogyric ratio are QM^{-1}, and therefore the SI unit is C kg^{-1}. I doubt, however, if many particle physicists use such simple units. They probably express magnetic moment in Bohr magnetons or nuclear magnetons and angular momentum in units of Planck's constant divided by 2π – but that is not our problem.

Let us calculate the magnetogyric ratio of a point charge and point mass moving in a circular orbit – rather like the electron moving around the proton in the simplest model of a hydrogen atom. We'll suppose that the angular speed in the orbit is ω and the radius of the orbit is a. The angular momentum is easy – it is just $ma^2\omega$. The frequency with which the particle (whose charge is Q) passes a given point in its orbit is $\omega/(2\pi)$, so the current is $Q\omega/(2\pi)$. The area of the orbit is πa^2 and so the magnetic moment of the orbiting particle is $\frac{1}{2}Q\omega a^2$. The magnetogyric ratio is therefore $Q/(2m)$.

The magnetogyric ratio will be the same as this in any spinning body in which the distributions of mass density and charge density inside the body are the same. Consider, however, the magnetogyric ratio of a charged, spinning metal sphere. The mass is distributed uniformly throughout the sphere, but the charge all resides on the surface. We may then expect the magnetogyric ratio to be rather larger than $q/(2m)$.

The angular momentum is easy. It is just $\frac{2}{5}ma^2\omega$. Now for the magnetic moment. Refer to figure VII.6.

FIGURE VII.6



The area of the elemental zone shown is $2\pi a^2 \sin\theta\, d\theta$. The area of the entire sphere is $4\pi a^2$, so the charge on the elemental zone is $\frac{1}{2}Q\sin\theta\, d\theta$. The zone is spinning, as is the entire sphere, at an angular speed ω, so the current is

$$\tfrac{1}{2}Q\sin\theta\, d\theta \times \omega/(2\pi) \;=\; \frac{Q\omega\sin\theta\, d\theta}{4\pi}. \qquad 7.9.1$$

The area enclosed by the elemental zone is $\pi a^2 \sin^2\theta$. The magnetic moment dp_m of the zone is the current times the area enclosed, which is

$$dp_m \;=\; \tfrac{1}{4}Q\omega a^2 \sin^3\theta\, d\theta. \qquad 7.9.2$$

The magnetic moment of the entire sphere is found by integrating this from $\theta = 0$ to π, whence

$$p_m \;=\; \tfrac{1}{3}Q\omega a^2. \qquad 7.9.3$$

The ratio of the magnetic moment to the angular momentum is therefore $5Q/(6m)$.

Those who are familiar with gyroscopic motion will know that if a spinning body of angular momentum **L** is subject to a torque **τ**, the angular momentum vector will not be constant in direction and indeed the rate of change of angular momentum will be equal to **τ.** Figure VII.7 is a reminder of the motion of a top in regular precession (that is, with no nutation).

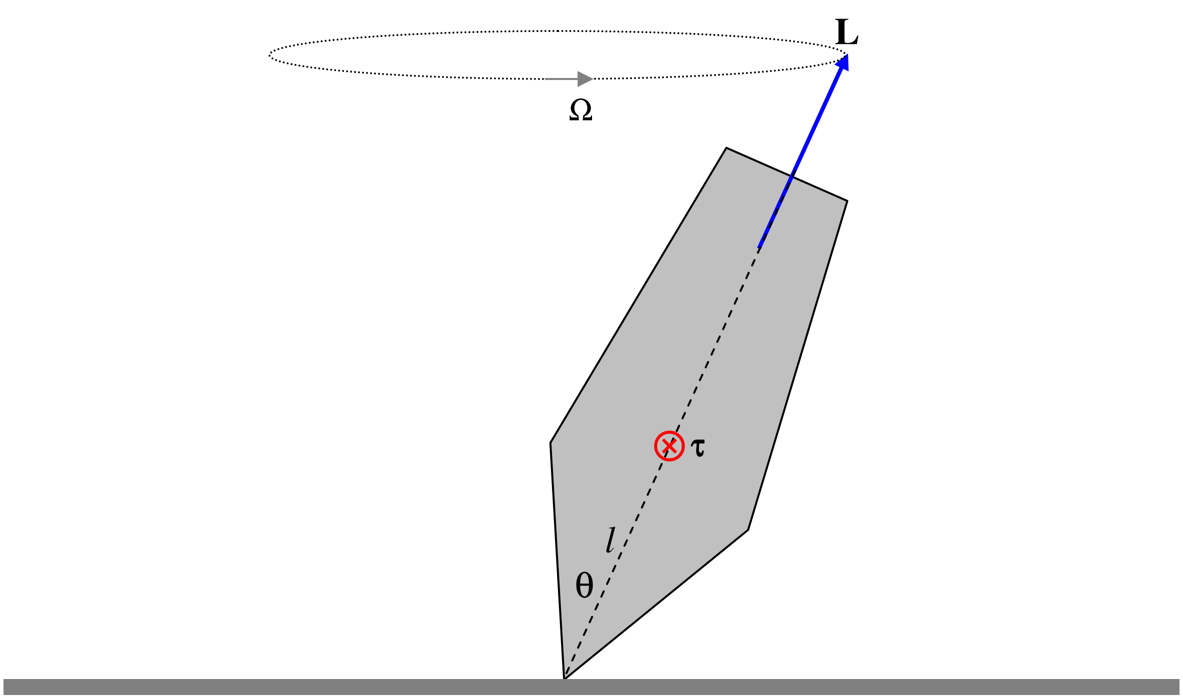


FIGURE VII.7

A study of Chapter 4 Section 4.10 of Classical Mechanics will be needed for a more detailed understanding of the motion of a top. The top is subject to a torque of magnitude $mgl \sin\theta$. The torque can be represented by a vector $\boldsymbol{\tau}$ directed into the plane of the paper. As drawn, the angular momentum vector $\mathbf{L}$ makes an angle θ with the gravitational field $\mathbf{g}$, and it precesses about the vertical with an angular velocity $\boldsymbol{\Omega}$, the three vectors $\boldsymbol{\tau}$, $\mathbf{L}$ and $\boldsymbol{\Omega}$ being related by $\boldsymbol{\tau} = \mathbf{L} \times \boldsymbol{\Omega}$. The magnitude of the angular momentum vector is therefore $\tau/(L \sin \theta)$. But $\tau = mgl \sin \theta$, so that the precessional frequency is mgl/L, independent of θ. Likewise a charged spinning body with a magnetic moment of $\mathbf{p}_\mathrm{m}$ is a magnetic field $\mathbf{B}$ experiences a torque $\boldsymbol{\tau} = \mathbf{p_m} \times \mathbf{B}$, which is of magnitude $p_\mathrm{m}B \sin \theta$, and consequently its angular momentum vector precesses around $\mathbf{B}$ at an angular speed $\dfrac{p_m}{L}B$, independent of θ. (Verify that this has dimensions T^{-1}.) The coefficient of B here is the magnetogyric ratio. The precessional speed can be measured very precisely, and hence the magnetogyric ratio can be measured correspondingly precisely. This phenomenon of "Larmor precession" is the basis of many interesting instruments and disciplines, such as the proton precession magnetometer, nuclear magnetic resonance spectroscopy and nuclear magnetic resonance imaging used in medicine. Because anything including the word "nuclear" is a politically incorrect phrase, the word "nuclear" is usually dropped, and nuclear magnetic resonance imaging is usually called just "magnetic resonance imaging", or MRI, which doesn't quite make sense, but at least is politically correct.

CHAPTER 8
ON THE ELECTRODYNAMICS OF MOVING BODIES

1. *Introduction*

First, I have shamelessly plagiarized the title of this chapter. I have stolen the title from that of one of the most famous physics research papers of the twentieth century – *Zur Elektrodynamik Bewegter Körper*, the paper in which Einstein described the Special Theory of Relativity in 1905. I shall be describing the motion of charged particles in electric and magnetic fields, but, unlike Einstein, I shall (unless I state otherwise – which *will* happen from time to time) be restricting the considerations of this chapter mostly to nonrelativistic speeds – that is to say speeds such that v^2/c^2 is much smaller than the level of precision one is interested in or can conveniently measure. Some relativistic aspects of electrodynamics are touched upon briefly in Chapter 15 of the Classical Mechanics notes in this series, but, apart from the fact that this chapter and Einstein's paper both deal with the motions of charged bodies in electric and magnetic fields, there will be little else in common.

Section 8.2 will deal with the motion of a charged particle in an electric field alone, and Section 8.3 will deal with the motion of a charged particle in a magnetic field alone. Section 8.4 will deal with the motion of a charged particle where both an electric and a magnetic field are present. That section may be a little more difficult than the others and may be omitted on a first reading by less experienced readers. Section 8.5 deals with the motion of a charged particle in a nonuniform magnetic field and is more difficult again.

8.2 *Charged Particle in an Electric Field*

There is really very little that can be said about a charged particle moving at nonrelativistic speeds in an electric field $\mathbf{E}$. The particle, of charge q and mass m, experiences a *force* $q\mathbf{E}$, and consequently it *accelerates* at a rate $q\mathbf{E}/m$. If it starts from rest, you can calculate how fast it is moving in time t, what distance it has travelled in time t, and how fast it is moving after it has covered a distance x, by all the usual first-year equations for uniformly accelerated motion in a straight line. If the charge is accelerated through a potential difference V, its loss of potential energy qV will equal its gain in kinetic energy $\frac{1}{2}mv^2$. Thus $v = \sqrt{2qV/m}$.

Let us calculate, using this nonrelativistic formula, the speed gained by an electron that is accelerated through 1, 10, 100, 1000, 10000, 100,000 and 1,000,000 volts, given that, for an electron, $e/m = 1.7588 \times 10^{11}$ C kg^{-1}. (The symbol for the electronic charge is usually written e. You might note here that that's a lot of coulombs per kilogram!). We'll also calculate v/c and v^2/c^2.

V volts	v m s^{-1}	v/c	v^2/c^2
1	5.931×10^5	1.978×10^{-3}	3.914×10^{-6}
10	1.876×10^6	6.256×10^{-3}	3.914×10^{-5}
100	5.931×10^6	1.978×10^{-2}	3.914×10^{-4}

1000	1.876×10^{7}	6.256×10^{-2}	3.914×10^{-3}
10000	5.931×10^{7}	1.978×10^{-1}	3.914×10^{-2}
100000	1.876×10^{8}	6.256×10^{-1}	3.914×10^{-1}
1000000	5.931×10^{8}	1.978	3.914

We can see that, even working to a modest precision of four significant figures, an electron accelerated through only a few hundred volts is reaching speeds at which v^2/c^2 is not quite negligible, and for less than a million volts, the electron is already apparently moving faster than light! Therefore for large voltages the formulas of special relativity should be used. Those who are familiar with special relativity (i.e. those who have read Chapter 15 of Classical Mechanics!), will understand that the relativistically correct relation between potential and kinetic energy is $qV = (\gamma - 1)m_0c^2$, and will be able to calculate the speeds *correctly* as in the following table. Those who are not familiar with relativity may be a bit lost here, but just take it as a warning that particles such as electrons with a very large charge-to-mass ratio rapidly reach speeds at which relativistic formulas need to be used. These figures are given here merely to give some idea of the magnitude of the potential differences that will accelerate an electron up to speeds where the relativistic formulas must be used.

V volts	v m s^{-1}	v/c	v^2/c^2
1	5.931×10^{5}	1.978×10^{-3}	3.914×10^{-6}
10	1.875×10^{6}	6.256×10^{-3}	3.914×10^{-5}
100	5.930×10^{6}	1.978×10^{-2}	3.912×10^{-4}
1000	1.873×10^{7}	6.247×10^{-2}	3.903×10^{-3}
10000	5.845×10^{7}	1.950×10^{-1}	3.803×10^{-2}
100000	1.644×10^{8}	5.482×10^{-1}	3.005×10^{-1}
1000000	2.821×10^{8}	0.941	0.855

If a charged particle is moving at constant speed in the x-direction, and it encounters a region in which there is an electric field in the y-direction (as in the Thomson e/m experiment, for example) it will accelerate in the y-direction while maintaining its constant speed in the x-direction. Consequently it will move in a parabolic trajectory just like a ball thrown in a uniform gravitational field, and all the familiar analysis of a parabolic trajectory will apply, except that instead of an acceleration g, the acceleration will be q/m.

8.3 *Charged Particle in a Magnetic Field*

We already know that an electric current **I** flowing in a region of space where there exists a magnetic field **B** will experience a force that is at right angles to both **I** and **B**, and the force per unit length, **F'**, is given by

$$\mathbf{F'} = \mathbf{I} \times \mathbf{B}, \qquad 8.3.1$$

and indeed we used this equation to define what we mean by **B**. Equation 8.3.1 is illustrated in figure VIII.1.

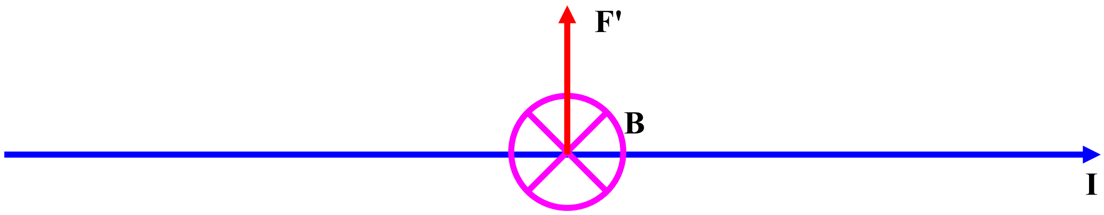


FIGURE VIII.1

The large cross in a circle is intended to indicate a magnetic field directed into the plane of the paper, and **I** and **F'** show the directions of the current and the force.

Now we might consider the current to comprise a stream of particles, n of them per unit length, each bearing a charge q, and moving with velocity $\mathbf{v}$ (speed υ). The current is then $nq\mathbf{v}$, and equation 8.3.1 then shows that the force on each particle is

$$\mathbf{F} = q\,\mathbf{v}\times\mathbf{B}\,. \qquad 8.3.2$$

This, then, is the equation that gives the force on a charged particle moving in a magnetic field, and the force is known as the *Lorentz force*.

It will be noted that there is a force on a charged particle in a magnetic field *only if the particle is moving*, and the force is at right angles to both **v** and **B.**

As to the question: "Who's to say if the particle is moving?" or "moving relative to what?" – that takes us into very deep waters indeed. For an answer, I refer you to the following paper: Einstein, A., *Zur Elektrodynamik Bewegter Körper*, Annalen der Physik **17**, 891 (1905).

Let us suppose that we have a particle, of charge q and mass m, moving with speed υ in the plane of the paper, and that there is a magnetic field **B** directed at right angles to the plane of the paper. (If you are reading this straight off the screen, then read "plane of the screen"!) The particle will experience a force of magnitude $q\upsilon B$ (because **v** and **B** are at right angles to each other), and this force is at right angles to the instantaneous velocity of the particle. Because the force is at right angles to the instantaneous velocity vector, the speed of the particle is unaffected. Its acceleration is constant in magnitude and therefore the particle moves in a *circle*, whose radius is determined by equating the force $q\upsilon B$ to the mass times the centripetal acceleration. That is $q\upsilon B = m\upsilon^2/r$, or

$$r = \frac{m\upsilon}{qB}\,. \qquad 8.3.3$$

If we are looking at the motion of some subatomic particle in a magnetic field, and we have reason to believe that the charge is equal to the electronic charge (or perhaps some small multiple of it), we see that the radius of the circular path tells us the *momentum* of the particle; that is, the product of its mass and speed. Equation 8.3.3 is quite valid for relativistic speeds, except that the mass that appears in the equation is then the relativistic mass, not the rest mass, so that the radius is a slightly more complicated function of speed and rest mass.

If **v** and **B** are not perpendicular to each other, we may resolve **v** into a component υ_1 perpendicular to **B** and a component υ_2 parallel to **B**. The particle will then move in a *helical* path, the radius of the helix being $m\upsilon_2/(qB)$, and the centre of the circle moving at speed υ_2 in the direction of **B**.

The angular speed ω of the particle in its circular path is $\omega = \upsilon / r$, which, in concert with equation 8.3.3, gives

$$\omega = \frac{q B}{m}. \qquad 8.3.4$$

This is called the *cyclotron angular speed* or the *cyclotron angular frequency.* You should verify that its dimensions are T^{-1}.

A *magnetron* is an evacuated cylindrical glass tube with two electrodes inside. One, the negative electrode (cathode) is a wire along the axis of the cylinder. This is surrounded by a hollow cylindrical anode of radius a. A uniform magnetic field is directed parallel to the axis of the cylinder. The cathode is heated (and emits electrons, of charge e and mass m) and a potential difference V is established across the electrodes. The electrons consequently reach a speed given by

$$eV = \tfrac{1}{2} m\upsilon^2. \qquad 8.3.5$$

Because of the magnetic field, they move in arcs of circles. As the magnetic field is increased, the radius of the circles become smaller, and, when the diameter of the circle is equal to the radius a of the anode, no electrons can reach the anode, and the current through the magnetron suddenly drops. This happens when

$$\tfrac{1}{2} a = \frac{m\upsilon}{eB}. \qquad 8.3.6$$

Elimination of υ from equations 8.3.5 and 8.3.6 shows that the current drops to zero when

$$B = \sqrt{\frac{8mV}{ea^2}}. \qquad 8.3.7$$

Those who are skilled in special relativity should try and do this with the relativistic formulas. In equation 8.3.5 the right hand side will have to be $(\gamma - 1)m_0c^2$, and in equation 8.3.6 m will have to be replaced with γm_0. I make the result

$$B = \frac{2\sqrt{2m_0c^2eV + e^2V^2}}{eac}. \qquad 8.3.8$$

For small potential differences, eV is very much less than m_0c^2, and equation 8.3.8 reduces to equation 8.3.5.

8.4 *Charged Particle in an Electric and a Magnetic Field*

The force on a charged particle in an electric and a magnetic field is

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}). \tag{8.4.1}$$

As an example, let us investigate the motion of a charged particle in uniform electric and magnetic fields that are at right angles to each other. Specifically, let us choose axes so that the magnetic field **B** is directed along the positive z-axis and the electric field is directed along the positive y-axis. (Draw this on a large diagram!) Try and imagine what the motion would be like. Suppose, for example, the motion is all in the yz-plane. Perhaps the particle will move round and round in a circle around an axis parallel to the magnetic field, but the centre of this circle will accelerate in the direction of the electric field. Well, you are right in that the particle does move in a circle around an axis parallel to **B**, and also that the centre of the circle does indeed move. But the rest of it isn't quite right. Before embarking on a mathematical analysis, see if you can imagine the motion a bit more accurately.

We'll suppose that at some instant the x, y and z components of the velocity of the particle are u , υ and w. We'll suppose that these velocity components are all nonrelativistic, which means that m is constant and not a function of the speed. The three components of the equation of motion (equation 8.4.1) are then

$$m\dot{u} = qB\upsilon\,, \tag{8.4.2}$$

$$m\dot{\upsilon} = -qBu + qE \tag{8.4.3}$$

and

$$m\dot{w} = 0. \tag{8.4.4}$$

For short, I shall write $q\,B/m = \omega$ (the cyclotron angular speed) and, noting that the dimensions of E/B are the dimensions of speed (verify this!), I shall write $E/B = V_{\rm D}$, where the significance of the subscript D will become apparent in due course. The equations of motion then become

$$\ddot{x} = \dot{u} = \omega\upsilon, \tag{8.4.5}$$

$$\ddot{y} = \dot{\upsilon} = -\omega(u - V_{\rm D}) \tag{8.4.6}$$

and

$$\ddot{z} = \dot{w} = 0. \tag{8.4.7}$$

To find the general solutions to these, we can, for example, let $X = u - V_{\rm D}$. Then equations 8.4.5 and 8.4.6 become $\dot{X} = \omega\upsilon$ and $\dot{\upsilon} = -\omega X$. From these, we obtain $\ddot{X} = -\omega^2 X$. The general solution of this is $X = A\sin(\omega t + \alpha)$, and so $u = A\sin(\omega t + \alpha) + V_\text{D}$. By integration and differentiation with respect to time we can find x and $\ddot{x}$ respectively. Thus we obtain:

$$x = -\frac{A}{\omega}\cos(\omega t + \alpha) + V_\text{D}t + D, \tag{8.4.8}$$

$$u = \dot{x} = A\sin(\omega t+\alpha) + V_\text{D} \tag{8.4.9}$$

and

$$\ddot{x} = A\omega\cos(\omega t+\alpha). \tag{8.4.10}$$

Similarly we can solve for y and z as follows:

$$y = \frac{A}{\omega}\sin(\omega t + \alpha) + F, \tag{8.4.11}$$

$$\upsilon = \dot{y} = A\cos(\omega t+\alpha), \tag{8.4.12}$$

$$\ddot{y} = -A\omega\sin(\omega t+\alpha), \tag{8.4.13}$$

$$z = w_0 t + z_0, \tag{8.4.14}$$

$$w = \dot{z} = w_0 \tag{8.4.15}$$

and

$$\ddot{z} = 0. \tag{8.4.16}$$

There are six arbitrary constants of integration, namely A , D , F , α , z_0 and w_0, whose values depend on the initial conditions (position and velocity at $t = 0$). Of these, z_0 and w_0 are just the initial values of z and w. Let us suppose that these are both zero and that all the motion takes place in the xy-plane.

In these equations A and α always occur in the combinations $A \sin \alpha$ and $A \cos \alpha$, and therefore for convenience I am going to let $A \sin \alpha = S$ and $A \cos \alpha = C$, and I am going to re-write equations 8.4.8, 8.4.9, 8.4.11 and 8.4.12 as

$$x = -\frac{1}{\omega}(C\cos\omega t - S\sin\omega t) + V_\text{D}t + D, \tag{8.4.17}$$

$$u = C\sin\omega t + S\cos\omega t + V_\text{D}, \tag{8.4.18}$$

$$y = \frac{1}{\omega}(C\sin\omega t + S\cos\omega t) + F, \tag{8.4.19}$$

and

$$\upsilon = C\cos\omega t - S\sin\omega t. \tag{8.4.20}$$

Let is suppose that the initial conditions are: at $t = 0$, $x = y = u = \upsilon = 0$. That is, the particle starts from rest at the origin. If the put these initial conditions in equations 8.4.17-20, we find that $C = 0$, $S = -V_{\rm D}$, $D = 0$ and $F = V_{\rm D}/\omega$. Equations 8.4.17 and 8.4.19, which give the equation to the path described by the particle, become

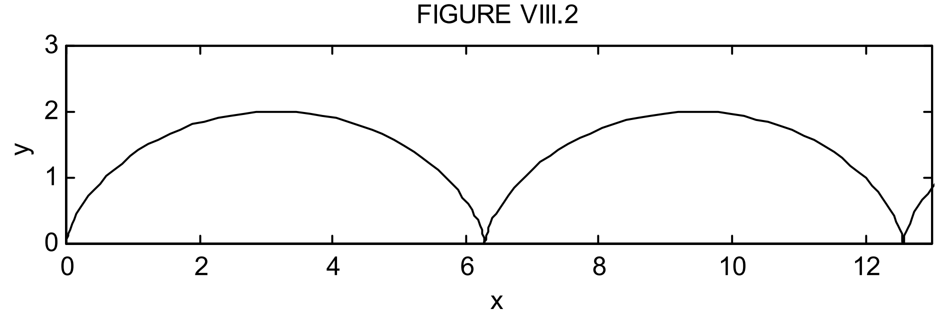
$$x = -\frac{V_{\rm D}}{\omega}\sin\omega t + V_{\rm D}t \qquad 8.4.21$$

and

$$y = \frac{V_{\rm D}}{\omega}(1 - \cos\omega t). \qquad 8.4.22$$

It is worth reminding ourselves here that the cyclotron angular speed is $\omega = qB/m$ and that $V_{\rm D} = E/B$, and therefore $\frac{V_{\rm D}}{\omega} = \frac{mE}{qB^2}$. These equations are the parametric equations of a *cycloid*. (For more on the cycloid, see Chapter 19 of the Classical Mechanics notes in this series.) The motion is a circular motion in which the centre of the circle *drifts* (hence the subscript D) in the x-direction at speed $V_{\rm D}$. The path is shown in figure VIII.2, drawn for distances in units of $\frac{V_{\rm D}}{\omega} = \frac{mE}{qB^2}$.

FIGURE VIII.2



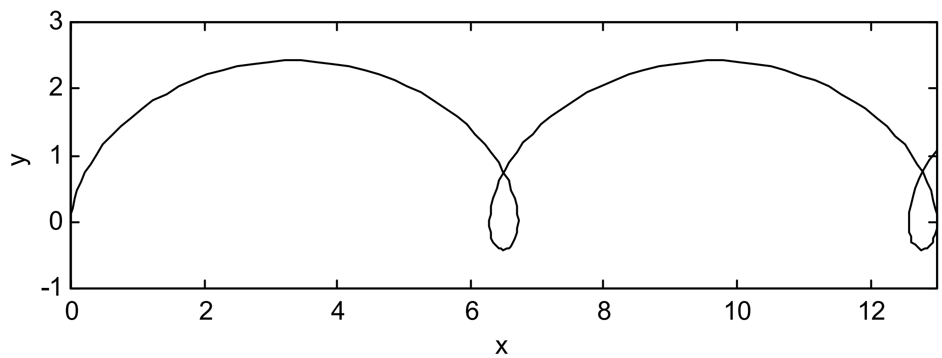
I leave it to the reader to try different initial conditions, such as one of u or v *not* initially zero. You can try with u_0 or v_0 equal to some multiple of fraction of V_D, and you can make the u_0 or v_0 positive or negative. Calculate the values of the constants D, F, C and S and draw the resulting path. You will always get some sort of cycloid. It may not be a simple cycloid as in our example, but it might be an *expanded cycloid* (i.e. small loops instead of cusps) or a *contracted cycloid*, which has neither loops nor cusps, but looks more or less sinusoidal. I'll try just one. I'll let $u_0 = 0$ and $v_0 = +V_D$. If I do that, I get

$$x = \frac{V_D}{\omega}(1 - \cos\omega t - \sin\omega t) + V_D t \qquad 8.4.23$$

and
$$y = \frac{V_D}{\omega}(1 - \cos\omega t + \sin\omega t). \qquad 8.4.24$$

This looks like this:

FIGURE VIII.3



8.5 *Motion in a Nonuniform Magnetic Field*

I give this as a rather more difficult example, not suitable for beginners, just to illustrate how one might calculate the motion of a charged particle in a magnetic field that is not uniform. I am going to suppose that we have an electric current I flowing (in a wire) in the positive z-direction up the z-axis. An electron of mass m and charge of magnitude e (i.e., its charge is $-e$) is wandering around in the vicinity of the current. The current produces a magnetic field, and consequently the electron, when it moves, experiences a Lorenz force. In the following table I write, in cylindrical coordinates, the components of the magnetic field produced by the current, the components of the Lorentz force on the electron, and the expressions in cylindrical coordinates for acceleration component. Some facility in classical mechanics will be needed to follow this.

	Field	Force	Acceleration
ρ	$B_\rho = 0$	$e\dot{z}B_\phi$	$\ddot{\rho} - \rho\dot{\phi}^2$
ϕ	$B_\phi = \dfrac{\mu_0 I}{2\pi\rho}$	0	$\rho\ddot{\phi} + 2\dot{\rho}\dot{\phi}$
z	$B_z = 0$	$-e\dot{\rho}B_\phi$	$\ddot{z}$

From this table we can write down the *equations of motion*, as follows, in which $S_{\rm C}$ is short for $\dfrac{\mu_0 eI}{2\pi m}$. This quantity has the dimensions of *speed* (verify!) and I am going to call it the *characteristic speed.* It has the numerical value 3.5176×10^4 I m s^{-1}, where I is in A. The equations of motion, then, are

Radial: $$\rho(\ddot{\rho} - \rho\dot{\phi}^2) = S_{\rm C}\dot{z} \qquad 8.5.1$$

Transverse (Azimuthal): $$\rho\ddot{\phi} + 2\dot{\rho}\dot{\phi} = 0 \qquad 8.5.2$$

Longitudinal: $$\rho\ddot{z} = -S_{\rm C}\dot{\rho}. \qquad 8.5.3$$

It will be convenient to define dimensionless velocity components:

$$u = \dot{\rho}/S_{\rm C}, \quad \upsilon = \rho\dot{\phi}/S_{\rm C}, \quad w = \dot{z}/S_{\rm C}. \qquad 8.5.4a,b,c$$

Suppose that initially, at time $t = 0$, their values are u_0, υ_0 and w_0, and also that the initial distance of the particle from the current is ρ_0. Further, introduce the dimensionless distance

$$x = \rho/\rho_0, \qquad 8.5.5$$

so that the initial value of x is 1. The initial values of ϕ and z may be taken to be zero by suitable choice of axes.

Integration of equations 8.5.2 and 3, with these initial conditions, yields

$$\dot{z} = S_C\left(w_0 - \ln(\rho/\rho_0)\right) \tag{8.5.6}$$

and

$$\rho^2\dot{\phi} = \rho_0 \upsilon_0 S_C\,; \tag{8.5.7}$$

or, in terms of the dimensionless variables,

$$w = w_0 - \ln x \tag{8.5.8}$$

and

$$\upsilon = \upsilon_0 / x. \tag{8.5.9}$$

We may write $\dot{\rho}\dfrac{d\dot{\rho}}{d\rho}$ for $\ddot{\rho}$ in equation 8.5.1, and substitution for $\dot{z}$ and $\dot{\phi}$ from equations 8.5.6 and 8.5.7 yields

$$u^2 = u_0^2 + \upsilon_0^2(1 - 1/x^2) + 2w_0 \ln x - (\ln x)^2. \tag{8.5.10}$$

Equations 8.5.8,9 and 10 give the velocity components of the electron as a function of its distance from the wire.

Equation 8.5.2 expresses the fact that there is no transverse (azimuthal) force. Its time integral, equation 8.5.7) expresses the consequence that the z-component of its angular momentum is conserved. Further, from equations 8.5.8,9 and 10, we find that

$$u^2 + \upsilon^2 + w^2 = u_0^2 + \upsilon_0^2 + w_0^2 = s^2, \text{ say,} \tag{8.5.11}$$

so that the speed of the electron is constant. This is as expected, since the force on the electron is always perpendicular to its velocity; the point of applicat6ion of the force does not move in the direction of the force, which therefore does no work, so that kinetic energy, and hence speed, is conserved.

The distance of the electron from the wire is bounded below and above. The lower and upper bounds, x_1 and x_2 are found from equation 8.5.10 by putting $u = 0$ and solving for x. Examples of these bounds are shown in the Table VIII.I for a variety of initial conditions.

TABLE VIII.1

BOUNDS OF THE MOTION

$\lvert u_0 \rvert$	$\lvert v_0 \rvert$	$\lvert w_0 \rvert$	x_1	x_2
0	0	–2	0.018	1.000
0	0	–1	0.135	1.000
0	0	0	1.000	1.000
0	0	1	1.000	7.389
0	0	2	1.000	54.598
0	1	–2	0.599	1.000
0	1	–1	1.000	1.000
0	1	0	1.000	2.501
0	1	1	1.000	11.149
0	1	2	1.000	69.132
0	2	–2	1.000	1.845
0	2	–1	1.000	3.137
0	2	0	1.000	7.249
0	2	1	1.000	25.398
0	2	2	1.000	125.009
1	0	–2	0.014	1.266
1	0	–1	0.089	1.513
1	0	0	0.368	2.718
1	0	1	0.661	11.181
1	0	2	0.790	69.135
1	1	–2	0.476	1.412
1	1	–1	0.602	1.919
1	1	0	0.726	4.024
1	1	1	0.809	15.345
1	1	2	0.857	85.581
1	2	–2	0.840	2.420
1	2	–1	0.873	4.052
1	2	0	0.896	9.259
1	2	1	0.912	31.458
1	2	2	0.925	148.409
2	0	–2	0.008	2.290
2	0	–1	0.039	3.442
2	0	0	0.135	7.389
2	0	1	0.291	25.433
2	0	2	0.437	125.014
2	1	–2	0.352	2.654
2	1	–1	0.409	4.212
2	1	0	0.474	9.332
2	1	1	0.542	31.478

$\lvert u_0 \rvert$	$\lvert \upsilon_0 \rvert$	$\lvert w_0 \rvert$	x_1	x_2
2	1	2	0.605	148.412
2	2	−2	0.647	4.183
2	2	−1	0.681	7.297
2	2	0	0.712	16.877
2	2	1	0.740	54.486
2	2	2	0.764	236.061

In analysing the motion in more detail, we can start with some particular initial conditions. One easy case is if $u_0 = \upsilon_0 = w_0 = 0$ – i.e. the electron starts at rest. In that case there will be no forces on it, and it remains at rest for all time. A less trivial initial condition is for $\upsilon_0 = 0$, but the other components not zero. In that case, equation 8.5.7 shows that ϕ is constant for all time. What this means is that the motion all takes place in a plane ϕ = constant, and there is no motion "around" the wire. This is just to be expected, because the ρ-component of the velocity gives rise to a z-component of the Lorenz force, and the z-component of the velocity gives rise to a Lorentz force towards the wire, and there is no component of force "around" (increasing ϕ) the wire. The electron, then, is going to move in the plane ϕ = constant at a constant speed $S = sS_\mathrm{C}$, where $s = \sqrt{u_0^2 + w_0^2}$. (Recall that u and w are dimensionless quantities, being the velocity components *in units of the characteristic speed* S_C.) I am going to coin the words *perineme* and *aponeme* to describe the least and greatest distances of the electrons from the wire – i.e. the bounds of the motion. These bounds can be found by setting $u = 0$ and $\upsilon_0 = 0$ in equation 8.5.10 (where we recall that $x = \rho/\rho_0$ - i.e. the ratio of the radial distance of the electron at some time to its initial radial distance). We obtain

$$\rho = \rho_0 e^{w_0 \pm s} \qquad 8.5.12$$

for the aponeme (upper sign) and perineme (lower sign) distances. From equation 8.5.8 we can deduce that the electron is moving at right angles to the wire (i.e. $w = 0$) when it is at a distance

$$\rho = \rho_0 e^{w_0}. \qquad 8.5.13$$

The form of the trajectory with $\upsilon_0 = 0$ is found by integrating equations 8.5.8 and 8.5.10. It is convenient to start the integration at perineme so that $u_0 = 0$ and $s = w_0$, and the initial value of x $(= \rho/\rho_0)$ is 1. For any other initial conditions, the perineme values of x and ρ can be found from equations 8.5.10 and 8.5.12 respectively. Equations 8.5.10 and 8.5.8 may them be written

$$t = \frac{\rho_0}{S_\mathrm{C}} \int_1^x \frac{dx}{[2s\ln x - (\ln x)^2]^{1/2}} \qquad 8.5.14$$

and $$z = St - \rho_0 \int_1^x \frac{\ln x\, dx}{[2s\ln x - (\ln x)^2]^{1/2}}\,. \qquad 8.5.15$$

There are singularities in the integrands at $x = 1$ and $\ln x = 2s$, and, in order to circumvent this difficulty it is convenient to introduce a variable θ defined by

$$\ln x = s(1 - \sin\theta). \qquad 8.5.16$$

Equations 8.5.14 and 15 then become

$$t = \frac{\rho_0 e^s}{S_C}\int_{\pi/2}^{\theta} e^{-s\sin\theta} d\theta \qquad 8.5.17$$

and $$z = \rho_0 s e^s \int_{\pi/2}^{\theta} \sin\theta\, e^{-s\sin\theta} d\theta. \qquad 8.5.18$$

Examples of these trajectories are shown in figure VIII.4, though I'm afraid you will have to turn your monitor on its side to view it properly. They are drawn for s = 0.25, 0.50, 1.00 and 2.00, where s is the ration of the constant electron speed to the characteristic speed S_C. The wire is supposed to be situated along the z-axis ($\rho = 0$) with the current flowing in the direction of positive z. The electron drifts in the opposite direction to the current. (A positively charged particle would drift in the same direction as the current.) Distances in the figure are expressed in terms of the perineme distance ρ_0. The shape of the path depends only on s (and not on ρ). For no speed does the path have a cusp. The radius of curvature R at any point is given by $R = \rho/s$.

Minima of ρ occur at $\rho = \rho_0$ and $\theta = (4n + 1)\pi/2$, where n is an integer;

Maxima of z occur at $\rho = \rho_0 e^s$ and $\theta = (4n + 2)\pi/2$;

Maxima of ρ occur at $\rho = \rho_0 e^{2s}$ and $\theta = (4n + 3)\pi/2$;

Minima of z occur at $\rho = \rho_0 e^s$ and $\theta = (4n + 4)\pi/2$.

The distance between successive loops and the period of each loop vary rapidly with electron speed, as is illustrated in Table VIII.2. In this table, s is the electron speed in units of the characteristic speed S_C, A_1 is the ratio of aponeme to perineme distance, A_2 is the ratio of interloop distance to perineme distance, A_3 is the ratio of period per loop to ρ_0/S_C, and A_4 is the drift speed in units of the characteristic speed S_C .

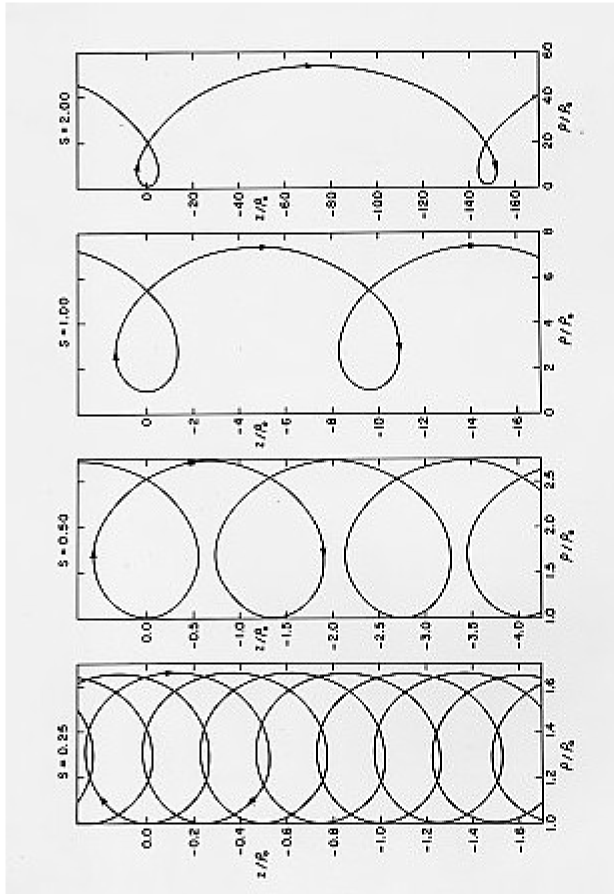


FIGURE IX.4

For example, for a current of 1 A, the characteristic speed is 3.5176×10^{4} m s^{-1}. If an electron is accelerated through 8.7940 V, it will gain a speed of 1.7588 m s^{-1}, which is 50 times the characteristic speed. If the electron starts off at this speed moving in the same direction as the current and 1 Å (10^{-10} m) from it, it will reach a maximum distance of 8.72×10^{10} megaparsecs (1 Mpc = 3.09×10^{22} m) from it, provided the Universe is euclidean. The distance between the loops will be 1.53×10^{12} Mpc, and the period will be 8.60×10^{20} years, after which the electron will have covered, at constant speed, a total distance of 1.55×10^{12} Mpc. The drift speed will be 1.741×10^{6} m s^{-1}.

TABLE VIII.2

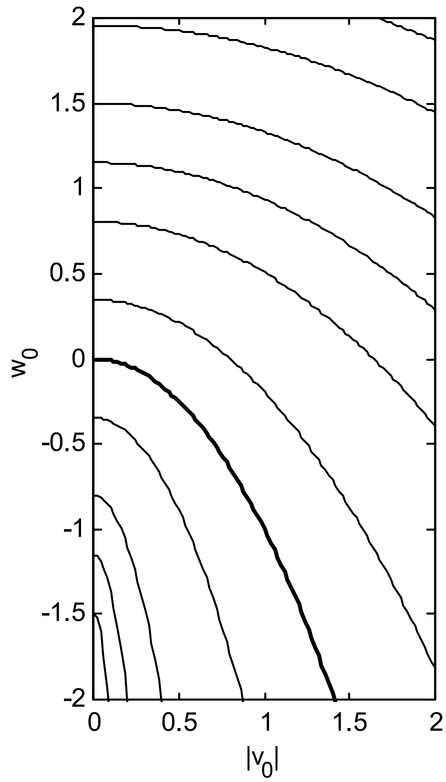
s	A_1	A_2	A_3	A_4
0.1	1.22	3.47×10^{-2}	6.96	4.99×10^{-3}
0.2	1.49	1.54×10^{-1}	7.75	1.99×10^{-2}
0.5	2.72	1.34	11.0	0.122
1.0	7.39	9.65	21.6	0.447
2.0	54.6	1.48×10^{2}	1.06×10^{2}	1.40
5.0	2.20×10^{4}	1.13×10^{5}	2.54×10^{4}	4.49
10.0	4.85×10^{8}	3.70×10^{9}	3.90×10^{8}	9.49
20.0	2.35×10^{17}	2.59×10^{18}	1.33×10^{17}	19.5
50.0	2.69×10^{43}	4.73×10^{44}	9.55×10^{42}	49.5

Let us now turn to consideration of cases where $\upsilon_0 \neq 0$, so that the motion of the electron is not restricted to a plane. At first glance is might be thought that since an azimuthal velocity component gives rise to no additional Lorenz force on the electron, the motion will hardly be affected by a nonzero υ_0, other than perhaps by a revolution around the wire. In particular, for given initial velocity components u_0 and w_0, the perineme and aponeme distances x_1 and x_2 might seem to be independent of υ_0 . Reference to Table VIII.1, however, shows that this is by no means so. The reason is that as the electron moves closer to or further from the wire, the changes in υ made necessary by conservation of the z-component of the angular momentum are compensated for by corresponding changes in u and w made necessary by conservation of kinetic energy.

Since the motion is bounded above and below, there will always be some time when $\dot{\rho} = 0$. There is no loss of generality if we shift the time origin so as to choose $\dot{\rho} = 0$ when $t = 0$ and $x = 1$. From this point, therefore, we shall consider only those trajectories for which $u_0 = 0$. In other words we shall follow the motion from a time $t = 0$ when the electron is at an apsis ($\dot{\rho} = 0$). [The plural of *apsis* is *apsides*. The word *apse* (plural *apses*) is often used in this connection, but it seems useful to maintain a distinction between the architectural term *apse* and the mathematical term *apsis.*] Whether this apsis is perineme (so that $\dot{\rho} = \rho_1$, $\upsilon_0 = \upsilon_1$, $w_0 = w_1$) or aponeme (so that $\dot{\rho} = \rho_2$, $\upsilon_0 = \upsilon_2$, $w_0 = w_2$) depends on the subsequent motion.

The electron starts, then, at a distance from the wire defined by $x = 1$. It is of interest to find the value of x at the next apsis, in terms of the initial velocity components υ_0 and w_0. This is found from equation 8.5.10 with u = 0 and u_0 = 0. The results are shown in figure VIII.4. This figure shows loci of constant next apsis distance, for values of x (going from bottom left to top right of the figure) of 0.05, 0.10, 0.20, 0.50, 1, 2, 5, 10, 20, 50, 100. The heavy curve is for x = 1. It will immediately be seen that, if $w_0 > -\upsilon_0^2$, (above the heavy curve) the value of x at the second apsis is greater than 1. (Recall that υ and w are dimensionless ratios, so there is no problem of dimensional imbalance in the inequality.) The electron was therefore initially at perineme and subsequently moves away from the wire. If on the other hand $w_0 < -\upsilon_0^2$, (below the heavy curve) the value of x at the second apsis is less than 1. The electron was therefore initially at aponeme and subsequently moves closer to the wire.

FIGURE VIII. 5



The case where $w_0 = -\upsilon_0^2$ is of special interest, for them perineme and aponeme distances are equal and indeed the electron stays at a constant distance from the wire at all times. It moves in a helical trajectory drifting in the opposite direction to the direction of the conventional current I. (A positively charged particle would drift in the same direction as I.) The pitch angle α of the helix (i.e. the angle between the instantaneous velocity and a plane normal to the wire) is given by

$$\tan\alpha = -w/\upsilon, \qquad 8.5.19$$

where w and υ are constrained by the equations

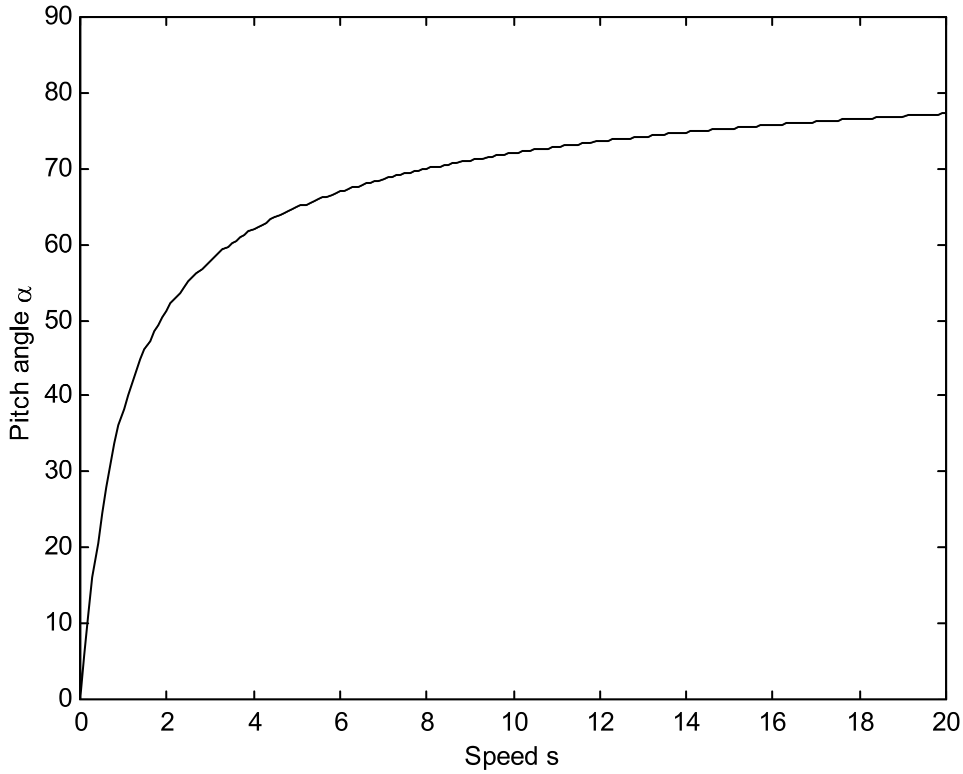
$$w_0 = -\upsilon_0^2 \qquad 8.5.20$$

and

$$\upsilon^2 + w^2 = s^2. \qquad 8.5.21$$

This implies that the pitch angle is determined solely by s, the ratio of the speed S of the electron to the characteristic speed S_C. On other words, the pitch angle is determined by the ratio of the electron speed S to the current I. The variation of pitch angle α with speed s is shown in figure VIII.6. This relation is entirely independent of the radius of the helix.

FIGURE VIII 6



If $w_0 \neq -\upsilon_0^2$, the electron no longer moves in a simple helix, and the motion must be calculated numerically for each case. It is convenient to start the calculation at perineme with initial conditions $u_0 = 0,\ w_0 > -\upsilon_0^2,\ x_0 = 1$. For other initial conditions, the perineme (and aponeme) values of u, υ, w and ρ can easily be found from equations 8.5.10 (with $u_0 = 0$), 8.5.8 and 8.5.9. Starting, then, from perineme, integrations of these equations take the respective forms

$$t = \frac{\rho_0}{S_C}\int_1^x [\upsilon_0^2(1 - 1/x^2) + 2w_0 \ln x - (\ln x)^2]^{-1/2} dx, \qquad 8.5.22$$

$$z = w_0 S_C t - \rho_0 \int_1^x [v_0^2(1 - 1/x^2) + 2w_0 \ln x - (\ln x)^2]^{-1/2} \ln x \, dx \qquad 8.5.23$$

and
$$\phi = v_0 \int_1^x [v_0^2(1 - 1/x^2) + 2w_0 \ln x - (\ln x)^2]^{-1/2} x^{-2} \, dx. \qquad 8.5.24$$

The integration of these equations is not quite trivial and is discussed in the Appendix (Section 8A).

In general the motion of the electron can be described qualitatively roughly as follows. The motion is bounded between two cylinders of radii equal to the perineme and aponeme distances, and the speed is constant. The electron moves around the wire in either a clockwise or a counterclockwise direction, but, once started, the sense of this motion does not change. The angular speed around the wire is greatest at perineme and least at aponeme, being inversely proportional to the square of the distance from the wire. Superimposed on the motion around the wire is a general drift in the opposite direction to that of the conventional current. However, for a brief moment near perineme the electron is temporarily moving in the same direction as the current.

An example of the motion is given in figures VIII.7 and 8 for initial velocity components $u_0 = 0$, $v_0 = w_0 = 1$. The aponeme distance is 11.15 times the perineme distance. The time interval between two perineme passages is 26.47 ρ_0/S_C . The time interval for a complete revolution around the wire ($\phi = 360^\circ$) is 68.05 ρ_0/S_C . In figure VIII.8, the conventional electric current is supposed to be flowing into the plane of the "paper" (computer screen), away from the reader. The portions of the electron trajectory where the electron is moving towards from the reader are drawn as a continuous line, and the brief portions near perineme where the electron is moving away from the reader are indicated by a dotted line. Time marks on the figure are at intervals of $5\rho_0/S_C$.

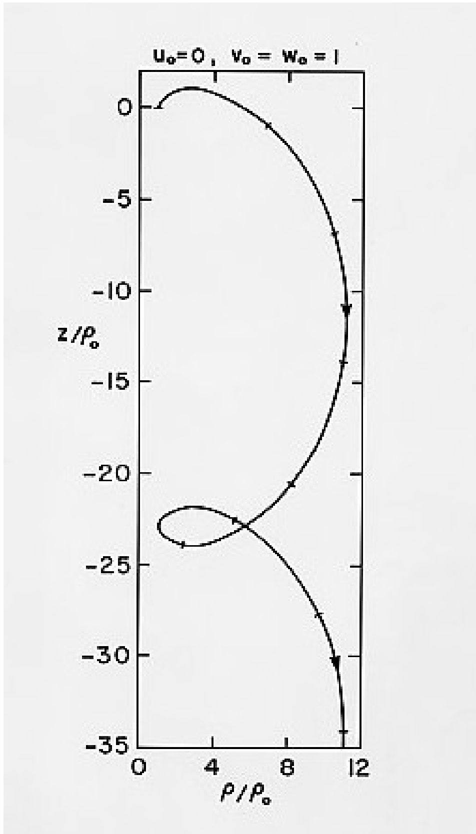


FIGURE VIII.7

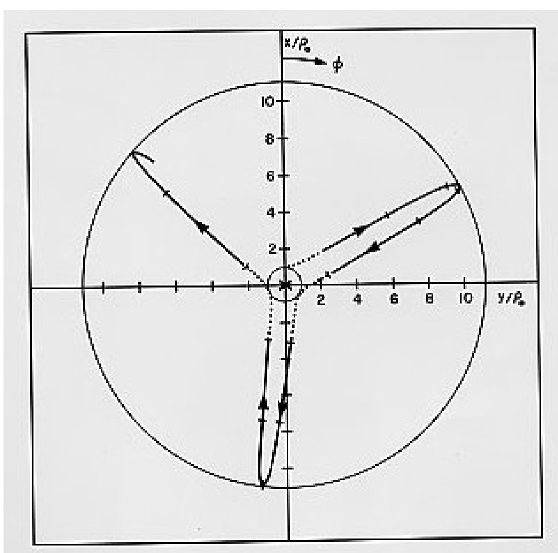


FIGURE VIII.8

8A *Appendix. Integration of the Equations*

Numerical integration of equations 8.5.22-24 is straightforward (by Simpson's rule, for example) except near perineme (x = 1) and aponeme ($x = x_2$), where the integrands become infinite. Near perineme, however, we can substitute $x = 1 + \xi$ and near aponeme we can substitute $x = x_2(1 - \xi)$, and we can expand the integrands as power series in ξ and integrate term by term. I gather here the following results for the intervals $x = 1$ to $x = 1 + \varepsilon$ and $x = x_2 - \varepsilon$ to $x = x_2$, where ε must be chosen to be sufficiently small that ε^4 is smaller than the precision required.

$$I_1 = \int_1^{1+\varepsilon} [\upsilon_0^2(1 - 1/x^2) + 2w_0 \ln x - (\ln x)^2]^{-1/2} dx = M(1 + \tfrac{1}{3}A_1\varepsilon + \tfrac{1}{5}B_1\varepsilon^2 + \tfrac{1}{7}C_1\varepsilon^3 + \ldots) \qquad \text{8A.1}$$

$$I_2 = \int_1^{1+\varepsilon} [\upsilon_0^2(1 - 1/x^2) + 2w_0 \ln x - (\ln x)^2]^{-1/2} \ln x dx = M(\tfrac{1}{3}\varepsilon + \tfrac{1}{5}D_1\varepsilon^2 + \tfrac{1}{7}E_1\varepsilon^3 + \ldots) \qquad \text{8A.2}$$

$$I_3 = \int_1^{1+\varepsilon} [\upsilon_0^2(1 - 1/x^2) + 2w_0 \ln x - (\ln x)^2]^{-1/2} x^{-2} dx = M(1 + \tfrac{1}{3}F_1\varepsilon + \tfrac{1}{5}G_1\varepsilon^2 + \tfrac{1}{7}H_1\varepsilon^3 + \ldots) \qquad \text{8A.3}$$

$$I_4 = \int_{x_2-\varepsilon}^{x_2} [\upsilon_0^2(1 - 1/x^2) + 2w_0 \ln x - (\ln x)^2]^{-1/2} dx = N[1 + \tfrac{1}{3}A_2\varepsilon/x_2 + \tfrac{1}{5}B_2(\varepsilon/x_2)^2 + \tfrac{1}{7}C_2(\varepsilon/x_2)^3 + \ldots) \qquad \text{8A.4}$$

$$I_5 = \int_{x_2-\varepsilon}^{x_2} [\upsilon_0^2(1 - 1/x^2) + 2w_0 \ln x - (\ln x)^2]^{-1/2} \ln x dx$$

$$= I_4 \ln x_2 - N[\tfrac{1}{3}\varepsilon/x_2 + \tfrac{1}{5}D_2(\varepsilon/x_2)^2 + \tfrac{1}{7}E_2(\varepsilon/x_2)^3 + \ldots] \qquad \text{8A.5}$$

$$I_6 = \int_{x_2-\varepsilon}^{x_2} [\upsilon_0^2(1 - 1/x^2) + 2w_0 \ln x - (\ln x)^2]^{-1/2} x^{-2} dx$$

$$= N[1 + \tfrac{1}{3}F_2\varepsilon/x_2 + \tfrac{1}{5}G_2(\varepsilon/x_2)^2 + \tfrac{1}{7}H_2(\varepsilon/x_2)^3 + \ldots]/x_2^2 \qquad \text{8A.6}$$

The constants are defined as follows.

$$M = \left(\frac{2\varepsilon}{\upsilon_0^2 + w_0}\right)^{1/2} \qquad \text{8A.7}$$

$$N = \left(\frac{2\varepsilon x_2}{\ln x_2 - (\upsilon_0/x_2)^2 - w_0}\right)^{1/2} \qquad \text{8A.8}$$

$$a_1 = -\frac{3\upsilon_0^2 + w_0 + 1}{2(\upsilon_0^2 + w_0)} \qquad \text{8A.9}$$

$$b_1 = \frac{4\upsilon_0^2 + \frac{2}{3}w_0 + 1}{2(\upsilon_0^2 + w_0)} \qquad \text{8A.10}$$

$$c_1 = -\frac{5\upsilon_0^2 + \frac{1}{2}w_0 + \frac{11}{12}}{2(\upsilon_0^2 + w_0)} \qquad \text{8A.11}$$

$$a_2 = \frac{3(\upsilon_0/x_2)^2 + w_0 - \ln x_2 + 1}{2\big((\upsilon_0/x_2)^2 + w_0 - \ln x_2\big)} \qquad \text{8A.12}$$

$$b_2 = \frac{4(\upsilon_0/x_2)^2 + \frac{2}{3}w_0 - \ln x_2 + 1}{2\big((\upsilon_0/x_2)^2 + w_0 - \ln x_2\big)} \qquad \text{8A.13}$$

$$c_2 = \frac{5(\upsilon_0/x_2)^2 + \frac{1}{2}w_0 - \frac{1}{2}\ln x_2 + \frac{11}{12}}{2\big((\upsilon_0/x_2)^2 + w_0 - \ln x_2\big)} \qquad \text{8A.14}$$

$$A_n = -\tfrac{1}{2}a_n \qquad \text{8A.15}$$

$$B_n = -\tfrac{1}{2}b_n + \tfrac{3}{8}a_n^2 \qquad \text{8A.16}$$

$$C_n = -\tfrac{1}{2}c_n + \tfrac{3}{4}a_n b_n - \tfrac{5}{16}a_n^3 \qquad \text{8A.17}$$

$$D_n = A_n + \tfrac{1}{2}(-1)^n \qquad \text{8A.18}$$

$$E_n = B_n + \tfrac{1}{2}(-1)^n A_n + \tfrac{1}{3} \qquad \text{8A.19}$$

$$F_n = A_n + 2(-1)^n \qquad \text{8A.20}$$

$$G_n = B_n + 2(-1)^n A_n + 3 \qquad \text{8A.21}$$

$$H_n = C_n + 2(-1)^n B_n + 3A_n + 4(-1)^n \qquad \text{8A.22}$$

$$n = 1, 2$$

CHAPTER 9
MAGNETIC POTENTIAL

9.1 *Introduction*

We are familiar with the idea that an electric field **E** can be expressed as minus the gradient of a potential function V. That is

$$\mathbf{E} = -\mathbf{grad}\, V = -\nabla V. \qquad 9.1.1$$

Note that V is not unique, because an arbitrary constant can be added to it. We can define a unique V by assigning a particular value of V to some point (such as zero at infinity).

Can we express the magnetic field **B** in a similar manner as the gradient of some potential function ψ, so that, for example, $\mathbf{B} = -\mathbf{grad}\,\psi = -\nabla\psi$? Before answering this, we note that there are some differences between **E** and **B** . Unlike **E**, the magnetic field **B** is *sourceless*; there are no sources or sinks; the magnetic field lines are closed loops. The force on a charge q in an electric field is $q\mathbf{E}$, and it depends only on where the charge is in the electric field – i.e. on its position. Thus the force is *conservative*, and we understand from any study of classical mechanics that only conservative forces can be expressed as the derivative of a potential function. The force on a charge q in a *magnetic* field is $q\mathbf{v} \times \mathbf{B}$. This force (the Lorentz force) does not depend only on the position of the particle, but also on its velocity (speed and direction). Thus the force is not conservative. This suggests that perhaps we cannot express the magnetic field merely as the gradient of a scalar potential function – and this is correct; we cannot.

9.2 *The Magnetic Vector Potential*

Although we cannot express the magnetic field as the gradient of a scalar potential function, we shall define a *vector* quantity **A** whose **curl** is equal to the magnetic field:

$$\mathbf{B} = \mathbf{curl}\,\mathbf{A} = \nabla \times \mathbf{A}\,. \qquad 9.2.1$$

Just as $\mathbf{E} = -\nabla V$ does not define V uniquely (because we can add an arbitrary constant to it, so, similarly, equation 9.2.1 does not define **A** uniquely. For, if ψ is some scalar quantity, we can always add $\nabla\psi$ to **A** without affecting **B**, because $\nabla \times \nabla\psi = \mathbf{curl\,grad}\,\psi = 0$.

The vector **A** is called the *magnetic vector potential*. Its dimensions are $\mathrm{MLT^{-1}Q^{-1}}$. Its SI units can be expressed as T m, or Wb $\mathrm{m^{-1}}$ or N $\mathrm{A^{-1}}$.

It might be briefly noted here that some authors define the magnetic vector potential from $\mathbf{H} = \mathbf{curl}\,\mathbf{A}$, though it is standard SI practice to define it from $\mathbf{B} = \mathbf{curl}\,\mathbf{A}$. Systems of units and definitions other than SI will be dealt with in Chapter 16.

Now in electrostatics, we have $\mathbf{E} = \dfrac{1}{4\pi\varepsilon}\dfrac{q}{r^2}\hat{\mathbf{r}}$ for the electric field near a point charge, and, with $\mathbf{E} = -\mathbf{grad}\ V$, we obtain for the potential $V = \dfrac{q}{4\pi\varepsilon r}$. In electromagnetism we have $\mathbf{dB} = \dfrac{\mu I}{4\pi r^2}\hat{\mathbf{r}} \times \mathbf{ds}$ for the contribution to the magnetic field near a circuit element $\mathbf{ds}$. Given that $\mathbf{B} = \mathbf{curl}\ \mathbf{A}$, can we obtain an expression for the magnetic vector potential from the current element? The answer is yes, if we recognize that $\hat{\mathbf{r}}/r^2$ can be written $-\nabla(1/r)$. (If this isn't obvious, go to the expression for $\nabla\psi$ in spherical coordinates, and put $\psi = 1/r$.) The Biot-Savart law becomes

$$\mathbf{dB} = -\frac{\mu I}{4\pi}\nabla(1/r) \times \mathbf{ds} = \frac{\mu I}{4\pi}\,\mathbf{ds} \times \nabla(1/r). \qquad 9.2.3$$

Since $\mathbf{ds}$ is independent of r, the nabla can be moved to the left of the cross product to give

$$\mathbf{dB} = \nabla \times \frac{\mu I}{4\pi r}\mathbf{ds}. \qquad 9.2.4$$

The expression $\dfrac{\mu I}{4\pi r}\mathbf{ds}$, then, is the contribution $\mathbf{dA}$ to the magnetic vector potential from the circuit element $\mathbf{ds}$. Of course an isolated circuit element cannot exist by itself, so, for the magnetic vector potential from a complete circuit, the line integral of this must be calculated around the circuit.

9.3 *Long, Straight, Current-carrying Conductor*

By way of example, let us use the expression $\mathbf{dA} = \dfrac{\mu I}{4\pi r}\mathbf{ds}$, to calculate the magnetic vector potential in the vicinity of a long, straight, current-carrying conductor ("wire" for short!). We'll suppose that the wire lies along the z-axis, with the current flowing in the direction of positive z. We'll work in cylindrical coordinates, and the symbols $\hat{\rho}, \hat{\phi}, \hat{\mathbf{z}}$ will denote the unit orthogonal vectors. After we have calculated $\mathbf{A}$, we'll try and calculate its **curl** to give us the magnetic field $\mathbf{B}$. We already know, of course, that for a straight wire the field is $\mathbf{B} = \dfrac{\mu I}{2\pi\rho}\hat{\phi}$, so this will serve as a check on our algebra.

Consider an element $\hat{\mathbf{z}}\,dz$ on the wire at a height z above the xy-plane. (The length of this element is dz; the unit vector $\hat{\mathbf{z}}$ just indicates its direction.) Consider also a point P in the xy-plane at a distance ρ from the wire. The distance of P from the element dz is $\sqrt{\rho^2 + z^2}$. The contribution to the magnetic vector potential is therefore

$$\mathbf{dA} = \hat{\mathbf{z}}\frac{\mu I}{4\pi}\cdot\frac{dz}{(\rho^2+z^2)^{1/2}}. \qquad 9.3.1$$

The total magnetic vector potential is therefore

$$\mathbf{A} = \hat{\mathbf{z}}\frac{\mu I}{2\pi}\cdot\int_0^\infty\frac{dz}{(\rho^2+z^2)^{1/2}}. \qquad 9.3.2$$

This integral is infinite, which at first may appear to be puzzling. Let us therefore first calculate the magnetic vector potential for a finite section of length $2l$ of the wire. For this section, we have

$$\mathbf{A} = \hat{\mathbf{z}}\frac{\mu I}{2\pi}\cdot\int_0^l\frac{dz}{(\rho^2+z^2)^{1/2}}. \qquad 9.3.3$$

To integrate this, let $z = \rho\tan\theta$, whence $\mathbf{A} = \hat{\mathbf{z}}\frac{\mu I}{2\pi}\cdot\int_0^\alpha\sec\theta d\theta$, where $l = \rho\tan\alpha$. From this we obtain $\mathbf{A} = \hat{\mathbf{z}}\frac{\mu I}{2\pi}\cdot\ln(\sec\alpha+\tan\alpha)$, whence

$$\mathbf{A} = \hat{\mathbf{z}}\frac{\mu I}{2\pi}\cdot\ln\left(\frac{\sqrt{l^2+\rho^2}+l}{\rho}\right). \qquad 9.3.4$$

For $l \gg \rho$ this becomes

$$\mathbf{A} = \hat{\mathbf{z}}\frac{\mu I}{2\pi}\cdot\ln\left(\frac{2l}{\rho}\right) = \hat{\mathbf{z}}\frac{\mu I}{2\pi}(\ln 2l - \ln\rho). \qquad 9.3.5$$

Thus we see that the magnetic vector potential in the vicinity of a straight wire is a vector field parallel to the wire. If the wire is of infinite length, the magnetic vector potential is infinite. For a finite length, the potential is given exactly by equation 9.3.4, and, very close to a long wire, the potential is given approximately by equation 9.3.5.

Now let us use equation 9.3.5 together with $\mathbf{B} = \mathbf{curl\ A}$, to see if we can find the magnetic field $\mathbf{B}$. We'll have to use the expression for **curl A** in cylindrical coordinates, which is

$$\mathbf{curl A} = \left(\frac{1}{\rho}\frac{\partial A_z}{\partial\phi} - \frac{\partial A_\phi}{\partial z}\right)\hat{\boldsymbol{\rho}} + \left(\frac{\partial A_\rho}{\partial z} - \frac{\partial A_z}{\partial\rho}\right)\hat{\boldsymbol{\phi}} + \frac{1}{\rho}\left(A_\phi+\rho\frac{\partial A_\phi}{\partial\rho}-\frac{\partial A_\rho}{\partial\phi}\right)\hat{\mathbf{z}}. \qquad 9.3.6$$

In our case, $\mathbf{A}$ has only a z-component, so this is much simplified:

$$\mathbf{curl A} = \frac{1}{\rho}\frac{\partial A_z}{\partial\phi}\hat{\boldsymbol{\rho}} - \frac{\partial A_z}{\partial\rho}\hat{\boldsymbol{\phi}}. \qquad 9.3.7$$

And since the z-component of **A** depends only on ρ, the calculation becomes trivial, and we obtain, as expected

$$\mathbf{B} = \frac{\mu I}{2\pi\rho}\hat{\boldsymbol{\phi}}. \qquad 9.3.8$$

This is an approximate result for very close to a long wire – but it is exact for any distance for an infinite wire. This may strike you as a long palaver to derive equation 9.3.8 – but the object of the exercise was not to derive equation 9.3.8 (which is trivial from Ampère's theorem), but to derive the expression for **A**. Calculating **B** subsequently was only to reassure us that our algebra was correct.

9.4 *Long Solenoid*

Let us place an infinitely long solenoid of n turns per unit length so that its axis coincides with the z-axis of coordinates, and the current I flows in the sense of increasing ϕ. In that case, we already know that the field inside the solenoid is uniform and is $\mu n I \hat{\mathbf{z}}$ inside the solenoid and zero outside. Since the field has only a z component, the vector potential **A** cam have only a ϕ-component.

We'll suppose that the radius of the solenoid is a. Now consider a circle of radius r (less than a) perpendicular to the axis of the solenoid (and hence to the field **B**). The magnetic flux through this circle (i.e. the surface integral of **B** across the circle) is $\pi r^2 B = \pi r^2 nI$. Now, as everybody knows, the surface integral of a vector field across a closed curve is equal to the line integral of its **curl** around the curve, and this is equal to $2\pi r A_\phi$. Thus, inside the solenoid the vector potential is

$$\mathbf{A} = \tfrac{1}{2}\mu n r I \hat{\boldsymbol{\phi}}. \qquad 9.4.1$$

It is left to the reader to argue that, outside the solenoid ($r > a$), the magnetic vector potential is

$$\mathbf{A} = \frac{\mu n a^2 I}{2r}\hat{\boldsymbol{\phi}}. \qquad 9.4.2$$

9.5 *Divergence*

Like the magnetic field itself, the lines of magnetic vector potential form closed loops (except in the case of the infinitely long straight conducting wire, in which case they are infinitely long straight lines). That is to say **A** has no sources or sinks, or, in other words, its divergence is everywhere zero:

$$\operatorname{div} \mathbf{A} = 0. \qquad 9.5.1$$

CHAPTER 10
ELECTROMAGNETIC INDUCTION

10.1 *Introduction*

In 1820, Oersted had shown that an electric current generates a magnetic field. But can a magnetic field generate an electric current? This was answered almost simultaneously and independently in 1831 by Joseph Henry in the United States and Michael Faraday in Great Britain. Faraday constructed an iron ring, about six inches in diameter. He wound two coils of wire tightly around the ring; one coil around one half (semicircle) of the ring, and the second coil around the second half of the ring. The two coils were not connected to one another other than by sharing the same iron core. One coil (which I'll refer to as the "primary" coil") was connected to a battery; the other coil (which I'll refer to as the "secondary" coil) was connected to a galvanometer. When the battery was connected to the primary coil a current, of course, flowed through the primary coil. This current generated a magnetic field throughout the iron core, so that there was a magnetic field inside each of the two coils. As long as the current in the primary coil remained constant, there was no current in the secondary coil. What Faraday observed was that at the instant when the battery was connected to the primary, and during that brief moment when the current in the primary was rising from zero, a current momentarily flowed in the secondary – but only while the current in the primary was changing. When the battery was disconnected, and during the brief moment when the primary current was falling to zero, again a current flowed in the secondary (but in the opposite direction to previously). Of course, while the primary current was changing, the magnetic field in the iron core was changing, and Faraday recognized that a current was generated in the secondary *while the magnetic flux through it was changing.* The strength of the current depended on the resistance of the secondary, so it is perhaps more fundamental to note that when the magnetic flux through a circuit changes, an electromotive force (EMF) is generated in the circuit, and the faster the flux changes, the greater the induced EMF. Quantitative measurements have long established that:

While the magnetic flux through a circuit is changing, an EMF is generated in the circuit which is equal to the rate of change of magnetic flux $\dot{\Phi}_B$ through the circuit.

This is generally called "**Faraday's Law** of Electromagnetic Induction". A complete statement of the laws of electromagnetic induction must also tell us the *direction* of the induced EMF, and this is generally given in a second statement usually known as "Lenz's Law of Electromagnetic Induction", which we shall describe in Section 10.2. When asked, therefore, for the laws of electromagnetic induction, both laws must be given: Faraday's, which deals with the magnitude of the EMF, and Lenz's, which deals with its direction.

You will note that the statement of Faraday's Law given above, says that the induced EMF is not merely "proportional" to the rate of change of magnetic B-flux, but is *equal* to it. You will therefore want to refer to the dimensions of electromotive force (SI unit: volt) and of *B*-flux (SI unit: weber) and verify that $\dot{\Phi}_B$ is indeed dimensionally similar to EMF. This alone does not tell you the constant of proportionality between the induced EMF and $\dot{\Phi}_B$, though the constant is in fact unity, as stated in Faraday's law. You may then ask: Is this value of 1 for the constant of proportionality between the EMF and $\dot{\Phi}_B$ an experimental value (and, if so, how close to 1 is it, and what is its currently determined best value), or is it expected theoretically to be exactly 1? Well, I suppose it has to be admitted that physics is an experimental science, so that from that point of view the constant has to be determined experimentally. But I shall advance an argument shortly to show not only that you would expect it to be exactly 1, but that the very phenomenon of electromagnetic induction is only to be expected from what we already knew (before embarking upon this chapter) about electricity and magnetism.

Incidentally, we recall that the SI unit for Φ_B is the weber (Wb). To some, this is not a very familiar unit and some therefore prefer to express Φ_B in T m^2. Yet again, consideration of Faraday's law tells us that a perfectly legitimate SI unit (which many prefer) for Φ_B is V s.

10.2 *Electromagnetic Induction and the Lorentz Force*

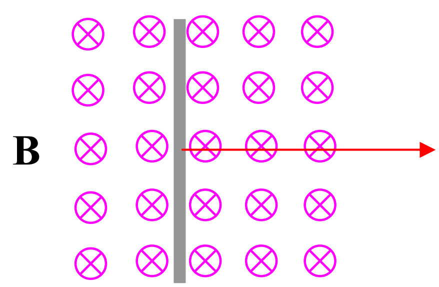


FIGURE X.1

Imagine that there is a uniform magnetic field directed into the plane of the paper (or your computer screen), as in figure X.1. Suppose there is a metal rod, as in the figure, and that the rod is being moved steadily to the right. We know that, within the metal, there are many free conduction electrons, not attached to any particular atom, but free to wander about inside the metal. As the metal rod is moved to the right, these free conduction electrons are also moving to the right and therefore they experience a Lorentz $q\ \mathbf{v} \times \mathbf{B}$ force, which moves them down (remember – electrons are negatively charged) towards the bottom end of the rod. Thus the movement of the rod through the magnetic field induces a potential difference across the ends of the rod. We have achieved electromagnetic induction, and, seen this way, there is nothing new: electromagnetic induction is nothing more than the Lorentz force on the conduction electrons within the metal.

You may speculate that, as an aircraft flies through Earth's magnetic field, a potential difference will be induced across the wingtips. You might try to imagine how you might set up an experiment to detect or measure this. You might also speculate that, as seawater flows up the English Channel, a potential difference is induced between England and France. You might also ask yourself: What if the rod were stationary, and the magnetic field were moving to the left? That's an interesting discussion for lunchtime: Can you imagine the magnetic field moving to the left? Who's to say whether the rod or the field is moving?

If we were somehow to connect the ends of the rod in figure X.1 to a closed circuit, we might cause a current to flow – and we would then have made an electric generator. Look at figure X.2.

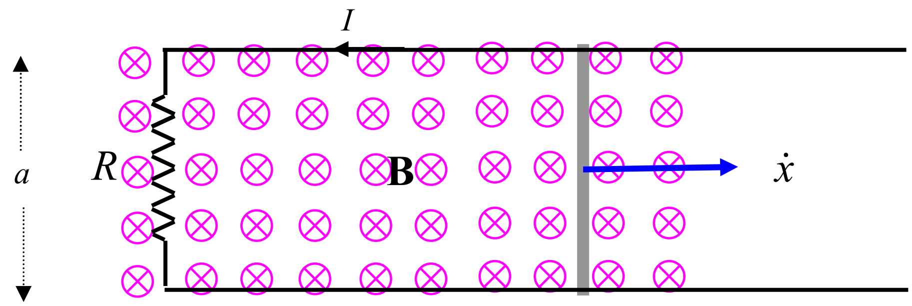


FIGURE X.2

We imagine that our metal bar is being pulled steadily to the right at speed v, and that it is in contact with, and sliding smoothly without friction upon, two rails a distance a apart, and that the rails are connected via a resistance R. As a consequence, a current I flows in the circuit in the direction shown, counterclockwise. (The current is, of course, made up of negative conduction electrons moving clockwise.) Now the magnetic field will exert a force on the current in the rod. The force on the rod will be $a\,\mathbf{I} \times \mathbf{B}$; that is aIB acting to the left. In order to keep the rod moving steadily at speed v to the right against this force, *work* will have to be done at a rate $aIBv$. The work will be dissipated in the resistance at a rate $I\ V$ where V is the induced EMF. Therefore the induced EMF is Bav. But av is the rate at which the area of the circuit is increasing, and Bav is the rate at which the magnetic B-flux through the circuit is increasing. Therefore the induced EMF is *equal to* the rate of change of magnetic flux through the circuit. Thus we have predicted Faraday's law quantitatively merely from what we already know about the forces on currents and charged particles in a magnetic field.

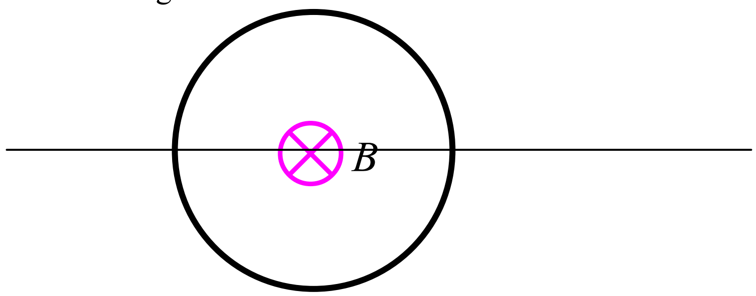
10.3 *Lenz's Law*

We can now address ourselves to the *direction* of the induced EMF. From our knowledge of the Lorentz force $q\,\mathbf{v} \times \mathbf{B}$ we see that the current flows counterclockwise, and that this results in a force on the rod that is in the opposite direction to its motion. But, even if we did not know this law, or had forgotten the formula, or if we didn't understand a vector product, we could see that this must be so. For, suppose that we move the rod to the right, and that, as a consequence, there will be a force also the right. Then the rod moves faster, and the force to the right is greater, and the rod moves yet faster, and so on. The rod would accelerate indefinitely, for the expenditure of no work. No – this cannot be right. The direction of the induced EMF must be such as to *oppose* the change of flux that causes it. This is merely a consequence of conservation of energy, and it can be stated as **Lenz's Law**:

When an EMF is induced in a circuit as a result of changing magnetic flux through the circuit, the direction of the induced EMF is such as to *oppose* the change of flux that causes it.

In our example of Section 10.2, we increased the magnetic flux through a circuit by increasing the area of the circuit. There are other ways of changing the flux through a circuit. For example, in figure X.3, we have a circular wire and a magnetic field perpendicular to the plane of the circle, directed into the plane of the drawing.

FIGURE X.3



We could increase the magnetic flux through the coil by increasing the strength of the field rather than by increasing the area of the coil. The rate of increase of the flux would then be $A\dot{B}$ rather than $\dot{A}B$. We could imagine increasing B, for example by moving a magnet closer to the coil, or by moving the coil into a region where the magnetic field was stronger; or, if the magnetic field is generated by an electromagnet somewhere, by increasing the current in the electromagnet. One way or another, we increase the strength of the field through the coil. An EMF is generated in the coil equal to the rate of change of magnetic flux, and consequently a current flows in the coil. In which direction does this induced current flow? It flows in such a direction as to *oppose* the increase in B that causes it. That is, the current flows counterclockwise in the coil. If this were not so, and the induced current were clockwise, this would still further increase the flux through the coil, and the current would increase further, and the flux would increase further, and so on. A runaway increase in the current and the field would result, and energy would not be conserved.

If we were in *decrease* the strength of the field through the coil, a current would flow clockwise in the coil – i.e. in such a sense as to tend to increase the field – i.e. to oppose the decrease in field that we are trying to impose. It may well occur to you at this stage that it is impossible to increase the current in a circuit instantaneously, and it takes a finite time to establish a new level of current. This is correct – a point to which we shall return later, when indeed we shall calculate just how long it does take.

Another way in which we could change the magnetic flux through a coil would be to *rotate* the coil in a magnetic field. For example, in figure X.4a, we see a magnetic field directed to the right, and a coil whose normal is perpendicular to the field. There is no magnetic flux through the coil. If we now rotate the coil, as in figure X.4b, the flux through the coil will increase, an EMF will be induced in the coil, equal to the rate of increase of flux, and a current will flow. The current will flow in a direction such that the magnetic moment of the coil will be as shown, which will result in an opposition to our imposed rotation on the coil, and the current will flow in the direction indicated by the symbols ⊙ and ⊗.

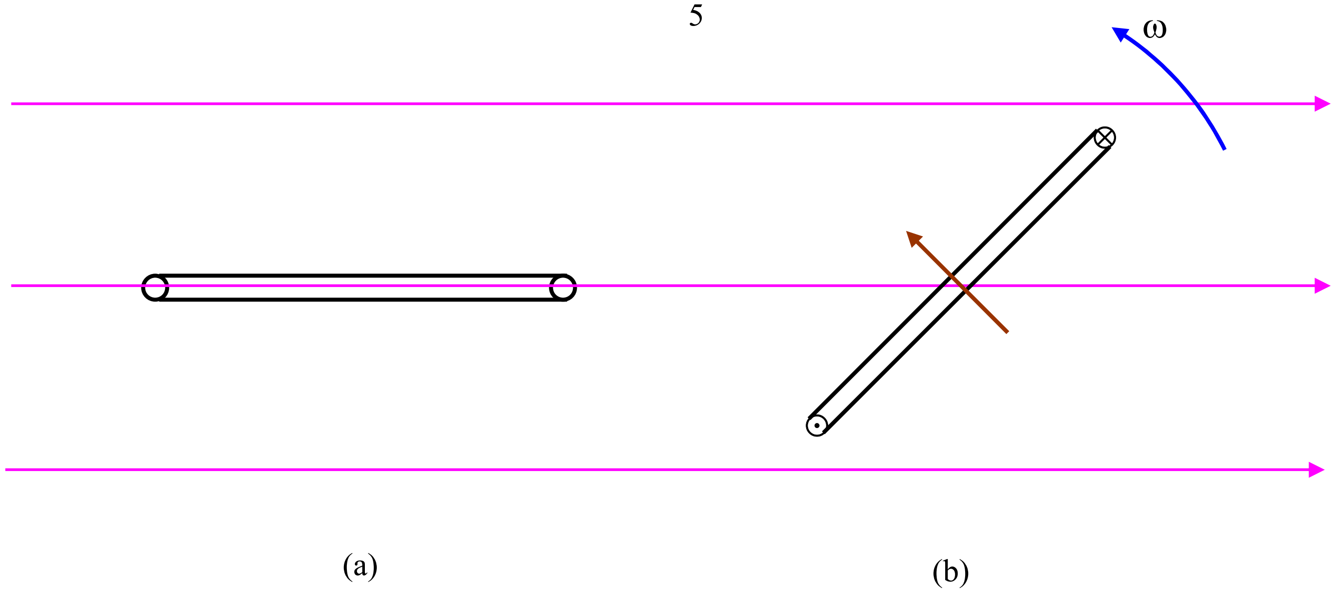


FIGURE X.4

If the flux through a coil changes at a rate $\dot{\Phi}_B$, and if the coil is not just a single turn but is made of N turns, the induced EMF will be $\dot{\Phi}_B$ *per turn*, so that the induced EMF in the coil as a whole will be $N\dot{\Phi}_B$.

10.4 *Ballistic Galvanometer and the Measurement of Magnetic Field*

A galvanometer is similar to a sensitive ammeter, differing mainly in that when no current passes through the meter, the needle is in the middle of the dial rather than at the left hand end. A galvanometer is used not so much to *measure* a current, but rather to *detect* whether or not a current is flowing, and in which direction. In the *ballistic* galvanometer, the motion of the needle is undamped, or as close to undamped as can easily be achieved. If a small quantity of electricity is passed through the ballistic galvanometer in a time that is short compared with the period of oscillation of the needle, the needle will jerk from its rest position, and then swing to and fro in lightly damped harmonic motion. (It would be simple harmonic motion if it could be completely undamped.) The amplitude of the motion, or rather the extent of the first swing, depends on the quantity of electricity that was passed through the galvanometer. It could be calibrated, for example, by discharging various capacitors through it, and making a table or graph of amplitude of swing versus quantity of electricity passed.

Now, if we have a small coil of area A, N turns, resistance R, we could place the coil perpendicular to a magnetic field B, and then connect the coil to a ballistic galvanometer. Then, suddenly (in a time that is short compared with the oscillation period of the galvanometer), remove the coil from the field (or rotate it through 90°) so that the flux through the coil goes from AB to zero. While the flux through the coil is changing, and EMF will be induced, equal to $NA\dot{B}$, and consequently a current will flow momentarily through the coil of magnitude

$$I = NA\dot{B}/(R + r), \qquad 10.4.1$$

where r is the resistance of the galvanometer. Integrate this with respect to time, with initial condition $Q = 0$ when $t = 0$, and we find for the total quantity of electricity that flows through the galvanometer

$$Q = NAB/(R + r). \qquad 10.4.2$$

Since Q can be measured from the amplitude of the galvanometer motion, the strength of the magnetic field, B is determined.

I mentioned that the ballistic galvanometer differs from that of an ordinary galvanometer or ammeter in that its motion is *undamped*. The motion of the needle in an ordinary ammeter is damped, so that the needle doesn't swing violently whenever the current is changed, and so that the needle moves promptly and purposefully towards its correct position. How is this damping achieved?

The coil of a moving-coil meter is wound around a small aluminium frame called a *former*. When the current through the ammeter coil is changed, the coil – and the former – swing round; but a current is induced in the former, which gives the former a magnetic moment in such a sense as to *oppose* and therefore dampen the motion. The resistance of the former is made just right so that critical damping is achieved, so that the needle reaches its equilibrium position in the least time without overshoot or swinging. The little aluminium former does not look as if it were an important part of the instrument – but in fact its careful design is very important!

10.5 *AC Generator*

This and the following sections will be devoted to *generators* and *motors*. I shall not be concerned with – and indeed am not knowledgeable about – the engineering design or practical details of real generators or motors, but only with the scientific principles involved. The "generators" and "motors" of this chapter will be highly idealized abstract concepts bearing little obvious resemblance to the real things. Need an engineering student, then, pay any attention to this? Well, of course, all real generators and motors obey and are designed around these very scientific principles, and they wouldn't work unless their designers and builders had a very clear knowledge and understanding of the basic principles.

The rod sliding on rails in a magnetic field described in Section 10.2 in fact was a D.C. (direct current) generator. I now describe an A.C. (alternating current) generator.

In figure X.5 we have a magnetic field **B**, and inside the field we have a coil of area **A** (yes – area is a vector) and N turns. The coil is being physically turned counterclockwise by some outside agency at an angular speed ω radians per second. I am not concerned with who, what or how it is

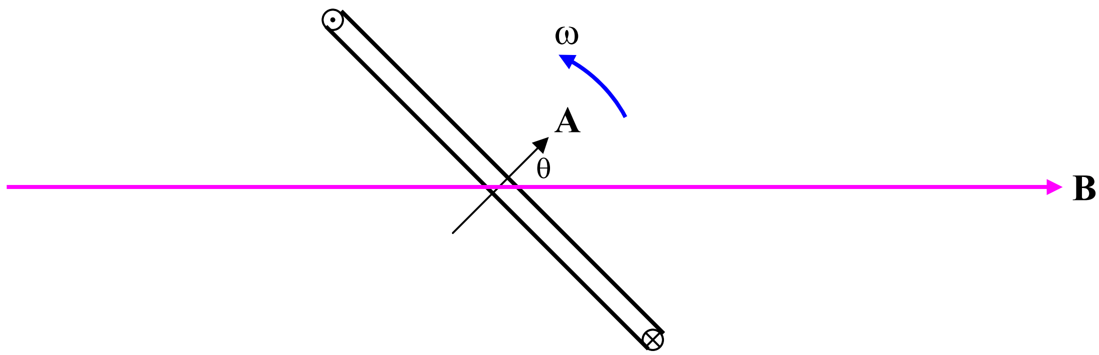


FIGURE X.5

being physically turned. For all I know, it might be turned by a little man turning a hand crank, or by a steam turbine driven by a coal- or oil-burning plant, or by a nuclear reactor, or it might be driven by a water turbine from a hydroelectric generating plant, or it might be turned by having something rubbing against the rim of your bicycle wheel. All I am interested in is that it is being mechanically turned at an angular speed ω. As the coil turns, the flux through it changes, and a current flows through the coil in a direction such that the magnetic moment generated for the coil is in the direction indicated for the area **A** in figure X.5, and also indicated by the symbols ⊙ and ⊗. This will result in an opposition to rotation of the coil; whoever or whatever is causing the coil to rotate will experience some opposition to his efforts and will have to do work. You can also deduce the direction of the induced current by considering the direction of the Lorentz force on the electrons in the wire of the coil.

At the instant illustrated in figure X.5, the flux through the coil is $AB\cos\theta$, or $AB\cos\omega t$, if we assume that $\theta = 0$ at $t = 0$. The rate of change of flux through the coil at this instant is the time derivative of this, or $-AB\,\omega\sin\omega t$. The magnitude of the induced EMF is therefore

$$V = NAB\omega\sin\omega t = \hat{V}\sin\omega t, \qquad 10.5.1$$

where $\hat{V}$ ("V-peak") is the peak or maximum EMF, given by

$$\hat{V} = NAB\omega. \qquad 10.5.2$$

Are you surprised that the peak EMF is proportional to N ? To A ? To B ? To ω ? Verify that $NAB\,\omega$ has the correct dimensions for $\hat{V}$.

The peak EMF occurs when the flux through the coil is *changing* most rapidly; this occurs when $\theta = 90^\circ$, at which time the coil is horizontal and the flux through it is zero.

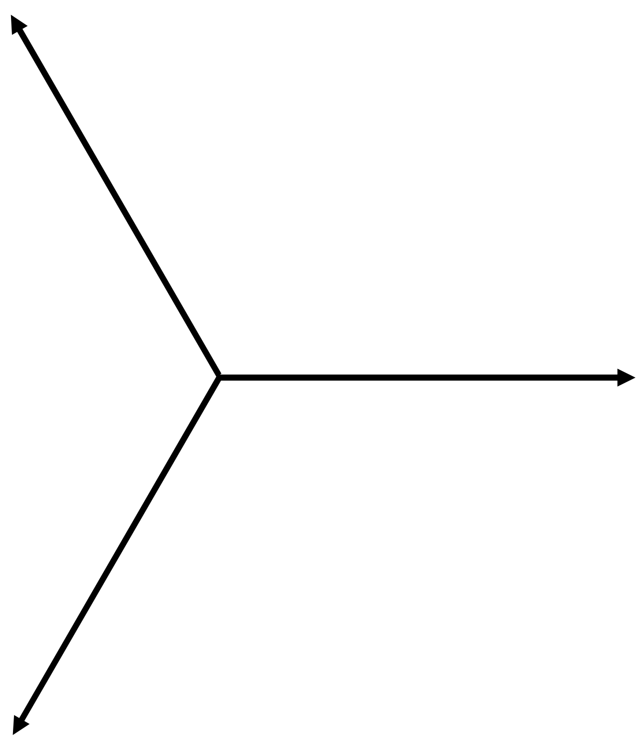
The leads from the coil can be connected to an external circuit via a pair of *slip rings* through which they can deliver current to the circuit.

The actual physical design of a generator is beyond the scope of this chapter and indeed of my expertise, though all depend on the physical principles herein described. In the "design" (such as it is) that I have described, the coil in which the EMF is induced is the *rotor* while the magnet is the *stator* – but this need not always be the case, and indeed designs are perfectly possible in which the magnet is the rotor and the coil the stator. In my design, too, I have assumed that there is but one coil – but there might be several in different planes. For example, you might have three coils whose planes make angles of 120° with each other. Each then generates a sinusoidal voltage, but the phase of each differs by 120° from the phases of the other two. This enables the delivery of power to three circuits. In a common arrangement these three circuits are not independent, but each is connected to a common line. The EMF in this common line is then made up three sine waves differing in phase by 120° :

$$V = \hat{V}[\sin \omega t + \sin(\omega t + 120^\circ) + \sin(\omega t + 240^\circ)]. \qquad 10.5.3$$

There are several ways in which you can see what this is like. For example, you could calculate this expression for numerous values of t and plot the function out as a graph. Or you could expand the expressions $V = \sin(\omega t + 120^\circ)$ and $V = \sin(\omega t + 240^\circ)$, and gather the various terms together to see what you get. (I recommend trying this.) Or you could simply add the three components in a *phase diagram*:

FIGURE X.6



It then becomes obvious that the sum is *zero*, and this line is the *neutral* line, the other three being *live* lines.

10.6 *AC Power*

When a current I flows through a resistance R, the rate of dissipation of electrical energy as heat is $I^2 R$. If an alternating potential difference $V = \hat{V}\sin\omega t$ is applied across a resistance, then an alternating current $I = \hat{I}\sin\omega t$ will flow through it, and the rate at which energy is dissipated as heat will also change periodically. Of interest is the *average* rate of dissipation of electrical energy as heat during a complete cycle of period $P = 2\pi/\omega$.

Let W = instantaneous rate of dissipation of energy, and $\overline{W}$ = average rate over a cycle of period $P = 2\pi/\omega$. Then

$$\overline{W}P = \int_0^P W dt = R\int_0^P I^2 dt = R\hat{I}^2\int_0^P \sin^2\omega t dt$$

$$= \tfrac{1}{2}R\hat{I}^2\int_0^P (1 - \cos 2\omega t)dt = \tfrac{1}{2}R\hat{I}^2[t - \tfrac{1}{2\omega}\sin 2\omega t]_0^{P=2\pi/\omega} = \tfrac{1}{2}R\hat{I}^2 P. \qquad 10.6.1$$

Thus

$$\overline{W} = \tfrac{1}{2}R\hat{I}^2. \qquad 10.6.2$$

The expression $\frac{1}{2}\hat{I}^2$ is the mean value of I^2 over a complete cycle. Its square root $\hat{I}/\sqrt{2} = 0.707\hat{I}$ is the *root mean square* value of the current, I_{RMS}. Thus the average rate of dissipation of electrical energy is

$$\overline{W} = RI^2_{\text{RMS}}. \qquad 10.6.3$$

Likewise, the RMS EMF (pardon all the abbreviations) over a complete cycle is $\hat{V}/\sqrt{2}$.

Often when an AC current or voltage is quoted, it is the RMS value that is meant rather than the peak value. I recommend that in writing or conversation you always *make it explicitly clear* which you mean.

Also of interest is the *mean* induced voltage $\overline{V}$ over half a cycle. (Over a full cycle, the mean voltage is, of course, zero.) We have

$$\overline{V}P/2 = \int_0^{P/2} V dt = \hat{V}\int_0^{P/2}\sin\omega t dt = \frac{\hat{V}}{\omega}[\cos\omega t]^0_{\frac{P}{2}=\frac{\pi}{2}}$$

$$= \frac{\hat{V}}{\omega}(1 - \cos\pi) = \frac{2\hat{V}}{\omega}. \qquad 10.6.4$$

Remembering that $P = 2\pi/\omega$, we see that

$$\overline{V} = \frac{2\hat{V}}{\pi} = 0.6366\hat{V} = \frac{2\sqrt{2}V_{\text{RMS}}}{\pi} = 0.9003V_{\text{RMS}}. \qquad 10.6.5$$

10.7 *Linear Motors and Generators*

Most (but not all!) real motors and generators are, of course, rotary. In this section I am going to describe highly idealized and imaginary linear motors and generators, only because the geometry is simpler than for rotary motors, and it is easier to explain certain principles. We'll move on the rotary motors afterwards.

In figure X.7 I compare a motor and a generator. In both cases there is supposed to be an external magnetic field (from some external magnet) directed away from the reader. A metal rod is resting on a pair of conducting rails.

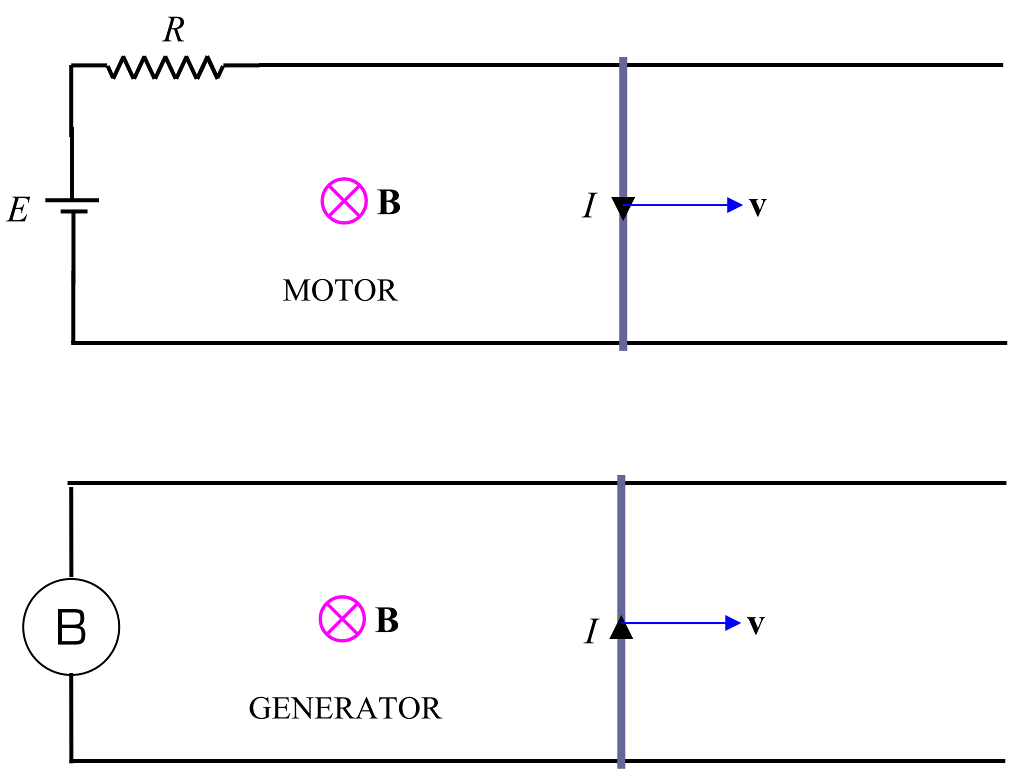


FIGURE X.7

In the motor, a battery is connected in the circuit, causing a current to flow clockwise around the circuit. The interaction between the current and the external magnetic field produces a force on the rod, moving it to the right.

In the generator, the rod is moved to the right by some externally applied force, and a current is induced counterclockwise. If the B inside the circle represents a light bulb, a current will flow through the bulb, and the bulb will light up.

Let us suppose that the rails are smooth and frictionless, and suppose that, in the motor, the rod isn't pulling any weight. That is to say, suppose that there is no mechanical load on the motor. How fast will the rod move? Since there is a force moving the rod to the right, will it continue to accelerate indefinitely to the right, with no limit to its eventual speed? No, this is not what happens. When the switch is first closed and the rod is stationary, a current will flow, given by $E = IR$, where E is the EMF of the battery and R is the total resistance of the circuit. However, when the rod has reached a speed υ, the area of the circuit is increasing at a rate $a\upsilon$, and a *back EMF* (which opposes the EMF of the battery), of magnitude $a\upsilon B$ is induced, so the net EMF in the circuit is now $E - a\upsilon B$ and the current is correspondingly reduced according to

$$E - a\upsilon B = IR. \qquad 10.7.6$$

Eventually the rod reaches a limiting speed of $E/(aB)$, at which point no further current is being taken from the battery, and the rod (sliding as it is on frictionless rails with no mechanical load) then obeys Newton's first law of motion – namely it will continue in its state of uniform motion, because no forces are no acting upon it.

Problem 1. Show that the speed increases with time according to

$$\upsilon = \frac{E}{aB}\left(1 - \exp\left(-\frac{(aB)^2 t}{mR}\right)\right), \qquad 10.7.7$$

where m is the mass of the rod.

Problem 2. Show that the time for the rod to reach half of its maximum speed is

$$t_{1/2} = \frac{mR\ln 2}{(a\,B)^2}. \qquad 10.7.8$$

Problem 3. Suppose that $E = 120$ V, $a = 1.6$ m, $m = 1.92$ kg and $R = 4\ \Omega$. If the rod reaches a speed of 300 m s^{-1} in 300 s, what is the strength of the magnetic field?

I'll give solutions to these problems at the end of this section. Until then – no peeking.

In a frictionless *rotary* motor, the situation would be similar. Initially the current would be E/R, but, when the motor is rotating with angular speed ω, the average back EMF is $2NAB\,\omega/\pi$ equation 10.5.5), and by the time this has reached the EMF of the battery, the frictionless, loadless coil carries on rotating at constant angular speed, taking no current from the battery.

Now let's go back to our linear motor consisting of a metal rod lying on two rails, but this time suppose that there is some mechanical resistance to the motion. This could be either because there is friction between the rod and the rails, or perhaps the rod is dragging a heavy weight behind it, or both. One way or another, let us suppose that the rod is subjected to a constant force F towards the left. As before, the relation between the current and the speed is given by equation 10.7.6, but, when a steady state has been reached, the electromagnetic force aIB pulling the rod to the right is equal to the mechanical load F dragging the rod to the left. That is, $E - a\upsilon B = IR$ and $F = aIB$. If we eliminate I between these two equations, we obtain

$$E - a\upsilon B = \frac{FR}{aB}, \qquad 10.7.9$$

or

$$\upsilon = \frac{E}{aB} - \frac{R}{(aB)^2}F. \qquad 10.7.10$$

This equation, which relates the speed at which the motor runs to the mechanical load, is called the *motor performance characteristic*. In our particular motor, the performance characteristic shows that the speed at which the motor runs decreases steadily as the load is increased., and the motor runs to a grinding halt for a load equal to aBE/R. (Verify that this has the dimensions of force.) The current is then E/R. This current may be quite large. If you physically prevent a real motor from turning by applying a mechanical torque to it so large that the motor cannot move, a large current will flow through the coil – large enough to heat and possibly fuse the coil. You will hear a sharp crack and see a little puff of smoke.

If we multiply equation 10.7.6 by I, we obtain

$$EI = aIB\upsilon + I^2R, \qquad 10.7.11$$

or

$$EI = F\upsilon + I^2R. \qquad 10.7.12$$

This shows that the power produced by the battery goes partly into doing external mechanical work, and the remainder is dissipated as heat in the resistance. Restrain the motor so that $\upsilon = 0$, and *all* of that EI goes into I^2R.

If you were physically to move the rod to the right at a speed faster than the equilibrium speed, the back EMF becomes greater than the battery EMF, and current flows back into the battery. The device is then a generator rather than a motor.

The nature of the performance characteristic varies with the details of motor design. You may not want a motor whose speed decreases so drastically with load. You may have to decide in advance what sort of performance characteristic you want the motor to have, depending on what tasks you want it to perform, and then you have to design the motor accordingly. We shall mention some possibilities in the next section.

Now – the promised solutions to the problems.

Solution to Problem 1.

When the speed of the rod is υ, the net EMF in the circuit is $E - aB\upsilon$, so the current is $(E - aB\upsilon)/R$, and so the force on the rod will be $aB(E - aB\upsilon)/R$ and the acceleration $d\upsilon/dt$ will be $aB(E - aB\upsilon)/(mR)$. The equation of motion is therefore

$$\frac{d\upsilon}{E - aB\upsilon} = \frac{aB}{mR}dt. \qquad 10.7.13$$

Integration, with $\upsilon = 0$ when $t = 0$, gives the required equation 10.7.7.

Solution to Problem 2.

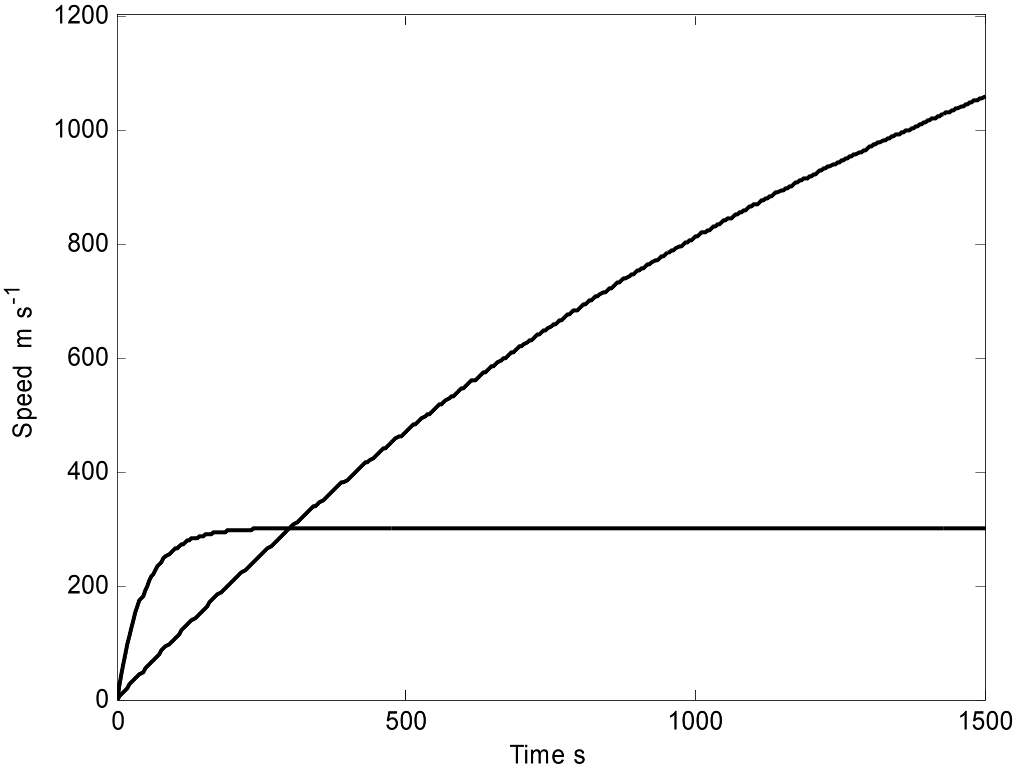
Just put $\upsilon = \dfrac{E}{2aB}$ in equation 10.7.7 and solve for t. Verify that the expression has the dimensions of time.

Solution to Problem 3.

Put the given numbers into equation 10.7.7 to get

$$B = \tfrac{1}{4}(1 - e^{-100B^2}) \qquad 10.7.14$$

and solve this for B. (Nice and easy. But if you are not experienced in solving equations such as this, the Newton-Raphson process is described in Chapter 1 of the Celestial Mechanics notes of this series. This equation would be good practice.) There are two possible answers, namely 0.043996 T and 0.249505 teslas. I draw the speed:time graphs for the two solutions below:



Numbers of interest for the two fields:

B (T)	v_∞ (m s^{-1})	$\bar{t}$ s
0.0440	1704.7	1074.29
0.2495	300.6	33.40

10.8 *Rotary Motors*

Most real motors, of course, are rotary motors, though all of the principles described for our highly idealized linear motor of the previous section still apply.

Current is fed into a coil (known as the *armature)* via a *split-ring commutator* and the coil therefore develops a magnetic moment. The coil is in a magnetic field, and it therefore experiences a torque. (Figure X.5) The coil rotates and soon its magnetic moment vector will be parallel to the field and there would be no further torque – except that, at that instant, the split-ring commutator reverses the direction of the current in the coil, and hence reverses the direction of the magnetic moment. Thus the coil continues to rotate until, half a period later, its new magnetic moment again lines up with the magnetic field, and the commutator again reverses the direction of the moment.

As in the case of the linear motor, the coil reaches a maximum angular speed, which depends on the mechanical load (this time a torque). and the relation between the maximum angular speed and the torque is the motor performance characteristic.

Also, as with a generator, there may be several coils (with a corresponding number of sections in the commutator), and it is also possible to design motors in which the armature is the stator and the magnet the rotor – but I am not particularly knowledgeable about the detailed engineering designs of real motors – except that all of them depend upon the same scientific principles.

In all of the foregoing, it has been assumed that the magnetic field is constant, as if produced by a permanent magnet. In real motors, the field is generally produced by an *electromagnet*. (Some types of iron retain their magnetism permanently unless deliberately demagnetized. Others become magnetized only when placed in a strong magnetic field such as produced by a solenoid, and they lose most of their magnetization as soon as the magnetizing field is removed.)

The field coils may be wound in series with the armature coil (a series-wound motor) or in parallel with it (a shunt-wound motor), or even partly in series and partly in parallel (a compound-wound motor). Each design has it own performance characteristic, depending on the use for which it is intended.

With a single coil rotating in a magnetic field, the induced back EMF varies periodically, the average value being, as we have seen, $2NAB\,\omega/\pi$. In practice the coil may be wound around many slots placed around the perimeter of a cylindrical core every few degrees, and there are a corresponding number of sections in the split-ring commutator. The back EMF is then less variable than with a single coil, and, although the formula $2NAB\,\omega/\pi$ is no longer appropriate, the back EMF is still proportional to $B\ \omega$. We can write the average back EMF as $KB\omega$, where the *motor constant* K depends on the detailed geometry of a particular design.

Shunt-wound Motor. In the shunt-wound motor, the field coil is wound in parallel to the armature coil. In this case, the back EMF generated in the armature does not affect the current in the field coil, so the motor operates rather as previously described for a constant field. That is, the motor performance characteristic, giving the equilibrium angular speed in terms of the mechanical load (torque, τ) is given by

$$\omega = \frac{E}{KB} - \frac{R}{(KB)^2}\tau. \qquad 10.8.1$$

Here, R is the armature resistance. In practice, there may be a variable resistance (rheostat) in series with the field coil, so that the current through the field coil – and hence the field strength – can be changed.

Series-wound Motor. The field coil is wound in series with the armature, and the motor performance characteristic is rather different that for the shunt-wound motor. If the magnet core does not saturate, then, to a linear approximation, the field is proportional to the current, and the back EMF is proportional to the product of the current I and the angular speed ω - so let's say that the back EMF is $kI\omega$. We then have

$$E - kI\omega = IR, \qquad 10.8.2$$

where E is the externally applied EMF (from a battery, for example) and R is the total resistance of field coil plus armature.

Multiply both sides by I:

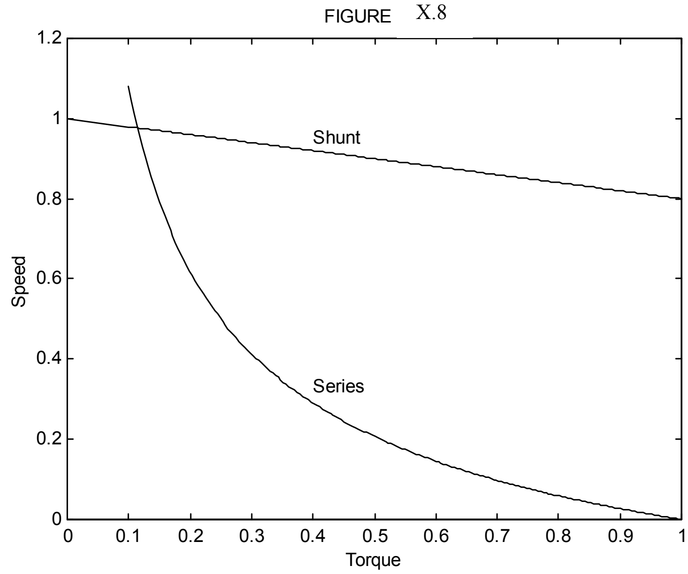
$$EI - kI^2\omega = I^2R. \qquad 10.8.3$$

The term EI is the power supplied by the battery and I^2R is the power dissipated as heat. Thus the rate of doing mechanical work is $kI^2\omega$, which shows that the torque exerted by the motor is $\tau = kI^2$. If we now substitute $\sqrt{\tau / k}$ for I in equation 10.8.2, we obtain the motor performance characteristic – i.e. the relation between ω and τ:

$$\omega = \frac{E}{\sqrt{k\tau}} - \frac{R}{k}. \qquad 10.8.4$$

In figure X.8 we show the performance characteristics, in arbitrary units, for shunt- and series-wound motors, based in our linear analysis, which assumes in both cases no saturation of the electromagnet iron core. The maximum possible torque in both cases is the torque that makes $\omega = 0$ in the corresponding performance characteristic, namely KBE/R for the shunt-wound motor and kE^2/R for the series-wound motor. The latter goes to infinity for zero load. This does not happen in practice, because we have made some assumptions that are not real (such as no saturation of the magnet core, and also there can never be literally zero load), but nevertheless the analysis is sufficient to show the general characteristics of the two types.

FIGURE X.8



The characteristics of the two may be combined in a compound-wound motor, depending on the intended application. For example, a tape-recorder requires constant speed, whereas a car starter requires a high starting torque.

10.9 *The Transformer*

Two coils are wound on a common iron core. The primary coil is connected to an AC (alternating current) generator of (RMS) voltage V_1. If there are N_1 turns in the primary coil, the primary current will be proportional to V_1 / N_1 and, provided the core is not magnetically saturated, the magnetic field will also be proportional to this. The voltage V_2 induced in the secondary coil (of N_2 turns) will be proportional to N_2 and to the field, and so we have

$$\frac{V_2}{V_1} = \frac{N_2}{N_1}. \qquad 10.9.1$$

We shall give a more detailed analysis of the transformer in a later chapter. However, one aspect which can be noted here is that the rapidly-changing magnetic field induces *eddy currents* in the iron core, and for this reason the core is usually constructed of thin laminated sheets (or sometimes wires) insulated from each other to reduce these energy-wasting eddy currents. Sometimes these laminations vibrate a little unless tightly bound together, and this is often responsible for the "hum" of a transformer.

10.10 *Mutual Inductance*

Consider two coils, not connected to one another, other than being *close together in space*. If the current changes in one of the coils, so will the magnetic field in the other, and consequently an EMF will be induced in the second coil. **Definition:** The ratio of the EMF V_2 induced in the second coil to the rate of change of current $\dot{I}_1$ in the first is called the *coefficient of mutual inductance M* between the two coils:

$$V_2 = M\dot{I}_1. \qquad 10.10.1$$

The dimensions of mutual inductance can be found from the dimensions of EMF and of current, and are readily found to be $\mathrm{ML^2Q^{-2}}$.

Definition: If an EMF of one volt is induced in one coil when the rate of change of current in the other is 1 amp per second, the coefficient of mutual inductance between the two is 1 *henry*, H.

Mental Exercise: If the current in coil 1 changes at a rate $\dot{I}_1$, the EMF induced in coil 2 is $M\dot{I}_1$. Now ask yourself this: If the current in coil 2 changes at a rate $\dot{I}_2$, is it true that the EMF induced in coil 1 will be $M\dot{I}_2$? (The answer is "yes" – but you are not excused the mental effort required to convince yourself of this.)

Example: Suppose that the primary coil is an infinite solenoid having n_1 turns per unit length wound round a core of permeability μ. Tightly would around this is a plain circular coil of N_2 turns. The solenoid and the coil wrapped tightly round it are of area A. We can calculate the mutual inductance of this arrangement as follows. The magnetic field in the primary is $\mu n_1 I$ so the flux through each coil is $\mu n_1 AI$. If the current changes at a rate $\dot{I}$, flux will change at a rate $\mu nA\dot{I}$, and the EMF induced in the secondary coil will be $\mu n_1 N_2 A\dot{I}$. Therefore the mutual inductance is

$$M = \mu n_1 N_2 A. \qquad 10.10.2$$

Several points:

1. Verify that this has the correct dimensions.

2 If the current in the solenoid changes in such a manner as to cause an increase in the magnetic field towards the right, the EMF induced in the secondary coil is such that, if it were connected to a closed circuit so that a secondary current flows, the direction of this current will produce a magnetic field towards the left – i.e. such as to oppose the rightward increase in B.

3. Because of the little mental effort you made a few minutes ago, you are now convinced that, if you were to change the current in the plane coil at a rate $\dot{I}$, the EMF induced in the solenoid would be $M\dot{I}$, where M is given by equation 10.10.2.

4. Equation 10.10.2 is the equation for the mutual inductance of the system, provided that the coil and the solenoid are *tightly coupled*. If the coil is rather loosely draped around the solenoid, or if the solenoid is not infinite in length, the mutual inductance would be rather less than given by equation 10.10.2. It would be, in fact, $k\mu n_1 N_2 A$, where k, a dimensionless number between 0 and 1, is the *coupling coefficient.*

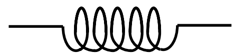
5. While we have hitherto expressed permeability in units of tesla metres per amp (T m A^{-1}) or some such combination, equation 10.10.2 shows that permeability can equally well be (and usually is) expressed in henrys per metre, H m^{-1}. Thus, we say that the permeability of free space is $\mu_0 = 4\pi \times 10^{-7}$ H m^{-1}.

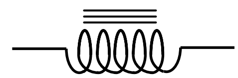
Exercise: A plane coil of 10 turns is tightly wound around a solenoid of diameter 2 cm having 400 turns per centimetre. The relative permeability of the core is 800. Calculate the mutual inductance.
(I make it 0.126 H.)

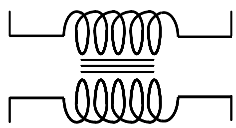
10.11 *Self Inductance*

In this section we are dealing with the self inductance of a *single coil* rather than the mutual inductance between two coils. If the current through a single coil changes, the magnetic field inside that coil will change; consequently a back EMF will be induced in the coil that will oppose the change in the magnetic field and indeed will oppose the change of current. **Definition:** The ratio of the back EMF to the rate of change of current is the *coefficient of self inductance L*. If the back EMF is 1 volt when the current changes at a rate of one amp per metre, the coefficient of self inductance is 1 henry.

Exercise: Show that the coefficient of self inductance (usually called simply the "inductance") of a long solenoid of length l and having n turns per unit length is $\mu n^2 Al$, where I'm sure you know what all the symbols stand for. Put some numbers in for an imaginary solenoid of your own choosing, and calculate its inductance in henrys.

The circuit symbol for inductance is 

If a coil has an iron core, this may be indicated in the circuit by 

The symbol for a transformer is 

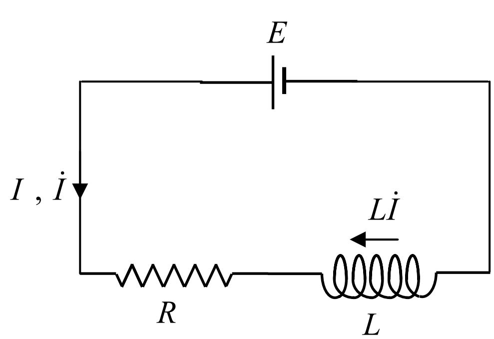
Finally, don't confuse self-inductance with self-indulgence.

10.12 *Growth of Current in a Circuit Containing Inductance*

It will have occurred to you that if the growth of current in a coil results in a back EMF which opposes the increase of current, current cannot change instantaneously in a circuit that contains inductance. This is correct. (Recall also that the potential difference in a circuit cannot change instantaneously in a circuit containing capacitance. Come to think of it, it is hardly possible for the capacitance or inductance of any circuit to be exactly zero; any real circuit must have some capacitance and inductance, even if very small.)

Consider the circuit of figure X.9. A battery of EMF E is in series with a resistance and an inductance. (A coil or solenoid or any inductor in general will have both inductance and resistance, so the R and the L in the figure may belong to one single item.) We have to be very careful about *signs* in what follows.

FIGURE X.9



When the circuit is closed (by a switch, for example) a current flows in the direction shown. by an arrow, which also indicates the direction of the *increase* of current. An EMF $L\dot{I}$ is induced in the opposite direction to $\dot{I}$. Thus, Ohm's law, or, if your prefer, Kirchhoff's second rule, applied to the circuit (watch the signs carefully) is

$$E - IR - L\dot{I} = 0. \qquad 10.12.1$$

Hence:

$$\int_0^I \frac{dI}{\frac{E}{R} - I} = \frac{R}{L}dt. \qquad 10.12.2$$

Warning: Some people find an almost irresistible urge to write this as $\int_0^I \frac{dI}{I - \frac{E}{R}} = -\frac{R}{L}dt.$

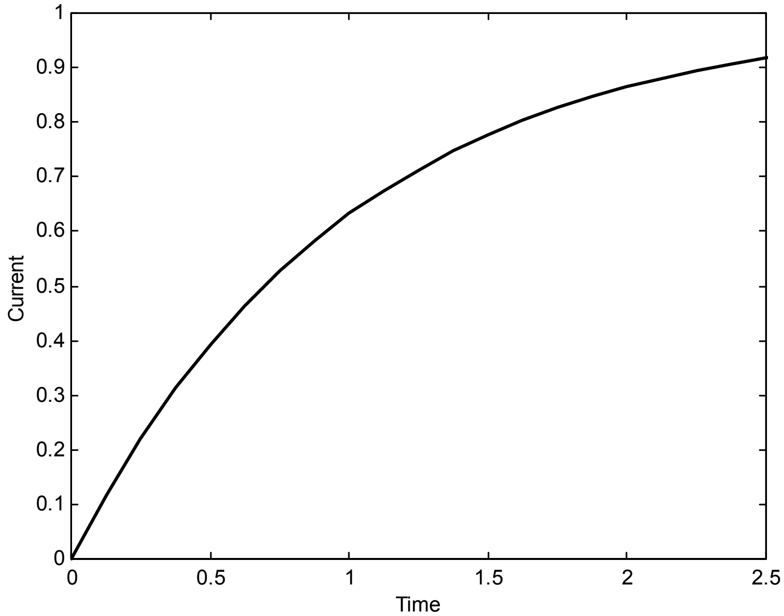
Don't!

You can anticipate that the left hand side is going to be a logarithm, so make sure that the denominator is positive. You may recall a similar warning when we were charging and discharging a capacitor through a resistance.

Integration of equation 10.12.2 results in the following equation for the growth of the current with time:

$$I = \frac{E}{R}\left(1 - e^{-(R/L)t}\right). \qquad 10.12.3$$

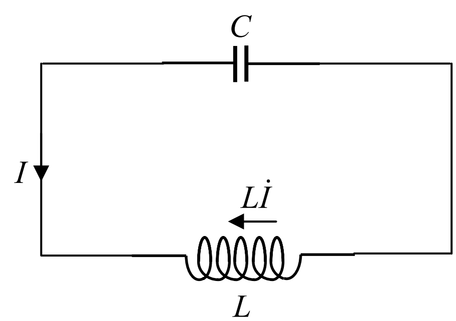
Thus the current asymptotically approaches its ultimate value of E/R, reaching 63% (i.e. $1 - e^{-1}$) of its ultimate value in a time L/R. In figure X.10, the current is shown in units of E/R, and the time in units of L/R. You should check that L/R, which is called the *time constant* of the circuit, has the dimensions of time.



10.13 *Discharge of a Capacitor through an Inductance*

The circuit is shown in figure X.11, and, once again, it is important to take care with the signs.

FIGURE X.11



If $+Q$ is the charge on the left hand plate of the capacitor at some time (and $-Q$ the charge on the right hand plate) the current I in the direction indicated is $-\dot{Q}$ and the potential difference across the plates is Q/C. The back EMF is in the direction shown, and we have

$$\frac{Q}{C} - L\dot{I} = 0, \qquad 10.13.1$$

or

$$\frac{Q}{C} + L\ddot{Q} = 0. \qquad 10.13.2$$

This can be written $$\ddot{Q} = -\frac{Q}{LC}, \qquad 10.13.3$$

which is simple harmonic motion of period $2\pi\sqrt{LC}$. (verify that this has dimensions of time.) Thus energy sloshes to and fro between storage as charge in the capacitor and storage as current in the inductor.

If there is resistance in the circuit, the oscillatory motion will be damped, the charge and current eventually approaching zero. But, even if there is no resistance, the oscillation does not continue for ever. While the details are beyond the scope of this chapter, being more readily dealt with in a discussion of electromagnetic radiation, the periodic changes in the charge in the capacitor and the current in the inductor, result in an oscillating electromagnetic field around the circuit, and in the generation of an electromagnetic wave, which carries energy away at a speed of $\sqrt{1/(\mu_0\varepsilon_0)}$. Verify that this has the dimensions of speed, and that it has the value 2.998×10^8 m s^{-1}. The motion in the circuit is damped just as if there were a resistance of $\sqrt{\mu_0/\varepsilon_0} = c\mu_0 = 1/(c\varepsilon_0)$ in the circuit. Verify that this has the dimensions of resistance and that it has a value of 376.7 Ω. This effective resistance is called the *impedance of free space.*

10.14 *Discharge of a Capacitor through an Inductance and a Resistance*

This results in *damped* oscillatory motion. It is discussed in detail in Chapter 11 of the Classical Mechanics part of these notes, especially Section 11.6.

10.15 *Energy Stored in an Inductance*

During the growth of the current in an inductor, at a time when the current is i and the rate of increase of current is $\dot{i}$, there will be a back EMF $L\dot{i}$. The rate of doing work against this back EMF is then $Li\dot{i}$. The work done in time dt is $Li\dot{i}dt = Lidi$, where di is the increase in current in time dt. The total work done when the current is increased from 0 to I is

$$L\int_0^I idi = \tfrac{1}{2}LI^2, \qquad 10.15.1$$

and this is the energy stored in the inductance. (Verify the dimensions.)

10.16 *Energy Stored in a Magnetic Field*

Recall your derivation (Section 10.11) that the inductance of a long solenoid is $\mu n^2 Al$. The energy stored in it, then, is $\frac{1}{2}\mu n^2 AlI^2$. The volume of the solenoid is Al, and the magnetic field is $B = \mu nI$, or $H = nI$. Thus we find that the energy stored per unit volume in a magnetic field is

$$\frac{B^2}{2\mu} = \tfrac{1}{2}BH = \tfrac{1}{2}\mu H^2. \qquad 10.16.2$$

In a vacuum, the energy stored per unit volume in a magnetic field is $\tfrac{1}{2}\mu_0 H^2$ - even though the vacuum is absolutely empty!

Equation 10.16.2 is valid in any isotropic medium, including a vacuum. In an anisotropic medium, $\mathbf{B}$ and $\mathbf{H}$ are not in general parallel – unless they are both parallel to a crystallographic axis. More generally, in an anisotropic medium, the energy per unit volume is $\tfrac{1}{2}\mathbf{B} \cdot \mathbf{H}$.

Verify that the product of B and H has the dimensions of energy per unit volume.

CHAPTER 11
DIMENSIONS

Although we have not yet met all of the quantities in use in electricity and magnetism, we have met most of the important ones. Of those yet to come, some, such as impedance and reactance, will obviously have the dimensions of resistance; some, such as reluctance and permeance, you will rarely come across; and some, such as magnetic susceptibility, will obviously be dimensionless. Now is therefore quite a convenient time to gather together the various quantities we have come across, together with their dimensions (i.e. the powers of M, L, T and Q of which they are composed) and their SI units. It will be a good time, too, for the reader to review the definitions of the various quantities and to verify, from their definitions, their dimensions. Let me know (jtatum@uvic.ca) if you find any mistakes in the following table.

	Powers of M	L	T	Q	SI unit
Force	1	1	–2	0	N
Work, energy	1	2	–2	0	J
Torque	1	2	–2	0	N m
Power	1	2	–3	0	W
Linear momentum, impulse	1	1	–1	0	kg m s^{-1} or N s
Rotational inertia	1	2	0	0	kg m^2
Angular momentum	1	2	–1	0	J s
Electric charge	0	0	0	1	C
Electric dipole moment	0	1	0	1	C m
Current	0	0	–1	1	A
Potential difference	1	2	–2	–1	V
Resistance	1	2	–1	–2	Ω
Resistivity	1	3	–1	–2	Ω m
Conductance	–1	–3	1	2	S
Conductivity	–1	–4	1	2	S m^{-1}
Capacitance	–1	–2	2	2	F
Electric field E	1	1	–2	–1	N C^{-1} or V m^{-1}
Electric field D	0	–2	0	1	C m^{-2}
Electric flux Φ_E	1	3	–2	1	V m
Electric flux Φ_D	0	0	0	1	C
Permittivity	–1	–3	2	2	F m^{-1}
Magnetic field B	1	0	–1	–1	T
Magnetic field H	0	–1	–1	1	A m^{-1}
Magnetic flux Φ_B	1	2	–1	–1	T m^2 or V s
Magnetic flux Φ_H	0	3	–1	1	A m
Permeability	1	1	0	–2	H m^{-1}
Magnetic vector potential	1	1	–1	–1	T m
Inductance	1	2	0	–2	H

CHAPTER 12
PROPERTIES OF MAGNETIC MATERIALS

12.1 *Introduction*

This chapter is likely to be a short one, not least because it is a subject in which my own knowledge is, to put it charitably, a little limited. A thorough understanding of why some materials are magnetic requires a full course in the physics of the solid state, a course that I could not possibly give. Nevertheless, there are a few basic concepts and ideas concerned with magnetic materials which everyone who is interested in electromagnetism should know, and it is the aim of this chapter to describe them in a very introductory way.

It may be worthwhile to remind ourselves of the ways in which we have defined the magnetic fields B and H. To define B, we noted that an electric current situated in a magnetic field experiences a force at right angles to the current, the magnitude and direction of this force depending on the direction of the current. We accordingly defined B as being equal to the maximum force per unit length experienced per unit current, the defining equation being $\mathbf{F'} = \mathbf{I} \times \mathbf{B}$.

Later, we asked ourselves about the strength of the magnetic field in the vicinity of an electric current. We introduced the Biot-Savart law, which says that the contribution to the magnetic field from an element ds of a circuit carrying a current I is proportional to $(I\,ds \sin\theta)/r^2$, and we called the constant of proportionality $\mu/(4\pi)$, where μ is the *permeability* of the material surrounding the current. We might equally well have approached it from another angle. For example, we might have noted that the magnetic field inside a solenoid is proportional to nI, and we could have denoted the constant of proportionality μ, the permeability of the material inside the solenoid.

We then defined H as being an alternative measure of the magnetic field, given by $H = B/\mu$.

In an isotropic medium, the vectors **B** and **H** are parallel, and the permeability is a scalar quantity. In an anisotropic crystal, **B** and **H** are not necessarily parallel, and the permeability is a tensor.

Some people see an analogy between the equation between the equation $B = \mu H$ and the equation $D = \epsilon E$ of electric fields. With our approach, however, I think most readers will see that, to the extent that there may be an analogy, the analogy is between $D = \epsilon E$ and $H = B/\mu$.

For example, consider a long solenoid, in the inside of which are two different magnetic materials in series, the first of permeability μ_1 and the second of greater permeability μ_2. The H-field everywhere inside the solenoid is just nI, regardless of what is inside it. Like **D**, the component of **H** perpendicular to the boundary between two media is continuous, whereas the perpendicular component of **B** is greater inside the material with the larger permeability. Likewise, if you were to consider, for example, two different media lying side-by-side in parallel, between the poles, for example, of a horseshoe magnet, the component of **B** parallel to the boundary between the media is continuous, and the parallel component of **H** is less in the medium of greater permeability.

In this chapter, we shall introduce a few new words, such as permeance and magnetization. We shall describe in a rather simple and introductory way five types of magnetism exhibited by various materials: *diamagnetism, paramagnetism, ferromagnetism, antiferromagnetism* and *ferrimagnetism.* And we shall discuss the phenomenon of *hysteresis.*

12.2 *Magnetic Circuits and Ohm's Law*

Some people find it helpful to see an analogy between a system of solenoids and various magnetic materials and a simple electrical circuit. They see it as a "magnetic circuit". I myself haven't found it to be particularly useful – but, as I mentioned, my experience in this field is less than extensive. I think it may be useful for some readers, however, at least to be introduced to the concept.

The magnetic field inside a long solenoid is given by $B = \mu nI = \mu NI / l$. Here, n is the number of turns per unit length, N is the total number of turns, and l is the length of the solenoid. If the cross-sectional area of the solenoid is A, the B-flux is $\Phi_B = \mu NIA / l$. This can be written

$$NI = \Phi_B \times \frac{l}{\mu A}. \qquad 12.2.1$$

The analogy which some people find useful is between this and Ohm's law:

$$V = IR. \qquad 12.2.2$$

The term NI, expressed in *ampere-turns*, is the *magnetomotive force* MMF.

The symbol Φ_B is the familiar B-flux, and is held to be analogous to current.

The term $l / (\mu A)$ is the *reluctance*, expressed in $\mathrm{H^{-1}}$. Reluctances add in series.

The reciprocal of the reluctance is the *permeance*, expressed in H. Permeances add in parallel.

Although the SI unit of permeance is the henry, permeance is not the same as the *inductance*. It will be recalled, for example, that the inductance of a long solenoid of N turns is $\frac{\mu A N^2}{l}$.

Continuing with the analogy, we recall that *resistivity* = $(A / l) \times$ resistance;

Similarly $\qquad$ *reluctivity* = $(A / l) \times$ reluctance = $1/\mu$. $(\mathrm{m\ H^{-1}})$

Also, the reciprocal of resistance is conductance.

Similarly the reciprocal of reluctance is *permeance.* (H)

And *conductivity* is $(l / A) \times$ conductance.

Likewise $(l / A) \times$ permeance is – what else? – permeability μ. $(\mathrm{H\ m^{-1}})$

I have mentioned these names partly for completeness and partly because it's fun to write some unusual and unfamiliar words such as permeance and reluctivity. I am probably not going to use these concepts further or give examples of their use. This is mostly because I am not as familiar with them myself as perhaps I ought to be, and I am sure that there are contexts in which these concepts are indeed highly useful. The next section introduces some more funny words, such as magnetization and susceptibility – but these are words that you *will* need to know and understand.

12.3 *Magnetization and Susceptibility*

The H-field inside a long solenoid is nI. If there is a vacuum inside the solenoid, the B-field is $\mu_0 H = \mu_0 nI$. If we now place an iron rod of permeability μ inside the solenoid, this doesn't change H, which remains nI. The B-field, however, is now $B = \mu H$. This is greater than $\mu_0 H$, and we can write

$$B = \mu_0(H + M). \qquad 12.3.1$$

The quantity M is called the *magnetization* of the material. In SI units it is expressed in A m^{-1}. We see that there are two components to B. There is the $\mu_0 H = \mu_0 nI$, which is the externally imposed field, and the component $\mu_0 M$, originating as a result of something that has happened within the material.

It might have occurred to you that you would have preferred to define the magnetization from $B = \mu_0 H + M$, so that the magnetization would be the excess of B over $\mu_0 H$. The equation $B = \mu_0 H + M$, would be analogous to the familiar $D = \epsilon_0 E + P$, and the magnetization would then be expressed in tesla rather than in A m^{-1}. This viewpoint does indeed have much to commend it, but so does $B = \mu_0(H + M)$. The latter is the recommended definition in the SI approach, and that is what we shall use here.

The ratio of the magnetization M ("the result") to H ("the cause"), which is obviously a measure of how susceptible the material is to becoming magnetized, is called the *magnetic susceptibility* χ_m of the material:

$$M = \chi_m H. \qquad 12.3.2$$

On combining this with equation 12.3.1 and $B = \mu H$, we readily see that the magnetic susceptibility (which is dimensionless) is related to the relative permeability $\mu_r = \mu / \mu_0$ by

$$\mu_r = 1 + \chi_m. \qquad 12.3.3$$

12.4 *Diamagnetism*

We mentioned in Section 12.1 that there are five types of magnetism exhibited by various materials. In this section we deal with the first of these, namely, diamagnetism.

Diamagnetic materials have a very weak *negative* susceptibility, typically of order -10^{-6}. That is to say, the relative permeability is slightly *less than* 1. Consequently, when a diamagnetic material is placed in a magnetic field, $B < \mu_0 H$.

If you are now hearing about this phenomenon for the first time, you may be a little surprised, and you will be expecting me to present a very short list of quite exotic materials known to be diamagnetic. So, here comes a *further* surprise: *All* materials are diamagnetic. Some materials may also be paramagnetic or ferromagnetic, and their positive paramagnetic or ferromagnetic susceptibilities may be larger than their negative diamagnetic susceptibility, so that their overall susceptibility is positive. But all materials are diamagnetic, even if their diamagnetism is hidden by their greater para- or ferromagnetism.

A proper account of the mechanism at the atomic level of the cause of diamagnetism requires a quantum mechanical treatment, but we can understand the phenomenon qualitatively classically. We just have to think of an atom as being a nucleus surrounded by electrons moving in orbits around the nucleus. When an atom (or a large collection of atoms in a macroscopic sample of matter) is placed in a magnetic field, a *current* is induced within the atom by electromagnetic induction. That is, the electrons are caused to orbit around the nucleus, and hence to give the atom a magnetic moment, in such a direction as to *oppose* the increase in the magnetic field that causes it. The result of this happening to all of the atoms in a macroscopic sample is that B will now be *less than* $\mu_0 H$, and the susceptibility will be negative. But, you may argue, these induced currents and their associated opposing magnetic moments will last only so long as the external field is *changing.* In fact it persists as long as the magnetizing field persists. The reason is as follows. In Chapter 10, we were dealing with wires and coils and resistors, and any current induced by a changing magnetic field was rapidly dissipated. For an electron in an orbit around a nucleus, however, there is no resistance, so, once it is set in motion, it will stay in motion. The same situation would arise if we were to induce a current in a loop of wire made of superconducting material whose resistance is zero. The current, once induced, continues, and is not dissipated away as heat.

12.5 *Paramagnetism*

Diamagnetism makes itself evident in atoms and molecules that have *no permanent magnetic moment.* Some atoms or molecules, however, do have a permanent magnetic moment, and such materials are *paramagnetic*. They must still be diamagnetic, but often the paramagnetism will outweigh the diamagnetism. The magnetic moment of an atom of a molecule is typically if order of a *Bohr magneton*. (See Chapter VII, Sections 21-23, of Stellar Atmospheres for more details about the Bohr magneton and the magnetic moments of atoms. All that we need note here is that a Bohr magneton is about 9.3×10^{-24} N m T^{-1}.) The presence of a permanent magnetic moment is often the result of *unpaired electron spins.* An example often quoted is the oxygen molecule O_2. Liquid oxygen indeed is paramagnetic. When a paramagnetic material is placed in a magnetic field, the magnetic moments experience a torque and they tend to orient themselves in the direction of the magnetic field, thus augmenting, rather than diminishing, B. Unsurprisingly the effect is greatest at low temperatures, where the random motion of atoms and molecules is low. At liquid helium temperatures (of order 1 K), susceptibilities can be of order $+10^{-3}$ or $+10^{-2}$, thus greatly exceeding the small negative susceptibility. At room temperature, paramagnetic susceptibilities are much less – typically about $+10^{-5}$, barely exceeding the diamagnetic susceptibility.

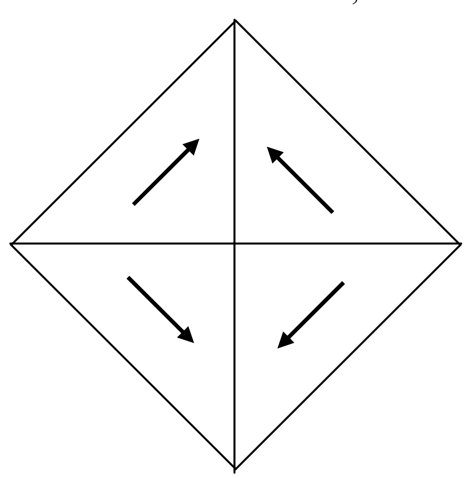
12.6 *Ferromagnetism*

What we normally think of as magnetic materials are technically *ferromagnetic.* The susceptibilities of ferromagnetic materials are typically of order $+10^3$ or 10^4 or even greater. However, the ferromagnetic susceptibility of a material is quite temperature sensitive, and, above a temperature known as the *Curie temperature*, the material ceases to become ferromagnetic, and it becomes merely paramagnetic.

Among the elements, only cobalt, iron and nickel are strongly ferromagnetic, their Curie temperatures being about 1400, 1040 and 630 K respectively. Gadolinium is ferromagnetic at low temperatures; its Curie temperature is about 289 K = 16 $^\circ$C. Dysprosium is ferromagnetic below its Curie temperature of about 105 K. There are many artificial alloys and ceramic materials which are ferromagnetic.

As with paramagnetic materials, the atoms have permanent magnetic moments, but with the difference that these moments are not randomly oriented but are strongly aligned to the crystallographic axes. Within a single crystal, there exist *domains,* within which all the magnetic moments are parallel and are aligned with a particular axis. In an adjacent domain, again all the moments are parallel to each other, but they may be aligned with a *different* axis, perhaps at right angles to the first domain, or perhaps aligned with the same axis but pointing in the opposite direction. Thus we have a number of domains, each highly magnetized, but with some domains magnetized in one direction and some in another. The domains are separated by domain boundaries, or "Bloch walls", perhaps a few hundred atoms thick, within which the orientation of the magnetic moments gradually changes from one domain to the next. Figure XII.1 is a schematic sketch of a crystal divided into four domains, with the magnetization in a different direction in each.

FIGURE XII.1



In figure XIII.2 I am exposing the crystal to a progressively stronger and stronger magnetic field, and we watch what happens to the domains, and, in figure XIII.3, to the magnetization of the crystal as a whole.

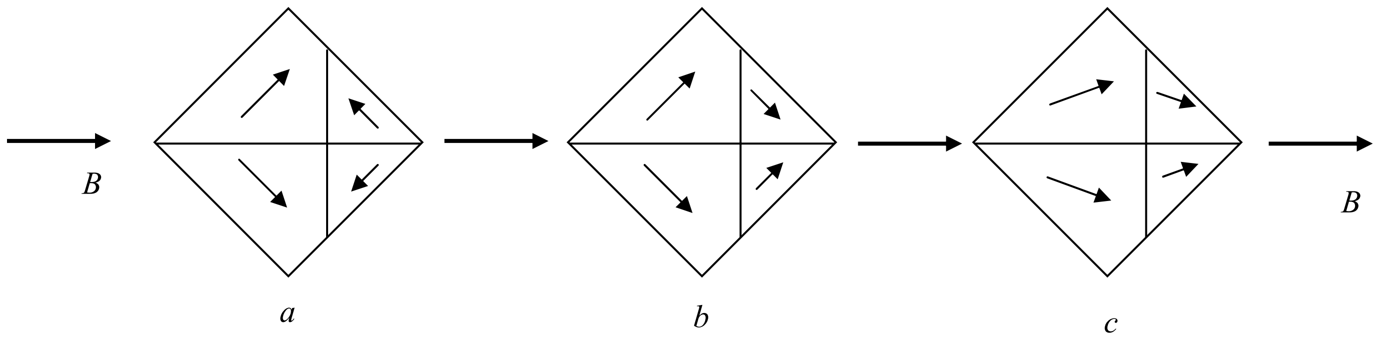
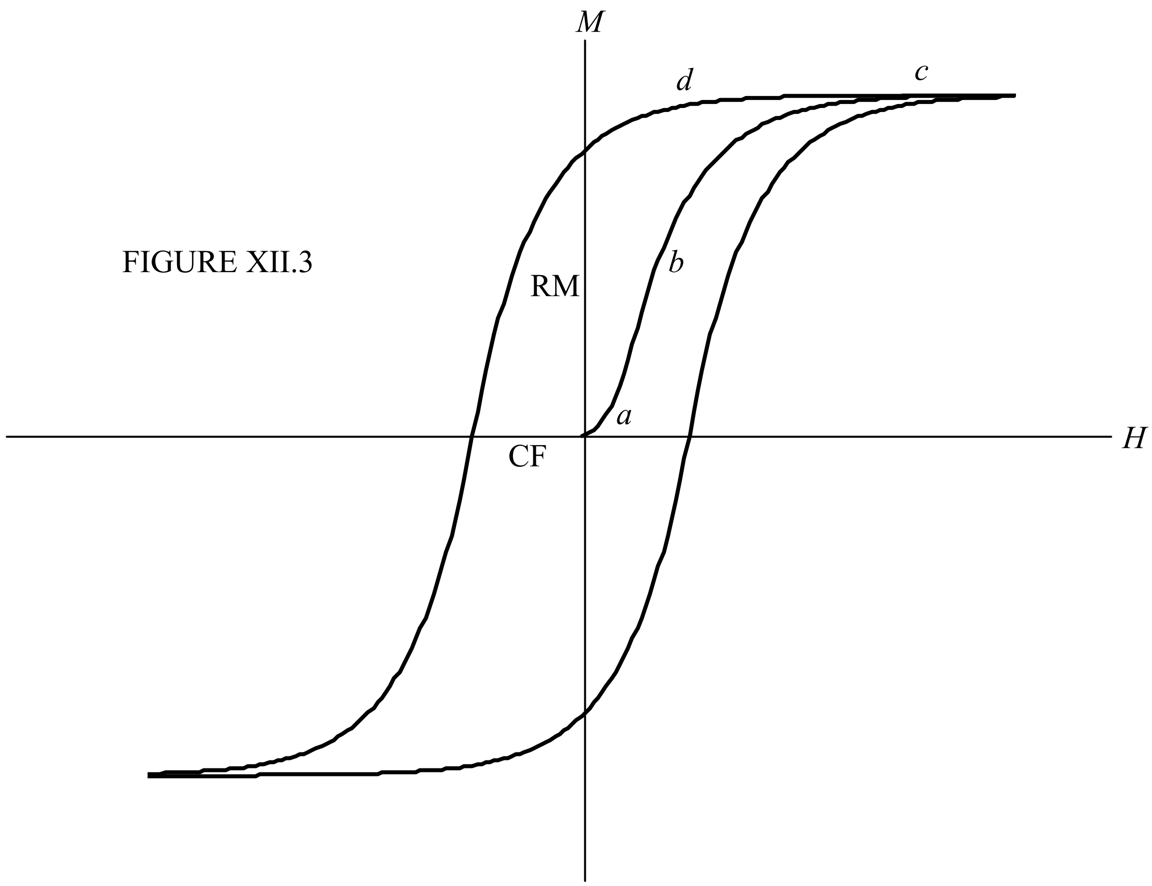


FIGURE XII.2

FIGURE XII.3



When we first apply a weak field (a), the Bloch walls (domain boundaries) move so that the favorably-oriented domains grow at the expense of the opposing domains, and the magnetization slowly increases. With stronger fields (*b*), suddenly all the magnetic moments (due to unpaired spins) within a single domain change direction almost in unison, so that an opposing domain suddenly becomes a favorable domain; this happens to one domain after another, until all domains are oriented favorably, and the magnetization of the specimen rapidly increases. For yet stronger fields (*c*), the magnetic moments, usually oriented parallel to a crystal axis, bend so that they are in the direction of the magnetizing field. When all of that is achieved, no further magnetization is possible, and the specimen is saturated.

Now, if the field is reduced, the magnetic moments relax and take up their normal positions parallel to a crystallographic axis. But, as the field is further reduced (*d*), there is no reason for the domains to reverse their polarity as happened at stage (*b*). That is, when stage (*b*) originally happened, this was an *irreversible process*. The demagnetization curve does not follow the magnetization curve in reverse. Consequently, when the magnetizing field has been reduced to zero, the specimen retains a *remanent magnetization* (indicated by RM in figure XIII.3), with all domains still favorably oriented. In order to reduce the magnetization to zero, you have to apply a field in the reverse direction. The reverse field needed to reduce the magnetization to zero is called the *coercive force* (indicated by CF in figure XIII.3).

As you repeatedly magnetize the specimen first on one direction and then the other, the graph of magnetization versus magnetizing field describes the *hysteresis* loop indicated in figure XII.3. Because of the irreversible process (*b*), magnetic energy is dissipated as heat during a complete cycle, the about of energy loss being proportional to the area of the hysteresis loop. The amount of the hysteresis depends on how freely the domain walls can move, which in turn depends on the physical and chemical constitution of the magnetic materials, particularly on the number of impurities present that can inhibit Bloch wall movement. For a permanent magnet, you need a material with a fat hysteresis loop, with a large remanent magnetization as well as a large coercive force, so that it cannot be demagnetized easily. For a transformer core, you need a material with a narrow hysteresis loop.

If you put a magnetic material inside a solenoid with alternating current inside the solenoid, the magnetization will repeatedly go around the hysteresis loop. If you now gradually decrease the amplitude of the current in the solenoid, the hysteresis loop will gradually become smaller and smaller, vanishing to a point (H and M both zero) when the current is reduced to zero. This provides a method of demagnetizing a specimen.

It is distressing how often one reads of the "remnant" magnetization. I have even encountered over-enthusiastic copy-editors who will change an author's correct spelling "remanent" to the incorrect "remnant". The difference is that "remnant" is a noun (as in a remnant of cloth) and "remanent", which is pronounced with three distinct syllables, is an adjective, meaning "remaining".

12.7 *Antiferromagnetism*

I include this largely for completeness, but I am obliged to be brief, because it is a subject I know little about. It is my understanding that it involves materials in which the atoms or ions or molecules have a permanent dipole moment (resulting from unpaired electron spins), as in paramagnetic and ferromagnetic materials, and the crystals have domain structure, as in ferromagnetic materials, but alternating ions within a domain have their magnetic moments oriented in opposite directions, so the domain as a whole has zero magnetization, or zero susceptibility. An example of an antiferromagnetic material is manganese oxide MnO, in which the Mn^{++} ion has a magnetic moment. Such materials are generally antiferromagnetic at low temperatures. As the temperature is increased, the domain structure breaks down and the material becomes paramagnetic – as also happens, of course, with ferromagnetic materials. But whereas the susceptibility of a ferromagnetic material decreases dramatically with rising temperature, until it become merely paramagnetic, the susceptibility of an antiferromagnetic material starts at zero, and its transformation to a paramagnetic material results in an *increase* (albeit a small increase) in its susceptibility. As the temperature is raised still further, the paramagnetic susceptibility drops (as is usual for paramagnetics), so there is presumably some temperature at which the susceptibility is a maximum.

12.8 *Ferrimagnetism*

This section will be shorter still, because I know even less about it! It is my understanding that, like ferromagnetics and antiferromagnetics, there is a domain structure, and, like antiferromagnetics, alternate magnetic moments are pointing in opposite directions. But this does not result is complete cancellation of the magnetization of a domain. This often results if the alternating atoms or ions within a domain are *different species*, *with unequal magnetic moments*.

CHAPTER 13
ALTERNATING CURRENTS

13.1 *Alternating current in an inductance*

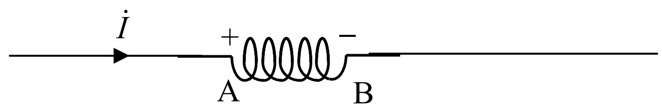


FIGURE XIII.1

In the figure we see a current increasing to the right and passing through an inductor. As a consequence of the inductance, a back EMF will be induced, with the signs as indicated. I denote the back EMF by $V = V_A - V_B$. The back EMF is given by $V = L\dot{I}$.

Now suppose that the current is an alternating current given by

$$I = \hat{I}\sin\omega t. \qquad 13.1.1$$

In that case $\dot{I} = \hat{I}\omega\cos\omega t$, and therefore the back EMF is

$$V = \hat{I}L\omega\cos\omega t, \qquad 13.1.2$$

which can be written

$$V = \hat{V}\cos\omega t, \qquad 13.1.3$$

where the peak voltage is

$$\hat{V} = L\omega\hat{I} \qquad 13.1.4$$

and, of course $V_{RMS} = L\omega I_{RMS}$.

The quantity $L\omega$ is called the *inductive reactance* X_L. It is expressed in ohms (check the dimensions), and, the higher the frequency, the greater the reactance. (The frequency ν is $\omega/(2\pi)$.)

Comparison of equations 13.1.1 and 13.1.3 shows that the current and voltage are out of phase, and that V leads on I by 90°, as shown in figure XIII.2.

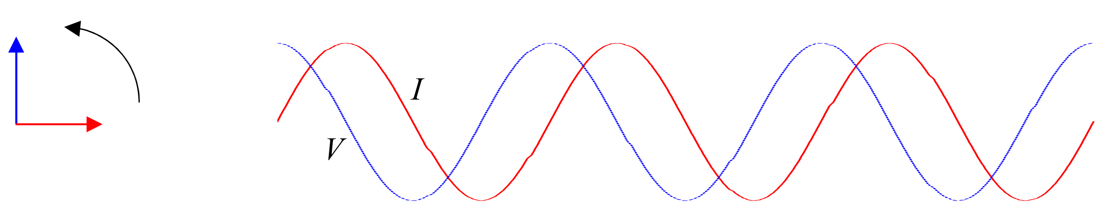


FIGURE XIII.2

13.2 *Alternating Voltage across a Capacitor*

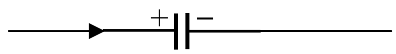


FIGURE XIII.3

At any time, the charge Q on the capacitor is related to the potential difference V across it by $Q = CV$. If there is a current in the circuit, then Q is changing, and $I = C\dot{V}$.

Now suppose that an alternating voltage given by

$$V = \hat{V} \sin \omega t \qquad 13.2.1$$

is applied across the capacitor.

In that case the current is $$I = C\omega\hat{V} \cos \omega t, \qquad 13.2.2$$

which can be written $$I = \hat{I} \cos \omega t, \qquad 13.2.3$$

where the peak current is $$\hat{I} = C\omega\hat{V} \qquad 13.2.4$$

and, of course $I_{\text{RMS}} = C\omega V_{\text{RMS}}$.

The quantity $1/(C\omega)$ is called the *capacitive reactance* X_C. It is expressed in ohms (check the dimensions), and, the higher the frequency, the smaller the reactance. (The frequency ν is $\omega/(2\pi)$.)

Comparison of equations 13.2.1 and 13.2.3 shows that the current and voltage are out of phase, and that V lags behind I by 90°, as shown in figure XIII.4.

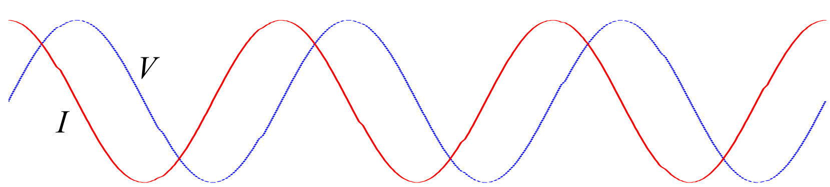


FIGURE XIII.4

13.3 *Complex Numbers*

I am now going to repeat the analyses of Sections 13.1 and 13.2 using the notation of complex numbers. In the context of alternating current theory, the imaginary unit is customarily given the symbol j rather than i, so that the symbol i is available, if need be, for electric currents. I am making the assumption that the reader is familiar with the basics of complex numbers; without that background, the reader may have difficulty with much of this chapter.

We start with the inductance. If the current is changing, there will be a back EMF given by $V = L\dot{I}$. If the current is changing as

$$I = \hat{I}e^{j\omega t}, \tag{13.3.1}$$

then $\dot{I} = \hat{I}j\omega e^{j\omega t} = j\omega I$. Therefore the voltage is given by

$$V = jL\omega I. \tag{13.3.2}$$

The quantity $jL\omega$ is called the *impedance* of the inductor, and is j times its reactance. Equation 13.3.2 (in particular the operator j on the right hand side) tells us that V leads on I by 90°.

Now suppose that an alternating voltage is applied across a capacitor. The charge on the capacitor at any time is $Q = CV$, and the current is $I = C\dot{V}$. If the voltage is changing as

$$V = \hat{V}e^{j\omega t}, \tag{13.3.3}$$

then $\dot{V} = \hat{V}j\omega e^{j\omega t} = j\omega V$. Therefore the current is given by

$$I = jC\omega V. \tag{13.3.4}$$

That is to say

$$V = -\frac{j}{C\omega}I. \tag{13.3.5}$$

The quantity $-j/(C\omega)$ is called the *impedance* of the capacitor, and is $-j$ times its reactance. Equation 13.3.5 (in particular the operator $-j$ on the right hand side) tells us that V lags behind I by 90°.

In summary:

Inductor:	Reactance = $L\omega$.	Impedance = $jL\omega$.	V leads on I.
Capacitor:	Reactance = $1/(C\omega)$.	Impedance = $-j/(C\omega)$.	V lags behind I.

It may be that at this stage you haven't got a very clear idea of the distinction between reactance (symbol X) and impedance (symbol Z) other than that one seems to be j or $-j$ times the other. The next section deals with a slightly more complicated situation, namely a resistor and an inductor in series. (In practice, it may be one piece of equipment, such as a solenoid, that has both resistance and inductance.) Paradoxically, you may find it easier to understand the distinction between impedance and reactance from this more complicated situation.

13.4 *Resistance and Inductance in Series*

The impedance is just the sum of the resistance of the resistor and the impedance of the inductor:

$$Z = R + jL\omega. \tag{13.4.1}$$

Thus the impedance is a *complex number*, whose real part R is the resistance and whose imaginary part $L\omega$ is the reactance. For a pure resistance, the impedance is real, and V and I are in phase. For a pure inductance, the impedance is imaginary (reactive), and there is a 90° phase difference between V and I.

The voltage and current are related by

$$V = IZ = (R + jL\omega)I. \tag{13.4.2}$$

Those who are familiar with complex numbers will see that this means that V leads on I, not by 90°, but by the *argument* of the complex impedance, namely $\tan^{-1}(L\omega / R)$. Further the ratio of the peak (or RMS) voltage to the peak (or RMS) current is equal to the *modulus* of the impedance, namely $\sqrt{R^2 + L^2\omega^2}$.

13.5 *Resistance and Capacitance in Series*

Likewise the impedance of a resistance and a capacitance in series is

$$Z = R - j/(C\omega). \tag{13.5.1}$$

The voltage and current are related, as usual, by $V = IZ$. Equation 13.5.1 shows that the voltage lags behind the current by $\tan^{-1}[1/(RC\omega)]$, and that $\hat{V}/\hat{I} = \sqrt{R^2 + 1/(C\omega)^2}$.

13.6 *Admittance*

In general, the impedance of a circuit is partly resistive and partly reactive:

$$Z = R + jX. \tag{13.6.2}$$

The real part is the resistance, and the imaginary part is the reactance. The relation between V and I is $V = IZ$. If the circuit is purely resistive, V and I are in phase. If is it purely reactive, V and I differ in phase by 90°. The reactance may be partly inductive and partly capacitive, so that

$$Z = R + j(X_{\mathrm{L}} - X_{\mathrm{C}}). \tag{13.6.3}$$

Indeed we shall describe such a system in detail in the next section.

The reciprocal of the impedance Z is the *admittance*, Y.

Thus
$$Y = \frac{1}{Z} = \frac{1}{R + jX}. \qquad 13.6.4$$

And of course, since $V = IZ$, $I = VY$.

Whenever we see a complex (or a purely imaginary) number in the denominator of an expression, we always immediately multiply top and bottom by the complex conjugate, so equation 13.6.4 becomes

$$Y = \frac{Z^*}{|Z|^2} = \frac{R - jX}{R^2 + X^2} . \qquad 13.6.5$$

This can be written

$$Y = G + jB, \qquad 13.6.6$$

where the real part, G, is the *conductance*:

$$G = \frac{R}{R^2 + X^2}, \qquad 13.6.7$$

and the imaginary part, B, is the *susceptance*:

$$B = -\frac{X}{R^2 + X^2}. \qquad 13.6.8$$

The SI unit for admittance, conductance and susceptance is the *siemens* (or the "mho" in informal talk).

I leave it to the reader to show that

$$R = \frac{G}{G^2 + B^2} \qquad 13.6.9$$

and
$$X = -\frac{B}{G^2 + B^2}. \qquad 13.6.10$$

13.7 *The RLC Series Acceptor Circuit*

A resistance, inductance and a capacitance in series is called an "acceptor" circuit, presumably because, for some combination of the parameters, the magnitude of the inductance is a minimum, and so current is accepted most readily. We see in figure XIII.5 an alternating voltage $V = \hat{V}e^{j\omega t}$ applied across such an R, L and C.

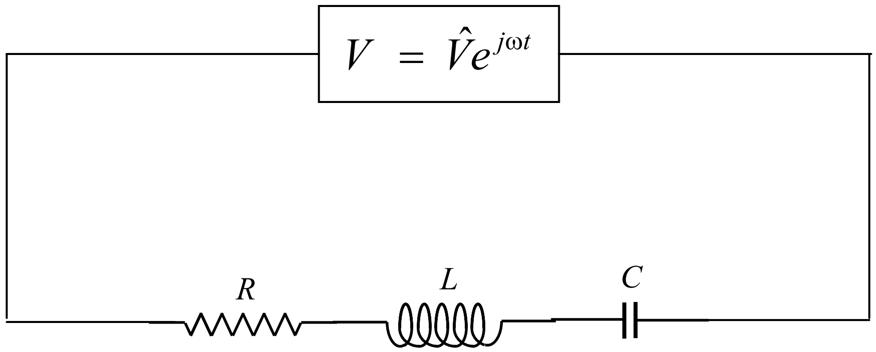


FIGURE XIII.5

The impedance is

$$Z = R + j\left(L\omega - \frac{1}{C\omega}\right). \qquad 13.7.1$$

We can see that the voltage leads on the current if the reactance is positive; that is, if the inductive reactance is greater than the capacitive reactance; that is, if $\omega > 1/\sqrt{LC}$. (Recall that the frequency, ν, is $\omega/(2\pi)$). If $\omega < 1/\sqrt{LC}$, the voltage lags behind the current. And if $\omega = 1/\sqrt{LC}$, the circuit is purely resistive, and voltage and current are in phase.

The magnitude of the impedance (which is equal to $\hat{V}/\hat{I}$) is

$$|Z| = \sqrt{R^2 + (L\omega - 1/(C\omega))^2}, \qquad 13.7.2$$

and this is least (and hence the current is greatest) when $\omega = 1/\sqrt{LC}$, the resonant frequency, which I shall denote by ω_0.

It is of interest to draw a graph of how the magnitude of the impedance varies with frequency for various values of the circuit parameters. I can reduce the number of parameters by defining the dimensionless quantities

$$\Omega = \omega/\omega_0 \qquad 13.7.3$$

$$Q = \frac{1}{R}\sqrt{\frac{L}{C}} \qquad 13.7.4$$

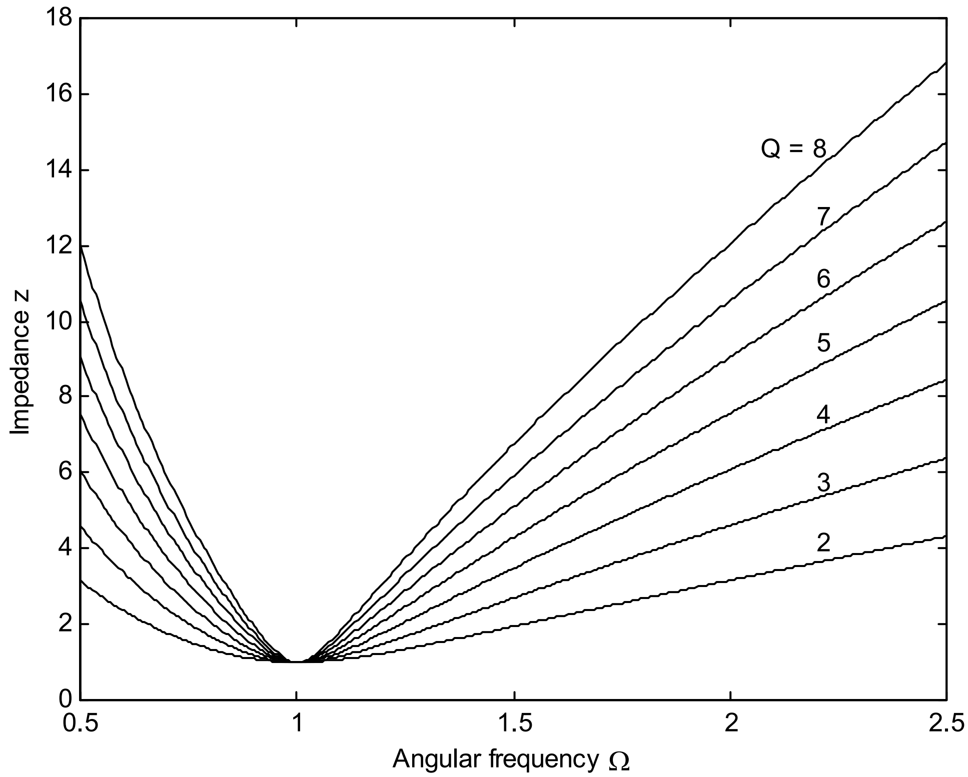
and $$z = \frac{|Z|}{R}. \qquad 13.7.4$$

You should verify that Q is indeed dimensionless. We shall see that the sharpness of the resonance depends on Q, which is known as the *quality factor* (hence the symbol Q). In terms of the dimensionless parameters, equation 13.7.2 becomes

$$z = \sqrt{1 + Q^2(\Omega - 1/\Omega)^2}\,. \qquad 13.7.5$$

This is shown in figure XIII.6, in which it can be seen that the higher the quality factor, the sharper the resonance.

FIGURE XIII.6

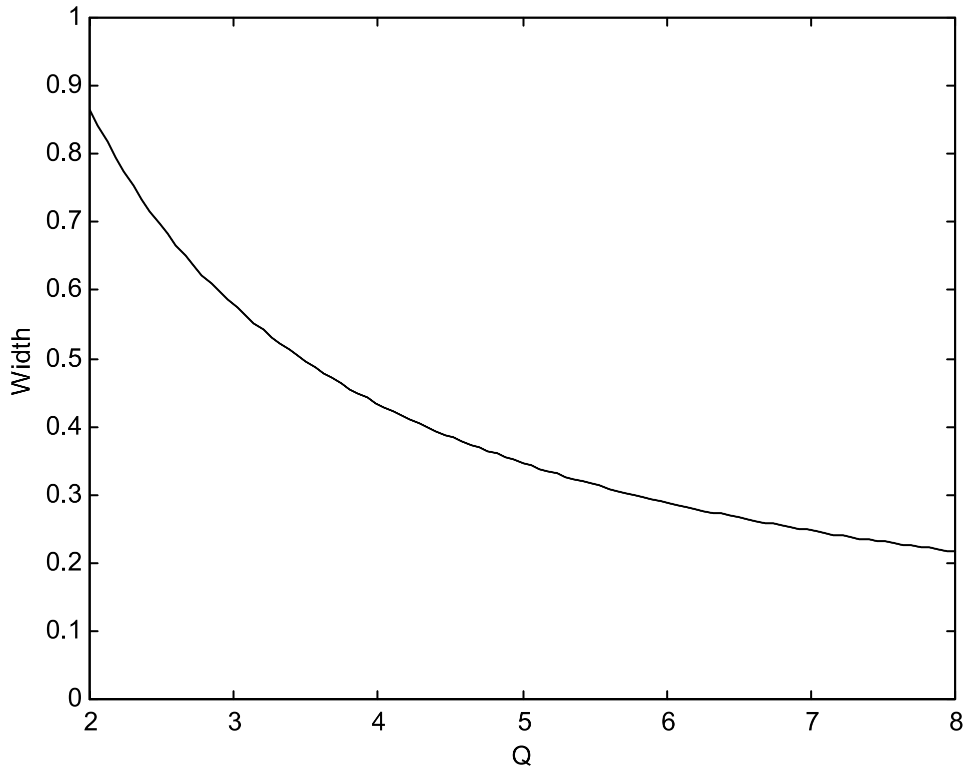


In particular, it is easy to show that the frequencies at which the impedance is twice its minimum value are given by the positive solutions of

$$\Omega^4 - \left(2 + \frac{3}{Q^2}\right)\Omega^2 + 1 = 0. \qquad 13.7.6$$

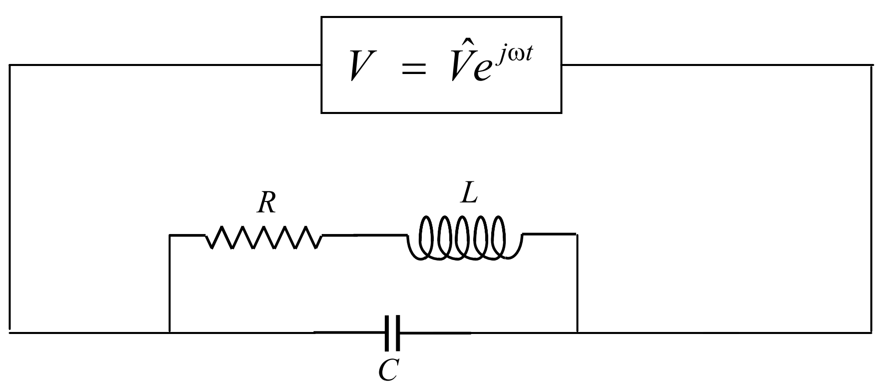
If I denote the smaller and larger of these solutions by Ω_- and Ω_+, then $\Omega_+ - \Omega_-$ will serve as a useful description of the width of the resonance, and this is shown as a function of quality factor in figure XIII.7.

FIGURE XIII.7



13.8 *The RLC Parallel Rejector Circuit*

In the circuit below, the magnitude of the *admittance* is least for certain values of the parameters. When you tune a radio set, you are changing the overlap area (and hence the capacitance) of the plates of a variable air-spaced capacitor so that the admittance is a minimum for a given frequency, so as to ensure the highest potential difference across the circuit. This resonance, as we shall see, does not occur for an angular frequency of exactly $1/\sqrt{LC}$, but at an angular frequency that is approximately this if the resistance is small.



The admittance is

$$Y = jC\omega + \frac{1}{R + jL\omega}. \qquad 13.8.1$$

After some routine algebra (multiply top and bottom by the conjugate; then collect real and imaginary parts), this becomes

$$Y = \frac{R + j\omega(L^2C\omega^2 + R^2C - L)}{R^2 + L^2\omega^2}. \qquad 13.8.2$$

The magnitude of the admittance is least when the susceptance is zero, which occurs at an angular frequency of

$$\omega_0^2 = \frac{1}{LC} - \frac{R^2}{L^2}. \qquad 13.8.3$$

If $R << \sqrt{L/C}$, this is approximately $1/\sqrt{LC}$.

13.9 *AC Bridges*

We have already met, in Chapter 4, Section 4.11, the Wheatstone bridge, which is a DC (direct current) bridge for comparing resistances, or for "measuring" an unknown resistance if it is compared with a known resistance. In the Wheatstone bridge (figure IV.9), balance is achieved when $\frac{R_1}{R_2} = \frac{R_3}{R_4}$. Likewise in a AC (alternating current) bridge, in which the power supply is an AC generator, and there are impedances (combinations of R, L and C) in each arm (figure XIII.8),

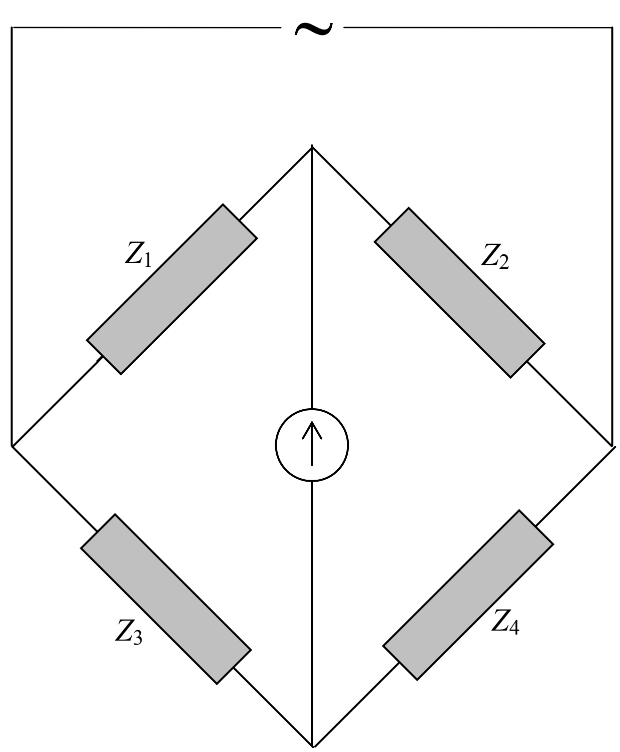


FIGURE XIII.8

balance is achieved when

$$\frac{Z_1}{Z_2} = \frac{Z_3}{Z_4} \qquad 13.9.1$$

or, of course, $\frac{Z_1}{Z_3} = \frac{Z_2}{Z_4}$. This means not only that the RMS potentials on both sides of the detector must be equal, but they must be *in phase*, so that the potentials are the same *at all times.* (I have drawn the "detector" as though it were a galvanometer, simply because that is easiest for me to draw. In practice, it might be a pair of earphones or an oscilloscope.) Each side of equation 13.9.1 is a complex number, and two complex numbers are equal if and only if their real and imaginary parts are separately equal. Thus equation 13.9.1 really represents two equations – which are necessary in order to satisfy the two conditions that the potentials on either side of the detector are equal in magnitude and in phase.

We shall look at three examples of AC bridges. It is not recommended that these be committed to memory. They are described only as examples of how to do the calculation.

13 9.1 *The Owen Bridge*

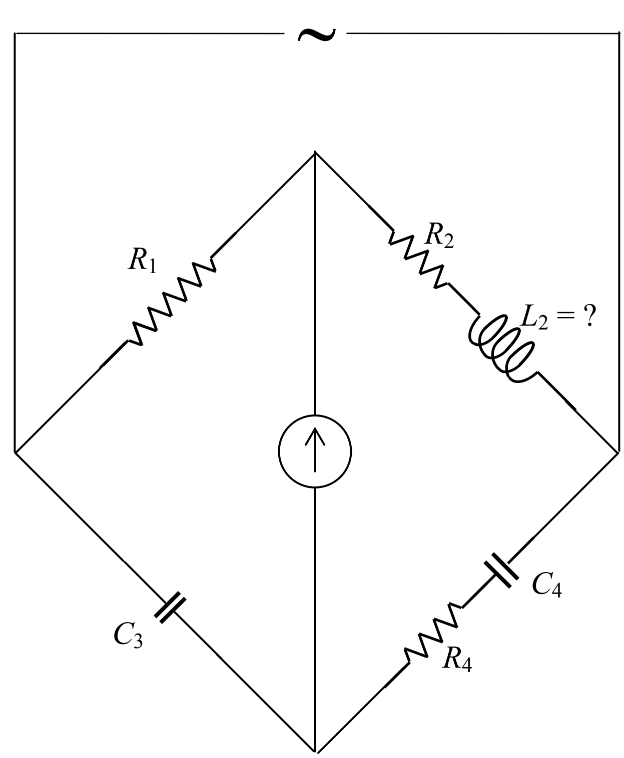


FIGURE XIII.9

This bridge can be used for measuring inductance. Note that the unknown inductance is the only inductance in the bridge. Reactance is supplied by the capacitors.

Equation 13.9.1 in this case becomes

$$\frac{R_1}{R_2 + jL_2\omega} = \frac{-j/(C_3\omega)}{R_4 - j/(C_4\omega)}. \qquad 13.9.2$$

That is,

$$R_1R_4 - j\frac{R_1}{C_4\omega} = \frac{L_2}{C_3} - j\frac{R_2}{C_3\omega}. \qquad 13.9.3$$

On equating real and imaginary parts separately, we obtain

$$L_2 = R_1R_4C_3 \qquad 13.9.4$$

and

$$\frac{R_1}{R_2} = \frac{C_4}{C_3}. \qquad 13.9.5$$

13 9.2 *The Schering Bridge*

This bridge can be used for measuring capacitance.

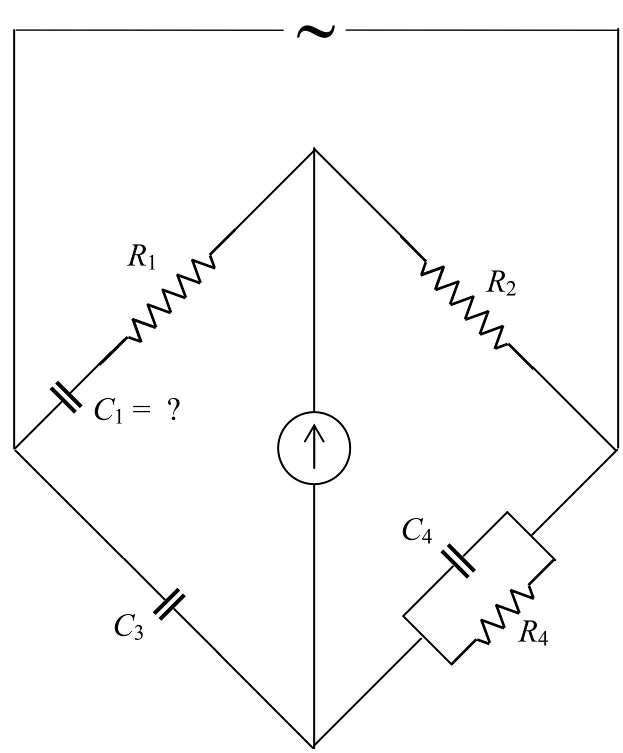


FIGURE XIII.10

The admittance of the fourth arm is $\frac{1}{R_4} + jC_4\omega$, and its impedance is the reciprocal of this. I leave the reader to balance the bridge and to show that

$$\frac{R_1}{R_2} = \frac{C_4}{C_3} \qquad 13.9.6$$

and

$$C_1 = \frac{C_3 R_4}{R_2}. \qquad 13.9.7$$

13 9.3 *The Wien Bridge*

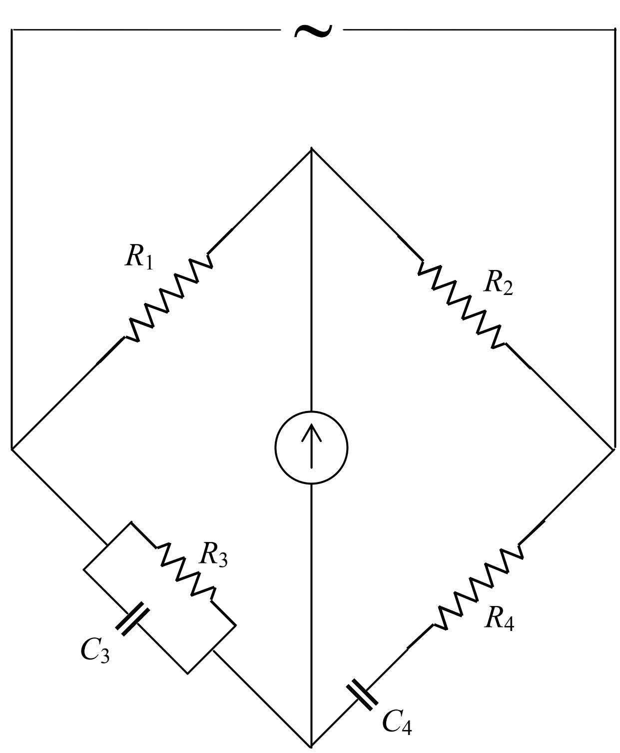


FIGURE XIII.11

This bridge can be used for measuring frequency.

The reader will, I think, be able to show that

$$\frac{R_4}{R_3} + \frac{C_3}{C_4} = \frac{R_2}{R_1} \qquad 13.9.8$$

and

$$\omega^2 = \frac{1}{R_3 R_4 C_3 C_4}\,. \qquad 13.9.10$$

13.10 *The Transformer*

We met the transformer briefly in Section 10.9. There we pointed out that the EMF induced in the secondary coil is equal to the number of turns in the secondary coil times the rate of change of magnetic flux; and the flux is proportional to the EMF applied to the primary times the number of turns in the primary. Hence we deduced the well known relation

$$\frac{V_2}{V_1} = \frac{N_2}{N_1} \qquad 13.10.1$$

relating the primary and secondary voltages to the number of turns in each. We now look at the transformer in more detail; in particular, we look at what happens when we connect the secondary coil to a circuit and take power from it.

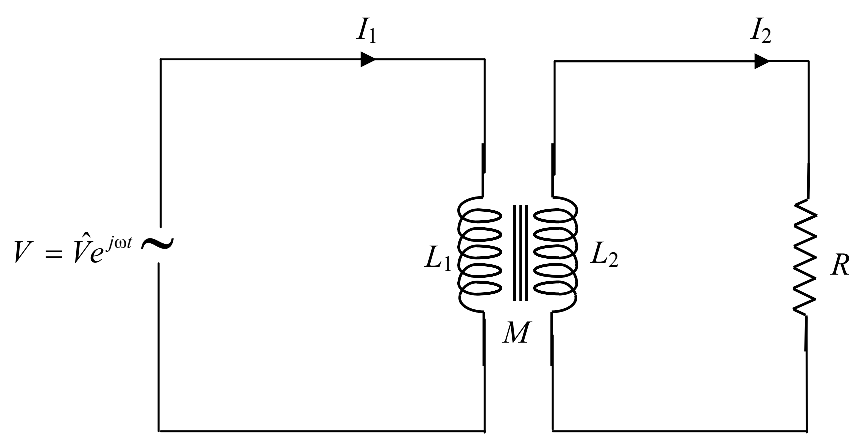


FIGURE XIII.12

In figure XIII.12, we apply an AC EMF $V = \hat{V}e^{j\omega t}$ to the primary circuit. The self inductance of the primary coil is L_1, and an alternating current I_1 flows in the primary circuit. The self inductance of the secondary coil is L_2, and the mutual inductance of the two coils is M. If the coupling between the two coils is very tight, then $M = \sqrt{L_1 L_2}$; otherwise it is less than this. I am supposing that the resistance of the primary circuit is much smaller than the reactance, so I am going to neglect it.

The secondary coil is connected to a resistance R. An alternating current I_2 flows in the secondary circuit.

Let us apply Ohm's law (or Kirchhoff's second rule) to each of the two circuits.

In the *primary* circuit, the applied EMF V is opposed by two back EMF's:

$$V = L_1\dot{I}_1 + M\dot{I}_2. \tag{13.10.2}$$

That is to say $$V = j\omega L_1 I_1 + j\omega M I_2. \tag{13.10.3}$$

Similarly for the *secondary* circuit:

$$0 = j\omega M I_1 + j\omega L_2 I_2 + RI_2. \tag{13.10.4}$$

These are two simultaneous equations for the currents, and we can (with a small effort) solve them for I_1 and I_2:

$$\left[\frac{RL_1}{M} + j\left(\frac{\omega L_1 L_2}{M} - \omega M\right)\right] I_1 = \left(\frac{L_2}{M} - j\frac{R}{\omega M}\right)V \tag{13.10.5}$$

and $$\left[R + j\left(\omega L_2 - \frac{\omega M^2}{L_1}\right)\right] I_2 = -\frac{MV}{L_1}. \tag{13.10.6}$$

This would be easier to understand if we were to do the necessary algebra to write these in the forms $I_1 = (a + jb)V$ and $I_2 = (c + jd)V$. We could then easily see the phase relationships between the current and V as well as the peak values of the currents. There is no reason why we should not try this, but I am going to be a bit lazy before I do it, and I am going to assume that we have a well designed transformer in which the secondary coil is really tightly wound around the primary, and $M = \sqrt{L_1 L_2}$. If you wish, you may carry on with a less efficient transformer, with $M = k\sqrt{L_1 L_2}$, where k is a coupling coefficient less than 1, but I'm going to stick with $M = \sqrt{L_1 L_2}$. In that case, equations 13.10.5 and 6 eventually take the forms

$$I_1 = \left(\frac{L_2}{L_1 R} - j\frac{1}{L_1\omega}\right)V = \left(\frac{N_2^2}{N_1^2 R} - j\frac{1}{L_1\omega}\right)V \tag{13.10.7}$$

and $$I_2 = -\frac{1}{R}\sqrt{\frac{L_2}{L_1}}V = -\frac{N_2}{N_1 R}V. \tag{13.10.8}$$

These equations will tell us, on examination, the magnitudes of the currents, and their phases relative to V.

Now look at the circuit shown in figure XIII.13.

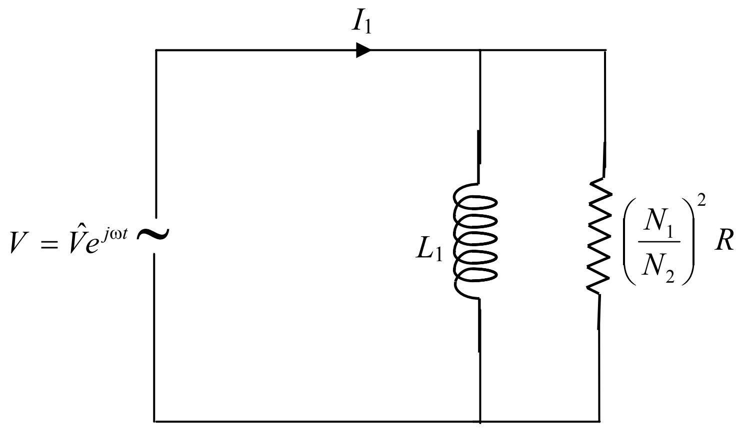


FIGURE XIII.13

In figure XIII.13 we have a resistance $R(N_1/N_2)^2$ in parallel with an inductance L_1. The admittances of these two elements are, respectively, $(N_2/N_1)^2/R$ and $-j/(L_1\omega)$, so the total admittance is $\frac{N_2^2}{N_1^2 R} - j\frac{1}{L_1\omega}\cdot$ Thus, as far as the relationship between current and voltage is concerned, the primary circuit of the transformer is precisely equivalent to the circuit drawn in figure XIII.13. To see the relationship between I_1 and V, we need look no further than figure XIII.13.

Likewise, equation 13.10.8 shows us that the relationship between I_2 and V is exactly as if we had an AC generator of EMF N_2E/N_1 connected across R, as in figure XIII.14.

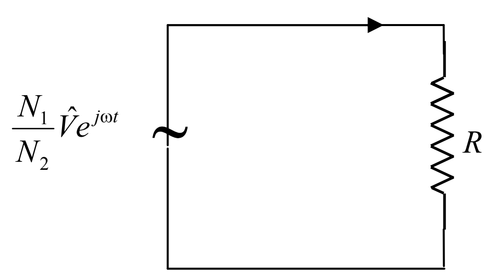


FIGURE XIII.14

Note that, if the secondary is short-circuited (i.e. if $R = 0$ and if the resistance of the secondary coil is literally zero) both the primary and secondary current become infinite. If the secondary circuit is left open (i.e. $R = \infty$), the secondary current is zero (as expected), and the primary current, also as expected, is not zero but is $-jV/(L_1\omega)$; That is to say, the current is of magnitude $V/(L_1\omega)$ and it lags behind the voltage by 90^{o}, just as if the secondary circuit were not there.

CHAPTER 14
LAPLACE TRANSFORMS

14.1 *Introduction*

If y(x) is a function of x, where x lies in the range 0 to ∞, then the function $\bar{y}(p)$, defined by

$$\bar{y}(p) = \int_0^\infty e^{-px} y(x)dx \ , \qquad 14.1.1$$

is called the *Laplace transform* of y(x). However, in this chapter, where we shall be applying Laplace transforms to electrical circuits, y will most often be a voltage or current that is varying with *time* rather than with "x". Thus I shall use t as our variable rather than x, and I shall use s rather than p (although it will be noted that, as yet, I have given no particular physical meaning to either p or to s.) Thus I shall define the Laplace transform with the notation

$$\bar{y}(s) = \int_0^\infty e^{-st} y(t)dt, \qquad 14.1.2$$

it being understood that t lies in the range 0 to ∞.

For short, I could write this as

$$\bar{y}(s) = \mathbf{L}\, y(t). \qquad 14.1.3$$

When we first learned differential calculus, we soon learned that there were just a few functions whose derivatives it was worth committing to memory. Thus we learned the derivatives of x^n, $\sin x$, e^x and a very few more. We found that we could readily find the derivatives of more complicated functions by means of a few simple rules, such as how to differentiate a product of two functions, or a function of a function, and so on. Likewise, we have to know only a very few basic Laplace transforms; there are a few simple rules that will enable us to calculate more complicated ones.

After we had learned differential calculus, we came across integral calculus. This was the inverse process from differentiation. We had to ask: What function would we have had to differentiate in order to arrive at this function? It was as though we were given the answer to a problem, and had to deduce what the question was. It will be a similar situation with Laplace transforms. We shall often be given a function $\bar{y}(s)$, and we shall want to know: what function y(t) is this the Laplace transform of? In other words, we shall need to know the *inverse Laplace transform*:

$$y(t) = \mathbf{L}^{-1}\bar{y}(s). \qquad 14.1.4$$

We shall find that facility in calculating Laplace transforms and their inverses leads to very quick ways of solving some types of differential equations – in particular the types of differential equations that arise in electrical theory. We can use Laplace transforms to see the relations between varying current and voltages in circuits containing resistance, capacitance and inductance.

However, these methods are quick and convenient only if we are in constant daily practice in dealing with Laplace transforms with easy familiarity. Few of us, unfortunately, have the luxury of calculating Laplace transforms and their inverses on a daily basis, and they lose many of their advantages if we have to refresh our memories and regain our skills every time we may want to use them. It may therefore be asked: Since we already know perfectly well how to do AC calculations using complex numbers, is there any point in learning what just amounts to another way of doing the same thing? There is an answer to that. The theory of AC circuits that we developed in Chapter 13 using complex numbers to find the relations between current and voltages dealt primarily with *steady state conditions*, in which voltages and current were varying sinusoidally. It did not deal with the *transient* effects that might happen in the first few moments after we switch on an electrical circuit, or situations where the time variations are *not sinusoidal*. The Laplace transform approach will deal equally well with steady state, sinusoidal, non-sinusoidal and transient situations.

14.2 *Table of Laplace Transforms*

It is easy, by using equation 14.1.2, to derive all of the transforms shown in the following table, in which $t > 0$. (Do it!)

$y(t)$	$\bar{y}(s)$
1	$1/s$
t	$1/s^2$
$\dfrac{t^{n-1}}{(n-1)!}$	$1/s^n$
$\sin at$	$\dfrac{a}{s^2 + a^2}$
$\cos at$	$\dfrac{s}{s^2 + a^2}$
$\sinh at$	$\dfrac{a}{s^2 - a^2}$
$\cosh at$	$\dfrac{s}{s^2 - a^2}$
e^{at}	$\dfrac{1}{s-a}$

This table can, of course, be used to find inverse Laplace transforms as well as direct transforms. Thus, for example, $\mathbf{L}^{-1}\dfrac{1}{s-1} = e^t$. In practice, you may find that you are using it more often to find inverse transforms than direct transforms.

These are really all the transforms that it is necessary to know – and they need not be committed to memory if this table is handy. For more complicated functions, there are rules for finding the transforms, as we shall see in the following sections, which introduce a number of theorems. Although I shall derive some of these theorems, I shall merely state others, though perhaps with an example. Many (not all) of them are straightforward to prove, but in any case I am more anxious to introduce their applications to circuit theory than to write a formal course on the mathematics of Laplace transforms.

After you have understood some of these theorems, you may well want to apply them to a number of functions and hence greatly expand your table of Laplace transforms with results that you will discover on application of the theorems.

14.3 *The First Integration Theorem*

The theorem is:

$$\mathbf{L}\int_0^t y(x)dx = \frac{\bar{y}(s)}{s}. \qquad 14.3.1$$

Before deriving this theorem, here's a quick example to show what it means. The theorem is most useful, as in this example, for finding an *inverse* Laplace transform. I.e. $\mathbf{L}^{-1}\dfrac{\bar{y}(s)}{s} = \int_0^t y(x)dx$.

Calculate

$$\mathbf{L}^{-1}\frac{1}{s(s-a)}.$$

Solution. From the table, we see that $\mathbf{L}^{-1}\dfrac{1}{s-a} = e^{at}$. The integration theorem tells us that $\mathbf{L}^{-1}\dfrac{1}{s(s-a)} = \int_0^t e^{ax}dx = (e^{at}-1)/a.$ You should now verify that this is the correct answer by substituting this in equation 14.1.2 and integrating – or (and!) using the table of Laplace transforms.

The proof of the theorem is just a matter of integrating by parts. Thus

$$\mathbf{L}\int_0^t y(x)\,dx = \int_0^\infty\left(\int_0^t y(x)dx\right)e^{-st}dt = -\frac{1}{s}\int_0^\infty\left(\int_0^t y(x)dx\right)d\left(e^{-st}\right)$$

$$= \left[-\frac{1}{s}e^{-st}\int_0^t y(x)dx\right]_{t=0}^\infty + \frac{1}{s}\int_0^\infty e^{-st}y(t)\,dt.$$

The expression in brackets is zero at both limits, and therefore the theorem is proved.

14.4 *The Second Integration Theorem (Dividing a Function by t)*

This theorem looks very like the first integration theorem, but "the other way round". It is

$$\mathbf{L}\left(\frac{y(t)}{t}\right) = \int_s^\infty \bar{y}(x)dx. \qquad 14.4.1$$

I'll leave it for the reader to derive the theorem. Here I just give an example of its use. Whereas the first integration theorem is most useful in finding inverse transforms, the second integration theorem is more useful for finding direct transforms.

Example: Calculate $\mathbf{L}\left(\frac{\sin at}{t}\right)$.

This means calculate $\int_0^\infty \frac{e^{-st}\sin at}{t}dt.$

While this integral can no doubt be done, you may find it a bit daunting, and the second integration theorem provides an alternative way of doing it, resulting in an easier integral.

Note that the right hand side of equation 14.4.1 is a function of s, not of x, which is just a dummy variable. The function $\bar{y}(x)$ is the Laplace transform, with x as argument, of $y(t)$. In our particular case, $y(t)$ is sin at, so that, from the table, $\bar{y}(x) = \frac{a}{a^2 + x^2}\cdot$ The second integration theorem , then, tells us that $\mathbf{L}\left(\frac{\sin at}{t}\right) = \int_s^\infty \frac{a}{a^2+x^2}dx.$ This is a much easier integral. It is $\left[\tan^{-1}\left(\frac{x}{a}\right)\right]_s^\infty = \frac{\pi}{2} - \tan^{-1}\left(\frac{s}{a}\right) = \tan^{-1}\left(\frac{a}{s}\right).$ You may want to add this result to your table of Laplace integrals. Indeed, you may already want to expand the table considerably by applying both integration theorems to several functions.

14.5 *Shifting Theorem*

This is a very useful theorem, and one that is almost trivial to prove. (Try it!) It is

$$\mathbf{L}\left(e^{-at}y(t)\right) = \bar{y}(s+a). \qquad 14.5.1$$

For example, from the table, we have $\mathbf{L}(t) = 1/s^2$. The shifting theorem tells us that

$\mathbf{L}\left(te^{-at}\right) = 1/(s+a)^2$. I'm sure you will now want to expand your table even more. Or you may want to go the other way, and cut down the table a bit! After all, you know that $\mathbf{L}(1) = 1/s$. The shifting theorem, then, tells you that $\mathbf{L}(e^{at}) = 1/(s-a)$, so that entry in the table is superfluous! Note that you can use the theorem to deduce either direct or inverse transforms.

14.6 *A Function Times t^n*

I'll just give this one with out proof:

For n a positive integer, $$L(t^n y) = (-1)^n \frac{d^2\overline{y}}{ds^n}. \qquad 14.6.1$$

Example: What is $\mathbf{L}(t^2 e^{-t})$?

Answer: For $y = e^{-t}$, $\overline{y} = 1/(s+1)$. $\therefore L(t^2e^{-t}) = 2/(s+1)^3$.

Before proceeding further, I strongly recommend that you now apply theorems 14.3.1, 14.4.1, 14.5.1 and 14.6.1 to the several entries in your existing table of Laplace transforms and greatly expand your table of Laplace transforms. For example, you can already add $(\sin at)/t$, te^{-at} and t^2e^{-t} to the list of functions for which you have calculated the Laplace transforms.

14.7 *Differentiation Theorem*

$$\mathbf{L}\left(\frac{d^n y}{dt^n}\right) = s^n\overline{y} - s^{n-1}y_0 - s^{n-2}\left(\frac{dy}{dt}\right)_0 - s^{n-3}\left(\frac{d^2y}{dt^2}\right)_0 - \dots \quad \dots - s\left(\frac{d^{n-2}y}{dt^{n-2}}\right)_0 - \left(\frac{d^{n-1}y}{dt^{n-1}}\right)_0. \qquad 14.7.1$$

This looks formidable, and you will be tempted to skip it – but don't, because it is essential! However, to make it more palatable, I'll point out that one rarely, if ever, needs derivatives higher than the second, so I'll re-write this for the first and second derivatives, and they will look much less frightening.

$$\mathbf{L}\dot{y} = s\overline{y} - y_0 \qquad 14.7.2$$

and

$$\mathbf{L}\ddot{y} = s^2\overline{y} - sy_0 - \dot{y}_0. \qquad 14.7.3$$

Here, the subscript zero means "evaluated at $t = 0$".

Equation 14.7.2 is easily proved by integration by parts:

$$\overline{y} = \mathbf{L}y = \int_0^\infty ye^{-st}dt = -\frac{1}{s}\int_0^\infty yde^{-st} = -\frac{1}{s}\left[ye^{-st}\right]_0^\infty + \frac{1}{s}\int_{t=0}^\infty e^{-st}dy$$

$$= \frac{1}{s}y_0 + \frac{1}{s}\int \dot{y}dt = \frac{1}{s}y_0 + \frac{1}{s}\mathbf{L}\dot{y}. \qquad 14.7.4$$

$$\therefore \qquad L\dot{y} = s\overline{y} - y_0. \qquad 14.7.5$$

From this, $L\ddot{y} = s\bar{\dot{y}} - \dot{y}_0 = sL\dot{y} - \dot{y}_0 = s(s\overline{y}-y_0) - \dot{y}_0 = s^2\overline{y} - sy_0 - \dot{y}_0.$ 14.7.6

Apply this over and over again, and you arrive at equation 14.7.1

14.8 *A First Order Differential Equation*

Solve $\dot{y} + 2y = 3te^t$, with initial condition $y_0 = 0$.

If you are in good practice with solving this type of equation, you will probably multiply it through by e^{2t}, so that it becomes

$$\frac{d}{dt}\left(ye^{2t}\right) = 3te^{3t},$$

from which $y = (t - \tfrac{1}{3})e^t + Ce^{-2t}.$

(You can now substitute this back into the original differential equation, to verify that it is indeed the correct solution.)

With the given initial condition, it is quickly found that $C = \tfrac{1}{3}$, so that the solution is

$$y = te^t - \tfrac{1}{3}e^t + \tfrac{1}{3}e^{-2t}.$$

Now, here's the same solution, using Laplace transforms.

We take the Laplace transform of both sides of the original differential equation:

$$s\overline{y} + 2\overline{y} = 3\mathbf{L}(te^t) = \frac{3}{(s-1)^2}.$$

Thus $\overline{y} = \dfrac{3}{(s+2)(s-1)^2}.$

Partial fractions: $\overline{y} = \dfrac{1}{3}\left(\dfrac{1}{s+2}\right) - \dfrac{1}{3}\left(\dfrac{1}{s-1}\right) + \dfrac{1}{(s-1)^2}.$

Inverse transforms: $y = \frac{1}{3}e^{-2t} - \frac{1}{3}e^{t} + te^{t}.$

You will probably admit that you can follow this, but will say that you can do this at speed only after a great deal of practice with many similar equations. But this is equally true of the first method, too.

14.9 *A Second Order Differential Equation*

Solve $$\ddot{y} - 4\dot{y} + 3y = e^{-t}$$

with initial conditions $y_0 = 1, \quad \dot{y}_0 = -1.$

You probably already know some method for solving this equation, so please go ahead and do it. Then, when you have finished, look at the solution by Laplace transforms.

Laplace transform: $s^2\overline{y} - s + 1 - 4(s\overline{y} - 1) + 3\overline{y} = 1/(s+1).$

(My! Wasn't that fast!)

A little algebra: $$\overline{y} = \frac{1}{(s-3)(s-1)(s+1)} + \frac{s-5}{(s-3)(s-1)}.$$

Partial fractions: $$\overline{y} = \frac{1}{8}\left(\frac{1}{s-3} - \frac{2}{s-1} + \frac{1}{s+1}\right) + \frac{2}{s-1} - \frac{1}{s-3},$$

or $$\overline{y} = \frac{1}{8}\left(\frac{1}{s+1}\right) + \frac{7}{4}\left(\frac{1}{s-1}\right) - \frac{7}{8}\left(\frac{1}{s-3}\right).$$

Inverse transforms: $y = \frac{1}{8}e^{-t} + \frac{7}{4}e^{t} - \frac{7}{8}e^{3t}$, and you can verify that this is correct by substitution in the original differential equation.

So: We have found a new way of solving differential equations. If (but only if) we have a lot of practice in manipulating Laplace transforms, and have used the various manipulations to prepare a slightly larger table of transforms from the basic table given above, and we can go from *t* to *s* and from *s* to *t* with equal facility, we can believe that our new method can be both fast and easy.

But, what has this to do with electrical circuits? Read on.

14.10 *Generalized Impedance*

We have dealt in Chapter 13 with a sinusoidally varying voltage applied to an inductance, a resistance and a capacitance in series. The equation that governs the relation between voltage and current is

$$V = L\dot{I} + RI + Q/C. \qquad 14.10.1$$

If we multiply by C, differentiate with respect to time, and write I for $\dot{Q}$, this becomes just

$$C\dot{V} = LC\ddot{I} + RC\dot{I} + I. \qquad 14.10.2$$

If we suppose that the applied voltage V is varying sinusoidally (that is, $V = \hat{V}e^{j\omega t}$, or, if you prefer, $V = \hat{V}\sin\omega t$), then the operator d^2/dt^2, or "double dot", is equivalent to multiplying by $-\omega^2$, and the operator d/dt, or "dot", is equivalent to multiplying by $j\omega$. Thus equation 14.10.2 is equivalent to

$$j\omega CV = -LC\omega^2 I + jRC\omega I + I. \qquad 14.10.3$$

That is,

$$V = [R + jL\omega + 1/jC\omega)]I. \qquad 14.10.4$$

The complex expression inside the brackets is the now familiar impedance Z, and we can write

$$V = IZ. \qquad 14.10.5$$

But what if V is not varying *sinusoidally*? Suppose that V is varying in some other manner, perhaps not even periodically? This might include, as one possible example, the situation where V is constant and not varying with time at all. But whether or not V varying with time, equation 14.10.2 is still valid – except that, unless the time variation is sinusoidally, we cannot substitute $j\omega$ for d/dt. We are faced with having to solve the differential equation 10.4.2.

But we have just learned a neat new way of solving differential equations of this type. We can take the Laplace transform of each side of the equation. Thus

$$C\overline{\dot{V}} = LC\overline{\ddot{I}} + RC\overline{\dot{I}} + \bar{I}. \qquad 14.10.6$$

Now we are going to make use of the differentiation theorem, equations 14.7.2 and 14.7.3.

$$C(s\overline{V} - V_0) = LC(s^2\bar{I} - sI_0 - \dot{I}_0) + RC(s\bar{I} - I_0) + \bar{I}. \qquad 14.10.7$$

Let us suppose that, at $t = 0$, V_0 and I_0 are both zero – i.e. before $t = 0$ a switch was open, and we close the switch at $t = 0$. Furthermore, since the circuit contains inductance, the current cannot change instantaneously, and, since it contains capacitance, the voltage cannot change instantaneously, so the equation becomes

$$\overline{V} = (R + Ls + 1/Cs)\bar{I}. \tag{14.10.8}$$

This is so regardless of the form of the variation of V: it could be sinusoidal, it could be constant, or it could be something quite different. This is a *generalized Ohm's law*. The *generalized impedance* of the circuit is $R + Ls + \frac{1}{Cs}\cdot$ Recall that in the complex number treatment of a steady-state sinusoidal voltage, the complex impedance was $R + jL\omega + \frac{1}{jC\omega}\cdot$

To find out how the current varies, all we have to do is to take the inverse Laplace transform of

$$\bar{I} = \frac{\overline{V}}{R + Ls + 1/(Cs)}. \tag{14.10.9}$$

We look at a couple of examples in the next sections.

14.11 *RLC Series Transient*

A battery of constant EMF V is connected to a switch, and an R, L and C in series. The switch is closed at time $t = 0$. We'll first solve this problem by "conventional" methods; then by Laplace transforms. The reader who is familiar with the mechanics of damped oscillatory motion, such as is dealt with in Chapter 11 of the Classical Mechanics notes of this series, may have an advantage over the reader for whom this topic is new – though not necessarily so!

"Ohm's law" is
$$V = Q/C + RI + L\dot{I}, \tag{14.11.1}$$

or
$$LC\ddot{Q} + RC\dot{Q} + Q = CV. \tag{14.11.2}$$

Those who are familiar with this type of equation will recognize that the general solution (complementary function plus particular integral) is

$$Q = Ae^{\lambda_1 t} + Be^{\lambda_2 t} + CV, \tag{14.11.3}$$

where
$$\lambda_1 = -\frac{R}{2L} + \sqrt{\frac{R^2}{4L^2} - \frac{1}{LC}} \text{ and } \lambda_2 = -\frac{R}{2L} - \sqrt{\frac{R^2}{4L^2} - \frac{1}{LC}}. \tag{14.11.4}$$

(Those who are not familiar with the solution of differential equations of this type should not give up here. Just go on to the part where we do this by Laplace transforms. You'll soon be streaking ahead of your more learned colleagues, who will be struggling for a while.)

Case I. $\dfrac{R^2}{4L^2} - \dfrac{1}{LC}$ is positive. For short I'm going to write equations 14.11.4 as

$$\lambda_1 = -a + k \quad \text{and} \quad \lambda_2 = -a - k. \tag{14.11.5}$$

Then
$$Q = Ae^{-(a-k)t} + Be^{-(a+k)t} + CV \tag{14.11.6}$$

and, by differentiation with respect to time,

$$I = -A(a-k)e^{-(a-k)t} - B(a+k)e^{-(a+k)t}. \tag{14.11.7}$$

At $t = 0$, Q and I are both zero, from which we find that

$$A = -\frac{(a+k)CV}{2k} \quad \text{and} \quad B = \frac{(a-k)CV}{2k}. \tag{14.11.8}$$

Thus
$$Q = \left[-\left(\frac{a+k}{2k}\right)e^{-(a-k)t} + \left(\frac{a-k}{2k}\right)e^{-(a+k)t} + 1\right]CV \tag{14.11.9}$$

and
$$I = \left[\left(\frac{a^2-k^2}{2k}\right)\left(e^{-(a-k)t} - e^{-(a+k)t}\right)\right]CV. \tag{14.11.10}$$

On recalling the meanings of a and k and the sinh function, and a little algebra, we obtain

$$I = \frac{V}{Lk}e^{-at}\sinh kt. \tag{14.11.11}$$

Exercise: Verify that this equation is dimensionally correct. Draw a graph of $I : t$. The current is, of course, zero at $t = 0$ and ∞. What is the maximum current, and when does it occur?

Case II. $\dfrac{R^2}{4L^2} - \dfrac{1}{LC}$ is zero. In this case, those who are in practice with differential equations will obtain for the general solution

$$Q = e^{\lambda t}(A + Bt) + CV, \tag{14.11.12}$$

where
$$\lambda = -R/(2L), \tag{14.11.13}$$

from which
$$I = \lambda(A + Bt)e^{\lambda t} + Be^{\lambda t}. \tag{14.11.14}$$

After applying the initial conditions that Q and I are initially zero, we obtain

$$Q = CV\left[1 - \left(1 - \frac{Rt}{2L}\right)e^{-Rt/(2L)}\right] \qquad 14.11.15$$

and $$I = \frac{V}{L}te^{-Rt/(2L)}. \qquad 14.11.16$$

As in case II, this starts and ends at zero and goes through a maximum, and you may wish to calculate what the maximum current is and when it occurs.

Case III. $\frac{R^2}{4L^2} - \frac{1}{LC}$ is negative. In this case, I am going to write equations 14.11.4 as

$$\lambda_1 = -a + j\omega \quad \text{and} \quad \lambda_2 = -a - j\omega, \qquad 14.11.17$$

where $$a = \frac{R}{2L} \quad \text{and} \quad \omega^2 = \frac{1}{LC} - \frac{R^2}{4L^2}. \qquad 14.11.18$$

All that is necessary, then, is to repeat the analysis for Case I, but to substitute $-\omega^2$ for k^2 and $j\omega$ for k, and, provided that you know that $\sinh j\omega t = j\sin\omega t$, you finish with

$$I = \frac{V}{L\omega}e^{-at}\sin\omega t. \qquad 14.11.19$$

This is lightly damped oscillatory motion.

Now let us try the same problem using Laplace transforms. Recall that we have a V in series with an R , L and C, and that initially Q, I and $\dot{I}$ are all zero. (The circuit contains capacitance, so Q cannot change instantaneously; it contains inductance, so I cannot change instantaneously.)

Immediately, automatically and with scarcely a thought, our first line is the generalized Ohm's law, with the Laplace transforms of V and I and the generalized impedance:

$$\overline{V} = [R + Ls + 1/(Cs)]\bar{I}. \qquad 14.11.20$$

Since V is constant, reference to the very first entry in your table of transforms shows that $\overline{V} = V/s$, and so

$$\bar{I} = \frac{V}{s[R + Ls + 1/(Cs)]} = \frac{V}{L(s^2 + bs + c)}, \qquad 14.11.21$$

where $$b = R/L \quad \text{and} \quad c = 1/(LC). \qquad 14.11.22$$

Case I. $b^2 > 4c$.

$$\bar{I} = \frac{V}{L}\left(\frac{1}{(s-\alpha)(s-\beta)}\right) = \frac{V}{L}\left(\frac{1}{\alpha-\beta}\right)\left(\frac{1}{s-\alpha} - \frac{1}{s-\beta}\right). \qquad 14.11.23$$

Here, of course, $2\alpha = -b + \sqrt{b^2 - 4c}$ and $2\beta = -b - \sqrt{b^2 - 4c}$ 14.11.24

On taking the inverse transforms, we find that

$$I = \frac{V}{L}\left(\frac{1}{\alpha-\beta)}\right)(e^{\alpha t} - e^{\beta t}). \qquad 14.11.25$$

From there it is a matter of routine algebra (do it!) to show that this is exactly the same as equation 14.11.11.

In order to arrive at this result, it wasn't at all necessary to know how to solve differential equations. All that was necessary was to understand generalized impedance and to look up a table of Laplace transforms.

Case II. $b^2 = 4c$.

In this case, equation 14.11.21 is of the form

$$\bar{I} = \frac{V}{L}\cdot\frac{1}{(s-\alpha)^2}, \qquad 14.11.26$$

where $\alpha = -\frac{1}{2}b$. If you have dutifully expanded your original table of Laplace transforms, as suggested, you will probably already have an entry for the inverse transform of the right hand side. If not, you know that the Laplace transform of t is $1/s^2$, so you can just apply the shifting theorem to see that the Laplace transform of $t\,e^{\alpha t}$ is $1/(s-\alpha)^2$. Thus

$$I = \frac{V}{L}te^{\alpha t} \qquad 14.11.27$$

which is the same as equation 14.11.16.

[Gosh – what could be quicker and easier than that!?]

Case III. $b^2 < 4c$.

This time, we'll complete the square in the denominator of equation 14.11. 21:

$$\bar{I} \;=\; \frac{V}{L}\cdot\frac{1}{(s+\tfrac{1}{2}b)^2 \;+\; (c-\tfrac{1}{4}b^2)} \;=\; \frac{V}{L\omega}\frac{\omega}{(s+\tfrac{1}{2}b)^2 \;+\; \omega^2}\,, \qquad 14.11.28$$

where I have introduced ω with obvious notation.

On taking the inverse transform (from our table, with a little help from the shifting theorem) we obtain

$$I \;=\; \frac{V}{L\omega}\cdot e^{-\frac{1}{2}bt}\sin\omega t\,, \qquad 14.11.29$$

which is the same as equation 14.11.19.

With this brief introductory chapter to the application of Laplace transforms to electrical circuitry, we have just opened a door by a tiny crack to glimpse the potential great power of this method. With practice, it can be used to solve complicated problems of many sorts with great rapidity. All we have so far is a tiny glimpse. I shall end this chapter with just one more example, in the hope that this short introduction will whet the reader's appetite to learn more about this technique.

14.12 *Another Example*

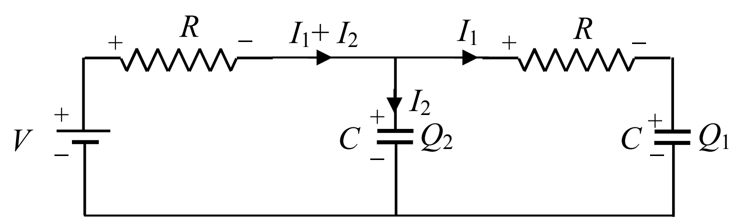


FIGURE XIV.1

The circuit in figure XIV.1 contains two equal resistances, two equal capacitances, and a battery. The battery is connected at time $t = 0$. Find the charges held by the capacitors after time t.

Apply Kirchhoff's second rule to each half:

$$(\dot{Q}_1 + \dot{Q}_2)RC \;+ Q_2 = CV, \qquad 14.12.1$$

and

$$\dot{Q}_1RC + Q_1 - Q_2 \;=\; 0. \qquad 14.12.2$$

Eliminate Q_2:

$$R^2C^2\ddot{Q}_1 + 3RCQ_1 + Q_1 \;= CV. \qquad 14.12.3$$

Transform, with Q_1 and $\dot{Q}_1$ initially zero:

$$(R^2C^2s^2 + 3RCs + 1)\overline{Q_1} = \frac{CV}{s}. \tag{14.12.4}$$

I.e.
$$R^2C\overline{Q_1} = \frac{1}{s(s^2 + 3as + a^2)} \cdot V, \tag{14.12.5}$$

where
$$a = 1/(RC). \tag{14.12.6}$$

That is
$$R^2C\overline{Q_1} = \frac{1}{s(s + 2.618a)(s + 0.382a)} V. \tag{14.12.7}$$

Partial fractions:
$$R^2C\overline{Q_1} = \left[\frac{1}{s} + \frac{0.1708}{s+2.618a} - \frac{1.1708}{s+0.382a}\right]\frac{V}{a^2}. \tag{14.12.8}$$

That is,
$$\overline{Q_1} = \left[\frac{1}{s} + \frac{0.1708}{s+2.618a} - \frac{1.1708}{s+0.382a}\right] CV. \tag{14.12.8}$$

Inverse transform:
$$Q_1 = \left[1 + 0.1708e^{-2.618\,t/(RC)} - 1.1708e^{-0.382\,t/(RC)}\right] \tag{14.12.9}$$

The current can be found by differentiation.

I leave it to the reader to eliminate Q_1 from equations 14.12.1 and 2 and hence to show that

$$Q_2 = \left[1 - 0.2764e^{-2.618\,t/(RC)} - 0.7236e^{-0.382\,t/(RC)}\right]. \tag{14.12.10}$$

CHAPTER 15
MAXWELL'S EQUATIONS

15.1 *Introduction*

One of Newton's great achievements was to show that all of the phenomena of classical mechanics can be deduced as consequences of three basic, fundamental laws, namely Newton's laws of motion. It was likewise one of Maxwell's great achievements to show that all of the phenomena of classical electricity and magnetism – all of the phenomena discovered by Oersted, Ampère, Henry, Faraday and others whose names are commemorated in several electrical units – can be deduced as consequences of four basic, fundamental equations. We describe these four equations in this chapter, and, in passing, we also mention Poisson's and Laplace's equations. We also show how Maxwell's equations predict the existence of electromagnetic waves that travel at a speed of 3×10^8 m s^{-1}. This is the speed at which light is measured to move, and one of the most important bases of our belief that light is an electromagnetic wave.

Before embarking upon this, we may need a reminder of two mathematical theorems, as well as a reminder of the differential equation that describes wave motion.

The two mathematical theorems that we need to remind ourselves of are:

The surface integral of a vector field over a closed surface is equal to the volume integral of its divergence.

The line integral of a vector field around a closed plane curve is equal to the surface integral of its curl.

A function $f(x - \upsilon t)$ represents a function that is moving with speed υ in the positive x-direction, and a function $g(x + \upsilon t)$ represents a function that is moving with speed υ in the negative x-direction. It is easy to verify by substitution that $y = Af + Bg$ is a solution of the differential equation

$$\frac{d^2 y}{dt^2} = \upsilon^2 \frac{d^2 y}{dx^2}. \qquad 15.1.1$$

Indeed it is the most general solution, since f and g are quite general functions, and the function y already contains the only two arbitrary integration constants to be expected from a second order differential equation. Equation 15.1.1 is, then, the differential equation for a wave in one dimension. For a function $\psi(x, y, z)$ in three dimensions, the corresponding wave equation is

$$\ddot{\psi} = \upsilon^2 \nabla^2 \psi. \qquad 15.1.2$$

It is easy to remember which side of the equation υ^2 is on from dimensional considerations.

One last small point before proceeding – I may be running out of symbols! I may need to refer to *surface charge density*, a scalar quantity for which the usual symbol is σ. I shall also need to refer to *magnetic vector potential*, for which the usual symbol is **A**. And I shall need to refer to *area*, for which either of the symbols A or σ are commonly used – or, if the vector nature of area is to be emphasized, **A** or **σ**. What I shall try to do, then, to avoid this difficulty, is to use **A** for magnetic vector potential, and **σ** for area, and I shall try to avoid using surface charge density in any equation. However, the reader is warned to be on the lookout and to be sure what each symbol means in a particular context.

15.2 *Maxwell's First Equation*

Maxwell's first equation, which describes the electrostatic field, is derived immediately from Gauss's theorem, which in turn is a consequence of Coulomb's inverse square law. Gauss's theorem states that the surface integral of the electrostatic field **D** over a closed surface is equal to the charge enclosed by that surface. That is

$$\int_{\text{surface}} \mathbf{D}\cdot\mathbf{d\sigma} = \int_{\text{volume}} \rho\, dv. \qquad 15.2.1$$

Here ρ is the charge per unit volume.

But the surface integral of a vector field over a closed surface is equal to the volume integral of its divergence, and therefore

$$\int_{\text{volume}} \text{div}\mathbf{D}\, dv = \int_{\text{volume}} \rho\, dv. \qquad 15.2.2$$

Therefore $$\text{div}\,\mathbf{D} = \rho\,, \qquad 15.2.3$$

or, in the nabla notation, $$\nabla\cdot\mathbf{D} = \rho. \qquad 15.2.4$$

This is the first of Maxwell's equations.

15.3 *Poisson's and Laplace's Equations*

Equation 15.2.4 can be written $\nabla\cdot\mathbf{E} = \rho/\varepsilon$. where ε is the permittivity. But **E** is minus the potential gradient; i.e. $\mathbf{E} = -\nabla V$. Therefore,

$$\nabla^2 V = -\rho/\varepsilon. \qquad 15.3.1$$

This is *Poisson's equation.* At a point in space where the charge density is zero, it becomes

$$\nabla^2 V = 0, \qquad 15.3.2$$

which is generally known as *Laplace's equation.* Thus, regardless of how many charged bodies there may be an a place of interest, and regardless of their shape or size, the potential at any point can be calculated from Poisson's or Laplace's equations. Courses in differential equations commonly discuss how to solve these equations for a variety of *boundary conditions* – by which is meant the size, shape and location of the various charged bodies and the charge carried by each. It perhaps just needs to be emphasized that Poisson's and Laplace's equations apply only for *static* fields.

15.4 *Maxwell's Second Equation*

Unlike the electrostatic field, magnetic fields have no sources or sinks, and the magnetic lines of force are closed curves. Consequently the surface integral of the magnetic field over a closed surface is zero, and therefore

$$\text{div}\,\mathbf{B} = 0\,, \qquad 15.4.1$$

or, in the nabla notation

$$\nabla \cdot \mathbf{B} = 0. \qquad 15.4.2$$

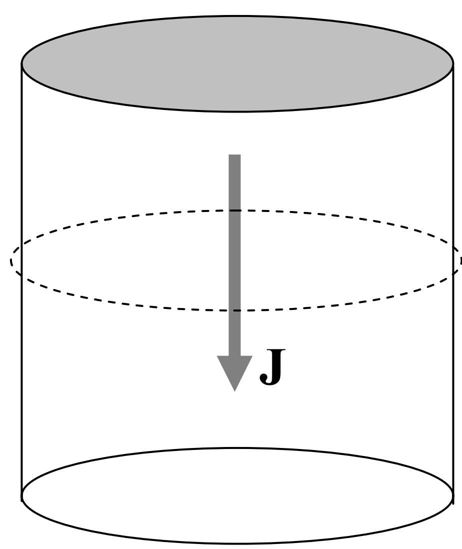
This is the second of Maxwell's equations.

15.5 *Maxwell's Third Equation*

This is derived from Ampère's theorem, which is that the line integral of the magnetic field **H** around a closed circuit is equal to the enclosed current.

Now there are two possible components to the "enclosed" current, one of which is obvious, and the other, I suppose, could also be said to be "obvious" once it has been pointed out! Let's deal with the immediately obvious one first, and look at figure XV.1.

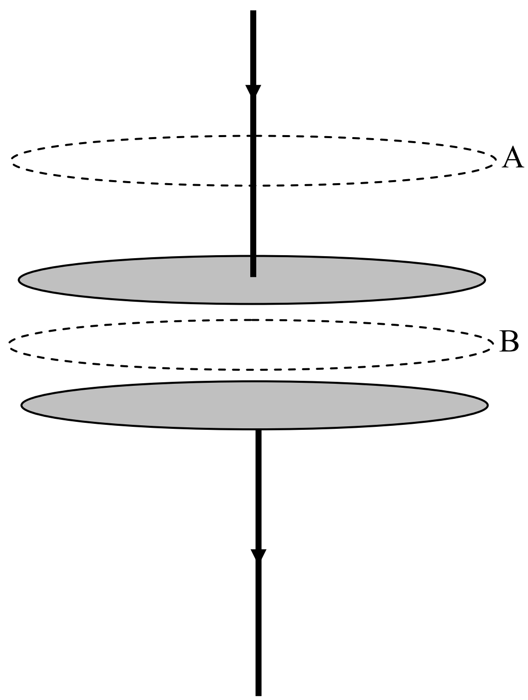
FIGURE XV.1



In figure XV.1, I am imagining a metal cylinder with current flowing from top to bottom (i.e. electrons flowing from bottom to top. It needn't be a metal cylinder, though. It could just be a volume of space with a stream of protons moving from top to bottom. In any case, the current density (which may vary with distance from the axis of the cylinder) is **J**, and the total current enclosed by the dashed circle is the integral of **J** throughout the cylinder. In a more general geometry, in which **J** is not necessarily perpendicular to the area of interest, and indeed in which the area need not be planar, this would be $\int \mathbf{J} \cdot \mathbf{d\sigma}$.

Now for the less obvious component to the "enclosed current". See figure XV.2.

FIGURE XV.2



In figure XV.2, I imagine two capacitor plates in the process of being charged. There is undoubtedly a current flowing in the connecting wires. There is a magnetic field at A, and the line integral of the field around the upper dotted curve is undoubtedly equal to the enclosed current. The current is equal to the rate at which charge is being built up on the plates. Electrons are being deposited on the lower plate and are leaving the upper plate. There is also a magnetic field at B (it doesn't suddenly stop!), and the field at B is just the same as the field at A, which is equal to the rate at which charge is being built up on the plates. The charge on the plates (which may not be uniform, and indeed won't be while the current is still flowing or if the plates are not infinite in extent) is equal to the integral of the charge density times the area. And the charge density on the plates, by Gauss's theorem, is equal to the electric field **D** between the plates. Thus the current is equal to the integral of $\dot{\mathbf{D}}$ over the surface of the plates. Thus the line integral of **H** around either of the dashed closed loops is equal to $\int \dot{\mathbf{D}} \cdot \mathbf{d\sigma}$.

In general, both types of current (the obvious one in which there is an obvious flow of charge, and the less obvious one, where the electric field is varying because of a real flow of charge elsewhere) contributes to the magnetic field, and so Ampère's theorem in general must read

$$\int_{\text{loop}} \mathbf{H} \cdot \mathbf{ds} \;=\; \int_{\text{area}} (\dot{\mathbf{D}} + \mathbf{J}) \cdot \mathbf{d\sigma}. \qquad 15.5.1$$

But the line integral of a vector field around a closed plane curve is equal to the surface integral of its curl, and therefore

$$\int_{\text{area}} \mathbf{curl\ H} \cdot \mathbf{d\sigma} \;=\; \int_{\text{area}} (\dot{\mathbf{D}} + \mathbf{J}) \cdot \mathbf{d\sigma}. \qquad 15.5.2$$

Thus we arrive at:

$$\mathbf{curl\ H} \;=\; \dot{\mathbf{D}} + \mathbf{J}, \qquad 15.5.3$$

or, in the nabla notation,

$$\nabla \times \mathbf{H} \;=\; \dot{\mathbf{D}} + \mathbf{J}. \qquad 15.5.4$$

This is the third of Maxwell's equations.

15.6 *The Magnetic Equivalent of Poisson's Equation*

This deals with a static magnetic field, where there is no electrostatic field or at least any electrostatic field is indeed static – i.e. not changing. In that case **curl H** $=$ **J**. Now the magnetic field can be derived from the curl of the magnetic vector potential, defined by the two equations

$$\mathbf{B} \;=\; \mathbf{curl\ A} \qquad 15.6.1$$

and

$$\text{div}\,\mathbf{A} \;=\; 0. \qquad 15.6.2$$

(See Chapter 9 for a reminder of this.) Together with $\mathbf{H} = \mathbf{B}/\mu$ (μ = permeability), this gives us

$$\mathbf{curl\ curl\ A} \;=\; \mu\mathbf{J}. \qquad 15.6.3$$

If we now remind ourselves of the jabberwockian-sounding vector differential operator equivalence

$$\mathbf{curl\ curl} \;\equiv\; \mathbf{grad}\ \text{div} \;-\; \text{nabla-squared}, \qquad 15.6.4$$

together with equation 15.6.2, this gives us

$$\nabla^2\mathbf{A} \;=\; -\mu\mathbf{J}. \qquad 15.6.5$$

I don't know if this equation has any particular name, but it plays the same role for static magnetic fields that Poisson's equation plays for electrostatic fields. No matter what the distribution of currents, the magnetic vector potential at any point must obey equation 15.6.5

15.7 *Maxwell's Fourth Equation*

This is derived from the laws of electromagnetic induction.

Faraday's and Lenz's laws of electromagnetic induction tell us that the E.M.F. induced in a closed circuit is equal to minus the rate of change of B-flux through the circuit. The E.M.F. around a closed circuit is the line integral of **E .ds** around the circuit, where **E** is the electric field. The line integral of **E** around the closed circuit is equal to the surface integral of its curl. The rate of change of B-flux through a circuit is the surface integral of $\dot{\mathbf{B}}$. Therefore

$$\mathbf{curl\,E} = -\dot{\mathbf{B}}, \qquad 15.7.1$$

or, in the nabla notation,

$$\nabla \times \mathbf{E} = -\dot{\mathbf{B}}. \qquad 15.7.2$$

This is the fourth of Maxwell's equations.

15.8 *Summary of Maxwell's and Poisson's Equations*

Maxwell's equations:

$$\nabla \bullet \mathbf{D} = \rho \qquad 15.8.1$$

$$\nabla \cdot \mathbf{B} = 0. \qquad 15.8.2$$

$$\nabla \times \mathbf{H} = \dot{\mathbf{D}} + \mathbf{J}. \qquad 15.8.3$$

$$\nabla \times \mathbf{E} = -\dot{\mathbf{B}}. \qquad 15.8.4$$

Sometimes you may see versions of these equations with factors such as 4π or c scattered liberally throughout them. If you do, my best advice is to white them out with a bottle of erasing fluid, or otherwise ignore them. I shall try to explain in Chapter 16 where they come from. They serve no scientific purpose, and are merely conversion factors between the many different systems of units that have been used in the past.

Poisson's equation for the potential in an electrostatic field:

$$\nabla^2 V = -\rho/\varepsilon. \qquad 15.8.5$$

The equivalent of Poisson's equation for the magnetic vector potential on a static magnetic field:

$$\nabla^2 \mathbf{A} = -\mu \mathbf{J}. \qquad 15.8.6$$

15.9 *Electromagnetic Waves*

Maxwell predicted the existence of electromagnetic waves, and these were generated experimentally by Hertz shortly afterwards. In addition, the predicted speed of the waves was 3×10^8 m s^{-1}, the same as the measured speed of light, showing that light is an electromagnetic wave.

In an isotropic, homogeneous, nonconducting, uncharged medium, where the permittivity and permeability are scalar quantities, Maxwell's equations can be written

$$\nabla \cdot \mathbf{E} = 0 \qquad 15.9.1$$

$$\nabla \cdot \mathbf{H} = 0 \qquad 15.9.2$$

$$\nabla \times \mathbf{H} = \varepsilon \dot{\mathbf{E}}. \qquad 15.9.3$$

$$\nabla \times \mathbf{E} = -\mu \dot{\mathbf{H}}. \qquad 15.9.4$$

Take the **curl** of equation 15.9.3, and make use of equation 15.6.4:

$$\mathbf{grad}\, \mathit{div}\, \mathbf{H} - \nabla^2 \mathbf{H} = \varepsilon \frac{\partial}{\partial t} \mathbf{curl}\, \mathbf{E}. \qquad 15.9.5$$

Substitute for *div* **H** and **curl E** from equations 15.9.2 and 15.9.4 to obtain

$$\nabla^2 \mathbf{H} = \varepsilon \mu \ddot{\mathbf{H}}. \qquad 15.9.6$$

Comparison with equation 15.1.2 shows that this is a wave of speed $1/\sqrt{\varepsilon\mu}$. (Verify that this has the dimensions of speed.)

In a similar manner the reader should easily be able to derive the equation

$$\nabla^2 \mathbf{E} = \varepsilon \mu \ddot{\mathbf{E}}. \qquad 15.9.7$$

In a vacuum, the speed is $1/\sqrt{\varepsilon_0 \mu_0}$. With $\mu_0 = 4\pi \times 10^{-7}$ H m^{-1} and $\varepsilon_0 = 8.854 \times 10^{-12}$ F m^{-1}, this comes to 2.998×10^8 m s^{-1}.

15.10 *Gauge Transformations*

We recall (equation 9.1.1) that a static electric field **E** can be derived from the negative of the gradient of a scalar potential function of space:

$$\mathbf{E} = -\mathbf{grad}\, V. \qquad 15.10.1$$

The zero of the potential is arbitrary. We can add any constant (with the dimensions of potential) to V. For example, if we define $V' = V + C$, where C is a constant (in the sense that it is not a function of x, y, z) then we can still calculate the electric field from $\mathbf{E} = -\mathbf{grad}\, V'$.

We also recall (equation 9.2.1) that a static magnetic field **B** can be derived from the **curl** of a magnetic vector potential function:

$$\mathbf{B} = \mathbf{curl}\, \mathbf{A}. \qquad 15.10.2$$

Let us also recall here the concept of the B-flux from equation 6.10.1:

$$\Phi_\mathrm{B} = \iint \mathbf{B} \cdot d\mathbf{A}. \qquad 15.10.3$$

It will be worth while here to recapitulate the dimensions and SI units of these quantities:

E	$\mathrm{MLT^{-2}Q^{-1}}$	$\mathrm{V\ m^{-1}}$		
B	$\mathrm{MT^{-1}Q^{-1}}$	T		
V	$\mathrm{ML^2T^{-2}Q^{-1}}$	V		
A	$\mathrm{MLT^{-1}Q^{-1}}$	T m	or	$\mathrm{Wb\ m^{-1}}$
Φ_B	$\mathrm{ML^2T^{-1}Q^{-1}}$	$\mathrm{T\ m^2}$	or	Wb

Equation 15.10.2 is also true for a nonstatic field. Thus a time-varying magnetic field can be represented by the **curl** of a time-varying magnetic vector potential. However, we know from the phenomenon of electromagnetic induction that a varying magnetic field has the same effect as an electric field, so that, if the fields are not static, the electric field is the result of an electrical potential gradient and a varying magnetic field, so that equation 15.10.1 holds only for static fields.

If we combine the Maxwell equation $\mathbf{curl}\,\mathbf{E} = -\dot{\mathbf{B}}$ with the equation for the definition of the magnetic vector potential $\mathbf{curl}\ \mathbf{A} = \mathbf{B}$, we obtain $\mathbf{curl}(\mathbf{E} + \dot{\mathbf{A}}) = \mathbf{0}$. Then, since **curl grad** of any scalar function is zero, we can define a potential function V such that

$$\mathbf{E} + \dot{\mathbf{A}} = -\mathbf{grad}\,V. \qquad 15.10.4$$

(We could have chosen a plus sign, but we choose a minus sign so that it reduces to the familiar $\mathbf{E} = -\mathbf{grad}\,V$ for a static field.) Thus equations 15.10.4 and 15.10.2 define the electric and magnetic potentials – or at least they define the **grad**ient of V and the **curl** of **A**. But we recall that, in the static case, we can add an arbitrary constant to V (as long as the constant is dimensionally similar to V), and the equation $\mathbf{E} = -\mathbf{grad}\,V'$, where $V' = V + C$, still holds. Can we find a suitable transformation for V and **A** such that equations 15.10.2 and 15.10.4 still hold in the nonstatic case? Such a transformation would be a *gauge transformation.*

Let χ be some arbitrary scalar function of space and time. I demand little of the form of χ; indeed I demand only two things. One is that it is a "well-behaved" function, in the sense that it is everywhere and at all times single-valued, continuous and differentiable. The other is that it should have dimensions $\mathrm{ML^2T^{-1}Q^{-1}}$. This is the same as the dimensions of magnetic B-flux, but I am not sure that it is particularly helpful to think of this. It *will*, however, be useful to note that the dimensions of **grad** χ and of $\dot{\chi}$ are, respectively, the same as the dimensions of magnetic vector potential (**A**) and of electric potential (V).

Let us make the transformations

$$\mathbf{A}' = \mathbf{A} - \mathbf{grad}\,\chi \qquad 15.10.5$$

and

$$V' = V + \dot{\chi}. \qquad 15.10.6$$

We shall see very quickly that this transformation (and we have a wide choice in the form of χ) preserves the forms of equations 15.10.2 and 15.10.4, and therefore this transformation (or, rather, these transformations, since χ can have any well-behaved form) are *gauge transformations.*

Thus $\mathbf{curl}\,\mathbf{A} = \mathbf{B}$ becomes $\mathbf{curl}\,(\mathbf{A}' + \mathbf{grad}\,\chi) = \mathbf{B}$. And since **curl grad** of any scalar field is zero, this becomes $\mathbf{curl}\,\mathbf{A}' = \mathbf{B}$.

Also, $\mathbf{grad}\,V = -(\mathbf{E} + \dot{\mathbf{A}})$

becomes $\mathbf{grad}\,(V' - \dot{\chi}) = -(\mathbf{E} + \dot{\mathbf{A}}' + \mathbf{grad}\dot{\chi})$, or $\mathbf{grad}\,V' = -(\mathbf{E} + \dot{\mathbf{A}}')$.

Thus the form of the equations is preserved. If we make a gauge transformation to the potentials such as equations 15.10.5 and 15.10.6, this does not change the fields **E** and **B**, so that the fields **E** and **B** are *gauge invariant.* Maxwell's equations in their usual form are expressed in terms of **E** and **B**, and are hence *gauge invariant*.

15.11 *Maxwell's Equations in Potential Form*

In their usual form, Maxwell's equations for an isotropic medium, written in terms of the fields, are

$$\text{div}\,\mathbf{D} = \rho \qquad 15.11.1$$

$$\text{div}\,\mathbf{B} = 0 \qquad 15.11.2$$

$$\mathbf{curl}\,\mathbf{H} = \dot{\mathbf{D}} + \mathbf{J} \qquad 15.11.3$$

$$\mathbf{curl}\,\mathbf{E} = -\dot{\mathbf{B}}. \qquad 15.11.4$$

If we write the fields in terms of the potentials:

$$\mathbf{E} = -\dot{\mathbf{A}} - \mathbf{grad}\,V \qquad 15.11.5$$

and

$$\mathbf{B} = \mathbf{curl}\,\mathbf{A}, \qquad 15.11.6$$

together with $\mathbf{D} = \varepsilon\mathbf{E}$ and $\mathbf{B} = \mu\mathbf{H}$, we obtain for the first Maxwell equation, after some vector calculus and algebra,

★ $$\nabla^2 V + \frac{\partial}{\partial t}(\text{div}\,\mathbf{A}) = -\frac{\rho}{\varepsilon}. \qquad 15.11.7$$

For the second equation, we merely verify that zero is equal to zero. (div **curl A** = 0.)

For the third equation, which requires a little more vector calculus and algebra, we obtain

★ $$\nabla^2\mathbf{A} - \varepsilon\mu\frac{\partial^2\mathbf{A}}{\partial t^2} = \mathbf{grad}\left(\text{div}\,\mathbf{A} + \varepsilon\mu\frac{\partial V}{\partial t}\right) - \mu\mathbf{J}. \qquad 15.11.8$$

The speed of electromagnetic waves in the medium is $1/\sqrt{\varepsilon\mu}$, and, in a vacuum, equation 15.11.8 becomes

★ $$\nabla^2\mathbf{A} - \frac{1}{c^2}\frac{\partial^2\mathbf{A}}{\partial t^2} = \mathbf{grad}\left(\text{div}\,\mathbf{A} + \frac{1}{c^2}\frac{\partial V}{\partial t}\right) - \mu_0\mathbf{J}, \qquad 15.11.9$$

where c is the speed of electromagnetic waves in a vacuum.

The fourth Maxwell equation, when written in terms of the potentials, tells us nothing new (try it), so equations 15.11.7 and 15.11.8 (or 15.11.9 *in vacuo*) are Maxwell's equations in potential form.

These equations look awfully difficult – but perhaps we can find a gauge transformation, using some form for χ, and subtracting **grad** ξ from **A** and adding $\dot{\xi}$ to V, which will make the equations much easier and which will still give the right answers for **E** and for **B**.

One of the things that make equations 15.11.7 and 15.11.9 look particularly difficult is that each equation contains both **A** and V; that is, we have two simultaneous differential equations to solve for the two potentials. It would be nice if we had one equation for **A** and one equation for V. This can be achieved, as we shall shortly see, if we can find a gauge transformation such that the potentials are related by

$$\operatorname{div}\mathbf{A} = -\frac{1}{c^2}\frac{\partial V}{\partial t}. \qquad 15.11.10$$

You should check that the two sides of this equation are dimensionally similar. What would be the SI units?

You'll see that this is chosen so as to make the "difficult" part of equation 15.11.9 zero.

If we make a gauge transformation and take the divergence of equation 15.10.5 and the time derivative of equation 15.10.6, we then see that condition 15.11.10 will be satisfied by a function χ that satisfies

$$\nabla^2\xi - \frac{1}{c^2}\frac{\partial^2\xi}{\partial t^2} = -\operatorname{div}\mathbf{A'} - \frac{1}{c^2}\frac{\partial V'}{\partial t}. \qquad 15.11.11$$

Don't worry – you don't have to solve this equation and find the function χ; you just have to be assured that some such function exists such that, when applied to the potentials, the potentials will be related by equation 15.11.10. Then, if you substitute equation 15.11.10 into Maxwell's equations in potential form (equations 15.11.7 and 15.11.9), you obtain the following forms for Maxwell's equations *in vacuo* in potential form, and the **A** and V are now separated:

★ $$\nabla^2 V - \frac{1}{c^2}\frac{\partial^2 V}{\partial t^2} = -\frac{\rho}{\varepsilon_0} \qquad 15.11.12$$

and ★ $$\nabla^2\mathbf{A} - \frac{1}{c^2}\frac{\partial^2\mathbf{A}}{\partial t^2} = -\mu_0\mathbf{J}. \qquad 15.11.13$$

And, since these equations were arrived at by a gauge transformation, their solutions, when differentiated, will give the right answers for the fields.

15.12 *Retarded Potential*

In a static situation, in which the charge density ρ, the current density **J**, the electric field **E** and potential V, and the magnetic field **B** and potential **A** are all constant in time (i.e. they are functions of x, y and z but not of t) we already know how to calculate, *in vacuo*, the electric potential from the electric charge densityand the magnetic potential from the current density. The formulas are

$$V(x,y,z) = \frac{1}{4\pi\varepsilon_0}\int\frac{\rho(x',y',z')\,d\upsilon'}{R} \qquad 15.12.1$$

and $$\mathbf{A}(x,y,z) = \frac{\mu_0}{4\pi}\int\frac{\mathbf{J}(x',y',z')\,d\upsilon'}{R}\,. \qquad 15.12.2$$

Here R is the distance between the point (x',y',z') and the point $(x,\ y,\ z)$, and υ' is a volume element at the point (x',y',z'). I can't remember if we have written these two equations in exactly that form before, but we have certainly used them, and given lots of examples of calculating V in Chapter 2, and one of calculating **A** in Section 9.3.

The question we are now going to address is whether these formulas are still valid in a nonstatic situation, in which the charge density ρ, the current density **J**, the electric field **E** and potential V, and the magnetic field **B** and potential **A** are all varying in time (i.e. they are functions of x, y, z and t). The answer is "yes, but…". The relevant formulas are indeed

$$V(x,y,z,t) = \frac{1}{4\pi\varepsilon_0}\int\frac{\rho(x',y',z',t')\,d\upsilon'}{R} \qquad 15.12.3$$

and $$\mathbf{A}(x,y,z,t) = \frac{\mu_0}{4\pi}\int\frac{\mathbf{J}(x',y',z',t')\,d\upsilon'}{R}\,, \qquad 15.12.4$$

…but notice the t' on the right hand side and the t on the left hand side! What this means is that, if $\rho(x',y',z',t')$ is the charge density at a point (x',y',z') at time t', equation 15.12.3 gives the correct potential at the point (x',y',z') at some *slightly later time t*, the time difference $t-t'$ being equal to the time R/c that it takes for an electromagnetic signal to travel from (x',y',z') to $(x,\ y,\ z)$. If the charge density at (x',y',z') changes, the information about this change cannot reach the point instantaneously; it takes a time R/c for the information to be transmitted from one point to another. The same considerations apply to the change in the magnetic potential when the current density changes, as described by equation 15.12.4. The potentials so calculated are called, naturally, the *retarded potentials.* While this result has been arrived at by a qualitative argument, in fact equations 15.12.3 and 4 can be obtained as a solution of the differential equations 15.11.12 and 13. Mathematically there is also a solution that gives an "advance potential" – that is, one in which $t'-t$ rather than $t-t'$ is equal to R/c. You can regard, if you wish, the retarded solution as the "physically acceptable" solution and discard the "advance" solution as not being physically significant. That is, the potential cannot predict in advance that the charge density is about to change, and so change its value before the charge density does. Alternatively one can think that the laws of physics, from the mathematical view at least, allow the universe to run equally well backward as well as forward, though in fact the arrow of time is such that cause must precede effect (a condition which, in relativity, leads to the conclusion that information cannot be transmitted from one place to another at a speed faster than the speed of light). One is also reminded that the laws of physics, from the mathematical view at least, allow the entropy of an isolated thermodynamical system to decrease (see Section 7.4 in the Thermodynamics part of these notes) – although in the real universe the arrow of time is such that the entropy in fact increases. Recall also the following passage from *Through the Looking-glass and What Alice Found There.*

Alice was just beginning to say, "There's a mistake somewhere -- -" when the Queen began screaming, so loud that she had to leave the sentence unfinished. "Oh, oh, oh!" shouted the Queen, shaking her hand about as if she wanted to shake it off. "My finger's bleeding! Oh, oh, oh, oh!"

Her screams were so exactly like the whistle of a steam-engine, that Alice had to hold both her hands over her ears.

"What *is* the matter?" she said, as soon as there was a chance of making herself heard. "Have you pricked your finger?"

"I haven't pricked it *yet*" the Queen said, "but I soon shall -- -oh, oh, oh!"

"When do you expect to do it?" Alice asked, feeling very much inclined to laugh.

"When I fasten my shawl again," the poor Queen groaned out: "the brooch will come undone directly. Oh, oh!" As she said the words the brooch flew open, and the Queen clutched wildly at it, and tried to clasp it again.

"Take care!" cried Alice. "You're holding it all crooked!" And she caught at the brooch; but it was too late: the pin had slipped, and the Queen had pricked her finger.

"That accounts for the bleeding, you see," she said to Alice with a smile. "Now you understand the way things happen here."

"But why don't you scream now?" Alice asked, holding her hands ready to put over her ears again. "Why, I've done all the screaming already," said the Queen. "What would be the good of having it all over again?"

Addendum. Coincidentally, just two days after having completed this chapter, I received the 2005 February issue of *Astronomy & Geophysics*, which included a fascinating article on the Arrow of Time. You might want to look it up. The reference is Davis, P., *Astronomy & Geophysics* (Royal Astronomical Society) **46**, 26 (2005).

-CHAPTER 16
CGS ELECTRICITY AND MAGNETISM

16.1 *Introduction*

We are accustomed to using MKS (metre-kilogram-second) units. A second, at one time defined as a fraction 1/86400 of a day, is now defined as 9 192 631 770 times the period of a hyperfine line emitted in the spectrum of the ^{133}Cs (caesium) atom. A metre was at one time defined as one ten-millionth of the length of a quadrant of Earth's surface measured from pole to equator. Later it was defined as the distance between two scratches on a platinum-iridium bar held on Paris. Still later, it was defined in terms of the wavelength of one or other of several spectral lines that have been used in the past for this purpose. At present, the metre is defined as the distance travelled by light *in vacuo* in a time of 1/(299 792 458) second. A kilogram is equal to the mass of a platinum-iridium cylinder held in Paris. The day may come when we are able to define a kilogram as the mass of so many electrons, but that day is not yet.

For electricity and magnetism, we extended the MKS system by adding an additional unit, the ampère, whose definition was given in Chapter 6, Section 6.2, to form the MKSA system. This in turn is a subset of SI (le Système International des Unités), which also includes the kelvin, the candela and the mole.

An older system of units, still used by some authors, was the CGS (centimetre-gram-second) system. In this system, a *dyne* is the *force* that will impart an acceleration of 1 cm s^{-2} to a mass of 1 gram. An *erg* is the *work* done when a force of one dyne moves its point of application through 1 cm in the line of action of the force. It will not take the reader a moment to see that a newton is equal to 10^5 dynes, and a joule is 10^7 ergs. As far as mechanical units are concerned, neither one system has any particular advantage over the other.

When it comes to electricity and magnetism, however, the situation is entirely different, and there is a huge difference between MKS and CGS. Part of the difficulty stems from the circumstance that electrostatics, magnetism and current electricity originally grew up as quite separate disciplines, each with its own system of units, and the connections between them were not appreciated or even discovered. It is not always realized that there are several version of CGS units used in electricity and magnetism, including hybrid systems, and countless conversion factors between one version and another. There are CGS *electrostatic* units (esu), to be used in electrostatics; CGS *electromagnetic* units (emu), to be used for describing magnetic quantities; and gaussian *mixed* units. In the gaussian mixed system, in equations that include both electrostatic quantities and magnetic quantities, the former were supposed to be expressed in esu and the latter in emu, and a conversion factor, given the symbol c, would appear in various parts of an equation to take account of the fact that some quantities were expressed in one system of units and others were expressed in another system. There was also the *practical* system of units, used in current electricity. In this, the ampère would be defined either in terms of the rate of electrolytic deposition of silver from a silver nitrate solution, or as exactly 10 CGS emu of current. The *ohm* would be defined in terms of the resistance of a column of mercury of defined dimensions, or again as exactly 10^9 emu of resistance. And a *volt* was 10^8 emu of potential difference. It will be seen already that, for every electrical quantity, several conversion factors between the different systems had to be known. Indeed, the MKSA system was devised specifically to avoid this proliferation of conversion factors.

Generally, the units in these CGS system have no particular names; one just talks about so many esu of charge, or so many emu of current. Some authors, however, give the names statcoulomb, statamp, statvolt, statohm ,etc., for the CGS esu of charge, current, potential difference and resistance, and abcoulomb, abamp, abvolt, abohm for the corresponding emu.

The difficulties by no means end there. For example, Coulomb's law is generally written as

$$F = \frac{Q_1 Q_2}{kr^2}. \qquad 16.1.1$$

It will immediately be evident from this that the permittivity defined by this equation differs by a factor of 4π from the permittivity that we are accustomed to. In the familiar equation generally used in conjunction with SI units, namely

$$F = \frac{Q_1 Q_2}{4\pi\epsilon r^2}, \qquad 16.1.2$$

the permittivity ε so defined is called the *rationalized* permittivity. The permittivity k of equation16.1.1 is the *unrationalized* permittivity. The two are related by $k = 4\pi\epsilon$. A difficulty with the unrationalized form is that a factor 4π appears in formulas describing uniform fields, and is absent from formulas describing situations with spherical symmetry.

Yet a further difficulty is that the magnitude of the CGS esu of charge is defined in such a way that the unrationalized free space permittivity has the numerical value 1 – and consequently it is normally left out of any equations in which it should appear. Thus equations as written often do not balance dimensionally, and one is deprived of dimensional analysis as a tool. Permittivity is regarded as a *dimensionless number*, and Coulomb's law for two charges *in vacuo* is written as

$$F = \frac{Q_1 Q_2}{r^2}. \qquad 16.1.3$$

The view is taken that electrical quantities can be expressed dimensionally in terms of mass, length and time only, and, from equation 16.1.3, it is asserted that the dimensions of electrical charge are

$$[Q] = \mathrm{M}^{1/2}\mathrm{L}^{3/2}\mathrm{T}^{-1}. \qquad 16.1.4$$

Because permittivity is regarded as a dimensionless quantity, the vectors **E** and **D** are regarded as dimensionally similar, and *in vacuo* they are *identical*. That is, *in vacuo*, there is no distinction between them.

When we come to CGS *electromagnetic units* all these difficulties reappear, except that, in the emu system, the free space *permeability* is regarded as a dimensionless number equal to 1, **B** and **H** are dimensionally similar, and *in vacuo* there is no distinction between them. The dimensions of electric charge in the CGS emu system are

$$[Q] = \mathrm{M}^{1/2}\mathrm{L}^{1/2}. \qquad 16.1.5$$

Thus the dimensions of charge are different in esu and in emu.

Two more highlights. The unit of capacitance in the CGS system is the centimeter, but in the CGS emu system, the centimetre is the unit of inductance.

Few users of CGS esu and emu fully understand the complexity of the system. Those who do so have long abandoned it for SI. CGS units are probably largely maintained by those who work with CGS units in a relatively narrow field and who therefore do not often have occasion to convert from one unit to another in this immensely complicated and physically unrealistic system.

Please don't blame *me* for this – I'm just the messenger!

In Sections 16.2, 16.3 and 16.4 I shall describe some of the features of the esu, emu and mixed systems. I shall not be giving a full and detailed exposition of CGS electricity, but I am just mentioning some of the highlights and difficulties. You are not going to like these sections very much, and will probably not make much sense of them. I suggest just skip through them quickly the first time, just to get some idea of what it's all about. *The practical difficulty that you are likely to come across in real life* is that you will come across equations and units written in CGS language, and you will want to know how to translate them into the SI language with which you are familiar. I hope to address that in Section 16.5, and to give you some way of translating a CGS formula into an SI formula that you can use and get the right answer from.

16.2 *The CGS Electrostatic System*

Definition. One CGS esu of charge (also known as the *statcoulomb*) is that charge which, if placed 1 cm from a similar charge *in vacuo*, will repel it with a force of 1 dyne.

The following exercises will be instructive.

Calculate, from the SI electricity that you already know, the force between two coulombs placed 1 cm from each other. From this, calculate how many CGS esu of charge there are in a coulomb. (I make it 1 coulomb = 2.998×10^9 esu. It will not escape your notice that this number is ten times the speed of light expressed in m s^{-1}.) Calculate the magnitude of the electronic charge in CGS esu. (I make it 4.8×10^{-10} esu. If, for example, you see that the potential energy of two electrons at a distance r apart is e^2/r, this is the number you must substitute for e. If you then express r in cm, the energy will be in ergs.)

Coulomb's law *in vacuo* is

$$F = \frac{Q_1 Q_2}{r^2}, \qquad 16.2.1$$

This differs from our accustomed formula (equation 16.1.2) in two ways. In the first place, we are using an *unrationalized* definition of permittivity, so that the familiar 4π is absent. Secondly, we are choosing units such that $4\pi\varepsilon_0$ has the numerical value 1, and so we are omitting it from the equation.

While some readers (and myself!) will object, and say that equation 16.2.1 does not balance dimensionally, and is valid only if the quantities are expressed in particular units, others will happily say that equation 16.2.1 shows that the dimensions of Q are $\mathrm{M}^{1/2}\mathrm{L}^{3/1}\mathrm{T}^{-1}$, and will not mind these extraordinary dimensions.

Electric field **E** is defined in the usual way, i.e. $\mathbf{F} = Q\mathbf{E}$, so that if the force on 1 esu of charge is 1 dyne, the field strength is 1 esu of electric field. (The esu of electric field has no name other than esu.) This is fine, but have you any idea how this is related to the SI unit of **E**, volt per metre? It requires a *great* deal of mental gymnastics to find out, so I'll just give the answer here, namely

$$1 \text{ CGS esu of } \mathbf{E} = 10^{-6}c \ \ \mathrm{V\ m^{-1}},$$

where

$$c = 2.997\,924\,58 \times 10^{10}.$$

Now the vector **D** is defined by $\mathbf{D} = k\mathbf{E}$, and since the permittivity is held to be a dimensionless number, **D** and **E** are held to be dimensionally similar ($\mathrm{M}^{1/2}\mathrm{L}^{-1/2}\mathrm{T}^{-1}$ in fact). Further, since the free space permittivity is 1, *in vacuo* there is no distinction between **D** and **E**, and either can substitute for the other. So the conversion between CGS esu and SI for **D** is the same as for **E**? No! In SI, we recognize **D** and **E** as being physically different quantities, and **D** is expressed in coulombs per square metre. *Awful* mental gymnastics are needed to find the conversion, but I'll give the answer here:

$$1 \text{ CGS esu of } \mathbf{D} = \frac{10^5}{4\pi c} \ \ \mathrm{C\ m^{-2}}.$$

Once again, please don't blame me – I'm just the messenger! And be warned – it is going to get worse – much worse.

Potential Difference

If the work required to move a charge of 1 esu from one point to another is 1 erg, the potential difference between the points is 1 esu of potential difference, or 1 statvolt.

It is often said that an esu of potential difference is 300 volts, but this is just an approximation. The exact conversion is

$$1 \text{ statvolt} = 10^{-8} c \text{ V}.$$

Capacitance

If the potential difference across the plate of a capacitor is one statvolt when the capacitor holds a charge of one statcoulomb, the capacitance of the capacitor is one centimetre. (No – that's not a misprint.)

$$1 \text{ cm} = 10^{9} c^{-2} \text{ F}.$$

Here is a sample of some formulas for use with CGS esu.

Potential at a distance r from a point charge Q *in vacuo* = Q/r.

Field at a distance r *in vacuo* from an infinite line charge of λ esu/cm = $2\lambda/r$.

Field *in vacuo* above an infinite charged plate bearing a surface charge density of σ esu/cm^2 = $2\pi\sigma$.

An electric dipole moment **p** is, as in SI, the maximum torque experienced by the dipole in unit electric field. A *debye* is 10^{-18} esu of dipole moment. The field at a distance r *in vacuo* along the axis of a dipole is $2p/r$.

Gauss's theorem: The total normal outward flux through a closed surface is 4π time the enclosed charge.

Capacitance of a plane parallel capacitor = $\dfrac{kA}{4\pi d}$.

Capacitance of an isolated sphere of radius a *in vacuo* = a. *Example*: What is the capacitance of a sphere of radius 1 cm? Answer: 1 cm. Easy, eh?

Energy per unit volume or an electric field = $E^2/(8\pi)$.

One more example before leaving esu. You will recall that, if a polarizable material is placed in an electrostatic field, the field **D** in the material is greater than $\varepsilon_0\mathbf{E}$ by the *polarization* **P** of the material. That is, $\mathbf{D} = \varepsilon\mathbf{E} + \mathbf{P}$. The equivalent formula for use with CGS esu is

$$\mathbf{D} = \mathbf{E} + 4\pi\mathbf{P}.$$

And since $\mathbf{P} = \chi_e\mathbf{E}$ and $\mathbf{D} = k\mathbf{E}$, it follows that

$$k = 1 + 4\pi\chi_e.$$

At this stage you may want a conversion factor between esu and SI for all quantities. I'll supply one a little later, but I want to describe emu first, and then we can construct a table given conversions between all three systems.

16.3 *The CGS Electromagnetic System*

If you have been dismayed by the problems of CGS esu, you don't yet know what is in store for you with CGS emu. Wait for it:

Definition. One CGS emu of magnetic pole strength is that pole which, if placed 1 cm from a similar pole *in vacuo*, will repel it with a force of 1 dyne.

The system is based on the proposition that there exists a "pole" at each end of a magnet, and that point poles repel each other according to an inverse square law. Magnetic field strength **H** is defined as the force experienced by a unit pole situated in the field. Thus, if a pole of strength m emu is situated in a field of strength **H**, it will experience a force $\mathbf{F} = m\mathbf{H}$.

Definition. If a pole of strength 1 emu experiences a force of 1 dyne when it is situated in a magnetic field, the strength of the magnetic field is 1 *oersted* (Oe). It will probably be impossible for the reader at this stage to try to work out the conversion factor between Oe and A m^{-1}, but, for the record

$$1 \text{ Oe} = \frac{250}{\pi} \text{ A m}^{-1}.$$

Now hold on tight, for the definition of the unit of *electric current*.

Definition: One emu of current (1 *abamp*) is that steady current, which, flowing in the arc of a circle of length 1 cm and of radius 1 cm (i.e. subtending 1 radian at the centre of the circle) gives rise to a magnetic field of 1 oersted at the centre of the circle.

This will involve quite an effort of the imagination. First you have to imagine a current flowing in an arc of a circle. Then you have to imagine measuring the field at the centre of the circle by measuring the force on a unit magnetic pole that you place there.

It follows that, if a current I abamp flows in a circle of radius a cm, the field at the centre is of the circle is

$$H = \frac{2\pi I}{a} \text{ Oe}\cdot$$

The conversion between emu of current (abamp) and ampères is

$$1 \text{ emu} = 10 \text{ A}.$$

The Biot-Savart law becomes

$$dH = \frac{I\,ds\,\sin\theta}{r^2}.$$

The field at a distance r *in vacuo* from a long straight current I is

$$H = \frac{2I}{r}.$$

Ampère's law says that the line integral of **H** around a closed plane curve is 4π times the enclosed current. The field inside a long solenoid of n turns per centimetre is

$$H = 4\pi nI.$$

So far, no mention of **B**, but it is now time to introduce it. Let us imagine that we have a long solenoid of n turns per cm, carrying a current of I emu, so that the field inside it is $4\pi nI$ Oe. Suppose that the cross-sectional area of the solenoid is A. Let us wrap a single loop of wire tightly around the outside of the solenoid, and then change the current at a rate $\dot{I}$ so that the field changes at a rate $\dot{H} = 4\pi n\dot{I}$. An EMF will be set up in the outside (secondary) coil of magnitude $A\dot{H}$. If we now insert an iron core inside the solenoid and repeat the experiment, we find that the induced EMF is much larger. It is larger by a (supposed dimensionless) factor called the *permeability* of the iron. Although this factor is called the permeability and the symbols used is often μ, I am going to use the symbol κ for it. The induced EMF is now A times $\kappa\dot{H}$. We denote the product of μ and H with the symbol B, so that $B = \kappa H$. The magnitude of B inside the solenoid is

$$B = 4\pi\kappa nI.$$

It will be evident from the familiar SI version $B = \mu nI$ that the CGS emu definition of the permeability differs from the SI definition by a factor 4π. The CGS emu definition is called an *unrationalized* definition; the SI definition is *rationalized.* The relation between them is $\mu = 4\pi\kappa$.

In CGS emu, the permeability of free space has the value 1. Indeed the supposedly dimensionless unrationalized permeability is what, in SI parlance, would be the *relative permeability.*

The CGS unit of B is the *gauss* (G), and 1 G $= 10^{-4}$ T.

It is usually held that κ is a dimensionless number, so that B and H have the same dimensions, and, in free space, B and H are *identical.* They are identical not only numerically, but there is physically no distinction between them. Because of this, the unit *oersted* is rarely heard, and it is common to hear the unit *gauss* used haphazardly to describe either B or H.

The scalar product of **B** and area is the magnetic flux, and its CGS unit, G cm^2, bears the name the *maxwell.* The rate of change of flux in maxwells per second will give you the induced EMF in emus (abvolts). An abvolt is 10^{-8} V.

The subject of *magnetic moment* has caused so much confusion in the literature that I shall devote an entire future chapter to it rather than try to do it here.

I end this section by giving the CGS emu version of *magnetization.* The familiar $\mathbf{B} = \mu_0(\mathbf{H} + \mathbf{M})$ becomes, in its CGS emu guise, $\mathbf{B} = \mathbf{H} + 4\pi\mathbf{M}$. The magnetic susceptibility χ_{m} is defined by $\mathbf{M} = \chi_{\mathrm{m}}\mathbf{H}$. Together with $\mathbf{B} = \kappa\mathbf{H}$, this results in $\kappa = 1 + 4\pi\chi_{\mathrm{m}}$.

16.4 *The Gaussian Mixed System*

A problem arises if we are dealing with a situation in which there are both "electrostatic" and "electromagnetic" quantities. The "mixed system", which is used *very frequently*, in CGS literature, uses esu for quantities that are held to be "electrostatic" and emu for quantities that are held to be "electromagnetic", and it seems to be up to each author to decide which quantities are to be regarded as "electrostatic" and which are "electromagnetic. Because different quantities are to be expressed in different sets of units within a single equation, the equation must include the conversion factor $c = 2.997\,924\,58 \times 10^{10}$ in strategic positions within the equation.

The most familiar example of this is the equation for the force **F** experienced by a charge Q when it is moving with velocity **v** in an electric field **E** and a magnetic field **B**. This equation is liable to appear either as

$$F = Q\left(\mathbf{E} + \frac{\mathbf{v} \times \mathbf{H}}{c}\right) \qquad 16.4.1$$

or as

$$F = Q\left(\mathbf{E} + \frac{\mathbf{v} \times \mathbf{B}}{c}\right). \qquad 16.4.2$$

It can appear in either of these forms because, if CGS emu are used, B and H are numerically equal *in vacuo*. The conversion factor c appears in these equations, because it is understood (by those who understand CGS units) that Q and E are to be expressed in esu, while B or H is to be expressed in emu, and the conversion factor c is necessary to convert it to esu.

It should be noted that in all previous chapters in these notes, equations balance dimensionally, and the equations are valid *in any coherent system of units, not merely SI*. Difficulties arise, of course, if you write an equation that is valid only so long as a particular set of units is used, and even more difficulties arise if some quantities are to be expressed in one system of units, and other quantities are to be expressed in another system of units.

An analogous situation is to be found in some of the older books on thermodynamics, where it is possible to find the following equation:

$$C_P - C_V = R/J. \qquad 16.4.3$$

This equation expresses the difference in the specific heat capacities of an ideal gas, measured at constant pressure and at constant volume. In equation 16.4.3, it is understood that C_P and C_V are to be expressed in calories per gram per degree, while the universal gas constant is to be expressed in ergs per gram per degree. The factor J is a conversion factor between erg and calories. Of course the sensible way to write the equation is merely

$$C_P - C_V = R. \qquad 16.4.4$$

This is valid *whatever* units are used, be they calories, ergs, joules, British Thermal Units or kWh, as long as all quantities are expressed in the same units. Yet it is truly extraordinary how many electrical equations are to be found in the literature, in which different units are to be used for dimensionally similar quantities.

Maxwell's equations may appear in several forms. I take one at random from a text written in CGS:

$$\mathrm{div}\mathbf{B} = 0, \qquad 16.4.5$$

$$\mathrm{div}\mathbf{D} = 4\pi\rho, \qquad 16.4.6$$

$$c\,\mathbf{curlH} = \dot{\mathbf{D}} + 4\pi\mathbf{J}, \qquad 16.4.7$$

$$c\,\mathbf{curlE} = -\dot{\mathbf{B}}. \qquad 16.4.8$$

The factor c occurs as a conversion factor, since some quantities are to be expressed in esu and some in emu. The 4π arises because of a different definition (unrationalized) of permeability. In some versions there may be no distinction between **B** and **H**, or between **E** and **D**, and the 4π and the c may appear in various places in the equations.

(It may also be remarked that, in the earlier papers, and in Maxwell's original writings, vector notation is not used, and the equations appear extremely cumbersome and all but incomprehensible to modern eyes.)

16.5 *Conversion Factors*

By this time, you are completely bewildered, and you want nothing to do with such a system. Indeed you may even be wondering if I made it all up, so irrational does it appear to be. You would like to ignore it all completely. But you cannot ignore it, because, in your reading, you keep coming across formulas that you need, but you don't know what units to use, or whether there should be a 4π in the formula, or whether there is a permittivity or permeability missing from the equation because the author happens to be using some set of units in which one or the other of these quantities has the numerical value 1, or whether the H in the equation should really be a B, or the E a D.

Is there anything I can do to help?

What I am going to do in this section is to list a number of conversion factors between the different systems of units. This may help a little, but it won't by any means completely solve the problem. Really to try and sort out what a CGS equation means requires some dimensional analysis, and I shall address that in section 16.6

In the conversion factors that I list in this section, the symbol c stands for the number 2.997 924 58 $\times\ 10^{10}$, which is numerically equal to the speed of light expressed in cm s^{-1}. The abbreviation "esu" will mean CGS electrostatic unit, and "emu" will mean CGS electromagnetic unit. A prefix "stat" to a unit implies that it is an esu; a prefix "ab" implies that it is an emu. I list the conversion factors for each quantity in the form "1 SI unit = so many esu = so many emu".

I might mention that people will say that "SI is full of conversion factors". The fact is that SI is a unified coherent set of units, and it has *no* conversion factors. Conversion factors are characteristic of CGS electricity and magnetism.

Quantity of Electricity (Electric Charge)

1 coulomb = $10^{-1}c$ statcoulomb = 10^{-1} emu

Electric Current

1 amp = $10^{-1}c$ esu = 10^{-1} abamp

Potential Difference

1 volt = $10^{8}/c$ statvolt = 10^{8} emu

Resistance

1 ohm = $10^{9}/c^{2}$ esu = 10^{9} abohm

Capacitance

1 farad = $10^{-9}c^{2}$ esu = 10^{-9} emu

Inductance

1 henry = $10^{9}/c^{2}$ esu = 10^{9} emu

Electric Field E

$$1 \text{ V m}^{-1} \;=\; 10^6/c \text{ esu} \;=\; 10^6 \text{ esu}$$

Electric Field D

$$1 \text{ C m}^{-2} \;=\; 4\pi \times 10^{-5}c \text{ esu} \;=\; 4\pi \times 10^{-5} \text{ emu}$$

Magnetic Field B

$$1 \text{ tesla} \;=\; 10^4/c \text{ esu} \;=\; 10^4 \text{ gauss}$$

Magnetic Field H

$$1 \text{ A m}^{-1} \;=\; 4\pi \times 10^{-3}c \text{ esu} \;=\; 4\pi \times 10^{-3} \text{ oersted}$$

Magnetic B-flux Φ_B

$$1 \text{ weber} \;=\; 10^8/c \text{ esu} \;=\; 10^8 \text{ maxwell}$$

16.6 *Dimensions*

A book says that the equivalent width W, in wavelength units, of a spectrum line, is related to the number of atoms per unit area in the line of sight, N, by

$$W \;=\; \frac{\pi e^2 N \lambda^2}{mc^2}. \qquad 16.6.1$$

Is this formula all right in *any* system of units? Can I use SI units on the right hand side, and get the answer in metres? Or must I use a particular set of units in order to get the right answer? And if so, which units?

A book says that the rate at which energy is radiated, P, from an accelerating charge is

$$P \;=\; \frac{2e^2 \ddot{x}^2}{c^3}. \qquad 16.6.2$$

Is this correct? Is c the speed of light, or is it merely a conversion factor between different units? Or is one of the cs a conversion factor, and the other two are the speed of light?

It *is* possible to find the answer to such bewildering questions, if we do a bit of dimensional analysis. So, before trying to answer these specific questions (which I shall do later as examples) I am going to present a table of dimensions. I already gave a table of dimensions of electrical quantities in Chapter 11, in terms of M, L, T and Q, but that table won't be particularly helpful in the present context.

I pointed out in Section 16.1 of the present chapter that Coulomb's law is often written in the form

$$F = \frac{Q_1 Q_2}{r^2}. \qquad 16.6.3$$

Consequently the dimensions of Q are held to be $[Q] = \mathrm{M}^{1/2}\mathrm{L}^{3/2}\mathrm{T}^{-1}$. But we know that a permittivity is missing from the denominator of equation 16.6.3, because the writer intends his formula to be restricted to a particular set of units such that k or $4\pi\epsilon_0 = 1$. In order to detect whether a permittivity has been omitted from an equation, we need a table in which the dimensions of electrical quantities are given not in terms of M, L, T and Q as in Chapter 11, but in terms of M, L, T and ε, and this is what I am just about to do. However, often it is the *permeability* that has been omitted from an equation, and, in order to detect whether this is so, I am also supplying a table in which the dimensions of electrical quantities are given in terms of M, L, T and μ.

If, from dimensional analysis, you find that an expression is dimensionally wrong by a power of the permittivity, insert $4\pi\epsilon_0$ in the appropriate part of the equation. If you find that an expression is dimensionally wrong by a power of the permeability, insert $\mu_0/(4\pi)$ in the appropriate part of the equation. If you find that the equation is wrong by LT^{-1}, insert or delete c as appropriate. Your equation will then balance dimensionally and will be ready for use in *any* coherent system of units, including SI. This procedure will probably work in most cases, but I cannot guarantee that it will work in all cases, because it cannot deal with those (frequent!) cases in which the formula given is plain wrong, whatever units are used!

	M	L	T	ε		M	L	T	μ
Electric charge	$\frac{1}{2}$	$\frac{3}{2}$	−1	$\frac{1}{2}$		$\frac{1}{2}$	$\frac{1}{2}$	0	$-\frac{1}{2}$
Electric dipole moment	$\frac{1}{2}$	$\frac{5}{2}$	−1	$\frac{1}{2}$		$\frac{1}{2}$	$\frac{3}{2}$	0	$-\frac{1}{2}$
Current	$\frac{1}{2}$	$\frac{3}{2}$	−2	$\frac{1}{2}$		$\frac{1}{2}$	$\frac{1}{2}$	−1	$-\frac{1}{2}$
Potential difference	$\frac{1}{2}$	$\frac{1}{2}$	−1	$-\frac{1}{2}$		$\frac{1}{2}$	$\frac{3}{2}$	−2	$\frac{1}{2}$
Resistance	0	−1	1	−1		0	1	−1	1
Resistivity	0	0	1	−1		0	2	−1	1
Conductance	0	1	−1	1		0	−1	1	−1
Conductivity	0	0	−1	1		0	−2	1	−1
Capacitance	0	1	0	1		0	−1	2	−1
Electric field E	$\frac{1}{2}$	$-\frac{1}{2}$	−1	$-\frac{1}{2}$		$\frac{1}{2}$	$\frac{1}{2}$	−2	$\frac{1}{2}$
Electric field D	$\frac{1}{2}$	$-\frac{1}{2}$	−1	$\frac{1}{2}$		$\frac{1}{2}$	$-\frac{3}{2}$	0	$-\frac{1}{2}$
Electric flux Φ_E	$\frac{1}{2}$	$\frac{3}{2}$	−1	$-\frac{1}{2}$		$\frac{1}{2}$	$-\frac{5}{2}$	−2	$\frac{1}{2}$
Electric flux Φ_D	$\frac{1}{2}$	$\frac{3}{2}$	−1	$\frac{1}{2}$		$\frac{1}{2}$	$\frac{1}{2}$	0	$-\frac{1}{2}$
Permittivity	0	0	0	1		0	−2	2	−1
Magnetic field B	$\frac{1}{2}$	$-\frac{3}{2}$	0	$-\frac{1}{2}$		$\frac{1}{2}$	$-\frac{1}{2}$	−1	$\frac{1}{2}$
Magnetic field H	$\frac{1}{2}$	$\frac{1}{2}$	−2	$\frac{1}{2}$		$\frac{1}{2}$	$-\frac{1}{2}$	−1	$-\frac{1}{2}$
Magnetic flux Φ_B	$\frac{1}{2}$	$\frac{1}{2}$	0	$-\frac{1}{2}$		$\frac{1}{2}$	$\frac{1}{2}$	−1	$\frac{1}{2}$
Magnetic flux Φ_H	$\frac{1}{2}$	$\frac{5}{2}$	−2	$\frac{1}{2}$		$\frac{1}{2}$	$\frac{1}{2}$	−1	$-\frac{1}{2}$
Permeability	0	−2	2	−1		0	0	0	1
Magnetic vector potential	$\frac{1}{2}$	$-\frac{1}{2}$	0	$-\frac{1}{2}$		$\frac{1}{2}$	$\frac{1}{2}$	−1	$\frac{1}{2}$
Inductance	0	−1	2	−1		0	1	0	1

Now let's look at the equation for equivalent width of a spectrum line:

$$W = \frac{\pi e^2 N \lambda^2}{mc^2}. \qquad 16.6.1$$

Here $[W]$ = L and $[N] = \mathrm{L}^{-2}$. By making use of the table we find that the dimensions of the right hand side are Lε. There is therefore a $4\pi\varepsilon_0$ missing from the denominator, and the equation should be

$$W = \frac{\pi e^2 N \lambda^2}{4\pi\varepsilon_0 mc^2}. \qquad 16.6.4$$

How about the rate at which energy is radiated from an accelerating charge?

$$P = \frac{2e^2\ddot{x}^2}{c^3}. \qquad 16.6.2$$

Power has dimensions ML^2T^{-3}, whereas the dimensions of the right hand side are $ML^2T^{-3}\varepsilon$, so again there is a $4\pi\varepsilon_0$ missing from the denominator and the formula should be

$$P = \frac{2e^2\ddot{x}^2}{4\pi\varepsilon_0 c^3}. \qquad 16.6.5$$

It is often the case that there is a $4\pi\varepsilon_0$ missing from the denominator is formulas that have an e^2 upstairs.

"Electromagnetic" formulas often give more difficulty. For example, one book says that the energy per unit volume in a magnetic field *in vacuo* is $\frac{B^2}{8\pi}$, while another says that is it is $\frac{H^2}{8\pi}\cdot$ Which is it (if indeed it is either)? Energy per unit volume has dimensions $ML^{-1}T^{-2}$. The dimensions of B^2 are $ML^{-1}T^{-2}\mu$. The equation given is therefore wrong dimensionally by permeability, and the equation should be divided by $\mu_0/(4\pi)$ to give $B^2/(2\mu_0)$, which is correct. On the other hand, the dimensions of H^2 are $ML^{-1}T^{-2}\mu^{-1}$, so perhaps we should *multiply* by $\mu_0/(4\pi)$? But this does not give a correct answer, and it exemplifies some of the many difficulties that are caused by writing formulas that do not balance dimensionally and are intended to be used only with a particular set of units. The situation is particularly difficult with respect to *magnetic moment*, a subject to which I shall devote the next chapter.

CHAPTER 17
MAGNETIC DIPOLE MOMENT

17.1 *Introduction*

A number of different units for expressing magnetic dipole moment (hereafter simply "magnetic moment") are commonly seen in the literature, including, for example, erg G^{-1}, G cm^3, Oe cm^3, T m^3, A m^2, J T^{-1}. It is not always obvious how to convert from one to another, nor is it obvious whether all quantities described as "magnetic moment" refer to the same physical concept or are dimensionally or numerically similar. It can be almost an impossibility, for example, to write down a list of the magnetic moments of the planets in order of increasing magnetic moment if one refers to the diverse literature in which the moments of each of the nine planets are quoted in different units. This chapter explores some of these aspects of magnetic moment.

In previous chapters, I have used the symbols p_e and p_m for electric and magnetic dipole moment. In this chapter I shall be concerned exclusively with magnetic moment, and so I shall dispense with the subscript m.

17.2 *The SI Definition of Magnetic Moment*

If a magnet is placed in an external magnetic field **B**, it will experience a torque. The magnitude of the torque depends on the orientation of the magnet with respect to the magnetic field. There are two oppositely-directed orientations in which the magnet will experience the greatest torque, and the magnitude of the magnetic moment is **defined** as the *maximum torque experienced by the magnet when placed in unit external magnetic field.* The magnitude and direction of the torque is given by the equation

$$\boldsymbol{\tau} = \mathbf{p}\times\mathbf{B}. \qquad 17.2.1$$

The SI unit for magnetic moment is clearly N m T^{-1}.

If an electric current I flows in a plane coil of area **A** (recall that area is a vector quantity – hence the boldface), the torque it will experience in a magnetic field is given by

$$\boldsymbol{\tau} = I\mathbf{A}\times\mathbf{B}. \qquad 17.2.2$$

This means that the magnetic moment of the coil is given by

$$\mathbf{p} = I\mathbf{A}. \qquad 17.2.3$$

Thus the unit A m^2 is also a correct SI unit for magnetic moment, though, unless the concept of "current in a coil" needs to be emphasized in a particular context, it is perhaps better to stick to N m T^{-1}.

While "J T^{-1}" is also formally dimensionally correct, it is perhaps better to restrict the unit "joule" to work or energy, and to use N m for torque. Although these are dimensionally similar, they are conceptually rather different. For this reason, the occasional practice seen in atomic physics of expressing magnetic moments in MeV T^{-1} is not entirely appropriate, however convenient it may sometimes seem to be in a field in which masses and momenta are often conveniently expressed in MeV/c^2 and MeV/c.

It is clear that the unit "T m^3", often seen for "magnetic moment" is not dimensionally correct for magnetic moment as defined above, so that, whatever quantity is being expressed by the often-seen "T m^3", it is not the conventionally defined concept of magnetic moment.

The *magnetization* **M** of a material is defined by the equation

$$\mathbf{B} = \mu_0(\mathbf{H} + \mathbf{M}) \qquad 17.2.4$$

Equations 17.2.2 and 17.2.4 for the definitions of magnetic moment and magnetization are consistent with the alternative concept of magnetization as "magnetic moment per unit volume".

17.3 *The Magnetic Field on the Equator of a Magnet*

By the "equator" of a magnet I mean a plane normal to its magnetic moment vector, passing through the mid-point of the magnet.

The magnetic field at a point at a distance r on the equator of a magnet may be expressed as a series of terms of successively higher powers of $1/r$ (the first term in the series being a term in r^{-3}), and the higher powers decrease rapidly with increasing distance. At large distances, the higher powers become negligible, so that, at a large distance from a small magnet, the magnitude of the magnetic field produced by the magnet is given approximately by

$$B = \frac{\mu_0}{4\pi}\frac{p}{r^3} . \qquad 17.3.1$$

For example, if the surface magnetic field on the equator of a planet has been measured, and the magnetic properties of the planet are being modelled in terms of a small magnet at the centre of the planet, the dipole moment can be calculated by multiplying the surface equatorial magnetic field by $\mu_0/(4\pi)$ times the cube of the radius of the planet. If B, μ_0 and r are expressed respectively in T, H m^{-1} and m, the magnetic moment will be in N m T^{-1}.

17.4 *CGS Magnetic Moment, and Lip Service to SI*

Equation 17.3.1 is the equation (written in the convention of quantity calculus, in which symbols stand for physical quantities rather than for their numerical values in some particular system of units) for the magnetic field at a large distance on the equator of a magnet. The equation is valid in <u>*any*</u> coherent system of units whatever, and its validity is not restricted to any particular system of units. Example of systems of units in which equation 17.3.1 are valid include SI, CGS EMU, and British Imperial Units.

If CGS EMU are used, the quantity $\mu_0/(4\pi)$ has the numerical value 1. Consequently, when working exclusively in CGS EMU, equation 17.3.1 is often written as

$$B = \frac{p}{r^3} . \qquad 17.4.1$$

This equation appears not to balance dimensionally. However, the equation is not written according to the conventions of quantity calculus, and the symbols do not stand for physical quantities. Rather, they stand for their numerical values in a particular system of units. Thus r is the distance in cm, B is the field in gauss, and p is the magnetic moment in dyne cm per gauss. However, because of the deceptive appearance of the equation, a common practice, for example, in calculating the magnetic moment of a planet is to measure its surface equatorial field, multiply it by the cube of the planet's radius, and then quote the magnetic moment in "G cm^3". While the numerical result is correct for the magnetic moment in CGS EMU, the units quoted are not.

While some may consider objections to incorrect units to be mere pedantry (and who would presumably therefore see nothing wrong with quoting a length in grams, as long as the actual number is correct), the situation becomes more difficult when a writer, wishing to pay lip service to SI, attempts to use equation 17.4.1 using SI units, by multiplying the surface equatorial field in T by the cube of the planet's radius, and then giving the magnetic moment in "T m^3", a clearly disastrous recipe!

Of course, some may use equation 17.4.1 as a *definition* of magnetic moment. If that is so, then the quantity so defined is clearly not the same quantity, physically, conceptually, dimensionally or numerically, as the quantity defined as magnetic moment in Section 17.2.

17.5 *Possible Alternative Definitions of Magnetic Moment*

Although the standard SI definition of magnetic moment is described in Section 17.2, and there is little reason for anyone who wishes to be understood by others to use any other, the previous paragraph suggested that there might be more than one choice as to how one wishes to define magnetic moment. Do we use equation 17.2.1 or equation 17.4.1 as the definition? (They are clearly different concepts.) Additional degrees of freedom as to how one might choose to define magnetic moment depend on whether we elect to use magnetic field H or magnetic field B in the definition, or whether the permeability is or is not to include the factor 4π in its definition – that is, whether we elect to use a “rationalized” or “unrationalized” definition of permeability.

If one chooses to define the magnetic moment as the maximum torque experienced in unit external magnetic field, there is still a choice as to whether by magnetic field we choose H or B. Thus we could define magnetic moment by either of the following two equations:

$$\tau = p_1 H \tag{17.5.1}$$

or

$$\tau = p_2 B. \tag{17.5.2}$$

Alternatively, we could choose to define the magnetic moment is terms of the field on the equator. In that case we have a choice of four. We can choose to use B or H for the magnetic field, and we can choose to exclude or include 4π in the denominator:

$$B = \frac{p_3}{r^3}, \tag{17.5.3}$$

$$H = \frac{p_4}{r^3}, \tag{17.5.4}$$

$$B = \frac{p_5}{4\pi r^3}, \tag{17.5.5}$$

$$H = \frac{p_6}{4\pi r^3}. \tag{17.5.6}$$

These six possible definitions of magnetic moment are clearly different quantities, and one may well wonder why to list them all. The reason is that *all* of them are to be found in current scientific literature. To give some hint as to the unnecessary complications introduced when authors depart from the simple SI definition, I list in Table XVII.1 the *dimensions* of each version of magnetic moment, the CGS EM unit, the SI unit, and the conversion factor between CGS and SI. The conversion factors cannot be obtained simply by referring to the *dimensions*, because this does not take into account the inclusion or exclusion of 4π in the permeability. The correct factors can be obtained from the *units*, for example by noting that 1 Oe = $10^{-3}/(4\pi)$ A m^{-1} and 1 G = 10^{-4} T.

TABLE XVII.1
DIMENSIONS, CGS AND SI UNITS, AND CONVERSION FACTORS
FOR MAGNETIC MOMENTS

	Dimensions	1 CGS EMU	=	Conversion factor	SI unit
p_1	$\mathrm{ML^3T^{-1}Q^{-1}}$	1 dyn cm $\mathrm{Oe^{-1}}$	=	$4\pi \times 10^{-10}$	N m $\mathrm{(A/m)^{-1}}$
p_2	$\mathrm{L^2T^{-1}Q}$	1 dyn cm $\mathrm{G^{-1}}$	=	10^{-3}	N m $\mathrm{T^{-1}}$
p_3	$\mathrm{ML^3T^{-1}Q^{-1}}$	1 G $\mathrm{cm^3}$	=	10^{-10}	T $\mathrm{m^3}$
p_4	$\mathrm{L^2T^{-1}Q}$	1 Oe $\mathrm{cm^3}$	=	$10^{-3}/4\pi$	A $\mathrm{m^2}$
p_5	$\mathrm{ML^3T^{-1}Q^{-1}}$	1 G $\mathrm{cm^3}$	=	10^{-10}	T $\mathrm{m^3}$
p_6	$\mathrm{L^2T^{-1}Q}$	1 Oe $\mathrm{cm^3}$	=	$10^{-3}/4\pi$	A $\mathrm{m^2}$

17.6 *Thirteen Questions*

We have seen that the SI definition of magnetic moment is unequivocally defined as the maximum torque experienced in unit external field. Nevertheless some authors prefer to think of magnetic moment as the product of the equatorial field and the cube of the distance. Thus there are two conceptually different concepts of magnetic moment, and, when to these are added minor details as to whether the magnetic field is B or H, and whether or not the permeability should include the factor 4π, six possible definitions of magnetic moment, described in Section 17.6, all of which are to be found in current literature, arise.

Regardless, however, how one chooses to define magnetic moment, whether the SI definition or some other unconventional definition, it should be easily possible to answer both of the following questions:

A. Given the magnitude of the equatorial field on the equator of a magnet, what is the maximum torque that that magnet would experience if it were placed in an external field?

B. Given the maximum torque that a magnet experiences when placed in an external field, what is the magnitude of the equatorial field produced by the magnet?

It must surely be conceded that a failure to be able to answer such basic questions indicates a failure to understand what is meant by magnetic moment.

I therefore now ask a series of thirteen questions. The first six are questions of type A, in which I use the six possible definitions of magnetic moment. The next six are similar questions of type B. And the last is an absurdly simple question, which anyone who believes he understands the meaning of magnetic moment should easily be able to answer.

1. The magnitude of the field in the equatorial plane of a magnet at a distance of 1 cm is 1 Oe.

What is the maximum torque that this magnet will experience in an external magnetic field of 1 Oe, and what is its magnetic moment?

Note that, in this question and the following seven there *must* be a unique answer for the *torque.* The answer you give for the *magnetic moment*, however, will depend on how you choose to define magnetic moment, and on whether you choose to give the answer in SI units or CGS EMU.

2. The magnitude of the field in the equatorial plane of a magnet at a distance of 1 cm is 1 Oe.

What is the maximum torque that this magnet will experience in an external magnetic field of 1 G, and what is its magnetic moment?

3. The magnitude of the field in the equatorial plane of a magnet at a distance of 1 cm is 1 G.

What is the maximum torque that this magnet will experience in an external magnetic field of 1 Oe, and what is its magnetic moment?

4. The magnitude of the field in the equatorial plane of a magnet at a distance of 1 cm is 1 G.

What is the maximum torque that this magnet will experience in an external magnetic field of 1 G, and what is its magnetic moment?

5. The magnitude of the field in the equatorial plane of a magnet at a distance of 1 m is 1 A m^{-1}.

What is the maximum torque that this magnet will experience in an external magnetic field of 1 A m^{-1}, and what is its magnetic moment?

6. The magnitude of the field in the equatorial plane of a magnet at a distance of 1 m is 1 A m^{-1}.

What is the maximum torque that this magnet will experience in an external magnetic field of 1 T, and what is its magnetic moment?

7. The magnitude of the field in the equatorial plane of a magnet at a distance of 1 m is 1 T.

What is the maximum torque that this magnet will experience in an external magnetic field of 1 A m^{-1}, and what is its magnetic moment?

8. The magnitude of the field in the equatorial plane of a magnet at a distance of 1 m is 1 T.

What is the maximum torque that this magnet will experience in an external magnetic field of 1 T, and what is its magnetic moment?

9. A magnet experiences a maximum torque of 1 dyn cm if placed in a field of 1 Oe. What is the magnitude of the field in the equatorial plane at a distance of 1 cm, and what is the magnetic moment?

Note that, in this question and the following three there *must* be a unique answer for B and a unique answer for H, though each can be expressed in SI or in CGS EMU, while the answer for the magnetic moment depends on which definition you adopt.

10. A magnet experiences a maximum torque of 1 dyn cm if placed in a field of 1 G. What is the magnitude of the field in the equatorial plane at a distance of 1 cm, and what is the magnetic moment?

11. A magnet experiences a maximum torque of 1 N m if placed in a field of 1 A m^{-1}. What is the magnitude of the field in the equatorial plane at a distance of 1 m, and what is the magnetic moment?

12. A magnet experiences a maximum torque of 1 N m if placed in a field of 1 T. What is the magnitude of the field in the equatorial plane at a distance of 1 m, and what is the magnetic moment?

I'll pose Question Number 13 a little later. In the meantime the answers to the first four questions are given in Table XVII.2, and the answers to Questions 5 – 12 are given in

Tables XVII.3 and 4. The sheer complexity of these answers to absurdly simple questions is a consequence of different usages by various authors of the meaning of “magnetic moment” and of departure from standard SI usage.

TABLE XVII.2
ANSWERS TO QUESTIONS 1 – 4 IN CGS EMU AND SI UNITS
The answers to the first four questions are identical

τ	=	1 dyn cm	=	10^{-7} N m
p_1	=	1 dyn cm Oe^{-1}	=	$4\pi \times 10^{-7}$ N m $(\mathrm{A/m})^{-1}$
p_2	=	1 dyn cm G^{-1}	=	10^{-3} N m $(\mathrm{T})^{-1}$
p_3	=	1 G cm^3	=	10^{-10} T m^3
p_4	=	1 Oe cm^3	=	$10^{-3}/(4\pi)$ A m^2
p_5	=	4π G cm^3	=	$4\pi \times 10^{-10}$ T m^3
p_6	=	4π Oe cm^3	=	10^{-3} A m^2

TABLE XVII.3
ANSWERS TO QUESTIONS 5 – 8 IN CGS EMU AND SI UNITS

		5	6	7	8	
τ	=	$(4\pi)^2$	$4\pi \times 10^7$	$4\pi \times 10^7$	10^{14}	dyn cm
	=	$(4\pi)^2 \times 10^{-7}$	4π	4π	10^7	N m
p_1	=	$4\pi \times 10^3$	$4\pi \times 10^3$	10^{10}	10^{10}	dyn cm Oe^{-1}
	=	$(4\pi)^2 \times 10^{-7}$	$(4\pi)^2 \times 10^{-7}$	4π	4π	N m $(\mathrm{A/m})^{-1}$
p_2	=	$4\pi \times 10^3$	$4\pi \times 10^3$	10^{10}	10^{10}	dyn cm G^{-1}
	=	4π	4π	10^7	10^7	N m T^{-1}
p_3	=	$4\pi \times 10^3$	$4\pi \times 10^3$	10^{10}	10^{10}	G cm^3
	=	$4\pi \times 10^{-7}$	$4\pi \times 10^{-7}$	1	1	T m^3
p_4	=	$4\pi \times 10^3$	$4\pi \times 10^3$	10^{10}	10^{10}	Oe cm^3
	=	1	1	$10^7/(4\pi)$	$10^7/(4\pi)$	A m^2
p_5	=	$(4\pi)^2 \times 10^3$	$(4\pi)^2 \times 10^3$	$4\pi \times 10^{10}$	$4\pi \times 10^{10}$	G cm^3
	=	$(4\pi)^2 \times 10^{-7}$	$(4\pi)^2 \times 10^{-7}$	4π	4π	T m^3
p_6	=	$(4\pi)^2 \times 10^3$	$(4\pi)^2 \times 10^3$	$4\pi \times 10^{10}$	$4\pi \times 10^{10}$	Oe cm^3
	=	4π	4π	10^7	10^7	A m^2

TABLE XVII.4
ANSWERS TO QUESTIONS 9 – 12 IN CGS EMU AND SI UNITS

		9	10	11	12	
B	=	1	1	$10^4/(4\pi)$	10^{-3}	G
	=	10^{-4}	10^{-4}	$1/(4\pi)$	10^{-7}	T
H	=	1	1	$10^4/(4\pi)$	10^{-3}	Oe
	=	$10^3/(4\pi)$	$10^3/(4\pi)$	$10^7/(4\pi)^2$	$1/(4\pi)$	A m^{-1}
p_1	=	1	1	$10^{10}/(4\pi)$	10^3	dyn cm Oe^{-1}
	=	$4\pi \times 10^{-10}$	$4\pi \times 10^{-10}$	1	$4\pi\times10^{-7}$	N m $(A/m)^{-1}$
p_2	=	1	1	$10^{10}/(4\pi)$	10^3	dyn cm G^{-1}
	=	10^{-3}	10^{-3}	$10^7/(4\pi)$	1	N m T^{-1}
p_3	=	1	1	$10^4/(4\pi)$	10^{-3}	G cm^3
	=	10^{-10}	10^{-10}	$10^{-6}/(4\pi)$	10^{-13}	T m^3
p_4	=	1	1	$10^4/(4\pi)$	10^{-3}	Oe cm^3
	=	$10^{-3}/(4\pi)$	$10^{-3}/(4\pi)$	$10/(4\pi)^2$	$10^{-6}/(4\pi)$	A m^2
p_5	=	4π	4π	10^4	$4\pi \times 10^{-3}$	G cm^3
	=	$4\pi \times 10^{-10}$	$4\pi \times 10^{-10}$	10^{-6}	$4\pi \times 10^{-13}$	T m^3
p_6	=	4π	4π	10^4	$4\pi \times 10^{-3}$	Oe cm^3
	=	10^{-3}	10^{-3}	$10/(4\pi)$	10^{-6}	A m^2

The thirteenth and last of these questions is as follows: Assume that Earth is a sphere of radius 6.4×10^6 m = 6.4×10^8 cm, and that the surface field at the magnetic equator is $B = 3 \times 10^{-5}$ T = 0.3 G, or $H = 75/\pi$ A m^{-1} = 0.3 Oe, what is the magnetic moment of Earth? It is hard to imagine a more straightforward question, yet it would be hard to find two people who would give the same answer.

The SI answer (which, to me, is the only answer) is

$$B = \frac{\mu_0}{4\pi}\frac{p}{r^3}, \quad \therefore\ p = \frac{4\pi r^3 B}{\mu_0} = \frac{4\pi\times(6.4\times10^6)^3\times3\times10^{-5}}{4\pi\times10^{-7}} = 7.86\times10^{22}\ \text{N m T}^{-1}.$$

This result correctly predicts that, if Earth were placed in an external field of 1 T, it would experience a maximum torque of 7.86×10^{22} N m, and this is the normal meaning of what is meant by magnetic moment.

A calculation in GCS might proceed thus:

$$B = \frac{p}{r^3}, \quad \therefore\ p = r^3 B = (6.4\times10^8)^3 \times 0.3 = 7.86\times10^{25}\ \text{G cm}^3.$$

Is this the same result as was obtained from the SI calculation? We can use the conversions 1 G = 10^{-4} T and 1 cm^3 = 10^{-6} m^3, and we obtain

$$p = 7.86 \times 10^{15}\ \text{T m}^3.$$

We arrive at a number that not only differs from the SI calculation by 10^7, but is expressed in quite different, dimensionally dissimilar, units.

Perhaps the CGS calculation should be

$$H = \frac{p}{r^3}, \quad \therefore\ p = r^3 H = (6.4\times10^8)^3 \times 0.3 = 7.86\times10^{25}\ \text{Oe cm}^3.$$

Now 1 Oe = $1000/(4\pi)$ A m^{-1} and 1 cm^3 = 10^{-6} m^3, and we obtain

$$p = 6.26 \times 10^{21}\ \text{A m}^2$$

This time we arrive at SI units that are dimensionally similar to N m T^{-1}, and which are perfectly correct SI units, but the magnetic moment is smaller than correctly predicted by the SI calculation by a factor of 12.6.

Yet again, we might do what appears to be frequently done by planetary scientists, and we can multiply the surface field in T by the cube of the radius in m to obtain

$$p = 7.86 \times 10^{15} \quad \text{T m}^3.$$

This arrives at the same result as one of the CGS calculations, but, whatever it is, it is not the magnetic moment in the sense of the greatest torque in a unit field. The quantity so obtained appears to be nothing more that the product of the surface equatorial field and the cube of the radius, and as such would appear to be a purposeless and meaningless calculation.

It would be a good deal more meaningful merely to multiply the surface value of H by 3. This in fact would give (correctly) the dipole moment divided by the volume of Earth, and hence it would be the average *magnetization* of Earth – a very meaningful quantity, which would be useful in comparing the magnetic properties of Earth with those of the other planets.

17.7 *Additional Remarks*

The units erg G^{-1} or J T^{-1} are frequently encountered for magnetic moment. These may be dimensionally correct, although ergs and joules (units of energy) are not quite the same things as dyn cm or N m as units of torque. It could be argued that magnetic moment could be defined from the expression $-\boldsymbol{p.B}$ for the potential energy of a magnet in a magnetic field. But the correct expression is actually constant $-\boldsymbol{p.B}$, the constant being zero only if you specify that the energy is taken to be zero when the magnetic moment and field vectors are perpendicular to each other. This seems merely to add yet further complications to what should be, but unfortunately is not, a concept of the utmost simplicity.

Nevertheless the use of ergs or joules rather than dyn cm or N m is not uncommon, and nuclear and particle physicists commonly convert joules to MeV. Magnetic moments of atomic nuclei are commonly quoted in nuclear magnetons, where a nuclear magneton is $e\hbar/(2m_p)$ and has the value 3.15×10^{-4} MeV T^{-1}. While one is never likely to want to express the magnetic moment of the planet Uranus in nuclear magetons, it is sobering to attempt to do so, given that the magnetic moment of Uranus is quoted as 0.42 Oe km^{-1}. While on the subject of Uranus, I have seen it stated that the magnetic quadrupole of Uranus is or the same order of magnitude as its magnetic dipole moment – though, since these are dimensionally dissimilar quantities, such a statement conveys no meaning.

Another exercise to illustrate the points I have been trying to make is as follows. From four published papers I find the following. The magnetic moment of Mercury is 1.2×10^{19} A m^2 in one paper, and 300 nT R_M^3 in another. The magnetic moment of Uranus is 4.2×10^{12} Oe km^3 in one paper, and 0.23 G R_U^3 in another. The radii of Mercury and Uranus are, respectively, 2.49×10^6 m and 2.63×10^7 m. Calculate the ratio of the magnetic moment of Uranus to that of Mercury. If you are by now completely confused, you are not alone.

17.8 *Conclusion*

Readers will by now probably be bewildered at the complexities described in this chapter. After all, there could scarcely be a simpler notion than that of the torque experienced by a magnet in a magnetic field, and there would seem to be no need for all of these complicated variations. You are right – there is no such need. All that need be known is summarized in Sections 17.2 and 17.3. The difficulty arises because authors of scientific papers are using almost all possible variations of what they think is meant by magnetic moment, and this has led to a thoroughly chaotic situation. All I can do is to hope that readers of these notes will be encouraged to use only the standard SI definition and units for magnetic moment, and to be aware of the enormous complications arising when they depart from these.

CHAPTER 10
ELECTROCHEMISTRY

For a long time I have resisted writing a chapter on electrochemistry in these notes on electricity and magnetism. The reason for this, quite frankly, is that I am not a chemist, I know relatively little about the subject, and I am not really qualified to write on it. However, a set of notes on electricity and magnetism with no mention at all of this huge topic is not very satisfactory, so I should perhaps attempt a little. I shall do little, however, other than merely introduce and define some words.

We can perhaps think of two sorts of cell with rather opposite purposes. In an electrolytic cell, we pass an electric current through a conducting liquid through two electrodes, which may be of the same or of different metals. The object may simply be to see what happens (i.e. scientific research); or it may be to deposit a metal from a salt in the electrolytic solution on to one of the electrodes, as, for example, in silver plating, or in the industrial manufacture of aluminium; or it may be to break up the electrolyte into its constituent elements, as, for example, in a classroom demonstration that water consists of two parts of hydrogen to one of oxygen. The process is called *electrolysis*; the Greek etymology of the word electrolysis suggests "loosening" by electricity.

The other sort is what we commonly call a "battery", such as a flashlight battery or a car battery. In a “battery”, we have an electrolyte and two metal poles (generally of different metals, or perhaps a metal and carbon). Because of chemical reactions inside the battery, there exists a small potential difference (usually about one or two volts) across the poles, and when the "battery" is connected to an external circuit, we can extract a continuous current from the battery. Strictly, we should call it a “cell” rather than a “battery”. A “battery” is a battery of several cells in series. Usually a flashlight holds a battery of two or three cells. A car battery is genuinely a battery of several connected cells and can correctly be called a battery. Unfortunately in common parlance we often refer to a single cell as a “battery”. In order to distinguish a cell in this sense from what I have called an "electrolytic cell", I shall refer to a cell from which we hope to extract a current as an "electrical cell". I hope these opposite terms "electrolytic cell" and "electric cell" do not prove too confusing. If they do, I'd be glad of suggestions. One suggestion that I have heard is to call an electric cell a "galvanic cell". Another is a "voltaic cell".

In an electrolytic cell, the *positive* electrode is called the *anode*, from a Greek derivation suggesting "up". The *negative* electrode is the *cathode*, from a Greek derivation suggesting "down". In the electrolyte, current is carried by positive ions and negative ions. The *positive ions*, which move toward the cathode, are called *cations*, and the *negative ions*, which move towards the anode, are called *anions*.

Do you find it confusing that the positive electrode is the anode, but the positive ion is the cation? And that the negative electrode is the cathode, but the negative ion is the anion? If you do, you are not alone. I find them confusing. Solution: I suggest that you call the positive electrode the *positive electrode*; the negative electrode the *negative electrode*; the positive ion the *positive ion*; and the negative ion the *negative ion*. That way there is no likelihood of your being misunderstood.

Now, when we come to electrical cells, it may be that the roles of the electrodes are reversed. What was a positive electrode in an electrolytic cell may be the negative side of an electrical cell. What are we going to call them? I suggest that, when we are talking about electrical cells we do not use the word "electrode" at all. We shall refer to the *positive pole* and the *negative pole* of an electrical cell.

Electrolysis of Water

I vaguely remember my first impressions of what is supposed to happen when two platinum electrodes are dipped into water and a current is passed into the water. My guess is that, if the water were very pure water (which is quite difficult to come by) very little would happen. Pure water has very few ions in it (we'll discuss just how many a little later) and its electrical conductivity is quite small – about 5.5×10^{-6} S m^{-1}. However, real water is not pure; it usually has enough impurities in it to supply plenty of ions and to make it a good conductor (and hence a danger in the presence of high-voltage equipment).

My early impression (not quite accurate) of what happens when an electric current passes through water was something like this. Water contains, in addition to billions of molecules of H_2O, a few ions formed through the partial dissociation of H_2O:

$$H_2O \leftrightarrow H^+ + OH^-$$

The hydrogen ions, which are the ~~cations~~ positive ions, move one way, and the hydroxyl ions, which are the negative ions, move the other way.

Well, I'm happy to believe in the existence (if not, maybe, the movement) of the hydroxyl ions, but not about the H^+ ions, which are bare protons with an enormous electrical field. I don't think bare protons exist in any liquid electrolyte, let alone water. I think the reaction is more like

$$H_2O + H_2O \leftrightarrow H_3O^+ + OH^-$$

The positive ion is actually a hydrated proton, also known as "hydronium".

My other problem is that I have difficulty in imagining these great clumsy ions trying to barge their way (the H_3O^+ ions one way, and the hydroxyl ions the other way) through the milling crowd of H_2O molecules. If you imagine the positive electrode to be at the left and the negative electrode at the right, so that the electric field is from left to right, I think what happens when an ion bumps into a neutral water molecule is something like this:

$$H_3O^+ + H_2O \rightarrow H_2O + H_3O^+$$

and

$$H_2O + OH^- \rightarrow OH^- + H_2O$$

In either case, a proton is swapped between the jostling bodies, and so the proton moves from left to right.

Be that as it may, the following reactions occur at the negative electrode
A fresh supply of hydroxyl ions is continuously being created while hydrogen molecules are being released in gaseous form

$$2H_2O + 2e^- \rightarrow 2OH^- + H_2\uparrow$$

Also, arriving hydrated protons are neutralized:

$$2H_3O^+ + 2e^- \rightarrow 2H_2O + H_2\uparrow$$

At the positive electrode, a fresh supply of hydrated protons is continuously being created while oxygen molecules are being released in gaseous form. This happens by electrons from H_2O molecules being transferred to the positive electrode, in several stages, but with net result:

$$6H_2O - 4e^- \rightarrow 4H_3O^+ + O_2\uparrow$$

In addition, arriving hydroxyl ions are neutralized:

$$4OH^- - 4e^- \rightarrow 2H_2O + O_2\uparrow$$

The whole process might seem a bit complicated, but, since we start off in water with twice as many hydrogen atoms as oxygen atoms, naturally the electrolysis results in hydrogen molecules being produced at twice the rate of oxygen molecules. Furthermore, the net result is the same as if we had supposed that there were indeed OH^- and H^+ ions moving through the water as in the simplest model.

One last point before we leave water. I mentioned that the electrical conductivity of pure water is very small, and the number of ions is very small, so that probably not much would happen if you tried to electrolyse pure water. Real water is rarely pure and it has enough impurity ions in it to allow an electrolysis experiment to proceed smoothly. In pure water at room temperature there are about 10^{-7} moles of hydrogen ions (hydrated protons) and a similar number of hydroxyl ions per litre. In SI units, that is also 10^{-7} kmole per m^3. If you add a little bit of acid, you of course very much increase the number of hydrogen ions, and correspondingly the number of hydroxyl ions drops. (Don't worry – the solution remains electrically neutral! If, for example, the acid is HCl, there will be lots of Cl^- ions.) The number of hydrogen ions per litre might be, for example, 10^{-5} mole per litre, or kmole per m^3. If you add a little alkali, you increase the number of hydroxyl ions and correspondingly decrease the number of hydrogen atoms, perhaps to 10^{-9} mole per litre, or kmole per m^3. (Don't worry – the solution remains electrically neutral! If, for example, the alkali is NaOH, there will be lots of Na^+ ions.)

The absolute value of that exponent (7 for the neutral solution, 5 for the acidic solution, 9 for the alkaline solution) is called the *p*H. I think the symbol *p* was originally chosen from the German *Potenz*, or potential or power. The *p*H is commonly used to describe the acidity or alkalinity of a solution. It is 7 for a neutral solution, less than 7 for an acidic solution, and greater than 7 for an alkaline solution. Actually, *p*H is just a rather rough indication of acidity, and there is often not much value in quoting a *p*H value to a large number of significant figures. For example, in hot water, more water molecules are dissociated, so there are more hydrogen ions (and of course more hydroxyl ions) so the *p*H value would nominally go down – though the hot water is no more acidic that it was when it was cold.

Electrolysis of silver nitrate.

There are countless examples of electrolysis that could be told in an encyclopaedia devoted to the subject, many of which are industrial applications. But I don't want to (indeed cannot) write a chemistry book. And indeed all that I need here is one example that is sufficient to introduce some scientific words. I choose the electrolysis of silver nitrate partly because it is simple; partly because at one time the practical definition of the amp was based on the rate of deposition of silver from silver nitrate solution; partly because it enables me to mention *Faraday's law* and the definition of the *faraday*, and even to mention *Avogadro's number*; partly because the electrolysis of silver nitrate can be used for very precise measurement of electric current, and partly because it is used in practice for silver plating.

The stoichiometric chemical formula for silver nitrate is $AgNO_3$, though in fact, both in the crystalline state and in solution, it consists of Ag^+ and NO_3^+ ions. During electrolysis of silver nitrate solution with silver electrodes, silver, of course, is deposited on the negative electrode. Oxygen is given off at the positive electrode. I mention this, because at one time (apart from the obvious practical use in silver plating), this was used as the practical definition of the unit the amp, or ampère. Indeed an "international amp" was that steady current which would deposit silver at a rate of 0.001118 grams of silver per second from a solution of silver nitrate. (This value is very close indeed to, though a tiny bit less than, the SI definition of the amp described in Chapter 6.) Oxygen is given off at the positive electrode.

Faraday measured the mass of many metals deposited from various electrolytes, and he enunciated what is now known as *Faraday's Law of Electrolysis.* In simple terms, and without using phrases that are more familiar to chemists than to physicists, Faraday's law could be stated as:

$$m \propto \frac{It\mu}{\upsilon}$$

That is to say, the mass m of a metal deposited on the cathode during electrolysis is proportional to the total quantity of electricity passed (that is It, where I is the current and t is the time), to the molar mass (i.e. the mass of a mole of the metal, popularly known as the "atomic weight") and inversely proportional to the "valence" or "valency" υ, which is just the charge on the ion. For example aluminium has a valence of 3. Its ion has a charge of plus 3 electronic charges, and consequently it needs three times the amount of electricity to deposit a mole of Al than to deposit, say, a mole of univalent Ag. The molar mass of a metal divided by its valence is called the *electrochemical equivalent* of the metal. It is the *mass of metal that would be deposited by a* *coulomb* of electricity. ($\frac{\mu}{\upsilon} = \frac{m}{It}\cdot$) A silver nitrate electrolytic cell is in fact a very suitable instrument for the highly precise measurement of electric current. In this connection it is called a *voltameter*, though I suppose "ammeter" would really be a better name!

So, how much electricity is required to deposit a mole of charge? (That is to say a mole of the metal ions times the charge on each.) The amount of electricity required to deposit a mole of charge (e.g. 108 grams of Ag, or 27 ÷ 3 = 9 grams of Al) from solution is called a *faraday*. It is about 96,485 C, which is the charge of a mole (Avogadro's number) of electrons.

This is worth knowing, for a good examination question for physics students might be "How is Avogadro's number determined?" This is obviously a very important thing for a physicist to know, but a physics student with only a modest background in chemistry might not immediately think of the answer, which is: by measuring the size of the faraday by measuring the mass of silver deposited from silver nitrate solution. (Of course you have to use the SI definition of the amp for this – it wouldn't do to use the definition of the "international amp"!)