1

Charge and Electric Fields

1.1 Charge as a Fundamental Property of Matter

All matter is comprised of fundamental sub-atomic particles which themselves have basic properties, such as mass, charge and spin. The sub-atomic particle which electrical engineers are probably most concerned with is the electron, which was discovered by the Nobel Prize winning physicist J.J. Thomson in Cambridge in 1897 (Thomson 1897). Among its fundamental properties, electrons have a mass of 9.109×10^{-31} kg and a charge of -1.602×10^{-19} C, the magnitude of which we call e. Of these two properties, we are probably more familiar with the concept of mass as the world around us is dominated by gravitational forces at the macroscopic scale. We understand that two particles which have non-zero mass both experience a force between them. We rationalize this by saying that a particle with mass produces a gravitational 'field' which extends spatially away from the particle. Another particle with a non-zero mass inside this field then experiences a force.

As engineers, we do not tend to worry about the exact mechanism by which this force is being exerted over some distance, leaving that important consideration to physicists. Instead, we simply apply the equations for force between masses, such as that exerted by the earth on all structures in civil engineering. Therefore, we should be content to accept the less familiar concept of charge on the same basis: a particle with a non-zero charge produces an *electric field* which extends spatially away from the particle, and another particle with a non-zero charge inside this field then experiences a force. We use the symbol Q or q to denote charge, and the SI unit of charge is the *coulomb* (C).

1.2 Electric Field and Flux

Fields are widely used in physics to describe regions of space in which an object with a particular property experiences a force. Therefore, an electric field is a region of space in which an object with charge q experiences a force $\mathbf{F}$. As force is a vector quantity, having both a magnitude and direction, electric field must also be a vector quantity $\mathbf{E}$, so that

$$\mathbf{F} = \mathbf{E}q \tag{1.1}$$

From this equation, it can be seen that the unit of electric field is N C^{-1}, although, as we will see in Section 1.3, the more common unit is V m^{-1}.

Electromagnetism for Engineers, First Edition. Andrew J. Flewitt.

Companion website: www.wiley.com/go/flewitt/electromagnetism

In everyday life we experience objects with both positive and negative charge, because while electrons have a charge of $-e$, protons, which are one of the sub-atomic constituents of the nucleus of atoms, have a positive charge of e. Therefore, the force acting on a positive charge at a particular point in an electric field will act in the opposite direction to the force on a negative charge at the same point. This is the reason why like charges repel each other whereas opposite charges attract. This is in contrast to mass, which is positive for all matter, and therefore the force between masses is always attractive.

To assist us in visualizing fields, we use the concept of *flux*. We imagine that the electric field is composed of lines of flux whose direction at a given point in space is the direction of the electric field at that point and whose number density per unit area relates to the magnitude (or intensity) of the electric field.

We know that charge produces electric fields, and therefore lines of electric flux begin on positive charges and end on negative charges. As the total sum of all charge in the universe is zero, it must be the case that every line of flux that begins on a positive charge must have a balancing negative charge somewhere to end on. We can now visualize the electric field around a small point charge $+q$ in free space (a vacuum) in Figure 1.1. If we assume that the balancing charge of $-q$ is uniformly distributed an infinite distance away, then lines of electric flux will radiate uniformly away from the point charge. This will cause the field to decrease with increasing radial distance r from the point charge as $1/r^2$, just as we find for gravitational fields around mass as well. This is the basis behind the *Coulomb law* for the magnitude of the force F that acts on a charge q_2 in an electric field of magnitude E_1 produced by another point charge q_1:

$$F = E_1 q_2 = \left(\frac{q_1}{4\pi\varepsilon_0 r^2}\right) q_2 \tag{1.2}$$

where r is the distance between the charges.

In Eq. (1.2) we have had to introduce a new quantity ε_0, which is the *permittivity of free space*. It is a fundamental constant with the value $8.854 \times 10^{-12}\ \mathrm{F\,m^{-1}}$, and it is required to yield a result for the force between two charges that is correct in SI units.

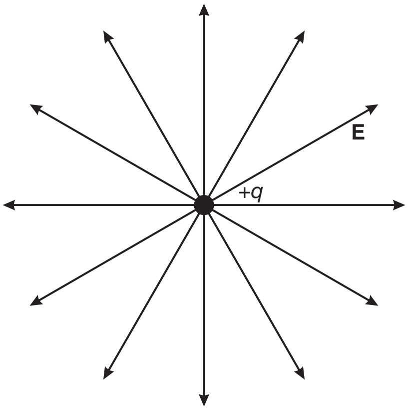

Figure 1.1 Lines of electric flux around a point charge $+q$ in free space assuming that the balancing charge is uniformly distributed an infinite distance away. The direction and area density of the flux lines are a measure of the electric field **E** at any point.

1.3 Electrical Potential

Let us imagine that we have two point charges of equal magnitude but opposite sign, $+q$ and $-q$, which initially exist at the same point in free space, so that there is no electric field around the charges. If we slowly move the charge $-q$ away from the $+q$ charge so that the distance r between them is increasing, then an electric field distribution will be created around and between the charges as shown in Figure 1.2. As there will be a force acting on the charge that is being moved, given by Eq. (1.2), there must be some change in energy – work is being done against the force that is attracting the charges together. In this case, mechanical energy is being converted into an electrical potential energy, which could be converted back into mechanical energy again by allowing the two charges to accelerate back towards each other once more. It is a key concept that whenever a field (whether electric, magnetic or gravitational) occupies a volume of space, then some potential energy has been stored.

We can use basic mechanics to relate electric field to potential energy. If we have a charge q, in an electric field of magnitude E, which is moved by a small distance δx in the direction of the electric field, then there will be a change in potential energy of the charge given by

$$\delta W = -F\delta x = -Eq\delta x \tag{1.3}$$

where F is the magnitude of the force acting on the charge due to the electric field. The change in potential energy is negative as the force acting on the charge is in the same direction as the movement. We define a new quantity, the *potential difference*, which is the change in energy per unit charge between two points in space. The potential difference is given the symbol V and has units of volts. Therefore, the small potential difference δV between the two points separated by the distance δx over which we have moved our charge q is

$$\delta V = \frac{\delta W}{q} \tag{1.4}$$

Equating δW in Eq. (1.3) and (1.4) gives

$$q\delta V = -Eq\delta x \tag{1.5}$$

and therefore, by basic calculus, we have the result that

$$E = -\frac{dV}{dx} \tag{1.6}$$

In other words, the electric field is the negative of the potential gradient. For readers who are familiar with vector calculus, we can rewrite this in three dimensions as

$$\mathbf{E} = -\nabla V \tag{1.7}$$

It should be noted that we can only ever talk about a potential difference *between two points*, for example the potential at a point a with respect to a point b which we could denote V_{ab}. The direction is significant, as $V_{ba} = -V_{ab}$. We often use arrows to denote a potential difference where we are considering the potential at the tip of the arrow with respect to the tail. If we know the electric field distribution between the two

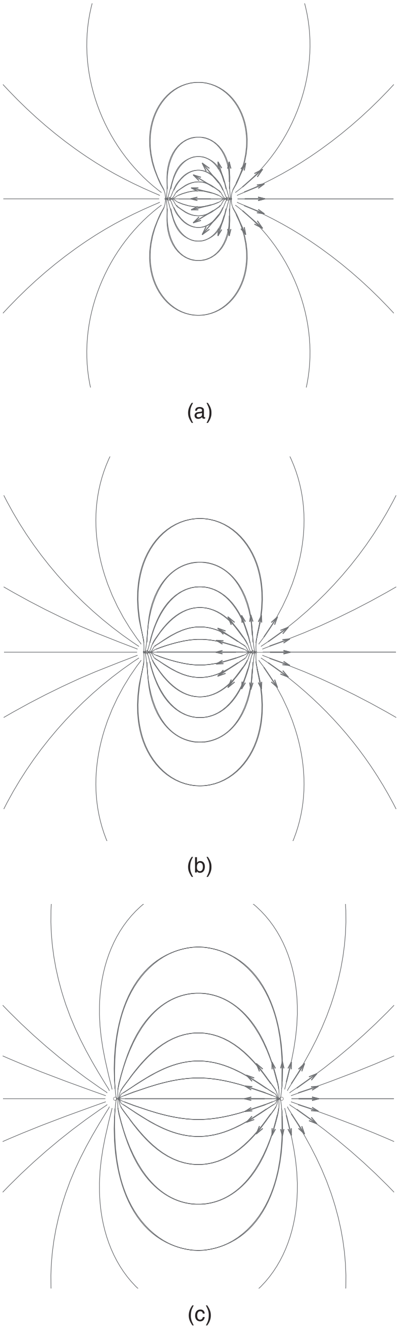

Figure 1.2 The electric field distribution around two point charges $+q$ (right) and $-q$ (left) separated by increasing distance from (a) to (c).

points, then we can evaluate this. A simple integration of the two sides of Eq. (1.6) would suggest that

$$V_{ab} = -\int_b^a E\,dx \tag{1.8}$$

However, this is a slight simplification as we know that the electric field is actually a vector quantity, and it is therefore only the component of a small movement dx parallel to the direction of the electric field that will lead to a change in potential. If we use vectors to express both the electric field **E** and the small movement **dx**, then the scalar (dot) product of the two yields exactly this result. Therefore, Eq. (1.8) is more generally expressed as

$$V_{ab} = -\int_b^a \mathbf{E} \cdot \mathbf{dx} \tag{1.9}$$

In practice, Eq. (1.8) is more commonly used in simple calculations. This is achieved by ensuring that a path between two points is chosen so that **E** and **dx** are always either parallel or perpendicular to each other so that the scalar product is either a simple product or zero at any point.

This picture is very similar to the situation for gravity, where we can define a gravitational potential difference as being the difference in gravitational potential energy per unit mass between two points. In practice, we often take the surface of the earth as a reference point from which to measure gravitational potential as a function of height h above the surface of the earth which, assuming the gravitational constant g to be uniform, is simply gh. Likewise, it is helpful to take a constant reference point from which to measure electric potential difference. The earth again provides a good practical reference point, as it is so large that small changes in the charge on the earth have no significant impact upon its electrical potential energy. For other situations, such as the point charge in free space shown in Figure 1.1, the earth is simply not present, and so we have to choose an alternative reference from which to measure potential difference. In this example, an infinite distance away from the charge ($r = \infty$) provides a good reference point. We can therefore use Eq. (1.8) to calculate an expression for the potential difference with respect to this reference $V(r)$ around the point charge in free space from the expression for the electric field around the charge in Eq. (1.2):

$$V(r) = -\int_\infty^r \frac{q_1}{4\pi\varepsilon_0 r^2} dr = \left[\frac{q_1}{4\pi\varepsilon_0 r}\right]_\infty^r = \frac{q_1}{4\pi\varepsilon_0 r} \tag{1.10}$$

Note that we have been able to use the scalar Eq. (1.8) rather than the vector Eq. (1.9) by choosing a radial path which is parallel to the radial electric field. To help visualize potentials, we often plot lines of equipotential, rather as we plot contour lines on a map of constant height and also therefore constant gravitational potential, and this is shown for the point charge in Figure 1.3. As electric field points down a potential gradient according to Eqs. (1.6) and (1.7), lines of electric flux always cut perpendicularly through lines of equipotential.

It is clear from Eq. (1.10) that the potential difference with respect to $r = \infty$ at any point in free space around the point charge is uniquely defined. This is always true as we are dealing with a linear system. Therefore, it does not matter what path is actually chosen over which to evaluate the potential difference between two points using the integral in Eq. (1.9) – the result will always be the same. It is for the same reason that if you walk up a hill, the change in your gravitational potential energy will not depend on which path you

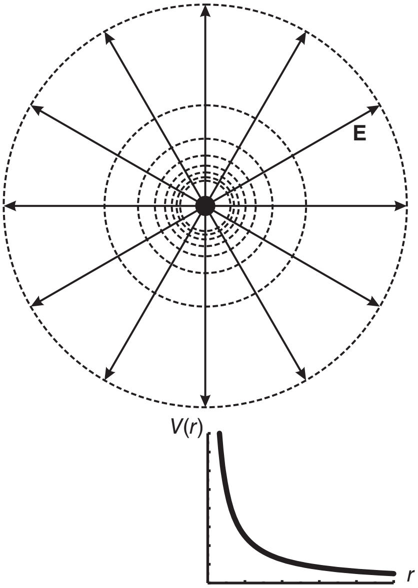

Figure 1.3 Lines of electric flux around a point charge $+q$ in free space with dashed lines of equipotential calculated with respect to $r = \infty$, with the graph of how potential varies as a function of distance r from the charge shown underneath on the same length scale.

take. Likewise, if we choose a closed path that ends back at its starting point, then there will be no potential difference between the start and end of the loop, which can be expressed mathematically from Eq. (1.9) as

$$\oint_C \mathbf{E} \cdot \mathbf{dx} = 0 \tag{1.11}$$

where $\oint_C$ indicates the integral round a closed loop.

1.4 The Gauss Law of Electrostatics in Free Space

Our starting point for the discussion about charge as a fundamental property of matter in Section 1.1 was that charge produces an electric field. In the following discussion we then refined this to say that lines of electric flux begin on positive charges and end on negative charges. The *Gauss law of electrostatics* takes this basic statement about the origin and nature of electric fields and expresses it in a rather elegant mathematical formulation, namely,

$$\oint_S \mathbf{E} \cdot \mathbf{dA} = \frac{q}{\varepsilon_0} \tag{1.12}$$

where the left-hand side is an integral over a closed *Gaussian surface* and q is the net charge enclosed by the surface. We will now consider the mathematical basis for this equation.

Let us imagine that we have an arbitrary surface in three dimensions, such as that in Figure 1.4. We could work out its surface area by laying it out onto a flat surface and measuring its area, but calculus and vector algebra allow us to take a more elegant approach. We could imagine dividing up the surface into lots of small elements of the surface, each of which approximates to a flat parallelogram, as shown in Figure 1.4. We define two vectors

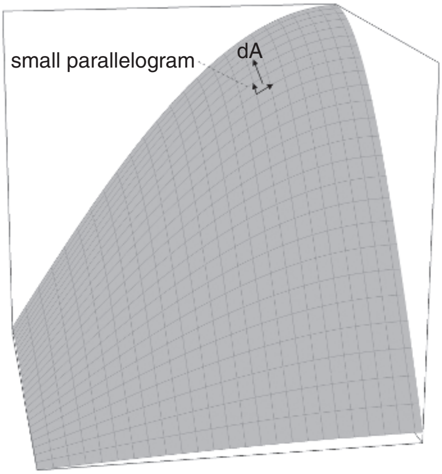

Figure 1.4 An arbitrary surface divided up into small parallelograms. For one parallelogram, the side vectors have been shown as small arrows and the resulting vector element of surface area **dA** is shown.

which point along adjacent edges of the small parallelogram; if we take the cross product of these two vectors then the result is a third vector which points perpendicularly away from the plane of the parallelogram and whose magnitude is equal to the area of the parallelogram. We call this vector **dA**. If we were to integrate the magnitude of all of the **dA** vectors over the whole surface, then we would have the whole surface area,

$$\int_{\text{surface}} |\mathbf{dA}| = \text{total surface area} \tag{1.13}$$

Let us now imagine that an electric field is passing through the surface. Taking the dot product of two vectors, such as **a** and **b**, gives the component of **a** in the direction of **b** multiplied by the magnitude of **b**, or, expressed mathematically,

$$\mathbf{a} \cdot \mathbf{b} = |\mathbf{b}|(|\mathbf{a}| \cos\theta) \tag{1.14}$$

where θ is the angle between **a** and **b**. Therefore, if we take the dot produce of **E** with any one of the small **dA** vectors, the result will be the component of **E** perpendicular to the surface multiplied by the area of the small surface element. Therefore, as $|\mathbf{E}|$ is the number of flux lines per unit area, $\mathbf{E} \cdot \mathbf{dA}$ is the number of flux lines passing perpendicularly through the small element of surface. We are taking a very similar approach to that used when calculating the force exerted by a ladder that is resting against a wall. For this mechanics problem we would resolve the force to calculate the component of force acting perpendicular to the surface of the wall. Here we are just resolving the electric field to take the component perpendicular to the surface.

We can therefore calculate the total number of electric flux lines passing perpendicularly through any surface by simply integrating $\mathbf{E} \cdot \mathbf{dA}$ over the surface,

$$\int_{\text{surface}} \mathbf{E} \cdot \mathbf{dA} = \text{total flux through surface} \tag{1.15}$$

Therefore, if the surface is closed (such as a box, sphere, or cylinder) then this integral would give the total flux passing out of the surface. We should note that the mathematical convention is that vector elements of surface area, **dA**, always point out of closed surfaces.

We should now be able to intuitively understand the meaning of the Gauss law of electrostatics as expressed in Eq. (1.12). We can call any closed surface that we can imagine in space a Gaussian surface. By taking the integral of $\mathbf{E} \cdot \mathbf{dA}$ over that surface, we are effectively counting flux lines. A line leaving the volume enclosed by the surface counts positively, while one entering counts negatively. As lines of electric field begin and end on charges, if there is no net charge enclosed within the Gaussian surface then the result of the integral must be zero. As many lines of flux leave through the Gaussian surface as enter. The result can only be non-zero if a line of flux has either begun or ended on a charge that is somewhere in the volume enclosed by the Gaussian surface. This is exactly what Eq. (1.12) states: that $\oint_S \mathbf{E} \cdot \mathbf{dA}$ is proportional to the net charge q enclosed by the surface. We only now need $1/\varepsilon_0$ as a constant of proportionality to ensure that the correct numerical values are produced.

1.5 Application of the Gauss Law

The Gauss law of electrostatics allows us to calculate the electric field produced by distributions of charge in space, and consequently the potential using Eqs. (1.6) or (1.7). The presence of the surface integral in Eq. (1.12) can make the Gauss law appear rather intimidating in the first instance. However, in practice we can often calculate the electric field produced by quite complex charge distributions through careful choice of the Gaussian surface. This may be achieved by ensuring that the electric field and vector elements of surface area are either parallel or perpendicular to each other, normally by reflecting the geometry of the charge distribution in that of the Gaussian surface.

For example, we can very easily derive the expression for the electric field around a point charge that is implicit in the Coulomb law (Eq. (1.2)). Intuitively, we would expect a point charge q to produce a spherically symmetric electric field, as shown in Figure 1.3. Therefore, we should choose a Gaussian surface that is a sphere of radius r centred on the charge, where the Gauss law will allow us to evaluate the electric field at that radius. As both the electric field $\mathbf{E}$ and the surface vectors $\mathbf{dA}$ both point radially outwards, the dot product of $\mathbf{E}$ and $\mathbf{dA}$ becomes just a simple product. Also, as can be clearly seen in Figure 1.3, the magnitude of the electric field $E(r)$ is a constant at a given radius. Therefore, for this scenario using Eq. (1.13) we have

$$\oint_S \mathbf{E} \cdot \mathbf{dA} = E(r) \oint_S |\mathbf{dA}| = E(r) \times \text{area} \tag{1.16}$$

As the area of the sphere is $4\pi r^2$, Eq. (1.12) reduces simply to

$$E(r) \cdot 4\pi r^2 = \frac{q}{\varepsilon_0} \tag{1.17}$$

which then rearranges to the familiar expression

$$E(r) = \frac{q}{4\pi\varepsilon_0 r^2} \tag{1.18}$$

We can also handle more complex charge distributions, such as an infinitely long line of charge with a charge per unit length of σ. In this case, the system has cylindrical symmetry,

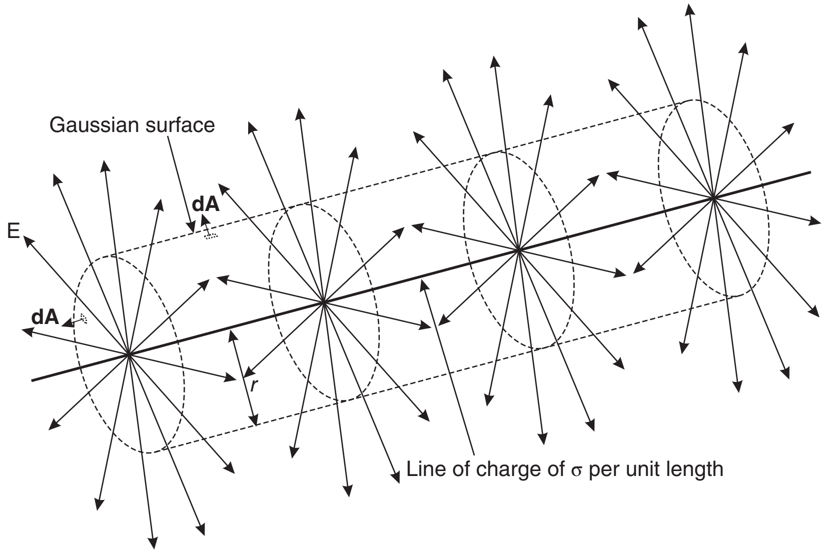

Figure 1.5 Lines of electric flux around a line of charge of $+\sigma$ per unit length in free space. A cylindrical Gaussian surface of length l and radius r is shown using a dashed line. Two elements of surface area **dA** are shown: one on an end of the cylinder pointing axially outwards and one on the curved surface pointing radially outwards.

and so we should choose a Gaussian surface which is a cylinder axially centred on the line of charge, with some arbitrary length l and a radius r, as shown in Figure 1.5. In this case, the vector elements of surface area on the end caps of the cylinder will point in the direction of the axial line of symmetry of the system, whereas the electric field will point radially outwards. Hence, $\mathbf{E} \cdot \mathbf{dA} = 0$, so these do not need to be considered further. Over the curved surface of the cylinder, **E** and the surface vectors **dA** both point radially outwards, and so we can use Eq. (1.16) again. The area of the curved surface of the cylinder is $2\pi rl$, and therefore the Gauss law reduces to

$$E(r) \cdot 2\pi rl = \frac{\sigma l}{\varepsilon_0} \tag{1.19}$$

where σl is the charge enclosed within the cylinder. This rearranges to

$$E(r) = \frac{\sigma}{2\pi\varepsilon_0 r} \tag{1.20}$$

Therefore, although Eq. (1.12) may appear rather complex at first sight, involving surface integrals of vector quantities, in many cases the actual application of the Gauss law only requires scalar multiplication of fields and areas.

1.6 Principle of Superposition

We have seen that the Gauss law of electrostatics can be used to calculate the electric field around geometrically ‘simple’ distributions of charge. We can then use this as a basis for calculating the field around more complex distributions using the *principle of superposition*.

The principle of superposition in its most general form as it applies across diverse branches of physics states that if there is a linear relationship between some stimulus and a response, then the total response due to many stimuli acting simultaneously is the same as the sum of the responses that each stimulus would have produced individually. In the case here, the stimulus is the charge which produces a response in the form of an electric field, and the Gauss law in Eq. (1.12) clearly shows that these are proportional to each other (in other words, they are linearly related). Therefore, we can break down a complex charge distribution into many small elements of charge, calculate the electric field distribution due to each of the elements and then simply add all the contributions to the electric field at each point in space to yield the total field at that point.

We could have used superposition to calculate the electric field around the line of charge shown in Figure 1.5 by dividing the line up into infinitesimally small elements of charge and then integrating the resulting fields at any point to determine the total field, but the symmetry of this charge distribution makes the approach using the Gauss law directly a significantly simpler means of deriving Eq. (1.20).

1.7 Electric Dipoles

Two charges of equal magnitude but opposite sign separated by a small distance are called a *dipole*. The electric field distribution around the dipole of two point charges $+q$ and $-q$ in Figure 1.2 has been calculated using the Principle of Superposition. Each of the two charges would independently produce their own electric field distribution, one pointing radially outwards and the other radially inwards, as shown in Figure 1.6. If we imagine the plane passing through all points that are equidistant between these two charges, then the components of the electric fields due to the two charges pointing parallel to this plane will cancel out when superposed, leaving only the perpendicular components which sum, as is clearly the case in Figures 1.2 and 1.6. If the two charges are separated by a distance d then, from Eqn. (1.18), the electric field at the equidistant point along the line between the two charges due to each of the two charges independently will be

$$E_+ = \frac{q}{4\pi\varepsilon_0(d/2)^2} \tag{1.21a}$$

$$E_- = \frac{-q}{4\pi\varepsilon_0(d/2)^2} \tag{1.21b}$$

To calculate the total electric field at this point, we would need to sum Eqs. (1.21a) and (1.21b). However, the two expressions have been calculated in different frames of reference: one pointing away from the positive charge and one pointing away from the negative charge. We have to use a common frame of reference when summing the two fields. In the frame of reference used for the positive charge, the electric field due to the negative charge at the equidistant point is $-E_-$, so the total field is

$$E = E_+ - E_- = \frac{2q}{\pi\varepsilon_0 d^2} \tag{1.22}$$

Therefore, the electric field at this equidistant point decreases as the distance between the dipole increases, and this can be seen in the density of the flux lines around this point

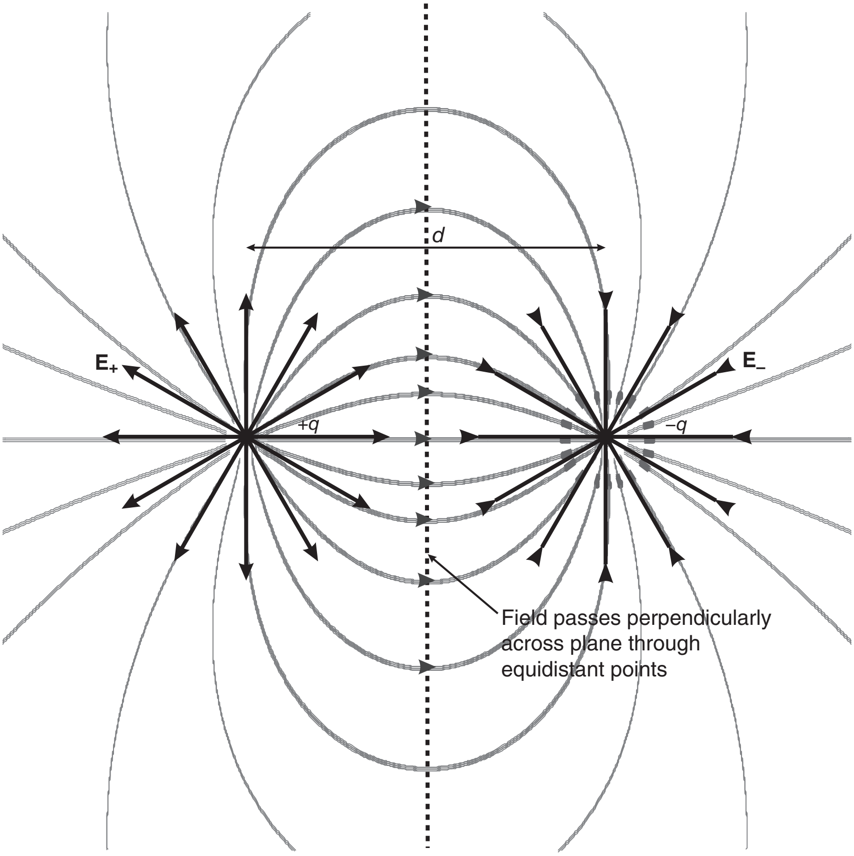

Figure 1.6 The electric fields $\mathbf{E}_+$ and $\mathbf{E}_-$ produced separately by each of the point charges in a dipole and the resultant total electric field calculated by superposition. The field passes perpendicularly through the plane of equidistance between the two charges.

shown in Figure 1.2, which are densest in Figure 1.2a where the dipole separation is smallest. At long distances away from the dipole (much greater than d) the two charges appear to be at the same point in space, and their electric fields will cancel out to leave no net field, and so the electric field is effectively concentrated in the region of space close to the two charges. This effect is also most pronounced in Figure 1.2a.

In practice, dipoles frequently appear in real situations as most objects have no net charge, and so the act of removing some charge by some distance from a particular object immediately creates a dipole between the removed charge and the original object. At the atomic scale, when an atom experiences an externally applied electric field, the electrons around the nucleus will be displaced from their equilibrium position, as shown in Figure 1.7. Just as we can think of a distributed mass acting from a ‘centre of mass’ point, so we can think of the electrons around the nucleus as acting from a ‘centre of charge’. The action of the electric field is therefore to move this effective centre of charge by a vector displacement $\mathbf{d}$ from the positive charge, resulting in the formation of a dipole.

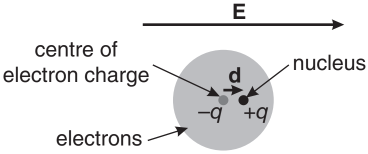

Figure 1.7 A simplistic diagram of the effect of an externally applied electric field **E** on a single atom, which results in an effective separation of the positive charge $+q$ on the nucleus and the negative charge $-q$ of the electrons by a displacement vector **d**, which results in a dipole moment $\mathbf{p} = q\mathbf{d}$.

A helpful quantity that is often used when performing calculations involving dipoles is the *dipole moment* **p**. It is a vector quantity that is simply defined as the product of the magnitude of the dipole charge q and the vector displacement of the positive charge from the negative charge **d**, that is,

$$\mathbf{p} = q\mathbf{d} \tag{1.23}$$

The dipole moment has units of C m.

Reference

Thomson, J.J. (1897). XL. Cathode rays. *Philosophical Magazine Series 5* 44 (269): 293–316.