I Electromagnetic Theory in the Optical Domain

Introduction to Part I

The first part of this book on optical imaging provides the reader with the necessary background in electromagnetic theory, relevant for solving optical problems. For a long time, optics was closely connected to mechanics, the oldest branch of science and engineering. The physical model for describing optical phenomena was largely inspired by mechanical analogues. Optical rays were represented as a stream of tiny particles, emitted by a source and propagating in a rectilinear manner, with very high speed. With respect to human vision Greek philosophers, for instance Plato, postulated the emission theory in which the eye emits beams of light that are reflected back from the environmental scene. This theory was later challenged by Euclid who wondered how one could see the very distant stars immediately after opening one's eyes during the night. It was not until the tenth century, in the work of Al-Haytham, that the eye was considered to receive independent optical rays from the outside scene, illuminated by other sources of light. The 'mechanic' nature of light has persisted through the ages, advocated among others by Descartes. A wave theory of light was put forward by Hooke and Huygens in the seventeenth century but it did not attract much attention. An eminent supporter of the particle or corpuscular theory of light was Newton. Numerous experiments on the colour of light itself and on the coloured fringes observed between two optical surfaces were performed by him in the years between 1665 and 1704 when his book *Opticks* was first published (see also **[258]**, the fourth edition of 1730) . His novel observations and experimental results were all explained in the framework of the corpuscular light theory.

The Descartes/Snell refraction law applied to Newton's mechanistic optical model requires a higher light propagation speed in glass than in air. This was made plausible by Newton by means of the attraction exerted by the glass material at the interface air/glass on the incident light corpuscles. Once inside the medium, the light corpuscles continue at the higher speed they have acquired at the transition from a less dense to a denser medium. To explain dispersion, Newton assumed that the red light particles have a different (larger) mass or shape than the blue light particles. As a consequence the red particles would experience a smaller increase of speed at the interface than the blue particles. The net effect is that refraction becomes smaller towards the red part of the spectrum. A conjecture by Newton that glasses all show the same ratio between dispersion and refraction angle was based on this assumption of a colour-dependent mass or shape of the light particles. Dispersion was thus caused by the nature of the light particles. The glass material, by means of its density, determines solely the average refraction angle. Newton's corpuscular theory was successful in explaining rectilinear propagation, refraction and reflection of light and also, to a lesser extent, the effect of diffraction (discovered by Grimaldi, published after his death in 1663 **[117]** and named 'inflection of light' by Newton).

To quantify the beam intensity of partially reflected and transmitted rays, Newton devised a theory of 'fits of easy reflection and transmission'. This property is carried by a particle from the source on, but it can be modified in the vicinity of, for instance, a glass medium. The impact of a light corpuscle on the glass interface creates a local 'wavelet' in the glass that propagates at reduced speed together with the light particle and leads to an enforcement or decrease of the total light phenomenon. An enforcement of the action of light particle and local wave leads to a 'fit of easy transmission' of the particle, the opposite to an inclination of the particle to be reflected. The distribution of the 'fits' over the corpuscles at emission from the source and their change of 'fit' at an interface were not well understood by Newton. The 'fit' property of a corpuscle was the subject of the first query (number 17) of an extended list of queries that was included by Newton in later editions of *Opticks*.

The corpuscular light theory had difficulty in explaining double refraction in a crystal of calcite, discovered by Bartholin in 1669. This strange phenomenon required at least a change in shape of the light corpuscles from spherical to flattened or rectangular. To explain the polarisation-dependent reflection and transmission coefficients at an interface between two media, the 'fits of easy reflection and transmission' of the corpuscles had to be further detailed in a rather artificial and ad hoc manner by the successors of Newton. Similar unsatisfactory assumptions about the nature of the light particles were needed to explain further experiments with polarised light by Wollaston, Malus and Brewster. In general, the corpuscular light theory was inadequate to deal with what we call today the *transverse* oscillatory nature of light.

The discovery of optical glasses with significantly different dispersion by Dollond in 1758 **[87]** was a first argument against the Newtonian light theory. Half a century later, Huygens' wave theory of light was revived by Young and Fresnel. An important extension was the notion of wavelength which immediately created the link with the colour of light. Fresnel's wave theory was very successful in accommodating the new experimental results with polarised light that were presented around 1810. Fresnel's memoir on *double refraction*, published in integral form in 1824 **[104]**, impressed the scientific community. The phenomenon of conical refraction, discovered shortly after Fresnel's untimely death, turned out to be seamlessly included in his theory. Finally, the coup de grâce for the classical corpuscular light theory was administered by the measurement of the speed of light in water, almost simultaneously by Fizeau and Foucault around 1850. It was only 75% of the speed of light in air instead of the $4 \cdot 10^8$ ms^{-1} that was required by the corpuscular light theory.

Fresnel's wave theory was a 'théorie mécanique', as stated by him in the above-mentioned memoir. Essential for the propagation of a wave is the transmission of the transverse wave motion by the molecules of the (luminiferous) aether. The all-pervading fluid of aether molecules had to be given special properties to permit the transmission of transverse wave movement into the propagation direction. Fresnel argued in his memoir that the optical polarisation experiments were so convincing that the aether fluid had to be given a mechanical property which is uncommon for a fluid, viz. a nonzero shear modulus. The existence of the aether and its relative movement with respect to moving bodies such as the planets was the subject of scientific discussion throughout the second half of the nineteenth century. The experiments by Michelson and Morley showed that no relative movement of the aether could be detected and that, most likely, it did not exist. For that reason, the original idea of Faraday that light was a high-frequency electromagnetic perturbation that could propagate in the absence of an aether medium rapidly gained ground. Since then, Maxwell's electromagnetic theory is considered to be the basis of optical wave phenomena. The twentieth century has brought further extensions of the optical theory, such as the quantum theory for black-body radiation, the quantum theory for the interaction of a photon with matter (photo-electric effect) and the quantum behaviour of the photon or assemblies of photons under the condition of low light levels.

Within the scope of this book on classical imaging optics, it is sufficient that the electromagnetic theory of light is taken as the basis for light propagation and imaging. In the first part of the book we focus on Maxwell's electromagnetic theory, applied to the domain of optical frequencies where in many instances the magnetic properties of a medium can be equated to those of vacuum. In the first chapter, after a general introduction to Maxwell's theory, we discuss the dipole source, Gaussian beam propagation and wave propagation at a perfectly smooth interface. To describe light fields emitted by a two- or three-dimensional object to be imaged, we study multilayer systems and the diffraction by periodic structures embedded in a multilayer. The second chapter of Part I is entirely devoted to wave propagation in anisotropic media, either exhibiting linear birefringence or circular birefringence. The phenomenon of conical refraction is treated in some detail. The third chapter is devoted to guided wave propagation at a planar surface. Special wave propagation properties are discussed associated with the so-called *metamaterials*. It is shown that a plane-parallel plate of an idealised metamaterial would behave as a 'perfect' imaging lens with virtually no limit on spatial resolution.

Cambridge University Press
978-1-108-42808-8 — Imaging Optics
Joseph Braat , Peter Török
Excerpt
More Information

1 Electromagnetic Wave Propagation in Isotropic Media

1.1 Introduction

In the beginning of the nineteenth century, Fresnel's quantitative extension of Huygens' wave theory enabled a detailed description of light propagation in isotropic and anisotropic media, including the diffraction effects arising at sharp edges, tiny holes in a screen or at small obstructions. The wave theory of Fresnel, based on a fine-tuned mechanical aether model to produce the observed optical effects, was quite powerful in describing light wave and light energy propagation. It was not able to explain the effects of magnetic fields on light propagation or reflection, the so-called Faraday and Kerr effects.

Maxwell's electromagnetic theory was needed to establish the firm foundation of light propagation in vacuum and matter. The classical Maxwell theory can be safely used in vacuum and when the material particles involved can be considered to have macroscopic dimensions and properties of which we only need to consider the average values. It is only at very low light levels and when the light interaction with the individual atoms and molecules has to be considered, that we have to switch to the full quantum theory of propagation, transmission, reflection, absorption and scattering of light. In this chapter we use the macroscopic Maxwell's equations as the starting point for wave propagation in the optical domain with the electric and magnetic field quantities represented by three-dimensional vectors. By imposing a simplified approximate solution to Maxwell's equations, we obtain the so-called scalar wave equation and the corresponding wave solution of which the magnitude is given by a single scalar quantity, the complex amplitude of the 'light disturbance'. A further simplification of the solution to Maxwell's equations leads to the ray model of light propagation and to Fermat's principle. It is customary to speak about geometrical optics when using this model, the light energy being propagated along geometrical trajectories that in many practical cases reduce to simple straight lines. Imaging theory based on geometrical optics is subject of Part II of this book, combined with the scalar wave propagation model ('physical optics') if this is necessary to improve the accuracy of the image intensity. In this chapter we treat the parts of electromagnetic theory that, in our view, are relevant for optical imaging. An in-depth treatment of electromagnetic theory can be found in well-known textbooks like **[36]**,**[37]**,**[160]**,**[328]**.

1.2 Maxwell's Equations as Experimental Laws

It is often forgotten, especially by theoreticians, that the four equations now known as Maxwell's equations, namely Gauss' law for electric fields, Gauss' law for magnetic fields, Ampère's law and Faraday's induction law, were once separate and purely experimental laws. It was not until Maxwell realised their relationship in 1861–62, and added the displacement current to Ampère's law, that modern electromagnetism was born.

1.2.1 Electric Field, Electric Flux and Electric Potential

It is perhaps simplest to gain an understanding of Maxwell's equations by first considering the electric and magnetic fields which they govern, since these quantities directly relate to measurable forces which we are familiar with. The definition of the electric field vector **E** originates from Coulomb's experiments on the forces between charges, published in 1785. Coulomb measured the force (the so-called Coulomb force) between charges q_1 and q_2 (units Coulomb [C]) and realised that the force was proportional to both charges and inversely proportional to the square of the distance r between them. The force $\mathbf{F}_E$ (units Newton [N]), which is of course a vector quantity, is parallel to the line connecting the two point charges q_1 and q_2. Coulomb observed that like charges repel each other, and hence the force is directed away from them, whereas opposite charges attract each other, and hence the force is directed towards the charges. If the unit vector along the line connecting the two charges and pointing away from them is denoted by $\hat{\mathbf{r}}$, then $\mathbf{F}_E \propto q_1 q_2 \hat{\mathbf{r}}/r^2$. The electric field due to the charge q_1 is then defined as the force between the two charges divided by the charge q_2, $\mathbf{E} \propto q_1 \hat{\mathbf{r}}/r^2$ (units [N/C] or [V/m], i.e. the force experienced by unit charge). It is therefore clear that electric field lines[1] must start and finish on charges. By convention electric field lines point away from a positive charge and hence towards a negative charge.

The flux of the electric field through a very small *open* surface (differential flux) is defined as the projection of the electric field vector **E** onto the outward surface unit normal $\hat{\mathbf{n}}$ times the area dA of the (differential) surface element. A nonzero net charge inside a *closed* surface $\mathcal{A}$ therefore gives rise to a non-vanishing net flux of electric field **E** through the surface of the volume:

$$\oiint_{\mathcal{A}} \mathbf{E} \cdot d\mathbf{A} = \frac{1}{\epsilon} \sum_i q_i \tag{1.2.1}$$

in the case of discrete charges q_i, or

$$\oiint_{\mathcal{A}} \mathbf{E} \cdot d\mathbf{A} = \frac{1}{\epsilon} \iiint \rho dV \tag{1.2.2}$$

in the case of a distribution of charges in the volume V of volume charge density ρ (units [C/m^3]). Here $d\mathbf{A} = \hat{\mathbf{n}} dA$ is the differential surface normal and ϵ is a constant of proportionality called the permittivity, the significance of which will become clear later. The summation on the right-hand side of Eq. (1.2.1) is over all charges *inside* the closed surface, while those outside the volume do not matter. The latter can readily be explained by the fact that the electric field due to charges outside the volume has a zero *net* flux. Electric fields arising due to a set of stationary charges are also called electrostatic fields. Equations (1.2.1) and (1.2.2) mean that *electrostatic fields are due to electric charges. Field lines do not form loops; they start and end on the charges*.

Electrostatic fields are conservative, which means that if a charge is moved in a closed loop in the presence of such a field then, even though there is in general instantaneous work done along the path, the net work done for the entire path is zero. This is because along a closed loop one can resolve the electric field into two components: one parallel to it and one perpendicular. There is no work done along the component of movement perpendicular to the electric field lines. When the displacement is parallel there is work done but positive work is cancelled exactly by negative work along some other segment of the path.

When a charge is moved along an open path in the presence of other stationary charges, work is done and hence the energy of the system changes. We call this energy the *electrostatic potential energy* U (units Joules [J]) and it is defined as the work that must be done against the electrostatic field produced by a charge q_1 to bring a charge q_2 from infinity, where the electrostatic field is zero, to a distance r from q_1. An associated quantity, the *electric potential*, Φ, is defined as $\Phi = U/q_2$ (units [J/C] or [V]). As mentioned before, when a charge is moved perpendicular to the electrostatic field there is no work done and therefore the electrostatic potential energy of the system does not change. Consequently, the potential Φ does not change either. Lines and surfaces of the same potential are called *equipotential* lines and surfaces, respectively. It is clear then that the electrostatic field vector **E** is perpendicular to equipotential lines and surfaces at every point. The normal of a surface at any given point can be calculated by taking the gradient of the surface which suggests that the electrostatic field vector can be determined from the potential Φ by taking the gradient too:

$$\mathbf{E} = -\nabla \Phi \, . \tag{1.2.3}$$

Although this might first seem counterintuitive as the electric field has three independent Cartesian components whereas the potential is scalar and so it has only one, it merely underlines the fact that not all electric field vectors describe electrostatic fields and that the Cartesian components of an electrostatic field are not independent of one another.

At this juncture it is worth interjecting a mathematical note. Conservative fields have non-vanishing flux but no circulation, i.e. they are said to be irrotational. Mathematically we characterise flux density by divergence and circulation density by curl. The simultaneous knowledge of the divergence and curl uniquely represents any well-behaved vector field as follows from the fundamental theorem of vector calculus. Irrotational fields have vanishing closed loop integrals

[1] The electric field vector is tangential to electric field lines at all points.

which also means that they can be represented by a *scalar* potential function. In the case of electrostatic fields, this scalar potential function is Φ. Divergenceless fields with non-vanishing circulation can be represented by a *vector* potential function as discussed in Section 1.5.1.

1.2.2 Magnetic Flux, Ampère's Law and Maxwell's Displacement Vector

As children we all experimented with bar magnets learning from experience that they have two poles, somewhat arbitrarily called the north and south pole. Like poles repel each other whilst opposite poles attract each other. When a bar magnet is broken in half, the two halves will each have their own north and south poles, which means that it does not seem to be possible to produce a stand-alone north or south pole. The quantity used to characterise the strength and direction of the magnetic field is the vector $\mathbf{B}$ (also referred to as the 'magnetic induction vector' or 'magnetic flux density'). It is, just as the electric field vector, derived from a measurement of a force; in this case from the force the magnetic field exerts on a moving charge q. Experimental evidence suggests that the force that a moving charge experiences in a homogeneous magnetic field is mutually orthogonal to both the magnetic field and the velocity of the charge and is proportional to q and the magnitude of $\mathbf{v}$ and $\mathbf{B}$: $\mathbf{F}_B = q\,\mathbf{v} \times \mathbf{B}$. The sum of the Coulomb force $\mathbf{F}_E$ and $\mathbf{F}_B$ is called the Lorentz force $\mathbf{F} = \mathbf{F}_E + \mathbf{F}_B = q(\mathbf{E} + \mathbf{v} \times \mathbf{B})$. It is seen that the magnetic field has units of Ns/Cm or Vs/m^2 but, more customarily, in the SI system of units, the unit of $\mathbf{B}$ is Tesla [T], though the older unit of Gauss [G] (1 G = 10^{-4} T) is still used.

If magnetic field lines are visualised by, for example, the sprinkling of iron filings on a paper placed on top of a magnet we find that they emerge from a pole of the magnet. Since poles always come in pairs and magnetic field lines also exist within magnets, it is an experimental fact that magnetic field lines always close on themselves. This should be contrasted with electric field lines which we found start and end on charges. Consequently, since it is only possible to put pairs of magnetic poles inside any closed volume, we can immediately write Gauss' law for the magnetic field as:

$$\oint\!\!\oint \mathbf{B} \cdot d\mathbf{A} = 0 \,, \tag{1.2.4}$$

which equation simply means that *there are no magnetic monopoles. Magnetic field lines are always closed.* It is interesting to point out that the absence of magnetic monopoles has caused considerable discomfort amongst physicists as it leads to an asymmetry of Maxwell's equations as shown later. In 1931 Dirac **[85]** showed that if magnetic monopoles existed it would require all electric charges to be quantised. Therefore, since electric charges are quantised, the existence of magnetic monopoles is fully consistent with Maxwell's equations.

Jean-Baptiste Biot and Félix Savart discovered that there is a magnetic field associated with current carrying wires whose magnitude $\mathbf{B}$ is proportional to the current I (unit Ampère [A]) in the wire and inversely proportional to the distance from the wire. The magnetic field circulates around the wire forming closed loops centred on the wire as shown in Fig. 1.1. The direction of the magnetic field was found to be perpendicular to both the wire and the direction from a point on the wire to the point of observation.

In 1826 André-Marie Ampère showed experimentally on the basis of Biot and Savart's work that the closed loop integral around the wire must be proportional to the current flowing in the wire. By defining current density $\mathbf{J}$ (unit A/m^2) as the current per unit area and assigning a direction to it along the conventional current flow he was able to write

$$\oint \mathbf{B} \cdot d\mathbf{l} = \mu \iint \mathbf{J} \cdot d\mathbf{A} = \mu I \,, \tag{1.2.5}$$

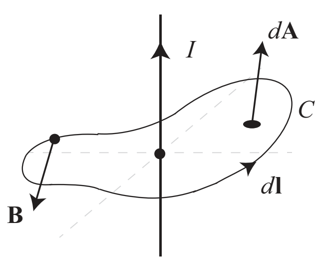

Figure 1.1: The law of Biot and Savart.

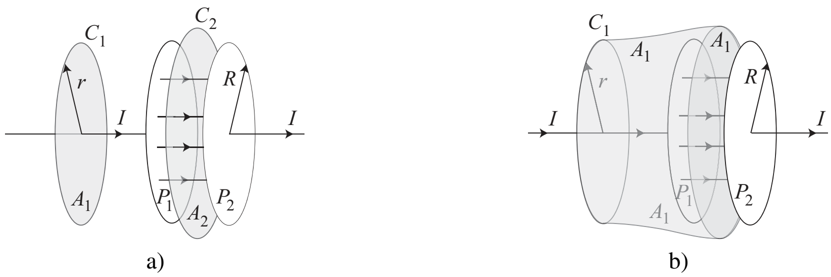

Figure 1.2: Illustrating Ampère's law and its extension by means of Maxwell's displacement field. A time varying current I charges a capacitor. P_1 and P_2 are the two capacitor plates, connected by a wire carrying a time-varying current I.
a) Ampère's law is first applied to the surface A_1 (circular planar integration curve C_1 with radius r) and then to surface A_2 (integration curve C_2).
b) Ampère's law is applied to the modified surface A_1 passing between the capacitor plates P_1 and P_2 and delimited by the circular integration curve C_1 with radius r.
The closing surfaces A_1 and A_2 have been grey-shaded in both drawings.

where μ is a constant of proportionality, called the permeability, the value of which depends on the definition of **B** and **J**, as discussed later. In Fig. 1.1 we show the integration curve C and an infinitesimal element $d\,\mathbf{l}$ of it. The positive direction of $d\,\mathbf{l}$ is connected to the direction of the outward normal of the surface element $d\mathbf{A}$ via the right-hand rule. It is important to note that the flux of the current density through the surface defined by the path along which the line integral is performed on the left-hand side must be taken into account. This surface does not have to be flat so the path does not need to be defined in a plane. This point will be further discussed below.

Maxwell used Ampère's law to calculate the magnetic field around a wire that carries a time varying current density to charge a capacitor as shown in Fig. 1.2 on the left-hand side. By arranging the first loop A_1 such that the so-defined surface is penetrated by the wire, Maxwell calculated the magnitude of the magnetic field at a distance r from the wire to be $B = \mu I/2\pi r$. He then chose the loop A_2 with surface as shown on the left-hand side in Fig. 1.2. Since the current density through the surface is zero, he concluded that the magnetic field between the electrodes must also be zero. Next, he considered the geometry shown on the right-hand side of Fig. 1.2. He again used the loop A_1 but now the associated surface was placed between the electrodes of the capacitor. Because there is no current density passing through the surface he obtained $B = 0$ again. However, this result contradicts that obtained using A_1 on the left-hand side. Therefore Maxwell asked what was so special about the volume between the electrodes of the capacitor. He inferred that, in addition to a current density, the time varying electric flux between the electrodes of the capacitor must also be responsible for generating magnetic fields. Therefore Maxwell inserted a correction term into Eq. (1.2.5) to obtain:

$$\oint \mathbf{B} \cdot d\mathbf{l} = \mu \left(\iint \mathbf{J} \cdot d\mathbf{A} + \epsilon \frac{\partial}{\partial t} \iint \mathbf{E} \cdot d\mathbf{A} \right) = \mu \iint \left(\mathbf{J} + \epsilon \frac{\partial}{\partial t} \mathbf{E} \right) \cdot d\mathbf{A} \,, \tag{1.2.6}$$

which is his extended version of Ampère's law. The line integral is performed over a closed path delimiting an open surface over which the right-hand side flux integral is evaluated. The equation states that *the circulation of magnetic field is due to the flux of a current density through a surface, whose circumference is where the circulation of the magnetic field is measured, and a time varying electric field flux through the same surface*. It is worth noting that the time varying electric field between the electrodes of the capacitor is not a conservative field and thus it is not irrotational. It is sometimes referred to as electrodynamic field.

In 1820 the Danish physicist Hans Christian Ørsted noticed that a compass deviates from its stable position if electric current flows through a wire placed in the vicinity of the compass. This was the first known experiment that connected electricity to magnetism. Michael Faraday, after seeing Ørsted's experiment, suggested that if electric current affects the compass then a magnetic field should produce a current. In order to prove this he set up two solenoids (the so-called Helmholtz coil), as shown in Fig. 1.3a. He then powered the one on the left from a battery and noticed that there was current induced in the solenoid on the right. However, he only experienced current when the switch was being flicked over. Once the switch was on, the current from the other solenoid disappeared. He hence concluded that changing (i.e. not steady) magnetic fields produce current in the other solenoid. The phenomenon is called electromagnetic induction. Heinrich Lenz later experimented to find the direction of the current that is produced by the changing magnetic field. He

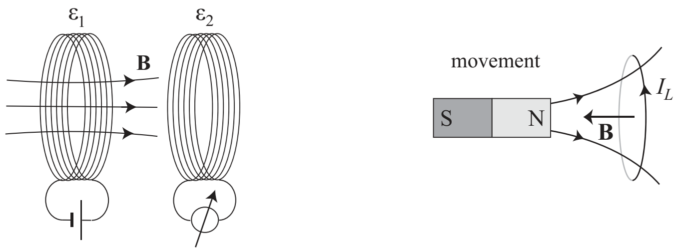

Figure 1.3: a) Faraday's experiment with solenoids.
b) Lenz's law demonstrated with a permanent magnet and a current loop.

found that the induced current in a current loop (shown with arrows in the figure above) is such that its magnetic field (denoted **B** in the figure) opposes the inducing magnetic field (see Fig. 1.3b). This is Lenz's law.

Consider now a wire loop permeated by a magnetic field of increasing magnitude. By Lenz's law, the current generated in the loop will flow such that the magnetic field it induces is opposed to the incoming magnetic field. The current must be produced by a potential difference so there has to be an electric field associated with that potential. The definition of the electromotive force (emf) [V], $\mathcal{E}_{\text{ind}}$, which is the potential in the wire is

$$\oint \mathbf{E} \cdot d\mathbf{l} = \mathcal{E}_{\text{ind}} \,. \tag{1.2.7}$$

The equation states what we mentioned briefly before: electrodynamic fields are not conservative, therefore they have a non-vanishing closed loop integral.

Faraday carried out a number of experiments with Helmholtz coils, as shown Fig. 1.3. He realised that the induced emf in the second coil, $\mathcal{E}_2$ is proportional to the change with time in the magnetic field produced by the first coil and also the area of the second coil. This permitted him to conclude that the quantity of importance is the change with time in the magnetic flux passing through the second coil. The magnetic flux is defined in a way similar to the flux of the electric field:

$$\Phi_B = \iint \mathbf{B} \cdot d\mathbf{A} \,. \tag{1.2.8}$$

Therefore

$$\mathcal{E}_2 = -\frac{\partial \Phi_B}{\partial t} = -\frac{\partial}{\partial t} \iint \mathbf{B} \cdot d\mathbf{A} \,, \tag{1.2.9}$$

or,

$$\oint \mathbf{E} \cdot d\mathbf{l} = -\frac{\partial}{\partial t} \iint \mathbf{B} \cdot d\mathbf{A} \,, \tag{1.2.10}$$

which states that *the induced emf, or circulation in the electrodynamic field, is due to time varying magnetic flux and it opposes that.* This is the fourth Maxwell's equation, Faraday's induction law.

1.2.3 Maxwell's Equations in a Material, Electric and Magnetic Polarisation

Maxwell's equations have been shown to successfully describe electromagnetic fields in vacuum and also in material media. The latter term might refer to a material that does not conduct electric current, often referred to as a dielectric. Electric fields applied to dielectrics will polarise materials. In the absence of an external electric field the atoms in dielectrics have their electron cloud evenly distributed around the nucleus, as shown in Fig. 1.4a. When an electric field **E** is applied in the direction given by Fig. 1.4b, the negative potential on the lower side gives rise to a repulsive force on the electrons and so the electron cloud will be predominantly located towards the more positive potential on the upper side of the drawing, thereby generating electric dipoles.[2] In the case of a capacitor having a dielectric material between

[2] The strength of a dipole is defined as the product of the separating distance $|\mathbf{d}|$ of the two charges with opposite sign and the magnitude q of each charge. The resulting quantity $\mathbf{p}_d = q\mathbf{d}$ is a vector and is called the dipole *moment* of the dipole. The moment vector points from the negative to the positive charge of the dipole. The strength of a dipole is expressed in units of [Cm], the net dipole moment **P** per unit volume in [Cm^{-2}]. A detailed treatment of the electromagnetic properties and the radiation pattern of an individual dipole is given in Subsection 1.6.2 of this chapter.

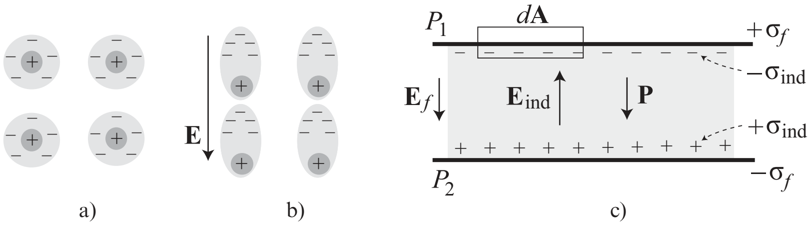

Figure 1.4: A capacitor with conducting surfaces P_1 and P_2 carrying charge densities σ_f. The dielectric material inside the capacitor has been grey-shaded.
a) Charge distribution in the unperturbed dielectric material.
b) Polarisation of the dielectric material inside the capacitor under the influence of an electric field **E**.
c) Electric fields, surface charges and polarisation in the capacitor.

the conducting plates (see Fig. 1.4c), surface charges are induced at the borders of the dielectric material, parallel to the capacitor plates P_1 and P_2. The associated surface charge density is denoted by $\sigma_{\rm ind}$ and gives rise to an induced electric field $\mathbf{E}_{\rm ind}$ in the capacitor. The total electric field **E** is the vector sum of the electric field $\mathbf{E}_f$ in the absence of the dielectric material and the induced electric field with the dielectric material in place. The field $\mathbf{E}_f$ is produced by the surface charges $+\sigma_f$ and $-\sigma_f$ on the capacitor plates P_1 and P_2, respectively. Under the influence of the field $\mathbf{E}_f$, the dielectric material inside the capacitor produces two thin layers with surface charges $-\sigma_{\rm ind}$ and $+\sigma_{\rm ind}$ on the upper and lower surface of the dielectric material, respectively.

The magnitude of the induced surface charges follows from the argument that in the bulk of the dielectric material the positive and negative charges are mutually displaced; their total charge, however, remains zero on average. At the upper and lower border of the dielectric this averaging to zero of the total charge does not happen. Given the direction of the applied field $\mathbf{E}_f$, negative charges are in excess at the upper border of the dielectric material, positive charges at the lower border. The dipole moment per unit volume in the dielectric is given by **P** and is the sum of the moments of N dipoles per m^3. If the dipoles are perfectly aligned through the entire volume, the total dipole moment **P** amounts to $Nq\mathbf{d}$. The vector **P** is commonly called the electric *polarisation*. We assume that it is linearly dependent on the external field provided this is small enough. Under the influence of the electric field in the capacitor each dipole axis **d** points in the downward direction in Fig. 1.4c. The charge movement due to the dipoles leads to an induced (negative) charge dQ on the upper surface of the dielectric which is given by $-N(dA)dq$ where dA is an infinitesimally small surface element on the upper surface of the dielectric. Division by dA yields the induced surface charge density $\sigma_{\rm ind}$ of which the magnitude is thus simply given by $|\mathbf{P}|$. The corresponding electric field **E** is calculated by means of Eq. (1.2.1), applied to the shoebox in Fig. 1.4c with upper and lower surface dA. If the lateral dimensions of the capacitor are much larger than its thickness, the electric fields inside the capacitor are aligned along the vertical direction as shown in Fig. 1.4c and it is permissible to write

$$\mathbf{E} = \mathbf{E}_f + \mathbf{E}_{\rm ind} = \frac{\sigma_f}{\epsilon_0} + \frac{\sigma_{\rm ind}}{\epsilon_0} = \mathbf{E}_f - \frac{\mathbf{P}}{\epsilon_0} = \mathbf{E}_f - \chi\mathbf{E} \, ,$$

or,

$$\mathbf{E}_f = (1+\chi)\mathbf{E} = \epsilon_r \mathbf{E} \, . \qquad (1.2.11)$$

Assuming linearity between the induced polarisation and the external field we have introduced a proportionality factor χ in Eq. (1.2.11) between $\mathbf{E}_{\rm ind}$ and the net field **E** such that $\mathbf{P} = \epsilon_0\chi\mathbf{E}$. The dimensionless quantity χ is called the *electric susceptibility* of the dielectric material. The equally dimensionless quantity ϵ_r is called the *relative permittivity* of the material medium and ϵ_0 is called the permittivity of vacuum, though it is only a constant of proportionality depending on the system of units. In the SI system, $\epsilon_0 = 8.854 \times 10^{-12}$ Fm^{-1}. It is also not unusual to use the *displacement field* or electric flux density **D** (unit [FV/m^2], [As/m^2] or [C/m^2]) defined formally as

$$\mathbf{D} = \epsilon\mathbf{E} = \epsilon_0\epsilon_r\mathbf{E} = \epsilon_0(1+\chi)\mathbf{E} = \epsilon_0\mathbf{E} + \epsilon_0\chi\mathbf{E} = \epsilon_0\mathbf{E} + \mathbf{P} \, . \qquad (1.2.12)$$

We note that the displacement field is not a fundamental field, meaning that it relates to a force measurement only via **E**. Note that this argument implicitly assumes that the dielectric material is linear and isotropic, meaning that the material is invariant to all rotational transformations. There are cases when the induced electric field vector is not antiparallel with the electric field inducing it. In this case χ, and consequently ϵ_r, becomes a tensor as discussed in Chapter 2. Apart from the global susceptibility χ of a material, we can also define the *polarisability* α of an elementary particle in the material, e.g. an atom or a molecule. The individual dipole moment $\mathbf{p}$ of such a particle, induced by the field $\mathbf{E}_{\text{ind}}$, equals $\mathbf{P}/N$ where N is the number of particles per unit volume. The induced dipole moment $\mathbf{p}$ of a single particle is defined as $\alpha\,\mathbf{E}$. It then follows that the polarisability is $\alpha = \epsilon_0\chi/N$ with unit [Cm^2V^{-1}].

In a similar fashion, magnetic materials also contain magnetic dipoles due to electron currents. Depending on the type of magnetic material, when an external magnetic field, usually denoted by $\mathbf{H}$ (unit [A/m]), is applied these magnetic dipoles can orient themselves to alter the effect of the inducing magnetic field. The induced magnetic field, denoted by $\mathbf{M}$, is called *magnetisation* or *magnetic polarisation*. The magnetic field $\mathbf{H}$ and the magnetic polarisation $\mathbf{M}$ together are responsible for the overall magnetic field:

$$\mathbf{B} = \mu_0(\mathbf{H} + \mathbf{M}) \,. \tag{1.2.13}$$

In diamagnetic and paramagnetic materials, the magnetisation is proportional to $\mathbf{H}$ with as constant of proportionality the *magnetic susceptibility*, χ_m, yielding $\mathbf{M} = \chi_m\mathbf{H}$ and Eq. (1.2.11) then reads

$$\mathbf{B} = \mu_0(\mathbf{H} + \chi_m\mathbf{H}) = \mu_0(1 + \chi_m)\mathbf{H} = \mu_0\mu_r\mathbf{H} = \mu\mathbf{H} \,, \tag{1.2.14}$$

where μ_r is the relative permeability (dimensionless). However, since the optical materials we are concerned with are not magnetically active in most cases, we shall restrict our discussions to $\mu_r = 1$.

1.3 Maxwell's Equations in the Optical Domain

As discussed in the previous section, the general laws governing electromagnetic phenomena are:
Coulomb's law or Gauss' law for electrostatics

$$\oiint \mathbf{D}\cdot d\mathbf{A} = \iiint \rho\, dV \,, \tag{1.3.1}$$

Gauss' law for magnetic fields

$$\oiint \mathbf{B}\cdot d\mathbf{A} = 0 \,, \tag{1.3.2}$$

Ampère–Maxwell law

$$\oint \frac{\mathbf{B}}{\mu}\cdot d\mathbf{l} = \iint \left(\mathbf{J} + \frac{\partial \mathbf{D}}{\partial t}\right)\cdot d\mathbf{A} \,, \tag{1.3.3}$$

Faraday's induction law

$$\oint \mathbf{E}\cdot d\mathbf{l} = -\frac{\partial}{\partial t}\iint \mathbf{B}\cdot d\mathbf{A} \,. \tag{1.3.4}$$

In the above integrals the inner products of vector quantities and line segments or surface elements imply the evaluation of the scalar product where the line segment $d\mathbf{l}$ points in the tangential direction and the surface element vector $d\mathbf{A}$ points in the direction of the outward normal to the surface. In Eqs. (1.3.1) and (1.3.2) the volume integral is over an arbitrary volume V that is bounded by a closed surface A over which the surface integral is evaluated. In Eqs. (1.3.3) and (1.3.4), the surface integral applies to an open surface A that is bounded by a curve l along which the line integral has to be carried out (see also Fig. 1.5 for the geometrical details). In the equations above, we consider $\mathbf{E}$, the electric field and $\mathbf{B}$, the magnetic induction, as the two basic quantities that determine the electromagnetic field.

The other medium-determined quantities occurring in Eqs. (1.3.1)–(1.3.4) are the current density $\mathbf{J}$, the scalar quantity ρ, the volume charge density (unit [C/m^3]) and the dielectric displacement or electric flux density (electric induction) $\mathbf{D}$. With the aid of the electric permittivity $\epsilon = \epsilon_0\epsilon_r$, the magnetic permeability $\mu = \mu_r\mu_0$ and the specific conductivity σ, we define the following relationships between the basic field vectors $\mathbf{E}$ and $\mathbf{B}$ and the other vector quantities $\mathbf{D}$, $\mathbf{H}$ and $\mathbf{J}$ via the so-called material equations or *constitutive relations*:

$$\mathbf{D} = \epsilon\,\mathbf{E} \,, \tag{1.3.5}$$

$$\mathbf{B} = \mu\,\mathbf{H} \,, \tag{1.3.6}$$

$$\mathbf{J} = \sigma\,\mathbf{E} \,, \tag{1.3.7}$$

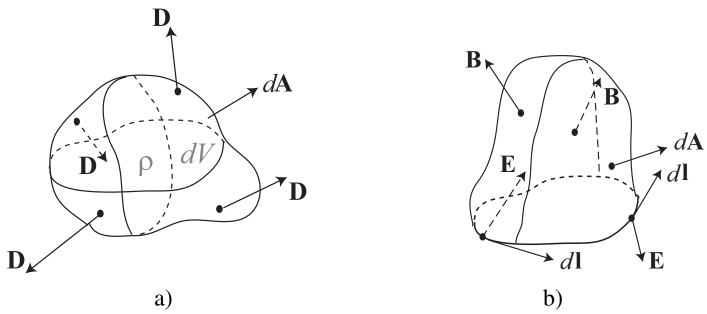

Figure 1.5: a) Illustrating the volume and closed surface integrals involved in the Coulomb/Gauss laws. The vector quantity $d\mathbf{A}$ is normal to the closed surface A. The vector $\mathbf{D}$ represents the electric displacement field vector (or electric flux density vector), ρ is the electric charge density and dV is an infinitesimal volume element.
b) The open surface integral and the line integral which appear in the Ampère and Faraday laws. The vector $d\mathbf{l}$ is tangent to the limiting curve of the open surface A. The vector $\mathbf{B}$ represents the magnetic flux density and $\mathbf{E}$ the electric field strength vector.

as follows from Eqs. (1.2.12) and (1.2.14). In homogeneous[3] and isotropic media, ϵ, μ and σ are constants. Moreover, in the optical domain, the conductivity of most dielectric materials of interest will be low or close to zero to guarantee high transmission through the medium; in many cases, it can be conveniently set to zero. In inhomogeneous isotropic media, ϵ, μ and σ are scalar functions of the position vector $\mathbf{r}$. In anisotropic media, these quantities become tensors. In the optical domain, we are generally allowed to equate the magnetic permeability of a medium to that of vacuum, μ_0. Recent developments in material engineering (metamaterials) show that this is not always necessarily the case. Some aspects of these recent material developments, like the possibility of 'perfect imaging', are treated in Chapter 3.

In order to transform the integral version of Maxwell's equations into their alternative differential form we apply the Gauss and Stokes vector integral theorems,

$$\iiint \nabla \cdot \mathbf{v}\, dV = \oiint \mathbf{v} \cdot d\mathbf{A}\,, \tag{1.3.8}$$

$$\oint \mathbf{v} \cdot d\mathbf{l} = \iint (\nabla \times \mathbf{v}) \cdot d\mathbf{A}\,, \tag{1.3.9}$$

where $\mathbf{v}$ is a general vector field, to Maxwell's equations above. On comparing Eqs. (1.3.1)–(1.3.4) with Eqs. (1.3.8) and (1.3.9), we obtain Maxwell's equations in *differential form*,

$$\nabla \cdot \mathbf{D} = \rho\,, \tag{1.3.10}$$

$$\nabla \cdot \mathbf{B} = 0\,, \tag{1.3.11}$$

$$\nabla \times \frac{\mathbf{B}}{\mu} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t}\,, \tag{1.3.12}$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}\,. \tag{1.3.13}$$

Note that there is a significant difference between Maxwell's equations expressed in integral and differential form. While the former are applicable to volumes of space, the latter are only applicable to points, with curl ($\nabla\times$) denoting *circulation density* and divergence ($\nabla\cdot$) denoting *flux density*. As we have seen in the introduction, it is more usual and perhaps sensible to use currents and charges with Maxwell's equations in their integral form but charge and current densities with the differential form. As noted above, because a vector function is uniquely characterised by the simultaneous knowledge of the circulation density (curl) and the divergence of the function, so the differential form of Maxwell's equations uniquely specifies both the electric and the magnetic field at a given position in space.

1.4 Electromagnetic Energy Density and Energy Transport

Electromagnetic fields create an energy density in space and can give rise to a flow of energy. In this section we establish the electric and magnetic energy density and the energy flow created by electromagnetic waves, expressed in terms of the

[3] By homogeneous we mean that ϵ and μ are not position-dependent.