

Chapter 1

Foundations of Scalar Diffraction Theory

Light can be described by two very different approaches: classical electrodynamics and quantum electrodynamics. In the classical treatment, electric and magnetic fields are continuous functions of space and time, and light comprises co-oscillating electric and magnetic wave fields. In the quantum treatment, photons are elementary particles with no mass nor charge, and light comprises one or more photons. There is rigorous theory behind each approach, and there is experimental evidence supporting both. Neither approach can be dismissed, which leads to the wave-particle duality of light. Generally, classical methods are used for macroscopic properties of light, while quantum methods are used for submicroscopic properties of light.

This book describes macroscopic properties, so it deals entirely with classical electrodynamics. When the wavelength λ of an electromagnetic wave is very small, approaching zero, the waves travel in straight lines with no bending around the edges of objects. That is realm of geometric optics. However, this book treats many situations in which geometric optics are inadequate to describe observed phenomena like diffraction. Therefore, the starting point is classical electrodynamics with solutions provided by scalar diffraction theory. Geometric optics is treated briefly in Sec. 6.5.

1.1 Basics of Classical Electrodynamics

Classical electrodynamics deals with relationships between electric fields, magnetic fields, static charge, and moving charge (i.e., current) in space and time based on the macroscopic properties of the materials in which the fields exist. We define each quantity here along with some basic relationships. This introduces the reader to the quantities in Maxwell's equations, which describe how electrically charged particles and objects give rise to electric and magnetic fields. Maxwell's equations are introduced here in their most general form, and then the discussion focuses on a specific case and solutions for oscillating electric and magnetic fields, which light comprises.

1.1.1 Sources of electric and magnetic fields

Electric charge, measured in coulombs, is a fundamental property of elementary particles and bulk materials. Classically, charge may be positive, negative, or zero. Further, charge is quantized, specifically the smallest possible nonzero amount of charge is the *elementary charge* $e = 1.602 \times 10^{-19}$ C. All nonzero amounts of charge are integer multiples of e . For bulk materials, the integer may be very large so that total charge can be treated as continuous rather than discrete. We denote the volume density of free charge, measured in coulombs per cubic meter, by $\rho(\mathbf{r}, t)$, where \mathbf{r} is a three-dimensional spatial vector, and t is time. Moving charge density is called free volume current density $\mathbf{J}(\mathbf{r}, t)$. Volume current density is measured in Ampères per square meter ($1 \text{ A} = 1 \text{ C/s}$). This represents the time rate at which charge passes through a surface of unit area. Finally, charge is conserved, meaning that the total charge of any system is constant. This is mathematically stated by the continuity equation

$$\nabla \cdot \mathbf{J}(\mathbf{r}, t) + \frac{\partial \rho(\mathbf{r}, t)}{\partial t} = 0. \quad (1.1)$$

Almost every material we encounter in life is composed of many, many atoms each with many positive and negative charges. Usually, the numbers of positive and negative charges are equal or nearly equal so that the whole material is electrically neutral. Still, such a material can give rise to electric or magnetic fields when the total charge and free current are zero. If the distribution of charge is not homogeneous or if the charges are circulating in tiny current loops, fields could be present.

The separation of charge is described by the electric dipole moment, which is the amount of separated charge times the separation distance. If a bulk material has its charge arranged in many tiny dipoles, it is said to be electrically polarized. The volume polarization density $\mathbf{P}(\mathbf{r}, t)$ is the density of electric dipole moments per unit volume, measured in coulombs per square meter.

Magnetization is a similar concept for moving charge. Charge circulating in a tiny current loop is described by magnetic dipole moment, which is the circulating current times the area of the loop. When a bulk material has internal current arranged in many tiny loops, it is said to be magnetized. The volume magnetization density $\mathbf{M}(\mathbf{r}, t)$ is the density of magnetic dipole moments per unit volume, measured in Ampères per meter.

1.1.2 Electric and magnetic fields

When a hypothetical charge, called a test charge, passes near a bulk material that has non-zero ρ , \mathbf{J} , \mathbf{P} , or \mathbf{M} , the charge experiences a force. This interaction is characterized by two vectors \mathbf{E} and \mathbf{B} . The electromagnetic force \mathbf{F} on a test particle at a given point and time is a function of these vector fields and the particle's charge q and velocity \mathbf{v} . The Lorentz force law describes this interaction as

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}). \quad (1.2)$$

If this empirical statement is valid (and, of course, countless experiments over the course of centuries have shown that it is), then two vector fields \mathbf{E} and \mathbf{B} are thereby defined throughout space and time, and these are called the “electric field” and “magnetic induction.”¹

Eq. (1.2) can be examined in a little more detail to provide more intuitive definitions of these fields. The electric field is the amount of force per unit of test charge when the test charge is stationary, given by

$$\mathbf{E} = \lim_{q \rightarrow 0^+} \frac{\mathbf{F}}{q} \Big|_{\mathbf{v}=0}. \quad (1.3)$$

This is called a push-and-pull force because the force is in either the same or opposite direction as the field, depending on the sign of the charge. Electric field is measured in units of volts per meter (1 V = 1 N m/C). The magnetic field is related to the amount of force per unit test charge given by

$$\mathbf{v} \times \mathbf{B} = \lim_{q \rightarrow 0^+} \frac{\mathbf{F} - q\mathbf{E}}{q} \Big|_{\mathbf{v} \neq 0}. \quad (1.4)$$

The force due to a magnetic field is called deflective because it is perpendicular to the particle’s velocity, which deflects its trajectory. Magnetic field is measured in units of Tesla [1 T = 1 N s/(C m)].

With this understanding of the fields, they now need to be related to the sources. This was accomplished through centuries of experimental measurements and theoretical and intuitive insight, resulting in

$$\nabla \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = \mathbf{0} \quad (1.5)$$

$$\nabla \times \mathbf{B} - \mu_0 \epsilon_0 \frac{\partial \mathbf{E}}{\partial t} = \mu_0 \left(\mathbf{J} + \frac{\partial \mathbf{P}}{\partial t} + \nabla \times \mathbf{M} \right). \quad (1.6)$$

These are two of Maxwell’s equations, the former being Faraday’s law and the latter being Ampère’s law with Maxwell’s correction. In Eq. (1.6), the sources on the right hand side include the free current \mathbf{J} and two terms due to bound currents. These are the polarization current $\partial \mathbf{P} / \partial t$ and the magnetization current $\nabla \times \mathbf{M}$.

These equations can be written in a more functionally useful form. Eq. (1.6) can be rewritten as

$$\nabla \times \left(\frac{\mathbf{B}}{\mu_0} - \mathbf{M} \right) = \mathbf{J} + \frac{\partial}{\partial t} (\epsilon_0 \mathbf{E} + \mathbf{P}). \quad (1.7)$$

Making the definitions

$$\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P} \quad (1.8)$$

$$\mathbf{H} = \frac{\mathbf{B}}{\mu_0} - \mathbf{M} \quad (1.9)$$

introduces the concepts of electric displacement \mathbf{D} and magnetic field \mathbf{H} , which are fields that account for the medium's response to the applied fields. Now the working form of these Maxwell equations becomes

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \quad (1.10)$$

$$\nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t}. \quad (1.11)$$

Further, when these are combined with conservation of charge expressed in Eq. (1.1), this leads to

$$\nabla \cdot \nabla \times \mathbf{H} = \nabla \cdot \mathbf{J} + \frac{\partial}{\partial t} \nabla \cdot \mathbf{D} \quad (1.12)$$

$$= -\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial t} \nabla \cdot \mathbf{D} \quad (1.13)$$

$$= 0. \quad (1.14)$$

Focusing on the right-hand side,

$$\frac{\partial}{\partial t} (\nabla \cdot \mathbf{D} - \rho) = 0 \quad (1.15)$$

$$\nabla \cdot \mathbf{D} - \rho = f(\mathbf{r}), \quad (1.16)$$

where $f(\mathbf{r})$ is an unspecified function of space but not time. Causality requires that $f(\mathbf{r}) = 0$ before the source is turned on, yielding Coulomb's law:

$$\nabla \cdot \mathbf{D} = \rho. \quad (1.17)$$

Similar manipulations yield

$$\nabla \cdot \mathbf{B} = 0. \quad (1.18)$$

This indicates that magnetic monopole charges do not exist. Finally, Eqs. (1.10), (1.11), (1.17), and (1.18) constitute Maxwell's equations.¹

In this model of macroscopic electrodynamics, Eqs. (1.10) and (1.11) are two independent vector equations. With three scalar components each, these are six independent scalar equations. Unfortunately, given knowledge of the sources, there are four unknown vector fields \mathbf{D} , \mathbf{B} , \mathbf{H} , and \mathbf{E} . Each has three scalar components for a total of twelve unknown scalars. With so many more unknown field components than equations, this is a poorly posed problem.

The key is to understand the medium in which the fields exist. This produces a means of relating \mathbf{P} to \mathbf{E} and \mathbf{M} to \mathbf{H} , which amount to six more scalar equations. For example, in simple media (linear, homogeneous, and isotropic),

$$\mathbf{P} = \epsilon_0 \chi_e \mathbf{E} \quad (1.19)$$

$$\mathbf{M} = \chi_m \mathbf{H}, \quad (1.20)$$

where χ_e is the electric susceptibility of the medium and χ_m is its magnetic susceptibility. Substituting these into Eqs. (1.8) and (1.9) yields

$$\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P} \quad (1.21)$$

$$= \epsilon_0 (1 + \chi_m) \mathbf{E} \quad (1.22)$$

$$= \epsilon \mathbf{E} \quad (1.23)$$

and

$$\mathbf{B} = \mu_0 (\mathbf{H} + \mathbf{M}) \quad (1.24)$$

$$= \mu_0 (1 + \chi_m) \mathbf{H} \quad (1.25)$$

$$= \mu \mathbf{H}, \quad (1.26)$$

where $\epsilon = (1 + \chi_e) \epsilon_0$ is the electric permittivity and $\mu = (1 + \chi_m) \mu_0$ is the magnetic permeability of the medium. Now this simplifies Eqs. (1.10) and (1.11) so that

$$\nabla \times \mathbf{E} = -\mu \frac{\partial \mathbf{H}}{\partial t} \quad (1.27)$$

$$\nabla \times \mathbf{H} = \mathbf{J} + \epsilon \frac{\partial \mathbf{E}}{\partial t}. \quad (1.28)$$

Now there are still six equations but only six unknowns (as long as the free current density \mathbf{J} is known). Finally, with a proper understanding of the materials, this is a well posed problem.

1.2 Simple Traveling-Wave Solutions to Maxwell's Equations

There are many solutions to Maxwell's equations, but there are only a few that can be written in closed form without an integral. This section begins with transforming Maxwell's four equations into two uncoupled wave equations. It continues with a few specific simple solutions such as the infinite-extent plane wave. A more general solution is left to the next section.

1.2.1 Obtaining a wave equation

This book deals with optical wave propagation through linear, isotropic, homogeneous, nondispersive, dielectric media in the absence of source charges and currents. In this case, the media discussed throughout the remainder of this book have

$$\epsilon = \text{a scalar, independent of } \lambda, \mathbf{r}, t \quad (1.29)$$

$$\mu = \mu_0 \quad (1.30)$$

$$\rho = 0 \quad (1.31)$$

$$\mathbf{J} = \mathbf{0}. \quad (1.32)$$

Taking the curl of Eq. (1.27) yields

$$\nabla \times (\nabla \times \mathbf{E}) = -\mu_0 \frac{\partial}{\partial t} (\nabla \times \mathbf{H}). \quad (1.33)$$

Then, substituting in Eq. (1.28) gives

$$\nabla \times (\nabla \times \mathbf{E}) = -\mu_0 \epsilon \frac{\partial^2}{\partial t^2} \mathbf{E}. \quad (1.34)$$

Now, applying the vector identity $\nabla \times (\nabla \times \mathbf{E}) = \nabla (\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E}$ leads to

$$\nabla (\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E} = -\mu_0 \epsilon \frac{\partial^2}{\partial t^2} \mathbf{E}. \quad (1.35)$$

Finally, substituting in Eqs. (1.17) and (1.23), and keeping in mind that ϵ is independent of position results in a wave differential equation:

$$\nabla^2 \mathbf{E} - \mu_0 \epsilon \frac{\partial^2}{\partial t^2} \mathbf{E} = \mathbf{0}. \quad (1.36)$$

Similar manipulations beginning with the curl of Eq. (1.28) yield

$$\nabla^2 \mathbf{B} - \mu_0 \epsilon \frac{\partial^2}{\partial t^2} \mathbf{B} = \mathbf{0}. \quad (1.37)$$

When the Laplacian is used on the Cartesian components of \mathbf{E} and \mathbf{B} , the result is six uncoupled but identical equations of the form

$$\left(\nabla^2 - \mu_0 \epsilon \frac{\partial^2}{\partial t^2} \right) U(x, y, z) = 0, \quad (1.38)$$

where the scalar $U(x, y, z)$ stands for any of the x -, y - or z - directed components of the vector fields \mathbf{E} and \mathbf{B} .

At this point, we can define index of refraction

$$n = \sqrt{\frac{\epsilon}{\epsilon_0}} \quad (1.39)$$

and the vacuum speed of light

$$c = \frac{1}{\sqrt{\mu_0 \epsilon_0}} \quad (1.40)$$

so that

$$\left(\nabla^2 - \frac{n^2}{c^2} \frac{\partial^2}{\partial t^2} \right) U(x, y, z) = 0. \quad (1.41)$$

The electric and magnetic fields that compose light are traveling wave fields. Therefore, fields with harmonic time dependence $\exp(-i2\pi\nu t)$ (where ν is the wave