

Classical Electromagnetism

Richard Fitzpatrick

Professor of Physics

The University of Texas at Austin

Contents

1	Maxwell's Equations	7
1.1	Introduction	7
1.2	Maxwell's Equations	7
1.3	Scalar and Vector Potentials	8
1.4	Dirac Delta Function	9
1.5	Three-Dimensional Dirac Delta Function	9
1.6	Solution of Inhomogeneous Wave Equation	10
1.7	Retarded Potentials	16
1.8	Retarded Fields	17
1.9	Electromagnetic Energy Conservation	19
1.10	Electromagnetic Momentum Conservation	20
1.11	Exercises	22
2	Electrostatic Fields	25
2.1	Introduction	25
2.2	Laplace's Equation	25
2.3	Poisson's Equation	26
2.4	Coulomb's Law	27
2.5	Electric Scalar Potential	28
2.6	Electrostatic Energy	29
2.7	Electric Dipoles	33
2.8	Charge Sheets and Dipole Sheets	34
2.9	Green's Theorem	37
2.10	Boundary Value Problems	40
2.11	Dirichlet Green's Function for Spherical Surface	43
2.12	Exercises	46
3	Potential Theory	49
3.1	Introduction	49

3.2	Associated Legendre Functions	49
3.3	Spherical Harmonics	50
3.4	Laplace's Equation in Spherical Coordinates	51
3.5	Poisson's Equation in Spherical Coordinates	52
3.6	Multipole Expansion	53
3.7	Axisymmetric Charge Distributions	56
3.8	Dirichlet Problem in Spherical Coordinates	57
3.9	Newmann Problem in Spherical Coordinates	58
3.10	Laplace's Equation in Cylindrical Coordinates	59
3.11	Poisson's Equation in Cylindrical Coordinates	64
3.12	Exercises	69
4	Electrostatics in Dielectric Media	73
4.1	Polarization	73
4.2	Boundary Conditions for \mathbf{E} and \mathbf{D}	75
4.3	Boundary Value Problems with Dielectrics	75
4.4	Energy Density Within Dielectric Medium	81
4.5	Force Density Within Dielectric Medium	82
4.6	Clausius-Mossotti Relation	85
4.7	Dielectric Liquids in Electrostatic Fields	86
4.8	Exercises	89
5	Magnetostatic Fields	93
5.1	Introduction	93
5.2	Biot-Savart Law	93
5.3	Continuous Current Distribution	93
5.4	Circular Current Loop	95
5.5	Localized Current Distribution	98
5.6	Exercises	100
6	Magnetostatics in Magnetic Media	103
6.1	Magnetization	103
6.2	Magnetic Susceptibility and Permeability	104
6.3	Ferromagnetism	105
6.4	Boundary Conditions for \mathbf{B} and \mathbf{H}	107
6.5	Permanent Ferromagnets	108
6.6	Uniformly Magnetized Sphere	110
6.7	Soft Iron Sphere in Uniform Magnetic Field	113
6.8	Magnetic Shielding	114
6.9	Magnetic Energy	116
6.10	Exercises	117
7	Wave Propagation in Uniform Dielectric Media	119

7.1	Introduction	119
7.2	Form of Dielectric Constant	120
7.3	Anomalous Dispersion and Resonant Absorption	122
7.4	Wave Propagation in Conducting Media	124
7.5	High Frequency Limit	126
7.6	Polarization of Electromagnetic Waves	127
7.7	Faraday Rotation	128
7.8	Wave Propagation in Magnetized Plasmas	131
7.9	Wave Propagation in Dispersive Media	132
7.10	Wave-Front Propagation	136
7.11	Sommerfeld Precursor	139
7.12	Method of Stationary Phase	144
7.13	Group Velocity	145
7.14	Brillouin Precursor	147
7.15	Signal Arrival	149
7.16	Exercises	150
8	Wave Propagation in Inhomogeneous Dielectric Media	153
8.1	Introduction	153
8.2	Laws of Geometric Optics	153
8.3	Fresnel Relations	155
8.4	Total Internal Reflection	161
8.5	Reflection by Conducting Surfaces	166
8.6	Ionospheric Radio Wave Propagation	167
8.7	WKB Approximation	168
8.8	Reflection Coefficient	172
8.9	Extension to Oblique Incidence	173
8.10	Ionospheric Pulse Propagation	177
8.11	Measurement of Ionospheric Electron Density Profile	179
8.12	Ionospheric Ray Tracing	180
8.13	Asymptotic Series	183
8.14	WKB Solution as Asymptotic Series	188
8.15	Stokes Constants	189
8.16	WKB Reflection Coefficient	194
8.17	Jeffries Connection Formula	197
8.18	Exercises	198
9	Radiation and Scattering	201
9.1	Introduction	201
9.2	Basic Antenna Theory	201
9.3	Antenna Directivity and Effective Area	204
9.4	Antenna Arrays	208
9.5	Thomson Scattering	210

9.6	Rayleigh Scattering	212
9.7	Exercises	213
10	Resonant Cavities and Waveguides	215
10.1	Introduction	215
10.2	Boundary Conditions	215
10.3	Cavities with Rectangular Boundaries	217
10.4	Quality Factor of a Resonant Cavity	218
10.5	Axially Symmetric Cavities	219
10.6	Cylindrical Cavities	222
10.7	Waveguides	223
10.8	Dielectric Waveguides	225
10.9	Exercises	230
11	Multipole Expansion	231
11.1	Introduction	231
11.2	Multipole Expansion of Scalar Wave Equation	231
11.3	Angular Momentum Operators	233
11.4	Multipole Expansion of Vector Wave Equation	234
11.5	Properties of Multipole Fields	237
11.6	Solution of Inhomogeneous Helmholtz Equation	240
11.7	Sources of Multipole Radiation	242
11.8	Radiation from Linear Centre-Fed Antenna	245
11.9	Spherical Wave Expansion of Vector Plane Wave	248
11.10	Mie Scattering	250
11.11	Exercises	254
12	Relativity and Electromagnetism	255
12.1	Introduction	255
12.2	Relativity Principle	255
12.3	Lorentz Transformation	256
12.4	Transformation of Velocities	260
12.5	Tensors	261
12.6	Physical Significance of Tensors	265
12.7	Space-Time	266
12.8	Proper Time	270
12.9	4-Velocity and 4-Acceleration	271
12.10	Current Density 4-Vector	272
12.11	Potential 4-Vector	273
12.12	Gauge Invariance	274
12.13	Retarded Potentials	274
12.14	Tensors and Pseudo-Tensors	276
12.15	Electromagnetic Field Tensor	278

12.16	Dual Electromagnetic Field Tensor	281
12.17	Transformation of Fields	283
12.18	Potential Due to a Moving Charge	284
12.19	Field Due to a Moving Charge	285
12.20	Relativistic Particle Dynamics	286
12.21	Force on a Moving Charge	288
12.22	Electromagnetic Energy Tensor	289
12.23	Accelerated Charges	292
12.24	Larmor Formula	295
12.25	Radiation Losses	299
12.26	Angular Distribution of Radiation	300
12.27	Synchrotron Radiation	301
12.28	Exercises	304

1 Maxwell's Equations

1.1 Introduction

This chapter gives a general overview of Maxwell's equations.

1.2 Maxwell's Equations

All classical (i.e., non-quantum) electromagnetic phenomena are governed by *Maxwell's equations*, which take the form

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0}, \quad (1.1)$$

$$\nabla \cdot \mathbf{B} = 0, \quad (1.2)$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \quad (1.3)$$

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{j} + \mu_0 \epsilon_0 \frac{\partial \mathbf{E}}{\partial t}. \quad (1.4)$$

Here, $\mathbf{E}(\mathbf{r}, t)$, $\mathbf{B}(\mathbf{r}, t)$, $\rho(\mathbf{r}, t)$, and $\mathbf{j}(\mathbf{r}, t)$ represent the *electric field-strength*, the *magnetic field-strength*, the *electric charge density*, and the *electric current density*, respectively. Moreover,

$$\epsilon_0 = 8.8542 \times 10^{-12} \text{ C}^2 \text{ N}^{-1} \text{ m}^{-2} \quad (1.5)$$

is the *electric permittivity of free space*, whereas

$$\mu_0 = 4\pi \times 10^{-7} \text{ N A}^{-2} \quad (1.6)$$

is the *magnetic permeability of free space*. As is well known, Equation (1.1) is equivalent to *Coulomb's law* (for the electric fields generated by point charges), Equation (1.2) is equivalent to the statement that magnetic monopoles do not exist (which implies that magnetic field-lines can never begin or end), Equation (1.3) is equivalent to *Faraday's law of electromagnetic induction*, and Equation (1.4) is equivalent to the *Biot-Savart law* (for the magnetic fields generated by line currents) augmented by the induction of magnetic fields by changing electric fields.

Maxwell's equations are linear in nature. In other words, if $\rho \rightarrow \alpha \rho$ and $\mathbf{j} \rightarrow \alpha \mathbf{j}$, where α is an arbitrary (spatial and temporal) constant, then it is clear from Equations (1.1)–(1.4) that $\mathbf{E} \rightarrow \alpha \mathbf{E}$ and $\mathbf{B} \rightarrow \alpha \mathbf{B}$. The linearity of Maxwell's equations accounts for the well-known fact that the electric fields generated by point charges, as well as the magnetic fields generated by line currents, are superposable.

Taking the divergence of Equation (1.4), and combining the resulting expression with Equation (1.1), we obtain

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{j} = 0. \quad (1.7)$$

In integral form, making use of the divergence theorem, this equation becomes

$$\frac{d}{dt} \int_V \rho dV + \int_S \mathbf{j} \cdot d\mathbf{S} = 0, \quad (1.8)$$

where V is a fixed volume bounded by a surface S . The volume integral represents the net electric charge contained within the volume, whereas the surface integral represents the outward flux of charge across the bounding surface. The previous equation, which states that the net rate of change of the charge contained within the volume V is equal to minus the net flux of charge across the bounding surface S , is clearly a statement of the *conservation of electric charge*. Thus, Equation (1.7) is the differential form of this conservation equation.

As is well known, a point electric charge q moving with velocity \mathbf{v} in the presence of an electric field \mathbf{E} and a magnetic field \mathbf{B} experiences a force

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}). \quad (1.9)$$

Likewise, a distributed charge distribution of charge density ρ and current density \mathbf{j} experiences a force density

$$\mathbf{f} = \rho \mathbf{E} + \mathbf{j} \times \mathbf{B}. \quad (1.10)$$

1.3 Scalar and Vector Potentials

We can automatically satisfy Equation (1.2) by writing

$$\mathbf{B} = \nabla \times \mathbf{A}, \quad (1.11)$$

where $\mathbf{A}(\mathbf{r}, t)$ is termed the *vector potential*. Furthermore, we can automatically satisfy Equation (1.3) by writing

$$\mathbf{E} = -\nabla\phi - \frac{\partial\mathbf{A}}{\partial t}, \quad (1.12)$$

where $\phi(\mathbf{r}, t)$ is termed the *scalar potential*.

The previous prescription for expressing electric and magnetic fields in terms of the scalar and vector potentials does not uniquely define the potentials. Indeed, it can be seen that if $\mathbf{A} \rightarrow \mathbf{A} - \nabla\psi$ and $\phi \rightarrow \phi + \partial\psi/\partial t$, where $\psi(\mathbf{r}, t)$ is an arbitrary scalar field, then the associated electric and magnetic fields are unaffected. The root of the problem lies in the fact that Equation (1.11) specifies the curl of the vector potential, but leaves the divergence of this vector field completely unspecified. We can make our prescription unique by adopting a convention that specifies the divergence of the vector potential—such a convention is usually called a *gauge condition*. It turns out that Maxwell's equations are Lorentz invariant. (See Chapter 12.) In other words, they take the same form in all inertial frames. Thus, it makes sense to adopt a gauge condition that is also Lorentz invariant. This leads us to the so-called *Lorenz gauge condition* (see Section 12.12),

$$\epsilon_0 \mu_0 \frac{\partial\phi}{\partial t} + \nabla \cdot \mathbf{A} = 0. \quad (1.13)$$

Equations (1.11)–(1.13) can be combined with Equations (1.1) and (1.4) to give

$$\frac{1}{c^2} \frac{\partial^2 \phi}{\partial t^2} - \nabla^2 \phi = \frac{\rho}{\epsilon_0}, \quad (1.14)$$

$$\frac{1}{c^2} \frac{\partial^2 \mathbf{A}}{\partial t^2} - \nabla^2 \mathbf{A} = \mu_0 \mathbf{j}, \quad (1.15)$$

where

$$c = \frac{1}{\sqrt{\epsilon_0 \mu_0}} = 2.988 \times 10^8 \text{ m s}^{-1} \quad (1.16)$$

is the *velocity of light in vacuum*. Thus, Maxwell's equations essentially boil down to Equations (1.14) and (1.15).

1.4 Dirac Delta Function

The *Dirac delta function*, $\delta(t - t')$, has the property

$$\delta(t - t') = 0 \quad \text{for } t \neq t'. \quad (1.17)$$

In addition, however, the function is singular at $t = t'$ in such a manner that

$$\int_{-\infty}^{\infty} \delta(t - t') dt' = 1. \quad (1.18)$$

It follows that

$$\int_{-\infty}^{\infty} f(t) \delta(t - t') dt' = f(t), \quad (1.19)$$

where $f(t)$ is an arbitrary function that is well behaved at $t = t'$. It is also easy to see that

$$\delta(t' - t) = \delta(t - t'). \quad (1.20)$$

1.5 Three-Dimensional Dirac Delta Function

The three-dimensional *Dirac delta function*, $\delta(\mathbf{r} - \mathbf{r}')$, has the property

$$\delta(\mathbf{r} - \mathbf{r}') = 0 \quad \text{for } \mathbf{r} \neq \mathbf{r}'. \quad (1.21)$$

In addition, however, the function is singular at $\mathbf{r} = \mathbf{r}'$ in such a manner that

$$\int_V \delta(\mathbf{r} - \mathbf{r}') dV = 1. \quad (1.22)$$

Here, V is any volume that contains the point $\mathbf{r} = \mathbf{r}'$. (Also, dV is an element of V expressed in terms of the components of \mathbf{r} , but independent of the components of \mathbf{r}' .) It follows that

$$\int_V f(\mathbf{r}) \delta(\mathbf{r} - \mathbf{r}') dV = f(\mathbf{r}'), \quad (1.23)$$

where $f(\mathbf{r})$ is an arbitrary function that is well behaved at $\mathbf{r} = \mathbf{r}'$. It is also easy to see that

$$\delta(\mathbf{r}' - \mathbf{r}) = \delta(\mathbf{r} - \mathbf{r}'). \quad (1.24)$$

We can show that

$$\nabla^2 \left(\frac{1}{|\mathbf{r} - \mathbf{r}'|} \right) = -4\pi \delta(\mathbf{r} - \mathbf{r}'). \quad (1.25)$$

(Here, ∇^2 is a Laplacian operator expressed in terms of the components of \mathbf{r} , but independent of the components of \mathbf{r}' .) We must first prove that

$$\nabla^2 \left(\frac{1}{|\mathbf{r} - \mathbf{r}'|} \right) = 0 \quad \text{for} \quad \mathbf{r} \neq \mathbf{r}', \quad (1.26)$$

in accordance with Equation (1.21). If $R = |\mathbf{r} - \mathbf{r}'|$ then this is equivalent to showing that

$$\frac{1}{R^2} \frac{d}{dR} \left[R^2 \frac{d}{dR} \left(\frac{1}{R} \right) \right] = 0 \quad (1.27)$$

for $R > 0$, which is indeed the case. (Here, R is treated as a radial spherical coordinate.) Next, we must show that

$$\int_V \nabla^2 \left(\frac{1}{|\mathbf{r} - \mathbf{r}'|} \right) dV = -4\pi, \quad (1.28)$$

in accordance with Equations (1.22) and (1.25). Suppose that S is a spherical surface, of radius R , centered on $\mathbf{r} = \mathbf{r}'$. Making use of the definition $\nabla^2 \phi \equiv \nabla \cdot \nabla \phi$, as well as the divergence theorem, we can write

$$\begin{aligned} \int_V \nabla^2 \left(\frac{1}{|\mathbf{r} - \mathbf{r}'|} \right) dV &= \int_V \nabla \cdot \nabla \left(\frac{1}{|\mathbf{r} - \mathbf{r}'|} \right) dV = \int_S \nabla \left(\frac{1}{|\mathbf{r} - \mathbf{r}'|} \right) \cdot d\mathbf{S} \\ &= 4\pi R^2 \frac{d}{dR} \left(\frac{1}{R} \right) = -4\pi. \end{aligned} \quad (1.29)$$

(Here, ∇ is a gradient operator expressed in terms of the components of \mathbf{r} , but independent of the components of \mathbf{r}' . Likewise, $d\mathbf{S}$ is a surface element involving the components of \mathbf{r} , but independent of the components of \mathbf{r}' .) Finally, if S is deformed into a general surface (without crossing the point $\mathbf{r} = \mathbf{r}'$) then the value of the volume integral is unchanged, as a consequence of Equation (1.26). Hence, we have demonstrated the validity of Equation (1.25).

1.6 Solution of Inhomogeneous Wave Equation

Equation (1.14), as well as the three Cartesian components of Equation (1.15), are inhomogeneous three-dimensional wave equations of the general form

$$\left(\frac{1}{c^2} \frac{\partial^2}{\partial t^2} - \nabla^2 \right) u = v, \quad (1.30)$$

where $u(\mathbf{r}, t)$ is an unknown potential, and $v(\mathbf{r}, t)$ a known source function. Let us investigate whether it is possible to find a unique solution of this type of equation.

Let us assume that the source function $v(\mathbf{r}, t)$ can be expressed as a Fourier integral,

$$v(\mathbf{r}, t) = \int_{-\infty}^{\infty} v_{\omega}(\mathbf{r}) e^{-i\omega t} d\omega. \quad (1.31)$$

The inverse transform is

$$v_{\omega}(\mathbf{r}) = \frac{1}{2\pi} \int_{-\infty}^{\infty} v(\mathbf{r}, t) e^{+i\omega t} dt. \quad (1.32)$$

Similarly, we can write the general potential $u(\mathbf{r}, t)$ as a Fourier integral,

$$u(\mathbf{r}, t) = \int_{-\infty}^{\infty} u_{\omega}(\mathbf{r}) e^{-i\omega t} d\omega, \quad (1.33)$$

with the corresponding inverse

$$u_{\omega}(\mathbf{r}) = \frac{1}{2\pi} \int_{-\infty}^{\infty} u(\mathbf{r}, t) e^{+i\omega t} dt. \quad (1.34)$$

Fourier transformation of Equation (1.30) yields

$$(\nabla^2 + k^2) u_{\omega} = -v_{\omega}, \quad (1.35)$$

where $k = \omega/c$.

Equation (1.35), which reduces to Poisson's equation (see Section 2.3),

$$\nabla^2 u_{\omega} = -v_{\omega}, \quad (1.36)$$

in the limit $k \rightarrow 0$, is known as *Helmholtz's equation*. Because Helmholtz's equation is linear, it is appropriate to attempt a Green's function method of solution. Let us try to find a Green's function, $G_{\omega}(\mathbf{r}, \mathbf{r}')$, such that

$$(\nabla^2 + k^2) G_{\omega}(\mathbf{r}, \mathbf{r}') = -\delta(\mathbf{r} - \mathbf{r}'). \quad (1.37)$$

The general solution to Equation (1.35) is then [cf., Equation (2.16)]

$$u_{\omega}(\mathbf{r}) = \int v_{\omega}(\mathbf{r}') G_{\omega}(\mathbf{r}, \mathbf{r}') dV'. \quad (1.38)$$

Let us adopt the spatial boundary condition $G_{\omega}(\mathbf{r}, \mathbf{r}') \rightarrow 0$ as $|\mathbf{r} - \mathbf{r}'| \rightarrow \infty$, so as to ensure that the potential goes to zero a long way from the source. Because Equation (1.37) is spherically symmetric about the point \mathbf{r}' , it is plausible that the Green's function itself is spherically symmetric: that is, $G_{\omega}(\mathbf{r} - \mathbf{r}') = G_{\omega}(|\mathbf{r} - \mathbf{r}'|)$. In this case, Equation (1.37) reduces to

$$\frac{1}{R} \frac{d^2(R G_{\omega})}{dR^2} + k^2 G_{\omega} = -\delta(\mathbf{R}), \quad (1.39)$$

where $\mathbf{R} = \mathbf{r} - \mathbf{r}'$, and $R = |\mathbf{R}|$. The most general solution to the above equation in the region $R > 0$ is¹

$$G_\omega(R) = \frac{A e^{+ikR} + B e^{-ikR}}{4\pi R}. \quad (1.40)$$

However, we know that Helmholtz's equation tends towards Poisson's equation in the limit $k \rightarrow 0$. It stands to reason that the Green's function for Helmholtz's equation much tend toward that for Poisson's equation in the same limit. Now, the Green's function for Poisson's equation, (1.36), satisfies

$$\nabla^2 G(\mathbf{r}, \mathbf{r}') = -\delta(\mathbf{r} - \mathbf{r}'), \quad (1.41)$$

as well as the usual constraint that $G(\mathbf{r}, \mathbf{r}') \rightarrow 0$ as $|\mathbf{r} - \mathbf{r}'| \rightarrow \infty$. It follows from Equation (1.25) that

$$G(\mathbf{r}, \mathbf{r}') = G(|\mathbf{r} - \mathbf{r}'|) = \frac{1}{4\pi |\mathbf{r} - \mathbf{r}'|} = \frac{1}{4\pi R}. \quad (1.42)$$

Thus, the condition that $G_\omega(R) \rightarrow G(R)$ as $k \rightarrow 0$ implies that $A + B = 1$.

Reconstructing $u(\mathbf{r}, t)$ from Equations (1.33), (1.38), and (1.40), we obtain

$$u(\mathbf{r}, t) = \frac{1}{4\pi} \int \int \frac{v_\omega(\mathbf{r}')}{R} \left[A e^{-i\omega(t-R/c)} + B e^{-i\omega(t+R/c)} \right] d\omega dV'. \quad (1.43)$$

It follows from Equation (1.31) that

$$u(\mathbf{r}, t) = \frac{A}{4\pi} \int \frac{v(\mathbf{r}', t - R/c)}{R} dV' + \frac{B}{4\pi} \int \frac{v(\mathbf{r}', t + R/c)}{R} dV'. \quad (1.44)$$

Now, the real-space Green's function for the inhomogeneous three-dimensional wave equation, (1.30), satisfies

$$\left(\frac{1}{c^2} \frac{\partial^2}{\partial t^2} - \nabla^2 \right) G(\mathbf{r}, \mathbf{r}'; t, t') = \delta(\mathbf{r} - \mathbf{r}') \delta(t - t'). \quad (1.45)$$

Hence, the most general solution of Equation (1.30) takes the form

$$u(\mathbf{r}, t) = \int \int v(\mathbf{r}', t') G(\mathbf{r}, \mathbf{r}'; t, t') dV' dt'. \quad (1.46)$$

Comparing Equations (1.44) and (1.46), we obtain

$$G(\mathbf{r}, \mathbf{r}'; t, t') = A G^{(+)}(\mathbf{r}, \mathbf{r}'; t, t') + B G^{(-)}(\mathbf{r}, \mathbf{r}'; t, t'), \quad (1.47)$$

where

$$G^{(\pm)}(\mathbf{r}, \mathbf{r}'; t, t') = \frac{\delta(t' - [t \mp |\mathbf{r} - \mathbf{r}'|/c])}{4\pi |\mathbf{r} - \mathbf{r}'|}, \quad (1.48)$$

and $A + B = 1$.

¹In principle, $A = A(\omega)$ and $B = B(\omega)$, with $A + B = 1$. However, we shall demonstrate, later on, that $B = 0$, otherwise causality is violated. It follows that $A = 1$. Thus, it is legitimate to assume, for the moment, that A and B are independent of ω .

The real-space Green's function specifies the response of the system to a point source located at position \mathbf{r}' that appears momentarily at time t' . According to the *retarded Green's function*, $G^{(+)}$, this response consists of a spherical wave, centered on the point \mathbf{r}' , that propagates forward in time. In order for the wave to reach position \mathbf{r} at time t , it must have been emitted from the source at \mathbf{r}' at the *retarded time* $t_r = t - |\mathbf{r} - \mathbf{r}'|/c$. According to the *advanced Green's function*, $G^{(-)}$, the response consists of a spherical wave, centered on the point \mathbf{r}' , that propagates backward in time. Clearly, the advanced potential is not consistent with our ideas about causality, which demand that an effect can never precede its cause in time. Thus, the Green's function that is consistent with our experience is

$$G(\mathbf{r}, \mathbf{r}'; t, t') = G^{(+)}(\mathbf{r}, \mathbf{r}'; t, t') = \frac{\delta(t' - [t - |\mathbf{r} - \mathbf{r}'|/c])}{4\pi |\mathbf{r} - \mathbf{r}'|}. \quad (1.49)$$

Incidentally, we are able to find solutions of the inhomogeneous wave equation, (1.30), that propagate backward in time because this equation is time symmetric (i.e., it is invariant under the transformation $t \rightarrow -t$).

In conclusion, the most general solution of the inhomogeneous wave equation, (1.30), that satisfies sensible boundary conditions at infinity, and is consistent with causality, is

$$u(\mathbf{r}, t) = \int \frac{v(\mathbf{r}', t - |\mathbf{r} - \mathbf{r}'|/c)}{4\pi |\mathbf{r} - \mathbf{r}'|} dV'. \quad (1.50)$$

This expression is sometimes written

$$u(\mathbf{r}, t) = \int \frac{[v(\mathbf{r}')] }{4\pi |\mathbf{r} - \mathbf{r}'|} dV', \quad (1.51)$$

where the rectangular bracket symbol $[]$ denotes that the terms inside the bracket are to be evaluated at the retarded time $t - |\mathbf{r} - \mathbf{r}'|/c$. Note, in particular, from Equation (1.50), that if there is no source [i.e., if $v(\mathbf{r}, t) = 0$] then there is no field [i.e., $u(\mathbf{r}, t) = 0$]. But, is expression (1.50) really the only solution of Equation (1.30) that satisfies sensible boundary conditions at infinity? In other words, is this solution really unique? Unfortunately, there is a weak link in our derivation—between Equations (1.38) and (1.39)—where we assumed, without proof, that the Green's function for Helmholtz's equation, subject to the boundary condition $G_\omega(\mathbf{r}, \mathbf{r}') \rightarrow 0$ as $|\mathbf{r} - \mathbf{r}'| \rightarrow \infty$, is spherically symmetric. Let us try to fix this problem.

With the benefit of hindsight, we can see that the Fourier-space Green's function

$$G_\omega = \frac{e^{+ikR}}{4\pi R} \quad (1.52)$$

corresponds to the retarded solution in real space, and is, therefore, the correct physical Green's function in Fourier space. The Fourier-space Green's function

$$G_\omega = \frac{e^{-ikR}}{4\pi R} \quad (1.53)$$

corresponds to the advanced solution in real space, and must, therefore, be rejected. We can select the retarded Green's function in Fourier space by imposing the following boundary condition at

infinity

$$\lim_{R \rightarrow \infty} R \left(\frac{\partial G_\omega}{\partial R} - i k G_\omega \right) = 0. \quad (1.54)$$

This is called the *Sommerfeld radiation condition*, and basically ensures that infinity is an absorber of radiation, but not a source. But, does this boundary condition uniquely select the spherically symmetric Green's function (1.52) as the solution of

$$(\nabla^2 + k^2) G_\omega(R, \theta, \varphi) = -\delta(\mathbf{R})? \quad (1.55)$$

Here, (R, θ, φ) are spherical polar coordinates. If it does then we can be sure that Equation (1.50) represents the unique solution of the inhomogeneous wave equation, (1.30), that is consistent with causality.

Let us suppose that there are two different solutions of Equation (1.55), both of which satisfy the boundary condition (1.54), and revert to the unique (see Section 2.3) Green's function for Poisson's equation, (1.42), in the limit $R \rightarrow 0$. Let us call these solutions u_1 and u_2 , and let us form the difference $w = u_1 - u_2$. Consider a surface Σ_0 which is a sphere of arbitrarily small radius centred on the origin. Consider a second surface Σ_∞ which is a sphere of arbitrarily large radius centred on the origin. Let V denote the volume enclosed by these surfaces. The difference function w satisfies the homogeneous Helmholtz equation,

$$(\nabla^2 + k^2) w = 0, \quad (1.56)$$

throughout V . According to the generalized (to deal with complex potentials) Green's theorem (see Section 2.9),

$$\int_V (w \nabla^2 w^* - w^* \nabla^2 w) dV = \left(\int_{\Sigma_0} + \int_{\Sigma_\infty} \right) \left(w \frac{\partial w^*}{\partial n} - w^* \frac{\partial w}{\partial n} \right) dS, \quad (1.57)$$

where $\partial/\partial n$ denotes a derivative normal to the surface in question. It is clear from Equation (1.56) that the volume integral is zero. It is also clear that the first surface integral is zero, because both u_1 and u_2 must revert to the Green's function for Poisson's equation in the limit $R \rightarrow 0$. Thus,

$$\int_{\Sigma_\infty} \left(w \frac{\partial w^*}{\partial n} - w^* \frac{\partial w}{\partial n} \right) dS = 0. \quad (1.58)$$

Equation (1.56) can be written

$$\frac{\partial^2(Rw)}{\partial R^2} + \frac{D(Rw)}{R^2} + k^2 R w = 0, \quad (1.59)$$

where D is the spherical harmonic operator

$$D = \frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial}{\partial \theta} \right) + \frac{1}{\sin^2 \theta} \frac{\partial^2}{\partial \varphi^2}. \quad (1.60)$$

The most general solution to Equation (1.59) takes the form

$$w(R, \theta, \varphi) = \sum_{l,m=0,\infty} [C_{l,m} h_l^{(1)}(kR) + D_{l,m} h_l^{(2)}(kR)] Y_{l,m}(\theta, \varphi). \quad (1.61)$$

Here, the $C_{l,m}$ and $D_{l,m}$ are arbitrary coefficients, the $Y_{l,m}$ are spherical harmonics (see Section 3.3), and

$$h_l^{(1,2)}(\rho) = \sqrt{\frac{\pi}{2\rho}} H_{l+1/2}^{1,2}(\rho), \quad (1.62)$$

where the $H_n^{1,2}$ are Hankel functions of the first and second kind.² It can be demonstrated that³

$$H_n^1(\rho) = \sqrt{\frac{2}{\pi\rho}} e^{+i[\rho-(n+1/2)\pi/2]} \sum_{m=0,1,2,\dots} \frac{(n,m)}{(-2i\rho)^m}, \quad (1.63)$$

$$H_n^2(\rho) = \sqrt{\frac{2}{\pi\rho}} e^{-i[\rho-(n+1/2)\pi/2]} \sum_{m=0,1,2,\dots} \frac{(n,m)}{(+2i\rho)^m}, \quad (1.64)$$

where

$$(n,m) = \frac{(4n^2 - 1)(4n^2 - 9) \cdots (4n^2 - \{2m - 1\}^2)}{2^{2m} m!} \quad (1.65)$$

and $(n,0) = 1$. Note that the summations in Equations (1.63) and (1.64) terminate after $n + 1/2$ terms.

The large- R behavior of the $h_l^{(2)}(kR)$ functions is clearly inconsistent with the Sommerfeld radiation condition, (1.54). It follows that all of the $D_{l,m}$ in Equations (1.61) are zero. The most general solution can now be expressed in the form

$$w(R, \theta, \varphi) = \frac{e^{+ikR}}{R} \sum_{n=0,\infty} \frac{f_n(\theta, \varphi)}{R^n}, \quad (1.66)$$

where the $f_n(\theta, \varphi)$ are various weighted sums of the spherical harmonics. Substitution of this solution into the differential equation (1.59) yields

$$e^{+ikR} \sum_{n=0,\infty} \left[-\frac{2ikn}{R^{n+1}} + \frac{n(n+1)}{R^{n+2}} + \frac{D}{R^{n+2}} \right] f_n = 0. \quad (1.67)$$

Replacing the index of summation n in the first term of the parentheses by $n + 1$, we obtain

$$e^{+ikR} \sum_{n=0,\infty} \frac{-2ik(n+1)f_{n+1} + [n(n+1) + D]f_n}{R^{n+2}} = 0, \quad (1.68)$$

which yields the recursion relation

$$2ik(n+1)f_{n+1} = [n(n+1) + D]f_n. \quad (1.69)$$

²J.D. Jackson, *Classical Electrodynamics*, 2nd Edition, (Wiley, 1962), p. 104.

³A. Sommerfeld, *Partial Differential Equations in Physics*, (Academic Press, New York, 1964), p. 117.

It follows that if $f_0 = 0$ then all of the f_n are equal to zero.

Let us now consider the surface integral (1.58). Because we are interested in the limit $R \rightarrow \infty$, we can replace w by the first term of its expansion in (1.66), so that

$$\int_{\Sigma_\infty} \left(w \frac{\partial w^*}{\partial n} - w^* \frac{\partial w}{\partial n} \right) dS = -2ik \int |f_0|^2 d\Omega = 0, \quad (1.70)$$

where $d\Omega$ is an element of solid angle. It is clear that $f_0 = 0$. This implies that $f_1 = f_2 = \dots = 0$, and, hence, that $w = 0$. Thus, there is only one solution of Equation (1.55) that is consistent with the Sommerfeld radiation condition, and this is given by Equation (1.52). We can now be sure that Equation (1.50) is the unique solution of Equation (1.30), subject to the boundary condition (1.54). This boundary condition ensures that infinity is an absorber of electromagnetic radiation, but not an emitter, which seems entirely reasonable.

1.7 Retarded Potentials

We are now in a position to solve Maxwell's equations. Recall, from Section 1.3, that Maxwell equations reduce to

$$\left(\frac{1}{c^2} \frac{\partial^2}{\partial t^2} - \nabla^2 \right) \phi = \frac{\rho}{\epsilon_0}, \quad (1.71)$$

$$\left(\frac{1}{c^2} \frac{\partial^2}{\partial t^2} - \nabla^2 \right) \mathbf{A} = \mu_0 \mathbf{j}. \quad (1.72)$$

We can solve these inhomogeneous three-dimensional waves equations using the appropriate Green's function, (1.49). In fact, making use of Equation (1.46), we find that

$$\phi(\mathbf{r}, t) = \frac{1}{4\pi \epsilon_0} \int \frac{\rho(\mathbf{r}', t - |\mathbf{r} - \mathbf{r}'|/c)}{|\mathbf{r} - \mathbf{r}'|} dV' \quad (1.73)$$

$$\mathbf{A}(\mathbf{r}, t) = \frac{\mu_0}{4\pi} \int \frac{\mathbf{j}(\mathbf{r}', t - |\mathbf{r} - \mathbf{r}'|/c)}{|\mathbf{r} - \mathbf{r}'|} dV'. \quad (1.74)$$

Alternatively, we can write

$$\phi(\mathbf{r}, t) = \frac{1}{4\pi \epsilon_0} \int \frac{[\rho(\mathbf{r}')] }{|\mathbf{r} - \mathbf{r}'|} dV', \quad (1.75)$$

$$\mathbf{A}(\mathbf{r}, t) = \frac{\mu_0}{4\pi} \int \frac{[\mathbf{j}(\mathbf{r}')] }{|\mathbf{r} - \mathbf{r}'|} dV'. \quad (1.76)$$

The above potentials are termed *retarded potentials* (because the integrands are evaluated at the retarded time). Finally, according to the discussion in the previous section, we can be sure that Equations (1.75) and (1.76) are the unique solutions to Equations (1.71) and (1.72), respectively, subject to sensible boundary conditions at infinity.

1.8 Retarded Fields

We have found the solution to Maxwell's equations in terms of retarded potentials. Let us now construct the associated retarded electric and magnetic fields using (see Section 1.3)

$$\mathbf{E} = -\nabla\phi - \frac{\partial\mathbf{A}}{\partial t}, \quad (1.77)$$

$$\mathbf{B} = \nabla \times \mathbf{A}. \quad (1.78)$$

It is helpful to write

$$\mathbf{R} = \mathbf{r} - \mathbf{r}', \quad (1.79)$$

where $R = |\mathbf{r} - \mathbf{r}'|$. The retarded time becomes $t_r = t - R/c$, and a general retarded quantity is written $[F(\mathbf{r}', t)] \equiv F(\mathbf{r}', t_r)$. Thus, we can express the retarded potential solutions of Maxwell's equations in the particularly compact form

$$\phi(\mathbf{r}, t) = \frac{1}{4\pi\epsilon_0} \int \frac{[\rho]}{R} dV', \quad (1.80)$$

$$\mathbf{A}(\mathbf{r}, t) = \frac{\mu_0}{4\pi} \int \frac{[\mathbf{j}]}{R} dV'. \quad (1.81)$$

It is easily seen that

$$\begin{aligned} \nabla\phi &= \frac{1}{4\pi\epsilon_0} \int \left([\rho] \nabla(R^{-1}) + \frac{[\partial\rho/\partial t]}{R} \nabla t_r \right) dV' \\ &= -\frac{1}{4\pi\epsilon_0} \int \left(\frac{[\rho]}{R^3} \mathbf{R} + \frac{[\partial\rho/\partial t]}{cR^2} \mathbf{R} \right) dV', \end{aligned} \quad (1.82)$$

where use has been made of

$$\nabla R = \frac{\mathbf{R}}{R}, \quad \nabla(R^{-1}) = -\frac{\mathbf{R}}{R^3}, \quad \nabla t_r = -\frac{\mathbf{R}}{cR}. \quad (1.83)$$

Likewise,

$$\begin{aligned} \nabla \times \mathbf{A} &= \frac{\mu_0}{4\pi} \int \left(\nabla(R^{-1}) \times [\mathbf{j}] + \frac{\nabla t_r \times [\partial\mathbf{j}/\partial t]}{R} \right) dV' \\ &= -\frac{\mu_0}{4\pi} \int \left(\frac{\mathbf{R} \times [\mathbf{j}]}{R^3} + \frac{\mathbf{R} \times [\partial\mathbf{j}/\partial t]}{cR^2} \right) dV'. \end{aligned} \quad (1.84)$$

Equations (1.77), (1.78), (1.82), and (1.84) can be combined to give

$$\mathbf{E} = \frac{1}{4\pi\epsilon_0} \int \left([\rho] \frac{\mathbf{R}}{R^3} + \left[\frac{\partial\rho}{\partial t} \right] \frac{\mathbf{R}}{cR^2} - \frac{[\partial\mathbf{j}/\partial t]}{c^2 R} \right) dV', \quad (1.85)$$

and

$$\mathbf{B} = \frac{\mu_0}{4\pi} \int \left(\frac{[\mathbf{j}] \times \mathbf{R}}{R^3} + \frac{[\partial\mathbf{j}/\partial t] \times \mathbf{R}}{cR^2} \right) dV'. \quad (1.86)$$

Suppose that our charges and currents vary on some characteristic timescale t_0 . Let us define $R_0 = ct_0$, which is the distance a light ray travels in time t_0 . We can evaluate Equations (1.85) and (1.86) in two asymptotic regions: the *near field* region $R \ll R_0$, and the *far field* region $R \gg R_0$. In the near field region,

$$\frac{|t - t_r|}{t_0} = \frac{R}{R_0} \ll 1, \quad (1.87)$$

so the difference between retarded time and standard time is relatively small. This allows us to expand retarded quantities in a Taylor series. Thus,

$$[\rho] \simeq \rho + \frac{\partial \rho}{\partial t} (t_r - t) + \frac{1}{2} \frac{\partial^2 \rho}{\partial t^2} (t_r - t)^2 + \dots, \quad (1.88)$$

giving

$$[\rho] \simeq \rho - \frac{\partial \rho}{\partial t} \frac{R}{c} + \frac{1}{2} \frac{\partial^2 \rho}{\partial t^2} \frac{R^2}{c^2} + \dots. \quad (1.89)$$

Expansion of the retarded quantities in the near field region yields

$$\mathbf{E}(\mathbf{r}, t) \simeq \frac{1}{4\pi \epsilon_0} \int \left(\frac{\rho \mathbf{R}}{R^3} - \frac{1}{2} \frac{\partial^2 \rho}{\partial t^2} \frac{\mathbf{R}}{c^2 R} - \frac{\partial \mathbf{j} / \partial t}{c^2 R} + \dots \right) dV', \quad (1.90)$$

$$\mathbf{B}(\mathbf{r}, t) \simeq \frac{\mu_0}{4\pi} \int \left(\frac{\mathbf{j} \times \mathbf{R}}{R^3} - \frac{1}{2} \frac{(\partial^2 \mathbf{j} / \partial t^2) \times \mathbf{R}}{c^2 R} + \dots \right) dV'. \quad (1.91)$$

In Equation (1.90), the first term on the right-hand side corresponds to Coulomb's law, the second term is the lowest order correction to Coulomb's law due to retardation effects, and the third term corresponds to Faraday induction. In Equation (1.91), the first term on the right-hand side is the Biot-Savart law, and the second term is the lowest order correction to the Biot-Savart law due to retardation effects. Note that the retardation corrections are only of order $(R/R_0)^2$. We might suppose, from looking at Equations (1.85) and (1.86), that the corrections should be of order R/R_0 . However, all of the order R/R_0 terms canceled out in the previous expansion.

In the far field region, $R \gg R_0$, Equations (1.85) and (1.86) are dominated by the terms that vary like R^{-1} , so that

$$\mathbf{E}(\mathbf{r}, t) \simeq -\frac{1}{4\pi \epsilon_0} \int \frac{[\partial \mathbf{j}_\perp / \partial t]}{c^2 R} dV', \quad (1.92)$$

$$\mathbf{B}(\mathbf{r}, t) \simeq \frac{\mu_0}{4\pi} \int \frac{[\partial \mathbf{j}_\perp / \partial t] \times \mathbf{R}}{c R^2} dV', \quad (1.93)$$

where

$$\mathbf{j}_\perp = \mathbf{j} - \frac{(\mathbf{j} \cdot \mathbf{R})}{R^2} \mathbf{R}. \quad (1.94)$$

Here, use has been made of $[\partial \rho / \partial t] = -[\nabla \cdot \mathbf{j}]$ and $[\nabla \cdot \mathbf{j}] \simeq -[\partial \mathbf{j} / \partial t] \cdot \mathbf{R} / (cR)$. Suppose that our charges and currents are localized to some finite region of space in the vicinity of the origin, and that the extent of the current-and-charge-containing region is much less than $|\mathbf{r}|$. It follows that retarded quantities can be written

$$[\rho(\mathbf{r}', t)] \simeq \rho(\mathbf{r}', t - r/c), \quad (1.95)$$

et cetera. Thus, the electric field reduces to

$$\mathbf{E}(\mathbf{r}, t) \simeq -\frac{1}{4\pi\epsilon_0} \frac{\left[\int \partial \mathbf{j}_\perp / \partial t \, dV' \right]}{c^2 r}, \quad (1.96)$$

whereas the magnetic field is given by

$$\mathbf{B}(\mathbf{r}, t) \simeq \frac{1}{4\pi\epsilon_0} \frac{\left[\int \partial \mathbf{j}_\perp / \partial t \, dV' \right] \times \mathbf{r}}{c^3 r^2}. \quad (1.97)$$

Here, $[\dots]$ merely denotes evaluation at the retarded time $t - r/c$. Note that

$$\frac{E}{B} = c, \quad (1.98)$$

and

$$\mathbf{E} \cdot \mathbf{B} = 0. \quad (1.99)$$

This configuration of electric and magnetic fields is characteristic of an electromagnetic wave. In fact, Equations (1.96) and (1.97) describe an electromagnetic wave propagating radially away from the charge and current containing region. The wave is clearly driven by time-varying electric currents. Now, charges moving with a constant velocity constitute a steady current, so a nonsteady current is associated with accelerating charges. We conclude that accelerating electric charges emit electromagnetic waves. The wave fields, (1.96) and (1.97), fall off like the inverse of the distance from the wave source. This behavior should be contrasted with that of Coulomb or Biot-Savart fields, which fall off like the inverse square of the distance from the source.

In conclusion, electric and magnetic fields look simple in the near field region (they are just Coulomb fields, etc.), and also in the far field region (they are just electromagnetic waves). Only in the intermediate region, $R \sim R_0$, do the fields get really complicated.

1.9 Electromagnetic Energy Conservation

Consider the fourth Maxwell equation:

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{j} + \epsilon_0 \mu_0 \frac{\partial \mathbf{E}}{\partial t}. \quad (1.100)$$

Forming the scalar product with the electric field, and rearranging, we obtain

$$-\mathbf{E} \cdot \mathbf{j} = -\frac{\mathbf{E} \cdot \nabla \times \mathbf{B}}{\mu_0} + \epsilon_0 \mathbf{E} \cdot \frac{\partial \mathbf{E}}{\partial t}, \quad (1.101)$$

which can be rewritten

$$-\mathbf{E} \cdot \mathbf{j} = -\frac{\mathbf{E} \cdot \nabla \times \mathbf{B}}{\mu_0} + \frac{\partial}{\partial t} \left(\frac{\epsilon_0 E^2}{2} \right). \quad (1.102)$$

Now,

$$\nabla \cdot (\mathbf{E} \times \mathbf{B}) \equiv \mathbf{B} \cdot \nabla \times \mathbf{E} - \mathbf{E} \cdot \nabla \times \mathbf{B}, \quad (1.103)$$

so

$$-\mathbf{E} \cdot \mathbf{j} = \nabla \cdot \left(\frac{\mathbf{E} \times \mathbf{B}}{\mu_0} \right) - \frac{\mathbf{B} \cdot \nabla \times \mathbf{E}}{\mu_0} + \frac{\partial}{\partial t} \left(\frac{\epsilon_0 E^2}{2} \right). \quad (1.104)$$

Making use of third Maxwell equation,

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \quad (1.105)$$

we obtain

$$-\mathbf{E} \cdot \mathbf{j} = \nabla \cdot \left(\frac{\mathbf{E} \times \mathbf{B}}{\mu_0} \right) + \mu_0^{-1} \mathbf{B} \cdot \frac{\partial \mathbf{B}}{\partial t} + \frac{\partial}{\partial t} \left(\frac{\epsilon_0 E^2}{2} \right), \quad (1.106)$$

which can be rewritten

$$-\mathbf{E} \cdot \mathbf{j} = \nabla \cdot \left(\frac{\mathbf{E} \times \mathbf{B}}{\mu_0} \right) + \frac{\partial}{\partial t} \left(\frac{\epsilon_0 E^2}{2} + \frac{B^2}{2\mu_0} \right). \quad (1.107)$$

Thus, we get

$$\frac{\partial U}{\partial t} + \nabla \cdot \mathbf{u} = -\mathbf{E} \cdot \mathbf{j}, \quad (1.108)$$

where U and \mathbf{u} are specified in Equations (1.109) and (1.110), respectively.

By comparison with Equation (1.7), we can recognize the previous expression as some sort of conservation equation. Here, U is the density of the conserved quantity, \mathbf{u} is the flux of the conserved quantity, and $-\mathbf{E} \cdot \mathbf{j}$ is the rate at which the conserved quantity is created per unit volume. However, $\mathbf{E} \cdot \mathbf{j}$ is the rate per unit volume at which electric charges gain energy via interaction with electromagnetic fields. Hence, $-\mathbf{E} \cdot \mathbf{j}$ is the rate per unit volume at which electromagnetic fields gain energy via interaction with charges. It follows that Equation (1.108) is a conservation equation for electromagnetic energy. Thus.

$$U = \frac{\epsilon_0 E^2}{2} + \frac{B^2}{2\mu_0} \quad (1.109)$$

can be interpreted as the electromagnetic energy density, and

$$\mathbf{u} = \frac{\mathbf{E} \times \mathbf{B}}{\mu_0} \quad (1.110)$$

as the electromagnetic energy flux. The latter quantity is usually called the *Poynting flux*, after its discoverer.

1.10 Electromagnetic Momentum Conservation

Let $g^{(i)}$ be the density of electromagnetic momentum directed parallel to the i th Cartesian axis. (Here, $i = 1$ corresponds to the x -axis, $i = 2$ to the y -axis, and $i = 3$ to the z -axis.) Furthermore, let $\mathbf{G}^{(i)}$ be the flux of such momentum. We would expect the conservation equation for electromagnetic momentum directed parallel to the i th Cartesian axis to take the form

$$\frac{\partial g^{(i)}}{\partial t} + \nabla \cdot \mathbf{G}^{(i)} = -(\rho \mathbf{E} + \mathbf{j} \times \mathbf{B})_i, \quad (1.111)$$

where the subscript i denotes a component of a vector parallel to the i th Cartesian axis. The term on the right-hand side is the rate per unit volume at which electromagnetic fields gain momentum parallel to the i th Cartesian axis via interaction with matter. Thus, the term is minus the rate at which matter gains momentum parallel to the i th Cartesian axis via interaction with electromagnetic fields. In other words, the term is minus the i th component of the force per unit volume exerted on matter by electromagnetic fields. [See Equation (1.10).] Equation (1.111) can be generalized to give

$$\frac{\partial \mathbf{g}}{\partial t} + \nabla \cdot \mathbf{G} = -(\rho \mathbf{E} + \mathbf{j} \times \mathbf{B}), \quad (1.112)$$

where \mathbf{g} is the electromagnetic momentum density (the i th Cartesian component of \mathbf{g} is thus $g^{(i)}$), and \mathbf{G} is a tensor (see Section 12.5) whose Cartesian components $G_{ij} = \mathbf{G}^{(i)} \cdot \mathbf{e}_j$, where \mathbf{e}_j is a unit vector parallel to the j th Cartesian axis, specify the flux of electromagnetic momentum parallel to the i th Cartesian axis across a plane surface whose normal is parallel to the j th Cartesian axis. Let us attempt to derive an expression of the form (1.112) from Maxwell's equations.

Maxwell's equations are as follows:

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0}, \quad (1.113)$$

$$\nabla \cdot \mathbf{B} = 0, \quad (1.114)$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \quad (1.115)$$

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{j} + \epsilon_0 \mu_0 \frac{\partial \mathbf{E}}{\partial t}. \quad (1.116)$$

We can take the vector product of Equation (1.116) divided by μ_0 with \mathbf{B} , and rearrange, to give

$$-\epsilon_0 \frac{\partial \mathbf{E}}{\partial t} \times \mathbf{B} = \frac{\mathbf{B} \times (\nabla \times \mathbf{B})}{\mu_0} + \mathbf{j} \times \mathbf{B}. \quad (1.117)$$

Next, we can take the vector product of \mathbf{E} with Equation (1.115) times ϵ_0 , rearrange, and add the result to the previous equation. We obtain

$$-\epsilon_0 \frac{\partial \mathbf{E}}{\partial t} \times \mathbf{B} - \epsilon_0 \mathbf{E} \times \frac{\partial \mathbf{B}}{\partial t} = \epsilon_0 \mathbf{E} \times (\nabla \times \mathbf{E}) + \frac{\mathbf{B} \times (\nabla \times \mathbf{B})}{\mu_0} + \mathbf{j} \times \mathbf{B}. \quad (1.118)$$

Making use of Equations (1.113) and (1.114), we get

$$\begin{aligned} -\frac{\partial}{\partial t} (\epsilon_0 \mathbf{E} \times \mathbf{B}) &= \epsilon_0 \mathbf{E} \times (\nabla \times \mathbf{E}) + \frac{\mathbf{B} \times (\nabla \times \mathbf{B})}{\mu_0} \\ &\quad - \epsilon_0 (\nabla \cdot \mathbf{E}) \mathbf{E} - \frac{1}{\mu_0} (\nabla \cdot \mathbf{B}) \mathbf{B} + \rho \mathbf{E} + \mathbf{j} \times \mathbf{B}. \end{aligned} \quad (1.119)$$

Now,

$$\nabla(E^2/2) \equiv \mathbf{E} \times (\nabla \times \mathbf{E}) + (\mathbf{E} \cdot \nabla) \mathbf{E}, \quad (1.120)$$

with a similar equation for \mathbf{B} . Hence, Equation (1.119) can be written

$$\begin{aligned} -\frac{\partial}{\partial t} (\epsilon_0 \mathbf{E} \times \mathbf{B}) &= \epsilon_0 \left[\nabla(E^2/2) - (\nabla \cdot \mathbf{E}) \mathbf{E} - (\mathbf{E} \cdot \nabla) \mathbf{E} \right] \\ &\quad + \frac{1}{\mu_0} \left[\nabla(B^2/2) - (\nabla \cdot \mathbf{B}) \mathbf{B} - (\mathbf{B} \cdot \nabla) \mathbf{B} \right] \\ &\quad + \rho \mathbf{E} + \mathbf{j} \times \mathbf{B}. \end{aligned} \quad (1.121)$$

Finally, when written in terms of components, the above equation becomes

$$\begin{aligned} -\frac{\partial}{\partial t} (\epsilon_0 \mathbf{E} \times \mathbf{B})_i &= \frac{\partial}{\partial x_j} (\epsilon_0 E^2 \delta_{ij}/2 - \epsilon_0 E_i E_j + B^2 \delta_{ij}/2\mu_0 - B_i B_j/\mu_0) \\ &\quad + (\rho \mathbf{E} + \mathbf{j} \times \mathbf{B})_i, \end{aligned} \quad (1.122)$$

because $[(\nabla \cdot \mathbf{E}) \mathbf{E}]_i \equiv (\partial E_j / \partial x_j) E_i$, and $[(\mathbf{E} \cdot \nabla) \mathbf{E}]_i \equiv E_j (\partial E_i / \partial x_j)$. Here, x_1 corresponds to x , x_2 to y , and x_3 to z . Furthermore, δ_{ij} is a Kronecker delta symbol (i.e., $\delta_{ij} = 1$ if $i = j$, and $\delta_{ij} = 0$ otherwise). Finally, we are making use of the Einstein summation convention (that repeated indices are summed from 1 to 3). Comparing the previous expression with Equation (1.112), we conclude that the momentum density of electromagnetic fields takes the form

$$\mathbf{g} = \epsilon_0 (\mathbf{E} \times \mathbf{B}), \quad (1.123)$$

whereas the corresponding momentum flux tensor has the Cartesian components

$$G_{ij} = \epsilon_0 (E^2 \delta_{ij}/2 - E_i E_j) + (B^2 \delta_{ij}/2 - B_i B_j)/\mu_0. \quad (1.124)$$

The momentum conservation equation, (1.112), is sometimes written

$$\rho \mathbf{E} + \mathbf{j} \times \mathbf{B} = \nabla \cdot \mathbf{T} - \frac{\partial}{\partial t} (\epsilon_0 \mathbf{E} \times \mathbf{B}), \quad (1.125)$$

where

$$T_{ij} = -G_{ij} = \epsilon_0 (E_i E_j - E^2 \delta_{ij}/2) + (B_i B_j - B^2 \delta_{ij}/2)/\mu_0 \quad (1.126)$$

is called the *Maxwell stress tensor*.

1.11 Exercises

- 1.1 Demonstrate that the energy contained in the magnetic field generated by a stationary current distribution $\mathbf{j}(\mathbf{r})$ in vacuum is given by

$$W = \frac{\mu_0}{8\pi} \int \int \frac{\mathbf{j}(\mathbf{r}) \cdot \mathbf{j}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} dV dV'.$$

- 1.2 A transverse plane wave is incident normally in vacuum on a perfectly absorbing flat screen. Show that the pressure exerted on the screen is equal to the electromagnetic energy density of the wave.

- 1.3 Consider an infinite parallel-plate capacitor. Let the lower plate lie at $z = -d/2$, and carry the charge density $-\sigma$. Likewise, let the upper plate lie at $z = +d/2$, and carry the charge density $+\sigma$. Calculate the electromagnetic momentum flux across the y - z plane. Hence, determine the direction and magnitude of the force per unit area that the plates exert on one another.
- 1.4 The equation of electromagnetic angular momentum conservation takes the general form

$$\frac{\partial \mathbf{L}}{\partial t} + \nabla \cdot \mathbf{M} = -\mathbf{r} \times (\rho \mathbf{E} + \mathbf{j} \times \mathbf{B}),$$

where \mathbf{L} is the electromagnetic angular momentum density, and the tensor \mathbf{M} is the electromagnetic angular momentum flux. Demonstrate that

$$\mathbf{L} = \mathbf{r} \times \mathbf{g},$$

and

$$\mathbf{M} = \mathbf{r} \times \mathbf{G},$$

where \mathbf{g} is the electromagnetic momentum density, and \mathbf{G} the electromagnetic momentum flux tensor.

- 1.5 A long solenoid of radius R , with N turn per unit length, carries a steady current I . Two hollow cylinders of length l are fixed coaxially such that they are free to rotate. The first cylinder, whose radius is $a < R$, carries the uniformly distributed electric charge Q . The second cylinder, whose radius is $b > R$, carries the uniformly distributed electric charge $-Q$. Both cylinders are initially stationary. When the current is switched off the cylinders start to rotate. Find the final angular momenta of the two cylinders, and demonstrate that the total angular momentum of the system is the same before and after the current is switched off.
- 1.6 Consider a system consisting of an electric charge e and a magnetic monopole g separated by a distance d . Demonstrate that the total angular momentum stored in the resulting electromagnetic fields is

$$L = \frac{\mu_0}{4\pi} e g.$$

[Hint: The radial magnetic field generated a distance r from a magnetic monopole of strength g is of magnitude $(\mu_0/4\pi)(g/r^2)$.]

2 Electrostatic Fields

2.1 Introduction

This chapter discusses electric fields generated by stationary charge distributions. Such fields are termed *electrostatic*. However, before commencing this discussion (in Section 2.4), it is convenient to review some useful mathematics.

2.2 Laplace's Equation

Laplace's equation is written

$$\nabla^2 \phi(\mathbf{r}) = 0, \quad (2.1)$$

where the function $\phi(\mathbf{r})$ is often referred to as a potential. Suppose that we wish to find a solution to this equation in some finite volume V , bounded by a closed surface S , subject to the boundary condition

$$\phi(\mathbf{r}) = 0, \quad (2.2)$$

when \mathbf{r} lies on S . Consider the vector identity

$$\nabla \cdot (\phi \nabla \phi) \equiv \phi \nabla^2 \phi + \nabla \phi \cdot \nabla \phi. \quad (2.3)$$

Integrating this expression over V , making use of the divergence theorem, we obtain

$$\int_S \phi \nabla \phi \cdot d\mathbf{S} = \int_V (\phi \nabla^2 \phi + \nabla \phi \cdot \nabla \phi) dV. \quad (2.4)$$

It follows from Equations (2.1) and (2.2) that

$$\int_V |\nabla \phi|^2 dV = 0, \quad (2.5)$$

which implies that $\nabla \phi = \mathbf{0}$ throughout V and on S . Hence, Equation (2.2) yields

$$\phi(\mathbf{r}) = 0 \quad (2.6)$$

throughout V and on S . We conclude that the only solution to Laplace's equation, (2.1), subject to the boundary condition (2.2), is the trivial solution (2.6). Finally, if we let the surface S tend to infinity then we deduce that the only solution to Laplace's equation, (2.1), subject to the boundary condition

$$\phi(\mathbf{r}) \rightarrow 0 \quad \text{as} \quad |\mathbf{r}| \rightarrow \infty, \quad (2.7)$$

is

$$\phi(\mathbf{r}) = 0 \quad (2.8)$$

for all \mathbf{r} .

Consider a potential $\phi(\mathbf{r})$ that satisfies Laplace's equation, (2.1), in some finite volume V , bounded by the closed surface S , subject to the boundary condition

$$\phi(\mathbf{r}) = \phi_S(\mathbf{r}), \quad (2.9)$$

when \mathbf{r} lies on S . Here, $\phi_S(\mathbf{r})$ is a known surface distribution. We can demonstrate that this potential is unique. Let $\phi_1(\mathbf{r})$ and $\phi_2(\mathbf{r})$ be two supposedly different potentials that both satisfy Laplace's equation throughout V , as well as the previous boundary condition on S . Let us form the difference $\phi_3(\mathbf{r}) = \phi_1(\mathbf{r}) - \phi_2(\mathbf{r})$. This function satisfies Laplace's equation throughout V , subject to the boundary condition

$$\phi_3(\mathbf{r}) = 0 \quad (2.10)$$

when \mathbf{r} lies on S . However, as we have already seen, this implies that $\phi_3(\mathbf{r}) = 0$ throughout V and on S . Hence, $\phi_1(\mathbf{r})$ and $\phi_2(\mathbf{r})$ are identical, and the potential $\phi(\mathbf{r})$ is therefore unique.

2.3 Poisson's Equation

Poisson's equation is written

$$\nabla^2 \phi(\mathbf{r}) = v(\mathbf{r}). \quad (2.11)$$

Here, the function $v(\mathbf{r})$ is conventionally referred to as a source. Suppose that we have to solve Equation (2.11) over all space, subject to the boundary condition

$$\phi(\mathbf{r}) \rightarrow 0 \quad \text{as} \quad |\mathbf{r}| \rightarrow \infty. \quad (2.12)$$

We can achieve this task by searching for a so-called Green's function, $G(\mathbf{r}, \mathbf{r}')$, that satisfies

$$\nabla^2 G(\mathbf{r}, \mathbf{r}') = \delta(\mathbf{r} - \mathbf{r}'), \quad (2.13)$$

subject to the boundary condition

$$G(\mathbf{r}, \mathbf{r}') \rightarrow 0 \quad \text{as} \quad |\mathbf{r}| \rightarrow \infty. \quad (2.14)$$

[Note that the source in Equation (2.11) is minus the source in the previously defined version of Poisson's equation, (1.36). Likewise, the Green's function (2.13) is minus the previously defined Green's function (1.41). These differences in sign are purely for the sake of convenience.] Once we have found the Green's function, the general solution to Equation (2.11), subject to the boundary condition (2.12), is given by

$$\phi(\mathbf{r}) = \int v(\mathbf{r}') G(\mathbf{r}, \mathbf{r}') dV', \quad (2.15)$$

where the integral is over all space. We can prove that this expression is indeed a solution to Equation (2.11) as follows:

$$\nabla^2 \phi(\mathbf{r}) = \int v(\mathbf{r}') \nabla^2 G(\mathbf{r}, \mathbf{r}') dV' = \int v(\mathbf{r}') \delta(\mathbf{r} - \mathbf{r}') dV' = v(\mathbf{r}). \quad (2.16)$$

Here, use has been made of Equations (1.23) and (1.24), as well as the fact that \mathbf{r} and \mathbf{r}' are independent variables. From Equation (2.14), the expression for $\phi(\mathbf{r})$ given in Equation (2.15) satisfies the boundary condition (2.12) provided that the volume integral on the right-hand side [of (2.15)] converges to a finite value.

According to Equation (1.25), a solution to Equation (2.13), subject to the boundary condition (2.14), is

$$G(\mathbf{r}, \mathbf{r}') = -\frac{1}{4\pi |\mathbf{r} - \mathbf{r}'|}. \quad (2.17)$$

Actually, we can prove that this is the only solution. Let there be two supposedly different functions, $G_1(\mathbf{r}, \mathbf{r}')$ and $G_2(\mathbf{r}, \mathbf{r}')$, that both satisfy Equation (2.13), subject to the boundary condition (2.14). Let us form the difference $G_3(\mathbf{r}, \mathbf{r}') = G_1(\mathbf{r}, \mathbf{r}') - G_2(\mathbf{r}, \mathbf{r}')$. It follows that

$$\nabla^2 G_3(\mathbf{r}, \mathbf{r}') = 0, \quad (2.18)$$

subject to the boundary condition

$$G_3(\mathbf{r}, \mathbf{r}') \rightarrow 0 \quad \text{as} \quad |\mathbf{r}| \rightarrow \infty. \quad (2.19)$$

However, as we saw in Section 2.2, the only solution to the previous two equations is

$$G_3(\mathbf{r}, \mathbf{r}') = 0 \quad (2.20)$$

for all \mathbf{r} (and \mathbf{r}'). Hence, the functions $G_1(\mathbf{r}, \mathbf{r}')$ and $G_2(\mathbf{r}, \mathbf{r}')$ are identical, and the Green's function (2.17) is unique. It follows from Equation (2.15) that the general solution to Poisson's equation, (2.11), subject to the boundary condition (2.12), is

$$\phi(\mathbf{r}) = -\frac{1}{4\pi} \int \frac{v(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} dV'. \quad (2.21)$$

Furthermore, this solution is unique.

2.4 Coulomb's Law

Coulomb's law is equivalent to the statement that the electric field $\mathbf{E}(\mathbf{r})$ generated by a point charge q' located at $\mathbf{r} = \mathbf{r}'$ is

$$\mathbf{E}(\mathbf{r}) = \frac{q'}{4\pi \epsilon_0} \frac{\mathbf{r} - \mathbf{r}'}{|\mathbf{r} - \mathbf{r}'|^3}. \quad (2.22)$$

The electric force \mathbf{F} exerted on a point charge q located at position vector \mathbf{r} is

$$\mathbf{F} = q \mathbf{E}(\mathbf{r}). \quad (2.23)$$

Hence,

$$\mathbf{F} = \frac{q q'}{4\pi \epsilon_0} \frac{\mathbf{r} - \mathbf{r}'}{|\mathbf{r} - \mathbf{r}'|^3}. \quad (2.24)$$

It follows that the electrostatic force acting between two point charges is inverse-square, central, proportional to the product of the charges, and repulsive if both charges are of the same sign.

Electric fields are superposable (see Section 1.2), which means that the electric field generated by N point charges, q_i , located at position vectors \mathbf{r}_i , for $i = 1, N$, is

$$\mathbf{E}(\mathbf{r}) = \sum_{i=1, N} \frac{q_i}{4\pi \epsilon_0} \frac{\mathbf{r} - \mathbf{r}_i}{|\mathbf{r} - \mathbf{r}_i|^3}. \quad (2.25)$$

In the continuum limit, the previous expression becomes

$$\mathbf{E}(\mathbf{r}) = \frac{1}{4\pi \epsilon_0} \int \frac{\rho(\mathbf{r}') (\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^3} dV', \quad (2.26)$$

where $\rho(\mathbf{r})$ is the charge density (i.e., the electric charge per unit volume), and the integral is over all space.

2.5 Electric Scalar Potential

It is easily demonstrated that

$$\frac{\mathbf{r} - \mathbf{r}'}{|\mathbf{r} - \mathbf{r}'|^3} = -\nabla \left(\frac{1}{|\mathbf{r} - \mathbf{r}'|} \right). \quad (2.27)$$

Hence, Equation (2.26) yields

$$\mathbf{E}(\mathbf{r}) = -\nabla \phi, \quad (2.28)$$

where

$$\phi(\mathbf{r}) = \frac{1}{4\pi \epsilon_0} \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} dV' \quad (2.29)$$

is the scalar potential. (See Section 1.3.) It follows from Equation (2.28) that

$$\nabla \times \mathbf{E} = \mathbf{0}. \quad (2.30)$$

In other words, an electric field generated by (stationary) charges is irrotational.

According to Equation (2.27), we can write Equation (2.26) in the form

$$\mathbf{E}(\mathbf{r}) = -\frac{1}{4\pi \epsilon_0} \int \rho(\mathbf{r}') \nabla \left(\frac{1}{|\mathbf{r} - \mathbf{r}'|} \right) dV'. \quad (2.31)$$

Hence,

$$\nabla \cdot \mathbf{E} = -\frac{1}{4\pi \epsilon_0} \int \rho(\mathbf{r}') \nabla^2 \left(\frac{1}{|\mathbf{r} - \mathbf{r}'|} \right) dV' = \frac{1}{\epsilon_0} \int \rho(\mathbf{r}') \delta(\mathbf{r} - \mathbf{r}') dV' = \frac{\rho(\mathbf{r})}{\epsilon_0}, \quad (2.32)$$

where use has been made of Equations (1.23)–(1.25). We deduce that

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0}, \quad (2.33)$$

which we recognize as the first Maxwell equation. (See Section 1.2.) The integral form of this equation, which follows from the divergence theorem,

$$\int_S \mathbf{E} \cdot d\mathbf{S} = \frac{1}{\epsilon_0} \int_V \rho(\mathbf{r}) dV, \quad (2.34)$$

is known as *Gauss' law*. Here, S is the bounding surface of volume V .

Equations (2.28) and (2.33) can be combined to give

$$\nabla^2 \phi = -\frac{\rho}{\epsilon_0}, \quad (2.35)$$

which we recognize as Poisson's equation, with $v(\mathbf{r}) = -\rho(\mathbf{r})/\epsilon_0$. (See Section 2.3.) Hence, Equation (2.21) yields

$$\phi(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} dV', \quad (2.36)$$

which is equivalent to Equation (2.29). Incidentally, according to the analysis of Sections 2.2 and 2.3, the previous expression represents the unique solution to Equation (2.35), subject to the boundary condition

$$\phi(\mathbf{r}) \rightarrow 0 \quad \text{as} \quad |\mathbf{r}| \rightarrow \infty. \quad (2.37)$$

2.6 Electrostatic Energy

Consider a collection of N static point charges q_i , located at position vectors \mathbf{r}_i , respectively (where i runs from 1 to N). Let us determine the electrostatic energy stored in such a collection. In other words, let us calculate the amount of work required to assemble the charges, starting from an initial state in which they are all at rest and very widely separated.

The work we would have to do against electrical forces in order to slowly move a charge q from point P to point Q is

$$W = \int_P^Q (-\mathbf{F}) \cdot d\mathbf{r} = -q \int_P^Q \mathbf{E} \cdot d\mathbf{r} = q \int_P^Q \nabla\phi \cdot d\mathbf{r} = q [\phi(Q) - \phi(P)], \quad (2.38)$$

where use has been made of Equations (2.23) and (2.28). Note that to move the charge we have to exert on it a force $-\mathbf{F}$, where \mathbf{F} is specified in Equation (2.23), in order to counteract the force exerted by the electric field. Recall that the scalar potential field generated by a point charge q , located at position \mathbf{r}' , is

$$\phi(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \frac{q}{|\mathbf{r} - \mathbf{r}'|}. \quad (2.39)$$

Let us build up our collection of charges one by one. It takes no work to bring the first charge from infinity, because there is no electric field to fight against. Let us clamp this charge in position at \mathbf{r}_1 . In order to bring the second charge into position at \mathbf{r}_2 , we have to do work against the electric field generated by the first charge. According to Equations (2.38) and Equations (2.39), this work is given by

$$W_2 = \frac{1}{4\pi\epsilon_0} \frac{q_2 q_1}{|\mathbf{r}_2 - \mathbf{r}_1|}. \quad (2.40)$$

Let us now bring the third charge into position. Because electric fields and scalar potentials are superposable, the work done while moving the third charge from infinity to \mathbf{r}_3 is simply the sum of the works done against the electric fields generated by charges 1 and 2 taken in isolation: that is,

$$W_3 = \frac{1}{4\pi\epsilon_0} \left(\frac{q_3 q_1}{|\mathbf{r}_3 - \mathbf{r}_1|} + \frac{q_3 q_2}{|\mathbf{r}_3 - \mathbf{r}_2|} \right). \quad (2.41)$$

Thus, the total work done in assembling the three charges is given by

$$W = \frac{1}{4\pi\epsilon_0} \left(\frac{q_2 q_1}{|\mathbf{r}_2 - \mathbf{r}_1|} + \frac{q_3 q_1}{|\mathbf{r}_3 - \mathbf{r}_1|} + \frac{q_3 q_2}{|\mathbf{r}_3 - \mathbf{r}_2|} \right). \quad (2.42)$$

This result can easily be generalized to N charges:

$$W = \frac{1}{4\pi\epsilon_0} \sum_{i=1, N} \sum_{j=1, i-1} \frac{q_i q_j}{|\mathbf{r}_i - \mathbf{r}_j|}. \quad (2.43)$$

The restriction that j must be less than i makes the above summation rather cumbersome. If we were to sum without restriction (other than $j \neq i$) then each pair of charges would be counted twice. It is convenient to do just this, and then to divide the result by two. Thus, we obtain

$$W = \frac{1}{2} \frac{1}{4\pi\epsilon_0} \sum_{i=1, N} \sum_{j=1, N}^{j \neq i} \frac{q_i q_j}{|\mathbf{r}_i - \mathbf{r}_j|}. \quad (2.44)$$

This expression specifies the electrostatic potential energy of a collection of point charges. We can think of this energy as the work required to bring stationary charges from infinity and assemble them in the required formation. Alternatively, it is the kinetic energy that would be released if the collection were dissolved, and the charges returned to infinity. Let us investigate how the potential energy of a collection of electric charges is stored.

Equation (2.44) can be written

$$W = \frac{1}{2} \sum_{i=1, N} q_i \phi_i, \quad (2.45)$$

where

$$\phi_i = \frac{1}{4\pi\epsilon_0} \sum_{j=1, N}^{j \neq i} \frac{q_j}{|\mathbf{r}_i - \mathbf{r}_j|} \quad (2.46)$$

is the scalar potential experienced by the i th charge due to the other charges in the distribution.

Let us now consider the potential energy of a continuous charge distribution. It is tempting to write

$$W = \frac{1}{2} \int \rho(\mathbf{r}) \phi(\mathbf{r}) dV, \quad (2.47)$$

by analogy with Equations (2.45) and (2.46), where

$$\phi(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} dV' \quad (2.48)$$

is the familiar scalar potential generated by a continuous charge distribution of charge density $\rho(\mathbf{r})$. Let us try this ansatz out. We know from Equation (2.33) that

$$\rho = \epsilon_0 \nabla \cdot \mathbf{E}, \quad (2.49)$$

so Equation (2.47) can be written

$$W = \frac{\epsilon_0}{2} \int \phi \nabla \cdot \mathbf{E} dV. \quad (2.50)$$

Making use of the vector identity,

$$\nabla \cdot (\mathbf{E} \phi) \equiv \phi \nabla \cdot \mathbf{E} + \mathbf{E} \cdot \nabla \phi, \quad (2.51)$$

as well as the fact that $\nabla \phi = -\mathbf{E}$, we obtain

$$W = \frac{\epsilon_0}{2} \left[\int \nabla \cdot (\mathbf{E} \phi) dV + \int E^2 dV \right]. \quad (2.52)$$

Application of the divergence theorem gives

$$W = \frac{\epsilon_0}{2} \left(\int_S \phi \mathbf{E} \cdot d\mathbf{S} + \int_V E^2 dV \right), \quad (2.53)$$

where V is some volume that contains all of the charges, and S is its bounding surface. Let us assume that V is a sphere, centered on the origin, and let us take the limit in which the radius r of this sphere goes to infinity. We know that, in general, the electric field a large distance r from a bounded charge distribution looks like the field of a point charge, and, therefore, falls off like $1/r^2$. Likewise, the potential falls off like $1/r$. However, the surface area of a sphere of radius r increases like r^2 . Hence, it is clear that, in the limit as $r \rightarrow \infty$, the surface integral in Equation (2.53) falls off like $1/r$, and is consequently zero. Thus, Equation (2.53) reduces to

$$W = \frac{\epsilon_0}{2} \int E^2 dV, \quad (2.54)$$

where the integral is over all space. This expression implies that the potential energy of a continuous charge distribution is stored in the electric field generated by the distribution, assuming that this field possesses the energy density (see Section 1.9)

$$U = \frac{\epsilon_0}{2} E^2. \quad (2.55)$$

We can easily check that Equation (2.54) is correct. Suppose that we have an amount of charge Q that is uniformly distributed within a sphere of radius a , centered on the origin. Let us imagine building up this charge distribution from a succession of thin spherical layers of infinitesimal thickness. At each stage, we gather a small amount of charge dq from infinity, and spread it over the surface of the sphere in a thin layer extending from r to $r + dr$. We continue this process until

the final radius of the sphere is a . If $q(r)$ is the sphere's charge when it has attained radius r then the work done in bringing a charge dq to its surface is

$$dW = \frac{1}{4\pi\epsilon_0} \frac{q(r) dq}{r}. \quad (2.56)$$

This follows from Equation (2.40), because the electric field generated outside a spherical charge distribution is the same as that of a point charge $q(r)$ located at its geometric center ($r = 0$). If the uniform charge density of the sphere is ρ then

$$q(r) = \frac{4\pi}{3} r^3 \rho, \quad (2.57)$$

and

$$dq = 4\pi r^2 \rho dr. \quad (2.58)$$

Thus, Equation (2.56) becomes

$$dW = \frac{4\pi}{3\epsilon_0} \rho^2 r^4 dr. \quad (2.59)$$

The total work needed to build up the sphere from nothing to radius a is plainly

$$W = \frac{4\pi}{3\epsilon_0} \rho^2 \int_0^a r^4 dr = \frac{4\pi}{15\epsilon_0} \rho^2 a^5. \quad (2.60)$$

This can also be written in terms of the total charge $Q = (4\pi/3) a^3 \rho$ as

$$W = \frac{3}{5} \frac{Q^2}{4\pi\epsilon_0 a}. \quad (2.61)$$

Now that we have evaluated the potential energy of a spherical charge distribution by the direct method, let us work it out using Equation (2.54). We shall assume that the electric field is both radial and spherically symmetric, so that $\mathbf{E} = E_r(r) \mathbf{e}_r$. Application of Gauss' law,

$$\int_S \mathbf{E} \cdot d\mathbf{S} = \frac{1}{\epsilon_0} \int_V \rho dV, \quad (2.62)$$

where V is a sphere of radius r , centered on the origin, gives

$$E_r(r) = \frac{Q}{4\pi\epsilon_0 a^3} r \quad (2.63)$$

for $r < a$, and

$$E_r(r) = \frac{Q}{4\pi\epsilon_0 r^2} \quad (2.64)$$

for $r \geq a$. Equations (2.54), (2.63), and (2.64) yield

$$W = \frac{Q^2}{8\pi\epsilon_0} \left(\frac{1}{a^6} \int_0^a r^4 dr + \int_a^\infty \frac{dr}{r^2} \right), \quad (2.65)$$

which reduces to

$$W = \frac{Q^2}{8\pi \epsilon_0 a} \left(\frac{1}{5} + 1 \right) = \frac{3}{5} \frac{Q^2}{4\pi \epsilon_0 a}. \quad (2.66)$$

Thus, Equation (2.54) gives the correct answer.

The reason that we have checked Equation (2.54) so carefully is that, on close inspection, it is found to be inconsistent with Equation (2.45), from which it was supposedly derived. For instance, the energy given by Equation (2.54) is manifestly positive definite, whereas the energy given by Equation (2.45) can be negative (it is certainly negative for a collection of two point charges of opposite sign). The inconsistency was introduced into our analysis when we replaced Equation (2.46) by Equation (2.48). In Equation (2.46), the self-interaction of the i th charge with its own electric field is specifically excluded, whereas it is included in Equation (2.48). Thus, the potential energies (2.45) and (2.54) are different because in the former we start from ready-made point charges, whereas in the latter we build up the whole charge distribution from scratch. Consequently, if we were to calculate the potential energy of a point charge distribution using Equation (2.54) then we would obtain the energy (2.45) plus the energy required to assemble the point charges. However, the latter energy is infinite. To see this, let us suppose, for the sake of argument, that our point charges actually consist of charge uniformly distributed in small spheres of radius b . According to Equation (2.61), the energy required to assemble the i th point charge is

$$W_i = \frac{3}{5} \frac{q_i^2}{4\pi \epsilon_0 b}. \quad (2.67)$$

We can think of this as the self-energy of the i th charge. Thus, we can write

$$W = \frac{\epsilon_0}{2} \int E^2 dV = \frac{1}{2} \sum_{i=1,N} q_i \phi_i + \sum_{i=1,N} W_i \quad (2.68)$$

which enables us to reconcile Equations (2.45) and (2.54). Unfortunately, if our point charges really are point charges then $b \rightarrow 0$, and the self-energy of each charge becomes infinite. Thus, the potential energies specified by Equations (2.45) and (2.54) differ by an infinite amount. We are forced to the conclusion that the idea of locating electrostatic potential energy in the electric field is inconsistent with the existence of point charges. One way out of this difficulty would be to say that elementary charges, such as electrons, are not point objects, but instead have finite spatial extents. Regrettably, there is no experimental evidence to back up this assertion. Alternatively, we could say that our classical theory of electromagnetism breaks down on very small length-scales due to quantum effects. Unfortunately, the quantum mechanical version of electromagnetism (which is called quantum electrodynamics) suffers from the same infinities in the self-energies of charged particles as the classical version. There is a prescription, called renormalization, for steering round these infinities, and getting finite answers that agree with experimental data to extraordinary accuracy. However, nobody really understands why this prescription works.

2.7 Electric Dipoles

Consider a charge q located at position vector \mathbf{r}' , and a charge $-q$ located at position vector $\mathbf{r}' - \mathbf{d}$. In the limit that $|\mathbf{d}| \rightarrow 0$, but $|q| |\mathbf{d}|$ remains finite, this combination of charges constitutes an *electric*

dipole, of dipole moment

$$\mathbf{p} = q \mathbf{d}, \quad (2.69)$$

located at position vector \mathbf{r}' . We have seen that the electric field generated at point \mathbf{r} by an electric charge q located at point \mathbf{r}' is

$$\mathbf{E}(\mathbf{r}) = -\nabla \left(\frac{q}{4\pi \epsilon_0} \frac{1}{|\mathbf{r} - \mathbf{r}'|} \right). \quad (2.70)$$

Hence, the electric field generated at point \mathbf{r} by an electric dipole of moment \mathbf{p} located at point \mathbf{r}' is

$$\mathbf{E}(\mathbf{r}) = -\nabla \left(\frac{q}{4\pi \epsilon_0} \frac{1}{|\mathbf{r} - \mathbf{r}'|} - \frac{q}{4\pi \epsilon_0} \frac{1}{|\mathbf{r} - \mathbf{r}' + \mathbf{d}|} \right). \quad (2.71)$$

However, in the limit that $|\mathbf{d}| \rightarrow 0$,

$$\frac{1}{|\mathbf{r} - \mathbf{r}' + \mathbf{d}|} \simeq \frac{1}{|\mathbf{r} - \mathbf{r}'|} - \frac{\mathbf{d} \cdot (\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^3}. \quad (2.72)$$

Thus, the electric field due to the dipole becomes

$$\mathbf{E}(\mathbf{r}) = -\nabla \left(\frac{1}{4\pi \epsilon_0} \frac{\mathbf{p} \cdot (\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^3} \right). \quad (2.73)$$

It follows from Equation (2.28) that the scalar electric potential due to the dipole is

$$\phi(\mathbf{r}) = \frac{1}{4\pi \epsilon_0} \frac{\mathbf{p} \cdot (\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^3} = -\frac{1}{4\pi \epsilon_0} \mathbf{p} \cdot \nabla \left(\frac{1}{|\mathbf{r} - \mathbf{r}'|} \right) = \frac{1}{4\pi \epsilon_0} \mathbf{p} \cdot \nabla' \left(\frac{1}{|\mathbf{r} - \mathbf{r}'|} \right). \quad (2.74)$$

(Here, ∇' is a gradient operator expressed in terms of the components of \mathbf{r}' , but independent of the components of \mathbf{r} .) Finally, because electric fields are superposable, the electric potential due to a volume distribution of electric dipoles is

$$\phi(\mathbf{r}) = \frac{1}{4\pi \epsilon_0} \int \mathbf{P}(\mathbf{r}') \cdot \nabla' \left(\frac{1}{|\mathbf{r} - \mathbf{r}'|} \right) dV', \quad (2.75)$$

where $\mathbf{P}(\mathbf{r})$ is the *electric polarization* (i.e., the electric dipole moment per unit volume), and the integral is over all space.

2.8 Charge Sheets and Dipole Sheets

The electric potential due to a *charge sheet* (i.e., a charge distribution that is confined to a surface) can be obtained from Equation (2.36) by replacing $\rho(\mathbf{r}') dV'$ with $\sigma(\mathbf{r}') dS'$. Here, $\sigma(\mathbf{r}')$ is the surface charge density (i.e., the charge per unit area) at position \mathbf{r}' . We obtain

$$\phi(\mathbf{r}) = \frac{1}{4\pi \epsilon_0} \int_S \frac{\sigma(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} dS', \quad (2.76)$$

where dS' is an element of the surface S , on which the charges are distributed, located at position vector \mathbf{r}' . Incidentally, we are assuming that the distribution is negligibly thin in the direction

normal to the surface. As is well known, application of Gauss' law to a thin pill-box aligned with an element of S tells us that there is a discontinuity in the normal electric field across the sheet. In fact,

$$(\mathbf{E}_2 - \mathbf{E}_1) \cdot \mathbf{n} = \frac{\sigma}{\epsilon_0}, \quad (2.77)$$

where \mathbf{n} is a unit normal at a given point on the sheet, \mathbf{E}_1 and \mathbf{E}_2 are the electric fields immediately to either side of the sheet at this point (\mathbf{E}_2 being the field on the side toward which \mathbf{n} is directed), and σ is the local charge density. As is also well known, integration of Equation (2.30) around a small loop that straddles the sheet reveals that there is no discontinuity in the tangential electric field across the sheet.

The electric potential due to a *dipole sheet* (i.e., a dipole distribution that is confined to a surface) can be obtained from Equation (2.75) by replacing $\mathbf{P}(\mathbf{r}') dV'$ with $\mathbf{D}(\mathbf{r}') dS'$. Here, $\mathbf{D}(\mathbf{r}')$ is the surface dipole density (i.e., the dipole moment per unit area) at position \mathbf{r}' . We obtain

$$\phi(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \int_S \mathbf{D}(\mathbf{r}') \cdot \nabla' \left(\frac{1}{|\mathbf{r} - \mathbf{r}'|} \right) dS'. \quad (2.78)$$

We are again assuming that the distribution is negligibly thin in the direction normal to the surface S on which the dipoles are distributed. Suppose that $\mathbf{D}(\mathbf{r}) = D(\mathbf{r}) \mathbf{n}$, where \mathbf{n} is a unit normal to the sheet at position vector \mathbf{r} . In other words, suppose that the constituent dipoles are all locally perpendicular to S . It follows that

$$\phi(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \int_S D(\mathbf{r}') \mathbf{n}' \cdot \nabla' \left(\frac{1}{|\mathbf{r} - \mathbf{r}'|} \right) dS', \quad (2.79)$$

where \mathbf{n}' is the unit normal to S at position \mathbf{r}' . Now, according to Equation (2.28) and (2.76), the normal electric field generated by a charge sheet is

$$\mathbf{E} \cdot \mathbf{n} = -\mathbf{n} \cdot \nabla \phi = -\frac{1}{4\pi\epsilon_0} \int_S \sigma(\mathbf{r}') \mathbf{n} \cdot \nabla \left(\frac{1}{|\mathbf{r} - \mathbf{r}'|} \right) dS' = \frac{1}{4\pi\epsilon_0} \int_S \sigma(\mathbf{r}') \mathbf{n} \cdot \nabla' \left(\frac{1}{|\mathbf{r} - \mathbf{r}'|} \right) dS'. \quad (2.80)$$

A comparison between Equations (2.77), (2.79), and (2.80) reveals that there is a discontinuity of the electric potential across a dipole sheet. In fact,

$$\phi_2 - \phi_1 = \frac{D}{\epsilon_0}, \quad (2.81)$$

where ϕ_1 and ϕ_2 are the potentials immediately to either side of a given point on the sheet [ϕ_2 being the potential on the side toward which \mathbf{n} (and, hence, \mathbf{D}) is directed], and D is the magnitude of the local dipole density. [We can neglect the distinction between \mathbf{n}' in Equation (2.79), and \mathbf{n} in Equation (2.80), because the discontinuous part of the electric field due to a current sheet (as well as the discontinuous part of the potential due to a dipole sheet) is generated locally.] Incidentally, there is no discontinuity in the normal electric field across a dipole sheet because the local charge density is zero. Hence, although the potential is discontinuous across a dipole sheet, the normal derivative of the potential is continuous. Likewise, there is no discontinuity in

the electric potential across a charge sheet because the local dipole density is zero. Thus, although the normal derivative of the potential is discontinuous across a charge sheet (because the normal electric field is discontinuous), the potential itself is continuous.

As an example, consider a charge sheet of uniform charge density σ that corresponds to the plane $x = 0$. Gauss' law, in combination with symmetry arguments, reveals that

$$\mathbf{E} = \begin{cases} (\sigma/2 \epsilon_0) \mathbf{e}_x & x > 0 \\ -(\sigma/2 \epsilon_0) \mathbf{e}_x & x < 0 \end{cases}, \quad (2.82)$$

which is in accordance with Equation (2.77). It follows from Equation (2.28), and the requirement that the electric potential be continuous across the sheet, that

$$\phi = \begin{cases} -(\sigma/2 \epsilon_0) x & x > 0 \\ (\sigma/2 \epsilon_0) x & x < 0 \end{cases}. \quad (2.83)$$

Consider, now, a dipole sheet of uniform dipole density $\mathbf{D} = D \mathbf{e}_x$ that corresponds to the plane $x = 0$. We can think of this sheet as a combination of two charge sheets: the first, of charge density σ , located at $x = d/2$, and the second, of charge density $-\sigma$, located at $x = -d/2$. In the limit $d \rightarrow 0$, but $d \sigma \rightarrow D$, the two charge sheets are equivalent to the dipole sheet. It follows, from the previous two equations, that the electric field, and potential, generated by the dipole sheet are

$$\mathbf{E} = \begin{cases} \mathbf{0} & x > 0 \\ \mathbf{0} & x < 0 \end{cases}, \quad (2.84)$$

and

$$\phi = \begin{cases} D/(2 \epsilon_0) & x > 0 \\ -D/(2 \epsilon_0) & x < 0 \end{cases}, \quad (2.85)$$

respectively. The latter equation is in accordance with Equation (2.81). Note that, although the dipole sheet does not generate an external electric field, its internal field accelerates any charge that crosses the sheet. In fact, assuming that $D > 0$, a positive charge gains energy by crossing the sheet from the region $x > 0$ to the region $x < 0$.

As a second example, consider a charge sheet of uniform charge density σ that lies on the surface of a sphere, of radius a , centered on the origin. Gauss' law, in combination with symmetry arguments, reveals that

$$\mathbf{E} = \begin{cases} \mathbf{0} & r < a \\ (\sigma/\epsilon_0) (a/r)^2 \mathbf{e}_r & r > a \end{cases}, \quad (2.86)$$

where r is a spherical polar coordinate. The above expression is again in accordance with Equation (2.77). It follows from Equation (2.28), and the requirement that the electric potential be continuous across the sheet (as well as zero at infinity), that

$$\phi = \begin{cases} \sigma a/\epsilon_0 & r < a \\ (\sigma a/\epsilon_0) (a/r) & r > a \end{cases}. \quad (2.87)$$

Consider, now, a dipole sheet of uniform dipole density $\mathbf{D} = D \mathbf{e}_r$ that lies on the surface of the sphere. We can think of this sheet as a combination of two charge sheets: the first, of charge density

$\sigma [a/(a+d/2)]^2$, located at $r = a+d/2$, and the second, of charge density $-\sigma [a/(a-d/2)]^2$, located at $r = a - d/2$. (The factors $[a/(a \pm d/2)]^2$ are needed to ensure that both sheets contain equal and opposite net charge.) In the limit $d \rightarrow 0$, but $d\sigma \rightarrow D$, the two charge sheets are equivalent to the dipole sheet. It follows, from the previous two equations, that the electric field, and potential, generated by the dipole sheet are

$$\mathbf{E} = \begin{cases} \mathbf{0} & r < a \\ \mathbf{0} & r > a \end{cases}, \quad (2.88)$$

and

$$\phi = \begin{cases} -D/\epsilon_0 & r < a \\ 0 & r > a \end{cases}, \quad (2.89)$$

respectively. The latter equation is in accordance with Equation (2.81). As before, the dipole sheet does not generate an external electric field, but its internal field is capable of accelerating a charge that crosses the sheet.

2.9 Green's Theorem

Consider the vector identity

$$\nabla \cdot (\psi \nabla \phi) \equiv \psi \nabla^2 \phi + \nabla \psi \cdot \nabla \phi, \quad (2.90)$$

where $\phi(\mathbf{r})$ and $\psi(\mathbf{r})$ are two arbitrary (but differentiable) vector fields. We can also write

$$\nabla \cdot (\phi \nabla \psi) \equiv \phi \nabla^2 \psi + \nabla \phi \cdot \nabla \psi. \quad (2.91)$$

Forming the difference between the previous two equation, we get

$$\nabla \cdot (\psi \nabla \phi - \phi \nabla \psi) = \psi \nabla^2 \phi - \phi \nabla^2 \psi. \quad (2.92)$$

Finally, integrating this expression over some volume V , bounded by the closed surface S , and making use of the divergence theorem, we obtain

$$\int_V (\psi \nabla^2 \phi - \phi \nabla^2 \psi) dV = \int_S (\psi \nabla \phi - \phi \nabla \psi) \cdot d\mathbf{S}. \quad (2.93)$$

This result is known as *Green's theorem*.

Changing the variable of integration, the above result can be rewritten

$$\int_V \left[\psi(\mathbf{r}') \nabla'^2 \phi(\mathbf{r}') - \phi(\mathbf{r}') \nabla'^2 \psi(\mathbf{r}') \right] dV' = \int_S \left[\psi(\mathbf{r}') \frac{\partial \phi(\mathbf{r}')}{\partial n'} - \phi(\mathbf{r}') \frac{\partial \psi(\mathbf{r}')}{\partial n'} \right] dS', \quad (2.94)$$

where $\partial \phi(\mathbf{r}')/\partial n'$ is shorthand for $\mathbf{n}' \cdot \nabla' \phi(\mathbf{r}')$, et cetera. Suppose that $\phi(\mathbf{r})$ is a solution to Poisson's equation,

$$\nabla^2 \phi = -\frac{\rho}{\epsilon_0}, \quad (2.95)$$

associated with some charge distribution, $\rho(\mathbf{r})$, that extends over all space, subject to the boundary condition

$$\phi(\mathbf{r}) \rightarrow 0 \quad \text{as} \quad |\mathbf{r}| \rightarrow \infty. \quad (2.96)$$

Suppose, further, that

$$\psi(\mathbf{r}, \mathbf{r}') = \frac{1}{4\pi |\mathbf{r} - \mathbf{r}'|}. \quad (2.97)$$

It follows from Equation (1.25), and symmetry, that

$$\nabla^2 \psi = \nabla'^2 \psi = -\delta(\mathbf{r} - \mathbf{r}'). \quad (2.98)$$

The previous five equations can be combined to give

$$\begin{aligned} \int_V \phi(\mathbf{r}') \delta(\mathbf{r} - \mathbf{r}') dV' &= \frac{1}{4\pi \epsilon_0} \int_V \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} dV' + \frac{1}{4\pi \epsilon_0} \int_S \frac{\sigma(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} dS' \\ &+ \frac{1}{4\pi \epsilon_0} \int_S D(\mathbf{r}') \frac{\partial}{\partial n'} \left(\frac{1}{|\mathbf{r} - \mathbf{r}'|} \right) dS', \end{aligned} \quad (2.99)$$

where

$$\sigma = \epsilon_0 \mathbf{n} \cdot \nabla \phi = -\epsilon_0 \mathbf{n} \cdot \mathbf{E}, \quad (2.100)$$

$$D = -\epsilon_0 \phi. \quad (2.101)$$

It follows, by comparison with Equations (2.36), (2.76), and (2.79), that the three terms on the right-hand side of Equation (2.99) are the electric potential generated by the charges distributed within S , the potential generated by the surface charge distribution, $\sigma(\mathbf{r})$, on S , and the potential generated by the surface dipole distribution, $D(\mathbf{r})$, on S , respectively.

Suppose that the point \mathbf{r} lies within S . In this case, Equation (2.99) yields

$$\phi(\mathbf{r}) = \frac{1}{4\pi \epsilon_0} \int_V \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} dV' + \frac{1}{4\pi \epsilon_0} \int_S \frac{\sigma(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} dS' + \frac{1}{4\pi \epsilon_0} \int_S D(\mathbf{r}') \frac{\partial}{\partial n'} \left(\frac{1}{|\mathbf{r} - \mathbf{r}'|} \right) dS'. \quad (2.102)$$

However, we know that the general solution to Equation (2.95), subject to the boundary condition (2.96), is

$$\phi(\mathbf{r}) = \frac{1}{4\pi \epsilon_0} \int_V \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} dV' + \frac{1}{4\pi \epsilon_0} \int_{\bar{V}} \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} dV'. \quad (2.103)$$

Here, V denotes the region of space lying within the closed surface S , whereas \bar{V} denotes the region lying outside S . A comparison of the previous two equations reveals that, for a point \mathbf{r} lying within S ,

$$\frac{1}{4\pi \epsilon_0} \int_{\bar{V}} \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} dV' = \frac{1}{4\pi \epsilon_0} \int_S \frac{\sigma(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} dS' + \frac{1}{4\pi \epsilon_0} \int_S D(\mathbf{r}') \frac{\partial}{\partial n'} \left(\frac{1}{|\mathbf{r} - \mathbf{r}'|} \right) dS'. \quad (2.104)$$

In other words, the electric potential (and, hence, the electric field) generated within S by the charges external to S is equivalent to that generated by the charge sheet $\sigma(\mathbf{r})$, and the dipole sheet

$D(\mathbf{r})$, distributed on S . Furthermore, because σ depends on the normal derivative of the potential at S , whereas D depends on the potential at S , it follows that we can completely determine the potential within S once we know the distribution of charges within S , and the values of the potential and its normal derivative on S . In fact, this is an overstatement, because the potential and its normal derivative are not independent of one another, but are related via Poisson's equation. In other words, a knowledge of the potential on S also implies a knowledge of its normal derivative, and vice versa. Hence, we can, actually, determine the potential within S from a knowledge of the distribution of charges inside S , and the distribution of either the potential, or its normal derivative, on S . The specification of the potential on S is known as a *Dirichlet boundary condition*. On the other hand, the specification of the normal derivative of the potential on S is called a *Neumann boundary condition*.

Suppose that the point \mathbf{r} lies outside S . In this case, Equation (2.99) yields

$$0 = \frac{1}{4\pi\epsilon_0} \int_V \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} dV' + \frac{1}{4\pi\epsilon_0} \int_S \frac{\sigma(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} dS' + \frac{1}{4\pi\epsilon_0} \int_S D(\mathbf{r}') \frac{\partial}{\partial n'} \left(\frac{1}{|\mathbf{r} - \mathbf{r}'|} \right) dS'. \quad (2.105)$$

In other words, outside S , the electric potential generated by the surface charge distribution $\sigma(\mathbf{r})$, combined with that generated by the surface dipole distribution $D(\mathbf{r})$, completely cancels out the electric potential (and, hence, the electric field) produced by the charges distributed within S . As an example of this type of cancellation, suppose that S is a spherical surface of radius a , centered on the origin. Within S , let there be a single charge q , located at the origin. The electric field and potential generated by this charge are

$$\mathbf{E} = \frac{q}{4\pi\epsilon_0 r^2} \mathbf{e}_r, \quad (2.106)$$

and

$$\phi = \frac{q}{4\pi\epsilon_0 r}, \quad (2.107)$$

respectively. It follows from Equations (2.100) and (2.101) that the densities of the charge and dipole sheets at $r = a$ that are needed to cancel out the effect of the central charge in the region $r > a$ are

$$\sigma = -\frac{q}{4\pi a^2}, \quad (2.108)$$

$$D = -\frac{q}{4\pi a}, \quad (2.109)$$

respectively. Making use of Equations (2.86)–(2.89), and (2.106)–(2.109), the electric field and potential generated by the combination of the charge at the origin, the charge sheet at $r = a$, and the dipole sheet at $r = a$, are

$$\mathbf{E} = \begin{cases} [q/(4\pi\epsilon_0 r^2)] \mathbf{e}_r & r < a \\ \mathbf{0} & r > a \end{cases}, \quad (2.110)$$

and

$$\phi = \begin{cases} q/(4\pi\epsilon_0 r) & r < a \\ 0 & r > a \end{cases}, \quad (2.111)$$

respectively. We can see that the charge and dipole sheet at $r = a$ do not affect the electric field, or the potential, due to the central charge in the region $r < a$, but completely cancel out this charge's field and potential in the region $r > a$.

2.10 Boundary Value Problems

Consider a volume V bounded by a surface S . Suppose that we wish to solve Poisson's equation,

$$\nabla^2 \phi = -\frac{\rho}{\epsilon_0}, \quad (2.112)$$

throughout V , subject to given Dirichlet or Neumann boundary conditions on S . The charge density distribution, $\rho(\mathbf{r})$, is assumed to be known throughout V . This type of problem is called a *boundary value problem*.

Similarly to the approach taken in Section 2.3, we can solve Poisson's equation by means of a Green's function, $G(\mathbf{r}, \mathbf{r}')$, that satisfies

$$\nabla'^2 G(\mathbf{r}, \mathbf{r}') = \delta(\mathbf{r} - \mathbf{r}'). \quad (2.113)$$

In fact, it follows from Equation (1.25) [because $\nabla^2(|\mathbf{r} - \mathbf{r}'|^{-1}) = \nabla'^2(|\mathbf{r} - \mathbf{r}'|^{-1})$, by symmetry] that

$$G(\mathbf{r}, \mathbf{r}') = -\frac{1}{4\pi|\mathbf{r} - \mathbf{r}'|} + F(\mathbf{r}, \mathbf{r}'), \quad (2.114)$$

where

$$\nabla'^2 F(\mathbf{r}, \mathbf{r}') = 0 \quad (2.115)$$

throughout V . Here, the function $F(\mathbf{r}, \mathbf{r}')$ is chosen in such a manner as to satisfy the boundary conditions on S . Making use of Green's theorem, (2.94), where $\psi(\mathbf{r}') = G(\mathbf{r}, \mathbf{r}')$, we find that

$$\phi(\mathbf{r}) = -\frac{1}{\epsilon_0} \int_V G(\mathbf{r}, \mathbf{r}') \rho(\mathbf{r}') dV' - \int_S \left[G(\mathbf{r}, \mathbf{r}') \frac{\partial \phi(\mathbf{r}')}{\partial n'} - \phi(\mathbf{r}') \frac{\partial G(\mathbf{r}, \mathbf{r}')}{\partial n'} \right] dS'. \quad (2.116)$$

Here, use has been made of Equations (1.23), (2.112), and (2.113). Note, incidentally, that the divergence theorem, combined with Equation (2.113), yields

$$\int_S \frac{\partial G(\mathbf{r}, \mathbf{r}')}{\partial n'} dS' = 1. \quad (2.117)$$

Consider the Dirichlet problem in which $\phi(\mathbf{r})$ is known on S , but $\partial\phi(\mathbf{r})/\partial n$ is unknown. We can construct an appropriate Green's function for this problem, $G_D(\mathbf{r}, \mathbf{r}')$, where

$$\nabla'^2 G_D(\mathbf{r}, \mathbf{r}') = \delta(\mathbf{r} - \mathbf{r}'), \quad (2.118)$$

by choosing the function $F(\mathbf{r}, \mathbf{r}')$ in Equation (2.114) in such a manner that

$$G_D(\mathbf{r}, \mathbf{r}') = 0 \quad (2.119)$$

when \mathbf{r}' lies on S . It then follows from Equation (2.116) that

$$\phi(\mathbf{r}) = -\frac{1}{\epsilon_0} \int_V G_D(\mathbf{r}, \mathbf{r}') \rho(\mathbf{r}') dV' + \int_S \phi(\mathbf{r}') \frac{\partial G_D(\mathbf{r}, \mathbf{r}')}{\partial n'} dS'. \quad (2.120)$$

Hence, the potential $\phi(\mathbf{r})$ is specified in terms of integrals over known functions throughout V and on S .

It is possible to prove that the Dirichlet Green's function is symmetric with respect to its arguments. In other words,

$$G_D(\mathbf{r}, \mathbf{r}') = G_D(\mathbf{r}', \mathbf{r}). \quad (2.121)$$

Making use of Green's theorem, (2.94), where $\psi(\mathbf{r}') = G_D(\mathbf{r}, \mathbf{r}')$ and $\phi(\mathbf{r}') = G_D(\mathbf{r}'', \mathbf{r}')$, we find that

$$\begin{aligned} \int_V [G_D(\mathbf{r}, \mathbf{r}') \nabla'^2 G_D(\mathbf{r}'', \mathbf{r}') - G_D(\mathbf{r}'', \mathbf{r}') \nabla'^2 G_D(\mathbf{r}, \mathbf{r}')] dV' \\ = \int_S \left[G_D(\mathbf{r}, \mathbf{r}') \frac{\partial G_D(\mathbf{r}'', \mathbf{r}')}{\partial n'} - G_D(\mathbf{r}'', \mathbf{r}') \frac{\partial G_D(\mathbf{r}, \mathbf{r}')}{\partial n'} \right] dS'. \end{aligned} \quad (2.122)$$

However, by definition, $\nabla'^2 G_D(\mathbf{r}, \mathbf{r}') = \delta(\mathbf{r} - \mathbf{r}')$, $\nabla'^2 G_D(\mathbf{r}'', \mathbf{r}') = \delta(\mathbf{r}'' - \mathbf{r}')$, and $G_D(\mathbf{r}, \mathbf{r}') = G_D(\mathbf{r}'', \mathbf{r}') = 0$ when \mathbf{r}' lies on S . Hence,

$$\int_V [G_D(\mathbf{r}, \mathbf{r}') \delta(\mathbf{r}'' - \mathbf{r}') - G_D(\mathbf{r}'', \mathbf{r}') \delta(\mathbf{r} - \mathbf{r}')] dV' = 0, \quad (2.123)$$

which yields

$$G_D(\mathbf{r}, \mathbf{r}'') = G_D(\mathbf{r}'', \mathbf{r}). \quad (2.124)$$

It is also possible to demonstrate that the Dirichlet Green's function is unique. Proceeding in the usual fashion, suppose that there are two different functions, $G_1(\mathbf{r}, \mathbf{r}')$ and $G_2(\mathbf{r}, \mathbf{r}')$, that both satisfy Equations (2.118) and (2.119). It follows that $G_3(\mathbf{r}, \mathbf{r}') = G_1(\mathbf{r}, \mathbf{r}') - G_2(\mathbf{r}, \mathbf{r}')$ satisfies

$$\nabla'^2 G_3(\mathbf{r}, \mathbf{r}') = 0 \quad (2.125)$$

throughout V , subject to the boundary condition

$$G_3(\mathbf{r}, \mathbf{r}') = 0 \quad (2.126)$$

when \mathbf{r}' lies on S . However, we saw in Section 2.2 that the only solution to this problem is $G_3(\mathbf{r}, \mathbf{r}') = 0$ for \mathbf{r}' in V or on S . Hence, the functions $G_1(\mathbf{r}, \mathbf{r}')$ and $G_2(\mathbf{r}, \mathbf{r}')$ are identical, and the Dirichlet Green's function is unique. It follows that the potential specified in Equation (2.120) is also unique.

Consider the Neumann problem in which $\partial\phi(\mathbf{r})/\partial n$ is known on S , but $\phi(\mathbf{r})$ is unknown. In this case, the obvious ansatz, $\partial G_N(\mathbf{r}, \mathbf{r}')/\partial n' = 0$ when \mathbf{r}' lies on S , is incorrect, because it is inconsistent with the constraint (2.117). The simplest ansatz that works is a choice of $F(\mathbf{r}, \mathbf{r}')$ in Equation (2.114) such that

$$\frac{\partial G_N(\mathbf{r}, \mathbf{r}')}{\partial n'} = 1 \int_S dS' \quad (2.127)$$

when \mathbf{r}' lies on S . Hence, Equation (2.116) yields

$$\phi(\mathbf{r}) = \langle \phi \rangle_S - \frac{1}{\epsilon_0} \int_V G_N(\mathbf{r}, \mathbf{r}') \rho(\mathbf{r}') dV' - \int_S G_N(\mathbf{r}, \mathbf{r}') \frac{\partial \phi(\mathbf{r}')}{\partial n'} dS', \quad (2.128)$$

where $\langle \phi \rangle_S = \int_S \phi(\mathbf{r}') dS' / \int_S dS'$ is the average value of the potential on S . This average value can be absorbed into the arbitrary constant that can always be added to a scalar potential. Thus, the potential is again specified in terms of integrals over known functions throughout V and on S .

It is possible to prove that the Neumann Green's function can be chosen in such a manner that it is symmetric with respect to its arguments. In other words,

$$G_N(\mathbf{r}, \mathbf{r}') = G_N(\mathbf{r}', \mathbf{r}). \quad (2.129)$$

Consider a Neumann Green's function that is not symmetric with respect to its arguments. That is, an asymmetric function $G_N(\mathbf{r}, \mathbf{r}')$ which satisfies

$$\nabla'^2 G_N(\mathbf{r}, \mathbf{r}') = \delta(\mathbf{r} - \mathbf{r}') \quad (2.130)$$

and

$$\frac{\partial G_N(\mathbf{r}, \mathbf{r}')}{\partial n'} = 1 \left/ \int_S dS' \right. \quad (2.131)$$

when \mathbf{r}' lies on S . Consider

$$\tilde{G}_N(\mathbf{r}, \mathbf{r}') = G_N(\mathbf{r}, \mathbf{r}') + F(\mathbf{r}), \quad (2.132)$$

where $F(\mathbf{r})$ is arbitrary. It follows that

$$\nabla'^2 \tilde{G}_N(\mathbf{r}, \mathbf{r}') = \delta(\mathbf{r} - \mathbf{r}') \quad (2.133)$$

and

$$\frac{\partial \tilde{G}_N(\mathbf{r}, \mathbf{r}')}{\partial n'} = 1 \left/ \int_S dS' \right. \quad (2.134)$$

when \mathbf{r}' lies on S . Hence, $\tilde{G}_N(\mathbf{r}, \mathbf{r}')$ is also a valid Neumann Green's function. Making use of Green's theorem, (2.94), where $\psi(\mathbf{r}') = \tilde{G}_N(\mathbf{r}, \mathbf{r}')$ and $\phi(\mathbf{r}') = \tilde{G}_N(\mathbf{r}'', \mathbf{r}')$, we find that

$$\tilde{G}_N(\mathbf{r}, \mathbf{r}'') - \tilde{G}_N(\mathbf{r}'', \mathbf{r}) = \int_S [\tilde{G}_N(\mathbf{r}, \mathbf{r}') - \tilde{G}_N(\mathbf{r}'', \mathbf{r}')] dS' \left/ \int_S dS' \right. . \quad (2.135)$$

We deduce that $\tilde{G}_N(\mathbf{r}, \mathbf{r}')$ is symmetric provided that $\int_S \tilde{G}_N(\mathbf{r}, \mathbf{r}') dS' / \int_S dS' = 0$. We can ensure that this is the case by choosing

$$F(\mathbf{r}) = - \int_S G_N(\mathbf{r}, \mathbf{r}') dS' \left/ \int_S dS' \right. . \quad (2.136)$$

Thus, given an asymmetric Neumann Green's function, it is always possible to construct a symmetric Green's function that satisfies

$$\nabla'^2 \tilde{G}_N(\mathbf{r}, \mathbf{r}') = \delta(\mathbf{r} - \mathbf{r}'), \quad (2.137)$$

$$\int_S \tilde{G}_N(\mathbf{r}, \mathbf{r}') dS' \left/ \int_S dS' \right. = 0, \quad (2.138)$$

and

$$\frac{\partial G_N(\mathbf{r}, \mathbf{r}')}{\partial n'} = 1 \left/ \int_S dS' \right. \quad (2.139)$$

when \mathbf{r}' lies on S .

We can also show that the symmetric Neumann Green's function is unique. Proceeding in the usual fashion, suppose that there are two different functions, $G_1(\mathbf{r}, \mathbf{r}')$ and $G_2(\mathbf{r}, \mathbf{r}')$, that both satisfy Equations (2.137)—(2.139). It follows that $G_3(\mathbf{r}, \mathbf{r}') = G_1(\mathbf{r}, \mathbf{r}') - G_2(\mathbf{r}, \mathbf{r}')$ satisfies

$$\nabla'^2 G_3(\mathbf{r}, \mathbf{r}') = 0 \quad (2.140)$$

throughout V , subject to the boundary condition

$$\frac{\partial G_3(\mathbf{r}, \mathbf{r}')}{\partial n'} = 0 \quad (2.141)$$

when \mathbf{r}' lies on S . Equation (2.4) can be written

$$\int_S \phi(\mathbf{r}') \frac{\partial \phi}{\partial n'} dS' = \int_V [\phi(\mathbf{r}') \nabla'^2 \phi + \nabla' \phi \cdot \nabla' \phi] dV'. \quad (2.142)$$

Suppose that $\phi(\mathbf{r}') = G_3(\mathbf{r}, \mathbf{r}')$. It follows that

$$\int_V |\nabla' G_3(\mathbf{r}, \mathbf{r}')|^2 dV' = 0, \quad (2.143)$$

which implies that

$$G_3(\mathbf{r}, \mathbf{r}') = F(\mathbf{r}), \quad (2.144)$$

where $F(\mathbf{r})$ is arbitrary. However, $G_3(\mathbf{r}, \mathbf{r}')$ also satisfies

$$\int_S G_3(\mathbf{r}, \mathbf{r}') dS' \left/ \int_S dS' \right. = F(\mathbf{r}) = 0. \quad (2.145)$$

Hence, $F(\mathbf{r}) = G_3(\mathbf{r}, \mathbf{r}') = 0$, and the Green's function is unique. It follows that the potential specified in Equation (2.128) is also unique (up to an arbitrary additive constant).

Finally, the fact that the Green's function for Poisson's equation, $G(\mathbf{r}, \mathbf{r}')$, is (or can be chosen to be) symmetric implies from Equation (2.113) that

$$\nabla'^2 G_D(\mathbf{r}, \mathbf{r}') = \nabla^2 G_D(\mathbf{r}, \mathbf{r}') = \delta(\mathbf{r} - \mathbf{r}'), \quad (2.146)$$

because $\delta(\mathbf{r}' - \mathbf{r}) = \delta(\mathbf{r} - \mathbf{r}')$.

2.11 Dirichlet Green's Function for Spherical Surface

As an example of a boundary value problem, suppose that we wish to solve Poisson's equation, subject to Dirichlet boundary conditions, in some domain V that lies between the spherical surfaces

$r = a$ and $r = \infty$, where r is a radial spherical coordinate. Let S and S' denote the former and latter surfaces, respectively. The Green's function for the problem, $G_D(\mathbf{r}, \mathbf{r}')$, must satisfy

$$\nabla^2 G_D(\mathbf{r}, \mathbf{r}') = \delta(\mathbf{r} - \mathbf{r}'), \quad (2.147)$$

for \mathbf{r}, \mathbf{r}' not outside V , and

$$G_D(\mathbf{r}, \mathbf{r}') = 0 \quad (2.148)$$

when \mathbf{r}' lies on S or on S' . The Green's function also has the symmetry property

$$G_D(\mathbf{r}', \mathbf{r}) = G_D(\mathbf{r}, \mathbf{r}'). \quad (2.149)$$

Let us try

$$G_D(\mathbf{r}, \mathbf{r}') = -\frac{1}{4\pi |\mathbf{r}' - \mathbf{r}|} + \frac{a}{4\pi r' |a^2 \mathbf{r}'/r'^2 - \mathbf{r}|}. \quad (2.150)$$

Note that the above function is symmetric with respect to its arguments, because $r' |a^2 \mathbf{r}'/r'^2 - \mathbf{r}| = r |a^2 \mathbf{r}/r^2 - \mathbf{r}'|$. It follows from Equation (1.25) that

$$\nabla^2 G_D(\mathbf{r}, \mathbf{r}') = \delta(\mathbf{r} - \mathbf{r}') - \frac{a}{r'} \delta(\mathbf{r} - a^2 \mathbf{r}'/r'^2). \quad (2.151)$$

However, if \mathbf{r}, \mathbf{r}' do not lie outside V then the argument of the latter delta function cannot be zero. Hence, for \mathbf{r}, \mathbf{r}' not outside V , this function takes the value zero, and the above expression reduces to

$$\nabla^2 G_D(\mathbf{r}, \mathbf{r}') = \delta(\mathbf{r} - \mathbf{r}'), \quad (2.152)$$

as required. Equation (2.150) can be written

$$G_D(\mathbf{r}, \mathbf{r}') = -\frac{1}{4\pi (r^2 - 2r r' \cos \gamma + r'^2)^{1/2}} + \frac{1}{4\pi (r^2 r'^2/a^2 - 2r r' \cos \gamma + a^2)^{1/2}}, \quad (2.153)$$

where $\cos \gamma = \mathbf{r} \cdot \mathbf{r}'/(r r')$. When written in this form, it becomes clear that $G_D(\mathbf{r}, \mathbf{r}') = 0$ when \mathbf{r}' lies on S (i.e., when $r' = a$) or on S' (i.e., when $r' = \infty$). We conclude that expression (2.153) is the unique Green's function for the Dirichlet problem within the domain V .

According to Equation (2.120), the electrostatic potential within the domain V is written

$$\phi(\mathbf{r}) = -\frac{1}{\epsilon_0} \int_V G_D(\mathbf{r}, \mathbf{r}') \rho(\mathbf{r}') dV' - \int_S \phi_S(\mathbf{r}') \frac{\partial G_D(\mathbf{r}, \mathbf{r}')}{\partial r'} dS'. \quad (2.154)$$

Here, $\rho(\mathbf{r})$ is the charge distribution within V (i.e., the region $r > a$), $\phi_S(\mathbf{r})$ is the potential on S (i.e., the surface $r = a$), and the potential at infinity (i.e., on the surface S') is assumed to be zero. Moreover, we have made use of the fact that $\partial/\partial n' = -\partial/\partial r'$ on S , because the unit vector \mathbf{n}' points radially inward. Finally, it is easily demonstrated that

$$\left. \frac{\partial G_D}{\partial r'} \right|_{r'=a} = \frac{a - r^2/a}{4\pi (r^2 - 2ra \cos \gamma + a^2)^{3/2}}. \quad (2.155)$$

Hence,

$$\begin{aligned}\phi(r, \theta, \varphi) &= \frac{1}{4\pi \epsilon_0} \int_a^\infty \int_0^\pi \int_0^{2\pi} \frac{\rho(r', \theta', \varphi') r'^2 \sin \theta' dr' d\theta' d\varphi'}{(r^2 - 2 r r' \cos \gamma + r'^2)^{1/2}} \\ &\quad - \frac{1}{4\pi \epsilon_0} \int_a^\infty \int_0^\pi \int_0^{2\pi} \frac{\rho(r', \theta', \varphi') r'^2 \sin \theta' dr' d\theta' d\varphi'}{(r^2 r'^2/a^2 - 2 r r' \cos \gamma + a^2)^{1/2}} \\ &\quad + \frac{1}{4\pi} \int_0^\pi \int_0^{2\pi} \frac{\phi_S(\theta', \varphi') (r^2 - a^2) a \sin \theta' d\theta' d\varphi'}{(r^2 - 2 r a \cos \gamma + a^2)^{3/2}},\end{aligned}\quad (2.156)$$

where r, θ, φ and r', θ', φ' are standard spherical coordinates, in terms of which,

$$\cos \gamma = \cos \theta \cos \theta' + \sin \theta \sin \theta' \cos(\varphi - \varphi'). \quad (2.157)$$

As a specific example, suppose that S represents the surface of a grounded spherical conductor. It follows that $\phi_S = 0$. Suppose that there is a single charge, q , in the domain V , located on the z -axis at $r = b > a$, $\theta = 0$, and $\varphi = 0$. It follows that

$$\rho(\mathbf{r}') dV' \rightarrow q \delta(r' - b) \delta(\theta') \delta(\varphi') dr' d\theta' d\varphi'. \quad (2.158)$$

Thus, Equations (2.156) and (2.157) yield

$$\phi(r, \theta) = \frac{q}{4\pi \epsilon_0} \left[\frac{1}{(r^2 - 2 r b \cos \theta + b^2)^{1/2}} - \frac{1}{(r^2 b^2/a^2 - 2 r b \cos \theta + a^2)^{1/2}} \right] \quad (2.159)$$

in the region $r > a$. Of course, $\phi = 0$ in the region $r < a$. It follows from Equations (2.28) and (2.77) that there is a charge sheet on the surface of the conductor (because the normal electric field is non-zero just above the surface, but zero just below) whose density is given by

$$\sigma(\theta) = -\epsilon_0 \left. \frac{\partial \phi}{\partial r} \right|_{r=a} = -\frac{q}{4\pi a} \frac{(b^2 - a^2)}{(a^2 - 2 a b \cos \theta + b^2)^{3/2}}. \quad (2.160)$$

Thus, the net charge induced on the surface of the conductor is

$$q' = \int \sigma dS = 2\pi a^2 \int_0^\pi \sigma(\theta) d\theta = -q \frac{a}{b}. \quad (2.161)$$

In Equation (2.159), the first term inside the square brackets clearly represents the electric potential generated by the charge q , which suggests that

$$\phi'(r, \theta) = -\frac{q}{4\pi \epsilon_0} \frac{1}{(r^2 b^2/a^2 - 2 r b \cos \theta + a^2)^{1/2}} \quad (2.162)$$

is the potential generated by the charges induced on the surface of the conductor. Hence, the attractive force acting on the charge q due to the induced charges is

$$\mathbf{F}' = -q \nabla \phi' \Big|_{r=b, \theta=0} = -\frac{q^2}{4\pi \epsilon_0} \frac{b a}{(b^2 - a^2)^2} \mathbf{e}_z. \quad (2.163)$$

Of course, an equal and opposite force acts on the conductor.

As a final example, suppose that there are no charges in the domain V , but that ϕ_S is prescribed on the surface S . It follows from Equation (2.156) that the electric potential on the z -axis (i.e., $r = z, \theta = 0$, and $\varphi = 0$) is

$$\phi(z) = \frac{1}{4\pi} \int_0^\pi \int_0^{2\pi} \frac{\phi_S(\theta', \varphi') (z^2 - a^2) a \sin \theta' d\theta' d\varphi'}{(z^2 - 2za \cos \theta' + a^2)^{3/2}} \quad (2.164)$$

for $|z| \geq a$. Let S correspond to the surface of two conducting hemispheres (separated by a thin insulator). Suppose that the upper (i.e., $0 \leq \theta < \pi/2$) hemisphere is held at the potential $+V$, whereas the lower (i.e., $\pi/2 < \theta \leq \pi$) hemisphere is held at the potential $-V$. It follows that

$$\phi(z) = \frac{V}{2} (z^2 - a^2) a \left[\int_0^1 \frac{d\mu}{(z^2 - 2za\mu + a^2)^{3/2}} - \int_{-1}^0 \frac{d\mu}{(z^2 - 2za\mu + a^2)^{3/2}} \right], \quad (2.165)$$

where $\mu = \cos \theta$, which yields

$$\phi(z) = \operatorname{sgn}(z) V \left[1 - \frac{z^2 - a^2}{|z|(z^2 + a^2)^{1/2}} \right] \quad (2.166)$$

for $|z| \geq a$.

2.12 Exercises

- 2.1 Prove the *mean value theorem*: for charge-free space the value of the electrostatic potential at any point is equal to the average of the potential over the surface of any sphere centered on that point.
- 2.2 Prove *Green's reciprocity theorem*: if ϕ is the potential due to a volume charge density ρ within a volume V and a surface charge density σ on the conducting surface S bounding the volume V , while ϕ' is the potential due to another charge distribution ρ' and σ' (non-simultaneously occupying the same volume and the same surface, respectively), then

$$\int_V \rho \phi' dV + \int_S \sigma \phi' dS = \int_V \rho' \phi dV + \int_S \sigma' \phi dS.$$

- 2.3 Two infinite grounded parallel conducting planes are separated by a distance d . A point charge q is placed between the planes. Use Green's reciprocity theorem to prove that the total charge induced on one of the planes is equal to $(-q)$ times the fractional perpendicular distance of the point charge from the other plane. [Hint: Choose as your comparison electrostatic problem with the same surfaces one whose charge densities and potential are known and simple.]
- 2.4 Consider two insulated conductors, labeled 1 and 2. Let ϕ_1 be the potential of the first conductor when it is uncharged and the second conductor holds a charge Q . Likewise, let

ϕ_2 be the potential of the second conductor when it is uncharged and the first conductor holds a charge Q . Use Green's reciprocity theorem to demonstrate that

$$\phi_1 = \phi_2.$$

2.5 Consider two insulated spherical conductors. Let the first have radius a . Let the second be sufficiently small that it can effectively be treated as a point charge, and let it also be located a distance $b > a$ from the center of the first. Suppose that the first conductor is uncharged, and that the second carries a charge q . What is the potential of the first conductor? [Hint: Consider the result proved in Exercise 2.1.]

2.6 Consider a set of N conductors distributed in a vacuum. Suppose that the i th conductor carries the charge Q_i and is at the scalar potential ϕ_i . It follows from the linearity of Maxwell's equations and Ohm's law that a linear relationship exists between the potentials and the charges: that is,

$$\phi_i = \sum_{j=1, N} p_{ij} Q_j.$$

Here, the p_{ij} are termed the *coefficients of potential*. Demonstrate that $p_{ij} = p_{ji}$ for all i, j . [Hint: Consider the result proved in Exercise 2.1.] Show that the total electrostatic potential energy of the charged conductors is

$$W = \frac{1}{2} \sum_{i, j=1, N} p_{ij} Q_i Q_j.$$

2.7 Demonstrate that the Green's function for Poisson's equation in two dimensions (i.e., $\partial/\partial z \equiv 0$) is

$$G(\mathbf{r}, \mathbf{r}') = \frac{\ln |\mathbf{r} - \mathbf{r}'|}{2\pi},$$

where $\mathbf{r} = (x, y)$, et cetera. Hence, deduce that the scalar potential field generated by the two-dimensional charge distribution $\rho(\mathbf{r})$ is

$$\phi(\mathbf{r}) = -\frac{1}{2\pi \epsilon_0} \int \rho(\mathbf{r}') \ln |\mathbf{r} - \mathbf{r}'| dV'.$$

2.8 A electric dipole of fixed moment \mathbf{p} is situated at position \mathbf{r} in a non-uniform external electric field $\mathbf{E}(\mathbf{r})$. Demonstrate that the net force on the dipole can be written $\mathbf{f} = -\nabla W$, where

$$W = -\mathbf{p} \cdot \mathbf{E}.$$

2.9 Demonstrate that the electric field generated by an electric dipole of dipole moment \mathbf{p} is

$$\mathbf{E}(\mathbf{r}) = \frac{3(\mathbf{p} \cdot \mathbf{r})\mathbf{r} - r^2 \mathbf{p}}{4\pi \epsilon_0 r^5},$$

where \mathbf{r} represents vector displacement relative to the dipole. Show that the potential energy of an electric dipole of moment \mathbf{p}_1 in the electric field generated by a second dipole of moment \mathbf{p}_2 is

$$W = \frac{r^2 (\mathbf{p}_1 \cdot \mathbf{p}_2) - 3 (\mathbf{p}_1 \cdot \mathbf{r}) (\mathbf{p}_2 \cdot \mathbf{r})}{4\pi \epsilon_0 r^5},$$

where \mathbf{r} is the displacement of one dipole from another.

2.10 Show that the torque on an electric dipole of moment \mathbf{p} in a uniform external electric field \mathbf{E} is

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

Hence, deduce that the potential energy of the dipole is

$$W = -\mathbf{p} \cdot \mathbf{E}.$$

2.11 Consider two coplanar electric dipoles with their centers a fixed distance apart. Show that if the angles the dipoles make with the line joining their centers are θ and θ' , respectively, and θ is held fixed, then

$$\tan \theta = -\frac{1}{2} \tan \theta'$$

in equilibrium.

3 Potential Theory

3.1 Introduction

This chapter discusses various techniques for solving Poisson's equation in multiple dimensions.

3.2 Associated Legendre Functions

The *associated Legendre functions*, $P_l^m(x)$, are the well-behaved solutions of the differential equation

$$\frac{d}{dx} \left[(1-x^2) \frac{dP_l^m}{dx} \right] + \left[l(l+1) - \frac{m^2}{1-x^2} \right] P_l^m = 0, \quad (3.1)$$

for x in the range $-1 \leq x \leq +1$. Here, l is a non-negative integer (known as the degree), and m is an integer (known as the order) lying in the range $-l \leq m \leq l$. The functions themselves take the form¹

$$P_l^m(x) = \frac{(-1)^{l+m}}{2^l l!} (1-x^2)^{m/2} \frac{d^{l+m}}{dx^{l+m}} (1-x^2)^l, \quad (3.2)$$

which implies that

$$P_l^{-m}(x) = (-1)^m \frac{(l-m)!}{(l+m)!} P_l^m(x). \quad (3.3)$$

Assuming that $0 \leq m \leq l$, the $P_l^m(x)$ satisfy the orthogonality condition

$$\int_{-1}^1 P_l^m(x) P_k^m(x) dx = \frac{2(l+m)!}{(2l+1)(l-m)!} \delta_{lk}, \quad (3.4)$$

where δ_{lk} is a Kronecker delta symbol.

The associated Legendre functions of order 0 (i.e., $m = 0$) are called *Legendre polynomials*, and are denoted the $P_l(x)$: that is, $P_l^0(x) = P_l(x)$. It follows that²

$$\int_{-1}^1 P_l(x) P_k(x) dx = \frac{2}{(2l+1)} \delta_{lk}. \quad (3.5)$$

It can also be shown that

$$\frac{1}{(1-2xt+t^2)^{1/2}} = \sum_{l=0, \infty} P_l(x) t^l, \quad (3.6)$$

provided $|t| < 1$ and $|x| \leq 1$.

¹J.D. Jackson, *Classical Electrodynamics*, 2nd Edition, (Wiley, 1962). Section 3.5.

²Ibid. Section 3.2.

All of the associated Legendre functions of degree less than 3 are listed below:

$$P_0^0(x) = 1, \quad (3.7)$$

$$P_1^{-1}(x) = (1/2)(1 - x^2)^{1/2}, \quad (3.8)$$

$$P_1^0(x) = x, \quad (3.9)$$

$$P_1^{+1}(x) = -(1 - x^2)^{1/2}, \quad (3.10)$$

$$P_2^{-2}(x) = (1/8)(1 - x^2), \quad (3.11)$$

$$P_2^{-1}(x) = (1/2)x(1 - x^2)^{1/2}, \quad (3.12)$$

$$P_2^0(x) = (1/2)(3x^2 - 1), \quad (3.13)$$

$$P_2^{+1}(x) = -3x(1 - x^2)^{1/2}, \quad (3.14)$$

$$P_2^{+2}(x) = 3(1 - x^2). \quad (3.15)$$

3.3 Spherical Harmonics

The spherical harmonics, $Y_{l,m}(\theta, \varphi)$, are the angular portions of the global solutions to Laplace's equation in standard spherical coordinates, r, θ, φ . Here, l is a non-negative integer (known as the degree), and m is an integer (known as the order) lying in the range $-l \leq m \leq l$. The spherical harmonics are well behaved and single valued functions that satisfy the differential equation

$$r^2 \nabla^2 Y_{l,m} + l(l+1) Y_{l,m} = 0, \quad (3.16)$$

and take the form³

$$Y_{l,m}(\theta, \varphi) = \left[\frac{(2l+1)(l-m)!}{4\pi(l+m)!} \right]^{1/2} P_l^m(\cos \theta) e^{im\varphi}. \quad (3.17)$$

It follows from Equation (3.3) that

$$Y_{l,-m} = (-1)^m Y_{l,m}^*. \quad (3.18)$$

The $Y_{l,m}(\theta, \varphi)$ satisfy the orthonormality constraint

$$\oint Y_{l,m}(\theta, \varphi) Y_{l',m'}^*(\theta, \varphi) d\Omega = \delta_{ll'} \delta_{mm'}, \quad (3.19)$$

where $d\Omega = \sin \theta d\theta d\varphi$ is an element of solid angle, and the integral is taken over all solid angle. Note that the spherical harmonics form a complete set of angular functions.

³Ibid. Section 3.5.

All of the spherical harmonics of degree less than 3 are listed below:

$$Y_{0,0}(\theta, \varphi) = \left(\frac{1}{4\pi}\right)^{1/2}, \quad (3.20)$$

$$Y_{1,-1}(\theta, \varphi) = \left(\frac{3}{8\pi}\right)^{1/2} \sin \theta e^{-i\varphi}, \quad (3.21)$$

$$Y_{1,0}(\theta, \varphi) = \left(\frac{3}{4\pi}\right)^{1/2} \cos \theta, \quad (3.22)$$

$$Y_{1,1}(\theta, \varphi) = -\left(\frac{3}{8\pi}\right)^{1/2} \sin \theta e^{+i\varphi}, \quad (3.23)$$

$$Y_{2,-2}(\theta, \varphi) = \left(\frac{15}{32\pi}\right)^{1/2} \sin^2 \theta e^{-2i\varphi}, \quad (3.24)$$

$$Y_{2,-1}(\theta, \varphi) = \left(\frac{15}{8\pi}\right)^{1/2} \sin \theta \cos \theta e^{-i\varphi}, \quad (3.25)$$

$$Y_{2,0}(\theta, \varphi) = \left(\frac{5}{16\pi}\right)^{1/2} (3 \cos^2 \theta - 1) \quad (3.26)$$

$$Y_{2,+1}(\theta, \varphi) = -\left(\frac{15}{8\pi}\right)^{1/2} \sin \theta \cos \theta e^{+i\varphi}, \quad (3.27)$$

$$Y_{2,+2}(\theta, \varphi) = \left(\frac{15}{32\pi}\right)^{1/2} \sin^2 \theta e^{+2i\varphi}. \quad (3.28)$$

Consider two spherical coordinate systems, r, θ, φ and r', θ', φ' , whose origins coincide, but whose polar axes subtend an angle γ with respect to one another. It follows that

$$\cos \gamma = \cos \theta \cos \theta' + \sin \theta \sin \theta' \cos(\varphi - \varphi'). \quad (3.29)$$

Moreover, the so-called *addition theorem* for spherical harmonics states that⁴

$$P_l(\cos \gamma) = \frac{4\pi}{2l+1} \sum_{m=-l,+l} Y_{l,m}^*(\theta', \varphi') Y_{l,m}(\theta, \varphi). \quad (3.30)$$

3.4 Laplace's Equation in Spherical Coordinates

Consider the general solution to Laplace's equation,

$$\nabla^2 \phi = 0, \quad (3.31)$$

⁴Ibid. Section 3.6.

in spherical coordinates. Let us write

$$\phi(r, \theta, \varphi) = \sum_{l=0, \infty} \sum_{m=-l, +l} \phi_{l,m}(r) Y_{l,m}(\theta, \varphi). \quad (3.32)$$

It follows from Equation (3.16) that

$$\sum_{l=0, \infty} \sum_{m=-l, +l} \left[\frac{d}{dr} \left(r^2 \frac{d\phi_{l,m}}{dr} \right) - l(l+1) \phi_{l,m} \right] Y_{l,m}(\theta, \varphi) = 0. \quad (3.33)$$

However, given that the spherical harmonics are mutually orthogonal [in the sense that they satisfy Equation (3.19)], we can separately equate the coefficients of each in the above equation, to give

$$\frac{d}{dr} \left(r^2 \frac{d\phi_{l,m}}{dr} \right) - l(l+1) \phi_{l,m} = 0, \quad (3.34)$$

for all $l \geq 0$ and $|m| \leq l$. It follows that

$$\phi_{l,m}(r) = \alpha_{l,m} r^l + \beta_{l,m} r^{-(l+1)}, \quad (3.35)$$

where the $\alpha_{l,m}$ and $\beta_{l,m}$ are arbitrary constants. Hence, the general solution to Laplace's equation in spherical coordinates is written

$$\phi(r, \theta, \varphi) = \sum_{l=0, \infty} \sum_{m=-l, +l} [\alpha_{l,m} r^l + \beta_{l,m} r^{-(l+1)}] Y_{l,m}(\theta, \varphi). \quad (3.36)$$

If the domain of solution includes the origin then all of the $\beta_{l,m}$ must be zero, in order to ensure that the potential remains finite at $r = 0$. On the other hand, if the domain of solution extends to infinity then all of the $\alpha_{l,m}$ (except $\alpha_{0,0}$) must be zero, otherwise the potential would be infinite at $r = \infty$.

3.5 Poisson's Equation in Spherical Coordinates

Consider the general solution to Poisson's equation,

$$\nabla^2 \phi = -\frac{\rho}{\epsilon_0}, \quad (3.37)$$

in spherical coordinates. According to Section 2.3, the general three-dimensional Green's function for Poisson's equation is

$$G(\mathbf{r}, \mathbf{r}') = -\frac{1}{4\pi |\mathbf{r} - \mathbf{r}'|}. \quad (3.38)$$

When expressed in terms of spherical coordinates, this becomes

$$G(\mathbf{r}, \mathbf{r}') = -\frac{1}{4\pi (r^2 - 2r r' \cos \gamma + r'^2)^{1/2}}, \quad (3.39)$$

where

$$\cos \gamma = \cos \theta \cos \theta' + \sin \theta \sin \theta' \cos(\varphi - \varphi'). \quad (3.40)$$

is the angle subtended between \mathbf{r} and \mathbf{r}' . According to Equation (3.6), we can write

$$G(\mathbf{r}, \mathbf{r}') = -\frac{1}{4\pi r} \sum_{l=0, \infty} \left(\frac{r'}{r}\right)^l P_l(\cos \gamma) \quad (3.41)$$

for $r' < r$, and

$$G(\mathbf{r}, \mathbf{r}') = -\frac{1}{4\pi r'} \sum_{l=0, \infty} \left(\frac{r}{r'}\right)^l P_l(\cos \gamma) \quad (3.42)$$

for $r' > r$. Thus, it follows from the spherical harmonic addition theorem, (3.30), that

$$G(\mathbf{r}, \mathbf{r}') = - \sum_{l=0, \infty} \sum_{m=-l, +l} \frac{1}{2l+1} \left(\frac{r'_<}{r'_>}\right)^l Y_{l,m}^*(\theta', \varphi') Y_{l,m}(\theta, \varphi), \quad (3.43)$$

where $r'_<$ represents the lesser of r and r' , whereas $r'_>$ represents the greater of r and r' .

According to Section 2.3, the general solution to Poisson's equation, (3.37), is

$$\phi(\mathbf{r}) = -\frac{1}{\epsilon_0} \int G(\mathbf{r}, \mathbf{r}') \rho(\mathbf{r}') dV'. \quad (3.44)$$

Thus, Equation (3.43) yields

$$\phi(\mathbf{r}) = \frac{1}{\epsilon_0} \sum_{l=0, \infty} \sum_{m=-l, +l} \frac{1}{2l+1} \left[r^l p_{l,m}^*(r) + \frac{q_{l,m}^*(r)}{r^{l+1}} \right] Y_{l,m}(\theta, \varphi), \quad (3.45)$$

where

$$p_{l,m}(r) = \int_r^\infty \oint \frac{1}{r'^{l+1}} \rho(r', \theta, \varphi) Y_{l,m}(\theta, \varphi) r'^2 d\Omega dr', \quad (3.46)$$

$$q_{l,m}(r) = \int_0^r \oint r'^l \rho(r', \theta, \varphi) Y_{l,m}(\theta, \varphi) r'^2 d\Omega dr'. \quad (3.47)$$

3.6 Multipole Expansion

Consider a bounded charge distribution that lies inside the sphere $r = a$. It follows that $\rho = 0$ in the region $r > a$. According to the previous three equations, the electrostatic potential in the region $r > a$ takes the form

$$\phi(\mathbf{r}) = \frac{1}{\epsilon_0} \sum_{l=0, \infty} \sum_{m=-l, +l} \frac{q_{l,m}^*}{2l+1} \frac{Y_{l,m}(\theta, \varphi)}{r^{l+1}}, \quad (3.48)$$

where the

$$q_{l,m}^* = \int r^l \rho(r, \theta, \varphi) Y_{l,m}(\theta, \varphi) dV \quad (3.49)$$

are known as the *multipole moments* of the charge distribution $\rho(\mathbf{r})$. Here, the integral is over all space. Incidentally, the type of expansion specified in Equation (3.48) is called a *multipole expansion*.

The most important $q_{l,m}^*$ are those corresponding to $l = 0$, $l = 1$, and $l = 2$, which are known as *monopole*, *dipole*, and *quadrupole* moments, respectively. For each l , the multipole moments $q_{l,m}^*$, for $m = -l$ to $+l$, form an l th-rank tensor with $2l + 1$ components. However, Equation (3.18) implies that

$$q_{l,m}^* = (-1)^m q_{l,-m}. \quad (3.50)$$

Hence, only $l + 1$ of these components are independent.

For $l = 0$, there is only one monopole moment. Namely,

$$q_{0,0}^* = \int \rho(\mathbf{r}') Y_{0,0}^*(\theta, \varphi) dV = \frac{1}{\sqrt{4\pi}} \int \rho(\mathbf{r}) dV = \frac{Q}{\sqrt{4\pi}}, \quad (3.51)$$

where Q is the net charge contained in the distribution, and use has been made of Equation (3.20). It follows from Equation (3.48) that, at sufficiently large r , the charge distribution acts like a point charge Q situated at the origin. That is,

$$\phi(\mathbf{r}) \simeq \phi_0(\mathbf{r}) = \frac{q_{0,0}^*}{\epsilon_0} \frac{Y_{0,0}(\theta, \varphi)}{r} = \frac{Q}{4\pi \epsilon_0 r}. \quad (3.52)$$

By analogy with Equation (2.69), the dipole moment of the charge distribution is written

$$\mathbf{p} = \int \rho(\mathbf{r}) \mathbf{r} dV. \quad (3.53)$$

The three Cartesian components of this vector are

$$p_x = \int \rho(\mathbf{r}) x dV = \int \rho(\mathbf{r}) r \sin \theta \cos \varphi dV, \quad (3.54)$$

$$p_y = \int \rho(\mathbf{r}) y dV = \int \rho(\mathbf{r}) r \sin \theta \sin \varphi dV, \quad (3.55)$$

$$p_z = \int \rho(\mathbf{r}) z dV = \int \rho(\mathbf{r}) r \cos \theta dV. \quad (3.56)$$

On the other hand, the spherical components of the dipole moment take the form

$$q_{1,-1}^* = \left(\frac{3}{8\pi}\right)^{1/2} \int \rho(\mathbf{r}) r \sin \theta e^{+i\varphi} dV = \left(\frac{3}{8\pi}\right)^{1/2} (p_x + i p_y), \quad (3.57)$$

$$q_{1,0}^* = \left(\frac{3}{4\pi}\right)^{1/2} \int \rho(\mathbf{r}) r \cos \theta dV = \left(\frac{3}{4\pi}\right)^{1/2} p_z, \quad (3.58)$$

$$q_{1,+1}^* = -\left(\frac{3}{8\pi}\right)^{1/2} \int \rho(\mathbf{r}) r \sin \theta e^{-i\varphi} dV = -\left(\frac{3}{8\pi}\right)^{1/2} (p_x - i p_y), \quad (3.59)$$

where use has been made of Equations (3.21)–(3.23). It can be seen that the three spherical dipole moments are independent linear combinations of the three Cartesian moments. The potential associated with the dipole moment is

$$\phi_1(\mathbf{r}) = \frac{1}{3\epsilon_0} \left(\frac{q_{1,-1}^* r Y_{1,-1} + q_{1,0}^* r Y_{1,0} + q_{1,+1}^* r Y_{1,+1}}{r^3} \right). \quad (3.60)$$

However, from Equations (3.21)–(3.23),

$$r Y_{1,-1} = \left(\frac{3}{8\pi} \right)^{1/2} (x - iy), \quad (3.61)$$

$$r Y_{1,0} = \left(\frac{3}{4\pi} \right)^{1/2} z, \quad (3.62)$$

$$r Y_{1,+1} = - \left(\frac{3}{8\pi} \right)^{1/2} (x + iy). \quad (3.63)$$

Hence,

$$\phi_1(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \frac{p_x x + p_y y + p_z z}{r^3} = \frac{1}{4\pi\epsilon_0} \frac{\mathbf{p} \cdot \mathbf{r}}{r^3}, \quad (3.64)$$

in accordance with Equation (2.74). Note, finally, that if the net charge, Q , contained in the distributions is non-zero then it is always possible to choose the origin of the coordinate system in such a manner that $\mathbf{p} = \mathbf{0}$.

The Cartesian components of the *quadrupole tensor* are defined

$$Q_{ij} = \int \rho(\mathbf{r}) (3 x_i x_j - r^2 \delta_{ij}) dV, \quad (3.65)$$

for $i, j = 1, 2, 3$. Here, $x_1 = x$, $x_2 = y$, and $x_3 = z$. Incidentally, because the quadrupole tensor is symmetric (i.e., $Q_{ji} = Q_{ij}$) and traceless (i.e., $Q_{11} + Q_{22} + Q_{33} = 0$), it only possesses five independent Cartesian components. The five spherical components of the quadrupole tensor take the form

$$q_{2,-2}^* = \left(\frac{5}{96\pi} \right)^{1/2} (Q_{11} + 2i Q_{12} - Q_{22}), \quad (3.66)$$

$$q_{2,-1}^* = \left(\frac{5}{24\pi} \right)^{1/2} (Q_{13} + i Q_{23}), \quad (3.67)$$

$$q_{2,0}^* = \left(\frac{5}{16\pi} \right)^{1/2} Q_{33}, \quad (3.68)$$

$$q_{2,+1}^* = - \left(\frac{5}{24\pi} \right)^{1/2} (Q_{13} - i Q_{23}), \quad (3.69)$$

$$q_{2,+2}^* = \left(\frac{5}{96\pi} \right)^{1/2} (Q_{11} - 2i Q_{12} - Q_{22}). \quad (3.70)$$

Moreover, the potential associated with the quadrupole tensor is

$$\phi_2(\mathbf{r}) = \frac{1}{5\epsilon_0} \sum_{m=-2,+2} \frac{q_{2,m}^* Y_{2,m}(\theta, \varphi)}{r^3} = \frac{1}{8\pi\epsilon_0} \sum_{i,j=1,3} \frac{Q_{ij} x_i x_j}{r^5}. \quad (3.71)$$

It follows, from the previous analysis, that the first three terms in the multipole expansion, (3.48), can be written

$$\phi(\mathbf{r}) \simeq \phi_0(\mathbf{r}) + \phi_1(\mathbf{r}) + \phi_2(\mathbf{r}) = \frac{Q}{4\pi\epsilon_0 r} + \frac{\mathbf{p} \cdot \mathbf{r}}{4\pi\epsilon_0 r^3} + \sum_{i,j=1,3} \frac{Q_{ij} x_i x_j}{8\pi\epsilon_0 r^5}. \quad (3.72)$$

Moreover, at sufficiently large r , these are always the dominant terms in the expansion.

3.7 Axisymmetric Charge Distributions

For the case of an axisymmetric charge distribution (i.e., a charge distribution that is independent of the azimuthal angle φ), we can neglect the spherical harmonics of non-zero order (i.e., the non-axisymmetric harmonics) in Equation (3.43), which reduces to the following expression for the general axisymmetric Green's function:

$$G(\mathbf{r}, \mathbf{r}') = -\frac{1}{4\pi} \sum_{l=0,\infty} \frac{r_{<}^l}{r_{>}^{l+1}} P_l(\cos \theta') P_l(\cos \theta). \quad (3.73)$$

Here, use have been made of the fact that [see Equation (3.17)]

$$Y_{l,0}(\theta, \varphi) = \left(\frac{2l+1}{4\pi} \right)^{1/2} P_l(\cos \theta). \quad (3.74)$$

In this case, the general solution to Poisson's equation, (3.45), reduces to

$$\phi(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \sum_{l=0,\infty} \left[r^l p_l(r) + \frac{q_l(r)}{r^{l+1}} \right] P_l(\cos \theta), \quad (3.75)$$

where

$$p_l(r) = \int_r^\infty \int_0^\pi \frac{1}{r'^{l+1}} \rho(r', \theta) P_l(\cos \theta) 2\pi r'^2 \sin \theta d\theta dr', \quad (3.76)$$

$$q_l(r) = \int_0^r \int_0^\pi r'^l \rho(r', \theta) P_l(\cos \theta) 2\pi r'^2 \sin \theta d\theta dr'. \quad (3.77)$$

Consider the potential generated by a charge q distributed uniformly in a thin ring of radius a that lies in the x - y plane, and is centered at the origin. It follows that

$$\rho(r, \theta) 2\pi r^2 \sin \theta d\theta dr \rightarrow q \delta(r-a) \delta(\theta-\pi/2) d\theta dr. \quad (3.78)$$

Hence, for $r < a$ we obtain $q_l = 0$ and $p_l = q P_l(0)/a^{l+1}$. On the other hand, for $r > a$ we get $p_l = 0$ and $q_l = q a^l P_l(0)$. Thus,

$$\phi(r, \theta) = \frac{q}{4\pi\epsilon_0} \sum_{l=0,\infty} \frac{r_{<}^l}{r_{>}^{l+1}} P_l(0) P_l(\cos \theta), \quad (3.79)$$

where $r_{<}$ represents the lesser of r and a , whereas $r_{>}$ represents the greater.

3.8 Dirichlet Problem in Spherical Coordinates

We saw in Section 2.10 that the solution to the Dirichlet problem, in which the charge density is specified within some volume V , and the potential given on the bounding surface S , takes the form

$$\phi(\mathbf{r}) = -\frac{1}{\epsilon_0} \int_V G_D(\mathbf{r}, \mathbf{r}') \rho(\mathbf{r}') dV' + \int_S \phi(\mathbf{r}') \frac{\partial G_D(\mathbf{r}, \mathbf{r}')}{\partial n'} dS', \quad (3.80)$$

where the Dirichlet Green's function is written

$$G_D(\mathbf{r}, \mathbf{r}') = -\frac{1}{4\pi |\mathbf{r} - \mathbf{r}'|} + F(\mathbf{r}, \mathbf{r}'). \quad (3.81)$$

Here, $F(\mathbf{r}, \mathbf{r}')$ is solution of Laplace's equation (i.e., $\nabla^2 F = 0$) which is chosen so as to ensure that $G_D(\mathbf{r}, \mathbf{r}') = 0$ when \mathbf{r} (or \mathbf{r}') lies on S . Thus, it follows from Sections 3.4 and 3.5 that

$$\begin{aligned} G_D(\mathbf{r}, \mathbf{r}') = & - \sum_{l=0, \infty} \sum_{m=-l, +l} \frac{1}{2l+1} \frac{r_{<}^l}{r_{>}^{l+1}} Y_{l,m}^*(\theta', \varphi') Y_{l,m}(\theta, \varphi) \\ & + \sum_{l=0, \infty} \sum_{m=-l, +l} \left[\alpha_{l,m}(r', \theta', \varphi') r^l + \frac{\beta_{l,m}(r', \theta', \varphi')}{r^{l+1}} \right] Y_{l,m}(\theta, \varphi), \end{aligned} \quad (3.82)$$

where the $\alpha_{l,m}$ and the $\beta_{l,m}$ are chosen in such a manner that the Green's function is zero when \mathbf{r} lies on S .

As a specific example, suppose that the volume V lies between the two spherical surfaces $r = a$ and $r = \infty$. The constraint that $G_D(\mathbf{r}, \mathbf{r}') \rightarrow 0$ as $r \rightarrow \infty$ implies that the $\alpha_{l,m}$ are all zero. On the other hand, the constraint $G_D(\mathbf{r}, \mathbf{r}') = 0$ when $r = a$ yields

$$\beta_{l,m} = \frac{1}{2l+1} \frac{a^{2l+1}}{r'^{l+1}} Y_{l,m}^*(\theta', \varphi'). \quad (3.83)$$

Hence, the unique Green's function for the problem becomes

$$G_D(\mathbf{r}, \mathbf{r}') = - \sum_{l=0, \infty} \sum_{m=-l, +l} \frac{1}{2l+1} \left(\frac{r_{<}^l}{r_{>}^{l+1}} - \frac{a^{2l+1}}{r_{<}^{l+1} r_{>}^{l+1}} \right) Y_{l,m}^*(\theta', \varphi') Y_{l,m}(\theta, \varphi). \quad (3.84)$$

Furthermore, it is readily demonstrated that

$$\left. \frac{\partial G_D}{\partial r'} \right|_{r'=a} = - \sum_{l=0, \infty} \sum_{m=-l, +l} \frac{a^{l-1}}{r^{l+1}} Y_{l,m}^*(\theta', \varphi') Y_{l,m}(\theta, \varphi). \quad (3.85)$$

It is convenient to write

$$\phi(r, \theta, \varphi) = \sum_{l=0, \infty} \sum_{m=-l, +l} \phi_{l,m}(r) Y_{l,m}(\theta, \varphi), \quad (3.86)$$

$$\rho(r, \theta, \varphi) = \sum_{l=0, \infty} \sum_{m=-l, +l} \rho_{l,m}(r) Y_{l,m}(\theta, \varphi). \quad (3.87)$$

It follows from Equation (3.19) that

$$\phi_{l,m}(r) = \oint \phi(r, \theta, \varphi) Y_{l,m}^*(\theta, \varphi) d\Omega, \quad (3.88)$$

$$\rho_{l,m}(r) = \oint \rho(r, \theta, \varphi) Y_{l,m}^*(\theta, \varphi) d\Omega. \quad (3.89)$$

Thus, Equations (3.80), (3.84) and (3.85) yield

$$\begin{aligned} \phi_{l,m}(r) = & \frac{1}{2l+1} \int_a^r \frac{\rho_{l,m}(r')}{\epsilon_0} \left(\frac{r'^l}{r'^{l+1}} - \frac{a^{2l+1}}{r'^{l+1} r^{l+1}} \right) r'^2 dr' \\ & + \frac{1}{2l+1} \int_r^\infty \frac{\rho_{l,m}(r')}{\epsilon_0} \left(\frac{r^l}{r'^{l+1}} - \frac{a^{2l+1}}{r'^{l+1} r^{l+1}} \right) r'^2 dr' + \phi_{l,m}(a) \left(\frac{a}{r} \right)^{l+1}. \end{aligned} \quad (3.90)$$

3.9 Neumann Problem in Spherical Coordinates

According to Section 2.10, the solution to the Neumann problem, in which the charge density is specified within some volume V , and the normal derivative of the potential given on the bounding surface S , takes the form

$$\phi(\mathbf{r}) = -\frac{1}{\epsilon_0} \int_V G_N(\mathbf{r}, \mathbf{r}') \rho(\mathbf{r}') dV' - \int_S G_N(\mathbf{r}, \mathbf{r}') \frac{\partial \phi(\mathbf{r}')}{\partial n'} dS', \quad (3.91)$$

where the Neumann Green's function is written

$$G_N(\mathbf{r}, \mathbf{r}') = -\frac{1}{4\pi |\mathbf{r} - \mathbf{r}'|} + F(\mathbf{r}, \mathbf{r}'). \quad (3.92)$$

Here, $F(\mathbf{r}, \mathbf{r}')$ is solution of Laplace's equation (i.e., $\nabla^2 F = 0$) which is chosen so as to ensure that

$$\int_S G_N(\mathbf{r}, \mathbf{r}') dS = 0, \quad (3.93)$$

and

$$\frac{\partial G_N(\mathbf{r}, \mathbf{r}')}{\partial n} = 1 \Big| \int_S dS. \quad (3.94)$$

The latter constraint holds when \mathbf{r} (or \mathbf{r}') lies on S . Note that we have chosen the arbitrary constant to which the potential $\phi(\mathbf{r})$ is undetermined such that $\langle \phi \rangle_S = 0$. It again follows from Sections 3.4 and 3.5 that

$$\begin{aligned} G_N(\mathbf{r}, \mathbf{r}') = & - \sum_{l=0, \infty} \sum_{m=-l, +l} \frac{1}{2l+1} \frac{r_{<}^l}{r_{>}^{l+1}} Y_{l,m}^*(\theta', \varphi') Y_{l,m}(\theta, \varphi) \\ & + \sum_{l=0, \infty} \sum_{m=-l, +l} \left[\alpha_{l,m}(r', \theta', \varphi') r^l + \frac{\beta_{l,m}(r', \theta', \varphi')}{r^{l+1}} \right] Y_{l,m}(\theta, \varphi), \end{aligned} \quad (3.95)$$

where the $\alpha_{l,m}$ and the $\beta_{l,m}$ are chosen in such a manner that the constraints (3.93) and (3.94) are satisfied.

As a specific example, suppose that the volume V lies inside the spherical surface $r = a$. The physical constraint that the Green's function remain finite at $r = 0$ implies that the $\beta_{l,m}$ are all zero. Applying the constraint (3.93) at $r = a$, we get

$$\alpha_{0,0}(r', \theta', \varphi') = \frac{Y_{0,0}^*(\theta', \varphi')}{a}. \quad (3.96)$$

Similarly, the constraint (3.94) leads to

$$\alpha_{l,m}(r', \theta', \varphi') = -\left(\frac{l+1}{2l+1}\right) \frac{r'^l}{a^{l+2}} Y_{l,m}^*(\theta', \varphi') \quad (3.97)$$

for $l > 0$. Hence, the unique Green's function for the problem becomes

$$\begin{aligned} G_D(\mathbf{r}, \mathbf{r}') &= -\left(\frac{1}{r_{>}} - \frac{1}{a}\right) Y_{0,0}^*(\theta', \varphi') Y_{0,0}(\theta, \varphi) \\ &\quad - \sum_{l=1, \infty} \sum_{m=-l, +l} \frac{1}{2l+1} \left(\frac{r_{<}^l}{r_{>}^{l+1}} + \frac{l+1}{l} \frac{r_{<}^l r_{>}^l}{a^{2l+1}}\right) Y_{l,m}^*(\theta', \varphi') Y_{l,m}(\theta, \varphi). \end{aligned} \quad (3.98)$$

Finally, expanding $\phi(\mathbf{r})$ and $\rho(\mathbf{r})$ in the forms (3.86) and (3.87), respectively, Equations (3.91) and (3.98) yield

$$\phi_{0,0}(r) = \int_0^r \frac{\rho_{0,0}(r')}{\epsilon_0} \left(\frac{1}{r} - \frac{1}{a}\right) r'^2 dr' + \int_r^a \frac{\rho_{0,0}(r')}{\epsilon_0} \left(\frac{1}{r'} - \frac{1}{a}\right) r'^2 dr', \quad (3.99)$$

and

$$\begin{aligned} \phi_{l,m}(r) &= \frac{1}{2l+1} \int_0^r \frac{\rho_{l,m}(r')}{\epsilon_0} \left(\frac{r'^l}{r^{l+1}} + \frac{l+1}{l} \frac{r'^l r^l}{a^{2l+1}}\right) r'^2 dr' \\ &\quad + \frac{1}{2l+1} \int_r^a \frac{\rho_{l,m}(r')}{\epsilon_0} \left(\frac{r^l}{r'^{l+1}} + \frac{l+1}{l} \frac{r'^l r^l}{a^{2l+1}}\right) r'^2 dr' + r \left. \frac{d\phi_{l,m}}{dr} \right|_{r=a} \frac{1}{l} \left(\frac{r}{a}\right)^l \end{aligned} \quad (3.100)$$

for $l > 0$.

3.10 Laplace's Equation in Cylindrical Coordinates

Suppose that we wish to solve Laplace's equation,

$$\nabla^2 \phi = 0, \quad (3.101)$$

within a cylindrical volume of radius a and height L . Let us adopt the standard cylindrical coordinates, r, θ, z . Suppose that the curved portion of the bounding surface corresponds to $r = a$,

while the two flat portions correspond to $z = 0$ and $z = L$, respectively. Suppose, finally, that the boundary conditions that are imposed at the bounding surface are

$$\phi(r, \theta, 0) = 0, \quad (3.102)$$

$$\phi(a, \theta, z) = 0, \quad (3.103)$$

$$\phi(r, \theta, L) = \Phi(r, \theta), \quad (3.104)$$

where $\Phi(r, \theta)$ is a given function. In other words, the potential is zero on the curved and bottom surfaces of the cylinder, and specified on the top surface.

In cylindrical coordinates, Laplace's equation is written

$$\frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial \phi}{\partial r} \right) + \frac{1}{r^2} \frac{\partial^2 \phi}{\partial \theta^2} + \frac{\partial^2 \phi}{\partial z^2} = 0. \quad (3.105)$$

Let us try a separable solution of the form

$$\phi(r, \theta, z) = R(r) Q(\theta) Z(z). \quad (3.106)$$

Proceeding in the usual manner, we obtain

$$\frac{d^2 Z}{dz^2} - k^2 Z = 0, \quad (3.107)$$

$$\frac{d^2 Q}{d\theta^2} + m^2 Q = 0, \quad (3.108)$$

$$\frac{d^2 R}{dr^2} + \frac{1}{r} \frac{dR}{dr} + \left(k^2 - \frac{m^2}{r^2} \right) R = 0. \quad (3.109)$$

Note that we have selected exponential, rather than oscillating, solutions in the z -direction [by writing $-k^2 Z$, instead of $+k^2 Z$, in Equation (3.107)]. As will become clear, this implies that the radial solutions oscillate, which is the appropriate choice for the particular set of boundary conditions under consideration. The solution to Equation (3.107), subject to the constraint that $Z(0) = 0$ [which follows from the first boundary condition, (3.102)] is

$$Z(z) = \sinh(kz). \quad (3.110)$$

The most general solution to Equation (3.108) is

$$Q(\theta) = \sum_{m=0, \infty} [A_m \cos(m\theta) + B_m \sin(m\theta)]. \quad (3.111)$$

Note that, to ensure that the potential is single-valued in θ , the constant m is constrained to be an integer. Finally, if we write $p = kr$ then Equation (3.109) becomes

$$\frac{d^2 R}{dp^2} + \frac{1}{p} \frac{dR}{dp} + \left(1 - \frac{m^2}{p^2} \right) R = 0. \quad (3.112)$$

This equation is known as *Bessel's equation*. The standard solution of this equation that is well behaved at $r = 0$ is⁵

$$J_m(p) = \frac{1}{\pi} \int_0^\pi \cos(p \sin \theta - m \theta) d\theta. \quad (3.113)$$

This solution, which is known as a *Bessel function*, has the properties that

$$J_m(p) \rightarrow \frac{1}{m!} \left(\frac{p}{2}\right)^m \quad \text{as } p \rightarrow 0, \quad (3.114)$$

$$J_m(p) \rightarrow \left(\frac{2}{\pi p}\right)^{1/2} \cos\left(p - m \frac{\pi}{2} - \frac{\pi}{4}\right) \quad \text{as } p \rightarrow \infty. \quad (3.115)$$

In other words, at small arguments the function has a power-law behavior, whereas at large arguments it takes the form of an oscillation of slowly decaying amplitude. It follows that

$$R(r) = J_m(kr). \quad (3.116)$$

Let j_{mn} denote the n th zero of the Bessel function $J_m(p)$. In other words, j_{mn} is the n th root (in order, as p increases from zero) of $J_m(p) = 0$. The values of the j_{mn} can be looked up in standard reference books.⁶ (For example, $j_{01} = 2.405$ and $j_{02} = 5.520$.) We can satisfy our second boundary condition, (3.103), by making the choice $k = k_{mn}$, where

$$k_{mn} = \frac{j_{mn}}{a}. \quad (3.117)$$

Thus, our separable solution becomes

$$\phi(r, \theta, z) = \sum_{m=0, \infty} \sum_{n=1, \infty} \sinh(j_{mn} z/a) J_m(j_{mn} r/a) [A_{mn} \cos(m\theta) + B_{mn} \sin(m\theta)]. \quad (3.118)$$

It is convenient to express the specified function $\Phi(r, \theta)$ in the form of a Fourier series: that is,

$$\Phi(r, \theta) = \sum_{m=0, \infty} [C_m(r) \cos(m\theta) + S_m(r) \sin(m\theta)]. \quad (3.119)$$

Our final boundary condition, (3.104), then yields

$$C_m(r) = \sum_{n=1, \infty} A_{mn} \sinh(j_{mn} L/a) J_m(j_{mn} r/a), \quad (3.120)$$

$$S_m(r) = \sum_{n=1, \infty} B_{mn} \sinh(j_{mn} L/a) J_m(j_{mn} r/a). \quad (3.121)$$

It remains to invert the previous two expressions to obtain the coefficients A_{mn} and B_{mn} . In fact, it is possible to demonstrate that if

$$f(p) = \sum_{n=1, \infty} a_{mn} J_m(j_{mn} p) \quad (3.122)$$

⁵M. Abramowitz, and I. Stegun (eds.), *Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables*, (Dover, New York, 1965). Chapter 9.

⁶Ibid.

then

$$a_{mn} = \frac{2}{J_{m+1}^2(j_{mn})} \int_0^1 p f(p) J_m(j_{mn} p) dp. \quad (3.123)$$

Hence,

$$A_{mn} = \frac{2}{a^2 J_{m+1}^2(j_{mn}) \sinh(j_{mn} L/a)} \int_0^a r C_m(r) J_m(j_{mn} r/a) dr, \quad (3.124)$$

$$B_{mn} = \frac{2}{a^2 J_{m+1}^2(j_{mn}) \sinh(j_{mn} L/a)} \int_0^a r S_m(r) J_m(j_{mn} r/a) dr, \quad (3.125)$$

and our solution is fully determined.

Consider the limit that $a \rightarrow \infty$. In this case, according to Equation (3.117), the allowed values of k become more and more closely spaced. Consequently, the sum over discrete k -values in (3.118) morphs into an integral over a continuous range of k -values. For instance, suppose that we wish to solve Laplace's equation in the region $z \geq 0$, subject to the boundary condition that $\phi \rightarrow 0$ as $z \rightarrow \infty$ and $r \rightarrow \infty$, with $\phi(r, \theta, 0) = \Phi(r, \theta)$, where $\Phi(r, \theta)$ is specified. In this case, we would choose $Z(z) = e^{-kz}$ in order to satisfy the boundary condition at large z . The choice $R(r) = J_m(kr)$ ensures that the potential is well behaved at $r = 0$, and automatically satisfies the boundary condition at large r . Hence, our general solution becomes

$$\phi(r, \theta, z) = \sum_{m=0, \infty} \int_0^{\infty} e^{-kz} J_m(kr) [A_m(k) \cos(m\theta) + B_m(k) \sin(m\theta)] dk. \quad (3.126)$$

If we write

$$\Phi(r, \theta) = \sum_{m=0, \infty} [C_m(r) \cos(m\theta) + S_m(r) \sin(m\theta)] \quad (3.127)$$

then the final boundary condition implies that

$$C_m(r) = \int_0^{\infty} J_m(kr) A_m(k) dk, \quad (3.128)$$

$$S_m(r) = \int_0^{\infty} J_m(kr) B_m(k) dk. \quad (3.129)$$

We can invert the previous two expressions by means of the identity

$$\int_0^{\infty} r J_m(kr) J_m(k'r) dr = \frac{1}{k} \delta(k - k'). \quad (3.130)$$

Hence, we obtain

$$A_m(k) = \int_0^{\infty} k r J_m(kr) C_m(r) dr, \quad (3.131)$$

$$B_m(k) = \int_0^{\infty} k r J_m(kr) S_m(r) dr, \quad (3.132)$$

and our solution is fully defined.

Suppose that we wish to solve Laplace's equation in a cylindrical volume of radius a and height L , subject to the boundary conditions

$$\phi(r, \theta, 0) = 0, \quad (3.133)$$

$$\phi(r, \theta, L) = 0, \quad (3.134)$$

$$\phi(a, \theta, z) = \Phi(\theta, z), \quad (3.135)$$

where $\Phi(\theta, z)$ is specified. In other words, the potential is zero on the two flat portions of the bounding surface, and given on the curved portion. We can again look for a separable solution of the form (3.106). Proceeding in the usual manner, we obtain

$$\frac{d^2 Z}{dz^2} + k^2 Z = 0, \quad (3.136)$$

$$\frac{d^2 Q}{d\theta^2} + m^2 Q = 0, \quad (3.137)$$

$$\frac{d^2 R}{dr^2} + \frac{1}{r} \frac{dR}{dr} - \left(k^2 + \frac{m^2}{r^2} \right) R = 0. \quad (3.138)$$

Note that we have selected oscillating, rather than exponential solutions in the z -direction [by writing $+k^2 Z$, instead of $-k^2 Z$, in Equation (3.136)]. This is the appropriate choice for the particular set of boundary conditions under consideration. The solution to Equation (3.136), subject to the constraints that $Z(0) = Z(L) = 0$ [which follow from the boundary conditions (3.133) and (3.134)] is

$$Z(k) = \sin(k_n z), \quad (3.139)$$

where

$$k_n = n \frac{\pi}{L}. \quad (3.140)$$

Here, n is a positive integer. The single-valued solution to Equation (3.137) is again

$$Q(\theta) = \sum_{m=0, \infty} [A_m \cos(m \theta) + B_m \sin(m \theta)]. \quad (3.141)$$

Finally, writing $p = k_n r$, Equation (3.138) takes the form

$$\frac{d^2 R}{dp^2} + \frac{1}{p} \frac{dR}{dp} - \left(1 + \frac{m^2}{p^2} \right) R = 0. \quad (3.142)$$

This equation is known as the *modified Bessel equation*. The standard solution of this equation that is well behaved at $r = 0$ is⁷

$$I_m(p) = \frac{1}{\pi} \int_0^\pi e^{p \cos \theta} \cos(m \theta) d\theta. \quad (3.143)$$

⁷Ibid.

This solution, which is known as a *modified Bessel function*, has the properties that

$$I_m(p) \rightarrow \frac{1}{m!} \left(\frac{p}{2}\right)^m \quad \text{as } p \rightarrow 0, \quad (3.144)$$

$$I_m(p) \rightarrow \frac{e^p}{\sqrt{2\pi p}} \quad \text{as } p \rightarrow \infty. \quad (3.145)$$

In other words, at small arguments the function has a power-law behavior, whereas at large arguments it grows exponentially. It follows that

$$R(r) = I_m(k_n r). \quad (3.146)$$

Thus, our separable solution becomes

$$\phi(r, \theta, z) = \sum_{m=0, \infty} \sum_{n=1, \infty} \sin(k_n z) I_m(k_n z) [A_{mn} \cos(m\theta) + B_{mn} \sin(m\theta)]. \quad (3.147)$$

If we express the function $\Phi(\theta, z)$ as a Fourier series in θ and z , so that

$$\Phi(\theta, z) = \sum_{m=0, \infty} \sum_{n=1, \infty} \sin(k_n z) [C_{mn} \cos(m\theta) + S_{mn} \sin(m\theta)], \quad (3.148)$$

then the boundary condition (3.135) yields

$$A_{mn} = \frac{C_{mn}}{I_m(k_n a)}, \quad (3.149)$$

$$B_{mn} = \frac{S_{mn}}{I_m(k_n a)}. \quad (3.150)$$

Hence, our solution is fully specified.

3.11 Poisson's Equation in Cylindrical Coordinates

Let us, finally, consider the solution of Poisson's equation,

$$\nabla^2 \phi = -\frac{\rho}{\epsilon_0}, \quad (3.151)$$

in cylindrical coordinates. Suppose that the domain of solution extends over all space, and the potential is subject to the simple boundary condition

$$\phi(\mathbf{r}) \rightarrow 0 \quad \text{as } |\mathbf{r}| \rightarrow \infty. \quad (3.152)$$

In this case, the solution is written (see Section 2.3)

$$\phi(\mathbf{r}) = - \int \frac{\rho(\mathbf{r}')}{\epsilon_0} G(\mathbf{r}, \mathbf{r}') dV', \quad (3.153)$$

where the integral is over all space, and $\mathbf{G}(\mathbf{r}, \mathbf{r}')$ is a symmetric Green's function [i.e., $\mathbf{G}(\mathbf{r}', \mathbf{r}) = G(\mathbf{r}, \mathbf{r}')$ —see Equation (2.17)] that satisfies

$$\nabla^2 G(\mathbf{r}, \mathbf{r}') = \delta(\mathbf{r} - \mathbf{r}'), \quad (3.154)$$

subject to the constraint [see Equation (2.17)]

$$G(\mathbf{r}, \mathbf{r}') \rightarrow 0 \quad \text{as } |\mathbf{r}| \rightarrow \infty. \quad (3.155)$$

In cylindrical coordinates,

$$\delta(\mathbf{r} - \mathbf{r}') = \frac{1}{r} \delta(r - r') \delta(\theta - \theta') \delta(z - z'). \quad (3.156)$$

This follows because, by definition (see Section 1.5),

$$\int_V \delta(\mathbf{r} - \mathbf{r}') dV = \int_V \delta(\mathbf{r} - \mathbf{r}') r dr d\theta dz = 1 \quad (3.157)$$

whenever \mathbf{r}' lies within the volume V . Thus, Equation (3.154) becomes

$$\frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial G}{\partial r} \right) + \frac{1}{r^2} \frac{\partial^2 G}{\partial \theta^2} + \frac{\partial^2 G}{\partial z^2} = \frac{1}{r} \delta(r - r') \delta(\theta - \theta') \delta(z - z'). \quad (3.158)$$

The well-known mathematical identities

$$\delta(\theta - \theta') = \frac{1}{2\pi} \sum_{m=-\infty, \infty} e^{im(\theta - \theta')}, \quad (3.159)$$

$$\delta(z - z') = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ik(z - z')} dk, \quad (3.160)$$

are conventionally used to invert Fourier series and Fourier transforms, respectively. In the present case, if we write

$$G(\mathbf{r}, \mathbf{r}') = \frac{1}{4\pi^2} \sum_{m=-\infty, \infty} \int_{-\infty}^{\infty} e^{ik(z - z')} e^{im(\theta - \theta')} g_m(r, r') dk \quad (3.161)$$

then, making use of these identities, Equation (3.158) becomes

$$\frac{1}{r} \frac{d}{dr} \left(r \frac{dg_m}{dr} \right) - \left(k^2 + \frac{m^2}{r^2} \right) g_m = \frac{1}{r} \delta(r - r'). \quad (3.162)$$

In the general case, when $r \neq r'$, the previous equation reduces to the modified Bessel equation,

$$\frac{1}{r} \frac{d}{dr} \left(r \frac{dg_m}{dr} \right) - \left(k^2 + \frac{m^2}{r^2} \right) g_m = 0. \quad (3.163)$$

As we saw in Section 3.10, the modified Bessel function $I_m(kr)$ [defined in Equation (3.143)] is a solution of the modified Bessel equation that is well behaved at $r = 0$, and badly behaved as $r \rightarrow \infty$. On the other hand, the modified Bessel function $K_m(kr)$, where⁸

$$K_m(p) = \int_0^\infty e^{-p \cosh t} \cosh(mt) dt, \quad (3.164)$$

is a solution that is badly behaved at $r = 0$, and well behaved as $r \rightarrow \infty$. In fact,

$$K_m(p) \rightarrow \infty \quad \text{as } p \rightarrow 0, \quad (3.165)$$

$$K_m(p) \rightarrow \sqrt{\frac{\pi}{2p}} e^{-p} \quad \text{as } p \rightarrow \infty. \quad (3.166)$$

We are searching for a solution of Equation (3.162) that is well behaved at $r = 0$ (because there is no reason for the potential to be infinite at $r = 0$) and goes to zero as $r \rightarrow \infty$, in accordance with the constraint (3.155). It follows that

$$g_m(r, r') = \begin{cases} \alpha(r') I_m(kr) & r < r' \\ \beta(r') K_m(kr) & r > r' \end{cases}. \quad (3.167)$$

However, given that $G(\mathbf{r}, \mathbf{r}')$ is a symmetric function, we expect $g_m(r, r')$ to also be symmetric: that is, $g_m(r', r) = g_m(r, r')$. Consequently,

$$g_m(r, r') = A I_m(kr_{<}) K_m(kr_{>}), \quad (3.168)$$

where $r_{<}$ is the lesser of r and r' , and $r_{>}$ the greater. Integration of Equation (3.162) across $r = r'$ yields

$$\left[\frac{dg_m}{dr} \right]_{r=r'_-}^{r=r'_+} = \frac{1}{r'}, \quad (3.169)$$

which implies that

$$A k [K'_m(kr') I_m(kr') - K_m(kr') I'_m(kr')] = \frac{1}{r'}, \quad (3.170)$$

where $'$ denotes differentiation with respect to argument. However, the modified Bessel functions $I_m(p)$ and $K_m(p)$ satisfy the well-known mathematical identity⁹

$$K_m(p) I'_m(p) - K'_m(p) I_m(p) = \frac{1}{p}. \quad (3.171)$$

Hence, we deduce that $A = -1$. Thus, our general Green's function becomes

$$G(\mathbf{r}, \mathbf{r}') = -\frac{1}{4\pi^2} \sum_{m=-\infty, \infty} \int_{-\infty}^{\infty} e^{ik(z-z')} e^{im(\theta-\theta')} I_m(kr_{<}) K_m(kr_{>}) dk. \quad (3.172)$$

⁸Ibid.

⁹Ibid.

The previous expression for the Green's function, in combination with Equation (3.153), leads to the following expressions for the general solution to Poisson's equation in cylindrical geometry, subject to the boundary condition (3.152):

$$\phi(r, \theta, z) = \sum_{m=-\infty, \infty} \phi_m(r, z) e^{im\theta}, \quad (3.173)$$

$$\phi_m(r, z) = \int_{-\infty}^{\infty} \Phi_m(r, k) e^{ikz} dk, \quad (3.174)$$

$$\Phi_m(r, k) = K_m(kr) \int_0^r R_m(r', k) I_m(kr') r' dr' + I_m(kr) \int_r^{\infty} R_m(r', k) K_m(kr') r' dr', \quad (3.175)$$

$$R_m(r', k) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \rho_m(r', z') e^{-ikz'} dz', \quad (3.176)$$

$$\rho_m(r', z') = \frac{1}{2\pi} \oint \frac{\rho(r', \theta', z')}{\epsilon_0} e^{-im\theta'} d\theta'. \quad (3.177)$$

Suppose that we wish to solve Poisson's equation within a finite cylindrical volume, V , bounded by the surfaces $z = 0$, $z = L$, and $r = a$. Let the boundary conditions imposed at the surface be

$$\phi(r, \theta, 0) = 0, \quad (3.178)$$

$$\phi(r, \theta, L) = 0, \quad (3.179)$$

$$\phi(a, \theta, z) = \Phi(\theta, z), \quad (3.180)$$

where $\Phi(r, \theta)$ is a specified function. According to Section 2.10, the solution to this Dirichlet problem is written

$$\phi(\mathbf{r}) = - \int_V G(\mathbf{r}, \mathbf{r}') \frac{\rho(\mathbf{r}')}{\epsilon_0} dV' + \int_S \phi(\mathbf{r}') \frac{\partial G(\mathbf{r}, \mathbf{r}')}{\partial n'} dS', \quad (3.181)$$

where S represents the bounding surface. Here, the Green's function is the symmetric solution to

$$\nabla^2 G(\mathbf{r}, \mathbf{r}') = \delta(\mathbf{r} - \mathbf{r}') \quad (3.182)$$

that satisfies

$$G(\mathbf{r}, \mathbf{r}') = 0 \quad (3.183)$$

when \mathbf{r} (or \mathbf{r}') lies on S .

As before, in cylindrical coordinates, Equation (3.182) is written

$$\frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial G}{\partial r} \right) + \frac{1}{r^2} \frac{\partial^2 G}{\partial \theta^2} + \frac{\partial^2 G}{\partial z^2} = \frac{1}{r} \delta(r - r') \delta(\theta - \theta') \delta(z - z'). \quad (3.184)$$

If we search for a separable solution of the form $R(r) Q(\theta) Z(z)$ then it is clear that

$$Z(z) = \sum_{n=1, \infty} Z_n \sin(k_n z), \quad (3.185)$$

where

$$k_n = n \frac{\pi}{L}, \quad (3.186)$$

is the appropriate expression for $Z(z)$ that satisfies the constraint $Z = 0$ when $z = 0$ and $z = L$. The Fourier series (3.185) can be inverted in the usual fashion to give

$$Z_n = \frac{2}{L} \int_0^L Z(z) \sin(k_n z) dz, \quad (3.187)$$

which implies that

$$\delta(z - z') = \frac{2}{L} \sum_{n=1, \infty} \sin(k_n z) \sin(k_n z'). \quad (3.188)$$

Thus, searching for a Green's function of the form

$$G(\mathbf{r}, \mathbf{r}') = \frac{1}{L\pi} \sum_{m=-\infty, \infty} \sum_{n=1, \infty} \sin(k_n z) \sin(k_n z') e^{im(\theta - \theta')} g_{m,n}(r, r'), \quad (3.189)$$

Equation (3.184) reduces to

$$\frac{1}{r} \frac{d}{dr} \left(r \frac{dg_{m,n}}{dr} \right) - \left(k_n^2 + \frac{m^2}{r^2} \right) g_{m,n} = \frac{1}{r} \delta(r - r'). \quad (3.190)$$

Of course, $g_{m,n}(r, r')$ must be well behaved at $r = 0$. Moreover, the constraint $G(\mathbf{r}, \mathbf{r}') = 0$ when $r = a$ implies that $g_{m,n}(a, r') = 0$. Hence,

$$g_{m,n}(r, r') = \begin{cases} \alpha(r') I_m(k_n r) & r < r' \\ \beta(r') [I_m(k_n r) K_m(k_n a) - I_m(k_n a) K_m(k_n r)] & r > r' \end{cases}. \quad (3.191)$$

Now, the Green's function must be continuous when $r = r'$ (otherwise, it would not be a symmetric function of \mathbf{r} and \mathbf{r}'): that is,

$$g_{m,n}(r = r'_+, r') = g_{m,n}(r = r'_-, r'). \quad (3.192)$$

This implies that

$$\alpha(r') I_m(k_n r') = \beta(r') [I_m(k_n r') K_m(k_n a) - I_m(k_n a) K_m(k_n r')]. \quad (3.193)$$

Integration of Equation (3.184) across $r = r'$ again gives (3.169), which leads to

$$\beta(r') = \frac{I_m(k_n r')}{I_m(k_n a)}, \quad (3.194)$$

where use has been made of Equations (3.171) and (3.193). It follows that

$$g_{m,n}(r, r') = - [I_m(k_n a) K_m(k_n r_>) - I_m(k_n r_>) K_m(k_n a)] \frac{I_m(k_n r_<)}{I_m(k_n a)}. \quad (3.195)$$

Our general expression for the Dirichlet Green's function becomes

$$G(\mathbf{r}, \mathbf{r}') = -\frac{1}{L\pi} \sum_{m=-\infty, \infty} \sum_{n=1, \infty} \sin(k_n z) \sin(k_n z') e^{im(\theta-\theta')} \\ [I_m(k_n a) K_m(k_n r_>) - I_m(k_n r_>) K_m(k_n a)] \frac{I_m(k_n r_<)}{I_m(k_n a)}. \quad (3.196)$$

It is easily demonstrated that

$$r' \left. \frac{\partial G(\mathbf{r}, \mathbf{r}')}{\partial r'} \right|_{r'=a} = \frac{1}{L\pi} \sum_{m=-\infty, \infty} \sum_{n=1, \infty} \sin(k_n z) \sin(k_n z') e^{im(\theta-\theta')} \frac{I_m(k_n r)}{I_m(k_n a)}. \quad (3.197)$$

Hence, making use of Equation (3.181), in combination with the previous two expressions, our general solution to the problem under discussion is specified by the following set of equations:

$$\phi(r, \theta, z) = \sum_{m=-\infty, \infty} \phi_m(r, z) e^{im\theta}, \quad (3.198)$$

$$\phi_m(r, z) = \sum_{n=0, \infty} \phi_{m,n}(r) \sin(k_n z), \quad (3.199)$$

$$\phi_{m,n}(r) = \left[K_m(k_n r) - \frac{I_m(k_n r) K_m(k_n a)}{I_m(k_n a)} \right] \int_0^r R_{m,n}(r') I_m(k_n r') r' dr' \\ + I_m(k_n r) \int_r^\infty R_{m,n}(r') \left[K_m(k_n r') - \frac{I_m(k_n r') K_m(k_n a)}{I_m(k_n a)} \right] r' dr' \\ + \frac{I_m(k_n r)}{I_m(k_n a)} \Phi_{m,n}, \quad (3.200)$$

$$R_{m,n}(r') = \frac{2}{L} \int_0^L \rho_m(r', z') \sin(k_n z') dz', \quad (3.201)$$

$$\rho_m(r', z') = \frac{1}{2\pi} \oint \frac{\rho(r', \theta', z')}{\epsilon_0} e^{-im\theta'} d\theta', \quad (3.202)$$

$$\Phi_{m,n} = \frac{2}{L} \int_0^L \Phi_m(z') \sin(k_n z') dz', \quad (3.203)$$

$$\Phi_m(z') = \frac{1}{2\pi} \oint \Phi(\theta', z') e^{-im\theta'} d\theta'. \quad (3.204)$$

3.12 Exercises

3.1 Two concentric spheres have radii a , b ($b > a$) and are each divided into two hemispheres by the same horizontal plane. The upper hemisphere of the inner sphere and the lower hemisphere of the outer sphere are maintained at potential V . The other hemispheres are at zero potential. Demonstrate that the potential in the region $a \leq r \leq b$ can be written

$$\phi(r, \theta) = \sum_{l=0, \infty} (\alpha_l r^l + \beta_l r^{-l-1}) P_l(\cos \theta),$$

where

$$\alpha_l = \frac{V}{2} \left[\frac{a^{l+1} - (-1)^l b^{l+1}}{a^{2l+1} - b^{2l+1}} \right] [P_{l-1}(0) - P_{l+1}(0)],$$

$$\beta_l = \frac{V}{2} \left[\frac{a^{-l} - (-1)^l b^{-l}}{a^{-2l-1} - b^{-2l-1}} \right] [P_{l-1}(0) - P_{l+1}(0)].$$

Here, r, θ, φ are conventional spherical coordinates whose origin coincides with the common center of the spheres, and are such that the dividing plane corresponds to $\theta = \pi/2$.

3.2 A spherical surface of radius R has charge uniformly distributed over its surface with density $Q/4\pi R^2$, except for a spherical cap at the north pole, defined by the cone $\theta = \alpha$. Here, r, θ, φ are conventional spherical coordinates whose origin coincides with the center of the surface.

(a) Show that the potential inside the spherical surface can be expressed as

$$\phi(r, \theta) = \frac{Q}{8\pi \epsilon_0} \sum_{l=0, \infty} \frac{1}{2l+1} [P_{l+1}(\cos \alpha) - P_{l-1}(\cos \alpha)] \frac{r^l}{R^{l+1}} P_l(\cos \theta),$$

where $P_{-1}(\cos \alpha) = -1$.

(b) Show that the electric field at the origin is

$$\mathbf{E}(\mathbf{0}) = \frac{Q}{16\pi \epsilon_0 R^2} \sin^2 \alpha \mathbf{e}_z.$$

(c) Show that in the limit $\alpha \rightarrow 0$,

$$\phi(r, \theta) \rightarrow \frac{Q}{4\pi \epsilon_0 R} - \frac{Q \alpha^2}{16\pi \epsilon_0 R} \sum_{l=0, \infty} \frac{r^l}{R^l} P_l(\cos \theta).$$

(d) Show that in the limit $\alpha \rightarrow \pi$,

$$\phi(r, \theta) \rightarrow \frac{Q(\pi - \alpha)^2}{16\pi \epsilon_0 R} \sum_{l=0, \infty} (-1)^l \frac{r^l}{R^l} P_l(\cos \theta).$$

3.3 The Dirichlet Green's function for the unbounded space between planes at $z = 0$ and $z = L$ allows a discussion of a point charge, or a distribution of charge, between parallel conducting planes held at zero potential.

(a) Using cylindrical coordinates, show that one form of the Green's function is

$$G(\mathbf{r}, \mathbf{r}') = -\frac{1}{\pi L} \sum_{n=1, \infty} \sum_{m=-\infty, \infty} \sin\left(\frac{n\pi}{L} z\right) \sin\left(\frac{n\pi}{L} z'\right) e^{im(\theta-\theta')}$$

$$I_m\left(\frac{n\pi}{L} r_{<}\right) K_m\left(\frac{n\pi}{L} r_{>}\right).$$

(b) Show that an alternative form of the Green's function is

$$G(\mathbf{r}, \mathbf{r}') = -\frac{1}{2\pi} \sum_{m=-\infty, \infty} \int_0^{\infty} \frac{\sinh(k z_{<}) \sinh[k(L - z_{>})]}{\sinh(kL)} J_m(kr) J_m(kr') e^{im(\theta - \theta')} dk.$$

3.4 From the results of the previous exercise, show that the potential due to a point charge q placed between two infinite parallel conducting planes held at zero potential can be written as

$$\phi(z, r) = \frac{q}{\pi \epsilon_0 L} \sum_{n=1, \infty} \sin\left(\frac{n\pi}{L} z_n\right) \sin\left(\frac{n\pi}{L} z\right) K_0\left(\frac{n\pi}{L} r\right),$$

where the planes are at $z = 0$ and $z = L$, and the charge is on the z -axis at $z = z_0$. Show that induced surface charge densities on the lower and upper planes are

$$\sigma_-(r) = -\frac{q}{\pi L} \sum_{n=1, \infty} \left(\frac{n\pi}{L}\right) \sin\left(\frac{n\pi}{L} z_0\right) K_0\left(\frac{n\pi}{L} r\right),$$

$$\sigma_+(r) = \frac{q}{\pi L} \sum_{n=1, \infty} \cos(n\pi) \left(\frac{n\pi}{L}\right) \sin\left(\frac{n\pi}{L} z_0\right) K_0\left(\frac{n\pi}{L} r\right),$$

respectively.

3.5 Show that the potential due to a conducting disk of radius a carrying a charge q is

$$\phi(r, z) = \frac{q}{4\pi \epsilon_0 a} \int_0^{\infty} e^{-k|z|} J_0(kr) \frac{\sin(ka)}{k} dk$$

in cylindrical coordinates (whose origin coincides with the center of the disk, and whose symmetry axis coincides with that of the disk.)

3.6 A conducting spherical shell of radius a is placed in a uniform electric field \mathbf{E} . Show that the force tending to separate two halves of the sphere across a diametral plane perpendicular to \mathbf{E} is given by

$$F = \frac{9}{4} \pi \epsilon_0 a^2 E^2.$$

4 Electrostatics in Dielectric Media

4.1 Polarization

The terrestrial environment is characterized by dielectric media (e.g., air and water) that are, for the most part, electrically neutral, because they are made up of neutral atoms and molecules. However, when the constituent atoms and molecules of such media are placed in an electric field they tend to polarize: that is, to develop electric dipole moments. Suppose that when a given neutral molecule is placed in an electric field, \mathbf{E} , the centre of charge of its constituent electrons (whose total charge is $-q$) is displaced by a distance $-\mathbf{r}$ with respect to the centre of charge of its nucleus (whose charge is $+q$). The dipole moment of the molecule is then $\mathbf{p} = q\mathbf{r}$. (See Section 2.7.) If there are N such molecules per unit volume then the electric polarization \mathbf{P} (i.e., the dipole moment per unit volume) is given by $\mathbf{P} = N\mathbf{p}$. More generally,

$$\mathbf{P}(\mathbf{r}) = \sum_i N_i \langle \mathbf{p}_i \rangle, \quad (4.1)$$

where $\langle \mathbf{p}_i \rangle$ is the average dipole moment of the i th type of molecule in the vicinity of point \mathbf{r} , and N_i is the average number of such molecules per unit volume at \mathbf{r} .

It is easily demonstrated [e.g., by integrating Equation (2.75) by parts, and then comparing the result with Equation (2.36)] that any divergence of the polarization field, $\mathbf{P}(\mathbf{r})$, gives rise to a charge density, $\rho_b(\mathbf{r})$, in the medium. In fact,

$$\rho_b = -\nabla \cdot \mathbf{P}. \quad (4.2)$$

This density is attributable to *bound charges* (i.e., charges that arise from the polarization of neutral atoms), and is usually distinguished from the charge density, ρ_f , due to *free charges*, which represents a net surplus or deficit of electrons in the medium. Thus, the total charge density, ρ , in the medium is

$$\rho = \rho_f + \rho_b. \quad (4.3)$$

It must be emphasized that both terms on the right-hand side of this equation represent real physical charge. Nevertheless, it is useful to make the distinction between bound and free charges, especially when it comes to working out the energy associated with electric fields in dielectric media.

Gauss' law takes the differential form

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0} = \frac{\rho_f + \rho_b}{\epsilon_0}. \quad (4.4)$$

This expression can be rearranged to give

$$\nabla \cdot \mathbf{D} = \rho_f, \quad (4.5)$$

where

$$\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P} \quad (4.6)$$

is termed the *electric displacement* (which should not be confused with dipole moment per unit area—see Section 2.8), and has the same dimensions as \mathbf{P} (i.e., dipole moment per unit volume). The divergence theorem tells us that

$$\oint_S \mathbf{D} \cdot d\mathbf{S} = \int_V \rho_f dV. \quad (4.7)$$

In other words, the flux of \mathbf{D} out of some closed surface S is equal to the total free charge enclosed within that surface. Unlike the electric field \mathbf{E} (which is the force acting on unit charge), or the polarization \mathbf{P} (which is the dipole moment per unit volume), the electric displacement \mathbf{D} has no clear physical meaning. The only reason for introducing this quantity is that it enables us to calculate electric fields in the presence of dielectric materials without first having to know the distribution of bound charges. However, this is only possible if we have a *constitutive relation* connecting \mathbf{E} and \mathbf{D} . It is conventional to assume that the induced polarization, \mathbf{P} , is directly proportional to the electric field, \mathbf{E} , so that

$$\mathbf{P} = \epsilon_0 \chi_e \mathbf{E}, \quad (4.8)$$

where χ_e is termed the medium's *electric susceptibility*. It follows that

$$\mathbf{D} = \epsilon_0 \epsilon \mathbf{E}, \quad (4.9)$$

where the dimensionless quantity

$$\epsilon = 1 + \chi_e \quad (4.10)$$

is known as the *relative dielectric constant* or *relative permittivity* of the medium. It follows from Equations (4.5) and (4.9) that

$$\nabla \cdot \mathbf{E} = \frac{\rho_f}{\epsilon_0 \epsilon}. \quad (4.11)$$

Thus, the electric fields produced by free charges in a dielectric medium are analogous to those produced by the same charges in a vacuum, except that they are reduced by a factor ϵ . This reduction can be understood in terms of a polarization of the medium's constituent atoms or molecules that produces electric fields in opposition to those of the free charges. One immediate consequence is that the capacitance of a capacitor is increased by a factor ϵ if the empty space between the electrodes is filled with a dielectric medium of dielectric constant ϵ (assuming that fringing fields can be neglected).

It must be understood that Equations (4.8)–(4.11) constitute an approximation that is generally found to hold under terrestrial conditions (provided the electric field-strength does not become too large) when dealing with isotropic media. For anisotropic media (e.g., crystals), Equation (4.9) generalizes to

$$\mathbf{D} = \epsilon_0 \boldsymbol{\epsilon} \cdot \mathbf{E}, \quad (4.12)$$

where $\boldsymbol{\epsilon}$ is a symmetric second-rank tensor known as the *dielectric tensor*. For strong electric fields, \mathbf{D} ceases to vary linearly with \mathbf{E} . Indeed, for sufficiently strong electric fields, neutral molecules are disrupted, and the medium becomes a plasma.

4.2 Boundary Conditions for \mathbf{E} and \mathbf{D}

If the region in the vicinity of a collection of free charges contains dielectric material of non-uniform dielectric constant then the electric field no longer has the same form as in a vacuum. Suppose, for example, that space is occupied by two dielectric media whose uniform dielectric constants are ϵ_1 and ϵ_2 . What are the matching conditions on \mathbf{E} and \mathbf{D} at the interface between the two media?

Imagine a Gaussian pill-box enclosing part of the interface. The thickness of the pill-box is allowed to tend towards zero, so that the only contribution to the outward flux of \mathbf{D} comes from its two flat faces. These faces are parallel to the interface, and lie in each of the two media. Their outward normals are $d\mathbf{S}_1$ (in medium 1) and $d\mathbf{S}_2$, where $d\mathbf{S}_1 = -d\mathbf{S}_2$. Assuming that there is no free charge inside the pill-box (which is reasonable in the limit that the volume of the box tends towards zero), Equation (4.7) yields

$$\mathbf{D}_1 \cdot d\mathbf{S}_1 + \mathbf{D}_2 \cdot d\mathbf{S}_2 = 0, \quad (4.13)$$

where \mathbf{D}_1 is the electric displacement in medium 1 at the interface with medium 2, et cetera. The above equation can be rewritten

$$(\mathbf{D}_2 - \mathbf{D}_1) \cdot \mathbf{n}_{21} = 0, \quad (4.14)$$

where \mathbf{n}_{21} is the normal to the interface, directed from medium 1 to medium 2. If the fields and charges are non-time-varying then Maxwell's equations yield

$$\nabla \times \mathbf{E} = \mathbf{0}, \quad (4.15)$$

which gives the familiar boundary condition (obtained by integrating around a small loop that straddles the interface)

$$(\mathbf{E}_2 - \mathbf{E}_1) \times \mathbf{n}_{21} = \mathbf{0}. \quad (4.16)$$

In other word, the normal component of the electric displacement, and the tangential component of the electric field, are both continuous across any interface between two dielectric media.

4.3 Boundary Value Problems with Dielectrics

Consider a point charge q embedded in a semi-infinite dielectric medium of dielectric constant ϵ_1 , and located a distance d from a plane interface that separates the first medium from another semi-infinite dielectric medium of dielectric constant ϵ_2 . Suppose that the interface coincides with the plane $z = 0$. We need to solve

$$\epsilon_1 \nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0} \quad (4.17)$$

in the region $z > 0$,

$$\epsilon_2 \nabla \cdot \mathbf{E} = 0 \quad (4.18)$$

in the region $z < 0$, and

$$\nabla \times \mathbf{E} = \mathbf{0} \quad (4.19)$$

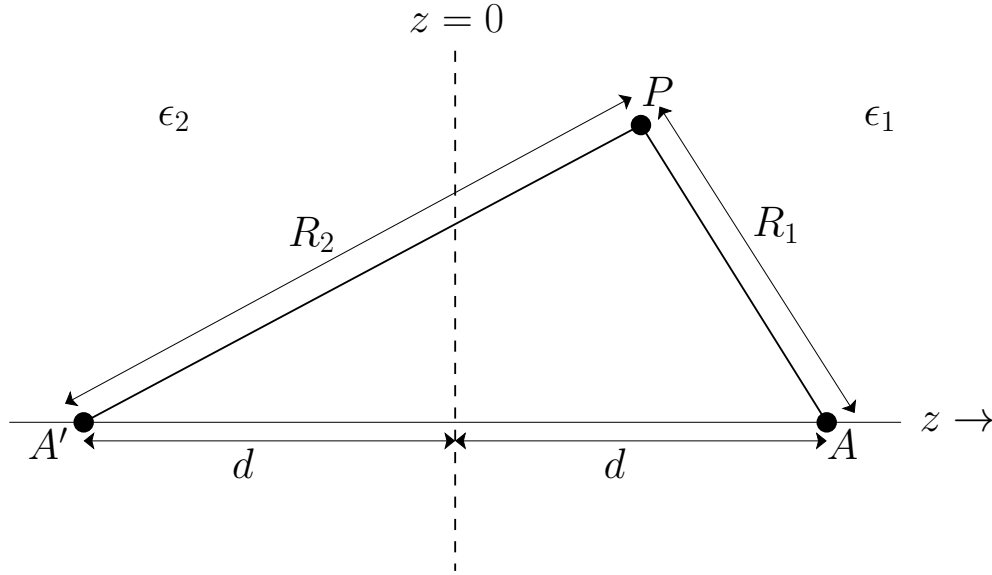


Figure 4.1: The method of images for a plane interface between two dielectric media.

everywhere, subject to the following constraints at $z = 0$:

$$\epsilon_1 E_z(z = 0_+) = \epsilon_2 E_z(z = 0_-), \quad (4.20)$$

$$E_x(z = 0_+) = E_x(z = 0_-), \quad (4.21)$$

$$E_y(z = 0_+) = E_y(z = 0_-). \quad (4.22)$$

In order to solve this problem, we shall employ a slightly modified form of the well-known method of images. Because $\nabla \times \mathbf{E} = 0$ everywhere, the electric field can be written in terms of a scalar potential: that is, $\mathbf{E} = -\nabla\phi$. Consider the region $z > 0$. Let us assume that the scalar potential in this region is the same as that obtained when the whole of space is filled with dielectric of dielectric constant ϵ_1 , and, in addition to the real charge q at position A , there is a second charge q' at the image position A' . (See Figure 4.1.) If this is the case then the potential at some point P in the region $z > 0$ is given by

$$\phi(z > 0) = \frac{1}{4\pi\epsilon_0\epsilon_1} \left(\frac{q}{R_1} + \frac{q'}{R_2} \right), \quad (4.23)$$

where $R_1 = \sqrt{r^2 + (d - z)^2}$ and $R_2 = \sqrt{r^2 + (d + z)^2}$. Here, r , θ , z are conventional cylindrical coordinates. Note that the potential (4.23) is clearly a solution of Equation (4.17) in the region $z > 0$: that is, it satisfies $\nabla \cdot \mathbf{E} = 0$, with the appropriate singularity at the position of the point charge q .

Consider the region $z < 0$. Let us assume that the scalar potential in this region is the same as that obtained when the whole of space is filled with a dielectric medium of dielectric constant ϵ_2 , and a charge q'' is located at the point A . If this is the case then the potential in this region is given

by

$$\phi(z < 0) = \frac{1}{4\pi\epsilon_0\epsilon_2} \frac{q''}{R_1}. \quad (4.24)$$

The above potential is clearly a solution of Equation (4.18) in the region $z < 0$: that is, it satisfies $\nabla \cdot \mathbf{E} = 0$, with no singularities.

It now remains to choose q' and q'' in such a manner that the constraints (4.20)–(4.22) are satisfied. The constraints (4.21) and (4.22) are obviously satisfied if the scalar potential is continuous across the interface between the two media: that is,

$$\phi(z = 0_+) = \phi(z = 0_-). \quad (4.25)$$

The constraint (4.20) implies a jump in the normal derivative of the scalar potential across the interface. In fact,

$$\epsilon_1 \frac{\partial\phi(z = 0_+)}{\partial z} = \epsilon_2 \frac{\partial\phi(z = 0_-)}{\partial z}. \quad (4.26)$$

The first matching condition yields

$$\frac{q + q'}{\epsilon_1} = \frac{q''}{\epsilon_2}, \quad (4.27)$$

whereas the second gives

$$q - q' = q''. \quad (4.28)$$

Here, use has been made of

$$\frac{\partial}{\partial z} \left(\frac{1}{R_1} \right)_{z=0} = -\frac{\partial}{\partial z} \left(\frac{1}{R_2} \right)_{z=0} = \frac{d}{(r^2 + d^2)^{3/2}}. \quad (4.29)$$

Equations (4.27) and (4.28) imply that

$$q' = -\left(\frac{\epsilon_2 - \epsilon_1}{\epsilon_2 + \epsilon_1} \right) q, \quad (4.30)$$

$$q'' = \left(\frac{2\epsilon_2}{\epsilon_2 + \epsilon_1} \right) q. \quad (4.31)$$

The polarization charge density is given by $\rho_b = -\nabla \cdot \mathbf{P}$. However, $\mathbf{P} = \epsilon_0 \chi_e \mathbf{E}$ inside either dielectric, which implies that

$$\nabla \cdot \mathbf{P} = \epsilon_0 \chi_e \nabla \cdot \mathbf{E} = 0, \quad (4.32)$$

except at the point charge q . Thus, there is zero bound charge density in either dielectric medium. At the interface, χ_e jumps discontinuously,

$$\Delta\chi_e = \epsilon_1 - \epsilon_2. \quad (4.33)$$

This implies that there is a bound charge sheet on the interface between the two dielectric media. In fact, it follows from Equation (4.2) that

$$\sigma_b = -(\mathbf{P}_2 - \mathbf{P}_1) \cdot \mathbf{n}_{21}, \quad (4.34)$$

where \mathbf{n}_{21} is a unit normal to the interface pointing from medium 1 to medium 2 (i.e., along the positive z -axis). Because

$$\mathbf{P}_i = \epsilon_0 (\epsilon_i - 1) \mathbf{E} = -\epsilon_0 (\epsilon_i - 1) \nabla \phi \quad (4.35)$$

in either medium, it is easily demonstrated that

$$\sigma_b(r) = -\frac{q}{2\pi \epsilon_1} \left(\frac{\epsilon_2 - \epsilon_1}{\epsilon_2 + \epsilon_1} \right) \frac{d}{(r^2 + d^2)^{3/2}}. \quad (4.36)$$

In the limit $\epsilon_2 \gg \epsilon_1$, the dielectric ϵ_2 behaves like a conducting medium (i.e., $\mathbf{E} \rightarrow \mathbf{0}$ in the region $z < 0$), and the bound surface charge density on the interface approaches that obtained in the case when the plane $z = 0$ coincides with a conducting surface.

The above argument can easily be generalized to deal with problems involving multiple point charges in the presence of multiple dielectric media whose interfaces form parallel planes.

Consider a second boundary value problem in which a slab of dielectric, of dielectric constant ϵ , lies between the planes $z = 0$ and $z = b$. Suppose that this slab is placed in a uniform z -directed electric field of strength E_0 . Let us calculate the field-strength E_1 inside the slab.

Because there are no free charges, and this is essentially a one-dimensional problem, it is clear from Equation (4.5) that the electric displacement D is the same in both the dielectric slab and the surrounding vacuum. In the vacuum region, $D = \epsilon_0 E_0$, whereas $D = \epsilon_0 \epsilon E_1$ in the dielectric. It follows that

$$E_1 = \frac{E_0}{\epsilon}. \quad (4.37)$$

In other words, the electric field inside the slab is reduced by polarization charges. As before, there is zero polarization charge density inside the dielectric. However, there is a uniform bound charge sheet on both surfaces of the slab. It is easily demonstrated that

$$\sigma_b(z = b) = -\sigma_b(z = 0) = \epsilon_0 \left(\frac{\epsilon - 1}{\epsilon} \right) E_0. \quad (4.38)$$

In the limit $\epsilon \gg 1$, the slab acts like a conductor, and $E_1 \rightarrow 0$.

Let us now generalize this result. Consider a dielectric medium whose dielectric constant ϵ varies with z . The medium is assumed to be of finite extent, and to be surrounded by a vacuum. It follows that $\epsilon(z) \rightarrow 1$ as $|z| \rightarrow \infty$. Suppose that this dielectric is placed in a uniform z -directed electric field of strength E_0 . What is the field $E(z)$ inside the dielectric?

We know that the electric displacement inside the dielectric is given by $D(z) = \epsilon_0 \epsilon(z) E(z)$. We also know from Equation (4.5) that, because there are no free charges, and this is essentially a one-dimensional problem,

$$\frac{dD(z)}{dz} = \epsilon_0 \frac{d[\epsilon(z) E(z)]}{dz} = 0. \quad (4.39)$$

Furthermore, $E(z) \rightarrow E_0$ as $|z| \rightarrow \infty$. It follows that

$$E(z) = \frac{E_0}{\epsilon(z)}. \quad (4.40)$$

Thus, the electric field is inversely proportional to the dielectric constant of the medium. The bound charge density within the medium is given by

$$\rho_b = \epsilon_0 \frac{dE(z)}{dz} = \epsilon_0 E_0 \frac{d}{dz} \left[\frac{1}{\epsilon(z)} \right]. \quad (4.41)$$

Consider a third, and final, boundary value problem in which a dielectric sphere of radius a , and dielectric constant ϵ , is placed in a z -directed electric field of strength E_0 (in the absence of the sphere). Let us calculate the electric field inside and around the sphere.

Because this is a static problem, we can write $\mathbf{E} = -\nabla\phi$. There are no free charges, so Equations (4.5) and (4.9) imply that

$$\nabla^2\phi = 0 \quad (4.42)$$

everywhere. The matching conditions (4.14) and (4.16) reduce to

$$\epsilon \left. \frac{\partial\phi}{\partial r} \right|_{r=a_-} = \left. \frac{\partial\phi}{\partial r} \right|_{r=a_+}, \quad (4.43)$$

$$\left. \frac{\partial\phi}{\partial\theta} \right|_{r=a_-} = \left. \frac{\partial\phi}{\partial\theta} \right|_{r=a_+}. \quad (4.44)$$

Furthermore,

$$\phi(r, \theta, \varphi) \rightarrow -E_0 r \cos\theta \quad (4.45)$$

as $r \rightarrow 0$: that is, the electric field asymptotes to uniform z -directed field of strength E_0 far from the sphere. Here, r, θ, φ are spherical coordinates centered on the sphere.

Let us search for an axisymmetric solution, $\phi = \phi(r, \theta)$. Because the solutions to Poisson's equation are unique, we know that if we can find such a solution that satisfies all of the boundary conditions then we can be sure that this is the correct solution. Equation (4.42) reduces to

$$\frac{1}{r} \frac{\partial^2(r\phi)}{\partial r^2} + \frac{1}{r^2 \sin\theta} \frac{\partial}{\partial\theta} \left(\sin\theta \frac{\partial\phi}{\partial\theta} \right) = 0. \quad (4.46)$$

Straightforward separation of the variables yields (see Section 3.7)

$$\phi(r, \theta) = \sum_{l=0, \infty} \left[A_l r^l + B_l r^{-(l+1)} \right] P_l(\cos\theta), \quad (4.47)$$

where l is a non-negative integer, the A_l and B_l are arbitrary constants, and the $P_l(x)$ are Legendre polynomials. (See Section 3.2.)

The Legendre polynomials form a complete set of angular functions, and it is easily demonstrated that the r^l and the $r^{-(l+1)}$ form a complete set of radial functions. It follows that Equation (4.47), with the A_l and B_l unspecified, represents a completely general (single-valued) axisymmetric solution to Equation (4.42). It remains to determine the values of the A_l and B_l that are consistent with the boundary conditions.

Let us divide space into the regions $r \leq a$ and $r > a$. In the former region

$$\phi(r, \theta) = \sum_{l=0, \infty} A_l r^l P_l(\cos\theta), \quad (4.48)$$

where we have rejected the $r^{-(l+1)}$ radial solutions because they diverge unphysically as $r \rightarrow 0$. In the latter region

$$\phi(r, \theta) = \sum_{l=0, \infty} [B_l r^l + C_l r^{-(l+1)}] P_l(\cos \theta). \quad (4.49)$$

However, it is clear from the boundary condition (4.45) that the only non-vanishing B_l is $B_1 = -E_0$. This follows because $P_1(\cos \theta) = \cos \theta$. The boundary condition (4.44) [which can be integrated to give $\phi(r = a_-) = \phi(r = a_+)$ for a potential that is single-valued in θ] gives

$$A_1 = -E_0 + \frac{C_1}{a^3}, \quad (4.50)$$

and

$$A_l = \frac{C_l}{a^{2l+1}} \quad (4.51)$$

for $l \neq 1$. Note that it is appropriate to match the coefficients of the $P_l(\cos \theta)$ because these functions are mutually orthogonal. (See Section 3.2.) The boundary condition (4.43) yields

$$\epsilon A_1 = -E_0 - 2 \frac{C_1}{a^3}, \quad (4.52)$$

and

$$\epsilon l A_l = -(l+1) \frac{C_l}{a^{2l+1}} \quad (4.53)$$

for $l \neq 1$. Equations (4.51) and (4.53) give $A_l = C_l = 0$ for $l \neq 1$. Equations (4.50) and (4.52) reduce to

$$A_1 = -\left(\frac{3}{2 + \epsilon}\right) E_0, \quad (4.54)$$

$$C_1 = \left(\frac{\epsilon - 1}{\epsilon + 2}\right) a^3 E_0. \quad (4.55)$$

The solution to the problem is therefore

$$\phi(r, \theta) = -\left(\frac{3}{2 + \epsilon}\right) E_0 r \cos \theta \quad (4.56)$$

for $r \leq a$, and

$$\phi(r, \theta) = -E_0 r \cos \theta + \left(\frac{\epsilon - 1}{\epsilon + 2}\right) E_0 \frac{a^3}{r^2} \cos \theta \quad (4.57)$$

for $r > a$.

Equation (4.56) is the potential of a uniform z -directed electric field of strength

$$E_1 = \frac{3}{2 + \epsilon} E_0. \quad (4.58)$$

Note that $E_1 < E_0$, provided that $\epsilon > 1$. Thus, the electric field-strength is reduced inside the dielectric sphere due to partial shielding by polarization charges. Outside the sphere, the potential

is equivalent to that of the applied field E_0 , plus the field of an electric dipole (see Section 2.7), located at the origin, and directed along the z -axis, whose dipole moment is

$$p = 4\pi \epsilon_0 \left(\frac{\epsilon - 1}{\epsilon + 2} \right) a^3 E_0. \quad (4.59)$$

This dipole moment can be interpreted as the volume integral of the polarization \mathbf{P} over the sphere. The polarization is

$$\mathbf{P} = \epsilon_0 (\epsilon - 1) E_1 \mathbf{e}_z = 3 \epsilon_0 \left(\frac{\epsilon - 1}{\epsilon + 2} \right) E_0 \mathbf{e}_z. \quad (4.60)$$

Because the polarization is uniform there is zero bound charge density inside the sphere. However, there is a bound charge sheet on the surface of the sphere, whose density is given by $\sigma_b = \mathbf{P} \cdot \mathbf{e}_r$ [see Equation (4.34)]. It follows that

$$\sigma_b(\theta) = 3 \epsilon_0 \left(\frac{\epsilon - 1}{\epsilon + 2} \right) E_0 \cos \theta. \quad (4.61)$$

The problem of a dielectric cavity of radius a inside a dielectric medium of dielectric constant ϵ , and in the presence of an applied electric field E_0 , parallel to the z -axis, can be treated in much the same manner as that of a dielectric sphere. In fact, it is easily demonstrated that the results for the cavity can be obtained from those for the sphere by making the transformation $\epsilon \rightarrow 1/\epsilon$. Thus, the field inside the cavity is uniform, parallel to the z -axis, and of magnitude

$$E_1 = \frac{3\epsilon}{2\epsilon + 1} E_0. \quad (4.62)$$

Note that $E_1 > E_0$, provided that $\epsilon > 1$. The field outside the cavity is the original field, plus that of a z -directed dipole, located at the origin, whose dipole moment is

$$p = -4\pi \epsilon_0 \left(\frac{\epsilon - 1}{2\epsilon + 1} \right) a^3 E_0. \quad (4.63)$$

Here, the negative sign implies that the dipole points in the opposite direction to the external field.

4.4 Energy Density Within Dielectric Medium

Consider a system of free charges embedded in a dielectric medium. The increase in the total energy when a small amount of free charge $\delta\rho_f$ is added to the system is given by

$$\delta U = \int_V \phi \delta\rho_f dV, \quad (4.64)$$

where the integral is taken over all space, and $\phi(\mathbf{r})$ is the electrostatic potential. Here, it is assumed that the original charges and the dielectric are held fixed, so that no mechanical work is performed. It follows from Equation (4.5) that

$$\delta U = \int_V \phi \nabla \cdot \delta\mathbf{D} dV, \quad (4.65)$$

where $\delta\mathbf{D}$ is the change in the electric displacement associated with the charge increment. Now, the above equation can also be written

$$\delta U = \int_V \nabla \cdot (\phi \delta\mathbf{D}) dV - \int_V \nabla\phi \cdot \delta\mathbf{D} dV, \quad (4.66)$$

giving

$$\delta U = \int_S \phi \delta\mathbf{D} \cdot d\mathbf{S} - \int_V \nabla\phi \cdot \delta\mathbf{D} dV, \quad (4.67)$$

where use has been made of the divergence theorem. If the dielectric medium is of finite spatial extent then we can neglect the surface term to give

$$\delta U = - \int_V \nabla\phi \cdot \delta\mathbf{D} dV = \int_V \mathbf{E} \cdot \delta\mathbf{D} dV. \quad (4.68)$$

This energy increment cannot be integrated unless \mathbf{E} is a known function of \mathbf{D} . Let us adopt the conventional approach, and assume that $\mathbf{D} = \epsilon_0 \epsilon \mathbf{E}$, where the dielectric constant ϵ is independent of the electric field. The change in energy associated with taking the displacement field from zero to $\mathbf{D}(\mathbf{r})$ at all points in space is given by

$$U = \int_0^{\mathbf{D}} \delta U = \int_0^{\mathbf{D}} \int_V \mathbf{E} \cdot \delta\mathbf{D} dV, \quad (4.69)$$

or

$$U = \int_V \int_0^E \frac{\epsilon_0 \epsilon \delta(E^2)}{2} dV = \frac{1}{2} \int_V \epsilon_0 \epsilon E^2 dV, \quad (4.70)$$

which reduces to

$$U = \frac{1}{2} \int_V \mathbf{E} \cdot \mathbf{D} dV. \quad (4.71)$$

Thus, the electrostatic energy density inside a dielectric is given by

$$W = \frac{1}{2} \mathbf{E} \cdot \mathbf{D}. \quad (4.72)$$

This is a standard result that is often quoted in textbooks. Nevertheless, it is important to realize that the above formula is only valid in dielectric media in which the electric displacement, \mathbf{D} , varies linearly with the electric field, \mathbf{E} .

4.5 Force Density Within Dielectric Medium

Equation (4.71) was derived by considering a virtual process in which true charges are added to a system of charges and dielectrics that are held fixed, so that no mechanical work is done against physical displacements. Consider a different virtual process in which the physical coordinates of the charges and dielectric are given a virtual displacement $\delta\mathbf{r}$ at each point in space, but no free charges are added to the system. Because we are dealing with a conservative system, the energy

expression (4.71) can still be employed, despite the fact that it was derived in terms of another virtual process. The variation in the total electrostatic energy δU when the system undergoes a virtual displacement $\delta \mathbf{r}$ is related to the electrostatic force density, \mathbf{f} , acting within the dielectric medium via

$$\delta U = - \int_V \mathbf{f} \cdot \delta \mathbf{r} dV. \quad (4.73)$$

If the medium is moving with a velocity field \mathbf{u} then the rate at which electrostatic energy is drained from the \mathbf{E} and \mathbf{D} fields is given by

$$\frac{dU}{dt} = - \int_V \mathbf{f} \cdot \mathbf{u} dV. \quad (4.74)$$

Consider the energy increment due to a change, $\delta \rho_f$, in the free charge distribution, and a change, $\delta \epsilon$, in the dielectric constant, which are both assumed to be caused by the virtual displacement. From Equation (4.71),

$$\delta U = \frac{1}{2 \epsilon_0} \int_V [D^2 \delta(1/\epsilon) + 2 \mathbf{D} \cdot \delta \mathbf{D} / \epsilon] dV, \quad (4.75)$$

or

$$\delta U = -\frac{\epsilon_0}{2} \int_V E^2 \delta \epsilon dV + \int_V \mathbf{E} \cdot \delta \mathbf{D} dV. \quad (4.76)$$

Here, the first term represents the energy increment due to the change in dielectric constant associated with the virtual displacement, whereas the second term corresponds to the energy increment caused by displacement of the free charges. The second term can be written

$$\int_V \mathbf{E} \cdot \delta \mathbf{D} dV = - \int_V \nabla \phi \cdot \delta \mathbf{D} dV = \int_V \phi \nabla \cdot \delta \mathbf{D} dV = \int_V \phi \delta \rho_f dV, \quad (4.77)$$

where surface terms have been neglected. Thus, Equation (4.76) implies that

$$\frac{dU}{dt} = \int_V \left(\phi \frac{\partial \rho_f}{\partial t} - \frac{\epsilon_0}{2} E^2 \frac{\partial \epsilon}{\partial t} \right) dV. \quad (4.78)$$

In order to arrive at an expression for the force density, \mathbf{f} , we need to express the time derivatives $\partial \rho / \partial t$ and $\partial \epsilon / \partial t$ in terms of the velocity field, \mathbf{u} . This can be achieved by adopting a dielectric equation of state: that is, a relation that specifies the dependence of the dielectric constant, ϵ , on the mass density, ρ_m . Let us assume that $\epsilon(\rho_m)$ is a known function. It follows that

$$\frac{D\epsilon}{Dt} = \frac{d\epsilon}{d\rho_m} \frac{D\rho_m}{Dt}, \quad (4.79)$$

where

$$\frac{D}{Dt} \equiv \frac{\partial}{\partial t} + \mathbf{u} \cdot \nabla \quad (4.80)$$

is the total time derivative (i.e., the time derivative in a frame of reference that is locally co-moving with the dielectric.) The hydrodynamic equation of continuity of the dielectric is

$$\frac{\partial \rho_m}{\partial t} + \nabla \cdot (\rho_m \mathbf{u}) = 0, \quad (4.81)$$

which implies that

$$\frac{D\rho_m}{Dt} = -\rho_m \nabla \cdot \mathbf{u}. \quad (4.82)$$

It follows that

$$\frac{\partial \epsilon}{\partial t} = -\frac{d\epsilon}{d\rho_m} \rho_m \nabla \cdot \mathbf{u} - \mathbf{u} \cdot \nabla \epsilon. \quad (4.83)$$

The conservation equation for the free charges is written

$$\frac{\partial \rho_f}{\partial t} + \nabla \cdot (\rho_f \mathbf{u}) = 0. \quad (4.84)$$

Thus, we can express Equation (4.78) in the form

$$\frac{dU}{dt} = \int_V \left[-\phi \nabla \cdot (\rho_f \mathbf{u}) + \frac{\epsilon_0}{2} E^2 \frac{d\epsilon}{d\rho_m} \rho_m \nabla \cdot \mathbf{u} + \frac{\epsilon_0}{2} E^2 \nabla \epsilon \cdot \mathbf{u} \right] dV. \quad (4.85)$$

Integrating the first term by parts, and neglecting any surface contributions, we obtain

$$- \int_V \phi \nabla \cdot (\rho_f \mathbf{u}) dV = \int_V \rho_f \nabla \phi \cdot \mathbf{u} dV. \quad (4.86)$$

Likewise,

$$\int_V \frac{\epsilon_0}{2} E^2 \frac{d\epsilon}{d\rho_m} \rho_m \nabla \cdot \mathbf{u} dV = - \int_V \frac{\epsilon_0}{2} \nabla \left(E^2 \frac{d\epsilon}{d\rho_m} \rho_m \right) \cdot \mathbf{u} dV. \quad (4.87)$$

Thus, Equation (4.85) becomes

$$\frac{dU}{dt} = \int_V \left[-\rho_f \mathbf{E} + \frac{\epsilon_0}{2} E^2 \nabla \epsilon - \frac{\epsilon_0}{2} \nabla \left(E^2 \frac{d\epsilon}{d\rho_m} \rho_m \right) \right] \cdot \mathbf{u} dV. \quad (4.88)$$

Comparing with Equation (4.74), we deduce that the force density inside the dielectric is given by

$$\mathbf{f} = \rho_f \mathbf{E} - \frac{\epsilon_0}{2} E^2 \nabla \epsilon + \frac{\epsilon_0}{2} \nabla \left(E^2 \frac{d\epsilon}{d\rho_m} \rho_m \right). \quad (4.89)$$

The first term in the above equation is the standard electrostatic force density (due to the presence of free charges). The second term represents a force that appears whenever an inhomogeneous dielectric is placed in an electric field. The last term, which is known as the *electrostriction* term, gives a force acting on a dielectric in an inhomogeneous electric field. Note that the magnitude of the electrostriction force density depends explicitly on the dielectric equation of state of the material, through $d\epsilon/d\rho_m$. The electrostriction term gives zero net force acting on any finite region of dielectric, provided we can integrate over a large enough portion of the dielectric that its extremities lie in a field-free region. For this reason, the term is frequently omitted, because in the calculation of the total forces acting on dielectric bodies it usually makes no contribution. Note, however, that if the electrostriction term is omitted then an incorrect pressure variation within the dielectric is obtained, even though the total force is given correctly.

4.6 Clausius-Mossotti Relation

Let us now investigate what a dielectric equation of state actually looks like. Suppose that a dielectric medium is made up of identical molecules that develop a dipole moment

$$\mathbf{p} = \alpha \epsilon_0 \mathbf{E} \quad (4.90)$$

when placed in an electric field \mathbf{E} . The constant α is called the *molecular polarizability*. If N is the number density of such molecules then the polarization of the medium is

$$\mathbf{P} = N \mathbf{p} = N \alpha \epsilon_0 \mathbf{E}, \quad (4.91)$$

or

$$\mathbf{P} = \frac{N_A \rho_m \alpha}{M} \epsilon_0 \mathbf{E}, \quad (4.92)$$

where ρ_m is the mass density, N_A is Avogadro's number, and M is the molecular weight. But, how does the electric field experienced by an individual molecule relate to the average electric field in the medium? This is not a trivial question because we expect the electric field to vary strongly (on atomic lengthscales) within the dielectric.

Suppose that the dielectric is polarized by a mean electric field \mathbf{E}_0 that is uniform (on macroscopic lengthscales), and directed along the z -axis. Consider one of the dielectric's constituent molecules. Let us draw a sphere of radius a around this particular molecule. The surface of the sphere is intended to represent the boundary between the microscopic and the macroscopic ranges of phenomena affecting the molecule. We shall treat the dielectric outside the sphere as a continuous medium, and the dielectric inside the sphere as a collection of polarized molecules. According to Equation (4.34), there is a bound surface charge of magnitude

$$\sigma_b(\theta) = -P \cos \theta \quad (4.93)$$

on the inside of the sphere's surface, where r, θ, φ are conventional spherical coordinates, and $\mathbf{P} = P \mathbf{e}_z = \epsilon_0 (\epsilon - 1) E_0 \mathbf{e}_z$ is the uniform polarization of the uniform dielectric outside the sphere. The magnitude of E_z at the molecule due to this surface charge is

$$E_z = -\frac{1}{4\pi \epsilon_0} \int_S \frac{\sigma_b(\theta) \cos \theta}{a^2} dS, \quad (4.94)$$

where $dS = 2\pi a^2 \sin \theta d\theta$ is an element of the surface. It follows that

$$E_z = \frac{P}{2\epsilon_0} \int_0^\pi \cos^2 \theta \sin \theta d\theta = \frac{P}{3\epsilon_0}. \quad (4.95)$$

It is easily demonstrated, from symmetry, that $E_\theta = E_\varphi = 0$ at the molecule. Thus, the field at the molecule due to the bound charges distributed on the inside of the sphere's surface is

$$\mathbf{E} = \frac{\mathbf{P}}{3\epsilon_0}. \quad (4.96)$$

The field due to the individual molecules within the sphere is obtained by summing over the dipole fields of these molecules. The electric field a distance \mathbf{r} from a dipole \mathbf{p} is (see Section 2.7)

$$\mathbf{E} = -\frac{1}{4\pi\epsilon_0} \left[\frac{\mathbf{p}}{r^3} - \frac{3(\mathbf{p} \cdot \mathbf{r})\mathbf{r}}{r^5} \right]. \quad (4.97)$$

It is assumed that the dipole moments of the molecules within the sphere are all the same, and also that the molecules are evenly distributed throughout the sphere. This being the case, the value of E_z at the molecule due to all of the other molecules within in the sphere,

$$E_z = -\frac{1}{4\pi\epsilon_0} \sum_{\text{mols}} \left[\frac{p_z}{r^3} - \frac{3(p_x x z + p_y y z + p_z z^2)}{r^5} \right], \quad (4.98)$$

is zero, because, for evenly distributed molecules,

$$\sum_{\text{mols}} x^2 = \sum_{\text{mols}} y^2 = \sum_{\text{mols}} z^2 = \frac{1}{3} \sum_{\text{mols}} r^2 \quad (4.99)$$

and

$$\sum_{\text{mols}} xy = \sum_{\text{mols}} yz = \sum_{\text{mols}} zx = 0. \quad (4.100)$$

It is also easily demonstrated that $E_\theta = E_\phi = 0$. Hence, the electric field at the molecule due to the other molecules within the sphere averages to zero.

It is clear that the net electric field experienced by an individual molecule is

$$\mathbf{E} = \mathbf{E}_0 + \frac{\mathbf{P}}{3\epsilon_0}, \quad (4.101)$$

which is larger than the average electric field, \mathbf{E}_0 , in the dielectric. The above analysis indicates that this effect is ascribable to the long range (rather than the short range) interactions of the molecule with the other molecules in the medium. Making use of Equation (4.92), as well as the definition $\mathbf{P} = \epsilon_0(\epsilon - 1)\mathbf{E}_0$, we obtain

$$\frac{\epsilon - 1}{\epsilon + 2} = \frac{N_A \rho_m \alpha}{3M}, \quad (4.102)$$

which is known as the *Clausius-Mossotti relation*. This expression is found to work very well for a wide class of dielectric liquids and gases. The Clausius-Mossotti relation also yields

$$\frac{d\epsilon}{d\rho_m} = \frac{(\epsilon - 1)(\epsilon + 2)}{3\rho_m}. \quad (4.103)$$

4.7 Dielectric Liquids in Electrostatic Fields

Consider the behavior of an uncharged dielectric liquid placed in an electrostatic field. If p is the pressure within the liquid when in equilibrium with the electrostatic force density \mathbf{f} then force balance requires that

$$\nabla p = \mathbf{f}. \quad (4.104)$$

It follows from Equation (4.89) that

$$\nabla p = -\frac{\epsilon_0}{2} E^2 \nabla \epsilon + \frac{\epsilon_0}{2} \nabla \left(E^2 \frac{d\epsilon}{d\rho_m} \rho_m \right) = \frac{\epsilon_0 \rho_m}{2} \nabla \left(E^2 \frac{d\epsilon}{d\rho_m} \right). \quad (4.105)$$

We can integrate this equation to give

$$\int_{p_1}^{p_2} \frac{dp}{\rho_m} = \frac{\epsilon_0}{2} \left(\left[E^2 \frac{d\epsilon}{d\rho_m} \right]_2 - \left[E^2 \frac{d\epsilon}{d\rho_m} \right]_1 \right), \quad (4.106)$$

where 1 and 2 refer to two general points in the liquid. Here, it is assumed that the liquid possesses an equation of state, so that $p = p(\rho_m)$. If the liquid is essentially incompressible (i.e., $\rho_m \approx \text{constant}$) then

$$p_2 - p_1 = \frac{\epsilon_0 \rho_m}{2} \left[E^2 \frac{d\epsilon}{d\rho_m} \right]_1^2. \quad (4.107)$$

Finally, if the liquid obeys the Clausius-Mossotti relation then

$$p_2 - p_1 = \left[\frac{\epsilon_0 E^2 (\epsilon - 1)(\epsilon + 2)}{2 \cdot 3} \right]_1^2. \quad (4.108)$$

According to Equations (4.58) and (4.108), if a sphere of dielectric liquid is placed in a uniform electric field \mathbf{E}_0 then the pressure inside the liquid takes the constant value

$$p = \frac{3}{2} \epsilon_0 E_0^2 \frac{\epsilon - 1}{\epsilon + 2}. \quad (4.109)$$

It is clear that the electrostatic forces acting on the dielectric are all concentrated at the edge of the sphere, and are directed radially inwards: that is, the dielectric is compressed by the external electric field. This is a somewhat surprising result because the electrostatic forces acting on a rigid conducting sphere are concentrated at the edge of the sphere, but are directed radially outwards. We might expect these two cases to give the same result in the limit $\epsilon \rightarrow \infty$. The reason that this does not occur is because a dielectric liquid is slightly compressible, and is, therefore, subject to an electrostriction force. There is no electrostriction force for the case of a completely rigid body. In fact, the force density inside a rigid dielectric (for which $\nabla \cdot \mathbf{u} = 0$) is given by Equation (4.89), with the third term (the electrostriction term) missing. It is easily demonstrated that the force exerted by an electric field on a rigid dielectric is directed outwards, and approaches that exerted on a rigid conductor in the limit $\epsilon \rightarrow \infty$.

As is well known, if a pair of charged (parallel plane) capacitor plates are dipped into a dielectric liquid then the liquid is drawn up between the plates to some extent. Let us examine this effect. We can, without loss of generality, assume that the transition from dielectric to vacuum takes place in a continuous manner. Consider the electrostatic pressure difference between a point A lying just above the surface of the liquid in between the plates, and a point B lying just above the surface of the liquid well away from the capacitor (where $E = 0$). The pressure difference is given by

$$p_A - p_B = - \int_A^B \mathbf{f} \cdot d\mathbf{r}. \quad (4.110)$$

Note, however, that the Clausius-Mossotti relation yields $d\epsilon/d\rho_m = 0$ at both A and B , because $\epsilon = 1$ in a vacuum [see Equation (4.103)]. Thus, it is clear from Equation (4.89) that the electrostriction term makes no contribution to the line integral (4.110). It follows that

$$p_A - p_B = \frac{\epsilon_0}{2} \int_A^B E^2 \nabla \epsilon \cdot d\mathbf{r}. \quad (4.111)$$

The only contribution to this integral comes from the vacuum/dielectric interface in the vicinity of point A (because ϵ is constant inside the liquid, and $E = 0$ in the vicinity of point B). Suppose that the electric field at point A has normal and tangential (to the surface) components E_\perp and E_\parallel , respectively. Making use of the boundary conditions that ϵE_\perp and E_\parallel are constant across a vacuum/dielectric interface, we obtain

$$p_A - p_B = \frac{\epsilon_0}{2} \left[E_\parallel^2 (\epsilon - 1) + \epsilon^2 E_\perp^2 (\epsilon) \int_1^\epsilon \frac{d\epsilon}{\epsilon^2} \right], \quad (4.112)$$

giving

$$p_A - p_B = \frac{\epsilon_0 (\epsilon - 1)}{2} \left[E_\parallel^2 + \frac{E_\perp^2}{\epsilon} \right]. \quad (4.113)$$

This electrostatic pressure difference can be equated to the hydrostatic pressure difference $\rho_m g h$ to determine the height, h , that the liquid rises between the plates. At first sight, the above analysis appears to suggest that the dielectric liquid is drawn upward by a surface force acting on the vacuum/dielectric interface in the region between the plates. In fact, this is far from being the case. A brief examination of Equation (4.108) shows that this surface force is actually directed downwards. According to Equation (4.89), the force which causes the liquid to rise between the plates is a volume force that develops in the region of non-uniform electric field at the base of the capacitor, where the field splays out between the plates. Thus, although we can determine the height to which the fluid rises between the plates without reference to the electrostriction force, it is, somewhat paradoxically, this force that is actually responsible for supporting the liquid against gravity.

Let us consider another paradox concerning the electrostatic forces exerted in a dielectric medium. Suppose that we have two charges embedded in a uniform dielectric of dielectric constant ϵ . The electric field generated by each charge is the same as that in a vacuum, except that it is reduced by a factor ϵ . We, therefore, expect the force exerted by one charge on the other to be the same as that in a vacuum, except that it is also reduced by a factor ϵ . Let us examine how this reduction in force comes about. Consider a simple example. Suppose that we take a parallel plate capacitor, and insert a block of solid dielectric between the plates. Suppose, further, that there is a small vacuum gap between the faces of the block and each of the capacitor plates. Let $\pm\sigma$ be the surface charge densities on each of the capacitor plates, and let $\pm\sigma_b$ be the bound charge densities that develop on the outer faces of the intervening dielectric block. The two layers of bound charge produce equal and opposite electric fields on each plate, and their effects therefore cancel each other. Thus, from the point of view of electrical interaction alone there would appear to be no change in the force exerted by one capacitor plate on the other when a dielectric slab is placed

between them (assuming that σ remains constant during this process). That is, the force per unit area (which is attractive) remains

$$f_s = \frac{\sigma^2}{2\epsilon_0}. \quad (4.114)$$

However, in experiments in which a capacitor is submerged in a dielectric liquid the force per unit area exerted by one plate on another is observed to decrease to

$$f_s = \frac{\sigma^2}{2\epsilon_0\epsilon}. \quad (4.115)$$

This apparent paradox can be explained by taking into account the difference in liquid pressure in the field-filled space between the plates, and the field-free region outside the capacitor. This pressure difference is balanced by internal elastic forces in the case of the solid dielectric discussed earlier, but is transmitted to the plates in the case of the liquid. We can compute the pressure difference between a point A on the inside surface of one of the capacitor plates, and a point B on the outside surface of the same plate using Equation (4.111). If we neglect end effects then the electric field is normal to the plates in the region between the plates, and is zero everywhere else. Thus, the only contribution to the line integral (4.111) comes from the plate/dielectric interface in the vicinity of point A . Using Equation (4.113), we find that

$$p_A - p_B = \frac{\epsilon_0}{2} \left(1 - \frac{1}{\epsilon}\right) E^2 = \frac{\sigma^2}{2\epsilon_0} \left(1 - \frac{1}{\epsilon}\right), \quad (4.116)$$

where E is the normal field-strength between the plates in the absence of dielectric. The sum of this pressure force and the purely electrical force (4.114) yields a net attractive force per unit area

$$f_s = \frac{\sigma^2}{2\epsilon_0\epsilon} \quad (4.117)$$

acting between the plates. Thus, any decrease in the forces exerted by charges on one another when they are immersed or embedded in a dielectric medium can only be understood in terms of mechanical forces transmitted between these charges by the medium itself.

4.8 Exercises

- 4.1 Starting from Equation (2.75), derive the result $\rho_b = -\nabla \cdot \mathbf{P}$.
- 4.2 Consider an electron of charge $-e$ moving in a circular orbit of radius a_0 about a charge $+e$ in a field directed at right angles to the plane of the orbit. Show that the polarizability α is approximately $4\pi a_0^3$.
- 4.3 A point charge q is located in free space a distance d from the center of a dielectric sphere of radius a ($a < d$) and dielectric constant ϵ . Find the potential at all points in space as an expansion in spherical harmonics. Calculate the rectangular components of the electric field in the vicinity of the center of the sphere.

- 4.4 A dielectric sphere of radius a and dielectric constant ϵ_1 is imbedded in an infinite dielectric block of dielectric constant ϵ_2 . The block is placed in a uniform electric field $\mathbf{E} = E_0 \mathbf{e}_z$. In other words, if $\epsilon_1 = \epsilon_2$ then the electric field would be $\mathbf{E} = E_0 \mathbf{e}_z$. Find the potential both inside and outside the sphere (assuming that $\epsilon_1 \neq \epsilon_2$), and the distribution of bound charges on the surface of the sphere.
- 4.5 An electric dipole of moment $\mathbf{p} = p \mathbf{e}_z$ lies at the center of a spherical cavity of radius a in a uniform dielectric material of relative dielectric constant ϵ . Find the electrostatic potential throughout space. Find the bound charge sheet density on the surface of the cavity.
- 4.6 A cylindrical coaxial cable consists of a thin inner conductor of radius a , surrounded by a dielectric sheath of dielectric constant ϵ_1 and outer radius b , surrounded by a second dielectric sheath of dielectric constant ϵ_2 and outer radius c , surrounded by a thin outer conductor. All components of the cable are touching. What is the capacitance per unit length of the cable?
- 4.7 A very long, right circular, cylindrical shell of dielectric constant ϵ and inner and outer radii a and b , respectively, is placed in a previously uniform electric field E_0 with its axis perpendicular to the field. The medium inside and outside the cylinder has a dielectric constant of unity. Determine the potential in the three regions, neglecting end effects. Discuss the limiting forms of your solutions for a solid dielectric cylinder in a uniform field, and a cylindrical cavity in a uniform dielectric.
- 4.8 Suppose that

$$\mathbf{D} = \epsilon_0 \boldsymbol{\epsilon} \cdot \mathbf{E},$$

where the dielectric tensor, $\boldsymbol{\epsilon}$, is constant (i.e., it is independent of \mathbf{E}). Demonstrate that

$$U = \int_0^{\mathbf{D}} \int_V \mathbf{E} \cdot \delta \mathbf{D} dV$$

can only be integrated to give

$$U = \frac{1}{2} \int_V \mathbf{E} \cdot \mathbf{D} dV$$

if $\boldsymbol{\epsilon}$ is symmetric. (Incidentally, because we generally expect a dielectric system to be conservative, this proves that $\boldsymbol{\epsilon}$ must be a symmetric tensor, otherwise the final energy of a dielectric system would not be independent of its past history.)

- 4.9 Show that for an *electret* (i.e., a material of fixed \mathbf{P}) the integral $\int \mathbf{E} \cdot \mathbf{D} dV$ over all space vanishes.
- 4.10 Two long, coaxial, cylindrical conducting surfaces of radii a and b ($b > a$) are lowered vertically into a liquid dielectric. If the liquid rises a mean height h between the electrodes when a potential difference V is established between them, show that the susceptibility of the liquid is

$$\chi_e = \frac{(b^2 - a^2) \rho_m g h \ln(b/a)}{\epsilon_0 V^2}$$

where ρ_m is the mass density of the liquid, g the acceleration due to gravity, and the susceptibility of air is neglected.

5 Magnetostatic Fields

5.1 Introduction

This chapter discusses magnetic fields generated by stationary current distributions. Such fields are conventionally termed *magnetostatic*.

5.2 Biot-Savart Law

According to the Biot-Savart law, the magnetic field generated at position vector \mathbf{r} by a current I_1 circulating around a thin loop, an element of which is located at position vector \mathbf{r}_1 , is

$$\mathbf{B}(\mathbf{r}) = \frac{\mu_0 I_1}{4\pi} \oint_1 \frac{d\mathbf{r}_1 \times (\mathbf{r} - \mathbf{r}_1)}{|\mathbf{r} - \mathbf{r}_1|^3}. \quad (5.1)$$

Suppose that a second current loop carries the current I_2 . The net magnetic force exerted on an element, $I_2 d\mathbf{r}_2$, of this loop, located at position vector \mathbf{r}_2 , is

$$d\mathbf{F}_{21} = I_2 d\mathbf{r}_2 \times \mathbf{B}(\mathbf{r}_2). \quad (5.2)$$

Hence, the net magnetic force exerted on loop 2 by loop 1 is

$$\mathbf{F}_{21} = \frac{\mu_0 I_1 I_2}{4\pi} \oint_1 \oint_2 \frac{d\mathbf{r}_2 \times (d\mathbf{r}_1 \times \mathbf{r}_{12})}{|\mathbf{r}_{12}|^3}, \quad (5.3)$$

where $\mathbf{r}_{12} = \mathbf{r}_2 - \mathbf{r}_1$.

5.3 Continuous Current Distribution

Making use of the fact that the magnetic fields generated by different current loops are superposable (see Section 1.2), Equation (5.1) can easily be generalized to deal with the magnetic field $\mathbf{B}(\mathbf{r})$ generated by a continuous current distribution of current density $\mathbf{j}(\mathbf{r})$. In fact,

$$\mathbf{B}(\mathbf{r}) = \frac{\mu_0}{4\pi} \int \frac{\mathbf{j}(\mathbf{r}') \times (\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^3} dV'. \quad (5.4)$$

For the case of a steady (i.e., $\partial/\partial t = 0$) current distribution, the charge conservation law (1.7) yields the constraint

$$\nabla \cdot \mathbf{j} = 0. \quad (5.5)$$

Given that [see Equation (2.27)]

$$\frac{\mathbf{r} - \mathbf{r}'}{|\mathbf{r} - \mathbf{r}'|^3} = -\nabla \left(\frac{1}{|\mathbf{r} - \mathbf{r}'|} \right), \quad (5.6)$$

Equation (5.4) can also be written

$$\mathbf{B} = \nabla \times \mathbf{A}, \quad (5.7)$$

where

$$\mathbf{A}(\mathbf{r}) = \frac{\mu_0}{4\pi} \int \frac{\mathbf{j}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} dV'. \quad (5.8)$$

Here, \mathbf{A} is termed the vector potential. (See Section 1.3.) It immediately follows that

$$\nabla \cdot \mathbf{B} = 0, \quad (5.9)$$

which is the second Maxwell equation. (See Section 1.2.) Now,

$$\begin{aligned} \nabla \cdot \mathbf{A}(\mathbf{r}) &= \frac{\mu_0}{4\pi} \int \mathbf{j}(\mathbf{r}') \cdot \nabla \left(\frac{1}{|\mathbf{r} - \mathbf{r}'|} \right) dV' = -\frac{\mu_0}{4\pi} \int \mathbf{j}(\mathbf{r}') \cdot \nabla' \left(\frac{1}{|\mathbf{r} - \mathbf{r}'|} \right) dV' \\ &= \frac{\mu_0}{4\pi} \int \frac{\nabla' \cdot \mathbf{j}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} dV', \end{aligned} \quad (5.10)$$

where we have integrated by parts, and neglected surface terms. Thus, according to Equation (5.5),

$$\nabla \cdot \mathbf{A} = 0. \quad (5.11)$$

In other words, the vector potential defined in Equation (5.8) automatically satisfies the time independent version of the Lorenz gauge condition, (1.13). Finally,

$$\nabla \times \mathbf{B} = \nabla \times (\nabla \times \mathbf{A}) \equiv \nabla(\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A} = -\nabla^2 \mathbf{A}, \quad (5.12)$$

where use has been made of Equations (5.7) and (5.11). It follows from Equations (5.8) and (1.25) that

$$\nabla \times \mathbf{B}(\mathbf{r}) = -\frac{\mu_0}{4\pi} \int \mathbf{j}(\mathbf{r}') \nabla^2 \left(\frac{1}{|\mathbf{r} - \mathbf{r}'|} \right) dV' = \mu_0 \int \mathbf{j}(\mathbf{r}') \delta(\mathbf{r} - \mathbf{r}') dV' = \mu_0 \mathbf{j}(\mathbf{r}), \quad (5.13)$$

or

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{j}, \quad (5.14)$$

which is the time independent form of the fourth Maxwell equation. (See Section 1.2.) The integral version of the previous equation, which follows from the curl theorem, is

$$\oint_C \mathbf{B} \cdot d\mathbf{r} = \mu_0 \int_S \mathbf{j} \cdot d\mathbf{S}. \quad (5.15)$$

This result is known as *Ampère's law*. Here, C is a closed curve spanned by a general surface S .

5.4 Circular Current Loop

Let us calculate the magnetic field generated by a thin circular loop of radius a , lying in the x - y plane, centered on the origin, and carrying the steady current I . Let r, θ, φ be spherical coordinates whose origin lies at the center of the loop, and whose symmetry axis is coincident with that of the loop. It follows that the distribution of current density in space is

$$\mathbf{j}(\mathbf{r}') = I \frac{\delta(r' - a)}{a} \delta(\cos \theta') \mathbf{e}_{\varphi'} = I \frac{\delta(r' - a)}{a} \delta(\cos \theta') (-\sin \varphi' \mathbf{e}_x + \cos \varphi' \mathbf{e}_y). \quad (5.16)$$

Because the geometry is cylindrically symmetric, we can, without loss of generality, choose the observation point to lie in the x - z plane (i.e., $\varphi = 0$). It follows from Equation (5.8) that

$$A_x = -\frac{\mu_0 I}{4\pi a} \int \frac{\sin \varphi' \delta(\cos \theta') \delta(r' - a)}{|\mathbf{r} - \mathbf{r}'|_{\varphi=0}} r'^2 dr' d\Omega', \quad (5.17)$$

$$A_y = \frac{\mu_0 I}{4\pi a} \int \frac{\cos \varphi' \delta(\cos \theta') \delta(r' - a)}{|\mathbf{r} - \mathbf{r}'|_{\varphi=0}} r'^2 dr' d\Omega', \quad (5.18)$$

$$A_z = 0, \quad (5.19)$$

where $|\mathbf{r} - \mathbf{r}'|_{\varphi=0} = [r^2 + r'^2 - 2rr'(\cos \theta \cos \theta' + \sin \theta \sin \theta' \cos \varphi')]^{1/2}$. It is clear that the integral for A_x averages to zero. Hence, only A_y , which corresponds to A_φ , is non-zero, and we can write

$$\mathbf{A} = A_\varphi(r, \theta) \mathbf{e}_\varphi, \quad (5.20)$$

where

$$A_\varphi(r, \theta) = \frac{\mu_0 I}{4\pi a} \int \frac{\cos \varphi' \delta(\cos \theta') \delta(r' - a)}{|\mathbf{r} - \mathbf{r}'|} r'^2 dr' d\Omega', \quad (5.21)$$

which reduces to

$$A_\varphi(r, \theta) = \frac{\mu_0 I a}{4\pi} \oint \frac{\cos \varphi' d\varphi'}{|\mathbf{r} - \mathbf{r}'|_{r'=a, \theta'=\pi/2, \varphi=0}} = \frac{\mu_0 I a}{4\pi} \oint \frac{\cos \varphi' d\varphi'}{(r^2 + a^2 - 2ar \sin \theta \cos \varphi')^{1/2}}. \quad (5.22)$$

The previous integral can be expressed in terms of complete elliptic integrals,¹ but this is not particularly illuminating. A better approach is to make use of the expansion of the Green's function for Poisson's equation in terms of spherical harmonics given in Section 3.5:

$$\frac{1}{4\pi |\mathbf{r} - \mathbf{r}'|} = \sum_{l=0, \infty} \sum_{m=-l, +l} \frac{1}{2l+1} \left(\frac{r_{<}^l}{r_{>}^{l+1}} \right) Y_{l,m}^*(\theta', \varphi') Y_{l,m}(\theta, \varphi). \quad (5.23)$$

Here, $r_{<}$ represents the lesser of r and r' , whereas $r_{>}$ represents the greater of r and r' . Hence,

$$A_\varphi(r, \theta) = \mu_0 I a \operatorname{Re} \sum_{l=0, \infty} \sum_{m=-l, +l} \frac{Y_{l,m}(\theta, 0)}{2l+1} \left(\frac{r_{<}^l}{r_{>}^{l+1}} \right) \oint Y_{l,m}(\pi/2, \varphi') e^{i\varphi'} d\varphi', \quad (5.24)$$

¹J.D. Jackson, *Classical Electrodynamics*, 2nd Edition, (Wiley, 1962). Section 5.5.

where $r_<$ now represents the lesser of r and a , whereas $r_>$ represents the greater of r and a . It follows from Equation (3.17) that

$$\oint Y_{l,m}^*(\theta', \varphi') e^{i\varphi'} d\varphi' = 2\pi Y_{l,m}^*(\theta', \varphi') e^{i\varphi'} \delta_{m1}. \quad (5.25)$$

Thus, Equation (5.24) yields

$$A_\varphi(r, \theta) = 2\pi \mu_0 I a \operatorname{Re} \sum_{l=0, \infty} \frac{Y_{l,1}(\theta, 0)}{2l+1} \left(\frac{r_<^l}{r_>^{l+1}} \right) Y_{l,1}^*(\pi/2, \varphi') e^{i\varphi'}. \quad (5.26)$$

However, according to Equation (3.17),

$$Y_{l,1}(\theta, 0) = \left[\frac{(2l+1)}{4\pi l(l+1)} \right]^{1/2} P_l^1(\cos \theta), \quad (5.27)$$

$$Y_{l,1}^*(\pi/2, \varphi') e^{i\varphi'} = \left[\frac{(2l+1)}{4\pi l(l+1)} \right]^{1/2} P_l^1(0). \quad (5.28)$$

Hence, we obtain

$$A_\varphi(r, \theta) = \frac{1}{2} \mu_0 I a \sum_{l=1,3,5,\dots} \frac{P_l^1(0)}{l(l+1)} \left(\frac{r_<^l}{r_>^{l+1}} \right) P_l^1(\cos \theta), \quad (5.29)$$

where we have made use of the fact that $P_l^1(0) = 0$ when l is even.² To be more exact,

$$A_\varphi(r, \theta) = \frac{1}{2} \mu_0 I \sum_{l=1,3,5,\dots} \frac{P_l^1(0)}{l(l+1)} \left(\frac{r}{a} \right)^l P_l^1(\cos \theta) \quad (5.30)$$

for $r < a$, and

$$A_\varphi(r, \theta) = \frac{1}{2} \mu_0 I \sum_{l=1,3,5,\dots} \frac{P_l^1(0)}{l(l+1)} \left(\frac{a}{r} \right)^{l+1} P_l^1(\cos \theta) \quad (5.31)$$

for $r > a$.

Now, according to Equations (5.7) and (5.20),

$$B_r = \frac{1}{r \sin \theta} \frac{\partial}{\partial \theta} (\sin \theta A_\varphi), \quad (5.32)$$

$$B_\theta = -\frac{1}{r} \frac{\partial}{\partial r} (r A_\varphi), \quad (5.33)$$

$$B_\varphi = 0. \quad (5.34)$$

Given that³

$$\frac{1}{\sin \theta} \frac{d}{d\theta} [\sin \theta P_l^1(\cos \theta)] = -l(l+1) P_l(\cos \theta), \quad (5.35)$$

²Ibid.

³Ibid.

we find that

$$B_r(r, \theta) = -\frac{\mu_0 I}{2a} \sum_{l=1,3,5,\dots} P_l^1(0) \left(\frac{r}{a}\right)^{l-1} P_l(\cos \theta), \quad (5.36)$$

$$B_\theta(r, \theta) = -\frac{\mu_0 I}{2a} \sum_{l=1,3,5,\dots} \frac{P_l^1(0)}{l} \left(\frac{r}{a}\right)^{l-1} P_l^1(\cos \theta) \quad (5.37)$$

in the region $r < a$. In particular, because $P_l(x) = -x$ and $P_1^1(x) = -(1-x^2)^{1/2}$, we obtain

$$B_r(\mathbf{0}) = \frac{\mu_0 I}{2a} \cos \theta, \quad (5.38)$$

$$B_\theta(\mathbf{0}) = -\frac{\mu_0 I}{2a} \sin \theta. \quad (5.39)$$

The previous two equations can be combined to give

$$\mathbf{B}(\mathbf{0}) = \frac{\mu_0 I}{2a} \mathbf{e}_z. \quad (5.40)$$

Of course, this result can be obtained in a more straightforward fashion via the direct application of the Biot-Savart law. We also have

$$B_r(r, \theta) = -\frac{\mu_0 I}{2a} \sum_{l=1,3,5,\dots} P_l^1(0) \left(\frac{a}{r}\right)^{l+2} P_l(\cos \theta), \quad (5.41)$$

$$B_\theta(r, \theta) = \frac{\mu_0 I}{2a} \sum_{l=1,3,5,\dots} \frac{P_l^1(0)}{l+1} \left(\frac{a}{r}\right)^{l+2} P_l^1(\cos \theta) \quad (5.42)$$

in the region $r > a$. A long way from the current loop (i.e., $r/a \rightarrow \infty$), we obtain

$$A_\varphi(r, \theta) = \frac{\mu_0}{4\pi} m \frac{\sin \theta}{r^2}, \quad (5.43)$$

$$B_r(r, \theta) = \frac{\mu_0}{4\pi} m \frac{2 \cos \theta}{r^3}, \quad (5.44)$$

$$B_\theta(r, \theta) = \frac{\mu_0}{4\pi} m \frac{\sin \theta}{r^3}, \quad (5.45)$$

where $m = I \pi a^2$.

Now, a small planar current loop of area A , carrying a current I , constitutes a *magnetic dipole* of moment

$$\mathbf{m} = I A \mathbf{n}. \quad (5.46)$$

Here, \mathbf{n} is a unit normal to the loop in the sense determined by the right-hand circulation rule (with the current determining the sense of circulation). It follows that in the limit $a \rightarrow 0$ and $I \pi a^2 \rightarrow m$ the current loop considered previously constitutes a magnetic dipole of moment $\mathbf{m} =$

$m \mathbf{e}_z$. Moreover, Equations (5.43)–(5.45) specify the non-zero components of the vector potential and the magnetic field generated by the dipole. It is easily seen from Equation (5.43) that

$$\mathbf{A} = \frac{\mu_0}{4\pi} \frac{\mathbf{m} \times \mathbf{r}}{r^3}. \quad (5.47)$$

Taking the curl of this expression, we obtain

$$\mathbf{B} = \frac{\mu_0}{4\pi} \left[\frac{3(\mathbf{m} \cdot \mathbf{r}) \mathbf{r} - r^2 \mathbf{m}}{r^5} \right], \quad (5.48)$$

which is consistent with Equations (5.44) and (5.45).

5.5 Localized Current Distribution

Consider the magnetic field generated by a current distribution that is localized in some relatively small region of space centered on the origin. From Equation (5.8), we have

$$\mathbf{A}(\mathbf{r}) = \frac{\mu_0}{4\pi} \int \frac{\mathbf{j}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} dV'. \quad (5.49)$$

Assuming that $r \gg r'$, so that our observation point lies well outside the distribution, we can write

$$\frac{1}{|\mathbf{r} - \mathbf{r}'|} = \frac{1}{r} + \frac{\mathbf{r} \cdot \mathbf{r}'}{r^3} + \dots. \quad (5.50)$$

Thus, the i th Cartesian component of the vector potential has the expansion

$$A_i(\mathbf{r}) = \frac{\mu_0}{4\pi} \frac{1}{r} \int j_i(\mathbf{r}') dV' + \frac{\mu_0}{4\pi} \frac{\mathbf{r}}{r^3} \cdot \int j_i(\mathbf{r}') \mathbf{r}' dV' + \dots \quad (5.51)$$

Consider the integral

$$K = \int (f \mathbf{j} \cdot \nabla' g + g \mathbf{j} \cdot \nabla' f) dV', \quad (5.52)$$

where $\mathbf{j}(\mathbf{r}')$ is a divergence-free [see Equation (5.5)] localized current distribution, and $f(\mathbf{r}')$ and $g(\mathbf{r}')$ are two well-behaved functions. Integrating the first term by parts, making use of the fact that $\mathbf{j}'(\mathbf{r}') \rightarrow \mathbf{0}$ as $|\mathbf{r}'| \rightarrow \infty$ (because the current distribution is localized), we obtain

$$K = \int [-g \nabla' \cdot (f \mathbf{j}) + g \mathbf{j} \cdot \nabla' f] dV' \quad (5.53)$$

Hence,

$$K = \int [-g \mathbf{j} \cdot \nabla' f - g f \nabla' \cdot \mathbf{j} + g \mathbf{j} \cdot \nabla' f] dV' = 0, \quad (5.54)$$

because $\nabla' \cdot \mathbf{j} = 0$. Thus, we have proved that

$$\int (f \mathbf{j} \cdot \nabla' g + g \mathbf{j} \cdot \nabla' f) dV' = 0. \quad (5.55)$$

Let $f = 1$ and $g = x'_i$ (where x'_i is the i th component of \mathbf{r}'). It immediately follows from Equation (5.55) that

$$\int j_i(\mathbf{r}') dV' = 0. \quad (5.56)$$

Likewise, if $f = x'_i$ and $g = x'_j$ then Equation (5.55) implies that

$$\int (x'_i j_j + x'_j j_i) dV' = 0. \quad (5.57)$$

According to Equations (5.51) and (5.56),

$$A_i(\mathbf{r}) = \frac{\mu_0}{4\pi} \frac{\mathbf{r}}{|\mathbf{r}|^3} \cdot \int j_i(\mathbf{r}') \mathbf{r}' dV' + \dots. \quad (5.58)$$

Now,

$$\mathbf{r} \cdot \int j_i(\mathbf{r}') \mathbf{r}' dV' = x_j \int x'_j j_i dV' = -\frac{1}{2} x_j \int (x'_i j_j - x'_j j_i) dV', \quad (5.59)$$

where use has been made of Equation (5.57), as well as the Einstein summation convention. Thus,

$$\mathbf{r} \cdot \int j_i \mathbf{r}' dV' = -\frac{1}{2} \int [(\mathbf{r} \cdot \mathbf{j}) \mathbf{r}' - (\mathbf{r} \cdot \mathbf{r}') \mathbf{j}]_i dV' = -\frac{1}{2} \left[\mathbf{r} \times \int (\mathbf{r}' \times \mathbf{j}) dV' \right]_i. \quad (5.60)$$

Hence, we obtain

$$\mathbf{A}(\mathbf{r}) = -\frac{\mu_0}{8\pi} \frac{\mathbf{r}}{|\mathbf{r}|^3} \times \int (\mathbf{r}' \times \mathbf{j}) dV' + \dots. \quad (5.61)$$

It is conventional to define the *magnetization*, or *magnetic moment density*, as

$$\mathbf{M}(\mathbf{r}) = \frac{1}{2} \mathbf{r} \times \mathbf{j}(\mathbf{r}). \quad (5.62)$$

The integral of this quantity is known as the *magnetic moment*:

$$\mathbf{m} = \frac{1}{2} \int \mathbf{r}' \times \mathbf{j}'(\mathbf{r}') dV'. \quad (5.63)$$

It immediately follows from Equation (5.61) that the vector potential a long way from a localized current distribution takes the form

$$\mathbf{A}(\mathbf{r}) = \frac{\mu_0}{4\pi} \frac{\mathbf{m} \times \mathbf{r}}{r^3}. \quad (5.64)$$

The corresponding magnetic field is

$$\mathbf{B}(\mathbf{r}) = \nabla \times \mathbf{A} = \frac{\mu_0}{4\pi} \left[\frac{3(\mathbf{m} \cdot \mathbf{r}) \mathbf{r} - r^2 \mathbf{m}}{r^5} \right]. \quad (5.65)$$

Thus, we have demonstrated that the magnetic field far from any localized current distribution takes the form of a magnetic dipole field whose moment is given by the integral (5.63).

Consider a localized current distribution that consists of a closed planar loop carrying the current I . If $d\mathbf{r}$ is a line element of the loop then Equation (5.63) reduces to

$$\mathbf{m} = I \oint \frac{1}{2} \mathbf{r} \times d\mathbf{r}. \quad (5.66)$$

However, $(1/2) \mathbf{r} \times d\mathbf{r} = d\mathbf{A}$, where $d\mathbf{A}$ is a triangular element of vector area defined by the two ends of $d\mathbf{r}$ and the origin. Thus, the loop integral gives the total vector area, \mathbf{A} , of the loop. It follows that

$$\mathbf{m} = I A \mathbf{n}, \quad (5.67)$$

where \mathbf{n} is a unit normal to the loop in the sense determined by the right-hand circulation rule (with the current determining the sense of circulation). Of course, Equation (5.67) is identical to Equation (5.46).

5.6 Exercises

- 5.1 Consider two thin current loops. Let loops 1 and 2 carry the currents I_1 and I_2 , respectively. The magnetic force exerted on loop 2 by loop 1 is [see Equation (5.3)]

$$\mathbf{F}_{21} = \frac{\mu_0 I_1 I_2}{4\pi} \oint_1 \oint_2 \frac{d\mathbf{r}_2 \times (d\mathbf{r}_1 \times \mathbf{r}_{12})}{|\mathbf{r}_{12}|^3},$$

where $\mathbf{r}_{12} = \mathbf{r}_2 - \mathbf{r}_1$. Here, \mathbf{r}_1 and \mathbf{r}_2 are the position vectors of elements of loops 1 and 2, respectively. Demonstrate that the previous expression can also be written

$$\mathbf{F}_{21} = -\frac{\mu_0 I_1 I_2}{4\pi} \oint_1 \oint_2 \frac{(d\mathbf{r}_1 \cdot d\mathbf{r}_2) \mathbf{r}_{12}}{|\mathbf{r}_{12}|^3}.$$

Hence, deduce that

$$\mathbf{F}_{12} = -\mathbf{F}_{21},$$

in accordance with Newton's third law of motion.

- 5.2 Consider the two current loops discussed in the previous question. The magnetic field generated at a general position vector \mathbf{r} by the current flowing around loop 1 is [see Equation (5.1)]

$$\mathbf{B}(\mathbf{r}) = \frac{\mu_0 I_1}{4\pi} \oint_1 \frac{d\mathbf{r}_1 \times (\mathbf{r} - \mathbf{r}_1)}{|\mathbf{r} - \mathbf{r}_1|^3}.$$

Demonstrate that

$$\mathbf{B} = \nabla \times \mathbf{A},$$

where

$$\mathbf{A}(\mathbf{r}) = \frac{\mu_0 I_1}{4\pi} \oint_1 \frac{d\mathbf{r}_1}{|\mathbf{r} - \mathbf{r}_1|}.$$

Show that the magnetic flux passing through loop 2, as a consequence of the current flowing around loop 1, is

$$\Phi_{21} = \frac{\mu_0 I_1}{4\pi} \oint_1 \oint_2 \frac{d\mathbf{r}_2 \cdot d\mathbf{r}_1}{|\mathbf{r}_1 - \mathbf{r}_2|}.$$

Hence, deduce that the mutual inductance of the two current loops takes the form

$$M = \frac{\mu_0}{4\pi} \oint_1 \oint_2 \frac{d\mathbf{r}_1 \cdot d\mathbf{r}_2}{|\mathbf{r}_2 - \mathbf{r}_1|}.$$

5.3 The vector potential of a magnetic dipole of moment \mathbf{m} is given by

$$\mathbf{A}(\mathbf{r}) = \frac{\mu_0}{4\pi} \frac{\mathbf{m} \times \mathbf{r}}{r^3}.$$

Show that the corresponding magnetic field is

$$\mathbf{B}(\mathbf{r}) = \frac{\mu_0}{4\pi} \left[\frac{3(\mathbf{r} \cdot \mathbf{m})\mathbf{r} - r^2\mathbf{m}}{r^5} \right].$$

5.4 Demonstrate that the torque acting on a magnetic dipole of moment \mathbf{m} placed in a uniform external magnetic field \mathbf{B} is

$$\boldsymbol{\tau} = \mathbf{m} \times \mathbf{B}.$$

Hence, deduce that the potential energy of the magnetic dipole is

$$W = -\mathbf{m} \cdot \mathbf{B}.$$

5.5 Consider two magnetic dipoles, \mathbf{m}_1 and \mathbf{m}_2 . Suppose that \mathbf{m}_1 is fixed, whereas \mathbf{m}_2 can rotate freely in any direction. Demonstrate that the equilibrium configuration of the second dipole is such that

$$\tan \theta_1 = -2 \tan \theta_2,$$

where θ_1 and θ_2 are the angles subtended by \mathbf{m}_1 and \mathbf{m}_2 , respectively, with the radius vector joining them.

6 Magnetostatics in Magnetic Media

6.1 Magnetization

All matter is built up out of atoms, and every atom contains moving electrons. The currents associated with these electrons are termed *atomic currents*. Each atomic current is a tiny closed circuit of atomic dimensions, and may therefore be appropriately described as a magnetic dipole. If the atomic currents of a given atom all flow in the same plane then the atomic dipole moment is directed normal to the plane (in the sense given by the right-hand circulation rule), and its magnitude is the product of the total circulating current and the area of the current loop. More generally, if $\mathbf{j}(\mathbf{r})$ is the atomic current density at point \mathbf{r} then the magnetic moment of the atom is [see Equation (5.63)]

$$\mathbf{m} = \frac{1}{2} \int \mathbf{r} \times \mathbf{j} dV, \quad (6.1)$$

where the integral is over the volume of the atom. If there are N such atoms or molecules per unit volume then the magnetization, \mathbf{M} , (i.e., the magnetic dipole moment per unit volume) is given by $\mathbf{M} = N \mathbf{m}$. More generally,

$$\mathbf{M}(\mathbf{r}) = \sum_i N_i \langle \mathbf{m}_i \rangle, \quad (6.2)$$

where $\langle \mathbf{m}_i \rangle$ is the average magnetic dipole moment of the i th type of molecule in the vicinity of point \mathbf{r} , and N_i is the average number of such molecules per unit volume at \mathbf{r} .

Consider a general medium that is made up of molecules that are polarizable, and possess a net magnetic moment. It is easily demonstrated that any circulation in the magnetization field $\mathbf{M}(\mathbf{r})$ gives rise to an effective current density \mathbf{j}_m in the medium. In fact,

$$\mathbf{j}_m = \nabla \times \mathbf{M}. \quad (6.3)$$

This current density is called the *magnetization current density*, and is usually distinguished from the *true current density*, \mathbf{j}_t , which represents the convection of free charges in the medium. In fact, there is a third type of current called a *polarization current*, which is due to the apparent convection of bound charges. It is easily demonstrated that the polarization current density, \mathbf{j}_p , is given by

$$\mathbf{j}_p = \frac{\partial \mathbf{P}}{\partial t}. \quad (6.4)$$

Thus, the total current density, \mathbf{j} , in the medium is given by

$$\mathbf{j} = \mathbf{j}_t + \nabla \times \mathbf{M} + \frac{\partial \mathbf{P}}{\partial t}. \quad (6.5)$$

It must be emphasized that all three terms on the right-hand side of the previous equation represent real physical currents, although only the first term is due to the motion of charges over more than atomic dimensions.

The fourth Maxwell equation, (1.4), takes the form

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{j} + \mu_0 \epsilon_0 \frac{\partial \mathbf{E}}{\partial t}, \quad (6.6)$$

which can also be written

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{j}_t + \mu_0 \nabla \times \mathbf{M} + \mu_0 \frac{\partial \mathbf{D}}{\partial t}, \quad (6.7)$$

where use has been made of the definition $\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P}$. The previous expression can be rearranged to give

$$\nabla \times \mathbf{H} = \mathbf{j}_t + \frac{\partial \mathbf{D}}{\partial t}, \quad (6.8)$$

where

$$\mathbf{H} = \frac{\mathbf{B}}{\mu_0} - \mathbf{M} \quad (6.9)$$

is termed the *magnetic intensity*, and has the same dimensions as \mathbf{M} (i.e., magnetic dipole moment per unit volume). In a steady-state situation, the curl theorem tell us that

$$\oint_C \mathbf{H} \cdot d\mathbf{r} = \int_S \mathbf{j}_t \cdot d\mathbf{S}. \quad (6.10)$$

In other words, the line integral of \mathbf{H} around some closed loop is equal to the flux of true current through any surface attached to that loop. Unlike the magnetic field \mathbf{B} (which specifies the force $e \mathbf{v} \times \mathbf{B}$ acting on a charge e moving with velocity \mathbf{v}), or the magnetization \mathbf{M} (which specifies the magnetic dipole moment per unit volume), the magnetic intensity \mathbf{H} has no clear physical meaning. The only reason for introducing it is that it enables us to calculate magnetic fields in the presence of magnetic materials without first having to know the distribution of magnetization currents. However, this is only possible if we possess a constitutive relation connecting \mathbf{B} and \mathbf{H} .

6.2 Magnetic Susceptibility and Permeability

In a large class of materials, there exists an approximately linear relationship between \mathbf{M} and \mathbf{H} . If the material is isotropic then

$$\mathbf{M} = \chi_m \mathbf{H}, \quad (6.11)$$

where the dimensionless quantity χ_m is known as the *magnetic susceptibility*. If χ_m is positive then the material is called *paramagnetic*, and the magnetic field is strengthened by the presence of the material. If χ_m is negative then the material is called *diamagnetic*, and the magnetic field is weakened in the presence of the material. The magnetic susceptibilities of paramagnetic and diamagnetic materials are generally extremely small. A few sample values are given in Table 6.1.

A linear relationship between \mathbf{M} and \mathbf{H} also implies a linear relationship between \mathbf{B} and \mathbf{H} . In fact, from Equation (6.9), we can write

$$\mathbf{B} = \mu \mathbf{H}, \quad (6.12)$$

where

$$\mu = \mu_0 (1 + \chi_m) \quad (6.13)$$

Material	χ_m
Aluminium	2.3×10^{-5}
Copper	-0.98×10^{-5}
Diamond	-2.2×10^{-5}
Tungsten	6.8×10^{-5}
Hydrogen (1 atm)	-0.21×10^{-8}
Oxygen (1 atm)	209.0×10^{-8}
Nitrogen (1 atm)	-0.50×10^{-8}

Table 6.1: Magnetic susceptibilities of some paramagnetic and diamagnetic materials at room temperature. Data obtained from the *Handbook of Chemistry and Physics*, Chemical Rubber Company Press, Boca Raton, FL.

is termed the magnetic *permeability* of the material in question. (Likewise, μ_0 is termed the *permeability of free space*.) It is clear from Table 6.1 that the permeabilities of common diamagnetic and paramagnetic materials do not differ substantially from that of free space. In fact, to all intents and purposes, the magnetic properties of such materials can be safely neglected (i.e., $\mu = \mu_0$).

6.3 Ferromagnetism

There is, however, a third class of magnetic materials called *ferromagnetic* materials. Such materials are characterized by a possible permanent magnetization, and generally have a profound effect on magnetic fields (i.e., $\mu/\mu_0 \gg 1$). Unfortunately, ferromagnetic materials do not exhibit a linear dependence between \mathbf{M} and \mathbf{H} , or between \mathbf{B} and \mathbf{H} , so that we cannot employ Equations (6.11) and (6.12) with constant values of χ_m and μ . It is still expedient to use Equation (6.12) as the definition of μ , with $\mu = \mu(\mathbf{H})$. However, this practice leads to complications under certain circumstances. In fact, the permeability of a ferromagnetic material, as defined by Equation (6.12), can vary through the entire range of possible values from zero to infinity, and may be either positive or negative. The most sensible approach is to consider each problem involving ferromagnetic materials separately, try to determine which region of the \mathbf{B} - \mathbf{H} diagram is important for the particular case in hand, and then make approximations appropriate to this region.

Let us, first, consider an unmagnetized sample of ferromagnetic material. If the magnetic intensity, which is initially zero, is increased monotonically, then the \mathbf{B} - \mathbf{H} relationship traces out a curve such as that shown in Figure 6.1. This is called a *magnetization curve*. It is evident that the permeabilities μ derived from the curve (according to the rule $\mu = B/H$) are always positive, and show a wide range of values. The maximum permeability occurs at the “knee” of the curve. In some materials this maximum permeability is as large as $10^5 \mu_0$. The reason for the knee in the curve is that the magnetization \mathbf{M} reaches a maximum value in the material, so that

$$\mathbf{B} = \mu_0 (\mathbf{H} + \mathbf{M}) \quad (6.14)$$

continues to increase at large \mathbf{H} only because of the $\mu_0 \mathbf{H}$ term. The maximum value of \mathbf{M} is called the *saturation magnetization* of the material.

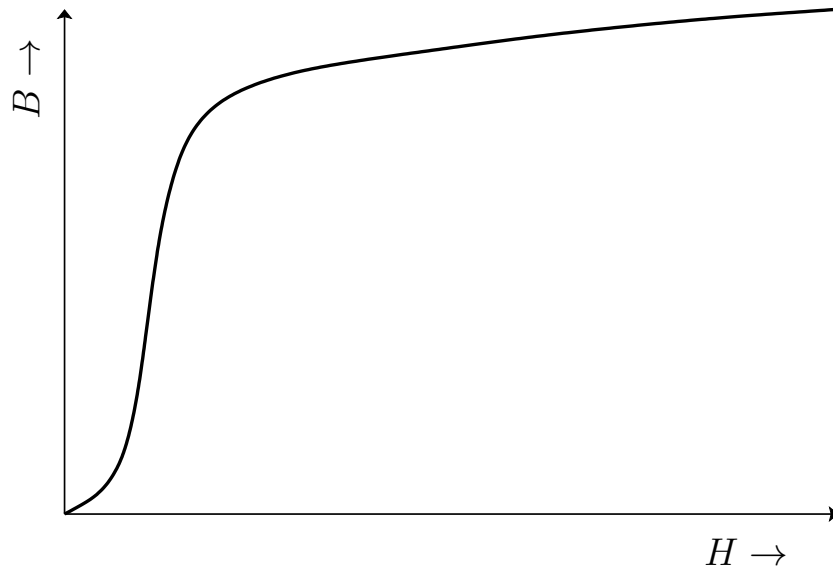


Figure 6.1: Schematic magnetization curve of an initially unmagnetized ferromagnet.

Next, consider a ferromagnetic sample magnetized by the previously described procedure. If the magnetic intensity \mathbf{H} is decreased then the \mathbf{B} - \mathbf{H} relation does not follow back down the curve of Figure 6.1, but instead moves along a new curve, shown in Figure 6.2, to the point R . The magnetization, once established, does not disappear with the removal of \mathbf{H} . In fact, it takes a reversed magnetic intensity to reduce the magnetization to zero. If \mathbf{H} continues to build up in the reversed direction then \mathbf{M} (and, hence, \mathbf{B}) becomes increasingly negative. Finally, if \mathbf{H} increases again then the operating point follows the lower curve of Figure 6.2. Thus, the \mathbf{B} - \mathbf{H} curve for increasing \mathbf{H} is quite different to that for decreasing \mathbf{H} . This phenomenon is known as *hysteresis*.

The loop shown in Figure 6.2 is called the hysteresis loop of the material in question. The value of \mathbf{B} at the point R is called the *retentivity* or *remanence*. The magnitude of \mathbf{H} at the point C is called the *coercivity*. It is evident that μ is negative in the second and fourth quadrants of the loop, and positive in the first and third quadrants. The shape of the hysteresis loop depends not only on the nature of the ferromagnetic material, but also on the maximum value of \mathbf{H} to which the material has been subjected. However, once this maximum value, \mathbf{H}_{\max} , becomes sufficient to produce saturation in the material, the hysteresis loop does not change shape with any further increase in \mathbf{H}_{\max} .

Ferromagnetic materials are used either to channel magnetic flux (e.g., around transformer circuits), or as sources of magnetic field (e.g., permanent magnets). For use as a permanent magnet, the material is first magnetized by placing it in a strong magnetic field. However, once the magnet is removed from the external field, it is subject to a demagnetizing \mathbf{H} . Thus, it is vitally important that a permanent magnet should possess both a large remanence and a large coercivity. As will become clear, later on, it is generally a good idea for the ferromagnetic materials used to channel magnetic flux around transformer circuits to possess small remanences and small coercivities.

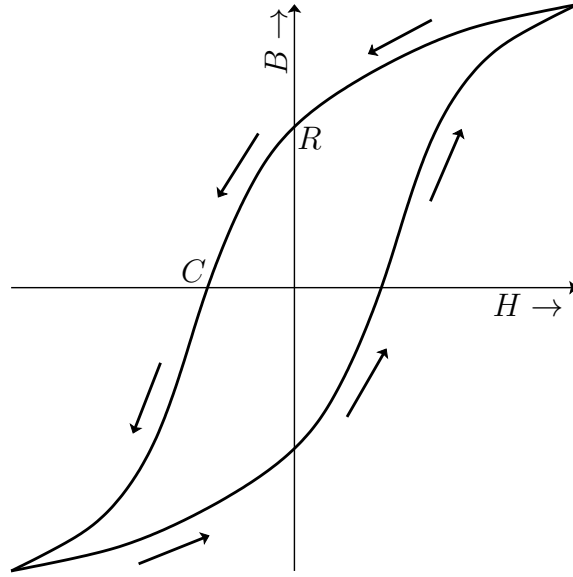


Figure 6.2: Typical hysteresis loop of a ferromagnetic material.

6.4 Boundary Conditions for \mathbf{B} and \mathbf{H}

Let us derive the matching conditions for \mathbf{B} and \mathbf{H} at the boundary between two magnetic media. The governing equations for a steady-state situation are

$$\nabla \cdot \mathbf{B} = 0, \quad (6.15)$$

and

$$\nabla \times \mathbf{H} = \mathbf{j}_t. \quad (6.16)$$

Integrating Equation (6.15) over a Gaussian pill-box enclosing part of the boundary surface between the two media gives

$$(\mathbf{B}_2 - \mathbf{B}_1) \cdot \mathbf{n}_{21} = 0, \quad (6.17)$$

where \mathbf{n}_{21} is the unit normal to this surface directed from medium 1 to medium 2. Integrating Equation (6.16) around a small loop that straddles the boundary surface yields

$$(\mathbf{H}_2 - \mathbf{H}_1) \times \mathbf{n}_{21} = 0, \quad (6.18)$$

assuming that there is no true current sheet flowing at the surface. In general, there is a magnetization current sheet flowing at the boundary surface whose density is given by

$$\mathbf{J}_m = \mathbf{n}_{21} \times (\mathbf{M}_2 - \mathbf{M}_1), \quad (6.19)$$

where \mathbf{M}_1 is the magnetization in medium 1 at the boundary, et cetera. It is clear that the normal component of the magnetic field, and the tangential component of the magnetic intensity, are both continuous across any boundary between magnetic materials.

6.5 Permanent Ferromagnets

Let us consider the magnetic field generated by a distribution of permanent ferromagnets. Suppose that the magnets in question are sufficiently “hard” that their magnetization is essentially independent of the applied field for moderate field-strengths. Such magnets can be treated as if they contain a fixed magnetization $\mathbf{M}(\mathbf{r})$.

Let us assume that there are no true currents in the problem, so that $\mathbf{j}_f = \mathbf{0}$. Let us also assume that we are dealing with a steady-state situation. Under these circumstances Equation (6.8) reduces to

$$\nabla \times \mathbf{H} = \mathbf{0}. \quad (6.20)$$

It follows that we can write

$$\mathbf{H} = -\nabla\phi_m, \quad (6.21)$$

where ϕ_m is called the *magnetic scalar potential*. Now, we know that

$$\nabla \cdot \mathbf{B} = \mu_0 \nabla \cdot (\mathbf{H} + \mathbf{M}) = 0. \quad (6.22)$$

Equations (6.21) and (6.22) combine to give

$$\nabla^2\phi_m = -\rho_m, \quad (6.23)$$

where

$$\rho_m = -\nabla \cdot \mathbf{M}. \quad (6.24)$$

Thus, the magnetostatic field, \mathbf{H} , is determined by Poisson’s equation. We can think of ρ_m as an effective magnetic charge density. Of course, this magnetic charge has no physical reality. We have only introduced it in order to make the problem of the steady magnetic field generated by a set of permanent magnets look formally the same as that of the steady electric field generated by a distribution of charges.

The unique solution of Poisson’s equation, subject to sensible boundary conditions at infinity, is well known (see Section 2.3):

$$\phi_m(\mathbf{r}) = \frac{1}{4\pi} \int \frac{\rho_m(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} dV'. \quad (6.25)$$

This solution yields

$$\phi_m(\mathbf{r}) = -\frac{1}{4\pi} \int \frac{\nabla' \cdot \mathbf{M}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} dV'. \quad (6.26)$$

If the magnetization field $\mathbf{M}(\mathbf{r})$ is well behaved and localized then we can integrate by parts to obtain

$$\phi_m(\mathbf{r}) = \frac{1}{4\pi} \int \mathbf{M}(\mathbf{r}') \cdot \nabla' \left(\frac{1}{|\mathbf{r} - \mathbf{r}'|} \right) dV'. \quad (6.27)$$

Now,

$$\nabla' \left(\frac{1}{|\mathbf{r} - \mathbf{r}'|} \right) = -\nabla \left(\frac{1}{|\mathbf{r} - \mathbf{r}'|} \right), \quad (6.28)$$

so our expression for the magnetic potential can be written

$$\phi_m(\mathbf{r}) = -\frac{1}{4\pi} \nabla \cdot \int \frac{\mathbf{M}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} dV'. \quad (6.29)$$

Far from the region of non-vanishing magnetization, the potential reduces to

$$\phi_m(\mathbf{r}) \simeq -\nabla \left(\frac{1}{4\pi r} \right) \cdot \int \mathbf{M}(\mathbf{r}') dV' \simeq \frac{\mathbf{m} \cdot \mathbf{r}}{4\pi r^3}, \quad (6.30)$$

where $\mathbf{m} = \int \mathbf{M} dV$ is the total magnetic moment of the distribution. This is the scalar potential of a dipole. (See Sections 3.6 and 5.5.) Thus, an arbitrary localized distribution of magnetization asymptotically produces a dipole magnetic field whose strength is determined by the net magnetic moment of the distribution.

It is often a good approximation to treat the magnetization field $\mathbf{M}(\mathbf{r})$ as a discontinuous quantity. In other words, $\mathbf{M}(\mathbf{r})$ is specified throughout the “hard” ferromagnets in question, and suddenly falls to zero at the boundaries of these magnets. Integrating Equation (6.24) over a Gaussian pill-box that straddles one of these boundaries leads to the conclusion that there is an effective magnetic surface charge density,

$$\sigma_m = \mathbf{n} \cdot \mathbf{M}, \quad (6.31)$$

on the surface of the ferromagnets, where \mathbf{M} is the surface magnetization, and \mathbf{n} is a unit outward directed normal to the surface. Under these circumstances, Equation (6.26) yields

$$\phi_m(\mathbf{r}) = -\frac{1}{4\pi} \int_V \frac{\nabla' \cdot \mathbf{M}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} dV' + \frac{1}{4\pi} \int_S \frac{\mathbf{M}(\mathbf{r}') \cdot d\mathbf{S}'}{|\mathbf{r} - \mathbf{r}'|}, \quad (6.32)$$

where V represents the volume occupied by the magnets and S is the bounding surface to V . Here, $d\mathbf{S}$ is an outward directed element of S . It is clear that the right-hand side of Equation (6.32) consists of a volume integral involving the volume magnetic charges $\rho_m = -\nabla \cdot \mathbf{M}$, and a surface integral involving the surface magnetic charges $\sigma_m = \mathbf{n} \cdot \mathbf{M}$. If the magnetization is uniform throughout the volume V then the volume integral vanishes, and only the surface integral makes a contribution.

We can also write $\mathbf{B} = \nabla \times \mathbf{A}$ in order to satisfy $\nabla \cdot \mathbf{B} = 0$ automatically. It follows from Equations (6.8) and (6.9) that

$$\nabla \times \mathbf{H} = \nabla \times (\mathbf{B}/\mu_0 - \mathbf{M}) = \mathbf{0}, \quad (6.33)$$

which gives

$$\nabla^2 \mathbf{A} = -\mu_0 \mathbf{j}_m, \quad (6.34)$$

because $\mathbf{j}_m = \nabla \times \mathbf{M}$. The unique solution to Equation (6.34), subject to sensible boundary conditions at infinity, is well known:

$$\mathbf{A}(\mathbf{r}) = \frac{\mu_0}{4\pi} \int \frac{\mathbf{j}_m(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} dV'. \quad (6.35)$$

Thus,

$$\mathbf{A}(\mathbf{r}) = \frac{\mu_0}{4\pi} \int \frac{\nabla' \times \mathbf{M}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} dV'. \quad (6.36)$$

If the magnetization field is discontinuous, it is necessary to add a surface integral to the previous expression. It is straightforward to show that

$$\mathbf{A}(\mathbf{r}) = \frac{\mu_0}{4\pi} \int_V \frac{\nabla' \times \mathbf{M}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} dV' + \frac{\mu_0}{4\pi} \int_S \frac{\mathbf{M}(\mathbf{r}') \times d\mathbf{S}'}{|\mathbf{r} - \mathbf{r}'|}. \quad (6.37)$$

It is clear that the previous expression consists of a volume integral involving the volume magnetization currents $\mathbf{j}_m = \nabla \times \mathbf{M}$, and a surface integral involving the surface magnetization currents $\mathbf{J}_m = \mathbf{M} \times \mathbf{n}$ [see Equation (6.19)]. However, if the magnetization field is uniform throughout V then only the surface integral makes a contribution.

6.6 Uniformly Magnetized Sphere

Consider a sphere of radius a , with a uniform permanent magnetization $\mathbf{M} = M_0 \mathbf{e}_z$, surrounded by a vacuum region. The simplest way of solving this problem is in terms of the scalar magnetic potential introduced in Equation (6.21). It follows from Equations (6.23) and (6.24) that ϕ_m satisfies Laplace's equation,

$$\nabla^2 \phi_m = 0, \quad (6.38)$$

because there is zero volume magnetic charge density in a vacuum, or a uniformly magnetized magnetic medium. However, according to Equation (6.31), there is a magnetic surface charge density,

$$\sigma_m = \mathbf{e}_r \cdot \mathbf{M} = M_0 \cos \theta, \quad (6.39)$$

on the surface of the sphere. Here, r and θ are spherical coordinates. One of the matching conditions at the surface of the sphere is that the tangential component of \mathbf{H} must be continuous. It follows from Equation (6.21) that the scalar magnetic potential must be continuous at $r = a$, so that

$$\phi_m(r = a_+, \theta) = \phi_m(r = a_-, \theta). \quad (6.40)$$

Integrating Equation (6.23) over a Gaussian pill-box straddling the surface of the sphere yields

$$\left[\frac{\partial \phi_m}{\partial r} \right]_{r=a^-}^{r=a^+} = -\sigma_m = -M_0 \cos \theta. \quad (6.41)$$

In other words, the magnetic charge sheet on the surface of the sphere gives rise to a discontinuity in the radial gradient of the magnetic scalar potential at $r = a$.

The most general axisymmetric solution to Equation (6.38) that satisfies physical boundary conditions at $r = a$ and $r = \infty$ is

$$\phi_m(r, \theta) = \sum_{l=0, \infty} A_l r^l P_l(\cos \theta) \quad (6.42)$$

for $r < a$, and

$$\phi_m(r, \theta) = \sum_{l=0, \infty} B_l r^{-(l+1)} P_l(\cos \theta) \quad (6.43)$$

for $r \geq a$. The boundary condition (6.40) yields

$$B_l = A_l a^{2l+1} \quad (6.44)$$

for all l . The boundary condition (6.41) gives

$$-\frac{(l+1)B_l}{a^{l+2}} - lA_l a^{l-1} = -M_0 \delta_{l1} \quad (6.45)$$

for all l , because $P_l(\cos \theta) = \cos \theta$. It follows that

$$A_l = B_l = 0 \quad (6.46)$$

for $l \neq 1$, and

$$A_1 = \frac{M_0}{3}, \quad (6.47)$$

$$B_1 = \frac{M_0 a^3}{3}. \quad (6.48)$$

Thus,

$$\phi_m(r, \theta) = \frac{M_0 a^2}{3} \frac{r}{a^2} \cos \theta \quad (6.49)$$

for $r < a$, and

$$\phi_m(r, \theta) = \frac{M_0 a^2}{3} \frac{a}{r^2} \cos \theta \quad (6.50)$$

for $r \geq a$. Because there is a uniqueness theorem associated with Poisson's equation (see Section 2.3), we can be sure that this axisymmetric potential is the only solution to the problem that satisfies physical boundary conditions at $r = 0$ and infinity.

In the vacuum region outside the sphere,

$$\mathbf{B} = \mu_0 \mathbf{H} = -\mu_0 \nabla \phi_m. \quad (6.51)$$

It is easily demonstrated from Equation (6.50) that

$$\mathbf{B}(r > a) = \frac{\mu_0}{4\pi} \left[-\frac{\mathbf{m}}{r^3} + \frac{3(\mathbf{m} \cdot \mathbf{r}) \mathbf{r}}{r^5} \right], \quad (6.52)$$

where

$$\mathbf{m} = \frac{4}{3} \pi a^3 \mathbf{M}. \quad (6.53)$$

This, of course, is the magnetic field of a magnetic dipole of moment \mathbf{m} . [See Section 5.5.] Not surprisingly, the net dipole moment of the sphere is equal to the integral of the magnetization \mathbf{M} (which is the dipole moment per unit volume) over the volume of the sphere.

Inside the sphere, we have $\mathbf{H} = -\nabla \phi_m$ and $\mathbf{B} = \mu_0 (\mathbf{H} + \mathbf{M})$, giving

$$\mathbf{H} = -\frac{\mathbf{M}}{3}, \quad (6.54)$$

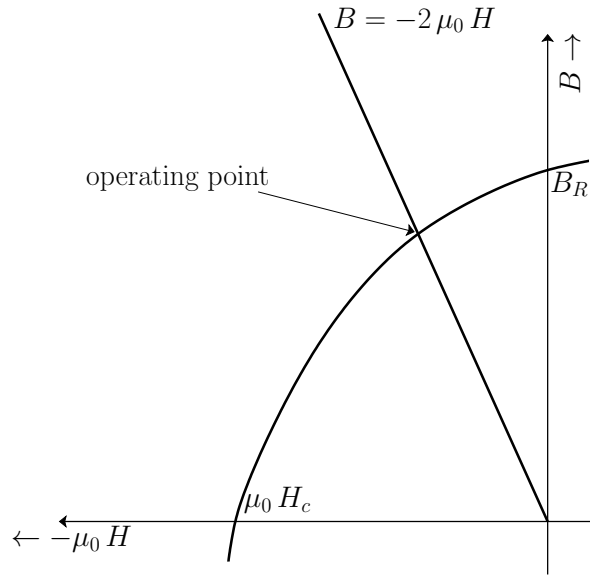


Figure 6.3: Schematic demagnetization curve for a permanent magnet.

and

$$\mathbf{B} = \frac{2}{3} \mu_0 \mathbf{M}. \quad (6.55)$$

Thus, both the \mathbf{H} and \mathbf{B} fields are uniform inside the sphere. Note that the magnetic intensity is oppositely directed to the magnetization. In other words, the \mathbf{H} field acts to demagnetize the sphere. How successful it is at achieving this depends on the shape of the hysteresis curve in the negative H and positive B quadrant. This curve is sometimes called the *demagnetization curve* of the magnetic material that makes up the sphere. Figure 6.3 shows a schematic demagnetization curve. The curve is characterized by two quantities: the retentivity B_R (i.e., the residual magnetic field strength at zero magnetic intensity) and the coercivity $\mu_0 H_c$ (i.e., the negative magnetic intensity required to demagnetize the material. The latter quantity is conventionally multiplied by μ_0 to give it the units of magnetic field-strength). The operating point (i.e., the values of B and $\mu_0 H$ inside the sphere) is obtained from the intersection of the demagnetization curve and the curve $B = \mu H$. It is clear from Equations (6.54) and (6.55) that

$$\mu = -2 \mu_0 \quad (6.56)$$

for a uniformly magnetized sphere in the absence of external fields. The magnetization inside the sphere is easily calculated once the operating point has been determined. In fact, $M_0 = B - \mu_0 H$. It is clear from Figure 6.3 that for a magnetic material to be a good permanent magnet it must possess both a large retentivity and a large coercivity. A material with a large retentivity but a small coercivity is unable to retain a significant magnetization in the absence of a strong external magnetizing field.

6.7 Soft Iron Sphere in Uniform Magnetic Field

The opposite extreme to a “hard” ferromagnetic material, which can maintain a large remnant magnetization in the absence of external fields, is a “soft” ferromagnetic material, for which the remnant magnetization is relatively small. Let us consider a somewhat idealized situation in which the remnant magnetization is negligible. In this situation, there is no hysteresis, so the \mathbf{B} - \mathbf{H} relation for the material reduces to

$$\mathbf{B} = \mu(B) \mathbf{H}, \quad (6.57)$$

where $\mu(B)$ is a single valued function. The most commonly occurring “soft” ferromagnetic material is soft iron (i.e., annealed, low impurity, iron).

Consider a sphere of soft iron placed in an initially uniform external field $\mathbf{B}_0 = B_0 \mathbf{e}_z$. The $\mu_0 \mathbf{H}$ and \mathbf{B} fields inside the sphere are most easily obtained by taking the solutions (6.54) and (6.55) (which are still valid), and superimposing on them the uniform field \mathbf{B}_0 . We are justified in doing this because the equations that govern magnetostatic problems are linear. Thus, inside the sphere we have

$$\mu_0 \mathbf{H} = \mathbf{B}_0 - \frac{1}{3} \mu_0 \mathbf{M}, \quad (6.58)$$

$$\mathbf{B} = \mathbf{B}_0 + \frac{2}{3} \mu_0 \mathbf{M}. \quad (6.59)$$

Combining Equations (6.57), (6.58), and (6.59) yields

$$\mu_0 \mathbf{M} = 3 \left(\frac{\mu - \mu_0}{\mu + 2\mu_0} \right) \mathbf{B}_0, \quad (6.60)$$

with

$$\mathbf{B} = \left(\frac{3\mu}{\mu + 2\mu_0} \right) \mathbf{B}_0, \quad (6.61)$$

where, in general, $\mu = \mu(B)$. Clearly, for a highly permeable material (i.e., $\mu/\mu_0 \gg 1$, which is certainly the case for soft iron) the magnetic field strength inside the sphere is approximately three times that of the externally applied field. In other words, the magnetic field is amplified inside the sphere.

The amplification of the magnetic field by a factor three in the high permeability limit is specific to a sphere. It can be shown that for elongated objects (e.g., rods), aligned along the direction of the external field, the amplification factor can be considerably larger than three.

It is important to realize that the magnetization inside a ferromagnetic material cannot increase without limit. The maximum possible value of \mathbf{M} is called the saturation magnetization, and is usually denoted \mathbf{M}_s . Most ferromagnetic materials saturate when they are placed in external magnetic fields whose strengths are greater than, or of order, one tesla. Suppose that our soft iron sphere first attains the saturation magnetization when the unperturbed external magnetic field strength is B_s . It follows from Equations (6.59) and (6.60) (with $\mu \gg \mu_0$) that

$$B = B_0 + 2 B_s \quad (6.62)$$

inside the sphere, for $B_0 > B_s$. In this case, the field amplification factor is

$$\frac{B}{B_0} = 1 + 2 \frac{B_s}{B_0}. \quad (6.63)$$

Thus, for $B_0 \gg B_s$ the amplification factor approaches unity. We conclude that if a ferromagnetic material is placed in an external field that greatly exceeds that required to cause saturation then the material effectively loses its magnetic properties, so that $\mu \simeq \mu_0$. Clearly, it is very important to avoid saturating the soft magnets used to channel magnetic flux around transformer circuits. This sets an upper limit on the magnetic field-strengths that can occur in such circuits.

6.8 Magnetic Shielding

There are many situations, particularly in experimental physics, where it is desirable to shield a certain region from magnetic fields. This goal can be achieved by surrounding the region in question by a material of high permeability. It is vitally important that a material used as a magnetic shield does not develop a permanent magnetization in the presence of external fields, otherwise the material itself may become a source of magnetic fields. The most effective commercially available magnetic shielding material is called *mu-metal*, and is an alloy of 5 percent copper, 2 percent chromium, 77 percent nickel, and 16 percent iron. The maximum permeability of mu-metal is about $10^5 \mu_0$. This material also possesses a particularly low retentivity and coercivity. Unfortunately, mu-metal is extremely expensive. Let us investigate how much of this material is actually required to shield a given region from an external magnetic field.

Consider a spherical shell of magnetic shielding, made up of material of permeability μ , placed in a formerly uniform magnetic field $\mathbf{B}_0 = B_0 \mathbf{e}_z$. Suppose that the inner radius of the shell is a , and the outer radius is b . Because there are no free currents in the problem, we can write $\mathbf{H} = -\nabla\phi_m$. Furthermore, because $\mathbf{B} = \mu \mathbf{H}$ and $\nabla \cdot \mathbf{B} = 0$, it is clear that the magnetic scalar potential satisfies Laplace's equation, $\nabla^2 \phi_m = 0$, throughout all space. The boundary conditions are that the potential must be well behaved at $r = 0$ and $r \rightarrow \infty$, and also that the tangential and the normal components of \mathbf{H} and \mathbf{B} , respectively, must be continuous at $r = a$ and $r = b$. The boundary conditions on \mathbf{H} merely imply that the scalar potential ϕ_m must be continuous at $r = a$ and $r = b$. The boundary conditions on \mathbf{B} yield

$$\mu_0 \frac{\partial \phi_m(r = a-, \theta)}{\partial r} = \mu \frac{\partial \phi_m(r = a+, \theta)}{\partial r}, \quad (6.64)$$

$$\mu_0 \frac{\partial \phi_m(r = b+, \theta)}{\partial r} = \mu \frac{\partial \phi_m(r = b-, \theta)}{\partial r}. \quad (6.65)$$

Let us try the following test solution for the magnetic potential:

$$\phi_m = -\frac{B_0}{\mu_0} r \cos \theta + \frac{\alpha}{r^2} \cos \theta \quad (6.66)$$

for $r > b$,

$$\phi_m = \left(\beta r + \frac{\gamma}{r^2} \right) \cos \theta \quad (6.67)$$

for $b \geq r \geq a$, and

$$\phi_m = \delta r \cos \theta \quad (6.68)$$

for $r < a$. This potential is certainly a solution of Laplace's equation throughout space. It yields the uniform magnetic field \mathbf{B}_0 as $r \rightarrow \infty$, and satisfies physical boundary conditions at $r = 0$ and infinity. Because there is a uniqueness theorem associated with Poisson's equation (see Section 2.3), we can be certain that this potential is the correct solution to the problem provided that the arbitrary constants α , β , et cetera, can be adjusted in such a manner that the boundary conditions at $r = a$ and $r = b$ are also satisfied.

The continuity of ϕ_m at $r = a$ and $r = b$ requires that

$$\beta a + \frac{\gamma}{a^2} = \delta a, \quad (6.69)$$

and

$$\beta b + \frac{\gamma}{b^2} = -\frac{B_0}{\mu_0} b + \frac{\alpha}{b^2}. \quad (6.70)$$

The boundary conditions (6.64) and (6.65) yield

$$\mu_0 \delta = \mu \left(\beta - \frac{2\gamma}{a^3} \right), \quad (6.71)$$

and

$$\mu_0 \left(-\frac{B_0}{\mu_0} - \frac{2\alpha}{b^3} \right) = \mu \left(\beta - \frac{2\gamma}{b^3} \right). \quad (6.72)$$

It follows that

$$\mu_0 \alpha = \left[\frac{(2\mu + \mu_0)(\mu - \mu_0)}{(2\mu + \mu_0)(\mu + 2\mu_0) - 2(a^3/b^3)(\mu - \mu_0)^2} \right] (b^3 - a^3) B_0, \quad (6.73)$$

$$\mu_0 \beta = - \left[\frac{3(2\mu + \mu_0)\mu_0}{(2\mu + \mu_0)(\mu + 2\mu_0) - 2(a^3/b^3)(\mu - \mu_0)^2} \right] B_0, \quad (6.74)$$

$$\mu_0 \gamma = - \left[\frac{3(\mu - \mu_0)\mu_0}{(2\mu + \mu_0)(\mu + 2\mu_0) - 2(a^3/b^3)(\mu - \mu_0)^2} \right] a^3 B_0, \quad (6.75)$$

$$\mu_0 \delta = - \left[\frac{9\mu\mu_0}{(2\mu + \mu_0)(\mu + 2\mu_0) - 2(a^3/b^3)(\mu - \mu_0)^2} \right] B_0. \quad (6.76)$$

Consider the limit of a thin, high permeability shell for which $b = a + d$, $d/a \ll 1$, and $\mu/\mu_0 \gg 1$. In this limit, the field inside the shell is given by

$$\mathbf{B} \simeq \frac{3}{2} \frac{\mu_0}{\mu} \frac{a}{d} \mathbf{B}_0. \quad (6.77)$$

Thus, given that $\mu \simeq 10^5 \mu_0$ for mu-metal, we can reduce the magnetic field-strength inside the shell by almost a factor of 1000 using a shell whose thickness is only 1/100 th of its radius. Note, however, that as the external field-strength, B_0 , is increased, the mu-metal shell eventually saturates, and μ/μ_0 gradually falls to unity. Thus, extremely strong magnetic fields (typically, $B_0 \gtrsim 1$ tesla) are hardly shielded at all by mu-metal, or similar magnetic materials.

6.9 Magnetic Energy

Consider an electrical conductor. Suppose that a battery with an electromotive field \mathbf{E}' is feeding energy into this conductor. The energy is either dissipated as heat, or is used to generate a magnetic field. Ohm's law inside the conductor gives

$$\mathbf{j}_t = \sigma (\mathbf{E} + \mathbf{E}'), \quad (6.78)$$

where \mathbf{j}_t is the true current density, σ is the conductivity, and \mathbf{E} is the inductive electric field. Taking the scalar product with \mathbf{j}_t , we obtain

$$\mathbf{E}' \cdot \mathbf{j}_t = \frac{j_t^2}{\sigma} - \mathbf{E} \cdot \mathbf{j}_t. \quad (6.79)$$

The left-hand side of this equation represents the rate at which the battery does work on the conductor. The first term on the right-hand side is the rate of Joule heating inside the conductor. We tentatively identify the remaining term with the rate at which energy is fed into the magnetic field. If all fields are quasi-stationary (i.e., slowly varying) then the displacement current can be neglected, and Equation (6.8) reduces to $\nabla \times \mathbf{H} = \mathbf{j}_t$. Substituting this expression into Equation (6.79) and integrating over all space, we get

$$\int \mathbf{E}' \cdot (\nabla \times \mathbf{H}) dV = \int \frac{(\nabla \times \mathbf{H})^2}{\sigma} dV - \int \mathbf{E} \cdot (\nabla \times \mathbf{H}) dV. \quad (6.80)$$

The last term can be integrated by parts using the identity

$$\nabla \cdot (\mathbf{E} \times \mathbf{H}) \equiv \mathbf{H} \cdot (\nabla \times \mathbf{E}) - \mathbf{E} \cdot (\nabla \times \mathbf{H}). \quad (6.81)$$

Making use of the divergence theorem, as well as Equation (1.3), we get

$$\int \mathbf{E} \cdot (\nabla \times \mathbf{H}) dV = - \int \mathbf{H} \cdot \frac{\partial \mathbf{B}}{\partial t} dV - \int (\mathbf{E} \times \mathbf{H}) \cdot d\mathbf{S}. \quad (6.82)$$

Because $\mathbf{E} \times \mathbf{H}$ falls off at least as fast as $1/r^5$ in electrostatic and quasi-stationary magnetic fields ($1/r^2$ comes from electric monopole fields, and $1/r^3$ from magnetic dipole fields), the surface integral in the previous expression can be neglected. Of course, this is not the case for radiation fields, for which \mathbf{E} and \mathbf{H} fall off like $1/r$. (See Section 1.8.) Thus, the constraint of "quasi-stationarity" effectively means that the fields vary sufficiently slowly that any radiation fields can be neglected.

The total power expended by the battery can now be written

$$\int \mathbf{E}' \cdot (\nabla \times \mathbf{H}) dV = \int \frac{(\nabla \times \mathbf{H})^2}{\sigma} dV + \int \mathbf{H} \cdot \frac{\partial \mathbf{B}}{\partial t} dV. \quad (6.83)$$

The first term on the right-hand side has already been identified as the energy loss rate due to Joule heating. The last term is obviously the rate at which energy is fed into the magnetic field. The variation δU in the magnetic field energy can therefore be written

$$\delta U = \int \mathbf{H} \cdot \delta \mathbf{B} dV. \quad (6.84)$$

In order to make Equation (6.84) integrable, we must assume a functional relationship between \mathbf{H} and \mathbf{B} . For a medium that magnetizes linearly, the integration can be carried out, in much the same manner as Equation (4.71), to give

$$U = \frac{1}{2} \int \mathbf{H} \cdot \mathbf{B} dV. \quad (6.85)$$

Thus, the magnetostatic energy density inside a linear magnetic material is given by

$$W = \frac{1}{2} \mathbf{H} \cdot \mathbf{B}. \quad (6.86)$$

Unfortunately, most interesting magnetic materials, such as ferromagnets, exhibit a nonlinear relationship between \mathbf{H} and \mathbf{B} . For such materials, Equation (6.84) can only be integrated between definite states, and the result, in general, depends on the past history of the sample. For ferromagnets, the integral of Equation (6.84) has a finite, non-zero value when \mathbf{B} is integrated around a complete magnetization cycle. This cyclic energy loss is given by

$$\Delta U = \int \oint \mathbf{H} \cdot d\mathbf{B} dV. \quad (6.87)$$

In other words, the energy expended per unit volume when a magnetic material is carried through a magnetization cycle is equal to the area of its hysteresis loop as plotted in a graph of B against H . Thus, it is particularly important to ensure that the magnetic used to form transformer cores possess hysteresis loops with comparatively small areas, otherwise the transformers are likely to be extremely inefficient.

6.10 Exercises

- 6.1 Given that the bound charge density associated with a polarization field $\mathbf{P}(\mathbf{r})$ is $\sigma_b = -\nabla \cdot \mathbf{P}$, use charge conservation to deduce that the current density due to bound charges is

$$\mathbf{j}_p = \frac{\partial \mathbf{P}}{\partial t}.$$

- 6.2 Given that $\nabla \times \mathbf{H} = \mathbf{0}$ in the absence of true currents, and $\mathbf{H} = \mathbf{B}/\mu_0 - \mathbf{M}$, demonstrate that the current density due to magnetization currents is

$$\mathbf{j}_m = \nabla \times \mathbf{M}.$$

- 6.3 A cylindrical hole of radius a is bored parallel to the axis of a cylindrical conductor of radius $b > a$ which carries a uniformly distributed current of density j running parallel to its axis. The distance between the center of the conductor and the center of the hole is x_0 . Find the \mathbf{B} field in the hole.

- 6.4 A sphere of radius a carries a uniform surface charge density σ . The sphere is rotated about a diameter with constant angular velocity ω . Find the vector potential and the \mathbf{B} field both inside and outside the sphere.
- 6.5 Find the \mathbf{B} and \mathbf{H} fields inside and outside a spherical shell of inner radius a and outer radius b which is magnetized permanently to a constant magnetization \mathbf{M} .
- 6.6 A long hollow, right cylinder of inner radius a and outer radius b , and of relative permeability μ , is placed in a region of initially uniform magnetic flux density \mathbf{B} at right-angles to the field. Find the flux density at all points in space. Neglect end effects.
- 6.7 A transformer consists of a thin uniform ring of ferromagnetic material of radius a , cross-sectional area A , and magnetic permeability μ . The primary circuit is wrapped N_1 times around one side of the ring, and the secondary N_2 times around the other side. Show that the mutual inductance between the two circuits is

$$M = \frac{\mu N_1 N_2 A}{2\pi a}.$$

Suppose that a thin gap of thickness $d \ll a$ is cut in a part of the ring in which there are no windings. What is the new mutual inductance of the two circuits? Suppose that the gap is filled with ferromagnetic material of permeability μ' . What, now, is the mutual inductance of the circuits? You may neglect flux-leakage (i.e., you may assume that magnetic field-lines do not leak out of the transformer core into the surrounding vacuum, except in the gap).

7 Wave Propagation in Uniform Dielectric Media

7.1 Introduction

As is easily demonstrated, the fields associated with an electromagnetic wave propagating through a uniform dielectric medium of dielectric constant ϵ satisfy

$$\left(\frac{\epsilon}{c^2} \frac{\partial^2}{\partial t^2} - \nabla^2 \right) \mathbf{E} = \mathbf{0}, \quad (7.1)$$

and

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}. \quad (7.2)$$

The plane wave solutions to these equations are well known:

$$\mathbf{E} = \mathbf{E}_0 \exp [i (\mathbf{k} \cdot \mathbf{r} - \omega t)], \quad (7.3)$$

$$\mathbf{B} = \mathbf{B}_0 \exp [i (\mathbf{k} \cdot \mathbf{r} - \omega t)], \quad (7.4)$$

where \mathbf{E}_0 and \mathbf{B}_0 are constant vectors,

$$\frac{\omega^2}{k^2} = \frac{c^2}{\epsilon}, \quad (7.5)$$

and

$$\mathbf{B}_0 = \frac{\mathbf{k} \times \mathbf{E}_0}{\omega}. \quad (7.6)$$

The phase velocity of the wave is given by

$$v = \frac{\omega}{k} = \frac{c}{n}, \quad (7.7)$$

where

$$n = \sqrt{\epsilon} \quad (7.8)$$

is the medium's *refractive index*. Thus, in a conventional dielectric medium (i.e., ϵ real and greater than unity), an electromagnetic wave propagates with a phase velocity that is slower than the velocity of light in vacuum.

In some dielectric media, the dielectric constant, ϵ , is complex. According to Equation (7.5), this leads to a complex wavevector, \mathbf{k} (assuming that the angular frequency is real). Thus, for a wave propagating in the x -direction, we obtain

$$\mathbf{E} = \mathbf{E}_0 \exp(i [\operatorname{Re}(k) x - \omega t]) \exp[-\operatorname{Im}(k) x]. \quad (7.9)$$

In other words, a complex dielectric constant leads to the attenuation (or amplification) of the wave, as it propagates through the medium.

Up to now, we have tacitly assumed that ϵ is the same for waves of all frequencies. In practice, ϵ varies (in some cases, strongly) with the wave frequency. Consequently, waves of different frequencies propagate through a dielectric medium at different phase velocities, leading to the dispersion of wave pulses. Moreover, there may exist frequency bands in which the waves are attenuated (i.e., absorbed). All of this makes the problem of determining the behavior of a wave pulse as it propagates through a dielectric medium a far from straightforward task. Of course, the solution to this problem for a wave pulse traveling through a vacuum is fairly trivial: that is, the pulse propagates at the velocity c without changing shape. What is the equivalent result for the case of a dielectric medium? This is a significant question, because most of our information regarding the universe is obtained from the study of electromagnetic waves emitted by distant objects. All of these waves have to propagate through dispersive media (e.g., the interstellar medium, the ionosphere, the atmosphere) before reaching us. It is, therefore, vitally important that we understand which aspects of these wave signals are predominantly determined by the wave sources, and which are strongly modified by the dispersive media through which the signals have propagated in order to reach us.

7.2 Form of Dielectric Constant

Consider an electromagnetic wave propagating through a transparent, isotropic, dielectric medium. The electric displacement inside the medium is given by

$$\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P}, \quad (7.10)$$

where \mathbf{P} is the electric polarization. Because electrons are much lighter than ions (or atomic nuclei), we would expect the former to displace further than the latter under the influence of an electric field. Thus, to a first approximation, the polarization, \mathbf{P} , is determined by the electron response to the wave. Suppose that the electrons displace an average distance \mathbf{s} from their rest positions in the presence of the wave. It follows that

$$\mathbf{P} = -N e \mathbf{s}, \quad (7.11)$$

where N is the number density of electrons, and $-e$ the electron charge.

Let us assume that the electrons are bound “quasi-elastically” to their rest positions, so that they seek to return to these positions when displaced from them by an electric field. It follows that \mathbf{s} satisfies a differential equation of the form

$$m \ddot{\mathbf{s}} + f \mathbf{s} = -e \mathbf{E}, \quad (7.12)$$

where m is the electron mass, $-f \mathbf{s}$ is the restoring force, and $\dot{}$ denotes a partial derivative with respect to time. The previous equation can also be written

$$\ddot{\mathbf{s}} + g \omega_0 \dot{\mathbf{s}} + \omega_0^2 \mathbf{s} = -\frac{e}{m} \mathbf{E}, \quad (7.13)$$

where

$$\omega_0^2 = \frac{f}{m} \quad (7.14)$$

is the characteristic oscillation frequency of the electrons. In almost all dielectric media, this frequency lies in the far ultraviolet region of the electromagnetic spectrum. Note that we have added a phenomenological damping term, $g \omega_0 \dot{\mathbf{s}}$, to Equation (7.13), in order to take into account the fact that an electron excited by an impulsive electric field does not oscillate for ever. In fact, electrons in dielectric media act like high-Q oscillators, which is another way of saying that the dimensionless damping constant, g , is typically much less than unity. Thus, an electron in a dielectric medium “rings” for a long time after being excited by an electromagnetic impulse.

Let us assume that the electrons oscillate in sympathy with the wave at the wave frequency, ω . It follows from Equation (7.13) that

$$\mathbf{s} = -\frac{(e/m) \mathbf{E}}{\omega_0^2 - \omega^2 - i g \omega \omega_0}. \quad (7.15)$$

Here, we have neglected the response of the electrons to the magnetic component of the wave. It is easily demonstrated that this is a good approximation provided the electrons do not oscillate with relativistic velocities (i.e., provided the amplitude of the wave is not too large—see Section 7.7). Thus, Equation (7.11) yields

$$\mathbf{P} = \frac{(N e^2/m) \mathbf{E}}{\omega_0^2 - \omega^2 - i g \omega \omega_0}. \quad (7.16)$$

Because, by definition,

$$\mathbf{D} = \epsilon_0 \epsilon \mathbf{E} = \epsilon_0 \mathbf{E} + \mathbf{P}, \quad (7.17)$$

it follows that

$$\epsilon(\omega) \equiv n^2(\omega) = 1 + \frac{(N e^2/\epsilon_0 m)}{\omega_0^2 - \omega^2 - i g \omega \omega_0}. \quad (7.18)$$

Thus, the index of refraction is indeed frequency dependent. Because ω_0 typically lies in the ultraviolet region of the spectrum (and $g \ll 1$), it is clear that the denominator, $\omega_0^2 - \omega^2 - i g \omega \omega_0 \simeq \omega_0^2 - \omega^2$, is positive throughout the visible spectrum, and is larger at the red than at the blue end of this spectrum. This implies that blue light is refracted more strongly than red light. This state of affairs, in which higher frequency waves are refracted more strongly than lower frequency waves, is termed *normal dispersion*. Incidentally, an expression, like the previous one, that (effectively) specifies the phase velocity of waves propagating through a dielectric medium, as a function of their frequency, is usually called a *dispersion relation*.

Let us now suppose that there are N molecules per unit volume, with Z electrons per molecule, and that, instead of a single oscillation frequency for all electrons, there are f_i electrons per molecule with oscillation frequency ω_i and damping constant g_i . It is easily demonstrated that Equation (7.18) generalizes to give

$$n^2(\omega) = 1 + \frac{N e^2}{\epsilon_0 m} \sum_i \frac{f_i}{\omega_i^2 - \omega^2 - i g_i \omega \omega_i}, \quad (7.19)$$

where the *oscillator strengths*, f_i , satisfy the sum rule,

$$\sum_i f_i = Z. \quad (7.20)$$

A more exact quantum mechanical treatment of the response of an atom, or molecule, to a low amplitude electromagnetic wave also leads to a dispersion relation of the previous form, except that the quantities f_i , ω_i , and g_i can, in principle, be calculated exactly. In practice, this is too difficult, except in very simple cases.

Because the damping constants, g_i , are generally small compared to unity, it follows from Equation (7.19) that $n(\omega)$ is a predominately real quantity at most wave frequencies. The factor $(\omega_i^2 - \omega^2)^{-1}$ is positive for $\omega < \omega_i$, and negative for $\omega > \omega_i$. Thus, at low frequencies (i.e., below the smallest ω_i) all of the terms appearing in the sum on the right-hand side of (7.19) are positive, and $n(\omega)$ is consequently greater than unity. As ω is raised, such that it exceeds successive ω_i values, more and more negative terms occur in the sum, until eventually the whole sum is negative, and $n(\omega)$ is less than unity. Hence, at very high frequencies, electromagnetic waves propagate through dielectric media with phase velocities that exceed the velocity of light in a vacuum. For $\omega \simeq \omega_i$, Equation (7.19) predicts strong variation of the refractive index with frequency. Let us examine this phenomenon more closely.

7.3 Anomalous Dispersion and Resonant Absorption

When ω is approximately equal to ω_i , the dispersion relation (7.19) reduces to

$$n^2 = n_i^2 + \frac{N e^2 f_i / \epsilon_0 m}{\omega_i^2 - \omega^2 - i g_i \omega \omega_i}, \quad (7.21)$$

where n_i is the average contribution in the vicinity of $\omega = \omega_i$ of all the other resonances (also included in n_i is the contribution 1 of the vacuum displacement current, which was previously written separately). The refractive index is clearly complex. For a wave propagating in the x -direction,

$$\mathbf{E} = \mathbf{E}_0 \exp[i(\omega/c)(\text{Re}(n)x - ct)] \exp[-(\omega/c)\text{Im}(n)x]. \quad (7.22)$$

Thus, the phase velocity of the wave is determined by the real part of the refractive index via

$$v = \frac{c}{\text{Re}(n)}. \quad (7.23)$$

Furthermore, a positive imaginary component of the refractive index leads to the attenuation of the wave as it propagates.

Let

$$a^2 = \frac{N e^2 f_i}{\epsilon_0 m \omega_i^2}, \quad (7.24)$$

$$x = \frac{\omega^2 - \omega_i^2}{\omega_i^2}, \quad (7.25)$$

$$y = \frac{[\text{Re}(n)]^2 - [\text{Im}(n)]^2}{a^2}, \quad (7.26)$$

$$z = \frac{2 \text{Re}(n) \text{Im}(n)}{a^2}, \quad (7.27)$$

where a , x , y , z are all dimensionless quantities. It follows from Equation (7.21) that

$$y = \frac{n_i^2}{a^2} - \frac{x}{x^2 + g_i^2(1+x)}, \quad (7.28)$$

$$z = \frac{g_i \sqrt{1+x}}{x^2 + g_i^2(1+x)}. \quad (7.29)$$

Let us adopt the physical ordering $g_i \ll 1$. In this case, the extrema of the function $y(x)$ occur at $x \approx \pm g_i$. In fact, it is easily demonstrated that

$$y_{\min} = y(x = g_i) = \frac{n_i^2}{a^2} - \frac{1}{2g_i}, \quad (7.30)$$

$$y_{\max} = y(x = -g_i) = \frac{n_i^2}{a^2} + \frac{1}{2g_i}. \quad (7.31)$$

The maximum value of the function $z(x)$ occurs at $x = 0$. In fact,

$$z_{\max} = \frac{1}{g_i}. \quad (7.32)$$

Note also that

$$z(x = \pm g_i) = \frac{1}{2g_i}. \quad (7.33)$$

Figure 7.1 shows a sketch of the functions $y(x)$ and $z(x)$. These curves are also indicative of the variation of $\text{Re}(n)$ and $\text{Im}(n)$, respectively, with frequency, ω , in the vicinity of the resonant frequency, ω_i . Recall that normal dispersion is associated with an increase in $\text{Re}(n)$ with increasing ω . The reverse situation is termed *anomalous dispersion*. It is clear, from the figure, that normal dispersion occurs everywhere, except at wave frequencies in the immediate neighborhood of the resonant frequency, ω_i . It is also clear that the imaginary part of the refractive index is only appreciable in those regions of the electromagnetic spectrum where anomalous dispersion takes place. A positive imaginary component of the refractive index implies that the wave is absorbed as it propagates through the medium. Consequently, the regions of the spectrum in which $\text{Im}(n)$ is appreciable are called regions of *resonant absorption*. Anomalous dispersion and resonant absorption take place in the vicinity of the i th resonance when $|\omega - \omega_i|/\omega_i < O(g_i)$. Because the damping constants, g_i , are, in practice, very small compared to unity, the regions of the spectrum in which resonant absorption takes place are strongly localized in the vicinity of the various resonant frequencies.

The dispersion relation (7.19) only takes electron resonances into account. Of course, there are also resonances associated with displacements of the ions (or atomic nuclei). The off-resonance contributions to the right-hand side of Equation (7.19) from the ions are typically smaller than those from the electrons by a factor of order m/M (where M is a typical ion mass). Nevertheless, the ion contributions are important, because they give rise to anomalous dispersion and resonant absorption close to the ion resonant frequencies. The ion resonances associated with the stretching

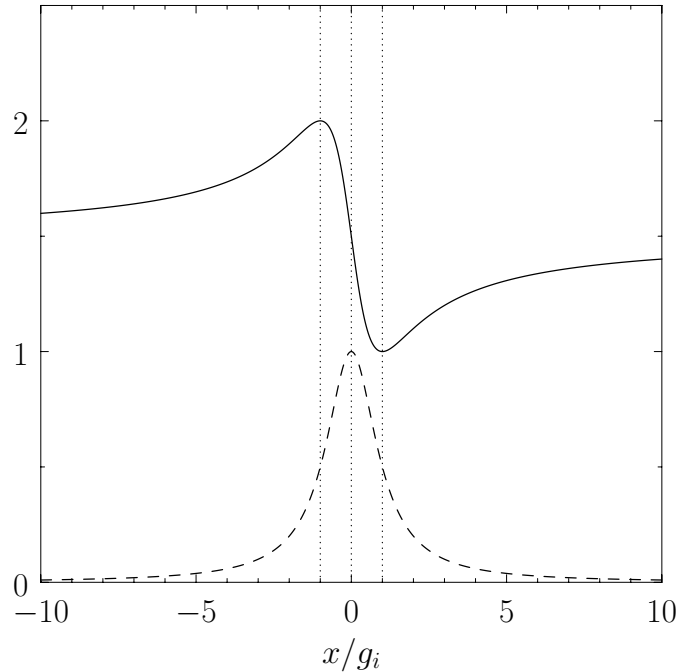


Figure 7.1: Sketch of the variation of the functions y and z with x . The solid and dashed curves shows y/g_i and z/g_i , respectively.

and bending of molecular bonds usually lie in the infrared region of the electromagnetic spectrum. Those resonances associated with molecular rotation (which only affect the dispersion relation if the molecule is polar) occur in the microwave region of the spectrum. Both air and water exhibit strong resonant absorption of electromagnetic waves in both the ultraviolet and infrared regions of the spectrum. In the former case, this is due to electron resonances, and in the latter to ion resonances. The visible region of the spectrum exists as a narrow window, lying between these two regions, in which there is comparatively little attenuation of electromagnetic waves.

7.4 Wave Propagation in Conducting Media

In the limit $\omega \rightarrow 0$, there is a significant difference in the response of a dielectric medium to an electromagnetic wave, depending on whether the lowest resonant frequency is zero or non-zero. For insulators, the lowest resonant frequency is different from zero. In this case, the low frequency refractive index is predominately real, and is also greater than unity. In a conducting medium, on the other hand, some fraction, f_0 , of the electrons are “free,” in the sense of having $\omega_0 = 0$. In this situation, the low frequency dielectric constant takes the form

$$\epsilon(\omega) \equiv n^2(\omega) = n_0^2 + i \frac{N e^2}{\epsilon_0 m \omega} \frac{f_0}{(\gamma_0 - i \omega)}, \quad (7.34)$$

where n_0 is the contribution to the refractive index from all of the other resonances, and $\gamma_0 = \lim_{\omega_0 \rightarrow 0} g_0 \omega_0$. Consider the Ampère-Maxwell equation,

$$\nabla \times \mathbf{B} = \mu_0 \left(\mathbf{j}_t + \frac{\partial \mathbf{D}}{\partial t} \right). \quad (7.35)$$

Here, \mathbf{j}_t is the true current: that is, the current carried by free, as opposed to bound, charges. Let us assume that the medium in question obeys Ohm's law, $\mathbf{j}_t = \sigma \mathbf{E}$, and has a "normal" dielectric constant n_0^2 . Here, σ is the conductivity. Assuming an $\exp(-i\omega t)$ time dependence of all field quantities, the previous equation yields

$$\nabla \times \mathbf{B} = -i \epsilon_0 \mu_0 \omega \left(n_0^2 + i \frac{\sigma}{\epsilon_0 \omega} \right) \mathbf{E}. \quad (7.36)$$

Suppose, however, that we do not explicitly use Ohm's law but, instead, attribute all of the properties of the medium to the dielectric constant. In this case, the effective dielectric constant of the medium is equivalent to the term in round brackets on the right-hand side of the previous equation: that is,

$$\epsilon(\omega) \equiv n^2(\omega) = n_0^2 + i \frac{\sigma}{\epsilon_0 \omega}. \quad (7.37)$$

A comparison of this term with Equation (7.34) yields the following expression for the conductivity,

$$\sigma = \frac{f_0 N e^2}{m(\gamma_0 - i\omega)}. \quad (7.38)$$

Thus, at low frequencies, conductors possess predominately real conductivities (i.e., the current remains in phase with the electric field). However, at higher frequencies, the conductivity becomes complex. At such frequencies, there is little meaningful distinction between a conductor and an insulator, because the "conductivity" contribution to $\epsilon(\omega)$ appears as a resonant amplitude, just like the other contributions. For a good conductor, such as copper, the conductivity remains predominately real for all frequencies up to, and including, those in the microwave region of the electromagnetic spectrum.

The conventional way in which to represent the complex refractive index of a conducting medium (in the low frequency limit) is to write it in terms of a real "normal" dielectric constant, $\epsilon = n_0^2$, and a real conductivity, σ . Thus, from Equation (7.37),

$$n^2(\omega) = \epsilon + i \frac{\sigma}{\epsilon_0 \omega}. \quad (7.39)$$

For a poor conductor (i.e., $\sigma/\epsilon \epsilon_0 \omega \ll 1$), we find that

$$k = n \frac{\omega}{c} \simeq \sqrt{\epsilon} \frac{\omega}{c} + i \frac{\sigma}{2\sqrt{\epsilon} \epsilon_0 c}. \quad (7.40)$$

In this limit, $\text{Re}(k) \gg \text{Im}(k)$, and the attenuation of the wave, which is governed by $\text{Im}(k)$ [see Equation (7.9)], is independent of the frequency. Thus, for a poor conductor, the wave acts like a

wave propagating through a conventional dielectric of dielectric constant ϵ , except that it attenuates gradually over a distance of very many wavelengths. For a good conductor (i.e., $\sigma/\epsilon\epsilon_0\omega \gg 1$), we obtain

$$k \simeq e^{i\pi/4} \sqrt{\mu_0 \sigma \omega}. \quad (7.41)$$

It follows from Equation (7.5) that

$$\frac{c B_0}{E_0} = \frac{k c}{\omega} = e^{i\pi/4} \sqrt{\frac{\sigma}{\epsilon_0 \omega}}. \quad (7.42)$$

Thus, the phase of the magnetic field lags that of the electric field by $\pi/4$ radians. Moreover, the magnitude of $c B_0$ is much larger than that of E_0 (because $\sigma/\epsilon_0\omega \gg \epsilon > 1$). It follows that the wave energy is almost entirely magnetic in nature. Clearly, an electromagnetic wave propagating through a good conductor has markedly different properties to a wave propagating through a conventional dielectric. For a wave propagating in the x -direction, the amplitudes of the electric and magnetic fields attenuate like $\exp(-x/d)$, where

$$d = \sqrt{\frac{2}{\mu_0 \sigma \omega}} \quad (7.43)$$

is termed the *skin depth*. It is apparent that an electromagnetic wave incident on a conducting medium will not penetrate more than a few skin depths into that medium.

7.5 High Frequency Limit

Consider the behavior of the dispersion relation (7.19) in the high frequency limit $\omega \gg \omega_i$ (for all i). In this case, the relation simplifies considerably to give

$$n^2(\omega) = 1 - \frac{\omega_p^2}{\omega^2}, \quad (7.44)$$

where the quantity

$$\omega_p = \sqrt{\frac{N Z e^2}{\epsilon_0 m}} \quad (7.45)$$

is called the *plasma frequency*. The wavenumber in the high frequency limit is given by

$$k = n \frac{\omega}{c} = \frac{(\omega^2 - \omega_p^2)^{1/2}}{c}. \quad (7.46)$$

This expression is only valid in dielectrics when $\omega \gg \omega_p$. Thus, the refractive index is real, and slightly less than unity, giving waves that propagate without attenuation at a phase velocity slightly larger than the velocity of light in vacuum. However, in certain ionized media (in particular, in tenuous plasmas, such as occur in the ionosphere) the electrons are free, and the damping is negligible. In this case, Equations (7.44) and (7.46) are valid even when $\omega < \omega_p$. It follows

that a wave can only propagate through a tenuous plasma when its frequency exceeds the plasma frequency (in which case, it has a real wavenumber). If wave frequency is less than the plasma frequency then, according to Equation (7.46), the wavenumber is purely imaginary, and the wave is unable to propagate. This phenomenon accounts for the fact that long-wave and medium-wave (terrestrial) radio signals can be received even when the transmitter lies over the horizon. The frequency of these waves is less than the plasma frequency of the ionosphere, which reflects them (see Chapter 8), so they become trapped between the ionosphere and the surface of the Earth (which is also a good reflector of radio waves), and can, in certain cases, travel many times around the Earth before being attenuated. Unfortunately, this scheme does not work very well for medium-wave signals at night. The problem is that the plasma frequency of the ionosphere is proportional to the square root of the number density of free ionospheric electrons. These free electrons are generated through the ionization of neutral molecules by ultraviolet radiation from the Sun. Of course, there is no radiation from the Sun at night, so the density of free electrons starts to drop as the electrons gradually recombine with ions in the ionosphere. Eventually, the plasma frequency of the ionosphere falls below the frequency of medium-wave radio signals, causing them to be transmitted through the ionosphere into outer space. The ionosphere appears almost completely transparent to high frequency signals such as TV and FM radio signals. Thus, this type of signal is not reflected by the ionosphere. Consequently, to receive such signals it is necessary to be in the line of sight of the relevant transmitter.

7.6 Polarization of Electromagnetic Waves

The electric component of an electromagnetic plane wave can oscillate in any direction normal to the direction of wave propagation (which is parallel to the \mathbf{k} -vector). Suppose that the wave is propagating in the z -direction. It follows that the electric field can oscillate in any direction that lies in the x - y plane. The actual direction of oscillation determines the *polarization* of the wave. For instance, a vacuum electromagnetic wave of angular frequency ω that is polarized in the x -direction has the associated electric field

$$\mathbf{E} = E_0 \cos(\omega t - k z) \mathbf{e}_x, \quad (7.47)$$

where $\omega = k c$. Likewise, a wave polarized in the y -direction has the electric field

$$\mathbf{E} = E_0 \cos(\omega t - k z) \mathbf{e}_y. \quad (7.48)$$

These two waves are termed *linearly polarized*, because the electric field vector oscillates in a straight-line. However, other types of polarization are possible. For instance, if we combine two linearly polarized waves of equal amplitude, one polarized in the x -direction, and one in the y -direction, that oscillate $\pi/2$ radians out of phase, then we obtain a *circularly polarized* wave:

$$\mathbf{E} = E_0 \cos(\omega t - k z) \mathbf{e}_x + E_0 \sin(\omega t - k z) \mathbf{e}_y. \quad (7.49)$$

This nomenclature arises from the fact that the tip of the electric field vector traces out a circle in the plane normal to the direction of wave propagation. To be more exact, the previous wave is a

right-hand circularly polarized wave, because if the thumb of the right hand points in the direction of wave propagation then the electric field vector rotates in the same sense as the fingers of this hand. Conversely, a left-hand circularly polarized wave takes the form

$$\mathbf{E} = E_0 \cos(\omega t - kz) \mathbf{e}_x - E_0 \sin(\omega t - kz) \mathbf{e}_y. \quad (7.50)$$

Finally, if the x - and y -components of the electric field in the previous two expressions have different (non-zero) amplitudes then we obtain right-hand and left-hand *elliptically polarized* waves, respectively. This nomenclature arises from the fact that the tip of the electric field vector traces out an ellipse in the plane normal to the direction of wave propagation.

7.7 Faraday Rotation

The electromagnetic force acting on an electron is given by

$$\mathbf{f} = -e (\mathbf{E} + \mathbf{v} \times \mathbf{B}). \quad (7.51)$$

If the \mathbf{E} and \mathbf{B} fields in question are due to an electromagnetic wave propagating through a dielectric medium then

$$|\mathbf{B}| = \frac{n}{c} |\mathbf{E}|, \quad (7.52)$$

where n is the refractive index. It follows that the ratio of the magnetic to the electric forces acting on the electron is nv/c . In other words, the magnetic force is completely negligible unless the wave amplitude is sufficiently high that the electron moves relativistically in response to the wave. This state of affairs is rare, but can occur when intense laser beams are made to propagate through plasmas.

Suppose, however, that the dielectric medium contains an externally generated magnetic field, \mathbf{B} . This can easily be made much stronger than the optical magnetic field. In this case, it is possible for a magnetic field to affect the propagation of low amplitude electromagnetic waves. The electron equation of motion (7.12) generalizes to

$$m \ddot{\mathbf{s}} + f \mathbf{s} = -e (\mathbf{E} + \dot{\mathbf{s}} \times \mathbf{B}), \quad (7.53)$$

where any damping of the motion has been neglected. Let \mathbf{B} be directed in the positive z -direction, and let the wave propagate in the same direction. These assumptions imply that the \mathbf{E} and \mathbf{s} vectors lie in the x - y plane. The previous equation reduces to

$$(\omega_0^2 - \omega^2) s_x - i \omega \Omega s_y = -\frac{e}{m} E_x, \quad (7.54)$$

$$(\omega_0^2 - \omega^2) s_y + i \omega \Omega s_x = -\frac{e}{m} E_y, \quad (7.55)$$

provided that all perturbed quantities have an $\exp(-i \omega t)$ time dependence. Here, $\omega_0 = \sqrt{f/m}$, and

$$\Omega = \frac{e B}{m} \quad (7.56)$$

is the *electron cyclotron frequency*. Let

$$E_{\pm} = E_x \pm i E_y, \quad (7.57)$$

and

$$s_{\pm} = s_x \pm i s_y. \quad (7.58)$$

Note that

$$E_x = \frac{1}{2}(E_+ + E_-), \quad (7.59)$$

$$E_y = \frac{1}{2i}(E_+ - E_-). \quad (7.60)$$

Equations (7.54) and (7.55) reduce to

$$(\omega_0^2 - \omega^2 - \omega \Omega) s_+ = -\frac{e}{m} E_+, \quad (7.61)$$

$$(\omega_0^2 - \omega^2 + \omega \Omega) s_- = -\frac{e}{m} E_-. \quad (7.62)$$

Defining $P_{\pm} = P_x \pm i P_y$, it follows from Equation (7.11) that

$$P_{\pm} = \frac{(N e^2/m) E_{\pm}}{\omega_0^2 - \omega^2 \mp \omega \Omega}. \quad (7.63)$$

Finally, from Equation (7.17), we can write

$$\epsilon_{\pm} \equiv n_{\pm}^2 = 1 + \frac{P_{\pm}}{\epsilon_0 E_{\pm}}, \quad (7.64)$$

giving

$$n_{\pm}^2(\omega) = 1 + \frac{(N e^2/\epsilon_0 m)}{\omega_0^2 - \omega^2 \mp \omega \Omega}. \quad (7.65)$$

According to the dispersion relation (7.65), the refractive index of a magnetized dielectric medium can take one of two possible values, which presumably correspond to two different types of wave propagating parallel to the z -axis. The first wave has the refractive index n_+ , and an associated electric field [see Equations (7.59) and (7.60)]

$$E_x = E_0 \cos[(\omega/c)(n_+z - ct)], \quad (7.66)$$

$$E_y = E_0 \sin[(\omega/c)(n_+z - ct)]. \quad (7.67)$$

This corresponds to a left-hand circularly polarized wave propagating in the z -direction at the phase velocity c/n_+ . The second wave has the refractive index n_- , and an associated electric field

$$E_x = E_0 \cos[(\omega/c)(n_-z - ct)], \quad (7.68)$$

$$E_y = -E_0 \sin[(\omega/c)(n_-z - ct)]. \quad (7.69)$$

This corresponds to a right-hand circularly polarized wave propagating in the z -direction at the phase velocity c/n_- . It is clear from Equation (7.65) that $n_+ > n_-$. We conclude that, in the presence of a z -directed magnetic field, a z -directed left-hand circularly polarized wave propagates at a phase velocity that is slightly less than that of the corresponding right-hand wave. It should be remarked that the refractive index is always real (in the absence of damping), so the magnetic field gives rise to no net absorption of electromagnetic radiation. This is not surprising because a magnetic field does no work on charged particles, and cannot therefore transfer energy from a wave propagating through a dielectric medium to the medium's constituent particles.

We have seen that right-hand and left-hand circularly polarized waves propagate through a magnetized dielectric medium at slightly different phase velocities. What does this imply for the propagation of a plane polarized wave? Let us add the left-hand wave whose electric field is given by Equations (7.66) and (7.67) to the right-hand wave whose electric field is given by Equations (7.68) and (7.69). In the absence of a magnetic field, $n_+ = n_- = n$, and we obtain

$$E_x = 2 E_0 \cos[(\omega/c)(n z - c t)], \quad (7.70)$$

$$E_y = 0. \quad (7.71)$$

This, of course, corresponds to a plane wave (polarized along the x -direction) propagating along the z -axis at the phase velocity c/n . In the presence of a magnetic field, we obtain

$$E_x = 2 E_0 \cos[(\omega/c)(n z - c t)] \cos[(\omega/2 c)(n_+ - n_-) z], \quad (7.72)$$

$$E_y = 2 E_0 \cos[(\omega/c)(n z - c t)] \sin[(\omega/2 c)(n_+ - n_-) z], \quad (7.73)$$

where

$$n = \frac{1}{2} (n_+ + n_-) \quad (7.74)$$

is the mean index of refraction. Equations (7.72) and (7.73) describe a plane wave whose angle of polarization with respect to the x -axis,

$$\chi = \tan^{-1}(E_y/E_x), \quad (7.75)$$

rotates as the wave propagates along the z -axis at the phase velocity c/n . In fact, the angle of polarization is given by

$$\chi = \frac{\omega}{2 c} (n_+ - n_-) z, \quad (7.76)$$

which clearly increases linearly with the distance traveled by the wave parallel to the magnetic field. This rotation of the plane of polarization of a linearly polarized wave propagating through a magnetized dielectric medium is known as *Faraday rotation* (because it was discovered by Michael Faraday in 1845).

Assuming that the cyclotron frequency, Ω , is relatively small compared to the wave frequency, ω , and also that ω does not lie close to the resonant frequency, ω_0 , it is easily demonstrated that

$$n \simeq 1 + \frac{(N e^2 / \epsilon_0 m)}{\omega_0^2 - \omega^2}, \quad (7.77)$$

and

$$n_+ - n_- \simeq \frac{N e^2}{\epsilon_0 m n} \frac{\omega \Omega}{(\omega_0^2 - \omega^2)^2}. \quad (7.78)$$

It follows that the rate at which the plane of polarization of an electromagnetic wave rotates as the distance traveled by the wave increases is

$$\frac{d\chi}{dl} = \frac{\kappa(\omega) N B_{\parallel}}{n(\omega)}, \quad (7.79)$$

where B_{\parallel} is the component of the magnetic field along the direction of propagation of the wave, and

$$\kappa(\omega) = \frac{e^3}{2 \epsilon_0 m^2 c} \frac{\omega^2}{(\omega_0^2 - \omega^2)^2}. \quad (7.80)$$

If the medium in question is a tenuous plasma then $n \simeq 1$, and $\omega_0 = 0$. Thus,

$$\frac{d\chi}{dl} \simeq \frac{e^3}{2 \epsilon_0 m^2 c} \frac{N B_{\parallel}}{\omega^2}. \quad (7.81)$$

In this case, the rate at which the plane of polarization rotates is proportional to the product of the electron number density and the parallel magnetic field-strength. Moreover, the plane of rotation rotates faster for low frequency waves than for high frequency waves. The total angle by which the plane of polarization is twisted after passing through a magnetized plasma is given by

$$\Delta\chi \simeq \frac{e^3}{2 \epsilon_0 m^2 c \omega^2} \int N(l) B_{\parallel}(l) dl, \quad (7.82)$$

assuming that N and B_{\parallel} vary on length-scales that are large compared to the wavelength of the radiation. This formula is regularly employed in radio astronomy to infer the magnetic field-strength in interstellar space.

7.8 Wave Propagation in Magnetized Plasmas

For a plasma (in which $\omega_0 = 0$), the dispersion relation (7.65) reduces to

$$n_{\pm}^2(\omega) = 1 - \frac{\omega_p^2}{\omega(\omega \mp \Omega)}. \quad (7.83)$$

The upper sign corresponds to a left-hand circularly polarized wave, and the lower sign to a right-hand polarized wave. Of course, Equation (7.83) is only valid for wave propagation parallel to the direction of the magnetic field. Wave propagation through the Earth's ionosphere is well described by the previous dispersion relation. There are wide frequency intervals where one of n_+^2 or n_-^2 is positive, and the other negative. At such frequencies, one state of circular polarization cannot propagate through the plasma. Consequently, a wave of that polarization incident on the plasma is totally reflected. (See Chapter 8). The other state of polarization is partially transmitted.

The behavior of $n_-^2(\omega)$ at low frequencies is responsible for a strange phenomenon known to radio hams as “whistlers.” As the wave frequency tends to zero, Equation (7.83) yields

$$n_-^2 \simeq \frac{\omega_p^2}{\omega \Omega}. \quad (7.84)$$

At such a frequency, n_+^2 is negative, so only right-hand polarized waves can propagate. The wavenumber of these waves is given by

$$k_- = n_- \frac{\omega}{c} \simeq \frac{\omega_p}{c} \sqrt{\frac{\omega}{\Omega}}. \quad (7.85)$$

Now, energy propagates through a dispersive medium at the group velocity (see Section 7.13)

$$v_g(\omega) = \frac{d\omega}{dk_-} \simeq 2c \frac{\sqrt{\omega \Omega}}{\omega_p}. \quad (7.86)$$

Thus, low frequency waves transmit energy at a slower rate than high frequency waves. A lightning strike in one hemisphere of the Earth generates a wide spectrum of radiation, some of which propagates along the dipolar field-lines of the Earth’s magnetic field in a manner described approximately by the dispersion relation (7.84). The high frequency components of the signal return to the surface of the Earth before the low frequency components (because they travel faster along the magnetic field). This gives rise to a radio signal that begins at a high frequency, and then “whistles” down to lower frequencies.

7.9 Wave Propagation in Dispersive Media

Let us investigate the propagation of electromagnetic radiation through a general dispersive medium by studying a simple one-dimensional problem. Suppose that our dispersive medium extends from $x = 0$, where it interfaces with a vacuum, to $x = \infty$. Suppose, further, that an electromagnetic wave is incident normally on the interface such that the field quantities at the interface only depend on x and t . The wave is then specified as a given function of t at $x = 0$. Because we are not interested in the reflected wave, let this function, $f(t)$, say, specify the wave amplitude just inside the surface of the dispersive medium. Suppose that the wave arrives at this surface at $t = 0$, and that

$$f(t) = \begin{cases} 0 & \text{for } t < 0, \\ \sin(2\pi t/\tau) & \text{for } t \geq 0. \end{cases} \quad (7.87)$$

How does the wave subsequently develop in the region $x > 0$? In order to answer this question, we must first of all decompose $f(t)$ into harmonic components of the form $\exp(-i\omega t)$ (i.e., Fourier harmonics). Unfortunately, if we attempt this using only real frequencies, ω , then we encounter convergence difficulties, because $f(t)$ does not vanish at $t = \infty$. For the moment, we can circumvent these difficulties by only considering finite (in time) wave-forms. In other words, we now imagine that $f(t) = 0$ for $t < 0$ and $t > T$. Such a wave-form can be thought of as the superposition of two

infinite (in time) wave-forms, the first beginning at $t = 0$, and the second at $t = T$ with the opposite phase, so that the two cancel for all time $t > T$.

According to standard Fourier transform theory,

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega \int_{-\infty}^{\infty} f(t') e^{-i\omega(t-t')} dt'. \quad (7.88)$$

Because $f(t)$ is a real function of t that is zero for $t < 0$ and $t > T$, we can write

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega \int_0^T f(t') \cos[\omega(t-t')] dt'. \quad (7.89)$$

Finally, it follows from symmetry (in ω) that

$$f(t) = \frac{1}{\pi} \int_0^{\infty} d\omega \int_0^T f(t') \cos[\omega(t-t')] dt'. \quad (7.90)$$

Equation (7.87) yields

$$f(t) = \frac{1}{\pi} \int_0^{\infty} d\omega \int_0^T \sin\left(\frac{2\pi t'}{\tau}\right) \cos[\omega(t-t')] dt', \quad (7.91)$$

or

$$f(t) = \frac{1}{2\pi} \int_0^{\infty} d\omega \left(\frac{\cos[2\pi t'/\tau + \omega(t-t')]}{\omega - 2\pi/\tau} - \frac{\cos[2\pi t'/\tau - \omega(t-t')]}{\omega + 2\pi/\tau} \right) \Big|_{t'=0}^{t'=T}. \quad (7.92)$$

Let us assume, for the sake of simplicity, that

$$T = N\tau, \quad (7.93)$$

where N is a positive integer. This ensures that $f(t)$ is continuous at $t = T$. Equation (7.92) reduces to

$$f(t) = \frac{2}{\tau} \int_0^{\infty} \frac{d\omega}{\omega^2 - (2\pi/\tau)^2} (\cos[\omega(t-T)] - \cos[\omega t]). \quad (7.94)$$

This expression can be written

$$f(t) = \frac{1}{\tau} \int_{-\infty}^{\infty} \frac{d\omega}{\omega^2 - (2\pi/\tau)^2} (\cos[\omega(t-T)] - \cos[\omega t]), \quad (7.95)$$

or

$$f(t) = \frac{1}{2\pi} \operatorname{Re} \int_{-\infty}^{\infty} \frac{d\omega}{\omega - 2\pi/\tau} [e^{-i\omega(t-T)} - e^{-i\omega t}]. \quad (7.96)$$

It is not entirely obvious that Equation (7.96) is equivalent to Equation (7.95). However, we can easily prove that this is the case by taking Equation (7.96), and then using the standard definition of a real part (i.e., half the sum of the quantity in question and its complex conjugate) to give

$$f(t) = \frac{1}{4\pi} \int_{-\infty}^{\infty} \frac{d\omega}{\omega - 2\pi/\tau} [e^{-i\omega(t-T)} - e^{-i\omega t}] + \frac{1}{4\pi} \int_{-\infty}^{\infty} \frac{d\omega}{\omega - 2\pi/\tau} [e^{+i\omega(t-T)} - e^{+i\omega t}]. \quad (7.97)$$

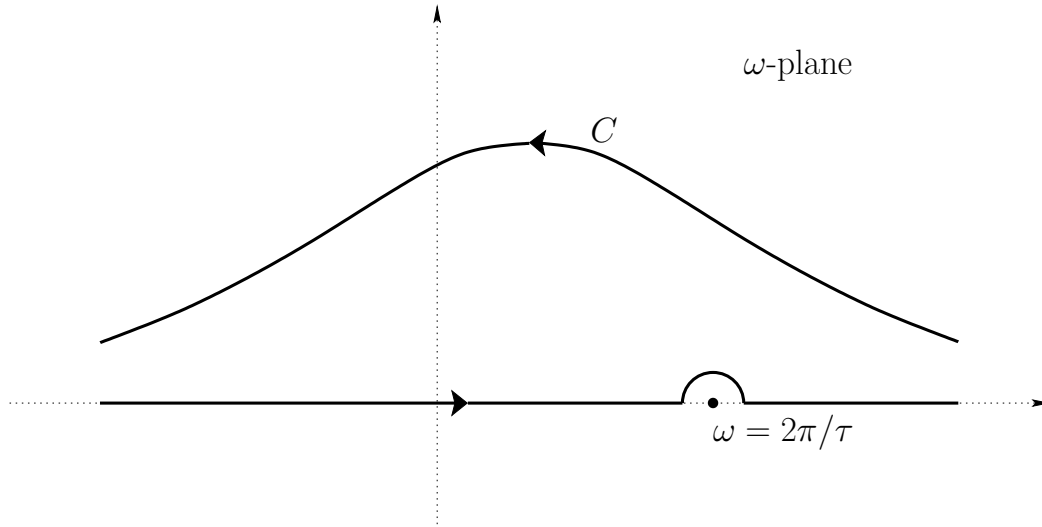


Figure 7.2: Sketch of the integration contours used to evaluate Equations (7.96) and (7.99).

Replacing the dummy integration variable ω by $-\omega$ in the second integral, and then making use of symmetry, it is easily seen that the previous expression reduces to Equation (7.95).

Equation (7.95) can be written

$$f(t) = \frac{2}{\tau} \int_{-\infty}^{\infty} d\omega \sin[\omega(t - T/2)] \frac{\sin(\omega T/2)}{\omega^2 - (2\pi/\tau)^2}. \quad (7.98)$$

Note that the integrand is finite at $\omega = 2\pi/\tau$, because, at this point, the vanishing of the denominator is compensated for by the simultaneous vanishing of the numerator. It follows that the integrand in Equation (7.96) is also not infinite at $\omega = 2\pi/\tau$, as long as we do not separate the two exponentials. Thus, we can replace the integration along the real axis through this point by a small semi-circle in the upper half of the complex plane. Once this has been done, we can deform the path still further, and can integrate the two exponentials in Equation (7.96) separately: that is,

$$f(t) = \frac{1}{2\pi} \operatorname{Re} \int_C e^{-i\omega t} \frac{d\omega}{\omega - 2\pi/\tau} - \frac{1}{2\pi} \operatorname{Re} \int_C e^{-i\omega(t-T)} \frac{d\omega}{\omega - 2\pi/\tau} \quad (7.99)$$

The contour C is sketched in Figure 7.2. Note that it runs from $+\infty$ to $-\infty$, which accounts for the change of sign between Equations (7.96) and (7.99).

We have already mentioned that a finite wave-form that is zero for $t < 0$ and $t > T$ can be thought of as the superposition of two out of phase infinite wave-forms, one starting at $t = 0$, and the other at $t = T$. It is plausible, therefore, that the first term in the previous expression corresponds to the infinite wave form starting at $t = 0$, and the second to the infinite wave form starting at $t = T$. If this is the case then the wave-form (7.87), which starts at $t = 0$ and ends at $t = \infty$, can be written

$$f(t) = \frac{1}{2\pi} \operatorname{Re} \int_C e^{-i\omega t} \frac{d\omega}{\omega - 2\pi/\tau}. \quad (7.100)$$

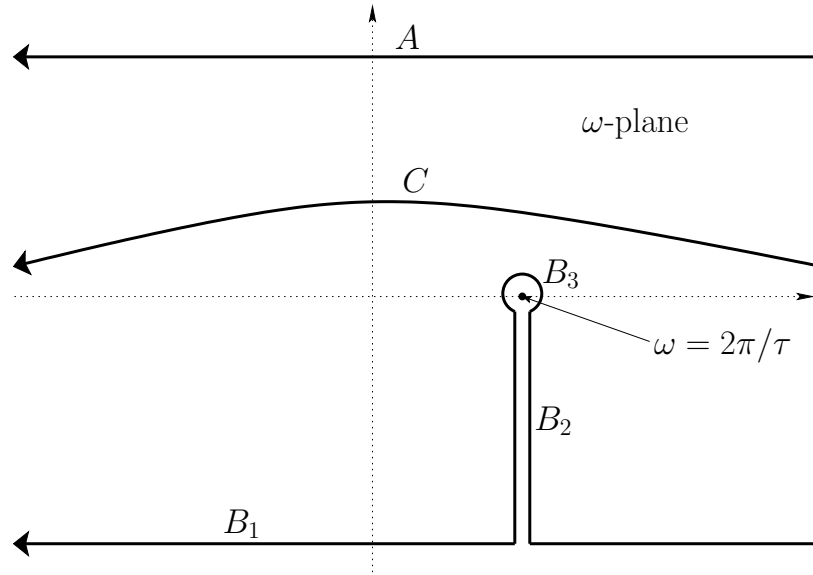


Figure 7.3: Sketch of the integration contours used to evaluate Equation (7.99).

Let us test this proposition. In order to do this, we must replace the original path of integration, C , by two equivalent paths.

First, consider $t < 0$. In this case, $-i\omega t$ has a negative real part in the upper half-plane that increases indefinitely with increasing distance from the axis. Thus, we can replace the original path of integration by the path A . (See Figure 7.3.) If we let A approach infinity in the upper half-plane then the integral clearly vanishes along this path. Consequently,

$$f(t) = 0 \quad (7.101)$$

for $t < 0$.

Next, consider $t > 0$. Now $-i\omega t$ has a negative real part in the lower half-plane, so the exponential vanishes at infinity in this half-plane. If we attempt to deform C to infinity in the lower half-plane then the path of integration “catches” on the singularity of the integrand at $\omega = 2\pi/\tau$. (See Figure 7.3.) The path of integration B therefore consists of three parts: 1) the part at infinity, B_1 , where the integral vanishes due to the exponential factor $e^{-i\omega t}$; 2) B_2 , the two parts leading to infinity, which cancel one another, and, thus, contribute nothing to the integral; 3) the path B_3 around the singularity. This latter contribution can easily be evaluated using the Cauchy residue theorem:

$$B_3 = \frac{1}{2\pi} \operatorname{Re} (2\pi i e^{-2\pi i t/\tau}) = \sin\left(\frac{2\pi t}{\tau}\right). \quad (7.102)$$

Thus, we have proved that expression (7.100) actually describes a wave-form, beginning at $t = 0$, whose subsequent motion is specified by Equation (7.87).

Equation (7.100) can immediately be generalized to give the wave motion in the region $x > 0$: that is,

$$f(x, t) = \frac{1}{2\pi} \operatorname{Re} \int_C e^{i(kx - \omega t)} \frac{d\omega}{\omega - 2\pi/\tau}. \quad (7.103)$$

This follows from standard wave theory, because we know that an untruncated wave motion at $x = 0$ of the form $e^{-i\omega t}$ takes the form $e^{i(kx - \omega t)}$ after moving a distance x into the dispersive medium, provided that k and ω are related by the appropriate dispersion relation. For a dielectric medium consisting of a single resonant species, this dispersion relation is written [see Equation (7.18)]

$$k^2 = \frac{\omega^2}{c^2} \left(1 + \frac{N e^2 / \epsilon_0 m}{\omega_0^2 - \omega^2 - i g \omega \omega_0} \right). \quad (7.104)$$

7.10 Wave-Front Propagation

It is helpful to define

$$s = t - \frac{x}{c}. \quad (7.105)$$

Let us consider the two cases $s < 0$ and $s > 0$ separately.

Suppose that $s < 0$. In this case, we distort the path C , used to evaluate the integral (7.103), into the path A shown in Figure 7.4. This is only a sensible thing to do if the real part of $i(kx - \omega t)$ is negative at infinity in the upper half-plane. Now, it is clear from the dispersion relation (7.104) that $k = \omega/c$ in the limit $|\omega| \rightarrow \infty$. Thus,

$$i(kx - \omega t) = -i\omega(t - x/c) = -i\omega s. \quad (7.106)$$

It follows that $i(kx - \omega t)$ possesses a large negative real part along path A provided that $s < 0$. Thus, Equation (7.103) yields

$$f(x, t) = 0 \quad (7.107)$$

for $s < 0$. In other words, it is impossible for the wave-front to propagate through the dispersive medium with a velocity greater than the velocity of light in a vacuum.

Suppose that $s > 0$. In this case, we distort the path C into the lower half-plane, because $i(kx - \omega t) = -i\omega s$ has a negative real part at infinity in this region. In doing this, the path becomes stuck not only at the singularity of the denominator at $\omega = 2\pi/\tau$, but also at the branch points of the expression for k . After a little algebra, the dispersion relation (7.104) yields

$$k = \frac{\omega}{c} \sqrt{\frac{\omega_{1+} - \omega}{\omega_{0+} - \omega}} \sqrt{\frac{\omega_{1-} - \omega}{\omega_{0-} - \omega}}, \quad (7.108)$$

where

$$\omega_{0\pm} = -i\rho \pm \sqrt{\omega_0^2 - \rho^2}, \quad (7.109)$$

and

$$\omega_{1\pm} = -i\rho \pm \sqrt{\omega_0^2 + \omega_p^2 - \rho^2}. \quad (7.110)$$

Here,

$$\omega_p = \sqrt{\frac{N e^2}{\epsilon_0 m}} \quad (7.111)$$

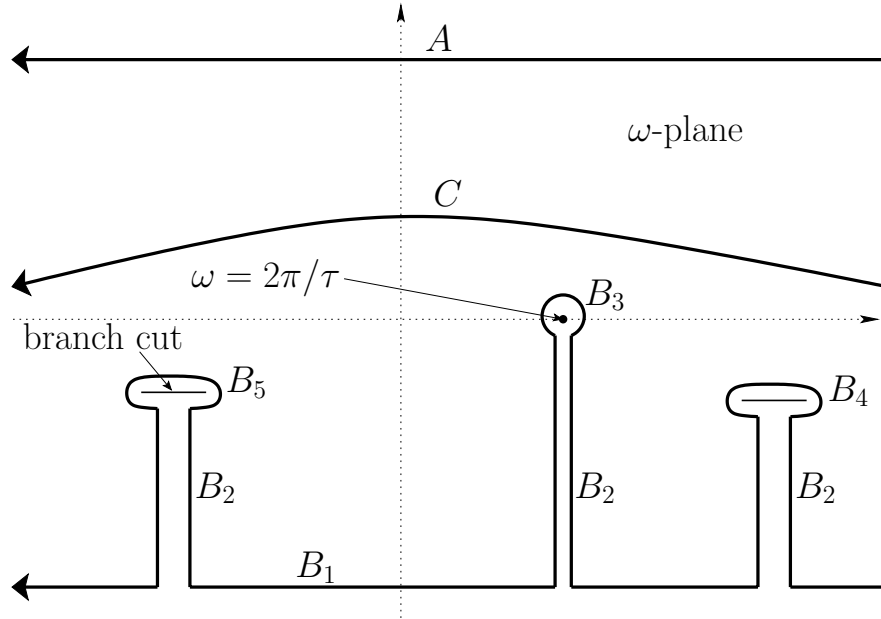


Figure 7.4: Sketch of the integration contours used to evaluate Equation (7.103).

is the plasma frequency, and

$$\rho = \frac{g\omega_0}{2} \ll \omega_0 \quad (7.112)$$

parameterizes the damping. In order to prevent multiple roots of Equation (7.108), it is necessary to place branch cuts between ω_{0+} and ω_{1+} , and also between ω_{0-} and ω_{1-} . (See Figure 7.4.)

The path of integration B is conveniently split into the parts B_1 through B_5 . (See Figure 7.4.) The contribution from B_1 is negligible, because the exponential in Equation (7.103) is vanishingly small on this part of the integration path. Likewise, the contribution from B_2 is zero, because its two sections always cancel one another. The contribution from B_3 follows from the residue theorem:

$$B_3 = \frac{1}{2\pi} \operatorname{Re} \left(2\pi i e^{i[k_\tau x - 2\pi t/\tau]} \right). \quad (7.113)$$

Here, k_τ denotes the value of k obtained from the dispersion relation (7.104) in the limit $\omega \rightarrow 2\pi/\tau$. Thus,

$$B_3 = e^{-\operatorname{Im}(k_\tau)x} \sin \left[2\pi \frac{t}{\tau} - \operatorname{Re}(k_\tau)x \right]. \quad (7.114)$$

In general, the contributions from B_4 and B_5 cannot be simplified further. For the moment, we denote them as

$$B_4 = \frac{1}{2\pi} \operatorname{Re} \oint_{B_4} e^{i(kx - \omega t)} \frac{d\omega}{\omega - 2\pi/\tau}, \quad (7.115)$$

and

$$B_5 = \frac{1}{2\pi} \operatorname{Re} \oint_{B_5} e^{i(kx - \omega t)} \frac{d\omega}{\omega - 2\pi/\tau}, \quad (7.116)$$

where the paths of integration circle the appropriate branch cuts. Altogether, we have

$$f(x, t) = e^{-\text{Im}(k_\tau)x} \sin\left[2\pi \frac{t}{\tau} - \text{Re}(k_\tau)x\right] + B_4 + B_5 \quad (7.117)$$

for $s > 0$.

Let us now look at the special case $s = 0$. For this value of s , we can change the original path of integration to one at infinity in either the upper or the lower half plane, because the integrand vanishes in each case, through no longer exponentially, but rather as $1/\omega^2$. We can see this from Equation (7.100), which can be written in the form

$$f(t) = \frac{1}{4\pi} \left(\int_C e^{-i\omega t} \frac{d\omega}{\omega - 2\pi/\tau} + \int_C e^{+i\omega t} \frac{d\omega}{\omega - 2\pi/\tau} \right). \quad (7.118)$$

Substitution of ω for $-\omega$ in the second integral yields

$$f(t) = \frac{1}{\tau} \int_C e^{-i\omega t} \frac{d\omega}{\omega^2 - (2\pi/\tau)^2}. \quad (7.119)$$

Now, applying dispersion theory, we obtain from the previous equation, just as we obtained Equation (7.103) from Equation (7.100),

$$f(x, t) = \frac{1}{\tau} \int_C e^{i(kx - \omega t)} \frac{d\omega}{\omega^2 - (2\pi/\tau)^2}. \quad (7.120)$$

Clearly, the integrand vanishes as $e^{-i\omega s}/\omega^2$ in the limit that $|\omega|$ becomes very large. Thus, it vanishes as $1/\omega^2$ for $s = 0$. Because we can calculate $f(x, t)$ using either path A or path B , we conclude that

$$f(x, t) = e^{-\text{Im}(k_\tau)x} \sin\left[2\pi \frac{t}{\tau} - \text{Re}(k_\tau)x\right] + B_4 + B_5 = 0 \quad (7.121)$$

for $s = 0$. Thus, there is continuity in the transition from the region $s < 0$ to the region $s > 0$.

We are now in a position to make some meaningful statements regarding the behavior of the signal at depth x within the dispersive medium. Prior to the time $t = x/c$, there is no wave motion. In other words, even if the phase velocity is superluminal, no electromagnetic signal can arrive earlier than one propagating at the velocity of light in vacuum, c . The wave motion for $t > x/c$ is conveniently divided into two parts: *free oscillations* and *forced oscillations*. The former are given by $B_4 + B_5$, and the latter by

$$e^{-\text{Im}(k_\tau)x} \sin\left[2\pi \frac{t}{\tau} - \text{Re}(k_\tau)x\right] = e^{-\text{Im}(k_\tau)x} \sin\left(\frac{2\pi}{\tau} \left[t - \frac{x}{v_p}\right]\right), \quad (7.122)$$

where

$$v_p = \frac{2\pi}{\tau \text{Re}(k_\tau)} \quad (7.123)$$

is termed the *phase velocity*. The forced oscillations have the same sine wave characteristics and oscillation frequency as the incident wave. However, the wave amplitude is diminished by the

damping coefficient, although, as we have seen, this is generally a negligible effect unless the frequency of the incident wave closely matches one of the resonant frequencies of the dispersive medium. The phase velocity v_p determines the velocity at which a point of constant phase (e.g., a peak or trough) of the forced oscillation signal propagates into the medium. However, the phase velocity has no effect on the velocity at which the forced oscillation wave-front propagates into the medium. This latter velocity is equivalent to the velocity of light in vacuum, c . The phase velocity v_p can be either greater or less than c , in which case peaks and troughs either catch up with or fall further behind the wave-front. Of course, peaks can never overtake the wave-front.

It is clear from Equations (7.109), (7.110), (7.115), and (7.116) that the free oscillations oscillate with real frequencies that lie somewhere between the resonant frequency, ω_0 , and the plasma frequency, ω_p . Furthermore, the free oscillations are damped in time like $\exp(-\rho t)$. The free oscillations, like the forced oscillations, begin at time $t = x/c$. At $t = x/c$, the free and forced oscillations exactly cancel one another [see Equation (7.121)]. As t increases, both the free and forced oscillations set in, but the former rapidly damp away, leaving only the forced oscillations. Thus, the free oscillations can be regarded as some sort of transient response of the medium to the incident wave, whereas the forced oscillations determine the time asymptotic response. The real frequency of the forced oscillations is that imposed externally by the incident wave, whereas the real frequency of the free oscillations is determined by the nature of the dispersive medium, quite independently of the frequency of the incident wave.

One slightly surprising result of the previous analysis is the prediction that the signal wave-front propagates into the dispersive medium at the velocity of light in vacuum, irrespective of the dispersive properties of the medium. Actually, this is a fairly obvious result. As is well described by Feynman in his famous Lectures on Physics, when an electromagnetic wave propagates through a dispersive medium, the electrons and ions that make up that medium oscillate in sympathy with the incident wave, and, in doing so, emit radiation. The radiation from the electrons and ions, as well as the incident radiation, travels at the velocity c . However, when these two radiation signals are superposed, the net effect is as if the incident signal propagates through the dispersive medium at a phase velocity that is different from c . Consider the wave-front of the incident signal, which clearly propagates into the medium with the velocity c . Prior to the arrival of this wave-front, the electrons and ions are at rest, because no information regarding the arrival of the incident wave at the surface of the medium can propagate faster than c . After the arrival of the wave-front, the electrons and ions are set into motion, and emit radiation which affects the apparent phase velocity of radiation that arrives somewhat later. But this radiation certainly cannot affect the propagation velocity of the wave-front itself, which has already passed by the time the electrons and ions are set into motion (because of their finite inertia).

7.11 Sommerfeld Precursor

Consider the situation immediately after the arrival of the signal: that is, when s is small and positive. Let us start from Equation (7.120), which can be written in the form

$$f(x, t) = \frac{1}{\tau} \int_C e^{i([k-\omega/c]x - \omega s)} \frac{d\omega}{\omega^2 - (2\pi/\tau)^2}. \quad (7.124)$$

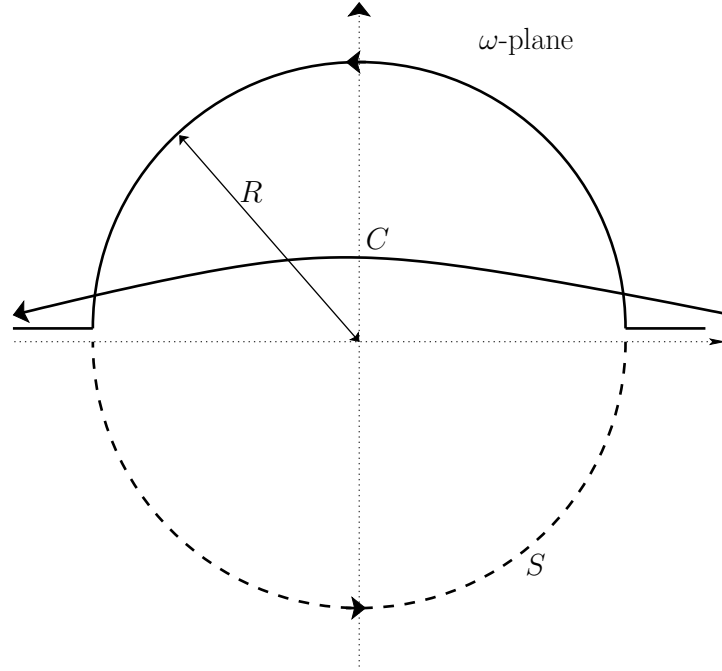


Figure 7.5: Sketch of the integration contour used to evaluate Equation (7.124).

We can deform the original path of integration C into a large semi-circle of radius R in the upper half-plane, plus two segments of the real axis, as shown in Figure 7.5. Because of the denominator $\omega^2 - (2\pi/\tau)^2$, the integrand tends to zero as $1/\omega^2$ on the real axis. We can add the path in the lower half-plane that is shown as a dotted line in the figure, because if the radius of the semi-circular portion of this lower path is increased to infinity then the integrand vanishes exponentially as $s > 0$. Therefore, we can replace our original path of integration by the entire circle S . Thus,

$$f(x, t) = \frac{1}{\tau} \oint_S e^{i([k-\omega/c]x-\omega s)} \frac{d\omega}{\omega^2 - (2\pi/\tau)^2} \quad (7.125)$$

in the limit that the radius of the circle S tends to infinity.

The dispersion relation (7.104) yields

$$k - \frac{\omega}{c} \simeq \frac{\omega}{c} \left(\sqrt{1 - \frac{\omega_p^2}{\omega^2}} - 1 \right) \simeq -\frac{\omega_p^2}{2c\omega} \quad (7.126)$$

in the limit $|\omega| \rightarrow \infty$. Using the abbreviation

$$\xi = \frac{\omega_p^2}{2c} x, \quad (7.127)$$

and, henceforth, neglecting $2\pi/\tau$ with respect to ω , we obtain

$$f(x, t) = f_1(\xi, t) \simeq \frac{1}{\tau} \oint_S \exp \left[i \left(-\frac{\xi}{\omega} - \omega s \right) \right] \frac{d\omega}{\omega^2} \quad (7.128)$$

from Equation (7.125). This expression can also be written

$$f_1(\xi, t) = \frac{1}{\tau} \oint_S \exp \left[-i \sqrt{\xi s} \left(\frac{1}{\omega} \sqrt{\frac{\xi}{s}} + \omega \sqrt{\frac{s}{\xi}} \right) \right] \frac{d\omega}{\omega^2}. \quad (7.129)$$

Let

$$\omega \sqrt{\frac{s}{\xi}} = e^{iu}. \quad (7.130)$$

It follows that

$$\frac{d\omega}{\omega} = i du, \quad (7.131)$$

giving

$$\frac{d\omega}{\omega^2} = i \sqrt{\frac{s}{\xi}} e^{-iu} du. \quad (7.132)$$

Substituting the angular variable u for ω in Equation (7.129), we obtain

$$f_1(\xi, t) = \frac{i}{\tau} \sqrt{\frac{s}{\xi}} \int_0^{2\pi} \exp(-2i \sqrt{\xi s} \cos u) e^{-iu} du. \quad (7.133)$$

Here, we have taken $\sqrt{\xi/s}$ as the radius of the circular integration path in the ω -plane. This is indeed a large radius because $s \ll 1$. From symmetry, Equation (7.133) simplifies to give

$$f_1(\xi, t) = \frac{i}{\tau} \sqrt{\frac{s}{\xi}} \int_0^{2\pi} \exp(-2i \sqrt{\xi s} \cos u) \cos u du. \quad (7.134)$$

The following mathematical identity is fairly well known,¹

$$J_n(z) = \frac{1}{2\pi i^n} \int_0^{2\pi} e^{iz \cos \theta} \cos(n\theta) d\theta, \quad (7.135)$$

where $J_n(z)$ is Bessel function of order n . It follows from Equation (7.134) that

$$f_1(\xi, t) = \frac{2\pi}{\tau} \sqrt{\frac{s}{\xi}} J_1(2\sqrt{\xi s}). \quad (7.136)$$

Here, we have made use of the fact that $J_1(-z) = -J_1(z)$.

The properties of Bessel functions are described in many standard references on mathematical functions (see, for instance, Abramowitz and Stegun). In the small argument limit, $z \ll 1$, we find that

$$J_1(z) = \frac{z}{2} + O(z^3). \quad (7.137)$$

¹M. Abramowitz, and I.A. Stegun, *Handbook of Mathematical Functions*, (Dover, New York, 1965). Equation 9.1.21.

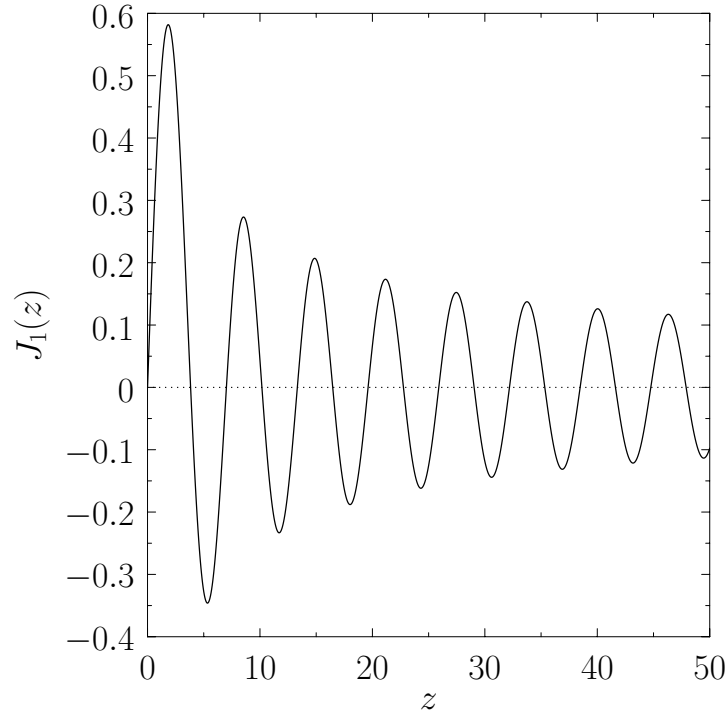


Figure 7.6: The Bessel function $J_1(z)$.

On the other hand, in the large argument limit, $z \gg 1$, we obtain

$$J_1(z) = \sqrt{\frac{2}{\pi z}} \cos(z - 3\pi/4) + \mathcal{O}(z^{-3/2}). \quad (7.138)$$

The behavior of $J_1(z)$ is further illustrated in Figure 7.6.

We are now in a position to make some quantitative statements regarding the signal that first arrives at a depth x within the dispersive medium. This signal propagates at the velocity of light in vacuum, and is called the *Sommerfeld precursor*. The first important point to note is that the amplitude of the Sommerfeld precursor is very small compared to that of the incident wave (whose amplitude is normalized to unity). We can easily see this because, in deriving Equation (7.136), we assumed that $|\omega| = \sqrt{\xi/s} \gg 2\pi/\tau$ on the circular integration path S . Because the magnitude of J_1 is always less than, or of order, unity, it is clear that $|f_1| \ll 1$. This is a comforting result, because in a naive treatment of wave propagation through a dielectric medium, the wave-front propagates at the group velocity v_g (which is less than c) and, therefore, no signal should reach a depth x within the medium before time x/v_g . We are finding that there is, in fact, a precursor that arrives at $t = x/c$, but that this signal is fairly weak. Note from Equation (7.127) that ξ is proportional to x . Consequently, the amplitude of the Sommerfeld precursor decreases as the inverse of the distance traveled by the wave-front through the dispersive medium [because $J_1(2\sqrt{\xi s})$ attains its maximum value when $s \sim 1/\xi$]. Thus, the Sommerfeld precursor is likely to become undetectable after the wave has traveled a long distance through the medium.

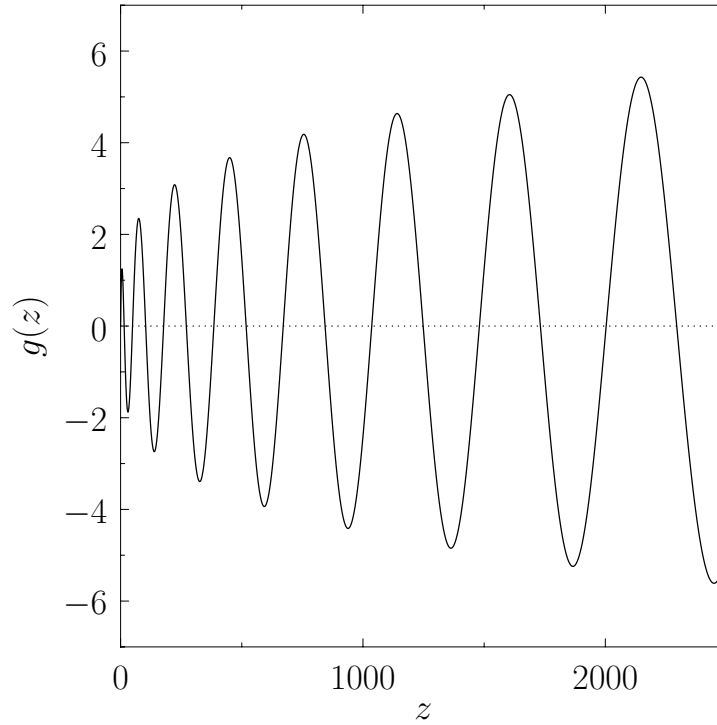


Figure 7.7: The Sommerfeld precursor.

Equation (7.136) can be written

$$f_1(\xi, t) = \frac{\pi}{\xi \tau} g(s/s_0), \quad (7.139)$$

where $s_0 = 1/(4\xi)$, and

$$g(z) = \sqrt{z} J_1(\sqrt{z}). \quad (7.140)$$

The normalized Sommerfeld precursor $g(z)$ is shown in Figure 7.7. It can be seen that both the amplitude and the oscillation period of the precursor gradually increase. The roots of $J_1(z)$ [i.e., the solutions of $J_1(z) = 0$] are spaced at distances of approximately π apart. Thus, the time interval for the m th half period of the precursor is approximately given by

$$\Delta t_m \sim \frac{m \pi^2}{2 \xi}. \quad (7.141)$$

Note that the initial period of oscillation,

$$\Delta t_0 \sim \frac{\pi^2}{2 \xi}, \quad (7.142)$$

is extremely small compared to the incident period τ . Moreover, the initial period of oscillation is completely independent of the frequency of the incident wave. In fact, Δt_0 depends only on the propagation distance x , and the dispersive power of the medium. The period also decreases with

increasing distance, x , traveled by the wave-front through the medium. So, when visible radiation is incident on a dispersive medium, it is quite possible for the first signal detected well inside the medium to lie in the X-ray region of the electromagnetic spectrum.

7.12 Method of Stationary Phase

Equation (7.120) can be written in the form

$$f(x, t) = \int_C e^{i\phi(\omega)} F(\omega) d\omega \quad (7.143)$$

where

$$F(\omega) = \frac{1}{\tau \omega^2 - (2\pi/\tau)^2}, \quad (7.144)$$

and

$$\phi(\omega) = k(\omega) x - \omega t. \quad (7.145)$$

Now, $F(\omega)$ is a relatively slowly varying function of ω (except in the immediate vicinity of the singular points, $\omega = \pm 2\pi/\tau$), whereas the phase $\phi(\omega)$ is generally large and rapidly varying. The rapid oscillations of $\exp(i\phi)$ over most of the range of integration means that the integrand averages to almost zero. Exceptions to this cancellation rule occur only at points where $\phi(\omega)$ is stationary: that is, where $\phi(\omega)$ has an extremum. The integral can therefore be estimated by finding all the points in the ω -plane where $\phi(\omega)$ has a vanishing derivative, evaluating (approximately) the integral in the neighborhood of each of these points, and summing the contributions. This procedure is known as the *method of stationary phase*.

Suppose that $\phi(\omega)$ has a vanishing first derivative at $\omega = \omega_s$. In the neighborhood of this point, $\phi(\omega)$ can be expanded as a Taylor series,

$$\phi(\omega) = \phi_s + \frac{1}{2} \phi_s'' (\omega - \omega_s)^2 + \dots \quad (7.146)$$

Here, the subscript s is used to indicate ϕ , or its second derivative, evaluated at $\omega = \omega_s$, whereas $'$ denotes a derivative with respect to ω . Because $F(\omega)$ is slowly varying, the contribution to the integral from this stationary phase point is approximately

$$f_s \simeq F(\omega_s) e^{i\phi_s} \int_{-\infty}^{\infty} \exp\left[\frac{i}{2} \phi_s'' (\omega - \omega_s)^2\right] d\omega. \quad (7.147)$$

It is tacitly assumed that the stationary point lies on the real axis in ω -space, so that locally the integral along the contour C is an integral along the real axis in the direction of decreasing ω . The previous expression can be written in the form

$$f_s \simeq -F(\omega_s) e^{i\phi_s} \sqrt{\frac{4\pi}{\phi_s''}} \int_0^{\infty} [\cos(\pi t^2/2) + i \sin(\pi t^2/2)] dt, \quad (7.148)$$

where

$$\frac{\pi}{2} t^2 = \frac{1}{2} \phi_s'' (\omega - \omega_s)^2. \quad (7.149)$$

The integrals in the previous expression are known as *Fresnel integrals*,² and can be shown to take the values

$$\int_0^\infty \cos(\pi t^2/2) dt = \int_0^\infty \sin(\pi t^2/2) dt = \frac{1}{2}. \quad (7.150)$$

It follows that

$$f_s \simeq -\sqrt{\frac{2\pi i}{\phi_s''}} F(\omega_s) e^{i\phi_s}. \quad (7.151)$$

It is easily demonstrated that the arc-length (in the ω -plane) of the section of the integration contour that makes a significant contribution to f_s is of order $\Delta\omega/\omega_s \sim 1/\sqrt{k(\omega_s)} x$. Thus, the arc-length is relatively short, provided that the wavelength of the signal is much less than the distance it has propagated into the dispersive medium. If there is more than one point of stationary phase in the range of integration then the integral is approximated as a sum of terms having the same form as the previous one.

Integrals of the form (7.143) can be calculated exactly using the *method of steepest descent*.³ The stationary phase approximation (7.151) agrees with the leading term of the method of steepest descent (which is far more difficult to implement than the method of stationary phase) provided that $\phi(\omega)$ is real (i.e., provided that the stationary point lies on the real axis). If ϕ is complex, however, then the stationary phase method can yield erroneous results. This suggests that the stationary phase method is likely to break down when the extremum point $\omega = \omega_s$ approaches any poles or branch cuts in the ω -plane.

7.13 Group Velocity

The point of stationary phase, defined by $\partial\phi/\partial\omega = 0$, satisfies the condition

$$\frac{c}{v_g} = \frac{ct}{x}, \quad (7.152)$$

where

$$v_g = \frac{d\omega}{dk} \quad (7.153)$$

is conventionally termed the *group velocity*. Thus, the signal seen at position x and time t is dominated by the frequency range whose group velocity v_g is equal to x/t . In this respect, the signal incident at the surface of the medium ($x = 0$) at time $t = 0$ can be said to propagate through the medium at the group velocity $v_g(\omega)$.

The simple one-resonance dielectric dispersion relation (7.104) yields

$$\frac{c}{v_g} \simeq n(\omega) \left(1 + \frac{\omega^2}{\omega_0^2 - \omega^2} + \frac{\omega^2}{\omega^2 - \omega_0^2 - \omega_p^2} \right) \quad (7.154)$$

²M. Abramowitz, and I.A. Stegun, *Handbook of Mathematical Functions*, (Dover, New York, 1965). Section 7.3.

³Léon Brillouin, *Wave Propagation and Group Velocity*, (Academic Press, New York, 1960).

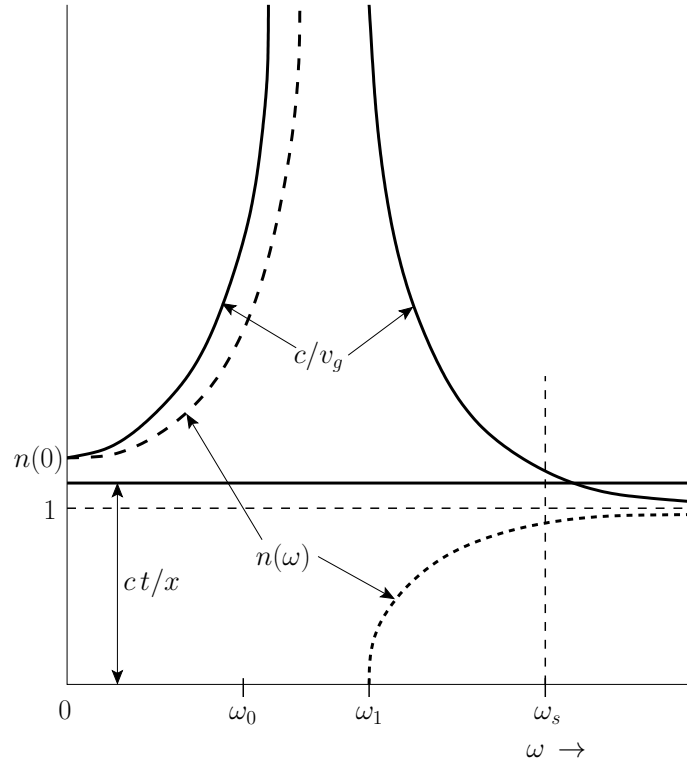


Figure 7.8: The typical variation of the functions $c/v_g(\omega)$ and $n(\omega)$. Here, $\omega_1 = (\omega_0^2 + \omega_p^2)^{1/2}$.

in the limit $g \rightarrow 0$, where

$$n(\omega) = \frac{ck}{\omega} = \sqrt{\frac{\omega_0^2 + \omega_p^2 - \omega^2}{\omega_0^2 - \omega^2}}. \quad (7.155)$$

The variation of c/v_g , and the refractive index n , with frequency is sketched in Figure 7.8. With $g = 0$, the group velocity is less than c for all ω , except for $\omega_0 < \omega < \omega_1 \equiv (\omega_0^2 + \omega_p^2)^{1/2}$, where it is purely imaginary. Note that the refractive index is also complex in this frequency range. The phase velocity $v_p = c/n$ is subluminal for $\omega < \omega_0$, imaginary for $\omega_0 \leq \omega \leq \omega_1$, and superluminal for $\omega > \omega_1$.

The frequency range that contributes to the amplitude at time t is determined graphically by finding the intersection of a horizontal line with ordinate ct/x with the solid curve in Figure 7.8. There is no crossing of the two curves for $t < t_0 \equiv x/c$. Thus, no signal can arrive before this time. For times immediately after $t = t_0$, the point of stationary phase is seen to be at $\omega \rightarrow \infty$. In this large- ω limit, the point of stationary phase satisfies

$$\omega_s \simeq \omega_p \sqrt{\frac{t_0}{2(t - t_0)}}. \quad (7.156)$$

Note that $\omega = -\omega_s$ is also a point of stationary phase. It is easily demonstrated that

$$\phi_s \simeq -2\sqrt{\xi(t - t_0)}, \quad (7.157)$$

and

$$\phi_s'' \simeq -2 \frac{(t - t_0)^{3/2}}{\xi^{1/2}}, \quad (7.158)$$

with

$$F(\omega_s) \simeq \frac{t - t_0}{\tau \xi}. \quad (7.159)$$

Here, ξ is given by Equation (7.127). The stationary phase approximation (7.151) yields

$$f_s \simeq \sqrt{\frac{\pi \xi^{1/2}}{(t - t_0)^{3/2}}} \frac{t - t_0}{\tau \xi} e^{-2i \sqrt{\xi(t-t_0)} + 3\pi i/4} + \text{c.c.}, \quad (7.160)$$

where c.c. denotes the complex conjugate of the preceding term (this contribution comes from the second point of stationary phase located at $\omega = -\omega_s$). The previous expression reduces to

$$f_s \simeq \frac{2\sqrt{\pi}}{\tau} \frac{(t - t_0)^{1/4}}{\xi^{3/4}} \cos\left[2\sqrt{\xi(t - t_0)} - 3\pi/4\right]. \quad (7.161)$$

It is readily shown that the previous formula is the same as expression (7.136) for the Sommerfeld precursor in the large argument limit $t - t_0 \gg 1/\xi$. Thus, the method of stationary phase yields an expression for the Sommerfeld precursor that is accurate at all times except those immediately following the first arrival of the signal.

7.14 Brillouin Precursor

As time progresses, the horizontal line ct/x in Figure 7.8 gradually rises, and the point of stationary phase moves to ever lower frequencies. In general, however, the amplitude remains relatively small. Only when the elapsed time reaches

$$t_1 = \frac{n(0)x}{c} > t_0 \quad (7.162)$$

is there a qualitative change. This time marks the arrival of a second precursor known as the *Brillouin precursor*. The reason for the qualitative change is evident from Figure 7.8. At $t = t_1$, the lower region of the c/v_g curve is intersected for the first time, and $\omega = 0$ becomes a point of stationary phase. It follows that the oscillation frequency of the Brillouin precursor is far less than that of the Sommerfeld precursor. Moreover, it is easily demonstrated that the second derivative of $k(\omega)$ vanishes at $\omega = 0$. This means that $\phi_s'' = 0$. The stationary phase result (7.151) gives an infinite answer in such circumstances. Of course, the amplitude of the Brillouin precursor is not infinite, but it is significantly larger than that of the Sommerfeld precursor.

In order to generalize the result (7.151) to deal with a stationary phase point at $\omega = 0$, it is necessary to expand $\phi(\omega)$ about this point, keeping terms up to ω^3 . Thus,

$$\phi(\omega) \simeq \omega(t_1 - t) + \frac{x}{6} k_0''' \omega^3, \quad (7.163)$$

where

$$k_0''' \equiv \left(\frac{d^3 k}{d\omega^3} \right)_{\omega=0} = \frac{3 \omega_p^2}{c n(0) \omega_0^4} \quad (7.164)$$

for the simple dispersion relation (7.104). The amplitude (7.143) is therefore given approximately by

$$f(x, t) \simeq F(0) \int_{-\infty}^{\infty} \exp \left[i \omega (t_1 - t) + i \frac{x}{6} k_0''' \omega^3 \right] d\omega. \quad (7.165)$$

This expression reduces to

$$f(x, t) = \frac{\tau}{\sqrt{2} \pi^2} \sqrt{\frac{|t - t_1|}{x k_0'''}} \int_0^{\infty} \cos \left[\frac{3}{2} z \left(\frac{v^3}{3} \pm v \right) \right] dv, \quad (7.166)$$

where

$$v = \sqrt{\frac{x k_0'''}{2 |t - t_1|}} \omega, \quad (7.167)$$

and

$$z = \frac{2 \sqrt{2} |t - t_1|^{3/2}}{3 \sqrt{x k_0'''}}. \quad (7.168)$$

The positive (negative) sign in the integrand is taken for $t < t_1$ ($t > t_1$).

The integral in Equation (7.166) is known as an *Airy integral*. It can be expressed in terms of Bessel functions of order 1/3, as follows:

$$\int_0^{\infty} \cos \left[\frac{3}{2} z \left(\frac{v^3}{3} + v \right) \right] dv = \frac{1}{\sqrt{3}} K_{1/3}(z), \quad (7.169)$$

and

$$\int_0^{\infty} \cos \left[\frac{3}{2} z \left(\frac{v^3}{3} - v \right) \right] dv = \frac{\pi}{3} [J_{1/3}(z) + J_{-1/3}(z)]. \quad (7.170)$$

From the well-known properties of Bessel functions, the precursor can be seen to have a growing exponential character for times earlier than $t = t_1$, and an oscillating character for $t > t_1$. The amplitude in the neighborhood of $t = t_1$ is plotted in Figure 7.9.

The initial oscillation period of the Brillouin precursor is crudely estimated (by setting $z \sim 1$) as

$$\Delta t_0 \sim (x k_0''')^{1/3}. \quad (7.171)$$

The amplitude of the Brillouin precursor is approximately

$$|f| \sim \frac{\tau}{(x k_0''')^{1/3}}. \quad (7.172)$$

Let us adopt the ordering

$$1/\tau \sim \omega_0 \sim \omega_p \ll \xi, \quad (7.173)$$

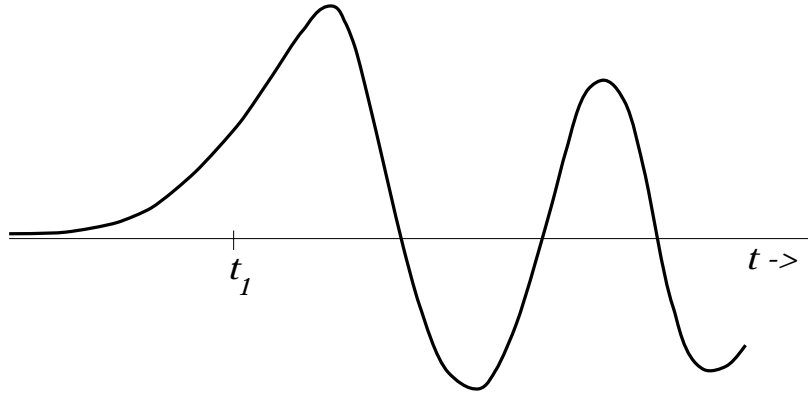


Figure 7.9: A sketch of the behavior of the Brillouin precursor as a function of time.

which corresponds to the majority of physical situations involving the propagation of electromagnetic radiation through dielectric media. It follows, from the previous results, plus the results of Section 7.11, that

$$(\Delta t_0 \omega_p)_{\text{brillouin}} \sim \left(\frac{\xi}{\omega_p} \right)^{1/3} \gg 1, \quad (7.174)$$

and

$$(\Delta t_0 \omega_p)_{\text{sommerfeld}} \sim \left(\frac{\omega_p}{\xi} \right) \ll 1. \quad (7.175)$$

Furthermore,

$$|f|_{\text{brillouin}} \sim \left(\frac{\omega_p}{\xi} \right)^{1/3} \ll 1, \quad (7.176)$$

and

$$|f|_{\text{sommerfeld}} \sim \left(\frac{\omega_p}{\xi} \right) \ll |f|_{\text{brillouin}}. \quad (7.177)$$

Thus, it is clear that the Sommerfeld precursor is essentially a low amplitude, high frequency signal, whereas the Brillouin precursor is a high amplitude, low frequency signal. Note that the amplitude of the Brillouin precursor, despite being significantly higher than that of the Sommerfeld precursor, is still much less than that of the incident wave.

7.15 Signal Arrival

Let us now try to establish at which time, t_2 , a signal first arrives at depth x inside the dielectric medium whose amplitude is comparable with that of the wave incident at time $t = 0$ on the surface of the medium ($x = 0$). Let us term this event the “arrival” of the signal. It is plausible, from the discussion in Section 7.12 regarding the stationary phase approximation, that signal arrival corresponds to the situation at which the point of stationary phase in ω -space corresponds to a pole of the function $F(\omega)$. In other words, when ω_s approaches the frequency $2\pi/\tau$ of the incident signal. It is certainly the case that the stationary phase approximation yields a particularly large

amplitude signal when $\omega_s \rightarrow 2\pi/\tau$. Unfortunately, as has already been discussed, the method of stationary phase becomes inaccurate under these circumstances. However, calculations involving the more robust method of steepest decent⁴ confirm that, in most cases, the signal amplitude first becomes significant when $\omega_s = 2\pi/\tau$. Thus, the signal arrival time is

$$t_2 = \frac{x}{v_g(2\pi/\tau)}, \quad (7.178)$$

where $v_g(2\pi/\tau)$ is the group velocity calculated using the frequency of the incident signal. It is clear from Figure 7.8 that

$$t_0 < t_1 < t_2. \quad (7.179)$$

Thus, the main signal arrives later than the Sommerfeld and Brillouin precursors.

7.16 Exercises

7.1 A general electromagnetic wave pulse propagating in the z -direction at velocity u is written

$$\begin{aligned} \mathbf{E} &= P(z - ut) \mathbf{e}_x + Q(z - ut) \mathbf{e}_y + R(z - ut) \mathbf{e}_z, \\ \mathbf{B} &= \frac{S(z - ut)}{u} \mathbf{e}_x + \frac{T(z - ut)}{u} \mathbf{e}_y + \frac{U(z - ut)}{u} \mathbf{e}_z, \end{aligned}$$

where P , Q , R , S , T , and U are arbitrary functions. In order to exclude electrostatic and magnetostatic fields, these functions are subject to the constraint that $\langle P \rangle = \langle Q \rangle = \langle R \rangle = \langle S \rangle = \langle T \rangle = \langle U \rangle = 0$, where

$$\langle P \rangle = \int_{-\infty}^{\infty} P(x) dx.$$

Suppose that the pulse propagates through a uniform dielectric medium of dielectric constant ϵ . Demonstrate from Maxwell's equation that $u = c/\sqrt{\epsilon}$, $R = U = 0$, $S = -Q$, and $T = P$. Incidentally, this result implies that a general wave pulse is characterized by two arbitrary functions, corresponding to the two possible independent polarizations of the pulse.

7.2 Show that the mean energy flux due to an electromagnetic wave of angular frequency ω propagating through a good conductor of conductivity σ can be written

$$\langle I \rangle = \frac{E^2}{\sqrt{8} Z},$$

where E is the peak electric field-strength, and $Z = (\epsilon_0 \omega / \sigma)^{1/2}$.

7.3 Consider an electromagnetic wave propagating in the positive z -direction through a good conductor of conductivity σ . Suppose that the wave electric field is

$$E_x(z, t) = E_0 e^{-z/d} \cos(\omega t - z/d),$$

⁴Ibid.

where d is the skin-depth. Demonstrate that the mean electromagnetic energy flux across the plane $z = 0$ matches the mean rate at which electromagnetic energy is dissipated per unit area due to Joule heating in the region $z > 0$.

- 7.4 A plane electromagnetic wave, linearly polarized in the x -direction, and propagating in the z -direction through an electrical conducting medium of conductivity σ and relative dielectric constant unity, is governed by

$$\begin{aligned}\frac{\partial H_y}{\partial t} &= -\frac{1}{\mu_0} \frac{\partial E_x}{\partial z}, \\ \frac{\partial E_x}{\partial t} &= -\frac{\sigma}{\epsilon_0} E_x - \frac{1}{\epsilon_0} \frac{\partial H_y}{\partial z},\end{aligned}$$

where $E_x(z, t)$ and $H_y(z, t)$ are the electric and magnetic components of the wave. Derive an energy conservation equation of the form

$$\frac{\partial \mathcal{E}}{\partial t} + \frac{\partial \mathcal{I}}{\partial z} = -\sigma E_x^2,$$

where \mathcal{E} is the electromagnetic energy per unit volume, and \mathcal{I} the electromagnetic energy flux. Give expressions for \mathcal{E} and \mathcal{I} . What does the right-hand side of the previous equation represent? Demonstrate that E_x obeys the wave-diffusion equation

$$\frac{\partial^2 E_x}{\partial t^2} + \frac{\sigma}{\epsilon_0} \frac{\partial E_x}{\partial t} = c^2 \frac{\partial^2 E_x}{\partial z^2},$$

where $c = 1/\sqrt{\epsilon_0 \mu_0}$. Consider the high frequency, low conductivity, limit $\omega \gg \sigma/\epsilon_0$. Show that a wave propagating into the medium varies as

$$\begin{aligned}E_x(z, t) &\simeq E_0 \cos[k(c t - z)] e^{-z/\delta}, \\ H_y(z, t) &\simeq Z_0^{-1} E_0 \cos[k(c t - z) - 1/(k \delta)] e^{-z/\delta},\end{aligned}$$

where $k = \omega/c$, $\delta = 2/(Z_0 \sigma)$, and $Z_0 = \sqrt{\mu_0/\epsilon_0}$. Demonstrate that $k \delta \ll 1$: that is, the wave penetrates many wavelengths into the medium.

- 7.5 Consider a uniform plasma of plasma frequency ω_p containing a uniform magnetic field $B_0 \mathbf{e}_z$. Show that left-hand circularly polarized electromagnetic waves can only propagate parallel to the magnetic field provided that $\omega > -\Omega/2 + \sqrt{\Omega^2/4 + \omega_p^2}$, where $\Omega = e B_0/m_e$ is the electron cyclotron frequency. Demonstrate that right-hand circularly polarized electromagnetic waves can only propagate parallel to the magnetic field provided that their frequencies do not lie in the range $\Omega \leq \omega \leq \Omega/2 + \sqrt{\Omega^2/4 + \omega_p^2}$.

8 Wave Propagation in Inhomogeneous Dielectric Media

8.1 Introduction

In this chapter, we extend the analysis of the previous chapter to investigate electromagnetic wave propagation through inhomogeneous dielectric media.

8.2 Laws of Geometric Optics

Suppose that the region $z < 0$ is occupied by a transparent dielectric medium of uniform refractive index n_1 , whereas the region $z > 0$ is occupied by a second transparent dielectric medium of uniform refractive index n_2 . (See Figure 8.1.) Let a plane light wave be launched toward positive z from a light source of angular frequency ω located at large negative z . Furthermore, suppose that this wave, which has the wavevector \mathbf{k}_i , is obliquely incident on the interface between the two media. We would expect the incident plane wave to be partially reflected and partially refracted (i.e., transmitted) by the interface. Let the reflected and refracted plane waves have the wavevectors \mathbf{k}_r and \mathbf{k}_t , respectively. (See Figure 8.1.) Hence, we can write

$$\psi(x, y, z, t) = \psi_i \cos(\omega t - \mathbf{k}_i \cdot \mathbf{r}) + \psi_r \cos(\omega t - \mathbf{k}_r \cdot \mathbf{r}) \quad (8.1)$$

in the region $z < 0$, and

$$\psi(x, y, z, t) = \psi_t \cos(\omega t - \mathbf{k}_t \cdot \mathbf{r}) \quad (8.2)$$

in the region $z > 0$. Here, $\psi(x, y, z, t)$ represents the magnetic component of the resultant light wave, ψ_i the amplitude of the incident wave, ψ_r the amplitude of the reflected wave, and ψ_t the amplitude of the refracted wave. All of the component waves have the same angular frequency, ω , because this property is ultimately determined by the wave source. Furthermore, if the magnetic component of an electromagnetic wave is specified then the electric component of the wave is fully determined, and vice versa. (See Section 7.1.)

In general, the wavefunction, ψ , must be continuous at $z = 0$, because there cannot be a discontinuity in either the normal or the tangential component of a magnetic field across an interface between two (non-magnetic) dielectric media. (The same is not true of an electric field, which can have a normal discontinuity across an interface between two dielectric media.) This explains why we have chosen ψ to represent the magnetic, rather than the electric, component of the wave. Thus, the matching condition at $z = 0$ takes the form

$$\psi_i \cos(\omega t - k_{ix} x - k_{iy} y) + \psi_r \cos(\omega t - k_{rx} x - k_{ry} y) = \psi_t \cos(\omega t - k_{tx} x - k_{ty} y). \quad (8.3)$$

This condition must be satisfied at all values of x , y , and t . This is only possible if

$$k_{ix} = k_{rx} = k_{tx}, \quad (8.4)$$

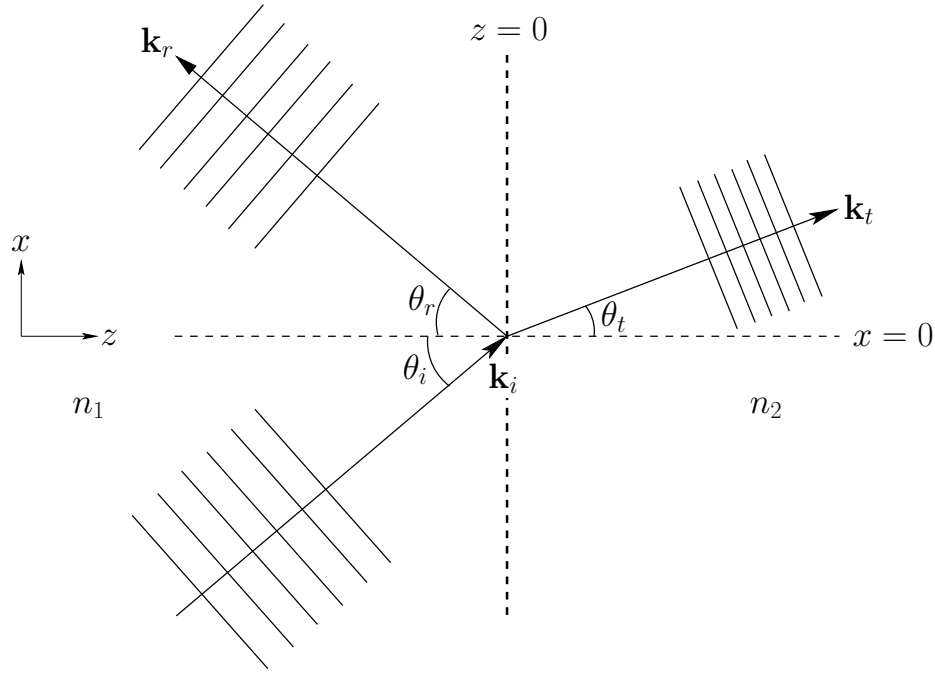


Figure 8.1: Reflection and refraction of a plane wave at a plane boundary.

and

$$k_{iy} = k_{ry} = k_{ty}. \quad (8.5)$$

Suppose that the direction of propagation of the incident wave lies in the x - z plane, so that $k_{iy} = 0$. It immediately follows, from Equation (8.5), that $k_{ry} = k_{ty} = 0$. In other words, the directions of propagation of the reflected and the refracted waves also lie in the x - z plane, which implies that \mathbf{k}_i , \mathbf{k}_r and \mathbf{k}_t are co-planar vectors. This constraint is implicit in the well-known laws of geometric optics.

Assuming that the previously mentioned constraint is satisfied, let the incident, reflected, and refracted wave normals subtend angles θ_i , θ_r , and θ_t with the z -axis, respectively. (See Figure 8.1.) It follows that

$$\mathbf{k}_i = n_1 k_0 (\sin \theta_i, 0, \cos \theta_i), \quad (8.6)$$

$$\mathbf{k}_r = n_1 k_0 (\sin \theta_r, 0, -\cos \theta_r), \quad (8.7)$$

$$\mathbf{k}_t = n_2 k_0 (\sin \theta_t, 0, \cos \theta_t), \quad (8.8)$$

where $k_0 = \omega/c$ is the vacuum wavenumber, and c the velocity of light in vacuum. Here, we have made use of the fact that wavenumber (i.e., the magnitude of the wavevector) of a light wave propagating through a dielectric medium of refractive index n is $n k_0$.

According to Equation (8.4), $k_{ix} = k_{rx}$, which yields

$$\sin \theta_i = \sin \theta_r, \quad (8.9)$$

and $k_{i,x} = k_{t,x}$, which reduces to

$$n_1 \sin \theta_i = n_2 \sin \theta_t. \quad (8.10)$$

The first of these relations states that the angle of incidence, θ_i , is equal to the angle of reflection, θ_r . This is the familiar *law of reflection*. Furthermore, the second relation corresponds to the equally familiar *law of refraction*, otherwise known as *Snell's law*.

Incidentally, the fact that a plane wave propagates through a uniform dielectric medium with a constant wavevector, and, therefore, a constant direction of motion, is equivalent to the well-known *law of rectilinear propagation*, which states that a light ray (i.e., the normal to a constant phase surface) propagates through a uniform medium in a straight-line.

It follows, from the previous discussion, that the laws of geometric optics (i.e., the law of rectilinear propagation, the law of reflection, and the law of refraction) are fully consistent with the wave properties of light, despite the fact that they do not seem to explicitly depend on these properties.

8.3 Fresnel Relations

The theory described in the previous section is sufficient to determine the directions of the reflected and refracted waves, when a light wave is obliquely incident on a plane interface between two dielectric media. However, it cannot determine the fractions of the incident energy that are reflected and refracted, respectively. In order to calculate the coefficients of reflection and transmission, it is necessary to take into account both the electric and the magnetic components of the various different waves. It turns out that there are two independent wave polarizations that behave slightly differently. The first of these is such that the magnetic components of the incident, reflected, and refracted waves are all parallel to the interface. The second is such that the electric components of these waves are all parallel to the interface.

Consider the first polarization. Let the interface correspond to the plane $z = 0$, let the region $z < 0$ be occupied by material of refractive index n_1 , and let the region $z > 0$ be occupied by material of refractive index n_2 . Suppose that the incident, reflected, and refracted waves are plane waves, of angular frequency ω , whose wavevectors lie in the x - z plane. (See Figure 8.1.) The equations governing the propagation of the wave are

$$\frac{\partial D_x}{\partial t} = -\frac{\partial H_y}{\partial z}, \quad (8.11)$$

$$\frac{\partial D_z}{\partial t} = \frac{\partial H_y}{\partial x}, \quad (8.12)$$

$$\frac{\partial H_y}{\partial t} = v^2 \left(\frac{\partial D_z}{\partial x} - \frac{\partial D_x}{\partial z} \right), \quad (8.13)$$

where

$$\mathbf{D} = \epsilon \epsilon_0 \mathbf{E} \quad (8.14)$$

is the electric displacement, $v = c/n$ the characteristic wave speed, and $n = \sqrt{\epsilon}$ the refractive index. Suppose that, as described in the previous section,

$$H_y(x, z, t) = \psi_i \cos(\omega t - n_1 k_0 \sin \theta_i x - n_1 k_0 \cos \theta_i z) \\ + \psi_r \cos(\omega t - n_1 k_0 \sin \theta_i x + n_1 k_0 \cos \theta_i z) \quad (8.15)$$

in the region $z < 0$, and

$$H_y(x, z, t) = \psi_t \cos(\omega t - n_1 k_0 \sin \theta_i x - n_2 k_0 \cos \theta_t z) \quad (8.16)$$

in the region $z > 0$. Here, $k_0 = \omega/c$ is the vacuum wavenumber, θ_i the angle of incidence, and θ_t the angle of refraction. (See Figure 8.1.) In writing the above expressions, we have made use of the law of reflection (i.e., $\theta_r = \theta_i$), as well as the law of refraction (i.e., $n_1 \sin \theta_i = n_2 \sin \theta_t$). The two terms on the right-hand side of Equation (8.15) correspond to the incident and reflected waves, respectively. The term on the right-hand side of Equation (8.16) corresponds to the refracted wave. Substitution of Equations (8.15)–(8.16) into the governing differential equations, (8.11)–(8.13), yields

$$D_x(x, z, t) = \frac{\psi_i \cos \theta_i}{v_1} \cos(\omega t - n_1 k_0 \sin \theta_i x - n_1 k_0 \cos \theta_i z) \\ - \frac{\psi_r \cos \theta_i}{v_1} \cos(\omega t - n_1 k_0 \sin \theta_i x + n_1 k_0 \cos \theta_i z), \quad (8.17)$$

$$D_z(x, z, t) = -\frac{\psi_i \sin \theta_i}{v_1} \cos(\omega t - n_1 k_0 \sin \theta_i x - n_1 k_0 \cos \theta_i z) \\ - \frac{\psi_r \sin \theta_i}{v_1} \cos(\omega t - n_1 k_0 \sin \theta_i x + n_1 k_0 \cos \theta_i z) \quad (8.18)$$

in the region $z < 0$, and

$$D_x(x, z, t) = \frac{\psi_t \cos \theta_t}{v_2} \cos(\omega t - n_1 k_0 \sin \theta_i x - n_2 k_0 \cos \theta_t z), \quad (8.19)$$

$$D_z(x, z, t) = -\frac{\psi_t \sin \theta_t}{v_1} \cos(\omega t - n_1 k_0 \sin \theta_i x - n_2 k_0 \cos \theta_t z) \quad (8.20)$$

in the region $z > 0$.

Now, both the normal and the tangential components of the magnetic intensity must be continuous at the interface. This implies that

$$[H_y]_{z=0_-}^{z=0_+} = 0, \quad (8.21)$$

which yields

$$\psi_i + \psi_r = \psi_t \quad (8.22)$$

Furthermore, the normal component of the electric displacement, as well as the tangential component of the electric field, must be continuous at the interface. In other words,

$$[D_z]_{z=0_-}^{z=0_+} = 0, \quad (8.23)$$

and

$$[E_x]_{z=0_+}^{z=0_-} = [D_x/(\epsilon \epsilon_0)]_{z=0_-}^{z=0_+} = 0. \quad (8.24)$$

The former of these conditions again gives Equation (8.22), whereas the latter yields

$$\psi_i - \psi_r = \frac{\alpha}{\beta} \psi_t. \quad (8.25)$$

Here,

$$\alpha = \frac{\cos \theta_t}{\cos \theta_i}, \quad (8.26)$$

$$\beta = \frac{v_1}{v_2} = \frac{n_2}{n_1}. \quad (8.27)$$

It follows that

$$\psi_r = \left(\frac{\beta - \alpha}{\beta + \alpha} \right) \psi_i, \quad (8.28)$$

$$\psi_t = \left(\frac{2\beta}{\beta + \alpha} \right) \psi_i. \quad (8.29)$$

The electromagnetic energy flux in the z -direction (i.e., normal to the interface) is

$$\mathcal{I}_z = E_x H_y. \quad (8.30)$$

Thus, the mean energy fluxes associated with the incident, reflected, and refracted waves are

$$\langle \mathcal{I}_z \rangle_i = \frac{\psi_i^2 \cos \theta_i}{2 \epsilon_0 c n_1}, \quad (8.31)$$

$$\langle \mathcal{I}_z \rangle_r = -\frac{\psi_r^2 \cos \theta_i}{2 \epsilon_0 c n_1}, \quad (8.32)$$

$$\langle \mathcal{I}_z \rangle_t = \frac{\psi_t^2 \cos \theta_t}{2 \epsilon_0 c n_2}, \quad (8.33)$$

respectively. The coefficients of reflection and transmission are defined

$$R = \frac{-\langle \mathcal{I}_z \rangle_r}{\langle \mathcal{I}_z \rangle_i}, \quad (8.34)$$

$$T = \frac{\langle \mathcal{I}_z \rangle_t}{\langle \mathcal{I}_z \rangle_i}, \quad (8.35)$$

respectively. Hence, it follows that

$$R = \left(\frac{\beta - \alpha}{\beta + \alpha} \right)^2, \quad (8.36)$$

$$T = \frac{4\alpha\beta}{(\beta + \alpha)^2} = 1 - R. \quad (8.37)$$

These expressions are known as *Fresnel relations*.

Let us now consider the second polarization, in which the electric components of the incident, reflected, and refracted waves are all parallel to the interface. In this case, the governing equations are

$$\frac{\partial H_x}{\partial t} = v^2 \frac{\partial D_y}{\partial z}, \quad (8.38)$$

$$\frac{\partial H_z}{\partial t} = -v^2 \frac{\partial D_y}{\partial x}, \quad (8.39)$$

$$\frac{\partial D_y}{\partial t} = -\frac{\partial H_z}{\partial x} + \frac{\partial H_x}{\partial z}. \quad (8.40)$$

If we make the transformations $H_y \rightarrow -v^2 D_y$, $D_x \rightarrow H_x$, $D_z \rightarrow H_z$, $\psi_{i,r,t} \rightarrow -v\psi_{i,r,t}$ then we can reuse the solutions that we derived for the other polarization. We find that

$$\begin{aligned} D_y(x, z, t) &= \frac{\psi_i}{v_1} \cos(\omega t - n_1 k_0 \sin \theta_i x - n_1 k_0 \cos \theta_i z) \\ &\quad + \frac{\psi_r}{v_1} \cos(\omega t - n_1 k_0 \sin \theta_i x + n_1 k_0 \cos \theta_i z), \end{aligned} \quad (8.41)$$

$$\begin{aligned} H_x(x, z, t) &= -\psi_i \cos \theta_i \cos(\omega t - n_1 k_0 \sin \theta_i x - n_1 k_0 \cos \theta_i z) \\ &\quad + \psi_r \cos \theta_i \cos(\omega t - n_1 k_0 \sin \theta_i x + n_1 k_0 \cos \theta_i z), \end{aligned} \quad (8.42)$$

$$\begin{aligned} H_z(x, z, t) &= \psi_i \sin \theta_i \cos(\omega t - n_1 k_0 \sin \theta_i x - n_1 k_0 \cos \theta_i z) \\ &\quad + \psi_r \sin \theta_i \cos(\omega t - n_1 k_0 \sin \theta_i x + n_1 k_0 \cos \theta_i z) \end{aligned} \quad (8.43)$$

in the region $z < 0$, and

$$D_y(x, z, t) = \frac{\psi_t}{v_2} \cos(\omega t - n_1 k_0 \sin \theta_i x - n_2 k_0 \cos \theta_t z), \quad (8.44)$$

$$H_x(x, z, t) = -\psi_t \cos \theta_t \cos(\omega t - n_1 k_0 \sin \theta_i x - n_2 k_0 \cos \theta_t z), \quad (8.45)$$

$$H_z(x, z, t) = \psi_t \sin \theta_t \cos(\omega t - n_1 k_0 \sin \theta_i x - n_2 k_0 \cos \theta_t z) \quad (8.46)$$

in the region $z > 0$. The first two matching conditions at the interface are that the normal and tangential components of the magnetic intensity are continuous. In other words,

$$[H_z]_{z=0_-}^{z=0_+} = 0, \quad (8.47)$$

$$[H_x]_{z=0_-}^{z=0_+} = 0. \quad (8.48)$$

The first of these conditions yields

$$\psi_i + \psi_r = \beta^{-1} \psi_t, \quad (8.49)$$

whereas the second gives

$$\psi_i - \psi_r = \alpha \psi_t. \quad (8.50)$$

The final matching condition at the interface is that the tangential component of the electric field is continuous. In other words,

$$[E_y]_{z=0_-}^{z=0_+} = [D_y/(\epsilon \epsilon_0)]_{z=0_-}^{z=0_+} = 0, \quad (8.51)$$

which again yields Equation (8.49). It follows that

$$\psi_r = \left(\frac{1 - \alpha\beta}{1 + \alpha\beta} \right) \psi_i, \quad (8.52)$$

$$\psi_t = \left(\frac{2\beta}{1 + \alpha\beta} \right) \psi_i. \quad (8.53)$$

The electromagnetic energy flux in the z -direction is

$$\mathcal{I}_z = -E_y H_x. \quad (8.54)$$

Thus, the mean energy fluxes associated with the incident, reflected, and refracted waves are

$$\langle \mathcal{I}_z \rangle_i = \frac{\psi_i^2 \cos \theta_i}{2 \epsilon_0 c n_1}, \quad (8.55)$$

$$\langle \mathcal{I}_z \rangle_r = -\frac{\psi_r^2 \cos \theta_i}{2 \epsilon_0 c n_1}, \quad (8.56)$$

$$\langle \mathcal{I}_z \rangle_t = \frac{\psi_t^2 \cos \theta_t}{2 \epsilon_0 c n_2}, \quad (8.57)$$

respectively. Hence, the coefficients of reflection and transmission are

$$R = \left(\frac{1 - \alpha\beta}{1 + \alpha\beta} \right)^2, \quad (8.58)$$

$$T = \frac{4\alpha\beta}{(1 + \alpha\beta)^2} = 1 - R, \quad (8.59)$$

respectively. These expressions are the Fresnel relations for the polarization in which the electric field is parallel to the interface.

It can be seen that, at oblique incidence, the Fresnel relations (8.36) and (8.37) for the polarization in which the magnetic field is parallel to the interface are different to the corresponding relations (8.58) and (8.59) for the polarization in which the electric field is parallel to the interface. This implies that the coefficients of reflection and transmission for these two polarizations are, in general, different.

Figure 8.2 shows the coefficients of reflection and transmission for oblique incidence from air ($n_1 = 1.0$) to glass ($n_2 = 1.5$). Roughly speaking, it can be seen that the coefficient of reflection rises, and the coefficient of transmission falls, as the angle of incidence increases. However, for the polarization in which the magnetic field is parallel to the interface, there is a particular angle of

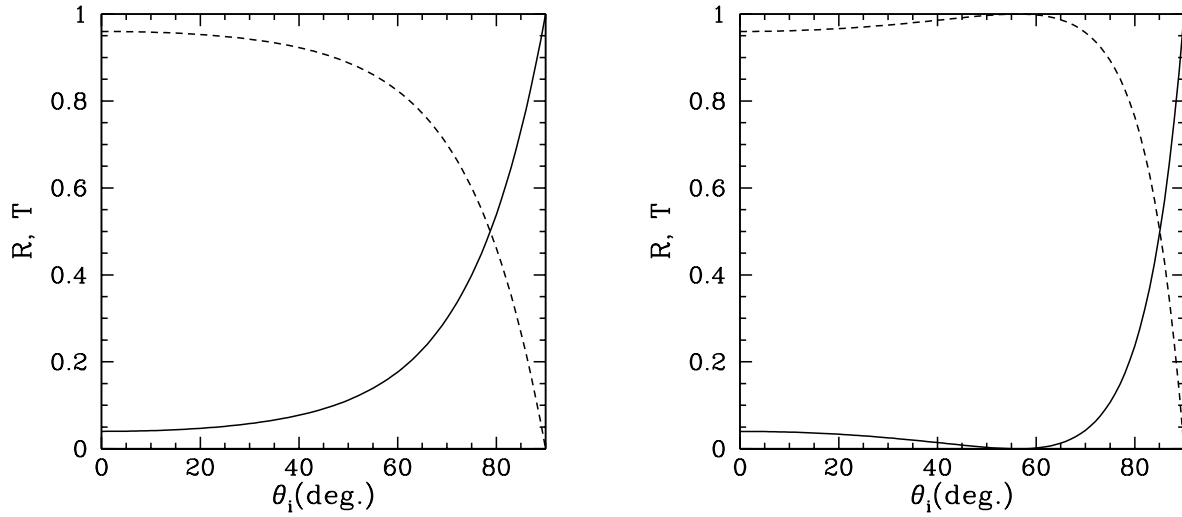


Figure 8.2: Coefficients of reflection (solid curves) and transmission (dashed curves) for oblique incidence from air ($n_1 = 1.0$) to glass ($n_2 = 1.5$). The left-hand panel shows the wave polarization for which the electric field is parallel to the boundary, whereas the right-hand panel shows the wave polarization for which the magnetic field is parallel to the boundary. The Brewster angle is 56.3° .

incidence, known as the *Brewster angle*, at which the reflected intensity is zero. There is no similar behavior for the polarization in which the electric field is parallel to the interface.

It follows from Equation (8.36) that the Brewster angle corresponds to the condition

$$\alpha = \beta, \quad (8.60)$$

or

$$\beta^2 = \frac{\cos^2 \theta_t}{\cos^2 \theta_i} = \frac{1 - \sin^2 \theta_t}{1 - \sin^2 \theta_i} = \frac{1 - \sin^2 \theta_i / \beta^2}{1 - \sin^2 \theta_i}, \quad (8.61)$$

where use has been made of Snell's law. The previous expression reduces to

$$\sin \theta_i = \frac{\beta}{\sqrt{1 + \beta^2}}, \quad (8.62)$$

or $\tan \theta_i = \beta = n_2/n_1$. Hence, the Brewster angle corresponds to $\theta_i = \theta_B$, where

$$\theta_B = \tan^{-1} \left(\frac{n_2}{n_1} \right). \quad (8.63)$$

If unpolarized light is incident on an air/glass (say) interface at the Brewster angle then the reflected light is 100 percent linearly polarized. (See Section 7.6.)

The fact that the coefficient of reflection for the polarization in which the electric field is parallel to the interface is generally greater than that for the other polarization (see Figure 8.2) implies that sunlight reflected from a horizontal water or snow surface is partially linearly polarized, with the horizontal polarization predominating over the vertical one. Such reflected light may be so intense as to cause glare. Polaroid sunglasses help reduce this glare by blocking horizontally polarized light.

8.4 Total Internal Reflection

According to Equation (8.10), when light is obliquely incident at an interface between two dielectric media, the angle of refraction θ_t is related to the angle of incidence θ_i according to

$$\sin \theta_t = \frac{n_1}{n_2} \sin \theta_i. \quad (8.64)$$

This formula presents no problems when $n_1 < n_2$. However, if $n_1 > n_2$ then the formula predicts that $\sin \theta_t$ is greater than unity when the angle of incidence exceeds some *critical angle* given by

$$\theta_c = \sin^{-1}(n_2/n_1). \quad (8.65)$$

In this situation, the analysis of the previous section requires modification.

Consider the polarization in which the magnetic field is parallel to the interface. We can write

$$H_y(x, z, t) = \psi_i \cos(\omega t - n_1 k_0 \sin \theta_i x - n_1 k_0 \cos \theta_i z) + \psi_r \cos(\omega t - n_1 k_0 \sin \theta_i x + n_1 k_0 \cos \theta_i z + \phi_r), \quad (8.66)$$

$$D_x(x, z, t) = \frac{\psi_i \cos \theta_i}{v_1} \cos(\omega t - n_1 k_0 \sin \theta_i x - n_1 k_0 \cos \theta_i z) - \frac{\psi_r \cos \theta_i}{v_1} \cos(\omega t - n_1 k_0 \sin \theta_i x + n_1 k_0 \cos \theta_i z + \phi_r), \quad (8.67)$$

$$D_z(x, z, t) = -\frac{\psi_i \sin \theta_i}{v_1} \cos(\omega t - n_1 k_0 \sin \theta_i x - n_1 k_0 \cos \theta_i z) - \frac{\psi_r \sin \theta_i}{v_1} \cos(\omega t - n_1 k_0 \sin \theta_i x + n_1 k_0 \cos \theta_i z + \phi_r) \quad (8.68)$$

in the region $z < 0$, and

$$H_y(x, z, t) = \psi_t e^{-n_2 k_0 \sinh \theta_t z} \cos(\omega t - n_1 k_0 \sin \theta_i x + \phi_t), \quad (8.69)$$

$$D_x(x, z, t) = \frac{\psi_t \sinh \theta_t}{v_2} e^{-n_2 k_0 \sinh \theta_t z} \sin(\omega t - n_1 k_0 \sin \theta_i x + \phi_t), \quad (8.70)$$

$$D_z(x, z, t) = -\frac{\psi_t \sin \theta_i}{v_1} e^{-n_2 k_0 \sinh \theta_t z} \cos(\omega t - n_1 k_0 \sin \theta_i x + \phi_t) \quad (8.71)$$

in the region $z > 0$. Here,

$$\cosh \theta_t = \sin \theta_t = \frac{n_1}{n_2} \sin \theta_i. \quad (8.72)$$

The matching conditions (8.21) and (8.23) both yield

$$\psi_i + \cos \phi_r \psi_r = \cos \phi_t \psi_t, \quad (8.73)$$

$$\sin \phi_r \psi_r = \sin \phi_t \psi_t, \quad (8.74)$$

whereas the matching condition (8.24) gives

$$\psi_i - \cos \phi_r \psi_r = \frac{\hat{\alpha}}{\beta} \sin \phi_t \psi_t, \quad (8.75)$$

$$\sin \phi_r \psi_r = \frac{\hat{\alpha}}{\beta} \cos \phi_t \psi_t. \quad (8.76)$$

Here,

$$\hat{\alpha} = \frac{\sinh \theta_t}{\cos \theta_i}. \quad (8.77)$$

It follows that

$$\tan \phi_r = \frac{2 \hat{\alpha} \beta}{\beta^2 - \hat{\alpha}^2}, \quad (8.78)$$

$$\tan \phi_t = \frac{\hat{\alpha}}{\beta}, \quad (8.79)$$

$$\psi_i = \psi_t, \quad (8.80)$$

$$\psi_t = \frac{2\beta}{(\beta^2 + \hat{\alpha}^2)^{1/2}} \psi_i. \quad (8.81)$$

Moreover,

$$\langle \mathcal{I}_z \rangle_i = -\langle \mathcal{I}_z \rangle_r = \frac{\psi_i^2 \cos \theta_i}{2 \epsilon_0 c n_1}, \quad (8.82)$$

and

$$\langle \mathcal{I}_z \rangle_t = 0. \quad (8.83)$$

The last result follows because H_y and D_x for the transmitted wave oscillate $\pi/2$ radians out of phase. Hence, when the angle of incidence exceeds the critical angle, the coefficient of reflection is unity, and the coefficient of transmission zero.

Consider the polarization in which the electric field is parallel to the interface. We can write

$$\begin{aligned} D_y(x, z, t) = & \frac{\psi_i}{v_1} \cos(\omega t - n_1 k_0 \sin \theta_i x - n_1 k_0 \cos \theta_i z) \\ & + \frac{\psi_r}{v_1} \cos(\omega t - n_1 k_0 \sin \theta_i x + n_1 k_0 \cos \theta_i z + \phi_r), \end{aligned} \quad (8.84)$$

$$\begin{aligned} H_x(x, z, t) = & -\psi_i \cos \theta_i \cos(\omega t - n_1 k_0 \sin \theta_i x - n_1 k_0 \cos \theta_i z) \\ & + \psi_r \cos \theta_i \cos(\omega t - n_1 k_0 \sin \theta_i x + n_1 k_0 \cos \theta_i z + \phi_r), \end{aligned} \quad (8.85)$$

$$\begin{aligned} H_z(x, z, t) = & \psi_i \sin \theta_i \cos(\omega t - n_1 k_0 \sin \theta_i x - n_1 k_0 \cos \theta_i z) \\ & + \psi_r \sin \theta_i \cos(\omega t - n_1 k_0 \sin \theta_i x + n_1 k_0 \cos \theta_i z + \phi_r) \end{aligned} \quad (8.86)$$

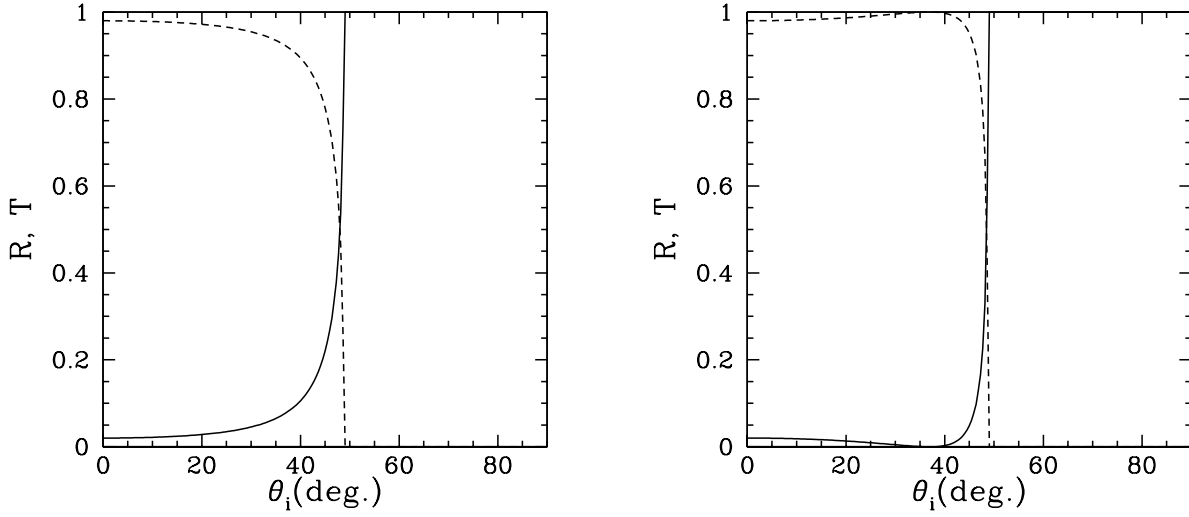


Figure 8.3: Coefficients of reflection (solid curves) and transmission (dashed curves) for oblique incidence from water ($n_1 = 1.33$) to air ($n_2 = 1.0$). The left-hand panel shows the wave polarization for which the electric field is parallel to the interface, whereas the right-hand panel shows the wave polarization for which the magnetic field is parallel to the interface. The critical angle is 48.8° .

in the region $z < 0$, and

$$D_y(x, z, t) = \frac{\psi_t}{v_2} e^{-n_2 k_0 \sinh \theta_t z} \cos(\omega t - n_1 k_0 \sin \theta_i x + \phi_t), \quad (8.87)$$

$$H_x(x, z, t) = -\psi_t \sinh \theta_t e^{-n_2 k_0 \sinh \theta_t z} \sin(\omega t - n_1 k_0 \sin \theta_i x + \phi_t), \quad (8.88)$$

$$H_z(x, z, t) = \psi_t \cosh \theta_t e^{-n_2 k_0 \sinh \theta_t z} \cos(\omega t - n_1 k_0 \sin \theta_i x + \phi_t) \quad (8.89)$$

in the region $z > 0$. The matching conditions (8.47) and (8.51) both yield

$$\psi_i + \cos \phi_r \psi_r = \beta^{-1} \cos \phi_t \psi_t \quad (8.90)$$

$$\sin \phi_r \psi_r = \beta^{-1} \sin \phi_t \psi_t, \quad (8.91)$$

whereas the matching condition (8.48) gives

$$\psi_i - \cos \phi_r \psi_r = \hat{\alpha} \sin \phi_t \psi_t, \quad (8.92)$$

$$\sin \phi_r \psi_r = \hat{\alpha} \cos \phi_t \psi_t. \quad (8.93)$$

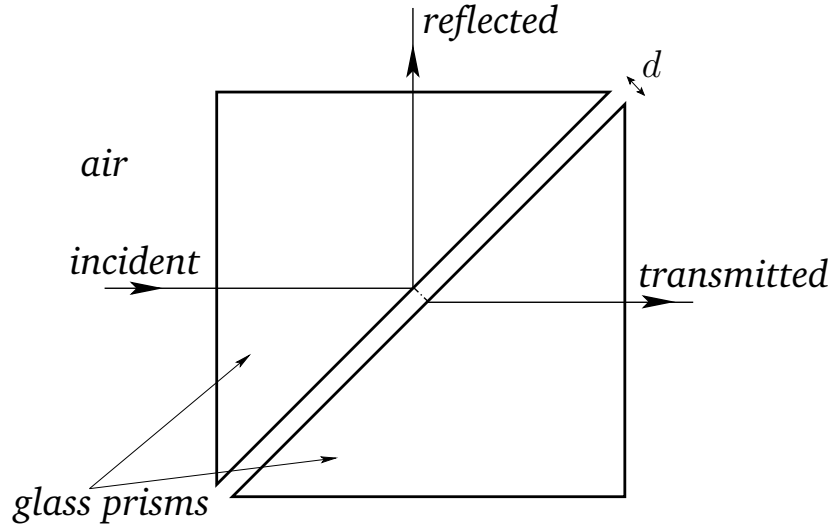


Figure 8.4: Frustrated total internal reflection.

It follows that

$$\tan \phi_r = \frac{2 \hat{\alpha} \beta}{1 - \hat{\alpha}^2 \beta^2}, \quad (8.94)$$

$$\tan \phi_t = \hat{\alpha} \beta, \quad (8.95)$$

$$\psi_i = \psi_t, \quad (8.96)$$

$$\psi_t = \frac{2 \beta}{(1 + \hat{\alpha}^2 \beta^2)^{1/2}} \psi_i. \quad (8.97)$$

As before, if the angle of incidence exceeds the critical angle, the coefficient of reflection is unity, and the coefficient of transmission zero.

According to the above analysis, when light is incident on an interface separating a medium of high refractive index from a medium of low refractive index, and the angle of incidence exceeds the critical angle, θ_c , the transmitted ray becomes evanescent (i.e., its amplitude decays exponentially), and all of the incident energy is reflected. This process is known as *total internal reflection*. Figure 8.3 shows the coefficients of reflection and transmission for oblique incidence from water ($n_1 = 1.33$) to air ($n_2 = 1.0$). In this case, the critical angle is $\theta_c = 48.8^\circ$.

When total internal reflection takes place, the evanescent transmitted wave penetrates a few wavelengths into the lower refractive index medium. The existence of the evanescent wave can be demonstrated using the apparatus pictured in Figure 8.4. This shows two right-angled glass prisms separated by a small air gap of width d . Light incident on the internal surface of the first prism is internally reflected (assuming that $\theta_c < 45^\circ$). However, if the spacing d is not too much larger than the wavelength of the light (in air) then the evanescent wave in the air gap still has a finite amplitude when it reaches the second prism. In this case, a detectable transmitted wave is excited in the second prism. The amplitude of this wave has an inverse exponential dependence on the

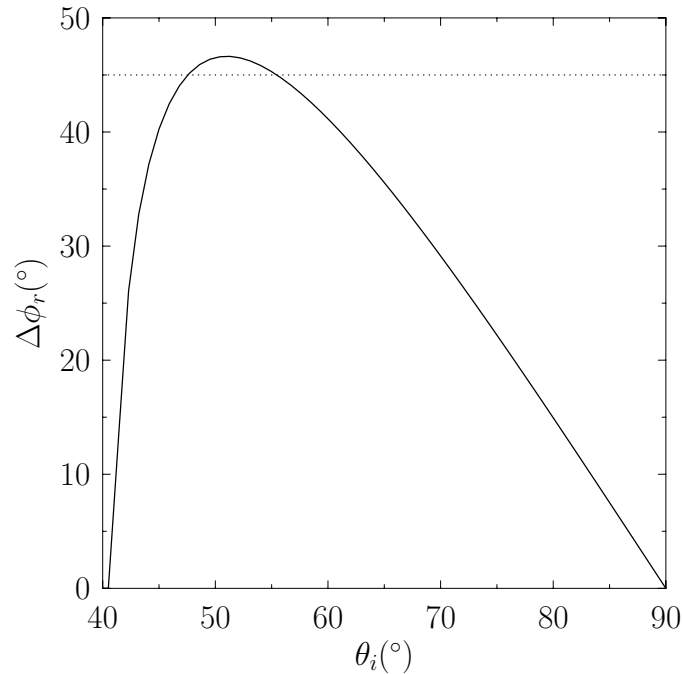


Figure 8.5: Phase advance introduced between the two different wave polarizations by total internal reflection at an interface between glass ($n_1 = 1.52$) and air ($n_2 = 1.0$).

width of the gap. This effect is called *frustrated total internal reflection*, and is analogous to the tunneling of wavefunctions through potential barriers in quantum mechanics.

According to Equations (8.78) and (8.94), total internal reflection produces a phase shift, ϕ_r , between the reflected and the incident waves. Moreover, this phase shift is different for the two possible wave polarizations. Hence, if unpolarized light is subject to total internal reflection then a phase advance, $\Delta\phi_r$, is introduced between the different polarizations. (The phase of the polarization in which the magnetic field is parallel to the interface is advanced with respect to that of the other polarization.) Figure 8.5 shows the phase advance due to total internal reflection at a glass/air interface, as a function of the angle of incidence. Here, the refractive indices of the glass and air are taken to be $n_1 = 1.52$ and $n_2 = 1.0$, respectively. It can be seen that there are two special values of the angle of incidence (i.e., 47.6° and 55.5°) at which the phase advance is $\pi/4$ radians.

The aforementioned phase advance on total internal reflection is exploited in the so-called *Fresnel rhomb* to convert linearly polarized light into circular polarized light. A Fresnel rhomb is a prism-like device (usually in the form of a right-parallelepiped) that is shaped such that light entering one of the small faces is internally reflected twice (once from each of the two sloped faces) before exiting through the other small face. (See Figure 8.6.) The angle of internal reflection is the same in each case, and is designed to produce a $\pi/4$ phase difference between the two wave polarizations. For the case of a prism made up of glass of refractive index 1.52, this is achieved by ensuring that the reflection angle is either 47.6° or 55.5° . The net result of sending light through the device is thus to introduce a $\pi/2$ phase difference between the two polarizations. If the incoming light is linearly polarized at 45° to the plane of the incident and reflected waves then the amplitudes

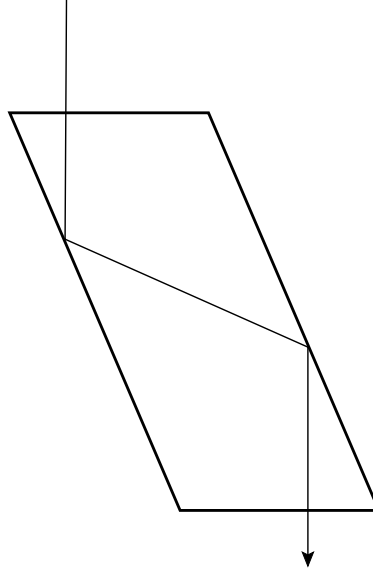


Figure 8.6: Path of light ray through Fresnel rhomb (schematic).

of the two wave polarizations are the same. This ensures that the $\pi/2$ phase difference introduced by the rhomb produces circularly (rather than elliptically) polarized light. (See Section 7.6.)

8.5 Reflection by Conducting Surfaces

Suppose that the region $z < 0$ is a vacuum, and the region $z > 0$ is occupied by a good conductor of conductivity σ . Consider a linearly polarized plane wave normally incident on the interface. Let the wave electric and magnetic fields in the vacuum region take the form

$$E_x(z, t) = E_i \cos[k_0 (c t - z)] + E_r \cos[k_0 (c t + z) + \phi_r], \quad (8.98)$$

$$H_y(z, t) = E_i Z_0^{-1} \cos[k_0 (c t - z)] - E_r Z_0^{-1} \cos[k_0 (c t + z) + \phi_r], \quad (8.99)$$

where $k_0 = \omega/c$ is the vacuum wavenumber. Here, E_i and E_r are the amplitudes of the incident and reflected waves, respectively, whereas $Z_0 = \sqrt{\mu_0/\epsilon_0}$. The wave electric and magnetic fields in the conductor are written

$$E_x(z, t) = E_t e^{-z/d} \cos(\omega t - z/d + \phi_t), \quad (8.100)$$

$$H_y(z, t) = E_t Z_0^{-1} \alpha^{-1} e^{-z/d} \cos(\omega t - z/d - \pi/4 + \phi_t), \quad (8.101)$$

where E_t is the amplitude of the evanescent wave that penetrates into the conductor, ϕ_t is the phase of this wave with respect to the incident wave, and

$$\alpha = \left(\frac{\epsilon_0 \omega}{\sigma} \right)^{1/2} \ll 1. \quad (8.102)$$

The appropriate matching conditions are the continuity of E_x and H_y at the vacuum/conductor interface ($z = 0$). In other words,

$$E_i \cos(\omega t) + E_r \cos(\omega t + \phi_r) = E_t \cos(\omega t + \phi_t), \quad (8.103)$$

$$\alpha [E_i \cos(\omega t) - E_r \cos(\omega t + \phi_r)] = E_t \cos(\omega t - \pi/4 + \phi_t). \quad (8.104)$$

Equations (8.103) and (8.104), which must be satisfied at all times, can be solved, in the limit $\alpha \ll 1$, to give

$$E_r \simeq -(1 - \sqrt{2}\alpha) E_i, \quad (8.105)$$

$$\phi_r \simeq -\sqrt{2}\alpha, \quad (8.106)$$

$$E_t \simeq 2\alpha E_i, \quad (8.107)$$

$$\phi_t \simeq \frac{\pi}{4} - \frac{\alpha}{\sqrt{2}}. \quad (8.108)$$

Hence, the coefficient of reflection becomes

$$R \simeq \left(\frac{E_r}{E_i}\right)^2 \simeq 1 - 2\sqrt{2}\alpha = 1 - \left(\frac{8\epsilon_0\omega}{\sigma}\right)^{1/2}. \quad (8.109)$$

According to the previous analysis, a good conductor reflects a normally incident electromagnetic wave with a phase shift of almost π radians (i.e., $E_r \simeq -E_i$). The coefficient of reflection is just less than unity, indicating that, while most of the incident energy is reflected by the conductor, a small fraction of it is absorbed.

High quality metallic mirrors are generally coated in silver, whose conductivity is 6.3×10^7 ($\Omega \text{ m}$)⁻¹. It follows, from Equation (8.109), that at optical frequencies ($\omega = 4 \times 10^{15}$ rad. s⁻¹) the coefficient of reflection of a silvered mirror is $R \simeq 93.3$ percent. This implies that about 7 percent of the light incident on the mirror is absorbed, rather than being reflected. This rather severe light loss can be problematic in instruments, such as astronomical telescopes, that are used to view faint objects.

8.6 Ionospheric Radio Wave Propagation

Let us investigate the propagation of an electromagnetic wave through a spatially non-uniform dielectric medium. As a specific example, consider the propagation of radio waves through the Earth's ionosphere. The refractive index of the ionosphere can be written [see Equation (7.34)]

$$n^2 = 1 - \frac{\omega_p^2}{\omega(\omega + i\nu)}, \quad (8.110)$$

where ν is a real positive constant that parameterizes the damping of electron motion (in fact, ν is the collision frequency of free electrons with other particles in the ionosphere), and

$$\omega_p = \sqrt{\frac{N e^2}{\epsilon_0 m}} \quad (8.111)$$

is the plasma frequency. In the previous formula, N is the density of free electrons in the ionosphere, and m is the electron mass. We shall assume that the ionosphere is horizontally stratified, so that $N = N(z)$, where the coordinate z measures height above the Earth's surface (the curvature of the Earth's surface is neglected in the following analysis). The ionosphere actually consists of two main layers; the E-layer, and the F-layer. We shall concentrate on the lower E-layer, which lies about 100 km above the surface of the Earth, and is about 50 km thick. The typical day-time number density of free electrons in the E-layer is $N \sim 3 \times 10^{11} \text{ m}^{-3}$. At night-time, the density of free electrons falls to about half this number. The typical day-time plasma frequency of the E-layer is, therefore, about 5 MHz. The typical collision frequency of free electrons in the E-layer is about 0.05 MHz. According to simplistic theory, any radio wave whose frequency lies below the day-time plasma frequency, 5 MHz, (i.e., any wave whose wavelength exceeds about 60 m) is reflected by the ionosphere during the day. Let us investigate in more detail how this process takes place. Note, incidentally, that $\nu \ll \omega$ for mega-Hertz frequency radio waves, so it follows from Equation (8.110) that n^2 is predominately real (i.e., under normal circumstances, electron collisions can be neglected).

The problem of radio wave propagation through the ionosphere was of great practical importance during the first half of the 20th century, because, during that period, long-wave radio waves were the principal means of military communication. Nowadays, the military have far more reliable methods of communication. Nevertheless, this subject area is still worth studying, because the principal tool used to deal with the problem of wave propagation through a non-uniform medium—the so-called WKB approximation—is of great theoretical importance. In particular, the WKB approximation is very widely used in quantum mechanics (in fact, there is a great similarity between the problem of wave propagation through a non-uniform medium, and the problem of solving Schrödinger's equation in the presence of a non-uniform potential).

Maxwell's equations for a wave propagating through a non-uniform, unmagnetized, dielectric medium are

$$\nabla \cdot \mathbf{E} = 0, \quad (8.112)$$

$$\nabla \cdot c \mathbf{B} = 0, \quad (8.113)$$

$$\nabla \times \mathbf{E} = i k c \mathbf{B}, \quad (8.114)$$

$$\nabla \times c \mathbf{B} = -i k n^2 \mathbf{E}, \quad (8.115)$$

where n is the non-uniform refractive index of the medium. It is assumed that all field quantities vary in time like $e^{-i\omega t}$, where $\omega = kc$. Note that, in the following, k is the wavenumber in free space, rather than the wavenumber in the dielectric medium.

8.7 WKB Approximation

Consider a radio wave that is vertically incident, from below, on a horizontally stratified ionosphere. Because the wave normal is initially aligned along the z -axis, and as $n = n(z)$, we expect

all field components to be functions of z only, so that

$$\frac{\partial}{\partial x} = \frac{\partial}{\partial y} = 0. \quad (8.116)$$

In this situation, Equations (8.112)–(8.115) reduce to $E_z = c B_z = 0$, with

$$-\frac{\partial E_y}{\partial z} = i k c B_x, \quad (8.117)$$

$$\frac{\partial c B_x}{\partial z} = -i k n^2 E_y, \quad (8.118)$$

and

$$\frac{\partial E_x}{\partial z} = i k c B_y, \quad (8.119)$$

$$-\frac{\partial c B_y}{\partial z} = -i k n^2 E_x. \quad (8.120)$$

Note that Equations (8.117)–(8.118) and (8.119)–(8.120) are isomorphic and completely independent of one another. It follows that, without loss of generality, we can assume that the wave is linearly polarized with its electric vector parallel to the y -axis. In other words, we need only consider the solution of Equations (8.117)–(8.118). The solution of Equations (8.119)–(8.120) is of exactly the same form, except that it describes a linear polarized wave with its electric vector parallel to the x -axis.

Equations (8.117)–(8.118) can be combined to give

$$\frac{d^2 E_y}{dz^2} + k^2 n^2 E_y = 0. \quad (8.121)$$

Incidentally, because E_y is a function of z only, we now use the total derivative sign d/dz , instead of the partial derivative sign $\partial/\partial z$. The solution of the previous equation for the case of a uniform medium, where n is constant, is straightforward:

$$E_y(z) = A e^{i\phi(z)}, \quad (8.122)$$

where A is a constant, and

$$\phi(z) = \pm k n z. \quad (8.123)$$

Note that the $e^{-i\omega t}$ time dependence of all wave quantities is taken as read during this investigation. The solution (8.122) represents a wave of constant amplitude A and phase $\phi(z)$. According to Equation (8.123), there are, in fact, two independent waves that can propagate through the medium in question. The upper sign corresponds to a wave that propagates vertically upwards, whereas the lower sign corresponds to a wave that propagates vertically downwards. Both waves propagate at the constant phase velocity c/n .

In general, if $n = n(z)$ then the solution to Equation (8.121) does not remotely resemble the wave-like solution (8.122). However, in the limit that $n(z)$ is a “slowly varying” function of z

(exactly how slowly varying is something that will be established later), we expect to recover wave-like solutions. Let us suppose that $n(z)$ is indeed a slowly varying function, and let us try substituting the wave solution (8.122) into Equation (8.121). We obtain

$$\left(\frac{d\phi}{dz}\right)^2 = k^2 n^2 + i \frac{d^2\phi}{dz^2}. \quad (8.124)$$

which is a non-linear differential equation that, in general, is very difficult to solve. However, if n is a constant then $d^2\phi/dz^2 = 0$. It is, therefore, reasonable to suppose that if $n(z)$ is a slowly varying function then the last term on the right-hand side of the previous equation can be regarded as small. Thus, to a first approximation, Equation (8.124) yields

$$\frac{d\phi}{dz} \simeq \pm k n, \quad (8.125)$$

and

$$\frac{d^2\phi}{dz^2} \simeq \pm k \frac{dn}{dz}. \quad (8.126)$$

It is clear from a comparison between Equations (8.124) and (8.126) that $n(z)$ can be regarded as a slowly varying function of z as long as its variation length-scale is far longer than the wavelength of the wave. In other words, provided $(dn/dz)/(kn^2) \ll 1$.

The second approximation to the solution is obtained by substituting Equation (8.126) into the right-hand side of Equation (8.124):

$$\frac{d\phi}{dz} \simeq \pm \left(k^2 n^2 \pm i k \frac{dn}{dz}\right)^{1/2}. \quad (8.127)$$

This gives

$$\frac{d\phi}{dz} \simeq \pm k n \left(1 \pm \frac{i}{kn^2} \frac{dn}{dz}\right)^{1/2} \simeq \pm k n + \frac{i}{2n} \frac{dn}{dz}, \quad (8.128)$$

where use has been made of the binomial expansion. The previous expression can be integrated to give

$$\phi(z) \simeq \pm k \int^z n(z') dz' + i \log(n^{1/2}). \quad (8.129)$$

Substitution of Equation (8.129) into Equation (8.122) yields the final result

$$E_y(z) \simeq A n^{-1/2}(z) \exp\left(\pm i k \int^z n(z') dz'\right). \quad (8.130)$$

It follows from Equation (4.183) that

$$c B_x(z) \simeq \mp A n^{1/2}(z) \exp\left(\pm i k \int^z n(z') dz'\right) - \frac{i A}{2 k n^{3/2}(z)} \frac{dn}{dz} \exp\left(\pm i k \int^z n(z') dz'\right). \quad (8.131)$$

Note that the second term on the right-hand side of the previous expression is small compared to the first, and can usually be neglected.

Let us test to what extent the expression (8.130) is a good solution of Equation (8.121) by substituting the former into the left-hand side of the latter. The result is

$$\frac{A}{n^{1/2}} \left[\frac{3}{4} \left(\frac{1}{n} \frac{dn}{dz} \right)^2 - \frac{1}{2n} \frac{d^2n}{dz^2} \right] \exp \left(\pm i k \int^z n(z') dz' \right), \quad (8.132)$$

which must be small compared with either term on the left-hand side of (8.121). Hence, the condition for expression (8.130) to be a good solution of Equation (8.121) becomes

$$\frac{1}{k^2} \left| \frac{3}{4} \left(\frac{1}{n^2} \frac{dn}{dz} \right)^2 - \frac{1}{2n^3} \frac{d^2n}{dz^2} \right| \ll 1. \quad (8.133)$$

The solution

$$E_y(z) \simeq A n^{-1/2}(z) \exp \left(\pm i k \int^z n(z') dz' \right), \quad (8.134)$$

$$c B_x(z) \simeq \mp A n^{1/2}(z) \exp \left(\pm i k \int^z n(z') dz' \right), \quad (8.135)$$

to the non-uniform wave equations (8.117)–(8.118) is usually called the *WKB solution*, in honor of G. Wentzel, H.A. Kramers, and L. Brillouin, who are credited with independently discovering it (in a quantum mechanical context) in 1926. Actually, H. Jeffries wrote a paper on this solution (in a wave propagation context) in 1923. Hence, some people call it the *WKB solution* (or even the *JWKB solution*). In fact, this solution was first discussed by Liouville and Green in 1837, and again by Rayleigh in 1912. In the following, we refer to (8.134)–(8.135) as the *WKB solution*, because this is what it is generally called. However, it should be understood that, in doing so, we are not making any statement as to the credit due to various scientists in discovering this solution.

As is well known, if a propagating electromagnetic wave is normally incident on an interface where the refractive index suddenly changes (for instance, when a light wave propagating in the air is normally incident on a glass slab) then there is generally significant reflection of the wave. However, according to the WKB solution, (8.134)–(8.135), if a propagating wave is normally incident on a medium in which the refractive index changes slowly along the direction of propagation of the wave then the wave is not reflected at all. This is true even if the refractive index varies very substantially along the path of the wave, as long as it varies slowly. The WKB solution implies that as the wave propagates through the medium its wavelength gradually changes. In fact, the wavelength at position z is approximately $\lambda(z) = 2\pi/k n(z)$. Equations (8.134)–(8.135) also imply that the amplitude of the wave gradually changes as it propagates. In fact, the amplitude of the electric field component is inversely proportional to $n^{1/2}$, whereas the amplitude of the magnetic field component is directly proportional to $n^{1/2}$. Note, however, that the energy flux in the z -direction, which is given by the the Poynting vector $-(E_y B_x^* + E_y^* B_x)/(4\mu_0)$, remains constant (assuming that n is predominately real).

Of course, the WKB solution, (8.134)–(8.135), is only approximate. In reality, a wave propagating into a medium in which the refractive index is a slowly varying function of position is subject to a small amount of reflection. However, it is easily demonstrated that the ratio of the reflected

amplitude to the incident amplitude is of order $(dn/dz)/(kn^2)$. Thus, as long as the refractive index varies on a much longer length-scale than the wavelength of the radiation, the reflected wave is negligibly small. This conclusion remains valid as long as the inequality (8.133) is satisfied. There are two main reasons why this inequality might fail to be satisfied. First of all, if there is a localized region in the dielectric medium where the refractive index suddenly changes (i.e., if there is an interface) then (8.133) is likely to break down in this region, allowing strong reflection of the incident wave. Secondly, the inequality obviously breaks down in the vicinity of a point where $n = 0$. We would, therefore, expect strong reflection of the incident wave from such a point.

8.8 Reflection Coefficient

Consider an ionosphere in which the refractive index is a slowly varying function of height z above the surface of the Earth. Let n^2 be positive for $z < z_0$, and negative for $z > z_0$. Suppose that an upgoing radio wave of amplitude E_0 is generated at ground level ($z = 0$). The complex amplitude of the wave in the region $0 < z < z_0$ is given by the upgoing WKB solution

$$E_y(z) = E_0 n^{-1/2}(z) \exp\left(i k \int_0^z n(z') dz'\right), \quad (8.136)$$

$$c B_x(z) = -E_0 n^{1/2}(z) \exp\left(i k \int_0^z n(z') dz'\right). \quad (8.137)$$

The upgoing energy flux is given by $-(E_y B_x^* + E_y^* B_x)/(4\mu_0) = (\epsilon_0/\mu_0)^{1/2} |E_0|^2/2$. In the region $z > z_0$, the WKB solution takes the form

$$E_y(z) = A e^{i\pi/4} |n(z)|^{-1/2} \exp\left(\pm k \int^z |n(z')| dz'\right), \quad (8.138)$$

$$c B_x(z) = \pm A e^{-i\pi/4} |n(z)|^{1/2} \exp\left(\pm k \int^z |n(z')| dz'\right), \quad (8.139)$$

where A is a constant. These solutions correspond to exponentially growing or decaying waves. Note that the magnetic components of the waves are in phase quadrature with the electric components. This implies that the Poynting fluxes of the waves are zero: in other words, the waves do not transmit energy. Thus, there is a non-zero incident energy flux in the region $z < z_0$, and zero energy flux in the region $z > z_0$. Clearly, the incident wave is either absorbed or reflected in the vicinity of the plane $z = z_0$ (where $n = 0$). In fact, as we shall prove later, the wave is reflected. The complex amplitude of the reflected wave in the region $0 < z < z_0$ is given by the downgoing WKB solution

$$E_y(z) = E_0 R n^{-1/2}(z) \exp\left(-i k \int_0^z n(z') dz'\right), \quad (8.140)$$

$$c B_x(z) = E_0 R n^{1/2}(z) \exp\left(-i k \int_0^z n(z') dz'\right), \quad (8.141)$$

where R is the coefficient of reflection. Suppose, for the sake of argument, that the plane $z = z_0$ acts like a perfect conductor, so that $E_y(z_0) = 0$. It follows that

$$R = -\exp\left(2ik \int_0^{z_0} n(z') dz'\right). \quad (8.142)$$

In fact, as we shall prove later, the correct answer is

$$R = -i \exp\left(2ik \int_0^{z_0} n(z') dz'\right). \quad (8.143)$$

Thus, there is only a $-\pi/2$ phase shift at the reflection point, instead of the $-\pi$ phase shift that would be obtained if the plane $z = z_0$ acted like a perfect conductor.

8.9 Extension to Oblique Incidence

We have discussed the WKB solution for radio waves propagating vertically through an ionosphere whose refractive index varies slowly with height. Let us now generalize this solution to take into account radio waves that propagate at an angle to the vertical axis.

The refractive index of the ionosphere is assumed to vary continuously with height, z . However, let us, for the sake of clarity, imagine that the ionosphere is replaced by a number of thin, discrete, homogeneous, horizontal strata. A continuous ionosphere corresponds to the limit in which the strata become innumerable and infinitely thin. Suppose that a plane wave is incident from below on the ionosphere. Let the wave normal lie in the x - z plane, and subtend an angle θ_I with the vertical axis. At the lower boundary of the first stratum, the wave is partially reflected and partially transmitted. The transmitted wave is then partially reflected and partially transmitted at the lower boundary of the second stratum, and so on. However, in the limit of many strata, in which the difference in refractive indices between neighboring strata is very small, the amount of reflection at the boundaries (which is proportional to the square of this difference) becomes negligible. In the n th stratum, let n_n be the refractive index, and let θ_n be the angle subtended between the wave normal and the vertical axis. According to Snell's law,

$$n_{n-1} \sin \theta_{n-1} = n_n \sin \theta_n. \quad (8.144)$$

Below the ionosphere $n = 1$, and so

$$n_n \sin \theta_n = \sin \theta_I. \quad (8.145)$$

For an electromagnetic wave in the n th stratum, a general field quantity depends on z and x via factors of the form

$$A \exp [ik n_n (\pm z \cos \theta_n + x \sin \theta_n)], \quad (8.146)$$

where A is a constant. The \pm signs denote upgoing and downgoing waves, respectively. When the operator $\partial/\partial x$ acts on the previous expression, it is equivalent to multiplication by $ik n_n \sin \theta_n =$

$ik \sin \theta_I$, which is independent of x and z . It is convenient to use the notation $S = \sin \theta_I$. Hence, we may write symbolically

$$\frac{\partial}{\partial x} \equiv ikS, \quad (8.147)$$

$$\frac{\partial}{\partial y} \equiv 0. \quad (8.148)$$

These results are true no matter how thin the strata are, so they must also hold for a continuous ionosphere. Note that, according to Snell's law, if the wave normal is initially parallel to the x - z plane then it will remain parallel to this plane as the wave propagates through the ionosphere. Equations (8.112)–(8.115) and (8.147)–(8.148) can be combined to give

$$-\frac{\partial E_y}{\partial z} = ikc B_x, \quad (8.149)$$

$$ikS E_y = ikc B_z, \quad (8.150)$$

$$\frac{\partial c B_x}{\partial z} - ikS c B_z = -ikn^2 E_y, \quad (8.151)$$

and

$$\frac{\partial E_x}{\partial z} - ikS E_z = ikc B_y, \quad (8.152)$$

$$-\frac{\partial c B_y}{\partial z} = -ikn^2 E_x, \quad (8.153)$$

$$ikS c B_y = -ikn^2 E_z. \quad (8.154)$$

As before, Maxwell's equations can be split into two independent groups, corresponding to two different polarizations of radio waves propagating through the ionosphere. For the first group of equations, the electric field is always parallel to the y -axis. The corresponding waves are, therefore, said to be *horizontally polarized*. For the second group of equations, the electric field always lies in the x - z plane. The corresponding waves are, therefore, said to be *vertically polarized*. (However, the term "vertically polarized" does not necessarily imply that the electric field is parallel to the vertical axis.) Note that the equations governing horizontally polarized waves are not isomorphic to those governing vertically polarized waves. Consequently, both types of waves must be dealt with separately.

For the case of horizontally polarized waves, Equations (8.150) and (8.151) yield

$$\frac{\partial c B_x}{\partial z} = -ikq^2 E_y, \quad (8.155)$$

where

$$q^2 = n^2 - S^2. \quad (8.156)$$

The previous equation can be combined with Equation (8.149) to give

$$\frac{\partial^2 E_y}{\partial z^2} + k^2 q^2 E_y = 0. \quad (8.157)$$

Equations (8.155) and (8.157) have exactly the same form as Equations (8.118) and (8.121), except that n^2 is replaced by q^2 . Hence, the analysis of Section 7.16 can be reused to find the appropriate WKB solution, which take the form

$$E_y(z) = A q^{-1/2}(z) \exp\left(\pm i k \int^z q(z') dz'\right), \quad (8.158)$$

$$c B_x(z) = \mp A q^{1/2}(z) \exp\left(\pm i k \int^z q(z') dz'\right), \quad (8.159)$$

where A is a constant. Of course, both expressions should also contain a multiplicative factor $e^{i(kSx - \omega t)}$, but this is usually omitted for the sake of clarity. By analogy with Equation (8.133), the previous WKB solution is valid as long as

$$\frac{1}{k^2} \left| \frac{3}{4} \left(\frac{1}{q^2} \frac{dq}{dz} \right)^2 - \frac{1}{2q^3} \frac{d^2q}{dz^2} \right| \ll 1. \quad (8.160)$$

This inequality clearly fails in the vicinity of $q = 0$, no matter how slowly q varies with z . Hence, $q = 0$, which is equivalent to $n^2 = S^2$, specifies the height at which reflection takes place. By analogy with Equation (8.143), the reflection coefficient at ground level ($z = 0$) is given by

$$R = -i \exp\left(2 i k \int_0^{z_0} q(z') dz'\right), \quad (8.161)$$

where z_0 is the height at which $q = 0$.

For the case of vertical polarization, Equations (8.152) and (8.154) yield

$$\frac{\partial E_x}{\partial z} = i k \frac{q^2}{n^2} c B_y. \quad (8.162)$$

This equation can be combined with Equation (8.153) to give

$$\frac{\partial^2 B_y}{\partial z^2} - \frac{1}{n^2} \frac{dn^2}{dz} \frac{\partial B_y}{\partial z} + k^2 q^2 B_y = 0. \quad (8.163)$$

Clearly, the differential equation that governs the propagation of vertically polarized waves is considerably more complicated than the corresponding equation for horizontally polarized waves.

The WKB solution for vertically polarized waves is obtained by substituting the wave-like solution

$$c B_y = A e^{i\phi(z)}, \quad (8.164)$$

where A is a constant, and $\phi(z)$ is the generalized phase, into Equation (8.163). The differential equation thereby obtained for the phase is

$$i \frac{d^2 \phi}{dz^2} - \left(\frac{d\phi}{dz} \right)^2 - \frac{i}{n^2} \frac{dn^2}{dz} \frac{d\phi}{dz} + k^2 q^2 \phi = 0. \quad (8.165)$$

Because the refractive index is assumed to be slowly varying, the first and third term in the previous equation are small, and so, to a first approximation,

$$\frac{d\phi}{dz} = \pm k q, \quad (8.166)$$

$$\frac{d^2 \phi}{dz^2} = \pm k \frac{dq}{dz}. \quad (8.167)$$

These expressions can be substituted into the first and third terms of Equation (8.165) to give the second approximation,

$$\frac{d\phi}{dz} = \pm \left[k^2 q^2 \pm i k \left(\frac{dq}{dz} - \frac{2q}{n} \frac{dn}{dz} \right) \right]^{1/2}. \quad (8.168)$$

The final two terms on the right-hand side of the previous equation are small, so expansion of the right-hand side by means of the binomial theorem yields

$$\frac{d\phi}{dz} = \pm k q + \frac{i}{2q} \frac{dq}{dz} - \frac{i}{n} \frac{dn}{dz}. \quad (8.169)$$

This expression can be integrated, and the result substituted into Equation (8.164), to give the WKB solution

$$c B_y(z) = A n(z) q^{-1/2}(z) \exp \left(\pm i k \int^z q(z') dz' \right). \quad (8.170)$$

The corresponding WKB solution for E_x is obtained from Equation (8.162):

$$E_x(z) = \pm A n^{-1}(z) q^{1/2}(z) \exp \left(\pm i k \int^z q(z') dz' \right). \quad (8.171)$$

Here, any terms involving derivatives of n and q have been neglected.

Substituting Equation (8.170) into the differential equation (8.163), and demanding that the remainder be small compared to the original terms in the equation, yields the following condition for the validity of the previous WKB solution:

$$\frac{1}{k^2} \left| \frac{3}{4} \left(\frac{1}{q^2} \frac{dq}{dz} \right)^2 - \frac{1}{2q^3} \frac{d^2 q}{dz^2} + \frac{1}{q^2} \left[\frac{1}{n} \frac{d^2 n}{dz^2} - 2 \left(\frac{1}{n} \frac{dn}{dz} \right)^2 \right] \right| \ll 1. \quad (8.172)$$

This criterion fails close to $q = 0$, no matter how slowly n and q vary with z . Hence, $q = 0$ gives the height at which reflection takes place. The condition also fails close to $n = 0$, which does not correspond to the reflection level. If, as is usually the case, the electron density in the ionosphere

increases monotonically with height then the level at which $n = 0$ lies above the reflection level (where $q = 0$). If the two levels are well separated then the reflection process is unaffected by the failure of the previous inequality at the level $n = 0$, and the reflection coefficient is given by Equation (8.161), just as is the case for the horizontal polarization. If, however, the level $n = 0$ lies close to the level $q = 0$ then the reflection coefficient may be affected, and a more accurate treatment of the differential equation (8.163) is required in order to obtain the true value of the reflection coefficient.

8.10 Ionospheric Pulse Propagation

Consider a radio wave generator that launches radio pulses vertically upwards into the ionosphere. For the sake of argument, we shall assume that these pulses are linearly polarized such that the electric field vector lies parallel to the y -axis. The pulse structure can be represented as

$$E_y(t) = \int_{-\infty}^{\infty} F(\omega) e^{-i\omega t} d\omega, \quad (8.173)$$

where $E_y(t)$ is the electric field produced by the generator (i.e., the field at $z = 0$). Suppose that the pulse is a signal of roughly constant (angular) frequency ω_0 that lasts a time T , where T is long compared to $1/\omega_0$. It follows that $F(\omega)$ possesses narrow maxima around $\omega = \pm\omega_0$. In other words, only those frequencies that lie very close to the central frequency, ω_0 , play a significant role in the propagation of the pulse.

Each component frequency of the pulse yields a wave that travels independently up into the ionosphere, in a manner specified by the appropriate WKB solution [see Equations (8.158)–(8.159)]. Thus, if Equation (8.173) specifies the signal at ground level ($z = 0$) then the signal at height z is given by

$$E_y(z, t) = \int_{-\infty}^{\infty} \frac{F(\omega)}{n^{1/2}(\omega, z)} e^{i\phi(\omega, z, t)} d\omega, \quad (8.174)$$

where

$$\phi(\omega, z, t) = \frac{\omega}{c} \int_0^z n(\omega, z') dz' - \omega t. \quad (8.175)$$

Here, we have made use of the definition $k = \omega/c$.

Equation (8.174) can be regarded as a contour integral in ω -space. The quantity $F/n^{1/2}$ is a relatively slowly varying function of ω , whereas the phase ϕ is a large and rapidly varying function. As described in Section 7.12, the rapid oscillations of $\exp(i\phi)$ over most of the path of integration ensure that the integrand averages almost to zero. However, this cancellation argument does not apply to those points on the integration path where the phase is stationary: that is, where $\partial\phi/\partial\omega = 0$. It follows that the left-hand side of Equation (8.174) averages to a very small value, except for those special values of z and t at which one of the points of stationary phase in ω -space coincides with one of the peaks of $F(\omega)$. The locus of these special values of z and t can be regarded as the equation of motion of the pulse as it propagates through the ionosphere. Thus, the equation of motion is specified by

$$\left(\frac{\partial\phi}{\partial\omega} \right)_{\omega=\omega_0} = 0, \quad (8.176)$$

which yields

$$t = \frac{1}{c} \int_0^z \left[\frac{\partial(\omega n)}{\partial \omega} \right]_{\omega=\omega_0} dz'. \quad (8.177)$$

Suppose that the z -velocity of a pulse of central frequency ω_0 at height z is given by $u_z(\omega_0, z)$. The differential equation of motion of the pulse is then $dt = dz/u_z$. This can be integrated, using the boundary condition $z = 0$ at $t = 0$, to give the full equation of motion:

$$t = \int_0^z \frac{dz'}{u_z}. \quad (8.178)$$

A comparison between Equations (8.177) and (8.178) yields

$$u_z(\omega_0, z) = c \int \left\{ \frac{\partial[\omega n(\omega, z)]}{\partial \omega} \right\}_{\omega=\omega_0}. \quad (8.179)$$

The velocity u_z corresponds to the group velocity of the pulse. (See Section 7.13.)

The dispersion relation (8.110) yields

$$n(\omega, z) = \left(1 - \frac{\omega_p^2(z)}{\omega^2} \right)^{1/2}, \quad (8.180)$$

in the limit that electron collisions are negligible. The phase velocity of radio waves of frequency ω propagating vertically through the ionosphere is given by

$$v_z(\omega, z) = \frac{c}{n(\omega, z)} = c \left(1 - \frac{\omega_p^2(z)}{\omega^2} \right)^{-1/2}. \quad (8.181)$$

According to Equations (8.179) and (8.180), the corresponding group velocity is

$$u_z(\omega, z) = c \left(1 - \frac{\omega_p^2(z)}{\omega^2} \right)^{1/2}. \quad (8.182)$$

It follows that

$$v_z u_z = c^2. \quad (8.183)$$

Note that as the reflection point $z = z_0$ [defined as the root of $\omega = \omega_p(z_0)$] is approached from below, the phase velocity tends to infinity, whereas the group velocity tends to zero.

Let τ be the time taken for the pulse to travel from the ground to the reflection level, and then back to the ground again. The product $c \tau/2$ is termed the *equivalent height of reflection*, and is denoted $h(\omega)$, because it is a function of the pulse frequency, ω . The equivalent height is the height at which an equivalent pulse traveling at the velocity c would have to be reflected in order to have the same travel time as the actual pulse. Because we know that a pulse of dominant frequency ω propagates at height z with the z -velocity $u_z(\omega, z)$ (this is true for both upgoing and downgoing pulses), and also that the pulse is reflected at the height $z_0(\omega)$, where $\omega = \omega_p(z_0)$, it follows that

$$\tau = 2 \int_0^{z_0(\omega)} \frac{dz}{u_z(\omega, z)}. \quad (8.184)$$

Hence,

$$h(\omega) = \int_0^{z_0(\omega)} \frac{c}{u_z(\omega, z)} dz. \quad (8.185)$$

The equivalent height of reflection, $h(\omega)$, is always greater than the actual height of reflection, $z_0(\omega)$, because the group velocity u_z is always less than the velocity of light. The previous equation can be combined with Equation (8.182) to give

$$h(\omega) = \int_0^{z_0(\omega)} \left(1 - \frac{\omega_p^2(z)}{\omega^2}\right)^{-1/2} dz. \quad (8.186)$$

Note that, despite the fact that the integrand diverges as the reflection point is approached, the integral itself remains finite.

8.11 Measurement of Ionospheric Electron Density Profile

The equivalent height of the ionosphere can be measured in a fairly straightforward manner, by timing how long it takes a radio pulse fired vertically upwards to return to ground level again. We can, therefore, determine the function $h(\omega)$ experimentally by performing this procedure many times with pulses of different central frequencies. But, is it possible to use this information to determine the number density of free electrons in the ionosphere as a function of height? In mathematical terms, the problem is as follows. Does a knowledge of the function

$$h(\omega) = \int_0^{z_0(\omega)} \frac{\omega}{[\omega^2 - \omega_p^2(z)]^{1/2}} dz, \quad (8.187)$$

where $\omega_p^2(z_0) = \omega^2$, necessarily imply a knowledge of the function $\omega_p^2(z)$? Recall that $\omega_p^2(z) \propto N(z)$.

Let $\omega^2 = v$ and $\omega_p^2(z) = u(z)$. Equation (8.187) then becomes

$$v^{-1/2} h(v^{1/2}) = \int_0^{z_0(v^{1/2})} \frac{dz}{[v - u(z)]^{1/2}}, \quad (8.188)$$

where $u(z_0) = v$, and $u(z) < v$ for $0 < z < z_0$. Let us multiply both sides of the previous equation by $(w - v)^{-1/2}/\pi$ and integrate from $v = 0$ to w . We obtain

$$\frac{1}{\pi} \int_0^w v^{-1/2} (w - v)^{-1/2} h(v^{1/2}) dv = \frac{1}{\pi} \int_0^w \left[\int_0^{z_0(v^{1/2})} \frac{dz}{(w - v)^{1/2} (v - u)^{1/2}} \right] dv. \quad (8.189)$$

Consider the double integral on the right-hand side of the previous equation. The region of v - z space over which this integral is performed is sketched in Figure 7.15. It can be seen that, as long as $z_0(v^{1/2})$ is a monotonically increasing function of v , we can swap the order of integration to give

$$\frac{1}{\pi} \int_0^{z_0(w^{1/2})} \left[\int_{u(z)}^w \frac{dv}{(w - v)^{1/2} (v - u)^{1/2}} \right] dz. \quad (8.190)$$

Here, we have used the fact that the curve $z = z_0(v^{1/2})$ is identical with the curve $v = u(z)$. Note that if $z_0(v^{1/2})$ is not a monotonically increasing function of v then we can still swap the order of integration, but the limits of integration are, in general, far more complicated than those indicated previously. The integral over v in the previous expression can be evaluated using standard methods (by making the substitution $v = w \cos^2 \theta + u \sin^2 \theta$): the result is simply π . Thus, expression (8.190) reduces to $z_0(w^{1/2})$. It follows from Equation (8.189) that

$$z_0(w^{1/2}) = \frac{1}{\pi} \int_0^w v^{-1/2} (w-v)^{-1/2} h(v^{1/2}) dv. \quad (8.191)$$

Making the substitutions $v = w \sin^2 \alpha$ and $w^{1/2} = \omega$, we obtain

$$z_0(\omega) = \frac{2}{\pi} \int_0^{\pi/2} h(\omega \sin \alpha) d\alpha. \quad (8.192)$$

By definition, $\omega = \omega_p$ at the reflection level $z = z_0$. Hence, the previous equation reduces to

$$z(\omega_p) = \frac{2}{\pi} \int_0^{\pi/2} h(\omega_p \sin \alpha) d\alpha. \quad (8.193)$$

Thus, we can obtain z as a function of ω_p (and, hence, ω_p as a function of z) by simply taking the appropriate integral of the experimentally determined function $h(\omega)$. Because $\omega_p(z) \propto [N(z)]^{1/2}$, this means that we can determine the electron number density profile in the ionosphere provided that we know the variation of the equivalent height with pulse frequency. The constraint that $z_0(\omega)$ must be a monotonically increasing function of ω translates to the constraint that $N(z)$ must be a monotonically increasing function of z . Note that we can still determine $N(z)$ from $h(\omega)$ for the case where the former function is non-monotonic, it is just a far more complicated procedure than that outlined previously. Incidentally, the mathematical technique by which we have inverted Equation (8.187), which specifies $h(\omega)$ as some integral over $\omega_p(z)$, to give $\omega_p(z)$ as some integral over $h(\omega)$, is known as *Abel inversion*.

8.12 Ionospheric Ray Tracing

Suppose that we possess a radio antenna that is capable of launching radio waves of constant frequency ω into the ionosphere at an angle to the vertical. Let us consider the paths traced out by these waves in the x - z plane. For the sake of simplicity, we shall assume that the waves are horizontally polarized, so that the electric field vector always lies parallel to the y -axis. The signal emitted by the antenna (located at $z = 0$) can be represented as

$$E_y(x) = \int_0^1 F(S) e^{ikSx} dS, \quad (8.194)$$

where $k = \omega/c$. Here, the $e^{-i\omega t}$ time dependence of the signal has been neglected for the sake of clarity. Suppose that the signal emitted by the antenna is mostly concentrated in a direction making

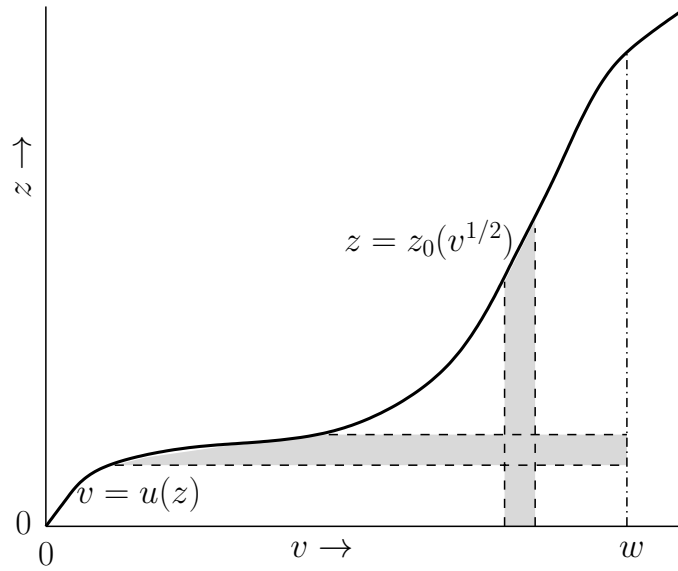


Figure 8.7: A sketch of the region of v - z space over which the integral on the right-hand side of Equation (8.187) is evaluated.

an angle θ_I with the vertical. It follows that $F(S)$ possesses a narrow maximum around $S = S_0$, where $S_0 = \sin \theta_I$.

If Equation (8.194) represents the signal at ground level then the signal at height z in the ionosphere is easily obtained by making use of the WKB solution for horizontally polarized waves described in Section 8.9. We obtain

$$E_y(x, z) = \int_0^1 \frac{F(S)}{q^{1/2}(z, S)} e^{i\phi(x, z, S)} dS, \quad (8.195)$$

where

$$\phi(x, z, S) = k \int_0^z q(z, S) dz + k S x. \quad (8.196)$$

Equation (8.195) is essentially a contour integral in S -space. The quantity $F/q^{1/2}$ is a relatively slowly varying function of S , whereas the phase ϕ is a large and rapidly varying function of S . As described in Section 7.12, the rapid oscillations of $\exp(i\phi)$ over most of the path of integration ensure that the integrand averages almost to zero. In fact, only those points on the path of integration where the phase is stationary (i.e., where $\partial\phi/\partial S = 0$) make a significant contribution to the integral. It follows that the left-hand side of Equation (8.195) averages to a very small value, except for those special values of x and z at which one of the points of stationary phase in S -space coincides with the peak of $F(S)$. The locus of these special values of x and z can clearly be regarded as the trajectory of the radio signal emitted by the antenna as it passes through the ionosphere. Thus, the signal trajectory is specified by

$$\left(\frac{\partial\phi}{\partial S} \right)_{S=S_0} = 0, \quad (8.197)$$

which yields

$$x = - \int_0^z \left(\frac{\partial q}{\partial S} \right)_{S=S_0} dz. \quad (8.198)$$

We can think of this equation as tracing the path of a ray of radio frequency radiation, emitted by the antenna at an angle θ_I to the vertical (where $S_0 = \sin \theta_I$), as it propagates through the ionosphere.

Now

$$q^2 = n^2 - S^2, \quad (8.199)$$

so the ray tracing equation becomes

$$x = S \int_0^z \frac{dz'}{\sqrt{n^2(z) - S^2}}, \quad (8.200)$$

where S is the sine of the initial (i.e., at the antenna) angle of incidence of the ray with respect to the vertical axis. Of course, Equation (8.200) only holds for upgoing rays. For downgoing rays, a simple variant of the previous analysis using the downgoing WKB solutions yields

$$x = S \int_0^{z_0(S)} \frac{dz}{\sqrt{n^2(z) - S^2}} + S \int_z^{z_0(S)} \frac{dz}{\sqrt{n^2(z) - S^2}}, \quad (8.201)$$

where $n(z_0) = S$. Thus, the ray ascends into the ionosphere after being launched from the antenna, reaches a maximum height z_0 above the surface of the Earth, and then starts to descend. The ray eventually intersects the Earth's surface again a horizontal distance

$$x_0 = 2S \int_0^{z_0(S)} \frac{dz}{\sqrt{n^2(z) - S^2}} \quad (8.202)$$

away from the antenna.

The angle ξ which the ray makes with the vertical is given by $\tan \xi = dx/dz$. It follows from Equations (8.200) and (8.201) that

$$\tan \xi = \pm \frac{S}{\sqrt{n^2(z) - S^2}} \quad (8.203)$$

where the upper and lower signs correspond to the upgoing and downgoing parts of the ray trajectory, respectively. Note that $\xi = \pi/2$ at the reflection point, where $n = S$. Thus, the ray is horizontal at the reflection point.

Let us investigate the reflection process in more detail. In particular, we wish to demonstrate that radio waves are reflected at the $q = 0$ surface, rather than being absorbed. We would also like to understand the origin of the $-\pi/2$ phase shift of radio waves at reflection which is evident in Equation (8.161). In order to achieve these goals, we shall need to review the mathematics of asymptotic series.

8.13 Asymptotic Series

It is often convenient to expand a function of the complex variable $f(z)$ as a series in inverse powers of z . For example,

$$f(z) = \phi(z) \left(A_0 + \frac{A_1}{z} + \frac{A_2}{z^2} + \cdots \right), \quad (8.204)$$

where $\phi(z)$ is a function whose behavior for large values of z is known. If $f(z)/\phi(z)$ is singular as $|z| \rightarrow \infty$ then the previous series clearly diverges. Nevertheless, under certain circumstances, the series may still be useful. In fact, this is the case if the difference between $f(z)/\phi(z)$ and the first $n + 1$ terms is of order $1/z^{n+1}$, so that for sufficiently large z this difference becomes vanishingly small. More precisely, the series is said to represent $f(z)/\phi(z)$ asymptotically, that is

$$f(z) \simeq \phi(z) \sum_{p=0, \infty} \frac{A_p}{z^p}, \quad (8.205)$$

provided that

$$\lim_{|z| \rightarrow \infty} \left(z^n \left[\frac{f(z)}{\phi(z)} - \sum_{p=0, n} \frac{A_p}{z^p} \right] \right) \rightarrow 0. \quad (8.206)$$

In other words, for a given value of n , the sum of the first $n + 1$ terms of the series may be made as close as desired to the ratio $f(z)/\phi(z)$ by making z sufficiently large. For each value of z and n there is an error in the series representation of $f(z)/\phi(z)$ which is of order $1/z^{n+1}$. Because the series actually diverges, there is an optimum number of terms in the series used to represent $f(z)/\phi(z)$ for a given value of z . Associated with this is an unavoidable error. As z increases, the optimal number of terms increases, and the error decreases.

Consider a simple example. The exponential integral is defined

$$E_1(x) = \int_x^\infty \frac{e^{-t}}{t} dt. \quad (8.207)$$

The asymptotic series for this function can be generated via a series of partial integrations. For example,

$$E_1(x) = \frac{e^{-x}}{x} - \int_x^\infty \frac{e^{-t}}{t^2} dt. \quad (8.208)$$

A continuation of this procedure yields

$$E_1(x) = \frac{e^{-x}}{x} \left[1 - \frac{1}{x} + \frac{2!}{x^2} - \frac{3!}{x^3} + \cdots + \frac{(-1)^n n!}{x^n} \right] + (-1)^{n+1} (n+1)! \int_x^\infty \frac{e^{-t}}{t^{n+2}} dt. \quad (8.209)$$

The infinite series obtained by taking the limit $n \rightarrow \infty$ diverges, because the Cauchy convergence test yields

$$\lim_{n \rightarrow \infty} \left| \frac{u_{n+1}}{u_n} \right| = \lim_{n \rightarrow \infty} \left(\frac{n}{x} \right) \rightarrow \infty. \quad (8.210)$$

Note that two successive terms in the series become equal in magnitude for $n = x$, indicating that the optimum number of terms for a given x is roughly the nearest integer to x . To prove that the series is asymptotic, we need to show that

$$\lim_{x \rightarrow 0} x^{n+1} e^x (-1)^{n+1} (n+1)! \int_x^\infty \frac{e^{-t}}{t^{n+2}} dt = 0. \quad (8.211)$$

This immediately follows, because

$$\int_x^\infty \frac{e^{-t}}{t^{n+2}} dt < \frac{1}{x^{n+2}} \int_x^\infty e^{-t} dt = \frac{e^{-x}}{x^{n+2}}. \quad (8.212)$$

Thus, the error involved in using the first n terms in the series is less than $(n+1)! e^{-x}/x^{n+2}$, which is the magnitude of the next term in the series. We can see that, as n increases, this estimate of the error first decreases, and then increases without limit. In order to visualize this phenomenon more exactly, let $f(x) = x \exp(x) E(x)$, and let

$$f_n(x) = \sum_{p=0, n} \frac{(-1)^p p!}{x^p} \quad (8.213)$$

be the asymptotic series representation of this function that contains $n+1$ terms. Figure 8.8 shows the relative error in the asymptotic series $|f_n(x) - f(x)|/f(x)$ plotted as a function of the approximate number of terms in the series, n , for $x = 10$. It can be seen that as n increases the error initially falls, reaches a minimum value at about $n = 10$, and then increases rapidly. Clearly, the optimum number of terms in the asymptotic series used to represent $f(10)$ is about 10.

Asymptotic series are fundamentally different to conventional power law expansions, such as

$$\sin z = z - \frac{z^3}{3!} + \frac{z^5}{5!} - \frac{z^7}{7!} + \dots \quad (8.214)$$

This series representation of $\sin z$ converges absolutely for all finite values of z . Thus, at any z , the error associated with the series can be made as small as is desired by including a sufficiently large number of terms. In other words, unlike an asymptotic series, there is no intrinsic, or unavoidable, error associated with a convergent series. It follows that a convergent power law series representation of a function is unique within the domain of convergence of the series. On the other hand, an asymptotic series representation of a function is not unique. It is perfectly possible to have two different asymptotic series representations of the same function, as long as the difference between the two series is less than the intrinsic error associated with each series. Furthermore, it is often the case that different asymptotic series are used to represent the same single-valued analytic function in different regions of the complex plane.

For example, consider the asymptotic expansion of the confluent hypergeometric function $F(a, c, z)$. This function is the solution of the differential equation

$$z F'' + (c - z) F' - a F = 0 \quad (8.215)$$

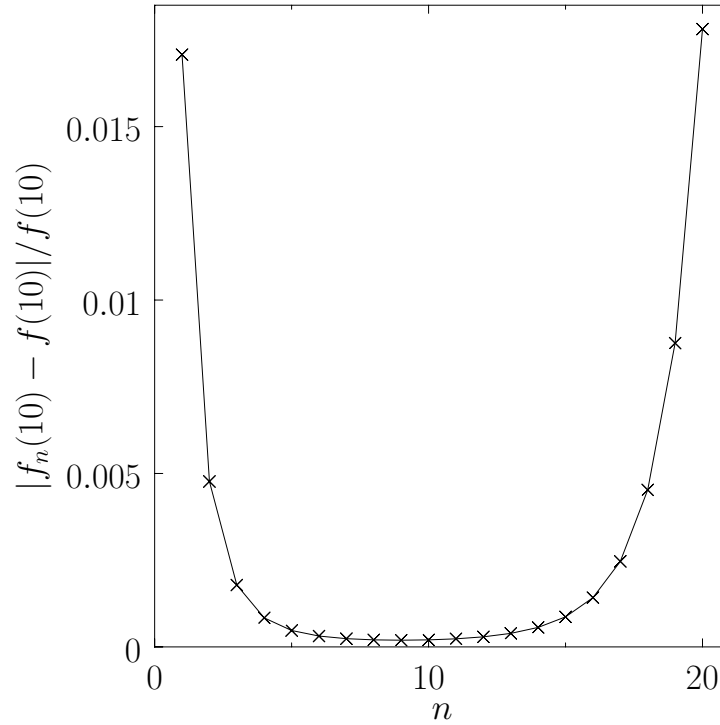


Figure 8.8: The relative error in a typical asymptotic series plotted as a function of the number of terms in the series.

which is analytic at $z = 0$ [in fact, $F(a, c, 0) = 1$]. Here, ' denotes d/dz . The asymptotic expansion of $F(a, c, z)$ takes the form:

$$\frac{\Gamma(a)\Gamma(c-a)}{\Gamma(c)} F(a, c, z) \simeq \Gamma(c-a) z^{a-c} e^z \left[1 + \mathcal{O}\left(\frac{1}{z}\right) \right] + \Gamma(a) z^{-a} e^{-i\pi a} \left[1 + \mathcal{O}\left(\frac{1}{z}\right) \right] \quad (8.216)$$

for $-\pi < \arg(z) < 0$, and

$$\frac{\Gamma(a)\Gamma(c-a)}{\Gamma(c)} F(a, c, z) \simeq \Gamma(c-a) z^{a-c} e^z \left[1 + \mathcal{O}\left(\frac{1}{z}\right) \right] + \Gamma(a) z^{-a} e^{i\pi a} \left[1 + \mathcal{O}\left(\frac{1}{z}\right) \right] \quad (8.217)$$

for $0 < \arg(z) < \pi$, and

$$\begin{aligned} \frac{\Gamma(a)\Gamma(c-a)}{\Gamma(c)} F(a, c, z) \simeq & \Gamma(c-a) z^{a-c} e^{-i2\pi(a-c)} e^z \left[1 + \mathcal{O}\left(\frac{1}{z}\right) \right] \\ & + \Gamma(a) z^{-a} e^{i\pi a} \left[1 + \mathcal{O}\left(\frac{1}{z}\right) \right] \end{aligned} \quad (8.218)$$

for $\pi < \arg(z) < 2\pi$, et cetera. Here,

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt \quad (8.219)$$

is a so-called Gamma function. This function has the property that $\Gamma(n + 1) = n!$, where n is a non-negative integer. It can be seen that the expansion consists of a linear combination of two asymptotic series (only the first term in each series is shown). For $|z| \gg 1$, the first series is exponentially larger than the second whenever $\text{Re}(z) > 0$. The first series is said to be *dominant* in this region, whereas the second series is said to be *subdominant*. Likewise, the first series is exponentially smaller than the second whenever $\text{Re}(z) < 0$. Hence, the first series is subdominant, and the second series dominant, in this region.

Consider a region in which one or other of the series is dominant. Strictly speaking, it is not mathematically consistent to include the subdominant series in the asymptotic expansion, because its contribution is actually less than the intrinsic error associated with the dominant series [this error is $O(1/z)$ times the dominant series, because we are only including the first term in this series]. Thus, at a general point in the complex plane, the asymptotic expansion simply consists of the dominant series. However, this is not the case in the immediate vicinity of the lines $\text{Re}(z) = 0$, which are called *anti-Stokes lines*. When an anti-Stokes line is crossed, a dominant series becomes subdominant, and vice versa. Thus, in the immediate vicinity of an anti-Stokes line neither series is dominant, so it is mathematically consistent to include both series in the asymptotic expansion.

The hypergeometric function $F(a, c, z)$ is a perfectly well-behaved, single-valued, analytic function in the complex plane. However, our two asymptotic series are, in general, multi-valued functions in the complex plane [see Equation (8.216)]. Can a single-valued function be represented asymptotically by a multi-valued function? The short answer is no. We have to employ different combinations of the two series in different regions of the complex plane in order to ensure that $F(a, c, z)$ remains single-valued. Equations (8.216)–(8.218) show how this is achieved. Basically, the coefficient in front of the subdominant series changes discontinuously at certain values of $\arg(z)$. This is perfectly consistent with $F(a, c, z)$ being an analytic function because the subdominant series is “invisible”: in other words, the contribution of the subdominant series to the asymptotic solution falls below the intrinsic error associated with the dominant series, so that it does not really matter if the coefficient in front of the former series changes discontinuously. Imagine tracing a large circle, centered on the origin, in the complex plane. Close to an anti-Stokes line, neither series is dominant, so we must include both series in the asymptotic expansion. As we move away from the anti-Stokes line, one series becomes dominant, which means that the other series becomes subdominant, and, therefore, drops out of our asymptotic expansion. Eventually, we approach a second anti-Stokes line, and the subdominant series reappears in our asymptotic expansion. However, the coefficient in front of the subdominant series, when it reappears, is different to that when the series disappeared. This new coefficient is carried across the second anti-Stokes line into the region where the subdominant series becomes dominant. In this new region, the dominant series becomes subdominant, and disappears from our asymptotic expansion. Eventually, a third anti-Stokes line is approached, and the series reappears, but, again, with a different coefficient in front. The jumps in the coefficients of the subdominant series are chosen in such a manner that if we perform a complete circuit in the complex plane then the value of the asymptotic expansion is the same at the beginning and the end points. In other words, the asymptotic expansion is single-valued, despite the fact that it is built up out of two asymptotic series that are not single-valued. The jumps in the coefficient of the subdominant series, which are needed to keep the asymptotic ex-

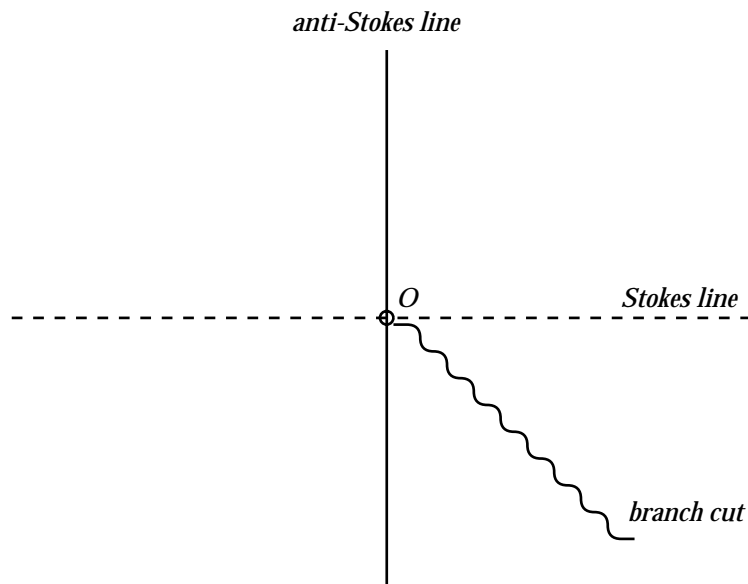


Figure 8.9: The location of the Stokes lines (dashed), the anti-Stokes lines (solid), and the branch cut (wavy) in the complex plane for the asymptotic expansion of the hypergeometric function.

pansion single-valued, are called *Stokes phenomena*, after the celebrated nineteenth century British mathematician Sir George Gabriel Stokes, who first drew attention to this effect.

Where exactly does the jump in the coefficient of the subdominant series occur? All we can really say is “somewhere in the region between two anti-Stokes lines where the series in question is subdominant.” The problem is that we only retained the first term in each asymptotic series. Consequently, the intrinsic error in the dominant series is relatively large, and we lose track of the subdominant series very quickly after moving away from an anti-Stokes line. However, we could include more terms in each asymptotic series. This would enable us to reduce the intrinsic error in the dominant series, and, thereby, expand the region of the complex plane in the vicinity of the anti-Stokes lines where we can see both the dominant and subdominant series. If we were to keep adding terms to our asymptotic series, so as to minimize the error in the dominant solution, we would eventually be forced to conclude that a jump in the coefficient of the subdominant series can only take place on those lines in the complex plane on which $\text{Im}(z) = 0$: these are called *Stokes lines*. This result was first proved by Stokes in 1857.¹ On a Stokes line, the magnitude of the dominant series achieves its maximum value with respect to that of the subdominant series. Once we know that a jump in the coefficient of the subdominant series can only take place at a Stokes line, we can retain the subdominant series in our asymptotic expansion in all regions of the complex plane. What we are basically saying is that, although, in practice, we cannot actually see the subdominant series very far away from an anti-Stokes line, because we are only retaining the first term in each asymptotic series, we could, in principle, see the subdominant series at all values of $\arg(z)$ provided that we retained a sufficient number of terms in our asymptotic series.

Figure 7.17 shows the location in the complex plane of the Stokes and anti-Stokes lines for

¹G.G. Stokes, *Trans. Camb. Phil. Soc.* **10**, 106–128 (1857).

the asymptotic expansion of the hypergeometric function. Also shown is a branch cut, which is needed to make z single-valued. The branch cut is chosen such that $\arg(z) = 0$ on the positive real axis. Every time we cross an anti-Stokes line, the dominant series becomes subdominant, and vice versa. Every time we cross a Stokes line, the coefficient in front of the dominant series stays the same, but that in front of the subdominant series jumps discontinuously [see Equations (8.216)–(8.218)]. Finally, the jumps in the coefficient of the subdominant series are such as to ensure that the asymptotic expansion is single-valued.

8.14 WKB Solution as Asymptotic Series

We have seen that the WKB solution

$$E_y(z) = n^{-1/2}(z) \exp\left(\pm i k \int^z n(z') dz'\right) \quad (8.220)$$

is an approximate solution of the differential equation

$$\frac{d^2 E_y}{dz^2} + k^2 n^2(z) E_y = 0 \quad (8.221)$$

in the limit that the typical wavelength, $2\pi/nk$, is much smaller than the typical variation length-scale of the refractive index. But, what sort of approximation is involved in writing this solution?

It is convenient to define the scaled variable

$$\hat{z} = \frac{z}{L}, \quad (8.222)$$

where L is the typical variation length-scale of the refractive index, $n(z)$. Equation (8.221) can then be written

$$w'' + h^2 q w = 0, \quad (8.223)$$

where $w(\hat{z}, h) \equiv E_y(L\hat{z})$, $q(\hat{z}) \equiv n^2(L\hat{z})$, $' \equiv d/d\hat{z}$, and $h = kL$. Note that, in general, $q(\hat{z})$, $q'(\hat{z})$, $q''(\hat{z})$, et cetera, are $O(1)$ quantities. The non-dimensional constant h is of order the ratio of the variation length-scale of the refractive index to the wavelength. Let us seek the solutions to Equation (8.223) in the limit $h \gg 1$.

We can write

$$w(\hat{z}, h) = \exp[i h \phi(\hat{z}, h)]. \quad (8.224)$$

Equation (8.223) transforms to

$$\frac{i}{h} \phi'' - (\phi')^2 + q = 0. \quad (8.225)$$

Expanding in powers of $1/h$, we obtain

$$\phi'(\hat{z}, h) = \pm q^{1/2}(\hat{z}) + \frac{i}{4h} \frac{q'(\hat{z})}{q(\hat{z})} + O\left(\frac{1}{h^2}\right), \quad (8.226)$$

which yields

$$w(\hat{z}, h) = q^{-1/4}(\hat{z}) \exp\left(\pm i h \int^{\hat{z}} q(\hat{z}') d\hat{z}'\right) \left[1 + O\left(\frac{1}{h}\right)\right]. \quad (8.227)$$

Of course, we immediately recognize this expression as a WKB solution.

Suppose that we keep expanding in powers of $1/h$ in Equation (8.226). The appropriate generalization of Equation (8.227) is a series solution of the form

$$w(\hat{z}, h) = q^{-1/4}(\hat{z}) \exp\left(\pm i h \int^{\hat{z}} q(\hat{z}') d\hat{z}'\right) \left[1 + \sum_{p=1, \infty} \frac{A_p(\hat{z})}{h^p}\right]. \quad (8.228)$$

This is, in fact, an asymptotic series in h . We can now appreciate that a WKB solution is just a highly truncated asymptotic series in h , in which only the first term in the series is retained.

But, why is it so important that we recognize that WKB solutions are highly truncated asymptotic series? The point is that the WKB method was initially rather controversial after it was popularized in the 1920s. Many scientists thought that the method was not mathematically rigorous. Let us try to understand the origin of the problem. Suppose that we have never heard of an asymptotic series. Looking at Equation (8.228), we would imagine that the expression in square brackets is a power law expansion in $1/h$. The WKB approximation involves neglecting all terms in this expansion except the first. This sounds fine, as long as h is much greater than unity. But, surely, to be mathematically rigorous, we have to check that the sum of all of the terms in the expansion that we are neglecting is small compared to the first term? However, if we attempt this then we discover, much to our consternation, that the expansion is divergent. In other words, the sum of all of the neglected terms is infinite! Thus, if we interpret Equation (8.228) as a conventional power law expansion in $1/h$ then the WKB method is clearly nonsense: in fact, the WKB solution would be the first approximation to infinity. However, once we appreciate that Equation (8.228) is actually an asymptotic series in h , the fact that the series diverges becomes irrelevant. If we retain the first n terms in the series then the series approximates the exact solution of Equation (8.223) with an intrinsic (fractional) error which is of order $1/h^n$ (i.e., the first neglected term in the series). The error is minimized at a particular value of h . As the number of terms in the series is increased, the intrinsic error decreases, and the value of h at which the error is minimized increases. In particular, we can see that there is an intrinsic error associated with a WKB solution that is of order $1/h$ times the solution.

It is amusing to note that if Equation (8.228) were not a divergent series then it would be impossible to obtain total reflection of the WKB solutions at the point $q = 0$. As we shall discover, such reflection is directly associated with the fact that the expansion (8.228) exhibits a Stokes phenomenon. It is, of course, impossible for a convergent power series expansion to exhibit a Stokes phenomenon.

8.15 Stokes Constants

We have seen that the differential equation

$$w'' + h^2 q(\hat{z}) w = 0, \quad (8.229)$$

where $' \equiv d/d\hat{z}$, possesses approximate WKB solutions of the form

$$(a, \hat{z}) = q^{-1/4}(\hat{z}) \exp\left(i h \int_a^{\hat{z}} q^{1/2}(\hat{z}') d\hat{z}'\right) \left[1 + \mathcal{O}\left(\frac{1}{h}\right)\right], \quad (8.230)$$

$$(\hat{z}, a) = q^{-1/4}(\hat{z}) \exp\left(-i h \int_a^{\hat{z}} q^{1/2}(\hat{z}') d\hat{z}'\right) \left[1 + \mathcal{O}\left(\frac{1}{h}\right)\right]. \quad (8.231)$$

Here, we have adopted an arbitrary phase reference level $\hat{z} = a$. The convenient notation (a, \hat{z}) is fairly self explanatory: a and \hat{z} refer to the lower and upper bounds of integration, respectively, inside the exponential. It follows that the other WKB solution can be written (\hat{z}, a) (because we can reverse the limits of integration inside the exponential to obtain minus the integral in \hat{z} from $\hat{z} = a$ to $\hat{z} = \hat{z}$).

Up to now, we have thought of \hat{z} as a real variable representing scaled height in the ionosphere. Let us now generalize our analysis somewhat, and think of \hat{z} as a complex variable. There is nothing in our derivation of the WKB solutions that depends crucially on \hat{z} being a real variable, so we expect these solutions to remain valid when \hat{z} is reinterpreted as a complex variable. Incidentally, we must now interpret $q(\hat{z})$ as some well-behaved function of the complex variable. An approximate general solution of the differential equation (8.229) in the complex \hat{z} -plane can be written as a linear superposition of the two WKB solutions (8.230)–(8.231).

The parameter h is assumed to be much larger than unity. It is clear from Equations (8.230)–(8.231) that in some regions of the complex plane one of the WKB solutions is going to be exponentially larger than the other. In such regions, it is not mathematically consistent to retain the smaller WKB solution in the expression for the general solution, because the contribution of the smaller WKB solution is less than the intrinsic error associated with the larger solution. Adopting the terminology introduced in Section 8.13, the larger WKB solution is said to be dominant, and the smaller solution is said to be subdominant. Let us denote the WKB solution (8.230) as $(a, \hat{z})_d$ in regions of the complex plane where it is dominant, and as $(a, \hat{z})_s$ in regions where it is subdominant. An analogous notation is adopted for the second WKB solution (8.231).

Suppose that $q(\hat{z})$ possesses a simple zero at the point $\hat{z} = \hat{z}_0$ (chosen to be the origin, for the sake of convenience). It follows that in the immediate neighborhood of the origin we can write

$$q(\hat{z}) = a_1 \hat{z} + a_2 \hat{z}^2 + \cdots, \quad (8.232)$$

where $a_1 \neq 0$. It is convenient to adopt the origin as the phase reference point (i.e., $a = 0$), so the two WKB solutions become $(0, \hat{z})$ and $(\hat{z}, 0)$. We can define *anti-Stokes lines* in the complex \hat{z} plane (see Section 8.13). These are lines that satisfy

$$\operatorname{Re}\left(i \int_0^{\hat{z}} q^{1/2}(\hat{z}') d\hat{z}'\right) = 0. \quad (8.233)$$

As we cross an anti-Stokes line, a dominant WKB solution becomes subdominant, and vice versa. Thus, $(0, \hat{z})_d \leftrightarrow (0, \hat{z})_s$ and $(\hat{z}, 0)_d \leftrightarrow (\hat{z}, 0)_s$. In the immediate vicinity of an anti-Stokes line the two WKB solutions have about the same magnitude, so it is mathematically consistent to include the contributions from both solutions in the expression for the general solution. In such a region, we

can drop the subscripts d and s , because the WKB solutions are neither dominant nor subdominant, and write the WKB solutions simply as $(0, \hat{z})$ and $(\hat{z}, 0)$.

It is clear from Equations (8.230)–(8.231) that the WKB solutions are not single-valued functions of \hat{z} , because they depend on $q^{1/2}(\hat{z})$, which is a double-valued function. Thus, if we wish to write an approximate analytic solution to the differential equation (8.229) then we cannot express this solution as the same linear combination of WKB solutions in all regions of the complex \hat{z} -plane. This implies that there must exist certain lines in the complex \hat{z} -plane across which the mix of WKB solutions in our expression for the general solution changes discontinuously. These lines are called *Stokes lines* (see Section 8.13), and satisfy

$$\operatorname{Im} \left(i \int_0^{\hat{z}} q^{1/2}(\hat{z}') d\hat{z}' \right) = 0. \quad (8.234)$$

As we cross a Stokes line, the coefficient of the dominant WKB solution in our expression for the general solution must remain unchanged, but the coefficient of the subdominant solution is allowed to change discontinuously. Incidentally, this is perfectly consistent with the fact that the general solution is analytic: the jump in our expression for the general solution due to the jump in the coefficient of the subdominant WKB solution is less than the intrinsic error in this expression due to the intrinsic error in the dominant WKB solution. Once we appreciate that the coefficient of the subdominant solution can only change at a Stokes line, we can retain both WKB solutions in our expression for the general solution throughout the complex \hat{z} -plane. In practice, we can only see a subdominant solution in the immediate vicinity of an anti-Stokes line, but if we were to evaluate the WKB solutions to higher accuracy [i.e., by retaining more terms in the asymptotic series in Equations (8.230)–(8.231)] then we could, in principle, follow a subdominant solution all the way to a neighboring Stokes line.

In the immediate vicinity of the origin

$$\int_0^{\hat{z}} q^{1/2}(\hat{z}') d\hat{z}' \simeq \frac{2\sqrt{a_1}}{3} \hat{z}^{3/2}. \quad (8.235)$$

It follows from Equations (8.233) and (8.234) that three Stokes lines and three anti-Stokes lines radiate from a zero of $q(\hat{z})$. The general arrangement of Stokes and anti-Stokes lines in the vicinity of a $q = 0$ point is sketched in Figure 8.10. Note that a branch cut must also radiate from the $q = 0$ point in order to uniquely specify the function $q^{1/2}(\hat{z})$. Thus, in general, seven lines radiate from a zero of $q(\hat{z})$, dividing the complex \hat{z} plane into seven domains (numbered 1 through 7).

Let us write our general solution as

$$w(\hat{z}, h) = A(0, \hat{z}) + B(\hat{z}, 0) \quad (8.236)$$

on the anti-Stokes line between domains 1 and 7, where A and B are arbitrary constants. Suppose that the WKB solution $(0, \hat{z})$ is dominant in domain 7. Thus, in domain 7 the general solution takes the form

$$w(7) = A(0, \hat{z})_d + B(\hat{z}, 0)_s. \quad (8.237)$$

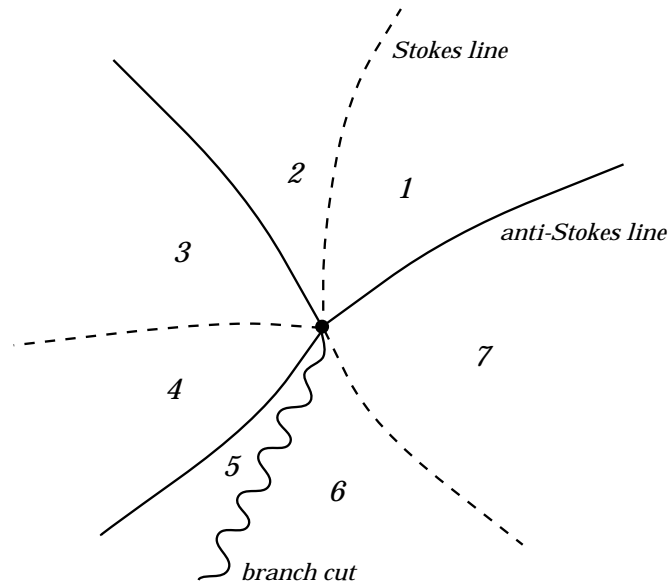


Figure 8.10: The arrangement of Stokes lines (dashed) and anti-Stokes lines (solid) around a simple zero of $q(\hat{z})$. Also shown is the branch cut (wavy line). All of the lines radiate from the point $q = 0$.

Let us move into domain 1. In doing so, we cross an anti-Stokes line, so the dominant solution becomes subdominant, and vice versa. Thus, in domain 1, the general solution takes the form

$$w(1) = A(0, \hat{z})_s + B(\hat{z}, 0)_d. \quad (8.238)$$

Let us now move into domain 2. In doing so, we cross a Stokes line, so the coefficient of the dominant solution, B , must remain constant, but the coefficient of the subdominant solution, A , is allowed to change. Suppose that the coefficient of the subdominant solution jumps by t times the coefficient of the dominant solution, where t is an undetermined constant. It follows that in domain 2 the general solution takes the form

$$w(2) = (A + tB)(0, \hat{z})_s + B(\hat{z}, 0)_d. \quad (8.239)$$

Let us now move into domain 3. In doing so, we cross an anti-Stokes line, so the the dominant solution becomes subdominant, and vice versa. Thus, in domain 3, the general solution takes the form

$$w(3) = (A + tB)(0, \hat{z})_d + B(\hat{z}, 0)_s. \quad (8.240)$$

Let us now move into domain 4. In doing so, we cross a Stokes line, so the coefficient of the dominant solution must remain constant, but the coefficient of the subdominant solution is allowed to change. Suppose that the coefficient of the subdominant solution jumps by u times the coefficient of the dominant solution, where u is an undetermined constant. It follows that in domain 4 the general solution takes the form

$$w(4) = (A + tB)(0, \hat{z})_d + (B + u[A + tB])(\hat{z}, 0)_s. \quad (8.241)$$

Let us now move into domain 5. In doing so, we cross an anti-Stokes line, so the the dominant solution becomes subdominant, and vice versa. Thus, in domain 5 the general solution takes the form

$$w(5) = (A + tB)(0, \hat{z})_s + (B + u[A + tB])(\hat{z}, 0)_d. \quad (8.242)$$

Let us now move into domain 6. In doing so, we cross the branch cut in an anti-clockwise direction. Thus, the argument of \hat{z} decreases by 2π . It follows from Equation (8.232) that $q^{1/2} \rightarrow -q^{1/2}$ and $q^{1/4} \rightarrow -iq^{1/4}$. The following rules for tracing the WKB solutions across the branch cut (in an anti-clockwise direction) ensure that the general solution is continuous across the cut [see Equations (8.230)–(8.231)]:

$$(0, \hat{z}) \rightarrow -i(\hat{z}, 0), \quad (8.243)$$

$$(\hat{z}, 0) \rightarrow -i(0, \hat{z}). \quad (8.244)$$

Note that the properties of dominance and subdominance are preserved when the branch cut is crossed. It follows that in domain 6 the general solution takes the form

$$w(6) = -i(A + tB)(\hat{z}, 0)_s - i(B + u[A + tB])(0, \hat{z})_d. \quad (8.245)$$

Let us, finally, move into domain 7. In doing so, we cross a Stokes line, so the coefficient of the dominant solution must remain constant, but the coefficient of the subdominant solution is allowed to change. Suppose that the coefficient of the subdominant solution jumps by v times the coefficient of the dominant solution, where v is an undetermined constant. It follows that in domain 7 the general solution takes the form

$$w(7) = -i(A + tB + v\{B + u[A + tB]\})(\hat{z}, 0)_s - i(B + u[A + tB])(0, \hat{z})_d. \quad (8.246)$$

Now, we expect our general solution to be an analytic function, so it follows that the solutions (8.237) and (8.246) must be identical. Thus, we can compare the coefficients of the two WKB solutions, $(\hat{z}, 0)_s$ and $(0, \hat{z})_d$. Because A and B are arbitrary, we can also compare the coefficients of A and B . Comparing the coefficients of $A(0, \hat{z})_d$, we find

$$1 = -iu. \quad (8.247)$$

Comparing the coefficients of $B(0, \hat{z})_d$ yields

$$0 = 1 + ut. \quad (8.248)$$

Comparing the coefficients of $A(\hat{z}, 0)_s$ gives

$$0 = 1 + vu. \quad (8.249)$$

Finally, comparing the coefficients of $B(\hat{z}, 0)_s$ yields

$$1 = -i(t + v + vut). \quad (8.250)$$

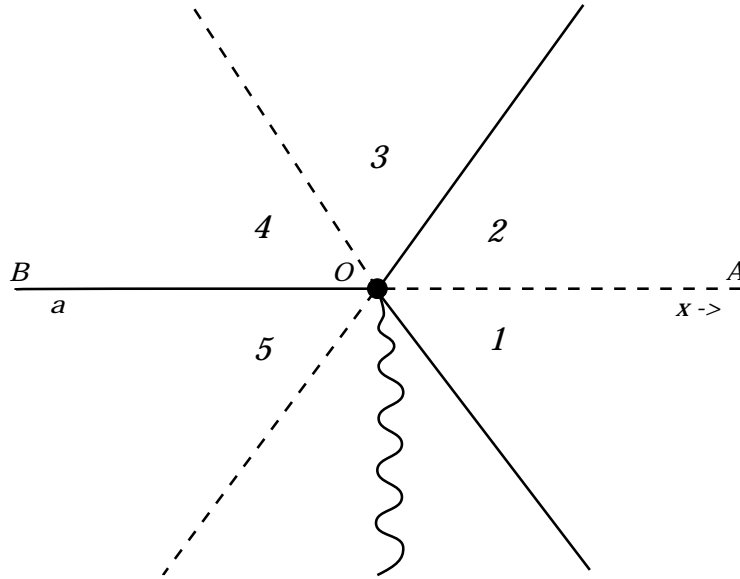


Figure 8.11: The arrangement of Stokes lines (dashed) and anti-Stokes lines (solid) in the complex \hat{z} plane. Also shown is the branch cut (wavy line).

Equations (8.247)–(8.250) imply that

$$t = u = v = i. \quad (8.251)$$

In other words, if we adopt the simple rule that every time we cross a Stokes line in an anti-clockwise direction the coefficient of the subdominant solution jumps by i times the coefficient of the dominant solution then this ensures that our expression for the general solution, (8.236), behaves as an analytic function. Here, the constant i is usually called a *Stokes constant*. Note that if we cross a Stokes line in a clockwise direction then the coefficient of the subdominant solution has to jump by $-i$ times the coefficient of the dominant solution in order to ensure that our general solution behaves as an analytic function.

8.16 WKB Reflection Coefficient

Let us write $\hat{z} = x + iy$, where x and y are real variables. Consider the solution of the differential equation

$$w'' + h^2 q(x) w = 0, \quad (8.252)$$

where $q(x)$ is a real function, h is a large number, $q > 0$ for $x < 0$, and $q < 0$ for $x > 0$. It is clear that $\hat{z} = 0$ represents a simple zero of $q(\hat{z})$. Here, we assume, as seems eminently reasonable, that we can find a well-behaved function of the complex variable $q(\hat{z})$ such that $q(\hat{z}) = q(x)$ along the real axis. The arrangement of Stokes and anti-Stokes lines in the immediate vicinity of the point $\hat{z} = 0$ is sketched in Figure 8.11. The argument of $q(\hat{z})$ on the positive x -axis is chosen to be $-\pi$. Thus, the argument of $q(\hat{z})$ on the negative x -axis is 0.

On OB , the two WKB solutions (8.230)–(8.231) can be written

$$(0, x) = q^{-1/4}(x) \exp\left(i h \int_0^x q^{1/2}(x') dx'\right), \quad (8.253)$$

$$(x, 0) = q^{-1/4}(x) \exp\left(-i h \int_0^x q^{1/2}(x') dx'\right). \quad (8.254)$$

Here, we can interpret $(0, x)$ as a wave propagating to the right along the x -axis, and $(x, 0)$ as a wave propagating to the left. On OA , the WKB solutions take the form

$$(0, x)_d = e^{i\pi/4} |q(x)|^{-1/4} \exp\left(+h \int_0^x |q(x')|^{1/2} dx'\right), \quad (8.255)$$

$$(x, 0)_s = e^{i\pi/4} |q(x)|^{-1/4} \exp\left(-h \int_0^x |q(x')|^{1/2} dx'\right). \quad (8.256)$$

Clearly, $(x, 0)_s$ represents an evanescent wave that decays to the right along the x -axis, whereas $(0, x)_d$ represents an evanescent wave that decays to the left. If we adopt the boundary condition that there is no incident wave from the region $x \rightarrow +\infty$, the most general asymptotic solution to Equation (8.252) on OA is written

$$w(x, h) = A (x, 0)_s, \quad (8.257)$$

where A is an arbitrary constant.

Let us assume that we can find an analytic solution, $w(\hat{z}, h)$, to the differential equation

$$w'' + h^2 q(\hat{z}) w = 0, \quad (8.258)$$

which satisfies $w(\hat{z}, h) = w(x, h)$ along the real axis, where $w(x, h)$ is the physical solution. From a mathematical point of view, this seems eminently reasonable. In the domains 1 and 2, the solution (8.257) becomes

$$w(1) = A (\hat{z}, 0)_s, \quad (8.259)$$

and

$$w(2) = A (\hat{z}, 0)_s, \quad (8.260)$$

respectively. Note that the solution is continuous across the Stokes line OA , because the coefficient of the dominant solution $(0, \hat{z})$ is zero: thus, the jump in the coefficient of the subdominant solution is zero times the Stokes constant, i . In other words, it is zero. Let us move into domain 3. In doing so, we cross an anti-Stokes line, so the solution becomes

$$w(3) = A (\hat{z}, 0)_d. \quad (8.261)$$

Let us now move into domain 4. In doing so, we cross a Stokes line. Applying the general rule derived in the preceding section, the solution becomes

$$w(4) = A (\hat{z}, 0)_d + i A (0, \hat{z})_s. \quad (8.262)$$

Finally, on OB the solution becomes

$$w(x, h) = A(x, 0) + iA(0, x). \quad (8.263)$$

Suppose that there is a point a on the negative x -axis where $q(x) = 1$. It follows from Equations (8.255) and (8.263) that we can write the asymptotic solution to Equation (8.252) as

$$w(x, h) = q^{-1/4}(x) \exp\left(ih \int_a^x q^{1/2}(x') dx'\right) - i \exp\left(2ih \int_a^0 q^{1/2}(x') dx'\right) q^{-1/4}(x) \exp\left(-ih \int_a^x q^{1/2}(x') dx'\right), \quad (8.264)$$

in the region $x < 0$, and

$$w(x, h) = \exp\left(ih \int_a^0 q^{1/2}(x') dx'\right) e^{-i\pi/4} |q(x)|^{-1/4} \exp\left(-h \int_0^x |q(x')|^{1/2} dx'\right) \quad (8.265)$$

in the region $x > 0$. Here, we have chosen

$$A = -i \exp\left(ih \int_a^0 q^{1/2}(x') dx'\right). \quad (8.266)$$

If we interpret x as a normalized altitude in the ionosphere, $q(x)$ as the square of the refractive index in the ionosphere, the point a as ground level, and w as the electric field strength of a radio wave propagating vertically upwards into the ionosphere, then Equation (8.264) tells us that a unit amplitude wave fired vertically upwards from ground level into the ionosphere is reflected at the level where the refractive index is zero. The first term in Equation (8.264) is the incident wave, and the second term is the reflected wave. The reflection coefficient (i.e., the ratio of the reflected to the incident wave at ground level) is given by

$$R = -i \exp\left(2ih \int_a^0 q^{1/2}(x') dx'\right). \quad (8.267)$$

Note that $|R| = 1$, so the amplitude of the reflected wave equals that of the incident wave. In other words, there is no absorption of the wave at the level of reflection. The phase shift of the reflected wave at ground level, with respect to that of the incident wave, is that associated with the wave propagating from ground level to the reflection level and back to ground level again, plus a $-\pi/2$ phase shift at reflection. According to Equation (8.265), the wave attenuates fairly rapidly (in the space of a few wavelengths) above the reflection level. Of course, Equation (8.267) is completely equivalent to Equation (8.143).

Note that the reflection of the incident wave at the point where the refractive index is zero is directly associated with the Stokes phenomenon. Without the jump in the coefficient of the subdominant solution, as we go from domain 3 to domain 4, there is no reflected wave on the OB axis. Note, also, that the WKB solutions (8.264) and (8.265) break down in the immediate vicinity of $q = 0$ (i.e., at the reflection point). Thus, it is possible to demonstrate that the incident wave is totally reflected at the point $q = 0$, with a $-\pi/2$ phase shift, without having to solve for the wave structure in the immediate vicinity of the reflection point. This demonstrates that the reflection of the incident wave at $q = 0$ is an intrinsic property of the WKB solutions, and does not depend on the detailed behavior of the wave in the region where the WKB solutions break down.

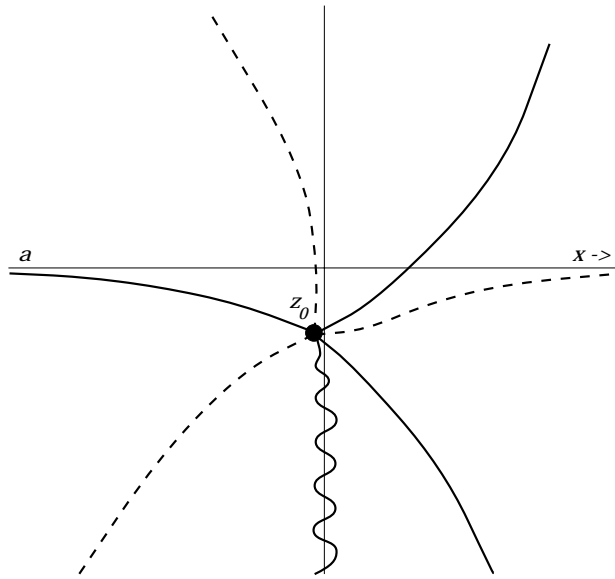


Figure 8.12: The arrangement of Stokes lines (dashed) and anti-Stokes lines (solid) in the complex \hat{z} plane. Also shown is the branch cut (wavy line).

8.17 Jeffries Connection Formula

In the preceding section, there is a tacit assumption that the square of the refractive index, $q(x) \equiv n^2(x)$, is a real function. However, as is apparent from Equation (8.110), this is only the case in the ionosphere as long as electron collisions are negligible. Let us generalize our analysis to take electron collisions into account. In fact, the main effect of electron collisions is to move the zero of $q(\hat{z})$ a short distance off the real axis (the distance is relatively short provided that we adopt the physical ordering $\nu \ll \omega$). The arrangement of Stokes and anti-Stokes lines around the new zero point, located at $\hat{z} = \hat{z}_0$, is sketched in Figure 8.12. Note that electron collisions only significantly modify the form of $q(\hat{z})$ in the immediate vicinity of the zero point. Thus, sufficiently far away from $\hat{z} = \hat{z}_0$ in the complex \hat{z} -plane, the WKB solutions, as well as the locations of the Stokes and anti-Stokes lines, are exactly the same as in the preceding section.

The WKB solutions (8.253) and (8.254) are valid all the way along the real axis, except for a small region close to the origin where electron collisions significantly modify the form of $q(\hat{z})$. Thus, we can still adopt the physically reasonable decaying solution (8.255) on the positive real axis. Let us trace this solution in the complex \hat{z} -plane until we reach the negative real axis. We can achieve this by moving in a semi-circle in the upper half-plane. Because we never move out of the region in which the WKB solutions (8.253) and (8.254) are valid, we conclude, by analogy with the preceding section, that the solution on the negative real axis is given by Equation (8.263). Of course, in all of the WKB solutions the point $\hat{z} = 0$ must be replaced by the new zero point $\hat{z} = \hat{z}_0$. The new formula for the reflection coefficient, which is just a straightforward generalization of

Equation (8.267), is

$$R = -i \exp\left(2i h \int_a^{\hat{z}_0} q^{1/2}(\hat{z}') d\hat{z}'\right). \quad (8.268)$$

This is called the *Jeffries connection formula*, after H. Jeffries, who discovered it in 1923. Thus, the general expression for the reflection coefficient is incredibly simple. We just integrate the WKB solution in the complex \hat{z} -plane from the phase reference level $\hat{z} = a$ to the zero point, square the result, and multiply by $-i$. Note that the path of integration between $\hat{z} = a$ and $\hat{z} = \hat{z}_0$ does not matter, because of Cauchy's theorem. Note, also, that because $q^{1/2}$ is, in general, complex along the path of integration, we no longer have $|R| = 1$. In fact, it is easily demonstrated that $|R| \leq 1$. Thus, when electron collisions are included in the analysis, we no longer obtain perfect reflection of radio waves from the ionosphere. Instead, some (small) fraction of the radio energy is absorbed at each reflection event. This energy is ultimately transferred to the particles in the ionosphere with which the electrons collide.

8.18 Exercises

- 8.1 Consider an electromagnetic wave propagating through a nonuniform dielectric medium whose dielectric constant ϵ is a function of \mathbf{r} . Demonstrate that the associated wave equations take the form

$$\begin{aligned} \nabla^2 \mathbf{E} - \frac{\epsilon}{c^2} \frac{\partial^2 \mathbf{E}}{\partial t^2} &= -\nabla \left(\frac{\nabla \epsilon \cdot \mathbf{E}}{\epsilon} \right), \\ \nabla^2 \mathbf{B} - \frac{\epsilon}{c^2} \frac{\partial^2 \mathbf{B}}{\partial t^2} &= -\frac{\nabla \epsilon \times (\nabla \times \mathbf{B})}{\epsilon}. \end{aligned}$$

- 8.2 Suppose that a light-ray is incident on the front (air/glass) interface of a uniform pane of glass of refractive index n at the Brewster angle. Demonstrate that the refracted ray is also incident on the rear (glass/air) interface of the pane at the Brewster angle.
- 8.3 Consider an electromagnetic wave obliquely incident on a plane boundary between two transparent magnetic media of permeabilities μ_1 and μ_2 . Find the coefficients of reflection and transmission as functions of the angle of incidence for the wave polarizations in which all electric fields are parallel to the boundary and all magnetic fields are parallel to the boundary. Is there a Brewster angle? If so, what is it? Is it possible to obtain total reflection? If so, what is the critical angle of incidence required to obtain total reflection?
- 8.4 A medium is such that the product of the phase and group velocities of electromagnetic waves is equal to c^2 at all wave frequencies. Demonstrate that the dispersion relation for electromagnetic waves takes the form

$$\omega^2 = k^2 c^2 + \omega_0^2,$$

where ω_0 is a constant.

- 8.5 Demonstrate that if the equivalent height of reflection in the ionosphere varies with the angular frequency of the wave as

$$h(\omega) = h_0 + \delta \left(\frac{\omega}{\omega_0} \right)^p,$$

where h_0 , δ , and ω_0 are positive constants, then $\omega_p(z) = 0$ for $z < h_0$, and

$$\omega_p(z) = \left[\frac{\pi \Gamma(1+p)}{\Gamma(1/2+p/2)\Gamma(1/2+p/2)} \right]^{1/p} \frac{\omega_0}{2} \left(\frac{z-h_0}{\delta} \right)^{1/p}$$

for $z \geq h_0$. Here, $\Gamma(z)$ is a Gamma function.

- 8.6 Suppose that the refractive index, $n(z)$, of the ionosphere is given by $n^2 = 1 - \alpha(z - h_0)$ for $z \geq h_0$, and $n^2 = 1$ for $z < h_0$, where α and h_0 are positive constants, and the Earth's magnetic field and curvature are both neglected. Here, z measures altitude above the Earth's surface.

- (a) A point transmitter sends up a wave packet at an angle θ to the vertical. Show that the packet returns to Earth a distance

$$x_0 = 2 h_0 \tan \theta + \frac{2}{\alpha} \sin 2\theta$$

from the transmitter. Demonstrate that if $\alpha h_0 < 1/4$ then for some values of x_0 the previous equation is satisfied by three different values of θ . In other words, wave packets can travel from the transmitter to the receiver via one of three different paths. Show that the critical case $\alpha h_0 = 1/4$ corresponds to $\theta = \pi/3$ and $x_0 = 6\sqrt{3}h_0$.

- (b) A point radio transmitter emits a pulse of radio waves uniformly in all directions. Show that the pulse first returns to the Earth a distance $4h_0(2/\alpha h_0 - 1)^{1/2}$ from the transmitter, provided that $\alpha h_0 < 2$.

9 Radiation and Scattering

9.1 Introduction

Let us briefly investigate the emission and reception of electromagnetic radiation by antenna systems, as well as the scattering of such radiation by charged particles.

9.2 Basic Antenna Theory

It is possible to solve exactly for the radiation pattern emitted by a linear antenna fed with a sinusoidal current pattern. Assuming that all fields and currents vary in time like $e^{-i\omega t}$, and adopting the Lorenz gauge, it is easily demonstrated that the vector potential obeys the inhomogeneous Helmholtz equation,

$$(\nabla^2 + k^2) \mathbf{A} = -\mu_0 \mathbf{j}, \quad (9.1)$$

where $k = \omega/c$. The Green's function for this equation, subject to the Sommerfeld radiation condition (which ensures that sources radiate waves instead of absorbing them), is

$$G(\mathbf{r}, \mathbf{r}') = -\frac{e^{ik|\mathbf{r}-\mathbf{r}'|}}{4\pi|\mathbf{r}-\mathbf{r}'|}. \quad (9.2)$$

(See Chapter 1.) Thus, we can invert Equation (9.1) to obtain

$$\mathbf{A}(\mathbf{r}) = \frac{\mu_0}{4\pi} \int \frac{\mathbf{j}(\mathbf{r}') e^{ik|\mathbf{r}-\mathbf{r}'|}}{|\mathbf{r}-\mathbf{r}'|} dV'. \quad (9.3)$$

The electric field in the source-free region,

$$\mathbf{E} = \frac{i}{k} \nabla \times c \mathbf{B}, \quad (9.4)$$

follows from the Ampère-Maxwell equation, as well as the definition $\mathbf{B} = \nabla \times \mathbf{A}$,

Now,

$$|\mathbf{r} - \mathbf{r}'| = r \left[1 - \frac{2\mathbf{n} \cdot \mathbf{r}'}{r} + \frac{r'^2}{r^2} \right]^{1/2}, \quad (9.5)$$

where $\mathbf{n} = \mathbf{r}/r$. Assuming that $r' \ll r$, this expression can be expanded binomially to give

$$|\mathbf{r} - \mathbf{r}'| = r \left[1 - \frac{\mathbf{n} \cdot \mathbf{r}'}{r} + \frac{r'^2}{2r^2} - \frac{1}{8} \left(\frac{2\mathbf{n} \cdot \mathbf{r}'}{r} \right)^2 + \dots \right], \quad (9.6)$$

where we have retained all terms up to order $(r'/r)^2$. The above expansion, which appears in the complex exponential of Equation (9.3), determines the phase of the radiation emitted by each element of the antenna. The quadratic terms in the expansion can be neglected provided they can be

shown to contribute a phase change that is significantly less than 2π . Thus, because the maximum possible value of r' is $d/2$, for a linear antenna that extends along the z -axis from $z = -d/2$ to $z = d/2$, the phase shift associated with the quadratic terms is insignificant as long as

$$r \gg \frac{k d^2}{16\pi} = \frac{d^2}{8\lambda}, \quad (9.7)$$

where $\lambda = 2\pi/k$ is the wavelength of the radiation. This constraint is known as the *Fraunhofer limit*.

In the Fraunhofer limit, we can approximate the phase variation of the complex exponential in Equation (9.3) as a linear function of r' :

$$|\mathbf{r} - \mathbf{r}'| \simeq r - \mathbf{n} \cdot \mathbf{r}'. \quad (9.8)$$

The denominator $|\mathbf{r} - \mathbf{r}'|$ in the integrand of Equation (9.3) can be approximated as r provided that the distance from the antenna is much greater than its length: that is,

$$r \gg d. \quad (9.9)$$

Thus, Equation (9.3) reduces to

$$\mathbf{A}(\mathbf{r}) \simeq \frac{\mu_0}{4\pi} \frac{e^{ikr}}{r} \int \mathbf{j}(\mathbf{r}') e^{-i\mathbf{k}\cdot\mathbf{n}\cdot\mathbf{r}'} dV' \quad (9.10)$$

when the constraints (9.7) and (9.9) are satisfied. If the additional constraint

$$kr \gg 1 \quad (9.11)$$

is also satisfied then the electromagnetic fields associated with Equation (9.10) take the form

$$\mathbf{B}(\mathbf{r}) \simeq ik \mathbf{n} \times \mathbf{A} = ik \frac{\mu_0}{4\pi} \frac{e^{ikr}}{r} \int \mathbf{n} \times \mathbf{j}(\mathbf{r}') e^{-i\mathbf{k}\cdot\mathbf{n}\cdot\mathbf{r}'} dV', \quad (9.12)$$

$$\mathbf{E}(\mathbf{r}) \simeq c \mathbf{B} \times \mathbf{n} = ic k (\mathbf{n} \times \mathbf{A}) \times \mathbf{n}. \quad (9.13)$$

These are clearly radiation fields, because they are mutually orthogonal, transverse to the radius vector, \mathbf{n} , and satisfy $E = cB \propto r^{-1}$. (See Section 1.8.) The three constraints (9.7), (9.9), and (9.11), can be summed up in the single inequality

$$d \ll \sqrt{\lambda r} \ll r. \quad (9.14)$$

The current density associated with a linear, sinusoidal, centre-fed antenna, aligned along the z -axis, is

$$\mathbf{j}(\mathbf{r}) = I \sin(kd/2 - k|z|) \delta(x) \delta(y) \mathbf{e}_z \quad (9.15)$$

for $|z| < d/2$, with $\mathbf{j}(\mathbf{r}) = 0$ for $|z| \geq d/2$. In this case, Equation (9.10) yields

$$\mathbf{A}(\mathbf{r}) = \frac{\mu_0 I}{4\pi} \frac{e^{ikr}}{r} \int_{-d/2}^{d/2} \sin(kd/2 - k|z|) e^{-ikz \cos\theta} dz \mathbf{e}_z, \quad (9.16)$$

where $\cos \theta = \mathbf{n} \cdot \mathbf{e}_z$. The result of this straightforward integration is

$$\mathbf{A}(\mathbf{r}) = \frac{\mu_0 I_0 2 e^{i k r}}{4\pi k r} \left[\frac{\cos(k d \cos \theta/2) - \cos(k d/2)}{\sin^2 \theta} \right] \mathbf{e}_z. \quad (9.17)$$

According to Equations (9.12) and (9.13), the electric component of the emitted radiation lies in the plane containing the antenna and the radius vector connecting the antenna to the observation point. The time-averaged power radiated by the antenna per unit solid angle is

$$\frac{dP}{d\Omega} = \frac{\operatorname{Re}(\mathbf{n} \cdot \mathbf{E} \times \mathbf{B}^*) r^2}{2\mu_0} = \frac{c k^2 \sin^2 \theta |A|^2 r^2}{2\mu_0}, \quad (9.18)$$

or

$$\frac{dP}{d\Omega} = \frac{\mu_0 c I_0^2}{8\pi^2} \left| \frac{\cos(k d \cos \theta/2) - \cos(k d/2)}{\sin \theta} \right|^2. \quad (9.19)$$

The angular distribution of power depends on the value of $k d$. In the long wavelength limit, $k d \ll 1$, the distribution reduces to

$$\frac{dP}{d\Omega} = \frac{\mu_0 c I_0^2}{128\pi^2} (k d)^2 \sin^2 \theta, \quad (9.20)$$

where $I_0 = I k d/2$ is the peak current in the antenna. It is easily shown, from Equation (9.15), that the associated current distribution in the antenna is linear: that is,

$$I(z) = I_0 (1 - 2|z|/d) \quad (9.21)$$

for $|z| < d/2$. This type of antenna corresponds to a short (compared to the wavelength) oscillating electric dipole, and is generally known as a *Hertzian dipole*. The total power radiated is

$$P = \frac{\mu_0 c I_0^2 (k d)^2}{48\pi}. \quad (9.22)$$

In order to maintain the radiation, power must be supplied continuously to the dipole from some generator. By analogy with the heating power produced in a resistor,

$$\langle P \rangle_{\text{heat}} = \langle I^2 \rangle R = \frac{I_0^2 R}{2}, \quad (9.23)$$

we can define the factor which multiplies $I_0^2/2$ in Equation (9.22) as the *radiation resistance* of the dipole antenna. Hence,

$$R_{\text{rad}} = \sqrt{\frac{\mu_0}{\epsilon_0}} \frac{(k d)^2}{24\pi} = 197 \left(\frac{d}{\lambda} \right)^2 \text{ ohms}. \quad (9.24)$$

Because we have assumed that $\lambda \gg d$, this radiation resistance is necessarily small. Typically, in a Hertzian dipole, the radiated power is swamped by ohmic losses that appear as heat. Thus, a “short” antenna is a very inefficient radiator. Practical antennas have dimensions that are comparable with the wavelength of the emitted radiation.

Probably the two most common practical antennas are the *half-wave antenna* ($kd = \pi$) and the *full-wave antenna* ($kd = 2\pi$). In the former case, Equation (9.19) reduces to

$$\frac{dP}{d\Omega} = \frac{\mu_0 c I^2 \cos^2(\pi \cos \theta/2)}{8\pi^2 \sin^2 \theta}. \quad (9.25)$$

In the latter case, Equation (9.19) yields

$$\frac{dP}{d\Omega} = \frac{\mu_0 c I^2 \cos^4(\pi \cos \theta/2)}{2\pi^2 \sin^2 \theta}. \quad (9.26)$$

The half-wave antenna radiation pattern is very similar to the characteristic $\sin^2 \theta$ pattern of a Hertzian dipole. However, the full-wave antenna radiation pattern is considerably sharper (i.e., it is more concentrated in the transverse directions $\theta = \pm\pi/2$).

The total power radiated by a half-wave antenna is

$$P = \frac{\mu_0 c I^2}{4\pi} \int_0^\pi \frac{\cos^2(\pi \cos \theta/2)}{\sin \theta} d\theta. \quad (9.27)$$

The integral can be evaluated numerically to give 1.2188. Thus,

$$P = 1.2188 \frac{\mu_0 c I^2}{4\pi}. \quad (9.28)$$

Note, from Equation (9.15), that I is equivalent to the peak current flowing in the antenna. Thus, the radiation resistance of a half-wave antenna is given by $P/(I^2/2)$, or

$$R_{\text{rad}} = \frac{0.6094}{\pi} \sqrt{\frac{\mu_0}{\epsilon_0}} = 73 \text{ ohms}. \quad (9.29)$$

This resistance is substantially larger than that of a Hertzian dipole [see Equation (9.24)]. In other words, a half-wave antenna is a far more efficient emitter of electromagnetic radiation than a Hertzian dipole. According to standard transmission line theory, if a transmission line is terminated by a resistor whose resistance matches the characteristic impedance of the line then all of the power transmitted down the line is dissipated in the resistor. On the other hand, if the resistance does not match the impedance of the line then some of the power is reflected and returned to the generator. We can think of a half-wave antenna, centre-fed by a transmission line, as a 73 ohm resistor terminating the line. The only difference is that the power absorbed from the line is radiated rather than dissipated as heat. Thus, in order to avoid problems with reflected power, the impedance of a transmission line feeding a half-wave antenna must be 73 ohms. Not surprisingly, 73 ohm impedance is one of the standard ratings for the co-axial cables used in amateur radio.

9.3 Antenna Directivity and Effective Area

We have seen that standard antennas emit more radiation in some directions than in others. Indeed, it is topologically impossible for an antenna to emit transverse waves uniformly in all directions

(for the same reason that it is impossible to comb the hair on a sphere in such a manner that there is no parting). One of the aims of antenna engineering is to design antennas that transmit most of their radiation in a particular direction. By a reciprocity argument, such an antenna, when used as a receiver, is preferentially sensitive to radiation incident from the same direction.

The *directivity* or *gain* of an antenna is defined as the ratio of the maximum value of the power radiated per unit solid angle to the average power radiated per unit solid angle: that is,

$$G = \frac{(dP/d\Omega)_{\max}}{P/4\pi}. \quad (9.30)$$

Thus, the directivity measures how much more intensely the antenna radiates in its preferred direction than a mythical “isotropic radiator” would when fed with the same total power. For a Hertzian dipole, the gain is $3/2$. For a half-wave antenna, the gain is 1.64. To achieve a directivity that is significantly greater than unity, the antenna size needs to be much larger than the wavelength. This is usually achieved using a phased array of half-wave, or full-wave, antennas.

Antennas can be used to receive, as well as emit, electromagnetic radiation. The incoming wave induces a voltage that can be detected in an electrical circuit connected to the antenna. In fact, this process is equivalent to the emission of electromagnetic waves by the antenna viewed in reverse. In the theory of electrical circuits, a receiving antenna is represented as an emf connected in series with a resistor. The emf, $V_0 \cos(\omega t)$, represents the voltage induced in the antenna by the incoming wave. The resistor, R_{rad} , represents the power re-radiated by the antenna (here, the real resistance of the antenna is neglected). Let us represent the detector circuit as a single load resistor R_{load} connected in series with the antenna. How can we choose R_{load} such that the maximum power is extracted from the incoming wave and transmitted to the load resistor? According to Ohm’s law,

$$V_0 \cos(\omega t) = I_0 \cos(\omega t) (R_{\text{rad}} + R_{\text{load}}), \quad (9.31)$$

where $I = I_0 \cos(\omega t)$ is the current induced in the circuit. The power input to the circuit is

$$P_{\text{in}} = \langle VI \rangle = \frac{V_0^2}{2(R_{\text{rad}} + R_{\text{load}})}. \quad (9.32)$$

The power transferred to the load is

$$P_{\text{load}} = \langle I^2 R_{\text{load}} \rangle = \frac{R_{\text{load}} V_0^2}{2(R_{\text{rad}} + R_{\text{load}})^2}. \quad (9.33)$$

Finally, the power re-radiated by the antenna is

$$P_{\text{rad}} = \langle I^2 R_{\text{rad}} \rangle = \frac{R_{\text{rad}} V_0^2}{2(R_{\text{rad}} + R_{\text{load}})^2}. \quad (9.34)$$

Note that $P_{\text{in}} = P_{\text{load}} + P_{\text{rad}}$. The maximum power transfer to the load occurs when

$$\frac{\partial P_{\text{load}}}{\partial R_{\text{load}}} = \frac{V_0^2}{2} \left[\frac{R_{\text{load}} - R_{\text{rad}}}{(R_{\text{rad}} + R_{\text{load}})^3} \right] = 0. \quad (9.35)$$

Thus, the maximum transfer rate corresponds to

$$R_{\text{load}} = R_{\text{res}}. \quad (9.36)$$

In other words, the resistance of the load circuit must match the radiation resistance of the antenna. For this optimum case,

$$P_{\text{load}} = P_{\text{rad}} = \frac{V_0^2}{8 R_{\text{rad}}} = \frac{P_{\text{in}}}{2}. \quad (9.37)$$

So, even in the optimum case, half of the power absorbed by the antenna is immediately re-radiated. If $R_{\text{load}} \neq R_{\text{res}}$ then more than half of the absorbed power is re-radiated. Clearly, an antenna that is receiving electromagnetic radiation is also emitting it. This is how the BBC (allegedly) catch people who do not pay their television license fee in the UK. They have vans that can detect the radiation emitted by a TV aerial while it is in use (they can even tell which channel you are watching!).

For a Hertzian dipole antenna interacting with an incoming wave whose electric field has an amplitude E_0 , we expect

$$V_0 = E_0 d/2. \quad (9.38)$$

Here, we have used the fact that the wavelength of the radiation is much longer than the length of the antenna, and that the relevant emf develops between the two ends and the centre of the antenna. We have also assumed that the antenna is properly aligned (i.e., the radiation is incident perpendicular to the axis of the antenna). The Poynting flux of the incoming wave is

$$\langle u_{\text{in}} \rangle = \frac{\epsilon_0 c E_0^2}{2}, \quad (9.39)$$

whereas the power transferred to a properly matched detector circuit is

$$P_{\text{load}} = \frac{E_0^2 d^2}{32 R_{\text{rad}}}. \quad (9.40)$$

Consider an idealized antenna in which all incoming radiation incident on some area A_{eff} is absorbed, and then magically transferred to the detector circuit with no re-radiation. Suppose that the power absorbed from the idealized antenna matches that absorbed from the real antenna. This implies that

$$P_{\text{load}} = \langle u_{\text{in}} \rangle A_{\text{eff}}. \quad (9.41)$$

The quantity A_{eff} , which is called the *effective area* of an antenna, is the area of the idealized antenna that absorbs as much net power from the incoming wave as the actual antenna. Alternatively, A_{eff} is the area of the incoming wavefront that is captured by the receiving antenna and fed to its load circuit. Thus,

$$P_{\text{load}} = \frac{E_0^2 d^2}{32 R_{\text{rad}}} = \frac{\epsilon_0 c E_0^2}{2} A_{\text{eff}}, \quad (9.42)$$

giving

$$A_{\text{eff}} = \frac{d^2}{16 \epsilon_0 c R_{\text{rad}}} = \frac{3}{8\pi} \lambda^2. \quad (9.43)$$

It is clear that the effective area of a Hertzian dipole antenna is of order the wavelength squared of the incoming radiation.

We can generalize from this analysis of a special case. The directivity of a Hertzian dipole is $3/2$. Thus, the effective area of the *isotropic radiator* (the mythical reference antenna against which directivities are measured) is

$$A_0 = \frac{2}{3} A_{\text{hd}} = \frac{\lambda^2}{4\pi}, \quad (9.44)$$

or

$$A_0 = \pi \lambda^2, \quad (9.45)$$

where $\lambda = \lambda/2\pi$. Here, we have used the formal definition of the effective area of an antenna: A_{eff} is that area which, when multiplied by the time-averaged Poynting flux of the incoming wave, equals the maximum power received by the antenna (when its orientation is optimal). Clearly, the effective area of an isotropic radiator is the same as the area of a circle whose radius is the reduced wavelength, λ .

We can take yet one more step, and conclude that the effective area of any antenna of directivity G is

$$A_{\text{eff}} = G \pi \lambda^2. \quad (9.46)$$

Of course, to realize this full capture area, the antenna must be orientated properly.

Let us calculate the coupling, or *insertion loss*, of an antenna-to-antenna communications link. Suppose that a generator delivers the power P_{in} to a transmitting antenna, which is aimed at a receiving antenna a distance r away. The (properly aligned) receiving antenna then captures and delivers the power P_{out} to its load circuit. From the definition of directivity, the transmitting antenna produces the time-averaged Poynting flux

$$\langle u \rangle = G_t \frac{P_{\text{in}}}{4\pi r^2} \quad (9.47)$$

at the receiving antenna. The received power is

$$P_{\text{out}} = \langle u \rangle G_r A_0. \quad (9.48)$$

Here, G_t is the gain of the transmitting antenna, and G_r is the gain of the receiving antenna. Thus,

$$\frac{P_{\text{out}}}{P_{\text{in}}} = G_t G_r \left(\frac{\lambda}{4\pi r} \right)^2 = \frac{A_t A_r}{\lambda^2 r^2}, \quad (9.49)$$

where A_t and A_r are the effective areas of the transmitting and receiving antennas, respectively. This result is known as the *Friis transmission formula*. Note that the insertion loss depends on the product of the gains of the two antennas. Thus, a properly aligned communications link has the same insertion loss operating in either direction.

A thin wire linear antenna might appear to be essentially one-dimensional. However, the concept of an effective area shows that it possesses a second dimension determined by the wavelength. For instance, for a half-wave antenna, the gain of which is 1.64, the effective area is

$$A_{\text{eff}} = 1.64 A_0 = \frac{\lambda}{2} (0.26 \lambda). \quad (9.50)$$

Thus, we can visualize the capture area as a rectangle that is the physical length of the antenna in one direction, and approximately one quarter of the wavelength in the other.

9.4 Antenna Arrays

Consider a linear array of N half-wave antennas arranged along the x -axis with a uniform spacing Δ . Suppose that each antenna is aligned along the z -axis, and also that all antennas are driven in phase. Let one end of the array coincide with the origin. The field produced in the radiation zone by the end-most antenna is given by [see Equation (9.17)]

$$\mathbf{A}(\mathbf{r}) = \frac{\mu_0 I}{4\pi} \frac{2 \cos(\pi \cos \theta/2)}{k r \sin^2 \theta} e^{i(kr - \omega t)} \mathbf{e}_z, \quad (9.51)$$

where I is the peak current flowing in each antenna. The fields produced at a given point in the radiation zone by successive elements of the array differ in phase by an amount $\alpha = k \Delta \sin \theta \cos \varphi$. Here, r, θ, φ are conventional spherical polar coordinates. Thus, the total field is given by

$$\mathbf{A}(\mathbf{r}) = \frac{\mu_0 I}{4\pi} \frac{2 \cos(\pi \cos \theta/2)}{k r \sin^2 \theta} \left[1 + e^{i\alpha} + e^{2i\alpha} + \dots + e^{(N-1)i\alpha} \right] e^{i(kr - \omega t)} \mathbf{e}_z. \quad (9.52)$$

The series in square brackets is a geometric progression in $\beta = \exp(i\alpha)$, the sum of which takes the value

$$1 + \beta + \beta^2 + \dots + \beta^{N-1} = \frac{\beta^N - 1}{\beta - 1}. \quad (9.53)$$

Thus, the term in square brackets becomes

$$\frac{e^{iN\alpha} - 1}{e^{i\alpha} - 1} = e^{i(N-1)\alpha/2} \frac{\sin(N\alpha/2)}{\sin(\alpha/2)}. \quad (9.54)$$

It follows from Equation (9.18) that the radiation pattern due to the array takes the form

$$\frac{dP}{d\Omega} = \left[\frac{\mu_0 c I^2 \cos^2(\pi \cos \theta/2)}{8\pi^2 \sin^2 \theta} \right] \left[\frac{\sin^2(N\alpha/2)}{\sin^2(\alpha/2)} \right]. \quad (9.55)$$

We can think of this formula as the product of the two factors in large parentheses. The first is just the standard radiation pattern of a half-wave antenna. The second arises from the arrangement of the array. If we retained the same array, but replaced the elements by something other than half-wave antennas, then the first factor would change, but not the second. If we changed the array, but not the elements, then the second factor would change, but the first would remain the same. Thus, the radiation pattern is the product of two independent factors, the *element function*, and the *array function*. This independence follows from the Fraunhofer approximation, (9.7), which justifies the linear phase shifts of Equation (9.8).

The array function in the present case is

$$f(\alpha) = \frac{\sin^2(N\alpha/2)}{\sin^2(\alpha/2)}, \quad (9.56)$$

where

$$\alpha = k \Delta \sin \theta \cos \varphi. \quad (9.57)$$

The function $f(\alpha)$ has nulls whenever the numerator vanishes: that is, whenever

$$\pm\alpha = \frac{2\pi}{N}, \frac{4\pi}{N}, \dots, \frac{(N-1)2\pi}{N}, \frac{(N+1)2\pi}{N}, \dots. \quad (9.58)$$

However, when $\pm\alpha = 0, 2\pi, \dots$, the denominator also vanishes, and the l'Hôpital limit is easily seen to be $f(0, 2\pi, \dots) \rightarrow N^2$. These limits are known as the *principal maxima* of the function. Secondary maxima occur approximately at the maxima of the numerator: that is, at

$$\pm\alpha = \frac{3\pi}{N}, \frac{5\pi}{N}, \dots, \frac{(2N-3)2\pi}{N}, \frac{(2N+3)2\pi}{N}, \dots. \quad (9.59)$$

There are $(N-2)$ secondary maxima between successive principal maxima.

Now, the maximum possible value of α is $k\Delta = 2\pi\Delta/\lambda$. Thus, when the element spacing Δ is less than the wavelength there is only one principal maximum (at $\alpha = 0$), directed perpendicular to the array (i.e., at $\varphi = \pm\pi/2$). Such a system is called a *broadside array*. The secondary maxima of the radiation pattern are called *side lobes*. In the direction perpendicular to the array, all elements contribute in phase, and the intensity is proportional to the square of the sum of the individual amplitudes. Thus, the peak intensity for an N element array is N^2 times the intensity of a single antenna. The angular half-width of the principle maximum (in φ) is approximately $\Delta\varphi \simeq \lambda/N\Delta$. Although the principal lobe clearly gets narrower in the azimuthal angle φ as N increases, the lobe width in the polar angle θ is mainly controlled by the element function, and is thus little affected by the number of elements. A radiation pattern which is narrow in one angular dimension, but broad in the other, is called a *fan beam*.

Arranging a set of antennas in a regular array has the effect of taking the azimuthally symmetric radiation pattern of an individual antenna and concentrating it into some narrow region of azimuthal angle of extent $\Delta\varphi \simeq \lambda/N\Delta$. The net result is that the gain of the array is larger than that of an individual antenna by a factor of order

$$\frac{2\pi N \Delta}{\lambda}. \quad (9.60)$$

It is clear that the boost factor is of order the linear extent of the array divided by the wavelength of the emitted radiation. Thus, it is possible to construct a very high gain antenna by arranging a large number of low gain antennas in a regular pattern, and driving them in phase. The optimum spacing between successive elements of the array is of order the wavelength of the radiation.

A linear array of antenna elements that are spaced $\Delta = \lambda/2$ apart, and driven with alternating phases, has its principal radiation maximum along $\varphi = 0$ and π , because the field amplitudes now add in phase in the plane of the array. Such a system is called an *end-fire array*. The direction of the principal maximum can be changed at will by introducing the appropriate phase shift between successive elements of the array. In fact, it is possible to produce a radar beam that sweeps around the horizon, without any mechanical motion of the array, by varying the phase difference between successive elements of the array electronically.

9.5 Thomson Scattering

When an electromagnetic wave is incident on a charged particle, the electric and magnetic components of the wave exert a Lorentz force on the particle, setting it into motion. Because the wave is periodic in time, so is the motion of the particle. Thus, the particle is accelerated and, consequently, emits radiation. More exactly, energy is absorbed from the incident wave by the particle, and re-emitted as electromagnetic radiation. Such a process is clearly equivalent to the scattering of the electromagnetic wave by the particle.

Consider a linearly polarized, monochromatic, plane wave incident on a particle of charge q . The electric component of the wave can be written

$$\mathbf{E} = \mathbf{e} E_0 e^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)}, \quad (9.61)$$

where E_0 is the peak amplitude of the electric field, \mathbf{e} is the polarization vector, and \mathbf{k} is the wave vector (of course, $\mathbf{e} \cdot \mathbf{k} = 0$). The particle is assumed to undergo small amplitude oscillations about an equilibrium position that coincides with the origin of the coordinate system. Furthermore, the particle's velocity is assumed to remain sub-relativistic, which enables us to neglect the magnetic component of the Lorentz force. The equation of motion of the charged particle is approximately

$$\mathbf{f} = q \mathbf{E} = m \ddot{\mathbf{s}}, \quad (9.62)$$

where m is the mass of the particle, \mathbf{s} is its displacement from the origin, and $\dot{}$ denotes $\partial/\partial t$. By analogy with Equation (9.20), the time-averaged power radiated per unit solid angle by an accelerating, non-relativistic, charged particle is given by

$$\frac{dP}{d\Omega} = \frac{q^2 \langle \ddot{\mathbf{s}}^2 \rangle}{16\pi^2 \epsilon_0 c^3} \sin^2 \theta, \quad (9.63)$$

where $\langle \dots \rangle$ denotes a time average. Here, we are effectively treating the oscillating particle as a short antenna. However,

$$\langle \ddot{\mathbf{s}}^2 \rangle = \frac{q^2}{m^2} \langle E^2 \rangle = \frac{q^2 E_0^2}{2m^2}. \quad (9.64)$$

Hence, the scattered power per unit solid angle becomes

$$\frac{dP}{d\Omega} = \left(\frac{q^2}{4\pi \epsilon_0 m c^2} \right)^2 \frac{\epsilon_0 c E_0^2}{2} \sin^2 \theta. \quad (9.65)$$

The time-averaged Poynting flux of the incident wave is

$$\langle u \rangle = \frac{\epsilon_0 c E_0^2}{2}. \quad (9.66)$$

It is convenient to define the *scattering cross-section* as the equivalent area of the incident wavefront that delivers the same power as that re-radiated by the particle: that is,

$$\sigma = \frac{\text{total re-radiated power}}{\langle u \rangle}. \quad (9.67)$$

By analogy, the *differential scattering cross-section* is defined

$$\frac{d\sigma}{d\Omega} = \frac{dP/d\Omega}{\langle u \rangle}. \quad (9.68)$$

It follows from Equations (9.65) and (9.66) that

$$\frac{d\sigma}{d\Omega} = \left(\frac{q^2}{4\pi \epsilon_0 m c^2} \right)^2 \sin^2 \theta. \quad (9.69)$$

The total scattering cross-section is then

$$\sigma = \int_0^\pi \frac{d\sigma}{d\Omega} 2\pi \sin \theta d\theta = \frac{8\pi}{3} \left(\frac{q^2}{4\pi \epsilon_0 m c^2} \right)^2. \quad (9.70)$$

The quantity θ , appearing in Equation (9.69), is the angle subtended between the direction of acceleration of the particle, and the direction of the outgoing radiation (which is parallel to the unit vector \mathbf{n}). In the present case, the acceleration is due to the electric field, so it is parallel to the polarization vector \mathbf{e} . Thus, $\cos \theta = \mathbf{e} \cdot \mathbf{n}$.

Up to now, we have only considered the scattering of linearly polarized radiation by a charged particle. Let us now calculate the angular distribution of scattered radiation for the commonly occurring case of randomly polarized incident radiation. It is helpful to set up a right-handed coordinate system based on the three mutually orthogonal unit vectors \mathbf{e} , $\mathbf{e} \times \hat{\mathbf{k}}$, and $\hat{\mathbf{k}}$, where $\hat{\mathbf{k}} = \mathbf{k}/k$. In terms of these unit vectors, we can write

$$\mathbf{n} = \sin \varphi \cos \psi \mathbf{e} + \sin \varphi \sin \psi \mathbf{e} \times \hat{\mathbf{k}} + \cos \varphi \hat{\mathbf{k}}, \quad (9.71)$$

where φ is the angle subtended between the direction of the incident radiation and that of the scattered radiation, and ψ is an angle that specifies the orientation of the polarization vector in the plane perpendicular to \mathbf{k} (assuming that \mathbf{n} is known). It is easily seen that

$$\cos \theta = \mathbf{e} \cdot \mathbf{n} = \cos \psi \sin \varphi, \quad (9.72)$$

so

$$\sin^2 \theta = 1 - \cos^2 \psi \sin^2 \varphi. \quad (9.73)$$

Averaging this result over all possible polarizations of the incident wave (i.e., over all possible values of the polarization angle ψ), we obtain

$$\overline{\sin^2 \theta} = 1 - \overline{\cos^2 \psi} \sin^2 \varphi = 1 - (\sin^2 \varphi)/2 = \frac{1 + \cos^2 \varphi}{2}. \quad (9.74)$$

Thus, the differential scattering cross-section for unpolarized incident radiation [obtained by substituting $\overline{\sin^2 \theta}$ for $\sin^2 \theta$ in Eq. (9.69)] is given by

$$\left(\frac{d\sigma}{d\Omega} \right)_{\text{unpolarized}} = \left(\frac{q^2}{4\pi \epsilon_0 m c^2} \right)^2 \left(\frac{1 + \cos^2 \varphi}{2} \right). \quad (9.75)$$

It is clear that the differential scattering cross-section is independent of the frequency of the incident wave, and is also symmetric with respect to forward and backward scattering. Moreover, the frequency of the scattered radiation is the same as that of the incident radiation. The total scattering cross-section is obtained by integrating over the entire solid angle of the polar angle φ and the azimuthal angle ψ . Not surprisingly, the result is exactly the same as Equation (9.70).

The classical scattering cross-section (9.75) is modified by quantum effects when the energy of the incident photons, $\hbar\omega$, becomes comparable with the rest mass of the scattering particle, mc^2 . The scattering of a photon by a charged particle is called *Compton scattering*, and the quantum mechanical version of the Compton scattering cross-section is known as the *Klein-Nishina cross-section*. As the photon energy increases, and eventually becomes comparable with the rest mass energy of the particle, the Klein-Nishina formula predicts that forward scattering of photons becomes increasingly favored with respect to backward scattering. The Klein-Nishina cross-section does, in general, depend on the frequency of the incident photons. Furthermore, energy and momentum conservation demand a shift in the frequency of scattered photons with respect to that of the incident photons.

If the charged particle in question is an electron then Equation (9.70) reduces to the well-known *Thomson scattering cross-section*

$$\sigma_{\text{Thomson}} = \frac{8\pi}{3} \left(\frac{e^2}{4\pi\epsilon_0 m_e c^2} \right)^2 = 6.65 \times 10^{-29} \text{ m}^2. \quad (9.76)$$

The quantity $e^2/(4\pi\epsilon_0 m_e c^2) = 2.8 \times 10^{-15} \text{ m}$ is called the *classical electron radius* (it is the radius of spherical shell of total charge e whose electrostatic energy equals the rest mass energy of the electron). Thus, when scattering radiation, the electron acts rather like a solid sphere whose radius is of order the classical electron radius.

9.6 Rayleigh Scattering

Let us now consider the scattering of electromagnetic radiation by a harmonically bound electron: for instance, an electron orbiting an atomic nucleus. We have seen in Section 7.3 that such an electron satisfies an equation of motion of the form

$$\ddot{\mathbf{s}} + \gamma_0 \dot{\mathbf{s}} + \omega_0^2 \mathbf{s} = -\frac{e}{m_e} \mathbf{E}, \quad (9.77)$$

where ω_0 is the characteristic oscillation frequency of the electron, and $\gamma_0 \ll \omega_0$ is the damping rate of such oscillations. Assuming an $e^{-i\omega t}$ time dependence of both \mathbf{s} and \mathbf{E} , we find that

$$\ddot{\mathbf{s}} = \frac{\omega^2}{\omega_0^2 - \omega^2 - i\gamma_0\omega} \frac{e}{m_e} \mathbf{E}. \quad (9.78)$$

It follows, by analogy with the analysis in the previous section, that the total scattering cross-section is given by

$$\sigma = \sigma_{\text{Thomson}} \frac{\omega^4}{(\omega_0^2 - \omega^2)^2 + (\gamma_0\omega)^2}. \quad (9.79)$$

The angular distribution of the radiation is the same as that in the case of a free electron.

The maximum value of the cross-section (9.79) is obtained when $\omega \simeq \omega_0$: that is, for resonant scattering. In this case, the scattering cross-section can become very large. In fact,

$$\sigma \simeq \sigma_{\text{Thomson}} \left(\frac{\omega_0}{\gamma_0} \right)^2, \quad (9.80)$$

which is generally far greater than the Thomson scattering cross-section.

For the case of strong binding, $\omega \ll \omega_0$, and Equation (9.79) reduces to

$$\sigma \simeq \sigma_{\text{Thomson}} \left(\frac{\omega}{\omega_0} \right)^4, \quad (9.81)$$

giving a scattering cross-section that depends on the inverse fourth power of the wavelength of the incident radiation. The cross-section (9.81) is known as the *Rayleigh scattering cross-section*, and is appropriate to the scattering of visible radiation by gas molecules. This is the basis of Rayleigh's famous explanation of the blue sky. The air molecules of the atmosphere preferentially scatter the shorter wavelength blue components out of "white" sunlight which grazes the atmosphere. Conversely, sunlight viewed directly through the long atmospheric path at sunset appears reddened. The Rayleigh scattering cross-section is much less than the Thomson scattering cross-section (for $\omega \ll \omega_0$). However, this effect is offset to some extent by the fact that the density of neutral molecules in a gas (e.g., the atmosphere) is much larger than the density of free electrons typically encountered in a plasma.

9.7 Exercises

- 9.1 Consider an electromagnetic wave propagating through a non-dielectric, non-magnetic medium containing free charge density ρ and free current density \mathbf{j} . Demonstrate from Maxwell's equations that the associated wave equations take the form

$$\begin{aligned} \nabla^2 \mathbf{E} - \frac{1}{c^2} \frac{\partial^2 \mathbf{E}}{\partial t^2} &= \frac{\nabla \rho}{\epsilon_0} + \mu_0 \frac{\partial \mathbf{j}}{\partial t}, \\ \nabla^2 \mathbf{B} - \frac{1}{c^2} \frac{\partial^2 \mathbf{B}}{\partial t^2} &= -\mu_0 \nabla \times \mathbf{j}. \end{aligned}$$

- 9.2 A spherically symmetric charge distribution undergoes purely radial oscillations. Show that no electromagnetic waves are emitted. [Hint: Show that there is no magnetic field.]

10 Resonant Cavities and Waveguides

10.1 Introduction

Let us briefly investigate the solution of the homogeneous wave equation in bounded regions; particularly in regions bounded by conductors. This type of boundary value problem is of great theoretical significance, and also has many practical applications.

10.2 Boundary Conditions

The general boundary conditions on the field vectors at an interface between medium 1 and medium 2 (say) are

$$\mathbf{n} \cdot (\mathbf{D}_1 - \mathbf{D}_2) = \tau, \quad (10.1)$$

$$\mathbf{n} \times (\mathbf{E}_1 - \mathbf{E}_2) = \mathbf{0}, \quad (10.2)$$

$$\mathbf{n} \cdot (\mathbf{B}_1 - \mathbf{B}_2) = 0, \quad (10.3)$$

$$\mathbf{n} \times (\mathbf{H}_1 - \mathbf{H}_2) = \mathbf{K}, \quad (10.4)$$

where τ is used for the interfacial surface charge density (to avoid confusion with the conductivity), and \mathbf{K} is the surface current density. Here, \mathbf{n} is a unit vector normal to the interface, directed from medium 2 to medium 1. We saw in Section 7.4 that, at normal incidence, the amplitude of an electromagnetic wave falls off very rapidly with distance inside the surface of a good conductor. In the limit of perfect conductivity (i.e., $\sigma \rightarrow \infty$), the wave does not penetrate into the conductor at all, in which case the internal tangential electric and magnetic fields vanish. This implies, from Equations (10.2) and (10.4), that the tangential component of \mathbf{E} vanishes just outside the surface of a good conductor, whereas the tangential component of \mathbf{H} may remain finite. Let us examine the behavior of the normal field components.

Let medium 1 be a conductor, of conductivity σ and dielectric constant ϵ_1 , for which $\sigma/\epsilon_1 \epsilon_0 \omega \gg 1$, and let medium 2 be a perfect insulator of dielectric constant ϵ_2 . The charge density that forms at the interface between the two media is related to the currents flowing inside the conductor. In fact, the conservation of charge requires that

$$\mathbf{n} \cdot \mathbf{j} = \frac{\partial \tau}{\partial t} = -i \omega \tau. \quad (10.5)$$

However, $\mathbf{n} \cdot \mathbf{j} = \mathbf{n} \cdot \sigma \mathbf{E}_1$, so it follows from Equation (10.1) that

$$\left(1 + \frac{i \omega \epsilon_0 \epsilon_1}{\sigma}\right) \mathbf{n} \cdot \mathbf{E}_1 = \frac{i \omega \epsilon_0 \epsilon_2}{\sigma} \mathbf{n} \cdot \mathbf{E}_2. \quad (10.6)$$

Thus, it is clear that the normal component of \mathbf{E} within the conductor also becomes vanishingly small as the conductivity approaches infinity.

If \mathbf{E} vanishes inside a perfect conductor then the curl of \mathbf{E} also vanishes, and the time rate of change of \mathbf{B} is correspondingly zero. This implies that there are no oscillatory fields whatever inside such a conductor, and that the fields just outside satisfy

$$\mathbf{n} \cdot \mathbf{D} = -\tau, \quad (10.7)$$

$$\mathbf{n} \times \mathbf{E} = \mathbf{0}, \quad (10.8)$$

$$\mathbf{n} \cdot \mathbf{B} = 0, \quad (10.9)$$

$$\mathbf{n} \times \mathbf{H} = -\mathbf{K}. \quad (10.10)$$

Here, \mathbf{n} is a unit normal at the surface of the conductor pointing into the conductor. Thus, the electric field is normal, and the magnetic field tangential, at the surface of a perfect conductor. For good conductors, these boundary conditions yield excellent representations of the geometrical configurations of the external fields, but they lead to the neglect of some important features of real fields, such as losses in cavities and signal attenuation in waveguides.

In order to estimate such losses, it is helpful to examine how the tangential and normal fields compare when σ is large but finite. Equations (7.6) and (7.41) imply that

$$\mathbf{H} = e^{i\pi/4} \sqrt{\frac{\sigma}{\mu_0 \omega}} \mathbf{n} \times \mathbf{E} \quad (10.11)$$

at the surface of a good conductor (provided that the wave propagates into the conductor). Let us assume, without obtaining a complete solution, that a wave with \mathbf{H} very nearly tangential and \mathbf{E} very nearly normal propagates parallel to the surface of a good conductor. According to the Faraday-Maxwell equation,

$$|H_{\parallel}| \simeq \frac{k}{\mu_0 \omega} |E_{\perp}| \quad (10.12)$$

just outside the surface, where k is the component of the wavevector parallel to the surface. However, Equation (10.11) implies that a tangential component of \mathbf{H} is accompanied by a small tangential component of \mathbf{E} . By comparing the previous two expressions, we obtain

$$\frac{|E_{\parallel}|}{|E_{\perp}|} \simeq k \sqrt{\frac{2}{\mu_0 \omega \sigma}} = \frac{d}{\lambda}, \quad (10.13)$$

where d is the skin depth [see Equation (7.43)] and $\lambda \equiv 1/k$. It is clear that the ratio of the tangential to the normal component of \mathbf{E} is of order the skin depth divided by the wavelength. It is readily demonstrated that the ratio of the normal to the tangential component of \mathbf{H} is of the same order of magnitude. Thus, we deduce that, in the limit of high conductivity, which implies vanishing skin depth, no fields penetrate into the conductor, and the boundary conditions are those given by Equations (10.7)–(10.10). Let us investigate the solution of the homogeneous wave equation subject to such constraints.

10.3 Cavities with Rectangular Boundaries

Consider a rectangular vacuum region totally enclosed by conducting walls. In this case, all of the field components satisfy the wave equation

$$\nabla^2 \psi - \frac{1}{c^2} \frac{\partial^2 \psi}{\partial t^2} = 0, \quad (10.14)$$

where ψ represents any component of \mathbf{E} or \mathbf{H} . The boundary conditions (10.7)–(10.10) require that the electric field at the boundary be normal to the conducting walls, whereas the magnetic field be tangential. If a , b , and c are the dimensions of the cavity, in the x , y , and z directions, respectively, then it is readily verified that the electric field components are

$$E_x(x, y, z, t) = E_1 \cos(k_1 x) \sin(k_2 y) \sin(k_3 z) e^{-i\omega t}, \quad (10.15)$$

$$E_y(x, y, z, t) = E_2 \sin(k_1 x) \cos(k_2 y) \sin(k_3 z) e^{-i\omega t}, \quad (10.16)$$

$$E_z(x, y, z, t) = E_3 \sin(k_1 x) \sin(k_2 y) \cos(k_3 z) e^{-i\omega t}, \quad (10.17)$$

where

$$k_1 = \frac{l\pi}{a}, \quad (10.18)$$

$$k_2 = \frac{m\pi}{b}, \quad (10.19)$$

$$k_3 = \frac{n\pi}{c}. \quad (10.20)$$

Here, l , m , n are non-negative integers. The allowed frequencies are given by

$$\frac{\omega^2}{c^2} = k_1^2 + k_2^2 + k_3^2 = \pi^2 \left(\frac{l^2}{a^2} + \frac{m^2}{b^2} + \frac{n^2}{c^2} \right). \quad (10.21)$$

It is clear from Equations (10.15)–(10.17) that at least two of the integers l , m , n must be different from zero in order to have non-vanishing fields. The magnetic fields, obtained by solving $\nabla \times \mathbf{E} = i\omega \mathbf{B}$, automatically satisfy the appropriate boundary conditions, and are in phase quadrature with the corresponding electric fields. Thus, the sum of the total electric and magnetic energies within the cavity is constant, although the two terms oscillate separately.

The amplitudes of the electric field components are not independent, but are related by the divergence condition $\nabla \cdot \mathbf{E} = 0$, which yields

$$k_1 E_1 + k_2 E_2 + k_3 E_3 = 0. \quad (10.22)$$

There are, in general, two linearly independent vectors \mathbf{E} that satisfy this condition, corresponding to two different polarizations. (The exception is when one of the integers l , m , n is zero, in which case \mathbf{E} is fixed in direction.) Each electric field vector is accompanied by a perpendicular magnetic field vector. The fields corresponding to a given set of integers l , m , and n constitute a particular

mode of oscillation of the cavity. It is evident from standard Fourier theory that the different modes are orthogonal (i.e., they are normal modes), and that they form a complete set. In other words, any general electric and magnetic fields that satisfy the boundary conditions (10.7)–(10.10) can be unambiguously decomposed into some linear combination of all of the various possible normal modes of the cavity. Because each normal mode oscillates at a specific frequency, it is clear that if we are given the electric and magnetic fields inside the cavity at time $t = 0$ then the subsequent behavior of the fields is uniquely determined for all time.

The conducting walls gradually absorb energy from the cavity, due to their finite resistivity, at a rate that can easily be calculated. For finite σ , the small tangential component of \mathbf{E} at the walls can be estimated using Equation (10.11):

$$\mathbf{E}_{\parallel} = e^{-i\pi/4} \sqrt{\frac{\mu_0 \omega}{\sigma}} \mathbf{H}_{\parallel} \times \mathbf{n}. \quad (10.23)$$

Now, the tangential component of \mathbf{H} at the walls is slightly different from that given by the ideal solution. However, this is a small effect, and can be neglected to leading order in σ^{-1} . The time averaged energy flux into the walls is then given by

$$\bar{\mathbf{N}} = \frac{1}{2} \operatorname{Re} (\mathbf{E}_{\parallel} \times \mathbf{H}_{\parallel}) = \frac{1}{2} \sqrt{\frac{\mu_0 \omega}{2\sigma}} H_{\parallel 0}^2 \mathbf{n} = \frac{H_{\parallel 0}^2}{2\sigma d} \mathbf{n}, \quad (10.24)$$

where $H_{\parallel 0}$ is the peak value of the tangential magnetic field at the walls that is predicted by the ideal solution. According to the boundary condition (10.10), $H_{\parallel 0}$ is equal to the peak value of the surface current density K_0 . It is helpful to define a surface resistance,

$$\bar{\mathbf{N}} = \bar{K}^2 R_s \mathbf{n} = \frac{1}{2} K_0^2 R_s \mathbf{n}, \quad (10.25)$$

where

$$R_s = \frac{1}{\sigma d}. \quad (10.26)$$

This approach makes it clear that the dissipation of energy in a resonant cavity is due to ohmic heating in a thin layer, whose thickness is of order the skin depth, covering the surface of the conducting walls.

10.4 Quality Factor of a Resonant Cavity

The quality factor Q of a resonant cavity is defined

$$Q = 2\pi \frac{\text{energy stored in cavity}}{\text{energy lost per cycle to walls}}. \quad (10.27)$$

For a specific normal mode of the cavity, this quantity is independent of the mode amplitude. By conservation of energy, the power dissipated via ohmic losses is minus the rate of change of the

stored energy, U . We can thus write a differential equation for the variation of U as a function of time:

$$\frac{dU}{dt} = -\frac{\omega_0}{Q} U, \quad (10.28)$$

where ω_0 is the oscillation frequency of the normal mode in question. The solution to the above equation is

$$U(t) = U(0) e^{-\omega_0 t/Q}. \quad (10.29)$$

This time dependence of the stored energy suggests that the oscillations of the electromagnetic fields inside the cavity are damped as follows:

$$E(t) = E_0 e^{-\omega_0 t/2Q} e^{-i(\omega_0 + \Delta\omega)t}, \quad (10.30)$$

where we have allowed for a shift $\Delta\omega$ of the resonant frequency, as well as for the damping. A damped oscillation such as that specified above does not consist of a pure frequency. Instead, it is made up of a superposition of frequencies centered on $\omega = \omega_0 + \Delta\omega$. Standard Fourier analysis yields

$$E(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} E(\omega) e^{-i\omega t} d\omega, \quad (10.31)$$

where

$$E(\omega) = \frac{1}{\sqrt{2\pi}} \int_0^{\infty} E_0 e^{-\omega_0 t/2Q} e^{i(\omega - \omega_0 - \Delta\omega)t} dt. \quad (10.32)$$

It follows that

$$|E(\omega)|^2 \propto \frac{1}{(\omega - \omega_0 - \Delta\omega)^2 + (\omega_0/2Q)^2}. \quad (10.33)$$

The above resonance curve has a full width at half-maximum equal to ω_0/Q . For a constant input voltage, the energy of oscillation within the cavity as a function of frequency follows this curve in the neighborhood of a particular resonant frequency. It can be seen that the ohmic losses, which determine Q for a particular mode, also determine the maximum amplitude of the oscillation when the resonance condition is exactly satisfied, as well as the width of the resonance (i.e., how far off the resonant frequency the system can be driven, and still yield a significant oscillation amplitude).

10.5 Axially Symmetric Cavities

The rectangular cavity that we have just discussed has many features in common with axially symmetric cavities of arbitrary cross-section. In every axially symmetric cavity, the allowed values of the wave vector, \mathbf{k} , and thus the allowed frequencies, are determined by the cavity geometry. We have seen that for each set of mode numbers, k_1, k_2, k_3 , in a rectangular cavity, there are, in general, two linearly independent modes: that is, the polarization remains arbitrary. We can take advantage of this fact to classify modes into two types, according to the orientation of the field vectors. Let us choose one type of mode such that the electric field vector lies in the cross-sectional plane, and the other such that the magnetic field vector lies in this plane. This classification into transverse electric (TE) and transverse magnetic (TM) modes turns out to be possible for all axially symmetric

cavities, although the rectangular cavity is unique in having one mode of each kind corresponding to each allowed frequency.

Suppose that the direction of symmetry is along the z -axis, and that the length of the cavity in this direction is L . The boundary conditions at $z = 0$ and $z = L$ demand that the z dependence of wave quantities be either $\sin(k_3 z)$ or $\cos(k_3 z)$, where $k_3 = n\pi/L$. In other words, all wave quantities satisfy

$$\left(\frac{\partial^2}{\partial z^2} + k_3^2\right)\psi = 0, \quad (10.34)$$

as well as

$$(\nabla^2 + k^2)\psi = 0, \quad (10.35)$$

where ψ stands for any component of \mathbf{E} or \mathbf{H} . The field equations

$$\nabla \times \mathbf{E} = i\omega\mu_0\mathbf{H}, \quad (10.36)$$

$$\nabla \times \mathbf{H} = -i\omega\epsilon_0\mathbf{E} \quad (10.37)$$

must also be satisfied.

Let us write each vector and each operator in the above equations as the sum of a transverse part, designated by the subscript s , and a component along z . We find that for the transverse fields

$$i\omega\mu_0\mathbf{H}_s = \nabla_s \times \mathbf{E}_z + \nabla_z \times \mathbf{E}_s, \quad (10.38)$$

$$-i\omega\epsilon_0\mathbf{E}_s = \nabla_s \times \mathbf{H}_z + \nabla_z \times \mathbf{H}_s. \quad (10.39)$$

When one of Equations (10.38)–(10.39) is used to substitute for the transverse field on the right-hand side of the other, and use is made of Equation (10.34), we obtain

$$\mathbf{E}_s = \frac{\nabla_s(\partial E_z/\partial z)}{k^2 - k_3^2} + \frac{i\omega\mu_0}{k^2 - k_3^2} \nabla_s \times \mathbf{H}_z, \quad (10.40)$$

$$\mathbf{H}_s = \frac{\nabla_s(\partial H_z/\partial z)}{k^2 - k_3^2} - \frac{i\omega\epsilon_0}{k^2 - k_3^2} \nabla_s \times \mathbf{E}_z. \quad (10.41)$$

Thus, all transverse fields can be expressed in terms of the z components of the fields, each of which satisfies the differential equation

$$\left[\nabla_s^2 + (k^2 - k_3^2)\right]A_z = 0, \quad (10.42)$$

where A_z stands for either E_z or H_z , and ∇_s^2 is the two-dimensional Laplacian operator in the transverse plane.

The conditions on E_z and H_z at the boundary (in the transverse plane) are quite different: E_z must vanish on the boundary, whereas the normal derivative of H_z must vanish to ensure that \mathbf{H}_s in Equation (10.41) satisfies the appropriate boundary condition. If the cross-section is a rectangle then these two conditions lead to the same eigenvalues of $(k^2 - k_3^2) = k_s^2 = k_1^2 + k_2^2$, as we have seen. Otherwise, they correspond to two different sets of eigenvalues, one for which E_z is permitted

but $H_z = 0$, and the other where the opposite is true. In every case, it is possible to classify the modes as transverse magnetic or transverse electric. Thus, the field components E_z and H_z play the role of independent potentials, from which the other field components of the TE and TM modes, respectively, can be derived using Equations (10.40)–(10.41).

The mode frequencies are determined by the eigenvalues of Equations (10.34) and (10.42). If we denote the functional dependence of E_z or H_z on the plane cross-section coordinates by $f(x, y)$ then we can write Equation (10.42) as

$$\nabla_s^2 f = -k_s^2 f. \quad (10.43)$$

Let us first show that $k_s^2 > 0$, and, hence, that $k > k_3$. Now,

$$f \nabla_s^2 f = \nabla_s \cdot (f \nabla_s f) - (\nabla_s f)^2. \quad (10.44)$$

It follows that

$$-k_s^2 \int_V f^2 dV + \int_V (\nabla_s f)^2 dV = \int_S f \nabla f \cdot d\mathbf{S}, \quad (10.45)$$

where the integration is over the transverse cross-section, V . If either f or its normal derivative is to vanish on the conducting surface, S , then

$$k_s^2 = \frac{\int_V (\nabla_s f)^2 dV}{\int_V f^2 dV} > 0. \quad (10.46)$$

We have already seen that $k_3 = n\pi/L$. The allowed values of k_s depend both on the geometry of the cross-section, and the nature of the mode.

For TM modes, $H_z = 0$, and the z dependence of E_z is given by $\cos(n\pi z/L)$. Equation (10.43) must be solved subject to the condition that f vanish on the boundaries of the plane cross-section, thus completing the determination of E_z and k . The transverse fields are then given by special cases of Equations (10.40)–(10.41):

$$\mathbf{E}_s = \frac{1}{k_s^2} \nabla_s \frac{\partial E_z}{\partial z}, \quad (10.47)$$

$$\mathbf{H} = \frac{i\omega\epsilon_0}{k_s^2} \mathbf{e}_z \times \nabla_s E_z. \quad (10.48)$$

For TE modes, in which $E_z = 0$, the condition that H_z vanish at the ends of the cylinder demands a $\sin(n\pi z/L)$ dependence on z , and a k_s which is such that the normal derivative of H_z is zero at the walls. Equations (10.40)–(10.41), for the transverse fields, then become

$$\mathbf{H}_s = \frac{1}{k_s^2} \nabla_s \frac{\partial H_z}{\partial z}, \quad (10.49)$$

$$\mathbf{E} = -\frac{i\omega\mu_0}{k_s^2} \mathbf{e}_z \times \nabla_s H_z, \quad (10.50)$$

and the mode determination is complete.

10.6 Cylindrical Cavities

Let us apply the methods of the previous section to the TM modes of a right circular cylinder of radius a . We can write

$$E_z(r, \varphi, z, t) = A f(r, \varphi) \cos(k_3 z) e^{-i\omega t}, \quad (10.51)$$

where $f(r, \varphi)$ satisfies the equation

$$\frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial f}{\partial r} \right) + \frac{1}{r^2} \frac{\partial^2 f}{\partial \varphi^2} + k_s^2 f = 0, \quad (10.52)$$

and r, φ, z are cylindrical coordinates. Let

$$f(r, \varphi) = g(r) e^{im\varphi}. \quad (10.53)$$

It follows that

$$\frac{1}{r} \frac{d}{dr} \left(r \frac{dg}{dr} \right) + \left(k_s^2 - \frac{m^2}{r^2} \right) g = 0, \quad (10.54)$$

or

$$z^2 \frac{d^2 g}{dz^2} + z \frac{dg}{dz} + (z^2 - m^2) g = 0, \quad (10.55)$$

where $z = k_s r$. The above equation can be recognized as Bessel's equation. The independent solutions of this equation are denoted $J_m(z)$ and $Y_m(z)$. The $J_m(z)$ are regular at $z = 0$, whereas the $Y_m(z)$ are singular. Moreover, both solutions are regular at large $|z|$.

Because the axis ($r = 0$) lies within the cavity, the radial eigenfunction must be regular at $r = 0$. This immediately rules out the $Y_m(k_s r)$ solutions. Thus, the most general solution for a TM mode is

$$E_z(r, \varphi, z, t) = A J_m(k_l r) e^{im\varphi} \cos(k_3 z) e^{-i\omega t}. \quad (10.56)$$

The k_l are the eigenvalues of k_s , and are determined by the solution of

$$J_m(k_l a) = 0. \quad (10.57)$$

The above constraint ensures that the tangential electric field is zero on the conducting walls surrounding the cavity ($r = a$).

The most general solution for a TE mode is

$$H_z(r, \varphi, z, t) = A J_m(k_l r) e^{im\varphi} \sin(k_3 z) e^{-i\omega t}. \quad (10.58)$$

In this case, the k_l are determined by the solution of

$$J'_m(k_l a) = 0, \quad (10.59)$$

where $'$ denotes differentiation with respect to argument. The above constraint ensures that the normal magnetic field is zero on the conducting walls surrounding the cavity. The oscillation frequencies of both TM and TE modes are given by

$$\frac{\omega^2}{c^2} = k^2 = k_l^2 + \frac{n^2 \pi^2}{L^2}. \quad (10.60)$$

l	j_{0l}	j'_{0l}	j_{1l}	j'_{1l}
1	2.4048	0.0000	3.8317	1.8412
2	5.5201	3.8317	7.0156	5.3314
3	8.6537	7.0156	10.173	8.5363
4	11.792	10.173	13.324	11.706

Table 10.1: The first few values of j_{0l} , j'_{0l} , j_{1l} and j'_{1l} .

If l is the ordinal number of a zero of a particular Bessel function of order m (defined such that l increases with increasing values of the argument) then each mode is characterized by three integers, l, m, n , as in the rectangular case. The l th zero of $J_m(z)$ is conventionally denoted j_{ml} [so, $J_m(j_{ml}) = 0$]. Likewise, the l th zero of $J'_m(z)$ is denoted j'_{ml} . Table 10.1 shows the first few zeros of $J_0(z)$, $J'_0(z)$, $J_1(z)$, and $J'_1(z)$. It is clear that, for fixed n and m , the lowest frequency mode (i.e., the mode with the lowest value of k_l) is a TE mode. The mode with the next highest frequency is a TM mode. The next highest frequency mode is a TE mode, and so on.

10.7 Waveguides

Let us consider the transmission of electromagnetic waves along the axis of a waveguide, which is simply a long, axially symmetric, hollow conductor with open ends. In order to represent a wave propagating along the z -direction, we express the dependence of field quantities on the spatial coordinates and time in the form

$$f(x, y) e^{i(k_g z - \omega t)}. \quad (10.61)$$

The *guide propagation constant*, k_g , is just the k_3 of previous sections, except that it is no longer restricted by the boundary conditions to take discrete values. The general considerations of Section 10.5 still apply, so that we can treat TM and TE modes separately. The solutions for $f(x, y)$ are identical to those for axially symmetric cavities already discussed. Although k_g is not restricted in magnitude, we note that for every eigenvalue of the transverse wave equation, k_s , there is a lowest value of k , namely $k = k_s$ (often designated k_c for waveguides), for which $k_g = \sqrt{k^2 - k_s^2}$ is real. This corresponds to the *cutoff frequency*, below which waves are not transmitted by the mode in question, and the fields fall off exponentially with increasing z . In fact, the waveguide dispersion relation for a particular mode can easily be shown to take the form

$$k_g = \frac{\sqrt{\omega^2 - \omega_c^2}}{c}, \quad (10.62)$$

where

$$\omega_c = k_c c \equiv k_s c \quad (10.63)$$

is the cutoff frequency. There is an absolute cutoff frequency associated with the mode of lowest frequency: that is, the mode with the lowest value of k_c .

For real k_g (i.e., $\omega > \omega_c$), it is clear from Equation (10.62) that the wave propagates along the guide at the phase velocity

$$u_p = \frac{\omega}{k_g} = \frac{c}{\sqrt{1 - \omega_c^2/\omega^2}}. \quad (10.64)$$

It is evident that this velocity is greater than that of electromagnetic waves in free space. The velocity is not constant, however, but depends on the frequency. The waveguide thus behaves as a dispersive medium. The group velocity of a wave pulse propagated along the guide is given by

$$u_g = \frac{d\omega}{dk_g} = c \sqrt{1 - \omega_c^2/\omega^2}. \quad (10.65)$$

It can be seen that u_g is always smaller than c , and also that

$$u_p u_g = c^2. \quad (10.66)$$

For a TM mode ($H_z = 0$), Equations (10.47)–(10.48) yield

$$\mathbf{E}_s = \frac{i k_g}{k_s^2} \nabla_s E_z, \quad (10.67)$$

$$\mathbf{H}_s = \frac{\omega \epsilon_0}{k_g} \mathbf{e}_z \times \mathbf{E}_s, \quad (10.68)$$

where use has been made of $\partial/\partial z = i k_g$. For TE modes ($E_z = 0$), Equations (10.49)–(10.50) give

$$\mathbf{H}_s = \frac{i k_g}{k_s^2} \nabla_s H_z, \quad (10.69)$$

$$\mathbf{E}_s = -\frac{\omega \mu_0}{k_g} \mathbf{e}_z \times \mathbf{H}_s. \quad (10.70)$$

The time-average z component of the Poynting vector, \mathbf{N} , is given by

$$\overline{N}_z = \frac{1}{2} |\mathbf{E}_s \times \mathbf{H}_s^*|. \quad (10.71)$$

It follows that

$$\overline{N}_z = \sqrt{\frac{\mu_0}{\epsilon_0}} \frac{1}{\sqrt{1 - \omega_c^2/\omega^2}} \frac{H_{s0}^2}{2} \quad (10.72)$$

for TE modes, and

$$\overline{N}_z = \sqrt{\frac{\mu_0}{\epsilon_0}} \sqrt{1 - \omega_c^2/\omega^2} \frac{H_{s0}^2}{2} \quad (10.73)$$

for TM modes. The subscript 0 denotes the peak value of a wave quantity.

For a given mode, waveguide losses can be estimated by integrating Equation (10.24) over the wall of the guide. The energy flow of a propagating wave attenuates as e^{-Kz} , where

$$K = \frac{\text{power loss per unit length of guide}}{\text{power transmitted through guide}}. \quad (10.74)$$

Thus,

$$K = \frac{1}{2\sigma d} \int (H_s^2 + H_z^2) dS \left/ \int \overline{N}_z dS \right., \quad (10.75)$$

where the numerator is integrated over unit length of the wall, and the denominator is integrated over the transverse cross-section of the guide. It is customary to define the *guide impedance*, Z_g , by writing

$$\int \overline{N}_z dS = \frac{Z_g}{2} \int H_{s0}^2 dS. \quad (10.76)$$

Here, both integrals are over the transverse cross-section of the guide. It follows from Equations (10.71) and (10.72) that

$$Z_g = \sqrt{\frac{\mu_0}{\epsilon_0}} \frac{1}{\sqrt{1 - \omega_c^2/\omega^2}} \quad (10.77)$$

for TE modes, and

$$Z_g = \sqrt{\frac{\mu_0}{\epsilon_0}} \sqrt{1 - \omega_c^2/\omega^2} \quad (10.78)$$

for TM modes. For both types of mode, $\mathbf{H}_s = (1/Z_g) \mathbf{e}_z \times \mathbf{E}_s$.

10.8 Dielectric Waveguides

We have seen that it is possible to propagate electromagnetic waves down a hollow conductor. However, other types of guiding structures are also possible. The general requirement for a guide of electromagnetic waves is that there be a flow of energy along the axis of the guiding structure, but not perpendicular to the axis. This implies that the electromagnetic fields are appreciable only in the immediate neighborhood of the guiding structure.

Consider a uniform cylinder of arbitrary cross-section made of some dielectric material, and surrounded by a vacuum. This structure can serve as a waveguide provided the dielectric constant of the material is sufficiently large. Note, however, that the boundary conditions satisfied by the electromagnetic fields are significantly different to those of a conventional waveguide. The transverse fields are governed by two equations: one for the region inside the dielectric, and the other for the vacuum region. Inside the dielectric, we have

$$\left[\nabla_s^2 + \left(\epsilon_1 \frac{\omega^2}{c^2} - k_g^2 \right) \right] \psi = 0. \quad (10.79)$$

In the vacuum region, we have

$$\left[\nabla_s^2 + \left(\frac{\omega^2}{c^2} - k_g^2 \right) \right] \psi = 0. \quad (10.80)$$

Here, $\psi(x, y) e^{ik_g z}$ stands for either E_z or H_z , ϵ_1 is the relative permittivity of the dielectric material, and k_g is the guide propagation constant. The propagation constant must be the same both inside and outside the dielectric in order to allow the electromagnetic boundary conditions to be satisfied at all points on the surface of the cylinder.

Inside the dielectric, the transverse Laplacian must be negative, so that the constant

$$k_s^2 = \epsilon_1 \frac{\omega^2}{c^2} - k_g^2 \quad (10.81)$$

is positive. Outside the cylinder the requirement of no transverse flow of energy can only be satisfied if the fields fall off exponentially (instead of oscillating). Thus,

$$k_t^2 = k_g^2 - \frac{\omega^2}{c^2} \quad (10.82)$$

must be positive.

The oscillatory solutions (inside) must be matched to the exponentiating solutions (outside). The boundary conditions are the continuity of normal \mathbf{B} and \mathbf{D} and tangential \mathbf{E} and \mathbf{H} on the surface of the tube. These boundary conditions are far more complicated than those in a conventional waveguide. For this reason, the normal modes cannot usually be classified as either pure TE or pure TM modes. In general, the normal modes possess both electric and magnetic field components in the transverse plane. However, for the special case of a dielectric cylinder tube of circular cross-section, the normal modes can have either pure TE or pure TM characteristics. Let us examine this case in detail.

Consider a dielectric cylinder of dielectric constant ϵ_1 whose transverse cross-section is a circle of radius a . For the sake of simplicity, let us only search for normal modes whose electromagnetic fields have no azimuthal variation. Equations (10.79) and (10.81) yield

$$\left(r^2 \frac{d^2}{dr^2} + r \frac{d}{dr} + r^2 k_s^2 \right) \psi = 0 \quad (10.83)$$

for $r < a$. The general solution to this equation is some linear combination of the Bessel functions $J_0(k_s r)$ and $Y_0(k_s r)$. However, because $Y_0(k_s r)$ is badly behaved at the origin ($r = 0$), the physical solution is $\psi(r) \propto J_0(k_s r)$.

Equations (10.80) and (10.82) yield

$$\left(r^2 \frac{d^2}{dr^2} + r \frac{d}{dr} - r^2 k_t^2 \right) \psi = 0. \quad (10.84)$$

which can be rewritten

$$\left(z^2 \frac{d^2}{dz^2} + z \frac{d}{dz} - z^2 \right) \psi = 0, \quad (10.85)$$

where $z = k_t r$. The above can be recognized as a type of modified Bessel equation, whose most general form is

$$\left[z^2 \frac{d^2}{dz^2} + z \frac{d}{dz} - (z^2 + m^2) \right] \psi = 0. \quad (10.86)$$

The two linearly independent solutions of the previous equation are denoted $I_m(z)$ and $K_m(z)$. Moreover, $I_m(z) \rightarrow \infty$ as $|z| \rightarrow \infty$, whereas $K_m(z) \rightarrow 0$. Thus, it is clear that the physical solution to Equation (10.84) (i.e., the one that decays as $|r| \rightarrow \infty$) is $\psi(r) \propto K_0(k_t r)$.

The physical solution is then

$$\psi(r) = J_0(k_s r) \quad (10.87)$$

for $r \leq a$, and

$$\psi(r) = A K_0(k_t r) \quad (10.88)$$

for $r > a$. Here, A is an arbitrary constant, and $\psi(r) e^{ik_g z}$ stands for either E_z or H_z . It follows from Equations (10.40)–(10.41) (using $\partial/\partial\theta = 0$) that

$$H_r = i \frac{k_g}{k_s^2} \frac{\partial H_z}{\partial r}, \quad (10.89)$$

$$E_\theta = -\frac{\omega \mu_0}{k_g} H_r, \quad (10.90)$$

$$H_\theta = i \frac{\omega \epsilon_0 \epsilon_1}{k_s^2} \frac{\partial E_z}{\partial r}, \quad (10.91)$$

$$E_r = \frac{k_g}{\omega \epsilon_0 \epsilon_1} H_\theta \quad (10.92)$$

for $r \leq a$. There are an analogous set of relations for $r > a$. The fact that the field components form two groups—that is, (H_r, E_θ) , which depend on H_z , and (H_θ, E_r) , which depend on E_z —implies that the normal modes take the form of either pure TE modes or pure TM modes.

For a TE mode ($E_z = 0$) we find that

$$H_z(r) = J_0(k_s r), \quad (10.93)$$

$$H_r(r) = -i \frac{k_g}{k_s} J_1(k_s r), \quad (10.94)$$

$$E_\theta(r) = i \frac{\omega \mu_0}{k_s} J_1(k_s r) \quad (10.95)$$

for $r \leq a$, and

$$H_z(r) = A K_0(k_t r), \quad (10.96)$$

$$H_r(r) = i A \frac{k_g}{k_t} K_1(k_t r), \quad (10.97)$$

$$E_\theta(r) = -i A \frac{\omega \mu_0}{k_t} K_1(k_t r) \quad (10.98)$$

for $r > a$. Here we have used the identities

$$J'_0(z) \equiv -J_1(z), \quad (10.99)$$

$$K'_0(z) \equiv -K_1(z), \quad (10.100)$$

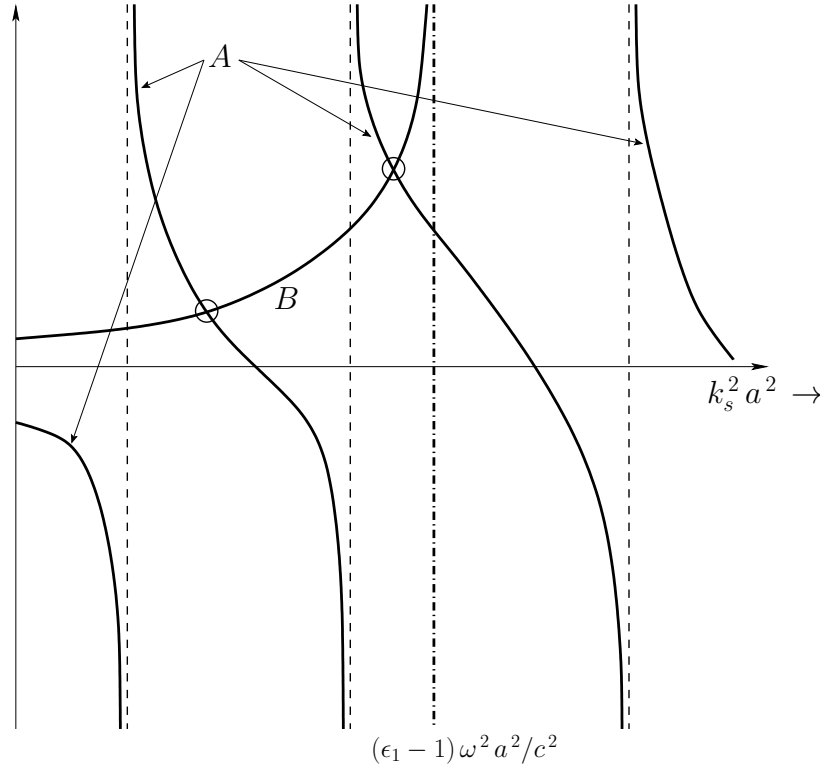


Figure 10.1: Graphical solution of the dispersion relation (10.103). The curve A represents $-J_1(k_s/a)/k_s J_0(k_s a)$. The curve B represents $K_1(k_t a)/k_t K_0(k_t a)$.

where $'$ denotes differentiation with respect to z . The boundary conditions require $H_z(r)$, $H_r(r)$, and $E_\theta(r)$ to be continuous across $r = a$. Thus, it follows that

$$A K_0(k_t a) = J_0(k_s a), \quad (10.101)$$

$$-A \frac{K_1(k_t a)}{k_t} = \frac{J_1(k_s a)}{k_s}. \quad (10.102)$$

Eliminating the arbitrary constant A between the above two equations yields the dispersion relation

$$\frac{J_1(k_s a)}{k_s J_0(k_s a)} + \frac{K_1(k_t a)}{k_t K_0(k_t a)} = 0, \quad (10.103)$$

where

$$k_t^2 + k_s^2 = (\epsilon_1 - 1) \frac{\omega^2}{c^2}. \quad (10.104)$$

Figure 10.1 shows a graphical solution of the above dispersion relation. The roots correspond to the crossing points of the two curves; $-J_1(k_s a)/k_s J_0(k_s a)$ and $K_1(k_t a)/k_t K_0(k_t a)$. The vertical asymptotes of the first curve are given by the roots of $J_0(k_s a) = 0$. The vertical asymptote of the second curve occurs when $k_t = 0$: that is, when $k_s^2 a^2 = (\epsilon_1 - 1) \omega^2 a^2 / c^2$. Note, from Equation (10.104), that k_t decreases as k_s increases. In Figure 10.1, there are two crossing points,

corresponding to two distinct propagating modes of the system. It is evident that if the point $k_t = 0$ corresponds to a value of $k_s a$ that is less than the first root of $J_0(k_s a) = 0$ then there is no crossing of the two curves, and, hence, there are no propagating modes. Because the first root of $J_0(z) = 0$ occurs at $z = 2.4048$ (see Table 10.1), the condition for the existence of propagating modes can be written

$$\omega > \omega_{01} = \frac{2.4048 c}{\sqrt{\epsilon_1 - 1} a}. \quad (10.105)$$

In other words, the mode frequency must lie above the cutoff frequency ω_{01} for the TE_{01} mode [here, the 0 corresponds to the number of nodes in the azimuthal direction, and the 1 refers to the first root of $J_0(z) = 0$]. It is also evident that, as the mode frequency is gradually increased, the point $k_t = 0$ eventually crosses the second vertical asymptote of $-J_1(k_s/a)/k_s J_0(k_s a)$, at which point the TE_{02} mode can propagate. As ω is further increased, more and more TE modes can propagate. The cutoff frequency for the TE_{0l} mode is given by

$$\omega_{0l} = \frac{j_{0l} c}{\sqrt{\epsilon_1 - 1} a}, \quad (10.106)$$

where j_{0l} is l th root of $J_0(z) = 0$ (in order of increasing z).

At the cutoff frequency for a particular mode, $k_t = 0$, which implies from Equation (10.82) that $k_g = \omega/c$. In other words, the mode propagates along the guide at the velocity of light in vacuum. At frequencies below this cutoff frequency, the system no longer acts as a guide, but rather as an antenna, with energy being radiated radially. For frequencies well above the cutoff, k_t and k_g are of the same order of magnitude, and are large compared to k_s . This implies that the fields do not extend appreciably outside the dielectric cylinder.

For a TM mode ($H_z = 0$) we find that

$$E_z(r) = J_0(k_s r), \quad (10.107)$$

$$H_\theta(r) = -i \frac{\omega \epsilon_0 \epsilon_1}{k_s} J_1(k_s r), \quad (10.108)$$

$$E_r(r) = -i \frac{k_g}{k_s} J_1(k_s r) \quad (10.109)$$

for $r \leq a$, and

$$E_z(r) = A K_0(k_t r), \quad (10.110)$$

$$H_\theta(r) = i A \frac{\omega \epsilon_0}{k_t} K_1(k_t r), \quad (10.111)$$

$$E_r(r) = i A \frac{k_g}{k_t} K_1(k_t r) \quad (10.112)$$

for $r > a$. The boundary conditions require $E_z(r)$, $H_\theta(r)$, and $D_r(r)$ to be continuous across $r = a$. Thus, it follows that

$$A K_0(k_t a) = J_0(k_s a), \quad (10.113)$$

$$-A \frac{K_1(k_t a)}{k_t} = \epsilon_1 \frac{J_1(k_s a)}{k_s}. \quad (10.114)$$

Eliminating the arbitrary constant A between the above two equations yields the dispersion relation

$$\frac{\epsilon_1 J_1(k_s a)}{k_s J_0(k_s a)} + \frac{K_1(k_t a)}{k_t K_0(k_t a)} = 0. \quad (10.115)$$

It is clear, from this dispersion relation, that the cutoff frequency for the TM_{0l} mode is exactly the same as that for the TE_{0l} mode. It is also clear that, in the limit $\epsilon_1 \gg 1$, the propagation constants are determined by the roots of $J_1(k_s a) \simeq 0$. However, this is exactly the same as the determining equation for TE modes in a metallic waveguide of circular cross-section (filled with dielectric of relative permittivity ϵ_1).

Modes with azimuthal dependence (i.e., $m > 0$) have longitudinal components of both \mathbf{E} and \mathbf{H} . This makes the mathematics somewhat more complicated. However, the basic results are the same as for $m = 0$ modes: that is, for frequencies well above the cutoff frequency the modes are localized in the immediate vicinity of the cylinder.

10.9 Exercises

- 10.1 Demonstrate that the electric and magnetic fields inside a waveguide are mutually orthogonal.
- 10.2 Consider a TE_{mm} mode in a rectangular waveguide of dimensions a and b . Calculate the mean electromagnetic energy per unit length, as well as the mean electromagnetic energy flux down the waveguide. Demonstrate that the ratio of the mean energy flux to the mean energy per unit length is equal to the group-velocity of the mode.

11 Multipole Expansion

11.1 Introduction

A study of the emission and scattering of electromagnetic radiation necessarily involves solving the vector wave equation. It turns out that the solutions of this equation in free space can be conveniently expressed as an expansion in orthogonal spherical waves. Let us examine this expansion, which is known as the *multipole expansion*.

11.2 Multipole Expansion of Scalar Wave Equation

Before considering the vector wave equation, let us consider the somewhat simpler scalar wave equation. A scalar field $\psi(\mathbf{r}, t)$ satisfying the homogeneous wave equation,

$$\nabla^2 \psi - \frac{1}{c^2} \frac{\partial^2 \psi}{\partial t^2} = 0, \quad (11.1)$$

can be Fourier analyzed in time,

$$\psi(\mathbf{r}, t) = \int_{-\infty}^{\infty} \psi(\mathbf{r}, \omega) e^{-i\omega t} d\omega, \quad (11.2)$$

with each Fourier harmonic satisfying the homogeneous Helmholtz wave equation,

$$(\nabla^2 + k^2) \psi(\mathbf{r}, \omega) = 0, \quad (11.3)$$

where $k^2 = \omega^2/c^2$. We can write the Helmholtz equation in terms of spherical coordinates r, θ, φ :

$$\left(\frac{1}{r^2} \frac{\partial}{\partial r} r^2 \frac{\partial}{\partial r} + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \sin \theta \frac{\partial}{\partial \theta} + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2}{\partial \varphi^2} + k^2 \right) \psi = 0. \quad (11.4)$$

As is well known, it is possible to solve this equation via separation of variables to give

$$\psi(\mathbf{r}, \omega) = \sum_{l=0, \infty} \sum_{m=-l, +l} f_{lm}(r) Y_{lm}(\theta, \varphi). \quad (11.5)$$

Here, we restrict our attention to physical solutions that are well-behaved in the angular variables θ and φ . The spherical harmonics $Y_{lm}(\theta, \varphi)$ satisfy the following equations:

$$-\frac{\partial^2 Y_{lm}}{\partial \varphi^2} = m^2 Y_{lm}, \quad (11.6)$$

$$-\left(\frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \sin \theta \frac{\partial}{\partial \theta} + \frac{1}{\sin^2 \theta} \frac{\partial^2}{\partial \varphi^2} \right) Y_{lm} = l(l+1) Y_{lm}, \quad (11.7)$$

where l is a non-negative integer, and m is an integer that satisfies the inequality $|m| \leq l$. The radial functions $f_{lm}(r)$ satisfy

$$\left[\frac{d^2}{dr^2} + \frac{2}{r} \frac{d}{dr} + k^2 - \frac{l(l+1)}{r^2} \right] f_{lm}(r) = 0. \quad (11.8)$$

With the substitution

$$f_{lm}(r) = \frac{u_{lm}(r)}{r^{1/2}}, \quad (11.9)$$

Equation (11.8) is transformed into

$$\left[\frac{d^2}{dr^2} + \frac{1}{r} \frac{d}{dr} + k^2 - \frac{(l+1/2)^2}{r^2} \right] u_{lm}(r) = 0, \quad (11.10)$$

which is a type of Bessel equation of half-integer order, $l + 1/2$. Thus, we can write the solution for $f_{lm}(r)$ as

$$f_{lm}(r) = \frac{A_{lm}}{r^{1/2}} J_{l+1/2}(kr) + \frac{B_{lm}}{r^{1/2}} Y_{l+1/2}(kr), \quad (11.11)$$

where A_{lm} and B_{lm} are arbitrary constants. The half-integer order Bessel functions $J_{l+1/2}(z)$ and $Y_{l+1/2}(z)$ have analogous properties to the integer order Bessel functions $J_m(z)$ and $Y_m(z)$. In particular, the $J_{l+1/2}(z)$ are well behaved in the limit $|z| \rightarrow 0$, whereas the $Y_{l+1/2}(z)$ are badly behaved.

It is convenient to define the *spherical Bessel functions*, $j_l(r)$ and $y_l(r)$, where

$$j_l(z) = \left(\frac{\pi}{2z} \right)^{1/2} J_{l+1/2}(z), \quad (11.12)$$

$$y_l(z) = \left(\frac{\pi}{2z} \right)^{1/2} Y_{l+1/2}(z). \quad (11.13)$$

It is also convenient to define the *spherical Hankel functions*, $h_l^{(1)}(r)$ and $h_l^{(2)}(r)$, where

$$h_l^{(1)}(z) = j_l(z) + i y_l(z), \quad (11.14)$$

$$h_l^{(2)}(z) = j_l(z) - i y_l(z). \quad (11.15)$$

Assuming that z is real, $h_l^{(2)}(z)$ is the complex conjugate of $h_l^{(1)}(z)$. It turns out that the spherical Bessel functions can be expressed in the closed form

$$j_l(z) = (-z)^l \left(\frac{1}{z} \frac{d}{dz} \right)^l \left(\frac{\sin z}{z} \right), \quad (11.16)$$

$$y_l(z) = -(-z)^l \left(\frac{1}{z} \frac{d}{dz} \right)^l \left(\frac{\cos z}{z} \right). \quad (11.17)$$

In the limit of small argument,

$$j_l(z) \rightarrow \frac{z^l}{(2l+1)!!} [1 + \mathcal{O}(z^2)], \quad (11.18)$$

$$y_l(z) \rightarrow -\frac{(2l-1)!!}{z^{l+1}} [1 + \mathcal{O}(z^2)], \quad (11.19)$$

where $(2l + 1)!! = (2l + 1)(2l - 1)(2l - 3) \cdots 5 \cdot 3 \cdot 1$. In the limit of large argument,

$$j_l(z) \rightarrow \frac{\sin(z - l\pi/2)}{z}, \quad (11.20)$$

$$y_l(z) \rightarrow -\frac{\cos(z - l\pi/2)}{z}, \quad (11.21)$$

which implies that

$$h_l^{(1)}(z) \rightarrow (-i)^{l+1} \frac{e^{+iz}}{z}, \quad (11.22)$$

$$h_l^{(2)}(z) \rightarrow (+i)^{l+1} \frac{e^{-iz}}{z}. \quad (11.23)$$

It follows, from the above discussion, that the radial functions $f_{lm}(r)$, specified in Equation (11.11), can also be written

$$f_{lm}(r) = A_{lm} h_l^{(1)}(kr) + B_{lm} h_l^{(2)}(kr). \quad (11.24)$$

Hence, the general solution of the homogeneous Helmholtz equation, (11.3), takes the form

$$\psi(\mathbf{r}, \omega) = \sum_{l=0, \infty} \sum_{m=-l, +l} [A_{lm} h_l^{(1)}(kr) + B_{lm} h_l^{(2)}(kr)] Y_{lm}(\theta, \varphi). \quad (11.25)$$

Moreover, it is clear from Equations (11.2) and (11.22)–(11.23) that, at large r , the terms involving the $h_l^{(1)}(kr)$ Hankel functions correspond to outgoing radial waves, whereas those involving the $h_l^{(2)}(kr)$ functions correspond to incoming radial waves.

11.3 Angular Momentum Operators

It is well known from quantum mechanics that Equation (11.7) can be written in the form

$$L^2 Y_{lm} = l(l + 1) Y_{lm}. \quad (11.26)$$

Here, the differential operator L^2 is given by

$$L^2 = L_x^2 + L_y^2 + L_z^2, \quad (11.27)$$

where

$$\mathbf{L} = -i \mathbf{r} \times \nabla \quad (11.28)$$

is $1/\hbar$ times the orbital angular momentum operator of wave mechanics.

The components of \mathbf{L} are conveniently written in the combinations

$$L_+ = L_x + i L_y = e^{i\varphi} \left(\frac{\partial}{\partial \theta} + i \cot \theta \frac{\partial}{\partial \varphi} \right), \quad (11.29)$$

$$L_- = L_x - i L_y = e^{-i\varphi} \left(-\frac{\partial}{\partial \theta} + i \cot \theta \frac{\partial}{\partial \varphi} \right), \quad (11.30)$$

$$L_z = -i \frac{\partial}{\partial \varphi}. \quad (11.31)$$

Note that \mathbf{L} only operates on angular variables, and is independent of r . It is evident from the definition (11.28) that

$$\mathbf{r} \cdot \mathbf{L} = 0. \quad (11.32)$$

It is easily demonstrated from Equations (11.29)–(11.31) that

$$L^2 = -\frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \sin \theta \frac{\partial}{\partial \theta} - \frac{1}{\sin^2 \theta} \frac{\partial^2}{\partial \varphi^2}. \quad (11.33)$$

The following results are well known in quantum mechanics:

$$L_+ Y_{lm} = \sqrt{(l-m)(l+m+1)} Y_{l,m+1}, \quad (11.34)$$

$$L_- Y_{lm} = \sqrt{(l+m)(l-m+1)} Y_{l,m-1}, \quad (11.35)$$

$$L_z Y_{lm} = m Y_{lm}. \quad (11.36)$$

In addition,

$$L^2 \mathbf{L} = \mathbf{L} L^2, \quad (11.37)$$

$$\mathbf{L} \times \mathbf{L} = i \mathbf{L}, \quad (11.38)$$

$$L_j \nabla^2 = \nabla^2 L_j, \quad (11.39)$$

where

$$\nabla^2 = \frac{1}{r^2} \frac{\partial}{\partial r} r^2 \frac{\partial}{\partial r} - \frac{L^2}{r^2}. \quad (11.40)$$

11.4 Multipole Expansion of Vector Wave Equation

Maxwell's equations in free space reduce to

$$\nabla \cdot \mathbf{E} = 0, \quad (11.41)$$

$$\nabla \cdot c \mathbf{B} = 0, \quad (11.42)$$

$$\nabla \times \mathbf{E} = i k c \mathbf{B}, \quad (11.43)$$

$$\nabla \times c \mathbf{B} = -i k \mathbf{E}, \quad (11.44)$$

assuming an $e^{-i\omega t}$ time dependence of all field quantities. Here, $k = \omega/c$. Eliminating \mathbf{E} between Equations (11.43) and (11.44), we obtain

$$(\nabla^2 + k^2) \mathbf{B} = 0, \quad (11.45)$$

$$\nabla \cdot \mathbf{B} = 0, \quad (11.46)$$

with \mathbf{E} given by

$$\mathbf{E} = \frac{i}{k} \nabla \times c \mathbf{B}. \quad (11.47)$$

Alternatively, \mathbf{B} can be eliminated to give

$$(\nabla^2 + k^2) \mathbf{E} = 0, \quad (11.48)$$

$$\nabla \cdot \mathbf{E} = 0, \quad (11.49)$$

with \mathbf{B} given by

$$c \mathbf{B} = -\frac{i}{k} \nabla \times \mathbf{E}. \quad (11.50)$$

It is clear that each Cartesian component of \mathbf{B} and \mathbf{E} satisfies the homogeneous Helmholtz wave equation, (11.3). Hence, according to the analysis of Section 11.2, these components can be written as a general expansion of the form

$$\psi(\mathbf{r}) = \sum_{l,m} \left[A_{lm}^{(1)} h_l^{(1)}(kr) + A_{lm}^{(2)} h_l^{(2)}(kr) \right] Y_{lm}(\theta, \varphi), \quad (11.51)$$

where ψ stands for any Cartesian component of \mathbf{E} or \mathbf{B} . Note, however, that the three Cartesian components of \mathbf{E} or \mathbf{B} are not entirely independent, because they must also satisfy the constraints $\nabla \cdot \mathbf{E} = 0$ and $\nabla \cdot \mathbf{B} = 0$. Let us examine how these constraints can be satisfied with the minimum of effort.

Consider the scalar $\mathbf{r} \cdot \mathbf{A}$, where \mathbf{A} is a well-behaved vector field. It is easily verified that

$$\nabla^2(\mathbf{r} \cdot \mathbf{A}) = \mathbf{r} \cdot (\nabla^2 \mathbf{A}) + 2 \nabla \cdot \mathbf{A}. \quad (11.52)$$

It follows from Equations (11.45)–(11.46) and (11.48)–(11.49) that the scalars $\mathbf{r} \cdot \mathbf{E}$ and $\mathbf{r} \cdot \mathbf{B}$ both satisfy the homogeneous Helmholtz wave equation: that is,

$$(\nabla^2 + k^2)(\mathbf{r} \cdot \mathbf{E}) = 0, \quad (11.53)$$

$$(\nabla^2 + k^2)(\mathbf{r} \cdot \mathbf{B}) = 0. \quad (11.54)$$

Thus, the general solutions for $\mathbf{r} \cdot \mathbf{E}$ and $\mathbf{r} \cdot \mathbf{B}$ can be written in the form (11.51).

Let us define a *magnetic multipole* field of order l, m as the solution of

$$\mathbf{r} \cdot c \mathbf{B}_{lm}^{(M)} = \frac{l(l+1)}{k} g_l(kr) Y_{lm}(\theta, \varphi), \quad (11.55)$$

$$\mathbf{r} \cdot \mathbf{E}_{lm}^{(M)} = 0, \quad (11.56)$$

where

$$g_l(kr) = A_l^{(1)} h_l^{(1)}(kr) + A_l^{(2)} h_l^{(2)}(kr). \quad (11.57)$$

The presence of the factor $l(l+1)/k$ in Equation (11.55) is for later convenience. Equation (11.50) yields

$$k \mathbf{r} \cdot c \mathbf{B} = -i \mathbf{r} \cdot (\nabla \times \mathbf{E}) = -i (\mathbf{r} \times \nabla) \cdot \mathbf{E} = \mathbf{L} \cdot \mathbf{E}, \quad (11.58)$$

where \mathbf{L} is given by Equation (11.28). Thus, with $\mathbf{r} \cdot c \mathbf{B}$ taking the form (11.55), the electric field associated with a magnetic multipole must satisfy

$$\mathbf{L} \cdot \mathbf{E}_{lm}^{(M)}(r, \theta, \varphi) = l(l+1) g_l(kr) Y_{lm}(\theta, \varphi), \quad (11.59)$$

as well as $\mathbf{r} \cdot \mathbf{E}_{lm}^{(M)} = 0$. Recall that the operator \mathbf{L} acts on the angular variables θ, φ only. This implies that the radial dependence of $\mathbf{E}_{lm}^{(M)}$ is given by $g_l(kr)$. It is easily seen from Equations (11.26) and (11.32) that the solution to Equations (11.56) and (11.59) can be written in the form

$$\mathbf{E}_{lm}^{(M)} = g_l(kr) \mathbf{L} Y_{lm}(\theta, \varphi). \quad (11.60)$$

It follows from the analysis Section 11.3 that the angular dependence of $\mathbf{E}_{lm}^{(M)}$ consists of a linear combination of $Y_{l,m-1}(\theta, \varphi)$, $Y_{lm}(\theta, \varphi)$, and $Y_{l,m+1}(\theta, \varphi)$ functions. Equation (11.60), together with

$$c \mathbf{B}_{lm}^{(M)} = -\frac{i}{k} \nabla \times \mathbf{E}_{lm}^{(M)}, \quad (11.61)$$

specifies the electromagnetic fields of a magnetic multipole of order l, m . According to Equation (11.32), the electric field (11.60) is transverse to the radius vector. Thus, magnetic multipole fields are sometimes termed *transverse electric* (TE) multipole fields.

The fields of an *electric*, or *transverse magnetic* (TM), multipole of order l, m satisfy

$$\mathbf{r} \cdot \mathbf{E}_{lm}^{(E)} = -\frac{l(l+1)}{k} f_l(kr) Y_{lm}(\theta, \varphi), \quad (11.62)$$

$$\mathbf{r} \cdot \mathbf{B}_{lm}^{(E)} = 0. \quad (11.63)$$

It follows that the fields of an electric multipole are

$$c \mathbf{B}_{lm}^{(E)} = f_l(kr) \mathbf{L} Y_{lm}(\theta, \varphi), \quad (11.64)$$

$$\mathbf{E}_{lm}^{(E)} = \frac{i}{k} \nabla \times c \mathbf{B}_{lm}^{(E)}. \quad (11.65)$$

Here, the radial function $f_l(kr)$ is an expression of the form (11.57).

The two sets of multipole fields, (11.60)–(11.61), and (11.64)–(11.65), form a complete set of vector solutions to Maxwell's equations in free space. Because the vector spherical harmonic $\mathbf{L} Y_{lm}$ plays an important role in the theory of multipole fields, it is convenient to introduce the normalized form

$$\mathbf{X}_{lm}(\theta, \varphi) = \frac{1}{\sqrt{l(l+1)}} \mathbf{L} Y_{lm}(\theta, \varphi). \quad (11.66)$$

It can be demonstrated that these forms possess the orthogonality properties

$$\oint \mathbf{X}_{l'm'}^* \cdot \mathbf{X}_{lm} d\Omega = \delta_{ll'} \delta_{mm'}, \quad (11.67)$$

$$\oint \mathbf{X}_{l'm'}^* \cdot (\mathbf{r} \times \mathbf{X}_{lm}) d\Omega = 0, \quad (11.68)$$

for all l, l', m , and m' .

By combining the two types of multipole fields, we can write the general solution to Maxwell's equations in free space as

$$c \mathbf{B} = \sum_{l,m} \left[a_E(l, m) f_l(kr) \mathbf{X}_{lm} - \frac{i}{k} a_M(l, m) \nabla \times g_l(kr) \mathbf{X}_{lm} \right], \quad (11.69)$$

$$\mathbf{E} = \sum_{l,m} \left[\frac{i}{k} a_E(l, m) \nabla \times f_l(kr) \mathbf{X}_{lm} + a_M(l, m) g_l(kr) \mathbf{X}_{lm} \right], \quad (11.70)$$

where the coefficients $a_E(l, m)$ and $a_M(l, m)$ specify the amounts of electric l, m and magnetic l, m multipole fields. The radial functions $f_l(kr)$ and $g_l(kr)$ are both of the form (11.57). The coefficients $a_E(l, m)$ and $a_M(l, m)$, as well as the relative proportions of the two types of Hankel functions in the radial functions $f_l(kr)$ and $g_l(kr)$, are determined by the sources and the boundary conditions.

Equations (11.69) and (11.70) yield

$$\mathbf{r} \cdot c \mathbf{B} = \frac{1}{k} \sum_{l,m} a_M(l, m) g_l(kr) \mathbf{L} \mathbf{X}_{lm} = \frac{1}{k} \sum_{l,m} a_M(l, m) g_l(kr) \sqrt{l(l+1)} Y_{lm}, \quad (11.71)$$

and

$$\mathbf{r} \cdot \mathbf{E} = -\frac{1}{k} \sum_{l,m} a_E(l, m) f_l(kr) \mathbf{L} \mathbf{X}_{lm} = -\frac{1}{k} \sum_{l,m} a_E(l, m) f_l(kr) \sqrt{l(l+1)} Y_{lm}, \quad (11.72)$$

where use has been made of Equations (11.26), (11.28), (11.32), and (11.66). It follows from the well-known orthogonality property of the spherical harmonics that

$$a_M(l, m) g_l(kr) = \frac{k}{\sqrt{l(l+1)}} \oint Y_{lm}^* \mathbf{r} \cdot c \mathbf{B} d\Omega, \quad (11.73)$$

$$a_E(l, m) f_l(kr) = -\frac{k}{\sqrt{l(l+1)}} \oint Y_{lm}^* \mathbf{r} \cdot \mathbf{E} d\Omega. \quad (11.74)$$

Thus, knowledge of $\mathbf{r} \cdot \mathbf{B}$ and $\mathbf{r} \cdot \mathbf{E}$ at two different radii in a source-free region permits a complete specification of the fields, including the relative proportions of the Hankel functions $h_l^{(1)}(kr)$ and $h_l^{(2)}(kr)$ present in the radial functions $f_l(kr)$ and $g_l(kr)$.

11.5 Properties of Multipole Fields

Let us examine some of the properties of the multipole fields (11.60)–(11.61) and (11.64)–(11.65). Consider, first of all, the so-called *near zone*, in which $kr \ll 1$. In this region, $f_l(kr)$ is dominated by $y_l(kr)$, which blows up as $kr \rightarrow 0$, and which has the asymptotic expansion (11.19), unless the coefficient of $y_l(kr)$ vanishes identically. Excluding this possibility, the limiting behavior of the magnetic field for an electric l, m multipole is

$$c \mathbf{B}_{lm}^{(E)} \rightarrow -\frac{k}{l} \mathbf{L} \frac{Y_{lm}}{r^{l+1}}, \quad (11.75)$$

where the proportionality constant is chosen for later convenience. To find the corresponding electric field, we must take the curl of the right-hand side of the above equation. The following operator identity is useful

$$\mathbf{i} \nabla \times \mathbf{L} \equiv \mathbf{r} \nabla^2 - \nabla \left(1 + r \frac{\partial}{\partial r} \right). \quad (11.76)$$

The electric field (11.65) can be written

$$\mathbf{E}_{lm}^{(E)} \rightarrow \frac{-\mathbf{i}}{l} \nabla \times \mathbf{L} \left(\frac{Y_{lm}}{r^{l+1}} \right). \quad (11.77)$$

Because Y_{lm}/r^{l+1} is a solution of Laplace's equation, it is annihilated by the first term on the right-hand side of (11.76). Consequently, for an electric l, m multipole, the electric field in the near zone becomes

$$\mathbf{E}_{lm}^{(E)} \rightarrow -\nabla \left(\frac{Y_{lm}}{r^{l+1}} \right). \quad (11.78)$$

This, of course, is an electrostatic multipole field. Such a field can be obtained in a more straightforward manner by observing that $\mathbf{E} \rightarrow -\nabla\phi$, where $\nabla^2\phi = 0$, in the near zone. Solving Laplace's equation by separation of variables in spherical coordinates, and demanding that ϕ be well behaved as $|\mathbf{r}| \rightarrow \infty$, yields

$$\phi(r, \theta, \varphi) = \sum_{l,m} \frac{Y_{lm}(\theta, \varphi)}{r^{l+1}}. \quad (11.79)$$

Note that (c times) the magnetic field (11.75) is smaller than the electric field (11.78) by a factor of order kr . Thus, in the near zone, (c times) the magnetic field associated with an electric multipole is much smaller than the corresponding electric field. For magnetic multipole fields, it is evident from Equations (11.60)–(11.61) and (11.64)–(11.65) that the roles of \mathbf{E} and $c\mathbf{B}$ are interchanged according to the transformation

$$\mathbf{E}^{(E)} \rightarrow -c\mathbf{B}^{(M)}, \quad (11.80)$$

$$c\mathbf{B}^{(E)} \rightarrow \mathbf{E}^{(M)}. \quad (11.81)$$

In the so-called *far zone*, or *radiation zone*, in which $kr \gg 1$, the multipole fields depend on the boundary conditions imposed at infinity. For definiteness, let us consider the case of outgoing waves at infinity, which is appropriate to radiation by a localized source. For this case, the radial function $f_l(kr)$ contains only the spherical Hankel function $h_l^{(1)}(kr)$. From the asymptotic form (11.22), it is clear that in the radiation zone the magnetic field of an electric l, m multipole varies as

$$c\mathbf{B}_{lm}^{(E)} \rightarrow (-\mathbf{i})^{l+1} \frac{e^{ikr}}{kr} \mathbf{L} Y_{lm}. \quad (11.82)$$

Using Equation (11.65), the corresponding electric field can be written

$$\mathbf{E}_{lm}^{(E)} = \frac{(-\mathbf{i})^l}{k^2} \left[\nabla \left(\frac{e^{ikr}}{r} \right) \times \mathbf{L} Y_{lm} + \frac{e^{ikr}}{r} \nabla \times \mathbf{L} Y_{lm} \right]. \quad (11.83)$$

Neglecting terms that fall off faster than $1/r$, the above expression reduces to

$$\mathbf{E}_{lm}^{(E)} = -(-i)^{l+1} \frac{e^{ikr}}{kr} \left[\mathbf{n} \times \mathbf{L} Y_{lm} - \frac{1}{k} (\mathbf{r} \nabla^2 - \nabla) Y_{lm} \right], \quad (11.84)$$

where use has been made of the identity (11.76), and $\mathbf{n} = \mathbf{r}/r$ is a unit vector pointing in the radial direction. The second term in square brackets is smaller than the first term by a factor of order $1/(kr)$, and can, therefore, be neglected in the limit $kr \gg 1$. Thus, we find that the electric field in the radiation zone takes the form

$$\mathbf{E}_{lm}^{(E)} = c \mathbf{B}_{lm}^{(E)} \times \mathbf{n}, \quad (11.85)$$

where $c \mathbf{B}_{lm}^{(E)}$ is given by Equation (11.82). These fields are typical radiation fields: that is, they are transverse to the radius vector, mutually orthogonal, fall off like $1/r$, and are such that $|\mathbf{E}| = c |\mathbf{B}|$. To obtain expansions for magnetic multipoles, we merely make the transformation (11.80)–(11.81).

Consider a linear superposition of electric l, m multipoles with different m values that all possess a common l value. Suppose that all multipoles correspond to outgoing waves at infinity. It follows from Equations (11.64)–(11.66) that

$$c \mathbf{B}_l = \sum_l a_E(l, m) h_l^{(1)}(kr) e^{-i\omega t} \mathbf{X}_{lm}, \quad (11.86)$$

$$\mathbf{E}_l = \frac{i}{k} \nabla \times c \mathbf{B}_l. \quad (11.87)$$

For harmonically varying fields, the time-averaged energy density is given by

$$u = \frac{\epsilon_0}{4} (\mathbf{E} \cdot \mathbf{E}^* + c \mathbf{B} \cdot c \mathbf{B}^*). \quad (11.88)$$

In the radiation zone, the two terms on the right-hand side of the above equation are equal. It follows that the energy contained in a spherical shell lying between radii r and $r + dr$ is

$$dU = \frac{\epsilon_0 dr}{2k^2} \sum_{m, m'} a_E^*(l, m') a_E(l, m) \oint \mathbf{X}_{lm'}^* \cdot \mathbf{X}_{lm} d\Omega, \quad (11.89)$$

where use has been made of the asymptotic form (11.22) of the spherical Hankel function $h_l^{(1)}(z)$. The orthogonality relation (11.67) leads to

$$\frac{dU}{dr} = \frac{\epsilon_0}{2k^2} \sum_m |a_E(l, m)|^2, \quad (11.90)$$

which is clearly independent of the radius. For a general superposition of electric and magnetic multipoles, the sum over m becomes a sum over l and m , whereas $|a_E|^2$ becomes $|a_E|^2 + |a_M|^2$. Thus, the net energy in a spherical shell situated in the radiation zone is an incoherent sum over all multipoles.

The time-averaged angular momentum density of harmonically varying electromagnetic fields is given by

$$\mathbf{m} = \frac{\epsilon_0}{2} \operatorname{Re} [\mathbf{r} \times (\mathbf{E} \times \mathbf{B}^*)]. \quad (11.91)$$

For a superposition of electric multipoles, the triple product can be expanded, and the electric field (11.87) substituted, to give

$$\mathbf{m} = \frac{\epsilon_0 c}{2k} \operatorname{Re} [\mathbf{B}^* (\mathbf{L} \cdot \mathbf{B})]. \quad (11.92)$$

Thus, the net angular momentum contained in a spherical shell lying between radii r and $r + dr$ (in the radiation zone) is

$$d\mathbf{M} = \frac{\epsilon_0 c dr}{2k^3} \operatorname{Re} \sum_{m,m'} a_E^*(l, m') a_E(l, m) \oint (\mathbf{L} \cdot \mathbf{X}_{lm'})^* \mathbf{X}_{lm} d\Omega. \quad (11.93)$$

It follows from Equations (11.26) and (11.66) that

$$\frac{d\mathbf{M}}{dr} = \frac{\epsilon_0 c}{2k^3} \operatorname{Re} \sum_{m,m'} a_E^*(l, m') a_E(l, m) \oint Y_{lm'}^* \mathbf{L} Y_{lm} d\Omega. \quad (11.94)$$

According to Equations (11.29)–(11.31), the Cartesian components of $d\mathbf{M}/dr$ can be written:

$$\begin{aligned} \frac{dM_x}{dr} = \frac{\epsilon_0 c}{4k^3} \operatorname{Re} \sum_m \left[\sqrt{(l-m)(l+m+1)} a_E^*(l, m+1) \right. \\ \left. + \sqrt{(l+m)(l-m+1)} a_E^*(l, m-1) \right] a_E(l, m), \end{aligned} \quad (11.95)$$

$$\begin{aligned} \frac{dM_y}{dr} = \frac{\epsilon_0 c}{4k^3} \operatorname{Im} \sum_m \left[\sqrt{(l-m)(l+m+1)} a_E^*(l, m+1) \right. \\ \left. - \sqrt{(l+m)(l-m+1)} a_E^*(l, m-1) \right] a_E(l, m), \end{aligned} \quad (11.96)$$

$$\frac{dM_z}{dr} = \frac{\epsilon_0 c}{2k^3} \sum_m m |a_E(l, m)|^2. \quad (11.97)$$

Thus, for a general l th order electric multipole that consists of a superposition of different m values, only the z component of the angular momentum takes a relatively simple form.

11.6 Solution of Inhomogeneous Helmholtz Equation

The inhomogeneous Helmholtz wave equation is conveniently solved by means of a Green's function, $G_\omega(\mathbf{r}, \mathbf{r}')$, that satisfies

$$(\nabla^2 + k^2) G_\omega(\mathbf{r}, \mathbf{r}') = -\delta(\mathbf{r} - \mathbf{r}'). \quad (11.98)$$

The solution of this equation, subject to the Sommerfeld radiation condition, which ensures that sources radiate waves instead of absorbing them, is written

$$G_\omega(\mathbf{r}, \mathbf{r}') = \frac{e^{ik|\mathbf{r}-\mathbf{r}'|}}{4\pi|\mathbf{r}-\mathbf{r}'|}. \quad (11.99)$$

(See Chapter 1.)

As is well known, the spherical harmonics satisfy the completeness relation

$$\sum_{l=0,\infty} \sum_{m=-l,+l} Y_{lm}^*(\theta', \varphi') Y_{lm}(\theta, \varphi) = \delta(\varphi - \varphi') \delta(\cos \theta - \cos \theta'). \quad (11.100)$$

Now, the three-dimensional delta function can be written

$$\delta(\mathbf{r} - \mathbf{r}') = \frac{\delta(r - r')}{r^2} \delta(\varphi - \varphi') \delta(\cos \theta - \cos \theta'). \quad (11.101)$$

It follows that

$$\delta(\mathbf{r} - \mathbf{r}') = \frac{\delta(r - r')}{r^2} \sum_{l=0,\infty} \sum_{m=-l,+l} Y_{lm}^*(\theta', \varphi') Y_{lm}(\theta, \varphi). \quad (11.102)$$

Let us expand the Green's function in the form

$$G_\omega(\mathbf{r}, \mathbf{r}') = \sum_{l,m} g_l(r, r') Y_{lm}^*(\theta', \varphi') Y_{lm}(\theta, \varphi). \quad (11.103)$$

Substitution of this expression into Equation (11.98) yields

$$\left[\frac{d^2}{dr^2} + \frac{2}{r} \frac{d}{dr} + k^2 - \frac{l(l+1)}{r^2} \right] g_l = -\frac{\delta(r - r')}{r^2}. \quad (11.104)$$

The appropriate boundary conditions are that $g_l(r)$ be finite at the origin, and correspond to an outgoing wave at infinity (i.e., $g \propto e^{ikr}$ in the limit $r \rightarrow \infty$). The solution of the above equation that satisfies these boundary conditions is

$$g_l(r, r') = A j_l(k r_<) h_l^{(1)}(k r_>), \quad (11.105)$$

where $r_<$ and $r_>$ are the greater and the lesser of r and r' , respectively. The appropriate discontinuity in slope at $r = r'$ is assured if $A = ik$, because

$$\frac{dh_l^{(1)}(z)}{dz} j_l(z) - h_l^{(1)}(z) \frac{dj_l(z)}{dz} = \frac{i}{z^2}. \quad (11.106)$$

Thus, the expansion of the Green's function becomes

$$\frac{e^{ik|\mathbf{r}-\mathbf{r}'|}}{4\pi|\mathbf{r}-\mathbf{r}'|} = ik \sum_{l=0,\infty} j_l(k r_<) h_l^{(1)}(k r_>) \sum_{m=-l,+l} Y_{lm}^*(\theta', \varphi') Y_{lm}(\theta, \varphi). \quad (11.107)$$

This is a particularly useful result, as we shall discover, because it easily allows us to express the general solution of the inhomogeneous wave equation as a multipole expansion.

11.7 Sources of Multipole Radiation

Let us now examine the connection between multipole fields and their sources. Suppose that there exist localized distributions of electric charge, $\rho(\mathbf{r}, t)$, true current, $\mathbf{j}(\mathbf{r}, t)$, and magnetization, $\mathbf{M}(\mathbf{r}, t)$. We assume that any time dependence can be analyzed into its Fourier components, and we therefore only consider harmonically varying sources, $\rho(\mathbf{r}) e^{-i\omega t}$, $\mathbf{j}(\mathbf{r}) e^{-i\omega t}$, and $\mathbf{M}(\mathbf{r}) e^{-i\omega t}$, where it is understood that we take the real parts of complex quantities.

Maxwell's equations can be written

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0}, \quad (11.108)$$

$$\nabla \cdot \mathbf{B} = 0, \quad (11.109)$$

$$\nabla \times \mathbf{E} - i k c \mathbf{B} = \mathbf{0}, \quad (11.110)$$

$$\nabla \times c \mathbf{B} + i k \mathbf{E} = \mu_0 c (\mathbf{j} + \nabla \times \mathbf{M}), \quad (11.111)$$

whereas the charge continuity equation takes the form

$$i \omega \rho = \nabla \cdot \mathbf{j}. \quad (11.112)$$

It is convenient to deal only with divergence-free fields. Thus, we use as our field variables, \mathbf{B} and

$$\mathbf{E}' = \mathbf{E} + \frac{i}{\epsilon_0 \omega} \mathbf{j}. \quad (11.113)$$

In the region external to the sources, \mathbf{E}' reduces to \mathbf{E} . When expressed in terms of these fields, Maxwell's equations become

$$\nabla \cdot \mathbf{E}' = 0, \quad (11.114)$$

$$\nabla \cdot \mathbf{B} = 0, \quad (11.115)$$

$$\nabla \times \mathbf{E}' - i k c \mathbf{B} = \frac{i}{\epsilon_0 \omega} \nabla \times \mathbf{j}, \quad (11.116)$$

$$\nabla \times c \mathbf{B} + i k \mathbf{E}' = \mu_0 c \nabla \times \mathbf{M}. \quad (11.117)$$

The curl equations can be combined to give two inhomogeneous Helmholtz wave equations:

$$(\nabla^2 + k^2) c \mathbf{B} = -\mu_0 c \nabla \times (\mathbf{j} + \nabla \times \mathbf{M}), \quad (11.118)$$

and

$$(\nabla^2 + k^2) \mathbf{E}' = -i k \mu_0 c \nabla \times \left(\mathbf{M} + \frac{\nabla \times \mathbf{j}}{k^2} \right). \quad (11.119)$$

These equations, together with $\nabla \cdot \mathbf{B} = 0$, and $\nabla \cdot \mathbf{E}' = 0$, as well as the curl equations giving \mathbf{E}' in terms of \mathbf{B} , and vice versa, are the generalizations of Equations (11.45)–(11.50) when sources are present.

Because the multipole coefficients in Equations (11.69)–(11.70) are determined, via Equations (11.73)–(11.74), from the scalars $\mathbf{r} \cdot c \mathbf{B}$ and $\mathbf{r} \cdot \mathbf{E}'$, it is sufficient to consider wave equations for these quantities, rather than the vector fields \mathbf{B} and \mathbf{E}' . From Equations (11.52), (11.118), (11.119), and the identity

$$\mathbf{r} \cdot (\nabla \times \mathbf{A}) = (\mathbf{r} \times \nabla) \cdot \mathbf{A} = i \mathbf{L} \cdot \mathbf{A}, \quad (11.120)$$

which holds for any vector field \mathbf{A} , we obtain the inhomogeneous wave equations

$$(\nabla^2 + k^2) \mathbf{r} \cdot c \mathbf{B} = -i \mu_0 c \mathbf{L} \cdot (\mathbf{j} + \nabla \times \mathbf{M}), \quad (11.121)$$

$$(\nabla^2 + k^2) \mathbf{r} \cdot \mathbf{E}' = k \mu_0 c \mathbf{L} \cdot \left(\mathbf{M} + \frac{\nabla \times \mathbf{j}}{k^2} \right). \quad (11.122)$$

Now, the Green's function for the inhomogeneous Helmholtz equation, subject to the boundary condition of outgoing waves at infinity, is given by Equation (11.99). It follows that Equations (11.121)–(11.122) can be inverted to give

$$\mathbf{r} \cdot c \mathbf{B}(\mathbf{r}) = \frac{i \mu_0 c}{4\pi} \int \frac{e^{ik|\mathbf{r}-\mathbf{r}'|}}{|\mathbf{r}-\mathbf{r}'|} \mathbf{L}' \cdot [\mathbf{j}(\mathbf{r}') + \nabla' \times \mathbf{M}(\mathbf{r}')] dV', \quad (11.123)$$

$$\mathbf{r} \cdot \mathbf{E}'(\mathbf{r}) = -\frac{k \mu_0 c}{4\pi} \int \frac{e^{ik|\mathbf{r}-\mathbf{r}'|}}{|\mathbf{r}-\mathbf{r}'|} \mathbf{L}' \cdot \left[\mathbf{M}(\mathbf{r}') + \frac{\nabla' \times \mathbf{j}(\mathbf{r}')}{k^2} \right] dV'. \quad (11.124)$$

In order to evaluate the multipole coefficients by means of Equations (11.73)–(11.74), we first observe that the requirement of outgoing waves at infinity implies that $A_l^{(2)} = 0$ in Equation (11.57). Thus, in Equations (11.69)–(11.70), we choose $f_l(kr) = g_l(kr) = h_l^{(1)}(kr)$ as the radial eigenfunctions of \mathbf{E} and \mathbf{B} in the source-free region. Next, let us consider the expansion (11.107) of the Green's function for the inhomogeneous Helmholtz equation. We assume that the point \mathbf{r} lies outside some spherical shell that completely encloses the sources. It follows that $r_< = r'$ and $r_> = r$ in all of the integrations. Making use of the orthogonality property of the spherical harmonics, it follows from Equation (11.107) that

$$\oint Y_{lm}^*(\theta, \varphi) \frac{e^{ik|\mathbf{r}-\mathbf{r}'|}}{4\pi|\mathbf{r}-\mathbf{r}'|} d\Omega = ik h_l^{(1)}(kr) j_l(kr') Y_{lm}^*(\theta', \varphi'). \quad (11.125)$$

Finally, Equations (11.73)–(11.74), and (11.123)–(11.125) yield

$$a_E(l, m) = \frac{\mu_0 c i k^3}{\sqrt{l(l+1)}} \int j_l(kr) Y_{lm}^* \mathbf{L} \cdot \left(\mathbf{M} + \frac{\nabla \times \mathbf{j}}{k^2} \right) dV, \quad (11.126)$$

$$a_M(l, m) = -\frac{\mu_0 c k^2}{\sqrt{l(l+1)}} \int j_l(kr) Y_{lm}^* \mathbf{L} \cdot (\mathbf{j} + \nabla \times \mathbf{M}) dV. \quad (11.127)$$

The previous two equations allow us to calculate the strengths of the various multipole fields, external to the source region, in terms of integrals over the source densities, \mathbf{j} and \mathbf{M} . These

equations can be transformed into more useful forms by means of the following arguments. The results

$$\mathbf{L} \cdot \mathbf{A} = i \nabla \cdot (\mathbf{r} \times \mathbf{A}), \quad (11.128)$$

$$\mathbf{L} \cdot (\nabla \times \mathbf{A}) = i \nabla^2 (\mathbf{r} \cdot \mathbf{A}) - i \frac{1}{r} \frac{\partial (r^2 \nabla \cdot \mathbf{A})}{\partial r} \quad (11.129)$$

follow from the definition of \mathbf{L} [see (11.28)], and simple vector identities. Substituting into Equation (11.126), we obtain

$$a_E(l, m) = -\frac{\mu_0 c k^3}{\sqrt{l(l+1)}} \int j_l(kr) Y_{lm}^* \left[\nabla \cdot (\mathbf{r} \times \mathbf{M}) + \frac{\nabla^2 (\mathbf{r} \cdot \mathbf{j})}{k^2} - i \frac{c}{kr} \frac{\partial (r^2 \rho)}{\partial r} \right] dV, \quad (11.130)$$

where use has been made of Equation (11.112). Use of Green's theorem on the second term in square brackets allows us to replace ∇^2 by $-k^2$ (because we can neglect surface terms, and $j_l(kr) Y_{lm}^*$ is a solution of the Helmholtz equation). A radial integration by parts on the third term (again neglecting surface terms) cause the radial derivative to operate on the spherical Bessel function. The resulting expression for the *electric multipole coefficient* is

$$a_E(l, m) = \frac{\mu_0 c k^2}{i \sqrt{l(l+1)}} \int Y_{lm}^* \left[c \rho \frac{d[r j_l(kr)]}{dr} + i k (\mathbf{r} \cdot \mathbf{j}) j_l(kr) - i k \nabla \cdot (\mathbf{r} \times \mathbf{M}) j_l(kr) \right] dV. \quad (11.131)$$

Similarly, Equation (11.127) leads to the following expression for the *magnetic multipole coefficient*:

$$a_M(l, m) = \frac{\mu_0 c k^2}{i \sqrt{l(l+1)}} \int Y_{lm}^* \left[\nabla \cdot (\mathbf{r} \times \mathbf{j}) j_l(kr) + \nabla \cdot \mathbf{M} \frac{d[r j_l(kr)]}{dr} - k^2 (\mathbf{r} \cdot \mathbf{M}) j_l(kr) \right] dV. \quad (11.132)$$

Both of the previous results are exact, and are valid for arbitrary wavelength and source size.

In the limit in which the source dimensions are small compared to a wavelength (i.e., $kr \ll 1$), the above expressions for the multipole coefficients can be considerably simplified. Using the asymptotic form (11.18), and retaining only lowest powers in kr for terms involving ρ , \mathbf{j} , and \mathbf{M} , we obtain the approximate electric multipole coefficient

$$a_E(l, m) \simeq \frac{\mu_0 c k^{l+2}}{i (2l+1)!!} \left(\frac{l+1}{l} \right)^{1/2} (Q_{lm} + Q'_{lm}), \quad (11.133)$$

where the multipole moments are

$$Q_{lm} = \int r^l Y_{lm}^* c \rho dV, \quad (11.134)$$

$$Q'_{lm} = \frac{-ik}{l+1} \int r^l Y_{lm}^* \nabla \cdot (\mathbf{r} \times \mathbf{M}) dV. \quad (11.135)$$

The moment Q_{lm} has the same form as a conventional electrostatic multipole moment. The moment Q'_{lm} is an induced electric multipole moment due to the magnetization. The latter moment is

generally a factor kr smaller than the former. For the magnetic multipole coefficient $a_M(l, m)$, the corresponding long wavelength approximation is

$$a_M(l, m) \simeq \frac{\mu_0 c}{(2l+1)!!} \left(\frac{l+1}{l} \right)^{1/2} (\mathcal{M}_{lm} + \mathcal{M}'_{lm}), \quad (11.136)$$

where the magnetic multipole moments are

$$\mathcal{M}_{lm} = -\frac{1}{l+1} \int r^l Y_{lm}^* \nabla \cdot (\mathbf{r} \times \mathbf{j}) dV, \quad (11.137)$$

$$\mathcal{M}'_{lm} = - \int r^l Y_{lm}^* \nabla \cdot \mathbf{M} dV. \quad (11.138)$$

Note that for a system with intrinsic magnetization, the magnetic moments \mathcal{M}_{lm} and \mathcal{M}'_{lm} are generally of the same order of magnitude. We conclude that, in the long wavelength limit, the electric multipole fields are determined by the charge density, ρ , whereas the magnetic multipole fields are determined by the magnetic moment densities, $\mathbf{r} \times \mathbf{j}$ and \mathbf{M} .

11.8 Radiation from Linear Centre-Fed Antenna

As an illustration of the use of a multipole expansion for a source whose dimensions are comparable to a wavelength, consider the radiation from a linear centre-fed antenna. We assume that the antenna runs along the z -axis, and extends from $z = -d/2$ to $z = d/2$. The current flowing along the antenna vanishes at the end points, and is an even function of z . Thus, we can write

$$I(z, t) = I(|z|) e^{-i\omega t}, \quad (11.139)$$

where $I(d/2) = 0$. Because the current flows radially, $\mathbf{r} \times \mathbf{j} = 0$. Furthermore, there is no intrinsic magnetization. Thus, according to Equations (11.137)–(11.138), all of the magnetic multipole coefficients, $a_M(l, m)$, vanish. In order to calculate the electric multipole coefficients, $a_E(l, m)$, we need expressions for the charge and current densities. In spherical polar coordinates, the current density \mathbf{j} can be written in the form

$$\mathbf{j}(\mathbf{r}) = \frac{I(r)}{2\pi r^2} [\delta(\cos \theta - 1) - \delta(\cos \theta + 1)] \mathbf{e}_r, \quad (11.140)$$

for $r < d/2$, where the delta functions cause the current to flow only upwards and downwards along the z -axis. From the continuity equation (11.112), the charge density is given by

$$\rho(\mathbf{r}) = \frac{1}{i\omega} \frac{dI(r)}{dr} \left[\frac{\delta(\cos \theta - 1) - \delta(\cos \theta + 1)}{2\pi r^2} \right], \quad (11.141)$$

for $r < d/2$.

The above expressions for \mathbf{j} and ρ can be substituted into Equation (11.133) to give

$$a_E(l, m) = \frac{\mu_0 c k^2}{2\pi \sqrt{l(l+1)}} \int_0^{d/2} \left\{ kr j_l(kr) I(r) - \frac{1}{k} \frac{dI(r)}{dr} \frac{d[r j_l(kr)]}{dr} \right\} dr \oint Y_{lm}^* [\delta(\cos \theta - 1) - \delta(\cos \theta + 1)] d\Omega. \quad (11.142)$$

The angular integral has the value

$$\oint Y_{lm}^* [\delta(\cos \theta - 1) - \delta(\cos \theta + 1)] d\Omega = 2\pi \delta_{m0} [Y_{l0}(0) - Y_{l0}(\pi)], \quad (11.143)$$

showing that only $m = 0$ multipoles are generated. This is hardly surprising, given the cylindrical symmetry of the antenna. The $m = 0$ spherical harmonics are even (odd) about $\theta = \pi/2$ for l even (odd). Hence, the only nonvanishing multipoles have l odd, and the angular integral reduces to

$$\oint Y_{lm}^* [\delta(\cos \theta - 1) - \delta(\cos \theta + 1)] d\Omega = \sqrt{4\pi(2l+1)}. \quad (11.144)$$

After some slight rearrangement, Equation (11.142) can be written

$$a_E(l, 0) = \frac{\mu_0 c k}{2\pi} \left[\frac{4\pi(2l+1)}{l(l+1)} \right]^{1/2} \int_0^{d/2} \left\{ -\frac{d}{dr} \left[r j_l(kr) \frac{dI}{dr} \right] + r j_l(kr) \left(\frac{d^2 I}{dr^2} + k^2 I \right) \right\} dr. \quad (11.145)$$

In order to evaluate the above integral, we need to specify the current $I(z)$ along the antenna. In the absence of radiation, the sinusoidal time variation at frequency ω implies a sinusoidal space variation with wavenumber $k = \omega/c$. However, the emission of radiation generally modifies the current distribution. The correct current $I(z)$ can only be found by solving a complicated boundary value problem. For the sake of simplicity, we assume that $I(z)$ is a known function: specifically,

$$I(z) = I \sin(kd/2 - k|z|), \quad (11.146)$$

for $z < d/2$, where I is the peak current. With a sinusoidal current, the second term in curly brackets in Equation (11.145) vanishes. The first term is a perfect differential. Consequently, Equations (11.145) and (11.146) yield

$$a_E(l, 0) = \frac{\mu_0 c I}{\pi d} \left[\frac{4\pi(2l+1)}{l(l+1)} \right]^{1/2} \left(\frac{kd}{2} \right)^2 j_l(kd/2), \quad (11.147)$$

for l odd.

Let us consider the special cases of a half-wave antenna (i.e., $kd = \pi$, so that the length of the antenna is half a wavelength) and a full-wave antenna (i.e., $kd = 2\pi$). For these two values of kd , the $l = 1$ electric multipole coefficient is tabulated in Table 11.1, along with the relative values for $l = 3$ and 5. It is clear, from the table, that the coefficients decrease rapidly in magnitude as l increases, and that higher l coefficients are more important the larger the source dimensions. However, even for a full-wave antenna, it is generally sufficient to retain only the $l = 1$ and $l = 3$

$k d$	$a_E(1, 0)$	$a_E(3, 0)/a_E(1, 0)$	$a_E(5, 0)/a_E(1, 0)$
π	$4 \sqrt{6\pi} (\mu_0 c I / 4\pi d)$	4.95×10^{-2}	1.02×10^{-3}
2π	$4\pi \sqrt{6\pi} (\mu_0 c I / 4\pi d)$	3.25×10^{-1}	3.09×10^{-2}

Table 11.1: The first few electric multipole coefficients for a half-wave and a full-wave antenna.

coefficients when calculating the angular distribution of the radiation. It is certainly adequate to keep only these two harmonics when calculating the total radiated power (which depends on the sum of the squares of the coefficients).

In the radiation zone, the multipole fields (11.69)–(11.70) reduce to

$$c \mathbf{B} \simeq \frac{e^{i(kr-\omega t)}}{kr} \sum_{l,m} (-i)^{l+1} [a_E(l, m) \mathbf{X}_{lm} + a_M(l, m) \mathbf{n} \times \mathbf{X}_{lm}], \quad (11.148)$$

$$\mathbf{E} \simeq c \mathbf{B} \times \mathbf{n}, \quad (11.149)$$

where use has been made of the asymptotic form (11.22). The time-averaged power radiated per unit solid angle is given by

$$\frac{dP}{d\Omega} = \frac{\text{Re}(\mathbf{n} \cdot \mathbf{E} \times \mathbf{B}^*) r^2}{2\mu_0}, \quad (11.150)$$

or

$$\frac{dP}{d\Omega} = \frac{1}{2\mu_0 c k^2} \left| \sum_{l,m} (-i)^{l+1} [a_E(l, m) \mathbf{X}_{lm} + a_M(l, m) \mathbf{n} \times \mathbf{X}_{lm}] \right|^2. \quad (11.151)$$

Retaining only the $l = 1$ and $l = 3$ electric multipole coefficients, the angular distribution of the radiation from the antenna takes the form

$$\frac{dP}{d\Omega} = \frac{|a_E(l, 0)|^2}{4\mu_0 c k^2} \left| \mathbf{LY}_{1,0} - \frac{a_E(3, 0)}{\sqrt{6} a_E(1, 0)} \mathbf{LY}_{3,0} \right|^2, \quad (11.152)$$

where use has been made of Equation (11.66). The various factors in the absolute square are

$$|\mathbf{LY}_{1,0}|^2 = \frac{3}{4\pi} \sin^2 \theta, \quad (11.153)$$

$$|\mathbf{LY}_{3,0}|^2 = \frac{63}{16\pi} \sin^2 \theta (5 \cos^2 \theta - 1)^2, \quad (11.154)$$

$$(\mathbf{LY}_{1,0})^* \cdot (\mathbf{LY}_{3,0}) = \frac{3\sqrt{21}}{8\pi} \sin^2 \theta (5 \cos^2 \theta - 1). \quad (11.155)$$

With these angular factors, Equation (11.151) becomes

$$\frac{dP}{d\Omega} = \lambda \frac{3\mu_0 c I^2}{\pi^3} \frac{3 \sin^2 \theta}{8\pi} \left| 1 - \sqrt{\frac{7}{8}} \frac{a_E(3, 0)}{a_E(1, 0)} (5 \cos^2 \theta - 1) \right|^2, \quad (11.156)$$

where λ equals 1 for a half-wave antenna, and $\pi^2/4$ for a full-wave antenna. The coefficient in front of $(5 \cos^2 \theta - 1)$ is 0.0463 and 0.304 for the half-wave and full-wave antenna, respectively. It turns out that the radiation pattern obtained from the two-term multipole expansion specified in the previous equation is almost indistinguishable from the exact result for the case of a half-wave antenna. For the case of a full-wave antenna, the two-term expansion yields a radiation pattern that differs from the exact result by less than 5 percent.

The total power radiated by the antenna is

$$P = \frac{1}{2\mu_0 c k^2} \sum_{l \text{ odd}} |a_E(l, 0)|^2, \quad (11.157)$$

where use has been made of Equation (11.90). It is evident from Table 11.1 that a two-term multipole expansion gives an accurate expression for the power radiated by both a half-wave and a full-wave antenna. In fact, a one-term multipole expansion gives a fairly accurate result for the case of a half-wave antenna.

It is clear, from the previous analysis, that the multipole expansion converges rapidly when the source dimensions are of order the wavelength of the radiation. It is also clear that if the source dimensions are much less than the wavelength then the multipole expansion is likely to be completely dominated by the term corresponding to the lowest value of l .

11.9 Spherical Wave Expansion of Vector Plane Wave

In discussing the scattering or absorption of electromagnetic radiation by localized systems, it is useful to be able to express a plane electromagnetic wave as a superposition of spherical waves.

Consider, first of all, the expansion of a scalar plane wave as a set of scalar spherical waves. This expansion is conveniently obtained from the expansion (11.107) for the Green's function of the scalar Helmholtz equation. Let us take the limit $|\mathbf{r}'| \rightarrow \infty$ of this equation. We can make the substitution $|\mathbf{r} - \mathbf{r}'| \simeq r' - \mathbf{n} \cdot \mathbf{r}$ on the left-hand-side, where \mathbf{n} is a unit vector pointing in the direction of \mathbf{r}' . On the right-hand side, $r_< = r$ and $r_> = r'$. Furthermore, we can use the asymptotic form (11.22) for $h_l^{(1)}(kr)$. Thus, we obtain

$$\frac{e^{ikr'}}{4\pi r'} e^{-i\mathbf{k} \cdot \mathbf{n} \cdot \mathbf{r}} = i k \frac{e^{ikr'}}{k r'} \sum_{l,m} (-i)^{l+1} j_l(kr) Y_{lm}^*(\theta', \varphi') Y_{lm}(\theta, \varphi). \quad (11.158)$$

Canceling the factor $e^{ikr'}/r'$ on either side of the above equation, and taking the complex conjugate, we get the following expansion for a scalar plane wave,

$$e^{i\mathbf{k} \cdot \mathbf{r}} = 4\pi \sum_{l=0,\infty} i^l j_l(kr) \sum_{m=-l,+l} Y_{lm}^*(\theta, \varphi) Y_{lm}(\theta', \varphi'), \quad (11.159)$$

where \mathbf{k} is a wavevector with the spherical coordinates k, θ', φ' . The well-known addition theorem for the spherical harmonics states that

$$P_l(\cos \gamma) = \frac{4\pi}{2l+1} \sum_{m=-l,+l} Y_{lm}^*(\theta, \varphi) Y_{lm}(\theta', \varphi'), \quad (11.160)$$

where γ is the angle subtended between the vectors \mathbf{r} and \mathbf{r}' . Consequently,

$$\cos \gamma = \cos \theta \cos \theta' + \sin \theta \sin \theta' \cos(\varphi - \varphi'). \quad (11.161)$$

It follows from Equations (11.159) and (11.160) that

$$e^{i\mathbf{k}\cdot\mathbf{r}} = \sum_{l=0,\infty} i^l (2l+1) j_l(kr) P_l(\cos \gamma), \quad (11.162)$$

or

$$e^{i\mathbf{k}\cdot\mathbf{r}} = \sum_{l=0,\infty} i^l \sqrt{4\pi(2l+1)} j_l(kr) Y_{l0}(\gamma), \quad (11.163)$$

because

$$Y_{l0}(\theta) = \sqrt{\frac{2l+1}{4\pi}} P_l(\cos \theta). \quad (11.164)$$

Let us now make an equivalent expansion for a circularly polarized plane wave incident along the z -axis:

$$\mathbf{E}(\mathbf{r}) = (\mathbf{e}_x \pm i\mathbf{e}_y) e^{ikz}, \quad (11.165)$$

$$c\mathbf{B}(\mathbf{r}) = \mathbf{e}_z \times \mathbf{E} = \mp i\mathbf{E}. \quad (11.166)$$

Because the plane wave is finite everywhere (including the origin), its multipole expansion (11.69)–(11.70) can only involve the well-behaved radial eigenfunctions $j_l(kr)$. Thus,

$$\mathbf{E} = \sum_{l,m} \left[a_{\pm}(l,m) j_l(kr) \mathbf{X}_{lm} + \frac{i}{k} b_{\pm}(l,m) \nabla \times j_l(kr) \mathbf{X}_{lm} \right], \quad (11.167)$$

$$c\mathbf{B} = \sum_{l,m} \left[\frac{-i}{k} a_{\pm}(l,m) \nabla \times j_l(kr) \mathbf{X}_{lm} + b_{\pm}(l,m) j_l(kr) \mathbf{X}_{lm} \right]. \quad (11.168)$$

To determine the coefficients $a_{\pm}(l,m)$ and $b_{\pm}(l,m)$, we make use of a slight generalization of the standard orthogonality properties (11.67)–(11.68) of the vector spherical harmonics:

$$\oint [f_l(r) \mathbf{X}_{l'm'}]^* \cdot [g_l(r) \mathbf{X}_{lm}] d\Omega = f_l^* g_l \delta_{l'l} \delta_{m'm}, \quad (11.169)$$

$$\oint [f_l(r) \mathbf{X}_{l'm'}]^* \cdot [\nabla \times g_l(r) \mathbf{X}_{lm}] d\Omega = 0. \quad (11.170)$$

The first of these follows directly from Equation (11.67). The second follows from Equations (11.32), (11.68), (11.76), and the identity

$$\nabla \equiv \frac{\mathbf{r}}{r} \frac{\partial}{\partial r} - \frac{i}{r^2} \mathbf{r} \times \mathbf{L}. \quad (11.171)$$

The coefficients $a_{\pm}(l,m)$ and $b_{\pm}(l,m)$ are obtained by taking the scalar product of Equations (11.167)–(11.168) with \mathbf{X}_{lm}^* and integrating over all solid angle, making use of the orthogonality relations

(11.169)–(11.170). This yields

$$a_{\pm}(l, m) j_l(kr) = \oint \mathbf{X}_{lm}^* \cdot \mathbf{E} d\Omega, \quad (11.172)$$

$$b_{\pm}(l, m) j_l(kr) = \oint \mathbf{X}_{lm}^* \cdot \mathbf{B} d\Omega. \quad (11.173)$$

Substitution of Equations (11.66) and (11.167) into Equation (11.172) gives

$$a_{\pm}(l, m) j_l(kr) = \oint \frac{(L_{\mp} Y_{lm})^*}{\sqrt{l(l+1)}} e^{ikz} d\Omega, \quad (11.174)$$

where the operators L_{\pm} are defined in Equations (11.29)–(11.30). Making use of Equations (11.34)–(11.36), the above expression reduces to

$$a_{\pm}(l, m) j_l(kr) = \frac{\sqrt{(l \pm m)(l \mp m + 1)}}{\sqrt{l(l+1)}} \oint Y_{l, m \mp 1}^* e^{ikz} d\Omega. \quad (11.175)$$

If the expansion (11.163) is substituted for e^{ikz} , and use is made of the orthogonality properties of the spherical harmonics, then we obtain the result

$$a_{\pm}(l, m) = i^l \sqrt{4\pi(2l+1)} \delta_{m, \pm 1}. \quad (11.176)$$

It is clear from Equations (11.166) and (11.173) that

$$b_{\pm}(l, m) = \mp i a_{\pm}(l, m). \quad (11.177)$$

Thus, the general expansion of a circularly polarized plane wave takes the form

$$\mathbf{E}(\mathbf{r}) = \sum_{l=1, \infty} i^l \sqrt{4\pi(2l+1)} \left[j_l(kr) \mathbf{X}_{l, \pm 1} \pm \frac{1}{k} \nabla \times j_l(kr) \mathbf{X}_{l, \pm 1} \right], \quad (11.178)$$

$$\mathbf{B}(\mathbf{r}) = \sum_{l=1, \infty} i^l \sqrt{4\pi(2l+1)} \left[\frac{-i}{k} \nabla \times j_l(kr) \mathbf{X}_{l, \pm 1} \mp i j_l(kr) \mathbf{X}_{l, \pm 1} \right]. \quad (11.179)$$

The expansion for a linearly polarized plane wave is easily obtained by taking the appropriate linear combination of the above two expansions.

11.10 Mie Scattering

Consider a plane electromagnetic wave incident on a spherical obstacle. In general, the wave is scattered, to some extent, by the obstacle. Thus, far away from the sphere, the electromagnetic fields can be expressed as the sum of a plane wave and a set of outgoing spherical waves. There may be absorption by the obstacle, as well as scattering. In this case, the energy flow away from the obstacle is less than the total energy flow towards it—the difference representing the absorbed energy.

The fields outside the sphere can be written as the sum of incident and scattered waves:

$$\mathbf{E}(\mathbf{r}) = \mathbf{E}_{\text{inc}} + \mathbf{E}_{\text{sc}}, \quad (11.180)$$

$$\mathbf{B}(\mathbf{r}) = \mathbf{B}_{\text{inc}} + \mathbf{B}_{\text{sc}}, \quad (11.181)$$

where \mathbf{E}_{inc} and \mathbf{B}_{inc} are given by (11.178)–(11.179). Because the scattered fields are outgoing waves at infinity, their expansions must be of the form

$$\mathbf{E}_{\text{sc}} = \frac{1}{2} \sum_{l=1, \infty} i^l \sqrt{4\pi(2l+1)} \left[\alpha_{\pm}(l) h_l^{(1)}(kr) \mathbf{X}_{l, \pm 1} \pm \frac{\beta_{\pm}(l)}{k} \nabla \times h_l^{(1)}(kr) \mathbf{X}_{l, \pm 1} \right], \quad (11.182)$$

$$c \mathbf{B}_{\text{sc}} = \frac{1}{2} \sum_{l=1, \infty} i^l \sqrt{4\pi(2l+1)} \left[\frac{-i \alpha_{\pm}(l)}{k} \nabla \times h_l^{(1)}(kr) \mathbf{X}_{l, \pm 1} \mp i \beta_{\pm}(l) h_l^{(1)}(kr) \mathbf{X}_{l, \pm 1} \right]. \quad (11.183)$$

The coefficients $\alpha_{\pm}(l)$ and $\beta_{\pm}(l)$ are determined by the boundary conditions on the surface of the sphere. In general, it is necessary to sum over all m harmonics in the above expressions. However, for the restricted class of spherically symmetric scatterers, only the $m = \pm 1$ harmonics need be retained (because only these harmonics occur in the spherical wave expansion of the incident plane wave [see Equations (11.178)–(11.179)], and a spherically symmetric scatterer does not couple different m harmonics).

The angular distribution of the scattered power can be written in terms of the coefficients $\alpha(l)$ and $\beta(l)$ using the scattered electromagnetic fields evaluated on the surface of a sphere of radius a surrounding the scatterer. In fact, it is easily demonstrated that

$$\frac{dP_{\text{sc}}}{d\Omega} = \frac{a^2}{2\mu_0} \text{Re} [\mathbf{n} \cdot \mathbf{E}_{\text{sc}} \times \mathbf{B}_{\text{sc}}^*]_{r=a} = -\frac{a^2}{2\mu_0} \text{Re} [\mathbf{E}_{\text{sc}} \cdot (\mathbf{n} \times \mathbf{B}_{\text{sc}}^*)]_{r=a}, \quad (11.184)$$

where \mathbf{n} is a radially directed outward normal. The differential scattering cross-section is defined as the ratio of $dP_{\text{sc}}/d\Omega$ to the incident flux $1/(\mu_0 c)$. Hence,

$$\frac{d\sigma_{\text{sc}}}{d\Omega} = -\frac{a^2}{2} \text{Re} [\mathbf{E}_{\text{sc}} \cdot (\mathbf{n} \times c \mathbf{B}_{\text{sc}}^*)]_{r=a}. \quad (11.185)$$

We need to evaluate this expression using the electromagnetic fields specified in Equations (11.178)–(11.183). The following identity, which can be established with the aid of Equations (11.28), (11.66), and (11.76), is helpful in this regard:

$$\nabla \times f(r) \mathbf{X}_{lm} = \frac{i \sqrt{l(l+1)}}{r} f(r) Y_{lm} \mathbf{n} + \frac{1}{r} \frac{d[r f(r)]}{dr} \mathbf{n} \times \mathbf{X}_{lm}. \quad (11.186)$$

For instance, using this result, we can write $\mathbf{n} \times c \mathbf{B}_{\text{sc}}$ in the form

$$\mathbf{n} \times c \mathbf{B}_{\text{sc}} = \frac{1}{2} \sum_{l=1, \infty} i^l \sqrt{4\pi(2l+1)} \left[\frac{i \alpha_{\pm}(l)}{k} \frac{1}{r} \frac{d[r h_l^{(1)}(kr)]}{dr} \mathbf{X}_{l, \pm 1} \mp i \beta_{\pm}(l) h_l^{(1)}(kr) \mathbf{n} \times \mathbf{X}_{l, \pm 1} \right]. \quad (11.187)$$

It can be demonstrated, after considerable algebra, that

$$\frac{d\sigma_{\text{sc}}}{d\Omega} = \frac{\pi}{2k^2} \left| \sum_l \sqrt{2l+1} [\alpha_{\pm}(l) \mathbf{X}_{l,\pm 1} \pm i\beta_{\pm}(l) \mathbf{n} \times \mathbf{X}_{l,\pm 1}] \right|^2. \quad (11.188)$$

In obtaining this formula, use has been made of the standard result

$$\frac{df_l(z)}{dz} f_{l'}^*(z) - \frac{df_{l'}^*(z)}{dz} f_l(z) = \frac{2i}{z^2}, \quad (11.189)$$

where $f_l(z) = i^l h_l^{(1)}(z)$. The total scattering cross-section is obtained by integrating Equation (11.188) over all solid angle, making use of the following orthogonality relations for the vector spherical harmonics [see Equations (11.67)–(11.68)]:

$$\oint \mathbf{X}_{l'm'}^* \cdot \mathbf{X}_{lm} d\Omega = \delta_{ll'} \delta_{mm'}, \quad (11.190)$$

$$\oint \mathbf{X}_{l'm'}^* \cdot (\mathbf{n} \times \mathbf{X}_{lm}) d\Omega = 0, \quad (11.191)$$

$$\oint (\mathbf{n} \times \mathbf{X}_{l'm'}^*) \cdot (\mathbf{n} \times \mathbf{X}_{lm}) d\Omega = \delta_{ll'} \delta_{mm'}. \quad (11.192)$$

Thus,

$$\sigma_{\text{sc}} = \frac{\pi}{2k^2} \sum_l (2l+1) [|\alpha_{\pm}(l)|^2 + |\beta_{\pm}(l)|^2]. \quad (11.193)$$

According to Equations (11.188) and (11.193), the total scattering cross-section is independent of the polarization of the incident radiation (i.e., it is the same for both the \pm signs). However, the differential scattering cross-section in any particular direction is, in general, different for different circular polarizations of the incident radiation. This implies that if the incident radiation is linearly polarized then the scattered radiation is elliptically polarized. Furthermore, if the incident radiation is unpolarized then the scattered radiation exhibits partial polarization, with the degree of polarization depending on the angle of observation.

The total power absorbed by the sphere is given by

$$P_{\text{abs}} = -\frac{a^2}{2\mu_0} \text{Re} \int [\mathbf{n} \cdot \mathbf{E} \times \mathbf{B}^*]_{r=a} d\Omega = \frac{a^2}{2\mu_0} \text{Re} \int [\mathbf{E} \cdot (\mathbf{n} \times \mathbf{B}^*)]_{r=a} d\Omega.$$

A similar calculation to that outlined above yields the following expression for the absorption cross-section,

$$\sigma_{\text{abs}} = \frac{\pi}{2k^2} \sum_l (2l+1) [2 - |\alpha_{\pm}(l) + 1|^2 - |\beta_{\pm}(l) + 1|^2]. \quad (11.194)$$

The total, or extinction, cross-section is the sum of σ_{sc} and σ_{abs} :

$$\sigma_{\text{t}} = -\frac{\pi}{k^2} \sum_l (2l+1) \text{Re} [\alpha_{\pm}(l) + \beta_{\pm}(l)]. \quad (11.195)$$

Not surprisingly, the above expressions for the various cross-sections closely resemble those obtained in quantum mechanics from partial wave expansions.

Let us now consider the boundary conditions at the surface of the sphere (whose radius is a , say). For the sake of simplicity, let us suppose that the sphere is a perfect conductor. In this case, the appropriate boundary condition is that the tangential electric field is zero at $r = a$. According to Equations (11.178), (11.180), and (11.186), the tangential electric field is given by

$$\mathbf{E}_{\text{tan}} = \sum_l i^l \sqrt{4\pi(2l+1)} \left\{ \left[j_l + \frac{\alpha_{\pm}(l)}{2} h_l^{(1)} \right] \mathbf{X}_{l,\pm 1} \pm \frac{1}{x} \frac{d}{dx} \left[x \left(j_l + \frac{\beta_{\pm}(l)}{2} h_l^{(1)} \right) \right] \mathbf{n} \times \mathbf{X}_{l,\pm 1} \right\}, \quad (11.196)$$

where $x = ka$, and all of the spherical Bessel functions have the argument x . Thus, the boundary condition yields

$$\alpha_{\pm}(l) + 1 = -\frac{h_l^{(2)}(ka)}{h_l^{(1)}(ka)}, \quad (11.197)$$

$$\beta_{\pm}(l) + 1 = -\left(\frac{[x h_l^{(2)}(x)]'}{[x h_l^{(1)}(x)]'} \right)_{x=ka}, \quad (11.198)$$

where $'$ denotes d/dx . Note that $\alpha_{\pm}(l) + 1$ and $\beta_{\pm}(l) + 1$ are both numbers of modulus unity. This implies, from Equation (11.194), that there is no absorption for the case of a perfectly conducting sphere (in general, there is some absorption if the sphere has a finite conductivity). We can write $\alpha_{\pm}(l)$ and $\beta_{\pm}(l)$ in the form

$$\alpha_{\pm}(l) = e^{2i\delta_l} - 1, \quad (11.199)$$

$$\beta_{\pm}(l) = e^{2i\delta'_l} - 1, \quad (11.200)$$

where the phase angles δ_l and δ'_l are called *scattering phase shifts*. It follows from Equations (11.197)–(11.198) that

$$\tan \delta_l = \frac{j_l(ka)}{y_l(ka)}, \quad (11.201)$$

$$\tan \delta'_l = \left(\frac{[x j_l(x)]'}{[x y_l(x)]'} \right)_{x=ka}. \quad (11.202)$$

Let us specialize to the limit $ka \ll 1$, in which the wavelength of the radiation is much greater than the radius of the sphere. The asymptotic expansions (11.18)–(11.19) yield

$$\begin{aligned} \alpha_{\pm}(l) &\simeq -\frac{2i(ka)^{2l+1}}{(2l+1)[(2l-1)!!]^2}, \\ \beta_{\pm}(l) &\simeq -\frac{(l+1)}{l} \alpha_{\pm}(l), \end{aligned} \quad (11.203)$$

for $l \geq 1$. It is clear that the scattering coefficients $\alpha_{\pm}(l)$ and $\beta_{\pm}(l)$ become small very rapidly as l increases. Thus, in the very long wavelength limit, only the $l = 1$ coefficients need be retained. It is easily seen that

$$\alpha_{\pm}(1) = -\frac{\beta_{\pm}(1)}{2} \simeq -\frac{2i}{3}(ka)^3. \quad (11.204)$$

In this limit, the differential scattering cross-section (11.188) reduces to

$$\frac{d\sigma_{sc}}{d\Omega} \simeq \frac{2\pi}{3} a^2 (ka)^4 |\mathbf{X}_{1,\pm 1} \mp 2i \mathbf{n} \times \mathbf{X}_{1,\pm 1}|^2. \quad (11.205)$$

It can be demonstrated that

$$|\mathbf{n} \times \mathbf{X}_{1,\pm 1}|^2 = |\mathbf{X}_{1,\pm 1}|^2 = \frac{3}{16\pi} (1 + \cos^2 \theta), \quad (11.206)$$

and

$$[\pm i (\mathbf{n} \times \mathbf{X}_{1,\pm 1}^*) \cdot \mathbf{X}_{1,\pm 1}] = -\frac{3\pi}{8} \cos \theta. \quad (11.207)$$

Thus, in long wavelength limit, the differential scattering cross-section reduces to

$$\frac{d\sigma_{sc}}{d\Omega} \simeq a^2 (ka)^4 \left[\frac{5}{8} (1 + \cos^2 \theta) - \cos \theta \right]. \quad (11.208)$$

The scattering is predominately backwards, and is independent of the state of polarization of the incident radiation. The total scattering cross-section is given by

$$\sigma_{sc} = \frac{10\pi}{3} a^2 (ka)^4. \quad (11.209)$$

This well-known result was first obtained by Mie and Debye. Note that the cross-section scales as the inverse fourth power of the wavelength of the incident radiation. This scaling is generic to all scatterers whose dimensions are much smaller than the wavelength. In fact, it was first derived by Rayleigh using dimensional analysis.

11.11 Exercises

11.1 An almost spherical surface defined by

$$R(\theta) = R_0 [1 + \beta P_2(\cos \theta)]$$

has inside it a uniform volume distribution of charge totaling Q . The small parameter β varies harmonically in time at the angular frequency ω . This corresponds to a surface waves on a sphere. Keeping only the lowest-order terms in β , and making the long-wavelength approximation, calculate the nonvanishing multipole moments, the angular distribution of radiation, and the total radiated power.

11.2 The uniform charge density of the previous exercise is replaced by a uniform magnetization parallel to the z -axis and having a total magnetic moment M . With the same approximations as in the previous exercise, calculate the nonvanishing multipole moments, the angular distribution of radiation, and the total radiated power.

12 Relativity and Electromagnetism

12.1 Introduction

In this chapter, we shall discuss Maxwell's equations in the light of Einstein's special theory of relativity.

12.2 Relativity Principle

Physical phenomena are conventionally described relative to some frame of reference that allows us to define fundamental quantities such as position and time. Of course, there are very many different ways of choosing a reference frame, but it is generally convenient to restrict our choice to the set of rigid inertial frames. A classical rigid reference frame is the imagined extension of a rigid body. For instance, the Earth determines a rigid frame throughout all space, consisting of all those points that remain rigidly at rest relative to the Earth, and to one another. We can associate an orthogonal Cartesian coordinate system S with such a frame by choosing three mutually orthogonal planes within it, and measuring x , y , and z as perpendicular distances from these planes. A time coordinate must also be defined, in order that the system can be used to specify events. A rigid frame, endowed with such properties, is called a *Cartesian frame*. The description given previously presupposes that the underlying geometry of space is Euclidian, which is reasonable provided that gravitational effects are negligible (as we shall assume to be the case). An *inertial frame* is a Cartesian frame in which free particles move without acceleration, in accordance with Newton's first law of motion. There are an infinite number of different inertial frames, moving with some constant velocity with respect to one another.

The key to understanding special relativity is Einstein's *Relativity Principle*, which states that:

All inertial frames are totally equivalent for the performance of all physical experiments.

In other words, it is impossible to perform a physical experiment that differentiates in any fundamental sense between different inertial frames. By definition, Newton's laws of motion take the same form in all inertial frames. Einstein generalized this result in his special theory of relativity by asserting that all laws of physics take the same form in all inertial frames.

Consider a wave-like disturbance. In general, such a disturbance propagates at a fixed speed with respect to the medium in which the disturbance takes place. For instance, sound waves (at STP) propagate at 343 meters per second with respect to air. So, in the inertial frame in which air is stationary, sound waves appear to propagate at 343 meters per second. Sound waves appear to propagate at a different speed in any inertial frame that is moving with respect to the air. However, this does not violate the relativity principle, because if the air were stationary in the second frame then sound waves would appear to propagate at 343 meters per second in that frame as well. In other words, exactly the same experiment (e.g., the determination of the speed of sound relative

to stationary air) performed in two different inertial frames of reference yields exactly the same result, in accordance with the relativity principle.

Consider, now, a wave-like disturbance that is self-regenerating, and does not require a medium through which to propagate. The most well-known example of such a disturbance is a light wave. Another example is a gravity wave. According to electromagnetic theory, the speed of propagation of a light wave through a vacuum is

$$c = \frac{1}{\sqrt{\epsilon_0 \mu_0}} = 2.99729 \times 10^8 \text{ meters per second,} \quad (12.1)$$

where ϵ_0 and μ_0 are physical constants that can be evaluated by performing two simple experiments which involve measuring the forces of attraction between two fixed charges and two fixed parallel current carrying wires. According to the relativity principle, these experiments must yield the same values for ϵ_0 and μ_0 in all inertial frames. Thus, the speed of light must be the same in all inertial frames. In fact, any disturbance that does not require a medium to propagate through must appear to travel at the same speed in all inertial frames, otherwise we could differentiate inertial frames using the apparent propagation speed of the disturbance, which would violate the relativity principle.

12.3 Lorentz Transformation

Consider two Cartesian frames $S(x, y, z, t)$ and $S'(x', y', z', t')$ in the standard configuration, in which S' moves in the x -direction of S with uniform velocity v , and the corresponding axes of S and S' remain parallel throughout the motion, having coincided at $t = t' = 0$. It is assumed that the same units of distance and time are adopted in both frames. Suppose that an event (e.g., the flashing of a light-bulb, or the collision of two point particles) has coordinates (x, y, z, t) relative to S , and (x', y', z', t') relative to S' . The “common sense” relationship between these two sets of coordinates is given by the *Galilean transformation*:

$$x' = x - vt, \quad (12.2)$$

$$y' = y, \quad (12.3)$$

$$z' = z, \quad (12.4)$$

$$t' = t. \quad (12.5)$$

This transformation is tried and tested, and provides a very accurate description of our everyday experience. Nevertheless, it must be wrong. Consider a light wave that propagates along the x -axis in S with velocity c . According to the Galilean transformation, the apparent speed of propagation in S' is $c - v$, which violates the relativity principle. Can we construct a new transformation that makes the velocity of light invariant between different inertial frames, in accordance with the relativity principle, but reduces to the Galilean transformation at low velocities, in accordance with our everyday experience?

Consider an event P , and a neighboring event Q , whose coordinates differ by dx, dy, dz, dt in S , and by dx', dy', dz', dt' in S' . Suppose that at the event P a flash of light is emitted, and that Q

is an event in which some particle in space is illuminated by the flash. In accordance with the laws of light propagation, and the invariance of the velocity of light between different inertial frames, an observer in S will find that

$$dx^2 + dy^2 + dz^2 - c^2 dt^2 = 0 \quad (12.6)$$

for $dt > 0$, and an observer in S' will find that

$$dx'^2 + dy'^2 + dz'^2 - c^2 dt'^2 = 0 \quad (12.7)$$

for $dt' > 0$. Any event near P whose coordinates satisfy either (12.6) or (12.7) is illuminated by the flash from P , and, therefore, its coordinates must satisfy both (12.6) and (12.7). Now, no matter what form the transformation between coordinates in the two inertial frames takes, the transformation between differentials at any fixed event P is linear and homogeneous. In other words, if

$$x' = F(x, y, z, t), \quad (12.8)$$

where F is a general function, then

$$dx' = \frac{\partial F}{\partial x} dx + \frac{\partial F}{\partial y} dy + \frac{\partial F}{\partial z} dz + \frac{\partial F}{\partial t} dt. \quad (12.9)$$

It follows that

$$\begin{aligned} dx'^2 + dy'^2 + dz'^2 - c^2 dt'^2 &= a dx^2 + b dy^2 + c dz^2 + d dt^2 \\ &\quad + g dx dt + h dy dt + k dz dt \\ &\quad + l dy dz + m dx dz + n dx dy, \end{aligned} \quad (12.10)$$

where a, b, c , et cetera, are functions of x, y, z , and t . We know that the right-hand side of the previous expression vanishes for all real values of the differentials that satisfy Equation (12.6). It follows that the right-hand side is a multiple of the quadratic in Equation (12.6): that is,

$$dx'^2 + dy'^2 + dz'^2 - c^2 dt'^2 = K(dx^2 + dy^2 + dz^2 - c^2 dt^2), \quad (12.11)$$

where K is a function of x, y, z , and t . [We can prove this by substituting into Equation (12.10) the following obvious zeros of the quadratic in Equation (12.6): $(\pm 1, 0, 0, 1)$, $(0, \pm 1, 0, 1)$, $(0, 0, \pm 1, 1)$, $(0, 1/\sqrt{2}, 1/\sqrt{2}, 1)$, $(1/\sqrt{2}, 0, 1/\sqrt{2}, 1)$, $(1/\sqrt{2}, 1/\sqrt{2}, 0, 1)$: and solving the resulting conditions on the coefficients.] Note that K at P is also independent of the choice of standard coordinates in S and S' . Because the frames are Euclidian, the values of $dx^2 + dy^2 + dz^2$ and $dx'^2 + dy'^2 + dz'^2$ relevant to P and Q are independent of the choice of axes. Furthermore, the values of dt^2 and dt'^2 are independent of the choice of the origins of time. Thus, without affecting the value of K at P , we can choose coordinates such that $P = (0, 0, 0, 0)$ in both S and S' . Because the orientations of the axes in S and S' are, at present, arbitrary, and because inertial frames are isotropic, the relation of S and S' relative to each other, to the event P , and to the locus of possible events Q , is now completely symmetric. Thus, we can write

$$dx^2 + dy^2 + dz^2 - c^2 dt^2 = K(dx'^2 + dy'^2 + dz'^2 - c^2 dt'^2), \quad (12.12)$$

in addition to Equation (12.11). It follows that $K = \pm 1$. $K = -1$ can be dismissed immediately, because the intervals $dx^2 + dy^2 + dz^2 - c^2 dt^2$ and $dx'^2 + dy'^2 + dz'^2 - c^2 dt'^2$ must coincide exactly when there is no motion of S' relative to S . Thus,

$$dx'^2 + dy'^2 + dz'^2 - c^2 dt'^2 = dx^2 + dy^2 + dz^2 - c^2 dt^2. \quad (12.13)$$

Equation (12.13) implies that the transformation equations between primed and unprimed coordinates must be linear. The proof of this statement is postponed until Section 12.7.

The linearity of the transformation allows the coordinate axes in the two frames to be orientated so as to give the standard configuration mentioned previously. Consider a fixed plane in S with the equation $lx + my + nz + p = 0$. In S' , this becomes (say) $l(a_1 x' + b_1 y' + c_1 z' + d_1 t' + e_1) + m(a_2 x' + \dots) + n(a_3 x' + \dots) + p = 0$, which represents a moving plane unless $ld_1 + md_2 + nd_3 = 0$. That is, unless the normal vector to the plane in S , (l, m, n) , is perpendicular to the vector (d_1, d_2, d_3) . All such planes intersect in lines that are fixed in both S and S' , and that are parallel to the vector (d_1, d_2, d_3) in S . These lines must correspond to the direction of relative motion of the frames. By symmetry, two such planes which are orthogonal in S must also be orthogonal in S' . This allows the choice of two common coordinate planes.

Under a linear transformation, the finite coordinate differences satisfy the same transformation equations as the differentials. It follows from Equation (12.13), assuming that the events $(0, 0, 0, 0)$ coincide in both frames, that for any event with coordinates (x, y, z, t) in S and (x', y', z', t') in S' , the following relation holds:

$$x^2 + y^2 + z^2 - c^2 t^2 = x'^2 + y'^2 + z'^2 - c^2 t'^2. \quad (12.14)$$

By hypothesis, the coordinate planes $y = 0$ and $y' = 0$ coincide permanently. Thus, $y = 0$ must imply $y' = 0$, which suggests that

$$y' = Ay, \quad (12.15)$$

where A is a constant. We can reverse the directions of the x - and z -axes in S and S' , which has the effect of interchanging the roles of these frames. This procedure does not affect Equation (12.15), but by symmetry we also have

$$y = Ay'. \quad (12.16)$$

It is clear that $A = \pm 1$. The negative sign can again be dismissed, because $y = y'$ when there is no motion between S and S' . The argument for z is similar. Thus, we have

$$y' = y, \quad (12.17)$$

$$z' = z, \quad (12.18)$$

as in the Galilean transformation.

Equations (12.14), (12.17) and (12.18) yield

$$x^2 - c^2 t^2 = x'^2 - c^2 t'^2. \quad (12.19)$$

Because $x' = 0$ must imply $x = vt$, we can write

$$x' = B(x - vt), \quad (12.20)$$

where B is a constant (possibly depending on v). It follows from the previous two equations that

$$t' = Cx + Dt, \quad (12.21)$$

where C and D are constants (possibly depending on v). Substituting Equations (12.20) and (12.21) into Equation (12.19), and comparing the coefficients of x^2 , xt , and t^2 , we obtain

$$B = D = \frac{1}{\pm(1 - v^2/c^2)^{1/2}}, \quad (12.22)$$

$$C = \frac{-v/c^2}{\pm(1 - v^2/c^2)^{1/2}}. \quad (12.23)$$

We must choose the positive sign in order to ensure that $x' \rightarrow x$ as $v/c \rightarrow 0$. Thus, collecting our results, the transformation between coordinates in S and S' is given by

$$x' = \frac{x - vt}{(1 - v^2/c^2)^{1/2}}, \quad (12.24)$$

$$y' = y, \quad (12.25)$$

$$z' = z, \quad (12.26)$$

$$t' = \frac{t - vx/c^2}{(1 - v^2/c^2)^{1/2}}. \quad (12.27)$$

This is the famous *Lorentz transformation*. It ensures that the velocity of light is invariant between different inertial frames, and also reduces to the more familiar Galilean transform in the limit $v \ll c$. We can solve Equations (12.24)–(12.27) for x , y , z , and t , to obtain the *inverse Lorentz transformation*:

$$x = \frac{x' + vt'}{(1 - v^2/c^2)^{1/2}}, \quad (12.28)$$

$$y = y', \quad (12.29)$$

$$z = z', \quad (12.30)$$

$$t = \frac{t' + vx'/c^2}{(1 - v^2/c^2)^{1/2}}. \quad (12.31)$$

Not surprisingly, the inverse transformation is equivalent to a Lorentz transformation in which the velocity of the moving frame is $-v$ along the x -axis, instead of $+v$.

12.4 Transformation of Velocities

Consider two frames, S and S' , in the standard configuration. Let \mathbf{u} be the velocity of a particle in S . What is the particle's velocity in S' ? The components of \mathbf{u} are

$$u_1 = \frac{dx}{dt}, \quad (12.32)$$

$$u_2 = \frac{dy}{dt}, \quad (12.33)$$

$$u_3 = \frac{dz}{dt}. \quad (12.34)$$

Similarly, the components of \mathbf{u}' are

$$u'_1 = \frac{dx'}{dt'}, \quad (12.35)$$

$$u'_2 = \frac{dy'}{dt'}, \quad (12.36)$$

$$u'_3 = \frac{dz'}{dt'}. \quad (12.37)$$

Now, we can write Equations (12.24)–(12.27) in the form $dx' = \gamma(dx - v dt)$, $dy' = dy$, $dz' = dz$, and $dt' = \gamma(dt - v dx/c^2)$, where

$$\gamma = (1 - v^2/c^2)^{-1/2} \quad (12.38)$$

is the well-known *Lorentz factor*. If we substitute these differentials into Equations (12.32)–(12.34), and make use of Equations (12.35)–(12.37), we obtain the transformation rules

$$u'_1 = \frac{u_1 - v}{1 - u_1 v/c^2}, \quad (12.39)$$

$$u'_2 = \frac{u_2}{\gamma(1 - u_1 v/c^2)}, \quad (12.40)$$

$$u'_3 = \frac{u_3}{\gamma(1 - u_1 v/c^2)}. \quad (12.41)$$

As in the transformation of coordinates, we can obtain the inverse transform by interchanging primed and unprimed symbols, and replacing $+v$ with $-v$. Thus,

$$u_1 = \frac{u'_1 + v}{1 + u'_1 v/c^2}, \quad (12.42)$$

$$u_2 = \frac{u'_2}{\gamma(1 + u'_1 v/c^2)}, \quad (12.43)$$

$$u_3 = \frac{u'_3}{\gamma(1 + u'_1 v/c^2)}. \quad (12.44)$$

Equations (12.42)–(12.44) can be regarded as giving the resultant, $\mathbf{u} = (u_1, u_2, u_3)$, of two velocities, $\mathbf{v} = (v, 0, 0)$ and $\mathbf{u}' = (u'_1, u'_2, u'_3)$, and are therefore usually referred to as the relativistic *velocity addition formulae*. The following relation between the magnitudes $u = (u_1^2 + u_2^2 + u_3^2)^{1/2}$ and $u' = (u'^2_1 + u'^2_2 + u'^2_3)^{1/2}$ of the velocities is easily demonstrated:

$$c^2 - u^2 = \frac{c^2 (c^2 - u'^2) (c^2 - v^2)}{(c^2 + u'_1 v)^2}. \quad (12.45)$$

If $u' < c$ and $v < c$ then the right-hand side is positive, implying that $u < c$. In other words, the resultant of two subluminal velocities is another subluminal velocity. It is evident that a particle can never attain the velocity of light relative to a given inertial frame, no matter how many subluminal velocity increments it is given. It follows that no inertial frame can ever appear to propagate with a superluminal velocity with respect to any other inertial frame (because we can track a given inertial frame using a particle which remains at rest at the origin of that frame).

According to Equation (12.45), if $u' = c$ then $u = c$, no matter what value v takes: that is, the speed of light is invariant between different inertial frames. Note that the Lorentz transform only allows one such invariant speed [i.e., the speed c that appears in Equations (12.24)–(12.27)]. Einstein's relativity principle tells us that any disturbance that propagates through a vacuum must appear to propagate at the same speed in all inertial frames. It is now evident that all such disturbances must propagate at the speed c . It follows immediately that all electromagnetic waves must propagate through the vacuum with this speed, irrespective of their wavelength. In other words, it is impossible for there to be any dispersion of electromagnetic waves propagating through a vacuum. Furthermore, gravity waves must also propagate at the speed c .

The Lorentz transformation implies that the propagation speeds of all physical effects are limited by c in deterministic physics. Consider a general process by which an event P causes an event Q at a velocity $U > c$ in some frame S . In other words, information about the event P appears to propagate to the event Q with a superluminal velocity. Let us choose coordinates such that these two events occur on the x -axis with (finite) time and distance separations $\Delta t > 0$ and $\Delta x > 0$, respectively. The time separation in some other inertial frame S' is given by [see Equation (12.27)]

$$\Delta t' = \gamma(\Delta t - v \Delta x/c^2) = \gamma \Delta t (1 - vU/c^2). \quad (12.46)$$

Thus, for sufficiently large $v < c$ we obtain $\Delta t' < 0$: that is, there exist inertial frames in which cause and effect appear to be reversed. Of course, this is impossible in deterministic physics. It follows, therefore, that information can never appear to propagate with a superluminal velocity in any inertial frame, otherwise causality would be violated.

12.5 Tensors

It is now convenient to briefly review the mathematics of *tensors*. Tensors are of primary importance in connection with coordinate transforms. They serve to isolate intrinsic geometric and physical properties from those that merely depend on coordinates.

A tensor of rank r in an n -dimensional space possesses n^r components which are, in general, functions of position in that space. A tensor of rank zero has one component, A , and is called a

scalar. A tensor of rank one has n components, (A_1, A_2, \dots, A_n) , and is called a *vector*. A tensor of rank two has n^2 components, which can be exhibited in matrix format. Unfortunately, there is no convenient way of exhibiting a higher rank tensor. Consequently, tensors are usually represented by a typical component: for instance, the tensor A_{ijk} (rank 3), or the tensor A_{ijkl} (rank 4), et cetera. The suffixes i, j, k, \dots are always understood to range from 1 to n .

For reasons that will become apparent later on, we shall represent tensor components using both superscripts and subscripts. Thus, a typical tensor might look like A^{ij} (rank 2), or B_j^i (rank 2), et cetera. It is convenient to adopt the *Einstein summation convention*. Namely, if any suffix appears twice in a given term, once as a subscript and once as a superscript, a summation over that suffix (from 1 to n) is implied.

To distinguish between various different coordinate systems, we shall use primed and multiply primed suffixes. A first system of coordinates (x^1, x^2, \dots, x^n) can then be denoted by x^i , a second system $(x^{1'}, x^{2'}, \dots, x^{n'})$ by $x^{i'}$, et cetera. Similarly, the general components of a tensor in various coordinate systems are distinguished by their suffixes. Thus, the components of some third rank tensor are denoted A_{ijk} in the x^i system, by $A_{i'j'k'}$ in the $x^{i'}$ system, et cetera.

When making a coordinate transformation from one set of coordinates, x^i , to another, $x^{i'}$, it is assumed that the transformation is non-singular. In other words, the equations that express the $x^{i'}$ in terms of the x^i can be inverted to express the x^i in terms of the $x^{i'}$. It is also assumed that the functions specifying a transformation are differentiable. It is convenient to write

$$\frac{\partial x^{i'}}{\partial x^i} = p_i^{i'}, \quad (12.47)$$

$$\frac{\partial x^i}{\partial x^{i'}} = p_{i'}^i. \quad (12.48)$$

Note that

$$p_{i'}^i p_j^{i'} = \delta_j^i, \quad (12.49)$$

by the chain rule, where δ_j^i (the *Kronecker delta*) equals 1 or 0 when $i = j$ or $i \neq j$, respectively.

The formal definition of a tensor is as follows:

1. An entity having components $A_{ij\dots k}$ in the x^i system and $A_{i'j'\dots k'}$ in the $x^{i'}$ system is said to behave as a *covariant tensor* under the transformation $x^i \rightarrow x^{i'}$ if

$$A_{i'j'\dots k'} = A_{ij\dots k} p_{i'}^i p_{j'}^j \dots p_{k'}^k. \quad (12.50)$$

2. Similarly, $A^{ij\dots k}$ is said to behave as a *contravariant tensor* under $x^i \rightarrow x^{i'}$ if

$$A^{i'j'\dots k'} = A^{ij\dots k} p_i^{i'} p_j^{j'} \dots p_k^{k'}. \quad (12.51)$$

3. Finally, $A_{k\dots l}^{i\dots j}$ is said to behave as a *mixed tensor* (contravariant in $i\dots j$ and covariant in $k\dots l$) under $x^i \rightarrow x^{i'}$ if

$$A_{k'\dots l'}^{i'\dots j'} = A_{k\dots l}^{i\dots j} p_i^{i'} \dots p_j^{j'} p_{k'}^k \dots p_{l'}^l. \quad (12.52)$$

When an entity is described as a tensor it is generally understood that it behaves as a tensor under all non-singular differentiable transformations of the relevant coordinates. An entity that only behaves as a tensor under a certain subgroup of non-singular differentiable coordinate transformations is called a *qualified tensor*, because its name is conventionally qualified by an adjective recalling the subgroup in question. For instance, an entity that only exhibits tensor behavior under Lorentz transformations is called a *Lorentz tensor*, or, more commonly, a *4-tensor*.

When applied to a tensor of rank zero (a scalar), the previous definitions imply that $A' = A$. Thus, a scalar is a function of position only, and is independent of the coordinate system. A scalar is often termed an invariant.

The main theorem of tensor calculus is as follows:

If two tensors of the same type are equal in one coordinate system then they are equal in all coordinate systems.

The simplest example of a contravariant vector (tensor of rank one) is provided by the differentials of the coordinates, dx^i , because

$$dx^{i'} = \frac{\partial x^{i'}}{\partial x^i} dx^i = dx^i p_i^{i'}. \quad (12.53)$$

The coordinates themselves do not behave as tensors under all coordinate transformations. However, because they transform like their differentials under linear homogeneous coordinate transformations, they do behave as tensors under such transformations.

The simplest example of a covariant vector is provided by the gradient of a function of position $\phi = \phi(x^1, \dots, x^n)$, because if we write

$$\phi_i = \frac{\partial \phi}{\partial x^i}, \quad (12.54)$$

then we have

$$\phi_{i'} = \frac{\partial \phi}{\partial x^{i'}} = \frac{\partial \phi}{\partial x^i} \frac{\partial x^i}{\partial x^{i'}} = \phi_i p_{i'}^i. \quad (12.55)$$

An important example of a mixed second-rank tensor is provided by the Kronecker delta introduced previously, because

$$\delta_j^i p_i^{i'} p_{j'}^j = p_j^{i'} p_{j'}^j = \delta_{j'}^{i'}. \quad (12.56)$$

Tensors of the same type can be added or subtracted to form new tensors. Thus, if A_{ij} and B_{ij} are tensors, then $C_{ij} = A_{ij} \pm B_{ij}$ is a tensor of the same type. Note that the sum of tensors at different points in space is not a tensor if the p 's are position dependent. However, under linear coordinate transformations the p 's are constant, so the sum of tensors at different points behaves as a tensor under this particular type of coordinate transformation.

If A^{ij} and B_{ijk} are tensors, then $C_{klm}^{ij} = A^{ij} B_{klm}$ is a tensor of the type indicated by the suffixes. The process illustrated by this example is called *outer multiplication* of tensors.

Tensors can also be combined by *inner multiplication*, which implies at least one dummy suffix link. Thus, $C_{kl}^j = A^{ij} B_{ikl}$ and $C_k = A^{ij} B_{ijk}$ are tensors of the type indicated by the suffixes.

Finally, tensors can be formed by *contraction* from tensors of higher rank. Thus, if A_{klm}^{ij} is a tensor then $C_{kl}^j = A_{ikl}^{ij}$ and $C_k = A_{kij}^{ij}$ are tensors of the type indicated by the suffixes. The most

important type of contraction occurs when no free suffixes remain: the result is a scalar. Thus, A_i^i is a scalar provided that A_i^j is a tensor.

Although we cannot usefully divide tensors, one by another, an entity like C^{ij} in the equation $A^j = C^{ij} B_i$, where A^i and B_i are tensors, can be formally regarded as the quotient of A^i and B_i . This gives the name to a particularly useful rule for recognizing tensors, the *quotient rule*. This rule states that if a set of components, when combined by a given type of multiplication with all tensors of a given type yields a tensor, then the set is itself a tensor. In other words, if the product $A^i = C^{ij} B_j$ transforms like a tensor for all tensors B_j then it follows that C^{ij} is a tensor.

Let

$$\frac{\partial A_{k\dots l}^{i\dots j}}{\partial x^m} = A_{k\dots l, m}^{i\dots j} \quad (12.57)$$

Then if $A_{k\dots l}^{i\dots j}$ is a tensor, differentiation of the general tensor transformation (12.52) yields

$$A_{k'\dots l', m'}^{i'\dots j'} = A_{k\dots l, m}^{i\dots j} p_i^{i'} \cdots p_j^{j'} p_{k'}^k \cdots p_{l'}^l p_{m'}^m + P_1 + P_2 + \cdots, \quad (12.58)$$

where P_1, P_2, \dots , et cetera, are terms involving derivatives of the p 's. Clearly, $A_{k\dots l}^{i\dots j}$ is not a tensor under a general coordinate transformation. However, under a linear coordinate transformation (p 's constant) $A_{k'\dots l', m'}^{i'\dots j'}$ behaves as a tensor of the type indicated by the suffixes, because the P_1, P_2, \dots , et cetera, all vanish. Similarly, all higher partial derivatives,

$$A_{k\dots l, mn}^{i\dots j} = \frac{\partial A_{k\dots l}^{i\dots j}}{\partial x^m \partial x^n} \quad (12.59)$$

et cetera, also behave as tensors under linear transformations. Each partial differentiation has the effect of adding a new covariant suffix.

So far, the space to which the coordinates x^i refer has been without structure. We can impose a structure on it by defining the distance between all pairs of neighboring points by means of a *metric*,

$$ds^2 = g_{ij} dx^i dx^j, \quad (12.60)$$

where the g_{ij} are functions of position. We can assume that $g_{ij} = g_{ji}$ without loss of generality. The previous metric is analogous to, but more general than, the metric of Euclidian n -space, $ds^2 = (dx^1)^2 + (dx^2)^2 + \cdots + (dx^n)^2$. A space whose structure is determined by a metric of the type (12.60) is called *Riemannian*. Because ds^2 is invariant, it follows from a simple extension of the quotient rule that g_{ij} must be a tensor. It is called the *metric tensor*.

The elements of the inverse of the matrix g_{ij} are denoted by g^{ij} . These elements are uniquely defined by the equations

$$g^{ij} g_{jk} = \delta_k^i. \quad (12.61)$$

It is easily seen that the g^{ij} constitute the elements of a contravariant tensor. This tensor is said to be conjugate to g_{ij} . The conjugate metric tensor is symmetric (i.e., $g^{ij} = g^{ji}$) just like the metric tensor itself.

The tensors g_{ij} and g^{ij} allow us to introduce the important operations of raising and lowering suffixes. These operations consist of forming inner products of a given tensor with g_{ij} or g^{ij} . For

example, given a contravariant vector A^i , we define its covariant components A_i by the equation

$$A_i = g_{ij} A^j. \quad (12.62)$$

Conversely, given a covariant vector B_i , we can define its contravariant components B^i by the equation

$$B^i = g^{ij} B_j. \quad (12.63)$$

More generally, we can raise or lower any or all of the free suffixes of any given tensor. Thus, if A_{ij} is a tensor we define $A^i{}_j$ by the equation

$$A^i{}_j = g^{ip} A_{pj}. \quad (12.64)$$

Note that once the operations of raising and lowering suffixes has been defined, the order of raised suffixes relative to lowered suffixes becomes significant.

By analogy with Euclidian space, we define the *squared magnitude* $(A)^2$ of a vector A^i with respect to the metric $g_{ij} dx^i dx^j$ by the equation

$$(A)^2 = g_{ij} A^i A^j = A_i A^i. \quad (12.65)$$

A vector A^i termed a *null vector* if $(A)^2 = 0$. Two vectors A^i and B^i are said to be *orthogonal* if their inner product vanishes: that is, if

$$g_{ij} A^i B^j = A_i B^i = A^i B_i = 0. \quad (12.66)$$

Finally, let us consider differentiation with respect to an invariant distance, s . The vector dx^i/ds is a contravariant tensor, because

$$\frac{dx^{i'}}{ds} = \frac{\partial x^{i'}}{\partial x^i} \frac{dx^i}{ds} = \frac{dx^i}{ds} p_i^{i'}. \quad (12.67)$$

The derivative $d(A^{i\dots j}_{k\dots l})/ds$ of some tensor with respect to s is not, in general, a tensor, because

$$\frac{d(A^{i\dots j}_{k\dots l})}{ds} = A^{i\dots j}_{k\dots l, m} \frac{dx^m}{ds}, \quad (12.68)$$

and, as we have seen, the first factor on the right-hand side is not generally a tensor. However, under linear transformations it behaves as a tensor, so under linear transformations the derivative of a tensor with respect to an invariant distance behaves as a tensor of the same type.

12.6 Physical Significance of Tensors

In this chapter, we shall only concern ourselves with coordinate transformations that transform an inertial frame into another inertial frame. This limits us to four classes of transformations: displacements of the coordinate axes, rotations of the coordinate axes, parity reversals (i.e., $x, y, z \rightarrow -x, -y, -z$), and Lorentz transformations.

One of the central tenets of physics is that experiments should be reproducible. In other words, if somebody performs a physical experiment today, and obtains a certain result, then somebody else performing the same experiment next week ought to obtain the same result, within the experimental errors. Presumably, in performing these hypothetical experiments, both experimentalists find it necessary to set up a coordinate frame. Usually, these two frames do not coincide. After all, the experiments are, in general, performed in different places and at different times. Also, the two experimentalists are likely to orientate their coordinate axes differently. Nevertheless, we still expect both experiments to yield the same result. What exactly do we mean by this statement? We do not mean that both experimentalists will obtain the same numbers when they measure something. For instance, the numbers used to denote the position of a point (i.e., the coordinates of the point) are, in general, different in different coordinate frames. What we do expect is that any physically significant interrelation between physical quantities (i.e., position, velocity, etc.) which appears to hold in the coordinate system of the first experimentalist will also appear to hold in the coordinate system of the second experimentalist. We usually refer to such interrelationships as *laws of physics*. So, what we are really saying is that the laws of physics do not depend on our choice of coordinate system. In particular, if a law of physics is true in one coordinate system then it is automatically true in every other coordinate system, subject to the proviso that both coordinate systems are inertial.

Recall that tensors are geometric objects that possess the property that if a certain interrelationship holds between various tensors in one particular coordinate system then the same interrelationship holds in any other coordinate system that is related to the first system by a certain class of transformations. It follows that the laws of physics are expressible as interrelationships between tensors. In special relativity, the laws of physics are only required to exhibit tensor behavior under transformations between different inertial frames: that is, under translations, rotations, and Lorentz transformations. Parity inversion is a special type of transformation, and will be dealt with separately later on. In general relativity, the laws of physics are required to exhibit tensor behavior under all non-singular coordinate transformations.

12.7 Space-Time

In special relativity, we are only allowed to use inertial frames to assign coordinates to events. There are many different types of inertial frames. However, it is convenient to adhere to those with *standard coordinates*. That is, spatial coordinates that are right-handed rectilinear Cartesians based on a standard unit of length, and time-scales based on a standard unit of time. We shall continue to assume that we are employing standard coordinates. However, from now on, we shall make no assumptions about the relative configuration of the two sets of spatial axes, and the origins of time, when dealing with two inertial frames. Thus, the most general transformation between two inertial frames consists of a Lorentz transformation in the standard configuration plus a translation (this includes a translation in time) and a rotation of the coordinate axes. The resulting transformation is called a *general Lorentz transformation*, as opposed to a Lorentz transformation in the standard configuration, which will henceforth be termed a *standard Lorentz transformation*.

In Section 12.3, we proved quite generally that corresponding differentials in two inertial

frames S and S' satisfy the relation

$$dx^2 + dy^2 + dz^2 - c^2 dt^2 = dx'^2 + dy'^2 + dz'^2 - c^2 dt'^2. \quad (12.69)$$

Thus, we expect this relation to remain invariant under a general Lorentz transformation. Because such a transformation is linear, it follows that

$$\begin{aligned} (x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2 - c^2 (t_2 - t_1)^2 = \\ (x'_2 - x'_1)^2 + (y'_2 - y'_1)^2 + (z'_2 - z'_1)^2 - c^2 (t'_2 - t'_1)^2, \end{aligned} \quad (12.70)$$

where (x_1, y_1, z_1, t_1) and (x_2, y_2, z_2, t_2) are the coordinates of any two events in S , and the primed symbols denote the corresponding coordinates in S' . It is convenient to write

$$-dx^2 - dy^2 - dz^2 + c^2 dt^2 = ds^2, \quad (12.71)$$

and

$$-(x_2 - x_1)^2 - (y_2 - y_1)^2 - (z_2 - z_1)^2 + c^2 (t_2 - t_1)^2 = s^2. \quad (12.72)$$

The differential ds , or the finite length s , defined by these equations is called the *interval* between the corresponding events. Equations (12.71) and (12.72) express the fact that the interval between two events is invariant, in the sense that it has the same value in all inertial frames. In other words, the interval between two events is invariant under a general Lorentz transformation.

Let us consider entities defined in terms of four variables,

$$x^1 = x, \quad x^2 = y, \quad x^3 = z, \quad x^4 = ct, \quad (12.73)$$

and which transform as tensors under a general Lorentz transformation. From now on, such entities will be referred to as *4-tensors*.

Tensor analysis cannot proceed very far without the introduction of a non-singular tensor g_{ij} , the so-called *fundamental tensor*, which is used to define the operations of raising and lowering suffixes. The fundamental tensor is usually introduced using a metric $ds^2 = g_{ij} dx^i dx^j$, where ds^2 is a differential invariant. We have already come across such an invariant, namely

$$\begin{aligned} ds^2 &= -dx^2 - dy^2 - dz^2 + c^2 dt^2 \\ &= -(dx^1)^2 - (dx^2)^2 - (dx^3)^2 + (dx^4)^2 \\ &= g_{\mu\nu} dx^\mu dx^\nu, \end{aligned} \quad (12.74)$$

where μ, ν run from 1 to 4. Note that the use of Greek suffixes is conventional in 4-tensor theory. Roman suffixes are reserved for tensors in three-dimensional Euclidian space—so-called *3-tensors*. The 4-tensor $g_{\mu\nu}$ has the components $g_{11} = g_{22} = g_{33} = -1$, $g_{44} = 1$, and $g_{\mu\nu} = 0$ when $\mu \neq \nu$, in all permissible coordinate frames. From now on, $g_{\mu\nu}$, as just defined, is adopted as the fundamental tensor for 4-tensors. $g_{\mu\nu}$ can be thought of as the metric tensor of the space whose points are the events (x^1, x^2, x^3, x^4) . This space is usually referred to as *space-time*, for obvious reasons. Note that space-time cannot be regarded as a straightforward generalization of Euclidian 3-space to four

dimensions, with time as the fourth dimension. The distribution of signs in the metric ensures that the time coordinate x^4 is not on the same footing as the three space coordinates. Thus, space-time has a non-isotropic nature which is quite unlike Euclidian space, with its positive definite metric. According to the relativity principle, all physical laws are expressible as interrelationships between 4-tensors in space-time.

A tensor of rank one is called a *4-vector*. We shall also have occasion to use ordinary vectors in three-dimensional Euclidian space. Such vectors are called *3-vectors*, and are conventionally represented by boldface symbols. We shall use the Latin suffixes i, j, k , et cetera, to denote the components of a 3-vector: these suffixes are understood to range from 1 to 3. Thus, $\mathbf{u} = u^i = dx^i/dt$ denotes a velocity vector. For 3-vectors, we shall use the notation $u^i = u_i$ interchangeably: that is, the level of the suffix has no physical significance.

When tensor transformations from one frame to another actually have to be computed, we shall usually find it possible to choose coordinates in the standard configuration, so that the standard Lorentz transform applies. Under such a transformation, any contravariant 4-vector, T^μ , transforms according to the same scheme as the difference in coordinates $x_2^\mu - x_1^\mu$ between two points in space-time. It follows that

$$T^{1'} = \gamma(T^1 - \beta T^4), \quad (12.75)$$

$$T^{2'} = T^2, \quad (12.76)$$

$$T^{3'} = T^3, \quad (12.77)$$

$$T^{4'} = \gamma(T^4 - \beta T^1), \quad (12.78)$$

where $\beta = v/c$. Higher rank 4-tensors transform according to the rules (12.50)–(12.52). The transformation coefficients take the form

$$p_{\mu'}^{\mu} = \begin{pmatrix} +\gamma & 0 & 0 & -\gamma\beta \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\gamma\beta & 0 & 0 & +\gamma \end{pmatrix}, \quad (12.79)$$

$$p_{\mu'}^{\mu} = \begin{pmatrix} +\gamma & 0 & 0 & +\gamma\beta \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ +\gamma\beta & 0 & 0 & +\gamma \end{pmatrix}. \quad (12.80)$$

Often the first three components of a 4-vector coincide with the components of a 3-vector. For example, the x^1, x^2, x^3 in $R^\mu = (x^1, x^2, x^3, x^4)$ are the components of \mathbf{r} , the position 3-vector of the point at which the event occurs. In such cases, we adopt the notation exemplified by $R^\mu = (\mathbf{r}, ct)$. The covariant form of such a vector is simply $R_\mu = (-\mathbf{r}, ct)$. The squared magnitude of the vector is $(R)^2 = R_\mu R^\mu = -r^2 + c^2 t^2$. The inner product $g_{\mu\nu} R^\mu Q^\nu = R_\mu Q^\mu$ of R^μ with a similar vector $Q^\mu = (\mathbf{q}, k)$ is given by $R_\mu Q^\mu = -\mathbf{r} \cdot \mathbf{q} + ct k$. The vectors R^μ and Q^μ are said to be orthogonal if $R_\mu Q^\mu = 0$.

Because a general Lorentz transformation is a linear transformation, the partial derivative of a 4-tensor is also a 4-tensor:

$$\frac{\partial A^{\nu\sigma}}{\partial x^\mu} = A^{\nu\sigma}{}_{,\mu}. \quad (12.81)$$

Clearly, a general 4-tensor acquires an extra covariant index after partial differentiation with respect to the contravariant coordinate x^μ . It is helpful to define a covariant derivative operator

$$\partial_\mu \equiv \frac{\partial}{\partial x^\mu} = \left(\nabla, \frac{1}{c} \frac{\partial}{\partial t} \right), \quad (12.82)$$

where

$$\partial_\mu A^{\nu\sigma} \equiv A^{\nu\sigma}{}_{,\mu}. \quad (12.83)$$

There is a corresponding contravariant derivative operator

$$\partial^\mu \equiv \frac{\partial}{\partial x_\mu} = \left(-\nabla, \frac{1}{c} \frac{\partial}{\partial t} \right), \quad (12.84)$$

where

$$\partial^\mu A^{\nu\sigma} \equiv g^{\mu\tau} A^{\nu\sigma}{}_{,\tau}. \quad (12.85)$$

The 4-divergence of a 4-vector, $A^\mu = (\mathbf{A}, A^0)$, is the invariant

$$\partial^\mu A_\mu = \partial_\mu A^\mu = \nabla \cdot \mathbf{A} + \frac{1}{c} \frac{\partial A^0}{\partial t}. \quad (12.86)$$

The four-dimensional Laplacian operator, or *d'Alembertian*, is equivalent to the invariant contraction

$$\square \equiv \partial_\mu \partial^\mu = -\nabla^2 + \frac{1}{c^2} \frac{\partial^2}{\partial t^2}. \quad (12.87)$$

Recall that we still need to prove (from Section 12.3) that the invariance of the differential metric,

$$ds^2 = dx'^2 + dy'^2 + dz'^2 - c^2 dt'^2 = dx^2 + dy^2 + dz^2 - c^2 dt^2, \quad (12.88)$$

between two general inertial frames implies that the coordinate transformation between such frames is necessarily linear. To put it another way, we need to demonstrate that a transformation that transforms a metric $g_{\mu\nu} dx^\mu dx^\nu$ with constant coefficients into a metric $g_{\mu'\nu'} dx^{\mu'} dx^{\nu'}$ with constant coefficients must be linear. Now,

$$g_{\mu\nu} = g_{\mu'\nu'} P_\mu^{\mu'} P_\nu^{\nu'}. \quad (12.89)$$

Differentiating with respect to x^σ , we get

$$g_{\mu'\nu'} P_{\mu\sigma}^{\mu'} P_\nu^{\nu'} + g_{\mu'\nu'} P_\mu^{\mu'} P_{\nu\sigma}^{\nu'} = 0, \quad (12.90)$$

where

$$P_{\mu\sigma}^{\mu'} = \frac{\partial p_\mu^{\mu'}}{\partial x^\sigma} = \frac{\partial^2 x^{\mu'}}{\partial x^\mu \partial x^\sigma} = p_{\sigma\mu}^{\mu'}, \quad (12.91)$$

et cetera. Interchanging the indices μ and σ yields

$$g_{\mu'\nu'} p_{\mu\sigma}^{\mu'} p_{\nu}^{\nu'} + g_{\mu'\nu'} p_{\sigma\mu}^{\mu'} p_{\nu\mu}^{\nu'} = 0. \quad (12.92)$$

Interchanging the indices ν and σ gives

$$g_{\mu'\nu'} p_{\sigma\mu}^{\mu'} p_{\nu\mu}^{\nu'} + g_{\mu'\nu'} p_{\mu\sigma}^{\mu'} p_{\nu\sigma}^{\nu'} = 0, \quad (12.93)$$

where the indices μ' and ν' have been interchanged in the first term. It follows from Equations (12.90), (12.92), and (12.93) that

$$g_{\mu'\nu'} p_{\mu\sigma}^{\mu'} p_{\nu}^{\nu'} = 0. \quad (12.94)$$

Multiplication by $p_{\sigma'}^{\nu'}$ yields

$$g_{\mu'\nu'} p_{\mu\sigma}^{\mu'} p_{\nu}^{\nu'} p_{\sigma'}^{\nu'} = g_{\mu'\sigma'} p_{\mu\sigma}^{\mu'} = 0. \quad (12.95)$$

Finally, multiplication by $g^{\nu'\sigma'}$ gives

$$g_{\mu'\sigma'} g^{\nu'\sigma'} p_{\mu\sigma}^{\mu'} = p_{\mu\sigma}^{\nu'} = 0. \quad (12.96)$$

This proves that the coefficients $p_{\mu}^{\nu'}$ are constants, and, hence, that the transformation is linear.

12.8 Proper Time

It is often helpful to write the invariant differential interval ds^2 in the form

$$ds^2 = c^2 d\tau^2. \quad (12.97)$$

The quantity $d\tau$ is called the *proper time*. It follows that

$$d\tau^2 = -\frac{dx^2 + dy^2 + dz^2}{c^2} + dt^2. \quad (12.98)$$

Consider a series of events on the world-line of some material particle. If the particle has speed u then

$$d\tau^2 = dt^2 \left(-\frac{dx^2 + dy^2 + dz^2}{c^2 dt^2} + 1 \right) = dt^2 \left(1 - \frac{u^2}{c^2} \right), \quad (12.99)$$

implying that

$$\frac{dt}{d\tau} = \gamma(u). \quad (12.100)$$

It is clear that $dt = d\tau$ in the particle's rest frame. Thus, $d\tau$ corresponds to the time difference between two neighboring events on the particle's world-line, as measured by a clock attached to the particle (hence, the name "proper time"). According to Equation (12.100), the particle's clock appears to run slow, by a factor $\gamma(u)$, in an inertial frame in which the particle is moving with velocity u . This is the celebrated *time dilation* effect.

Let us consider how a small 4-dimensional volume element in space-time transforms under a general Lorentz transformation. We have

$$d^4x' = \mathcal{J} d^4x, \quad (12.101)$$

where

$$\mathcal{J} = \frac{\partial(x^1, x^2, x^3, x^4)}{\partial(x^{1'}, x^{2'}, x^{3'}, x^{4'})} \quad (12.102)$$

is the Jacobian of the transformation: that is, the determinant of the transformation matrix $p_{\mu}^{\mu'}$. A general Lorentz transformation is made up of a standard Lorentz transformation plus a displacement and a rotation. Thus, the transformation matrix is the product of that for a standard Lorentz transformation, a translation, and a rotation. It follows that the Jacobian of a general Lorentz transformation is the product of that for a standard Lorentz transformation, a translation, and a rotation. It is well known that the Jacobians of the latter two transformations are unity, because they are both volume preserving transformations that do not affect time. Likewise, it is easily seen [e.g., by taking the determinant of the transformation matrix (12.79)] that the Jacobian of a standard Lorentz transformation is also unity. It follows that

$$d^4x' = d^4x \quad (12.103)$$

for a general Lorentz transformation. In other words, a general Lorentz transformation preserves the volume of space-time. Because time is dilated by a factor γ in a moving frame, the volume of space-time can only be preserved if the volume of ordinary 3-space is reduced by the same factor. As is well-known, this is achieved by *length contraction* along the direction of motion by a factor γ .

12.9 4-Velocity and 4-Acceleration

We have seen that the quantity dx^{μ}/ds transforms as a 4-vector under a general Lorentz transformation [see Equation (12.67)]. Because $ds \propto d\tau$ it follows that

$$U^{\mu} = \frac{dx^{\mu}}{d\tau} \quad (12.104)$$

also transforms as a 4-vector. This quantity is known as the *4-velocity*. Likewise, the quantity

$$A^{\mu} = \frac{d^2x^{\mu}}{d\tau^2} = \frac{dU^{\mu}}{d\tau} \quad (12.105)$$

is a 4-vector, and is called the *4-acceleration*.

For events along the world-line of a particle traveling with 3-velocity \mathbf{u} , we have

$$U^{\mu} = \frac{dx^{\mu}}{d\tau} = \frac{dx^{\mu}}{dt} \frac{dt}{d\tau} = \gamma(u) (\mathbf{u}, c), \quad (12.106)$$

where use has been made of Equation (12.100). This gives the relationship between a particle's 3-velocity and its 4-velocity. The relationship between the 3-acceleration and the 4-acceleration is less straightforward. We have

$$A^\mu = \frac{dU^\mu}{d\tau} = \gamma \frac{dU^\mu}{dt} = \gamma \frac{d}{dt}(\gamma \mathbf{u}, \gamma c) = \gamma \left(\frac{d\gamma}{dt} \mathbf{u} + \gamma \mathbf{a}, c \frac{d\gamma}{dt} \right), \quad (12.107)$$

where $\mathbf{a} = d\mathbf{u}/dt$ is the 3-acceleration. In the rest frame of the particle, $U^\mu = (\mathbf{0}, c)$ and $A^\mu = (\mathbf{a}, 0)$. It follows that

$$U_\mu A^\mu = 0 \quad (12.108)$$

(note that $U_\mu A^\mu$ is an invariant quantity). In other words, the 4-acceleration of a particle is always orthogonal to its 4-velocity.

12.10 Current Density 4-Vector

Let us now consider the laws of electromagnetism. We wish to demonstrate that these laws are compatible with the relativity principle. In order to achieve this, it is necessary for us to make an assumption about the transformation properties of electric charge. The assumption that we shall make, which is well substantiated experimentally, is that charge, unlike mass, is invariant. That is, the charge carried by a given particle has the same measure in all inertial frames. In particular, the charge carried by a particle does not vary with the particle's velocity.

Let us suppose, following Lorentz, that all charge is made up of elementary particles, each carrying the invariant amount e . Suppose that n is the number density of such charges at some given point and time, moving with velocity \mathbf{u} , as observed in a frame S . Let n_0 be the number density of charges in the frame S_0 in which the charges are momentarily at rest. As is well known, a volume of measure V in S has measure $\gamma(u)V$ in S_0 (because of length contraction). Because observers in both frames must agree on how many particles are contained in the volume, and, hence, on how much charge it contains, it follows that $n = \gamma(u)n_0$. If $\rho = en$ and $\rho_0 = en_0$ are the charge densities in S and S_0 , respectively, then

$$\rho = \gamma(u)\rho_0. \quad (12.109)$$

The quantity ρ_0 is called the *proper density*, and is obviously Lorentz invariant.

Suppose that x^μ are the coordinates of the moving charge in S . The *current density 4-vector* is constructed as follows:

$$J^\mu = \rho_0 \frac{dx^\mu}{d\tau} = \rho_0 U^\mu. \quad (12.110)$$

Thus,

$$J^\mu = \rho_0 \gamma(u) (\mathbf{u}, c) = (\mathbf{j}, \rho c), \quad (12.111)$$

where $\mathbf{j} = \rho \mathbf{u}$ is the current density 3-vector. Clearly, charge density and current density transform as the time-like and space-like components of the same 4-vector.

Consider the invariant 4-divergence of J^μ :

$$\partial_\mu J^\mu = \nabla \cdot \mathbf{j} + \frac{\partial \rho}{\partial t}. \quad (12.112)$$

We know that one of the caveats of Maxwell's equations is the charge conservation law

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{j} = 0. \quad (12.113)$$

It is clear that this expression can be rewritten in the manifestly Lorentz invariant form

$$\partial_\mu J^\mu = 0. \quad (12.114)$$

This equation tells us that there are no net sources or sinks of electric charge in nature: that is, electric charge is neither created nor destroyed.

12.11 Potential 4-Vector

There are many ways of writing the laws of electromagnetism. However, the most obviously Lorentz invariant way is to write them in terms of the vector and scalar potentials. When written in this fashion, Maxwell's equations reduce to

$$\left(-\nabla^2 + \frac{1}{c^2} \frac{\partial^2}{\partial t^2}\right) \phi = \frac{\rho}{\epsilon_0}, \quad (12.115)$$

$$\left(-\nabla^2 + \frac{1}{c^2} \frac{\partial^2}{\partial t^2}\right) \mathbf{A} = \mu_0 \mathbf{j}, \quad (12.116)$$

where ϕ is the scalar potential, and \mathbf{A} the vector potential. Note that the differential operator appearing in these equations is the Lorentz invariant d'Alembertian, defined in Equation (12.87). Thus, the previous pair of equations can be rewritten in the form

$$\square \phi = \frac{\rho c}{c \epsilon_0}, \quad (12.117)$$

$$\square c \mathbf{A} = \frac{\mathbf{j}}{c \epsilon_0}. \quad (12.118)$$

Maxwell's equations can be written in Lorentz invariant form provided that the entity

$$\Phi^\mu = (c \mathbf{A}, \phi) \quad (12.119)$$

transforms as a contravariant 4-vector. This entity is known as the *potential 4-vector*. It follows from Equations (12.111), (12.115), and (12.116) that

$$\square \Phi^\mu = \frac{J^\mu}{c \epsilon_0}. \quad (12.120)$$

Thus, the field equations that govern classical electromagnetism can all be summed up in a single 4-vector equation.

12.12 Gauge Invariance

The electric and magnetic fields are obtained from the vector and scalar potentials according to the prescription

$$\mathbf{E} = -\nabla\phi - \frac{\partial\mathbf{A}}{\partial t}, \quad (12.121)$$

$$\mathbf{B} = \nabla \times \mathbf{A}. \quad (12.122)$$

These fields are important because they determine the electromagnetic forces exerted on charged particles. Note that the previous prescription does not uniquely determine the two potentials. It is possible to make the following transformation, known as a gauge transformation, that leaves the fields unaltered:

$$\phi \rightarrow \phi + \frac{\partial\psi}{\partial t}, \quad (12.123)$$

$$\mathbf{A} \rightarrow \mathbf{A} - \nabla\psi, \quad (12.124)$$

where $\psi(\mathbf{r}, t)$ is a general scalar field. It is necessary to adopt some form of convention, generally known as a gauge condition, to fully specify the two potentials. In fact, there is only one gauge condition that is consistent with Equations (12.114). This is the Lorenz gauge condition,

$$\frac{1}{c^2} \frac{\partial\phi}{\partial t} + \nabla \cdot \mathbf{A} = 0. \quad (12.125)$$

Note that this condition can be written in the Lorentz invariant form

$$\partial_\mu \Phi^\mu = 0. \quad (12.126)$$

This implies that if the Lorenz gauge holds in one particular inertial frame then it automatically holds in all other inertial frames. A general gauge transformation can be written

$$\Phi^\mu \rightarrow \Phi^\mu + c \partial^\mu \psi. \quad (12.127)$$

Note that, even after the Lorenz gauge has been adopted, the potentials are undetermined to a gauge transformation using a scalar field, ψ , that satisfies the sourceless wave equation

$$\square\psi = 0. \quad (12.128)$$

However, if we adopt sensible boundary conditions in both space and time then the only solution to the previous equation is $\psi = 0$.

12.13 Retarded Potentials

The solutions to Equations (12.117) and (12.118) take the form:

$$\phi(\mathbf{r}, t) = \frac{1}{4\pi\epsilon_0} \int \frac{[\rho(\mathbf{r}')] }{|\mathbf{r} - \mathbf{r}'|} dV', \quad (12.129)$$

$$\mathbf{A}(\mathbf{r}, t) = \frac{\mu_0}{4\pi} \int \frac{[\mathbf{j}(\mathbf{r}')] }{|\mathbf{r} - \mathbf{r}'|} dV'. \quad (12.130)$$

The previous equations can be combined to form the solution of the 4-vector wave equation (12.120),

$$\Phi^\mu = \frac{1}{4\pi \epsilon_0 c} \int \frac{[J^\mu]}{r} dV. \quad (12.131)$$

Here, the components of the 4-potential are evaluated at some event P in space-time, r is the distance of the volume element dV from P , and the square brackets indicate that the 4-current is to be evaluated at the retarded time: that is, at a time r/c before P .

But, does the right-hand side of Equation (12.131) really transform as a contravariant 4-vector? This is not a trivial question, because volume integrals in 3-space are not, in general, Lorentz invariant due to the length contraction effect. However, the integral in Equation (12.131) is not a straightforward volume integral, because the integrand is evaluated at the retarded time. In fact, the integral is best regarded as an integral over events in space-time. The events that enter the integral are those which intersect a spherical light wave launched from the event P and evolved backwards in time. In other words, the events occur before the event P , and have zero interval with respect to P . It is clear that observers in all inertial frames will, at least, agree on which events are to be included in the integral, because both the interval between events, and the absolute order in which events occur, are invariant under a general Lorentz transformation.

We shall now demonstrate that all observers obtain the same value of dV/r for each elementary contribution to the integral. Suppose that S and S' are two inertial frames in the standard configuration. Let unprimed and primed symbols denote corresponding quantities in S and S' , respectively. Let us assign coordinates $(0, 0, 0, 0)$ to P , and (x, y, z, ct) to the retarded event Q for which r and dV are evaluated. Using the standard Lorentz transformation, (12.24)–(12.27), the fact that the interval between events P and Q is zero, and the fact that both t and t' are negative, we obtain

$$r' = -ct' = -c\gamma \left(t - \frac{vx}{c^2} \right), \quad (12.132)$$

where v is the relative velocity between frames S' and S , γ is the Lorentz factor, and $r^2 = x^2 + y^2 + z^2$, et cetera. It follows that

$$r' = r\gamma \left(-\frac{ct}{r} + \frac{vx}{cr} \right) = r\gamma \left(1 + \frac{v}{c} \cos \theta \right), \quad (12.133)$$

where θ is the angle (in 3-space) subtended between the line PQ and the x -axis.

We now know the transformation for r . What about the transformation for dV ? We might be tempted to set $dV' = \gamma dV$, according to the usual length contraction rule. However, this is incorrect. The contraction by a factor γ only applies if the whole of the volume is measured at the same time, which is not the case in the present problem. Now, the dimensions of dV along the y - and z -axes are the same in both S and S' , according to Equations (12.24)–(12.27). For the x -dimension these equations give $dx' = \gamma(dx - v dt)$. The extremities of dx are measured at times differing by dt , where

$$dt = -\frac{dr}{c} = -\frac{dx}{c} \cos \theta. \quad (12.134)$$

Thus,

$$dx' = \gamma \left(1 + \frac{v}{c} \cos \theta \right) dx, \quad (12.135)$$

giving

$$dV' = \gamma \left(1 + \frac{v}{c} \cos \theta \right) dV. \quad (12.136)$$

It follows from Equations (12.133) and (12.136) that $dV'/r' = dV/r$. This result clearly remains valid even when S and S' are not in the standard configuration.

Thus, dV/r is an invariant and, therefore, $[J^\mu] dV/r$ is a contravariant 4-vector. For linear transformations, such as a general Lorentz transformation, the result of adding 4-tensors evaluated at different 4-points is itself a 4-tensor. It follows that the right-hand side of Equation (12.131) is indeed a contravariant 4-vector. Thus, this 4-vector equation can be properly regarded as the solution to the 4-vector wave equation (12.120).

12.14 Tensors and Pseudo-Tensors

The totally antisymmetric fourth rank tensor is defined

$$\epsilon^{\alpha\beta\gamma\delta} = \begin{cases} +1 & \text{for } \alpha, \beta, \gamma, \delta \text{ any even permutation of } 1, 2, 3, 4 \\ -1 & \text{for } \alpha, \beta, \gamma, \delta \text{ any odd permutation of } 1, 2, 3, 4 \\ 0 & \text{otherwise} \end{cases}. \quad (12.137)$$

The components of this tensor are invariant under a general Lorentz transformation, because

$$\epsilon^{\alpha\beta\gamma\delta} p_\alpha^{\alpha'} p_\beta^{\beta'} p_\gamma^{\gamma'} p_\delta^{\delta'} = \epsilon^{\alpha'\beta'\gamma'\delta'} |p_\mu^{\mu'}| = \epsilon^{\alpha'\beta'\gamma'\delta'}, \quad (12.138)$$

where $|p_\mu^{\mu'}|$ denotes the determinant of the transformation matrix, or the Jacobian of the transformation, which we have already established is unity for a general Lorentz transformation. We can also define a totally antisymmetric third rank tensor ϵ^{ijk} which stands in the same relation to 3-space as $\epsilon^{\alpha\beta\gamma\delta}$ does to space-time. It is easily demonstrated that the elements of ϵ^{ijk} are invariant under a general translation or rotation of the coordinate axes. The totally antisymmetric third rank tensor is used to define the cross product of two 3-vectors,

$$(\mathbf{a} \times \mathbf{b})^i = \epsilon^{ijk} a_j b_k, \quad (12.139)$$

and the curl of a 3-vector field,

$$(\nabla \times \mathbf{A})^i = \epsilon^{ijk} \frac{\partial A_k}{\partial x^j}. \quad (12.140)$$

The following two rules are often useful in deriving vector identities

$$\epsilon^{ijk} \epsilon_{iab} = \delta_a^j \delta_b^k - \delta_b^j \delta_a^k, \quad (12.141)$$

$$\epsilon^{ijk} \epsilon_{ijb} = 2 \delta_b^k. \quad (12.142)$$

Up to now, we have restricted ourselves to three basic types of coordinate transformation: namely, translations, rotations, and standard Lorentz transformations. An arbitrary combination of these three transformations constitutes a general Lorentz transformation. Let us now extend

our investigations to include a fourth type of transformation known as a *parity inversion*: that is, $x, y, z, \rightarrow -x, -y, -z$. A reflection is a combination of a parity inversion and a rotation. As is easily demonstrated, the Jacobian of a parity inversion is -1 , unlike a translation, rotation, or standard Lorentz transformation, which all possess Jacobians of $+1$.

The prototype of all 3-vectors is the difference in coordinates between two points in space, \mathbf{r} . Likewise, the prototype of all 4-vectors is the difference in coordinates between two events in space-time, $R^\mu = (\mathbf{r}, ct)$. It is not difficult to appreciate that both of these objects are invariant under a parity transformation (in the sense that they correspond to the same geometric object before and after the transformation). It follows that any 3- or 4-tensor which is directly related to \mathbf{r} and R^μ , respectively, is also invariant under a parity inversion. Such tensors include the distance between two points in 3-space, the interval between two points in space-time, 3-velocity, 3-acceleration, 4-velocity, 4-acceleration, and the metric tensor. Tensors that exhibit tensor behavior under translations, rotations, special Lorentz transformations, and are invariant under parity inversions, are termed *proper tensors*, or sometimes *polar tensors*. Because electric charge is clearly invariant under such transformations (i.e., it is a proper scalar), it follows that 3-current and 4-current are proper vectors. It is also clear from Equation (12.120) that the scalar potential, the vector potential, and the potential 4-vector, are proper tensors.

It follows from Equation (12.137) that $\epsilon^{\alpha\beta\gamma\delta} \rightarrow -\epsilon^{\alpha\beta\gamma\delta}$ under a parity inversion. Tensors such as this, which exhibit tensor behavior under translations, rotations, and special Lorentz transformations, but are not invariant under parity inversions (in the sense that they correspond to different geometric objects before and after the transformation), are called *pseudo-tensors*, or sometimes *axial tensors*. Equations (12.139) and (12.140) imply that the cross product of two proper vectors is a pseudo-vector, and the curl of a proper vector field is a pseudo-vector field.

One particularly simple way of performing a parity transformation is to exchange positive and negative numbers on the three Cartesian axes. A proper vector is unaffected by such a procedure (i.e., its magnitude and direction are the same before and after). On the other hand, a pseudo-vector ends up pointing in the opposite direction after the axes are renumbered.

What is the fundamental difference between proper tensors and pseudo-tensors? The answer is that all pseudo-tensors are defined according to a handedness convention. For instance, the cross product between two vectors is conventionally defined according to a right-hand rule. The only reason for this is that the majority of human beings are right-handed. Presumably, if the opposite were true then cross products, et cetera, would be defined according to a left-hand rule, and would, therefore, take minus their conventional values. The totally antisymmetric tensor is the prototype pseudo-tensor, and is, of course, conventionally defined with respect to a right-handed spatial coordinate system. A parity inversion converts left into right, and vice versa, and, thereby, effectively swaps left- and right-handed conventions.

The use of conventions in physics is perfectly acceptable provided that we recognize that they are conventions, and are consistent in our use of them. It follows that laws of physics cannot contain mixtures of tensors and pseudo-tensors, otherwise they would depend our choice of handedness convention.¹

¹Here, we are assuming that the laws of physics do not possess an intrinsic handedness. This is certainly the case for mechanics and electromagnetism. However, the weak interaction does possess an intrinsic handedness: that is, it

Let us now consider electric and magnetic fields. We know that

$$\mathbf{E} = -\nabla\phi - \frac{\partial\mathbf{A}}{\partial t}, \quad (12.143)$$

$$\mathbf{B} = \nabla \times \mathbf{A}. \quad (12.144)$$

We have already seen that the scalar and the vector potential are proper scalars and vectors, respectively. It follows that \mathbf{E} is a proper vector, but that \mathbf{B} is a pseudo-vector (because it is the curl of a proper vector). In order to fully appreciate the difference between electric and magnetic fields, let us consider a thought experiment first proposed by Richard Feynman. Suppose that we are in radio contact with a race of aliens, and are trying to explain to them our system of physics. Suppose, further, that the aliens live sufficiently far away from us that there are no common objects that we can both see. The question is this: could we unambiguously explain to these aliens our concepts of electric and magnetic fields? We could certainly explain electric and magnetic lines of force. The former are the paths of charged particles (assuming that the particles are subject only to electric fields), and the latter can be mapped out using small test magnets. We could also explain how we put arrows on electric lines of force to convert them into electric field-lines: the arrows run from positive charges (i.e., charges with the same sign as atomic nuclei) to negative charges. This explanation is unambiguous provided that our aliens live in a matter (rather than an anti-matter) dominated part of the universe. But, could we explain how we put arrows on magnetic lines of force in order to convert them into magnetic field-lines? The answer is no. By definition, magnetic field-lines emerge from the north poles of permanent magnets and converge on the corresponding south poles. The definition of the north pole of a magnet is simply that it possesses the same magnetic polarity as the south (geographic) pole of the Earth. This is obviously a convention. In fact, we could redefine magnetic field-lines to run from the south poles to the north poles of magnets without significantly altering our laws of physics (we would just have to replace \mathbf{B} by $-\mathbf{B}$ in all our equations). In a parity inverted universe, a north pole becomes a south pole, and vice versa, so it is hardly surprising that $\mathbf{B} \rightarrow -\mathbf{B}$.

12.15 Electromagnetic Field Tensor

Let us now investigate whether we can write the components of the electric and magnetic fields as the components of some proper 4-tensor. There is an obvious problem here. How can we identify the components of the magnetic field, which is a pseudo-vector, with any of the components of a proper-4-tensor? The former components transform differently under parity inversion than the latter components. Consider a proper-3-tensor whose covariant components are written B_{ik} , and which is antisymmetric:

$$B_{ij} = -B_{ji}. \quad (12.145)$$

This immediately implies that all of the diagonal components of the tensor are zero. In fact, there are only three independent non-zero components of such a tensor. Could we, perhaps, use these

is fundamentally different in a parity inverted universe. So, the equations governing the weak interaction do actually contain mixtures of tensors and pseudo-tensors.

components to represent the components of a pseudo-3-vector? Let us write

$$B^i = \frac{1}{2} \epsilon^{ijk} B_{jk}. \quad (12.146)$$

It is clear that B^i transforms as a contravariant pseudo-3-vector. It is easily seen that

$$B^{ij} = B_{ij} = \begin{pmatrix} 0 & B_z & -B_y \\ -B_z & 0 & B_x \\ B_y & -B_x & 0 \end{pmatrix}, \quad (12.147)$$

where $B^1 = B_1 \equiv B_x$, et cetera. In this manner, we can actually write the components of a pseudo-3-vector as the components of an antisymmetric proper-3-tensor. In particular, we can write the components of the magnetic field \mathbf{B} in terms of an antisymmetric proper magnetic field 3-tensor which we shall denote B_{ij} .

Let us now examine Equations (12.143) and (12.144) more carefully. Recall that $\Phi_\mu = (-c \mathbf{A}, \phi)$ and $\partial_\mu = (\nabla, c^{-1} \partial/\partial t)$. It follows that we can write Equation (12.143) in the form

$$E_i = -\partial_i \Phi_4 + \partial_4 \Phi_i. \quad (12.148)$$

Likewise, Equation (12.144) can be written

$$c B^i = \frac{1}{2} \epsilon^{ijk} c B_{jk} = -\epsilon^{ijk} \partial_j \Phi_k. \quad (12.149)$$

Let us multiply this expression by ϵ_{iab} , making use of the identity

$$\epsilon_{iab} \epsilon^{ijk} = \delta_a^j \delta_b^k - \delta_b^j \delta_a^k. \quad (12.150)$$

We obtain

$$\frac{c}{2} (B_{ab} - B_{ba}) = -\partial_a \Phi_b + \partial_b \Phi_a, \quad (12.151)$$

or

$$c B_{ij} = -\partial_i \Phi_j + \partial_j \Phi_i, \quad (12.152)$$

because $B_{ij} = -B_{ji}$.

Let us define a proper-4-tensor whose covariant components are given by

$$F_{\mu\nu} = \partial_\mu \Phi_\nu - \partial_\nu \Phi_\mu. \quad (12.153)$$

It is clear that this tensor is antisymmetric:

$$F_{\mu\nu} = -F_{\nu\mu}. \quad (12.154)$$

This implies that the tensor only possesses six independent non-zero components. Maybe it can be used to specify the components of \mathbf{E} and \mathbf{B} ?

Equations (12.148) and (12.153) yield

$$F_{4i} = \partial_4 \Phi_i - \partial_i \Phi_4 = E_i. \quad (12.155)$$

Likewise, Equations (12.152) and (12.153) imply that

$$F_{ij} = \partial_i \Phi_j - \partial_j \Phi_i = -c B_{ij}. \quad (12.156)$$

Thus,

$$F_{i4} = -F_{4i} = -E_i, \quad (12.157)$$

$$F_{ij} = -F_{ji} = -c B_{ij}. \quad (12.158)$$

In other words, the completely space-like components of the tensor specify the components of the magnetic field, whereas the hybrid space and time-like components specify the components of the electric field. The covariant components of the tensor can be written

$$F_{\mu\nu} = \begin{pmatrix} 0 & -c B_z & +c B_y & -E_x \\ +c B_z & 0 & -c B_x & -E_y \\ -c B_y & +c B_x & 0 & -E_z \\ +E_x & +E_y & +E_z & 0 \end{pmatrix}. \quad (12.159)$$

Not surprisingly, $F_{\mu\nu}$ is usually called the *electromagnetic field tensor*. The previous expression, which appears in all standard textbooks, is very misleading. Taken at face value, it is simply wrong. We cannot form a proper-4-tensor from the components of a proper-3-vector and a pseudo-3-vector. The expression only makes sense if we interpret B_x (say) as representing the component B_{23} of the proper magnetic field 3-tensor B_{ij} .

The contravariant components of the electromagnetic field tensor are given by

$$F^{i4} = -F^{4i} = +E^i, \quad (12.160)$$

$$F^{ij} = -F^{ji} = -c B^{ij}, \quad (12.161)$$

or

$$F^{\mu\nu} = \begin{pmatrix} 0 & -c B_z & +c B_y & +E_x \\ +c B_z & 0 & -c B_x & +E_y \\ -c B_y & +c B_x & 0 & +E_z \\ -E_x & -E_y & -E_z & 0 \end{pmatrix}. \quad (12.162)$$

Let us now consider two of Maxwell's equations:

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0}, \quad (12.163)$$

$$\nabla \times \mathbf{B} = \mu_0 \left(\mathbf{j} + \epsilon_0 \frac{\partial \mathbf{E}}{\partial t} \right). \quad (12.164)$$

Recall that the 4-current is defined $J^\mu = (\mathbf{j}, \rho c)$. The first of these equations can be written

$$\partial_i E^i = \partial_i F^{i4} + \partial_4 F^{44} = \frac{J^4}{c \epsilon_0}. \quad (12.165)$$

because $F^{44} = 0$. The second of these equations takes the form

$$\epsilon^{ijk} \partial_j (c B_k) - \partial_4 E^i = \epsilon^{ijk} \partial_j (1/2 \epsilon_{kab} c B^{ab}) + \partial_4 F^{4i} = \frac{J^i}{c \epsilon_0}. \quad (12.166)$$

Making use of Equation (12.150), the previous expression reduces to

$$\frac{1}{2} \partial_j (c B^{ij} - c B^{ji}) + \partial_4 F^{4i} = \partial_j F^{ji} + \partial_4 F^{4i} = \frac{J^i}{c \epsilon_0}. \quad (12.167)$$

Equations (12.165) and (12.167) can be combined to give

$$\partial_\mu F^{\mu\nu} = \frac{J^\nu}{c \epsilon_0}. \quad (12.168)$$

This equation is consistent with the equation of charge continuity, $\partial_\mu J^\mu = 0$, because of the anti-symmetry of the electromagnetic field tensor.

12.16 Dual Electromagnetic Field Tensor

We have seen that it is possible to write the components of the electric and magnetic fields as the components of a proper-4-tensor. Is it also possible to write the components of these fields as the components of some pseudo-4-tensor? It is obvious that we cannot identify the components of the proper-3-vector \mathbf{E} with any of the components of a pseudo-tensor. However, we can represent the components of \mathbf{E} in terms of those of an antisymmetric pseudo-3-tensor E_{ij} by writing

$$E^i = \frac{1}{2} \epsilon^{ijk} E_{jk}. \quad (12.169)$$

It is easily demonstrated that

$$E^{ij} = E_{ij} = \begin{pmatrix} 0 & E_z & -E_y \\ -E_z & 0 & E_x \\ E_y & -E_x & 0 \end{pmatrix}, \quad (12.170)$$

in a right-handed coordinate system.

Consider the *dual electromagnetic field tensor*, $G^{\mu\nu}$, which is defined

$$G^{\mu\nu} = \frac{1}{2} \epsilon^{\mu\nu\alpha\beta} F_{\alpha\beta}. \quad (12.171)$$

This tensor is clearly an antisymmetric pseudo-4-tensor. We have

$$G^{4i} = \frac{1}{2} \epsilon^{4ijk} F_{jk} = -\frac{1}{2} \epsilon^{ijk4} F_{jk} = \frac{1}{2} \epsilon^{ijk} c B_{jk} = c B^i, \quad (12.172)$$

plus

$$G^{ij} = \frac{1}{2} (\epsilon^{ijk4} F_{k4} + \epsilon^{ij4k} F_{4k}) = \epsilon^{ijk} F_{k4}, \quad (12.173)$$

where use has been made of $F_{\mu\nu} = -F_{\nu\mu}$. The previous expression yields

$$G^{ij} = -\epsilon^{ijk} E_k = -\frac{1}{2} \epsilon^{ijk} \epsilon_{kab} E^{ab} = -E^{ij}. \quad (12.174)$$

It follows that

$$G^{i4} = -G^{4i} = -c B^i, \quad (12.175)$$

$$G^{ij} = -G^{ji} = -E^{ij}, \quad (12.176)$$

or

$$G^{\mu\nu} = \begin{pmatrix} 0 & -E_z & +E_y & -c B_x \\ +E_z & 0 & -E_x & -c B_y \\ -E_y & +E_x & 0 & -c B_z \\ +c B_x & +c B_y & +c B_z & 0 \end{pmatrix}. \quad (12.177)$$

The previous expression is, again, slightly misleading, because E_x stands for the component E^{23} of the pseudo-3-tensor E^{ij} , and not for an element of the proper-3-vector \mathbf{E} . Of course, in this case, B_x really does represent the first element of the pseudo-3-vector \mathbf{B} . Note that the elements of $G^{\mu\nu}$ are obtained from those of $F^{\mu\nu}$ by making the transformation $c B^{ij} \rightarrow E^{ij}$ and $E^i \rightarrow -c B^i$.

The covariant elements of the dual electromagnetic field tensor are given by

$$G_{i4} = -G_{4i} = +c B_i, \quad (12.178)$$

$$G_{ij} = -G_{ji} = -E_{ij}, \quad (12.179)$$

or

$$G_{\mu\nu} = \begin{pmatrix} 0 & -E_z & +E_y & +c B_x \\ +E_z & 0 & -E_x & +c B_y \\ -E_y & +E_x & 0 & +c B_z \\ -c B_x & -c B_y & -c B_z & 0 \end{pmatrix}. \quad (12.180)$$

The elements of $G_{\mu\nu}$ are obtained from those of $F_{\mu\nu}$ by making the transformation $c B_{ij} \rightarrow E_{ij}$ and $E_i \rightarrow -c B_i$.

Let us now consider the two Maxwell equations

$$\nabla \cdot \mathbf{B} = 0, \quad (12.181)$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}. \quad (12.182)$$

The first of these equations can be written

$$-\partial_i (c B^i) = \partial_i G^{i4} + \partial_4 G^{44} = 0, \quad (12.183)$$

because $G^{44} = 0$. The second equation takes the form

$$\epsilon^{ijk} \partial_j E_k = \epsilon^{ijk} \partial_j (1/2 \epsilon_{kab} E^{ab}) = \partial_j E^{ij} = -\partial_4 (c B^i), \quad (12.184)$$

or

$$\partial_j G^{ji} + \partial_4 G^{4i} = 0. \quad (12.185)$$

Equations (12.183) and (12.185) can be combined to give

$$\partial_\mu G^{\mu\nu} = 0. \quad (12.186)$$

Thus, we conclude that Maxwell's equations for the electromagnetic fields are equivalent to the following pair of 4-tensor equations:

$$\partial_\mu F^{\mu\nu} = \frac{J^\nu}{c \epsilon_0}, \quad (12.187)$$

$$\partial_\mu G^{\mu\nu} = 0. \quad (12.188)$$

It is obvious from the form of these equations that the laws of electromagnetism are invariant under translations, rotations, special Lorentz transformations, parity inversions, or any combination of these transformations.

12.17 Transformation of Fields

The electromagnetic field tensor transforms according to the standard rule

$$F^{\mu'\nu'} = F^{\mu\nu} p_\mu^{\mu'} p_\nu^{\nu'}. \quad (12.189)$$

This easily yields the celebrated rules for transforming electromagnetic fields:

$$E'_\parallel = E_\parallel, \quad (12.190)$$

$$B'_\parallel = B_\parallel, \quad (12.191)$$

$$\mathbf{E}'_\perp = \gamma (\mathbf{E}_\perp + \mathbf{v} \times \mathbf{B}), \quad (12.192)$$

$$\mathbf{B}'_\perp = \gamma (\mathbf{B}_\perp - \mathbf{v} \times \mathbf{E}/c^2), \quad (12.193)$$

where \mathbf{v} is the relative velocity between the primed and unprimed frames, and the perpendicular and parallel directions are, respectively, perpendicular and parallel to \mathbf{v} .

At this stage, we may conveniently note two important invariants of the electromagnetic field. They are

$$\frac{1}{2} F_{\mu\nu} F^{\mu\nu} = c^2 B^2 - E^2, \quad (12.194)$$

and

$$\frac{1}{4} G_{\mu\nu} F^{\mu\nu} = c \mathbf{E} \cdot \mathbf{B}. \quad (12.195)$$

The first of these quantities is a proper-scalar, and the second a pseudo-scalar.

12.18 Potential Due to a Moving Charge

Suppose that a particle carrying a charge e moves with uniform velocity \mathbf{u} through a frame S . Let us evaluate the vector potential, \mathbf{A} , and the scalar potential, ϕ , due to this charge at a given event P in S .

Let us choose coordinates in S so that $P = (0, 0, 0, 0)$ and $\mathbf{u} = (u, 0, 0)$. Let S' be that frame in the standard configuration with respect to S in which the charge is (permanently) at rest at (say) the point (x', y', z') . In S' , the potential at P is the usual potential due to a stationary charge,

$$\mathbf{A}' = \mathbf{0}, \quad (12.196)$$

$$\phi' = \frac{e}{4\pi\epsilon_0 r'}, \quad (12.197)$$

where $r' = \sqrt{x'^2 + y'^2 + z'^2}$. Let us now transform these equations directly into the frame S . Because $A^\mu = (c\mathbf{A}, \phi)$ is a contravariant 4-vector, its components transform according to the standard rules (12.75)–(12.78). Thus,

$$cA_1 = \gamma \left(cA'_1 + \frac{u}{c} \phi' \right) = \frac{\gamma u e}{4\pi\epsilon_0 c r'}, \quad (12.198)$$

$$cA_2 = cA'_2 = 0, \quad (12.199)$$

$$cA_3 = cA'_3 = 0, \quad (12.200)$$

$$\phi = \gamma \left(\phi' + \frac{u}{c} cA'_1 \right) = \frac{\gamma e}{4\pi\epsilon_0 r'}, \quad (12.201)$$

because $\beta = -u/c$ in this case. It remains to express the quantity r' in terms of quantities measured in S . The most physically meaningful way of doing this is to express r' in terms of retarded values in S . Consider the retarded event at the charge for which, by definition, $r' = -ct'$ and $r = -ct$. Using the standard Lorentz transformation, (12.24)–(12.27), we find that

$$r' = -ct' = -c\gamma(t - ux/c^2) = r\gamma(1 + u_r/c), \quad (12.202)$$

where $u_r = ux/r = \mathbf{r} \cdot \mathbf{u}/r$ denotes the radial velocity of the charge in S . We can now rewrite Equations (12.198)–(12.201) in the form

$$\mathbf{A} = \frac{\mu_0 e}{4\pi} \frac{[\mathbf{u}]}{[r + \mathbf{r} \cdot \mathbf{u}/c]}, \quad (12.203)$$

$$\phi = \frac{e}{4\pi\epsilon_0} \frac{1}{[r + \mathbf{r} \cdot \mathbf{u}/c]}, \quad (12.204)$$

where the square brackets, as usual, indicate that the enclosed quantities must be retarded. For a uniformly moving charge, the retardation of \mathbf{u} is, of course, superfluous. However, because

$$\Phi^\mu = \frac{1}{4\pi\epsilon_0 c} \int \frac{[J^\mu]}{r} dV, \quad (12.205)$$

it is clear that the potentials depend only on the (retarded) velocity of the charge, and not on its acceleration. Consequently, the expressions (12.203) and (12.204) give the correct potentials for an arbitrarily moving charge. They are known as the *Liénard-Wiechert potentials*.

12.19 Field Due to a Moving Charge

Although the fields generated by a uniformly moving charge can be calculated from the expressions (12.203) and (12.204) for the potentials, it is simpler to calculate them from first principles.

Let a charge e , whose position vector at time $t = 0$ is \mathbf{r} , move with uniform velocity \mathbf{u} in a frame S whose x -axis has been chosen in the direction of \mathbf{u} . We require to find the field strengths \mathbf{E} and \mathbf{B} at the event $P = (0, 0, 0, 0)$. Let S' be that frame in standard configuration with S in which the charge is permanently at rest. In S' , the field is given by

$$\mathbf{B}' = \mathbf{0}, \quad (12.206)$$

$$\mathbf{E}' = -\frac{e}{4\pi\epsilon_0} \frac{\mathbf{r}'}{r'^3}. \quad (12.207)$$

This field must now be transformed into the frame S . The direct method, using Equations (12.190)–(12.193), is somewhat simpler here, but we shall use a somewhat indirect method because of its intrinsic interest.

In order to express Equations (12.206) and (12.207) in tensor form, we need the electromagnetic field tensor $F^{\mu\nu}$ on the left-hand side, and the position 4-vector $R^\mu = (\mathbf{r}, ct)$ and the scalar $e/(4\pi\epsilon_0 r'^3)$ on the right-hand side. (We regard r' as an invariant for all observers.) To get a vanishing magnetic field in S' , we multiply on the right by the 4-velocity $U^\mu = \gamma(u)(\mathbf{u}, c)$, thus tentatively arriving at the equation

$$F^{\mu\nu} = \frac{e}{4\pi\epsilon_0 c r'^3} U^\mu R^\nu. \quad (12.208)$$

Recall that $F^{4i} = -E^i$ and $F^{ij} = -cB^{ij}$. However, this equation cannot be correct, because the antisymmetric tensor $F^{\mu\nu}$ can only be equated to another antisymmetric tensor. Consequently, let us try

$$F^{\mu\nu} = \frac{e}{4\pi\epsilon_0 c r'^3} (U^\mu R^\nu - U^\nu R^\mu). \quad (12.209)$$

This is found to give the correct field at P in S' , as long as R^μ refers to any event whatsoever at the charge. It only remains to interpret Equation (12.209) in S . It is convenient to choose for R^μ that event at the charge at which $t = 0$ (not the retarded event). Thus,

$$F^{jk} = -cB^{jk} = \frac{e}{4\pi\epsilon_0 c r'^3} \gamma(u) (u^j r^k - u^k r^j), \quad (12.210)$$

giving

$$B_i = \frac{1}{2} \epsilon_{ijk} B^{jk} = -\frac{\mu_0 e}{4\pi r'^3} \gamma(u) \epsilon_{ijk} u^j r^k, \quad (12.211)$$

or

$$\mathbf{B} = -\frac{\mu_0 e \gamma}{4\pi r'^3} \mathbf{u} \times \mathbf{r}. \quad (12.212)$$

Likewise,

$$F^{4i} = -E^i = \frac{e \gamma}{4\pi\epsilon_0 r'^3} r^i, \quad (12.213)$$

or

$$\mathbf{E} = -\frac{e\gamma}{4\pi\epsilon_0 r'^3} \mathbf{r}. \quad (12.214)$$

Lastly, we must find an expression for r'^3 in terms of quantities measured in S at time $t = 0$. If t' is the corresponding time in S' at the charge then we have

$$r'^2 = r^2 + c^2 t'^2 = r^2 + \frac{\gamma^2 u^2 x^2}{c^2} = r^2 \left(1 + \frac{\gamma^2 u_r^2}{c^2} \right). \quad (12.215)$$

Thus,

$$\mathbf{E} = -\frac{e}{4\pi\epsilon_0 r^3} \frac{\gamma}{(1 + u_r^2 \gamma^2 / c^2)^{3/2}} \mathbf{r}, \quad (12.216)$$

$$\mathbf{B} = -\frac{\mu_0 e}{4\pi r^3} \frac{\gamma}{(1 + u_r^2 \gamma^2 / c^2)^{3/2}} \mathbf{u} \times \mathbf{r} = \frac{1}{c^2} \mathbf{u} \times \mathbf{E}. \quad (12.217)$$

Note that \mathbf{E} acts in line with the point which the charge occupies at the instant of measurement, despite the fact that, owing to the finite speed of propagation of all physical effects, the behavior of the charge during a finite period before that instant can no longer affect the measurement. Note also that, unlike Equations (12.203) and (12.204), the previous expressions for the fields are not valid for an arbitrarily moving charge, nor can they be made valid by merely using retarded values. For whereas acceleration does not affect the potentials, it does affect the fields, which involve the derivatives of the potential.

For low velocities, $u/c \rightarrow 0$, Equations (12.216) and (12.217) reduce to the well-known Coulomb and Biot-Savart fields. However, at high velocities, $\gamma(u) \gg 1$, the fields exhibit some interesting behavior. The peak electric field, which occurs at the point of closest approach of the charge to the observation point, becomes equal to γ times its non-relativistic value. However, the duration of appreciable field strength at the point P is decreased. A measure of the time interval over which the field is appreciable is

$$\Delta t \sim \frac{b}{\gamma c}, \quad (12.218)$$

where b is the distance of closest approach (assuming $\gamma \gg 1$). As γ increases, the peak field increases in proportion, but its duration goes in the inverse proportion. The time integral of the field is independent of γ . As $\gamma \rightarrow \infty$, the observer at P sees electric and magnetic fields that are indistinguishable from the fields of a pulse of plane polarized radiation propagating in the x -direction. The direction of polarization is along the radius vector pointing towards the particle's actual position at the time of observation.

12.20 Relativistic Particle Dynamics

Consider a particle that, in its instantaneous rest frame S_0 , has mass m_0 and constant acceleration in the x -direction a_0 . Let us transform to a frame S , in the standard configuration with respect to S_0 , in which the particle's instantaneous velocity is u . What is the value of a , the particle's instantaneous x -acceleration, in S ?

The easiest way in which to answer this question is to consider the acceleration 4-vector [see Equation (12.107)]

$$A^\mu = \gamma \left(\frac{d\gamma}{dt} \mathbf{u} + \gamma \mathbf{a}, c \frac{d\gamma}{dt} \right). \quad (12.219)$$

Using the standard transformation, (12.75)–(12.78), for 4-vectors, we obtain

$$a_0 = \gamma^3 a, \quad (12.220)$$

$$\frac{d\gamma}{dt} = \frac{u a_0}{c^2}. \quad (12.221)$$

Equation (12.220) can be written

$$f = m_0 \gamma^3 \frac{du}{dt}, \quad (12.222)$$

where $f = m_0 a_0$ is the constant force (in the x -direction) acting on the particle in S_0 .

Equation (12.222) is equivalent to

$$f = \frac{d(mu)}{dt}, \quad (12.223)$$

where

$$m = \gamma m_0. \quad (12.224)$$

Thus, we can account for the ever decreasing acceleration of a particle subject to a constant force [see Equation (12.220)] by supposing that the inertial mass of the particle increases with its velocity according to the rule (12.224). Henceforth, m_0 is termed the *rest mass*, and m the *inertial mass*.

The rate of increase of the particle's energy E satisfies

$$\frac{dE}{dt} = f u = m_0 \gamma^3 u \frac{du}{dt}. \quad (12.225)$$

This equation can be written

$$\frac{dE}{dt} = \frac{d(mc^2)}{dt}, \quad (12.226)$$

which can be integrated to yield Einstein's famous formula

$$E = m c^2. \quad (12.227)$$

The 3-momentum of a particle is defined

$$\mathbf{p} = m \mathbf{u}, \quad (12.228)$$

where \mathbf{u} is its 3-velocity. Thus, by analogy with Equation (12.223), Newton's law of motion can be written

$$\mathbf{f} = \frac{d\mathbf{p}}{dt}, \quad (12.229)$$

where \mathbf{f} is the 3-force acting on the particle.

The 4-momentum of a particle is defined

$$P^\mu = m_0 U^\mu = \gamma m_0 (\mathbf{u}, c) = (\mathbf{p}, E/c), \quad (12.230)$$

where U^μ is its 4-velocity. The 4-force acting on the particle obeys

$$\mathcal{F}^\mu = \frac{dP^\mu}{d\tau} = m_0 A^\mu, \quad (12.231)$$

where A^μ is its 4-acceleration. It is easily demonstrated that

$$\mathcal{F}^\mu = \gamma \left(\mathbf{f}, c \frac{dm}{dt} \right) = \gamma \left(\mathbf{f}, \frac{\mathbf{f} \cdot \mathbf{u}}{c} \right), \quad (12.232)$$

because

$$\frac{dE}{dt} = \mathbf{f} \cdot \mathbf{u}. \quad (12.233)$$

12.21 Force on a Moving Charge

The electromagnetic 3-force acting on a charge e moving with 3-velocity \mathbf{u} is given by the well-known formula

$$\mathbf{f} = e(\mathbf{E} + \mathbf{u} \times \mathbf{B}). \quad (12.234)$$

When written in component form this expression becomes

$$f_i = e(E_i + \epsilon_{ijk} u^j B^k), \quad (12.235)$$

or

$$f_i = e(E_i + B_{ij} u^j), \quad (12.236)$$

where use has been made of Equation (12.147).

Recall that the components of the \mathbf{E} and \mathbf{B} fields can be written in terms of an antisymmetric electromagnetic field tensor $F_{\mu\nu}$ via

$$F_{i4} = -F_{4i} = -E_i, \quad (12.237)$$

$$F_{ij} = -F_{ji} = -c B_{ij}. \quad (12.238)$$

Equation (12.236) can be written

$$f_i = -\frac{e}{\gamma c} (F_{i4} U^4 + F_{ij} U^j), \quad (12.239)$$

where $U^\mu = \gamma(\mathbf{u}, c)$ is the particle's 4-velocity. It is easily demonstrated that

$$\frac{\mathbf{f} \cdot \mathbf{u}}{c} = \frac{e}{c} \mathbf{E} \cdot \mathbf{u} = \frac{e}{c} E_i u^i = \frac{e}{\gamma c} (F_{4i} U^i + F_{44} U^4). \quad (12.240)$$

Thus, the 4-force acting on the particle,

$$\mathcal{F}_\mu = \gamma \left(-\mathbf{f}, \frac{\mathbf{f} \cdot \mathbf{u}}{c} \right), \quad (12.241)$$

can be written in the form

$$\mathcal{F}_\mu = \frac{e}{c} F_{\mu\nu} U^\nu. \quad (12.242)$$

The skew symmetry of the electromagnetic field tensor ensures that

$$\mathcal{F}_\mu U^\mu = \frac{e}{c} F_{\mu\nu} U^\mu U^\nu = 0. \quad (12.243)$$

This is an important result, because it ensures that electromagnetic fields do not change the rest mass of charged particles. In order to appreciate this, let us assume that the rest mass m_0 is not a constant. Because

$$\mathcal{F}_\mu = \frac{d(m_0 U_\mu)}{d\tau} = m_0 A_\mu + \frac{dm_0}{d\tau} U_\mu, \quad (12.244)$$

we can use the standard results $U_\mu U^\mu = c^2$ and $A_\mu U^\mu = 0$ to give

$$\mathcal{F}_\mu U^\mu = c^2 \frac{dm_0}{d\tau}. \quad (12.245)$$

Thus, if rest mass is to remain an invariant, it is imperative that all laws of physics predict 4-forces acting on particles that are orthogonal to the particles' instantaneous 4-velocities. The laws of electromagnetism pass this test.

12.22 Electromagnetic Energy Tensor

Consider a continuous volume distribution of charged matter in the presence of an electromagnetic field. Let there be n_0 particles per unit proper volume (that is, unit volume determined in the local rest frame), each carrying a charge e . Consider an inertial frame in which the 3-velocity field of the particles is \mathbf{u} . The number density of the particles in this frame is $n = \gamma(u) n_0$. The charge density and the 3-current due to the particles are $\rho = en$ and $\mathbf{j} = en\mathbf{u}$, respectively. Multiplying Equation (12.242) by the proper number density of particles, n_0 , we obtain an expression

$$f_\mu = c^{-1} F_{\mu\nu} J^\nu \quad (12.246)$$

for the 4-force f_μ acting on unit proper volume of the distribution due to the ambient electromagnetic fields. Here, we have made use of the definition $J^\mu = en_0 U^\mu$. It is easily demonstrated, using some of the results obtained in the previous section, that

$$f^\mu = \left(\rho \mathbf{E} + \mathbf{j} \times \mathbf{B}, \frac{\mathbf{E} \cdot \mathbf{j}}{c} \right). \quad (12.247)$$

The previous expression remains valid when there are many charge species (e.g., electrons and ions) possessing different number density and 3-velocity fields. The 4-vector f^μ is usually called the *Lorentz force density*.

We know that Maxwell's equations reduce to

$$\partial_\mu F^{\mu\nu} = \frac{J^\nu}{c \epsilon_0}, \quad (12.248)$$

$$\partial_\mu G^{\mu\nu} = 0, \quad (12.249)$$

where $F^{\mu\nu}$ is the electromagnetic field tensor, and $G^{\mu\nu}$ is its dual. As is easily verified, Equation (12.249) can also be written in the form

$$\partial_\mu F_{\nu\sigma} + \partial_\nu F_{\sigma\mu} + \partial_\sigma F_{\mu\nu} = 0. \quad (12.250)$$

Equations (12.246) and (12.248) can be combined to give

$$f_\nu = \epsilon_0 F_{\nu\sigma} \partial_\mu F^{\mu\sigma}. \quad (12.251)$$

This expression can also be written

$$f_\nu = \epsilon_0 \left[\partial_\mu (F^{\mu\sigma} F_{\nu\sigma}) - F^{\mu\sigma} \partial_\mu F_{\nu\sigma} \right]. \quad (12.252)$$

Now,

$$F^{\mu\sigma} \partial_\mu F_{\nu\sigma} = \frac{1}{2} F^{\mu\sigma} (\partial_\mu F_{\nu\sigma} + \partial_\sigma F_{\mu\nu}), \quad (12.253)$$

where use has been made of the antisymmetry of the electromagnetic field tensor. It follows from Equation (12.250) that

$$F^{\mu\sigma} \partial_\mu F_{\nu\sigma} = -\frac{1}{2} F^{\mu\sigma} \partial_\nu F_{\sigma\mu} = \frac{1}{4} \partial_\nu (F^{\mu\sigma} F_{\mu\sigma}). \quad (12.254)$$

Thus,

$$f_\nu = \epsilon_0 \left[\partial_\mu (F^{\mu\sigma} F_{\nu\sigma}) - \frac{1}{4} \partial_\nu (F^{\mu\sigma} F_{\mu\sigma}) \right]. \quad (12.255)$$

The previous expression can also be written

$$f_\nu = -\partial_\mu T^\mu{}_\nu, \quad (12.256)$$

where

$$T^\mu{}_\nu = \epsilon_0 \left[F^{\mu\sigma} F_{\sigma\nu} + \frac{1}{4} \delta_\nu^\mu (F^{\rho\sigma} F_{\rho\sigma}) \right] \quad (12.257)$$

is called the *electromagnetic energy tensor*. Note that $T^\mu{}_\nu$ is a proper-4-tensor. It follows from Equations (12.159), (12.162), and (12.194) that

$$T^i{}_j = \epsilon_0 E^i E_j + \frac{B^i B_j}{\mu_0} - \delta_j^i \frac{1}{2} \left(\epsilon_0 E^k E_k + \frac{B^k B_k}{\mu_0} \right), \quad (12.258)$$

$$T^i{}_4 = -T^4{}_i = \frac{\epsilon^{ijk} E_j B_k}{\mu_0 c}, \quad (12.259)$$

$$T^4{}_4 = \frac{1}{2} \left(\epsilon_0 E^k E_k + \frac{B^k B_k}{\mu_0} \right). \quad (12.260)$$

Equation (12.256) can also be written

$$f^\nu = -\partial_\mu T^{\mu\nu}, \quad (12.261)$$

where $T^{\mu\nu}$ is a symmetric tensor whose elements are

$$T^{ij} = -\epsilon_0 E^i E^j - \frac{B^i B^j}{\mu_0} + \delta^{ij} \frac{1}{2} \left(\epsilon_0 E^2 + \frac{B^2}{\mu_0} \right), \quad (12.262)$$

$$T^{i4} = T^{4i} = \frac{(\mathbf{E} \times \mathbf{B})^i}{\mu_0 c}, \quad (12.263)$$

$$T^{44} = \frac{1}{2} \left(\epsilon_0 E^2 + \frac{B^2}{\mu_0} \right). \quad (12.264)$$

Consider the time-like component of Equation (12.261). It follows from Equation (12.247) that

$$\frac{\mathbf{E} \cdot \mathbf{j}}{c} = -\partial_i T^{i4} - \partial_4 T^{44}. \quad (12.265)$$

This equation can be rearranged to give

$$\frac{\partial U}{\partial t} + \nabla \cdot \mathbf{u} = -\mathbf{E} \cdot \mathbf{j}, \quad (12.266)$$

where $U = T^{44}$ and $\mathbf{u}^i = c T^{i4}$, so that

$$U = \frac{\epsilon_0 E^2}{2} + \frac{B^2}{2\mu_0}, \quad (12.267)$$

and

$$\mathbf{u} = \frac{\mathbf{E} \times \mathbf{B}}{\mu_0}. \quad (12.268)$$

The right-hand side of Equation (12.266) represents the rate per unit volume at which energy is transferred from the electromagnetic field to charged particles. It is clear, therefore, that Equation (12.266) is an energy conservation equation for the electromagnetic field. (See Section 1.9.) The proper-3-scalar U can be identified as the energy density of the electromagnetic field, whereas the proper-3-vector \mathbf{u} is the energy flux due to the electromagnetic field: that is, the *Poynting flux*.

Consider the space-like components of Equation (12.261). It is easily demonstrated that these reduce to

$$\frac{\partial \mathbf{g}}{\partial t} + \nabla \cdot \mathbf{G} = -\rho \mathbf{E} - \mathbf{j} \times \mathbf{B}, \quad (12.269)$$

where $G^{ij} = T^{ij}$ and $g^i = T^{4i}/c$, or

$$G^{ij} = -\epsilon_0 E^i E^j - \frac{B^i B^j}{\mu_0} + \delta^{ij} \frac{1}{2} \left(\epsilon_0 E^2 + \frac{B^2}{\mu_0} \right), \quad (12.270)$$

and

$$\mathbf{g} = \frac{\mathbf{u}}{c^2} = \epsilon_0 \mathbf{E} \times \mathbf{B}. \quad (12.271)$$

Equation (12.269) is basically a *momentum conservation equation* for the electromagnetic field. (See Section 1.10.) The right-hand side represents the rate per unit volume at which momentum is transferred from the electromagnetic field to charged particles. The symmetric proper-3-tensor G^{ij} specifies the flux of electromagnetic momentum parallel to the i th axis crossing a surface normal to the j th axis. The proper-3-vector \mathbf{g} represents the momentum density of the electromagnetic field. It is clear that the energy conservation law (12.266) and the momentum conservation law (12.269) can be combined together to give the relativistically invariant energy-momentum conservation law (12.261).

12.23 Accelerated Charges

Let us calculate the electric and magnetic fields observed at position x^i and time t due to a charge e whose retarded position and time are $x^{i'}$ and t' , respectively. From now on (x^i, t) is termed the *field point* and $(x^{i'}, t')$ is termed the *source point*. It is assumed that we are given the retarded position of the charge as a function of its retarded time: i.e., $x^{i'}(t')$. The retarded velocity and acceleration of the charge are

$$u^i = \frac{dx^{i'}}{dt'}, \quad (12.272)$$

and

$$\dot{u}^i = \frac{du^{i'}}{dt'}, \quad (12.273)$$

respectively. The radius vector \mathbf{r} is defined to extend from the retarded position of the charge to the field point, so that $r^i = x^i - x^{i'}$. (Note that this is the opposite convention to that adopted in Sections 12.18 and 12.19). It follows that

$$\frac{d\mathbf{r}}{dt'} = -\mathbf{u}. \quad (12.274)$$

The field and the source point variables are connected by the retardation condition

$$r(x^i, x^{i'}) = [(x^i - x^{i'})(x_i - x_{i'})]^{1/2} = c(t - t'). \quad (12.275)$$

The potentials generated by the charge are given by the Liénard-Wiechert formulae,

$$\mathbf{A}(x^i, t) = \frac{\mu_0 e \mathbf{u}}{4\pi s}, \quad (12.276)$$

$$\phi(x^i, t) = \frac{e}{4\pi \epsilon_0 s}, \quad (12.277)$$

where $s = r - \mathbf{r} \cdot \mathbf{u}/c$ is a function both of the field point and the source point variables. Recall that the Liénard-Wiechert potentials are valid for accelerating, as well as uniformly moving, charges.

The fields \mathbf{E} and \mathbf{B} are derived from the potentials in the usual manner:

$$\mathbf{E} = -\nabla\phi - \frac{\partial\mathbf{A}}{\partial t}, \quad (12.278)$$

$$\mathbf{B} = \nabla \times \mathbf{A}. \quad (12.279)$$

However, the components of the gradient operator ∇ are partial derivatives at constant time, t , and not at constant time, t' . Partial differentiation with respect to the x^i compares the potentials at neighboring points at the same time, but these potential signals originate from the charge at different retarded times. Similarly, the partial derivative with respect to t implies constant x^i , and, hence, refers to the comparison of the potentials at a given field point over an interval of time during which the retarded coordinates of the source have changed. Because we only know the time variation of the particle's retarded position with respect to t' we must transform $\partial/\partial t|_{x^i}$ and $\partial/\partial x^i|_t$ to expressions involving $\partial/\partial t'|_{x^i}$ and $\partial/\partial x^i|_{t'}$.

Now, because x^i is assumed to be given as a function of t' , we have

$$r(x^i, x^i(t')) \equiv r(x^i, t') = c(t - t'), \quad (12.280)$$

which is a functional relationship between x^i , t , and t' . Note that

$$\left(\frac{\partial r}{\partial t'}\right)_{x^i} = -\frac{\mathbf{r} \cdot \mathbf{u}}{r}. \quad (12.281)$$

It follows that

$$\frac{\partial r}{\partial t} = c \left(1 - \frac{\partial t'}{\partial t}\right) = \frac{\partial r}{\partial t'} \frac{\partial t'}{\partial t} = -\frac{\mathbf{r} \cdot \mathbf{u}}{r} \frac{\partial t'}{\partial t}, \quad (12.282)$$

where all differentiation is at constant x^i . Thus,

$$\frac{\partial t'}{\partial t} = \frac{1}{1 - \mathbf{r} \cdot \mathbf{u}/rc} = \frac{r}{s}, \quad (12.283)$$

giving

$$\frac{\partial}{\partial t} = \frac{r}{s} \frac{\partial}{\partial t'}. \quad (12.284)$$

Similarly,

$$\nabla r = -c \nabla t' = \nabla' r + \frac{\partial r}{\partial t'} \nabla t' = \frac{\mathbf{r}}{r} - \frac{\mathbf{r} \cdot \mathbf{u}}{r} \nabla t', \quad (12.285)$$

where ∇' denotes differentiation with respect to x^i at constant t' . It follows that

$$\nabla t' = -\frac{\mathbf{r}}{sc}, \quad (12.286)$$

so that

$$\nabla = \nabla' - \frac{\mathbf{r}}{sc} \frac{\partial}{\partial t'}. \quad (12.287)$$

Equation (12.278) yields

$$\frac{4\pi \epsilon_0}{e} \mathbf{E} = \frac{\nabla s}{s^2} - \frac{\partial}{\partial t} \frac{\mathbf{u}}{sc^2}, \quad (12.288)$$

or

$$\frac{4\pi \epsilon_0}{e} \mathbf{E} = \frac{\nabla' s}{s^2} - \frac{\mathbf{r}}{s^3 c} \frac{\partial s}{\partial t'} - \frac{r}{s^2 c^2} \dot{\mathbf{u}} + \frac{r \mathbf{u}}{s^3 c^2} \frac{\partial s}{\partial t'}. \quad (12.289)$$

However,

$$\nabla' s = \frac{\mathbf{r}}{r} - \frac{\mathbf{u}}{c}, \quad (12.290)$$

and

$$\frac{\partial s}{\partial t'} = \frac{\partial r}{\partial t'} - \frac{\mathbf{r} \cdot \dot{\mathbf{u}}}{c} + \frac{\mathbf{u} \cdot \mathbf{u}}{c} = -\frac{\mathbf{r} \cdot \mathbf{u}}{r} - \frac{\mathbf{r} \cdot \dot{\mathbf{u}}}{c} + \frac{u^2}{c}. \quad (12.291)$$

Thus,

$$\frac{4\pi \epsilon_0}{e} \mathbf{E} = \frac{1}{s^2 r} \left(\mathbf{r} - \frac{r \mathbf{u}}{c} \right) + \frac{1}{s^3 c} \left(\mathbf{r} - \frac{r \mathbf{u}}{c} \right) \left(\frac{\mathbf{r} \cdot \mathbf{u}}{r} - \frac{u^2}{c} + \frac{\mathbf{r} \cdot \dot{\mathbf{u}}}{c} \right) - \frac{r}{s^2 c^2} \dot{\mathbf{u}}, \quad (12.292)$$

which reduces to

$$\frac{4\pi \epsilon_0}{e} \mathbf{E} = \frac{1}{s^3} \left(\mathbf{r} - \frac{r \mathbf{u}}{c} \right) \left(1 - \frac{u^2}{c^2} \right) + \frac{1}{s^3 c^2} \left(\mathbf{r} \times \left[\left(\mathbf{r} - \frac{r \mathbf{u}}{c} \right) \times \dot{\mathbf{u}} \right] \right). \quad (12.293)$$

Similarly,

$$\frac{4\pi}{\mu_0 e} \mathbf{B} = \nabla \times \frac{\mathbf{u}}{s} = -\frac{\nabla' s \times \mathbf{u}}{s^2} - \frac{\mathbf{r}}{s c} \times \left(\frac{\dot{\mathbf{u}}}{s} - \frac{\mathbf{u}}{s^2} \frac{\partial s}{\partial t'} \right), \quad (12.294)$$

or

$$\frac{4\pi}{\mu_0 e} \mathbf{B} = -\frac{\mathbf{r} \times \mathbf{u}}{s^2 r} - \frac{\mathbf{r}}{s c} \times \left[\frac{\dot{\mathbf{u}}}{s} + \frac{\mathbf{u}}{s^2} \left(\frac{\mathbf{r} \cdot \mathbf{u}}{r} + \frac{\mathbf{r} \cdot \dot{\mathbf{u}}}{c} - \frac{u^2}{c} \right) \right], \quad (12.295)$$

which reduces to

$$\frac{4\pi}{\mu_0 e} \mathbf{B} = \frac{\mathbf{u} \times \mathbf{r}}{s^3} \left(1 - \frac{u^2}{c^2} \right) + \frac{1}{s^3 c r} \mathbf{r} \times \left(\mathbf{r} \times \left[\left(\mathbf{r} - \frac{r \mathbf{u}}{c} \right) \times \dot{\mathbf{u}} \right] \right). \quad (12.296)$$

A comparison of Equations (12.293) and (12.296) yields

$$\mathbf{B} = \frac{\mathbf{r} \times \mathbf{E}}{r c}. \quad (12.297)$$

Thus, the magnetic field is always perpendicular to \mathbf{E} and the retarded radius vector \mathbf{r} . Note that all terms appearing in the previous formulae are retarded.

The electric field is composed of two separate parts. The first term in Equation (12.293) varies as $1/r^2$ for large distances from the charge. We can think of $\mathbf{r}_u = \mathbf{r} - r \mathbf{u}/c$ as the *virtual present radius vector*: that is, the radius vector directed from the position the charge would occupy at time t if it had continued with uniform velocity from its retarded position to the field point. In terms of \mathbf{r}_u , the $1/r^2$ field is simply

$$\mathbf{E}_{\text{induction}} = \frac{e}{4\pi \epsilon_0} \frac{1 - u^2/c^2}{s^3} \mathbf{r}_u. \quad (12.298)$$

We can rewrite the expression (12.216) for the electric field generated by a uniformly moving charge in the form

$$\mathbf{E} = \frac{e}{4\pi \epsilon_0 r_0^3} \frac{1 - u^2/c^2}{(1 - u^2/c^2 + u_r^2/c^2)^{3/2}} \mathbf{r}_0, \quad (12.299)$$

where \mathbf{r}_0 is the radius vector directed from the present position of the charge at time t to the field point, and $u_r = \mathbf{u} \cdot \mathbf{r}_0/r_0$. For the case of uniform motion, the relationship between the retarded radius vector \mathbf{r} and the actual radius vector \mathbf{r}_0 is simply

$$\mathbf{r}_0 = \mathbf{r} - \frac{r}{c} \mathbf{u}. \quad (12.300)$$

It is straightforward to demonstrate that

$$s = r_0 \sqrt{1 - u^2/c^2 + u_r^2/c^2} \quad (12.301)$$

in this case. Thus, the electric field generated by a uniformly moving charge can be written

$$\mathbf{E} = \frac{e}{4\pi \epsilon_0} \frac{1 - u^2/c^2}{s^3} \mathbf{r}_0. \quad (12.302)$$

Because $\mathbf{r}_u = \mathbf{r}_0$ for the case of a uniformly moving charge, it is clear that Equation (12.298) is equivalent to the electric field generated by a uniformly moving charge located at the position the charge would occupy if it had continued with uniform velocity from its retarded position.

The second term in Equation (12.293),

$$\mathbf{E}_{\text{radiation}} = \frac{e}{4\pi \epsilon_0 c^2} \frac{\mathbf{r} \times (\mathbf{r}_u \times \dot{\mathbf{u}})}{s^3}, \quad (12.303)$$

is of order $1/r$, and, therefore, represents a radiation field. Similar considerations hold for the two terms of Equation (12.296).

12.24 Larmor Formula

Let us transform to the inertial frame in which the charge is instantaneously at rest at the origin at time $t = 0$. In this frame, $u \ll c$, so that $\mathbf{r}_u \simeq \mathbf{r}$ and $s \simeq r$ for events that are sufficiently close to the origin at $t = 0$ that the retarded charge still appears to travel with a velocity that is small compared to that of light. It follows from the previous section that

$$\mathbf{E}_{\text{radiation}} \simeq \frac{e}{4\pi \epsilon_0 c^2} \frac{\mathbf{r} \times (\mathbf{r} \times \dot{\mathbf{u}})}{r^3}, \quad (12.304)$$

$$\mathbf{B}_{\text{radiation}} \simeq \frac{e}{4\pi \epsilon_0 c^3} \frac{\dot{\mathbf{u}} \times \mathbf{r}}{r^2}. \quad (12.305)$$

Let us define spherical polar coordinates whose axis points along the direction of instantaneous acceleration of the charge. It is easily demonstrated that

$$E_\theta \simeq \frac{e}{4\pi \epsilon_0 c^2} \frac{\sin \theta}{r} \dot{u}, \quad (12.306)$$

$$B_\phi \simeq \frac{e}{4\pi \epsilon_0 c^3} \frac{\sin \theta}{r} \dot{u}. \quad (12.307)$$

These fields are identical to those of a radiating dipole whose axis is aligned along the direction of instantaneous acceleration. The radial Poynting flux is given by

$$\frac{E_\theta B_\phi}{\mu_0} = \frac{e^2}{16\pi^2 \epsilon_0 c^3} \frac{\sin^2 \theta}{r^2} \dot{u}^2. \quad (12.308)$$

We can integrate this expression to obtain the instantaneous power radiated by the charge

$$P = \frac{e^2}{6\pi \epsilon_0 c^3} \dot{u}^2. \quad (12.309)$$

This is known as *Larmor's formula*. Note that zero net momentum is carried off by the fields (12.306) and (12.307).

In order to proceed further, it is necessary to prove two useful theorems. The first theorem states that if a 4-vector field T^μ satisfies

$$\partial_\mu T^\mu = 0, \quad (12.310)$$

and if the components of T^μ are non-zero only in a finite spatial region, then the integral over 3-space,

$$I = \int T^4 d^3x, \quad (12.311)$$

is an invariant. In order to prove this theorem, we need to use the 4-dimensional analog of Gauss's theorem, which states that

$$\int_V \partial_\mu T^\mu d^4x = \oint_S T^\mu dS_\mu, \quad (12.312)$$

where dS_μ is an element of the 3-dimensional surface S bounding the 4-dimensional volume V . The particular volume over which the integration is performed is indicated in Figure 12.1. The surfaces A and C are chosen so that the spatial components of T^μ vanish on A and C . This is always possible because it is assumed that the region over which the components of T^μ are non-zero is of finite extent. The surface B is chosen normal to the x^4 -axis, whereas the surface D is chosen normal to the $x^{4'}$ -axis. Here, the x^μ and the $x^{\mu'}$ are coordinates in two arbitrarily chosen inertial frames. It follows from Equation (12.312) that

$$\int T^4 dS_4 + \int T^{4'} dS_{4'} = 0. \quad (12.313)$$

Here, we have made use of the fact that $T^\mu dS_\mu$ is a scalar and, therefore, has the same value in all inertial frames. Because $dS_4 = -d^3x$ and $dS_{4'} = d^3x'$ it follows that $I = \int T^4 d^3x$ is an invariant under a Lorentz transformation. Incidentally, taking the limit in which the two inertial frames are identical, the previous argument also demonstrates that I is constant in time.

The second theorem is an extension of the first. Suppose that a 4-tensor field $Q^{\mu\nu}$ satisfies

$$\partial_\mu Q^{\mu\nu} = 0, \quad (12.314)$$

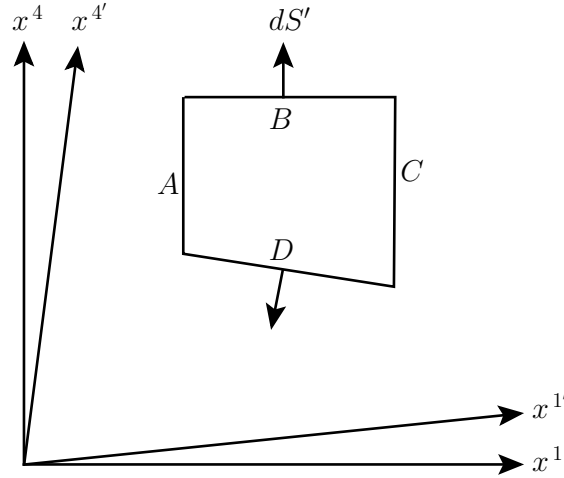


Figure 12.1: Application of Gauss' theorem.

and has components which are only non-zero in a finite spatial region. Let A_μ be a 4-vector whose coefficients do not vary with position in space-time. It follows that $T^\mu = A_\nu Q^{\mu\nu}$ satisfies Equation (12.310). Therefore,

$$I = \int A_\nu Q^{4\nu} d^3x \quad (12.315)$$

is an invariant. However, we can write

$$I = A_\mu B^\mu, \quad (12.316)$$

where

$$B^\mu = \int Q^{4\mu} d^3x. \quad (12.317)$$

It follows from the quotient rule that if $A_\mu B^\mu$ is an invariant for arbitrary A_μ then B^μ must transform as a constant (in time) 4-vector.

These two theorems enable us to convert differential conservation laws into integral conservation laws. For instance, in differential form, the conservation of electrical charge is written

$$\partial_\mu J^\mu = 0. \quad (12.318)$$

However, from Equation (12.313) this immediately implies that

$$Q = \frac{1}{c} \int J^4 d^3x = \int \rho d^3x \quad (12.319)$$

is an invariant. In other words, the total electrical charge contained in space is both constant in time, and the same in all inertial frames.

Suppose that S is the instantaneous rest frame of the charge. Let us consider the electromagnetic energy tensor $T^{\mu\nu}$ associated with all of the radiation emitted by the charge between times $t = 0$ and $t = dt$. According to Equation (12.261), this tensor field satisfies

$$\partial_\mu T^{\mu\nu} = 0, \quad (12.320)$$

apart from a region of space of measure zero in the vicinity of the charge. Furthermore, the region of space over which $T^{\mu\nu}$ is non-zero is clearly finite, because we are only considering the fields emitted by the charge in a small time interval, and these fields propagate at a finite velocity. Thus, according to the second theorem,

$$P^\mu = \frac{1}{c} \int T^{4\mu} d^3x \quad (12.321)$$

is a 4-vector. It follows from Section 12.22 that we can write $P^\mu = (d\mathbf{p}, dE/c)$, where $d\mathbf{p}$ and dE are the total momentum and energy carried off by the radiation emitted between times $t = 0$ and $t = dt$, respectively. As we have already mentioned, $d\mathbf{p} = 0$ in the instantaneous rest frame S . Transforming to an arbitrary inertial frame S' , in which the instantaneous velocity of the charge is u , we obtain

$$dE' = \gamma(u) (dE + u dp^1) = \gamma dE. \quad (12.322)$$

However, the time interval over which the radiation is emitted in S' is $dt' = \gamma dt$. Thus, the instantaneous power radiated by the charge,

$$P' = \frac{dE'}{dt'} = \frac{dE}{dt} = P, \quad (12.323)$$

is the same in all inertial frames.

We can make use of the fact that the power radiated by an accelerating charge is Lorentz invariant to find a relativistic generalization of the Larmor formula, (12.309), which is valid in all inertial frames. We expect the power emitted by the charge to depend only on its 4-velocity and 4-acceleration. It follows that the Larmor formula can be written in Lorentz invariant form as

$$P = -\frac{e^2}{6\pi\epsilon_0 c^3} A_\mu A^\mu, \quad (12.324)$$

because the 4-acceleration takes the form $A^\mu = (\dot{\mathbf{u}}, 0)$ in the instantaneous rest frame. In a general inertial frame,

$$-A_\mu A^\mu = \gamma^2 \left(\frac{d\gamma}{dt} \mathbf{u} + \gamma \dot{\mathbf{u}} \right)^2 - \gamma^2 c^2 \left(\frac{d\gamma}{dt} \right)^2, \quad (12.325)$$

where use has been made of Equation (12.107). Furthermore, it is easily demonstrated that

$$\frac{d\gamma}{dt} = \gamma^3 \frac{\mathbf{u} \cdot \dot{\mathbf{u}}}{c^2}. \quad (12.326)$$

It follows, after a little algebra, that the relativistic generalization of Larmor's formula takes the form

$$P = \frac{e^2}{6\pi\epsilon_0 c^3} \gamma^6 \left[\dot{\mathbf{u}}^2 - \frac{(\mathbf{u} \times \dot{\mathbf{u}})^2}{c^2} \right]. \quad (12.327)$$

12.25 Radiation Losses

Radiation losses often determine the maximum achievable energy in a charged particle accelerator. Let us investigate radiation losses in various different types of accelerator device using the relativistic Larmor formula.

For a linear accelerator, the motion is one dimensional. In this case, it is easily demonstrated that

$$\frac{dp}{dt} = m_0 \gamma^3 \dot{u}, \quad (12.328)$$

where use has been made of Equation (12.326), and $p = \gamma m_0 u$ is the particle momentum in the direction of acceleration (the x -direction, say). Here, m_0 is the particle rest mass. Thus, Equation (12.327) yields

$$P = \frac{e^2}{6\pi \epsilon_0 m_0^2 c^3} \left(\frac{dp}{dt} \right)^2. \quad (12.329)$$

The rate of change of momentum is equal to the force exerted on the particle in the x -direction, which, in turn, equals the change in the energy, E , of the particle per unit distance. Consequently,

$$P = \frac{e^2}{6\pi \epsilon_0 m_0^2 c^3} \left(\frac{dE}{dx} \right)^2. \quad (12.330)$$

Thus, in a linear accelerator the radiated power depends on the external force acting on the particle, and not on the actual energy or momentum of the particle. It is obvious, from the previous formula, that light particles, such as electrons, are going to radiate a lot more than heavier particles, such as protons. The ratio of the power radiated to the power supplied by the external sources is

$$\frac{P}{dE/dt} = \frac{e^2}{6\pi \epsilon_0 m_0^2 c^3} \frac{1}{u} \frac{dE}{dx} \simeq \frac{e^2}{6\pi \epsilon_0 m_0 c^2} \frac{1}{m_0 c^2} \frac{dE}{dx}, \quad (12.331)$$

because $u \simeq c$ for a highly relativistic particle. It is clear, from the previous expression, that the radiation losses in an electron linear accelerator are negligible unless the gain in energy is of order $m_e c^2 = 0.511$ MeV in a distance of $e^2/(6\pi \epsilon_0 m_e c^2) = 1.28 \times 10^{-15}$ meters. That is 3×10^{14} MeV/meter. Typical energy gains are less than 10 MeV/meter. It follows, therefore, that radiation losses are completely negligible in linear accelerators, whether for electrons, or for other heavier particles.

The situation is quite different in circular accelerator devices, such as the synchrotron and the betatron. In such machines, the momentum \mathbf{p} changes rapidly in direction as the particle rotates, but the change in energy per revolution is small. Furthermore, the direction of acceleration is always perpendicular to the direction of motion. It follows from Equation (12.327) that

$$P = \frac{e^2}{6\pi \epsilon_0 c^3} \gamma^4 \dot{u}^2 = \frac{e^2}{6\pi \epsilon_0 c^3} \frac{\gamma^4 u^4}{\rho^2}, \quad (12.332)$$

where ρ is the orbit radius. Here, use has been made of the standard result $\dot{u} = u^2/\rho$ for circular motion. The radiative energy loss per revolution is given by

$$\delta E = \frac{2\pi \rho}{u} P = \frac{e^2}{3 \epsilon_0 c^3} \frac{\gamma^4 u^3}{\rho}. \quad (12.333)$$

For highly relativistic ($u \simeq c$) electrons, this expression yields

$$\delta E(\text{MeV}) = 8.85 \times 10^{-2} \frac{[E(\text{GeV})]^4}{\rho(\text{meters})}. \quad (12.334)$$

In the first electron synchrotrons, $\rho \sim 1$ meter, $E_{\text{max}} \sim 0.3$ GeV. Hence, $\delta E_{\text{max}} \sim 1$ keV per revolution. This was less than, but not negligible compared to, the energy gain of a few keV per turn. For modern electron synchrotrons, the limitation on the available radio-frequency power needed to overcome radiation losses becomes a major consideration, as is clear from the E^4 dependence of the radiated power per turn.

12.26 Angular Distribution of Radiation

In order to calculate the angular distribution of the energy radiated by an accelerated charge, we must think carefully about what is meant by the rate of radiation of the charge. This quantity is actually the amount of energy lost by the charge in a retarded time interval dt' during the emission of the signal. Thus,

$$P(t') = -\frac{dE}{dt'}, \quad (12.335)$$

where E is the energy of the charge. The Poynting vector

$$\frac{\mathbf{E}_{\text{rad}} \times \mathbf{B}_{\text{rad}}}{\mu_0} = \epsilon_0 c E_{\text{rad}}^2 \frac{\mathbf{r}}{r}, \quad (12.336)$$

where use has been made of $\mathbf{B}_{\text{rad}} = (\mathbf{r} \times \mathbf{E}_{\text{rad}})/rc$ [see Equation (12.297)], represents the energy flux per unit actual time, t . Thus, the energy loss rate of the charge into a given element of solid angle $d\Omega$ is

$$\frac{dP(t')}{d\Omega} d\Omega = -\frac{dE(\theta, \varphi)}{dt'} d\Omega = \frac{dE(\theta, \varphi)}{dt} \frac{dt}{dt'} r^2 d\Omega = \epsilon_0 c E_{\text{rad}}^2 \frac{r}{r} r^2 d\Omega, \quad (12.337)$$

where use has been made of Equation (12.283). Here, θ and φ are angular coordinates used to locate the element of solid angle. It follows from Equation (12.303) that

$$\frac{dP(t')}{d\Omega} = \frac{e^2 r}{16\pi^2 \epsilon_0 c^3} \frac{[\mathbf{r} \times (\mathbf{r}_u \times \dot{\mathbf{u}})]^2}{s^5}. \quad (12.338)$$

Consider the special case in which the direction of acceleration coincides with the direction of motion. Let us define spherical polar coordinates whose axis points along this common direction. It is easily demonstrated that, in this case, the previous expression reduces to

$$\frac{dP(t')}{d\Omega} = \frac{e^2 \dot{u}^2}{16\pi^2 \epsilon_0 c^3} \frac{\sin^2 \theta}{[1 - (u/c) \cos \theta]^5}. \quad (12.339)$$

In the non-relativistic limit, $u/c \rightarrow 0$, the radiation pattern has the familiar $\sin^2 \theta$ dependence of dipole radiation. In particular, the pattern is symmetric in the forward ($\theta < \pi/2$) and backward ($\theta >$

$\pi/2$) directions. However, as $u/c \rightarrow 1$, the radiation pattern becomes more and more concentrated in the forward direction. The angle θ_{\max} for which the intensity is a maximum is

$$\theta_{\max} = \cos^{-1} \left[\frac{1}{3u/c} \left(\sqrt{1 + 15u^2/c^2} - 1 \right) \right]. \quad (12.340)$$

This expression yields $\theta_{\max} \rightarrow \pi/2$ as $u/c \rightarrow 0$, and $\theta_{\max} \rightarrow 1/(2\gamma)$ as $u/c \rightarrow 1$. Thus, for a highly relativistic charge, the radiation is emitted in a narrow cone whose axis is aligned along the direction of motion. In this case, the angular distribution (12.339) reduces to

$$\frac{dP(t')}{d\Omega} \simeq \frac{2e^2 \dot{u}^2}{\pi^2 \epsilon_0 c^3} \gamma^8 \frac{(\gamma\theta)^2}{[1 + (\gamma\theta)^2]^5}. \quad (12.341)$$

The total power radiated by the charge is obtained by integrating Equation (12.339) over all solid angles. We obtain

$$P(t') = \frac{e^2 \dot{u}^2}{8\pi \epsilon_0 c^3} \int_0^\pi \frac{\sin^3 \theta d\theta}{[1 - (u/c) \cos \theta]^5} = \frac{e^2 \dot{u}^2}{8\pi \epsilon_0 c^3} \int_{-1}^{+1} \frac{(1 - \mu^2) d\mu}{[1 - (u/c)\mu]^5}. \quad (12.342)$$

It is easily verified that

$$\int_{-1}^{+1} \frac{(1 - \mu^2) d\mu}{[1 - (u/c)\mu]^5} = \frac{4}{3} \gamma^6. \quad (12.343)$$

Hence,

$$P(t') = \frac{e^2}{6\pi \epsilon_0 c^3} \gamma^6 \dot{u}^2, \quad (12.344)$$

which agrees with Equation (12.327), provided that $\mathbf{u} \times \dot{\mathbf{u}} = 0$.

12.27 Synchrotron Radiation

Synchrotron radiation (i.e., radiation emitted by a charged particle constrained to follow a circular orbit by a magnetic field) is of particular importance in astrophysics, because much of the observed radio frequency emission from supernova remnants and active galactic nuclei is thought to be of this type.

Consider a charged particle moving in a circle of radius a with constant angular velocity ω_0 . Suppose that the orbit lies in the x - y plane. The radius vector pointing from the centre of the orbit to the retarded position of the charge is defined

$$\boldsymbol{\rho} = a (\cos \phi, \sin \phi, 0), \quad (12.345)$$

where $\phi = \omega_0 t'$ is the angle subtended between this vector and the x -axis. The retarded velocity and acceleration of the charge take the form

$$\mathbf{u} = \frac{d\boldsymbol{\rho}}{dt'} = u (-\sin \phi, \cos \phi, 0), \quad (12.346)$$

$$\dot{\mathbf{u}} = \frac{d\mathbf{u}}{dt'} = -\dot{u} (\cos \phi, \sin \phi, 0), \quad (12.347)$$

where $u = a\omega_0$ and $\dot{u} = a\omega_0^2$. The observation point is chosen such that the radius vector \mathbf{r} , pointing from the retarded position of the charge to the observation point, is parallel to the y - z plane. Thus, we can write

$$\mathbf{r} = r(0, \sin\alpha, \cos\alpha), \quad (12.348)$$

where α is the angle subtended between this vector and the z -axis. As usual, we define θ as the angle subtended between the retarded radius vector \mathbf{r} and the retarded direction of motion of the charge \mathbf{u} . It follows that

$$\cos\theta = \frac{\mathbf{u} \cdot \mathbf{r}}{ur} = \sin\alpha \cos\phi. \quad (12.349)$$

It is easily seen that

$$\dot{\mathbf{u}} \cdot \mathbf{r} = -\dot{u}r \sin\alpha \sin\phi. \quad (12.350)$$

A little vector algebra shows that

$$[\mathbf{r} \times (\mathbf{r}_u \times \dot{\mathbf{u}})]^2 = -(\mathbf{r} \cdot \dot{\mathbf{u}})^2 r^2 (1 - u^2/c^2) + \dot{u}^2 r^4 (1 - \mathbf{r} \cdot \mathbf{u}/rc)^2, \quad (12.351)$$

giving

$$[\mathbf{r} \times (\mathbf{r}_u \times \dot{\mathbf{u}})]^2 = \dot{u}^2 r^4 \left[\left(1 - \frac{u}{c} \cos\theta\right)^2 - \left(1 - \frac{u^2}{c^2}\right) \tan^2\phi \cos^2\theta \right]. \quad (12.352)$$

Making use of Equation (12.337), we obtain

$$\frac{dP(t')}{d\Omega} = \frac{e^2 \dot{u}^2}{16\pi^2 \epsilon_0 c^3} \frac{[1 - (u/c) \cos\theta]^2 - (1 - u^2/c^2) \tan^2\phi \cos^2\theta}{[1 - (u/c) \cos\theta]^5}. \quad (12.353)$$

It is convenient to write this result in terms of the angles α and ϕ , instead of θ and ϕ . After a little algebra we obtain

$$\frac{dP(t')}{d\Omega} = \frac{e^2 \dot{u}^2}{16\pi^2 \epsilon_0 c^3} \frac{[1 - (u^2/c^2)] \cos^2\alpha + [(u/c) - \sin\alpha \cos\phi]^2}{[1 - (u/c) \sin\alpha \cos\phi]^5}. \quad (12.354)$$

Let us consider the radiation pattern emitted in the plane of the orbit: that is, $\alpha = \pi/2$, with $\cos\phi = \cos\theta$. It is easily seen that

$$\frac{dP(t')}{d\Omega} = \frac{e^2 \dot{u}^2}{16\pi^2 \epsilon_0 c^3} \frac{[(u/c) - \cos\theta]^2}{[1 - (u/c) \cos\theta]^5}. \quad (12.355)$$

In the non-relativistic limit, the radiation pattern has a $\cos^2\theta$ dependence. Thus, the pattern is like that of dipole radiation where the axis is aligned along the instantaneous direction of acceleration. As the charge becomes more relativistic, the radiation lobe in the forward direction (i.e., $0 < \theta < \pi/2$) becomes more more focused and more intense. Likewise, the radiation lobe in the backward direction (i.e., $\pi/2 < \theta < \pi$) becomes more diffuse. The radiation pattern has zero intensity at the angles

$$\theta_0 = \cos^{-1}(u/c). \quad (12.356)$$

These angles demark the boundaries between the two radiation lobes. In the non-relativistic limit, $\theta_0 = \pm\pi/2$, so the two lobes are of equal angular extents. In the highly relativistic limit, $\theta_0 \rightarrow \pm 1/\gamma$,

so the forward lobe becomes highly concentrated about the forward direction ($\theta = 0$). In the latter limit, Equation (12.355) reduces to

$$\frac{dP(t')}{d\Omega} \simeq \frac{e^2 \dot{u}^2}{2\pi^2 \epsilon_0 c^3} \gamma^6 \frac{[1 - (\gamma\theta)^2]^2}{[1 + (\gamma\theta)^2]^5}. \quad (12.357)$$

Thus, the radiation emitted by a highly relativistic charge is focused into an intense beam, of angular extent $1/\gamma$, pointing in the instantaneous direction of motion. The maximum intensity of the beam scales like γ^6 .

Integration of Equation (12.354) over all solid angle (making use of $d\Omega = \sin\alpha \, d\alpha \, d\phi$) yields

$$P(t') = \frac{e^2}{6\pi \epsilon_0 c^3} \gamma^4 \dot{u}^2, \quad (12.358)$$

which agrees with Equation (12.327), provided that $\mathbf{u} \cdot \dot{\mathbf{u}} = 0$. This expression can also be written

$$\frac{P}{m_0 c^2} = \frac{2}{3} \frac{\omega_0^2 r_0}{c} \beta^2 \gamma^4, \quad (12.359)$$

where $r_0 = e^2/(4\pi \epsilon_0 m_0 c^2) = 2.82 \times 10^{-15}$ meters is the *classical electron radius*, m_0 is the rest mass of the charge, and $\beta = u/c$. If the circular motion takes place in an orbit of radius a , perpendicular to a magnetic field \mathbf{B} , then ω_0 satisfies $\omega_0 = eB/(m_0 \gamma)$. Thus, the radiated power is

$$\frac{P}{m_0 c^2} = \frac{2}{3} \left(\frac{eB}{m_0} \right)^2 \frac{r_0}{c} (\beta \gamma)^2, \quad (12.360)$$

and the radiated energy ΔE per revolution is

$$\frac{\Delta E}{m_0 c^2} = \frac{4\pi r_0}{3 a} \beta^3 \gamma^4. \quad (12.361)$$

Let us consider the frequency distribution of the emitted radiation in the highly relativistic limit. Suppose, for the sake of simplicity, that the observation point lies in the plane of the orbit (i.e., $\alpha = \pi/2$). Because the radiation emitted by the charge is beamed very strongly in the charge's instantaneous direction of motion, a fixed observer will only see radiation (at some later time) when this direction points almost directly towards the point of observation. This occurs once every rotation period, when $\phi \simeq 0$, assuming that $\omega_0 > 0$. Note that the point of observation is located many orbit radii away from the centre of the orbit along the positive y -axis. Thus, our observer sees short periodic pulses of radiation from the charge. The repetition frequency of the pulses (in radians per second) is ω_0 . Let us calculate the duration of each pulse. Because the radiation emitted by the charge is focused into a narrow beam of angular extent $\Delta\theta \sim 1/\gamma$, our observer only sees radiation from the charge when $\phi \lesssim \Delta\theta$. Thus, the observed pulse is emitted during a time interval $\Delta t' = \Delta\theta/\omega_0$. However, the pulse is received in a somewhat shorter time interval

$$\Delta t = \frac{\Delta\theta}{\omega_0} \left(1 - \frac{u}{c} \right), \quad (12.362)$$

because the charge is slightly closer to the point of observation at the end of the pulse than at the beginning. The previous equation reduces to

$$\Delta t \simeq \frac{\Delta\theta}{2\omega_0\gamma^2} \sim \frac{1}{\omega_0\gamma^3}, \quad (12.363)$$

because $\gamma \gg 1$ and $\Delta\theta \sim 1/\gamma$. The width $\Delta\omega$ of the pulse in frequency space obeys $\Delta\omega\Delta t \sim 1$. Hence,

$$\Delta\omega = \gamma^3 \omega_0. \quad (12.364)$$

In other words, the emitted frequency spectrum contains harmonics up to γ^3 times that of the cyclotron frequency, ω_0 .

12.28 Exercises

- 12.1 Consider two Cartesian reference frames, S and S' , in the standard configuration. Suppose that S' moves with constant velocity $v < c$ with respect to S along their common x -axis. Demonstrate that the Lorentz transformation between coordinates in the two frames can be written

$$\begin{aligned} x' &= x \cosh \varphi - ct \sinh \varphi, \\ y' &= y, \\ z' &= z, \\ ct' &= ct \cosh \varphi - x \sinh \varphi, \end{aligned}$$

where $\tanh \varphi = v/c$. Show that the previous transformation is equivalent to a rotation through an angle $i\varphi$, in the $x-ict$ plane, in (x, y, z, ict) space.

- 12.2 Show that, in the standard configuration, two successive Lorentz transformations with velocities v_1 and v_2 are equivalent to a single Lorentz transformation with velocity

$$v = \frac{v_1 + v_2}{1 + v_1 v_2/c^2}.$$

- 12.3 Let \mathbf{r} and \mathbf{r}' be the displacement vectors of some particle in the Cartesian reference frames S and S' , respectively. Suppose that frame S' moves with velocity \mathbf{v} with respect to frame S . Demonstrate that a general Lorentz transformation takes the form

$$\begin{aligned} \mathbf{r}' &= \mathbf{r} + \left[\frac{(\gamma - 1) \mathbf{r} \cdot \mathbf{v}}{v^2} - \gamma t \right] \mathbf{v}, \\ t' &= \gamma \left(t - \frac{\mathbf{r} \cdot \mathbf{v}}{c^2} \right), \end{aligned} \quad (12.365)$$

where $\gamma = (1 - v^2/c^2)^{-1/2}$. If $\mathbf{u} = d\mathbf{r}/dt$ and $\mathbf{u}' = d\mathbf{r}'/dt'$ are the particle's velocities in the two reference frames, respectively, demonstrate that a general velocity transformation is written

$$\mathbf{u}' = \frac{\mathbf{u} + [(\gamma - 1)\mathbf{u} \cdot \mathbf{v}/c^2 - \gamma]\mathbf{v}}{\gamma(1 - \mathbf{u} \cdot \mathbf{v}/c^2)}.$$

- 12.4 Let v be the Earth's approximately constant orbital speed. Demonstrate that the direction of starlight incident at right-angles to the Earth's instantaneous direction of motion appears slightly shifted in the Earth's instantaneous rest frame by an angle $\theta = \sin^{-1}(v/c)$. This effect is known as the *abberation* of starlight. Estimate the magnitude of θ (in arc seconds).
- 12.5 Let \mathbf{E} and \mathbf{B} be the electric and magnetic field, respectively, in some Cartesian reference frame S . Likewise, let \mathbf{E}' and \mathbf{B}' be the electric and magnetic field, respectively, in some other Cartesian frame S' , which moves with velocity \mathbf{v} with respect to S . Demonstrate that the general transformation of fields takes the form

$$\begin{aligned}\mathbf{E}' &= \gamma \mathbf{E} + \frac{1 - \gamma}{v^2} (\mathbf{v} \cdot \mathbf{E}) \mathbf{v} + \gamma (\mathbf{v} \times \mathbf{B}), \\ \mathbf{B}' &= \gamma \mathbf{B} + \frac{1 - \gamma}{v^2} (\mathbf{v} \cdot \mathbf{B}) \mathbf{v} - \frac{\gamma}{c^2} (\mathbf{v} \times \mathbf{E}),\end{aligned}$$

where $\gamma = (1 - v^2/c^2)^{-1/2}$.

- 12.6 A particle of rest mass m and charge e moves relativistically in a uniform magnetic field of strength B . Show that the particle's trajectory is a helix aligned along the direction of the field, and that the particle drifts parallel to the field at a uniform velocity, and gyrates in the plane perpendicular to the field with constant angular velocity

$$\Omega = \frac{eB}{\gamma m}.$$

Here, $\gamma = (1 - v^2/c^2)^{-1/2}$, and v is the particle's (constant) speed.

- 12.7 Let $P = \mathbf{E} \cdot \mathbf{B}$ and $Q = c^2 B^2 - E^2$. Prove the following statements, assuming that E and B are not both zero.
- At any given event, \mathbf{E} is perpendicular to \mathbf{B} either in all frames of reference, or in none. Moreover, each of the three relations $E > cB$, $E = cB$, and $E < cB$ holds in all frames or in none.
 - If $P = Q = 0$ then the field is said to be null. For a null field, \mathbf{E} is perpendicular to \mathbf{B} , and $E = cB$, in all frames.
 - If $P = 0$ and $Q \neq 0$ then there are infinitely many frames (with a common relative direction of motion) in which $E = 0$ or $B = 0$, according as $Q > 0$ or $Q < 0$, and none other. Precisely one of these frames moves in the direction $\mathbf{E} \times \mathbf{B}$, its velocity being E/B or $c^2 B/E$, respectively.

- (d) If $P \neq 0$ then there are infinitely many frames (with a common direction of motion) in which \mathbf{E} is parallel to \mathbf{B} , and none other. Precisely one of these moves in the direction $\mathbf{E} \times \mathbf{B}$, its velocity being given by the smaller root of the quadratic equation $\beta^2 - R\beta + 1 = 0$, where $\beta = v/c$, and $R = (E^2 + c^2 B^2)/|\mathbf{E} \times c\mathbf{B}|$. In order for β to be real we require $R > 2$. Demonstrate that this is always the case.

12.8 In the rest frame of a conducting medium, the current density satisfies Ohm's law $\mathbf{j}' = \sigma \mathbf{E}'$, where σ is the conductivity, and primes denote quantities in the rest frame.

- (a) Taking into account the possibility of convection currents, as well as conduction currents, show that the covariant generalization of Ohm's law is

$$J^\mu - \frac{1}{c^2}(U_\nu J^\nu) U^\mu = \frac{\sigma}{c} F^{\mu\nu} U_\nu,$$

where U^μ is the 4-velocity of the medium, J^μ the 4-current, and $F^{\mu\nu}$ the electromagnetic field tensor.

- (b) Show that if the medium has a velocity $\mathbf{v} = c\boldsymbol{\beta}$ with respect to some inertial frame then the 3-vector current in that frame is

$$\mathbf{j} = \gamma \sigma [\mathbf{E} + \boldsymbol{\beta} \times c\mathbf{B} - (\boldsymbol{\beta} \cdot \mathbf{E})\boldsymbol{\beta}] + \rho \mathbf{v}$$

where ρ is the charge density observed in the inertial frame.

12.9 Consider the relativistically covariant form of Maxwell's equations in the presence of magnetic monopoles. Demonstrate that it is possible to define a proper-4-current

$$J^\mu = (\mathbf{j}, \rho c),$$

and a pseudo-4-current

$$J_m = (\mathbf{j}_m, \rho_m c),$$

where \mathbf{j} and ρ are the flux and density of electric charges, respectively, whereas \mathbf{j}_m and ρ_m are the flux and density of magnetic monopoles, respectively. Show that the conservation laws for electric charges and magnetic monopoles take the form

$$\partial_\mu J^\mu = 0,$$

$$\partial_\mu J_m^\mu = 0,$$

respectively. Finally, if $F^{\mu\nu}$ is the electromagnetic field tensor, and $G^{\mu\nu}$ its dual, show that Maxwell's equations are equivalent to

$$\partial_\mu F^{\mu\nu} = \frac{J^\nu}{\epsilon_0 c},$$

$$\partial_\mu G^{\mu\nu} = \frac{J_m^\nu}{\epsilon_0 c}.$$

12.10 Prove that the electromagnetic energy tensor satisfies the following two identities:

$$T^\mu{}_\mu = 0,$$

and

$$T^\mu{}_\sigma T^\sigma{}_\nu = \frac{I^2}{4} \delta^\mu{}_\nu,$$

where

$$I^2 = \left(\frac{B^2}{\mu_0} - \epsilon_0 E^2 \right)^2 + \frac{4\epsilon_0}{\mu_0} (\mathbf{E} \cdot \mathbf{B})^2.$$

12.11 A charge e moves in simple harmonic motion along the z axis, such that its retarded position is $z(t') = a \cos(\omega_0 t')$.

(a) Show that the instantaneous power radiated per unit solid angle is

$$\frac{dP(t')}{d\Omega} = \frac{e^2 c \beta^4}{16\pi^2 \epsilon_0 a^2} \frac{\sin^2 \theta \cos^2(\omega_0 t')}{[1 + \beta \cos \theta \sin(\omega_0 t')]^5}$$

where $\beta = a \omega_0 / c$, and θ is a standard spherical polar coordinate.

(b) By time averaging, show that the average power radiated per unit solid angle is

$$\frac{dP}{d\Omega} = \frac{e^2 c \beta^4}{128\pi^2 \epsilon_0 a^2} \left[\frac{4 + \beta^2 \cos^2 \theta}{(1 - \beta^2 \cos^2 \theta)^{7/2}} \right] \sin^2 \theta.$$

(c) Sketch the angular distribution of the radiation for non-relativistic and ultra-relativistic motion.

12.12 The trajectory of a relativistic particle of charge e and rest mass m in a uniform magnetic field \mathbf{B} is a helix aligned with the field. Let the pitch angle of the helix be α (so, $\alpha = 0$ corresponds to circular motion). By arguments similar to those used for synchrotron radiation, show that an observer far from the charge would detect radiation with a fundamental frequency

$$\omega_0 = \frac{\Omega}{\cos^2 \alpha},$$

where $\Omega = e B / (\gamma m)$, and that the spectrum would extend up to frequencies of order

$$\omega_c = \gamma^3 \Omega \cos \alpha.$$