

Introduction to Electrostatics

Charles Augustin de Coulomb

(1736 - 1806)

December 23, 2000

Contents

1	Coulomb's Law	2
2	Electric Field	4
3	Gauss's Law	7
4	Differential Form of Gauss's Law	9
5	An Equation for $\nabla \times \mathbf{E}$; the Scalar Potential	10
5.1	Conservative Potentials	11
6	Poisson's and Laplace's Equations	13
7	Energy in the Electric Field; Capacitance; Forces	15
7.1	Conductors	20
7.2	Forces on Charged Conductors	22

8	Green's Theorem	27
8.1	Green's Theorem	29
8.2	Applying Green's Theorem 1	30
8.3	Applying Green's Theorem 2	31
8.3.1	Greens Theorem with Dirichlet B.C.	34
8.3.2	Greens Theorem with Neumann B.C.	36

We shall follow the approach of Jackson, which is more or less historical. Thus we start with classical electrostatics, pass on to magnetostatics, add time dependence, and wind up with Maxwell's equations. These are then expressed within the framework of special relativity. The remainder of the course is devoted to a broad range of interesting and important applications.

This development may be contrasted with the more formal and elegant approach which starts from the Maxwell equations plus special relativity and then proceeds to work out electrostatics and magnetostatics - as well as everything else - as special cases. This is the method of *e.g.*, Landau and Lifshitz, *The Classical Theory of Fields*.

The first third of the course, *i.e.*, Physics 707, deals with physics which should be familiar to everyone; what will perhaps not be familiar are the mathematical techniques and functions that will be introduced in order to solve certain kinds of problems. These are of considerable usefulness and therefore will be important to us.

1 Coulomb's Law

By performing experiments on small charged bodies (ideally, point charges), Charles Augustin de Coulomb, working around the time of the American and French revolutions (1785), was able to empirically infer that the force between two static charged particles is proportional

to the inverse square of the distance between them. The following has since become known¹ have as *Coulomb's Law*: Given two static charges q_1 and q_2 , there is a force acting on each of them which is:

1. Proportional to the product of the magnitudes of the charges, being attractive for unlike charges and repulsive for like charges
2. Inversely proportional to the square of the distance between the charges.
3. Directed along the line between the charges.

In the form of an equation, the law states that

$$\mathbf{F}_{21} = k \frac{q_1 q_2}{|\mathbf{x}_2 - \mathbf{x}_1|^2} \frac{\mathbf{x}_2 - \mathbf{x}_1}{|\mathbf{x}_2 - \mathbf{x}_1|} \quad (1)$$

where charge q_i is located at \mathbf{x}_i , \mathbf{F}_{21} is the force on charge 2 produced by charge 1, and k is a positive constant; vectors are denoted by boldface type.

In addition, the force satisfies a superposition law (or principle) in that the force \mathbf{F} on a charge q in the presence of a number of other charges q_i at \mathbf{x}_i , $i = 1, \dots, n$, is simply the sum of the forces arising from each of the latter as though it were the only other charge present².

Thus,

$$\mathbf{F} = kq \sum_{i=1}^n \frac{q_i (\mathbf{x} - \mathbf{x}_i)}{|\mathbf{x} - \mathbf{x}_i|^3}, \quad (2)$$

¹Numerous others, such as Henry Cavendish, also may legitimately have some claim to the law.

²As we shall see, the principle of superposition follows from the linearity of Maxwell's Equations

given that q is at \mathbf{x} .

The constant k has units and magnitude which depend on the system of units employed. We shall adopt cgs Gaussian units. The units of mass length and time are, respectively grams (g), centimeters (cm), and seconds (sec). The unit of charge is the statcoulomb (*statcoul*) which is defined by the statement that the force between two charges, each of one *statcoul*, one cm apart is one dyne (*dyn*). Then $k = 1 \text{ dyn} - \text{cm}^2 / (\text{statcoul})^2$. In practice one may treat k as having dimension unity while charge has dimension of $M^{1/2}L^{3/2}/T$.

2 Electric Field

It is customary and useful to introduce the concept of the electric field at this point. This is a vector field, *i.e.*, a vector function of \mathbf{x} . It is written as $\mathbf{E}(\mathbf{x})$ and is defined as the force that would be experienced by a charge q at \mathbf{x} , divided by q^3 . Thus, for a distribution of charges q_i at \mathbf{x}_i , $i=1,2,\dots,n$,

$$\mathbf{E}(\mathbf{x}) = \sum_{i=1}^n \frac{q_i(\mathbf{x} - \mathbf{x}_i)}{|\mathbf{x} - \mathbf{x}_i|^3} \quad (3)$$

The electric field has the property of being independent of the ‘test’ charge q ; it is a function of the charge distribution which gives rise to the force on the test charge, and, of course, of the test charge’s position. This object has dimension Q/L^2 or $M^{1/2}/L^{1/2}T$.

³This definition is not complete. The field has other attributes as well since it carries momentum and energy: *i.e.* photons

At this point let us introduce the charge density $\rho(\mathbf{x})$ which is the charge per unit volume at, or very close to, \mathbf{x} . This object is needed if we would like to integrate over a source distribution instead of summing over its constituent charges. Thus a sum is replaced by an equivalent integral,

$$\sum_{i=1}^n q_i \rightarrow \int d^3x \rho(\mathbf{x}). \quad (4)$$

The charge density has dimension Q/L^3 . In terms of ρ , the expression for the electric field can be written as

$$\mathbf{E}(\mathbf{x}) = \int d^3x' \rho(\mathbf{x}') \frac{\mathbf{x} - \mathbf{x}'}{|\mathbf{x} - \mathbf{x}'|^3} \quad (5)$$

In the particular case of a distribution of discrete point charges, it is possible to recover the sum in Eq. (3) by writing the charge density in an appropriate way. To do so we introduce the Dirac delta function $\delta(x - a)$. It is defined by the integral

$$f(a) = \int dx f(x) \delta(x - a) \quad (6)$$

where $f(x)$ is an arbitrary continuous function of x , and the range of integration includes the point $x = a$. A special case is $f(x) = 1$ which leads to

$$\int dx \delta(x - a) = 1, \quad (7)$$

demonstrating the normalization of the delta function. From the arbitrariness of $f(x)$, we may conclude that $\delta(x - a)$ is zero when x is not a and sufficiently singular at $x = a$ to give the normalization property.

In other words, it is in essence the charge density of a point charge (in one dimension) located at $x = a$.

Some important relations involving delta functions are as follows:

$$\int_{a_1}^{a_2} f(x) \frac{d\delta(x-a)}{dx} dx = - \left. \frac{df(x)}{dx} \right|_{x=a} \quad (8)$$

and

$$\int_{a_1}^{a_2} \delta[f(x)] dx = \sum_{i=1}^N \left[1 / \left| \frac{df(x)}{dx} \right|_{x_i} \right] \quad (9)$$

In the final expression the x_i are the 0's of $f(x)$ between a_1 and a_2 .

A delta function in three dimensions may be built as a product of three one-dimensional delta functions. In Cartesian coordinates,

$$\delta(\mathbf{x}) = \delta(x)\delta(y)\delta(z) \quad (10)$$

This function has the property that

$$\int d^3x f(\mathbf{x})\delta(\mathbf{x} - \mathbf{x}_0) = f(\mathbf{x}_0) \quad (11)$$

Returning to electrostatics, we can see that the charge density of a collection of point charges can be written as a sum of delta functions:

$$\rho(\mathbf{x}) = \sum_{i=1}^n q_i \delta(\mathbf{x} - \mathbf{x}_i) \quad (12)$$

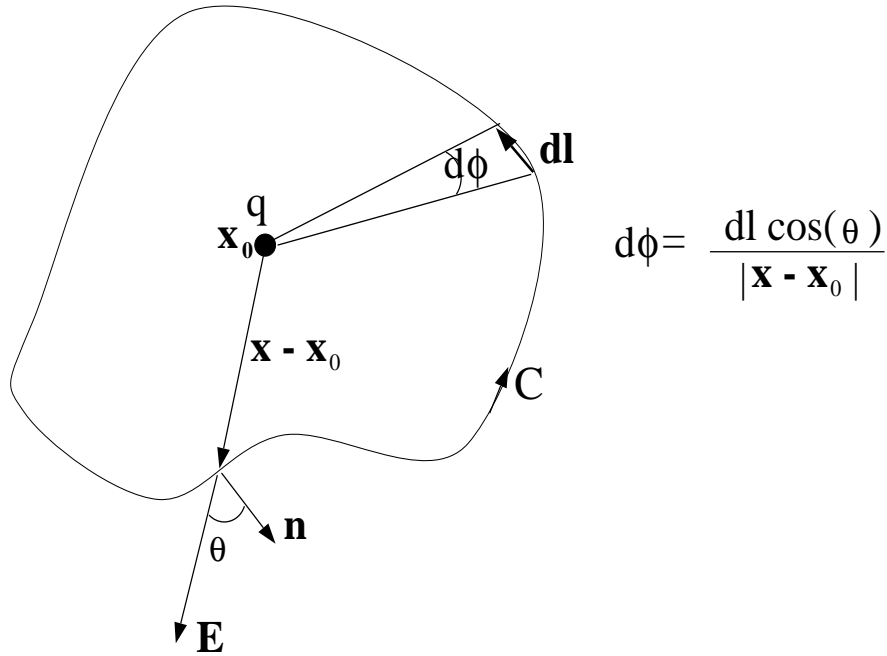
Thus

$$\begin{aligned} \mathbf{E}(\mathbf{x}) &= \int d^3x' \rho(\mathbf{x}') (\mathbf{x} - \mathbf{x}') / |\mathbf{x} - \mathbf{x}'|^3 \\ &= \sum_{i=1}^n \int d^3x' q_i \delta(\mathbf{x}' - \mathbf{x}_i) (\mathbf{x} - \mathbf{x}') / |\mathbf{x} - \mathbf{x}'|^3 \\ &= \sum_{i=1}^n q_i (\mathbf{x} - \mathbf{x}_i) / |\mathbf{x} - \mathbf{x}_i|^3. \end{aligned} \quad (13)$$

3 Gauss's Law

Although Coulomb's Law is quite sufficient for finding electric fields and forces, the integral form in which we have expressed it is not always the most useful approach to a problem. Another integral form, called Gauss's Law, is often more useful.

Let us look first at a **two-dimensional** version of this law. Consider a point charge q located within a closed path C . In two dimensions, the field produced by this charge is $2q(\mathbf{x} - \mathbf{x}_0)/|\mathbf{x} - \mathbf{x}_0|^2$. Consider now the integral around C of that component of \mathbf{E} which is normal to the path. This normal component is $\mathbf{E} \cdot \mathbf{n} = q\cos\theta/r$, where r is the distance from the charge to the integration point on the loop. However, $dl \cos\theta/r$ is just the infinitesimal angle $d\phi$ subtended by dl at the charge. Hence we just need to integrate $d\phi$ around the loop.



$$\oint_C dl \mathbf{E} \cdot \mathbf{n} = \oint_C dl \frac{q \cos \theta}{|\mathbf{x} - \mathbf{x}_0|} = q \int d\phi \quad (14)$$

Since the charge is inside, the integral is 2π ; if it were outside, the integral would be 0 because over one part of the path, $\cos \theta$ is positive and over another part it is negative with the two parts cancelling one another when the integration is completed. Thus one finds that

$$\int_C dl [\mathbf{n} \cdot \mathbf{E}(\mathbf{x})] = \begin{cases} 2\pi q, & q \text{ inside of } C \\ 0, & q \text{ outside of } C \end{cases} \quad (15)$$

The **three-dimensional** case works out much the same way. The field varies as $1/r^2$ and so one finds that $d^2x \cos \theta / r^2$ is the solid angle element $d\Omega$ subtended by the infinitesimal area element d^2x of S at the position of the charge. Integration over the surface thus reduces to

integration over the solid angle subtended by the surface at the charge, and this is 4π if the charge is inside of the surface and 0 otherwise,

$$\int_S d^2x [\mathbf{E}(\mathbf{x}) \cdot \mathbf{n}] = \begin{cases} 4\pi q, & q \text{ inside of } S \\ 0, & q \text{ outside of } S \end{cases} \quad (16)$$

Next, the superposition principle allows us to add up the fields arising from an arbitrary collection of charges, with Gauss's Law holding for each bit of charge. As a consequence, we may say that

$$\int_S d^2x [\mathbf{E}(\mathbf{x}) \cdot \mathbf{n}] = 4\pi Q \quad (17)$$

where Q is the total charge contained inside of the surface,

$$Q = \int_V d^3x \rho(\mathbf{x}). \quad (18)$$

4 Differential Form of Gauss's Law

A differential form of this law may be found by applying the divergence theorem which states that, for a general vector field $C(\mathbf{x})$,

$$\int_S d^2x [C(\mathbf{x}) \cdot \mathbf{n}] = \int_V d^3x [\nabla \cdot C(\mathbf{x})]. \quad (19)$$

Let us apply this equation to Gauss's Law:

$$4\pi \int_V d^3x \rho(\mathbf{x}) = \int_S d^2x [\mathbf{E}(\mathbf{x}) \cdot \mathbf{n}] = \int_V d^3x [\nabla \cdot \mathbf{E}(\mathbf{x})] \quad (20)$$

or

$$\int_V d^3x [\nabla \cdot \mathbf{E}(\mathbf{x}) - 4\pi\rho(\mathbf{x})] = 0 \quad (21)$$

Because V is completely arbitrary, we may equate the integrand to zero and find

$$\nabla \cdot \mathbf{E}(\mathbf{x}) = 4\pi\rho(\mathbf{x}) \quad (22)$$

which is the differential form of Gauss's Law.

In the process of obtaining this equation from Coulomb's Law, we have lost some of the information contained in it. Merely specifying the divergence of a vector field is not sufficient to determine the field. Hence we need an additional equation to supplement Gauss's Law.

5 An Equation for $\nabla \times \mathbf{E}$; the Scalar Potential

Let us start once again from Coulomb's Law:

$$\mathbf{E}(\mathbf{x}) = \int d^3x' \rho(\mathbf{x}') \frac{\mathbf{x} - \mathbf{x}'}{|\mathbf{x} - \mathbf{x}'|^3} \quad (23)$$

But one may write part of the integrand as a gradient,

$$\frac{\mathbf{x} - \mathbf{x}'}{|\mathbf{x} - \mathbf{x}'|^3} = -\nabla \left(\frac{1}{|\mathbf{x} - \mathbf{x}'|} \right), \quad (24)$$

where the gradient is taken with respect to the variable \mathbf{x} . Hence

$$\mathbf{E}(\mathbf{x}) = -\nabla \int d^3x' \frac{\rho(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|}, \quad (25)$$

which is to say, \mathbf{E} can be written as the (negative) gradient of a scalar function of \mathbf{x} . This function we shall call the scalar potential and denote by $\Phi(\mathbf{x})$:

$$\Phi(\mathbf{x}) \equiv \int d^3x' \frac{\rho(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|}; \quad (26)$$

$$\mathbf{E}(\mathbf{x}) = -\nabla\Phi(\mathbf{x}). \quad (27)$$

From this statement it follows immediately that $\nabla \times \mathbf{E}(\mathbf{x}) = 0$ because the curl of the gradient of a scalar function is always zero.

To summarize,

$$\nabla \cdot \mathbf{E}(\mathbf{x}) = 4\pi\rho(\mathbf{x}) \quad (28)$$

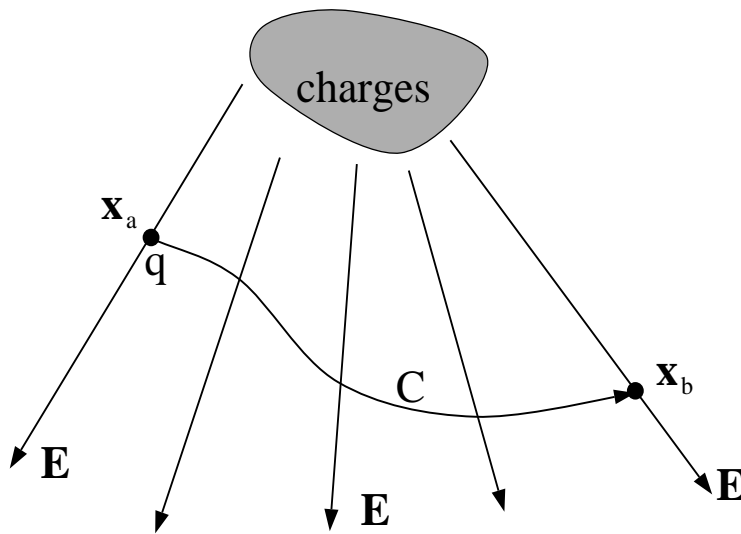
and

$$\nabla \times \mathbf{E}(\mathbf{x}) = 0. \quad (29)$$

5.1 Conservative Potentials

From our derivation of the curl equation, we can see that this simple result follows from the fact that the force (or electric field) is central and depends only on the distance between charges. Such a force is also called conservative, and the potential function is related in a simple way to the energy of a charge in an electric field.

To find this relation, consider that a set of fixed source charges produce a field \mathbf{E} and that a charge q is placed at point \mathbf{x}_a . Here it experiences an electric field force $\mathbf{F} = q\mathbf{E}(\mathbf{x}_a)$ and so an equal and opposite force $\mathbf{F}_{ext} = -\mathbf{F} = -q\mathbf{E}(\mathbf{x}_a)$ must be applied by some external agent to keep it in position.



If we now move the charge along a path C from \mathbf{x}_a to \mathbf{x}_b , the work done by the external agent is found by integrating the force along the path,

$$W_{a \rightarrow b} = - \int_C d\mathbf{l} \cdot \mathbf{F}(\mathbf{x}), \quad (30)$$

or

$$W_{a \rightarrow b} = q \int_C d\mathbf{l} \cdot \nabla \Phi(\mathbf{x}) = q \int_C d\Phi = q[\Phi(\mathbf{x}_b) - \Phi(\mathbf{x}_a)] \quad (31)$$

This result shows that $q\Phi(\mathbf{x})$ can be interpreted as the potential energy of charge q in the electrostatic field at point \mathbf{x} , aside from a constant defining the zero of potential energy. In going from \mathbf{x}_a to \mathbf{x}_b , work $q[\Phi(\mathbf{x}_b) - \Phi(\mathbf{x}_a)]$ is done on the charge, and so the change in the energy of the system (composed of the charge q and the sources of the field) is just this work.

Notice especially that the work does not depend on the path C except through the endpoints. This statement can always be made

of conservative systems. In particular, the integral of the work done around a closed path is 0,

$$\int_C d\mathbf{l} \cdot \mathbf{E}(\mathbf{x}) = 0 \quad (32)$$

It is instructive to apply Stokes' Theorem to this relation. His theorem states that, for an arbitrary vector field \mathbf{A} , and a closed path C with a surface S "linking" the path (which means that S is an open surface with edges coinciding with C),

$$\int_C d\mathbf{l} \cdot \mathbf{A} = \int_S d^2x [\nabla \times \mathbf{A}(\mathbf{x})] \cdot \mathbf{n} \quad (33)$$

where \mathbf{n} is a unit normal to the surface in the right-hand sense relative to the direction in which the path is traversed. As applied to the electric field, we have

$$0 = \int_C d\mathbf{l} \cdot \mathbf{E}(\mathbf{x}) = \int_S d^2x [\nabla \times \mathbf{E}(\mathbf{x})] \cdot \mathbf{n}. \quad (34)$$

Because C is arbitrary and can in particular be any infinitesimal closed loop, this relation implies that the integrand is zero, $\nabla \times \mathbf{E}(\mathbf{x}) = 0$. Thus the statement that \mathbf{E} is a conservative field and $\nabla \times \mathbf{E}(\mathbf{x}) = 0$ are equivalent.

6 Poisson's and Laplace's Equations

The differential equations we have determined for \mathbf{E} are sufficient to find it uniquely, given appropriate boundary conditions and the charge

density, but they do not necessarily provide the simplest approach to the solution of an electrostatics problem. Often, it is best to solve for Φ from which \mathbf{E} follows easily. Since $\nabla \cdot \mathbf{E} = 4\pi\rho$, and $\mathbf{E} = -\nabla\Phi$, we have

$$\nabla \cdot \nabla\Phi(\mathbf{x}) \equiv \nabla^2\Phi(\mathbf{x}) = -4\pi\rho(\mathbf{x}) \text{ Poisson's Equation,} \quad (35)$$

which is *Poisson's Equation*; the operator ∇^2 is the *Laplacian operator*. In those regions of space where the charge density vanishes, we find the simpler equation,

$$\nabla^2\Phi(\mathbf{x}) = 0 \text{ Laplace's Equation,} \quad (36)$$

which is *Laplace's Equation*.

Consider the effect of operating with ∇^2 on the integral expression for Φ :

$$-4\pi\rho(\mathbf{x}) = \nabla^2\Phi(\mathbf{x}) = \int d^3x' \rho(\mathbf{x}') \nabla^2 \left(\frac{1}{|\mathbf{x} - \mathbf{x}'|} \right), \quad (37)$$

or

$$\rho(\mathbf{x}) = - \int d^3x' \rho(\mathbf{x}') \left(\frac{1}{4\pi} \nabla^2 \frac{1}{|\mathbf{x} - \mathbf{x}'|} \right). \quad (38)$$

However, we defined $\delta(\mathbf{x} - \mathbf{x}')$ as

$$f(\mathbf{x}) = \int d^3x' f(\mathbf{x}') \delta(\mathbf{x} - \mathbf{x}') \quad (39)$$

for general $f(\mathbf{x})$. Since $\rho(\mathbf{x})$ can be quite general, the quantity in large parentheses above satisfies the condition placed on the delta function; hence we conclude that

$$\nabla^2 \left(\frac{1}{|\mathbf{x} - \mathbf{x}'|} \right) = -4\pi\delta(\mathbf{x} - \mathbf{x}') \quad (40)$$

which is only appropriate because $1/|\mathbf{x} - \mathbf{x}'|$ is the potential of a unit point charge, and $\delta(\mathbf{x} - \mathbf{x}')$ is the corresponding charge density. Thus, Eq. (40) expresses Poisson's equation for a unit point charge located at \mathbf{x}' .

As an exercise we may derive this result in a different way. Consider

$$\nabla^2 \left(\frac{1}{|\mathbf{x}|} \right) = \nabla^2 \left(\frac{1}{r} \right) = \frac{1}{r} \frac{d^2}{dr^2} \left(r \frac{1}{r} \right) = 0, \quad (41)$$

except possibly at $r = 0$ where r/r is undefined. To determine what happens here, we integrate $\nabla^2(1/r)$ over a small sphere centered on the origin:

$$\begin{aligned} \int_V d^3x \nabla^2(1/r) &= \int_V d^3x \nabla \cdot \nabla(1/r) = \\ \int_S d^2x [\nabla(1/r)] \cdot \mathbf{n} &= - \int_S d^2x (1/r^2) = - \int r^2 dr \sin \theta d\theta d\phi (1/r^2) = -4\pi \end{aligned} \quad (42)$$

Thus we have shown the following:

$$\begin{aligned} (i) \quad \nabla^2(1/r) &= 0, \quad r \neq 0 \\ (ii) \quad \int_V d^3x \nabla^2(1/r) &= -1/4\pi, \quad r = 0 \in V. \end{aligned} \quad (43)$$

These results tell us that $\nabla^2(1/r) = -4\pi\delta(\mathbf{x})$.

7 Energy in the Electric Field; Capacitance; Forces

The energy of the static electric field, or of a static charge distribution, is of some importance. Let us start our investigation by constructing the energy of interaction of n point charges q_i located at \mathbf{x}_i . As we have

seen, the work required to move a charge q from one point to another in an applied electric field is q times the difference in the electric potentials at the end points. If we suppose that this potential is produced by our collection of point charges, then it is given by

$$\Phi(\mathbf{x}) = \sum_{i=1}^n [q_i/|\mathbf{x} - \mathbf{x}_i|] \quad (44)$$

and the work done to bring q from infinitely far away, where $\Phi(\mathbf{x}) = 0$, to point \mathbf{x} is

$$W = q \sum_{i=1}^n [q_i/|\mathbf{x} - \mathbf{x}_i|]. \quad (45)$$

This is therefore the increase in the total energy of the system of charges when a charge is added to it at some particular point.

We may use this result to calculate the energy of the collection of charges by bringing them in one at a time from points at infinity where they are assumed to be widely separated. The first charge is brought in to \mathbf{x}_1 , and this costs no energy because $\Phi = 0$ when there are no other charges present. The second charge costs energy

$$W_2 = \frac{q_1 q_2}{|\mathbf{x}_1 - \mathbf{x}_2|}. \quad (46)$$

The third then costs

$$W_3 = q_3 \sum_{j=1}^2 \frac{q_j}{|\mathbf{x}_3 - \mathbf{x}_j|}, \quad (47)$$

and so on. The amount of work which must be done to bring in the i^{th} particle is

$$W_i = q_i \sum_{j=1}^{i-1} \frac{q_j}{|\mathbf{x}_i - \mathbf{x}_j|}. \quad (48)$$

If we add up these energies to find the total work done, it is

$$W = \sum_{i=2}^n \sum_{j=1}^{i-1} \frac{q_i q_j}{|\mathbf{x}_i - \mathbf{x}_j|} \quad (49)$$

which is the sum over all pairs, each pair taken once; it may also be written as

$$W = \frac{1}{2} \sum'_{i,j} \frac{q_i q_j}{|\mathbf{x}_i - \mathbf{x}_j|}; \quad (50)$$

the prime on the summation sign means that the terms with $i = j$ are omitted. In this sum, we include each pair i, j with $i \neq j$ twice and so have to multiply by a factor of $1/2$. Given a continuous charge distribution, the same argument can be applied using as the elementary charges infinitesimal charge elements located in infinitesimal volume elements. The result must be

$$W = \frac{1}{2} \int d^3x d^3x' \frac{\rho(\mathbf{x})\rho(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|} \quad (51)$$

where the integrations are unrestricted and include the points $\mathbf{x} = \mathbf{x}'$ because the interaction energy of an infinitesimal continuously distributed charge element with itself vanishes in the limit of zero extent. However, if the charge distribution contains finite point charges, represented by delta functions in $\rho(\mathbf{x})$, then one has to omit the interaction of each of these charges with itself, as in the original sum, Eq. (50), in order to obtain a finite result.

The expression for W can be cast into a number of other useful forms. Recall that

$$\Phi(\mathbf{x}) = \int d^3x' \frac{\rho(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|}; \quad (52)$$

substitution into the expression for W gives

$$W = \frac{1}{2} \int d^3x \rho(\mathbf{x}) \Phi(\mathbf{x}). \quad (53)$$

Further, $\rho(\mathbf{x}) = -\nabla^2 \Phi(\mathbf{x})/4\pi$, so

$$W = -\frac{1}{8\pi} \int d^3x \Phi(\mathbf{x}) \nabla^2 \Phi(\mathbf{x}). \quad (54)$$

Let us now do an integration by parts in three dimensions. This operation is easy to achieve by making use of the divergence theorem; for a vector field $f(\mathbf{x})\mathbf{A}(\mathbf{x})$ consider the integral

$$\begin{aligned} \int_V d^3x \nabla \cdot [f(\mathbf{x})\mathbf{A}(\mathbf{x})] &= \int_S d^2x f(\mathbf{x})[\mathbf{A}(\mathbf{x}) \cdot \mathbf{n}] \\ &= \int_V d^3x [\nabla f(\mathbf{x}) \cdot \mathbf{A}(\mathbf{x})] + \int_V d^3x f(\mathbf{x})[\nabla \cdot \mathbf{A}(\mathbf{x})] \end{aligned} \quad (55)$$

or

$$\int_V d^3x f(\mathbf{x})[\nabla \cdot \mathbf{A}(\mathbf{x})] = -\int_V d^3x [\nabla f(\mathbf{x})] \cdot \mathbf{A}(\mathbf{x}) + \int_S d^2x f(\mathbf{x})[\mathbf{A}(\mathbf{x}) \cdot \mathbf{n}], \quad (56)$$

where V and S are related in the usual way. As applied to the integral for W , this useful formula gives, letting V be all space,

$$W = -\frac{1}{8\pi} \int d^3x \Phi(\mathbf{x}) \nabla \cdot \nabla \Phi(\mathbf{x}) \quad (57)$$

$$= \frac{1}{8\pi} \int d^3x \nabla \Phi(\mathbf{x}) \cdot \nabla \Phi(\mathbf{x}) - \int d^2x \Phi(\mathbf{x})[\nabla \Phi(\mathbf{x}) \cdot \mathbf{n}] \quad (58)$$

or

$$W = \frac{1}{8\pi} \int d^3x [\mathbf{E}(\mathbf{x}) \cdot \mathbf{E}(\mathbf{x})]. \quad (59)$$

The surface integral has vanished because Φ falls off at distances r which are large compared to the extent of the charge distribution at least as fast as $1/r$. Hence the integrand in the surface integral falls off at least as fast as $1/r^3$ while the area of the surface at distance r varies as r^2 . This integral therefore falls off at least as fast as $1/r$ and so vanishes when the surface is at infinity.

An interesting and plausible interpretation of the final expression is that the integrand is the energy density $u(\mathbf{x})$ of the electric field,

$$u(\mathbf{x}) = \frac{1}{8\pi} \mathbf{E}(\mathbf{x}) \cdot \mathbf{E}(\mathbf{x}). \quad (60)$$

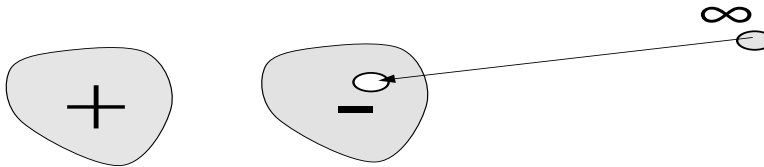
This is only an interpretation, however. All we really know is that the total energy is the integral of this quantity over all space. The idea is plausible because $u(\mathbf{x})$ so defined is everywhere positive or zero (a negative energy density would be disturbing). Note, too, that our other expression for the energy as an integral over a single position variable has an integrand that can be both positive and negative which makes it unreasonable to interpret that integrand, $\rho(\mathbf{x})\Phi(\mathbf{x})/2$, as the energy density.

Given Eq. (59) for W , we can see that the energy will be positive definite. Yet the energy of, *e.g.*, a pair of point charges q and $-q$ at \mathbf{x} and \mathbf{x}' is negative, $-q^2/|\mathbf{x} - \mathbf{x}'|$. The reason is that the expression we have for the energy of a set of point charges does not include the (infinite) energy required to assemble each of the point charges in the first place, but Eq. (59) would include this (positive) energy. A more

concrete example involves two oppositely charged masses. The energy required to bring them together from infinity is negative,



whereas the energy required to assemble the entire charge distribution



at its final location is positive.

7.1 Conductors

Consider now the special case that our electrostatic system consists of a collection of n electrically isolated conductors; for our present purposes, a conductor may be defined as an object which cannot support an electric field (because it contains “free” charges which move under the influence of a field until there is no field). Thus the interior of a conductor is an equipotential. Using Eq. (53), we see that for such a system,

$$W = \frac{1}{2} \int d^3x \rho(\mathbf{x})\Phi(\mathbf{x}) = \frac{1}{2} \sum_{i=1}^n Q_i V_i \quad (61)$$

where Q_i and V_i are, respectively, the charge and potential on the i^{th} conductor. Now, because the potential is a linear function of charge

(superposition theorem), it is true that

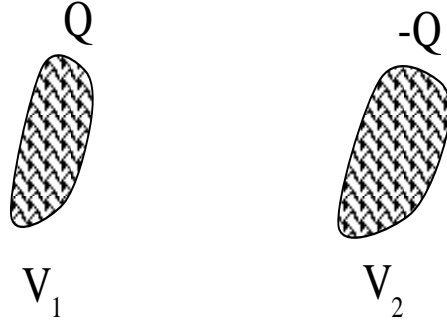
$$V_i = \sum_{j=1}^n p_{ij} Q_j. \quad (62)$$

The coefficients p_{ij} are independent of the charges; they depend only on the distribution and shapes of the conductors and are called the *coefficients of potential*. To see that this relation is valid, one need only think of the potentials produced on each conductor given charge Q_j on the j^{th} conductor and zero charge on all others; then superpose the solutions to each of the problems of this kind. Inversion of Eq. (62) yields the charges Q_i as linear combinations of the potentials V_j ,

$$Q_i = \sum_{j=1}^n C_{ij} V_j. \quad (63)$$

The coefficients C_{ij} are called *coefficients of capacitance*; the diagonal elements C_{ii} are more commonly referred to simply as *capacitances* while the off-diagonal ones C_{ij} are called *coefficients of electrostatic induction* and are not to be confused with the inductances introduced in connection with Faraday's Law. The capacitance of a single conductor, C_{ii} , is thus the total charge on that conductor when it is maintained at unit potential while all other conductors are held at zero potential.

As an example the capacitance of a pair of conductors with equal and opposite charge is defined as the ratio of the charge on one conductor to the potential difference between them when all other conductors are maintained at zero potential.



$$\begin{pmatrix} Q \\ -Q \end{pmatrix} = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix} \begin{pmatrix} V_1 \\ V_2 \end{pmatrix} \quad (64)$$

$$\begin{pmatrix} V_1 \\ V_2 \end{pmatrix} = \frac{\begin{pmatrix} C_{22} & -C_{21} \\ -C_{12} & C_{11} \end{pmatrix} \begin{pmatrix} Q \\ -Q \end{pmatrix}}{C_{11}C_{22} - C_{12}C_{21}} \quad (65)$$

The capacitance $C(1, 2) = Q/|V_1 - V_2|$ turns out to be

$$C(1, 2) = (C_{11}C_{22} - C_{12}^2)/(C_{11} + C_{22} + 2C_{12}). \quad (66)$$

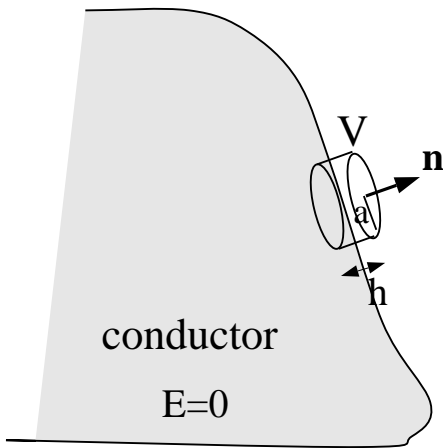
The energy of the system of conductors may be written in terms of potentials and the coefficients C_{ij} as

$$W = \frac{1}{2} \sum_{i=1}^n Q_i V_i = \frac{1}{2} \sum_{i,j=1}^n C_{ij} V_i V_j. \quad (67)$$

7.2 Forces on Charged Conductors

Another useful application of the expressions for the energy is in the calculation of forces on charged conductors. Consider the surface of a

conductor. The field at the surface can be inferred from $\nabla \cdot \mathbf{E}(\mathbf{x}) = 4\pi\rho(\mathbf{x})$ and $\nabla \times \mathbf{E}(\mathbf{x}) = 0$. Consider an integral of the first of these equations over a “pillbox” or short right circular cylinder oriented with the “drumhead” parallel to the surface of a conductor and situated half inside and half outside of the conductor.



$$\int_V d^3x \nabla \cdot \mathbf{E} = \int_S d^2x \mathbf{E} \cdot \mathbf{n} = \mathbf{E}_n(\mathbf{x})\pi a^2 = 4\pi \int_V d^3x \rho(\mathbf{x}) = \pi a^2 4\pi\sigma(\mathbf{x}) \quad (68)$$

Using the divergence theorem, we may convert to a surface integral. Given that the height h of the cylinder is much smaller than its radius, $h \ll a$, the only important contribution to the surface integral must come from the drumheads. But $\mathbf{E}(\mathbf{x}) = 0$ on the one inside of the conductor, so we pick up only the contribution from the component of \mathbf{E} normal to the surface of the conductor on the outside. Given that a is much smaller than distances over which the field varies appreciably, we get simply $\pi a^2 \mathbf{E}_n(\mathbf{x})$ where \mathbf{x} is a point just outside of the surface

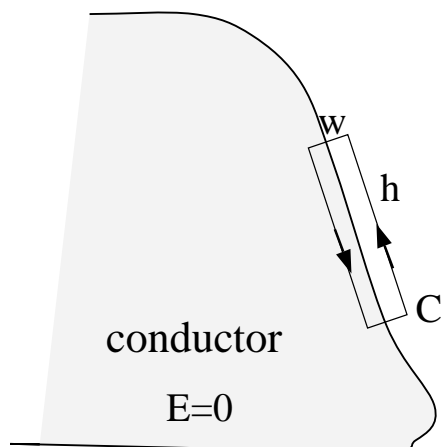
of the conductor and the subscript n designates the outward (from the conductor) normal component of the field. The volume integral of $\rho(\mathbf{x})$, on the other hand, yields the total charge within the pillbox. Because $h \ll a$, we will get a vanishingly small contribution from any finite volume density of charge. However, we get an important contribution from a *surface* charge density; we shall denote such a thing by $\sigma(\mathbf{x})$. It can be written as a volume charge density by using a δ function:

$$\rho(\mathbf{x}) = \sigma(\mathbf{x})\delta(\xi(\mathbf{x})) \quad (69)$$

where ξ is the normal distance of the point \mathbf{x} from the surface of the conductor. When this surface charge density is integrated over the volume of the pillbox, it reduces to a surface integral of the surface charge density over a disc on the conductor surface and having the area of the crosssection of the pillbox, *i.e.*, πa^2 . Hence one finds $\pi a^2(4\pi\sigma(\mathbf{x}))$. Putting the two sides of the equation together, we find the following relation between the charge density on the surface of a conductor and the normal component of the electric field just outside of the conductor:

$$\mathbf{E}_n(\mathbf{x}) = 4\pi\sigma(\mathbf{x}) \quad (70)$$

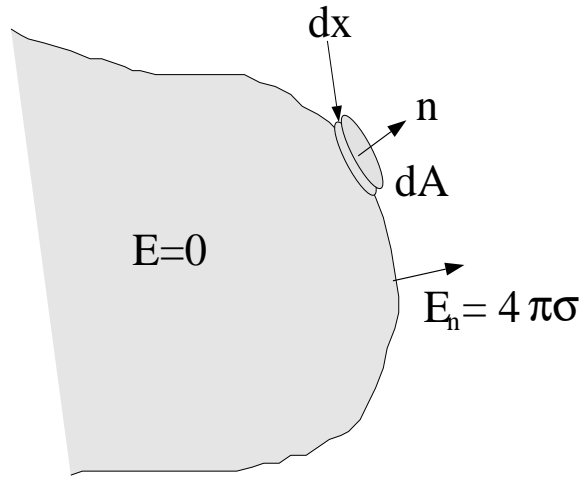
We may also find an equation for the **tangential component** of the electric field at the surface of a conductor. Consider the line integral of $d\mathbf{l} \cdot \mathbf{E}(\mathbf{x})$ around a rectangle which straddles the conductor's surface.



$$\int_S d^2x (\nabla \times \mathbf{E}) \cdot \mathbf{n} = \oint_C d\mathbf{l} \cdot \mathbf{E} = \mathbf{E}_t h = 0 \implies \mathbf{E}_t = 0 \quad (71)$$

The width w of the rectangle, which is its size in the direction normal to the interface, is much smaller than its height h , which is its size parallel to the interface. The dominant contribution to the line integral, which is zero, comes from the two sides parallel to the interface. On the inside, $\mathbf{E}(\mathbf{x})$ is zero, so we have only the integral along the side which is exterior to the conductor. Since the whole integral is zero, the integral along this single side must be zero, and hence we can conclude that the tangential component of $\mathbf{E}(\mathbf{x})$, or $\mathbf{E}_t(\mathbf{x})$, just outside of a conductor must vanish,

$$\mathbf{E}_t(\mathbf{x}) = 0 \quad \text{at the surface of a conductor.} \quad (72)$$



Now we are in a position to consider **the force on the surface of a conductor**. We use the method of virtual work. Imagine moving a small element dA of the conductor's surface, along with the charge on it, a distance dx from its initial position in the direction normal to the surface. It will sweep out a volume $dAdx$. In this volume, to a first approximation ⁴, there will no longer be any electric field (since the field is zero within a conductor) while the electric field elsewhere will be unchanged. Hence there will be a change in the field energy of

$$dW = -dAdx \frac{1}{8\pi} E_n^2 = -2\pi\sigma^2 dAdx + \mathcal{O}(dAd^2x, d^2Adx) \quad (73)$$

where Eq. (70) has been used for the normal component of the electric field. Energy conservation demands that the amount of work done on the system in making this displacement is equal to dW . It is also dx times the negative of the electric force acting on the area element dA .

⁴This approximation improves as the ratio of dx to \sqrt{A} goes to zero

Thus,

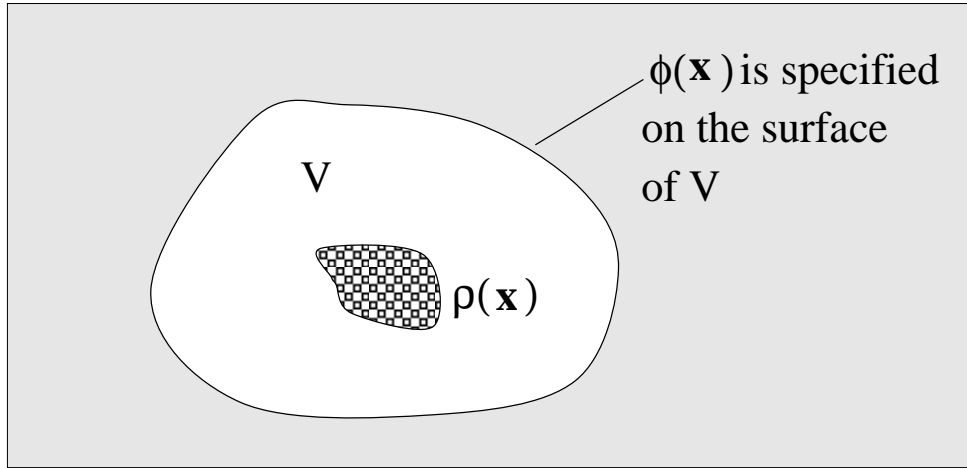
$$-Fdx = dW \text{ or } F = 2\pi\sigma^2 dA. \quad (74)$$

Hence the force per unit area on the surface of the conductor is $2\pi\sigma^2$; it is directed normally outward from the conductor (“negative pressure”).

A second way of looking at this problem is to calculate the force directly. It must be the charge on dA , or σdA , times the electric field which acts on this charge, *i.e.*, that part of the electric field at the surface which is produced by charges other than those on dA . This field is just $2\pi\sigma$ (Why?), so the force comes out as before.

8 Green’s Theorem

In everything we have discussed thus far, we have assumed that $\rho(\mathbf{x})$ is known, and that there is a simple boundary condition on $\phi(\mathbf{x})$ at infinity (that it must at least as fast as $1/r$). *This is not generally true!* Usually, we only know $\rho(\mathbf{x})$ within some finite volume V , and the value of $\phi(\mathbf{x})$ on the corresponding surface S .



For these problems Green's theorem and functions are useful. The simplest greens function is that for free space

$$G_{free}(\mathbf{x}, \mathbf{x}') = \frac{1}{|\mathbf{x} - \mathbf{x}'|}, \quad (75)$$

and the corresponding potential is

$$\phi(\mathbf{x}) = \int_V d^3x' G_{free}(\mathbf{x}, \mathbf{x}') \rho(\mathbf{x}'). \quad (76)$$

Note that $G_{free}(\mathbf{x}, \mathbf{x}')$ gives the response at \mathbf{x} , due to a unit point charge at \mathbf{x}' . For the response due to a collection of charges $\rho(\mathbf{x})$, superposition yields the integral above.

For the more general problem of a collection of charges and boundary conditions on a surface, we might expect that this integral relation will become

$$\phi(\mathbf{x}) = \int_V d^3x' G(\mathbf{x}, \mathbf{x}') \rho(\mathbf{x}') + \text{surface term} \quad (77)$$

Before we can proceed further, we must develop some formalism.

8.1 Green's Theorem

To this end, let us develop Green's Theorem, starting from the divergence theorem as applied to a vector field \mathbf{A} :

$$\int_V d^3x [\nabla \cdot \mathbf{A}(\mathbf{x})] = \int_S d^2x [\mathbf{A}(\mathbf{x}) \cdot \mathbf{n}] \quad (78)$$

where S is the surface bounding the domain V . Consider the special case $\mathbf{A}(\mathbf{x}) = \phi(\mathbf{x})\nabla\psi(\mathbf{x})$; then

$$\nabla \cdot \mathbf{A}(\mathbf{x}) = \nabla\phi(\mathbf{x}) \cdot \nabla\psi(\mathbf{x}) + \phi(\mathbf{x})\nabla^2\psi(\mathbf{x}) \quad (79)$$

and

$$\mathbf{n} \cdot \mathbf{A}(\mathbf{x}) = \phi(\mathbf{x})[\mathbf{n} \cdot \nabla\psi(\mathbf{x})] = \phi(\mathbf{x})\frac{\partial\psi(\mathbf{x})}{\partial n} \quad (80)$$

where $\frac{\partial\psi}{\partial n}$ is the outward normal derivative of ψ at the surface. Substitution into the divergence theorem produces

$$\int_V d^3x [\nabla\phi(\mathbf{x}) \cdot \nabla\psi(\mathbf{x}) + \phi(\mathbf{x})\nabla^2\psi(\mathbf{x})] = \int_S d^2x \phi(\mathbf{x})\frac{\partial\psi(\mathbf{x})}{\partial n}, \quad (81)$$

a result known as *Green's first identity*. We may also start from a vector field $\mathbf{A}(\mathbf{x}) = \psi(\mathbf{x})\nabla\phi(\mathbf{x})$ and wind up with

$$\int_V d^3x [\nabla\psi(\mathbf{x}) \cdot \nabla\phi(\mathbf{x}) + \psi(\mathbf{x})\nabla^2\phi(\mathbf{x})] = \int_S d^2x \psi(\mathbf{x})\frac{\partial\phi(\mathbf{x})}{\partial n}. \quad (82)$$

Subtract the second expression from the first and obtain

$$\int_V d^3x [\phi(\mathbf{x})\nabla^2\psi(\mathbf{x}) - \psi(\mathbf{x})\nabla^2\phi(\mathbf{x})] = \int_S d^2x \left[\phi(\mathbf{x})\frac{\partial\psi(\mathbf{x})}{\partial n} - \psi(\mathbf{x})\frac{\partial\phi(\mathbf{x})}{\partial n} \right], \quad (83)$$

which is *Green's second identity*, also known as *Green's Theorem*.

8.2 Applying Green's Theorem 1

We next make a particular choice of the scalar functions and also change the integration variable to \mathbf{x}' :

$$\phi(\mathbf{x}') = \Phi(\mathbf{x}') \text{ and } \psi(\mathbf{x}') = 1/|\mathbf{x} - \mathbf{x}'| \equiv 1/R. \quad (84)$$

In the latter function, \mathbf{x} is to be regarded as a parameter which will eventually become the point at which we evaluate the potential. Substitution into Green's Theorem gives

$$\int_V d^3x' [\Phi(\mathbf{x}')\nabla'^2(1/R) - (1/R)\nabla'^2\Phi(\mathbf{x}')] \quad (85)$$

$$= \int_S d^2x' \left[\Phi(\mathbf{x}')\frac{\partial(1/R)}{\partial n'} - \frac{1}{R}\frac{\partial\Phi(\mathbf{x}')}{\partial n'} \right], \quad (86)$$

or

$$-4\pi\Phi(\mathbf{x}) + 4\pi \int_V d^3x' \frac{\rho(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|} = \int_S d^2x' \left[\Phi(\mathbf{x}')\frac{\partial(1/R)}{\partial n'} - \frac{1}{R}\frac{\partial\Phi(\mathbf{x}')}{\partial n'} \right] \quad (87)$$

where we have assumed \mathbf{x} is inside of V . With a little rearrangement, the final equation can be written as

$$\Phi(\mathbf{x}) = \int_V d^3x' \frac{\rho(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|} + \frac{1}{4\pi} \int_S d^2x' \left[\frac{1}{R}\frac{\partial\Phi(\mathbf{x}')}{\partial n'} - \Phi(\mathbf{x}')\frac{\partial(1/R)}{\partial n'} \right] \quad (88)$$

The first term on the right is the familiar volume integral over the charge density, but notice that it no longer is over all space. The charge outside of V does, of course, contribute to $\Phi(\mathbf{x})$; its contribution is now taken into account by the integral over the surface surrounding V . Note, too, that if the point \mathbf{x} is outside of V , then the left-hand side (LHS)

of the equation is zero. Further, if V is all space so that S is at infinity, then the surface integral will vanish and we recover the volume integral over all space. Note also that if $\rho(\mathbf{x}') = 0$ for all \mathbf{x}' in V , then the potential is found simply from the surface integral of Φ and its normal derivative.

8.3 Applying Green's Theorem 2

This equation is not the best one of its kind. It is in fact (as we shall see) possible to find $\Phi(\mathbf{x})$ from the charge density in V and from **either** $\Phi(\mathbf{x}')$ **or** $\partial\Phi(\mathbf{x}')/\partial n'$ on the surface S ; that is, it is not necessary to know both of these things on the surface. When Φ is specified, that is called *Dirichlet boundary conditions*; when the normal derivative of Φ is given, that is called *Neumann boundary conditions*. Various combinations are also possible, such as Dirichlet conditions on part of S and Neumann conditions on the remainder, a case known as *mixed boundary conditions*.

specification	Boundary Condition
$\phi(\mathbf{x}')$ specified on S	Dirichlet
$\partial\Phi(\mathbf{x}')/\partial n'$ specified on S	Neumann
$\phi(\mathbf{x}')$ specified on part of S $\partial\Phi(\mathbf{x}')/\partial n'$ specified on remainder	Mixed

Let us demonstrate that it is sufficient to know either $\Phi(\mathbf{x}')$ or its normal derivative on S in order to obtain a unique solution to

$$\nabla^2\Phi(\mathbf{x}) = -4\pi\rho(\mathbf{x}) \quad (89)$$

with some volume V bounded by S . Start by supposing that either type of boundary condition is given and that there are two distinct solutions Φ_1 and Φ_2 . Define $U = \Phi_1 - \Phi_2$. This function is such that

$$\nabla^2U(\mathbf{x}) = 0 \text{ inside of } V \quad (90)$$

and

$$\begin{aligned} \text{either } U(\mathbf{x}) &= 0, \text{ for } \mathbf{x} \text{ on } S \text{ (Dirichlet)} \\ \text{or } \frac{\partial U(\mathbf{x})}{\partial n} &= 0, \text{ for } \mathbf{x} \text{ on } S \text{ (Neumann)}. \end{aligned} \quad (91)$$

In Green's first identity Eq. (81), let $\psi = \phi = U$:

$$\int_V d^3x [U(\mathbf{x})\nabla^2U(\mathbf{x}) + \nabla U(\mathbf{x}) \cdot \nabla U(\mathbf{x})] = \int_S d^2x U(\mathbf{x}) \frac{\partial U(\mathbf{x})}{\partial n}. \quad (92)$$

Now, since U satisfies the Laplace equation, the first term in brackets on the left vanishes. From the boundary condition for either the Neumann or the Dirichlet problem, the surface integral also vanishes. Hence we have just

$$\int_V d^3x |\nabla U(\mathbf{x})|^2 = 0 \quad (93)$$

from which it follows that $\nabla U(\mathbf{x}) = 0$ in V . Therefore $U(\mathbf{x})$ is a constant in V and so the two solutions Φ_1 and Φ_2 are the same up to a constant. For Dirichlet conditions, this constant is zero since the two functions

are the same on the boundary. For Neumann conditions it is arbitrary and amounts only to a choice of the zero of potential.

The preceding proof is also valid for the case of mixed boundary conditions because the surface integral vanishes in this case also. Finally, one cannot in general specify *both* $\Phi(\mathbf{x})$ and $\frac{\partial\Phi}{\partial n}$ everywhere on S (*Cauchy* boundary conditions); either one alone is sufficient to determine a unique solution and the two solutions so determined are not necessarily the same. However, if Cauchy boundary conditions are given on just an appropriate part of S , that can be sufficient to give a unique solution for the potential.

In the light of what we have learned, it is evident that our integral expression for $\Phi(\mathbf{x})$, Eq. (88), which involves surface integrals of both the potential and its normal derivative, is not a very effective way to solve an electrostatic boundary value problem; it requires more input information than is actually needed to determine a solution and so is an integral equation as opposed to a solution in the form of an integral. If we had made a better choice of $\psi(\mathbf{x}')$ at the outset, we could have come up with a better result. Let's try again, choosing for $\psi(\mathbf{x}')$ a function we shall call $G(\mathbf{x}, \mathbf{x}')$; it is given by $1/|\mathbf{x} - \mathbf{x}'|$ plus an as yet undetermined function $F(\mathbf{x}, \mathbf{x}')$ which is to be a solution of the Laplace equation in V ,

$$\nabla'^2 F(\mathbf{x}, \mathbf{x}') = 0 \tag{94}$$

for \mathbf{x} and \mathbf{x}' in V . Since

$$G(\mathbf{x}, \mathbf{x}') = \frac{1}{|\mathbf{x} - \mathbf{x}'|} + F(\mathbf{x}, \mathbf{x}'), \quad (95)$$

it is the case that

$$\nabla'^2 G(\mathbf{x}, \mathbf{x}') = -4\pi\delta(\mathbf{x} - \mathbf{x}') \quad (96)$$

for \mathbf{x} and \mathbf{x}' in V . Physically, the function G , viewed as a function of \mathbf{x}' , is a solution of Poisson's equation in V given a unit point charge at \mathbf{x} . The function $F(\mathbf{x}, \mathbf{x}')$ is a solution of Laplace's equation; we shall presently determine its properties further by requiring that it satisfy certain conditions on S . Using $G(\mathbf{x}, \mathbf{x}')$ for $\psi(\mathbf{x}')$ and $\Phi(\mathbf{x})$ for $\phi(\mathbf{x})$ in Green's Theorem, we can easily show that

$$\Phi(\mathbf{x}) = \int_V d^3x' G(\mathbf{x}, \mathbf{x}')\rho(\mathbf{x}') + \frac{1}{4\pi} \int_S d^2x' \left[G(\mathbf{x}, \mathbf{x}') \frac{\partial\Phi(\mathbf{x}')}{\partial n'} - \Phi(\mathbf{x}') \frac{\partial G(\mathbf{x}, \mathbf{x}')}{\partial n'} \right]. \quad (97)$$

8.3.1 Greens Theorem with Dirichlet B.C.

Now consider in turn two different sets of boundary conditions for G . First, require that $G(\mathbf{x}, \mathbf{x}') = 0$ for \mathbf{x}' on S and denote this function by G_D . Then the preceding equation becomes

$$\Phi(\mathbf{x}) = \int_V d^3x' G_D(\mathbf{x}, \mathbf{x}')\rho(\mathbf{x}') - \frac{1}{4\pi} \int_S d^2x' \frac{\partial G_D(\mathbf{x}, \mathbf{x}')}{\partial n'} \Phi(\mathbf{x}'). \quad (98)$$

This is a useful equation when we have a Dirichlet problem with $\Phi(\mathbf{x}')$ specified for \mathbf{x}' on S . Then we have in principle the information we need

to complete the integration and so find $\Phi(\mathbf{x})$ in which case Eq. (98) is an integral solution for $\Phi(\mathbf{x})$ as opposed to an integral equation for the potential.

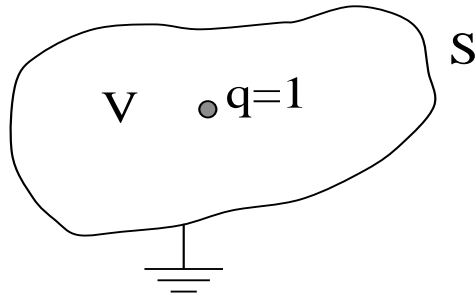
The question naturally arises, does the function G_D exist? That is, is it possible to find the Dirichlet Green's function G_D which is specified by the conditions

$$\nabla^2 G_D(\mathbf{x}, \mathbf{x}') = -4\pi\delta(\mathbf{x} - \mathbf{x}'); \quad \mathbf{x}, \mathbf{x}' \text{ in } V \quad (99)$$

and

$$G_D(\mathbf{x}, \mathbf{x}')|_{\mathbf{x}' \text{ on } S} = 0? \quad (100)$$

The answer is that this function does exist; further, it is unique. The preceding two conditions are sufficient to determine it completely. We know this without resorting to a mathematical proof because we can see that $G_D(\mathbf{x}, \mathbf{x}')$ is just the scalar potential at \mathbf{x}' given a unit point charge at \mathbf{x} inside of a cavity with conducting walls coincident with S and held at zero potential.



This is the physical interpretation of the Dirichlet Green's function. Notice in particular that this is a strongly geometry-dependent function (it depends on S very much) but it is not dependent on any other

properties of the system. In other words, we can solve any Dirichlet problem for a given geometry if we can solve the “point charge with grounded conducting surfaces” problem for the same geometry in the sense that we can reduce the solution to a quadrature, *i.e.*, to an integral.

An important property of the Dirichlet Green’s function is that it is invariant under interchange of \mathbf{x} and \mathbf{x}' , $G(\mathbf{x}, \mathbf{x}') = G(\mathbf{x}', \mathbf{x})$. To demonstrate this property, let $\phi(\mathbf{y}) = G(\mathbf{x}, \mathbf{y})$ and $\psi(\mathbf{y}) = G(\mathbf{x}', \mathbf{y})$. Then insert these functions into Green’s Theorem (with \mathbf{y} as the integration variable)

$$\int_V d^3y [G(\mathbf{x}, \mathbf{y})\nabla^2 G(\mathbf{x}', \mathbf{y}) - G(\mathbf{x}', \mathbf{y})\nabla^2 G(\mathbf{x}, \mathbf{y})] = \int_S d^2y \left[G(\mathbf{x}, \mathbf{y})\frac{\partial G(\mathbf{x}', \mathbf{y})}{\partial n} - G(\mathbf{x}', \mathbf{y})\frac{\partial G(\mathbf{x}, \mathbf{y})}{\partial n} \right], \quad (101)$$

and make use of the properties of the Dirichlet Green’s function that $\nabla_{\mathbf{y}}^2 G(\mathbf{x}, \mathbf{y}) = -4\pi\delta(\mathbf{x} - \mathbf{y})$ and $G(\mathbf{x}, \mathbf{y}) = 0$ for \mathbf{y} on S. The result is that $G(\mathbf{x}, \mathbf{x}') = G(\mathbf{x}', \mathbf{x})$.

8.3.2 Greens Theorem with Neumann B.C.

The second case of boundary conditions we consider on G is $\partial G(\mathbf{x}, \mathbf{x}')/\partial n' = 0$ for \mathbf{x}' on S. Then application of Green’s Theorem (Eq. (97)) leads to

$$\Phi(\mathbf{x}) = \int_V d^3x' G(\mathbf{x}, \mathbf{x}')\rho(\mathbf{x}') + \frac{1}{4\pi} \int_S d^2x' \frac{\partial \Phi(\mathbf{x}')}{\partial n'} G(\mathbf{x}, \mathbf{x}'). \quad (102)$$

Unfortunately, this G does not exist as we may show by applying Gauss's Law,

$$0 = \int_S d^2x' \frac{\partial G(\mathbf{x}, \mathbf{x}')}{\partial n'} = \int_S d^2x' [\mathbf{n}' \cdot \nabla' G(\mathbf{x}, \mathbf{x}')] = \int_V d^3x' \nabla'^2 G(\mathbf{x}, \mathbf{x}') = -4\pi. \quad (103)$$

Clearly, we cannot have a G with zero normal derivative everywhere on S . The next simplest possibility is that

$$\left. \frac{\partial G_N(\mathbf{x}, \mathbf{x}')}{\partial n'} \right|_{\mathbf{x}' \text{ on } S} = -\frac{4\pi}{S}, \quad (104)$$

where S is the area of the surface. Given such a function, we can use it in Green's Theorem and will be led to the following integral expression for the scalar potential:

$$\Phi(\mathbf{x}) = \langle \Phi \rangle_S + \int_V d^3x' G_N(\mathbf{x}, \mathbf{x}') \rho(\mathbf{x}') + \frac{1}{4\pi} \int_S d^2x' \frac{\partial \Phi(\mathbf{x}')}{\partial n'} G_N(\mathbf{x}, \mathbf{x}') \quad (105)$$

where $\langle \Phi \rangle_S$ is the average of the potential over the surface S ,

$$\langle \Phi \rangle_S \equiv \frac{1}{S} \int_S d^2x' \Phi(\mathbf{x}') \quad (106)$$

One can understand the necessity of the presence of this term from the fact that the Neumann boundary condition problem can only be solved up to an arbitrary constant.

The Dirichlet Green's function is the one that we shall use most often as one more commonly specifies the potential on the boundary than the normal component of the electric field.

Boundary-value Problems in Electrostatics I

Karl Friedrich Gauss

(1777 - 1855)

December 23, 2000

Contents

1	Method of Images	1
1.1	Point Charge Above a Conducting Plane	2
1.2	Point Charge Between Multiple Conducting Planes	4
1.3	Point Charge in a Spherical Cavity	5
1.4	Conducting Sphere in a Uniform Applied Field	13
2	Green's Function Method for the Sphere	16
3	Orthogonal Functions and Expansions; Separation of Variables	19
3.1	Fourier Series	22
3.2	Separation of Variables	24
3.3	Rectangular Coordinates	25
3.4	Fields and Potentials on Edges	29

4	Examples	34
4.1	Two-dimensional box with Neumann boundaries	34
4.2	Numerical Solution of Laplace's Equation	37
4.3	Derivation of Eq. 35: A Mathematica Session	40

In this chapter we shall solve a variety of boundary value problems using techniques which can be described as commonplace.

1 Method of Images

This method is useful given sufficiently simple geometries. It is closely related to the Green's function method and can be used to find Green's functions for these same simple geometries. We shall consider here only conducting (equipotential) bounding surfaces which means the boundary conditions take the form of $\Phi(\mathbf{x}) = \text{constant}$ on each electrically isolated conducting surface. The idea behind this method is that the solution for the potential in a finite domain V with specified charge density and potentials on its surface S can be the *same* within V as the solution for the potential given the same charge density inside of V but a quite different charge density elsewhere. Thus we consider two distinct electrostatics problems. The first is the "real" problem in which we are given a charge density $\rho(\mathbf{x})$ in V and some boundary conditions on the surface S . The second is a "fictitious problem" in which the charge density inside of V is the same as for the real problem and in

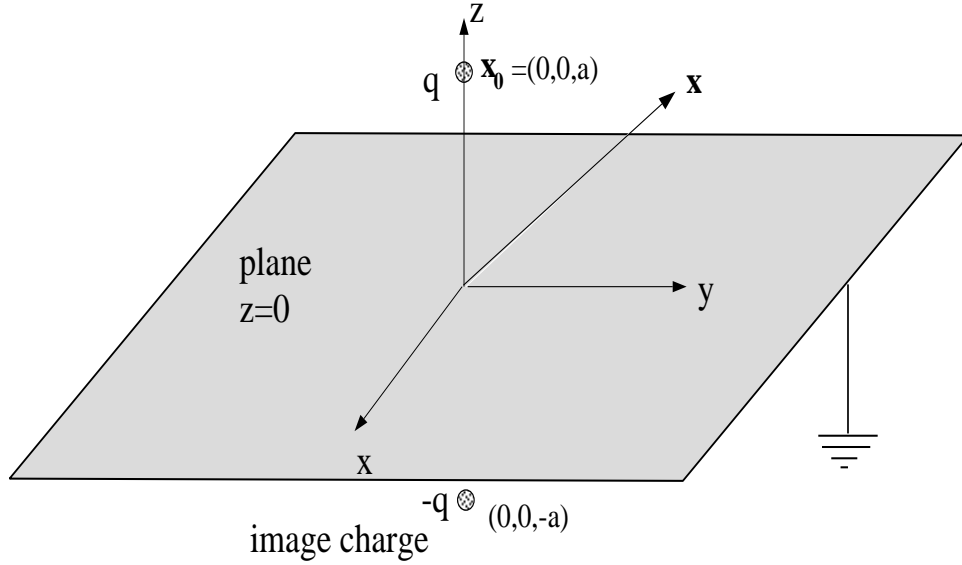
which there is some undetermined charge distribution elsewhere; this is to be chosen such that the solution to the second problem satisfies the boundary conditions specified in the first problem. Then the solution to the second problem is also the solution to the first problem inside of V (but not outside of V). If one has found the initially undetermined exterior charge in the second problem, called *image charge*, then the potential is found simply from Coulomb's Law,

$$\Phi(\mathbf{x}) = \int d^3x' \frac{\rho_2(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|}; \quad (1)$$

ρ_2 is the total charge density of the second problem.

1.1 Point Charge Above a Conducting Plane

This may sound confusing, but it is made quite clear by a simple example. Suppose that we have a point charge q located at a point $\mathbf{x}_0 = (0, 0, a)$ in Cartesian coordinates. Further, a grounded conductor occupies the half-space $z < 0$, which means that we have the Dirichlet boundary condition at $z = 0$ that $\Phi(x, y, 0) = 0$; also, $\Phi(\mathbf{x}) \rightarrow 0$ as $r \rightarrow \infty$. The first thing that we must do is determine some image charge located in the half-space $z < 0$ such that the potential of the image charge plus the real charge (at \mathbf{x}_0) produces zero potential on the $z = 0$ plane. With just a little thought one realizes that a single image charge $-q$ located at the point $\mathbf{x}'_0 = (0, 0, -a)$ is what is required.



All points on the $z = 0$ plane are equidistant from the real charge and its image, and so the two charges produce cancelling potentials at each of these points. The solution to the problem is therefore

$$\Phi(\mathbf{x}) = q \left[\frac{1}{|\mathbf{x} - \mathbf{x}_0|} - \frac{1}{|\mathbf{x} - \mathbf{x}'_0|} \right]. \quad (2)$$

This function satisfies the correct Poisson equation in the domain $z > 0$ and also satisfies the correct boundary conditions at $z = 0$; therefore it is the (unique) solution. It is important to realize, however, that it is *not* the correct solution in the space $z < 0$; here, the real potential is zero because this domain is inside of the grounded conductor.

In the real system, there is some surface charge density $\sigma(x, y)$ on the conductor; to determine what this is, we have only to evaluate the normal component of the electric field at the surface of the conductor,

$$\mathbf{E}_n(x, y, 0) = - \left. \frac{\partial \Phi(\mathbf{x})}{\partial z} \right|_{z=0} = - \frac{2qa}{(\rho^2 + a^2)^{3/2}}, \quad (3)$$

where $\rho = \sqrt{x^2 + y^2}$. The surface charge density is just this field, divided by 4π , $E_n = \sigma/4\pi$.

From this example we can also see why this technique has the name ‘method of images.’ The image charge is precisely the mirror image in the $z = 0$ plane of the real charge.

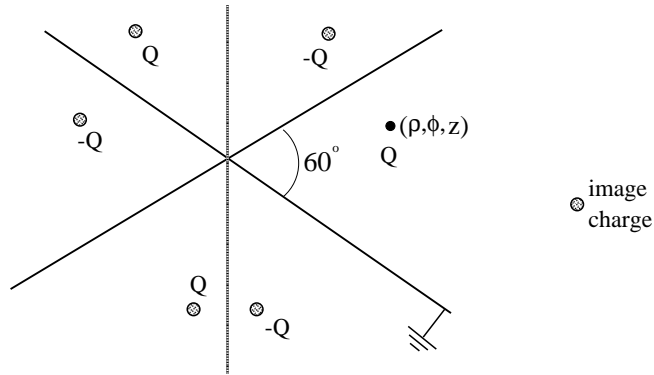
As a by-product of our solution, we have also got the Dirichlet Green’s function for the semi-infinite half-space $z > 0$; it is

$$G(\mathbf{x}, \mathbf{x}') = \left(\frac{1}{|\mathbf{x} - \mathbf{x}'|} - \frac{1}{|\mathbf{x} - \mathbf{x}'_i|} \right) \quad (4)$$

where \mathbf{x}'_i is the mirror image of \mathbf{x}' in the $z = 0$ plane. Hence we can solve, by doing appropriate integrals, any problem in which we are given some $\rho(\mathbf{x})$ in the domain $z > 0$ and an arbitrary potential $\Phi(x, y, 0)$.

1.2 Point Charge Between Multiple Conducting Planes

A simple extension of the problem above is one with a point charge between two intersecting conducting planes. For example, consider two grounded conducting planes that intersect at an angle of 60° forming a wedge, with point charge Q placed at (ρ, ϕ, z) within the wedge.



To solve this problem, we again use image charges to satisfy the boundary conditions. There are five image charges, as indicated in the figure above. They all share the same value of ρ and z as the real charge, and their azimuthal angles are given in the table below.

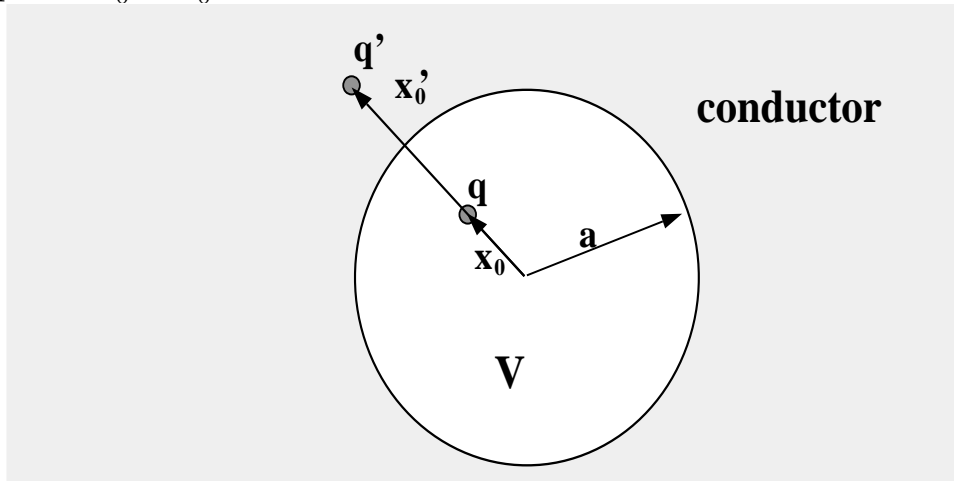
charge	angle
$-Q$	$\frac{2\pi}{3} - \phi$
$+Q$	$\frac{2\pi}{3} + \phi$
$-Q$	$-\frac{2\pi}{3} - \phi$
$+Q$	$-\frac{2\pi}{3} + \phi$
$-Q$	$-\phi$

1.3 Point Charge in a Spherical Cavity

It is also sometimes possible to use the image method when the boundary S involves curved surfaces. However, just as curved mirrors produced distorted images, so do curved surfaces make the image of a point

charge more complicated¹. Let's do the simplest problem of this kind. Suppose that we have a spherical cavity of radius a inside of a conductor; within this cavity is a point charge q located a distance r_0 from the center of the sphere which is also chosen as the origin of coordinates. Thus the charge is at point $\mathbf{x}_0 = r_0 \mathbf{n}_0$ where \mathbf{n}_0 is a unit vector pointing in the direction from the origin to the charge.

We need to find the image(s) of the charge in the spherical surface which encloses it. The simplest possible set of images would be a single charge q' ; if there is such a solution, symmetry considerations tell us that the image must be located on the line passing through the origin and going in the direction of \mathbf{n}_0 . Let us therefore put an image charge q' at point $\mathbf{x}'_0 = r'_0 \mathbf{n}_0$.



¹Consider the example of the right-hand side-view mirror of a car. Here the mirror is concave, and images appear to be much farther away than they actually are

The potential produced by this charge and the real one at \mathbf{x}_0 is

$$\Phi(\mathbf{x}) = \frac{q}{|\mathbf{x} - \mathbf{x}_0|} + \frac{q'}{|\mathbf{x} - \mathbf{x}'_0|}. \quad (5)$$

Now we must choose, if possible, q' and r'_0 such that $\Phi(\mathbf{x}) = 0$ for \mathbf{x} on the cavity's spherical surface, $\mathbf{x} = a\mathbf{n}$ where the direction of \mathbf{n} is arbitrary. The potential at such a point may be written as

$$\Phi(a\mathbf{n}) = \frac{q/a}{|\mathbf{n} - (r_0/a)\mathbf{n}_0|} + \frac{q'/r'_0}{|\mathbf{n}_0 - (a/r'_0)\mathbf{n}|}. \quad (6)$$

The from the figure below, it is clear that denominators are equal if $r_0/a = a/r'_0$, and the numerators are equal and opposite if $q/a = -q'/r'_0$. (The “other” solution, $r'_0 = r_0$ and $q' = -q$ is no solution at all since then the image charge would be within the volume V and cancel the real charge. We must have $r'_0 > 1$)



$$\mathbf{d}_2 = \mathbf{d}_1 \text{ if } (a/r'_0) = (r_0/a)$$

Hence, we make Φ zero on S by choosing

$$r'_0 = a^2/r_0 \text{ and } q' = -q(a/r_0). \quad (7)$$

Thus we have the solution for a point charge in a spherical cavity with an equipotential surface:

$$\Phi(\mathbf{x}) = q \left[\frac{1}{|\mathbf{x} - \mathbf{x}_0|} - \frac{a/r_0}{|\mathbf{x} - (a^2/r_0^2)\mathbf{x}_0|} \right]; \quad (8)$$

we have also found the Dirichlet Green's function for the interior of a sphere of radius a :

$$G(\mathbf{x}, \mathbf{x}') = \frac{1}{|\mathbf{x} - \mathbf{x}'|} - \frac{a/r}{|\mathbf{x}' - (a^2/r^2)\mathbf{x}|}. \quad (9)$$

The solution of the “inverse” problem which is a point charge outside of a conducting sphere is the same, with the roles of the real charge and the image charge reversed. The preceding equations for $\Phi(\mathbf{x})$ and $G(\mathbf{x}, \mathbf{x}')$ are valid except that r , r_0 , and r' are all larger than a .

Let's look at a few more features of the solution for the charge inside of the spherical cavity. First, what is σ , the charge density on the surface of the cavity. From Gauss's Law, we know that the charge density is the normal component of the electric field out of the conductor at its surface divided by 4π . This is the negative of the radial component in spherical polar coordinates, so

$$\sigma = -\frac{E_r}{4\pi} = \frac{1}{4\pi} \frac{\partial \Phi}{\partial r}. \quad (10)$$

If we define the z -direction to be the direction of \mathbf{n}_0 , then the potential at an arbitrary point within the sphere is

$$\Phi(\mathbf{x}) = q \left[\frac{1}{(r^2 + r_0^2 - 2rr_0 \cos \theta)^{1/2}} - \frac{(a/r_0)}{(r^2 + (a^4/r_0^2) - 2r(a^2/r_0) \cos \theta)^{1/2}} \right] \quad (11)$$

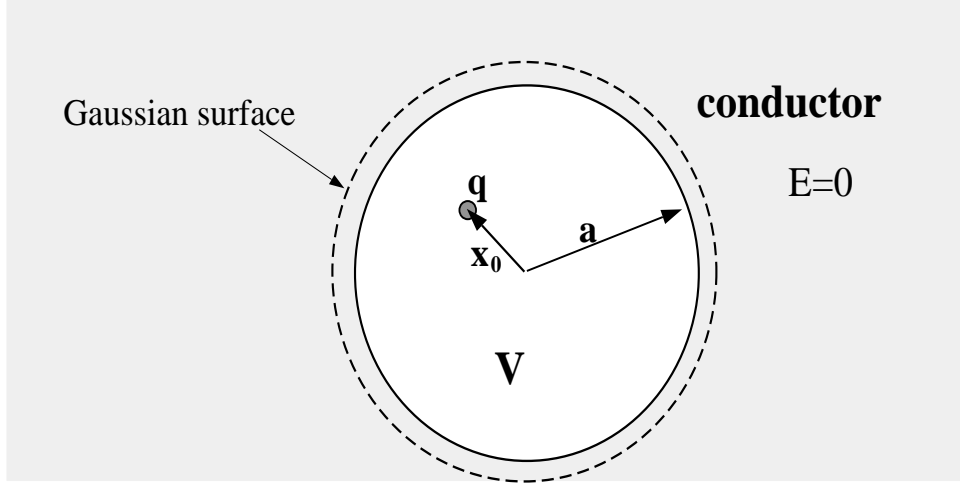
where θ is the usual polar angle between the z -axis, or direction of \mathbf{x}_0 , and the direction of \mathbf{x} , the field point. The radial component of the electric field at $r = a$ is

$$\begin{aligned}
E_r &= - \left. \frac{\partial \Phi(\mathbf{x})}{\partial r} \right|_{r=a} \\
&= -q \left[-\frac{1}{2} \frac{2a - 2r \cos \theta}{(a^2 + r_0^2 - 2ar_0 \cos \theta)^{3/2}} + \frac{a}{r_0} \frac{1}{2} \frac{2a - 2(a^2/r_0) \cos \theta}{(a^2 + (a^4/r_0^2) - 2(a^3/r_0) \cos \theta)^{3/2}} \right] \\
&= q \left[\frac{a - r_0 \cos \theta}{(a^2 + r_0^2 - 2ar_0 \cos \theta)^{3/2}} - \frac{r_0^2}{a^2} \frac{a - (a^2/r_0) \cos \theta}{(a^2 + r_0^2 - 2ar_0 \cos \theta)^{3/2}} \right] \\
&= \frac{q}{a^2} \left[\frac{1 - r_0^2/a^2}{(1 + (r_0^2/a^2) - 2(r_0/a) \cos \theta)^{3/2}} \right] \quad (12)
\end{aligned}$$

If we introduce $\epsilon = r_0/a$, then the surface charge density may be written concisely as

$$\sigma = -\frac{q}{4\pi a^2} \frac{1 - \epsilon^2}{(1 + \epsilon^2 - 2\epsilon \cos \theta)^{3/2}}. \quad (13)$$

The **total charge on the surface** may be found by integrating over σ . But it may be obtained more easily by invoking Gauss's Law; if we integrate the normal component of $\mathbf{E}(\mathbf{x})$ over a closed surface which lies entirely in conducting material and which also encloses the cavity, we know that we will get zero, because the field in the conductor is zero.



$$0 = \int_S d^2x \mathbf{E} \cdot \mathbf{n} = 4\pi Q = 4\pi \left(q + \int_S d^2x \sigma \right)$$

charge within this surface. What is inside is the charge q in the cavity and the surface charge on the conductor. The implication is that the total surface charge is equal to $-q$. It is perhaps useful to actually do the integral over the surface as a check that we have gotten the charge density there right:

$$\begin{aligned} Q_i &= \int_S d^2x \sigma(\mathbf{x}) \\ &= -\frac{q}{2}(1 - \epsilon^2) \int_{-1}^1 \frac{du}{(1 + \epsilon^2 - 2\epsilon u)^{3/2}} \\ &= -\frac{q}{2} \frac{1 - \epsilon^2}{2\epsilon} \left[\frac{2}{(1 + \epsilon^2 - 2\epsilon)^{1/2}} - \frac{2}{(1 + \epsilon^2 + 2\epsilon)^{1/2}} \right] \\ &= -\frac{q}{2} \frac{1 - \epsilon^2}{\epsilon} \left(\frac{1}{1 - \epsilon} - \frac{1}{1 + \epsilon} \right) = -q. \end{aligned} \quad (14)$$

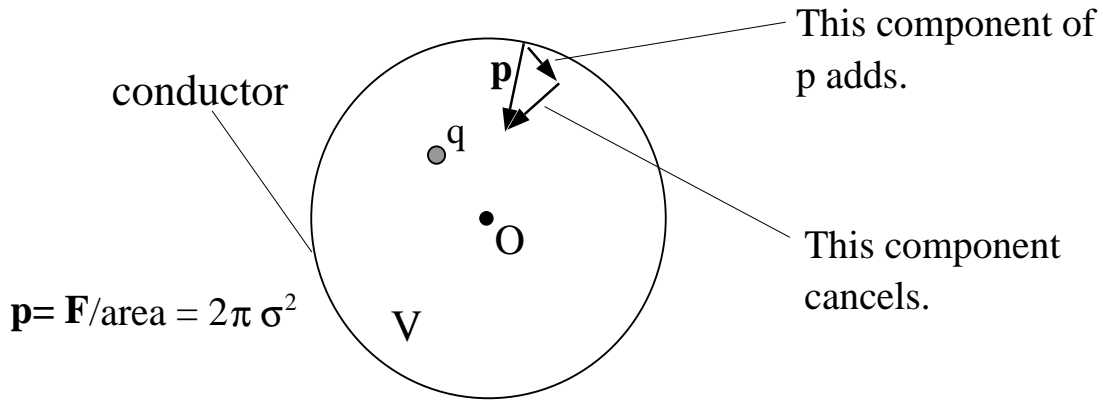
Notice that $|\sigma|$ is largest in the direction of \mathbf{n}_0 and is

$$|\sigma_{max}| = -\frac{q}{4\pi a^2} \frac{1 + \epsilon}{(1 - \epsilon)^2}. \quad (15)$$

In the opposite direction, the magnitude of the charge density is at its minimum which is

$$|\sigma_{min}| = -\frac{q}{4\pi a^2} \frac{1 - \epsilon}{(1 + \epsilon)^2}. \quad (16)$$

The **total force on the charge** may also be computed. This is the negative of the total force on the conductor. Now, we know that the force per unit area on the surface of the conductor is $2\pi\sigma^2$ and is directed normal to the conductor's surface into the cavity. Because of the rotational invariance of the system around the direction of \mathbf{n}_0 , only the component of the force along this direction need be computed; the other components will average to zero when integrated over the surface.



Hence we find

$$\begin{aligned} |\mathbf{F}_n| &= 2\pi a^2 \int_0^{2\pi} d\phi \int_0^\pi \sin\theta d\theta \sigma^2(\theta) \cos\theta \\ &= \frac{4\pi^2 a^2 q^2}{16\pi^2 a^4} (1 - \epsilon^2)^2 \int_{-1}^1 \frac{udu}{(1 + \epsilon^2 - 2\epsilon u)^3} = \frac{1}{4} \frac{q^2}{a^2} (1 - \epsilon^2)^2 \int_{-1}^1 \frac{udu}{(1 + \epsilon^2 - 2\epsilon u)^3} \\ &= \frac{1}{4} \frac{q^2}{a^2} (1 - \epsilon^2)^2 \frac{1}{4\epsilon^2} \left[-\frac{1}{1 + \epsilon^2 - 2\epsilon u} + \frac{1 + \epsilon^2}{2(1 + \epsilon^2 - 2\epsilon u)^2} \right]_{-1}^1 \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{4} \frac{q^2 (1 - \epsilon^2)^2}{a^2 4\epsilon^2} \left[\frac{-1}{(1 - \epsilon)^2} + \frac{1}{(1 + \epsilon)^2} + \frac{1 + \epsilon^2}{2(1 - \epsilon)^4} - \frac{1 + \epsilon^2}{2(1 + \epsilon)^4} \right] \\
&= \frac{q^2 (1 - \epsilon^2)^2}{4a^2 4\epsilon^2} \left[\frac{-4\epsilon}{(1 - \epsilon^2)^2} + \frac{(1 + \epsilon^2)(8\epsilon + 8\epsilon^3)}{2(1 - \epsilon^2)^4} \right] \\
&= \frac{q^2}{4a^2} \frac{-4\epsilon(1 - \epsilon^2)^2 + 4\epsilon(1 + \epsilon^2)^2}{4\epsilon^2(1 - \epsilon^2)^2} = \frac{q^2}{a^2} \frac{\epsilon}{(1 - \epsilon^2)^2} \quad (17)
\end{aligned}$$

The direction of this force is such that the charge is attracted toward the point on the cavity wall that is closest to it.

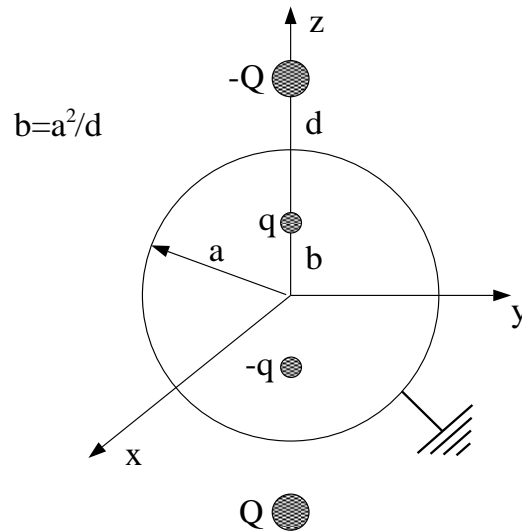
We may also ask what is the “force” between the charge and its image. The distance between them is $r'_0 - r_0 = a\epsilon(1/\epsilon^2 - 1) = a(1 - \epsilon^2)/\epsilon$, and the product of the charges is $qq' = -q^2/\epsilon$, so

$$|\mathbf{F}| = \frac{q^2}{a^2} \frac{\epsilon}{(1 - \epsilon^2)^2} \quad (18)$$

which is the same as the real force between the charge and the surface. One is led to ask whether the real force on the charge is always the same as that between the charge and its images. The answer is yes. The electric field produced by the real surface charge at the position of the real charge is the same as that produced by the image charge at the real charge, and so the same force will arise in both systems. It is generally much easier to calculate the force between the real charge and its images than the force between the real charge and the surface charges.

1.4 Conducting Sphere in a Uniform Applied Field

Consider next the example of a grounded conducting sphere, which means that $\Phi(\mathbf{x}) = 0$ on the sphere, placed in a region of space where there was initially a uniform electric field $\mathbf{E}_0 = E_0 \hat{\mathbf{z}}$ produced by some far away fixed charges. Here, $\hat{\mathbf{z}}$ is a unit vector pointing in the z -direction. We approach this problem by replacing it with another one which will become equivalent to the first one in some limit. Let the sphere be centered at the origin and let there be not a uniform applied field but rather a charge Q placed at the point $(0, 0, -d)$ and another charge $-Q$ placed at the point $(0, 0, d)$ in Cartesian coordinates.



The resulting potential configuration is easily solved by the image method; there are images of the charges $\pm Q$ in the sphere at $(0, 0, -a^2/d)$ and at $(0, 0, a^2/d)$; they have size $-Qa/d$ and Qa/d , respectively. The potential produced by these four charges is zero on the surface of the

sphere. Thus we have solved the problem of a grounded sphere in the presence of two symmetrically located equal and opposite charges. We could equally well think of the sphere as isolated (not electrically connected to anything) and neutral, because the total image charge is zero.

Now we want to think about what happens if we let Q become increasingly large and at the same time move the real charges farther and farther away from the sphere in such a way that the field they produce at the origin is constant. This field is $\mathbf{E}(\mathbf{x}) = (2Q/d^2)\hat{\mathbf{z}}$, so if Q is increased at a rate proportional to d^2 , the field at the origin is unaffected. As d becomes very large in comparison with the radius a of the sphere, not only will the applied field at the origin have this value, but it will have very nearly this value everywhere in the vicinity of the sphere. The difference becomes negligible in the limit $d/a \rightarrow \infty$. Hence we recover the configuration presented in the original problem of a sphere placed in a uniform applied field. If we pick $E_0 = 2Q/d^2$, or, more appropriately, $Q = E_0 d^2/2$, we have the solution in the limit of $d \rightarrow \infty$:

$$\begin{aligned} \Phi(\mathbf{x}) &= \lim_{d \rightarrow \infty} \left[\frac{E_0 d^2/2}{(d^2 + r^2 + 2rd \cos \theta)^{1/2}} - \frac{E_0 d^2 a/2d}{(a^4/d^2 + r^2 + 2r(a^2/d) \cos \theta)^{1/2}} \right] \\ &\quad + \lim_{d \rightarrow \infty} \left[-\frac{E_0 d^2/2}{(d^2 + r^2 - 2rd \cos \theta)^{1/2}} + \frac{E_0 d^2 a/2d}{(a^4/d^2 + r^2 - 2r(a^2/d) \cos \theta)^{1/2}} \right] \\ &= \lim_{d \rightarrow \infty} \left[\pm \frac{E_0 d/2}{(1 \pm 2(r/d) \cos \theta + r^2/d^2)^{1/2}} \mp \frac{E_0 da/2r}{(1 \pm (a^2/rd) \cos \theta + a^4/d^2 r^2)^{1/2}} \right] \end{aligned}$$

$$= -E_0 r \cos \theta + \frac{E_0 a^3}{r^2} \cos \theta \quad (19)$$

The first term, $-E_0 r \cos \theta$, is the potential of the applied constant field, \mathbf{E}_0 . The second is the potential produced by the induced surface charge density on the sphere. This has the characteristic form of an *electric dipole* field, of which we shall hear more presently. The *dipole moment* \mathbf{p} associated with any charge distribution is defined by the equation

$$\mathbf{p} = \int d^3x \mathbf{x} \rho(\mathbf{x}); \quad (20)$$

in the present case the dipole moment of the sphere may be found either from the surface charge distribution or from the image charge distribution. Taking the latter tack, we find

$$\begin{aligned} \mathbf{p} &= \int d^3x \mathbf{x} \frac{E_0 da}{2} [-\delta(z + a^2/d)\delta(y)\delta(x) + \delta(z - a^2/d)\delta(y)\delta(x)] \\ &= \frac{E_0 da}{2} [(a^2/d)\hat{\mathbf{z}} + (a^2/d)\hat{\mathbf{z}}] = E_0 a^3 \hat{\mathbf{z}}. \end{aligned} \quad (21)$$

Comparison with the expression for the potential shows that the dipolar part of the potential may be written as

$$\Phi(\mathbf{x}) = \mathbf{p} \cdot \mathbf{x} / r^3 \quad (22)$$

The charge density on the surface of the sphere may be found in the usual way:

$$4\pi\sigma = E_r = - \left. \frac{\partial \Phi}{\partial r} \right|_{r=a} = E_0 \cos \theta + \frac{2E_0}{a^3} a^3 \cos \theta = 3E_0 \cos \theta. \quad (23)$$

Hence,

$$\sigma(\theta) = \frac{3}{4\pi} E_0 \cos \theta. \quad (24)$$

2 Green's Function Method for the Sphere

Next, let us do an example of the use of the Green's function method by considering a Dirichlet potential problem inside of a sphere. The task is to calculate the potential distribution inside of an empty ($\rho(\mathbf{x}) = 0$, $x \in V$) spherical cavity of radius a , given some specified potential distribution $V(\theta, \phi)$ on the surface of the sphere. We can immediately invoke the Green's function expression

$$\Phi(\mathbf{x}) = -\frac{1}{4\pi} \int_S d^2x' \Phi(\mathbf{x}') \frac{\partial G(\mathbf{x}, \mathbf{x}')}{\partial n'}, \quad (25)$$

and we already know that,

$$G(\mathbf{x}, \mathbf{x}') = \frac{1}{|\mathbf{x} - \mathbf{x}'|} - \frac{a/r'}{|\mathbf{x} - (a^2/r'^2)\mathbf{x}'|} \quad (26)$$

since $G(\mathbf{x}, \mathbf{x}')$ is the potential at \mathbf{x} due to a unit point charge at \mathbf{x}' ($\mathbf{x}, \mathbf{x}' \in V$), and we have just solved this problem. If we let γ be the angle between \mathbf{x} and \mathbf{x}' ,

$$G(\mathbf{x}, \mathbf{x}') = \frac{1}{(r^2 + r'^2 - 2rr' \cos \gamma)^{1/2}} - \frac{a/r'}{(r^2 + (a^4/r'^2) - 2r(a^2/r') \cos \gamma)^{1/2}}. \quad (27)$$

Then

$$\left. \frac{\partial G(\mathbf{x}, \mathbf{x}')}{\partial n'} \right|_S = \left. \frac{\partial G(\mathbf{x}, \mathbf{x}')}{\partial r'} \right|_{r'=a}$$

$$\begin{aligned}
&= -\frac{1}{2} \left[\frac{2a - 2r \cos \gamma}{(r^2 + a^2 - 2ra \cos \gamma)^{3/2}} - a \frac{2ar^2 - 2ra^2 \cos \gamma}{(r^2 a^2 + a^4 - 2ra^3 \cos \gamma)^{3/2}} \right] \\
&= -\frac{a(1 - r^2/a^2)}{(r^2 + a^2 - 2ra \cos \gamma)^{3/2}} = -\frac{1}{a^2} \frac{(1 - \epsilon^2)}{(1 + \epsilon^2 - 2\epsilon \cos \gamma)^{3/2}} \quad (28)
\end{aligned}$$

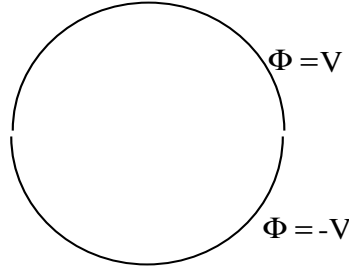
where $\epsilon = r/a$. For simplicity, let us suppose that $\rho(\mathbf{x}) = 0$ inside of the sphere. Then

$$\Phi(\mathbf{x}) = \frac{1}{4\pi} \int_0^{2\pi} d\phi' \int_0^\pi \frac{\sin \theta' d\theta' V(\theta', \phi') (1 - \epsilon^2)}{(1 + \epsilon^2 - 2\epsilon \cos \gamma)^{3/2}}. \quad (29)$$

In terms of θ, ϕ and θ' and ϕ' ,

$$\cos \gamma = \cos \theta \cos \theta' + \sin \theta \sin \theta' \cos(\phi - \phi'). \quad (30)$$

This integral can rarely be done in closed form in terms of simple functions; however, it is generally a simple matter to carry out the integrals numerically. As an example, consider



$$V(\theta, \phi) = \begin{cases} V, & 0 \leq \theta \leq \pi/2 \\ -V, & \pi/2 \leq \theta \leq \pi. \end{cases} \quad (31)$$

Then the answer will not depend on ϕ , so we may arbitrarily set ϕ equal to zero and proceed. Using $\epsilon \equiv r/a$, we have

$$\Phi(\epsilon, \theta) = \frac{V}{4\pi} (1 - \epsilon^2) \int_0^{2\pi} d\phi' \left[\int_0^{\pi/2} \frac{\sin \theta' d\theta'}{(1 + \epsilon^2 - 2\epsilon \cos \gamma)^{3/2}} - \int_{\pi/2}^\pi \frac{\sin \theta' d\theta'}{(1 + \epsilon^2 - 2\epsilon \cos \gamma)^{3/2}} \right] \quad (32)$$

The integral is still difficult in the general case. For $\theta = 0$, it is easier:

$$\Phi(\epsilon, 0) = \frac{V}{4\pi}(1 - \epsilon^2)2\pi \left[\int_0^1 \frac{du}{(1 + \epsilon^2 - 2\epsilon u)^{3/2}} - \int_{-1}^0 \frac{du}{(1 + \epsilon^2 - 2\epsilon u)^{3/2}} \right] \quad (33)$$

These integrals are easily completed with the result that

$$\Phi(\epsilon, 0) = \frac{V}{\epsilon} \left(1 - \frac{1 - \epsilon^2}{\sqrt{1 + \epsilon^2}} \right). \quad (34)$$

An alternative approach, valid for $r/a \ll 1$, is to expand the integrand in powers of ϵ and then to complete the integration term by term. This is straightforward with a symbolic manipulator but tedious by hand. Either way, a solution in powers of ϵ is generated.

$$\Phi(\epsilon, \theta) = \frac{3V}{2} \left[\epsilon \cos \theta - \frac{7}{12} \epsilon^3 \left(\frac{5}{2} \cos^3 \theta - \frac{3}{2} \cos \theta \right) + O(\epsilon^5) \right]. \quad (35)$$

The alert student will recognize that the functions of $\cos \theta$ that are being generated are *Legendre polynomials*;

$$\begin{aligned} P_1(\cos \theta) &= \cos \theta, \\ P_3(\cos \theta) &= \frac{5}{2} \cos^3 \theta - \frac{3}{2} \cos \theta, \end{aligned} \quad (36)$$

etc. Note that only terms which are odd in $\cos \theta$ enter into the sum, due to the symmetry of the boundary conditions.

3 Orthogonal Functions and Expansions; Separation of Variables

We turn now to a quite different, much more systematic approach to the solution Laplace's equation

$$\nabla^2\Phi(\mathbf{x}) = 0 \tag{37}$$

as a boundary value problem. It is implemented by expanding the solution in some domain V using complete sets of orthogonal functions

$$\Phi(\eta, \xi, \nu) = \sum_{nlm} A_n(\eta) B_l(\xi) C_m(\nu) \tag{38}$$

and determining the coefficients in the expansion by requiring that the solution take on the proper values on the boundaries. For simple geometries for which Laplace's equation separates (spheres, cylinders, rectangular parallelepipeds) this method can always be utilized². Before launching into a description of how one proceeds in specific cases (or geometries), let us take a few minutes to review the terminology of orthogonal function expansions and some basic facts.

Suppose that we have a set of functions $U_n(\eta)$, $n = 1, 2, \dots$ which are orthogonal on the interval $a \leq \eta \leq b$, by which we mean that

$$\int_a^b d\eta U_n^*(\eta) U_m(\eta) = 0, \text{ if } m \neq n; \tag{39}$$

²It can also be very tedious.

the superscript * denotes complex conjugation. Further, the functions $U_n(\eta)$ are normalized on the interval,

$$\int_a^b d\eta U_n^*(\eta)U_n(\eta) = \int_a^b d\eta |U_n(\eta)|^2 = 1. \quad (40)$$

Combining these equations we have

$$\int_a^b d\eta U_n^*(\eta)U_m(\eta) = \begin{cases} 0, & n \neq m \\ 1, & n = m \end{cases} = \delta_{nm}. \quad (41)$$

The functions $U_n(\eta)$ are said to be *orthonormal*; δ_{nm} is called a *Kronecker delta function*.

Next, we attempt to expand, on the interval $a \leq \eta \leq b$, an arbitrary function $f(\eta)$ as a linear combination of the functions $U_n(\eta)$, which are referred to as *basis functions*. Keeping just N terms in the expansion, one has

$$f(\eta) \approx \sum_{n=1}^N a_n U_n(\eta). \quad (42)$$

We need a criterion for choosing the coefficients in the expansion; a standard criterion is to minimize the mean square error E which may be defined as follows:

$$\begin{aligned} E &= \int_a^b d\eta |f(\eta) - \sum_{n=1}^N a_n U_n(\eta)|^2 \\ &= \int_a^b d\eta \left(f^*(\eta) - \sum_{n=1}^N a_n^* U_n^*(\eta) \right) \left(f(\eta) - \sum_{m=1}^N a_m U_m(\eta) \right). \end{aligned} \quad (43)$$

The conditions for an extremum are

$$\left(\frac{\partial E}{\partial a_k} \right)_{a_k^*} = 0 = \left(\frac{\partial E}{\partial a_k^*} \right)_{a_k}. \quad (44)$$

where a_k and a_k^* have been treated as independent variables ³ Application of these conditions leads to

$$\begin{aligned} 0 &= \int_a^b d\eta \left(f^*(\eta) - \sum_{n=1}^N a_n^* U_n^*(\eta) \right) U_k(\eta) \\ &= \int_a^b d\eta \left(f(\eta) - \sum_{n=1}^N a_n U_n(\eta) \right) U_k^*(\eta) \end{aligned} \quad (45)$$

or, making use of the orthonormality of the basis functions,

$$a_k = \int_a^b d\eta f(\eta) U_k^*(\eta). \quad (46)$$

with a_n^* given by the complex conjugate of this relation. If the basis functions are orthogonal but not normalized, then one finds

$$a_k = \frac{\int_a^b d\eta f(\eta) U_k^*(\eta)}{\int_a^b d\eta |U_k(\eta)|^2}. \quad (47)$$

The set of basis functions $U_n(\eta)$ is said to be complete if the mean square error can be made arbitrarily small by keeping a sufficiently large number of terms in the sum. Then one says that the sum converges in the mean to the given function. If we are a bit careless, we can then write

$$\begin{aligned} f(\eta) &= \sum_n a_n U_n(\eta) = \sum_n \int_a^b d\eta' f(\eta') U_n^*(\eta') U_n(\eta) \\ &= \int_a^b d\eta' \left(\sum_n U_n(\eta) U_n^*(\eta') \right) f(\eta'), \end{aligned} \quad (48)$$

from which it is evident that

$$\sum_n U_n(\eta) U_n^*(\eta') = \delta(\eta - \eta') \quad (49)$$

³This is always possible, since a_k and a_k^* are linearly related to $\text{Re}(a_k)$ and $\text{Im}(a_k)$.

for a complete set of functions. This equation is called the *completeness* or *closure relation*.

We may easily generalize to a space of arbitrary dimension. For example, in two dimensions we may have the space of η and ζ with $a \leq \eta \leq b$, and $c \leq \zeta \leq d$ and complete sets of orthonormal functions $U_n(\eta)$ and $V_m(\zeta)$ on the respective intervals. Then the arbitrary function $f(\eta, \zeta)$ has the expansion

$$f(\eta, \zeta) = \sum_{n,m} A_{nm} U_n(\eta) V_m(\zeta), \quad (50)$$

where

$$A_{nm} = \int_a^b d\eta \int_a^b d\zeta f(\eta, \zeta) U_n^*(\eta) V_m^*(\zeta). \quad (51)$$

3.1 Fourier Series

Returning to the one-dimensional case, suppose that the interval is infinite, $-\infty < \eta < \infty$. Then the index n of the functions $U_n(\eta)$ may become a continuous index, $U_n(\eta) \rightarrow U(\eta; \rho)$. A familiar example of this is the Fourier integral which is the limit of a Fourier series when the interval on which functions are expanded becomes infinite. Consider that we have the interval $-a/2 < \eta < a/2$. Then the Fourier series may be built from the basis functions

$$U_m(\eta) = \frac{1}{\sqrt{a}} e^{i2\pi m\eta/a}, \quad \text{with } m = 0, \pm 1, \pm 2, \dots; \quad (52)$$

these functions form a complete orthonormal set. The expansion of $f(\eta)$ is

$$f(\eta) = \frac{1}{\sqrt{a}} \sum_{m=-\infty}^{\infty} A_m e^{i2\pi m\eta/a} \quad (53)$$

with

$$A_m = \frac{1}{\sqrt{a}} \int_{-a/2}^{a/2} d\eta f(\eta) e^{-i2\pi m\eta/a}. \quad (54)$$

The closure relation is

$$\frac{1}{a} \sum_m e^{i2\pi m(\eta-\eta')/a} = \delta(\eta - \eta'). \quad (55)$$

Now define $k \equiv 2\pi m/a$ or $m = ka/2\pi$. Also, define $A_m = \sqrt{2\pi/a} A(k)$. Note that for $a \rightarrow \infty$, k takes on a set of values that approach a continuum. Thus

$$f(\eta) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} \frac{a}{2\pi} dk e^{ik\eta} \sqrt{\frac{2\pi}{a}} A(k) = \frac{1}{\sqrt{2\pi}} \int dk e^{ik\eta} A(k), \quad (56)$$

while

$$\sqrt{\frac{2\pi}{a}} A(k) = \frac{1}{\sqrt{a}} \int d\eta f(\eta) e^{-ik\eta} \quad (57)$$

or

$$A(k) = \frac{1}{\sqrt{2\pi}} \int d\eta f(\eta) e^{-ik\eta} \quad (58)$$

while the closure relation now reads

$$\frac{1}{2\pi} \int dk e^{ik(\eta-\eta')} = \delta(\eta - \eta'), \quad (59)$$

thus $e^{ik\eta}$ form a complete set (this is also a useful representation of the Dirac delta function).

Note that we can also write this equation as

$$\frac{1}{2\pi} \int d\eta e^{i\eta(k-k')} = \delta(k - k') \quad (60)$$

which is the orthonormalization expression of the complete set of functions $U(\eta; k)$ on the infinite η interval. These functions are

$$U(\eta; k) = \frac{1}{\sqrt{2\pi}} e^{i\eta k}. \quad (61)$$

3.2 Separation of Variables

We are going to attempt to find solutions to boundary value problems in three dimensions by writing the solution as a sum of products of three one-dimensional functions,

$$\Phi(\eta, \zeta, \nu) = \sum_{n,l,m} A_{nlm} E_n(\eta) Z_l(\zeta) N_m(\nu). \quad (62)$$

We will do this for the particular cases of rectangular, cylindrical, and spherical polar coordinates. Now, if the functions E , Z , and N are members of complete sets on appropriate intervals, we can certainly write any three-dimensional function as a linear combination of such products. Because we are looking for special three-dimensional functions, however, that is, solutions to the Laplace equation, we do not actually have to employ complete sets of functions of all three variables. To determine just what we do have to use, we will try to demand that each term in the sum is itself a solution to the Laplace equation, which is more restrictive than just requiring the sum to be a solution. It turns

out that this is possible in the Cartesian, cylindrical, and spherical coordinate systems and also in eight more (see Landau and Lifshitz for more information)! The simplification that takes place when one makes this *separation of variables* is that each of the functions of a single variable has to be a solution of a relatively simple second order ordinary differential equation rather than a partial differential equation.

3.3 Rectangular Coordinates

Let us look for a solution of Laplace's equation in the form of a product of functions of x , y , and z ,

$$\Phi(\mathbf{x}) = X(x)Y(y)Z(z). \quad (63)$$

Substitution into Laplace's equation $\nabla^2\phi(\mathbf{x}) = 0$ yields

$$Y(y)Z(z)\frac{d^2X(x)}{dx^2} + X(x)Z(z)\frac{d^2Y(y)}{dy^2} + X(x)Y(y)\frac{d^2Z(z)}{dz^2} = 0. \quad (64)$$

Dividing by Φ we find

$$\frac{1}{X}\frac{d^2X}{dx^2} + \frac{1}{Y}\frac{d^2Y}{dy^2} + \frac{1}{Z}\frac{d^2Z}{dz^2} = 0. \quad (65)$$

Each term on the LHS of this equation depends on a single variable; consequently, since the equation must remain true when any one variable is varied with the others held fixed, it must be the case that each term is a constant, independent of the variable. Since the three terms add to zero, at least one must be a positive constant, and at least one

must be a negative constant. Let us suppose that two are negative, and one, positive. Thus we have

$$\frac{1}{X} \frac{d^2 X}{dx^2} = -\alpha^2; \quad \frac{1}{Y} \frac{d^2 Y}{dy^2} = -\beta^2; \quad \frac{1}{Z} \frac{d^2 Z}{dz^2} = \gamma^2 = \alpha^2 + \beta^2. \quad (66)$$

For negative constants, the solutions are oscillatory; when they are positive, solutions are exponential:

$$Z(z) \sim e^{\pm\gamma z}; \quad X(x) \sim e^{\pm i\alpha x}; \quad Y(y) \sim e^{\pm i\beta y}, \quad (67)$$

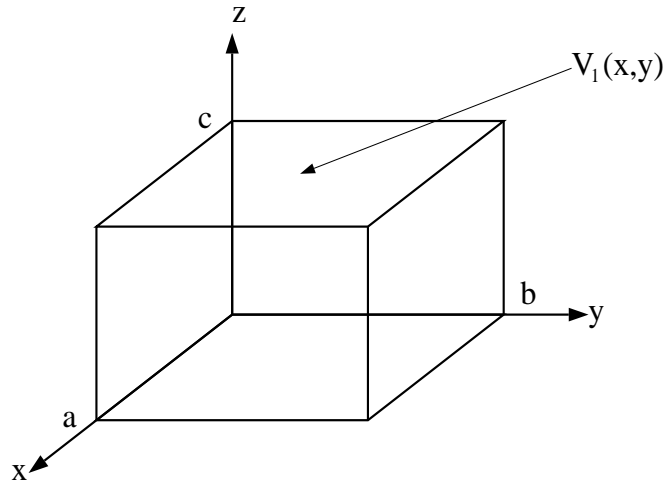
or, equivalently,

$$\begin{aligned} X(x) &\sim \sin(\alpha x) \text{ or } \cos(\alpha x) \\ Y(y) &\sim \sin(\beta y) \text{ or } \cos(\beta y) \\ Z(z) &\sim \sinh(\gamma z) \text{ or } \cosh(\gamma z). \end{aligned} \quad (68)$$

Now, α and β can be any real constants whatsoever, which means that by taking linear combinations of solutions of the kind outlined above, we can construct **any** function of x and y at some particular value of z .

This is just what we need to solve boundary value problems with planar surfaces. For example, suppose that we need to solve the Laplace equation inside of a rectangular parallelepiped of edge lengths a , b , and c with the potential given on the surface. We can find a solution by considering six distinct problems and superposing the six solutions to them. Each of these six problems has on one face (a different one in each problem) of the box the same potential as that given in the original

problem while on the other five faces the potential is zero. Summing the six solutions gives a potential which has the same values on each face of the box as given in the original problem. Let's see how to solve one of these six problems; the others follow trivially. For this problem we may suppose that the faces of the box are given by the planes $z = 0, c$, $x = 0, a$, and $y = 0, b$. Let the potential on the face $z = c$ be $\Phi(x, y, c) = V(x, y)$ while $\Phi(\mathbf{x}) \equiv 0$ on the other five faces.



$$\Phi(x, y, c) = V(x, y) \quad \Phi(\mathbf{x}) \equiv 0 \text{ on the other faces} \quad (69)$$

In order to satisfy the B.C., we must choose the constants α , β and γ so that we have a complete set of functions ⁴ on the face with the non-trivial boundary condition. Our expansion for the potential now

⁴As we have seen, the sinusoidal function form a complete set, the hyperbolic functions do not

takes the form

$$\Phi(x, y, z) = \sum_{\alpha, \beta} A_{\alpha\beta} \sinh(\gamma_{\alpha\beta} z) (\sin \alpha x \sin \beta y) \quad (70)$$

where α and β are such that $\alpha a = n\pi$ and $\beta b = m\pi$ which makes the basis functions of x and y orthogonal and complete on the domain of the constant- z face of the box. Thus,

$$\Phi(x, y, z) = \sum_{n,m=1}^{\infty} A_{nm} \sinh(\gamma_{nm} z) \sin\left(\frac{n\pi}{a}x\right) \sin\left(\frac{m\pi}{b}y\right) \quad (71)$$

where

$$\gamma_{nm} = \pi \left[\left(\frac{n}{a}\right)^2 + \left(\frac{m}{b}\right)^2 \right]^{1/2}. \quad (72)$$

Notice that only the sine functions are used and also only the hyperbolic sine. The reason for the latter is that the potential must vanish at $z = 0$; this condition rules out the use of the hyperbolic cosine which is not zero at zero argument. The cosine could be used but is not needed as the sine functions with arguments introduced above are complete on the appropriate intervals.

The coefficients in the sum for the solution are determined by looking at the potential on the top face of the box:

$$V(x, y) = \sum_{nm} A_{nm} \sinh(\gamma_{nm} c) \sin(n\pi x/a) \sin(m\pi y/b). \quad (73)$$

Multiply by $\sin(l\pi x/a) \sin(p\pi y/b)$ and integrate⁵ over the face of the box:

$$\int_0^a dx \int_0^b dy V(x, y) \sin(l\pi x/a) \sin(p\pi y/b) = A_{lp} \sinh(\gamma_{lp} c) \frac{1}{4} ab, \quad (74)$$

⁵Here we use the relation $\int_0^a \sin^2(n\pi x/a) = a/2$

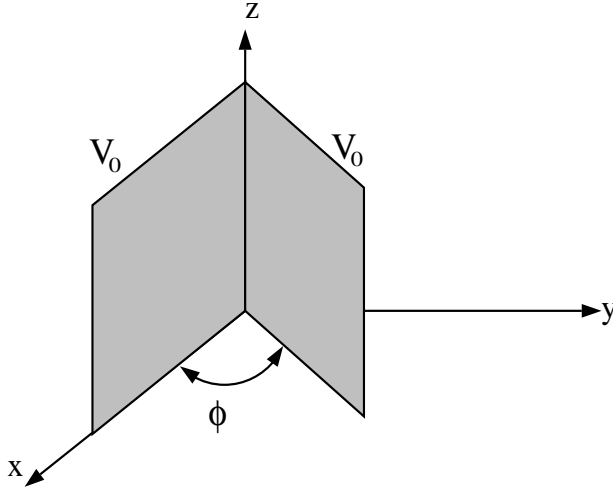
or

$$A_{lp} = \frac{4}{ab \sinh(\gamma_{lp}c)} \int_0^a dx \int_0^b dy V(x, y) \sin(l\pi x/a) \sin(p\pi y/b) \quad (75)$$

In this manner one can do any Dirichlet problem on a rectangular parallelepiped in the form of an infinite series.

3.4 Fields and Potentials on Edges

What we will never find very accurately from the expansion devised in the preceding section is the behavior of the potential and field close to an edge of the box where many terms must be kept to have decent convergence of the series. However, in these regions we may devise a quite different approximation which converges well very close to the edge. Suppose then that one is very close to such an edge where the boundary may be considered to consist of two infinite intersecting planes. Let the edge be coincident with the z-axis with the planes lying at constant values of the azimuthal angle ϕ .



The solution will then depend only on ϕ and ρ where $\rho = \sqrt{x^2 + y^2}$. In these variables, the Laplace equation is

$$\frac{1}{\rho} \frac{\partial}{\partial \rho} \left(\rho \frac{\partial \Phi}{\partial \rho} \right) + \frac{1}{\rho^2} \frac{\partial^2 \Phi}{\partial \phi^2} = 0. \quad (76)$$

Once again we use separation of variables, writing

$$\Phi(\rho, \phi) = R(\rho)\Psi(\phi). \quad (77)$$

Substitution into the Laplace equation and division by Φ itself yields the equation

$$\frac{1}{\rho R(\rho)} \frac{d}{d\rho} \left(\rho \frac{dR(\rho)}{d\rho} \right) + \frac{1}{\rho^2 \Psi(\phi)} \frac{d^2 \Psi(\phi)}{d\phi^2} = 0. \quad (78)$$

If we multiply by ρ^2 , we find that the first term on the LHS depends only on ρ and the second one depends only on ϕ ; consequently they must be equal and opposite constants,

$$\frac{\rho}{R} \frac{d}{d\rho} \left(\rho \frac{dR}{d\rho} \right) = C, \quad C = \text{constant}, \quad (79)$$

and

$$\frac{1}{\Psi} \frac{d^2 \Psi}{d\phi^2} = -C. \quad (80)$$

What must be the sign of C ? If $C > 0$, the $\Psi(\phi)$ is an oscillatory function while $R(\rho)$ is not. But if $C < 0$, then the converse is true. If our boundary value problem has Φ equal to some constant on the edges of a wedge with the surfaces of the wedge at $\phi = 0$ and $\phi = \beta$, then we will need to have an oscillatory $\Psi(\phi)$. Hence choose $C \geq 0$. Write $C \equiv \nu^2$, where ν is real. There is the special case when $\nu = 0$, for which $\Psi(\phi) = a + b\phi$ and $R(\nu) = c + d \ln \rho$. When $C > 0$, then

$$\Psi(\phi) = A \sin(\nu\phi) + B \cos(\nu\phi) = A' \sin(\nu\phi + \phi_0), \quad (81)$$

and $R(\rho)$ is the general solution of Eq. (105). Try $R = a\rho^p$; substitution into the differential equation gives

$$ap^2 \rho^{p-1} - a\nu^2 \rho^{p-1} = 0 \quad (82)$$

from which we find $p = \pm\nu$. The most general solution is

$$R(\rho) = a\rho^\nu + b\rho^{-\nu}, \quad (83)$$

and so a single term in the expansion for Φ is

$$\Phi(\rho, \phi) = (A\rho^\nu + B\rho^{-\nu}) \sin(\nu\phi + \phi_0) \quad (84)$$

where A , B , and ϕ_0 are constants to be determined by some boundary conditions.

There is also still the question of allowed values of ν . Let us specify that on the sides of the wedge, $\Phi(\rho, 0) = \Phi(\rho, \beta) = V_0$. To match this, we use $\nu = 0$ with $b = d = 0$ and $ac = V_0$. Then the boundary condition is matched on the edges of the wedge. Further, we must pick $\nu \neq 0$ (and ϕ_0) so that

$$\sin(0 + \phi_0) = 0 \quad (85)$$

and

$$\sin(\nu\beta + \phi_0) = 0; \quad (86)$$

we can easily see that $\phi_0 = 0$ and $\nu\beta = n\pi$, $n = 1, 2, \dots$ will work. Now add up solutions of the kind generated, that is, solutions with different values of n and undetermined coefficients, to find the most general solution (of this kind),

$$\Phi(\rho, \phi) = V_0 + \sum_{n=1}^{\infty} (A_n \rho^{n\pi/\beta} + B_n \rho^{-n\pi/\beta}) \sin\left(\frac{n\pi\phi}{\beta}\right). \quad (87)$$

where the constant term V_0 corresponds to $\nu = 0$.

If the physical region includes the origin ($\rho \rightarrow 0$), then we cannot have any negative powers of ρ because they will lead to singularities in Φ at the origin; physically, we know that that won't happen. Hence all B_n are zero (And that is also why we didn't keep the $\ln \rho$ part of the $\nu = 0$ solution). Thus we have

$$\Phi(\rho, \phi) = V_0 + \sum_{n=1}^{\infty} A_n \rho^{n\pi/\beta} \sin(n\pi\phi/\beta). \quad (88)$$

The remaining coefficients are determined by boundary conditions on a

surface that closes the system; for example a surface specified by $\rho = \rho_0$ for $0 \leq \phi \leq \beta$.

Without concerning ourselves with the details of fitting the expansion to such a function, we can still see what are the interesting qualitative features of the potential and fields for ρ very small, which means $\rho \ll \rho_0$. There the potential will be dominated by the term proportional to the smallest power of ρ , which is the $n = 1$ term,

$$\Phi(\rho, \phi) \approx V_0 + A_1 \rho^{\pi/\beta} \sin(\pi\phi/\beta), \text{ at small } \rho, \quad (89)$$

assuming that $A_1 \neq 0$. Taking appropriate derivatives of the potential, we may find the components of the electric field,

$$E_\rho = -\frac{\partial\Phi}{\partial\rho} = -\frac{\pi A_1}{\beta} \rho^{\pi/\beta-1} \sin(\pi\phi/\beta) \quad (90)$$

and

$$E_\phi = -\frac{1}{\rho} \frac{\partial\Phi}{\partial\phi} = -\frac{\pi A_1}{\beta} \rho^{\pi/\beta-1} \cos(\pi\phi/\beta). \quad (91)$$

Also, the charge density on the conductor close to the origin is found from

$$\sigma(\rho, 0) = \frac{E_\phi(\rho, 0)}{4\pi} = -\frac{A_1}{4\beta} \rho^{\pi/\beta-1} \quad (92)$$

and

$$\sigma(\rho, \beta) = -\frac{E_\phi(\rho, \beta)}{4\pi} = -\frac{A_1}{4\beta} \rho^{\pi/\beta-1}. \quad (93)$$

Depending on whether $\beta < \pi$ or $\beta > \pi$, one gets dramatically different fields and charge densities as $\rho \rightarrow 0$. For $\beta < \pi$, $\pi/\beta - 1 > 0$ and

fields and σ vanish as ρ goes to 0. But for $\beta > \pi$, $\pi/\beta - 1 < 0$, and consequently they become very large.

Of course, no real conductor has a perfectly sharp point; there is some rounding on a scale of length δ , leading to a maximum field of order

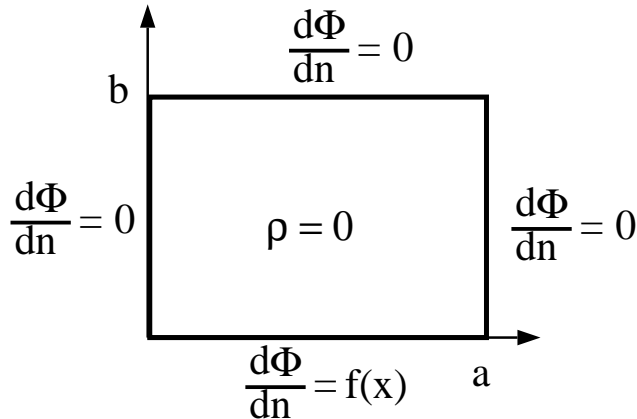
$$E_{max} \approx \frac{A_1}{4\beta} \delta^{\pi/\beta-1} \sim \frac{V_0}{R} \left(\frac{R}{\delta}\right)^{1-\pi/\beta} \sim E_0 \left(\frac{R}{\delta}\right)^{1-\pi/\beta}, \quad (94)$$

where R is the overall size of the system, that is, the distance from the point or wedge to ground. For a potential difference of, say 10^4 *statv*, $R = 1$ *km*, $\delta = 1$ *mm*, and $\beta = 2\pi$, we have $E_{max} \sim 30$ *statv/cm* or 9000 *v/cm*.

4 Examples

4.1 Two-dimensional box with Neumann boundaries

Consider the following 2-dimensional boundary value problem.



Find $\Phi(\mathbf{x})$ inside the rectangle (Note that due to the Neumann B.C. $\Phi(\mathbf{x})$ can only be determined up to an arbitrary additive constant). Show that we must have $\int_0^a dx f(x) = 0$.

This problem is very similar to that discussed in Sec. III.C. The difference is that this is in 2-d instead of 3-d, and has Neumann rather than Dirichlet B.C. Thus, we search for solutions of

$$\nabla^2 \Phi(x, y) = 0$$

in the form

$$\Phi(\mathbf{x}) = X(x)Y(y)$$

subject to the boundary conditions indicated above. Combining these equations yields

$$\frac{1}{X} \frac{d^2 X}{dx^2} + \frac{1}{Y} \frac{d^2 Y}{dy^2} = 0.$$

As in class, for arbitrary x and y , the only way to satisfy this equation is for both parts to be constant. Thus

$$\frac{1}{X} \frac{d^2 X}{dx^2} = -\alpha^2; \quad \frac{1}{Y} \frac{d^2 Y}{dy^2} = \alpha^2.$$

We choose this sign convention for the constant α so that we can easily satisfy the B.C.. Thus the solutions take the form

$$X(x) = A \sin(\alpha x) + B \cos(\alpha x); \quad Y(y) = C \sinh\{\alpha(b-y)\} + D \cosh\{\alpha(b-y)\}$$

Here we choose the form $\sinh\{\alpha(b-y)\}$ rather than $\sinh\{\alpha y\}$ with an eye toward satisfying the B.C. easily.

We can eliminate some of these coefficients by imposing the simple B.C.

$$\left. \frac{dX}{dx} \right|_{x=0} = \left. \frac{dX}{dx} \right|_{x=a} = 0$$

Clearly, to satisfy the first of these $A = 0$, and to satisfy the second $\alpha = n\pi/a$. Thus

$$X_n(x) = \cos(n\pi x/a)$$

Similarly, $\left. \frac{dY}{dy} \right|_{y=b} = 0$ indicates that $C = 0$. Thus

$$\Phi(x, y) = \sum_{n=0}^{\infty} a_n \cos(n\pi x/a) \cosh\left(\frac{n\pi}{a}(b-y)\right)$$

The set $\{a_n\}$ are determined by the remaining B.C.

$$\frac{\partial \Phi}{\partial n} = - \left. \frac{\partial \Phi}{\partial y} \right|_{y=0} = f(x)$$

$$f(x) = \sum_n \left\{ \frac{n\pi}{a} \sinh(n\pi b/a) a_n \right\} \cos(n\pi x/a)$$

or if we identify $b_n = \frac{n\pi}{a} \sinh(n\pi b/a) a_n$

$$f(x) = \sum_n b_n \cos(n\pi x/a)$$

Now if $f(x)$ is a regular function defined on the interval $0 < x < a$, then it may be represented as a cosine sum over all terms. However, the sum above is incomplete since it does not include the b_0 term. Thus we can only solve the problem if $b_0 = 0$, or equivalently, $\int dx f(x) \propto b_0 = 0$. Physically, what does this mean? (Hint, consider Gauss' law in 2-d, and the fact that the rectangle encloses no charge since $\nabla^2 \Phi = 0$).

The remaining b_n may be determined in the usual way.

$$\int_0^a dx f(x) \cos(l\pi x/a) = \sum_{n=1}^{\infty} b_n \int_0^a dx \cos(n\pi x/a) \cos(l\pi x/a)$$

then as

$$\int_0^a \cos(n\pi x/a) \cos(l\pi x/a) = \delta_{l,n}$$

and using the relation between a_l and b_l above we get

$$a_l = \frac{2}{l\pi \sinh(l\pi b/a)} \int_0^a dx f(x) \cos(l\pi x/a)$$

which define Φ through

$$\Phi(x, y) = \Phi_o + \sum_{n=1}^{\infty} a_n \cos(n\pi x/a) \cosh\left(\frac{n\pi}{a}(b-y)\right).$$

Now suppose that $f(x)$ is defined as $f(x) = E_o(1 - 2x/a)$.

$$a_l = \frac{2}{l\pi \sinh(l\pi b/a)} \int_0^a dx E_o(1 - 2x/a) \cos(l\pi x/a)$$

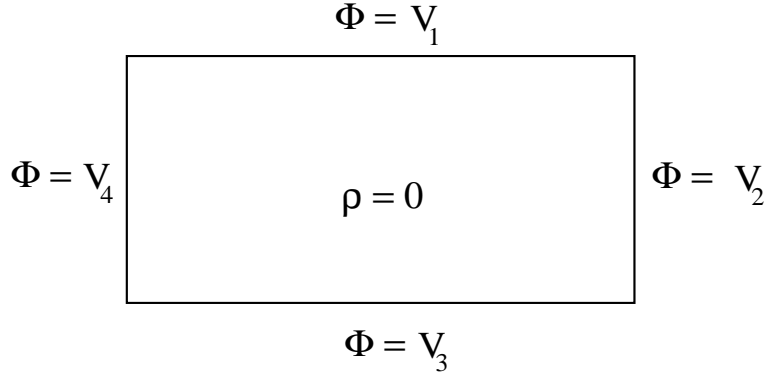
or, after integrating,

$$a_l = \frac{2}{l\pi \sinh(l\pi b/a)} \frac{2a}{l^2\pi^2} (1 - (-1)^l)$$

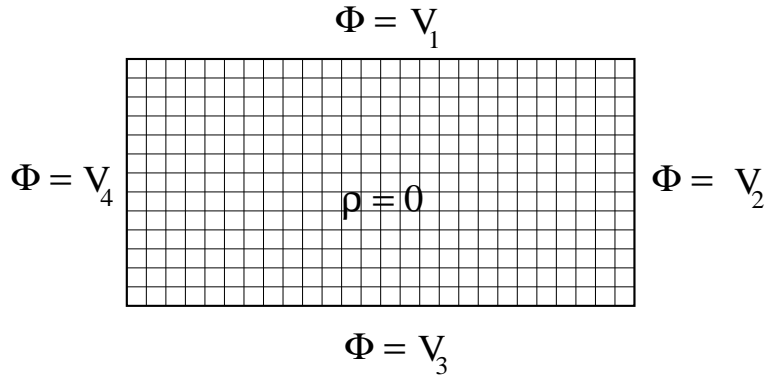
4.2 Numerical Solution of Laplace's Equation

As we discussed earlier, it is possible to solve Laplace's equation through separation of variables and special functions only for a restricted set of problems with separable geometries. When this is not the case, i.e. when the bounding surfaces, or charge distribution (when solving Poisson's equation) involve complex geometries, we generally solve for the potential numerically.

To illustrate how this is done, consider the following (exactly solvable) problem of a two-dimensional box with Dirichlet boundary conditions.



In order to solve Laplace's equation numerically with these boundary conditions, we will introduce a regular grid of width δ and dimensions $N \times M$ in the xy plane.



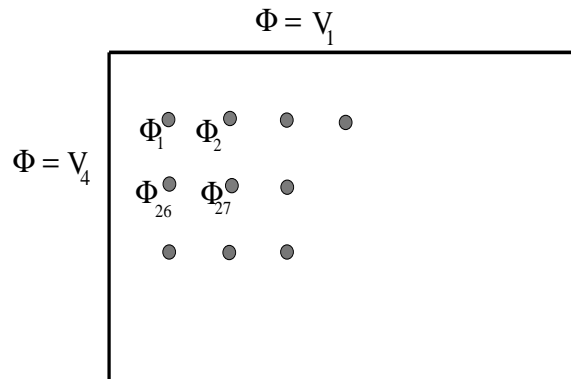
It is then a simple matter to approximately solve

$$\begin{aligned} \nabla^2 \Phi(\mathbf{x}) &= \frac{\partial^2 \Phi(\mathbf{x})}{\partial x^2} + \frac{\partial^2 \Phi(\mathbf{x})}{\partial y^2} = 0 \\ &\approx \frac{\Phi(x + \delta, y) - 2\Phi(x, y) + \Phi(x - \delta, y)}{\delta^2} + \frac{\Phi(x, y + \delta) - 2\Phi(x, y) + \Phi(x, y - \delta)}{\delta^2} \end{aligned} \quad (9)$$

or, equivalently,

$$\Phi(x + \delta, y) + \Phi(x - \delta, y) + \Phi(x, y + \delta) + \Phi(x, y - \delta) - 4\Phi(x, y) = 0 \quad (96)$$

Since Laplace's equation is now written as a finite difference equation of the potentials, it is a simple matter to enforce the boundary conditions by replacing each occurrence of $\Phi(\mathbf{x})$ above, when \mathbf{x} is the location of a boundary, by the appropriate boundary value. For example, consider a point in the upper left-hand quarter of the rectangle. This is illustrated below, where the potential at each internal grid point is denoted by Φ_n where integers n index each grid point.



At the point $n = 1$, the finite-difference form of Laplace's equation is

$$\Phi_2 + V_4 + V_1 + \Phi_{26} - 4\Phi_1 = 0$$

or

$$\Phi_2 + \Phi_{26} - 4\Phi_1 = -V_4 - V_1.$$

In fact, we obtain one such linear equation for each grid point. The equations are coupled of course in that each Φ_n occurs in at least two other equations. Thus, we have a set of $N \times M$ linear equations to solve

for the potential. This is done numerically by a variety of methods (see the homework). This method is easily extended to handle irregular boundaries, a mixture of Dirichlet and Neumann boundary conditions, as well as finite (non-singular) charge densities in the solution of Poisson's equation.

4.3 Derivation of Eq. 35: A Mathematica Session

In section II, we explore the problem of the potential between two hemispheres maintained at opposite potentials. Using the Dirichlet Greens function method, we had reduced the problem to quadrature.

$$\Phi(\epsilon, \theta) = \frac{V}{4\pi}(1-\epsilon^2) \int_0^{2\pi} d\phi' \left[\int_0^{\pi/2} \frac{\sin \theta' d\theta'}{(1 + \epsilon^2 - 2\epsilon \cos \gamma)^{3/2}} - \int_{\pi/2}^{\pi} \frac{\sin \theta' d\theta'}{(1 + \epsilon^2 - 2\epsilon \cos \gamma)^{3/2}} \right] \quad (97)$$

The integral is still difficult in the general case; however, alternative approach, valid for $\epsilon = r/a \ll 1$, is to expand the integrand in powers of ϵ and then to complete the integration term by term. This is straightforward but tedious and generates a solution in powers of ϵ . This is quite tedious to perform by hand, but is straightforward with a symbolic manipulator like Mathematica.

```
In[1]:= integrand= Sin[thetap]/(1+ep^2-2 ep Cos[gamma])^(3/2)
```

```
Sin[thetap]
```

Out[1]= -----

$$\frac{1}{(1 + \epsilon^2 - 2 \epsilon \cos[\gamma])^{3/2}}$$

In[2]:= integrand=integrand/.Cos[gamma]->Cos[theta] Cos[thetap] +
 Sin[theta] Sin[thetap] Cos[phip]

Out[2]= Sin[thetap] /

$$\frac{1}{(1 + \epsilon^2 - 2 \epsilon (\cos[\theta] \cos[\text{thetap}] + \cos[\text{phip}] \sin[\theta] \sin[\text{thetap}]))^{3/2}}$$

In[3]:= In[3]:= integrand1=Series[integrand,{ep,0,6}];

In[4]:= integrand1=Simplify[integrand1];

In[5]:= integrand1=Expand[integrand1];

In[6]:= answer1=Integrate[integrand1,{thetap,0,Pi/2}]-
 Integrate[integrand1,{thetap,Pi/2,Pi}];

In[7]:= answer1=Simplify[answer1];

In[8]:= answer2=Simplify[Integrate[answer1,{phi,0,2 Pi}]]

Out[8]=
$$6 \text{ Pi Cos}[\text{theta}] \text{ ep} + \frac{5 \text{ Pi (15 Cos}[\text{theta}] - 7 \text{ Cos}[3 \text{ theta}]) \text{ ep}^3}{16}$$

$$> \frac{21 \text{ Pi (130 Cos}[\text{theta}] - 35 \text{ Cos}[3 \text{ theta}] + 33 \text{ Cos}[5 \text{ theta}]) \text{ ep}^5}{512} + 0$$

In[9]:= Legrules={Cos[5x_]-> 16 (Cos[x])^5 - 20 (Cos[x])^3 + 5 Cos[x],
Cos[3x_]-> 4 (Cos[x])^3 - Cos[x] }

Out[9]= {Cos[5 (x_)] -> 5 Cos[x]^3 - 20 Cos[x]^5 + 16 Cos[x]^5 ,
 > Cos[3 (x_)] -> -Cos[x]^3 + 4 Cos[x]^3 }

In[10]:= answer2= Simplify[answer2/.Legrules]

$$\text{Out}[10]= 6 \text{ Pi Cos}[\text{theta}] \text{ ep} + \frac{5 \text{ Pi (Cos}[\text{theta}] - 7 \text{ Cos}[3 \text{ theta}]) \text{ ep}^3}{16} +$$

$$> \frac{21 \text{ Pi (60 Cos}[\text{theta}] - 35 \text{ Cos}[3 \text{ theta}] + 33 \text{ Cos}[5 \text{ theta}]) \text{ ep}^5}{512} + 0[$$

In[11]:= answer2= Simplify[answer2 V (1-ep^2)/(4 Pi)]

$$\text{Out}[11]= \frac{3 \text{ V Cos}[\text{theta}] \text{ ep}}{2} - \frac{7 \text{ V (13 Cos}[\text{theta}] + 5 \text{ Cos}[3 \text{ theta}]) \text{ ep}^3}{64} +$$

$$> \frac{11 \text{ V (100 Cos}[\text{theta}] + 35 \text{ Cos}[3 \text{ theta}] + 63 \text{ Cos}[5 \text{ theta}]) \text{ ep}^5}{512} + 0[$$

which is the answer we found, Eq. (35).

Boundary Value Problems in Electrostatics II

Friedrich Wilhelm Bessel

(1784 - 1846)

December 23, 2000

Contents

1	Laplace Equation in Spherical Coordinates	2
1.1	Legendre Equation and Polynomials	4
1.2	Solution of Boundary Value Problems with Azimuthal Symmetry . . .	12
1.2.1	Example: A Sphere With a Specified Potential	13
1.2.2	Example: Hemispheres of Opposite Potential	14
1.2.3	Example: Potential of an Isolated Charge	17
1.3	Behavior of Fields in Conical Holes and Near Sharp Points	19
1.4	Associated Legendre Polynomials; Spherical Harmonics	23
1.5	The Addition Theorem	26
1.6	Expansion of the Green's Function in Spherical Harmonics	30
2	Laplace Equation in Cylindrical Coordinates; Bessel Functions	33
2.1	Example I	43
2.2	Example II	45
2.3	B.V.P. on Large Cylinders	47
2.4	Green's Function Expansion in Cylindrical Coordinates	48

Although this is a new chapter, we continue to do things begun in the previous chapter. In particular, the first topic is the separation of variable method in spherical polar coordinates.

1 Laplace Equation in Spherical Coordinates

The Laplacian operator in spherical coordinates is

$$\nabla^2 = \frac{1}{r} \frac{\partial^2}{\partial r^2} r + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \sin \theta \frac{\partial}{\partial \theta} + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2}{\partial \phi^2}. \quad (1)$$

This is also a coordinate system in which it is possible to find a solution in the form of a product of three functions of a single variable each:

$$\Phi(r, \theta, \phi) = R(r)P(\theta)Q(\phi) = U(r)P(\theta)Q(\phi)/r. \quad (2)$$

Operate on Φ with ∇^2 , and set the result equal to zero to find

$$\frac{PQ}{r} \frac{d^2U}{dr^2} + \frac{UQ}{r^2 \sin \theta} \frac{d}{d\theta} \left(\sin \theta \frac{dP}{d\theta} \right) + \frac{UP}{r^3 \sin^2 \theta} \frac{d^2Q}{d\phi^2} = 0 \quad (3)$$

Multiply by $r^3 \sin^2 \theta / UPQ$ to find

$$\frac{r^2 \sin^2 \theta}{U} \frac{d^2U}{dr^2} + \frac{\sin \theta}{P} \frac{d}{d\theta} \left(\sin \theta \frac{dP}{d\theta} \right) + \frac{1}{Q} \frac{d^2Q}{d\phi^2} = 0. \quad (4)$$

The first two terms are independent of ϕ while the third depends only on this variable. Thus the third must be a constant as must the sum of the first two; the first of these conditions is

$$\frac{1}{Q} \frac{d^2Q}{d\phi^2} = C \text{ or } \frac{d^2Q}{d\phi^2} = CQ, \quad (5)$$

from which it follows that $Q \sim e^{\sqrt{C}\phi}$.

Now, a change in ϕ by 2π corresponds to no change whatsoever in spatial position; therefore, we must have $Q(\phi + 2\pi) = Q(\phi)$ because a function describing a measurable quantity must be a single-valued function of position. Hence we can conclude that $\sqrt{C} = im$ where m is an integer so that $e^{im2\pi} = 1$. Thus $C = -m^2$, and $Q(\phi) \rightarrow Q_m(\phi) = e^{im\phi}$, with $m = 0, \pm 1, \pm 2, \dots$. We recognize that the functions Q_m can be used to construct a Fourier series and are a complete orthogonal set on the interval $\phi_0 \leq \phi \leq \phi + 2\pi$.

Returning now to Laplace's equation, Eq. (4), and using $-m^2$ for $\frac{1}{Q} \frac{d^2 Q}{d\phi^2}$, we find

$$\frac{r^2 \sin^2 \theta}{U} \frac{d^2 U}{dr^2} + \frac{\sin \theta}{P} \frac{d}{d\theta} \left(\sin \theta \frac{dP}{d\theta} \right) - m^2 = 0 \quad (6)$$

or

$$\frac{r^2}{U} \frac{d^2 U}{dr^2} + \frac{1}{P \sin \theta} \frac{d}{d\theta} \left(\sin \theta \frac{dP}{d\theta} \right) - \frac{1}{\sin^2 \theta} m^2 = 0. \quad (7)$$

In this expression we recognize that the first term depends only on r and next two, only on θ , so we as usual conclude that each term must be separately a constant and that the constants must add to zero. The first equation extracted by this device is

$$\frac{d^2 U}{dr^2} = \frac{A}{r^2} U \quad (8)$$

where A is the constant. It is a standard convention to write A as $l(l+1)$ which is still quite general if l is allowed to be complex. Thus

the preceding equation becomes

$$\frac{d^2U}{dr^2} = \frac{l(l+1)}{r^2}U. \quad (9)$$

The solutions of this ordinary, second-order, linear, differential equation are two in number and are $U \sim r^{l+1}$ and $U \sim 1/r^l$. Before commenting further on that, let us go on to the equation for $P(\theta)$.

1.1 Legendre Equation and Polynomials

Substitution of $l(l+1)$ for the first term in Eq. (7) produces

$$\frac{1}{\sin \theta} \frac{d}{d\theta} \left(\sin \theta \frac{dP}{d\theta} \right) + \left(l(l+1) - \frac{m^2}{\sin^2 \theta} \right) P = 0. \quad (10)$$

This is the *generalized Legendre Equation*; it is commonly written in terms of a different variable, namely $u \equiv \cos \theta$. Then one has

$$\frac{d}{d\theta} \left(\sin \theta \frac{dP}{d\theta} \right) = \frac{du}{d\theta} \frac{d}{du} \left(\sqrt{1-u^2} \frac{du}{d\theta} \frac{dP}{du} \right) = -\sqrt{1-u^2} \frac{d}{du} \left(-(1-u^2) \frac{dP}{du} \right) \quad (11)$$

and

$$l(l+1) - \frac{m^2}{\sin^2 \theta} = l(l+1) - \frac{m^2}{1-u^2}; \quad (12)$$

hence,

$$\frac{d}{du} \left((1-u^2) \frac{dP}{du} \right) + \left(l(l+1) - \frac{m^2}{1-u^2} \right) P = 0 \quad (13)$$

is the form of the generalized Legendre equation using u as the variable.

The interval of interest to us is $0 \leq \theta \leq \pi$ which is $-1 \leq u \leq 1$.

We shall discuss first the special case of $m = 0$ which corresponds to $Q(\phi) = 1$, or a system for which $\Phi(\mathbf{x})$ is independent of ϕ ; we shall

call such a potential *azimuthally invariant*; there are many interesting systems which are more or less of this type. The equation for P is

$$\frac{d}{du} \left((1 - u^2) \frac{dP}{du} \right) + l(l + 1)P = 0, \quad (14)$$

which is called the *Legendre equation*.

A standard procedure for solving this equation (and other similar second-order differential equations) is to assume that the solution can be written as a power series. Then there must be a smallest power α in the series, so we can write

$$P(u) = u^\alpha \sum_{j=0}^{\infty} a_j u^j = \sum_{j=0}^{\infty} a_j u^{j+\alpha} \quad (15)$$

from which we may evaluate the derivatives as

$$\frac{dP}{du} = \sum_{j=0}^{\infty} (j + \alpha) a_j u^{j+\alpha-1}, \quad (16)$$

$$\frac{d^2P}{du^2} = \sum_{j=0}^{\infty} (j + \alpha)(j + \alpha - 1) a_j u^{j+\alpha-2}, \quad (17)$$

and

$$-\frac{d}{du} u^2 \frac{dP}{du} = - \sum_{j=0}^{\infty} (j + \alpha)(j + \alpha + 1) a_j u^{j+\alpha}. \quad (18)$$

Substitution into the Legendre equation gives

$$\sum_{j=0}^{\infty} [(\alpha + j)(\alpha + j - 1) u^{\alpha+j-2} a_j - (\alpha + j)(\alpha + j + 1) u^{\alpha+j} a_j + l(l + 1) u^{\alpha+j} a_j] = 0, \quad (19)$$

or, if we shift the zero of j in each term so as to isolate individual powers of u ,

$$\alpha(\alpha - 1) u^{\alpha-2} a_0 + \alpha(\alpha + 1) a_1 u^{\alpha-1} +$$

$$\sum_{j=0}^{\infty} [(\alpha + j + 2)(\alpha + j + 1)a_{j+2} + (l(l + 1) - (\alpha + j)(\alpha + j + 1)) a_j] u^{\alpha+j} = 0 \quad (20)$$

The only way that this power series can vanish for all u on the interval is to have the coefficient of each power of u vanish separately.

Thus I will list the coefficients:

j	Coefficient of $u^{\alpha+j}$
-2	$a_0\alpha(\alpha - 1)$
-1	$a_1\alpha(\alpha + 1)$
$j \geq 0$	$[(\alpha + j + 2)(\alpha + j + 1)a_{j+2} + (l(l + 1) - (\alpha + j)(\alpha + j + 1)) a_j]$

The coefficient of the leading (smallest) power $j = -2$ is zero if $\alpha = 0$ or 1; $a_0 = 0$ is not an option because by definition the first term in the expansion has a nonvanishing coefficient. Thus we find at this juncture two possible allowed values of α .

$$\alpha = 0, 1 \quad (21)$$

In order that the coefficient of the next power $j = -1$ of u vanish we must have either $\alpha = 0$ or $a_1 = 0$ (we can also have both); we can't have $\alpha = -1$ because of our first condition. Finally, the condition that the coefficient of $u^{\alpha+j}$ vanish for $j \geq 0$ is

$$a_{j+2} = \frac{(\alpha + j)(\alpha + j + 1) - l(l + 1)}{(\alpha + j + 1)(\alpha + j + 2)} a_j. \quad (22)$$

This is call the *recurrence relation*.

Consider this relation when j is very large, much larger than 1. In this limit it simplifies to the statement that $a_{j+2} = a_j(1 + \mathcal{O}(1/j))$ which will produce a power series (for large powers) in the form of a sum of terms proportional to u^{2j} , all with the same coefficient. At $u \rightarrow 1$, this sum will not converge, and so P is singular at $u = 1$. This is not an allowed behavior for a solution to the Laplace equation, so we cannot have such a function representing the potential. What must therefore happen is that the series terminates which means that there must be some j such that $a_j \neq 0$ while $a_{j+2} = 0$. Examining Eq. (22), we see that this j is such that

$$(\alpha + j)(\alpha + j + 1) - l(l + 1) = 0. \quad (23)$$

This condition requires that $\alpha + j = l$ which is a condition on l ; since α is 0 or 1, and j is a non-negative integer, we see that l must be an integer equal to or larger than α .

$$l \in \mathcal{Z} \quad l \geq \alpha \quad (24)$$

Now, our recurrence relation gives us a_{j+2} from a_j ; hence, starting from a_0 , we can get only the even coefficients a_j , j even, and starting from a_1 , we get the odd coefficients. Lets consider the odd series and the even series separately. First consider the **even series**. Since at termination of the series $\alpha + j = l$, we see that l is even when $\alpha = 0$ and l is odd when $\alpha = 1$. Thus the even series terminates when

$$\alpha = 0 \text{ and } l \text{ even} \quad (25)$$

$$\alpha = 1 \text{ and } l \text{ odd}$$

By similar arguments applied to the **odd series**, we can see that it terminates when

$$\alpha = 0 \text{ and } l \text{ odd} \tag{26}$$

$$\alpha = 1 \text{ and } l \text{ even}$$

Since l cannot be both odd and even, we can only have an even or an odd series (they are actually equivalent). The other must be zero. Since by convention, we choose $a_0 \neq 0$, it must be that the odd series vanishes, and thus $a_1 = 0$. Remembering that l is odd when $\alpha = 1$ and is even when $\alpha = 0$, we see that the solutions are polynomials of degree l . They are known as *Legendre polynomials*. It is easy to generate a few of them, aside from normalization, starting from $l = 0$ and using the recurrence relation. As for normalization, they are traditionally chosen to be such that $P(1) = 1$. Let us add a subscript l to P to designate the particular Legendre polynomial. The first few are

l	$P_l(u)$
0	$P_0(u) = 1$
1	$P_1(u) = u$
2	$P_2(u) = \frac{3}{2}u^2 - \frac{1}{2}$
3	$P_3(u) = \frac{5}{2}u^3 - \frac{3}{2}u$
4	$P_4(u) = \frac{35}{8}u^4 - \frac{15}{4}u^2 + \frac{3}{8}$

There are other ways to generate the Legendre polynomials. For example, one has *Rodrigues' formula* which is

$$P_l(u) = \frac{1}{2^l l!} \frac{d^l}{du^l} (u^2 - 1)^l; \quad (27)$$

it is easy to see that this generates a polynomial of degree l ; one may show that it is a solution to the Legendre equation by direct substitution into that equation. Thus, it must be the Legendre polynomial (one should also check normalization). If one expands the factor $(u^2 - 1)^l$ in Rodrigues' formula using the binomial expansion and then takes the derivatives, she/he will find that

$$P_l(u) = \frac{(-1)^l}{2^l} \sum_{n \geq l/2}^l \frac{(-1)^n (2n)!}{n!(l-n)!(2n-l)!} u^{2n-l} \quad (28)$$

Another way of generating $P_l(u)$ is via the generating function

$$T(u, x) = (1 - 2ux + x^2)^{-1/2} = \sum_{l=0}^{\infty} x^l P_l(u). \quad (29)$$

If one takes l derivatives of this function with respect to x and then sets $x = 0$, the result is $l!P_l(u)$.

The Legendre functions have many properties that we will need to make use of from time to time. For summaries of these, see *e.g.*, , the section on Legendre functions in Abramowitz and Stegun starting on p. 332 and also the section on orthogonal polynomials starting on p. 771. Here we summarize some of the most significant properties. First, orthogonality and normalization. Consider the integral

$$\int_{-1}^1 du P_l(u) P_{l'}(u) = \frac{1}{2^{l+l'} l! l'!} \int_{-1}^1 du \frac{d^l}{du^l} (u^2 - 1)^l \frac{d^{l'}}{du^{l'}} (u^2 - 1)^{l'}. \quad (30)$$

Suppose, without loss of generality, that $l' \geq l$ and start by integrating by parts,

$$\int_{-1}^1 du P_l(u) P_{l'}(u) = \frac{1}{2^{l+l'} l! l'!} \left[\frac{d^l}{du^l} (u^2 - 1)^l \frac{d^{l'-1}}{du^{l'-1}} (u^2 - 1)^{l'} \Big|_{-1}^1 - \int_{-1}^1 du \frac{d^{l+1}}{du^{l+1}} (u^2 - 1)^l \frac{d^{l'-1}}{du^{l'-1}} (u^2 - 1)^{l'} \right] \quad (31)$$

The first term in brackets vanishes because $(u^2 - 1)$ is zero at the end points of the interval. Continuing from the right-hand side, integrate in like fashion $l' - 1$ more times. The result is

$$\int_{-1}^1 du P_l(u) P_{l'}(u) = (-1)^{l'} \frac{1}{2^{l+l'} l! l'!} \int_{-1}^1 du \frac{d^{l+l'} (u^2 - 1)^l}{du^{l+l'}} (u^2 - 1)^{l'}. \quad (32)$$

Now, $d^{l+l'} (u^2 - 1)^l / du^{l+l'} = 0$ if $l' > l$, and so the integral is zero in this case. If $l' = l$, we have

$$\begin{aligned} \int_{-1}^1 du P_l(u) P_l(u) &= (-1)^l \frac{(2l)!}{2^{2l} (l!)^2} \int_{-1}^1 du (u^2 - 1)^l \\ &= (-1)^l \frac{(2l)!}{2^{2l} (l!)^2} 2(-1)^l \frac{(2l)!!}{(2l+1)!!} = 2 \frac{(2^l l!)^2 (2l-1)!!}{2^{2l} (l!)^2 (2l+1)!!} = \frac{2}{2l+1}. \end{aligned} \quad (33)$$

Thus we have derived the relation

$$\int_{-1}^1 du P_l(u) P_{l'}(u) = \frac{2}{2l+1} \delta_{ll'} \quad (34)$$

which expresses the orthogonality and normalization of the Legendre polynomials.

Consider next recurrence relations. These provide, among other things, a good way to generate values of Legendre polynomials on

computers. A number of recurrence relations can be derived using Rodrigues' formula and the Legendre equation. Consider, for example,

$$\begin{aligned}
\frac{dP_{l+1}}{du} &= \frac{1}{2^{l+1}(l+1)!} \frac{d^{l+2}}{du^{l+2}} (u^2 - 1)^{l+1} \\
&= \frac{l+1}{2^l(l+1)!} \frac{d^{l+1}}{du^{l+1}} ((u^2 - 1)^l u) \\
&= \frac{1}{2^l l!} \frac{d^l}{du^l} [(u^2 - 1)^l + 2lu^2(u^2 - 1)^{l-1}] \\
&= \frac{1}{2^l l!} \frac{d^l}{du^l} [(u^2 - 1)^l + 2l(u^2 - 1)^l + 2l(u^2 - 1)^{l-1}] \\
&= (2l+1)P_l(u) + \frac{dP_{l-1}(u)}{du}, \tag{35}
\end{aligned}$$

or

$$\frac{dP_{l+1}(u)}{du} - (2l+1)P_l(u) - \frac{dP_{l-1}(u)}{du} = 0. \tag{36}$$

From this relation and the Legendre equation

$$\frac{d}{du} \left((1-u^2) \frac{dP_l}{du} \right) + l(l+1)P_l = 0 \tag{37}$$

one may derive additional standard recurrence relations for the Legendre polynomials. Several of these are

$$\begin{aligned}
(l+1)P_{l+1} - u(2l+1)P_l + lP_{l-1} &= 0 \\
(1-u^2) \frac{dP_l}{du} + luP_l - lP_{l-1} &= 0 \\
\frac{dP_{l+1}}{du} - u \frac{dP_l}{du} - (l+1)P_l &= 0. \tag{38}
\end{aligned}$$

These may be used to advantage in numerous applications such as doing integrals of products of two Legendre polynomials and a power of u .

1.2 Solution of Boundary Value Problems with Azimuthal Symmetry

Using what we have learned in the previous two sections, we are now in a position to construct a general solution to the Laplace equation in spherical coordinates under conditions of azimuthal invariance, that is, when $\Phi(\mathbf{x})$ is independent of ϕ . The most general form that a solution can have is

$$\Phi(r, \theta) = \sum_{l=0}^{\infty} (A_l r^l + B_l r^{-(l+1)}) P_l(\cos \theta). \quad (39)$$

The Legendre polynomials form a complete set on the interval $-1 \leq u \leq 1$ or $0 \leq \cos \theta \leq \pi$. Thus any specified ϕ -independent potential on a spherical surface can be expressed as a sum of P_l 's. If the volume in which a solution is to be found includes the origin, then none of the terms $\sim r^{-(l+1)}$ can be included in the sum as they are singular at the origin, and the potential will not be singular there. Similarly, if the volume extends to $r \rightarrow \infty$, then no terms $\sim r^l$ are allowed. In the former case, the conclusion is that $B_l = 0$ for all l , and in the latter case, all $A_l = 0$.

We consider now some examples.

1.2.1 Example: A Sphere With a Specified Potential

An isolated sphere of radius a is centered at the origin. By unspecified means, the potential on its surface is maintained at

$$\Phi(a, \theta, \phi) = V_o \cos^3(\theta)$$

where θ is the polar angle. Find $\Phi(r, \theta, \phi)$ for all $r > a$.

This problem is azimuthally symmetric. Thus, in general,

$$\Phi(r, \theta) = \sum_{l=0}^{\infty} (A_l r^l + B_l r^{-(l+1)}) P_l(\cos \theta).$$

Since our volume contains all $r > a$, physics demands that $A_l = 0$ for all l . The constants B_l are then determined by matching the terms in the series the boundary condition on the surface of the sphere. Recall that

$P_0(x)$	1
$P_1(x)$	x
$P_2(x)$	$\frac{1}{2}(3x^2 - 1)$
$P_3(x)$	$\frac{1}{2}(5x^3 - 3x)$
$P_4(x)$	$\frac{1}{8}(35x^4 - 30x^2 + 3)$
$P_5(x)$	$\frac{1}{8}(63x^5 - 70x^3 + 15x)$

So that,

$$\Phi(a, \theta, \phi) = V_o \cos^3(\theta) = V_o \left(\frac{2}{5} P_3(\cos(\theta)) + \frac{3}{5} P_1(\cos(\theta)) \right)$$

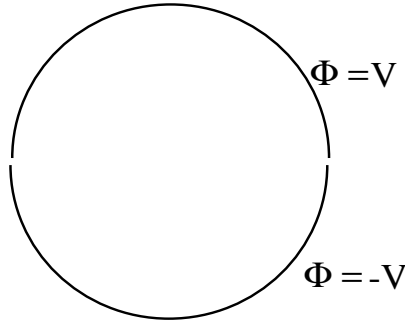
Thus

$$\Phi(r, \theta, \phi) = V_o \left(\frac{2}{5} \left(\frac{a}{r} \right)^4 P_3(\cos(\theta)) + \frac{3}{5} \left(\frac{a}{r} \right)^2 P_1(\cos(\theta)) \right).$$

1.2.2 Example: Hemispheres of Opposite Potential

For the first, suppose that we need to solve the Laplace equation inside of a sphere of radius a given that on the surface, the potential is specified as follows:

$$\Phi(a, \theta) = \begin{cases} V, & 0 \leq \theta \leq \pi/2 \\ -V, & \pi/2 \leq \theta \leq \pi. \end{cases} \quad (40)$$



Then the expansion must take the form

$$\Phi(r, \theta) = V \sum_{l=0}^{\infty} A_l \left(\frac{r}{a} \right)^l P_l(\cos \theta). \quad (41)$$

Notice the introduction of the factor V on the right-hand side, along with the use of the powers of a in the sum. These are included for convenience. The scale of the potential and hence the size of a leading term in the sum is set by V which also gives the correct dimensions to the terms in the sum; it is thus natural to put this factor in each

term. The powers of a are included for the same reasons; r is of order a and has the same dimensions so that leading coefficients A_l are of order unity and have dimension unity.

On the spherical surface, we have

$$\Phi(a, \theta) = V \sum_{l=0}^{\infty} A_l P_l(\cos \theta). \quad (42)$$

In order to find a given coefficient A_n , we multiply this equation by $P_n(\cos \theta)$ and integrate over $\cos \theta$, recalling that $d \cos \theta = -\sin \theta d\theta$. Making use of the orthogonality and normalization of the Legendre polynomials, we find

$$\int_0^\pi d\theta \Phi(a, \theta) P_n(\cos \theta) \sin \theta = V \sum_{l=0}^{\infty} A_l \left(\frac{2}{2l+1} \right) \delta_{ln} = V \left(\frac{2}{2n+1} \right) A_n, \quad (43)$$

or

$$\begin{aligned} A_n &= \frac{2n+1}{2} \left[\int_0^{\pi/2} d\theta P_n(\cos \theta) \sin \theta - \int_{\pi/2}^\pi d\theta P_n(\cos \theta) \sin \theta \right] \\ &= \frac{2n+1}{2} \left[\int_0^1 du P_n(u) - \int_{-1}^0 du P_n(u) \right] \end{aligned} \quad (44)$$

Now use the inversion property of the Legendre polynomials, $P_n(u) = (-1)^n P_n(-u)$ to conclude that

$$A_n = \frac{2n+1}{2} \left[\int_0^1 du P_n(u) - (-1)^n \int_0^1 du P_n(u) \right] = \begin{cases} 0 & n \text{ even} \\ (2n+1) \int_0^1 du P_n(u) & n \text{ odd.} \end{cases} \quad (45)$$

To complete the integral for the case of odd n we use a recurrence

relation

$$\frac{dP_{n+1}}{du} = (2n+1)P_n + \frac{dP_{n-1}}{du} \quad (46)$$

so that:

$$A_n = (2n+1) \int_0^1 du P_n(u) = \int_0^1 du \left[\frac{dP_{n+1}}{du} - \frac{dP_{n-1}}{du} \right] = P_{n-1}(0) - P_{n+1}(0) \quad (47)$$

where we make use of the fact that $P_n(1) = 1$, independent of n . Further, for even l ,

$$P_l(0) = (-1)^{l/2} \frac{(l-1)!!}{l!!} \quad (48)$$

where the “double factorial” sign means $l!! = l(l-2)(l-4)\dots(2 \text{ or } 1)$.

Hence

$$\begin{aligned} P_{n-1}(0) - P_{n+1}(0) &= (-1)^{(n-1)/2} \left[\frac{(n-2)!!}{(n-1)!!} + \frac{n!!}{(n+1)!!} \right] \\ &= (-1)^{(n-1)/2} \frac{(n-2)!!}{(n+1)!!} (n+1-n) = (-1)^{(n-1)/2} \frac{(n-2)!!}{(n+1)!!} (2n+1). \end{aligned} \quad (49)$$

Now set $n = 2m+1$, $m = 0, 1, 2, \dots$, and have

$$\Phi(r, \theta) = V \sum_{m=0}^{\infty} B_m \left(\frac{r}{a} \right)^{2m+1} P_{2m+1}(\cos \theta) \quad (50)$$

where

$$B_m = (-1)^m \frac{(2m-1)!!}{(2m+2)!!} (4m+3). \quad (51)$$

The first few terms in the expansion are

$$\Phi(r, \theta) = \frac{3}{2} V \left\{ \frac{r}{a} P_1(\cos \theta) - \frac{7}{12} \left(\frac{r}{a} \right)^3 P_3 + \frac{11}{24} \left(\frac{r}{a} \right)^5 P_5 - \frac{25}{64} \left(\frac{r}{a} \right)^7 P_7 + \dots \right\} \quad (52)$$

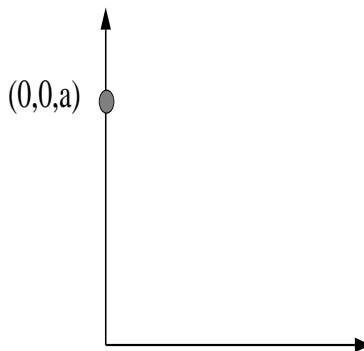
1.2.3 Example: Potential of an Isolated Charge

Another method of finding the coefficients in the expansion makes use of the fact that the expansion is unique. If, for example, we are able to find the potential at fixed $\theta = \theta_0$ for all r ,

$$\Phi(\theta_0, r) = g(r) = \sum_{l=0}^{\infty} (A_l r^l + B_l r^{-(l+1)}) P_l(\cos \theta_0) \quad (53)$$

then we can infer the form of the expansion by expanding $g(r)$ in powers of r and recognizing that the coefficient of r^l *must* be $P_l(\cos \theta_0)$ times a coefficient A_l while the coefficient of $(1/r)^{-(l+1)}$ must be $B_l P_l(\cos \theta_0)$. The most convenient value of θ_0 to use is certainly 0 or π since we know immediately the value of $P_l(\cos \theta)$ in these instances.

Consider the following specific example: Suppose that there is a charge q at a position $\mathbf{x} = a\hat{\mathbf{z}}$



in which case we know that the potential is

$$\Phi(r, \theta) = \frac{q}{\sqrt{r^2 + a^2 - 2ar \cos \theta}}. \quad (54)$$

For $\theta = 0$, we have simply $\phi(\mathbf{x}) = q/|r - a|$. At $r < a$ in particular,

this function has a simple power series expansion,

$$\Phi(r, 0) = \frac{q}{a-r} = \frac{q}{a} \frac{1}{1-r/a} = \frac{q}{a} \left\{ 1 + \frac{r}{a} + \frac{r^2}{a^2} + \frac{r^3}{a^3} + \dots \right\}. \quad r < a \quad (55)$$

Hence, associating P_l with $(r/a)^l$, we have

$$\Phi(r, \theta) = \frac{q}{a} \sum_{l=0}^{\infty} \left(\frac{r}{a}\right)^l P_l(\cos \theta); \quad r < a \quad (56)$$

the point is, the uniqueness of the expansion in terms of Legendre polynomials tells us that this must be the solution. A similar expansion done for $r > a$ yields

$$\Phi(r, \theta) = \frac{q}{r} \sum_{l=0}^{\infty} \left(\frac{a}{r}\right)^l P_l(\cos \theta) \quad r > a. \quad (57)$$

There are two points that are worth making in connection with these expansions. First, as stated earlier, there is a generating function $T(u, x)$ for the Legendre polynomials; see Eq. (29). We have just derived it; that is, it is Φ , Eq. (54), equal to the sum in Eq. (56). Also, we have obtained a convenient and useful expansion for the potential of a point charge; in more general notation, we have derived

$$\frac{1}{|\mathbf{x} - \mathbf{x}'|} = \sum_{l=0}^{\infty} \frac{r_{<}^l}{r_{>}^{l+1}} P_l(\cos \gamma) \quad (58)$$

where γ is the angle between \mathbf{x} and \mathbf{x}' while $r_{<}$ ($r_{>}$) is the smaller (larger) of $|\mathbf{x}|$ and $|\mathbf{x}'|$.

1.3 Behavior of Fields in Conical Holes and Near Sharp Points

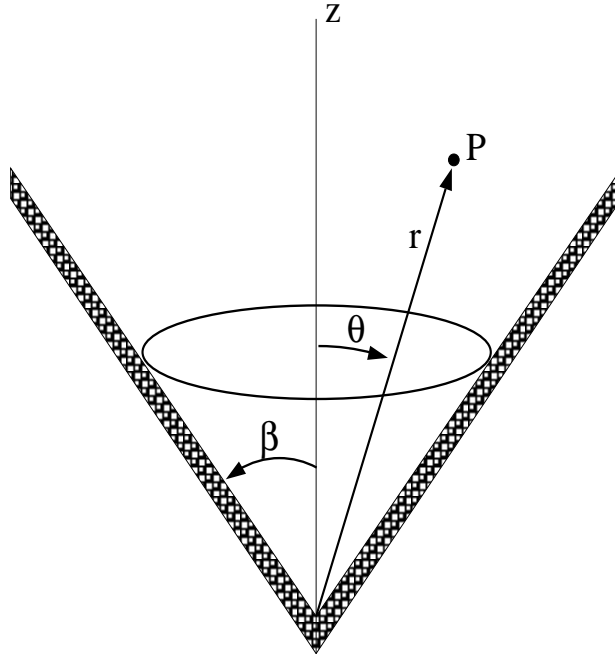
The field in the vicinity of the apex of a cone-shaped tip or depression can also be investigated using the separation of variables method in spherical coordinates. The solution for the potential is of the form, for r small enough,

$$\Phi(r, \theta) \sim r^\nu P_\nu(\cos \theta) \quad (59)$$

where $P_\nu(u)$ is a solution to the Legendre equation

$$\frac{d}{du}(1 - u^2)\frac{dP_\nu}{du} + \nu(\nu + 1)P_\nu = 0 \quad (60)$$

with ν to be determined.



For the geometry shown, the solution must be well-behaved as $\theta \rightarrow 0$, or $u = \cos \theta \rightarrow 1$, but not necessarily as $\theta \rightarrow \pi$ or $u = -1$. Introduce

the variable $y \equiv \frac{1}{2}(1 - u)$ or $u = 1 - 2y$; then Eq. (60) becomes

$$-\frac{1}{2} \frac{d}{dy} (1 - (1 - 2y)^2) \left(-\frac{1}{2} \frac{dP_\nu}{dy} \right) + \nu(\nu + 1)P_\nu = 0 \quad (61)$$

or

$$\frac{d}{dy} \left(y(1 - y) \frac{dP_\nu}{dy} \right) + \nu(\nu + 1)P_\nu = 0. \quad (62)$$

Let us look once again for a solution in the form of a power series expansion,

$$P_\nu = y^\alpha \sum_{j=0}^{\infty} a_j y^j, \quad (63)$$

with $0 \leq y \leq y_0 \leq 1$. Then

$$\frac{dP_\nu}{dy} = \sum_{j=0}^{\infty} (\alpha + j) a_j y^{\alpha+j-1}, \quad (64)$$

$$y(1 - y) \frac{dP_\nu}{dy} = \sum_{j=0}^{\infty} (\alpha + j) a_j (y^{\alpha+j} - y^{\alpha+j+1}), \quad (65)$$

and

$$\frac{d}{dy} \left(y(1 - y) \frac{dP_\nu}{dy} \right) = \sum_{j=0}^{\infty} a_j [(\alpha + j)y^{\alpha+j-1} - (\alpha + j + 1)y^{\alpha+j}] (\alpha + j). \quad (66)$$

Now combine these equations to find

$$\sum_{j=0}^{\infty} [a_j(\alpha + j)^2 y^{\alpha+j-1} + a_j((\alpha + j)(\alpha + j + 1) + \nu(\nu + 1)) y^{\alpha+j}] = 0 \quad (67)$$

or, isolating individual powers of y ,

$$a_0 \alpha^2 y^{\alpha-1} = 0 \quad (68)$$

which implies that $\alpha = 0$, and

$$a_{j+1} = a_j \frac{j(j+1) - \nu(\nu+1)}{(j+1)^2}. \quad (69)$$

If one lets $\nu = l$, a non-negative integer, the result is just the Legendre polynomials (no surprise), viewed as functions of y . More generally, for any real $\nu > 0$, one finds that the solutions are *Legendre functions of the first kind of order ν* .

$$P_\nu = \sum_{j=0}^{\infty} a_j(\nu) y^j, \quad (70)$$

These are well-behaved (that is, they are not singular) functions of y for $y < 1$ corresponding to $u > -1$ and are singular at $y = 1$.

For $1 > \nu > 0$, $P_\nu(y)$ has a single zero; for $2 > \nu > 1$, $P_\nu(y)$ has two zeroes, etc. This is important because if we have a cone of half-angle β with equipotential surfaces, we need ν to be such that $P_\nu((1 - \cos \beta)/2) = 0$. There will thus be a sequence of allowed values of ν , which we designate by ν_k , $k = 1, 2, 3, \dots$, which are such that $y_\beta \equiv \frac{1}{2}(1 - \cos \beta) \equiv k^{\text{th}}$ zero of P_ν .

The general solution at finite values of r , and including the point $r = 0$, is

$$\Phi(r, \theta) = \sum_{k=1}^{\infty} A_k r^{\nu_k} P_{\nu_k}(\cos \theta). \quad (71)$$

For small r , the leading term is the one with the smallest power of r , that is the $k = 1$ term. Hence we may approximate the sum sufficiently close to the origin by its leading term

$$\Phi(r, \theta) \approx A r^{\nu_1} P_{\nu_1}(\cos \theta). \quad (72)$$

The dominant contribution to the electric field in this region comes from this term; we have, by the usual $\mathbf{E}(\mathbf{x}) = -\nabla\Phi(\mathbf{x})$,

$$E_r = \frac{d\Phi}{dr} = -\nu_1 A r^{\nu_1-1} P_{\nu_1}(\cos\theta) \quad (73)$$

and

$$E_\theta = -\frac{1}{r} \frac{d\Phi}{d\theta} = A \sin\theta r^{\nu_1-1} \left. \frac{dP_{\nu_1}(u)}{du} \right|_{\cos\theta} \quad (74)$$

The behavior of ν_1 as a function of β is shown below. For β less than

about 0.8π , one has¹ $\nu_1 \approx \frac{2.405}{\beta} - \frac{1}{2}$, while for β larger than about the same number, $\nu_1 \approx \left[2 \ln\left(\frac{2}{\pi-\beta}\right)\right]^{-1}$. As $\beta \rightarrow \pi$, $\nu_1 \rightarrow 0$ and so $E_r \sim E_\theta \sim 1/r$ in this limit. The enhancement of a field near *e.g.*, a lightning rod is thus $\sim (R/\delta)$ if R is the size of the system and δ is the radius of curvature of the tip of the rod. Recall that in two dimensions

¹This relation comes from study of the properties of the Legendre functions.

we found an enhancement of order $(R/\delta)^{1/2}$. The enhancement is much more pronounced in three dimensions; a three dimensional tip is a much sharper thing than an edge.

1.4 Associated Legendre Polynomials; Spherical Harmonics

Let us now return to the more general case of a solution to Laplace's equation (i.e. a potential) which depends on the azimuthal angle ϕ . Then we must have the functions of ϕ $e^{im\phi}$, or, equivalently, $\sin m\phi$ and $\cos m\phi$, and the differential equation we have to face on the space of θ is

$$\frac{d}{du}(1-u^2)\frac{dP}{du} + \left(l(l+1) - \frac{m^2}{1-u^2} \right) P = 0. \quad (75)$$

The solutions are not finite polynomials in u in general but can be expressed as infinite power series. They are only "well-behaved" on the interval $-1 \leq u \leq 1$ when $l \geq |m|$, with l an integer. Then there is just one well-behaved solution which is known as the *associated Legendre function of degree l and order m* . For $m \geq 0$, the associated Legendre function can be written in terms of the Legendre polynomial of the same degree as

$$P_l^m(u) = (-1)^m (1-u^2)^{m/2} \frac{d^m}{dx^m} (P_l(u)); \quad (76)$$

one can read all about this in Abramowitz and Stegun on pages 332 to 353. Making use of Rodrigues' formula for the Legendre polynomials,

we see that

$$P_l^m(u) = (-1)^m \frac{(1-u^2)^{m/2}}{2^l l!} \frac{d^{l+m}}{du^{l+m}} [(u^2-1)^l]. \quad (77)$$

This last formula is also valid for negative m ²; comparing the two cases, one may see that

$$P_l^{-m}(u) = (-1)^m \frac{(l-m)!}{(l+m)!} P_l^m(u). \quad (78)$$

As for the Legendre polynomials, there is a generating function for the associated Legendre functions as well as a variety of recurrence relations. For example, a recurrence relation in degree is given by

$$(2l+1)uP_l^m(u) = (l-m+1)P_{l+1}^m(u) + (l+m)P_{l-1}^m(u) \quad (79)$$

and one in order is

$$P_l^{m+1} + \frac{2mu}{\sqrt{1-u^2}} P_l^m(u) + (l-m+1)(l+m)P_l^{m-1}(u) = 0 \quad (80)$$

Out of all of this, what is of importance to us is that the product

$$(Ar^l + Br^{-l-1})P_l^m(\cos\theta)e^{im\phi} \quad (81)$$

is a solution of the Laplace equation and that the set of functions $e^{im\phi}P_l^m(\cos\theta)$ with $l = 0, 1, 2, \dots$, and $m = -l, -l+1, \dots, l-1, l$ form a complete orthogonal set on the two-dimensional domain $0 \leq \theta \leq \pi$ and $0 \leq \phi \leq 2\pi$. As usual, completeness is difficult to demonstrate

²That's in part a matter of definition.

but orthogonality is quite straightforward using the formulae we have already written down. Consider the integral

$$I = \int_0^{2\pi} d\phi \int_0^\pi d\theta \sin \theta e^{-im\phi} P_l^m(\cos \theta) e^{im'\phi} P_{l'}^{m'}(\cos \theta) = 2\pi \delta_{mm'} \int_{-1}^1 du P_l^m(u) P_{l'}^m(u) \quad (82)$$

Assume $l' \geq l$, $m \geq 0$, and write $P_{l'}^m$ in terms of $P_{l'}^{-m}$:

$$\begin{aligned} I &= 2\pi \delta_{mm'} (-1)^m \frac{(l' + m)!}{(l' - m)!} \int_{-1}^1 du P_{l'}^{-m}(u) P_l^m(u) \\ &= 2\pi \delta_{mm'} (-1)^m \frac{(l' + m)!}{(l' - m)!} \frac{1}{2^{l+l'} l! l'!} \int_{-1}^1 du \frac{d^{l'-m}}{du^{l'-m}} (u^2 - 1)^{l'} \frac{d^{l+m}}{du^{l+m}} (u^2 - 1)^l \\ &= 2\pi \delta_{mm'} (-1)^{l'} \frac{(l' + m)!}{(l' - m)!} \frac{1}{2^{l+l'} l! l'!} \int_{-1}^1 du (u^2 - 1)^{l'} \frac{d^{l+l'}}{du^{l+l'}} (u^2 - 1)^l \\ &= 2\pi \delta_{ll'} \delta_{mm'} \frac{(l + m)! (2l)! (2l)!!}{(l - m)! 2^{2l} (l!)^2 (2l + 1)!!} 2 = \frac{4\pi}{2l + 1} \delta_{ll'} \delta_{mm'} \frac{(l + m)!}{(l - m)!} \end{aligned} \quad (83)$$

Thus we may construct an orthonormal set of functions on the surface of the unit sphere; these are called *spherical harmonics* and are defined as

$$Y_{l,m}(\theta, \phi) \equiv \sqrt{\frac{2l + 1}{4\pi} \frac{(l - m)!}{(l + m)!}} P_l^m(\cos \theta) e^{im\phi}, \quad (84)$$

with $m = -l, -l + 1, \dots, l - 1, l$ and $l = 0, 1, 2, \dots$. These functions have the property that

$$Y_{l,m}^*(\theta, \phi) = (-1)^m Y_{l,-m}(\theta, \phi). \quad (85)$$

The condition of orthonormality is

$$\int_0^{2\pi} d\phi \int_0^\pi \sin \theta d\theta Y_{l,m}^*(\theta, \phi) Y_{l',m'}(\theta, \phi) = \delta_{ll'} \delta_{mm'}. \quad (86)$$

The completeness relation (not derived as usual) is

$$\sum_{l=0}^{\infty} \sum_{m=-l}^m Y_{l,m}^*(\theta', \phi') Y_{l,m}(\theta, \phi) = \delta(\phi - \phi') \delta(\cos \theta - \cos \theta'). \quad (87)$$

A general function $g(\theta, \phi)$ is expanded in terms of the spherical harmonics as

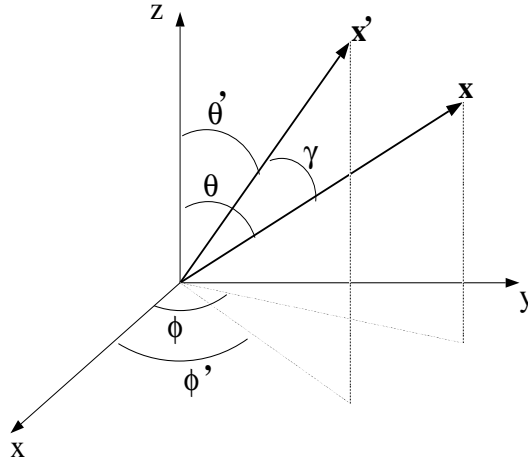
$$g(\theta, \phi) = \sum_{l,m} A_{lm} Y_{l,m}(\theta, \phi) \quad (88)$$

with

$$A_{lm} = \int_0^{2\pi} d\phi \int_0^\pi d\theta \sin \theta g(\theta, \phi) Y_{l,m}^*(\theta, \phi). \quad (89)$$

1.5 The Addition Theorem

In applications we will occasionally have need to know the function $P_l(\cos \gamma)$ where γ is the angle between two vectors \mathbf{x} and \mathbf{x}' ; it will prove to be useful to be able to write this function in terms of the variables $\theta, \phi, \theta',$ and ϕ' .

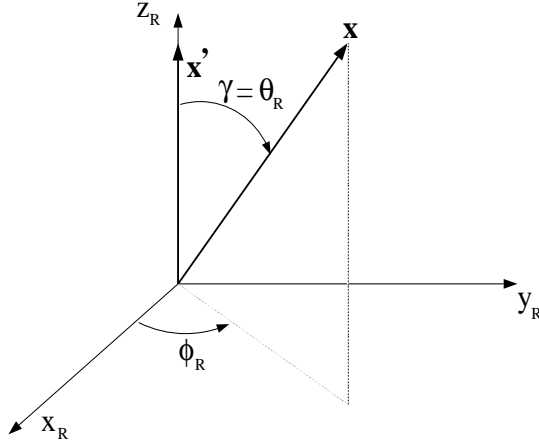


It must be possible to do so in terms of any complete sets of functions of these variables such as the spherical harmonics. In fact, the expansion

is

$$P_l(\cos \gamma) = \frac{4\pi}{2l+1} \sum_{m=-l}^l Y_{l,m}^*(\theta', \phi') Y_{l,m}(\theta, \phi). \quad (90)$$

We shall derive this expression as an example of the use of spherical harmonics and their properties. First, let us set up a second coordinate system rotated relative to the original one in such a way that its polar axis lies along the direction of \mathbf{x}' . In this system, the vector \mathbf{x} has components (r_R, θ_R, ϕ_R) .



Further, $\theta_R = \gamma$. Next, we may regard $P_l(\cos \gamma)$ as a function of θ and ϕ for fixed θ' and ϕ' and so can certainly expand it as

$$P_l(\cos \gamma) = \sum_{l'=0}^{\infty} \sum_{m'=-l}^l A_{l'm'}(\theta', \phi') Y_{l'm'}(\theta, \phi). \quad (91)$$

Similarly, in terms of spherical harmonics whose arguments are coordinates in the rotated system, it is easy to see that

$$P_l(\cos \gamma) = \sqrt{\frac{4\pi}{2l+1}} Y_{l,0}(\theta_R, \phi_R). \quad (92)$$

Now, the spherical harmonics satisfy the differential equation

$$\nabla^2 Y_{l,m}(\theta, \phi) + \frac{l(l+1)}{r^2} Y_{l,m}(\theta, \phi) = 0 \quad (93)$$

and they also satisfy this equation with variables θ_R, ϕ_R . But the Laplacian operator $\nabla^2 = \nabla \cdot \nabla$ is a scalar object which is invariant under coordinate rotations which means we can write it in the unrotated frame while writing the spherical harmonic in the rotated frame:

$$\nabla^2 Y_{l,m}(\theta_R, \phi_R) + \frac{l(l+1)}{r^2} Y_{l,m}(\theta_R, \phi_R) = 0. \quad (94)$$

Now recall that

$$Y_{l,0}(\theta_R, \phi_R) = \sqrt{\frac{2l+1}{4\pi}} P_l(\cos \gamma) = \sqrt{\frac{2l+1}{4\pi}} \sum_{l'=0}^{\infty} \sum_{m'=-l'}^{l'} A_{l'm'}(\theta', \phi') Y_{l'm'}(\theta, \phi). \quad (95)$$

If we plug this into the differential equation above, we obtain:

$$\sum_{l'=0}^{\infty} \sum_{m'=-l'}^{l'} A_{l'm'}(\theta', \phi') \left[\nabla^2 Y_{l'm'}(\theta, \phi) + \frac{l(l+1)}{r^2} Y_{l'm'}(\theta, \phi) \right] = 0 \quad (96)$$

or

$$\sum_{l'=0}^{\infty} \sum_{m'=-l'}^{l'} A_{l'm'}(\theta', \phi') \left[-\frac{l'(l'+1)}{r^2} + \frac{l(l+1)}{r^2} \right] Y_{l'm'}(\theta, \phi) = 0. \quad (97)$$

This equation can be true only if $l = l'$ **or** if $A_{l'm'} = 0$. Thus we have demonstrated that $A_{l'm'} = 0$ for $l' \neq l$, and the expansion of P_l reduces to

$$P_l(\cos \gamma) = \sum_{m=-l}^l A_{lm}(\theta', \phi') Y_{l,m}(\theta, \phi). \quad (98)$$

The coefficients in this expansion are found in the usual way for an orthogonal function expansion,

$$A_{lm} = \int d\Omega P_l(\cos \gamma) Y_{l,m}^*(\theta, \phi) = \int d\Omega_R P_l(\cos \theta_R) Y_{l,m}^*(\theta, \phi). \quad (99)$$

Following the same line of reasoning, we may express $\sqrt{4\pi/(2l+1)} Y_{l,m}^*(\theta, \phi)$ as a sum of the form

$$\sqrt{\frac{4\pi}{2l+1}} Y_{l,m}^*(\theta, \phi) = \sum_{m'=-l}^l B_{lm'}(m) Y_{lm'}(\theta_R, \phi_R) \quad (100)$$

where B_{l0} in particular is

$$\begin{aligned} B_{l0}(m) &= \int d\Omega_R \sqrt{\frac{4\pi}{2l+1}} Y_{l,m}^*(\theta, \phi) Y_{l,0}^*(\theta_R, \phi_R) \\ &= \int d\Omega_R Y_{l,m}^*(\theta, \phi) P_l(\cos \theta_R) \equiv A_{lm}. \end{aligned} \quad (101)$$

However, from Eq. (76), it is clear that when $u = 1$ $P_l^m(u) = 0$ when $m \neq 0$, and $P_l^m(u) = P_l(u)$ when $m = 0$, thus

$$\sqrt{\frac{4\pi}{2l+1}} Y_{l,m}^*(\theta, \phi)|_{\theta_R=0} = B_{l0}(m) \sqrt{\frac{2l+1}{4\pi}} P_l(1) = B_{l0}(m) \sqrt{\frac{2l+1}{4\pi}}, \quad (102)$$

so

$$A_{lm} = B_{l0}(m) = \frac{4\pi}{2l+1} Y_{l,m}^*(\theta, \phi)|_{\theta_R=0} = Y_{l,m}^*(\theta', \phi') \frac{4\pi}{2l+1}, \quad (103)$$

where the last step follow since when $\theta_R = 0$, $\mathbf{x} = \mathbf{x}'$. Thus,

$$P_l(\cos \gamma) = \frac{4\pi}{2l+1} \sum_{m=-l}^l Y_{l,m}^*(\theta', \phi') Y_{l,m}(\theta, \phi). \quad (104)$$

Thus ends our demonstration of the spherical harmonic addition theorem.

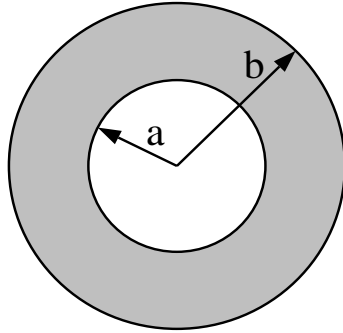
An application of this theorem is that we can write $\frac{1}{|\mathbf{x} - \mathbf{x}'|}$ as an expansion in spherical harmonics:

$$\frac{1}{|\mathbf{x} - \mathbf{x}'|} = \sum_{l=0}^{\infty} \frac{r_{<}^l}{r_{>}^{l+1}} P_l(\cos \gamma) = \sum_{l=0}^{\infty} \left\{ \frac{4\pi}{2l+1} \frac{r_{<}^l}{r_{>}^{l+1}} \sum_{m=-l}^l [Y_{l,m}^*(\theta', \phi') Y_{l,m}(\theta, \phi)] \right\}. \quad (105)$$

This expansion is often useful when faced with common integrals in electrostatics such as $\int d^3x' \rho(\mathbf{x}')/|\mathbf{x} - \mathbf{x}'|$.

1.6 Expansion of the Green's Function in Spherical Harmonics

More as an example of the use of the addition theorem than anything else, let us devise an expansion for the Dirichlet Green's function for the region V bounded by $r = a$ and $r = b$, $a < b$.



This function can be written as

$$G(\mathbf{x}, \mathbf{x}') = \frac{1}{|\mathbf{x} - \mathbf{x}'|} + F(\mathbf{x}, \mathbf{x}') \quad (106)$$

where $\nabla^2 F = 0$ in V . Thus it must be possible to write

$$G(\mathbf{x}, \mathbf{x}') = \sum_{l=0}^{\infty} \frac{4\pi}{2l+1} \frac{r_{<}^l}{r_{>}^{l+1}} \sum_{m=-l}^l Y_{l,m}^*(\theta', \phi') Y_{l,m}(\theta, \phi)$$

$$+ \sum_{l=0}^{\infty} \sum_{m=-l}^l \left(A_{lm} \frac{r^l}{b^{l+1}} + B_{lm} \frac{a^l}{r^{l+1}} \right) Y_{l,m}(\theta, \phi) \quad (107)$$

where A_{lm} and B_{lm} can be functions of \mathbf{x}' . The first term on the right-hand side is $\frac{1}{|\mathbf{x}-\mathbf{x}'|}$ and the second is a general solution of the Laplace equation. The coefficients are determined by requiring that the boundary conditions on $G(\mathbf{x}, \mathbf{x}')$ are satisfied ($G(\mathbf{x}, \mathbf{x}') = 0$ for \mathbf{x} on one of the two bounding spherical surfaces). At $r = a$ ($r_{<} = r = a$, $r_{>} = r'$), we have

$$0 = \sum_{l,m} \left\{ \frac{4\pi}{2l+1} \frac{a^l}{r'^{l+1}} Y_{l,m}^*(\theta', \phi') + A_{lm} \frac{a^l}{b^{l+1}} + B_{lm} \frac{1}{a} \right\} Y_{l,m}(\theta, \phi) = 0, \quad (108)$$

from which it follows, using the orthogonality of the spherical harmonics, that

$$\frac{4\pi}{2l+1} \frac{a^l}{r'^{l+1}} Y_{l,m}^*(\theta', \phi') + A_{lm} \frac{a^l}{b^{l+1}} + B_{lm} \frac{1}{a} = 0. \quad (109)$$

By similar means applied at $r = b$ ($r_{<} = r'$, $r_{>} = r = b$), one may show that

$$\frac{4\pi}{2l+1} \frac{r'^l}{b^{l+1}} Y_{l,m}^*(\theta', \phi') + A_{lm} \frac{1}{b} + B_{lm} \frac{a^l}{b^{l+1}} = 0. \quad (110)$$

These present us with two linear equations that may be solved for A_{lm} and B_{lm} ; the solutions are

$$A_{lm} = \frac{4\pi}{(2l+1)} Y_{l,m}^*(\theta', \phi') \left(\frac{a^{2l+1}}{b^l r'^{l+1}} - \frac{r'^l}{b^l} \right) / \left[1 - \left(\frac{a}{b} \right)^{2l+1} \right] \quad (111)$$

and

$$B_{lm} = \frac{4\pi}{2l+1} Y_{l,m}^*(\theta', \phi') \left(\frac{a^{l+1} r'^l}{b^{2l+1}} - \frac{a^{l+1}}{r'^{l+1}} \right) / \left[1 - \left(\frac{a}{b} \right)^{2l+1} \right]. \quad (112)$$

From these and the expansion Eq. (107), we find that the Green's function is

$$G(\mathbf{x}, \mathbf{x}') = \sum_{l,m} \frac{4\pi}{(2l+1) \left[1 - \left(\frac{a}{b}\right)^{2l+1}\right]} Y_{l,m}^*(\theta', \phi') Y_{l,m}(\theta, \phi) \left\{ \left[1 - \left(\frac{a}{b}\right)^{2l+1}\right] \frac{r_{<}^l}{r_{>}^{l+1}} + \left(\frac{a}{b}\right)^{2l+1} \frac{r^l}{r^{l+1}} - \frac{r'^l r^l}{b^{2l+1}} + \left(\frac{a}{b}\right)^{2l+1} \frac{r'^l}{r^{l+1}} - \frac{a^{2l+1}}{r^{l+1} r'^{l+1}} \right\} \quad (113)$$

This result, if we can call it that, can be written in a somewhat more compact form by factoring the quantity $\{\dots\}$. Suppose that $r_{>} = r'$ and $r_{<} = r$; then

$$\begin{aligned} \{\dots\} &= \left(r^l - \frac{a^{2l+1}}{r^{l+1}}\right) \left(\frac{1}{r'^{l+1}} - \frac{r'^l}{b^{2l+1}}\right) \\ &\equiv \left(r_{<}^l - \frac{a^{2l+1}}{r_{<}^{l+1}}\right) \left(\frac{1}{r_{>}^{l+1}} - \frac{r_{>}^l}{b^{2l+1}}\right). \end{aligned} \quad (114)$$

If, on the other hand, $r_{>} = r$ and $r_{<} = r'$, then

$$\{\dots\} = \left(r'^l - \frac{a^{2l+1}}{r'^{l+1}}\right) \left(\frac{1}{r^{l+1}} - \frac{r^l}{b^{2l+1}}\right) \equiv \left(r_{<}^l - \frac{a^{2l+1}}{r_{<}^{l+1}}\right) \left(\frac{1}{r_{>}^{l+1}} - \frac{r_{>}^l}{b^{2l+1}}\right). \quad (115)$$

Comparing these results, we see that we may in general write

$$G(\mathbf{x}, \mathbf{x}') = \sum_{l,m} \frac{4\pi/(2l+1)}{1 - \left(\frac{a}{b}\right)^{2l+1}} Y_{l,m}^*(\theta', \phi') Y_{l,m}(\theta, \phi) \left(r_{<}^l - \frac{a^{2l+1}}{r_{<}^{l+1}}\right) \left(\frac{1}{r_{>}^{l+1}} - \frac{r_{>}^l}{b^{2l+1}}\right). \quad (116)$$

Notice that the Green's functions for the interior of a sphere of radius b and for the exterior of a sphere of radius a are easily obtained by taking the limits $a \rightarrow 0$ and $b \rightarrow \infty$, respectively. In the former case,

for example, one finds

$$\begin{aligned}
G(\mathbf{x}, \mathbf{x}') &= \sum_{l,m} \frac{4\pi}{2l+1} Y_{l,m}^*(\theta', \phi') Y_{l,m}(\theta, \phi) \left(\frac{r_{\leq}^l}{r_{>}^{l+1}} - \frac{r_{\leq}^l}{b^{2l+1}} \right) \\
&= \frac{1}{|\mathbf{x} - \mathbf{x}'|} - \sum_{l,m} \frac{4\pi}{2l+1} Y_{l,m}^*(\theta', \phi') Y_{l,m}(\theta, \phi) \frac{b}{r'} \frac{r^l}{(b^2/r')^{l+1}} \\
&= \frac{1}{|\mathbf{x} - \mathbf{x}'|} - \frac{b/r'}{|\mathbf{x} - \mathbf{x}'_R|} \quad (117)
\end{aligned}$$

where $\mathbf{x}'_R = (b^2/r', \theta', \phi')$ in spherical coordinates.

2 Laplace Equation in Cylindrical Coordinates; Bessel Functions

In cylindrical coordinates the Laplacian is

$$\nabla^2 = \frac{\partial^2}{\partial \rho^2} + \frac{1}{\rho} \frac{\partial}{\partial \rho} + \frac{1}{\rho^2} \frac{\partial^2}{\partial \phi^2} + \frac{\partial^2}{\partial z^2}. \quad (118)$$

We once again look for solutions of the Laplace equation in the form of products of functions of a single variable,

$$\Phi(\mathbf{x}) = R(\rho)Q(\phi)Z(z); \quad (119)$$

Following the usual procedure (substitute into the Laplace equation; divide by appropriate functions to obtain terms which appear to depend on a single variable; argue that such terms must be constants; etc.), we wind up with the following three ordinary differential equations:

$$\frac{d^2 Q}{d\phi^2} + \nu^2 Q = 0, \quad \nu = 0, \pm 1, \pm 2, \dots, \quad (120)$$

$$\frac{d^2 Z}{dz^2} - k^2 Z = 0, \quad (121)$$

and

$$\frac{d^2 R}{d\rho^2} + \frac{1}{\rho} \frac{dR}{d\rho} + \left(k^2 - \frac{\nu^2}{\rho^2} \right) R = 0 \quad (122)$$

where the value of k is yet to be determined, and ν is determined as indicated by the same argument as in the case of spherical coordinates; the functions $Q(\phi)$ are the same as in spherical coordinates also, $Q(\phi) \sim e^{i\nu\phi}$.

The choice of k is specified by the sort of boundary conditions one has. One could imagine having to satisfy quite arbitrary conditions on an end face $z = c$ where c is constant; alternatively, one may have to fit some function on a side wall $\rho = c$. In the former case, one wants to have functions of ρ which form a complete set on an appropriate interval of ρ ; and in the latter case, one wants functions of z to form a complete set on some interval of z ; in both cases we will need a complete set of functions of ϕ , which we have. Now, looking at the equations for R and Z , we can see that the latter function in particular is going to be simple exponentials of kz ; for k real, these do not form a complete set; for k imaginary, they are sines and cosines and can form a complete set. We may not recognize it yet, but a similar thing happens to R ; for k imaginary, it is roughly exponential in character and we cannot get a complete set of functions in this way. But for k real, the functions R are oscillatory (although not sines and cosines) and can form a complete

set.

k	$Z(z)$	$R(\rho)$
real	incomplete ($e^{\pm kz}$)	complete (oscillatory)
imaginary	complete ($e^{i\pm k z}$)	incomplete

The functions of z in either case (k real or imaginary) are familiar to us and do not require further discussion. The functions of R , although probably known to all of us at least vaguely, are much less familiar so we will spend some time presenting their most important, to us, properties. Let's start by defining a dimension-free variable $x = k\rho$. Then Eq. (122) becomes

$$k^2 \frac{d^2 R}{dx^2} + k^2 \frac{1}{x} \frac{dR}{dx} + k^2 \left(1 - \frac{\nu^2}{x^2}\right) R = 0 \quad (123)$$

or

$$\frac{d^2 R}{dx^2} + \frac{1}{x} \frac{dR}{dx} + \left(1 - \frac{\nu^2}{x^2}\right) R = 0 \quad (124)$$

which is *Bessel's Equation*. Its solutions are *Bessel functions of order ν* . In our particular case, ν is an integer, although this need not be true in general. If k is imaginary, then x is imaginary, so we must deal with Bessel functions of imaginary argument; viewed as functions of a real variable $|x|$, these are known as *modified Bessel functions*.

For a given ν , there are two linearly independent solutions of Bessel's equation. Their choice is somewhat arbitrary since any linear combination of them is also a solution. One possible and common way to

choose them is as follows:

$$J_\nu(x) = \left(\frac{x}{2}\right)^\nu \sum_{j=0}^{\infty} \frac{(-1)^j (x/2)^{2j}}{j! \Gamma(j + \nu + 1)} \quad (125)$$

and

$$J_{-\nu}(x) = \left(\frac{x}{2}\right)^{-\nu} \sum_{j=0}^{\infty} \frac{(-1)^j (x/2)^{2j}}{j! \Gamma(j - \nu + 1)} \quad (126)$$

where

$$\Gamma(z) \equiv \int_0^{\infty} dt t^{z-1} e^{-t} \quad (127)$$

is the *gamma function* which, for z a real, positive integer n , is $\Gamma(n) = (n - 1)!$. For $z = 0$ or a negative integer, it is singular.

The two Bessel functions introduced above are linearly independent solutions of Bessel's equation so long as ν is not an integer. It is easy to verify that they are solutions by direct substitution into the differential equation. If, however, ν is an integer, they become identical (aside from a possible sign difference) and so do not provide us with everything we need in this case, which is the important one for us. Another function, which is a solution and which is linearly independent of either of the two solutions introduced above (taken one at a time) is given by

$$N_\nu(x) \equiv \frac{J_\nu(x) \cos(\nu\pi) - J_{-\nu}(x)}{\sin(\nu\pi)}. \quad (128)$$

This is the *Neumann function*; it is also called the Bessel function of the second kind of order ν .³

³Why does this work? Consider the limit $\nu \rightarrow m$ by La Hopital's rule. To formally show that N_m and J_m are independent, one must calculate the Wronskian and show that $W[N_m, J_m] \neq 0$. Note that in Abramowitz and Stegun's book, this function is written as $Y_\nu(x)$; see p. 358.

For $\nu = n$, a non-negative integer, the Neumann function has a series representation which is

$$N_n(x) \equiv -\frac{1}{\pi} \left(\frac{x}{2}\right)^{-n} \sum_{j=0}^{n-1} \frac{(n-j-1)!}{j!} \left(\frac{x}{2}\right)^{2j} + \frac{2}{\pi} \ln(x/2) J_n(x) - \frac{1}{\pi} \left(\frac{x}{2}\right)^n \sum_{j=0}^{\infty} \left\{ [\psi(j+1) + \psi(n+j+1)] \frac{(x/2)^{2j} (-1)^j}{j!(n+j)!} \right\} \quad (129)$$

where $\psi(y) = d(\ln \Gamma(y))/dy$ is known as the *digamma* or *psi function*.

Finally, for some purposes it is more useful to use *Bessel functions of the third kind*, also called *Hankel functions*; these are given by

$$H_\nu^{(1)} = J_\nu(x) + iN_\nu(x) \quad (130)$$

and

$$H_\nu^{(2)} = J_\nu(x) - iN_\nu(x). \quad (131)$$

It is not easy to see what are the properties of the various kinds of Bessel functions from the expansions we have written down so far. As it turns out, their behavior is really quite simple; many of the important features are laid bare by their behavior at small and large arguments. For $x \ll 1$ and real, non-negative ν , one finds

$$J_\nu(x) = \frac{1}{\Gamma(\nu+1)} \left(\frac{x}{2}\right)^\nu [1 - O(x^2)] \quad (132)$$

and

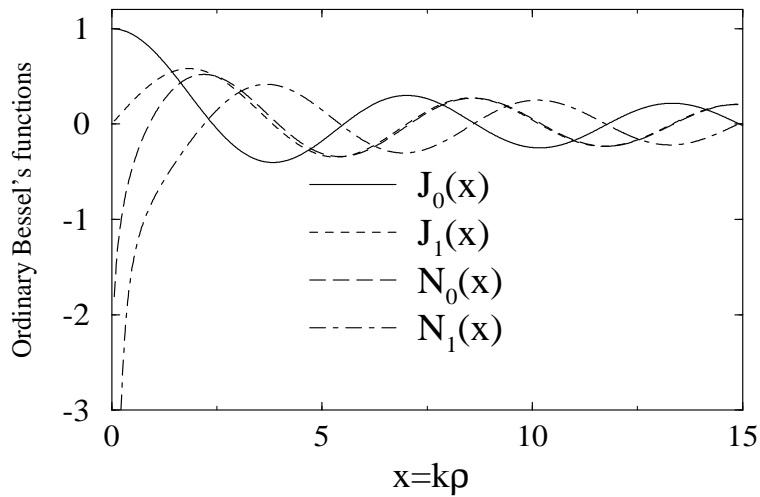
$$N_\nu(x) = \begin{cases} \frac{2}{\pi} [\ln(x/2) + 0.5772 + \dots] & \nu = 0 \\ -\frac{\Gamma(\nu)}{\pi} \left(\frac{2}{x}\right)^\nu & \nu \neq 0 \end{cases} \quad (133)$$

For $x \gg \gg 1$, ν ,

$$J_\nu(x) \sim \sqrt{\frac{2}{\pi x}} \cos\left(x - \frac{\nu\pi}{2} - \frac{\pi}{4}\right) \quad (134)$$

and

$$N_\nu(x) \sim \sqrt{\frac{2}{\pi x}} \sin\left(x - \frac{\nu\pi}{2} - \frac{\pi}{4}\right). \quad (135)$$



Notice in particular that as $x \rightarrow 0$, the Bessel functions of the first kind are well-behaved (finite) whereas the Neumann functions are singular; $N_0(x)$ has only a logarithmic singularity while the higher-order functions are progressively more singular. At large (real) argument, on the other hand, both $J_\nu(x)$ and $N_\nu(x)$ are finite and oscillatory. Hence, the Bessel functions (by which we mean the $J_\nu(x)$) of non-negative order are allowable as solutions of the Laplace equation at all values of ρ ; the Neumann functions, on the other hand, are not allowable on a domain which includes the point $\rho = 0$.

Bessel functions of all kinds satisfy certain recurrence relations. It is a straightforward if tedious matter to show by direct substitution of the series expansions that they obey the following:

$$\Omega_{\nu+1} - \frac{2\nu}{x}\Omega_{\nu} + \Omega_{\nu-1} = 0 \quad (136)$$

and

$$\Omega_{\nu+1} + 2\frac{d\Omega_{\nu}}{dx} - \Omega_{\nu-1} = 0. \quad (137)$$

By taking the sum and difference of these relations, we find also

$$\Omega_{\nu\pm 1} = \frac{\nu}{x}\Omega_{\nu} \mp \frac{d\Omega_{\nu}}{dx}. \quad (138)$$

These are valid for all three kinds of Bessel functions.

The Bessel function $J_{\nu}(x)$ can form a complete orthogonal set on an interval $0 \leq x \leq x_0$ in much the same way as the sine function $\sin(n\pi x/x_0)$ does (Note that x_0 is a zero of the sine function.) Similarly, let us denote the n^{th} zero of $J_{\nu}(x)$ by $x_{\nu n}$ and then form the functions $J_{\nu}(x_{\nu n}y)$, with $n = 1, 2, \dots$. Then it turns out that for fixed ν , these functions provide a complete orthogonal set on the interval $0 \leq y \leq 1$. As usual, we shall not demonstrate completeness. Orthogonality can be demonstrated by making use of the Bessel equation and recurrence relations. It is a useful exercise to do so. Start from the Bessel equation for $J_{\nu}(xy)$,

$$\frac{1}{y} \frac{d}{dy} \left(y \frac{dJ_{\nu}(xy)}{dy} \right) + \left(x^2 - \frac{\nu^2}{y^2} \right) J_{\nu}(xy) = 0 \quad (139)$$

Multiply this equation by $yJ_\nu(x'y)$ and integrate from 0 to 1:

$$\int_0^1 dy \left\{ J_\nu(x'y) \frac{d}{dy} \left(y \frac{dJ_\nu(xy)}{dy} \right) + yJ_\nu(x'y) \left(\mathbf{x}^2 - \frac{\nu^2}{y^2} \right) J_\nu(xy) \right\} = 0 \quad (140)$$

or if we integrate the first term by parts:

$$J_\nu(x'y)y \frac{dJ_\nu(xy)}{dy} \Big|_0^1 - \int_0^1 dy \frac{dJ_\nu(x'y)}{dy} y \frac{dJ_\nu(xy)}{dy} + \int_0^1 dy y J_\nu(x'y) J_\nu(xy) \left(\mathbf{x}^2 - \frac{\nu^2}{y^2} \right) = 0 \quad (141)$$

Similarly, if we start from the differential equation for $J_\nu(x'y)$ and perform the same manipulations, we find the same equation with x and x' interchanged. Subtract the second equation from the first to find

$$J_\nu(x')x \frac{dJ_\nu(u)}{du} \Big|_x - J_\nu(x)x' \frac{dJ_\nu(u)}{du} \Big|_{x'} + (x^2 - x'^2) \int_0^1 y dy J_\nu(xy) J_\nu(x'y) = 0 \quad (142)$$

If we let $x = x_{\nu n}$ and $x' = x_{\nu n'}$, two distinct zeros of the Bessel function, then the integrated terms vanish and we may conclude that the Bessel functions $J_\nu(x_{\nu n}y)$ and $J_\nu(x_{\nu n'}y)$ are orthogonal when integrated over y from 0 to one, provided a factor of y is included in the integrand.

We still have to determine normalization in the case $n = n'$. In the preceding equation, let $x' = x_{\nu n}$ and rearrange the terms to have

$$-J_\nu(x)x_{\nu n} \frac{dJ_\nu(u)}{du} \Big|_{x_{\nu n}} = - \int_0^1 dy y (x^2 - x_{\nu n}^2) J_\nu(x_{\nu n}y) J_\nu(xy), \quad (143)$$

or

$$\int_0^1 dy y J_\nu(x_{\nu n}y) J_\nu(xy) = \frac{J_\nu(x)x_{\nu n} (dJ_\nu(u)/du) \Big|_{x_{\nu n}}}{x^2 - x_{\nu n}^2}. \quad (144)$$

Use L'Hôpital's Rule to evaluate the limit of this expression as $x \rightarrow x_{\nu n}$:

$$\int_0^1 dy y [J_\nu(x_{\nu n} y)]^2 = \frac{1}{2} \left. \frac{dJ_\nu(x)}{dx} \right|_{x_{\nu n}} \left. \frac{dJ_\nu(u)}{du} \right|_{x_{\nu n}} = \frac{1}{2} [J'_\nu(x_{\nu n})]^2, \quad (145)$$

where the prime ' denotes a derivative with respect to argument. Now employ the recurrence relation

$$x J'_\nu(x) = \nu J_\nu(x) - x J_{\nu+1}(x) \quad (146)$$

to find $J'_\nu(x_{\nu n}) = -J_{\nu+1}(x_{\nu n})$, from which the normalization integral becomes

$$\int_0^1 dy y [J_\nu(x_{\nu n} y)]^2 = \frac{1}{2} [J_{\nu+1}(x_{\nu n})]^2 \quad (147)$$

The expansion of an arbitrary function of ρ on the interval $0 \leq \rho \leq a$ may be written as

$$f(\rho) = \sum_{n=1}^{\infty} A_n J_\nu(\rho x_{\nu n}/a) \quad (148)$$

with coefficients which may be determined from the orthonormalization properties of the basis functions as

$$A_n = \frac{\int_0^a \rho d\rho f(\rho) J_\nu(\rho x_{\nu n}/a)}{\frac{a^2}{2} [J_{\nu+1}(x_{\nu n})]^2}. \quad (149)$$

This type of expansion is termed a *Fourier-Bessel series*. The completeness relation for the basis functions is

$$\sum_{n=1}^{\infty} \frac{J_\nu(\rho x_{\nu n}/a) J_\nu(\rho' x_{\nu n}/a)}{(a^2/2) [J_{\nu+1}(x_{\nu n})]^2} = \delta(\rho^2/2 - \rho'^2/2) \equiv \frac{1}{\rho} \delta(\rho - \rho'). \quad (150)$$

It is also of importance to consider the case of imaginary k , $k = i\kappa$ with real κ . Then the functions of z are oscillatory, being of the

form $Z(z) \sim e^{\pm i\kappa z}$, and the functions of ρ will be Bessel functions of imaginary argument, *e.g.*, $J_\nu(i\kappa\rho)$. For given ν^2 , there are two linearly independent solutions which are conventionally chosen to be J_ν and $H_\nu^{(1)}$, the reason being that they have particularly simple behaviors at large and small arguments. Let us introduce the *modified Bessel functions* $I_\nu(x)$ and $K_\nu(x)$,

$$I_\nu(x) \equiv i^{-\nu} J_\nu(ix) \quad (151)$$

$$K_\nu(x) \equiv \frac{\pi}{2} i^{\nu+1} H_\nu^{(1)}(ix). \quad (152)$$

These have the forms at small argument, $x \ll 1$,

$$I_\nu(x) = \frac{1}{\Gamma(\nu+1)} \left(\frac{x}{2}\right)^\nu \quad (153)$$

and

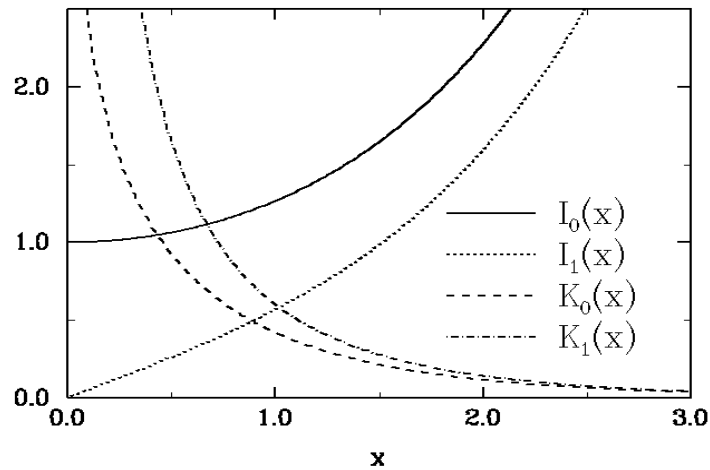
$$K_\nu(x) = \begin{cases} -\ln(x/2) + 0.5772 + \dots & \nu = 0 \\ \frac{\Gamma(\nu)}{2} \left(\frac{2}{x}\right)^\nu & \nu \neq 0 \end{cases} \quad (154)$$

while at large argument, $x \gg \gg 1, \nu$

$$I_\nu(x) = \frac{1}{\sqrt{2\pi x}} e^x \left[1 + O\left(\frac{1}{x}\right)\right] \quad (155)$$

and

$$K_\nu(x) = \sqrt{\frac{\pi}{2x}} e^{-x} \left[1 + O\left(\frac{1}{x}\right)\right]. \quad (156)$$

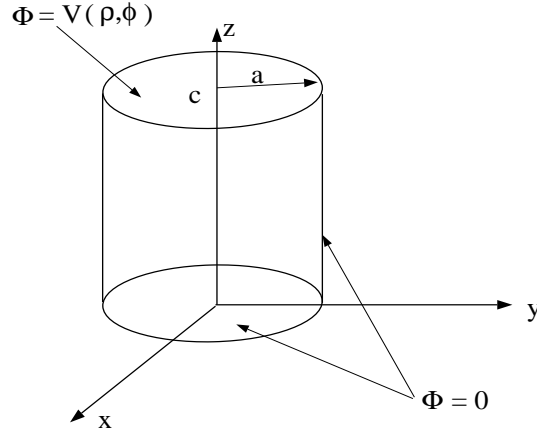


From these equations we see that $I_\nu(x)$ is well-behaved for $x < \infty$, corresponding to $\rho < \infty$, while $K_\nu(x)$ is well-behaved for $x > 0$ or $\rho > 0$, from which we can decide which function(s) to use in expanding any given potential problem.

Let's look at some examples of expansions in cylindrical coordinates.

2.1 Example I

Consider a charge-free right-circular cylinder bounded by S given by $\rho = a$, $z = 0$, and $z = c$. Let $\Phi(\mathbf{x})$ be zero on S except for the top face $z = c$ where $\Phi(\rho, \phi, c) = V(\rho, \phi)$ with V given.



For this distribution of boundary potential, we need a complete set of functions of the space $0 \leq \phi \leq 2\pi$ and $0 \leq \rho \leq a$. Thus we take (k real) $Z(z)$ to be damped exponentials (sinh and cosh) and $R(\rho)$ to be ordinary Bessels functions.

$$\Phi \sim e^{im\phi} (AJ_m(k\rho) + BN_m(k\rho)) (\cos(kz) \pm \sin(kz)) \quad (157)$$

It will be convenient if each of these functions is equal to zero when $\rho = a$ and also if each one is zero when $z = 0$. With just a little thought,

1. No Neumann functions N_m since they diverge at $\rho = 0$
2. No cosh since it is finite at $z = 0$, and hence would not satisfy the B.C.
3. Since J_m and J_{-m} are not independent, use $J_{|m|}$.

we realize that we want to use the hyperbolic sine function of z and the Bessel function of the first kind for R . Our expansion is thus of the

form

$$\Phi(\rho, \phi, z) = \sum_{m=-\infty}^{\infty} \sum_{n=1}^{\infty} A_{mn} e^{im\phi} J_{|m|}(x_{mn}\rho/a) \sinh(x_{mn}z/a) \quad (158)$$

where x_{mn} is the n^{th} zero of $J_{|m|}(x)$. Each term in the sum is itself a solution of the Laplace equation; each one satisfies the boundary conditions on $z = 0$ and $\rho = a$, and, for given n , we have a complete set of functions of ϕ while for given m , we have a complete set of functions of ρ .

The coefficients in the expansion are determined from the condition that Φ reduce to the given potential V on the top face of the cylinder. Making use of the orthogonality of the basis functions of both ϕ and ρ , we have

$$\begin{aligned} & \int_0^a \rho d\rho \int_0^{2\pi} d\phi V(\rho, \phi) J_{|m|}(x_{mn}\rho/a) e^{-im\phi} \\ &= A_{mn} 2\pi (a^2/2) [J_{|m|+1}(x_{mn})]^2 \sinh(x_{mn}c/a) \end{aligned} \quad (159)$$

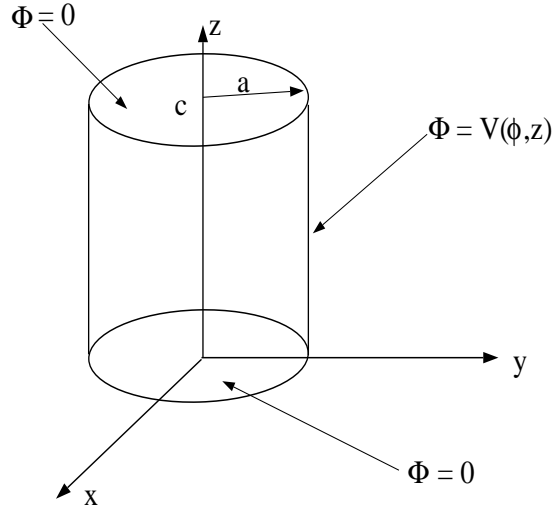
or

$$A_{mn} = \frac{\int_0^{2\pi} d\phi \int_0^a \rho d\rho e^{-im\phi} J_{|m|}(x_{mn}\rho/a) V(\rho, \phi)}{\pi a^2 \sinh(x_{mn}c/a) [J_{|m|+1}(x_{mn})]^2}. \quad (160)$$

For any given function V , one may now attempt to complete the integrals.

2.2 Example II

Consider the same geometry as in the first example but now with boundary condition $\Phi = 0$ on the constant- z faces and some given value $V(\phi, z)$ on the surface at $\rho = a$.



For this system we need a complete set of functions on the domain $0 \leq z \leq c$ and $0 \leq \phi \leq 2\pi$ which means picking k imaginary, $k = i\kappa$. The appropriate functions of z are sin and cos, and the appropriate functions of ρ are the modified Bessels Functions.

$$\Phi \sim e^{im\phi} (\sin(kz) \pm \cos(kz)) (AI(k\rho) + BK(k\rho)) \quad (161)$$

1. We may eliminate the K modified Bessels functions since they diverge when $\rho \rightarrow 0$.
2. Since I_m is not independent of I_{-m} , we use $I_{|m|}$.
3. The cos function of z cannot be zero at both $z = 0$ and $z = c$, and so may be eliminated.
4. Take $k = n\pi/c$ so that $\sin(n\pi z/c)|_{z=c} = 0$

Thus the expansion is

$$\Phi(\rho, \phi, z) = \sum_{m=-\infty}^{\infty} \sum_{n=1}^{\infty} A_{mn} e^{im\phi} \sin(n\pi z/c) I_{|m|}(n\pi\rho/c) \quad (162)$$

with coefficients given by

$$A_{mn} = \frac{\int_0^c dz \int_0^{2\pi} d\phi e^{-im\phi} \sin(n\pi z/c) V(\phi, z)}{\pi c I_{|m|}(n\pi a/c)} \quad (163)$$

2.3 B.V.P. on Large Cylinders

By applying the same considerations, one may solve other boundary-value problems on cylinders. A case of special interest, and requiring special treatment, is one in which $a \rightarrow \infty$; then $k_{\nu n} \equiv x_{\nu n}/a$ becomes a continuous variable and instead of a Fourier-Bessel series, we come up with an integral. The orthogonality condition is

$$\int_0^{\infty} x dx J_m(kx) J_m(k'x) = \frac{1}{k} \delta(k - k') \quad (164)$$

and the completeness relation is the same, with different names for the variables,

$$\int_0^{\infty} k dk J_m(kx) J_m(kx') = \frac{1}{x} \delta(x - x'). \quad (165)$$

To see how this comes to be, consider the completeness relation on a finite interval,

$$\sum_{n=1}^{\infty} \frac{J_m(x_{mn}\rho/a) J_m(x_{mn}\rho'/a)}{(a^2/2)[J_{m+1}(x_{mn})]^2} = \frac{1}{\rho} \delta(\rho - \rho') \quad (166)$$

and then let $a \rightarrow \infty$, defining x_{mn}/a as k while noting that the interval between roots of the Bessel function at large argument is π . Also, the

asymptotic form of the Bessel function, valid at large argument, is

$$\begin{aligned} J_{m+1}(x_{mn}) &\sim \sqrt{\frac{2}{\pi x_{mn}}} \cos[n\pi + (m - 1/2)\pi/2 - (m + 1)\pi/2 - \pi/4] \\ &= \sqrt{\frac{2}{\pi x_{mn}}} \cos[(n - 1)\pi] = -(-1)^n \sqrt{\frac{2}{\pi x_{mn}}} \end{aligned} \quad (167)$$

Using this substitution in the closure relation for the finite interval and taking the limit of large a , one finds that the sum becomes the integral, Eq. (165). Of course, this is not a rigorous derivation because the asymptotic expression is not arbitrarily accurate for all roots.

2.4 Green's Function Expansion in Cylindrical Coordinates

We can expand the Dirichlet Green's function in cylindrical coordinates in much the same manner as we did in spherical coordinates. We shall go through the derivation to expose a somewhat different approach from what we employed in the latter case. We consider a domain between two infinitely long right-circular cylindrical surfaces. Then G must vanish on these surfaces and it must also satisfy a Poisson equation,

$$\begin{aligned} \nabla^2 G(\mathbf{x}, \mathbf{x}') &= -4\pi \delta(\rho^2/2 - \rho'^2/2) \delta(\phi - \phi') \delta(z - z') \\ &= -\frac{4\pi}{\rho} \delta(\rho - \rho') \delta(\phi - \phi') \delta(z - z'). \end{aligned} \quad (168)$$

Let us write the delta functions of ϕ and z using closure relations,

$$\delta(z - z') = \frac{1}{2\pi} \int_{-\infty}^{\infty} dk e^{ik(z-z')} = \frac{1}{\pi} \int_0^{\infty} dk \cos[k(z - z')] \quad (169)$$

and

$$\delta(\phi - \phi') = \frac{1}{2\pi} \sum_{m=-\infty}^{\infty} e^{im(\phi-\phi')}. \quad (170)$$

Similarly, expand the ϕ -dependence and z -dependence of G using the same basis functions,

$$G(\mathbf{x}, \mathbf{x}') = \sum_{m=-\infty}^{\infty} \int_0^{\infty} \frac{dk}{2\pi^2} g_m(k, \rho, \rho') e^{im(\phi-\phi')} \cos[k(z-z')]. \quad (171)$$

Now operate on this expansion with the Laplacian using cylindrical coordinates:

$$\begin{aligned} \nabla^2 G(\mathbf{x}, \mathbf{x}') &= \sum_{m=-\infty}^{\infty} \int_0^{\infty} \frac{dk}{2\pi^2} \left(\frac{d^2}{d\rho^2} + \frac{1}{\rho} \frac{d}{d\rho} - \left[\frac{m^2}{\rho^2} + k^2 \right] \right) g_m e^{im(\phi-\phi')} \cos[k(z-z')] \\ &= -\frac{4\pi}{\rho} \delta(\rho - \rho') \sum_{m=-\infty}^{\infty} \int_0^{\infty} \frac{dk}{2\pi^2} e^{im(\phi-\phi')} \cos[k(z-z')] \end{aligned} \quad (172)$$

Multiply by members of the basis sets, *i.e.*, $e^{-im'(\phi-\phi')}$ and $\cos[k'(z-z')]$ and integrate over the appropriate intervals of $\phi - \phi'$ and $z - z'$ to find a differential equation for g_m ,

$$\frac{d^2 g_m}{d\rho^2} + \frac{1}{\rho} \frac{d g_m}{d\rho} - \left(k^2 + \frac{m^2}{\rho^2} \right) g_m = -\frac{4\pi}{\rho} \delta(\rho - \rho'). \quad (173)$$

For $\rho \neq \rho'$, this is Bessel's equation with solutions (viewed as functions of ρ) which are Bessel functions of imaginary argument, or, as we have described, modified Bessel functions of argument $k\rho$. Because of the delta function inhomogeneous term, the solution for $\rho < \rho'$ is different from the solution for $\rho > \rho'$. Hence we may write that, for $\rho < \rho'$,

$$g_m(k, \rho, \rho') = A_{<}(\rho') K_m(k\rho) + B_{<}(\rho') I_m(k\rho), \quad (174)$$

and, for $\rho > \rho'$,

$$g_m(k, \rho, \rho') = A_{>}(\rho')K_m(k\rho) + B_{>}(\rho')I_m(k\rho). \quad (175)$$

The various coefficients are functions of ρ' and must in fact be linear combinations of $K_m(\rho')$ and $I_m(\rho')$ because $G(\mathbf{x}, \mathbf{x}')$ is also a solution of $\nabla'^2 G(\mathbf{x}, \mathbf{x}') = 0$ when $\rho \neq \rho'$; another way to see this same point is to recall that $G(\mathbf{x}, \mathbf{x}') = G(\mathbf{x}', \mathbf{x})$. Finally, the coefficients are further constrained by the condition that the Green's function must vanish when ρ becomes equal to the radius of either the inner or outer cylinder.

Let us at this point restrict our attention to a special (and simple) limiting case which is the infinite space. The radius of the inner cylinder is 0 and that of the outer one becomes infinite in this limit. Then we have to have a function g_m which remains finite as $\rho \rightarrow 0$ which can only be $I_m(k\rho)$; also, we must have g_m vanish as $\rho \rightarrow \infty$, which can only be $K_m(k\rho)$. Thus we have

$$g_m(k, \rho, \rho') = \begin{cases} A_{<}(\rho')I_m(k\rho) & \rho < \rho' \\ A_{>}(\rho')K_m(k\rho) & \rho > \rho'. \end{cases} \quad (176)$$

The symmetry condition on G tells us that $A_{<}(\rho') = AK_m(k\rho')$ while $A_{>}(\rho') = AI_m(k\rho')$. All of these conditions are included in the statement

$$g_m(k, \rho, \rho') = AI_m(k\rho_{<})K_m(k\rho_{>}) \quad (177)$$

where $\rho_{<} (\rho_{>})$ is the smaller (larger) of ρ and ρ' .

The remaining constant in the determination of g_m can be found from the required normalization of G . Let us integrate Eq. (173) across the point $\rho = \rho'$:

$$\int_{\rho'-\epsilon}^{\rho'+\epsilon} d\rho \left(\frac{d^2}{d\rho^2} + \frac{1}{\rho} \frac{d}{d\rho} - \left[k^2 + \frac{m^2}{\rho^2} \right] \right) g_m = -\frac{4\pi}{\rho'} \quad (178)$$

If we take the limit of $\epsilon \rightarrow 0$ and realize that g_m is continuous while its first derivative is not, we find that this equation gives

$$\lim_{\epsilon \rightarrow 0} \frac{dg_m}{d\rho} \Big|_{\rho'-\epsilon}^{\rho'+\epsilon} = A \left[I_m(k\rho')k \frac{dK_m(x)}{dx} \Big|_{k\rho'} - K_m(k\rho')k \frac{dI_m(x)}{dx} \Big|_{k\rho'} \right] = -\frac{4\pi}{\rho'}, \quad (179)$$

or

$$A[I_m(x)K'_m(x) - K_m(x)I'_m(x)] = -\frac{4\pi}{x}; \quad (180)$$

the primes denote derivatives with respect to the argument x . The quantity in [...] here is the *Wronskian* of I_m and K_m . One may learn by consulting, *e.g.*, the section on Bessel functions in Abramowitz and Stegun, that Bessel functions have simple Wronskians:

$$I_m(x)K'_m(x) - K_m(x)I'_m(x) \equiv W[I_m(x), K_m(x)] = -\frac{1}{x}. \quad (181)$$

Comparison of the two preceding equations leads one to conclude that $A = 4\pi$. Hence our expansion of $G(\mathbf{x}, \mathbf{x}')$, which is just $1/|\mathbf{x} - \mathbf{x}'|$, is

$$G(\mathbf{x}, \mathbf{x}') = \frac{1}{|\mathbf{x} - \mathbf{x}'|} = \frac{2}{\pi} \sum_{m=-\infty}^{\infty} \int_0^{\infty} dk e^{im(\phi-\phi')} \cos[k(z-z')] I_m(k\rho_{<}) K_m(k\rho_{>}) \quad (182)$$

which may also be written entirely in terms of real functions as

$$\frac{1}{|\mathbf{x} - \mathbf{x}'|} = \frac{4}{\pi} \int_0^\infty dk \cos[k(z - z')] \\ \times \left\{ \frac{1}{2} I_0(k\rho_<) K_0(k\rho_>) + \sum_{m=1}^\infty \cos[m(\phi - \phi')] I_m(k\rho_<) K_m(k\rho_>) \right\}. \quad (183)$$

This turns out to be a useful expansion of $1/|\mathbf{x} - \mathbf{x}'|$; it also provides a starting point for the derivation of some other equally useful expansions. For example, if we let $\mathbf{x}' = 0$, then $\rho_< = 0$ and all I_m vanish except for $m = 0$, while $1/|\mathbf{x} - \mathbf{x}'| = 1/|\mathbf{x}| = 1/\sqrt{\rho^2 + z^2}$, so we find, using also $I_0(0) = 1$,

$$\frac{1}{\sqrt{\rho^2 + z^2}} = \frac{2}{\pi} \int_0^\infty dk \cos(kz) K_0(k\rho). \quad (184)$$

Other useful identities may be obtained.

Multipoles; Macroscopic Media; Dielectrics

Henry Cavendish

(1731 - 1810)

December 23, 2000

Contents

1	Multipole Expansion: An Alternate Approach	2
1.1	Interpretation of the Moments	5
1.2	Dipole Field	8
2	Energy of the Charge Distribution	10
2.1	Example: Dipole Energies	12
2.2	Example: Quadrupole Energies	12
3	Dipoles in Nature: Permanent and Induced	13
3.1	Permanent Dipoles	14
3.2	Induced Dipoles	14
	3.2.1 Static Models	15
	3.2.2 Dynamic Model	16
4	Dielectric Materials	17
4.1	Statistical Mechanics	18
	4.1.1 Induced dipoles	19

4.1.2	Permanent Dipoles	20
4.1.3	Both	21
4.2	Macroscopic Electrostatics; Dielectrics	21
4.2.1	Electric Displacement	27
4.2.2	Summary and Discussion	29
5	Boundary-Value Problems in Dielectrics	33
5.1	Example: Point Charge Near a Boundary	35
5.2	Example: Dielectric Sphere in a Uniform Field	40
5.2.1	The Inverse Problem	43
5.3	Clausius-Mossotti equation	44
6	Electrostatic Energy in Dielectrics	46
6.1	Force on a Dielectric	49
6.2	Forces on a Dielectric Revisited	52
7	Example: Dielectrophoresis	56
A	Multipole Expansion: with Spherical Harmonics	64

In this chapter, we shall first develop the multipole expansion for the electrostatic potential and field. This is useful not only for expressing the field produced by a localized distribution of charge but is also a helpful preliminary investigation for the business of describing the electrostatics of materials containing a large number of charges and which are not conductors. These are called *dielectrics*. After developing a means of describing their electrostatic properties, we shall turn to boundary value problems in systems comprising dielectrics and conductors.

1 Multipole Expansion: An Alternate Approach

In this section we will develop the multipole expansion for a charge distribution by an alternate means to that used in Jackson (the method used in Jackson is discussed in the appendix to this chapter).

We begin by writing the general expression for the potential due to a finite charge distribution $\rho(\mathbf{x})$,

$$\Phi(\mathbf{x}) = \int d^3x' \rho(\mathbf{x}') \frac{1}{|\mathbf{x} - \mathbf{x}'|} \quad (1)$$

Let us consider the case where the origin is within the charge distribution and where $|\mathbf{x}| = r$ is large compared to the size of the charge distribution. Then we may expand the denominator in the integrand.

$$\frac{1}{|\mathbf{x} - \mathbf{x}'|} = \frac{1}{\sqrt{r^2 - 2\mathbf{x} \cdot \mathbf{x}' + r'^2}} = \frac{1}{r\sqrt{1 - 2\mathbf{x} \cdot \mathbf{x}'/r^2 + r'^2/r^2}}. \quad (2)$$

Using the Taylor series expansion

$$\frac{1}{\sqrt{1-x}} = 1 + \frac{1}{2}x + \frac{3}{8}x^2 + \dots \quad (3)$$

We get

$$\frac{1}{|\mathbf{x} - \mathbf{x}'|} = \frac{1}{r} \left(1 + \frac{1}{2} (2\mathbf{x} \cdot \mathbf{x}'/r^2 - r'^2/r^2) + \frac{3}{8} (4(\mathbf{x} \cdot \mathbf{x}')^2/r^4) \right) + \text{order}((r'/r)^3) \quad (4)$$

$$\frac{1}{|\mathbf{x} - \mathbf{x}'|} = \frac{1}{r} + \frac{\mathbf{x} \cdot \mathbf{x}'}{r^3} + \frac{1}{2r^5} (3(\mathbf{x} \cdot \mathbf{x}')^2 - r'^2 r^2) + \dots \quad (5)$$

With this expansion, we can rewrite $\Phi(\mathbf{x})$ as

$$\begin{aligned} \Phi(\mathbf{x}) &= \int d^3x' \frac{\rho(\mathbf{x}')}{r} + \\ &\quad \int d^3x' \frac{(\mathbf{x} \cdot \mathbf{x}')\rho(\mathbf{x}')}{r^3} + \\ &\quad \int d^3x' \frac{\rho(\mathbf{x}')}{2r^5} (3(\mathbf{x} \cdot \mathbf{x}')^2 - r'^2 r^2) + \dots \end{aligned}$$

$$\begin{aligned} \Phi(\mathbf{x}) &= \frac{1}{r} \int d^3x' \rho(\mathbf{x}') \quad \text{monopole} + \\ &\quad \frac{\mathbf{x}}{r^3} \cdot \int d^3x' \mathbf{x}' \rho(\mathbf{x}') \quad \text{dipole} + \\ &\quad \frac{1}{2r^5} \int d^3x' \rho(\mathbf{x}') (3(\mathbf{x} \cdot \mathbf{x}')^2 - r'^2 r^2) \quad \text{quadrupole} + \dots \end{aligned}$$

or

$$\begin{aligned} \Phi(\mathbf{x}) &= \frac{1}{r} \int d^3x' \rho(\mathbf{x}') + \\ &\quad \frac{\mathbf{x}}{r^3} \cdot \int d^3x' \mathbf{x}' \rho(\mathbf{x}') + \\ &\quad \frac{1}{2} \sum_{i,j} \frac{x_i x_j}{2r^5} \int d^3x' \rho(\mathbf{x}') (3x'_i x'_j - r'^2 \delta_{ij}) + \dots \quad (6) \end{aligned}$$

where the sum is over the 3 coordinates of space. If we follow the conventional designation of these terms, then

$$\Phi(\mathbf{x}) = q/r + \frac{\mathbf{x} \cdot \mathbf{p}}{r^3} + \frac{1}{2} \sum_{ij} Q_{ij} \frac{x_i x_j}{r^5} + \dots \quad (7)$$

where

$$q = \int d^3 x' \rho(\mathbf{x}') \quad \text{Monopole Moment} \quad (8)$$

$$\mathbf{p} = \int d^3 x' \rho(\mathbf{x}') \mathbf{x}' \quad \text{Dipole Moment} \quad (9)$$

$$Q_{ij} = \int d^3 x' \rho(\mathbf{x}') (3x'_i x'_j - r'^2 \delta_{ij}) \quad \text{Quadrupole Moment} \quad (10)$$

Note that the matrix Q_{ij} is real and symmetric ($Q_{ij} = Q_{ji}$). Thus only six of its elements are independent. In fact, only 5 are, since there is an additional constraint that $\text{Tr}(Q) = 0$.

$$\text{Tr}(Q) = \sum_i Q_{ii} = \int d^3 x' \rho(\mathbf{x}') \sum_i (3x'_i x'_i - r'^2 \delta_{ii}) \quad (11)$$

then as

$$\sum_i 3x'_i x'_i = 3r'^2 \quad (12)$$

$$\sum_i r'^2 \delta_{ii} = 3r'^2 \quad (13)$$

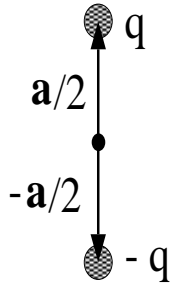
$$\text{Tr}(Q) = \int d^3 x' \rho(\mathbf{x}') (3r'^2 - 3r'^2) = 0 \quad (14)$$

Thus it must be that $Q_{33} = -Q_{11} - Q_{22}$ and only two of the diagonal components are independent. This is important, since we will relate Q to the set of five Y_2^m .

1.1 Interpretation of the Moments

What is the interpretation of these terms? The *monopole moment* is just the total charge of the distribution. Thus the monopole term gives the potential due to the charge as a whole. Since the monopole term in the potential falls off like $1/r$ at large r , it will dominate the far field potential whenever q is finite. The *dipole moment* is the first moment of the charge distribution; and refers to how the charge is distributed in space. Similarly, the quadrupole moment is a second moment of the distribution.

Let's consider the dipole in some detail, with the model shown below.



$$\Phi(\mathbf{x}) = \frac{q}{|\mathbf{x} - \mathbf{a}/2|} + \frac{-q}{|\mathbf{x} + \mathbf{a}/2|} \quad (15)$$

For $|\mathbf{x}| = r \gg a$ we can expand

$$\frac{1}{|\mathbf{x} \mp \mathbf{a}/2|} = \frac{1}{r} \left(1 \pm \frac{\mathbf{x} \cdot \mathbf{a}}{2r^2} \right) + \text{order}(a^2/r^2) \quad (16)$$

$$\Phi(\mathbf{x}) \approx \frac{q}{r} \left(1 + \frac{\mathbf{x} \cdot \mathbf{a}}{2r^2} \right) - \frac{q}{r} \left(1 - \frac{\mathbf{x} \cdot \mathbf{a}}{2r^2} \right)$$

$$\approx \frac{q}{r^3} \mathbf{x} \cdot \mathbf{a} \quad (17)$$

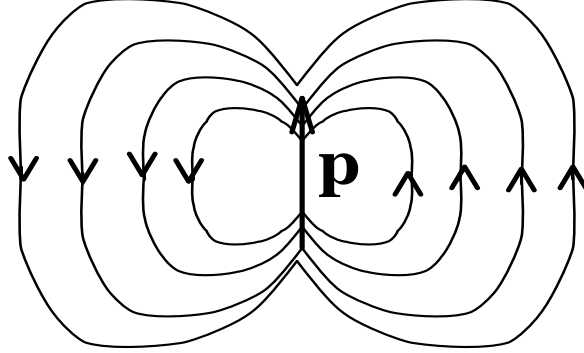
If $\mathbf{a} \rightarrow 0$ in the diagram (such that $q\mathbf{a} = \mathbf{p} = \text{constant}$), then the higher order terms vanish, and this result becomes exact. In such a limit ($\mathbf{a} \rightarrow 0, q\mathbf{a} = \mathbf{p}$) we obtain a point dipole.

Now consider the field due to a dipole $\mathbf{p} = p\hat{\mathbf{z}}$

$$\Phi(\mathbf{x}) = \frac{p \cos \theta}{r^2} \quad (18)$$

$$E_r = -\frac{\partial \Phi}{\partial r} = \frac{2p \cos \theta}{r^3} \quad (19)$$

$$E_\theta = -\frac{1}{r} \frac{\partial \Phi}{\partial \theta} = \frac{p \sin \theta}{r^3}, \quad (20)$$



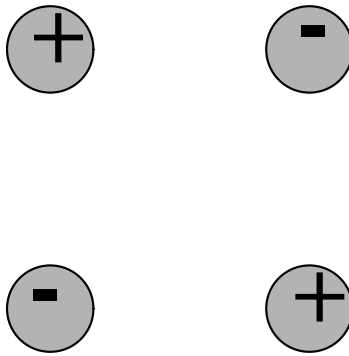
or more formally (in Cartesian coordinates).

$$\begin{aligned} E_{\mathbf{p}} &= -\nabla \left(\frac{\mathbf{p} \cdot \mathbf{x}}{r^3} \right) = -\sum_i \mathbf{e}_i \frac{\partial}{\partial x_i} \sum_j \frac{p_j x_j}{r^3} \\ &= -\sum_i \mathbf{e}_i \left(\frac{p_i}{r^3} - 3 \sum_j \frac{p_j x_j x_i}{r^4} \right) \end{aligned} \quad (21)$$

where we have used the fact that $\frac{\partial r}{\partial x_i} = \frac{x_i}{r}$.

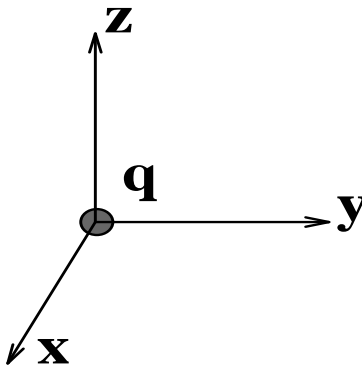
$$E_{\mathbf{p}} = \frac{3\mathbf{x}(\mathbf{p} \cdot \mathbf{x})}{r^5} - \frac{\mathbf{p}}{r^3} \quad (22)$$

In a similar fashion the potential term involving the quadrupole moment may be interpreted as due to an assembly of four charges (hence the name).



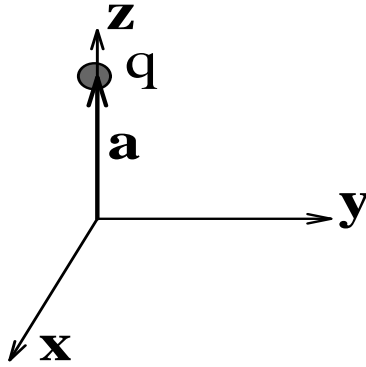
Higher-order moments (octapole, hexadecapole, etc.) may be generated in a like fashion.

It is important to note that the interpretation of the moments depends strongly upon the origin. For example consider a point charge located at the origin.



$$\Phi(\mathbf{x}) = q/r \quad (23)$$

It has only a monopole term. Now displace the charge by a vector \mathbf{a} .



$$\Phi(\mathbf{x}) = \frac{q}{|\mathbf{x} - \mathbf{a}|} = q/r + q \frac{\mathbf{a} \cdot \mathbf{x}}{r^3} + \frac{q}{2} \frac{3(\mathbf{a} \cdot \mathbf{x})^2 - a^2 r^2}{r^5} + \dots \quad (24)$$

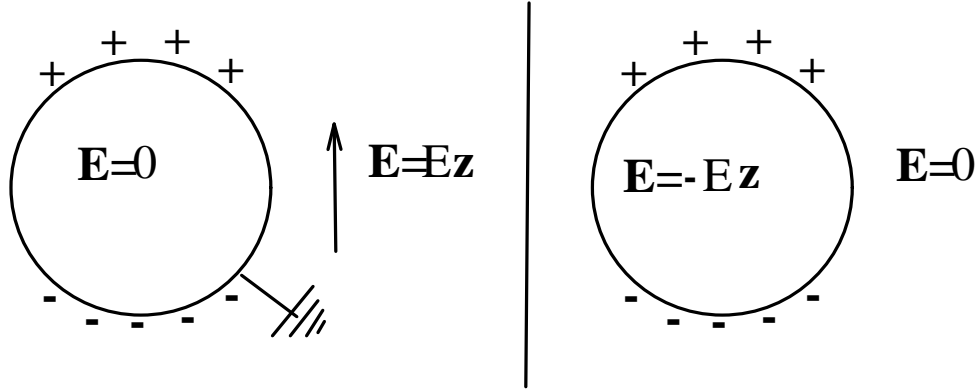
This has moments to all orders! *watch your origin!*

1.2 Dipole Field

It is interesting to ask what is going on at the origin, where our expansion fails. Let's look at the particular case of the dipole field, assuming a point dipole at $r = 0$. For any $r > 0$, we know that the potential is as given in Eq. (17). Once before we found such a potential when we solved the problem of a conducting sphere of radius a in a uniform external applied field \mathbf{E}_0 . What we found was that the potential outside of the sphere is

$$\Phi(\mathbf{x}) = -E_0 r \cos \theta + E_0 a^3 \frac{\cos \theta}{r^2} \quad (25)$$

while inside of the sphere the potential is a constant and the field is zero.



If one removes the applied field but retains the field produced by the charges on the surface of the sphere, then the potential for $r > a$ is simply $E_0 a^3 \cos \theta / r^2$ which may also be written as $\mathbf{p} \cdot \mathbf{x} / r^3$ with $\mathbf{p} = E_0 a^3 \hat{\mathbf{z}}$. By the superposition principle, the field at $r < a$ is now $-\mathbf{E}_0 = -\mathbf{p} / a^3$. What this means is that the surface charge on the sphere has a dipole moment and, remarkably, no other multipole moments. Now let us fix \mathbf{p} while letting $a \rightarrow 0$. The region $r < a$ shrinks, while the field inside gets bigger. As the region shrinks to zero, the field strength at the origin (*i.e.*, inside the sphere) diverges. The integral of the field over the spherical domain $r < a$ is, however, a constant and equal to $-(4\pi/3)\mathbf{p}$. Consequently, in the limit of vanishing a , this field may be represented by a delta function, $-(4\pi/3)\mathbf{p}\delta(\mathbf{x})$. The total field of a point dipole of moment \mathbf{p} is thus the dipolar field, Eq. (21), for $r > 0$

plus a delta-function piece at the origin,

$$\mathbf{E}(\mathbf{x}) = \frac{3\mathbf{n}(\mathbf{p} \cdot \mathbf{n}) - \mathbf{p}}{r^3} - \frac{4\pi}{3}\mathbf{p}\delta(\mathbf{x}). \quad (26)$$

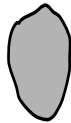
Our derivation of this result is not completely general since it is based on the limiting form of the solution to one particular problem involving a sphere; the result is, however, quite correct for any point dipole. See Jackson, Chapter 4, Section 1, for a more complete discussion of this point.

2 Energy of the Charge Distribution

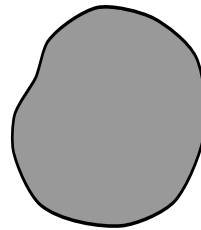
In this section we consider the energy of a localized charge distribution $\rho(\mathbf{x})$ in an external applied electric field $\mathbf{E}(\mathbf{x})$ which may be described through its potential $\Phi(\mathbf{x})$. This energy is, as we know from Chapter 1,

$$W = \int d^3x \rho(\mathbf{x})\Phi(\mathbf{x}). \quad (27)$$

Source of E



To calculate the energy of a charge distribution in an external field, we must ignore the self field



Not source of E

Notice that there is no factor of 1/2; that is because we are finding the interaction energy of a charge distribution with a field which is *not*

produced by that same charge distribution and so we do not double count the energy in Eq. (24) by omitting this factor.

Now if we assume that Φ changes slowly over the region where ρ is appreciable, then we can expand the potential Φ around the origin of coordinates using a Taylor series:

$$\begin{aligned}\Phi(\mathbf{x}) &= \Phi(0) + \mathbf{x} \cdot \nabla\Phi(\mathbf{x})|_{\mathbf{x}=0} + \frac{1}{2} \sum_{i,j=1}^3 x_i x_j \left. \frac{\partial^2\Phi(\mathbf{x})}{\partial x_i \partial x_j} \right|_{\mathbf{x}=0} + \dots \\ &= \Phi(0) - \mathbf{x} \cdot \mathbf{E}(0) - \frac{1}{2} \sum_{i,j=1}^3 x_i x_j \left. \frac{\partial E_j(\mathbf{x})}{\partial x_i} \right|_{\mathbf{x}=0} + \dots\end{aligned}\quad (28)$$

where $\mathbf{E}(\mathbf{x})$ is the external applied field. Now, this field is such that its sources are far away, or at least are zero in the region where the charge distribution $\rho(\mathbf{x})$ is located. Therefore $\nabla \cdot \mathbf{E}(\mathbf{x}) = 0$ in this region and so we can add a term proportional to $\nabla \cdot \mathbf{E}(\mathbf{x})$ to the potential $\Phi(\mathbf{x})$ without changing the result of the integral in Eq. (24). We choose this term to be

$$\frac{1}{6} r^2 \nabla \cdot \mathbf{E}(\mathbf{x})|_{\mathbf{x}=0} = \frac{1}{6} \sum_{i,j=1}^3 r^2 \delta_{ij} \left. \frac{\partial E_j(\mathbf{x})}{\partial x_i} \right|_{\mathbf{x}=0} . \quad (29)$$

Hence we have

$$\Phi(\mathbf{x}) = \Phi(0) - \mathbf{x} \cdot \mathbf{E}(0) - \frac{1}{6} \sum_{i,j=1}^3 (3x_i x_j - r^2 \delta_{ij}) \left. \frac{\partial E_j(\mathbf{x})}{\partial x_i} \right|_{\mathbf{x}=0} , \quad (30)$$

plus higher-order terms. If we substitute this expansion into the expression for the energy, we find

$$W = q\Phi(0) - \mathbf{p} \cdot \mathbf{E}(0) - \frac{1}{6} \sum_{i,j=1}^3 Q_{ij} \left. \frac{\partial E_j(\mathbf{x})}{\partial x_i} \right|_{\mathbf{x}=0} + \dots \quad (31)$$

2.1 Example: Dipole Energies

As an example making use of this result, suppose that we have a dipole of moment \mathbf{p}_1 at point \mathbf{x}_1 in the presence of a second dipole of moment \mathbf{p}_2 at \mathbf{x}_2 . Then the energy of interaction is



$$W = -\mathbf{p}_1 \cdot \mathbf{E}_2(\mathbf{x}_1) = \frac{-3(\mathbf{p}_2 \cdot \mathbf{n})(\mathbf{p}_1 \cdot \mathbf{n}) + \mathbf{p}_1 \cdot \mathbf{p}_2}{|\mathbf{x}_1 - \mathbf{x}_2|^3} \quad (32)$$

where $\mathbf{n} = (\mathbf{x}_1 - \mathbf{x}_2)/|\mathbf{x}_1 - \mathbf{x}_2|$ is a unit vector pointing from the second dipole to the first (or vice versa).

2.2 Example: Quadrupole Energies

A second example has to do with the coupling of a nucleus' electric quadrupole moment to an external field (such as that from the electrons). By choosing the origin in an appropriate fashion, one can guarantee that any nucleus (any object with a non-zero net charge, in fact) has no dipole moment. Hence the first interesting term in the nucleus' interaction with an external field is the electric quadrupole interaction. Further, a nucleus in an angular momentum eigenstate $|J, M\rangle$ will have a charge density which is invariant under rotation around the

z -axis,

$$\rho \propto Y_J^M Y_J^{M*} \propto e^{iM\phi} e^{-iM\phi}$$

leading to a diagonal electric quadrupole moment tensor (the matrix of Q_{lm} 's) which is such that $Q_{xx} = Q_{yy}$. Since the trace of this tensor (or matrix) is zero, this means that¹ $Q_{xx} = Q_{yy} = -Q_{zz}/2$. The upshot is that the interaction of the nuclear quadrupole with the applied field is

$$W = -\frac{1}{4} Q_{zz} \left. \frac{\partial E_z(\mathbf{x})}{\partial z} \right|_{\mathbf{x}=0}. \quad (33)$$

Bear in mind that the moment Q_{zz} is a function of the internal state of the nucleus and in particular of its angular momentum states. The quadrupolar coupling thus provides a way to lift the degeneracy associated with the different quantum numbers M for the z -component of angular momentum.

3 Dipoles in Nature: Permanent and Induced

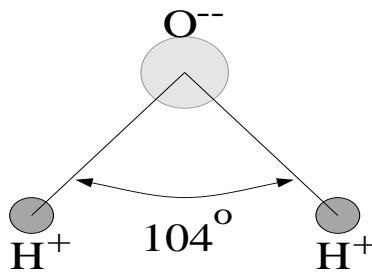
Why are dipoles so interesting?? The reason is that many atoms and molecules have dipole moments which affects their chemical and electrical properties.

¹Or maybe that means this: $Q_{11} = Q_{22} = -Q_{33}/2$.

3.1 Permanent Dipoles

An example of a molecule with a permanent electric dipole moment is water H_2O .

$$|\mathbf{p}| = 1.86 \times 10^{-18} \text{esu} - \text{cm} = 1.86 \text{Debyes}$$



This dipole moment corresponds to approximately one electron charge separated across the size of the molecule. Other polar molecules have similar dipole moments (NH_3 : 1.47 debyes, HCl : 1.03 debyes).

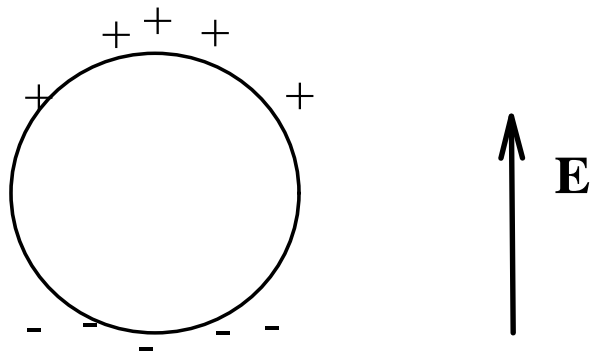
Atoms or Nuclei cannot have permanent dipole moments, since they are in states of good angular momentum l : the dipole moment of an electron in such a state vanishes. In contrast the molecules mentioned above are in sp hybrid orbitals, so that l is not a good quantum number, and thus a dipole moment is allowed.

3.2 Induced Dipoles

Atoms and molecules that lack permanent dipoles *can* have induced dipole moments when placed in an external electric field.

3.2.1 Static Models

We have already seen the effect of an external field inducing a dipole moment in a metallic sphere.



$$\Phi(\mathbf{x}) = \Phi_{external}(\mathbf{x}) + \Phi_{induced}(\mathbf{x}) \quad (34)$$

$$\Phi_{induced}(\mathbf{x}) = \frac{\mathbf{p} \cdot \mathbf{x}}{r^3} \quad \mathbf{p} = E_o a^3 \hat{\mathbf{z}} \quad (35)$$

The induced dipole moment is proportional to the external electric field.

If we define α to be the polarizability of the body, then

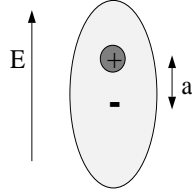
$$\mathbf{p} = \alpha \mathbf{E} \quad (36)$$

where, in this case, $\alpha = a^3$. We see that in general the polarizability of the order of magnitude of the volume of the body. Thus for an atom

$$\alpha_{atom} \approx \text{atomic volume} \approx 10^{-24} \text{cm}^3 \quad (37)$$

This is consistent with experiment.

To see what this means in realistic terms, consider an atom placed in a relatively large electric field $E = 100$ statvolts/cm.



Lets assume that the induced dipole moment is

$$\mathbf{p} = \alpha \mathbf{E} = e \mathbf{a} \quad (38)$$

where e is an electronic charge and $|\mathbf{a}|$ is the distance which separates the charge. Then

$$a = \frac{\alpha E}{e} \approx \frac{(10^{-24} \text{cm}^3)(100 \text{statvolts/cm})}{4.8 \times 10^{-10} \text{esu}} \approx 2 \times 10^{-13} \text{cm} \quad (39)$$

This is of nuclear dimensions (2 fermi's). Thus the atom is quite rigid to polarization. To have a distortion of the order of and Angstrom, we need a field of order $E = 5 \times 10^6 \text{statvolts/cm}$. This type of field strength is only available with a laser.

3.2.2 Dynamic Model

We may also calculate α for an atom using a simple dynamical model. Suppose that the electron is bound to the ion by a spring, so that, if displaced from equilibrium, it feels a restoring force.

$$\mathbf{F}_{restore} = -m\omega_o^2 \mathbf{x} \quad (40)$$

where m is the electronic mass, and ω_o the frequency of oscillation. If we apply an external electric field \mathbf{E} , the displacement \mathbf{x} of the electron

from equilibrium will grow until $\mathbf{F}_{restore}$ is equal and opposite to the electronic force on the electron.

$$-(-e\mathbf{E}) = -m\omega_o^2\mathbf{x} \quad \text{so} \quad \mathbf{x} = \frac{-e\mathbf{E}}{m\omega_o^2} \quad (41)$$

The induced dipole moment is then

$$\mathbf{p} = -e\mathbf{x} = \frac{e^2}{m\omega_o^2}\mathbf{E} = \alpha\mathbf{E} \quad (42)$$

So that the atomic polarizability is

$$\alpha = \frac{e^2}{m\omega_o^2} \quad (43)$$

Now, we expect that $\omega_o \approx$ angular frequency of oscillation, which is approximately the frequency of the light which is emitted by atoms. For a wavelength of 3000 Angstroms $\omega_o \approx 6 \times 10^{15} s^{-1}$, giving

$$\alpha \approx 6 \times 10^{-24} \text{cm}^3.$$

This is in accord with our previous estimate.

4 Dielectric Materials

The electrostatic properties of some insulating materials may be modeled by a collection of dipole molecules, each with a dipole moment. Higher order moments are usually neglected. Our main interest here is not in the dipole moments of individual atoms or molecules, but rather the dipole moments of atoms or molecules in a solid. In such a medium,

we expect that there will be no net permanent dipole moment. This is for two reasons:

- (1) If there is a permanent dipole moment in the atoms or molecules which make up the system, then the orientation of them will be random. Thus the average dipole moment $\langle \mathbf{p} \rangle$ will be zero.
- (2) If there is no permanent dipole moment of the component atoms or molecules, then in the absence of an external field, each will have no dipole moment, and thus the average dipole moment $\langle \mathbf{p} \rangle$ will also be zero

4.1 Statistical Mechanics

Now suppose that we do apply an external electric field, what will $\langle \mathbf{p} \rangle$ be then? What effect will thermal fluctuations have? We must again consider the ensemble of molecules for two different cases.

4.1.1 Induced dipoles

If each molecule has an induced dipole moment, then the Hamiltonian for each molecule is

$$U = -\mathbf{p} \cdot \mathbf{E} + \frac{1}{2}m\omega_0^2 r^2. \quad (44)$$

We can thus find the thermal average value of \mathbf{x} by averaging it over the distribution $\exp(-U/kT)$, where k is Boltzmann's constant and T is the temperature:

$$\langle \mathbf{x} \rangle = \frac{\int d^3x e^{-U/kT} \mathbf{x}}{\int d^3x e^{-U/kT}} \quad (45)$$

Let \mathbf{E} define the z -direction and have

$$\begin{aligned} \langle \mathbf{x} \rangle &= \frac{\int d^3x e^{(eEr \cos \theta - \frac{1}{2}m\omega_0^2 r^2)/kT} r \cos \theta \hat{\mathbf{z}}}{\int d^3x e^{(eEr \cos \theta - \frac{1}{2}m\omega_0^2 r^2)/kT}} \\ &= \frac{\int dz e^{(eEz - \frac{1}{2}m\omega_0^2 z^2)/kT} z \hat{\mathbf{z}}}{\int dz e^{(eEz - \frac{1}{2}m\omega_0^2 z^2)/kT}} \\ &= \frac{\int du e^{-\frac{1}{2}m\omega_0^2 u^2/kT} (u + eE/m\omega_0^2) \hat{\mathbf{z}}}{\int du e^{-\frac{1}{2}m\omega_0^2 u^2/kT}} \end{aligned} \quad (46)$$

where $u \equiv z - eE/m\omega_0^2$. The remaining integrals cancel nicely and we find that

$$\langle \mathbf{x} \rangle = (eE/m\omega_0^2) \hat{\mathbf{z}} \quad \mathbf{p} = (e^2 E/m\omega_0^2) \hat{\mathbf{z}} \quad \alpha = (e^2/m\omega_0^2) \quad (47)$$

the same as before we introduced thermal fluctuations in the separation. Thus thermal effects vanish.

4.1.2 Permanent Dipoles

For permanent dipoles (remember H_2O) we may do something similar using a Boltzmann distribution $\exp(-U/kT)$, as in the previous example but this time U is simply

$$U = -\mathbf{p} \cdot \mathbf{E} \quad (48)$$

with \mathbf{p} fixed in magnitude. Thus, letting \mathbf{E} define the z -direction again, we have

$$\langle \mathbf{p} \rangle = \frac{\int d\Omega e^{pE \cos \theta/kT} p \cos \theta \hat{\mathbf{z}}}{\int d\Omega e^{pE \cos \theta/kT}} = kT \hat{\mathbf{z}} \frac{d}{dE} \ln \left[\int_{-1}^1 du e^{pEu/kT} \right]. \quad (49)$$

The integral is easy; upon taking the derivative and simplifying the result insofar as possible, one finds

$$\langle \mathbf{p} \rangle = p \hat{\mathbf{z}} (\coth(pE/kT) - kT/pE). \quad (50)$$

As $T \rightarrow 0$, this becomes $p \hat{\mathbf{z}}$, meaning that the dipole is perfectly aligned with the field. For large T , $kT \gg pE$, we may expand the hyperbolic cosine and find the leading term

$$\langle \mathbf{p} \rangle = \frac{1}{3} \frac{p^2 E}{kT} \hat{\mathbf{z}}. \quad (51)$$

This is the most frequently encountered situation at *e.g.*, room temperature; it leads to a polarizability which is

$$\alpha = \frac{1}{3} \frac{p^2}{kT}. \quad (52)$$

4.1.3 Both

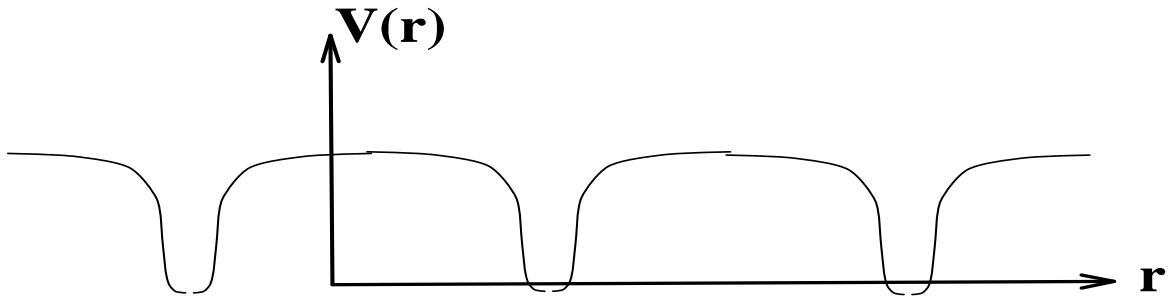
Finally, if a molecule has both a permanent moment and the possibility of being polarized, then the polarizability consists of both a temperature-independent term and one which varies inversely as the temperature,

$$\alpha = \frac{e^2}{m\omega_0^2} + \frac{p^2}{3kT}. \quad (53)$$

In each of the model calculations, where the dipoles are induced or permanent or both, the mean dipole moment induced in the material by an external field is proportional to that field. This is the basic assumption which we will use to explore the electrostatics of dielectric materials.

4.2 Macroscopic Electrostatics; Dielectrics

Before this section, we have considered only one kind of macroscopic material, conductors. Within conductors, there is no electric field, we said, because a conductor is an equipotential. If we had bothered to think a bit about that statement, we would have realized that it is a statement which applies only in some average sense. If one looks at the microscopic structure of a conductor or any other material, one finds electrons and nuclei with very strong electric fields reflecting the forces that act between these objects.



There is no electric field only in some *macroscopic* sense, that is, only if one averages over some region with a size large compared to an atomic size (which can still be much smaller than the size of a macroscopic probe whose size is at least of order 1μ).

Now we want to do the same with other materials, *i.e.*, nonconductors or insulators. Such materials are termed *dielectrics*. We concern ourselves again only with the macroscopic electric field, which is the true electric field averaged over some small domain, but it will no longer be zero, so that we must work a little harder to understand how to describe these materials.

Start by supposing that a piece of material is subjected to an externally applied electric field. This field will alter the multipole moments of the constituents of the material, which we shall call molecules (They could also be atoms or ions), yielding a net polarization of the material.

Now let's calculate the potential due to this polarization. If we regard it as a sum over the dipoles of each molecule, then

$$\Phi(\mathbf{x}) = \sum_j \frac{\mathbf{p}_j \cdot (\mathbf{x} - \mathbf{x}_j)}{|\mathbf{x} - \mathbf{x}_j|^3} \quad (54)$$

For now assume the the molecules are neutral, so that there is no

monopole term. In addition, assume that quadrupole and higher terms in the series are negligible. If we define the polarization vector

$$\mathbf{P}(\mathbf{x}) = \text{dipole moment per unit volume}, \quad (55)$$

then this becomes,

$$\begin{aligned} \Phi(\mathbf{x}) &= \int d^3x' \left[\frac{\mathbf{P}(\mathbf{x}') \cdot (\mathbf{x} - \mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^3} \right] \\ &= \int d^3x' \left[\mathbf{P}(\mathbf{x}') \cdot \nabla' \left(\frac{1}{|\mathbf{x} - \mathbf{x}'|} \right) \right]. \end{aligned} \quad (56)$$

where we have used the expression

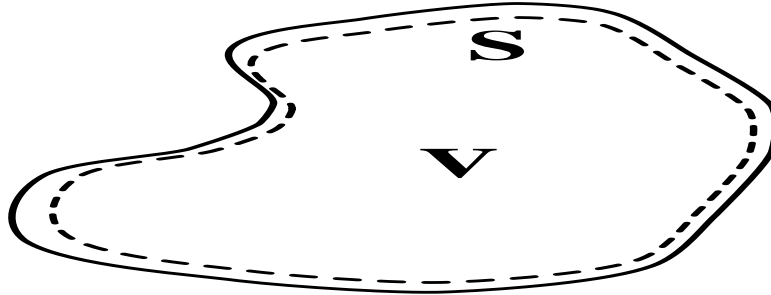
$$\frac{(\mathbf{x} - \mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^3} = -\nabla' \frac{1}{|\mathbf{x} - \mathbf{x}'|} = \nabla' \frac{1}{|\mathbf{x} - \mathbf{x}'|} \quad (57)$$

If we integrate by parts, we get the form

$$\Phi(\mathbf{x}) = \int d^3x' \nabla' \cdot \left(\frac{\mathbf{P}(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|} \right) - \int d^3x' \frac{(\nabla' \cdot \mathbf{P}(\mathbf{x}'))}{|\mathbf{x} - \mathbf{x}'|} \quad (58)$$

There are two ways to regard this expression. Assume we have a volume V with $\mathbf{P}(\mathbf{x}')$ finite inside and zero outside.

First case. Let V be bounded by a surface S just inside the volume. Then using the divergence theorem, the equation above becomes.



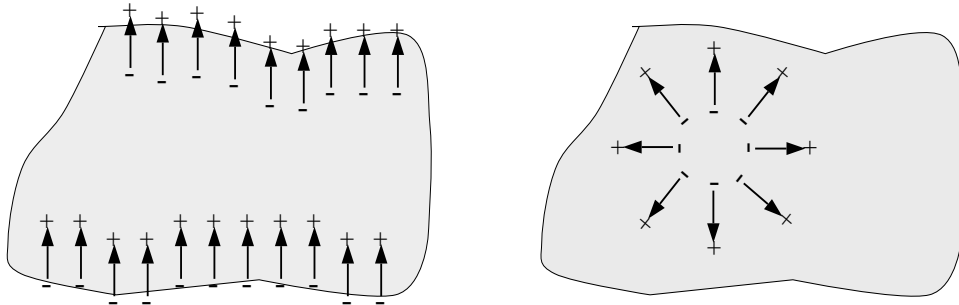
$$\Phi(\mathbf{x}) = \int_S d^2x' \frac{\mathbf{n}' \cdot \mathbf{P}(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|} - \int_V d^3x' \frac{(\nabla' \cdot \mathbf{P}(\mathbf{x}'))}{|\mathbf{x} - \mathbf{x}'|} \quad (59)$$

Thus we can define the surface and volume polarization charge densities:

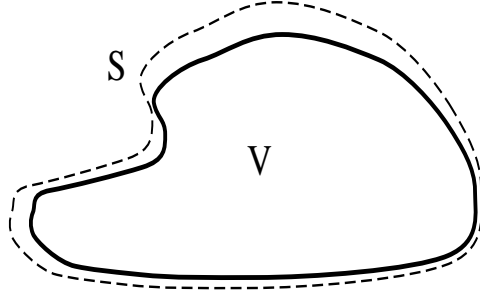
$$\sigma_p(\mathbf{x}) = \mathbf{P}_n(\mathbf{x}) \text{ or } \sigma_p(\mathbf{x}) = \mathbf{P}(\mathbf{x}) \cdot \mathbf{n} \quad (60)$$

$$\rho_p(\mathbf{x}) = -\nabla \cdot \mathbf{P}(\mathbf{x}) \quad (61)$$

These have simple physical interpretations. For example, in the figure below on the left, the material has a constant finite \mathbf{P} throughout its volume, so that at the surface, charge congregates since all of the dipoles are aligned. Also, in the figure on the right a certain location within a material the dipoles point radially outward (yielding a positive divergence). At the center of this region, where the tails of the dipoles are concentrated, there is an excess of negative charge (hence the $-$ sign in $\rho_p(\mathbf{x}) = -\nabla \cdot \mathbf{P}(\mathbf{x})$).



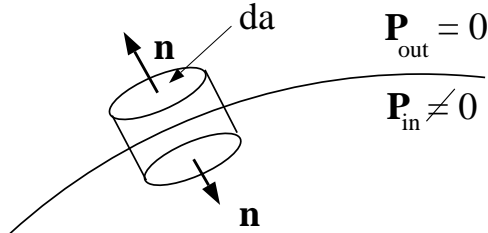
Second case. Let V be bounded by a surface S just *outside* of the region of finite polarization. Then



$$\Phi(\mathbf{x}) = \int_S d^2x' \frac{\mathbf{n}' \cdot \mathbf{P}(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|} - \int_V d^3x' \frac{(\nabla' \cdot \mathbf{P}(\mathbf{x}'))}{|\mathbf{x} - \mathbf{x}'|} \quad (62)$$

$$= \int_V d^3x' \frac{\rho_p(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|} \quad (63)$$

Although the surface charge does not appear explicitly, it still must be there. It may be obtained from the discontinuity in $\mathbf{P}(\mathbf{x})$ at the surface of the polarization region. To see this consider a small pill box enclosing a small section of the volume and surface of the polarized material. From the divergence theorem



$$\int_V d^3x \nabla \cdot \mathbf{P}(\mathbf{x}) = \int_S d^2x \mathbf{n} \cdot \mathbf{P}(\mathbf{x}), \quad (64)$$

or if the region is small enough

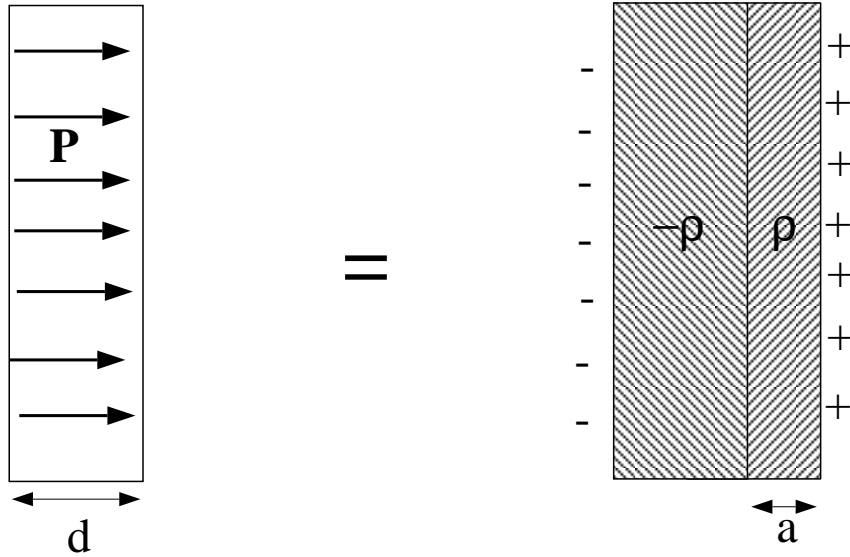
$$(-\mathbf{n} \cdot \mathbf{P}_{in} + \mathbf{n} \cdot \mathbf{P}_{out}) da = \nabla \cdot \mathbf{P} d^3x \quad (65)$$

so that

$$\mathbf{n} \cdot \mathbf{P} da = -\nabla \cdot \mathbf{P} d^3x, \quad (66)$$

Thus $-\nabla \cdot \mathbf{P}$ must have a delta-function at the surface, and we still have the surface polarization charge.

To understand the surface polarization charge, consider a uniformly polarized slab of dielectric.



$$\mathbf{p} = \begin{cases} \text{constant} & \text{inside} \\ 0 & \text{outside} \end{cases} \quad (67)$$

We may actually regard this as two overlapping slabs, one of uniform positive charge $+\rho$, and one of uniform negative charge $-\rho$, separated by a small distance \mathbf{a} . The whole is then electrically neutral, with uniform polarization

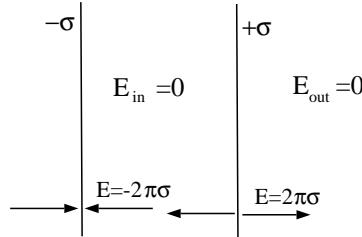
$$\mathbf{P} = \rho \mathbf{a} = \text{dipole moment per unit volume.} \quad (68)$$

The charge density is then

$$\rho_p = -\nabla \cdot \mathbf{P} = 0 \text{ Inside the slab} \quad (69)$$

$$\sigma_p = \mathbf{P} \cdot \mathbf{n} = \begin{cases} P > 0 & \text{on rhs} \\ -P < 0 & \text{on lhs} \end{cases} \quad (70)$$

The potential due to the slab is just that of two oppositely charged sheets separated by a distance d . The corresponding electric field is just obtained by summing that due to each sheet



$$\mathbf{E} = \begin{cases} 0 & \text{outside the slab} \\ -4\pi\mathbf{P} & \text{inside the slab} \end{cases} \quad (71)$$

(We used $E_{in} = 2\pi\sigma + 2\pi\sigma = 4\pi\sigma$ and $\sigma = P$)

4.2.1 Electric Displacement

Thus far we have assumed that the dielectric is neutral. If there are free charges present as well, then the total charge density is

$$\rho_{tot} = \rho_{free} - \nabla \cdot \mathbf{P} \quad (72)$$

Then as \mathbf{E} is generated by *all* charges, we have

$$\nabla \cdot \mathbf{E} = 4\pi\rho_{tot} = 4\pi(\rho_{free} - \nabla \cdot \mathbf{P}) \quad (73)$$

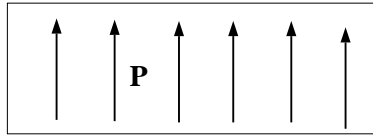
or,

$$\nabla \cdot (\mathbf{E} + 4\pi\mathbf{P}) = 4\pi\rho_{free} \quad (74)$$

The vector $(\mathbf{E} + 4\pi\mathbf{P})$ is generated by free charges only. We will define the *electric displacement* \mathbf{D} as this field

$$\mathbf{D} = \mathbf{E} + 4\pi\mathbf{P} \quad \nabla \cdot \mathbf{D} = 4\pi\rho_{free} \quad (75)$$

As an example, consider the uniformly polarized slab



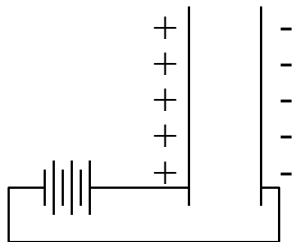
, here

$$\rho_{free} = 0 \quad \text{everywhere} \quad (76)$$

Thus $\mathbf{E} + 4\pi\mathbf{P} = 0$, and

$$\mathbf{E} = \begin{cases} 0 & \text{where } \mathbf{P} = 0 \\ -4\pi\mathbf{P} & \text{where } \mathbf{P} \neq 0 \end{cases} \quad (77)$$

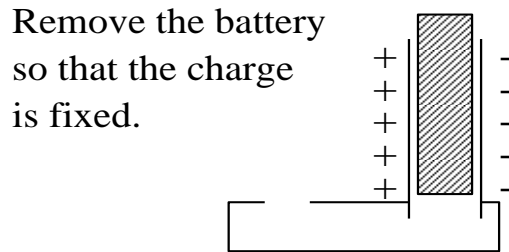
Another simple example using \mathbf{D} with the same slab geometry is the parallel plate capacitor.



If we have two charged plates in a vacuum, then $\mathbf{P} = 0$ everywhere, and

$$\mathbf{D} = \mathbf{E} = \begin{cases} 0 & \text{outside} \\ -4\pi\sigma_{free} & \text{between the plates} \end{cases}$$

If we now slide a dielectric slab between the plates, then we expect it to obtain a uniform polarization, giving rise to surface charges.



However, \mathbf{D} responds to only free charges, thus it is unchanged by the introduction of the dielectric slab. \mathbf{E} responds to all charges, so it *is* changed. Since $\mathbf{E} = \mathbf{D} - 4\pi\mathbf{P}$, we see that \mathbf{E} decreases in magnitude inside the dielectric, and since \mathbf{D} , \mathbf{P} , and \mathbf{E} are parallel:

$$\frac{E_{\text{dielectric}}}{E_{\text{vacuum}}} = \frac{E_{\text{dielectric}}}{D} = \frac{E_{\text{dielectric}}}{E_{\text{dielectric}} + 4\pi P}$$

4.2.2 Summary and Discussion

At this point a summary of the dielectric equations will be useful.

$\mathbf{P}(\mathbf{x}) =$ dipole moment per unit volume,

$$\Phi(\mathbf{x}) = \int d^3x' \frac{\rho_{total}(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|}$$

$$\rho_{total} = \rho_{free} - \nabla \cdot \mathbf{P}$$

$$\mathbf{E}(\mathbf{x}) = -\nabla\Phi(\mathbf{x}) \implies \nabla \times \mathbf{E}(\mathbf{x}) = 0$$

$$\mathbf{D}(\mathbf{x}) = \mathbf{E}(\mathbf{x}) + 4\pi\mathbf{P}(\mathbf{x})$$

$$\nabla \cdot \mathbf{D} = 4\pi\rho_{free}$$

Several points must be made in relation to these. The first is that they do not form a sufficient set from which we can solve for $\mathbf{E}(\mathbf{x})$ as we have no way of writing \mathbf{E} in terms of \mathbf{D} or vice versa. The defining relation does not help as we don't know \mathbf{P} . What is needed is a *constitutive relation* which can be of the form $\mathbf{D} = \mathbf{D}(\mathbf{E})$ or $\mathbf{P} = \mathbf{P}(\mathbf{E})$. If there is no nonanalytic behavior entering this relation, then one can expand components of \mathbf{P} as a power series in components of \mathbf{E} . If \mathbf{E} is not too large, then only the linear term in these expansions need be kept,

$$P_i = \sum_{j=1}^3 \chi_{ij} E_j. \quad (78)$$

where the nine numbers χ_{ij} are the components of the *electric susceptibility tensor*. When this is a good approximation, one says that the dielectric is *linear*. In disordered materials as well as highly ordered ones with a high degree of symmetry (cubic crystals, for example), this tensor reduces to a single non-zero number,

$$\chi_{ij} = \chi_e \delta_{ij}; \quad (79)$$

χ_e is called simply the electric susceptibility and such materials are said to be *isotropic*. Finally, if a material is uniform in its electrical properties, χ_e will be a constant, independent of position; then the

material is said to be *homogeneous*. If all of these things are true, the dielectric material is as simple as it can be.

For a linear, isotropic, homogeneous dielectric, the connection between \mathbf{E} and \mathbf{D} is

$$\mathbf{D}(\mathbf{x}) = \mathbf{E}(\mathbf{x}) + 4\pi\chi_e\mathbf{E}(\mathbf{x}) \equiv \epsilon\mathbf{E}(\mathbf{x}) \quad (80)$$

where

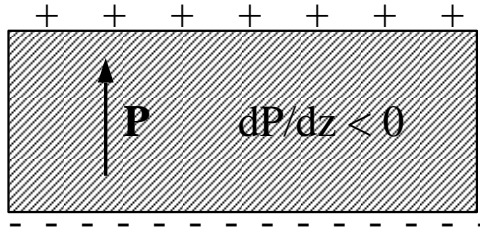
$$\epsilon = 1 + 4\pi\chi_e \quad (81)$$

is the *dielectric constant* of the material.

A second point is that the electric displacement is neither fish nor fowl, that is, neither field (force on a test charge) nor source. Look again at the integral expression for $\Phi(\mathbf{x})$; from it we see that the negative of the divergence of $\mathbf{P}(\mathbf{x})$ must be a (macroscopic) charge density; it is called the *polarization charge density*,

$$\rho_p(\mathbf{x}) = -\nabla \cdot \mathbf{P}(\mathbf{x}). \quad (82)$$

To see how this can be so, imagine a polarization which points in the z -direction and decreases in this direction so that its divergence is negative. Because of the variation of $\mathbf{P}(\mathbf{x})$, the molecules at smaller z are more polarized than those at slightly larger z , meaning that less positive charge “sticks out” on the larger- z side of the former than negative charge sticks out on the smaller- z side of the latter.



Hence there is a net positive charge density in the region between the two sets of molecules, and this is the polarization charge density. This argument leads one to believe that the total polarization charge must be zero. One can easily show by an application of the divergence theorem that it is indeed zero.

Having understood that the polarization leads to a charge density, how then may we understand the electric displacement? It is a linear combination of a macroscopic field (representing the force on a test charge) and of the polarization, whose divergence is a charge density. The polarization is itself source, being the dipole moment density of the constituent molecules of the material. Hence the displacement is neither field (\mathbf{E}) nor source (\mathbf{P}). Its usefulness lies in the fact that problems involving macroscopic electrostatics, and especially boundary value problems, are conveniently approached by making use of both the electric field and the electric displacement.

Another point that should be mentioned has to do with the higher multipole moments. We have seen how it is essential to keep the sources associated with the electric dipole moments of the molecules. What of

the higher multipole moments? One may show that they contribute negligibly at the macroscopic level.

Finally, there is the question of solving for the macroscopic electric field. Given a medium such that Eq. (40) is valid, we may use $\mathbf{D} = \epsilon\mathbf{E}$ and have the field equations

$$\nabla \times \mathbf{E}(\mathbf{x}) = 0 \text{ and } \nabla \cdot \mathbf{E}(\mathbf{x}) = 4\pi(\rho(\mathbf{x})/\epsilon); \quad (83)$$

these are the same as we have been working with right along except that the charge density is rescaled by a factor of $1/\epsilon$; hence all of the lore that we have learned may be applied to solve for the macroscopic field.

5 Boundary-Value Problems in Dielectrics

In this section we shall solve a few representative boundary-value problems involving dielectrics. Since $\nabla \times \mathbf{E} = 0$, $\mathbf{E} = -\nabla\Phi(\mathbf{x})$, $\mathbf{D} = \epsilon\mathbf{E}$, and $\nabla \cdot \mathbf{D} = 4\pi\rho$ in a dielectric, we may write

$$\nabla^2\Phi(\mathbf{x}) = 4\pi\rho/\epsilon.$$

Thus, all the methods we have learned (images, greens functions, series expansion etc.) will all work if properly modified.

There is of course the question of **boundary conditions**. At an interface between two materials (dielectric-vacuum, dielectric-dielectric, dielectric-conductor, etc.), we have a choice. We can either learn how

to solve for the field in a system with non-homogeneous properties, or we can split the system up into pieces in each of which the material properties are uniform and then solve a boundary value problem. The latter course is the simpler if the interfaces may be treated as abrupt.

The appropriate boundary or continuity conditions may be found from the basic differential equations for \mathbf{D} and \mathbf{E} . Applying the divergence theorem and Stokes' theorem as we did once before, one can show that the appropriate boundary conditions are

$$\nabla \cdot \mathbf{D} = 4\pi\rho_{\text{free}}$$

$$\int_V \nabla \cdot \mathbf{D} d^3\mathbf{x} = \int \mathbf{D} \cdot \mathbf{n} d^2\mathbf{x}$$

$$(\mathbf{D}_2 - \mathbf{D}_1) \cdot \mathbf{n} = 4\pi\sigma_{\text{free}}$$

$$\mathbf{D}_{2n} = \mathbf{D}_{1n} + 4\pi\sigma_{\text{free}}$$

$$\nabla \times \mathbf{E} = 0$$

$$\int_C d^2\mathbf{x} (\nabla \times \mathbf{E}) \cdot \mathbf{n} = \oint_C \mathbf{E} \cdot d\mathbf{l} = 0$$

$$(\mathbf{E}_2 - \mathbf{E}_1) \cdot \mathbf{t} = 0$$

$$\mathbf{E}_{2t} = \mathbf{E}_{1t}$$

$$[\mathbf{D}_2(\mathbf{x}) - \mathbf{D}_1(\mathbf{x})] \cdot \mathbf{n} = 4\pi\sigma \text{ and } [\mathbf{E}_2(\mathbf{x}) - \mathbf{E}_1(\mathbf{x})] \times \mathbf{n} = 0 \quad (84)$$

which say that the discontinuity in the normal component of \mathbf{D} is equal

to 4π times the surface charge density (*not* including the surface charge density arising from the polarization) and that the tangential component of \mathbf{E} is continuous. The unit normal in the equation for \mathbf{D} points into medium 2 from medium 1.

5.1 Example: Point Charge Near a Boundary

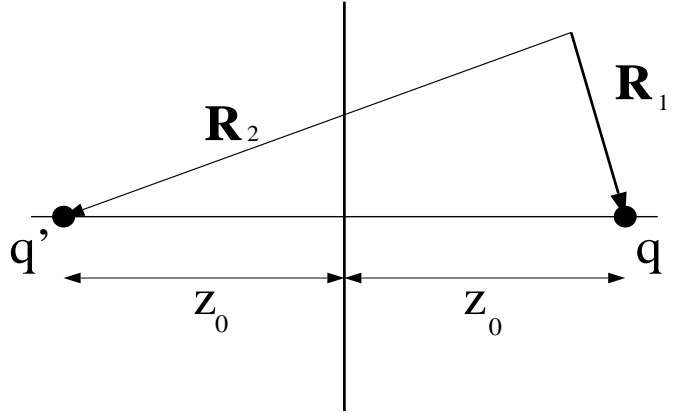
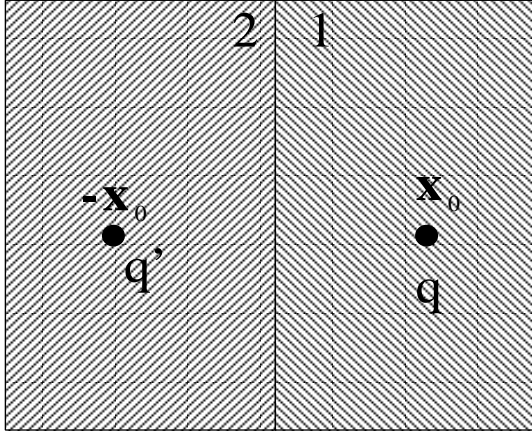
Consider that we have two dielectric materials; the first, with dielectric constant ϵ_1 , occupies the half-space $z > 0$, and the second, with ϵ_2 , occupies the half-space $z < 0$. Let there be a point charge q inside of the first dielectric at point $\mathbf{x}_0 = (0, 0, z_0)$.

Without a boundary, we can solve the problem easily. Since \mathbf{D} is unchanged by the dielectric,

$$\mathbf{D} = -\nabla \left(\frac{q}{R} \right) = \epsilon \mathbf{E} \quad \text{thus} \quad \mathbf{E} = -\frac{1}{\epsilon} \nabla \left(\frac{q}{R} \right) \quad \text{and} \quad \Phi = \frac{q}{\epsilon R} \quad (85)$$

where R is the distance between the charge and where the electric displacement is evaluated

We will try to solve for the electric field using the method of images. For the region $z > 0$, following our earlier success with this approach, let us locate an image charge q' at the image position $\mathbf{x}_i = (0, 0, -z_0)$.



The potential produced by these two charges, embedded in a medium which **everywhere** has the properties of the first medium, is

$$\Phi_1(\mathbf{x}) = \frac{1}{\epsilon_1} \left(\frac{q}{R_1} + \frac{q'}{R_2} \right), \quad (86)$$

where R_1 and R_2 are, respectively, the distances of the field point from x_0 and x_i ; this becomes our potential in the region $z > 0$ for the real system.

For the region $z < 0$, we imagine in the fictitious system that there is a charge q'' at the location of the real charge, embedded in a medium whose dielectric constant is everywhere ϵ_2 . The potential of such a system is

$$\Phi_2(\mathbf{x}) = \frac{1}{\epsilon_2} \frac{q''}{R_1}. \quad (87)$$

This becomes our potential in the region $z < 0$.

Now we try to pick the image charges in such a way that the boundary conditions are satisfied. these conditions involve the following

derivatives:

$$\frac{\partial}{\partial z} \left(\frac{1}{R_1} \right) \Big|_{z=0} = -\frac{\partial}{\partial z} \left(\frac{1}{R_2} \right) \Big|_{z=0} = \frac{z_0}{(\rho^2 + z_0^2)^{3/2}} \quad (88)$$

and

$$\frac{\partial}{\partial \rho} \left(\frac{1}{R_1} \right) \Big|_{z=0} = \frac{\partial}{\partial \rho} \left(\frac{1}{R_2} \right) \Big|_{z=0} = -\frac{\rho}{(\rho^2 + z_0^2)^{3/2}}. \quad (89)$$

Using these one finds that the condition of continuous normal component of \mathbf{D} or, $D_{1z} = D_{2z}$ leads to (since $\sigma_{free} = 0$)

$$q - q' = q'' \quad (90)$$

and that the condition of continuous tangential component of \mathbf{E} or $E_{1\rho} = E_{2\rho}$ leads to

$$\frac{1}{\epsilon_1}(q + q') = \frac{1}{\epsilon_2}q''. \quad (91)$$

The solution of these two linear equations is

$$\begin{aligned} q' &= \left(\frac{\epsilon_1 - \epsilon_2}{\epsilon_1 + \epsilon_2} \right) q \\ q'' &= \left(\frac{2\epsilon_2}{\epsilon_1 + \epsilon_2} \right) q. \end{aligned} \quad (92)$$

Hence the potential on the right side, $z > 0$, is

$$\Phi_1(\rho, z) = \frac{q}{\sqrt{\rho^2 + (z - z_0)^2}} + \frac{q(\epsilon_1 - \epsilon_2)/(\epsilon_1 + \epsilon_2)}{\sqrt{\rho^2 + (z + z_0)^2}} \quad (93)$$

while that on the left, $z < 0$, is

$$\Phi_2(\mathbf{x}) = \frac{2q\epsilon_2/(\epsilon_1 + \epsilon_2)}{\sqrt{\rho^2 + (z - z_0)^2}}. \quad (94)$$

consequences The constitutive relations yield several interesting results for this problem. Since $\nabla\mathbf{D} = 4\pi\rho$, where ρ is the free charge density, it must be that

$$\nabla \cdot \mathbf{D} = 0 \quad \text{except at the real charge}$$

Then as $\mathbf{D} = \epsilon\mathbf{E}$,

$$\nabla \cdot \mathbf{E} = 0 \quad \text{except at the real charge}$$

Then since $\mathbf{D} = \mathbf{E} + 4\pi\mathbf{P}$, it must be that

$$\nabla \cdot \mathbf{P} = 0 \quad \text{except at the real charge}$$

Thus the polarization charge density is zero except at the real charge! (this is consistent with the potential for a point charge retaining $1/r$ behavior).

This line of reasoning breaks down at the surface between the two dielectrics, since there $\nabla\epsilon \neq 0$ corresponding to the polarization surface charge $\mathbf{P} \cdot \mathbf{n}$ discussed earlier. Thus

$$\sigma_p = (\mathbf{P}_2 - \mathbf{P}_1) \cdot \mathbf{n} \quad (95)$$

where \mathbf{n} is the unit normal outward from medium 2 into medium 1. If we apply this to our present example, we first need to find the polarizations. These are given by, for $i = 1, 2$,

$$\mathbf{P}_i = \chi_i \mathbf{E}_i = \frac{\epsilon_i - 1}{4\pi} \mathbf{E}_i. \quad (96)$$

and so

$$\sigma_p = \frac{\epsilon_2 - 1}{4\pi} \mathbf{E}_{2z} - \frac{\epsilon_1 - 1}{4\pi} \mathbf{E}_{1z}, \quad (97)$$

evaluated at $z = 0$. Here,

$$\mathbf{E}_{1z} = -\frac{1}{\epsilon_1} \frac{(q - q')z_0}{(\rho^2 + z_0^2)^{3/2}} \quad (98)$$

and

$$\mathbf{E}_{2z} = -\frac{1}{\epsilon_2} \frac{q''d_0}{(\rho^2 + z_0^2)^{3/2}}. \quad (99)$$

The polarization surface-charge density is then

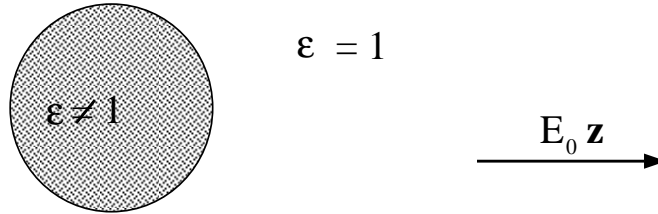
$$\begin{aligned} \sigma_p = -P_{1z} + P_{2z} &= \frac{z_0}{(\rho^2 + z_0^2)^{3/2} 4\pi} \left\{ \frac{\epsilon_1 - 1}{\epsilon_1 12} (-q + q') - \frac{\epsilon_2 - 1}{\epsilon_2} q'' \right\} \\ &= \frac{qz_0}{4\pi(\rho^2 + z_0^2)^{3/2}} \left\{ \frac{\epsilon_1 - 1}{\epsilon_1} \frac{2\epsilon_2}{\epsilon_1 + \epsilon_2} - \frac{\epsilon_2 - 1}{\epsilon_1} \frac{2\epsilon_1}{\epsilon_1 + \epsilon_2} \right\} \\ &= \frac{qz_0(\epsilon_1 - \epsilon_2)}{2\pi\epsilon_1(\epsilon_1 + \epsilon_2)(\rho^2 + z_0^2)^{3/2}}. \end{aligned} \quad (100)$$

An important limiting case is given by $\epsilon_1 = 1$ and $\epsilon_2 \rightarrow \infty$, in which case the material at $z < 0$ cannot support an electric field and behaves like a conductor. Then our system reduces to a point charge outside of a conductor, for which we already know that the answer is

$$\sigma_p = -\frac{qz_0}{2\pi(\rho^2 + z_0^2)^{3/2}}. \quad (101)$$

5.2 Example: Dielectric Sphere in a Uniform Field

Our second example is a dielectric sphere placed in a uniform externally applied field.



Rather than use the image charge method, this time we shall make use of an orthogonal function expansion. Because $\nabla \times \mathbf{E}(\mathbf{x}) = 0$ everywhere, we can write the electric field as the gradient of a scalar potential. Further, for a uniform medium $\mathbf{D}(\mathbf{x}) = \epsilon \mathbf{E}(\mathbf{x})$ with constant ϵ , so, from $\nabla \cdot \mathbf{D}(\mathbf{x}) = 4\pi\rho(\mathbf{x})$, we know that $\nabla \cdot \mathbf{E}(\mathbf{x}) = 0$ where there is no macroscopic charge density $\rho(\mathbf{x})$. Given that the radius of the sphere is a , we have such conditions for $r < a$ and for $r > a$. Hence the scalar potential satisfies the Laplace equation in these two regimes, and we can expand it in the usual way in spherical coordinates. The symmetries

in the problem imply that the solution is independent of ϕ , so we need to use a Legendre polynomial expansion. Thus, for $r < a$, we have

$$\Phi_{<}(r, \theta) = E_0 a \sum_{l=0}^{\infty} A_l \left(\frac{r}{a}\right)^l P_l(\cos \theta) \quad (102)$$

and, for $r > a$,

$$\Phi_{>}(r, \theta) = -E_0 r \cos \theta + E_0 a \sum_{l=0}^{\infty} B_l \left(\frac{a}{r}\right)^{l+1} P_l(\cos \theta). \quad (103)$$

The first term in the second of these expansions is the potential associated with the applied field; the others come from the sources induced on the dielectric sphere (polarization charge). The boundary conditions that must be applied are (i) tangential E and (ii) normal D continuous; these are

$$\left. \frac{\partial \Phi_{<}}{\partial \theta} \right|_{r=a} = \left. \frac{\partial \Phi_{>}}{\partial \theta} \right|_{r=a} \quad (104)$$

and

$$\epsilon \left. \frac{\partial \Phi_{<}}{\partial r} \right|_{r=a} = \left. \frac{\partial \phi_{>}}{\partial r} \right|_{r=a}, \quad (105)$$

where ϵ is the dielectric constant of the sphere. By proceeding with the solution in the by now familiar way for orthogonal function expansions, one finds that all A_l and B_l are zero except for $l = 1$. For $l = 1$ the conditions are

$$A_1 = -1 + B_1 \text{ and } \epsilon A_1 = -1 - 2B_1. \quad (106)$$

These are easily solved to yield

$$A_1 = -3/(2 + \epsilon) \text{ and } B_1 = (\epsilon - 1)/(\epsilon + 2). \quad (107)$$

Hence, the potential is

$$\Phi_{<}(r, \theta) = -\frac{3}{\epsilon + 2} E_0 r \cos \theta \quad (108)$$

and

$$\Phi_{>}(r, \theta) = -E_0 r \cos \theta + E_0 a \left(\frac{\epsilon - 1}{\epsilon + 2} \right) \left(\frac{a}{r} \right)^2 \cos \theta. \quad (109)$$

Notice that if $\epsilon \rightarrow \infty$, we recover the result for the conducting sphere.

From this result, and since $\Phi(\mathbf{x})_{dipole} = \mathbf{p} \cdot \mathbf{x}/r^3$, we can see that the sphere has a dipole moment which is

$$\mathbf{p} = E_0 a^3 \left(\frac{\epsilon - 1}{\epsilon + 2} \right) \hat{\mathbf{z}}. \quad (110)$$

The electric field inside of the sphere is a constant, and so is \mathbf{D} ,

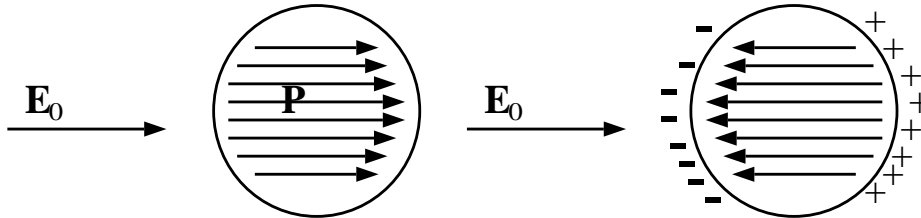
$$\mathbf{D} = \epsilon \mathbf{E} = \epsilon \left(\frac{3}{\epsilon + 2} \right) E_0 \hat{\mathbf{z}} \equiv \mathbf{E} + 4\pi \mathbf{P}, \quad (111)$$

so

$$4\pi \mathbf{P} = \mathbf{D} - \mathbf{E} = 3 \left(\frac{\epsilon - 1}{\epsilon + 2} \right) E_0 \hat{\mathbf{z}}, \quad (112)$$

or

$$\mathbf{P} = \frac{3}{4\pi} \left(\frac{\epsilon - 1}{\epsilon + 2} \right) E_0 \hat{\mathbf{z}}. \quad (113)$$



Dielectric sphere in a uniform field, showing the polarization on the left and the polarization charge with its associated, apposing, electric field on the right.

Although there is no macroscopic charge density anywhere in the system, there is polarization charge density. There is no volume polarization charge density because \mathbf{P} has zero divergence. However, there is a surface charge density; it is given by

$$\sigma_p = \mathbf{P} \cdot \hat{\mathbf{r}} = P_r = \frac{3}{4\pi} E_0 \left(\frac{\epsilon - 1}{\epsilon + 2} \right) \cos \theta. \quad (114)$$

As an application of the polarizable sphere problem, consider a water drop in air. For this system, roughly

$$\epsilon_{rmair} \approx 1 \quad \epsilon_{rmwater} \approx 81. \quad (115)$$

Water is a dielectric composed of permanent dipoles. The polarizability α ($\mathbf{p} = \alpha\mathbf{E}$) of the water drop is then

$$\alpha_{H_2O} = \frac{81 - 1}{81 + 2} a^3 \approx a^3 \quad (116)$$

Water drops look like metallic spheres to a static E-field.

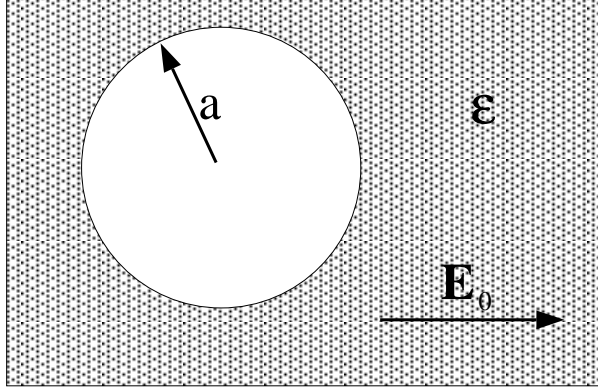
5.2.1 The Inverse Problem

The inverse problem of a dielectric with a spherical cavity is easy to solve because one has only to change ϵ into $1/\epsilon$ in the results found here. The reason is that the relative dielectric constant of the cavity to that of the surrounding medium is $1/\epsilon$. In this way we find

$$\Phi_{<}(r, \theta) = -\frac{3\epsilon}{2\epsilon + 1} r E_0 \cos \theta \quad (117)$$

and

$$\Phi_{>}(r, \theta) = -E_0 r \cos \theta - E_0 a \frac{\epsilon - 1}{2\epsilon + 1} \left(\frac{a}{r}\right)^2 \cos \theta. \quad (118)$$



5.3 Clausius-Mossotti equation

In writing α above note that the electric field in

$$\mathbf{p} = \alpha \mathbf{E}$$

is the external field *not including the field of the induced dipole itself*.

However, in

$$\mathbf{P} = \chi_e \mathbf{E}$$

the field \mathbf{E} *does* include the the field due to the dipoles in \mathbf{P} . If we can relate these two, we can calculate the relation between α (a microscopic quantity), and the macroscopic quantities χ_e and ϵ . If we define two different \mathbf{E} s

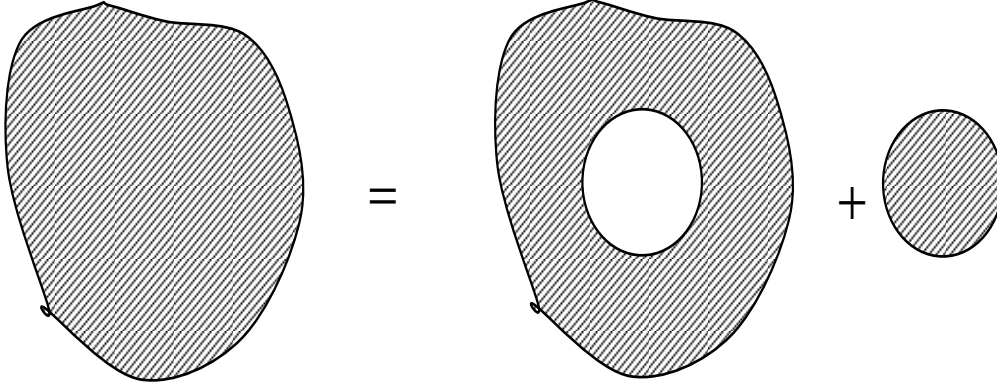
$$\mathbf{p} = \alpha \mathbf{E}_{loc} = \alpha(\text{electric field at the site of the molecule})$$

$$\mathbf{P} = \chi_e \mathbf{E}_{med} = \chi_e(\text{electric field in the medium})$$

If n is the number of molecules per unit volume, then $\mathbf{P} = n\mathbf{p}$ and

$$\mathbf{P} = n\alpha\mathbf{E}_{loc}.$$

With these equations we can find a relation between α and χ_e if we can relate \mathbf{E}_{loc} and \mathbf{E}_{med} . We note that \mathbf{E}_{loc} is the field at the site of a molecule if the molecule is removed. To calculate this, we will consider a spherical cavity in a dielectric medium. If the cavity were filled with dielectric, then the field at the center would be \mathbf{E}_{med} , so



$$\mathbf{E}_{med} = \mathbf{E}_{loc} + \mathbf{E}_{sphere}$$

where \mathbf{E}_{sphere} is the field at the center of a uniformly polarized dielectric sphere due *only* to the polarization. From Eq. (111) it is clear that the field inside the sphere (due to *both* the polarization and the external field) is

$$\mathbf{E}_{inside\ sphere} = \frac{3}{2 + \epsilon}\mathbf{E}_0 = \mathbf{E}_0 + \frac{1 - \epsilon}{2 + \epsilon}\mathbf{E}_0$$

Thus that due only to the polarization is

$$\mathbf{E}_{sphere} = \frac{1 - \epsilon}{2 + \epsilon}\mathbf{E}_0$$

Comparing this to Eq. (113) we see that

$$\mathbf{E}_{sphere} = -\frac{4\pi}{3}\mathbf{P}$$

Thus

$$\mathbf{E}_{loc} = \mathbf{E}_{med} - \mathbf{E}_{sphere} = \mathbf{E}_{med} + \frac{4\pi}{3}\mathbf{P}.$$

Now, if we multiply by $n\alpha$ and solve for \mathbf{P} , we get

$$\mathbf{P} = \frac{n\alpha}{1 - \frac{4\pi}{3}n\alpha}\mathbf{E}_{med}$$

Then since $\mathbf{P} = \chi_e\mathbf{E}_{med}$ and $\epsilon = 1 + 4\pi\chi_e$ we get

$$\frac{4\pi}{3}n\alpha = \frac{\epsilon - 1}{\epsilon + 2} \quad \text{Clasius - Mossotti Equation} \quad (119)$$

6 Electrostatic Energy in Dielectrics

In free space we derived the energy of a distribution of charge $\rho(\mathbf{x})$ by assembling the distribution little by little, bringing infinitesimal pieces of charge in from infinity. Following this reasoning we found that

$$W = \frac{1}{2} \int d^3x \rho(\mathbf{x})\Phi(\mathbf{x})$$

This is in general *not* true in the presence of dielectrics (however, as we will see, it may be true in some cases). In the presence of dielectrics work must also be done to induce polarization in the dielectric, and it is not clear if this work is included in the equation above.

When dielectrics are present we shall use a somewhat different argument (which still corresponds to the same procedure). Suppose that

there is initially some macroscopic charge density $\rho(\mathbf{x})$, potential $\Phi(\mathbf{x})$, and fields $\mathbf{E}(\mathbf{x})$ and $\mathbf{D}(\mathbf{x})$. The imagine that some infinitesimal change in the charge density, $\delta\rho(\mathbf{x})$, is made. To first order in $\delta\rho$, the change in energy of the system is

$$\delta W = \int_V d^3x \Phi(\mathbf{x})\delta\rho(\mathbf{x}) \quad (120)$$

where the integration is done over that region of space where the integrand is non-zero. The point is that this is the interaction energy of $\delta\rho(\mathbf{x})$ with the sources already present (and which produce $\Phi(\mathbf{x})$); the interaction energy of $\delta\rho$ with itself is second-order in small (infinitesimal) quantities. (this form is consistent with the fact that W is a natural thermodynamic function of the charges, not the potential).

The change in \mathbf{D} which arises as a consequence of the change $\delta\rho$ in the charge density is related to the latter by the equation $\nabla \cdot (\mathbf{D} + \delta\mathbf{D}) = 4\pi(\rho + \delta\rho)$, or

$$\nabla \cdot (\delta\mathbf{D}(\mathbf{x})) = 4\pi\delta\rho(\mathbf{x}), \quad (121)$$

so we may write the change in the energy as

$$\delta W = \int_V d^3x \Phi(\mathbf{x})\frac{1}{4\pi}(\nabla \cdot \delta\mathbf{D}(\mathbf{x})). \quad (122)$$

We next do an integration by parts in the by now familiar way.

$$\begin{aligned} \delta W &= \frac{1}{4\pi} \int_V d^3x \nabla \cdot (\Phi(\mathbf{x})\delta\mathbf{D}(\mathbf{x})) - \frac{1}{4\pi} \int_V d^3x \nabla\Phi(\mathbf{x}) \cdot \delta\mathbf{D}(\mathbf{x}) \\ &= \frac{1}{4\pi} \int_S d^2x \Phi(\mathbf{x})\delta\mathbf{D}(\mathbf{x}) \cdot \mathbf{n} + \frac{1}{4\pi} \int_V d^3x \mathbf{E}(\mathbf{x}) \cdot \delta\mathbf{D}(\mathbf{x}). \end{aligned} \quad (123)$$

The surface integral is zero for a localized charge distribution if V includes all space. Thus we have simply

$$\delta W = \frac{1}{4\pi} \int d^3x \mathbf{E}(\mathbf{x}) \cdot \delta \mathbf{D}(\mathbf{x}). \quad (124)$$

Now we must introduce some statement about the properties of the medium. If it is **linear** ($\mathbf{D} = \epsilon \mathbf{E}$), then

$$\mathbf{E} \cdot \delta \mathbf{D} = \mathbf{E} \cdot (\epsilon \delta \mathbf{E}) = \frac{1}{2} \epsilon \delta (\mathbf{E} \cdot \mathbf{E}) = \frac{1}{2} \delta (\mathbf{E} \cdot \mathbf{D})$$

so that

$$[\mathbf{E}(\mathbf{x}) \cdot \delta \mathbf{D}(\mathbf{x})] = \frac{1}{2} \delta [\mathbf{E}(\mathbf{x}) \cdot \mathbf{D}(\mathbf{x})]. \quad (125)$$

and so

$$\delta W = \frac{1}{8\pi} \int d^3x \delta [\mathbf{E}(\mathbf{x}) \cdot \mathbf{D}(\mathbf{x})]. \quad (126)$$

If we now integrate from zero field up to the final fields (a *functional integration*),

$$W = \frac{1}{8\pi} \int d^3x \int_0^{\mathbf{D}} \delta [\mathbf{E}(\mathbf{x}) \cdot \mathbf{D}(\mathbf{x})]$$

we find

$$W = \frac{1}{8\pi} \int d^3x \mathbf{E}(\mathbf{x}) \cdot \mathbf{D}(\mathbf{x}). \quad (127)$$

This result is valid **only** for linear media.

There are several amusing consequences obtainable from this relation. First, by writing $\mathbf{E}(\mathbf{x}) = -\nabla \Phi(\mathbf{x})$ and integrating by parts, we obtain

$$W = \frac{-1}{8\pi} \int d^3x \nabla \cdot (\Phi(\mathbf{x}) \mathbf{D}(\mathbf{x})) + \frac{1}{8\pi} \int d^3x \Phi(\mathbf{x}) \nabla \cdot \mathbf{D}(\mathbf{x})$$

Through the divergence theorem, the first term yields a surface term which vanishes at infinity. The second term becomes

$$W = \frac{1}{2} \int d^3x \rho(\mathbf{x})\Phi(\mathbf{x}); \quad (128)$$

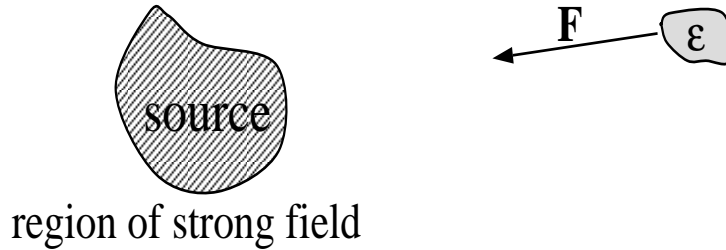
Thus for a **linear** dielectric, the original formula is valid.

6.1 Force on a Dielectric

From the above, it is clear that W , as written, is a function of the free charges, their positions, and of the positions of the dielectrics through $\epsilon(\mathbf{x})$ (which may vary from point to point). Thus we may write

$$W(\mathbf{x}, \rho) = \frac{1}{8\pi} \int d^3x \frac{|\mathbf{D}(\mathbf{x})|^2}{\epsilon(\mathbf{x})} \quad \text{linear only} \quad (129)$$

From this, it is clear that if the free charges (which produce \mathbf{D}) are fixed, and we move one of the dielectrics, then the energy is *reduced* if the change makes ϵ increase in the region where $\mathbf{D}(\mathbf{x})$ is finite. In particular, *the energy is reduced by having a dielectric move from a region of low field to one of high field. Thus the force on such a linear dielectric must always be such as to draw it into a region of greater fields.*



If the free charges are held fixed, then since W depends on the positions and magnitude of the charges and dielectrics, it follows that the force on a dielectric is

$$F_\eta = - \left(\frac{\partial W}{\partial \eta} \right)_Q \quad (130)$$

where F_η is the η -component of the force on the dielectric. This is because the most stable state of the system is that with the minimum W .

In this calculation, it is important that the energy was a natural function of the charges and positions of the charges and dielectrics. Then we could evaluate the total differential

$$dW = \left(\frac{\partial W}{\partial \eta} \right)_Q d\eta + \left(\frac{\partial W}{\partial Q} \right)_\eta dQ$$

to obtain the force. This is analogous to the situation in elementary thermodynamics where the energy U is a natural function of the volume and temperature $U(V, T)$. If we wanted to obtain a potential which was a function of the entropy S and V (suppose for example S is change in

such a way as to keep T fixed, i.e. a system in a heat bath), then we made a Legendre transformation $F = U - TS$, and the most stable state of the system is that with the minimum F (For an elegant discussion of elementary thermodynamics see *Thermodynamics* by Enrico Fermi, about 77pp.)

Thus, since W is a natural function of the positions and charges, it is not appropriate for the case where the potentials are held fixed. We need the potential which is a natural function of the positions and potentials. As in the paragraph above, the way to remedy this is a Legendre transformation to a new function W' , defined by

$$W' = W - \frac{1}{4\pi} \int d^3x \mathbf{E}(\mathbf{x}) \cdot \mathbf{D}(\mathbf{x}) \quad (131)$$

This is a general expression (not just for the linear case) where $\mathbf{E} = -\nabla\Phi(\mathbf{x})$ is a natural function of the potentials, and \mathbf{D} is a function of the charges (since $\nabla \cdot \mathbf{D} = 4\pi\rho$). A differential change in W' is given by

$$\delta W' = \delta W - \frac{1}{4\pi} \int d^3x \mathbf{E}(\mathbf{x}) \cdot \delta\mathbf{D}(\mathbf{x}) - \frac{1}{4\pi} \int d^3x \mathbf{D}(\mathbf{x}) \cdot \delta\mathbf{E}(\mathbf{x}). \quad (132)$$

Then, since

$$\begin{aligned} \delta W &= \frac{1}{4\pi} \int d^3x \mathbf{E}(\mathbf{x}) \cdot \delta\mathbf{D}(\mathbf{x}) \\ \delta W' &= -\frac{1}{4\pi} \int d^3x \mathbf{D}(\mathbf{x}) \cdot \delta\mathbf{E}(\mathbf{x}) \end{aligned} \quad (133)$$

Then since $\mathbf{E}(\mathbf{x})$ is a natural function of the potential, W' is a natural function of potentials and positions, as desired.

Thus the force on a dielectric in the presence of fixed potentials (i.e. conductors connected to a battery) is

$$F_\eta = - \left(\frac{\partial W'}{\partial \eta} \right)_\Phi \quad (134)$$

In the linear case, we can evaluate this in terms of W , since

$$W' = W - \frac{1}{4\pi} \int d^3x \mathbf{E}(\mathbf{x}) \cdot \mathbf{D}(\mathbf{x}) = -\frac{1}{8\pi} \int d^3x \mathbf{E}(\mathbf{x}) \cdot \mathbf{D}(\mathbf{x}) = -W \quad (135)$$

Thus in **the linear case only**,

$$F_\eta = + \left(\frac{\partial W}{\partial \eta} \right)_\Phi \quad (136)$$

6.2 Forces on a Dielectric Revisited

These force formulae may also be derived in a more pedestrian manner.

We can derive one for the change in a system's energy when a piece of dielectric is moved from one place to another under conditions of constant macroscopic charge density. Consider that initially the macroscopic fields, charge density, potential, and polarization are \mathbf{E}_0 , W_0 , \mathbf{D}_0 , ρ_0 , Φ_0 , and \mathbf{P}_0 . Let the final ones have subscript 1 instead of 0. Then

$$W_1 = \frac{1}{8\pi} \int d^3x \mathbf{E}_1(\mathbf{x}) \cdot \mathbf{D}_1(\mathbf{x}) \text{ and } W_0 = \frac{1}{8\pi} \int d^3x \mathbf{E}_0(\mathbf{x}) \cdot \mathbf{D}_0(\mathbf{x}), \quad (137)$$

so

$$\Delta W \equiv W_1 - W_0 = \frac{1}{8\pi} \int d^3x [\mathbf{E}_1 \cdot \mathbf{D}_1 - \mathbf{E}_0 \cdot \mathbf{D}_0]. \quad (138)$$

By adding and subtracting identical terms, we can turn this expression into

$$\Delta W = \frac{1}{8\pi} \int d^3x [\mathbf{E}_0 + \mathbf{E}_1] \cdot [\mathbf{D}_1 - \mathbf{D}_0] + \frac{1}{8\pi} \int d^3x [\mathbf{E}_1 \cdot \mathbf{D}_0 - \mathbf{E}_0 \cdot \mathbf{D}_1]. \quad (139)$$

By doing an integration by parts (in the usual way), one can show that the first term is zero if $\rho_1 = \rho_0$, so we have

$$\Delta W = \frac{1}{8\pi} \int d^3x [\mathbf{E}_1 \cdot \mathbf{D}_0 - \mathbf{E}_0 \cdot \mathbf{D}_1]. \quad (140)$$

As an example of the use of this formula, imagine that a dielectric having $\epsilon = \epsilon_1$ is moved in from infinity to occupy some domain V where formerly there was empty space. Everywhere else there is vacuum. Then $\mathbf{D}_1 = \epsilon_1 \mathbf{E}_1$ in V , and $\mathbf{D}_1 = \mathbf{E}_1$ elsewhere. Also, $\mathbf{E}_0 = \mathbf{D}_0$ everywhere. Our formula for the change in energy gives

$$\Delta W = -\frac{1}{8\pi} \int_V d^3x (\epsilon_1 - 1) \mathbf{E}_1 \cdot \mathbf{E}_0 = -\frac{1}{2} \int d^3x \mathbf{P}_1 \cdot \mathbf{E}_0. \quad (141)$$

This is the energy of the dielectric object placed in an external field \mathbf{E}_0 . The factor of 1/2 distinguishes it from the energy of a permanent dipole placed in an external field which we derived earlier. It has to do with the fact that in the present case the field has to do work to polarize the dielectric in the first place.

We may also devise a formula for the force on a piece of dielectric. In the case that the (macroscopic) charge is fixed, no work is done moving any charge and so we have a conservative system in the sense that the change in the field energy must be equal to the work that an external agent does on the dielectric when the latter is moved. This force (recall our earlier arguments of this kind) is equal and opposite to the electric field force on the dielectric so we wind up concluding that

$$F_\eta = - \left(\frac{\partial W}{\partial \eta} \right)_Q \quad (142)$$

where the “ Q ” means that the derivative with respect to displacement in the η direction is taken at constant sources (constant $\rho(\mathbf{x})$). The result is the force in the direction of η , and this is the usual expression for a conservative system.

A more difficult case is one in which there is an external source of energy. A very common case of this kind which frequently arises in electrostatics involves a set of conducting objects or surfaces on which the macroscopic charge ρ resides, and keeping these surfaces at fixed potentials when the dielectric is moved. The latter is easily achieved by connecting the conductors to fixed voltage sources (batteries). To see what happens, imagine making a small displacement of the dielectric in two steps. First, move it by $d\eta$ while maintaining $\rho(\mathbf{x})$ fixed. Then restore $\Phi(\mathbf{x})$ to its original value at those points where there is non-zero macroscopic charge density by adjusting this charge density as necessary. We can calculate the change in field energy during either of these steps by applying the general formula for a linear system

$$W = \frac{1}{2} \int d^3x \rho(\mathbf{x})\Phi(\mathbf{x}), \quad (143)$$

which gives, for small changes in ρ and Φ ,

$$\delta W = \frac{1}{2} \int d^3x [\delta\rho(\mathbf{x})\Phi(\mathbf{x}) + \rho(\mathbf{x})\delta\Phi(\mathbf{x})]. \quad (144)$$

In the first step described above, there is no change in ρ , so

$$\delta W_1 = \frac{1}{2} \int d^3x \rho(\mathbf{x})\delta\Phi_1(\mathbf{x}). \quad (145)$$

This is the same as δW_Q since charges are fixed in this step:

$$\delta W_1 = \delta W_Q. \quad (146)$$

In step 2, $\rho(\mathbf{x})$ is adjusted so that $\Phi(\mathbf{x})$ returns to its initial value everywhere where the charge density does not vanish. In this step

$$\delta W_2 = \frac{1}{2} \int d^3x \rho(\mathbf{x}) \delta \Phi_2(\mathbf{x}) + \frac{1}{2} \int d^3x \delta \rho(\mathbf{x}) \Phi(\mathbf{x}). \quad (147)$$

However, at points where $\rho(\mathbf{x})$ is non-zero, $\delta \Phi_2(\mathbf{x}) = -\delta \Phi_1(\mathbf{x})$ because in the second step we restore the potential to its original value at these points. Hence we can rewrite δW_2 as

$$\delta W_2 = -\frac{1}{2} \int d^3x \rho(\mathbf{x}) \delta \Phi_1(\mathbf{x}) + \frac{1}{2} \int d^3x \delta \rho(\mathbf{x}) \Phi(\mathbf{x}). \quad (148)$$

There is a second way to see what δW_2 is; in this step we make an infinitesimal change in the charge density, $\delta \rho(\mathbf{x})$, and the change in energy accompanying this adjustment is, to first order in infinitesimals,

$$\delta W_2 = \int d^3x \Phi(\mathbf{x}) \delta \rho(\mathbf{x}). \quad (149)$$

By comparing the two equations we have for δW_2 , we learn that

$$\frac{1}{2} \int d^3x \Phi(\mathbf{x}) \delta \rho(\mathbf{x}) = -\frac{1}{2} \int d^3x \rho(\mathbf{x}) \delta \Phi_1(\mathbf{x}). \quad (150)$$

Using this relation in Eq. (113), we find that

$$\delta W_2 = - \int d^3x \rho(\mathbf{x}) \delta \Phi_1(\mathbf{x}) \quad (151)$$

and this is the same as $-2\delta W_1$. Consequently we can say that the total change in energy, which we shall call δW_V (the “V” signifies constant potentials at points where $\delta\rho(\mathbf{x})$ is non-zero), is

$$\delta W_V = \delta W_1 + \delta W_2 = -\delta W_1 = -\delta W_Q. \quad (152)$$

Consequently,

$$F_\eta = -\left(\frac{\partial W}{\partial \eta}\right)_Q = +\left(\frac{\partial W}{\partial \eta}\right)_V. \quad (153)$$

In other words, if we can calculate the energy as a function of η , the position of the dielectric, at constant potentials where $\rho(\mathbf{x}) \neq 0$, we can find the force on the dielectric by taking the **positive** derivative of this energy with respect to the dielectric’s position.

7 Example: Dielectrophoresis

A spherical dielectric particle of radius a and $\epsilon = \epsilon_1$ is placed in a dielectric fluid ($\epsilon_2 \neq \epsilon_1$) contained within an annulus with conducting walls.

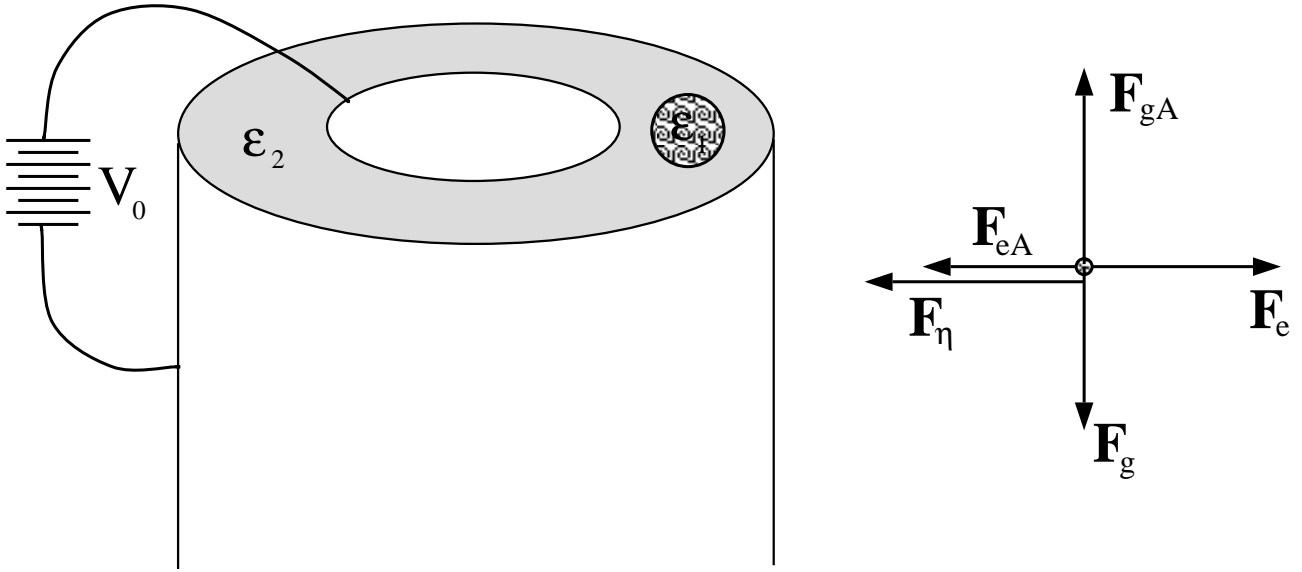


Fig.1. Dielectric fluid and particle between two conducting cylinders

The annulus is maintained at a relative potential V_0 . Assuming that $a \ll$ any other dimension in the problem, and that the densities of the particle and fluid are the same:

- Show that the **net** force on the particle is

$$\mathbf{F}_{net} = -\frac{1}{3}a^3AV_0^2r^{-3}[\ln(R_{out}/R_{in})]^{-2}\hat{\mathbf{r}},$$

where

$$A = \frac{\epsilon_2(4\epsilon_2 + 5) - \epsilon_1(\epsilon_2 - 1)}{(2\epsilon_2 + \epsilon_1)^2}(\epsilon_1 - \epsilon_2).$$

- Discuss how \mathbf{F}_{net} depends upon (1) ϵ_2 relative to ϵ_1 and (2) ϵ_1 for fixed ϵ_2 .
- The drift velocity \mathbf{v} of the particle is given by

$$\mathbf{F}_{net} = 6\pi\eta a\mathbf{v},$$

- (1) Discuss the dependence of \mathbf{v} on a and ϵ_1 for fixed ϵ_2 , and (2) Suggest possible uses for this setup (dielectrophoresis).

Solution. To solve this we must first find the field between the cylinders in the absence of the particle. To accomplish this, we use Gauss' law, the constitutive relations, and the fact that \mathbf{D} and \mathbf{E} are purely radial.

$$\nabla \cdot \mathbf{D} = 4\pi\rho \quad ; \quad \mathbf{D} = \epsilon_2\mathbf{E}$$

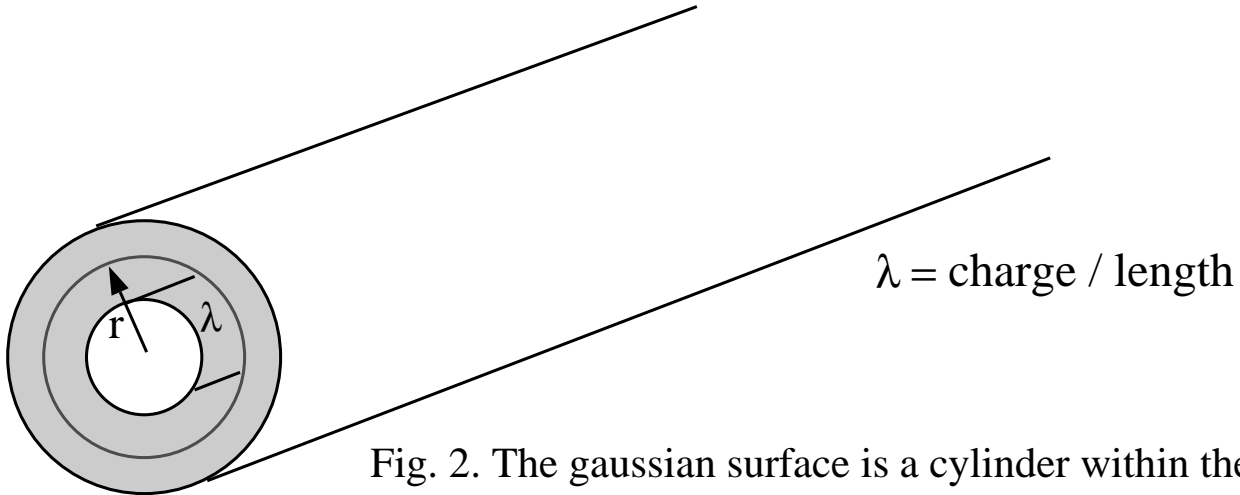


Fig. 2. The gaussian surface is a cylinder within the annulus

$$l2\pi r\epsilon_2 E_r = 4\pi\lambda l \quad ; \quad E_r = \frac{2\lambda}{\epsilon_2 r}$$

$$V_0 = \int_{R_{in}}^{R_{out}} dr \frac{\partial\Phi}{\partial r} = -\frac{2\lambda}{\epsilon_2} \ln(R_{out}/R_{in})$$

When we solve for $\lambda(V_0)$, we find that

$$\mathbf{E}_{cyl} = -\frac{V_0}{r \ln(R_{out}/R_{in})} \hat{\mathbf{r}}$$

Now we must solve for the field within the volume of the particle. Since $a \ll R$, we will assume that \mathbf{E}_{cyl} is essentially uniform over the

diameter of the particle. This problem then becomes very similar to one we solved in class, that of a dielectric sphere in a uniform external field \mathbf{E}_0 .



Fig. 3. A dielectric sphere in a uniform external field.

Recall that for this problem,

$$\mathbf{E}_{sphere} = \frac{3}{\epsilon + 2} \mathbf{E}_0$$

where \mathbf{E}_{sphere} is the electric field within the sphere. However, since the dielectric constant only enters through the boundary condition in a relative way

$$D_{1n} = D_{2n} \Rightarrow \epsilon_1 E_{1n} = \epsilon_2 E_{2n} \Rightarrow \frac{\epsilon_1}{\epsilon_2} E_{1n} = E_{2n},$$

it must be that

$$\mathbf{E}_{sphere} = \frac{3\epsilon_2}{2\epsilon_2 + \epsilon_1} \mathbf{E}_{cyl}.$$

The polarization of the sphere is then

$$\mathbf{P} = \frac{\epsilon_1 - 1}{4\pi} \mathbf{E}_{sphere} = \frac{1}{4\pi} \frac{3\epsilon_2(\epsilon_1 - 1)}{2\epsilon_2 + \epsilon_1} \mathbf{E}_{cyl},$$

which corresponds to a dipole moment of the sphere

$$\mathbf{p} = \frac{4\pi}{3} a^3 \mathbf{P} = a^3 \frac{\epsilon_2(\epsilon_1 - 1)}{2\epsilon_2 + \epsilon_1} \mathbf{E}_{cyl}$$

Thus the electrostatic force on the sphere is given by

$$\mathbf{F}_{elec} = (\mathbf{p} \cdot \nabla) \mathbf{E}_{sphere} = a^3 \frac{3\epsilon_2^2(\epsilon_1 - 1)}{(2\epsilon_2 + \epsilon_1)^2} (\mathbf{E}_{cyl} \cdot \nabla) \mathbf{E}_{cyl}$$

$$\mathbf{F}_{elec} = (\mathbf{p} \cdot \nabla) \mathbf{E}_{sphere} = \frac{a^3 3\epsilon_2^2(\epsilon_1 - 1)}{2 (2\epsilon_2 + \epsilon_1)^2} \nabla |\mathbf{E}_{cyl}|^2$$

In addition to this force, there is an electric Archimedes force. This force is due to the fact that the particle displaces some of the dielectric fluid. We can calculate it by applying Archimedes principle just like we do for gravitational forces. The electric Archimedes force is minus the force that the sphere of displaced fluid experiences.

$$\mathbf{F}_{elec-Arch} = -\mathbf{F}_{elec}(\epsilon_1 \rightarrow \epsilon_2)$$

As indicated, we can get this force by replacing ϵ_1 by ϵ_2 in the force equation above.

$$\mathbf{F}_{elec-Arch} = -\frac{a^3}{6} (\epsilon_2 - 1) \nabla |\mathbf{E}_{cyl}|^2$$

Thus the net electric force on the sphere is

$$\mathbf{F}_{net} = \frac{a^3}{2} \left[\frac{3\epsilon_2^2(\epsilon_1 - 1)}{(2\epsilon_2 + \epsilon_1)^2} - \frac{\epsilon_2 - 1}{3} \right] \nabla |\mathbf{E}_{cyl}|^2.$$

Now evaluating the gradient, we get

$$\mathbf{F}_{net} = -\frac{a^3 V_0^3}{r^3 (\ln(R_{out}/R_{in}))^2} \left[\frac{3\epsilon_2^2(\epsilon_1 - 1)}{(2\epsilon_2 + \epsilon_1)^2} - \frac{\epsilon_2 - 1}{3} \right],$$

or, after a bit of algebra,

$$\mathbf{F}_{net} = -\frac{a^3 A V_0^3}{3r^3 (\ln(R_{out}/R_{in}))^2}$$

where

$$A = \frac{\epsilon_2(4\epsilon_2 + 5) - \epsilon_1(\epsilon_2 - 1)}{(2\epsilon_2 + \epsilon_1)^2}(\epsilon_1 - \epsilon_2)$$

Interpretation. This setup has practical applications for sorting bits of dielectric particles with different ϵ_1 . Thus we should consider A for fixed ϵ_2 and different ϵ_1 of the different particles. If $A > 0$, then the force is inward, toward stronger fields, and if $A < 0$, then the force is outward, toward weaker fields. The zeroes of A occur when

$$\epsilon_2 = \epsilon_1 \quad ; \quad \epsilon_1 = \frac{\epsilon_2(4\epsilon_2 + 5)}{\epsilon_2 - 1},$$

and the maximum for fixed ϵ_2

$$\frac{\partial A}{\partial \epsilon_1} = 0 \Rightarrow \epsilon_1 = 2(\epsilon_2 + 1).$$

Now lets consider a numerical example. Let $\epsilon_2 = 2.0$ and vary ϵ_1 .

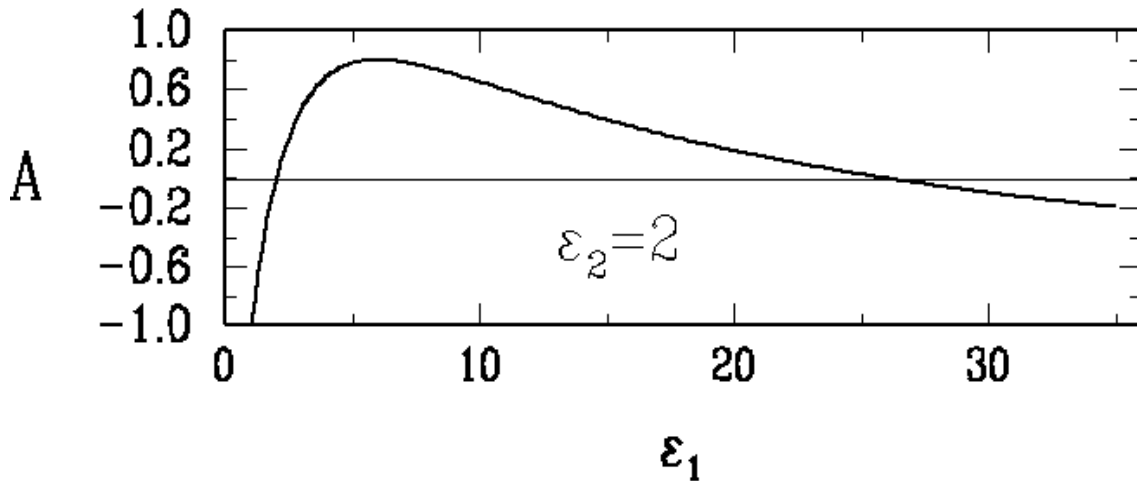


Fig. 4. Sketch of A

As shown in the figure, A is not a monotonic function of ϵ_1 . In fact there are three regions

$\epsilon_1 < \epsilon_2$	force is outward
$\epsilon_2 < \epsilon_1 < \epsilon_2(4\epsilon_2 + 5)/(\epsilon_2 - 1)$	force is inward
$\epsilon_1 > \epsilon_2(4\epsilon_2 + 5)/(\epsilon_2 - 1)$	force is outward

Thus, mildly and highly polar particles move out, while particles of intermediate polarizability move in. At first it seems that this makes no sense; however, consider the following explanation:

Since the potential within the annulus is held fixed by an external source, the appropriate thermodynamic potential for this system is

$$W'(r, \Phi) = -\frac{1}{8\pi} \int_V d^3x \epsilon(\mathbf{x}) |\mathbf{E}(\mathbf{x})|^2$$

The most stable state of the system is the one which minimizes this potential. Thus the most stable state of the system is obtained by having the largest field where $\epsilon(\mathbf{x})$ is largest. Reconsider the three regions.

- (1) $\epsilon_1 < \epsilon_2$ The particle has a lower ϵ than that of the fluid, thus it is expelled to regions of low field. Thus, $\epsilon(\mathbf{x})$ is maximized where \mathbf{E} is large.
- (2) $\epsilon_2 < \epsilon_1 < \epsilon_2(4\epsilon_2 + 5)/(\epsilon_2 - 1)$ Now the particle has a higher ϵ than the fluid

thus it moves to regions of large field

$$(3) \quad \epsilon_1 > \epsilon_2(4\epsilon_2 + 5)/(\epsilon_2 - 1)$$

The field inside the particle goes to zero as $\epsilon_2 \rightarrow \infty$ (the sphere approaches a metal), and the potential is minimized by having the sphere go to regions of low field.

Now consider the motion of the particle. The drift velocity of the particle is given by

$$\mathbf{F}_{net} = 6\pi\eta\mathbf{v}.$$

If we use the following reasonable parameters

parameter	value	cgs value
a	0.1 mm	0.01 cm
ϵ_1	4.6	4.6
ϵ_2	2.3	2.3
η	6.5×10^{-4} (MKSA)	6.5×10^{-3} cgs
R_{in}	0.5 mm	0.05 cm
R_{out}	1 cm	1 cm
V_0	4.0×10^3 volts	13.34 statvolts
r	0.5 cm	0.5 cm

The value of ϵ_2 corresponds to that of organic solvents, in this case Benzene. Solving with these parameters, we find that $v = 0.031$ cm./sec.

Slow.

We could use this setup to separate living cells from dead ones due to the very different water content (since $\epsilon_{water} \approx 81$).

A Multipole Expansion: with Spherical Harmonics

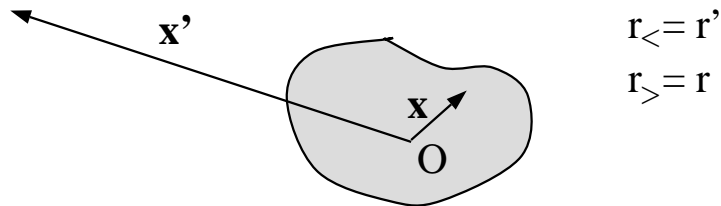
In this appendix, we will discuss the multipole expansion as it is done in Jackson.

Consider the potential $\Phi(\mathbf{x})$ produced by some localized charged distribution $\rho(\mathbf{x})$,

$$\Phi(\mathbf{x}) = \int d^3x' \rho(\mathbf{x}') \frac{1}{|\mathbf{x} - \mathbf{x}'|}. \quad (154)$$

Substitute the spherical harmonic expansion for $\frac{1}{|\mathbf{x} - \mathbf{x}'|}$ to have

$$\Phi(\mathbf{x}) = \sum_{l,m} \frac{4\pi}{2l+1} \left[\int d^3x' \rho(\mathbf{x}') Y_{l,m}^*(\theta', \phi') \frac{r_{<}^l}{r_{>}^{l+1}} \right] Y_{l,m}(\theta, \phi). \quad (155)$$



Now, if the origin of coordinates is chosen to be around the center of the charge distribution, and if the field point \mathbf{x} is such that r is larger

than the distance of any source point (where $\rho(\mathbf{x}') \neq 0$) \mathbf{x}' from the origin, then it is true that for all \mathbf{x}' of importance in the integral, $r > r'$ and so $r_{<} = r'$ and $r_{>} = r$. Thus,

$$\Phi(\mathbf{x}) = \sum_{l,m} \frac{4\pi}{2l+1} \left[\int d^3x' r'^l Y_{l,m}^*(\theta', \phi') \rho(\mathbf{x}') \right] \frac{Y_{l,m}(\theta, \phi)}{r^{l+1}}. \quad (156)$$

This result may be written in the form

$$\Phi(\mathbf{x}) = \sum_{l=0}^{\infty} \sum_{m=-l}^l \frac{4\pi}{2l+1} q_{lm} \frac{Y_{l,m}(\theta, \phi)}{r^{l+1}} \equiv \sum_{l,m} \Phi_{lm}(\mathbf{x}), \quad (157)$$

where

$$q_{lm} \equiv \int d^3x r^l Y_{l,m}^*(\theta, \phi) \rho(\mathbf{x}) \quad (158)$$

is known as a *multipole moment* of the charge distribution. These moments, which satisfy the identity

$$q_{l,m} = (-1)^m q_{l,-m}^* \quad (159)$$

by virtue of the same property of the spherical harmonics, completely determine the field outside of the domain where the charge is located. Note, however, that they do not contain enough information to tell us what the actual charge distribution is. The moments of greatest interest are the ones with small values of l . We can understand this statement from the fact that the moment q_{lm} is proportional to, as seen from Eq. (5), a^l , where a is the size of the charge distribution. Hence the potential produced by this moment is proportional to $(q'/r)(a/r)^l$ where q' is a characteristic charge in the distribution (The actual total

charge may vanish). This contribution to the potential becomes very small for large l given that r is significantly larger than a .

The components of the electric field associated with the l, m multipole are

$$\begin{aligned} E_r &= -\frac{\partial\Phi_{lm}}{\partial r} = \frac{4\pi(l+1)}{2l+1}q_{lm}\frac{Y_{l,m}(\theta,\phi)}{r^{l+2}} \\ E_\theta &= -\frac{1}{r}\frac{\partial\Phi_{lm}}{\partial\theta} = -\frac{4\pi}{2l+1}q_{lm}\frac{1}{r^{l+2}}\frac{\partial Y_{l,m}(\theta,\phi)}{\partial\theta} \\ E_\phi &= -\frac{1}{r\sin\theta}\frac{\partial\Phi_{lm}}{\partial\phi} = -\frac{4\pi}{2l+1}q_{lm}\frac{im}{r^{l+2}\sin\theta}Y_{l,m}(\theta,\phi). \end{aligned} \quad (160)$$

The leading moments are

$$q_{00} = \int d^3x \rho(\mathbf{x}) \frac{1}{\sqrt{4\pi}} = \frac{q}{\sqrt{4\pi}} \quad (161)$$

where q is precisely the total charge of the system. This term is the *monopole moment* of the charge distribution; it is fundamentally just the total charge. Similarly,

$$q_{10} = \sqrt{\frac{3}{4\pi}} \int d^3x \rho(\mathbf{x}) r \cos\theta = \sqrt{\frac{3}{4\pi}} \int d^3x \rho(\mathbf{x}) z \quad (162)$$

and

$$q_{11} = -q_{1,-1}^* = -\sqrt{\frac{3}{8\pi}} \int d^3x \rho(\mathbf{x}) r \sin\theta e^{-i\phi} = -\sqrt{\frac{3}{8\pi}} \int d^3x \rho(\mathbf{x}) (x - iy). \quad (163)$$

From these equations we can see that the information contained in the coefficients q_{1m} is the same as what is contained in the components of the electric dipole moment \mathbf{p} of the charge distribution,

$$\mathbf{p} \equiv \int d^3x \rho(\mathbf{x}) \mathbf{x}. \quad (164)$$

The explicit connection is

$$\mathbf{p} = \left\{ \frac{1}{2} \sqrt{\frac{8\pi}{3}} (q_{11} - q_{1,-1}) \hat{\mathbf{x}} - \frac{1}{2} \sqrt{\frac{8\pi}{3}} i (q_{11} + q_{1,-1}) \hat{\mathbf{y}} + \sqrt{\frac{4\pi}{3}} q_{10} \hat{\mathbf{z}} \right\} \quad (165)$$

The $l = 2$ moments, called *electric quadrupole moments*, are easily shown to be

$$\begin{aligned} q_{22} &= \frac{1}{4} \sqrt{\frac{15}{2\pi}} \int d^3x \rho(\mathbf{x}') (x - iy)^2 \\ q_{21} &= -\sqrt{\frac{15}{8\pi}} \int d^3x \rho(\mathbf{x}) (x - iy) z \\ q_{20} &= \frac{1}{2} \sqrt{\frac{5}{4\pi}} \int d^3x \rho(\mathbf{x}) (3z^2 - r^2). \end{aligned} \quad (166)$$

These multipole moments are traditionally written in terms of the components of the *traceless quadrupole moment tensor*, defined by

$$Q_{ij} \equiv \int d^3x \rho(\mathbf{x}) (3x_i x_j - r^2 \delta_{ij}); \quad (167)$$

the subscripts i and j stand for Cartesian components x , y , and z , or 1,2,3. With a little algebra, one can show that

$$\begin{aligned} q_{22} &= \frac{1}{12} \sqrt{\frac{15}{2\pi}} (Q_{11} - 2iQ_{12} - Q_{22}) \\ q_{21} &= -\frac{1}{3} \sqrt{\frac{15}{8\pi}} (Q_{13} - iQ_{23}) \\ q_{20} &= \frac{1}{2} \sqrt{\frac{5}{4\pi}} Q_{33} \end{aligned} \quad (168)$$

It seems a little strange to be replacing at most five independent numbers (contained in the moments q_{2m}) by nine numbers Q_{ij} ; however,

the quadrupole moment tensor is symmetric, $Q_{ij} = Q_{ji}$, reducing the number of possible independent components to six, and it also has, as its name suggests and as may be shown easily from the definition, zero trace so that $Q_{33} = -Q_{11} - Q_{22}$ and only two of the diagonal components are independent. Thus the tensor can have at most five independent components also.

Static and Stationary Magnetic Fields

André-Marie Ampère

(1775 - 1836)

December 23, 2000

Contents

1	Introduction and Definitions	2
1.1	Magnetic Induction	2
1.2	Current Density and Conservation	3
2	Ampère's Law	5
2.1	Induction of an Arbitrary Current Density	8
2.2	An Alternate Form of Ampère's law	9
2.3	Example: Force Between Parallel Wires	10
3	Differential Equations of Magnetostatics	11
4	Vector and Scalar Potentials	15
4.1	Scalar Potential	15
4.2	Vector Potential and Gauge Invariance	16
4.3	Example: A Circular Current Loop	19
5	The Field of a Localized Current Distribution	22

6	Forces on a Localized Current Distribution	28
7	Macroscopic Magnetostatics	31
7.1	Magnetization Current Density	33
7.2	Magnetic Field	35
7.3	Boundary Conditions	38
8	Examples of Boundary-Value Problems in Magnetostatics	41
8.1	Uniformly Magnetized Sphere	41
8.1.1	Scalar Potential for the Induction	42
8.1.2	Scalar Potential for the Field	44
8.1.3	Direct Calculation of \mathbf{B}	46
8.2	Shielding by a Paramagnetic Cylinder	48

1 Introduction and Definitions

As far as anyone knows, there is no such thing as a free magnetic charge or *magnetic monopole*, although there are people who look for them (and occasionally claim to have found one); certainly they may exist. Because no known phenomena require their existence, we shall develop magnetostatics and eventually electrodynamics assuming that they do not exist. In this case there is a fundamental difference between electrostatics and magnetostatics, explaining in part why the two subjects developed independently and were regarded as distinct rather than different limits or aspects of one type of phenomenon (electromagnetic phenomena).

1.1 Magnetic Induction

In the absence of monopole moments, the most fundamental source of magnetic effects is the *magnetic dipole*. In the presence of other magnetic materials, a point dipole will experience some force. One defines the *magnetic flux density* or *magnetic induction* \mathbf{B} in terms of the torque \mathbf{N} exerted on the dipole. Given that the dipole moment is μ , the defining relation is

$$\mathbf{N} \equiv \mu \times \mathbf{B}. \tag{1}$$

Thus, as for electrostatics, the basic field of magnetostatics is defined by the effect produced on an elementary source.

1.2 Current Density and Conservation

Among the first known manifestations of magnetic phenomena were the forces observed to act on some materials (magnets) as a consequence of the earth's magnetic field. In 1819, Hans Christian Oersted (1777-1851) found that very similar effects could be produced by placing a magnet close to a current-carrying wire, indicating a connection between electrical current and magnetism. Much of what we have to say about magnetostatics will involve the use of currents as sources of \mathbf{B} , so let us say a few words about the properties of stationary, *i.e.*, time-independent, currents. We shall write the *current density* as $\mathbf{J}(\mathbf{x})$; it has dimensions of charge/area-time and is by definition such that a component J_i is the amount of charge that crosses unit area in unit time given that the normal to the surface is in the i -direction. Given a charge density $\rho(\mathbf{x})$ moving at a velocity \mathbf{v} , there is a current density

$$\mathbf{J}(\mathbf{x}) = \rho(\mathbf{x})\mathbf{v}. \quad (2)$$

It is an experimental fact that charge is conserved. We may determine a continuity or conservation equation which expresses this fact.

Consider

$$\int_V d^3x \nabla \cdot \mathbf{J}(\mathbf{x}) = \int_S d^2x \mathbf{J} \cdot \mathbf{n}. \quad (3)$$

The surface integral gives the rate at which charge flows out of the domain V through the surface S . Because charge is conserved, this must be equal to the negative of the rate at which the total charge inside of V changes:

$$\int_S d^2x \mathbf{J}(\mathbf{x}) \cdot \mathbf{n} = -\frac{\partial}{\partial t} \left(\int_V d^3x \rho(\mathbf{x}) \right). \quad (4)$$

Assuming that V is independent of time, we may move the derivative with respect to time inside of the integral and so have

$$\int_V d^3x \nabla \cdot \mathbf{J}(\mathbf{x}) = \int_S d^2x \mathbf{J}(\mathbf{x}) \cdot \mathbf{n} = -\int_V d^3x \frac{\partial \rho(\mathbf{x})}{\partial t}. \quad (5)$$

Now argue in the usual fashion: Because V is an arbitrary domain, this equation can only be true if the integrands on the two sides are the same everywhere. Hence we have

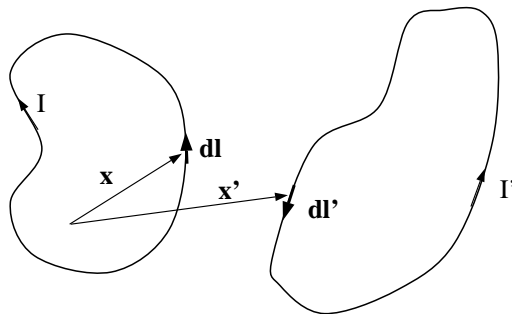
$$\nabla \cdot \mathbf{J}(\mathbf{x}) + \frac{\partial \rho(\mathbf{x})}{\partial t} = 0. \quad (6)$$

This equation is true so long as charge is *locally* conserved, meaning that the only way for charge to appear in V (or to disappear from V) is by flowing across the boundary. The equation has the typical form of a continuity equation which is that the divergence of the current density of some quantity plus the time derivative of the density of that quantity

equals zero. If there are sources (or sinks) of the quantity in question, there is an additional term in the equation expressing the contribution of these sources.

2 Ampère's Law

Even as currents in wires produce forces on magnetic materials, so do they produce forces on other current loops. Félix Savart (1791-1841) and Jean-Baptiste Biot (1774-1862) began experiments on these forces soon after Oersted's discovery, as did André-Marie Ampère. Ampère continued his experiments for some years and published his collected results in 1825. The basic law emerging from Ampère's work deals with the forces acting between closed current loops. Suppose that we have a current I in one loop and a current I' in a second. Let \mathbf{F} be the force acting on the loop carrying current I .



Then *Ampère's Law* may be expressed as follows:

$$\mathbf{F} = kII' \iint \frac{d\mathbf{l} \times [d\mathbf{l}' \times (\mathbf{x} - \mathbf{x}')] }{|\mathbf{x} - \mathbf{x}'|^3} \quad (7)$$

where the integrals over \mathbf{l} and \mathbf{l}' are, respectively, around the loops carrying currents I and I' ; \mathbf{x} and \mathbf{x}' are the position vectors of the integration points. The constant k depends on the units employed. For our units, with current expressed as *statcoul/sec*, or *statamp*, k has dimensions of (inverse speed)² (remember that charge has dimensions of $M^{1/2}L^{3/2}/T$ in our units). Hence one writes $k \equiv 1/c^2$ where c is a speed. From appropriate experiments one may find that its numerical value in cgs units is $c = 2.998 \times 10^{10} \text{ cm/sec}$. We recognize this as the speed of light, but that is, for the moment, not important.

At this juncture we may introduce the magnetic induction by writing the force as

$$\mathbf{F} \equiv \frac{I}{c} \int d\mathbf{l} \times \mathbf{B}(\mathbf{x}) \quad (8)$$

where $\mathbf{B}(\mathbf{x})$ is the magnetic induction produced by the other loop's current. It is not yet clear apparent this \mathbf{B} is the same as the one introduced in Eq. (1); nevertheless, it is, as we shall see presently.

Comparison of Eq. (7) with Eq. (8) shows that the magnetic induction produced by the loop carrying current I' may be written as an integral over that loop,

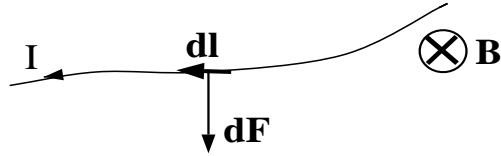
$$\mathbf{B}(\mathbf{x}) = \frac{I'}{c} \int \frac{d\mathbf{l}' \times (\mathbf{x} - \mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^3}; \quad (9)$$

this equation is often called the *Biot-Savart Law*.

One also writes these equations in differential form, although that

may introduce some inaccuracies and even misunderstanding. First, the force acting on just an infinitesimal piece of the loop carrying current I is

$$d\mathbf{F}(\mathbf{x}) = \frac{I}{c} d\mathbf{l} \times \mathbf{B}(\mathbf{x}). \quad (10)$$



The correct interpretation of this equation is that it expresses the force on the element $d\mathbf{l}$ of the loop carrying current I which is produced by the current I' in the other loop; $\mathbf{B}(\mathbf{x})$ is the magnetic induction produced by this other loop. There will be additional forces on the element $d\mathbf{l}$ produced by the current in other parts of its own loop.

Another equation one frequently sees is an expression for the infinitesimal magnetic induction produced at a point \mathbf{x} by an infinitesimal element of a source loop. Given that the source loop is, as above, the one carrying current I' , this expression is

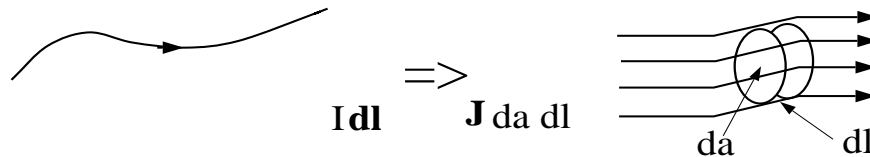
$$d\mathbf{B}(\mathbf{x}) = \frac{I'}{c} \frac{d\mathbf{l}' \times (\mathbf{x} - \mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^3}. \quad (11)$$

This is, however, not a correct statement in that the element $d\mathbf{l}'$ of this circuit acting alone does not produce such a magnetic induction. First, it is impossible to have such a source acting alone if there is no time dependence in the sources and fields; the flowing charge which gives the

current I' in $d\mathbf{l}'$ has to come from somewhere and go somewhere and so if this element is the entire source current distribution, then there must be some time dependence associated with the accumulation and depletion of charge at the two ends of the element. When this time dependence is included, Eq. (11) will not give the magnetic induction correctly.

2.1 Induction of an Arbitrary Current Density

The preceding results are capable of generalization to arbitrary current distributions (not just filaments). One has to replace $I d\mathbf{l}$ by $\mathbf{J}(\mathbf{x}) da dl$ where $\mathbf{J}(\mathbf{x})$ is the current density, da is the cross-sectional area of the filament, and dl is the magnitude of $d\mathbf{l}$. Then note that $da dl = d^3x$, a volume element.



Hence one finds that the flux produced by an extended current distribution is

$$\mathbf{B}(\mathbf{x}) = \frac{1}{c} \int d^3x' \frac{\mathbf{J}(\mathbf{x}') \times (\mathbf{x} - \mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^3} \quad (12)$$

while the force on an extended current distribution $\mathbf{J}(\mathbf{x})$ produced by

some externally applied $\mathbf{B}(\mathbf{x})$ is

$$\mathbf{F} = \frac{1}{c} \int d^3x \mathbf{J}(\mathbf{x}) \times \mathbf{B}(\mathbf{x}). \quad (13)$$

The corresponding force density at point \mathbf{x} is

$$d\mathbf{F} = \frac{1}{c} d^3x \mathbf{J}(\mathbf{x}) \times \mathbf{B}(\mathbf{x}), \quad (14)$$

and the expression for $d\mathbf{B}$ (not to be trusted) becomes

$$d\mathbf{B}(\mathbf{x}) = \frac{1}{c} d^3x' \frac{\mathbf{J}(\mathbf{x}') \times (\mathbf{x} - \mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^3}. \quad (15)$$

Finally, let's add to our arsenal of equations one for the torque \mathbf{N} felt by a current distribution $\mathbf{J}(\mathbf{x})$ acted upon by an externally applied magnetic induction $\mathbf{B}(\mathbf{x})$. Given an object experiencing a force \mathbf{F} , the torque relative to a point O is just $\mathbf{x} \times \mathbf{F}$ where \mathbf{x} is the location of the object relative to O . Hence the torque on the current distribution is

$$\mathbf{N} = \frac{1}{c} \int d^3x [\mathbf{x} \times (\mathbf{J}(\mathbf{x}) \times \mathbf{B}(\mathbf{x}))]. \quad (16)$$

2.2 An Alternate Form of Ampère's law

Before going on to additional formalism, let us express Ampère's law, Eq. (7), in a more symmetric form by applying some vector manipulations. First, we will make use of the general identity $\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = (\mathbf{A} \cdot \mathbf{C})\mathbf{B} - (\mathbf{A} \cdot \mathbf{B})\mathbf{C}$ to have

$$\mathbf{F} = \frac{II'}{c^2} \iint \frac{[d\mathbf{l} \cdot (\mathbf{x} - \mathbf{x}')]d\mathbf{l}' - [d\mathbf{l} \cdot d\mathbf{l}'](\mathbf{x} - \mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^3}. \quad (17)$$

According to Newton, the force of the loop carrying current I should be equal and opposite to the force on the other loop, implying certain symmetries in the integrand above. The second term in the numerator changes sign under interchange of \mathbf{x} and \mathbf{x}' , but the first does not. Let us study the latter more closely. Consider

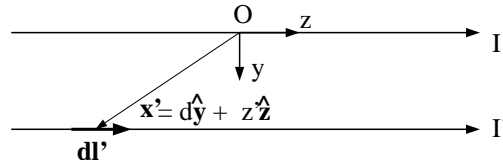
$$\int d\mathbf{l}' \int d\mathbf{l} \cdot \left[-\nabla \left(\frac{1}{|\mathbf{x} - \mathbf{x}'|} \right) \right] = - \int d\mathbf{l}' \int d\mathbf{l} \left(\frac{1}{|\mathbf{x} - \mathbf{x}'|} \right) \equiv 0 \quad (18)$$

where the last step follows from the fact we integrate over a closed path¹. Thus we find that the force \mathbf{F} may be written simply as

$$\mathbf{F} = -\frac{II'}{c^2} \iint \frac{d\mathbf{l} \cdot d\mathbf{l}'}{|\mathbf{x} - \mathbf{x}'|^3} (\mathbf{x} - \mathbf{x}'). \quad (19)$$

2.3 Example: Force Between Parallel Wires

Let us look at an example. Suppose we have two parallel wires a distance d apart carrying currents I and I' and wish to find the force per unit length acting on one of them. At a point \mathbf{x} on the one carrying current I , there is an induction B from the other one which is directed perpendicular to the plane containing the wires. This field is given by the integral over the sources in the other wire,



¹For example, $\oint d\phi = \phi(x) - \phi(x)$, where x is an arbitrary point along the contour

$$|\mathbf{B}| = B = \frac{I'}{c} \int \frac{d\mathbf{l}' \times (\mathbf{x} - \mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^3} = \frac{I'd}{c} \int_{-\infty}^{\infty} \frac{dz'}{(d^2 + z'^2)^{3/2}} = \frac{I'}{cd} \int_{-\infty}^{\infty} \frac{du}{(1 + u^2)^{3/2}} = \frac{2I'}{cd}. \quad (20)$$

The consequent force on the current I in length dz is

$$|d\mathbf{F}| = \frac{I}{c} dz B = \frac{2II'}{c^2 d}. \quad (21)$$

The direction of \mathbf{F} is such that the wires are attracted toward each other if the currents are in the same direction and they are repelled if the currents are in opposite directions. In doing this calculations we have conveniently ignored the fact that we didn't deal with closed current loops; in principal the wires have to be bent into closed loops somewhere far away from where we have calculated the force. With a little work one can convince himself that for sufficiently large loops, the contributions from the part that we ignored will be as small as desired.

3 Differential Equations of Magnetostatics

Even as we found equations for the divergence and curl of the electric field, so can we find such equations for the magnetic induction. Our starting point is the integral expression for the induction produced by a source distribution $\mathbf{J}(\mathbf{x})$,

$$\mathbf{B}(\mathbf{x}) = \frac{1}{c} \int d^3x' \mathbf{J}(\mathbf{x}') \times \frac{(\mathbf{x} - \mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^3}, \quad (22)$$

or, proceeding in the same manner as we have done before,

$$\mathbf{B}(\mathbf{x}) = -\frac{1}{c} \int d^3x' \mathbf{J}(\mathbf{x}') \times \nabla \left(\frac{1}{|\mathbf{x} - \mathbf{x}'|} \right). \quad (23)$$

Now apply the identity $\nabla \times (f\mathbf{A}) = (\nabla \times \mathbf{A})f + \nabla f \times \mathbf{A}$:

$$\nabla \times \left(\frac{\mathbf{J}(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|} \right) = (\nabla \times \mathbf{J}(\mathbf{x}')) \frac{1}{|\mathbf{x} - \mathbf{x}'|} + \nabla \left(\frac{1}{|\mathbf{x} - \mathbf{x}'|} \right) \times \mathbf{J}(\mathbf{x}'). \quad (24)$$

The curl (taken with respect to components of \mathbf{x}) of $\mathbf{J}(\mathbf{x}')$ is zero, so, upon substituting into Eq. (24), we find

$$\mathbf{B}(\mathbf{x}) = \frac{1}{c} \int d^3x' \nabla \times \left(\frac{\mathbf{J}(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|} \right) = \frac{1}{c} \nabla \times \left(\int d^3x' \frac{\mathbf{J}(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|} \right). \quad (25)$$

Because $\mathbf{B}(\mathbf{x})$ is the curl of a vector field, its divergence must be zero,

$$\nabla \cdot \mathbf{B}(\mathbf{x}) = 0. \quad (26)$$

This is our equation for the divergence of the magnetic induction. It tells us that there are no magnetic charges.

To find a curl equation for $\mathbf{B}(\mathbf{x})$, we take the curl of Eq. (25),

$$\nabla \times \mathbf{B}(\mathbf{x}) = \frac{1}{c} \nabla \times \left[\nabla \times \left(\int d^3x' \frac{\mathbf{J}(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|} \right) \right]. \quad (27)$$

Now employ the identity

$$\nabla \times (\nabla \times \mathbf{A}) = \nabla(\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A}, \quad (28)$$

valid for any vector field $\mathbf{A}(\mathbf{x})$; the last term on the right-hand side of this identity is to be interpreted as

$$\nabla^2 \mathbf{A} \equiv (\nabla^2 A_x) \hat{\mathbf{x}} + (\nabla^2 A_y) \hat{\mathbf{y}} + (\nabla^2 A_z) \hat{\mathbf{z}}; \quad (29)$$

it is important that this relation is written in Cartesian coordinates.

Using the identity, we find

$$\nabla \times \left[\nabla \times \left(\frac{\mathbf{J}(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|} \right) \right] = \nabla \left[\mathbf{J}(\mathbf{x}') \cdot \nabla \left(\frac{1}{|\mathbf{x} - \mathbf{x}'|} \right) \right] - \mathbf{J}(\mathbf{x}') \nabla^2 \left(\frac{1}{|\mathbf{x} - \mathbf{x}'|} \right). \quad (30)$$

The second term is just $4\pi\mathbf{J}(\mathbf{x}')\delta(\mathbf{x} - \mathbf{x}')$, so we have

$$\nabla \times \mathbf{B}(\mathbf{x}) = \frac{4\pi}{c}\mathbf{J}(\mathbf{x}) + \frac{1}{c}\nabla \left[\int d^3x' \mathbf{J}(\mathbf{x}') \cdot \nabla \left(\frac{1}{|\mathbf{x} - \mathbf{x}'|} \right) \right]. \quad (31)$$

The remaining integral may be completed as follows:

$$\begin{aligned} \nabla \int d^3x' \mathbf{J}(\mathbf{x}') \cdot \nabla \left(\frac{1}{|\mathbf{x} - \mathbf{x}'|} \right) &= -\nabla \int d^3x' \mathbf{J}(\mathbf{x}') \cdot \nabla' \left(\frac{1}{|\mathbf{x} - \mathbf{x}'|} \right) \\ &= -\nabla \int d^3x' \nabla' \cdot \left(\frac{\mathbf{J}(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|} \right) + \nabla \int d^3x' \frac{\nabla' \cdot \mathbf{J}(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|}. \end{aligned} \quad (32)$$

The first term in the final expression can be converted to a surface integral which then vanishes for a localized current distribution which lies totally within the domain enclosed by the surface. The integrand of the second term involves $\nabla' \cdot \mathbf{J}(\mathbf{x}') = -\partial\rho(\mathbf{x}')/\partial t \equiv 0$ for a steady-state current distribution. Consequently we have the curl equation

$$\nabla \times \mathbf{B}(\mathbf{x}) = \frac{4\pi}{c}\mathbf{J}(\mathbf{x}) \quad \text{Sometimes called Ampere's Law} \quad (33)$$

The curl and divergence equations, $\nabla \cdot \mathbf{B}(\mathbf{x}) = 0$ and $\nabla \times \mathbf{B}(\mathbf{x}) = (4\pi/c)\mathbf{J}(\mathbf{x})$ plus an appropriate statement about the behavior of $\mathbf{B}(\mathbf{x})$ on a boundary tell us all we need to know to find the magnetic induction for a given set of sources $\mathbf{J}(\mathbf{x})$. There are, of course, also integral

versions of these differential equations. One finds the most common forms of them from the Stokes theorem and the divergence theorem.

$$\int_V d^3x \nabla \cdot \mathbf{B}(\mathbf{x}) = \int_S d^2x \mathbf{B}(\mathbf{x}) \cdot \mathbf{n} = 0 \quad (34)$$

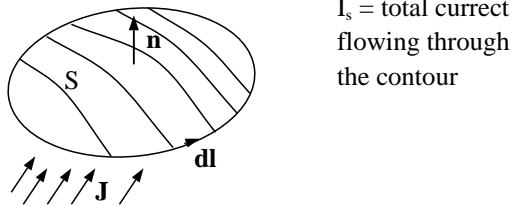
and

$$\int_S d^2x (\nabla \times \mathbf{B}(\mathbf{x})) \cdot \mathbf{n} = \int_C d\mathbf{l} \cdot \mathbf{B}(\mathbf{x}) = \frac{4\pi}{c} \int_S d^2x \mathbf{J}(\mathbf{x}) \cdot \mathbf{n}. \quad (35)$$

The last of these is also commonly written as

$$\int_C d\mathbf{l} \cdot \mathbf{B}(\mathbf{x}) = \frac{4\pi}{c} I_s \quad (36)$$

where I_s is the total current passing through the surface S in the direction of a right-hand normal relative to the direction in which the line integral around C is done.



This relation is frequently called Ampère's law. One interesting feature of this equation is that the result is independent of the actual surface S employed so long as it is an open surface that ends on the path C . The amount of charge passing through all such surfaces per unit time *i.e.*, I_s , is independent of S because $\nabla \cdot \mathbf{J}(\mathbf{x}) = 0$.

4 Vector and Scalar Potentials

In all electromagnetic systems it is possible to devise a potential which is a vector field for $\mathbf{B}(\mathbf{x})$; it is called a *vector potential*. Sometimes, it is also possible to devise a scalar potential for $\mathbf{B}(\mathbf{x})$. We shall consider the latter first.

4.1 Scalar Potential

As we have seen in the study of electrostatics, one can write a vector field as the gradient of a scalar if the vector field has zero divergence. This is the case for $\mathbf{B}(\mathbf{x})$ in those regions of space where $\mathbf{J}(\mathbf{x}) = 0$. Thus, in a source-free domain, we can write

$$\mathbf{B}(\mathbf{x}) = -\nabla\Phi_M(\mathbf{x}). \quad (37)$$

The *magnetic scalar potential* satisfies a differential equation which follows from the requirement that $\nabla \cdot \mathbf{B}(\mathbf{x}) = 0$; it is

$$\nabla^2\Phi_M(\mathbf{x}) = 0. \quad (38)$$

Thus, wherever there is a magnetic scalar potential, it satisfies the Laplace equation. In order to solve for it, we may apply any of the techniques we learned for finding the electrostatic potential in charge-free regions of space. Hence no more will be said about the magnetic scalar potential in general in this chapter.

4.2 Vector Potential and Gauge Invariance

Consider now the vector potential for a magnetostatic field. Because $\nabla \cdot \mathbf{B}(\mathbf{x}) = 0$, it is possible to write the magnetic induction as the curl of another vector field (since $\nabla \cdot (\nabla \times \mathbf{A})$). We have in fact already constructed such a vector field because in Eq. (25) we have written

$$\mathbf{B}(\mathbf{x}) = \nabla \times \left(\frac{1}{c} \int d^3x' \frac{\mathbf{J}(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|} \right). \quad (39)$$

The field within the parentheses is a vector potential for $\mathbf{B}(\mathbf{x})$. We shall write it as $\mathbf{A}(\mathbf{x})$. It is, in contrast to the electrostatic scalar potential, not unique because one can always add to it the gradient of any scalar field $\chi(\mathbf{x})$ and have a vector field whose curl is still $\mathbf{B}(\mathbf{x})$. That is, given

$$\mathbf{A}(\mathbf{x}) = \frac{1}{c} \int d^3x' \frac{\mathbf{J}(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|}, \quad (40)$$

which is such that $\mathbf{B}(\mathbf{x}) = \nabla \times \mathbf{A}(\mathbf{x})$, we can write

$$\mathbf{A}'(\mathbf{x}) = \mathbf{A}(\mathbf{x}) + \nabla\chi(\mathbf{x}), \quad (41)$$

where $\chi(\mathbf{x})$ is arbitrary. Then it is true that $\mathbf{B}(\mathbf{x}) = \nabla \times \mathbf{A}'(\mathbf{x})$ because the curl of the gradient of a scalar field is zero.

By writing $\mathbf{B}(\mathbf{x}) = \nabla \times \mathbf{A}(\mathbf{x})$, we have automatically satisfied the requirement that $\nabla \cdot \mathbf{B}(\mathbf{x}) = 0$; Hence we may find a single (vector) equation for the vector potential by substituting $\mathbf{B} = \nabla \times \mathbf{A}$ into Ampère's law:

$$\nabla \times \mathbf{B}(\mathbf{x}) = \nabla \times (\nabla \times \mathbf{A}(\mathbf{x})) = \frac{4\pi}{c} \mathbf{J}(\mathbf{x}), \quad (42)$$

or, using $\nabla \times (\nabla \times \mathbf{A}) = \nabla(\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A}$,

$$\nabla^2 \mathbf{A}(\mathbf{x}) - \nabla(\nabla \cdot \mathbf{A}(\mathbf{x})) = -\frac{4\pi}{c} \mathbf{J}(\mathbf{x}). \quad (43)$$

This equation would be considerably simpler if we could make the divergence of $\mathbf{A}(\mathbf{x})$ vanish. It is possible to do this by using a vector potential constructed with an appropriate choice of $\chi(\mathbf{x})$. The underlying mathematical point is the following: so far, the only condition we have placed on the vector potential is that its curl should be the magnetic induction. We are free to choose it so that its divergence conforms to our wishes because a vector field is sufficiently flexible that its curl and divergence can both be specified arbitrarily. Consider then the divergence of the most general vector potential. We write this potential as

$$\mathbf{A}(\mathbf{x}) = \frac{1}{c} \int d^3x' \frac{\mathbf{J}(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|} + \nabla \chi(\mathbf{x}). \quad (44)$$

Then

$$\nabla \cdot \mathbf{A}(\mathbf{x}) = \frac{1}{c} \int d^3x' \mathbf{J}(\mathbf{x}') \cdot \nabla \left(\frac{1}{|\mathbf{x} - \mathbf{x}'|} \right) + \nabla^2 \chi(\mathbf{x}). \quad (45)$$

Now change the ∇ operator to $-\nabla'$ and then do an integration by parts. The surface integral may be discarded if the volume integral is over all space and $\mathbf{J}(\mathbf{x}')$ is localized (so that it vanishes on the surface of the volume of integration). What one then has left is

$$\nabla \cdot \mathbf{A}(\mathbf{x}) = \frac{1}{c} \int d^3x' \frac{\nabla' \cdot \mathbf{J}(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|} + \nabla^2 \chi(\mathbf{x}) = \nabla^2 \chi(\mathbf{x}), \quad (46)$$

where the last step follows from the fact that a time-independent set of sources give a current distribution having zero divergence.

From this result we can see two things: first, the vector potential has zero divergence if we forget about $\chi(\mathbf{x})$, *i.e.*, if we set it equal to zero. Second, we can make the divergence be any scalar function $f(\mathbf{x})$ we want by setting $\nabla^2\chi(\mathbf{x}) = f(\mathbf{x})$. We would then have to solve for $\chi(\mathbf{x})$ which is in principle straightforward because the differential equation for χ is just the Poisson equation and we know how to solve that. The solution is

$$\chi(\mathbf{x}) = -\frac{1}{4\pi} \int d^3x' \frac{f(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|}. \quad (47)$$

Specifying the divergence of $\mathbf{A}(\mathbf{x})$ is called choosing the *gauge* of the vector potential, and changing the function $\chi(\mathbf{x})$ is called making a *gauge transformation*. Notice that one can make a transformation without changing the gauge (*i.e.*, without changing $\nabla \cdot \mathbf{A}$); to do this one must change χ by a function which satisfies the Laplace equation. The particular gauge specified by $\nabla \cdot \mathbf{A}(\mathbf{x}) = 0$ is called the *Coulomb gauge*.

Returning to our original point, we have found that we can pick a vector potential with zero divergence. In particular,

$$\mathbf{A}(\mathbf{x}) = \frac{1}{c} \int d^3x' \frac{\mathbf{J}(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|} \quad (48)$$

has this property. In this, the Coulomb gauge, the vector potential

satisfies the equations

$$\nabla^2 \mathbf{A}(\mathbf{x}) = -\frac{4\pi}{c} \mathbf{J}(\mathbf{x}), \quad (49)$$

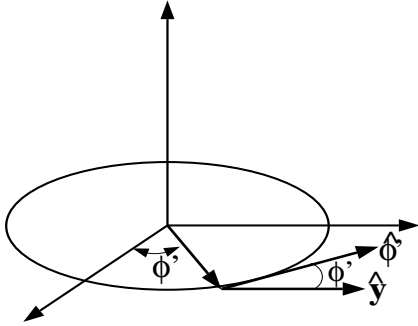
which **must** be interpreted as

$$\nabla^2 A_i(\mathbf{x}) = -\frac{4\pi}{c} J_i(\mathbf{x}) \quad (50)$$

where the subscript i denotes any Cartesian component of \mathbf{A} .

4.3 Example: A Circular Current Loop

Consider a circular loop of radius a carrying a current I . Let the loop lie in the $z = 0$ plane and be centered at the origin of coordinates.



Then we may write $\mathbf{J}(\mathbf{x}) = J_\phi(\mathbf{x})\hat{\phi}$ with

$$\mathbf{J}_\phi(\mathbf{x}) = \frac{I}{a} \delta(\cos \theta) \delta(r - a). \quad (51)$$

It is a natural temptation to do the following INCORRECT thing:

write $\mathbf{A}(\mathbf{x}) = A_\phi(\mathbf{x})\hat{\phi}$ (correct so far) with

$$A_\phi(\mathbf{x}) = \frac{1}{c} \int d^3x' \frac{J_\phi(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|}. \quad (52)$$

This integral is not correct because the unit vector $\hat{\phi}$ at the field point \mathbf{x} is not the same as the unit vector $\hat{\phi}'$ at the field point \mathbf{x}' .

What we can do instead is to make use of the azimuthal symmetry of the system and trivially generalize to all values of ϕ after having first evaluated the vector potential at some particular value of ϕ . Let us look at the potential at $\phi = 0$, *i.e.*, in the x - z plane. Here, the vector potential will be in the y direction, so we can say that

$$A_\phi(r, \theta) = A_y(r, \theta, \phi = 0) = \frac{1}{c} \int d^3x' \frac{J_y(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|} \quad (53)$$

at $\phi = 0$. Now, $J_y(\mathbf{x}') = J_{\phi'} \cos \phi'$, so

$$A_\phi(r, \phi) = \frac{I}{ac} \int d^3x' \frac{\delta(r' - a)\delta(\cos \theta') \cos \phi'}{(r^2 + a^2 - 2ar \cos \gamma)^{1/2}} \quad (54)$$

where γ is the angle between the directions of \mathbf{x} and \mathbf{x}' , $\cos \gamma = \sin \theta \cos \phi'$.

The integral is an elliptic integral; rather than deal with its arcane properties, we will expand it in the usual way:

$$\begin{aligned} A_\phi(r, \theta) &= \frac{I}{ca} \int d^3x' \frac{\delta(r' - a)\delta(\cos \theta') \cos \phi'}{|\mathbf{x} - \mathbf{x}'|} \\ &= Re \left\{ \frac{I}{ca} \int d^3x' \frac{\delta(r' - a)\delta(\cos \theta')}{|\mathbf{x} - \mathbf{x}'|} e^{i\phi'} \right\} \\ &= Re \left\{ \frac{I}{ca} \int d^3x' \delta(r' - a)\delta(\cos \theta') e^{i\phi'} \sum_{l,m} \frac{4\pi}{2l+1} \frac{r_{<}^l}{r_{>}^{l+1}} Y_{l,m}^*(\theta', \phi') Y_{l,m}(\theta, \phi) |_{\phi=0} \right\} \\ &= Re \left\{ \frac{Ia}{c} \sum_{l,m} \frac{(l-m)!}{(l+m)!} P_l^m(\cos \theta) P_l^m(0) \frac{r_{<}^l}{r_{>}^{l+1}} \int_0^{2\pi} d\phi' e^{i\phi'} e^{-im\phi'} \right\} \end{aligned}$$

$$= \frac{2\pi I a}{c} \sum_{l=1}^{\infty} \frac{P_l^1(0) P_l^1(\cos \theta)}{l(l+1)} \frac{r_{<}^l}{r_{>}^{l+1}} \quad (55)$$

where $r_{<}$ and $r_{>}$ refer to the smaller and larger of r and a . Next one notes that $P_l^1(0) = 0$ for l even and

$$P_l^1(0) = (-1)^{(l+1)/2} l! / [(l-1)!!]^2 \quad (56)$$

for l odd; set $l = 2n + 1$ and have

$$\begin{aligned} A_{\phi}(r, \theta) &= \frac{2\pi I a}{c} \sum_{n=0}^{\infty} \frac{(-1)^{n+1} (2n+1)!}{(2n+1)(2n+2)[(2n)!!]^2} \frac{r_{<}^{2n+1}}{r_{>}^{2n+2}} P_{2n+1}^1(\cos \theta) \\ &= \frac{\pi I a}{c} \sum_{n=0}^{\infty} \frac{(-1)^{n+1} (2n-1)!!}{2^n (n+1)!} \frac{r_{<}^{2n+1}}{r_{>}^{2n+2}} P_{2n+1}^1(\cos \theta). \end{aligned} \quad (57)$$

At large r , *i.e.*, $r \gg a$, we can keep just the leading term ($n=0$) and find

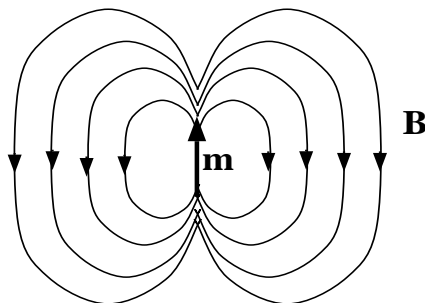
$$A_{\phi}(r, \theta) = \frac{\pi I a}{c} \frac{a}{r^2} \sin \theta. \quad (58)$$

The corresponding components of the magnetic induction are

$$B_r(r, \theta) = \frac{1}{r \sin \theta} \frac{\partial}{\partial \theta} (\sin \theta A_{\phi}) = \frac{2\pi I a^2}{c r^3} \cos \theta, \quad (59)$$

and

$$B_{\theta}(r, \theta) = -\frac{1}{r} \frac{\partial}{\partial r} (r A_{\phi}) = \frac{\pi I a^2}{c r^3} \sin \theta. \quad (60)$$



We recognize these as having the same form as the field of an electric dipole, so that, to order a/r , the current loop is treated as a *magnetic dipole* with *magnetic dipole moment* $\mathbf{m} \equiv (\pi I a^2/c)\hat{\mathbf{z}}$; the corresponding vector potential, Eq. (58), is

$$\mathbf{A}(\mathbf{x}) = A_\phi(r, \theta)\hat{\phi} = \frac{|\mathbf{m}| \sin \theta}{r^2}\hat{\phi} = \frac{\mathbf{m} \times \mathbf{x}}{|\mathbf{x}|^3}. \quad (61)$$

We may compare this potential to that of an electric dipole which is

$$\Phi(\mathbf{x}) = \frac{\mathbf{p} \cdot \mathbf{x}}{|\mathbf{x}|^3}. \quad (62)$$

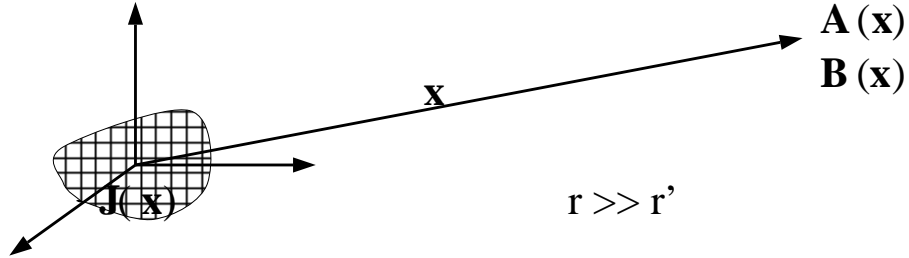
5 The Field of a Localized Current Distribution

The example of the preceding section is a special case of the field of a localized current distribution. Let us now suppose that we have some such distribution $\mathbf{J}(\mathbf{x}')$ around the origin of coordinates. We shall find the potential of this distribution in the dipole approximation. To this

end we must expand the function

$$\frac{1}{|\mathbf{x} - \mathbf{x}'|} \approx \frac{1}{|\mathbf{x}|} \left[1 + \frac{\mathbf{x} \cdot \mathbf{x}'}{|\mathbf{x}|^2} + \dots \right] \quad (63)$$

in powers of r'/r .



Of particular interest to us is the first-order term in the brackets.

Using this expansion, we find that the vector potential is

$$\mathbf{A}(\mathbf{x}) = \frac{1}{c} \int d^3x' \mathbf{J}(\mathbf{x}') \left[\frac{1}{|\mathbf{x}|} + \frac{\mathbf{x} \cdot \mathbf{x}'}{|\mathbf{x}|^3} + \dots \right]. \quad (64)$$

The first term $\sim \int d^3x' \mathbf{J}(\mathbf{x}')$ is zero for a localized steady current dis-

tribution (Why?)² ; the second one is

$$\mathbf{A}_d(\mathbf{x}) = \frac{1}{c|\mathbf{x}|^3} \int d^3x' (\mathbf{x} \cdot \mathbf{x}') \mathbf{J}(\mathbf{x}'). \quad (65)$$

However, $(\mathbf{x} \cdot \mathbf{x}') \mathbf{J} = -\mathbf{x} \times (\mathbf{x}' \times \mathbf{J}) + (\mathbf{x} \cdot \mathbf{J}) \mathbf{x}'$, so

$$\mathbf{A}_d(\mathbf{x}) = \frac{1}{c|\mathbf{x}|^3} \left\{ \int d^3x' (\mathbf{x} \cdot \mathbf{J}(\mathbf{x}')) \mathbf{x}' - \mathbf{x} \times \int d^3x' (\mathbf{x}' \times \mathbf{J}(\mathbf{x}')) \right\}. \quad (66)$$

Consider the j^{th} component of the first integral:

$$\int d^3x' x'_j (\mathbf{x} \cdot \mathbf{J}(\mathbf{x}')) = \sum_{i=1}^3 \int d^3x' x'_j x_i J_i(\mathbf{x}'). \quad (67)$$

Now, it is the case that

$$\nabla' \cdot (x'_i \mathbf{J}(\mathbf{x}')) = (\nabla' x'_i) \cdot \mathbf{J}(\mathbf{x}') + x'_i (\nabla' \cdot \mathbf{J}(\mathbf{x}')) = J_i(\mathbf{x}'), \quad (68)$$

since the divergence of \mathbf{J} is zero. This allows us to write

$$\int d^3x' (\mathbf{x} \cdot \mathbf{J}(\mathbf{x}')) x'_j = \sum_{i=1}^3 x_i \int d^3x' [\nabla' \cdot (x'_i \mathbf{J}(\mathbf{x}'))] x'_j$$

²The short answer is just that in lieu of sources or sinks of charge the total current density in a current distribution must be zero. More formally, consider the 3-vector with components U_i

$$U_i = 0 = \int_V d^3x x_i \nabla \cdot \mathbf{J} = \int_V d^3x [\nabla \cdot (x_i \mathbf{J}) - \mathbf{J} \cdot \nabla x_i]$$

Using the divergence theorem, this becomes

$$\int_S d^2x x_i \mathbf{J} \cdot \mathbf{n} - \int_V d^3x \mathbf{J} \cdot \hat{\mathbf{x}}_i$$

which follows since $\nabla x_i = \hat{\mathbf{x}}_i$. Then if we take the surface to infinity, where there is no current density, we obtain the desired result

$$0 = - \int_V d^3x J_i$$

Parts integration:

$$= - \sum_{i=1}^3 x_i \int d^3 x' (x'_i \mathbf{J}(\mathbf{x}')) \cdot (\nabla' x'_j) = - \int d^3 x' (\mathbf{x} \cdot \mathbf{x}') J_j(\mathbf{x}'). \quad (69)$$

Generalizing now to all three components of this integral, we find

$$\int d^3 x' (\mathbf{x} \cdot \mathbf{J}(\mathbf{x}')) \mathbf{x}' = - \int d^3 x' (\mathbf{x} \cdot \mathbf{x}') \mathbf{J}(\mathbf{x}'). \quad (70)$$

Comparison with the expressions for \mathbf{A}_d shows that

$$\mathbf{A}_d(\mathbf{x}) = - \frac{1}{2c} \frac{1}{|\mathbf{x}|^3} \mathbf{x} \times \left[\int d^3 x' (\mathbf{x}' \times \mathbf{J}(\mathbf{x}')) \right] \equiv \frac{\mathbf{m} \times \mathbf{x}}{|\mathbf{x}|^3} = \nabla \times \left(\frac{\mathbf{m}}{|\mathbf{x}|} \right) \quad (71)$$

where the *magnetic dipole moment* of the current distribution is defined as

$$\mathbf{m} \equiv \frac{1}{2c} \int d^3 x' (\mathbf{x}' \times \mathbf{J}(\mathbf{x}')). \quad (72)$$

The *magnetic moment density* of the distribution is defined by

$$\mathcal{M}(\mathbf{x}) \equiv \frac{1}{2c} \mathbf{x} \times \mathbf{J}(\mathbf{x}) \quad (73)$$

so that the magnetic moment is the integral of this density over all space.

The field produced by the source in the magnetic dipole approximation is

$$\mathbf{B}_d(\mathbf{x}) = \nabla \times \mathbf{A}_d(\mathbf{x}) = \nabla \times \left(\frac{\mathbf{m} \times \mathbf{x}}{|\mathbf{x}|^3} \right). \quad (74)$$

Using the identity

$$\nabla \times (\mathbf{A} \times \mathbf{B}) = \mathbf{A}(\nabla \cdot \mathbf{B}) - \mathbf{B}(\nabla \cdot \mathbf{A}) + (\mathbf{B} \cdot \nabla) \mathbf{A} - (\mathbf{A} \cdot \nabla) \mathbf{B}, \quad (75)$$

we can write this equation as

$$\mathbf{B}_d(\mathbf{x}) = \frac{3\mathbf{n}(\mathbf{n} \cdot \mathbf{m}) - \mathbf{m}}{|\mathbf{x}|^3} \quad (76)$$

where $\mathbf{n} = \mathbf{x}/|\mathbf{x}|$.

This is, as was the case for the electric dipole, only the field outside of the current distribution. For a point magnetic dipole, it would be the field away from the location of the dipole (the origin), and at the origin there is a delta-function field as for the point electric dipole. We may find the magnitude and direction of this singular field by a more careful analysis of what happens as $r \rightarrow 0$. To this end it is useful to write the vector potential of the dipole as

$$\mathbf{A}_d(\mathbf{x}) = \nabla \times \left(\frac{\mathbf{m}}{|\mathbf{x}|} \right). \quad (77)$$

Then we can write

$$\mathbf{B}(\mathbf{x}) = \nabla \times \mathbf{A}(\mathbf{x}) = \nabla \times \left[\nabla \times \left(\frac{\mathbf{m}}{r} \right) \right] = \nabla \left[\nabla \cdot \left(\frac{\mathbf{m}}{r} \right) \right] - \nabla^2 \left(\frac{\mathbf{m}}{r} \right). \quad (78)$$

The last term on the right-hand side is just $4\pi\mathbf{m}\delta(\mathbf{x})$. The first one we have seen before, as it is the same as the electric field of an electric dipole; we already know what singularity is contained therein but will figure it out again as an exercise. Start by integrating this term over a small sphere of radius ϵ centered at the origin and then take the limit

as $\epsilon \rightarrow 0$:

$$\int_{r<\epsilon} d^3x \nabla \left[\nabla \cdot \left(\frac{\mathbf{m}}{r} \right) \right] = \int_{r=\epsilon} d^2x \mathbf{n} \left[\nabla \cdot \left(\frac{\mathbf{m}}{r} \right) \right], \quad (79)$$

where we have used the identity, valid for any scalar function $f(\mathbf{x})$:

$$\int d^3x \nabla f(\mathbf{x}) = \int d^2x \mathbf{n} f(\mathbf{x}); \quad (80)$$

S is the surface enclosing the domain V and \mathbf{n} is the usual outward unit normal. Continuing, we have

$$\int_{r=\epsilon} d^2x \mathbf{n} \left[\nabla \cdot \left(\frac{\mathbf{m}}{r} \right) \right] = - \int_{r=\epsilon} d^2x \hat{\mathbf{r}} \left(\frac{\mathbf{m} \cdot \mathbf{x}}{r^3} \right) = -\frac{4\pi}{3} \mathbf{m}. \quad (81)$$

Hence we find that $\nabla[\nabla \cdot (\mathbf{m}/r)]$ contains the singular piece $(4\pi/3)\mathbf{m}\delta(\mathbf{x})$. Putting it into Eq. (78), we conclude that the delta-function piece of the magnetic field is $(8\pi/3)\mathbf{m}\delta(\mathbf{x})$, and hence the total field of the magnetic dipole is

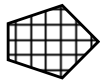
$$\mathbf{B}(\mathbf{x}) = \frac{3(\mathbf{n} \cdot \mathbf{m})\mathbf{n} - \mathbf{m}}{r^3} + \frac{8\pi}{3}\mathbf{m}\delta(\mathbf{x}). \quad (82)$$

The consequences of the presence of the delta-function piece are observed in atomic hydrogen where the magnetic moment of the electron interacts with that of the nucleus, or proton. Without this interaction, all total-spin states of the atom would be degenerate. As a consequence of the interaction, the “triplet” or “spin-one” states are raised slightly in energy relative to the “singlet” or “spin-zero” state. The splitting is small even on the scale of atomic energies, being about 10^{-17}erg or

10^{-5} *ev*. A photon which has the same energy as this splitting, and which is thus produced or absorbed by a transition between the two atomic energy levels, has a wavelength of 21 *cm*. The delta-function part of the field also plays an important role in the scattering of neutrons from magnetic materials.

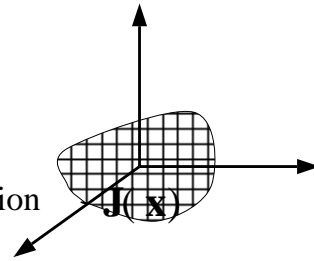
6 Forces on a Localized Current Distribution

We shall look next at the interaction of a localized current distribution with an externally applied field, again using procedures reminiscent of the multipole expansion for a localized charge distribution. First, we expand the applied field around some suitably chosen origin (at the center of the current distribution),



Source of
 $\mathbf{B}(\mathbf{x})$

We will assume that \mathbf{B} varies slowly over the region where \mathbf{J} is finite.



$$\mathbf{B}(\mathbf{x}) = \mathbf{B}(0) + (\mathbf{x} \cdot \nabla')\mathbf{B}(\mathbf{x}')|_{\mathbf{x}'=0} + \dots; \quad (83)$$

The force that the field exerts on a localized current distribution located around the origin is then expanded as follows:

$$\begin{aligned}\mathbf{F} &= \frac{1}{c} \int d^3x (\mathbf{J}(\mathbf{x}) \times \mathbf{B}(\mathbf{x})) \\ &= -\frac{1}{c} \mathbf{B}(0) \times \int d^3x \mathbf{J}(\mathbf{x}) + \frac{1}{c} \int d^3x \mathbf{J}(\mathbf{x}) \times [(\mathbf{x} \cdot \nabla') \mathbf{B}(\mathbf{x}')|_{\mathbf{x}'=0}] + \dots\end{aligned}\quad (84)$$

Now, the first integral in the last line vanishes for a localized steady-state current distribution (there can't be any net flow of charge in any direction), and we can manipulate the integrand in the final integral as follows:

$$(\mathbf{x} \cdot \nabla') \mathbf{B}(\mathbf{x}') = \nabla'(\mathbf{x} \cdot \mathbf{B}(\mathbf{x}')) - \mathbf{x} \times (\nabla' \times \mathbf{B}(\mathbf{x}')), \quad (85)$$

and we may suppose that \mathbf{B} is due entirely to external sources so that $\nabla' \times \mathbf{B}(\mathbf{x}') = 0$ for \mathbf{x}' around the origin. Thus we find that the force is

$$\begin{aligned}\mathbf{F} &= \frac{1}{c} \int d^3x \mathbf{J}(\mathbf{x}) \times [\nabla'(\mathbf{x} \cdot \mathbf{B}(\mathbf{x}'))|_{\mathbf{x}'=0}] \\ &= -\frac{1}{c} \int d^3x \nabla' \times [(\mathbf{x} \cdot \mathbf{B}(\mathbf{x}'))|_{\mathbf{x}'=0} \mathbf{J}(\mathbf{x})] \\ &= -\frac{1}{c} \nabla' \times \int d^3x [\mathbf{x} \cdot \mathbf{B}(\mathbf{x}'))|_{\mathbf{x}'=0} \mathbf{J}(\mathbf{x})].\end{aligned}\quad (86)$$

Now, we can write the last integral as

$$\int d^3x (\mathbf{x} \cdot \mathbf{B}(\mathbf{x}')) \mathbf{J}(\mathbf{x}) = - \int d^3x [\mathbf{B}(\mathbf{x}') \times (\mathbf{x} \times \mathbf{J}(\mathbf{x})) - \mathbf{x}(\mathbf{B}(\mathbf{x}') \cdot \mathbf{J}(\mathbf{x}))]. \quad (87)$$

Further, it is true that

$$\int d^3x \mathbf{x}(\mathbf{B}(\mathbf{x}') \cdot \mathbf{J}(\mathbf{x})) = - \int d^3x (\mathbf{x} \cdot \mathbf{B}(\mathbf{x}')) \mathbf{J}(\mathbf{x}); \quad (88)$$

we can demonstrate this fact by considering the i^{th} component of the integral.

$$\begin{aligned}
-\int d^3x (\mathbf{x} \cdot \mathbf{B}(\mathbf{x}')) J_i(\mathbf{x}) &= -\int d^3x (\mathbf{x} \cdot \mathbf{B}(\mathbf{x}')) \nabla \cdot (x_i \mathbf{J}(\mathbf{x})) \\
&= \int d^3x [\nabla(\mathbf{x} \cdot \mathbf{B}(\mathbf{x}'))] \cdot (x_i \mathbf{J}(\mathbf{x})) = \int d^3x \mathbf{B}(\mathbf{x}') \cdot (x_i \mathbf{J}(\mathbf{x})) \\
&= \int d^3x x_i (\mathbf{B}(\mathbf{x}') \cdot \mathbf{J}(\mathbf{x})). \quad (89)
\end{aligned}$$

Hence, from Eqs. (84), (85), and (86) we find that

$$\begin{aligned}
\mathbf{F} &= \frac{1}{2c} \nabla' \times \left\{ \int d^3x [\mathbf{B}(\mathbf{x}') \times (\mathbf{x} \times \mathbf{J}(\mathbf{x}))] \right\} \Big|_{\mathbf{x}'=0} \\
&= \frac{1}{2c} \nabla' \times \left\{ \mathbf{B}(\mathbf{x}') \times \int d^3x (\mathbf{x} \times \mathbf{J}(\mathbf{x})) \right\} \Big|_{\mathbf{x}'=0} \\
&= \nabla' \times (\mathbf{B}(\mathbf{x}') \times \mathbf{m}) \Big|_{\mathbf{x}'=0} = (\mathbf{m} \cdot \nabla') \mathbf{B}(\mathbf{x}') \Big|_{\mathbf{x}'=0} - \mathbf{m} (\nabla' \cdot \mathbf{B}(\mathbf{x}')) \Big|_{\mathbf{x}'=0} \\
&= \nabla' (\mathbf{m} \cdot \mathbf{B}(\mathbf{x}')) \Big|_{\mathbf{x}'=0} - \mathbf{m} \times (\nabla' \times \mathbf{B}(\mathbf{x}')) \Big|_{\mathbf{x}'=0}. \quad (90)
\end{aligned}$$

Along the way in this derivation we have made use of the facts that the divergence and curl of \mathbf{B} are zero in the region near the origin. The final result has the form of the gradient of a scalar function,

$$\mathbf{F} = -\nabla(-\mathbf{m} \cdot \mathbf{B}(\mathbf{x})) \quad (91)$$

where the gradient is to be evaluated at the center of the current distribution. Notice in particular that there is no force if the applied magnetic induction is uniform. More generally, the force is in the direction of the gradient of the component of \mathbf{B} in the direction of \mathbf{m} .

The torque \mathbf{N} on the current distribution may be found using similar manipulations.

$$\begin{aligned}\mathbf{N} &= \frac{1}{c} \int d^3x \mathbf{x} \times (\mathbf{J}(\mathbf{x}) \times \mathbf{B}(\mathbf{x})) \approx \frac{1}{c} \int d^3x \mathbf{x} \times (\mathbf{J}(\mathbf{x}) \times \mathbf{B}(0)) \\ &= \frac{1}{c} \int d^3x [(\mathbf{x} \cdot \mathbf{B})\mathbf{J}(\mathbf{x}) - (\mathbf{x} \cdot \mathbf{J}(\mathbf{x}))\mathbf{B}] = -\frac{1}{2c} \int d^3x \mathbf{B} \times (\mathbf{x} \times \mathbf{J}(\mathbf{x})) \\ &\quad - \frac{1}{c} \mathbf{B} \int d^3x (\mathbf{x} \cdot \mathbf{J}(\mathbf{x})) = \mathbf{m} \times \mathbf{B} - \frac{1}{c} \mathbf{B} \int d^3x (\mathbf{x} \cdot \mathbf{J}(\mathbf{x}))\end{aligned}\tag{92}$$

However,

$$\nabla \cdot (r^2 \mathbf{J}(\mathbf{x})) = 2\mathbf{x} \cdot \mathbf{J}(\mathbf{x}) + r^2(\nabla \cdot \mathbf{J}(\mathbf{x}));\tag{93}$$

the final term here is zero, so it is the case that

$$\int d^3x (\mathbf{x} \cdot \mathbf{J}(\mathbf{x})) = \frac{1}{2} \int d^3x \nabla \cdot (r^2 \mathbf{J}(\mathbf{x})) = 0,\tag{94}$$

the final step following from the fact that the current distribution is localized. Thus the torque on the localized source reduces to

$$\mathbf{N} = \mathbf{m} \times \mathbf{B},\tag{95}$$

in the dipole approximation. Compare this result with Eq. (1).

7 Macroscopic Magnetostatics

If one has a macroscopic object with lots of microscopic currents flowing around in it (electrons on molecules give many small current loops), then it is best to carry out suitable averages over volumes small compared to anything macroscopic but large compared to molecules and

to calculate only the average quantities. The elementary procedure, in which many complications are blissfully ignored, is to break the sources up into free and bound:

1. the *macroscopic current density*, written as $\mathbf{J}(\mathbf{x})$
2. the *magnetization* $\mathbf{M}(\mathbf{x})$ which is defined as the magnetic dipole moment per unit volume.

The latter comes from the molecules, atoms, ions, etc., in a material medium. The macroscopic current density comes from moving charges which are not bound on molecules, *i.e.*, not localized. Notice that within this approach we can accommodate sources which cannot actually be described as the motion of charges, or currents; that is, the spin magnetic moments of “elementary” particles such as the electron. These may be treated as point dipoles and so simply contribute to the magnetization.

Given these sources, use of Eqs. (48) and (71) tell us that the (macroscopic) vector potential is

$$\mathbf{A}(\mathbf{x}) = \frac{1}{c} \int d^3x' \left[\frac{\mathbf{J}(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|} + \frac{c\mathbf{M}(\mathbf{x}') \times (\mathbf{x} - \mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^3} \right]. \quad (96)$$

We can manipulate the term involving the magnetization in a manner which should be familiar:

$$\mathbf{A}(\mathbf{x}) = \frac{1}{c} \int d^3x' \left\{ \frac{\mathbf{J}(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|} + \frac{c\mathbf{M}(\mathbf{x}') \times (\mathbf{x} - \mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^3} \right\}$$

$$\begin{aligned}
&= \frac{1}{c} \int d^3x' \left\{ \frac{\mathbf{J}(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|} + c\mathbf{M}(\mathbf{x}') \times \nabla' \left(\frac{1}{|\mathbf{x} - \mathbf{x}'|} \right) \right\} \\
&= \frac{1}{c} \int d^3x' \left\{ \frac{\mathbf{J}(\mathbf{x}') + c(\nabla' \times \mathbf{M}(\mathbf{x}'))}{|\mathbf{x} - \mathbf{x}'|} \right\} - \int d^3x' \nabla' \times \left(\frac{\mathbf{M}(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|} \right) \quad (97)
\end{aligned}$$

However, there is an identity which allows us to convert the integral of a curl to a surface integral:

$$\int_V d^3x \nabla \times \mathbf{V} = \oint_S d^2x \mathbf{n} \times \mathbf{V}. \quad (98)$$

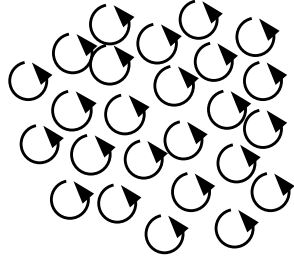
In the present instance, when the last integral in Eq. (97) is converted to a surface integral, it will vanish because if the surface encloses all of the magnetic materials, \mathbf{M} will be zero on the boundary. Hence we have the result


$$\mathbf{A}(\mathbf{x}) = \frac{1}{c} \int d^3x' \frac{\mathbf{J}(\mathbf{x}') + c[\nabla' \times \mathbf{M}(\mathbf{x}')] }{|\mathbf{x} - \mathbf{x}'|} \quad (99)$$

for the macroscopic vector potential.

7.1 Magnetization Current Density

The form of this equation implies that $c\nabla \times \mathbf{M}(\mathbf{x})$ is a current density associated with spatial variations of the magnetization. It is possible to imagine how the curl of the magnetization yields a current source, if we consider each magnetic dipole as originating from a vanishingly small current loop.



 Each microscopic current loop represents a dipole. The net circulation of the loops (the current density) is given by the curl of the loop density

This may or may not really be the case. If the magnetization is the consequence of the motions of bound charges on molecules, then it is indeed a current. But if the magnetization is the consequence of the intrinsic dipole moments of elementary particles such as the electron, then it is not reasonable to think of the curl of the magnetization as a current density. Nevertheless, we shall define the *magnetization current density* $\mathbf{J}_M(\mathbf{x})$ as³

$$\mathbf{J}_M(\mathbf{x}) = c\nabla \times \mathbf{M}(\mathbf{x}). \quad (100)$$

In terms of the magnetization current density, we have

$$\mathbf{A}(\mathbf{x}) = \frac{1}{c} \int d^3x' \frac{\mathbf{J}(\mathbf{x}') + \mathbf{J}_M(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|}. \quad (101)$$

The differential field equations obeyed by the *macroscopic magnetic induction*, which is the curl of the macroscopic vector potential,

$$\mathbf{B}(\mathbf{x}) = \nabla \times \mathbf{A}(\mathbf{x}), \quad (102)$$

³Notice that this object has zero divergence, as a steady-state current density should.

are easy to write down by referring to our earlier derivation; they are

$$\nabla \times \mathbf{B}(\mathbf{x}) = \frac{4\pi}{c}(\mathbf{J}(\mathbf{x}) + \mathbf{J}_M(\mathbf{x})) \quad (103)$$

and, of course,

$$\nabla \cdot \mathbf{B}(\mathbf{x}) = 0. \quad (104)$$

7.2 Magnetic Field

The difficulty one faces in solving these equations is much the same as faced in macroscopic electrostatics. One generally does not know magnetization until after solving for the induction. Further, the relation between $\mathbf{B}(\mathbf{x})$ and $\mathbf{M}(\mathbf{x})$ depends on the material. It is customary to define an additional macroscopic field $\mathbf{H}(\mathbf{x})$, called the *magnetic field*,

$$\mathbf{H}(\mathbf{x}) \equiv \mathbf{B}(\mathbf{x}) - 4\pi\mathbf{M}(\mathbf{x}). \quad (105)$$

Then Eq. (103) can be rewritten as

$$\nabla \times \mathbf{B}(\mathbf{x}) = \frac{4\pi}{c}(\mathbf{J}(\mathbf{x}) + c\nabla \times \mathbf{M}(\mathbf{x})) \quad (106)$$

or

$$\nabla \times (\mathbf{B}(\mathbf{x}) - 4\pi\mathbf{M}(\mathbf{x})) = \frac{4\pi}{c}\mathbf{J}(\mathbf{x}) \quad (107)$$

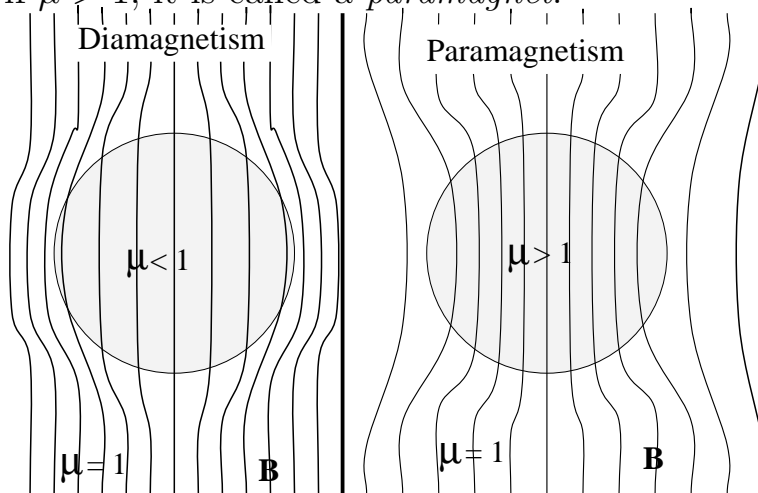
or

$$\nabla \times \mathbf{H}(\mathbf{x}) = \frac{4\pi}{c}\mathbf{J}(\mathbf{x}). \quad (108)$$

We haven't resolved anything by making this definition, of course; we have simply phrased the problem in a somewhat different form. The advantage in introducing the magnetic field is that it obeys a relatively simple differential equation and that, in many materials, there is a simple approximate relation between \mathbf{B} and \mathbf{H} . The simplest materials are, like dielectrics, linear, isotropic, and uniform, in which case the relation between \mathbf{B} and \mathbf{H} is

$$\mathbf{H}(\mathbf{x}) = \frac{1}{\mu} \mathbf{B}(\mathbf{x}) \quad (109)$$

where μ is a constant known as the *magnetic permeability*. This positive constant can be smaller or larger than unity, leading to two classifications of magnetic materials: If $\mu < 1$, the material is said to be a *diamagnet*; if $\mu > 1$, it is called a *paramagnet*.



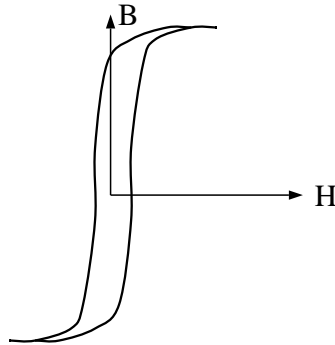
For linear, isotropic materials, it is also common to introduce the *mag-*

netic susceptibility χ_m ,

$$\chi_m \equiv \frac{\mu - 1}{4\pi} \quad (110)$$

so that $\mathbf{M} = \chi_m \mathbf{H}$.

Many materials have dramatically different relations among \mathbf{B} , \mathbf{M} , and \mathbf{H} . *Ferromagnets* are a prime example; they can have a finite magnetic induction with zero magnetic field as well as the converse. In addition, as shown in the figure below the relation between these two fields in ferromagnets is usually not single-valued; the particular magnetic field one finds for a given value of \mathbf{B} depends on the “history” of the sample, meaning it depends on what external fields it was subjected to prior to determining \mathbf{B} and \mathbf{H} . We will not concern ourselves with the origins of this behavior.



7.3 Boundary Conditions

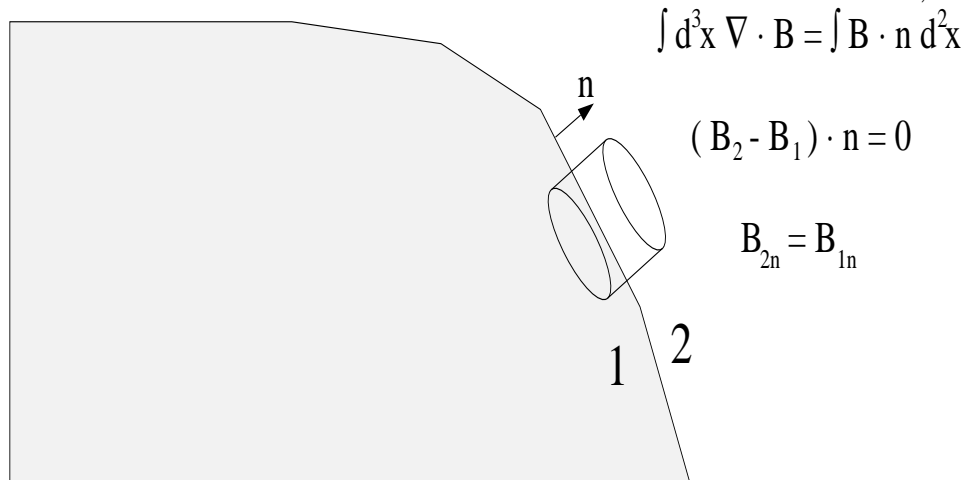
We have the general differential equations of magnetostatics

$$\nabla \times \mathbf{H}(\mathbf{x}) = \frac{4\pi}{c} \mathbf{J}(\mathbf{x}) \quad (111)$$

and

$$\nabla \cdot \mathbf{B}(\mathbf{x}) = 0. \quad (112)$$

These are valid for any system, independent of the particular relation between the magnetic induction and the magnetic field. From them we can derive general boundary or continuity conditions. From the divergence equation, we infer that the normal component of the magnetic induction is continuous at an interface between two materials,

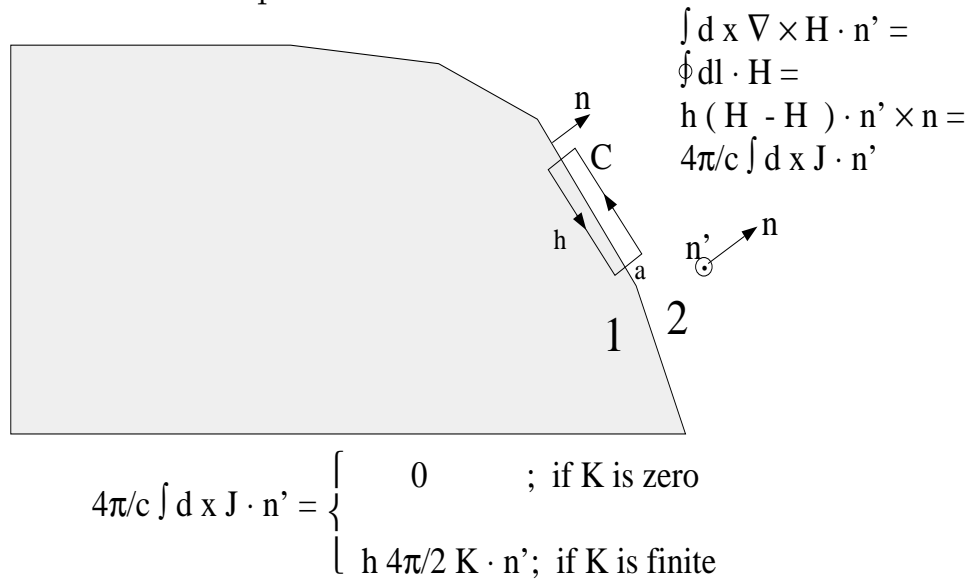


$$(\mathbf{B}_2 - \mathbf{B}_1) \cdot \mathbf{n} = 0 \quad (113)$$

where the subscripts 1 and 2 refer to the induction in each of two

materials meeting at an interface, and \mathbf{n} is a unit vector normal to the interface and pointing into medium 2.

From the curl equation, one finds that the tangential components of \mathbf{H} can be discontinuous. The discontinuity is related to the current flowing along (parallel to) the interface. Consider Stokes' theorem as applied to the curl equation.



Thus,

$$\int_S d^2x [\nabla \times \mathbf{H}(\mathbf{x})] \cdot \mathbf{n}' = \oint_C d\mathbf{l} \cdot \mathbf{H}(\mathbf{x}) = \frac{4\pi}{c} \int_S d^2x \mathbf{J}(\mathbf{x}) \cdot \mathbf{n}' \quad (114)$$

where the unit vector \mathbf{n}' is directed normal to the surface S over which the integration is done. Using a rectangle of dimensions a by h , where $a \ll h$ and a is directed perpendicular to the interface while h is parallel to it, we find that the line integral comes down to

$$\oint_C d\mathbf{l} \cdot \mathbf{H}(\mathbf{x}) = h(\mathbf{H}_2 - \mathbf{H}_1) \cdot (\mathbf{n}' \times \mathbf{n}) \quad (115)$$

which is to say, we find the discontinuity in a particular tangential component of \mathbf{H} across the interface. The integral over the current, on the other hand, gives, for a finite, slowly varying current,

$$\frac{4\pi}{c} \int_S d^2x \mathbf{J}(\mathbf{x}) \cdot \mathbf{n}' \sim ah\mathbf{J} \cdot \mathbf{n}'; \quad (116)$$

this is proportional to a and so is insignificant for $a \ll h$ assuming a finite well-behaved current density. In this case, there is no discontinuity in the tangential components of \mathbf{H} ,

$$(\mathbf{H}_2 - \mathbf{H}_1) \times \mathbf{n} = 0. \quad (117)$$

There is also the possibility of a singular term in $\mathbf{J}(\mathbf{x})$ which would be of the form

$$\mathbf{J}_s(\mathbf{x}) = \mathbf{K}(\mathbf{x})\delta(\xi) \quad (118)$$

where ξ is the distance from the interface and the vector \mathbf{K} points in a direction parallel to the interface. This vector is a *surface-current density* and has dimensions of charge per unit length per unit time. This expression is, of course, an idealization. When currents run along the surface of a material, they typically are not localized precisely at the surface (*i.e.*, , within an atomic size) but are spread over a surface layer of thickness ranging from a few hundred Angstroms to some microns. If the length a is significantly larger than this layer's thickness, then it is reasonable to talk about a surface-current density and quite acceptable

to regard it as being localized at the interface. Proceeding on this basis, we find that the continuity condition becomes

$$h(\mathbf{H}_2 - \mathbf{H}_1) \cdot (\mathbf{n}' \times \mathbf{n}) = h \frac{4\pi}{c} \mathbf{K} \cdot \mathbf{n}'. \quad (119)$$

This relation can be written in the more general form

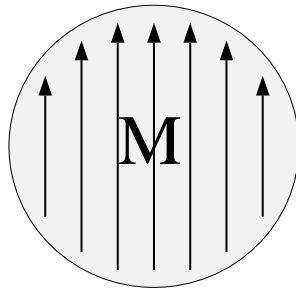
$$\mathbf{n} \times (\mathbf{H}_2 - \mathbf{H}_1) = \frac{4\pi}{c} \mathbf{K} \quad (120)$$

as may be shown by (1) realizing that $\mathbf{K} \cdot \mathbf{n} = 0$ and (2) taking the inner product of Eq. (120) with any vector (such as \mathbf{n}') lying in the plane of the interface.

8 Examples of Boundary-Value Problems in Magnetostatics

8.1 Uniformly Magnetized Sphere

The principal example is a uniformly magnetized sphere meaning a material that maintains a constant magnetization \mathbf{M}_0 in the absence of any applied field.



We shall let the direction of \mathbf{M} be the z-direction and so have

$$\mathbf{M}(\mathbf{x}) = \begin{cases} M_0 \hat{\mathbf{z}} & r < a \\ 0 & r > a. \end{cases} \quad (121)$$

In the region $r > a$, $\mathbf{B} = \mathbf{H}$. For $r < a$, $\mathbf{B} = \mathbf{H} + 4\pi\mathbf{M}$. However, \mathbf{M} is constant here and so has no curl or divergence. This means that the curls of both \mathbf{B} and \mathbf{H} are zero everywhere except at the boundary of the sphere.

Consider the magnetization current density, $\mathbf{J}_M = c\nabla \times \mathbf{M}$; it is non-zero only at $r = a$ where it is singular. By applying Stokes' theorem to this equation in the same manner as was done to find the continuity condition on the tangential components of \mathbf{H} at an interface, one finds that there is a magnetization surface-current density \mathbf{K}_M given by

$$\mathbf{K}_M = c\mathbf{n} \times (\mathbf{M}_2 - \mathbf{M}_1) \quad (122)$$

where \mathbf{n} is the unit normal at the surface pointing into material 2. In the present application, $\mathbf{M}_1 = \mathbf{M}_0$ and $\mathbf{M}_2 = 0$. Since $\hat{\mathbf{r}} \times \hat{\mathbf{z}} = -\sin\theta\hat{\phi}$, we have $\mathbf{K}_M = K_M\hat{\phi}$ where $K_M = cM_0\sin\theta$.

8.1.1 Scalar Potential for the Induction

Now that we have identified the sources, let's look at some methods of solution. First, consider a scalar potential approach. Because the curl of \mathbf{B} is zero for $r < a$ and $r > a$, we can devise scalar potentials for

the magnetic induction in these two regions. Because $\nabla \cdot \mathbf{B}(\mathbf{x}) = 0$, the potentials satisfy the Laplace equation. Further, the system is invariant under rotation around the z -axis, implying that the potential is independent of ϕ . Hence it must be possible to write

$$\Phi_{<}(\mathbf{x}) = M_0 a \sum_l A_l \left(\frac{r}{a}\right)^l P_l(\cos \theta) \text{ for } r < a \quad (123)$$

and

$$\Phi_{>}(\mathbf{x}) = M_0 a \sum_l C_l \left(\frac{a}{r}\right)^{l+1} P_l(\cos \theta) \text{ for } r > a. \quad (124)$$

Given that $\mathbf{B} = -\nabla\Phi(\mathbf{x})$, the condition that the normal component of \mathbf{B} is continuous at $r = a$ becomes (making use of the orthogonality of the Legendre polynomials in the usual way)

$$lA_l = -(l+1)C_l \quad (125)$$

for all l . The other boundary condition is that the tangential component of \mathbf{H} , or H_θ is continuous. Since $\mathbf{B} = \mathbf{H} + 4\pi\mathbf{M}$, and since the θ -component of the magnetization is $-M_0 \sin \theta$ (and $\sin \theta = -dP_l(\cos \theta)/d\theta$, this second condition leads to

$$A_l = C_l \text{ for } l \text{ not equal to } 1 \quad (126)$$

and

$$C_1 = A_1 + 4\pi. \quad (127)$$

From these equations it is easy to see that $A_l = C_l = 0$, $l \neq 1$, while $C_1 = 4\pi/3$ and $A_1 = -8\pi/3$.

Having found the potential, we may compute the fields. The magnetic induction at $r > a$ has the familiar dipolar form; the field within the sphere is a constant:

$$\mathbf{B}_{<} = \frac{8\pi}{3}\mathbf{M}_0 \text{ and } \mathbf{H}_{<} = -\frac{4\pi}{3}\mathbf{M}_0. \quad (128)$$

8.1.2 Scalar Potential for the Field

A second approach to this problem is to devise a scalar potential for the magnetic field. Since $\mathbf{J}(\mathbf{x}) = 0$ everywhere, the curl of \mathbf{H} is zero everywhere and so there is a potential Φ_H for \mathbf{H} **everywhere** such that $\mathbf{H}(\mathbf{x}) = -\nabla\Phi_H(\mathbf{x})$. The divergence of \mathbf{H} obeys

$$\nabla \cdot \mathbf{H}(\mathbf{x}) = -\nabla^2\Phi_H(\mathbf{x}) = \nabla \cdot \mathbf{B}(\mathbf{x}) - 4\pi\nabla \cdot \mathbf{M}(\mathbf{x}) = -4\pi\nabla \cdot \mathbf{M}(\mathbf{x}). \quad (129)$$

Looking upon this as a Poisson equation, we can immediately see that the solution for the potential is

$$\Phi_H(\mathbf{x}) = -\int d^3x' \frac{\nabla' \cdot \mathbf{M}(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|} = -\int d^3x' \left[\nabla' \cdot \left(\frac{\mathbf{M}(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|} \right) - \mathbf{M}(\mathbf{x}') \cdot \nabla' \left(\frac{1}{|\mathbf{x} - \mathbf{x}'|} \right) \right] \quad (130)$$

The first term in the final expression may be converted to a surface integral which is seen to be zero from the fact that the magnetization is non-zero only within the sphere. Hence

$$\begin{aligned} \Phi_H(\mathbf{x}) &= M_0\hat{\mathbf{z}} \cdot \int_{r' < a} d^3x' \nabla' \left(\frac{1}{|\mathbf{x} - \mathbf{x}'|} \right) = -\nabla \cdot \left[M_0\hat{\mathbf{z}} \int_{r' < a} \frac{d^3x'}{|\mathbf{x} - \mathbf{x}'|} \right] \\ &= -\nabla \cdot \left[M_0\hat{\mathbf{z}} \int_0^a r'^2 dr' 4\pi \frac{1}{r_{>}} \right] \end{aligned} \quad (131)$$

where $r_>$ is the larger of a and r . For $r > a$, we find

$$\Phi_H(\mathbf{x}) = -\nabla \cdot \left(M_0 \hat{\mathbf{z}} 4\pi \frac{a^3}{3r} \right) = \frac{4\pi}{3} a^3 M_0 \frac{\cos \theta}{r^2}, \quad (132)$$

and for $r < a$,

$$\Phi_H(\mathbf{x}) = -\nabla \cdot \left[M_0 \hat{\mathbf{z}} 4\pi \left(\frac{a^2}{2} - \frac{r^2}{6} \right) \right] = \frac{4\pi}{3} M_0 z. \quad (133)$$

Notice how the preceding calculation avoids having to think about what happens at the surface of the sphere; the integration by parts leaves us with a simpler integration over the magnetization. It is, of course, possible to evaluate the divergence of the magnetization and complete the integral directly without using the integration by parts. To this end, write the magnetization as

$$\mathbf{M}(\mathbf{x}) = M_0 \hat{\mathbf{z}} \theta(a - r) \quad (134)$$

where the θ -function is a step function,

$$\theta(x) = \begin{cases} 1, & x > 0 \\ 0, & x < 0. \end{cases} \quad (135)$$

Then

$$\nabla \cdot \mathbf{M}(\mathbf{x}) = M_0 \frac{\partial}{\partial z} \theta(a - r) = -M_0 \delta(a - r) \frac{\partial r}{\partial z} \quad (136)$$

since

$$\frac{d\theta(x)}{dx} = \delta(x). \quad (137)$$

Hence

$$\Phi_H(\mathbf{x}) = \int d^3x' \frac{\nabla' \cdot \mathbf{M}(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|} = M_0 a^2 \int d\Omega' \frac{\cos \theta'}{|\mathbf{x} - \mathbf{x}'|} \quad (138)$$

$$= M_0 a^2 \frac{4\pi}{3} \frac{3}{4\pi} \frac{4\pi}{3} \frac{r_{<}}{r_{>}^2} \cos \theta = \frac{4\pi}{3} M_0 a^2 \frac{r_{<}}{r_{>}^2} \cos \theta. \quad (139)$$

Here, \mathbf{x}' , after integrating over r' , has the magnitude a .

8.1.3 Direct Calculation of \mathbf{B}

Another approach to this problem is to calculate the vector potential for $\mathbf{B}(\mathbf{x})$ directly. Since $\mathbf{J}(\mathbf{x}) = 0$ everywhere, only the curl of the magnetization acts as a source of this potential. Thus we have

$$\begin{aligned} \mathbf{A}(\mathbf{x}) &= \frac{1}{c} \int d^3x' \frac{c \nabla' \times \mathbf{M}(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|} = \int d^3x' \left[\nabla' \times \left(\frac{\mathbf{M}(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|} \right) - \nabla' \left(\frac{1}{|\mathbf{x} - \mathbf{x}'|} \right) \times \mathbf{M}(\mathbf{x}') \right] \\ &= \oint_S d^2x' \mathbf{n}' \times \left(\frac{\mathbf{M}(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|} \right) + \nabla \times \left(\int d^3x' \frac{\mathbf{M}(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|} \right) \\ &= \nabla \times M_0 \hat{\mathbf{z}} 4\pi \begin{cases} a^3/3r, & r > a \\ a^2/2 - r^2/6, & r < a. \end{cases} \quad (140) \end{aligned}$$

One may easily work out the curl to find

$$\mathbf{A}(\mathbf{x}) = \frac{4\pi}{3} M_0 a^3 \frac{\sin \theta}{r^2} \hat{\phi} \text{ for } r > a \quad (141)$$

and

$$\mathbf{A}(\mathbf{x}) = \frac{4\pi}{3} M_0 r \sin \theta \hat{\phi} \text{ for } r < a. \quad (142)$$

These expressions may be summarized as

$$\mathbf{A}(\mathbf{x}) = \frac{4\pi M_0}{3} a^2 \left(\frac{r_{<}}{r_{>}^2} \right) \sin \theta \hat{\phi}. \quad (143)$$

The preceding example is a typical one involving a permanently magnetized material. Then one knows $\mathbf{M}(\mathbf{x})$ and so can proceed with a solution of the equations for \mathbf{B} or \mathbf{H} via the route of one's choice. A quite different sort of problem is one involving linear, isotropic magnetic materials with $\mathbf{J}(\mathbf{x}) = 0$. These are entirely equivalent to problems in macroscopic electrostatics (Laplace equation problems) with which we have extensive experience. The point is that one has

$$\nabla \times \mathbf{H}(\mathbf{x}) = 0 \quad (144)$$

which means there is a potential $\Phi_H(\mathbf{x})$ such that

$$\mathbf{H}(\mathbf{x}) = -\nabla \Phi_H(\mathbf{x}). \quad (145)$$

Further, since $\nabla \cdot \mathbf{B}(\mathbf{x}) = 0$, and $\mathbf{B}(\mathbf{x}) = \mu \mathbf{H}(\mathbf{x})$, we have

$$0 = \nabla \cdot \mathbf{B}(\mathbf{x}) = \nabla \cdot (\mu \mathbf{H}(\mathbf{x})) = (\nabla \mu) \mathbf{H}(\mathbf{x}) + \mu \nabla \cdot \mathbf{H}(\mathbf{x}) = -(\nabla \mu) \nabla \Phi_H(\mathbf{x}) - \mu \nabla^2 \Phi_H(\mathbf{x}), \quad (146)$$

or

$$\nabla^2 \Phi_H(\mathbf{x}) = -\frac{\nabla \mu}{\mu} \nabla \Phi_H(\mathbf{x}). \quad (147)$$

In any region of space where μ is a constant, this is just the Laplace equation,

$$\nabla^2 \Phi_H(\mathbf{x}) = 0. \quad (148)$$

Hence we may write down an appropriate solution of the Laplace equation for $\Phi_H(\mathbf{x})$ in each region of space where the magnetic properties are uniform and then make sure the continuity conditions on the tangential components of $\mathbf{H}(\mathbf{x})$ and on the normal components of $\mathbf{B}(\mathbf{x})$ are satisfied on the boundaries between such regions. The techniques used are the same as for boundary value problems in electrostatics.

Notice that in problems of this kind, we can equally well make a scalar potential for $\mathbf{B}(\mathbf{x})$ in each of the regions where μ is constant.

8.2 Shielding by a Paramagnetic Cylinder

A standard example is magnetic shielding in which a shell of magnetic material with a very large value of μ (a strongly paramagnetic material) is placed around some region of space. Suppose, for example, that a long cylindrical shell of inner radius a and outer radius b is placed around the z -axis and that this system is subjected to a transverse applied field $\mathbf{H}_0 = H_0 \hat{\mathbf{x}}$. Then we have three regions, $\rho < a$, $a < \rho < b$, and $b < \rho$, in which we may make potentials Φ_H which satisfy the Laplace equation. These potentials will have the forms

(i) $\rho < a$:

$$\Phi_H(\mathbf{x}) = H_0 A \rho \cos \phi \tag{149}$$

(ii) $a < \rho < b$:

$$\Phi_H(x) = H_0 [C\rho \cos \phi + Da^2 \cos \phi/\rho] \quad (150)$$

(iii) $b < \rho$:

$$\Phi_H(\mathbf{x}) = H_0 [-\rho \cos \phi + Eb^2 \cos \phi/\rho]. \quad (151)$$

Here, we have guessed (on the basis of long experience) that the only terms in the potential will be the ones with the same dependence on ϕ as the potential of the applied field. The latter is $-H_0x = -H_0\rho \cos \phi$.

Continuing, one has the boundary conditions at $\rho = a$ and $\rho = b$ which are that H_ϕ and B_ρ must be continuous. These lead to the four equations

$$\begin{aligned} A &= C + D \\ C + D\frac{a^2}{b^2} &= -1 + E \\ A &= \mu C - \mu D \\ &\text{and} \\ \mu C - \mu D\frac{a^2}{b^2} &= -1 - E. \end{aligned} \quad (152)$$

The solution for A in particular is

$$A = - \left(\frac{4\mu}{(\mu + 1)^2 - (a^2/b^2)(\mu - 1)^2} \right), \quad (153)$$

from which one finds that the field inside of the shield is

$$\mathbf{H}(\mathbf{x}) = \mathbf{B}(\mathbf{x}) = \frac{4\mu}{(\mu + 1)^2 - (a^2/b^2)(\mu - 1)^2} \mathbf{H}_0. \quad (154)$$

The field inside is decreased relative to the applied field by a factor of A . If one employs a “high- μ ” material with a permeability of order 10^4 or 10^5 , then $a \sim 4/\{\mu[1 - (a^2/b^2)]\}$ which will be much smaller than unity if a is significantly smaller than b . This is the limit in which the sleeve acts as a good shield against the applied magnetic field.

Maxwell's Equations

James Clerk Maxwell

(1831 - 1879)

November 9, 2001

$$\begin{aligned}\nabla \cdot \mathbf{B}(\mathbf{x}, t) &= 0, & \nabla \times \mathbf{H}(\mathbf{x}, t) &= \frac{4\pi}{c} \mathbf{J}(\mathbf{x}, t) + \frac{1}{c} \frac{\partial \mathbf{D}(\mathbf{x}, t)}{\partial t} \\ \nabla \times \mathbf{E}(\mathbf{x}, t) &= -\frac{1}{c} \frac{\partial \mathbf{B}(\mathbf{x}, t)}{\partial t}, & \nabla \cdot \mathbf{D}(\mathbf{x}, t) &= 4\pi \rho(\mathbf{x}, t)\end{aligned}$$

Contents

1	Faraday's Law of Induction	2
2	Energy in the Magnetic Field	7
2.1	Example: Motion of a permeable Bit in a Fixed \mathbf{J}	10
2.2	Energy of a Current distribution in an External Field	12
3	Maxwell's Displacement Current; Maxwell's Equations	15
4	Vector and Scalar Potentials	17
5	Gauge Transformations	19
5.1	Lorentz Gauge	19
5.2	Coulomb Transverse Gauge	21

6	Green's Functions for the Wave Equation	23
7	Derivation of Macroscopic Electromagnetism	28
8	Poynting's Theorem; Energy and Momentum Conservation	28
8.1	Energy Conservation	29
8.2	Momentum Conservation	31
8.3	Need for Field Momentum	35
8.4	Example: Force on a Conductor	36
9	Conservation Laws for Macroscopic Systems	37
10	Poynting's Theorem for Harmonic Fields; Impedance, Admittance, etc.	37
11	Transformations: Reflection, Rotation, and Time Reversal	42
11.1	Transformation Properties of Physical Quantities	43
12	Do Maxwell's Equations Allow Magnetic Monopoles?	48
A	Helmholtz' Theorem	50

Our first task in this chapter is to put time into the equations of electromagnetism. There are traditionally two steps in this process. The first of them is to develop Faraday's Law of Induction which is the culmination of a series of experiments performed by Michael Faraday (1791 - 1867) around 1830. Faraday studied the current "induced" in one closed circuit when a second nearby current-carrying circuit either was moved or had its current varied as a function of time. He also did experiments in which the second circuit was replaced by a permanent magnet in motion. The general conclusion of these experiments is that if the "magnetic flux" through a closed loop or circuit changes with time, an induced voltage or electromotive force (abbreviated by $\mathcal{E}mf$) appears in the circuit.

1 Faraday's Law of Induction

Define the *magnetic flux* through, or "linking" a closed loop C as

$$F = \int_S d^2x \mathbf{B} \cdot \mathbf{n} \quad (1)$$

where S is an open surface that ends on the curve C and \mathbf{n} is the usual unit right-hand normal (see below) to the surface. So long as $\nabla \cdot \mathbf{B} = 0$, this integral is the same for all such surfaces.

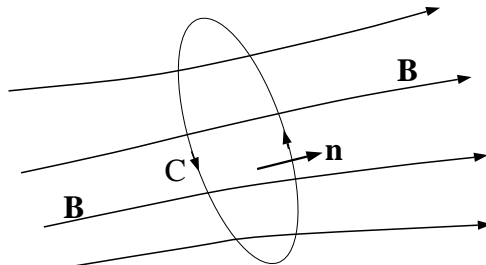


Figure 1: Orientation of \mathbf{n} and C.

Define next the $\mathcal{E}mf$ in the loop as

$$\mathcal{E} = \oint_C d\mathbf{l} \cdot \mathbf{E}' \quad (2)$$

where \mathbf{E}' is the electric field in that “frame of reference” in which the loop is at rest.¹ The path C is traversed in such a direction that the unit normal \mathbf{n} is the right-hand normal relative to this direction. We may now write *Faraday’s Law of Induction* as

$$\mathcal{E} = -k \frac{dF}{dt} \quad (3)$$

where t is the time and k is a positive constant.

Let us make two points in relation to this equation.

First, the minus sign has the consequence, when taken along with our definitions of the direction of \mathbf{n} , etc., that the induced electromotive force will try to drive a current through the loop in such a direction as to produce a flux through the loop that will be opposite in sign to the change in F that gave rise to the $\mathcal{E}mf$ in the first place. Thus if one tries to increase the flux F through the loop by manipulating some external currents or magnets, the resulting induced current will produce a $\mathbf{B}(\mathbf{x})$ which acts to counter the externally applied magnetic induction. This particular aspect of Faraday’s Law is also known as *Lenz’s Law*.

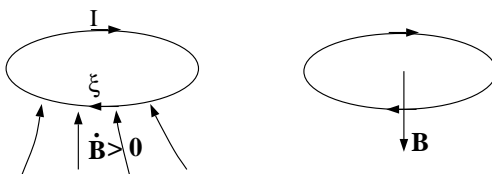


Figure 2: Induced current, and the resulting induced magnetic induction.

Second, in Gaussian units, the constant k has dimensions of T/L or inverse speed. It is in fact $1/c$ where c is the constant that appears in Ampère’s Law and in the Lorentz force law. This is not a separate experimental fact but may be deduced from classical notions of relativity and the laws of electromagnetism as we currently understand them. We may demonstrate this claim; consider the time rate of change of flux through a loop C that is moving with some constant velocity \mathbf{v} relative to the

¹Notice that this definition cannot cover the case of a rotating loop since such a loop is not at rest in any one inertial (unaccelerated) frame.

(lab) frame in which we are measuring \mathbf{x} and t . We have

$$\frac{dF}{dt} = \frac{d}{dt} \left(\int_S d^2x \mathbf{B}(\mathbf{x}, t) \cdot \mathbf{n} \right) = \int_S d^2x \frac{\partial \mathbf{B}(\mathbf{x}, t)}{\partial t} \cdot \mathbf{n} + \left(\frac{dF}{dt} \right)_2 \quad (4)$$

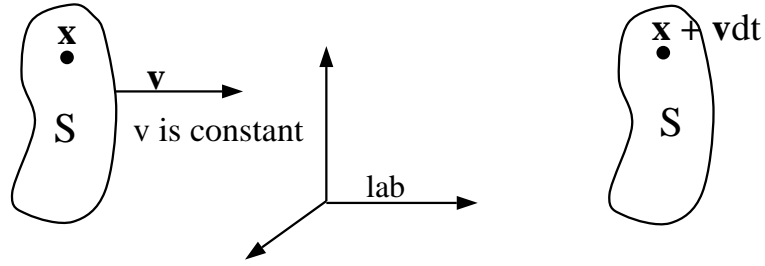


Figure 3: Loop moving relative to the lab frame

where the final term on the right accounts for the fact that the surface S over which we must integrate changes with time (it moves). To evaluate this term, we note that the distance the loop moves in time dt is $\mathbf{v}dt$ and that S is displaced by the same amount. A point on S initially at \mathbf{x} goes to $\mathbf{x} + \mathbf{v}dt$ in time dt , so if $\mathbf{B}(\mathbf{x})$ is sampled before dt elapses, then $\mathbf{B}(\mathbf{x} + \mathbf{v}dt)$ is sampled afterwards. We can expand the latter as

$$\mathbf{B}(\mathbf{x} + \mathbf{v}dt) = \mathbf{B}(\mathbf{x}) + dt(\mathbf{v} \cdot \nabla)\mathbf{B}(\mathbf{x}) + \dots \quad (5)$$

For an infinitesimal time element, we can ignore the higher-order terms. Hence, dF_2 , which is the integral of the change $[\mathbf{B}(\mathbf{x} + \mathbf{v}dt) - \mathbf{B}(\mathbf{x})] \cdot \mathbf{n}$ over the surface S (at time t) is easy to formulate:

$$\left(\frac{dF}{dt} \right)_2 = \int_S d^2x (\mathbf{v} \cdot \nabla)[\mathbf{B}(\mathbf{x}) \cdot \mathbf{n}]. \quad (6)$$

Now recall the vector identity $\nabla \times (\mathbf{v} \times \mathbf{B}) = \mathbf{v}(\nabla \cdot \mathbf{B}) - (\mathbf{v} \cdot \nabla)\mathbf{B}$. If we use the law $\nabla \cdot \mathbf{B}(\mathbf{x}) = 0$, we can use this identity to find that

$$\left(\frac{dF}{dt} \right)_2 = - \int_S d^2x \mathbf{n} \cdot [\nabla \times (\mathbf{v} \times \mathbf{B})] = - \oint_C d\mathbf{l} \cdot (\mathbf{v} \times \mathbf{B}), \quad (7)$$

the last step following from Stokes theorem. Using Eqs. (4) and (7) in Faraday's Law of Induction², we find

$$\oint_C d\mathbf{l} \cdot [\mathbf{E}' - k(\mathbf{v} \times \mathbf{B})] = -k \int_S d^2x \left(\frac{\partial \mathbf{B}}{\partial t} \cdot \mathbf{n} \right) \quad (8)$$

where \mathbf{E}' is the electric field in the frame where the loop is at rest; this frame moves at velocity \mathbf{v} relative to the one where \mathbf{B} is measured (along with \mathbf{x} and t).

Now consider a second circuit which is at rest in the frame where \mathbf{B} , \mathbf{x} , and t are measured and which coincides with the first loop at the particular time t .

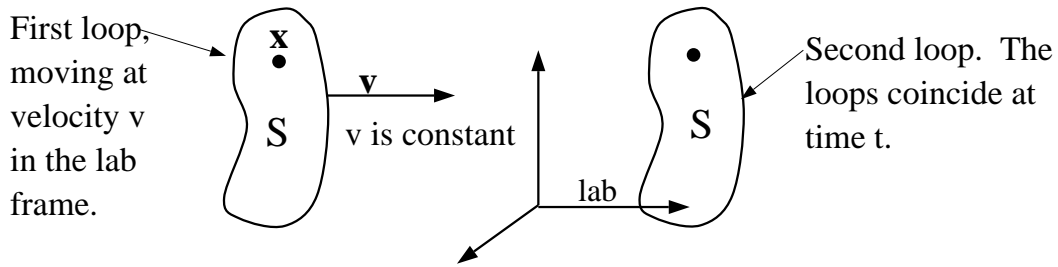


Figure 3b: Consider a second loop stationary in the lab frame

For this loop Faraday's Law says that

$$\oint_C d\mathbf{l} \cdot \mathbf{E} = -k \int_S d^2x \left(\frac{\partial \mathbf{B}}{\partial t} \cdot \mathbf{n} \right) \quad (9)$$

where \mathbf{E} is the electric field in the lab frame. Comparing Eqs. (8) and (9), we see that

$$\oint_C d\mathbf{l} \cdot [\mathbf{E}' - k(\mathbf{v} \times \mathbf{B})] = \oint_C d\mathbf{l} \cdot \mathbf{E}. \quad (10)$$

This relation tells us that the integrands are equal, give or take a vector field \mathbf{V} which has the property that the line integral of its component along the line is zero when taken around the loop C . This condition plus the arbitrariness of C tells us the $\nabla \times \mathbf{V} = 0$. This field can depend on \mathbf{v} , so we shall write it as $\mathbf{V}(\mathbf{v})$. Taking these statements together, we have the relation

$$\mathbf{E}' = \mathbf{E} + k(\mathbf{v} \times \mathbf{B}) + \mathbf{V}(\mathbf{v}). \quad (11)$$

² $\frac{dF}{dt} = -1/k \oint_C d\mathbf{l} \cdot \mathbf{E}'$

This relation gives us the transformation of the electric field from the lab frame to the rest frame of the moving circuit.

There is a second way to get this transformation. Consider a charge q following some trajectory $\mathbf{x}(t)$ under the influence of \mathbf{E} and \mathbf{B} , as seen from the “lab” frame (the one where the unprimed quantities are measured).

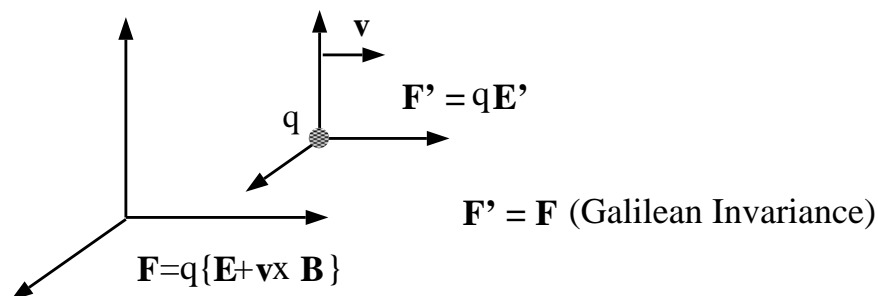


Figure 4: By Galilean invariance (see footnote) the forces $\mathbf{F}=\mathbf{F}'$

If at time t the particle is at some point \mathbf{x} and has velocity \mathbf{v} , the same as the velocity of the moving circuit, then it feels a force \mathbf{F} given by

$$\mathbf{F} = q[\mathbf{E} + \frac{1}{c}(\mathbf{v} \times \mathbf{B})]. \quad (12)$$

But in its rest frame, where the electric field is \mathbf{E}' , it feels a force

$$\mathbf{F}' = q\mathbf{E}'. \quad (13)$$

Now, according to classical, or Galilean, relativity, \mathbf{F}' is the same as \mathbf{F} , so, upon comparing the expressions for the force, we see that

$$\mathbf{E}' = \mathbf{E} + \frac{1}{c}(\mathbf{v} \times \mathbf{B}). \quad (14)$$

Comparison with Eq. (11) gives $k = 1/c$ and $\mathbf{V}(\mathbf{v}) = 0$. Using this result for k in Faraday’s Law, we find that it is fully specified by

$$\mathcal{E} = -\frac{1}{c} \frac{dF}{dt} \text{ or } \oint_C d\mathbf{l} \cdot \mathbf{E}(\mathbf{x}) = -\frac{1}{c} \int_S d^2x \left(\frac{\partial \mathbf{B}}{\partial t} \cdot \mathbf{n} \right) \quad (15)$$

for a stationary path.³ If we take the integral relation and apply Stokes theorem, we find

$$\int_S d^2x \left(\nabla \times \mathbf{E} + \frac{1}{c} \frac{\partial \mathbf{B}}{\partial t} \right) \cdot \mathbf{n} = 0. \quad (16)$$

Because S is an arbitrary open surface, this relation must be true for all such surfaces. That can only be if the integrand is everywhere zero, leading us to the differential equation

$$\nabla \times \mathbf{E}(\mathbf{x}, t) = -\frac{1}{c} \frac{\partial \mathbf{B}(\mathbf{x}, t)}{\partial t} \quad (17)$$

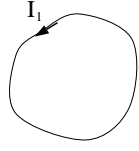
which is the differential equation statement of Faraday's Law.

2 Energy in the Magnetic Field

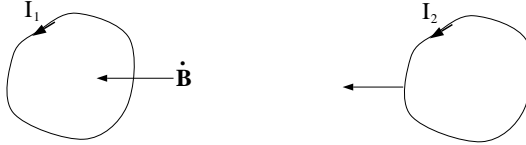
Given Faraday's Law, we are in a position to calculate the energy required to produce a certain current distribution \mathbf{J} starting from a state with $\mathbf{J} = 0$ even though we do not as yet know all of the time-dependent terms in the field equations. In this section, we shall determine what is this energy. The mechanism that requires work to be done is as follows: If we attempt to make a change in any existing current distribution, there will be time-dependent sources (the current) with an accompanying time-dependent magnetic induction. The latter must in turn produce electromagnetic forces, or electric fields, against which work must be done not only in order to change the currents but also simply in order to maintain them. By examining this work, we can determine the change in the "magnetic energy" of the system.

To get started, consider a single loop or circuit carrying current I_1 . Given a changing flux dF_1/dt through this loop, there is an $\mathcal{E}_1 = -c^{-1}dF_1/dt$. If we wish to maintain the current in the face of this electromotive force, we must counter the latter by introducing an

³It should be remarked that all of this makes sense only to order v/c since in the next order, v^2/c^2 , Galilean relativity fails. However, the conclusion that $k = 1/c$ must remain valid since k is a constant (according to experiments), independent of the relative size of any velocities.



The total current can be constructed from many current loops. We will start with one, and bring in another from infinity. To do this, we must change the flux through the first one, inducing an EMF in it.



Maintaining the current in the first loop requires work to overcome the induced EMF.

Figure 5: Energy required to construct a current distribution

external agent which maintains a voltage $V_1 = -\mathcal{E}_1 = \frac{1}{c} \frac{dF_1}{dt}$ around the loop. This agent thus does work at a rate $V_1 I_1 = -\mathcal{E}_1 I_1$. The total work that it does is

$$\delta W_1 = - \int dt I_1 \mathcal{E}_1 = \frac{I_1}{c} \int dt \frac{dF_1}{dt} = \frac{I_1 \delta F_1}{c} \quad (18)$$

where δF_1 is the total change of the flux through the loop. We may also express our result as

$$\delta W_1 = \frac{I_1}{c} \int_S d^2x (\delta \mathbf{B} \cdot \mathbf{n}) \quad (19)$$

where S and \mathbf{n} are related to the loop in the usual way, and $\delta \mathbf{B}$ is the change in the magnetic induction. If we write $\delta \mathbf{B} = \nabla \times \delta \mathbf{A}$, which is possible so long as $\nabla \cdot \delta \mathbf{B} = 0$, we have

$$\delta W_1 = \frac{I_1}{c} \int_S d^2x (\nabla \times \delta \mathbf{A}) \cdot \mathbf{n} = \frac{I_1}{c} \oint_C d\mathbf{l} \cdot \delta \mathbf{A} \quad (20)$$

where Stokes theorem has been invoked.

Now let us generalize. A current distribution \mathbf{J} can be thought of as the sum of many infinitesimal loops. Given many such loops, the total work done by the external agent(s) will be the sum of the works done on each of the loops. The relation between the current in a loop, I , and \mathbf{J} is $|\mathbf{J}|(d\sigma)$ where $d\sigma$ is the (infinitesimal) cross-sectional

area of the loop. Thus $I d\mathbf{l} \rightarrow \mathbf{J} d\sigma d\mathbf{l}$ or $\mathbf{J} d^3x$. Integrating over individual loops and summing over all loops is equivalent to integrating over all space, so we find that the change in energy accompanying a change $\delta\mathbf{A}(\mathbf{x})$ in the vector potential (reflecting an infinitesimal change $\delta\mathbf{B}(\mathbf{x})$ in the magnetic induction) is

$$\delta W = \frac{1}{c} \int d^3x [\mathbf{J}(\mathbf{x}) \cdot \delta\mathbf{A}(\mathbf{x})]. \quad (21)$$

The change in the magnetic induction has to be infinitesimal for this expression to be valid because we did not ask how the current density must be changed to produce it. Note that this form indicates that W is a natural thermodynamic function of the potential or magnetic flux, rather than the sources

Now let us write the current density in terms of the fields. If we think we are doing macroscopic electromagnetism, then⁴ $\mathbf{J} = c(\nabla \times \mathbf{H})/4\pi$ and we can proceed as follows:

$$\begin{aligned} \delta W &= \frac{1}{4\pi} \int_V d^3x [(\nabla \times \mathbf{H}) \cdot \delta\mathbf{A}] = \frac{1}{4\pi} \int_V d^3x [\nabla \cdot (\mathbf{H} \times \delta\mathbf{A}) + \mathbf{H} \cdot (\nabla \times \delta\mathbf{A})] \\ &= \frac{1}{4\pi} \oint_S d^2x (\mathbf{H} \times \delta\mathbf{A}) \cdot \mathbf{n} + \frac{1}{4\pi} \int_V d^3x (\mathbf{H} \cdot \delta\mathbf{B}) \\ &= \frac{1}{4\pi} \int d^3x (\mathbf{H} \cdot \delta\mathbf{B}). \quad (22) \end{aligned}$$

The final step in this argument is achieved by letting the domain of integration be all space and assuming that the fields \mathbf{H} and $\delta\mathbf{A}$ fall off fast enough far away that the surface integral vanishes. This is in fact true for a set of localized sources in the limit that changes are made very slowly.

In the final step of our derivation we want to integrate $\delta\mathbf{B}$ up to some final \mathbf{B} starting from zero magnetic induction or zero current density. We can only do this functional integral if we know how \mathbf{H} depends on \mathbf{B} . For a **linear** medium or set of

⁴This relation is only true for static phenomena; since we are changing the fields with time, it is not valid. However, if the change is accomplished sufficiently slowly that $\frac{\partial \mathbf{D}}{\partial t}$ may be neglected, then the corrections to Ampère's Law are so small as to have negligible influence on our argument. Notice too that use of this equation demands that changes in $\mathbf{J}(\mathbf{x})$ be done in such a way that $\nabla \cdot \mathbf{J} = 0$.

media, we know that

$$\mathbf{H} \cdot \delta \mathbf{B} = \frac{1}{2} \delta(\mathbf{H} \cdot \mathbf{B}) \quad (23)$$

and so

$$\delta W = \frac{1}{8\pi} \int d^3x \delta(\mathbf{B} \cdot \mathbf{H}) \quad (24)$$

which integrates to

$$W = \frac{1}{8\pi} \int d^3x \mathbf{B}(\mathbf{x}) \cdot \mathbf{H}(\mathbf{x}) \quad (25)$$

provided we define $W \equiv 0$ in the state with $\mathbf{B} \equiv 0$.

Our result, Eq. (25), can be written in other forms; a particularly useful one is obtained by writing $\mathbf{B} = \nabla \times \mathbf{A}$ and doing a parts integration in the familiar way. Assuming that one can discard the resulting surface term (valid for localized sources), we find

$$W = \frac{1}{2c} \int d^3x \mathbf{J}(\mathbf{x}) \cdot \mathbf{A}(\mathbf{x}). \quad (26)$$

In analogy with the electrostatic case, one conventionally defines the magnetic energy density to be

$$w \equiv \frac{1}{8\pi} (\mathbf{H}(\mathbf{x}) \cdot \mathbf{B}(\mathbf{x})). \quad (27)$$

2.1 Example: Motion of a permeable Bit in a Fixed \mathbf{J}

Let us look at a specific example of the use of the expression(s) for the energy. Suppose that there is some initial current distribution \mathbf{J}_0 which produces fields \mathbf{H}_0 and \mathbf{B}_0 and energy W_0 . Then we have

$$W_0 = \frac{1}{8\pi} \int d^3x \mathbf{B}_0 \cdot \mathbf{H}_0 = \frac{1}{2c} \int d^3x \mathbf{J}_0 \cdot \mathbf{A}_0. \quad (28)$$

Now move some permeable materials around without changing the macroscopic current density \mathbf{J}_0 .

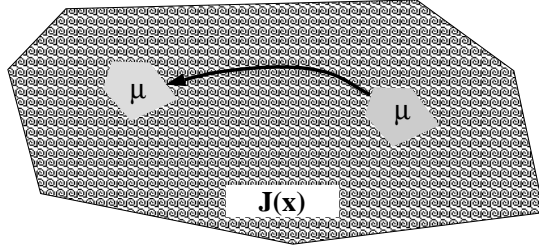


Figure 6: Motion of a permeable bit in a fixed current distribution

After doing so we have a new energy which is

$$W_1 = \frac{1}{8\pi} \int d^3x \mathbf{B}_1 \cdot \mathbf{H}_1 = \frac{1}{2c} \int d^3x \mathbf{J}_0 \cdot \mathbf{A}_1 \quad (29)$$

where the new fields, energy, etc., are designated by the subscript ‘1’. The change in energy may be written as

$$\begin{aligned} W_1 - W_0 &= \frac{1}{2c} \int d^3x \mathbf{J}_0 \cdot (\mathbf{A}_1 - \mathbf{A}_0) = \frac{1}{8\pi} \int d^3x [(\nabla \times \mathbf{H}_0) \cdot \mathbf{A}_1 - (\nabla \times \mathbf{H}_1) \cdot \mathbf{A}_0] \\ &= \left(\frac{1}{8\pi} \int d^3x \nabla \cdot [\mathbf{H}_0 \times (\mathbf{A}_1 - \mathbf{A}_0)] + \int d^3x [\mathbf{H}_0 \cdot (\nabla \times \mathbf{A}_1) - \mathbf{H}_1 \cdot (\nabla \times \mathbf{A}_0)] \right) \\ &= \frac{1}{8\pi} \left(\oint_S d^2x \mathbf{n} \cdot [\mathbf{H}_0 \times (\mathbf{A}_1 - \mathbf{A}_0)] + \int d^3x [\mathbf{H}_0 \cdot \mathbf{B}_1 - \mathbf{H}_1 \cdot \mathbf{B}_0] \right) \end{aligned} \quad (30)$$

The surface term may be discarded for localized sources. Assuming further that all materials are isotropic so that $\mathbf{B} = \mu\mathbf{H}$, we find

$$W_1 - W_0 = \frac{1}{8\pi} \int d^3x (\mu_1 - \mu_0)(\mathbf{H}_0 \cdot \mathbf{H}_1) \quad (31)$$

where μ_1 is the final value of the permeability (a function of position) and μ_0 is the initial value. If in addition $\mu_0 = 1$, a value appropriate for non-permeable materials or empty space, and $\mu_1 \neq 1$ only in some particular domain V , then

$$W_1 - W_0 = \frac{1}{2} \int_V d^3x (\mathbf{M}_1 \cdot \mathbf{B}_0) \quad (32)$$

where we have made use of the facts that $\mathbf{M}_1 = (\mu_1 - 1)\mathbf{H}_1/4\pi$ and $\mathbf{B}_0 = \mathbf{H}_0$.

It is instructive (maybe it’s just confusing) to compare the answer with other things that we have seen. First, we calculated previously the change in the electrostatic energy when a piece of dielectric is introduced into a previously empty space

in the presence of some fixed charges. We found

$$W_1 - W_0 = -\frac{1}{2} \int_V d^3x \mathbf{P}_1 \cdot \mathbf{E}_0. \quad (33)$$

The difference in sign between the electrostatic and magnetostatic energies is a reflection of the fact that in the magnetic system we maintained fixed currents and in the electrostatic system, fixed charges. In the former case external agents must do work to maintain the currents, and in the latter one, no work need be done to maintain the fixed charges. Hence in the latter case, the force on the dielectric may be found by using the argument that in a conservative system, the force is the negative gradient of the energy (which is given above) with respect to the displacement of the dielectric. **If**, in the magnetic system, the work done on the system by the current-maintaining external agents turns out to be precisely twice as large as the change in the system's energy, then the force on the permeable material will be the (positive) gradient of the energy with respect to the displacement of the material. This is, in fact, the case, as we shall see below.

2.2 Energy of a Current distribution in an External Field

As a second example, consider the energy of a current distribution in an external field.

$$W = \frac{1}{c} \int d^3x \mathbf{J}(\mathbf{x}) \cdot \mathbf{A}(\mathbf{x}). \quad (34)$$

where $\mathbf{A}(\mathbf{x})$ is due to sources other than $\mathbf{J}(\mathbf{x})$, which do not overlap the region where $\mathbf{J}(\mathbf{x})$ is finite. (note the lack of a factor of $\frac{1}{2}$, why?)

Source of
the vector
potential

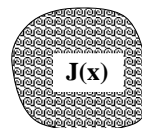


Figure 7: The source of the vector potential is far removed from the current distribution.

Now assume that $\mathbf{A}(\mathbf{x})$ changes little over the region where $\mathbf{J}(\mathbf{x})$ is finite, and thus expand $\mathbf{A}(\mathbf{x})$ about the origin of the current distribution.

$$\mathbf{A}(\mathbf{x}) = \mathbf{A}(0) + \mathbf{x} \cdot \nabla \mathbf{A}(\mathbf{x})|_{\mathbf{x}=0} + \dots$$

through the now familiar manipulations, we end up with

$$W = \mathbf{m} \cdot \mathbf{B}(0) + \dots$$

where \mathbf{m} is the dipole moment of the current distribution

This appears to make no sense at all!! Recall last quarter, we found that the force on a permanent dipole is

$$\mathbf{F} = \nabla(\mathbf{M} \cdot \mathbf{B}_0)$$

which acts to minimize the potential $-\mathbf{m} \cdot \mathbf{B}$, rather than $+\mathbf{m} \cdot \mathbf{B}$!

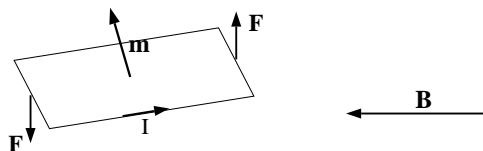


Figure 8: Forces on a current loop.

Consider a current loop. Clearly the force tries to make \mathbf{m} and \mathbf{B} parallel. Have we misplaced a “-”-sign? Our expression for W is correct; however, we do not have a conservative situation, since in calculating the force on the loop, we assumed that the current I is constant. However, rotating the loop changes the flux through it which induces an $\mathcal{E}mf$ which opposes the changing flux. Thus I will not be constant unless external work is done to make it so. In such a non-conservative situation, the force is not the negative gradient of the energy. For this particular situation, it must be that the the force is given by the *positive* gradient of the energy.

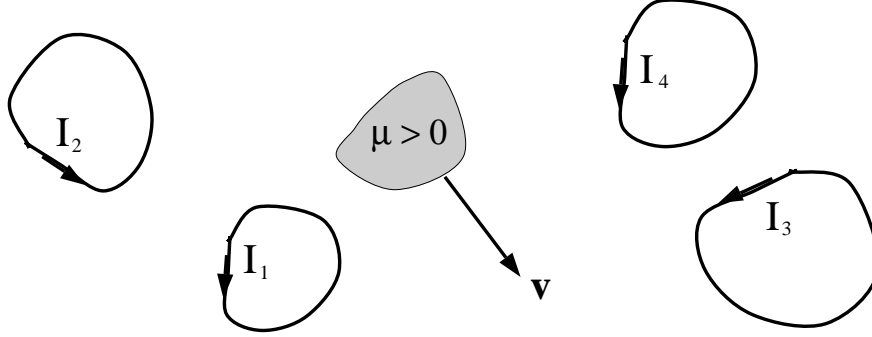
To demonstrate that the last statement above is correct, we consider a set of circuits C_i , $i = 1, 2, \dots, n$, with currents I_i . The energy of this system is

$$W = \frac{1}{2c} \int d^3x \mathbf{J} \cdot \mathbf{A} = \frac{1}{2c} \sum_i I_i \oint_{C_i} d\mathbf{l} \cdot \mathbf{A} = \frac{1}{2c} \sum_i \int_{S_i} d^2x \mathbf{n} \cdot (\nabla \times \mathbf{A})$$

$$= \frac{1}{2c} \sum_i I_i \int_{S_i} d^2x (\mathbf{n} \cdot \mathbf{B}) = \frac{1}{2c} \sum_i I_i F_i \quad (35)$$

where F_i is the flux linking the i^{th} circuit.

Now suppose that a piece of permeable material moves through the system at velocity \mathbf{v} .



It will experience a force which we may calculate by demanding total energy conservation. The various energies that must be considered are as follows:

1. The energy transferred to the moving object from the magnetic field. In time dt , this energy is

$$dW_m = \mathbf{F} \cdot \mathbf{v} dt = F_\eta v dt \quad (36)$$

given that \mathbf{v} is in the η -direction.

2. The field energy; for fixed currents in the circuits, this energy changes by

$$dW = \frac{1}{2c} \sum_i I_i dF_i \quad (37)$$

in time dt .

3. The energy transferred to the magnetic field (or circuits) by some external agents whose business it is to maintain the currents. The $\mathcal{E}mf$'s in the circuits are $\mathcal{E}_i = -c^{-1} dF_i/dt$ and so the external agents do work on the i^{th} circuit at a rate $-\mathcal{E}_i I_i$ in order to maintain the currents. The work done on these external

agents in time dt is therefore

$$dW_e = \sum_i \mathcal{E}_i I_i dt = -\frac{1}{c} \sum_i I_i dF_i = -2dW \quad (38)$$

Now invoke conservation of energy which demands that the sum of all of the preceding infinitesimal energy changes must be zero

$$dW_m + dW + dW_e = 0 \quad (39)$$

or, using Eqs. (36) and (38),

$$F_\eta v dt = dW. \quad (40)$$

Further, $v dt = d\eta$, so we find that

$$F_\eta = + \left(\frac{\partial W}{\partial \eta} \right)_{\mathbf{J}} \quad (41)$$

as suggested earlier.

What would the force be if the flux F was held constant?

3 Maxwell's Displacement Current; Maxwell's Equations

Let us summarize the equations of electromagnetism as we now have them:

$$\begin{aligned} \nabla \cdot \mathbf{B} &= 0 \\ \nabla \cdot \mathbf{D} &= 4\pi\rho \\ \nabla \times \mathbf{H} &= \frac{4\pi}{c} \mathbf{J} \\ \nabla \times \mathbf{E} &= -\frac{1}{c} \frac{\partial \mathbf{B}}{\partial t}. \end{aligned} \quad (42)$$

These are not internally consistent. Consider the divergence of Ampère's Law:

$$\nabla \cdot (\nabla \times \mathbf{H}) = 0 = \frac{4\pi}{c} (\nabla \cdot \mathbf{J}) \quad (43)$$

which implies that $\nabla \cdot \mathbf{J} = 0$. We know, however, that for general time-dependent phenomena the divergence of the current density is not necessarily zero; indeed, we have seen that charge conservation requires

$$\nabla \cdot \mathbf{J} + \frac{\partial \rho}{\partial t} = 0. \quad (44)$$

In consequence of this requirement, we must, at the very least, add a (time-dependent) term to Ampère's Law which will give, in distinction to Eq. (43),

$$\nabla \cdot (\nabla \times \mathbf{H}) = \frac{4\pi}{c} \left[\nabla \cdot \mathbf{J} + \frac{\partial \rho}{\partial t} \right]. \quad (45)$$

This new term which must be a vector field \mathbf{X} , has to be such that

$$\nabla \cdot \mathbf{X} = \frac{4\pi}{c} \frac{\partial \rho}{\partial t}. \quad (46)$$

We need not look beyond the things we have already learned to find a plausible candidate for this term. We have a (static) equation which reads

$$\rho = \frac{1}{4\pi} \nabla \cdot \mathbf{D}; \quad (47)$$

if we accept this as correct, we have

$$\frac{4\pi}{c} \frac{\partial \rho}{\partial t} = \frac{1}{c} \frac{\partial}{\partial t} \nabla \cdot \mathbf{D} = \nabla \cdot \left(\frac{1}{c} \frac{\partial \mathbf{D}}{\partial t} \right). \quad (48)$$

The simplest possible resolution of the inconsistency in the field equations is thus to choose \mathbf{X} to be $c^{-1} \partial \mathbf{D} / \partial t$, which would turn Ampère's Law into the relation

$$\nabla \times \mathbf{H} = \frac{4\pi}{c} \mathbf{J} + \frac{1}{c} \frac{\partial \mathbf{D}}{\partial t}. \quad (49)$$

This adjustment was made⁵ by J. C. Maxwell in 1864, and the resulting set of differential field equations has since become known as *Maxwell's Equations*:

$$\nabla \cdot \mathbf{B}(\mathbf{x}, t) = 0, \quad (50)$$

⁵He delivered a paper containing this statement in 1864, but he'd had the idea at least as early as 1861.

$$\nabla \cdot \mathbf{D}(\mathbf{x}, t) = 4\pi\rho(\mathbf{x}, t), \quad (51)$$

$$\nabla \times \mathbf{E}(\mathbf{x}, t) = -\frac{1}{c} \frac{\partial \mathbf{B}(\mathbf{x}, t)}{\partial t}, \quad (52)$$

and

$$\nabla \times \mathbf{H}(\mathbf{x}, t) = \frac{4\pi}{c} \mathbf{J}(\mathbf{x}, t) + \frac{1}{c} \frac{\partial \mathbf{D}(\mathbf{x}, t)}{\partial t}. \quad (53)$$

The term that Maxwell added was called by him the *displacement current* \mathbf{J}_D :

$$\mathbf{J}_D \equiv \frac{1}{4\pi} \frac{\partial \mathbf{D}}{\partial t}. \quad (54)$$

This term has the appearance of an additional current entering Ampère's Law so that the latter reads

$$\nabla \times \mathbf{H} = \frac{4\pi}{c} (\mathbf{J} + \mathbf{J}_D). \quad (55)$$

Notice that $\nabla \cdot (\mathbf{J} + \mathbf{J}_D) = 0$. We shall not emphasize the “current” interpretation of \mathbf{J}_D because it is misleading; the displacement current does not describe a flow of charge and is not a true current density.

We close this section with two comments. **First**, the Maxwell equations must be regarded as empirically justified. In the years since Maxwell's final adjustment of the field equations of electromagnetism, they have been subjected to extensive experimental tests and have been found to be correct for classical phenomena (no quantum effects or general relativistic effects); with proper interpretation, they even have considerable validity within the realm of quantum phenomena. **Second**, the equations as we have written them are for macroscopic electromagnetism. The more fundamental version of these equations has \mathbf{B} in place of \mathbf{H} and \mathbf{E} in place of \mathbf{D} ; then the sources ρ and \mathbf{J} are the total charge and current densities.

4 Vector and Scalar Potentials

For time-dependent phenomena, which are fully described by the Maxwell equations, one can still write the fields \mathbf{E} and \mathbf{B} in terms of a scalar potential and a vector

potential. Because the divergence of the magnetic induction is still zero, one continues to be able to find a vector potential $\mathbf{A}(\mathbf{x}, t)$ which has the property that

$$\mathbf{B}(\mathbf{x}, t) = \nabla \times \mathbf{A}(\mathbf{x}, t); \quad (56)$$

this potential is not unique. Further, since the curl of the electric field is not zero in general, we cannot write $\mathbf{E}(\mathbf{x}, t)$ as the gradient of a scalar; however, from Faraday's Law, Eq. (52), and from Eq. (56), we have

$$\nabla \times \left(\mathbf{E}(\mathbf{x}, t) + \frac{1}{c} \frac{\partial \mathbf{A}(\mathbf{x}, t)}{\partial t} \right) = 0 \quad (57)$$

and so we can write the combination of fields in parentheses as the gradient of a scalar function $\Phi(\mathbf{x}, t)$:

$$\mathbf{E}(\mathbf{x}, t) + \frac{1}{c} \frac{\partial \mathbf{A}(\mathbf{x}, t)}{\partial t} = -\nabla \Phi(\mathbf{x}, t) \quad (58)$$

or

$$\mathbf{E}(\mathbf{x}, t) = -\nabla \Phi(\mathbf{x}, t) - \frac{1}{c} \frac{\partial \mathbf{A}(\mathbf{x}, t)}{\partial t}. \quad (59)$$

Equations (56) and (59) tell us how to find \mathbf{B} and \mathbf{E} from potentials; these potentials must themselves satisfy certain field equations that can be derived from the Maxwell equations involving the sources ρ and \mathbf{J} ; the “homogeneous” or source-free equations $\nabla \cdot \mathbf{B} = 0$ and $\nabla \times \mathbf{E} = c^{-1} \partial \mathbf{B} / \partial t$ are then automatically satisfied. Letting $\mathbf{H} = \mathbf{B}$ and $\mathbf{D} = \mathbf{E}$ for simplicity, we have, upon substituting Eqs. (56) and (59) into Eqs. (51) and (53),

$$\nabla^2 \Phi + \frac{1}{c} \frac{\partial}{\partial t} (\nabla \cdot \mathbf{A}) = -4\pi \rho \quad (60)$$

and

$$\nabla \times (\nabla \times \mathbf{A}) + \frac{1}{c^2} \frac{\partial^2 \mathbf{A}}{\partial t^2} + \nabla \left(\frac{1}{c} \frac{\partial \Phi}{\partial t} \right) = \frac{4\pi}{c} \mathbf{J}. \quad (61)$$

These equations clearly do not have particularly simple pleasing or symmetric forms. However, we have some flexibility left in the choice of the potentials because we can choose the vector potential's divergence in an arbitrary fashion.

5 Gauge Transformations

Suppose that we have some \mathbf{A} and Φ which give \mathbf{E} and \mathbf{B} correctly. Let us add to \mathbf{A} the gradient of a scalar function $\chi(\mathbf{x}, t)$, thereby obtaining \mathbf{A}' :

$$\mathbf{A}' \equiv \mathbf{A} + \nabla\chi. \quad (62)$$

The field \mathbf{A}' has the same curl as \mathbf{A} , and hence $\mathbf{B} = \nabla \times \mathbf{A}'$. However, the electric field is **not** given by $-\nabla\Phi - c^{-1}\partial\mathbf{A}'/\partial t$; we must therefore change Φ to Φ' where Φ' is chosen to that

$$\mathbf{E} = -\nabla\Phi' - \frac{1}{c}\frac{\partial\mathbf{A}'}{\partial t}. \quad (63)$$

To this end we consider $\Phi' = \Phi + \psi$ where ψ is some scalar function of \mathbf{x} and t . Our requirement, Eq. (63), is that

$$\mathbf{E} = -\nabla\Phi - \nabla\psi - \frac{1}{c}\frac{\partial\mathbf{A}}{\partial t} - \frac{1}{c}\frac{\partial(\nabla\chi)}{\partial t}. \quad (64)$$

However, we know that \mathbf{E} is given according to Eq. (59), so, combining this relation and Eq. (64), we find that ψ must satisfy the equation

$$\nabla\psi = -\frac{1}{c}\frac{\partial(\nabla\chi)}{\partial t}. \quad (65)$$

A clear possible choice of ψ is $\psi = -c^{-1}\partial\chi/\partial t$.

What we have learned is that, given potentials \mathbf{A} and Φ , we may make a *gauge transformation* to equally acceptable potentials \mathbf{A}' and Φ' given by

$$\begin{aligned} \mathbf{A}' &= \mathbf{A} + \nabla\chi \\ \Phi' &= \Phi - \frac{1}{c}\frac{\partial\chi}{\partial t} \end{aligned} \quad (66)$$

where $\chi(\mathbf{x}, t)$ is an arbitrary scalar function of position and time.

5.1 Lorentz Gauge

Now let's look again at the differential equations, Eqs. (60) and (61), for the potentials; these may be written as

$$\nabla^2\Phi + \frac{1}{c}\frac{\partial}{\partial t}(\nabla \cdot \mathbf{A}) = -4\pi\rho$$

$$\nabla^2 \mathbf{A} - \frac{1}{c^2} \frac{\partial^2 \mathbf{A}}{\partial t^2} - \nabla \left(\nabla \cdot \mathbf{A} + \frac{1}{c} \frac{\partial \Phi}{\partial t} \right) = -\frac{4\pi}{c} \mathbf{J} \quad (67)$$

where we have used the identity $\nabla \times (\nabla \times \mathbf{A}) = \nabla(\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A}$. These have a much more pleasing form if \mathbf{A} and Φ satisfy the *Lorentz condition*

$$\nabla \cdot \mathbf{A} + \frac{1}{c} \frac{\partial \Phi}{\partial t} = 0. \quad (68)$$

Supposing for the moment that such a choice is possible, we make use of Eq. (68) in Eq. (67) and find

$$\begin{aligned} \nabla^2 \Phi - \frac{1}{c^2} \frac{\partial^2 \Phi}{\partial t^2} &= -4\pi\rho \\ \nabla^2 \mathbf{A} - \frac{1}{c^2} \frac{\partial^2 \mathbf{A}}{\partial t^2} &= -\frac{4\pi}{c} \mathbf{J}. \end{aligned} \quad (69)$$

These are very pleasing in that there are distinct equations for Φ and \mathbf{A} , driven by ρ and \mathbf{J} , respectively; furthermore, all equations have the form of the classical wave equation,

$$\square^2 \psi(\mathbf{x}, t) = -4\pi f(\mathbf{x}, t) \quad (70)$$

where

$$\square^2 \equiv \nabla^2 - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \quad (71)$$

is known as the *D'Alembertian* operator.

Consider next whether it is generally possible to find potentials that satisfy the Lorentz condition. Suppose we have some potentials \mathbf{A}_0 and Φ_0 for a given set of sources \mathbf{J} and ρ . We make a gauge transformation to new potentials \mathbf{A} and Φ ,

$$\begin{aligned} \mathbf{A} &= \mathbf{A}_0 + \nabla \chi \\ \Phi &= \Phi_0 - \frac{1}{c} \frac{\partial \chi}{\partial t}, \end{aligned} \quad (72)$$

where the gauge function χ is to be chosen so that the Lorentz condition is satisfied by the new potentials. The condition on χ is thus

$$0 = \nabla \cdot \mathbf{A} + \frac{1}{c} \frac{\partial \Phi}{\partial t} = \nabla \cdot \mathbf{A}_0 + \nabla^2 \chi + \frac{1}{c} \frac{\partial \Phi_0}{\partial t} - \frac{1}{c^2} \frac{\partial^2 \chi}{\partial t^2}, \quad (73)$$

or

$$\left(\nabla^2 - \frac{1}{c^2} \frac{\partial^2}{\partial t^2}\right) \chi = -\nabla \cdot \mathbf{A}_0 - \frac{1}{c} \frac{\partial \Phi_0}{\partial t}. \quad (74)$$

The function that we seek is thus itself a solution of the classical wave equation with a “source” which is, aside from some constant factor, just $\nabla \cdot \mathbf{A}_0 + c^{-1}(\partial\Phi_0/\partial t)$. Such a function always exists. In fact, there are many solutions to this wave equation which means that there are many sets of potentials Φ and \mathbf{A} which satisfy the Lorentz condition. Potentials satisfying the Lorentz condition are said to be in the *Lorentz gauge*.

5.2 Coulomb Transverse Gauge

Another gauge which can be useful is the *Coulomb* or *transverse gauge*. It is defined by the condition that $\nabla \cdot \mathbf{A} = 0$. The beauty of this gauge is that in it the scalar potential satisfies the Poisson equation,

$$\nabla^2 \Phi(\mathbf{x}, t) = -4\pi\rho(\mathbf{x}, t). \quad (75)$$

We know the solution of this equation:

$$\Phi(\mathbf{x}, t) = \int d^3x' \frac{\rho(\mathbf{x}', t)}{|\mathbf{x} - \mathbf{x}'|}. \quad (76)$$

Notice that the time is the same at both the source point \mathbf{x}' and the field point \mathbf{x} . There is nothing unacceptable about this because the scalar potential is not a measurable quantity.

The vector potential in the Coulomb gauge is less satisfying; it obeys the wave equation

$$\square^2 \mathbf{A} = -\frac{4\pi}{c} \mathbf{J} + \frac{1}{c} \frac{\partial}{\partial t} (\nabla \Phi). \quad (77)$$

In practice one may solve for the potentials in the transverse gauge, given the sources ρ and \mathbf{J} , by first finding the scalar potential from the integral Eq. (76) and then using the result in Eq. (77) and solving the wave equation (see the following

section). One should not be surprised to learn that the “source” term in Eq. (77) involving the scalar potential can be made to look like a current; consider

$$\begin{aligned} \frac{1}{c} \frac{\partial}{\partial t} (\nabla \Phi) &= \frac{1}{c} \frac{\partial}{\partial t} \left[\nabla \left(\int d^3 x' \frac{\rho(\mathbf{x}', t)}{|\mathbf{x} - \mathbf{x}'|} \right) \right] = -\nabla \int d^3 x' \frac{\nabla' \cdot \mathbf{J}(\mathbf{x}', t)}{|\mathbf{x} - \mathbf{x}'|} \\ &= -\frac{4\pi}{c} \left(\frac{1}{4\pi} \nabla \int d^3 x' \frac{\nabla' \cdot \mathbf{J}(\mathbf{x}', t)}{|\mathbf{x} - \mathbf{x}'|} \right) \end{aligned} \quad (78)$$

The negative of the quantity within the parentheses (...) is called the *longitudinal current density*. More generally, the *longitudinal* and *transverse* components \mathbf{J}_l and \mathbf{J}_t of a vector field such as \mathbf{J} are defined by the conditions⁶

$$\mathbf{J}_l + \mathbf{J}_t = \mathbf{J}, \quad \nabla \times \mathbf{J}_l = 0, \quad \text{and} \quad \nabla \cdot \mathbf{J}_t = 0. \quad (79)$$

In other words, J_l and J_t satisfy the equations

$$\begin{aligned} \nabla \cdot \mathbf{J}_l &= \nabla \cdot \mathbf{J} & \nabla \cdot \mathbf{J}_t &= 0 \\ \nabla \times \mathbf{J}_l &= 0 & \nabla \times \mathbf{J}_t &= \nabla \times \mathbf{J}, \end{aligned} \quad (80)$$

which means that \mathbf{J}_l carries all of the divergence of the current density and \mathbf{J}_t has all of the curl.

From these equations and our knowledge of electrostatics we can see immediately that

$$\mathbf{J}_l(\mathbf{x}, t) = -\frac{1}{4\pi} \nabla \left(\int d^3 x' \frac{\nabla' \cdot \mathbf{J}(\mathbf{x}', t)}{|\mathbf{x} - \mathbf{x}'|} \right) \quad (81)$$

which may also be written as

$$\mathbf{J}_l(\mathbf{x}, t) = -\frac{1}{4\pi} \nabla \left(\nabla \cdot \int d^3 x' \frac{\mathbf{J}(\mathbf{x}', t)}{|\mathbf{x} - \mathbf{x}'|} \right), \quad (82)$$

and from that of magnetostatics we see that

$$\mathbf{J}_t(\mathbf{x}, t) = \nabla \times \left(\frac{1}{4\pi} \int d^3 x' \frac{\nabla' \times \mathbf{J}(\mathbf{x}', t)}{|\mathbf{x} - \mathbf{x}'|} \right) \quad (83)$$

⁶In the appendix it is shown that such a decomposition, into longitudinal and transverse parts, of a vector function of position is always possible

which may also be written as

$$\mathbf{J}_t(\mathbf{x}, t) = \nabla \times \left[\frac{1}{4\pi} \nabla \times \left(\int d^3x' \frac{\mathbf{J}(\mathbf{x}', t)}{|\mathbf{x} - \mathbf{x}'|} \right) \right]. \quad (84)$$

Comparison of Eq. (81) with Eqs. (77) and (76) shows that the wave equation for \mathbf{A} in the Coulomb or transverse gauge is

$$\square^2 \mathbf{A} = -\frac{4\pi}{c} (\mathbf{J} - \mathbf{J}_l) = -\frac{4\pi}{c} \mathbf{J}_t. \quad (85)$$

This equation lends justification to the term “transverse gauge.” In this gauge, the vector potential is driven by the transverse part of the current in the same sense that the vector potential is driven by the entire current in the Lorentz gauge.

6 Green’s Functions for the Wave Equation

We’re going to spend quite a lot of time looking for solutions of the classical wave equation

$$\square^2 \psi(\mathbf{x}, t) = -4\pi f(\mathbf{x}, t); \quad (86)$$

Therefore, it could be useful to have Green’s functions for the D’Alembertian operator, meaning functions $G(\mathbf{x}, t; \mathbf{x}', t')$ which satisfy the equation

$$\square^2 G(\mathbf{x}, t; \mathbf{x}', t') = -4\pi \delta(\mathbf{x} - \mathbf{x}') \delta(t - t') \quad (87)$$

subject to some boundary conditions.

Compare this equation with that for the Green’s function for the Laplacian operator, $\nabla^2 G(\mathbf{x}, \mathbf{x}') = -4\pi \delta(\mathbf{x} - \mathbf{x}')$. The solution for the latter is, as we know, $G(\mathbf{x}, \mathbf{x}') = 1/|\mathbf{x} - \mathbf{x}'|$ in an infinite space; physically, it is the scalar potential at \mathbf{x} produced by a unit point charge located at \mathbf{x}' . The function $G(\mathbf{x}, t; \mathbf{x}', t')$, by contrast has dimensions $1/LT$ and may be thought of as the “response” at the space-time point (\mathbf{x}, t) to a unit “source” at the space-time point (\mathbf{x}', t') , that is, to a source which exists only for a moment at a single space point. If this source is a pulse of

current or an evanescent point charge, then the Green's function would be the vector or scalar potential in the Lorentz gauge produced by that pulse.

As for the question of boundary conditions, we shall keep things as simple as possible by considering an infinite system and applying the boundary condition that $G(\mathbf{x}, t; \mathbf{x}', t') \rightarrow 0$ for $|\mathbf{x}| \rightarrow \infty$. There is also a boundary condition on the behavior of $G(\mathbf{x}, t; \mathbf{x}', t')$ as a function of t ; this is usually called an initial condition. We shall not make any specific statement now of the initial conditions but will keep them in mind as our derivation progresses.

We shall solve the wave equation for $G(\mathbf{x}, t; \mathbf{x}', t')$ by appealing to Fourier analysis. Assuming that the Fourier transform $q(\mathbf{k}, \omega)$ of a function $f(\mathbf{x}, t)$ exists, then one can write that function in terms of the transform as

$$f(\mathbf{x}, t) = \int d^3k d\omega e^{i(\mathbf{k}\cdot\mathbf{x}-\omega t)} q(\mathbf{k}, \omega). \quad (88)$$

This expression may be thought of as an expansion of $f(\mathbf{x}, t)$ using basis functions $e^{i(\mathbf{k}\cdot\mathbf{x}-\omega t)}$. These form a complete orthonormal set; the orthonormality condition is

$$\int d^3x dt e^{-i(\mathbf{k}'\cdot\mathbf{x}-\omega't')} e^{i(\mathbf{k}\cdot\mathbf{x}-\omega t)} = (2\pi)^4 \delta(\mathbf{k} - \mathbf{k}') \delta(\omega - \omega') \quad (89)$$

and the completeness relation is much the same,

$$\int d^3k d\omega e^{-i(\mathbf{k}\cdot\mathbf{x}'-\omega t')} e^{i(\mathbf{k}\cdot\mathbf{x}-\omega t)} = (2\pi)^4 \delta(\mathbf{x} - \mathbf{x}') \delta(t - t'). \quad (90)$$

If we substitute this last relation into the right-hand side of Eq. (87) and also expand $G(\mathbf{x}, t; \mathbf{x}', t')$ on the left-hand side as⁷

$$G(\mathbf{x}, t; \mathbf{x}', t') = \int d^3k d\omega g(\mathbf{k}, \omega) e^{i[\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')-\omega(t-t')]}, \quad (91)$$

then one finds that

$$\begin{aligned} \square^2 G(\mathbf{x}, t; \mathbf{x}', t') &= \int d^3k d\omega \left(-k^2 + \frac{\omega^2}{c^2} \right) g(\mathbf{k}, \omega) e^{i[\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')-\omega(t-t')]} \\ &= -4\pi \delta(\mathbf{x} - \mathbf{x}') \delta(t - t') = -\frac{4\pi}{(2\pi)^4} \int d^3k d\omega e^{i[\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')-\omega(t-t')]}. \end{aligned} \quad (92)$$

⁷From Eq. (85) one may infer that $G(\mathbf{x}, t; \mathbf{x}', t')$ can be written as a function of $\mathbf{x} - \mathbf{x}'$ and of $t - t'$ in an infinite space.

The orthonormality of the basis functions may now be used to argue that the integrands on the two sides of this equation must be equal; hence,

$$g(\mathbf{k}, \omega) = \frac{4\pi}{(2\pi)^4} \frac{1}{k^2 - \omega^2/c^2}. \quad (93)$$

Now we have only to do the Fourier transform to find $G(\mathbf{x}, t; \mathbf{x}', t')$. Consider first the frequency integral,

$$I(\mathbf{k}, t - t') = \int_{-\infty}^{\infty} d\omega \left(\frac{e^{-i\omega(t-t')}}{k^2 - \omega^2/c^2} \right). \quad (94)$$

This integral can have different values depending on how we handle the poles in the integrand at $\omega = \pm ck$. Let's see what are the possibilities: If $t < t'$, we can extend the path of integration at $|\omega| \rightarrow \infty$ around a semi-circle in the upper half-plane where $\Re \omega > 0$ without getting an additional non-zero contribution to the integral.

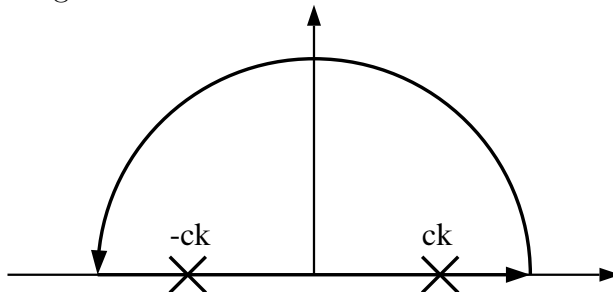


Figure 10: Appropriate Contour when $t < t'$

Hence, in this case the integral in Eq. (94) is just $2\pi i$ times the sum of the residues of the poles of the integrand in the upper half-plane. The only poles in the integrand as it stands are the ones at $\omega = \pm ck$. The path specified in the integral in fact runs right across both of them. However, if we make infinitesimal deformations of the path so that the poles are either inside of or outside of the contour, we will still obtain a function $G(\mathbf{x}, t; \mathbf{x}', t')$ which is a solution of Eq. (87). Thus the solution that we are obtaining is not unique.

Why not? The answer lies in the fact that we are dealing with a differential equation which is second-order in time, and there are two independent solutions which we can discriminate by specifying some initial condition. Consider what happens if

we deform the contour so as to go just above the poles on the real- ω axis. Then, for $t < t'$, we close the contour in the upper half-plane and find that the integral around the closed contour is zero (there are no poles inside of it). This means that $G(\mathbf{x}, t; \mathbf{x}', t') = 0$ for $t < t'$. But if $t > t'$, then we must close the contour in the lower half-plane and will find that $G(\mathbf{x}, t; \mathbf{x}', t') \neq 0$. This Green's function is called the *retarded Green's function* and its essential character is that it is non-zero only for times t which are **later** than the time t' of the source pulse which “produces” it. Instead of deforming the contour, we may change the integrand in such a way as to put the poles an infinitesimal distance below the real- ω axis and so have

$$I(\mathbf{k}, t - t') = \int_{-\infty}^{\infty} d\omega \left(\frac{e^{-i\omega(t-t')}}{k^2 - (\omega + i\epsilon)^2/c^2} \right) \quad (95)$$

where the integral is to be evaluated in the limit that ϵ is a positive infinitesimal constant. In this way we obtain a Green's function that vanishes at all times t earlier than the time t' .

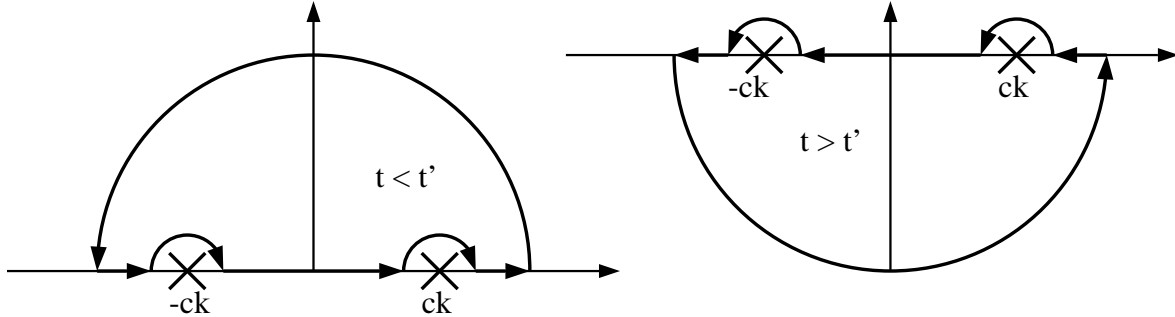


Figure 11: Contours for the retarded Green's function.

The second possibility is that we displace the poles so that they lie slightly above the real- ω axis. Then we will get a Green's function that will vanish for all $t > t'$ but not for times t earlier than t' ; this one is called the *advanced Green's function* and to have an expression for it, we need only change $+\epsilon$ to $-\epsilon$ in Eq. (95).

Now that we have determined how to handle the poles in the integrand, let us proceed with the evaluation of the retarded Green's function in particular. For $t < t'$, it is identically zero, and for $t > t'$, we close the contour in the lower half-plane and

find

$$\begin{aligned}
I(\mathbf{k}, t - t') &= -c^2 \oint_C d\omega \frac{e^{-i\omega(t-t')}}{[\omega - (ck - i\epsilon)][\omega - (-ck - i\epsilon)]} \\
&= -c^2(-2\pi i) \left[\frac{e^{-ick(t-t')}}{2ck} + \frac{e^{ick(t-t')}}{-2ck} \right] = \frac{2\pi c}{k} \sin[ck(t - t')].
\end{aligned} \tag{96}$$

Now we may complete the determination of $G(\mathbf{x}, t; \mathbf{x}', t')$ by doing the integration over the wavevector:

$$G(R, \tau) = \frac{1}{4\pi^3} \int d^3k \frac{2\pi c}{k} \sin(ck\tau) e^{i\mathbf{k}\cdot\mathbf{R}} \tag{97}$$

where $\tau \equiv t - t'$ and $\mathbf{R} \equiv (\mathbf{x} - \mathbf{x}')$. Thus,

$$\begin{aligned}
G(R, \tau) &= \frac{c}{2\pi^2} \int_0^\infty k dk \int_{-1}^1 du \int_0^{2\pi} d\phi \sin(ck\tau) e^{ikRu} \\
&= \frac{c}{\pi} \int_0^\infty k dk \left(\frac{e^{ikR} - e^{-ikR}}{ikR} \right) \sin(ck\tau) \\
&= -\frac{c}{4\pi R} \int_{-\infty}^\infty dk \left(e^{ikR} - e^{-ikR} \right) \left(e^{ick\tau} - e^{-ick\tau} \right) \\
&= -\frac{c}{2R} [\delta(R + c\tau) + \delta(-R - c\tau) - \delta(R - c\tau) - \delta(-R + c\tau)] = \frac{\delta(\tau - R/c)}{R};
\end{aligned} \tag{98}$$

the final step here follows from the fact that this expression is only correct for $\tau > 0$; for $\tau < 0$ $G \equiv 0$. In more detail, our result for the retarded Green's function is

$$G^{(+)}(\mathbf{x}, t; \mathbf{x}', t') = \frac{\delta(t - t' - |\mathbf{x} - \mathbf{x}'|/c)}{|\mathbf{x} - \mathbf{x}'|}. \tag{99}$$

By similar manipulations one may show that the advanced Green's function is

$$G^{(-)}(\mathbf{x}, t; \mathbf{x}', t') = \frac{\delta(t' - t - |\mathbf{x} - \mathbf{x}'|/c)}{|\mathbf{x} - \mathbf{x}'|}. \tag{100}$$

Given the appropriate Green's function, we can write down a solution to the classical wave equation with matching initial conditions and the appropriate boundary conditions as $|\mathbf{x}| \rightarrow \infty$. Suppose that the wave equation is

$$\square^2 \psi(\mathbf{x}, t) = -4\pi f(\mathbf{x}, t); \tag{101}$$

the solution is

$$\psi(\mathbf{x}, t) = \int d^3x' dt' G(\mathbf{x}, t; \mathbf{x}', t') f(\mathbf{x}', t') \quad (102)$$

as may be seen by operating on this equation with \square^2 :

$$\begin{aligned} \square^2 \psi(\mathbf{x}, t) &= \int d^3x' dt' \square^2 G(\mathbf{x}, t; \mathbf{x}', t') f(\mathbf{x}', t') \\ &= \int d^3x' dt' [-4\pi \delta(\mathbf{x} - \mathbf{x}') \delta(t - t')] f(\mathbf{x}', t') = -4\pi f(\mathbf{x}, t). \end{aligned} \quad (103)$$

The fact that $G(\mathbf{x}, t; \mathbf{x}', t')$ is proportional to a delta function means that one can always complete the integration over time trivially. For the retarded Green's function in particular, one finds that

$$\psi(\mathbf{x}, t) = \int d^3x' \frac{f(\mathbf{x}', t - |\mathbf{x} - \mathbf{x}'|/c)}{|\mathbf{x} - \mathbf{x}'|} \quad (104)$$

which has a fairly obvious interpretation.

Finally, it is worth pointing out that one may always add to $\psi(\mathbf{x}, t)$ a solution of the homogeneous wave equation.

7 Derivation of Macroscopic Electromagnetism

Regretfully omitted because of time constraints.

8 Poynting's Theorem; Energy and Momentum Conservation

A conservation law is a statement to the effect that some quantity q (such as charge) is a constant for an isolated system. Often, it is possible to express the law mathematically in the form

$$\nabla \cdot \mathbf{J}_q + \frac{\partial \rho_q}{\partial t} = \left(\frac{d\rho_q}{dt} \right)_e \quad (105)$$

Here, the density of the conserved quantity is ρ_q , its current density is \mathbf{J}_q , and the term on the right-hand side of the equation represents the contribution to the density at a particular space-time point arising from external sources or sinks. As we have already seen for the particular case of electrical charge, the divergence term represents the flow of the conserved quantity away from a point in space while the partial time derivative represents the rate at which its density is changing at that point. For electrical charge, there are no (known) sources or sinks and so the term on the right-hand side is zero.

8.1 Energy Conservation

It is possible to derive several such conservation laws from the Maxwell equations. These include the charge conservation law⁸ and also ones that are interpreted as energy, momentum, and angular momentum conservation. To get started, consider the rate per unit volume at which the fields transfer energy to the charged particles, or sources. The magnetic induction does no work since it is directed normal to the velocity of a particle, and so we have only the electric field which does work at the rate $\mathbf{J} \cdot \mathbf{E}$ per unit volume. This is, in the context of Eq. (105), the term $-(d\rho/dt)$ representing the transfer of energy to external (to the fields) sources or sinks.

The total power \mathcal{P}_m transferred to the sources within some domain V is found by integrating over that domain,

$$\mathcal{P}_m = \int_V d^3x (\mathbf{J} \cdot \mathbf{E}). \quad (106)$$

Now what we would like to do is perform some manipulations on $\mathbf{J} \cdot \mathbf{E}$ designed to remove all reference to the current density, leaving only electromagnetic fields; further, we want to make the expression look like the left-hand side of Eq. (105).

⁸Not surprising since the equations were explicitly constructed so as to be consistent with charge conservation.

That is not so hard to do, starting with the generalization of Ampère's law:

$$\begin{aligned}
\mathbf{J} \cdot \mathbf{E} &= \frac{1}{4\pi} \left[c\mathbf{E} \cdot (\nabla \times \mathbf{H}) - \mathbf{E} \cdot \frac{\partial \mathbf{D}}{\partial t} \right] \\
&= \frac{1}{4\pi} \left[-c\nabla \cdot (\mathbf{E} \times \mathbf{H}) + c\mathbf{H} \cdot (\nabla \times \mathbf{E}) - \mathbf{E} \cdot \frac{\partial \mathbf{D}}{\partial t} \right] \\
&= -\frac{c}{4\pi} \nabla \cdot (\mathbf{E} \times \mathbf{H}) - \frac{1}{4\pi} \left(\mathbf{H} \cdot \frac{\partial \mathbf{B}}{\partial t} + \mathbf{E} \cdot \frac{\partial \mathbf{D}}{\partial t} \right). \tag{107}
\end{aligned}$$

In the last step, Faraday's law has been employed. The result is promising; there is a divergence and the remainder is almost a time derivative. It becomes a time derivative if the materials in the system have linear properties so that

$$\mathbf{H} \cdot \frac{\partial \mathbf{B}}{\partial t} = \frac{1}{2} \frac{\partial}{\partial t} (\mathbf{B} \cdot \mathbf{H}) \quad \text{and} \quad \mathbf{E} \cdot \frac{\partial \mathbf{D}}{\partial t} = \frac{1}{2} \frac{\partial}{\partial t} (\mathbf{E} \cdot \mathbf{D}). \tag{108}$$

Then Eq. (107) becomes

$$-\mathbf{J} \cdot \mathbf{E} = \frac{c}{4\pi} \nabla \cdot (\mathbf{E} \times \mathbf{H}) + \frac{1}{8\pi} \frac{\partial}{\partial t} (\mathbf{E} \cdot \mathbf{D} + \mathbf{B} \cdot \mathbf{H}) \tag{109}$$

which is of the very form we seek. The interpretation of this equation is that the *Poynting vector* \mathbf{S} , defined by

$$\mathbf{S} \equiv \frac{c}{4\pi} (\mathbf{E} \times \mathbf{H}) \tag{110}$$

is the energy current density of the electromagnetic field and that the *energy density* u , defined by

$$u \equiv \frac{1}{8\pi} (\mathbf{E} \cdot \mathbf{D} + \mathbf{B} \cdot \mathbf{H}) \tag{111}$$

is, indeed, the energy density of the electromagnetic field. In terms of these quantities the conservation law is

$$-\mathbf{J} \cdot \mathbf{E} = \nabla \cdot \mathbf{S} + \frac{\partial u}{\partial t}. \tag{112}$$

Notice, however, that the energy current density is not unique. Because only its divergence enters into the conservation law, we may add to \mathbf{S} the curl of any vector field and would still have Eq. (112). The distinction is not important for measurable

quantities because one always measures the rate of energy flow through a closed surface (the surface of the detector),

$$\oint_S d^2x (\mathbf{n} \cdot \mathbf{S}) \equiv \int_V d^3x (\nabla \cdot \mathbf{S}), \quad (113)$$

so that only the divergence of \mathbf{S} is important. Finally, note that if the conservation law is integrated over some volume V , it can be expressed as

$$\int_V d^3x \frac{\partial u}{\partial t} = - \oint_S d^2x (\mathbf{n} \cdot \mathbf{S}) - \int_V d^3x (\mathbf{J} \cdot \mathbf{E}) \quad (114)$$

with the interpretation that rate of change of field energy within the domain V is equal to the rate at which field energy flows in through the surface of the domain plus the rate at which the sources within V transfer energy to the field.

8.2 Momentum Conservation

We can find a similar-looking momentum conservation law. We start from an expression for the rate per unit volume at which the fields transfer momentum to the sources; this is just the force density \mathbf{f} ,

$$\mathbf{f} = \rho \mathbf{E} + \frac{1}{c} (\mathbf{J} \times \mathbf{B}). \quad (115)$$

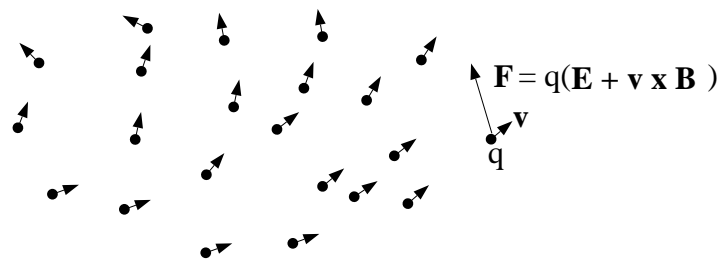


Figure 12: A swarm of charged particles.

Now use the Maxwell equations to remove all appearance of the sources ρ and \mathbf{J} :

$$\mathbf{f} = \left(\frac{\nabla \cdot \mathbf{D}}{4\pi} \right) \mathbf{E} + \frac{1}{c} \left(\frac{c}{4\pi} \nabla \times \mathbf{H} - \frac{1}{4\pi} \frac{\partial \mathbf{D}}{\partial t} \right) \times \mathbf{B}$$

$$\begin{aligned}
&= \frac{1}{4\pi} \left[\mathbf{E}(\nabla \cdot \mathbf{D}) + (\nabla \times \mathbf{H}) \times \mathbf{B} - \frac{1}{c} \left(\frac{\partial \mathbf{D}}{\partial t} \times \mathbf{B} \right) \right] \\
&= \frac{1}{4\pi} \left[\mathbf{E}(\nabla \cdot \mathbf{D}) - \mathbf{B} \times (\nabla \times \mathbf{H}) - \frac{1}{c} \frac{\partial}{\partial t} (\mathbf{D} \times \mathbf{B}) + \frac{1}{c} \left(\mathbf{D} \times \frac{\partial \mathbf{B}}{\partial t} \right) \right] \quad (116) \\
&= -\frac{1}{4\pi} \left\{ \frac{1}{c} \frac{\partial (\mathbf{D} \times \mathbf{B})}{\partial t} + [\mathbf{B} \times (\nabla \times \mathbf{H}) - \mathbf{H}(\nabla \cdot \mathbf{B}) + \mathbf{D} \times (\nabla \times \mathbf{E}) - \mathbf{E}(\nabla \cdot \mathbf{D})] \right\}
\end{aligned}$$

where we've used Faraday's law and the field equation $\nabla \cdot \mathbf{B} = 0$. Now we have to get all of the terms in the square brackets [...] in the final expression to look like a divergence. There is one complication which is that the equation is a vector equation, so we need a divergence which yields a vector, not a scalar.

Alternatively, we can look at Eq. (116) as three scalar equations, one for each Cartesian component of the force density. Then the term in brackets contains three components U_i and we want to write each component as the divergence of some vector field \mathbf{V}_i , $U_i = \nabla \cdot \mathbf{V}_i$. If we expand the vector fields as

$$\mathbf{V}_i = \sum_{j=1}^3 V_{ji} \boldsymbol{\epsilon}_j, \quad (117)$$

with

$$U_i = \nabla \cdot \mathbf{V}_i = \sum_j \frac{\partial V_{ji}}{\partial x_j}, \quad (118)$$

then we have nine scalar functions V_{ji} which can be put into a 3×3 matrix. Let us go a step further and define a *dyadic* $\bar{\mathbf{V}}$ by its inner products with the complete set of basis vectors $\boldsymbol{\epsilon}_i$:

$$\boldsymbol{\epsilon}_j \cdot \bar{\mathbf{V}} \cdot \boldsymbol{\epsilon}_i \equiv V_{ji}. \quad (119)$$

A convenient way to write $\bar{\mathbf{V}}$ in terms of its components is

$$\bar{\mathbf{V}} = \sum_{i,j=1}^3 \boldsymbol{\epsilon}_j V_{ji} \boldsymbol{\epsilon}_i, \quad (120)$$

with the understanding that when an inner product is taken of $\bar{\mathbf{V}}$ with a vector, the dot product is taken with the left or right $\boldsymbol{\epsilon}_i$ depending on whether the other vector

lies to the left or right of $\bar{\mathbf{V}}$. In particular,

$$\nabla \cdot \bar{\mathbf{V}} = \sum_k \epsilon_k \frac{\partial}{\partial x_k} \sum_{i,j} \epsilon_j V_{ji} \epsilon_i = \sum_{i,j} \frac{\partial V_{ji}}{\partial x_j} \epsilon_i = \sum_i U_i \epsilon_i. \quad (121)$$

Now what we need, referring back to Eq. (116), is some $\bar{\mathbf{V}}$ such that

$$\nabla \cdot \bar{\mathbf{V}} = \mathbf{B} \times (\nabla \times \mathbf{H}) - \mathbf{H}(\nabla \cdot \mathbf{B}) + \mathbf{D} \times (\nabla \times \mathbf{E}) - \mathbf{E}(\nabla \cdot \mathbf{D}) \quad (122)$$

where we shall restrict ourselves to the case of vacuum, or at least constant μ and ϵ . First, we have an identity,

$$\mathbf{B} \times (\nabla \times \mathbf{B}) = \frac{1}{2} \nabla (\mathbf{B} \cdot \mathbf{B}) - (\mathbf{B} \cdot \nabla) \mathbf{B} \quad (123)$$

which allows us to write

$$\begin{aligned} \mathbf{B} \times (\nabla \times \mathbf{H}) - \mathbf{H}(\nabla \cdot \mathbf{B}) &= \frac{1}{\mu} \left[\frac{1}{2} \nabla (\mathbf{B} \cdot \mathbf{B}) - (\mathbf{B} \cdot \nabla) \mathbf{B} - \mathbf{B}(\nabla \cdot \mathbf{B}) \right] \\ &= \frac{1}{\mu} \sum_{i,j} \left[\frac{1}{2} \delta_{ij} \epsilon_j \frac{\partial B^2}{\partial x_i} - B_i \frac{\partial B_j}{\partial x_i} \epsilon_j - B_j \epsilon_j \frac{\partial B_i}{\partial x_i} \right] \\ &= \frac{1}{\mu} \sum_k \left[\epsilon_k \frac{\partial}{\partial x_k} \right] \cdot \left[\sum_{i,j} \epsilon_i \left(\frac{1}{2} \delta_{ij} B^2 - B_i B_j \right) \epsilon_j \right] \\ &= \nabla \cdot \left[\sum_{i,j} \frac{1}{\mu} \epsilon_i \left(\frac{1}{2} \delta_{ij} B^2 - B_i B_j \right) \epsilon_j \right] \end{aligned} \quad (124)$$

which is indeed the divergence of a dyadic. By similar means one may demonstrate that

$$\mathbf{D} \times (\nabla \times \mathbf{E}) - \mathbf{E}(\nabla \cdot \mathbf{D}) = \nabla \cdot \left[\sum_{i,j} \epsilon \epsilon_i \left(\frac{1}{2} \delta_{ij} E^2 - E_i E_j \right) \epsilon_j \right]. \quad (125)$$

Putting these back into Eq. (124), we may write

$$-\mathbf{f} = \frac{\epsilon}{4\pi c} \frac{\partial}{\partial t} [(\mathbf{E} \times \mathbf{B})] - \nabla \cdot \bar{\mathbf{T}} \quad (126)$$

where the components T_{ij} of the *Maxwell stress tensor* are

$$T_{ij} \equiv \frac{1}{4\pi} \left[\epsilon E_i E_j + \frac{1}{\mu} B_i B_j - \frac{1}{2} \delta_{ij} \left(\epsilon E^2 + \frac{1}{\mu} B^2 \right) \right]. \quad (127)$$

The Maxwell stress tensor is a symmetric tensor (in a uniform, isotropic, linear medium) and has the physical interpretation that $-T_{ij}$ is the j -component of the current density of the i -component of momentum. Note, however, that $\bar{\mathbf{T}}$ is not unique in that if we redefine it to include the curl of another dyadic, Eq. (116) would still hold.

Let us define also a vector field \mathbf{g} ,

$$\mathbf{g} \equiv \frac{1}{4\pi c}(\mathbf{D} \times \mathbf{B}) \quad (128)$$

which is interpreted as the momentum density of the electromagnetic field, or the momentum per unit volume. Notice that for $\mathbf{D} = \epsilon\mathbf{E}$ and $\mathbf{B} = \mu\mathbf{H}$,

$$\mathbf{g} = \frac{\epsilon\mu}{4\pi c}(\mathbf{E} \times \mathbf{H}) = \frac{\mu\epsilon}{c^2}\mathbf{S}; \quad (129)$$

there is thus a simple connection between the energy current density and the momentum density of the field. Our conservation law is now simply written as

$$-\mathbf{f} = \frac{\partial \mathbf{g}}{\partial t} - \nabla \cdot \bar{\mathbf{T}}; \quad (130)$$

if integrated over some domain V , it may also be expressed as

$$-\int_V d^3x \mathbf{f} = \int_V d^3x \frac{\partial \mathbf{g}}{\partial t} - \int_S d^2x (\mathbf{n} \cdot \bar{\mathbf{T}}). \quad (131)$$

Finally, let us define

$$\frac{d\mathbf{P}_m}{dt} \equiv \int_V d^3x \mathbf{f} \quad \text{and} \quad \frac{d\mathbf{P}_f}{dt} \equiv \int_V d^3x \frac{\partial \mathbf{g}}{\partial t}; \quad (132)$$

these are rather obviously meant to be the time rate of change of the mechanical momentum⁹ and of the field momentum in V . Now we can write the conservation law as

$$\frac{d\mathbf{P}_m}{dt} + \frac{d\mathbf{P}_f}{dt} = \int_S d^2x \mathbf{n} \cdot \bar{\mathbf{T}}. \quad (133)$$

⁹Note, however, that it includes only the rate of change of mechanical momentum as a consequence of the forces applied to the particles by the electromagnetic field and does not include the change in mechanical momentum which comes about because particles are entering and leaving the domain V .

8.3 Need for Field Momentum

We defined our fields, the electric field and the magnetic induction, in terms of the force and torque, respectively, that they inflict upon an elementary source. However, as we have seen above, this definition is incomplete since the fields also carry energy and momentum. From a quantum point of view, this is obvious since we can think of the fields as resulting from the exchange of real and virtual photons each with momentum $\hbar\mathbf{k}$ and energy $\hbar\omega$. However, from a 19'th century perspective the need of the field to carry energy and momentum (especially momentum) is less obvious.

However, Newton's law of action and reaction requires that the field carry momentum. First consider two completely isolated particles in free space. If the only force exerted on either particle is from its counterpart, then the net momentum \mathbf{P} is conserved when the forces are equal and opposite.

$$\frac{d\mathbf{P}}{dt} = \frac{d}{dt} (\mathbf{P}_1 + \mathbf{P}_2) = m_1 \frac{d\mathbf{v}_1}{dt} + m_2 \frac{d\mathbf{v}_2}{dt} = \mathbf{F}_1 + \mathbf{F}_2$$

I.e. the momentum is conserved when $\mathbf{F}_1 = -\mathbf{F}_2$

Now consider the situation where the particles are charged, and have trajectories as shown in the figure below.

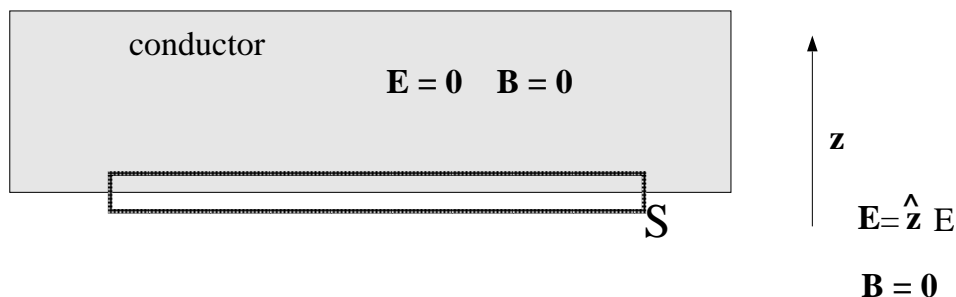


Clearly the forces due to the electric field are equal and opposite, but those due to the magnetic fields are not. In fact, charge 1 exerts a magnetic force on charge 2, but charge 2 does not exert a magnetic force on charge 1. Momentum is not conserved by the particles (and the forces on the particles) alone. The excess momentum must

be carried by the field! Thus, the field is required to have a momentum density by Newtonian law.

8.4 Example: Force on a Conductor

As an example, consider a conductor in the presence of an external electric field ($\mathbf{B} = 0$).



Since $\mathbf{B} = 0$ everywhere, $\mathbf{E} \times \mathbf{B} = 0$, and there is no electromagnetic momentum density. By conservation of momentum, the force on the conducting surface is then given

$$\left(\frac{d\mathbf{P}}{dt}\right)_i = F_i = \sum_j \int_S d^2x n_i T_{ij} \quad (134)$$

If we take the surface of integration as shown in the figure, then we need only consider the surface in the xy plane for which \mathbf{n} is in the $-\hat{\mathbf{z}}$ direction. Thus we only need T_{ij} when $i = z$. Since $\mathbf{E} = \hat{\mathbf{z}}E$, and since

$$T_{ij} = \frac{1}{4\pi} \left(E_i E_j - \frac{1}{2} \delta_{ij} E^2 \right) \quad (135)$$

when $\mathbf{B} = 0$, we have

$$T_{zx} = T_{zy} = 0 \quad T_{zz} = \frac{1}{8\pi} E^2. \quad (136)$$

Thus,

$$F_z = - \int_S d^2x \frac{1}{8\pi} E^2 = - \frac{A}{8\pi} E^2 \quad (137)$$

where A is the area of the conductor presented to the E -field. Thus the force on the conductor is downward

9 Conservation Laws for Macroscopic Systems

Regretfully omitted.

10 Poynting's Theorem for Harmonic Fields; Impedance, Admittance, etc.

Poynting's theorem has many important applications at the practical (electrical engineering) level. In this section we briefly make a foray into such an application. We shall restrict attention to harmonic fields which means ones that are harmonic (sine) functions of time. To this end we write the time-dependent part of the fields, as well as the sources, as $e^{-i\omega t}$ where ω is the angular frequency, and the physical field or source so represented is the real part of the complex mathematical function that we are using. For example,

$$\mathbf{E}(\mathbf{x}, t) = \Re [\mathbf{E}(\mathbf{x})e^{-i\omega t}] \equiv \frac{1}{2} [\mathbf{E}(\mathbf{x})e^{-i\omega t} + \mathbf{E}^*(\mathbf{x})e^{i\omega t}]. \quad (138)$$

Similarly,

$$\begin{aligned} \mathbf{J}(\mathbf{x}, t) \cdot \mathbf{E}(\mathbf{x}, t) &= \frac{1}{4} [\mathbf{J}(\mathbf{x})e^{-i\omega t} + \mathbf{J}^*(\mathbf{x})e^{i\omega t}] \cdot [\mathbf{E}(\mathbf{x})e^{-i\omega t} + \mathbf{E}^*(\mathbf{x})e^{i\omega t}] \\ &= \frac{1}{2} \Re [\mathbf{J}^*(\mathbf{x}) \cdot \mathbf{E}(\mathbf{x}) + \mathbf{J}(\mathbf{x}) \cdot \mathbf{E}(\mathbf{x})e^{-2i\omega t}]. \end{aligned} \quad (139)$$

The time-average of the product is particularly simple; the term which varies as $e^{-2i\omega t}$ has a zero time-average and so we are left with

$$\langle \mathbf{J}(\mathbf{x}, t) \cdot \mathbf{E}(\mathbf{x}, t) \rangle = \frac{1}{2} \Re [\mathbf{J}^*(\mathbf{x}) \cdot \mathbf{E}(\mathbf{x})]. \quad (140)$$

One can easily see that there is a general rule for the time-average of the product of two harmonic functions. Given $Q(\mathbf{x}, t)$ and $R(\mathbf{x}, t)$,

$$Q(\mathbf{x}, t) = \Re [Q(\mathbf{x})e^{-i\omega t}] \quad R(\mathbf{x}, t) = \Re [R(\mathbf{x})e^{-i\omega t}], \quad (141)$$

the time average of the product may be expressed as

$$\langle Q(\mathbf{x}, t)R(\mathbf{x}, t) \rangle = \frac{1}{2} \Re [R^*(\mathbf{x})Q(\mathbf{x})] = \frac{1}{2} \Re [R(\mathbf{x})Q^*(\mathbf{x})]. \quad (142)$$

We start by supposing that all fields are harmonic. Then they all have the complex form $\mathbf{F}(\mathbf{x})e^{-i\omega t}$ and their time derivatives are given by

$$\frac{\partial(\mathbf{F}(\mathbf{x})e^{-i\omega t})}{\partial t} = -i\omega\mathbf{F}(\mathbf{x})e^{-i\omega t} \quad (143)$$

so that the Maxwell equations for the complex amplitudes (just the position-dependent parts of the fields) are

$$\begin{aligned} \nabla \cdot \mathbf{B}(\mathbf{x}) &= 0 & \nabla \times \mathbf{E}(\mathbf{x}) &= -i\frac{\omega}{c}\mathbf{B}(\mathbf{x}) \\ \nabla \cdot \mathbf{D}(\mathbf{x}) &= 4\pi\rho(\mathbf{x}) & \nabla \times \mathbf{H}(\mathbf{x}) + i\frac{\omega}{c}\mathbf{D}(\mathbf{x}) &= \frac{4\pi}{c}\mathbf{J}(\mathbf{x}). \end{aligned} \quad (144)$$

Notice that there is no problem in generalizing the Maxwell equations to complex fields because the equations involve linear combinations, with real coefficients, of complex objects. One thus has two sets of equations, one for the real parts of these objects and one for the imaginary parts. The set for the real parts comprises the “true” Maxwell equations.

For the remainder of this section, the symbols \mathbf{E} , \mathbf{B} , etc, stand for the complex amplitudes $\mathbf{E}(\mathbf{x})$, $\mathbf{B}(\mathbf{x})$, We can rederive the Poynting theorem for these by starting from the inner product $\mathbf{J}^* \cdot \mathbf{E}$ and proceeding as in the original derivation. The result is

$$\mathbf{J}^* \cdot \mathbf{E} = \frac{c}{4\pi} \left[-\nabla \cdot (\mathbf{E} \times \mathbf{H}^*) - i\frac{\omega}{c}(\mathbf{E} \cdot \mathbf{D}^* - \mathbf{B} \cdot \mathbf{H}^*) \right]. \quad (145)$$

Define now the (complex) Poynting vector and energy densities

$$\mathbf{S} \equiv \frac{c}{8\pi}(\mathbf{E} \times \mathbf{H}^*), \quad (146)$$

and

$$w_e \equiv \frac{1}{16\pi}(\mathbf{E} \cdot \mathbf{D}^*) \quad w_m \equiv \frac{1}{16\pi}(\mathbf{B} \cdot \mathbf{H}^*). \quad (147)$$

Notice that the real part of \mathbf{S} is just the time-averaged real Poynting vector while the real parts of the energy densities are the time-averaged energy densities. More generally, the energy densities can be complex functions, depending on the relations between \mathbf{D} and \mathbf{E} and between \mathbf{B} and \mathbf{H} . If the two members of each pair of fields are in phase with one another, then the corresponding energy density is real. Similarly, if \mathbf{E} and \mathbf{H} are in phase, then \mathbf{S} is also real.

In terms of our densities, Poynting's theorem for harmonic fields becomes

$$\frac{1}{2}\mathbf{J}^* \cdot \mathbf{E} = -\nabla \cdot \mathbf{S} - 2i\omega(w_e - w_m). \quad (148)$$

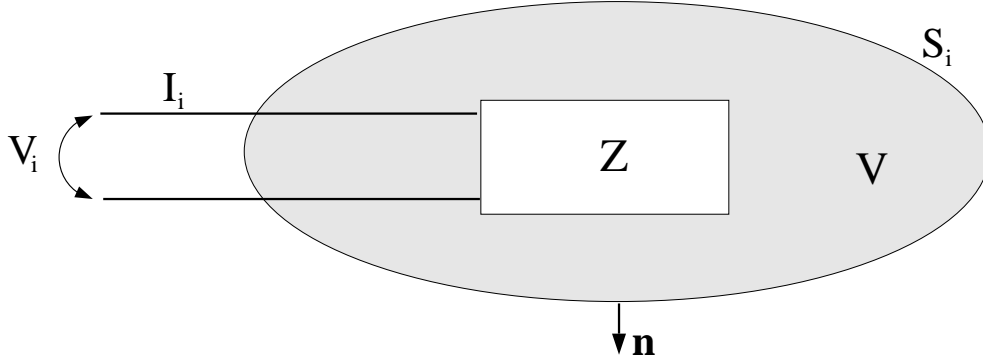
If we integrate this expression over some domain V and apply the divergence theorem to the term involving the Poynting vector, we find the relation

$$\frac{1}{2} \int_V d^3x \mathbf{J}^* \cdot \mathbf{E} + 2i\omega \int_V d^3x (w_e - w_m) + \oint_S d^2x \mathbf{S} \cdot \mathbf{n} = 0. \quad (149)$$

The interpretation of this equation is that the real part expresses the time-averaged conservation of energy. The imaginary part also has a meaning in connection with energy and its flow.

Consider first the simplest case of real w_e and w_m . Then the energy densities drop out of the real part of this equation and what it (the real part) tells us is that the time-average rate of doing work on the sources in V is equal to the time-averaged flow of energy (expressed by the Poynting vector) into V through the surface S . If the energy densities are not real, then there is an additional real piece in Eq. (149) so that the work done on the sources in V is not equal to the energy that comes in through S ; this case corresponds to having “lossy” materials within V which dissipate additional energy.

Now let's suppose that there is some electromagnetic device within V , i.e., surrounded by S . Let it have two input terminals which are its only material communication with the rest of the world. At these terminals there are some input current I_i and voltage V_i which we suppose are harmonic and which may also be written in the form Eq. (138).



Then the (complex) input power is $I_i^* V_i / 2$, meaning that the time-averaged input power is the real part of this quantity. Using our interpretation of the Poynting vector, we can express the input power in terms of a surface integral of the normal component of \mathbf{S} ,

$$\frac{1}{2} I_i^* V_i = - \int_{S_i} d^2 x \mathbf{S} \cdot \mathbf{n} \quad (150)$$

where the surface integral is done just over the cross-section of the (presumed) coaxial cable feeding power into the device; it is assumed that for such a cable, the input fields are confined to the region within the shield on the cable and so the integral over the remainder of the surface S surrounding the device has no contribution from the incident fields.

If we now combine this equation with Eq. (149), we find that we can write

$$\frac{1}{2} I_i^* V_i = \frac{1}{2} \int_V d^3 x \mathbf{J}^* \cdot \mathbf{E} + 2i\omega \int_V d^3 x (w_e - w_m) + \int_{S-S_i} d^2 x \mathbf{S} \cdot \mathbf{n}. \quad (151)$$

The surface integral in this expression gives the power passing through the surface S , excluding the part through which the input power comes. The real part of this integral is the power radiated by the device.

Now let us define the *input impedance* Z of the device,

$$V_i \equiv Z I_i; \quad (152)$$

the impedance is complex and so can be written as

$$Z \equiv R - iX \quad (153)$$

where the *resistance* R and the *reactance* X are real. From Eq. (151) we find expressions for these:

$$R = \frac{1}{|I_i|^2} \left\{ \text{Re} \left[\int_V d^3x \mathbf{J}^* \cdot \mathbf{E} + 2 \int_{S-S_i} d^2x \mathbf{S} \cdot \mathbf{n} \right] + 4\omega \text{Im} \left[\int_V d^3x (w_m - w_e) \right] \right\} \quad (154)$$

and

$$X = \frac{1}{|I_i|^2} \left\{ 4\omega \text{Re} \left[\int_V d^3x (w_m - w_e) \right] - \text{Im} \left[\int_V d^3x \mathbf{J}^* \cdot \mathbf{E} + 2 \int_{S-S_i} d^2x \mathbf{S} \cdot \mathbf{n} \right] \right\}. \quad (155)$$

By deforming the surface so that it lies far away from the device, one may make the integral over $\mathbf{S} \cdot \mathbf{n}$ purely real so that it does not contribute to the reactance; then it is only a part of the resistance and is the so-called “radiation resistance” which will be present if the device radiates a significant amount of power.

Our result has a simple and pleasing form at low frequencies. Then radiation is negligible and so the contributions of the surface integral may be ignored. Also, we may drop the term in the resistance proportional to ω . Then, assuming the current density and electric field are related by $\mathbf{J} = \sigma \mathbf{E}$ where σ is the (real) *electrical conductivity*, and assuming real energy densities, we find

$$R = \frac{1}{|I_i|^2} \int_V d^3x \sigma |\mathbf{E}|^2 \quad (156)$$

and

$$X = \frac{4\omega}{|I_i|^2} \int_V d^3x (w_m - w_e) \quad (157)$$

The last equation may be used to establish contact between our expressions, based on the electromagnetic field equations, and some standard and fundamental relations in elementary circuit theory. If there is an inductance (magnetic energy-storing device) in the “black-box,” then the integral of the magnetic energy may be expressed (see the first two or three problems at the end of Chap. 6 of Jackson) as $L|I_i|^2/4$, and so we find the familiar (if one knows anything about circuits) result that $X = L\omega$. But if there is a capacitor, the energy becomes $|Q_i|^2/4C$ where the charge Q_i is obtained by integrating the current over time; that gives $|Q_i|^2 = |I_i|^2/\omega^2$ and so $X = -1/\omega C$, another familiar tenet of elementary circuit theory.

11 Transformations: Reflection, Rotation, and Time Reversal

Before entering into discussion of the specific transformations of interest, we give a brief review of orthogonal transformations. Introduce a 3×3 matrix a with components a_{ij} and use it to transform a position vector $\mathbf{x} = (x_1, x_2, x_3) = (x, y, z)$ into a new vector \mathbf{x}' :

$$x'_i = \sum_j a_{ij} x_j. \quad (158)$$

An orthogonal transformation is one that leaves the length of the vector unchanged,

$$\sum_i (x'_i)^2 = \sum_i x_i^2. \quad (159)$$

Using this condition, one may show that a must satisfy the conditions

$$\sum_i a_{ij} a_{ik} = \delta_{jk} \quad (160)$$

and

$$\det(a) = \pm 1. \quad (161)$$

Orthogonal transformations with $\det(a) = +1$ are simple rotations. The other ones are combinations of a rotation and an inversion¹⁰; these are called *improper rotations*.

It is common to refer to a collection of three objects ψ_i , $i = 1, 2, 3$, which transform, under orthogonal transformations, in the same way as the components of \mathbf{x} , as a *vector*, a *polar vector*, or a *rank-one tensor*. A collection of nine objects q_{ij} , $i, j = 1, 2, 3$, which transform in the same way as the nine objects $x_i x_j$ is called a *rank-two tensor*. And so on. An object which is *invariant*, that is, which is unchanged under an orthogonal transformation, is called a *scalar* or *rank-zero tensor*; the length of a vector is such an object.

One also defines *pseudotensors* of each rank. A *rank-p pseudotensor* comprises a set of 3^p objects which transform in the same way as a rank-p tensor under ordinary

¹⁰An inversion is a transformation $\mathbf{x}' = -\mathbf{x}$.

rotations but which transform with an extra change of sign relative to a rank- p tensor under improper rotations. One also uses the terms *pseudoscalar* for a rank-0 pseudotensor and *pseudovector* or *axial vector* for a rank-1 pseudotensor. Notice that under inversion, for which a is just the negative of the unit 3×3 matrix, a vector changes sign, $\mathbf{x}' = -\mathbf{x}$, while a pseudovector is invariant. This statement can be generalized: Under inversion, a tensor T of rank n transforms to T' with

$$T' = (-1)^n T. \quad (162)$$

A pseudotensor P of the same rank, on the other hand, transforms according to

$$P' = (-1)^{(n+1)} P \quad (163)$$

under inversion.

11.1 Transformation Properties of Physical Quantities

It is important to realize that objects which we are accustomed to referring to as “vectors,” such as \mathbf{B} , are not necessarily vectors in the sense introduced here; indeed, it is one of our tasks in this section to find out just what sorts of tensor are the various physical quantities we have been studying. Consider for example the charge density. Suppose that we have a system with a certain $\rho(\mathbf{x})$ and that we rotate it; then ρ becomes ρ' and \mathbf{x} becomes \mathbf{x}' .

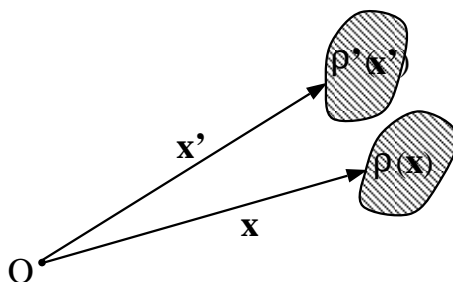


Figure 14: Under a rotation $\rho(\mathbf{x}) = \rho'(\mathbf{x}')$

The question is, how is $\rho'(\mathbf{x}')$ related to $\rho(\mathbf{x})$? It is easy to see, since \mathbf{x}' is what \mathbf{x} becomes as a consequence of the rotation, that $\rho'(\mathbf{x}') = \rho(\mathbf{x})$. Under an inversion

also, this relation is true. Hence we conclude that the charge density is a scalar or rank-0 tensor.

An example of a vector or rank-1 tensor is, of course, \mathbf{x} . Similarly, from this fact one may show that the operator ∇ is a rank-1 tensor (Differential operators can also be tensors or components of tensors);

$$\frac{\partial}{\partial x'_i} = \sum_j a_{ij} \frac{\partial}{\partial x_j}. \quad (164)$$

What then is $\nabla\rho$? From the (known) transformation properties of ρ and of ∇ , it is easy to show that it is a rank-1 tensor. The gradient of any scalar function is a rank-1 tensor. Similarly, one may show that the inner product of two rank-1 tensors, or vectors, is a scalar as is the inner product of two rank-1 pseudotensors; the inner product of a rank-1 tensor and a rank-1 pseudotensor is a pseudoscalar; and the gradient of a pseudoscalar is a rank-1 pseudotensor.

All of the foregoing are quite easy to demonstrate. A little harder is the crossproduct of two vectors (rank-1 tensors). Consider that \mathbf{b} and \mathbf{c} are rank-1 tensors. Their cross product may be written as

$$\mathbf{u} = \mathbf{b} \times \mathbf{c} \quad (165)$$

with a Cartesian component given by

$$u_i = \sum_{j,k} \epsilon_{ijk} b_j c_k \quad (166)$$

where

$$\epsilon_{ijk} \equiv \begin{cases} +1 & \text{if } (i, j, k) = (1, 2, 3), (2, 3, 1), (3, 1, 2) \\ -1 & \text{if } (i, j, k) = (2, 1, 3), (1, 3, 2), (3, 2, 1) \\ 0 & \text{otherwise.} \end{cases} \quad (167)$$

If we **define** ϵ_{ijk} to be given by this equation in all frames, then we can **show** that it is a rank-3 pseudotensor. Alternatively, we can use Eq. (164) to specify it in a single frame, **define** it to be a rank-3 pseudotensor, and then **show** that it is given by Eq. (164) in **any** frame. However one chooses to do it, one can use this object, called

Table 1: Rotation, inversion, and time-reversal properties of some common mechanical quantities.

Function	Rank	Inversion Symmetry	Time-reversal Symmetry
\mathbf{x}	1	– (vector)	+
$\mathbf{v} = d\mathbf{x}/dt$	1	– (vector)	–
$\mathbf{p} = m\mathbf{v}$	1	– (vector)	–
$\mathbf{L} = \mathbf{x} \times m\mathbf{v}$	1	+ (pseudovector)	–
$\mathbf{F} = d\mathbf{p}/dt$	1	– (vector)	+
$\mathbf{N} = \mathbf{x} \times \mathbf{F}$	1	+ (pseudovector)	+
$T = p^2/2m$	0	+ (scalar)	+
V	0	+ (scalar)	+

the *completely antisymmetric unit rank-3 pseudotensor*, and the assumed transformation properties of \mathbf{b} and \mathbf{c} (rank-1 tensors) to determine the transformation properties of the crossproduct. What one finds is that

$$u'_i = \det(a) \sum_j a_{ij} u_j \quad (168)$$

which means that \mathbf{u} is a pseudovector or a rank-1 pseudotensor.

The transformations considered so far have all dealt with space; to them we wish to add the time-reversal transformation. The question to ask of a given entity is how it changes if time is reversed. Imagine making a videotape of the entity's behavior and then running the tape backwards. If, in this viewing, the quantity is the same at a given point on the tape as when the tape is running forward, then the quantity is *even* or *invariant* under time reversal. If its sign has been reversed, then it is *odd* under time reversal. For example, the position $\mathbf{x}(t)$ of an object is even under time reversal; the velocity of the object, however, is odd.

In Table 1, we catalog some familiar mechanical functions according to their rotation, inversion, and time-reversal symmetries.

Table 2: Rotation, inversion and time-reversal properties of some common electromagnetic quantities.

function	rank	inversion symmetry	time-reversal symmetry
ρ	0	+ (scalar)	+
\mathbf{J}	1	- (vector)	-
$\mathbf{E}, \mathbf{D}, \mathbf{P}$	1	- (vector)	+
$\mathbf{B}, \mathbf{H}, \mathbf{M}$	1	+ (pseudovector)	-
\mathbf{S}, \mathbf{g}	1	- (vector)	-
$\bar{\mathbf{T}}$	2	+ (tensor)	+

We may make the same sort of table for various electromagnetic quantities, basing our analysis on the Maxwell equations, which we assume to be the correct equations of electromagnetism. Given that ρ is a scalar and that ∇ is a vector, the equation $\nabla \cdot \mathbf{E} = 4\pi\rho$ tells us that the electric field is a vector; further, it is even under time reversal (since ρ and ∇ are both even). Similarly, \mathbf{D} and \mathbf{P} must be vectors and even under time reversal. Moving on to Faraday's Law, $\nabla \times \mathbf{E} = -c^{-1}\partial\mathbf{B}/\partial t$, from our knowledge of the properties of the gradient, the cross product, and the electric field, we see that \mathbf{B} is a pseudovector and that it is odd under time reversal; \mathbf{H} and \mathbf{M} have the same properties. Finally, Ampère's Law, $\nabla \times \mathbf{B} = (4\pi/c)\mathbf{J} + c^{-1}\partial\mathbf{E}/\partial t$ is consistent with these determinations and with the statement that \mathbf{J} is a vector, odd under time reversal, which follows from the fact that $\mathbf{J} = \rho\mathbf{v}$. Finally, \mathbf{S} and \mathbf{g} are vectors with odd time-reversal symmetry while the Maxwell stress tensor is a rank-2 tensor, even under time reversal. These properties are summarized in Table 2.

The usefulness of these expressions lies in the belief that acceptable equations of physics should be invariant under various symmetry operations. The Maxwell equations and the classical equations of mechanics (Newton's Laws), for example, are invariant under time reversal and under orthogonal transformations, meaning that each term in any given equation transforms in the same way as all of the other terms

in that equation¹¹. If we believe that this should be true of all elementary equations of classical physics, then there are certain implied constraints on the form of the equations. Consider as an example the relation between \mathbf{P} and \mathbf{E} . Supposing that one can make an expansion of a component of \mathbf{P} in terms of the components of \mathbf{E} , we have

$$P_i = \sum_j \alpha_{ij} E_j + \sum_{jk} \beta_{ijk} E_j E_k + \sum_{jkl} \gamma_{ijkl} E_j E_k E_l + \dots \quad (169)$$

where, since \mathbf{P} and \mathbf{E} are both rank-1 tensors, invariant under time reversal, it follows, using the invariance argument, that the coefficients α_{ij} are the components of a rank-2 tensor, invariant under time reversal; the β_{ijk} are components of a rank-3 tensor, invariant under time reversal; and the γ_{ijkl} are components of a rank-4 tensor, also invariant under time reversal.

If we now add some statement about the properties of the medium, we can get further conditions. In the simplest case of an isotropic material, it must be the case that each of these tensors is invariant under orthogonal transformations. This condition severely limits their forms; in particular, it means that $\alpha_{ij} = \alpha \delta_{ij}$. We can see this by appealing to the transformation properties of second rank tensors. Thus, α_{ij} must transform like $x_i x_j$, or

$$\alpha'_{nm} = a_{ni} a_{mj} \alpha_{ij} \quad (170)$$

Since the medium is isotropic, we require that $\alpha'_{ij} = \alpha_{ij}$. The only way to satisfy both of these conditions of transformation is if

$$\alpha'_{nm} = \alpha a_{ni} a_{mi} = \alpha \delta_{nm} \quad (171)$$

The same type of thing cannot be done with β , so that $\beta_{ijk} = 0$, but we can perform similar manipulations on γ so the coefficients γ_{ijkl} are such as to produce

$$\sum_{jkl} \gamma_{ijkl} E_i E_j E_l = \gamma (\mathbf{E} \cdot \mathbf{E}) \mathbf{E} \quad (172)$$

¹¹Of course there are some equations, like Ohm's Law, which describe truly irreversible processes for which time reversal invariance does not hold.

where γ is a constant. Thus, through third-order terms, the expansion of \mathbf{P} in terms of \mathbf{E} must have the form

$$\mathbf{P} = \alpha\mathbf{E} + \gamma E^2\mathbf{E}. \quad (173)$$

The general forms of many other less obvious relations may be determined by similar considerations.

12 Do Maxwell's Equations Allow Magnetic Monopoles?

The answer is yes, but only in a restricted, and trivial, sense. If there were magnetic charges of density ρ_m and an associated magnetic current density \mathbf{J}_m , with a corresponding conservation law

$$\frac{\partial \rho_m}{\partial t} + \nabla \cdot \mathbf{J}_m = 0, \quad (174)$$

then the field equations would read

$$\begin{aligned} \nabla \cdot \mathbf{B} &= 4\pi\rho_m & \nabla \times \mathbf{H} &= \frac{4\pi}{c}\mathbf{J} + \frac{1}{c}\frac{\partial \mathbf{D}}{\partial t} \\ \nabla \cdot \mathbf{D} &= 4\pi\rho & \nabla \times \mathbf{E} &= -\frac{4\pi}{c}\mathbf{J}_m - \frac{1}{c}\frac{\partial \mathbf{B}}{\partial t}. \end{aligned} \quad (175)$$

In fact, the Maxwell equations as we understand them can be put into this form by making a particular kind of transformation, called a *duality transformation* of the fields and sources. Introduce

$$\begin{aligned} \mathbf{E} &= \mathbf{E}' \cos \eta + \mathbf{H}' \sin \eta & \mathbf{D} &= \mathbf{D}' \cos \eta + \mathbf{B}' \sin \eta \\ \mathbf{H} &= -\mathbf{E}' \sin \eta + \mathbf{H}' \cos \eta & \mathbf{B} &= -\mathbf{D}' \sin \eta + \mathbf{B}' \cos \eta \\ \rho &= \rho' \cos \eta + \rho'_m \sin \eta & \mathbf{J} &= \mathbf{J}' \cos \eta + \mathbf{J}'_m \sin \eta \\ \rho_m &= -\rho' \sin \eta + \rho'_m \cos \eta & \mathbf{J}_m &= -\mathbf{J}' \sin \eta + \mathbf{J}'_m \cos \eta. \end{aligned} \quad (176)$$

where η is an arbitrary real constant.

If one now substitutes these into the generalized field equations, one finds, upon separating the coefficients of $\sin \eta$ from those of $\cos \eta$ (These must be independent

because η is arbitrary), that the primed fields and sources obey an identical set of field equations. What this means is that the Maxwell equations (with no magnetic sources) may be thought of as a special case of the generalized field equations, one in which η is chosen so that ρ_m and \mathbf{J}_m are equal to zero. From the form of the transformations for the sources, we see that this is possible if the ratio of ρ to ρ_m for each source (particle) is the same as that for all of the other sources (particles). Hence it is meaningless to say that there are no magnetic monopoles; the real question is whether all elementary particles have the same ratio of electric to magnetic charge. If they do, then Maxwell's equations are correct and correspond, as stated above, to a particular choice of η in the more general field equations.

If one subjects the electron and proton to scrutiny regarding the question of whether they have the same ratio of electric to magnetic charge, one finds that if one defines (by choice of η) the magnetic charge of the electron to be zero, then experimentally the magnetic charge of the proton is known to be smaller than 10^{-24} of its electric charge. That's pretty good evidence for its being zero.

But there remains the question whether there are other kinds of particles, not yet discovered, which have a different ratio ρ/ρ_m than do electrons and protons. Dirac, for example, has given a simple and clever argument which shows that the quantization of the electric charge follows from the mere existence of an electrically uncharged magnetic monopole. Moreover, the argument gives the magnetic charge g of the monopole as $g = nhc/4\pi e$ where n is any integer and h is Planck's constant. This is, in comparison with the electric charge, very large so that it ought to be in principle easy to detect a "Dirac monopole" should there be any of them around. So far, none has been reliably detected.

A Helmholtz' Theorem

Any vector function of position $\mathbf{C}(x)$ can be written as the sum of two vector functions such that the divergence vanishes for one and the curl vanishes for the other. In other words, the decomposition

$$\mathbf{C}(\mathbf{x}) = \mathbf{D}(\mathbf{x}) + \mathbf{F}(\mathbf{x}) \quad (177)$$

is always possible, where

$$\nabla \cdot \mathbf{D} = 0 \quad \nabla \times \mathbf{F} = 0 \quad (178)$$

Proof. We may satisfy the two conditions for \mathbf{F} and \mathbf{D} , by writing

$$\mathbf{D} = \nabla \times \mathbf{A} \quad \mathbf{F} = -\nabla\Phi. \quad (179)$$

Then taking the curl and divergence of these equations respectively, we can write

$$\nabla^2\Phi = -\nabla \cdot \mathbf{C} \quad \nabla \times (\nabla \times \mathbf{A}) = \nabla \times \mathbf{C}. \quad (180)$$

We already know how to solve these solutions (at least in Cartesian coordinates).

$$\Phi(\mathbf{x}') = \frac{1}{4\pi} \int_V d^3x \frac{\nabla \cdot \mathbf{C}(\mathbf{x})}{|\mathbf{x} - \mathbf{x}'|} \quad \mathbf{A}(\mathbf{x}') = \frac{1}{4\pi} \int_V d^3x \frac{\nabla \times \mathbf{C}(\mathbf{x})}{|\mathbf{x} - \mathbf{x}'|} \quad (181)$$

Since \mathbf{D} and \mathbf{F} can now be found from these potentials, we have demonstrated the decomposition claimed by Helmholtz' Theorem, and thus proved it.

An interesting corollary of this theorem is that a vector function is completely determined if its curl and divergence are known everywhere. The field $\mathbf{F} = -\nabla\Phi$, where produced by a point source, is longitudinal to the vector from the source to the point where the field is evaluated. The field $\mathbf{D} = \nabla \times \mathbf{A}$ is transverse to the vector from the source to the field point. Thus \mathbf{F} is typically called the longitudinal, and \mathbf{D} the transverse, part of \mathbf{C} .

Plane Waves and Wave Propagation

Augustin Jean Fresnel

(1788 - 1827)

November 9, 2001

Contents

1	Plane Waves in Uniform Linear Isotropic Nonconducting Media	2
1.1	The Wave Equation	2
1.2	Conditions Imposed by Maxwell's Equations	4
2	Polarization	6
3	Boundary Conditions; Waves at an Interface	9
3.1	Kinematic Conditions	10
3.2	Conditions from Maxwell's Equations	12
3.2.1	Polarization of \mathbf{E}_0 Perpendicular to the Plane	15
3.2.2	Polarization of \mathbf{E}_0 Parallel to the Plane	16
3.3	Parallel Interfaces	17
4	Reflection and Transmission Coefficients	19
5	Examples	21
5.1	Polarization by Reflection	21
5.2	Total Internal Reflection	23

6	Models of Dielectric Functions	26
6.1	Dielectric Response of Free Electrons	30
7	A Model for the Ionosphere	31
8	Waves in a Dissipative Medium	35
8.1	Reflection of a Wave Normally Incident on a Conductor	39
9	Superposition of Waves; Pulses and Packets	41
9.1	A Pulse in the Ionosphere	46
10	Causality and the Dielectric Function	47
11	Arrival of a Signal in a Dispersive Medium	53
A	Waves in a Conductor	57

In this chapter we start by considering plane waves in infinite or semi-infinite media. We shall look at their properties in both insulating and conducting materials and shall give some thought to the possible properties of materials of different kinds. We will also look at the reflection and refraction of waves at planar boundaries between different materials, a topic which forms the basis for much of physical optics. If time allows, we shall also look at some of the more abstract aspects of wave propagation having to do with causality and signal propagation.

1 Plane Waves in Uniform Linear Isotropic Non-conducting Media

1.1 The Wave Equation

One of the most important predictions of the Maxwell equations is the existence of electromagnetic waves which can transport energy. The simplest solutions are plane waves in infinite media, and we shall explore these now.

Consider a material in which

$$\mathbf{B} = \mu\mathbf{H} \quad \mathbf{D} = \epsilon\mathbf{E} \quad \mathbf{J} = \rho = 0. \quad (1)$$

Then the Maxwell equations read

$$\nabla \cdot \mathbf{E} = 0 \quad \nabla \cdot \mathbf{B} = 0 \quad \nabla \times \mathbf{E} = -\frac{1}{c} \frac{\partial \mathbf{B}}{\partial t} \quad \nabla \times \mathbf{B} = \frac{\mu\epsilon}{c} \frac{\partial \mathbf{E}}{\partial t}. \quad (2)$$

Now we do several simple manipulations that will become second nature. First take the curl of one of the curl equations, e.g., Faraday's law, to find

$$\nabla \times (\nabla \times \mathbf{E}) = \nabla(\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E} = -\frac{1}{c} \frac{\partial}{\partial t} (\nabla \times \mathbf{B}) = -\frac{\mu\epsilon}{c^2} \frac{\partial^2 \mathbf{E}}{\partial t^2}, \quad (3)$$

where the generalized Ampère's law was employed in the last step. Because the divergence of \mathbf{E} is zero, this equation may be written as

$$\left(\nabla^2 - \frac{\mu\epsilon}{c^2} \frac{\partial^2}{\partial t^2} \right) \mathbf{E} = 0. \quad (4)$$

Identical manipulations starting from Ampère’s law rather than Faraday’s law also lead to

$$\left(\nabla^2 - \frac{\mu\epsilon}{c^2} \frac{\partial^2}{\partial t^2}\right) \mathbf{B} = 0. \quad (5)$$

Thus any Cartesian component of \mathbf{E} or \mathbf{B} obeys a classical wave equation of the form

$$\left(\nabla^2 - \frac{1}{v^2} \frac{\partial^2}{\partial t^2}\right) \psi(\mathbf{x}, t) = 0, \quad (6)$$

where $v = c/\sqrt{\mu\epsilon}$.

There is a simple set of complex traveling wave solutions to this equation. They are of the form

$$u_{\mathbf{k}}(\mathbf{x}, t) = e^{i(\mathbf{k}\cdot\mathbf{x} - \omega t)} \quad (7)$$

where $\omega = vk$ and \mathbf{k} is any real vector.¹ Notice that the derivatives of this function are

$$\begin{aligned} \nabla u_{\mathbf{k}} &= i\mathbf{k}u_{\mathbf{k}} \\ \nabla^2 u_{\mathbf{k}} &= -k^2 u_{\mathbf{k}} \\ \frac{\partial u_{\mathbf{k}}}{\partial t} &= -i\omega u_{\mathbf{k}} \\ \frac{\partial^2 u_{\mathbf{k}}}{\partial t^2} &= -\omega^2 u_{\mathbf{k}}. \end{aligned} \quad (8)$$

Hence

$$\left(\nabla^2 - \frac{1}{v^2} \frac{\partial^2}{\partial t^2}\right) u_{\mathbf{k}} = \left(-k^2 + \frac{\omega^2}{v^2}\right) u_{\mathbf{k}} = 0, \quad (9)$$

demonstrating that we do indeed have a solution of the wave equation.

This solution is a wave “traveling” in the direction of \mathbf{k} in the sense that a point of constant phase, meaning $\mathbf{k} \cdot \mathbf{x} - \omega t = \text{constant}$, moves along this direction with a speed v which is ω/k . Furthermore, we have a *plane wave*, by which we mean that a surface of constant phase is a plane; in particular, the surfaces of constant phase are just planes perpendicular to \mathbf{k} .

¹This vector is real if ϵ and μ are real; they can be complex, in which case there are still solutions of this form with complex \mathbf{k} .

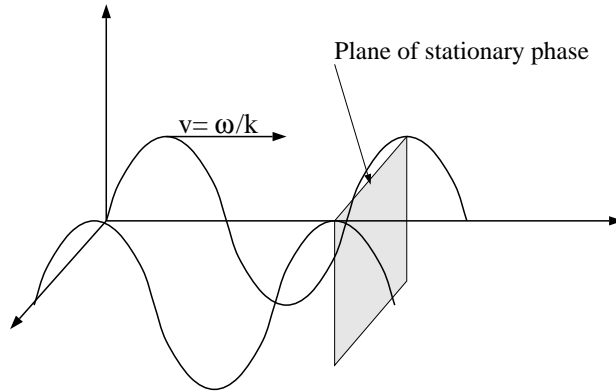


Fig.1: A point of stationary phase moves with velocity $|v| = \omega/k$

1.2 Conditions Imposed by Maxwell's Equations

Next, let us see how the electromagnetic fields can be described in terms of these scalar plane waves. Let us look for an electric field and a magnetic induction with the forms

$$\mathbf{E}(\mathbf{x}, t) = \mathbf{E}_0 e^{i(\mathbf{k} \cdot \mathbf{x} - \omega t)} \quad \mathbf{B}(\mathbf{x}, t) = \mathbf{B}_0 e^{i(\mathbf{k} \cdot \mathbf{x} - \omega t)} \quad (10)$$

with the understanding that the true fields are the real parts of these complex expressions.

In addition to satisfying the wave equation, the complex fields must be solutions of the Maxwell equations. Let us see what additional constraints are thereby imposed. Consider first the divergence equations; these require that

$$0 = \nabla \cdot \mathbf{B}(\mathbf{x}, t) = \nabla \cdot [\mathbf{B}_0 e^{i(\mathbf{k} \cdot \mathbf{x} - \omega t)}] = i\mathbf{k} \cdot \mathbf{B}_0 e^{i(\mathbf{k} \cdot \mathbf{x} - \omega t)} \quad (11)$$

and

$$0 = \nabla \cdot \mathbf{E}(\mathbf{x}, t) = \nabla \cdot [\mathbf{E}_0 e^{i(\mathbf{k} \cdot \mathbf{x} - \omega t)}] = i\mathbf{k} \cdot \mathbf{E}_0 e^{i(\mathbf{k} \cdot \mathbf{x} - \omega t)}. \quad (12)$$

Or

$$\mathbf{k} \cdot \mathbf{B}_0 = 0 \quad \text{and} \quad \mathbf{k} \cdot \mathbf{E}_0 = 0. \quad (13)$$

These conditions mean that \mathbf{B}_0 and \mathbf{E}_0 must be perpendicular to \mathbf{k} , which is to say, parallel to the surfaces of constant phase and perpendicular to the direction in which

the surface of constant phase is moving. Such an electromagnetic wave is called a *transverse wave*. Notice that this nomenclature is consistent with our definition in the last chapter of a transverse vector field as one having zero divergence.

There are further conditions on the amplitudes \mathbf{E}_0 and \mathbf{B}_0 from the other Maxwell equations. From the Ampère law one has

$$\nabla \times \mathbf{B}(\mathbf{x}, t) = \frac{\mu\epsilon}{c} \frac{\partial \mathbf{E}(\mathbf{x}, t)}{\partial t} \quad (14)$$

which leads to

$$i\mathbf{k} \times \mathbf{B}_0 = -\frac{i\omega\epsilon\mu}{c} \mathbf{E}_0 \quad (15)$$

or

$$\mathbf{E}_0 = -\frac{\mathbf{k} \times \mathbf{B}_0}{k\sqrt{\mu\epsilon}} = -\frac{\mathbf{n} \times \mathbf{B}_0}{\sqrt{\epsilon\mu}} \quad (16)$$

where $\mathbf{n} = \mathbf{k}/k$ is a unit vector in the direction of propagation of the wave. From Faraday's Law and similar manipulations one finds the further, and final condition that

$$\mathbf{B}_0 = \sqrt{\mu\epsilon}(\mathbf{n} \times \mathbf{E}_0); \quad (17)$$

however, one may also find this relation from Eq. (16) and the condition that $\mathbf{n} \cdot \mathbf{B}_0 = 0$ and so it is not an additional constraint. Alternatively, one may derive Eq. (16) from Eq. (17) and the condition $\mathbf{n} \cdot \mathbf{E}_0 = 0$. As a consequence, one may, for example, write

$$\mathbf{E}(\mathbf{x}, t) = \mathbf{E}_0 e^{i(\mathbf{k} \cdot \mathbf{x} - \omega t)} \quad (18)$$

where the only condition on \mathbf{E}_0 is $\mathbf{n} \cdot \mathbf{E}_0 = 0$. Then $\mathbf{B}(\mathbf{x}, t)$ follows from Eq. (17) and is

$$\mathbf{B}(\mathbf{x}, t) = \sqrt{\mu\epsilon}(\mathbf{n} \times \mathbf{E}_0) e^{i(\mathbf{k} \cdot \mathbf{x} - \omega t)}. \quad (19)$$

Alternatively, we may start by writing

$$\mathbf{B}(\mathbf{x}, t) = \mathbf{B}_0 e^{i(\mathbf{k} \cdot \mathbf{x} - \omega t)} \quad (20)$$

where \mathbf{B}_0 is orthogonal to \mathbf{k} , $\mathbf{n} \cdot \mathbf{B}_0 = 0$. Then $\mathbf{E}(\mathbf{x}, t)$ is given from Eq. (16) as

$$\mathbf{E}(\mathbf{x}, t) = -\frac{\mathbf{n} \times \mathbf{B}_0}{\sqrt{\epsilon\mu}} e^{i(\mathbf{k} \cdot \mathbf{x} - \omega t)}. \quad (21)$$

From these conditions, and those obtained in the previous paragraph, we may conclude that \mathbf{E} , \mathbf{B} and \mathbf{k} form a mutually orthogonal set.

Before leaving this section, let's look at the time-averaged energy density and Poynting vector in such electromagnetic waves. We shall write them in terms of the amplitude \mathbf{E}_0 . First,

$$\langle \mathbf{S} \rangle = \frac{c}{8\pi} \Re [\mathbf{E}(\mathbf{x}, t) \times \mathbf{H}^*(\mathbf{x}, t)] = \frac{c}{8\pi} \sqrt{\frac{\epsilon}{\mu}} \Re [\mathbf{E}_0 \times (\mathbf{n} \times \mathbf{E}_0^*)] = \frac{c}{8\pi} \sqrt{\frac{\epsilon}{\mu}} |\mathbf{E}_0|^2 \mathbf{n}. \quad (22)$$

Similarly,

$$\langle u \rangle = \frac{1}{16\pi} \Re [\mathbf{E}(\mathbf{x}, t) \cdot \mathbf{D}^*(\mathbf{x}, t) + \mathbf{B}(\mathbf{x}, t) \cdot \mathbf{H}^*(\mathbf{x}, t)] = \frac{\epsilon}{8\pi} |\mathbf{E}_0|^2. \quad (23)$$

The time-averaged momentum density is:

$$\langle \mathbf{g} \rangle = \frac{1}{8\pi c} \Re [\mathbf{E}(\mathbf{x}, t) \times \mathbf{H}^*(\mathbf{x}, t)] = \frac{\sqrt{\epsilon/\mu}}{8\pi c} |E_0|^2 \mathbf{n}. \quad (24)$$

The evaluation of the time-averaged Maxwell stress tensor is left as an exercise.

2 Polarization

In this section we address the question of the most general possible monochromatic plane wave, which amounts to asking what are the possible choices of \mathbf{E}_0 . Let us specify that $\mathbf{k} = k\boldsymbol{\epsilon}_3$ and suppose that we have an orthogonal right-handed set of real unit basis vectors $\boldsymbol{\epsilon}_i$, $i = 1, 2, 3$. Then it must be the case that $\mathbf{E}_0 \cdot \boldsymbol{\epsilon}_3 = 0$ which means that the most general amplitude \mathbf{E}_0 can be expanded as

$$\mathbf{E}_0 = E_{01}\boldsymbol{\epsilon}_1 + E_{02}\boldsymbol{\epsilon}_2. \quad (25)$$

The scalar amplitudes in this expansion can be complex so we have in all four real amplitudes which we may choose with complete abandon. Let us write the complex scalar amplitudes in polar form,

$$E_{01} = E_1 e^{i\phi_1} \qquad E_{02} = E_2 e^{i\phi_2} \quad (26)$$

where E_i and ϕ_i , $i = 1, 2$, are real. Further, introduce

$$E_0 = (E_1^2 + E_2^2)^{1/2} \quad \text{and} \quad \phi = \phi_2 - \phi_1. \quad (27)$$

Then the complex field becomes

$$\mathbf{E}(\mathbf{x}, t) = E_0 \alpha_1 \left(\boldsymbol{\epsilon}_1 + (\alpha_2/\alpha_1) e^{i\phi} \boldsymbol{\epsilon}_2 \right) e^{i(\mathbf{k}\boldsymbol{\epsilon}_3 \cdot \mathbf{x} - \omega t)} e^{i\phi_1} \quad (28)$$

where $\alpha_i = E_i/E_0$ and $\alpha_1^2 + \alpha_2^2 = 1$. In this form, the wave is seen to have just two interesting parameters, α_2/α_1 and $\phi_2 - \phi_1$; these specify the relative phase and amplitude of the two components of the vector amplitude. The other two parameters simply to set the overall magnitude of the field and its absolute phase².

Look at the real part of the complex wave as a function of time at a point in space which is conveniently taken to be the origin. Aside from the overall magnitude and phase, the wave looks like

$$\mathbf{E} \sim \boldsymbol{\epsilon}_1 \cos(\omega t) + (\alpha_2/\alpha_1) \boldsymbol{\epsilon}_2 \cos(\omega t - \phi). \quad (29)$$

If we map out the path traced by the tip of this vector in the space of ϵ_1 and ϵ_2 , we find in general an ellipse. The ellipse is characterized by two parameters, equivalent to α_2/α_1 and ϕ , these being its eccentricity (the ratio of the semi-minor to the semi-major axis) and the amount by which the major axis is rotated relative to some fixed direction such as that of ϵ_1 . Such a wave is said to be *elliptically polarized*, the term “polarization” referring to the behavior of the electric field at a point as a function of time. There are two limiting special cases. One is when the eccentricity is unity in which case the ellipse becomes a circle and the wave is said to be *circularly polarized*; the second is when the eccentricity becomes zero so that the ellipse reduces to a line and the wave is *linearly polarized*.

²These will, of course, be interesting if the wave meets another wave; but they are not interesting if there is no other wave.

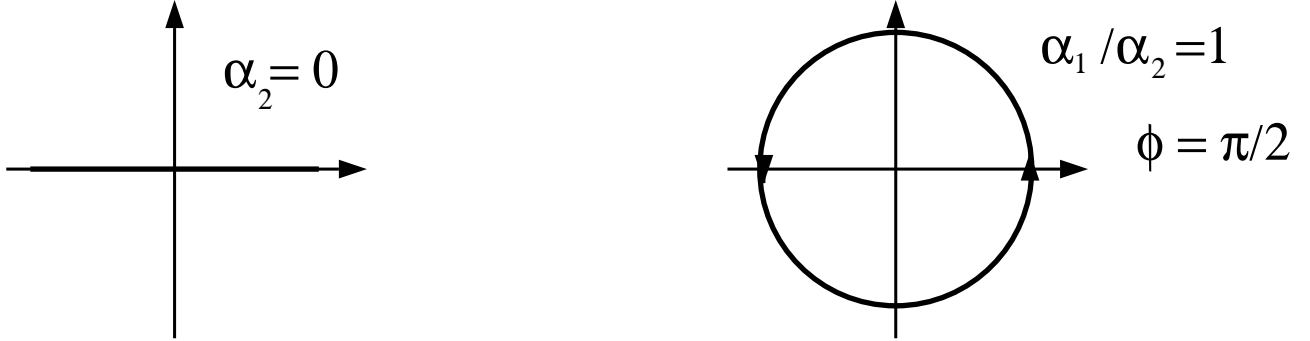


Fig.2: linearly ($\alpha_2 = 0$) and circularly ($\alpha_2/\alpha_1 = 1$ $\phi = \pi/2$) polarized

Often one uses a set of complex basis vectors in which ϵ_1 and ϵ_2 are replaced by vectors ϵ_{\pm} defined by

$$\epsilon_{\pm} \equiv \frac{1}{\sqrt{2}}(\epsilon_1 \pm i\epsilon_2). \quad (30)$$

These have the properties

$$\epsilon_{\pm} \cdot \epsilon_{\pm} = 0 \quad \epsilon_{\pm} \cdot \epsilon_{\mp}^* = 0 \quad \epsilon_{\pm} \cdot \epsilon_{\pm}^* = 1, \quad (31)$$

and it is possible to write the electric field of a general plane wave as

$$\mathbf{E}(\mathbf{x}, t) = (E_+ \epsilon_+ + E_- \epsilon_-) e^{i(\mathbf{k} \cdot \mathbf{x} - \omega t)}, \quad (32)$$

where E_+ and E_- are arbitrary complex constants. If just one of these is non-zero and is written in polar form, then, aside from phase, the complex electric field at a point is

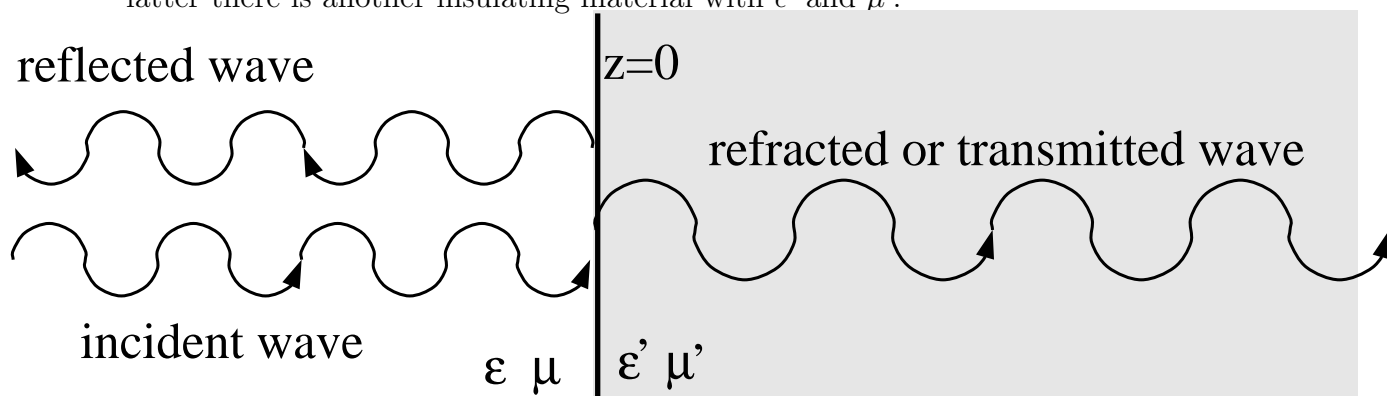
$$\mathbf{E} = |E_{\pm}| \frac{1}{\sqrt{2}} (\epsilon_1 \pm i\epsilon_2) e^{-i\omega t}. \quad (33)$$

The real part then varies as $\epsilon_1 \cos(\omega t) \pm \epsilon_2 \sin(\omega t)$ which is a circularly polarized wave. In the case of the upper sign, one says that the wave is *left-circularly polarized* or that it has *positive helicity*; in the case of the lower sign, it is *right-circularly polarized* or has *negative helicity*. In writing the general wave in terms of these basis vectors, we have expressed it as a superposition of positive and negative helicity waves with amplitudes E_+ and E_- , respectively.

3 Boundary Conditions; Waves at an Interface

In this section, we shall find out what plane waves must look like in semi-infinite media or when there is a planar boundary between two nonconducting materials such as air (or vacuum) and glass. We will need appropriate continuity conditions on the fields at the interface. These may be derived from general kinematic considerations, and from Maxwell equations.

The basic example from which all cases may be inferred is that of a planar interface located at $z = 0$ dividing space into two regions, $z < 0$ and $z > 0$. In the former, we assume an insulating material with dielectric constant ϵ and permeability μ ; in the latter there is another insulating material with ϵ' and μ' .



Now suppose that from the left, or $z < 0$, there is an *incident* wave which has electromagnetic fields

$$\mathbf{E}(\mathbf{x}, t) = \mathbf{E}_0 e^{i(\mathbf{k}\cdot\mathbf{x} - \omega t)}, \quad \mathbf{B}(\mathbf{x}, t) = \sqrt{\mu\epsilon} \frac{\mathbf{k} \times \mathbf{E}(\mathbf{x}, t)}{k}. \quad (34)$$

Also, $k = \omega \sqrt{\mu\epsilon}/c$, and $\mathbf{k} \cdot \hat{\mathbf{z}} > 0$ so that the wave is approaching the interface. Finally, \mathbf{E}_0 is such that $\mathbf{k} \cdot \mathbf{E}_0 = 0$.

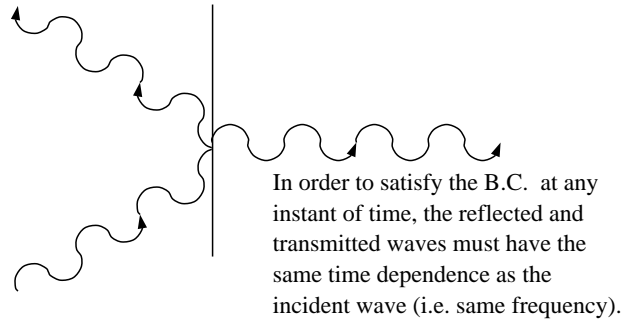
The incident wave is a solution of the Maxwell equations in the region $z < 0$. At the interface, however, it is not a solution; there must be other waves present in order to satisfy the Maxwell equations (or boundary conditions) here. To phrase it another way, when the incident wave hits the interface, additional waves, called *transmitted* (or *refracted*) and *reflected* waves must be generated. The refracted waves are the

ones that propagate into the medium at $z > 0$; the reflected waves are the ones that propagate back into the other medium.

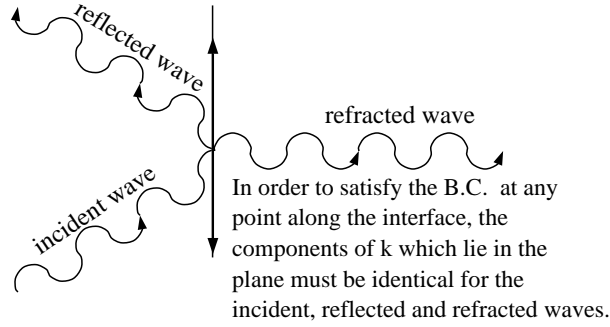
3.1 Kinematic Conditions

We can, from quite general considerations, learn a lot about the properties of the reflected and refracted waves.

First, in order that the continuity conditions remain satisfied at all times, given that they are satisfied at one instant of time, these waves must have the same time dependence as the incident wave. This statement follows from the linear nature of the field equations (each term in the equations is proportional to some component of one of the fields). Hence, all fields vary in time as $e^{-i\omega t}$.



Second, the continuity conditions must be satisfied at all points on the interface or $z = 0$ plane. Suppose that they are satisfied at one particular point, such as $\mathbf{x} = 0$. Then, in order that they remain so for other points on the interface, each wave must vary in the same fashion as each of the other waves as one moves in the plane of the interface. This statement follows, as does the first one, from the linearity of the field equations. Now, since the dependence of a plane wave on position is $\exp(i\mathbf{k} \cdot \mathbf{x})$, this condition means that all waves (incident, reflected, and refracted) must have wave vectors whose components lying in the plane of the interface are identical.



We can express this condition as

$$\mathbf{n} \times \mathbf{k} = \mathbf{n} \times \mathbf{k}' = \mathbf{n} \times \mathbf{k}'' \quad (35)$$

where \mathbf{k}' and \mathbf{k}'' are, respectively, the wave vectors of any refracted and reflected waves. This relation may also be written as

$$k \sin i = k'' \sin r'' = k' \sin r \quad (36)$$

where i , r'' , and r are the angles between the wavevectors of the incident, reflected, and transmitted waves and the normal to the interface. They are called the *angle of incidence*, the *angle of reflection*, and the *angle of refraction*.

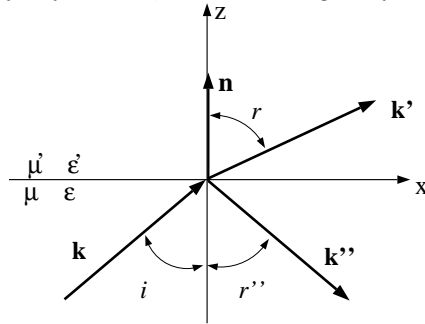


Figure 6: Definition of the angles i , r'' , and r

Finally, any reflected wave is a solution of the same wave equation as the incident wave; consequently, it has a wave number $k'' = k$. Any transmitted wave, however, has wave number $k' = \omega \sqrt{\mu' \epsilon'} / c$, so $k' \neq k$. If we combine these statements with Eq. (36), we can see that $r'' = i$, *the angle of incidence equals the angle of reflection*.

For a transmitted wave, however, the wave equation is such that $k' = \omega\sqrt{\mu'\epsilon'}/c$ which is not k ; in fact, $k'/n' = k/n$ where

$$n \equiv \sqrt{\mu\epsilon} \quad \text{and} \quad n' \equiv \sqrt{\mu'\epsilon'} \quad (37)$$

are the *indices of refraction* in the two materials. Using these definitions in Eq. (36) we find

$$n \sin i = n' \sin r \quad (38)$$

which is known in optics as *Snell's Law*.

3.2 Conditions from Maxwell's Equations

Notice that we derived Snell's law and the statement $i = r''$ without using explicitly the continuity conditions; we had only to use the fact that there **are** linear continuity conditions. Hence these properties are called kinematic properties (they don't depend on the particular dynamics of the fields which are given by the Maxwell equations) and they are applicable to much more than just electromagnetic phenomena.

To fully develop the rules of reflection and refraction for electromagnetic waves, we must use the Maxwell equations to tell us the specific relations among the fields and then must apply the continuity conditions at a specific point on the interface, such as $\mathbf{x} = 0$, and at a specific time, such as $t = 0^3$.

$$\nabla \cdot \mathbf{D} = 4\pi\rho, \quad \nabla \cdot \mathbf{B} = 0, \quad \nabla \times \mathbf{H} = \frac{4\pi}{c}\mathbf{J} + \frac{1}{c}\frac{\partial\mathbf{D}}{\partial t}, \quad \nabla \times \mathbf{E} = -\frac{1}{c}\frac{\partial\mathbf{B}}{\partial t}. \quad (39)$$

³For other points and times we know that the conditions will be satisfied by making sure the kinematic conditions derived above are satisfied.

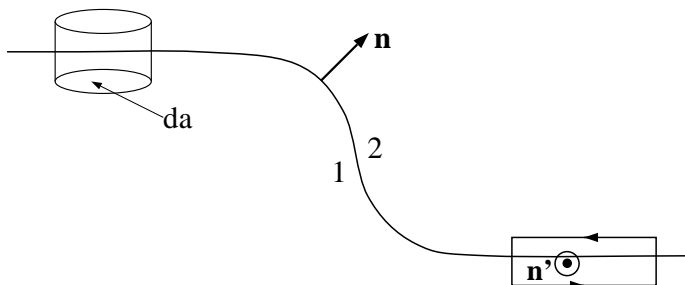


Figure 3: Integration surfaces used for the B.C.

Application of the divergence theorem to the two divergence equations using the familiar pillbox construction leads, as for the static case, to the continuity conditions

$$(\mathbf{D}_1 - \mathbf{D}_2) \cdot \mathbf{n} = 4\pi\sigma \quad (40)$$

$$(\mathbf{B}_2 - \mathbf{B}_1) \cdot \mathbf{n} = 0 \quad (41)$$

where \mathbf{n} is a unit outward normal from material 1 and σ is the macroscopic surface-charge density. Application of Stokes' theorem to the curl equations in the "usual" way leads to

$$\mathbf{n} \times (\mathbf{E}_2 - \mathbf{E}_1) = 0 \quad (42)$$

$$\mathbf{n} \times (\mathbf{H}_2 - \mathbf{H}_1) = \frac{4\pi}{c} \mathbf{K} \quad (43)$$

where \mathbf{K} is the macroscopic surface-current density lying inside of the loop C to which Stokes' theorem is applied. Notice that the time derivatives in Faraday's law and Ampère's law do not contribute to the continuity conditions.⁴

For uncharged insulators, the surface sources σ and \mathbf{K} are always zero; then the continuity conditions are especially simple and state that the normal components of \mathbf{D} and \mathbf{B} are continuous as are the tangential components of \mathbf{H} and \mathbf{E} .

At $\mathbf{x} = 0$, $t = 0$, the fields of an incident wave, a single transmitted wave, and a single reflected wave⁵ may be written as follows:

⁴We have assumed that there are no singular parts of the time derivatives localized at the interface; were there any such contributions, they would show up in the continuity conditions.

⁵We don't know at this point that we need only one reflected and one transmitted wave to obtain a solution to the boundary value problem. By construction, we will see that such is the case.

Incident wave:

$$\mathbf{E} = \mathbf{E}_0 \quad \mathbf{B} = \frac{n}{k}(\mathbf{k} \times \mathbf{E}_0) \quad (44)$$

Reflected wave:

$$\mathbf{E} = \mathbf{E}_0'' \quad \mathbf{B} = \frac{n}{k}(\mathbf{k}'' \times \mathbf{E}_0'') \quad (45)$$

Transmitted wave:

$$\mathbf{E} = \mathbf{E}_0' \quad \mathbf{B} = \frac{n'}{k'}(\mathbf{k}' \times \mathbf{E}_0') = \frac{n}{k}(\mathbf{k}' \times \mathbf{E}_0'). \quad (46)$$

We suppose that we are given n , n' , \mathbf{k} , and \mathbf{E}_0 ; we need to find \mathbf{k}' , \mathbf{k}'' , \mathbf{E}_0' , and \mathbf{E}_0'' . The wave vectors follow from the kinematic relations; they all lie in the plane containing the normal to the interface and the incident wave vector, called the *plane of incidence* and make angles with the normal as discussed above. As for the amplitudes, they are found from the continuity conditions:

1. D_n continuous:

$$\epsilon(\mathbf{E}_0 + \mathbf{E}_0'') \cdot \mathbf{n} = \epsilon' \mathbf{E}_0' \cdot \mathbf{n} \quad (47)$$

2. B_n continuous:

$$(\mathbf{k} \times \mathbf{E}_0 + \mathbf{k}'' \times \mathbf{E}_0'') \cdot \mathbf{n} = (\mathbf{k}' \times \mathbf{E}_0') \cdot \mathbf{n} \quad (48)$$

3. \mathbf{E}_t continuous:

$$(\mathbf{E}_0 + \mathbf{E}_0'') \times \mathbf{n} = \mathbf{E}_0' \times \mathbf{n} \quad (49)$$

4. \mathbf{H}_t continuous:

$$\frac{1}{\mu}(\mathbf{k} \times \mathbf{E}_0 + \mathbf{k}'' \times \mathbf{E}_0'') \times \mathbf{n} = \frac{1}{\mu'}(\mathbf{k}' \times \mathbf{E}_0') \times \mathbf{n}. \quad (50)$$

It is a messy bit of algebra to solve these equations in the general case. The task can be made simpler by writing the incident wave's electric field as a linear combination of two linearly polarized waves, which is always possible. One solves each of these cases separately. The appropriate sum of the two solutions is then the solution of the original problem. Once again, the linearity of the field equations leads to enormous simplification of the algebra. The two cases that we are going to treat are

1. polarization of \mathbf{E}_0 perpendicular to the plane of incidence and
2. polarization parallel to the plane of incidence.

3.2.1 Polarization of \mathbf{E}_0 Perpendicular to the Plane

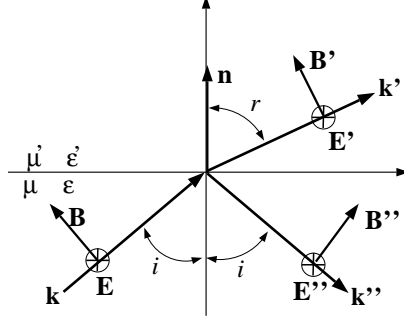


Figure 7: Polarization of \mathbf{E}_0 perpendicular to the plane of incidence

The figure sets the conventions for the first case. They are such that $\mathbf{E}_0 = E_0 \hat{\mathbf{y}}$, $\mathbf{E}'_0 = E'_0 \hat{\mathbf{y}}$, and $\mathbf{E}''_0 = E''_0 \hat{\mathbf{y}}$. Remember also that $k'' = k$, $k/n = k'/n'$, and $n \sin i = n' \sin r$. Now apply the four continuity conditions. The first gives nothing because there is no normal component of the electric displacement or electric field; the second gives $E_0 + E''_0 = E'_0$; the third gives the same constraint as the second; and the fourth results in $(k/\mu) \cos i (E_0 - E''_0) = (k'/\mu') \cos r E'_0$. Since $k' = kn'/n$ and $n = \sqrt{\mu\epsilon}$, we can write the latter as $\sqrt{\epsilon/\mu} \cos i (E_0 - E''_0) = \sqrt{\epsilon'/\mu'} \cos r E'_0$. In addition, $\cos r = \sqrt{1 - \sin^2 r} = \sqrt{1 - (n/n')^2 \sin^2 i}$. Combining these relations we find the two conditions

$$E'_0 - E''_0 = E_0 \quad \sqrt{\frac{\epsilon'}{\mu'}} \sqrt{1 - \left(\frac{n}{n'}\right)^2 \sin^2 i} E'_0 + \sqrt{\frac{\epsilon}{\mu}} \cos i E''_0 = \sqrt{\frac{\epsilon}{\mu}} \cos i E_0. \quad (51)$$

Notice that these are written entirely in terms of the angle of incidence; the angle of refraction does not appear. Their solution is easily shown to be

$$E'_0 = \frac{2n \cos i}{n \cos i + (\mu/\mu') \sqrt{n'^2 - n^2 \sin^2 i}} E_0$$

and

$$E_0'' = \frac{n \cos i - (\mu/\mu')\sqrt{n'^2 - n^2 \sin^2 i}}{n \cos i + (\mu/\mu')\sqrt{n'^2 - n^2 \sin^2 i}} E_0 \quad (52)$$

3.2.2 Polarization of \mathbf{E}_0 Parallel to the Plane

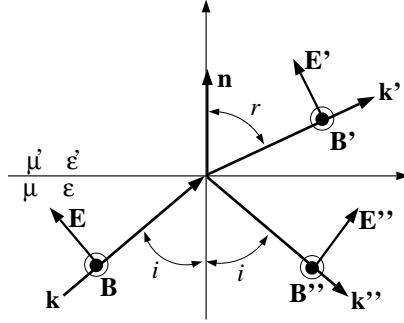


Figure 8: Polarization of \mathbf{E}_0 parallel to the plane of incidence

The second case, polarization in the plane of incidence may be similarly analyzed. The figure shows the conventions for this case. They are such that $\mathbf{E}_0 = E_0(\sin i \hat{\mathbf{z}} - \cos i \hat{\mathbf{x}})$, $\mathbf{E}'_0 = E'_0(\sin r \hat{\mathbf{z}} - \cos r \hat{\mathbf{x}})$, and $\mathbf{E}''_0 = E''_0(\sin i \hat{\mathbf{z}} + \cos i \hat{\mathbf{x}})$. The first boundary condition implies that $\epsilon \sin i (E_0 + E''_0) = \epsilon' \sin r E'_0$; the second gives nothing; the third gives $\cos i (-E_0 + E''_0) = -\cos r E'_0$; and the fourth gives a condition that is redundant with the first when Snell's law is invoked. Thus we may write the two conditions, after removing all occurrences of r as in the first case, as

$$\sqrt{\frac{\epsilon}{\mu}}(E_0 + E''_0) = \sqrt{\frac{\epsilon'}{\mu'}} E'_0 \quad \cos i (E_0 + E''_0) = \sqrt{1 - (n/n')^2 \sin^2 i} E'_0. \quad (53)$$

Their solution is

$$E'_0 = \frac{2nn' \cos i}{(\mu/\mu')n'^2 \cos i + n\sqrt{n'^2 - n^2 \sin^2 i}} E_0$$

and

$$E_0'' = \frac{(\mu/\mu')n'^2 \cos i - n\sqrt{n'^2 - n^2 \sin^2 i}}{(\mu/\mu')n'^2 \cos i + n\sqrt{n'^2 - n^2 \sin^2 i}} E_0 \quad (54)$$

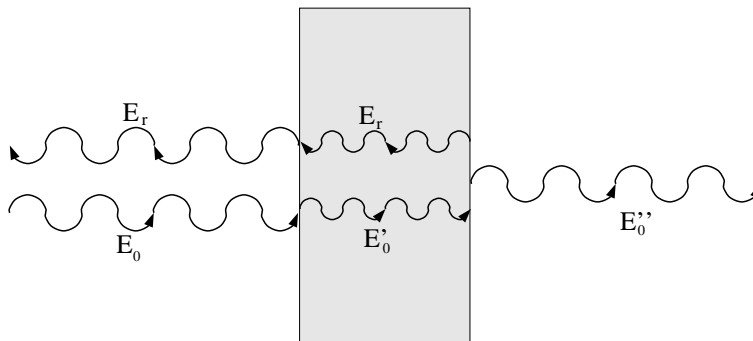
Our solutions to the reflection-refraction problem have the following characteristics by design. First, as mentioned above, they involve only the angle of incidence,

the angle of refraction having been removed wherever it appeared by using Snell's law; second, the material properties enter through the permeabilities and indices of refraction as opposed to the permeabilities and dielectric constants. The reason is that for most of the materials one encounters, $\mu = \mu' = 1$ and so the permeabilities drop out of the relations. Second, one is generally more likely to be given an index of refraction than a dielectric constant and so expressing the amplitudes in terms of n makes them more readily applicable.⁶

Equations (52) and (54) are known as *Fresnel's equations*; with them we can calculate the reflection and transmission of a plane wave at a planar interface for arbitrary initial polarization. Such an incident wave gives rise to a single reflected plane wave and a single transmitted plane wave, meaning that there is just one reflected wave vector \mathbf{k}'' and one transmitted wave vector \mathbf{k}' .

3.3 Parallel Interfaces

With a little thought we may see how to generalize to the case of two (or more) parallel interfaces. Consider the figure showing two parallel interfaces separating three materials. If we follow the consequences of an incident plane wave from the first material on one side we can see that the reflection processes within the middle material of the “sandwich” generate many plane waves in here, but that these waves have just two distinct wave vectors.



⁶Of course, the relation between n and ϵ is sufficiently simple that there is really no great difference.

Figure 9: Plane wave incident on a sandwich.

Also, all waves transmitted into the third material have the same wave vector, and the “reflected” waves in the first medium all have a single wave vector. Hence one finds that in the first medium, there are just two waves with electric fields

$$\mathbf{E} = \mathbf{E}_0 e^{i(\mathbf{k} \cdot \mathbf{x} - \omega t)} \quad \mathbf{E}_r = \mathbf{E}_{r0} e^{i(\mathbf{k}_r \cdot \mathbf{x} - \omega t)}; \quad (55)$$

in the middle medium there are again just two distinct waves with fields

$$\mathbf{E}' = \mathbf{E}'_0 e^{i(\mathbf{k}' \cdot \mathbf{x} - \omega t)} \quad \mathbf{E}'_r = \mathbf{E}'_{r0} e^{i(\mathbf{k}'_r \cdot \mathbf{x} - \omega t)} \quad (56)$$

and in the third medium there is just one plane wave with field

$$\mathbf{E}'' = \mathbf{E}''_0 e^{i(\mathbf{k}'' \cdot \mathbf{x} - \omega t)}. \quad (57)$$

To find the four amplitudes \mathbf{E}_{r0} , \mathbf{E}'_0 , \mathbf{E}'_{r0} , and \mathbf{E}''_0 , one must apply the boundary conditions at the two interfaces, leading to four distinct linear relations involving these amplitudes and that of the incident wave, \mathbf{E}_0 . Solving these equations, one finds the amplitudes of all waves in terms of that of the incident wave.

Returning briefly to Fresnel’s equations for reflection and refraction at a single interface, let us look at the special case of normal incidence, $i = 0$. then $r = 0$ also, and the first set (polarization normal to the plane of incidence) of Fresnel equations tells us that ⁷

$$E'_0 = \frac{2n}{n + (\mu/\mu')n'} E_0 \quad E''_0 = \frac{n - (\mu/\mu')n'}{n + (\mu/\mu')n'} E_0. \quad (58)$$

These are simple results, especially when $\mu = \mu'$. They clearly tell us that when the two materials have comparable indices of refraction and permeabilities, the wave is mostly transmitted and when they have very different properties (an engineer would

⁷Actually, both sets of Fresnel equations are applicable for normal incidence. The second set, however, will produce a result with some signs switched as a consequence of the different conventions used for the directions of the electric fields in the two cases.

call that impedance mismatch), reflection is the rule. Notice also that if $n'\mu/\mu' > n$, the reflected amplitude is opposite in sign to the incident one, meaning that the electric field of the reflected wave is phase shifted by π radians relative to that of the incident one under these circumstances.

4 Reflection and Transmission Coefficients

In this section we look at the power or energy transmitted and reflected at an interface between two insulators. To do so, we must evaluate the time-averaged power in the incident, reflected, and transmitted waves which is done by calculating the Poynting vector. The energy current density toward or away from the interface is then given by the component of the Poynting vector in the direction normal to the interface. In the second medium, where there is just a single (refracted) wave, the normal component of \mathbf{S} is unambiguously the transmitted power per unit area. But in the first medium, the total electromagnetic field is the sum of the fields of the incident and reflected waves. In evaluating $\mathbf{E} \times \mathbf{H}$, one finds three kinds of terms. There is one which is the cross-product of the fields in the incident wave, and its normal component gives the incident power per unit area. A second is the cross-product of the fields in the reflected wave, giving the reflected power. But there are also two cross-terms involving the electric field of one of the plane waves and the magnetic field of the other one. It turns out that the time-average of the normal component of these terms is zero, so that they may be ignored in the present context. Bearing this in mind, we have the following quantities of interest:

The time-averaged incident power per unit area:

$$\mathcal{P} = \langle \mathbf{S} \rangle \cdot \mathbf{n} = \frac{c}{8\pi} \sqrt{\frac{\epsilon}{\mu}} |\mathbf{E}_0|^2 \frac{\mathbf{k} \cdot \mathbf{n}}{k} \quad (59)$$

The time-averaged reflected power per unit area:

$$\mathcal{P}'' = - \langle \mathbf{S}'' \rangle \cdot \mathbf{n} = \frac{c}{8\pi} \sqrt{\frac{\epsilon}{\mu}} |\mathbf{E}_0''|^2 \frac{\mathbf{k}'' \cdot \mathbf{n}}{k} \quad (60)$$

The time-averaged transmitted power per unit area:

$$\mathcal{P}' = \langle \mathbf{S}' \rangle \cdot \mathbf{n} = \frac{c}{8\pi} \sqrt{\frac{\epsilon'}{\mu'}} |\mathbf{E}'_0|^2 \frac{\mathbf{k}' \cdot \mathbf{n}}{k'} d \quad (61)$$

The *reflection coefficient* R and the *transmission coefficient* T are defined as the ratios of the reflected and transmitted power to the incident power.

We may calculate the reflection and transmission coefficients for the cases of polarization perpendicular and parallel to the plane of incidence by using the Fresnel equations. If an incident wave has general polarization so that its fields are linear combinations of these two special cases, then there is once again the possibility of cross terms in the power involving an electric field with one type of polarization and a magnetic field with the other type. Fortunately, these turn out to vanish, so that one may treat the two polarizations individually.

For the case of **polarization perpendicular to the plane of incidence**, we use the Fresnel equations (52) and (54) for the reflected and transmitted amplitudes and have

$$T = \frac{\sqrt{\frac{\epsilon'}{\mu'}} \frac{4n^2 \cos^2 i \cos r}{(n \cos i + (\mu/\mu') \sqrt{n'^2 - n^2 \sin^2 i})^2}}{\sqrt{\frac{\epsilon}{\mu}} \cos i} \quad (62)$$

Making use of the relations $n = \sqrt{\epsilon\mu}$, $n' = \sqrt{\epsilon'\mu'}$, $\sin r = (n/n') \sin i$, and $\cos i = \sqrt{1 - \sin^2 i}$, one finds that

$$T = \frac{4n(\mu/\mu') \cos i \sqrt{n'^2 - n^2 \sin^2 i}}{[n \cos i + (\mu/\mu') \sqrt{n'^2 - n^2 \sin^2 i}]^2} \quad (63)$$

By similar means one can write the reflection coefficient as

$$R = \frac{[n \cos i - (\mu/\mu') \sqrt{n'^2 - n^2 \sin^2 i}]^2}{[n \cos i + (\mu/\mu') \sqrt{n'^2 - n^2 \sin^2 i}]^2} \quad (64)$$

By inspection one can see that $R + T = 1$ which expresses the conservation of energy; what is not transmitted is reflected.

The case of **polarization in the plane of incidence** is treated similarly. One finds

$$T = \frac{4nn'^2(\mu/\mu') \cos i \sqrt{n'^2 - n^2 \sin^2 i}}{[(\mu/\mu')n'^2 \cos i + n \sqrt{n'^2 - n^2 \sin^2 i}]^2} \quad (65)$$

and

$$R = \frac{[(\mu/\mu')n'^2 \cos i - n\sqrt{n'^2 - n^2 \sin^2 i}]^2}{[(\mu/\mu')n'^2 \cos i + n\sqrt{n'^2 - n^2 \sin^2 i}]^2}. \quad (66)$$

Once again, $R + T = 1$.

5 Examples

5.1 Polarization by Reflection

From inspection of Fresnel's equations, we can see that the relative amounts of transmitted and reflected amplitude depend on the state of polarization and are distinctly not the same for both polarizations.

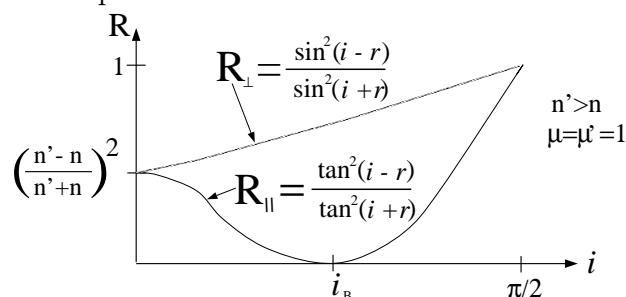


Figure 10: Reflection Coefficient when $n' > n$, and $\mu' = \mu = 1$

That means that in the general case, the polarizations of the transmitted and reflected waves will not be the same as that of the incident one. A very special case has to do with the reflected wave given incident polarization in the plane of incidence. We see that the reflected amplitude will vanish if⁸

$$n'^2 \cos i = n\sqrt{n'^2 - n^2 \sin^2 i}. \quad (67)$$

Squaring this relation we find

$$n'^4 \cos^2 i = n^2 n'^2 - n^4 \sin^2 i = n'^4 (1 - \sin^2 i) \text{ or } \sin^2 i = \frac{n'^2}{n'^2 + n^2} \text{ or } \tan i = \frac{n'}{n}. \quad (68)$$

⁸We let $\mu = \mu'$ in this section unless explicitly stated otherwise; keeping the permeability around usually contributes nothing but extra work and obfuscation.

This special angle of incidence is called the *Brewster angle*,

$$i_B = \arctan(n'/n); \tag{69}$$

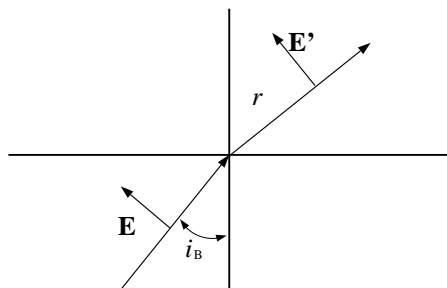


Figure 11: No reflected wave when $i = i_B$ and the field is polarized in the plane.

a wave polarized in the plane of incidence and incident on the interface at the Brewster angle is completely transmitted with no reflected wave. If a wave of general polarization is incident at the Brewster angle, then the reflected wave is completely (linearly) polarized perpendicular to the plane of incidence. Hence this phenomenon provides a method for obtaining a linearly polarized wave from an unpolarized one. More generally, if the angle of incidence is reasonably close to the Brewster angle, the reflected light is to a large degree polarized perpendicular to the plane of incidence. This fact is utilized by polarizing sun glasses which screen out most of the light polarized parallel to the surface of the earth, which is to say, most of the light reflected by the earth.

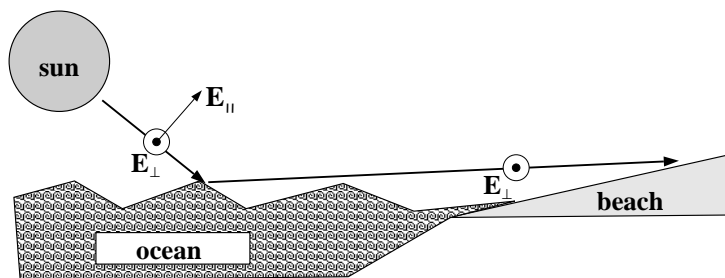


Figure 12: Light reflected from the ocean (glare) is largely polarized along the horizon, and may be removed with polarized sunglasses.

5.2 Total Internal Reflection

As a second example we look at the phenomenon of total internal reflection which is the opposite of the one just considered in that no energy is transmitted across an interface under appropriate conditions. Suppose that $n > n'$. As shown in the figure, this means that $r > i$.

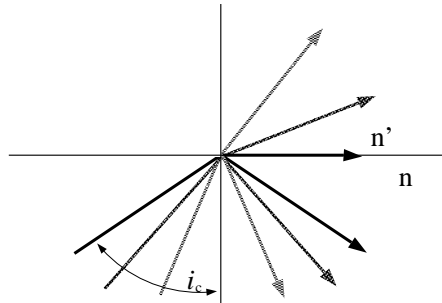


Figure 13: A series of angles when $n > n'$.

Now consider an incident wave with i large enough that $n \sin i > n'$. How can we have a refracted wave with r such that Snell's law, $n \sin i = n' \sin r$ is satisfied? Recall our argument for Snell's law; it was based on the fact that the wave vector \mathbf{k}' of the refracted wave had to have a component k'_t parallel to the interface equal to the same component of the incident wave. Given that $n \sin i > n'$, this condition means that k'_t is **larger** than $\omega n'/c$ which is supposed to be the magnitude of \mathbf{k}' , according to the wave equation. But there is a way around this. The condition that comes from the wave equation is that, if k'_t and k'_n are respectively the components of \mathbf{k}' tangential and normal to the interface, then $k'^2_t + k'^2_n = \omega^2 n'^2/c^2$. If $k'_t > \omega n'/c$, we can satisfy this condition by having k'_n be imaginary. In particular,

$$k'_n = \pm i \frac{n\omega}{c} \sqrt{\sin^2 i - (n'/n)^2}. \quad (70)$$

The choice of sign has to be such as to produce a wave that damps away to nothing in the second medium; otherwise it becomes exceedingly large (which is unphysical behavior) as one moves far away from the interface. Now that we have figured out what is \mathbf{k}' ; that is, $k'_t = (n\omega/c) \sin i$ and k'_n is given by Eq. (70), we can see the

character of the transmitted electric field. It is

$$\mathbf{E}' \sim e^{ik'_t x} e^{-|k'_n|z} e^{-i\omega t} \quad (71)$$

where $\hat{\mathbf{x}}$ is the direction of the tangential component of \mathbf{k} .

The Poynting vector for a wave of this sort has no component directed normal to the interface although there is one parallel to the interface. To see this, take \mathbf{E} to be in the y -direction.

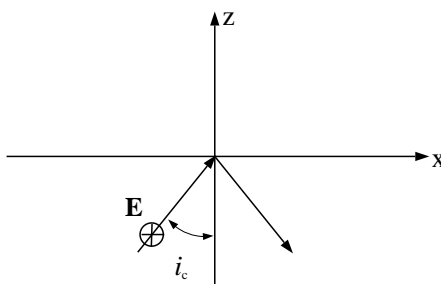


Figure 14: Polarization \perp to the plane of incidence.

Then

$$\mathbf{E}' = \mathbf{E}'_0 e^{i(k'_t x - \omega t)} e^{-|k'_n|z}$$

so that

$$S'_z = \frac{c}{8\pi} \Re(\mathbf{E}' \times \mathbf{B}'^*)_z = \frac{c}{8\pi} \Re(E'_y \times B'_x{}^*) .$$

We may use Faraday's law to relate \mathbf{E} to \mathbf{B}

$$\nabla \times \mathbf{E} = -\frac{1}{c} \frac{\partial \mathbf{B}}{\partial t} \rightarrow \frac{i\omega}{c} B'_x = |k'_n| E'_y .$$

Thus,

$$S'_z = \frac{c}{8\pi} \Re \left\{ -E'_y \frac{c|k'_n|}{-i\omega} E'^*_y \right\} = 0$$

Thus, as shown in the figure below, when $i > i_c$, the power is totally reflected.

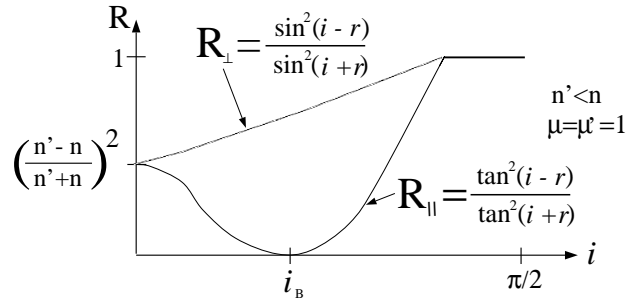


Figure 15: Reflection Coefficient when $n > n'$, and $\mu' = \mu = 1$

What we have is therefore a *surface wave* confined to the region close to the interface and transporting energy parallel to it. Moreover, by evaluating the Poynting vector of the reflected and incident waves, one finds that as much energy is reflected from the interface as is incident upon it. Hence we have the phenomenon of perfect or total reflection of the incident wave. This phenomenon is utilized in fiber optics; an electromagnetic wave is propagated inside of a thin tube of some material having a large index of refraction and surrounded by another material having a much smaller index. Wherever the wave is incident upon the wall of the tube, it is completely reflected.

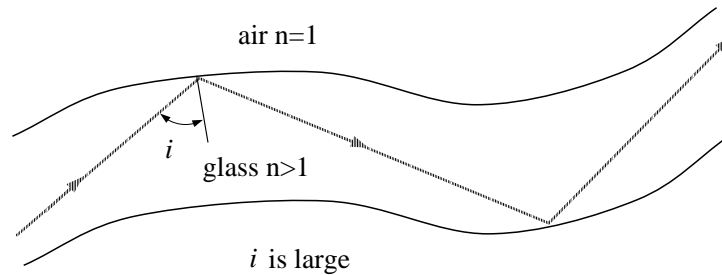


Figure 16: Total internal reflection occurs within a fiber optic tube.

There is some natural attenuation of the wave because of imperfect dielectric properties of the material itself or its coating; nevertheless, a beam of light, for example, can be transmitted long distances and around many curves (as long as they aren't too sharp) in such a "pipe."

6 Models of Dielectric Functions

The dielectric “constant” of almost any material is in fact a function of frequency, meaning that it has different values for waves of different frequencies.

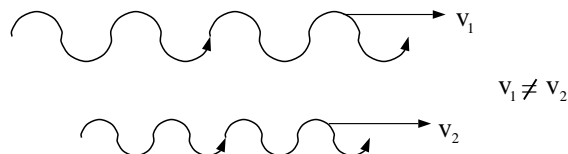


Figure 17: In a dispersive medium waves of different frequencies have different phase velocities $v = c/\sqrt{\epsilon(\omega)\mu}$.

We can make a simple model of the dielectric “function” of an insulating material as follows: Suppose that the charges which primarily respond to an electric field are electrons bound on atoms or molecules. Let one such electron be harmonically bound, meaning that the binding forces are treated as linear in the displacement of the charge from its equilibrium position. Also, let there be a damping force proportional to the velocity \mathbf{v} of the electron. Then, if the mass and charge of the electron are m and $-e$, the equation of motion of the electron under the influence of an electric field $\mathbf{E}(\mathbf{x}, t)$ is

$$m \left(\frac{d^2 \mathbf{x}}{dt^2} + \gamma \frac{d\mathbf{x}}{dt} + \omega_0^2 \mathbf{x} \right) = -e\mathbf{E}(\mathbf{x}, t). \quad (72)$$

The harmonic restoring force is expressed through a “natural” frequency of oscillation ω_0 of the electron. We have ignored the possible influence of a magnetic induction $\mathbf{B}(\mathbf{x}, t)$ on the electron’s motion. Typically this force is much smaller than the electric field force because the electron’s speed is much smaller than c ; there can be exceptions, however, and one of them is explored below.

Next, the typical magnitude of the electron’s displacement $|\mathbf{x}|$ is on the order of an atomic size.

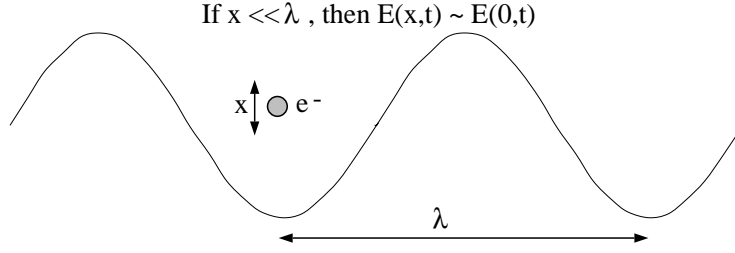


Figure 18: If the wavelength of the incident wave is much larger than the electronic displacement, then we may neglect the spacial dependence of $\mathbf{E}(\mathbf{x}, t)$.

If the electric field $\mathbf{E}(\mathbf{x}, t)$ is that of visible or even ultraviolet light, then the displacement is much smaller than distances over which $\mathbf{E}(\mathbf{x}, t)$ varies significantly, meaning that we can approximate $\mathbf{E}(\mathbf{x}, t) \approx \mathbf{E}(0, t) = \mathbf{E}_0 \exp(-i\omega t)$. In this limit, the solution we seek is of the form $\mathbf{x}(t) = \mathbf{x}_0 \exp(-i\omega t)$. Substituting into the equation of motion, we find that the equation for the amplitude of the motion is

$$m(-\omega^2 - i\omega\gamma + \omega_0^2)\mathbf{x}_0 = -e\mathbf{E}_0 \quad (73)$$

or

$$\mathbf{x}_0 = \frac{-e\mathbf{E}_0}{m(\omega_0^2 - i\omega\gamma - \omega^2)}. \quad (74)$$

The amplitude of the dipole moment associated with the motion of this electron is $\mathbf{p}_0 = -ex_0$. To find the polarization, we need to compute the dipole moments of all electrons in some finite volume of material. These electrons will not all have the same damping or natural frequencies, so let us say that there are n molecules per unit volume with z electrons each. If f_i of the electrons on each molecule have resonant frequency ω_i and damping constant γ_i , then we get a polarization or dipole moment per unit volume which varies harmonically with an amplitude

$$\mathbf{P}_0 = e^2\mathbf{E}_0n \sum_j \left(\frac{f_j}{m(\omega_j^2 - i\omega\gamma_j - \omega^2)} \right); \quad (75)$$

this is also the relation between $\mathbf{E}(\mathbf{x}, t)$ and $\mathbf{P}(\mathbf{x}, t)$. If we further say that $\mathbf{E}(\mathbf{x}, t)$

is the macroscopic field⁹, then we can write $\mathbf{D} = \mathbf{E} + 4\pi\mathbf{P} = \epsilon\mathbf{E}$ with the preceding expression for the polarization. The result is an expression for $\epsilon(\omega)$:

$$\epsilon(\omega) = 1 + \frac{4\pi ne^2}{m} \sum_j \left(\frac{f_j}{\omega_j^2 - i\omega\gamma_j - \omega^2} \right) \quad (76)$$

or

$$\epsilon(\omega) = 1 + \frac{4\pi nze^2}{m} \sum_j \left(\frac{f_j}{z} \right) \left(\frac{\omega_j^2 - \omega^2 + i\omega\gamma_j}{(\omega_j^2 - \omega^2)^2 + \omega^2\gamma_j^2} \right) \equiv \epsilon_1 + i\epsilon_2 \quad (77)$$

where ϵ_1 and ϵ_2 are real.

In a typical term of the sum, different regimes of the relative sizes of ω , ω_j , and γ_j give rise to very different behaviors. The resonant frequencies are, when Planck's constant is thrown in, comparable to binding energies of electrons which are on the order of a few electron-volts, so that ω_j is of order 10^{15} sec^{-1} , much the same as optical frequencies. The damping constants tend to be somewhat smaller, perhaps of order 10^{12} sec^{-1} (see below). Starting from low frequencies, $\omega \ll \omega_j^2$ and also $\omega\gamma_j \ll \omega_j^2$, then we can approximate the dielectric function as

$$\epsilon(\omega) \approx 1 + \frac{4\pi ne^2}{m} \sum \frac{f_j}{\omega_j^2} \quad (78)$$

which is a constant. Now, as ω increases from a low value, the real part of ϵ will also increase (slowly at first); when it gets to within about γ_j of the smallest ω_j , there is a resonance (the electron is being "pushed" by the electric field at a frequency close to its natural frequency) which will show up in ϵ_1 as a sudden rise, fall, and rise. After this, ϵ_1 is again roughly constant. There are as many such resonances as there are distinct resonant frequencies or terms in the sum over j .

The rapid variation of the dielectric function in the vicinity of a resonance also produces a rapidly varying index of refraction, meaning that waves with relatively

⁹In this we follow Jackson, but remember the Clausius-Mossotti relation from last quarter; we argued that the electric field which produces the polarization should be the local field and not the macroscopic field. It is not difficult to make the necessary corrections to what is given here.

Figure 1: Real and imaginary parts of ϵ near resonances

close frequencies propagate with quite different speeds. The frequency regime where ϵ_1 decreases with increasing ω is known as a region of *anomalous dispersion*.

The imaginary part of ϵ also behaves in an interesting fashion near a resonance. Because the denominator of the resonant term in $\epsilon(\omega)$ gets quite small at $\omega = \omega_j$ while the numerator for the imaginary part does not get small, there is a pronounced peak in ϵ_2 here. The smaller the value of γ_j , the bigger the peak. A large imaginary part of the dielectric function produces strong damping or *absorption* of the wave, so a region of anomalous dispersion is also a region of strong absorption, termed *resonant absorption*.

Finally, for ω very large in comparison with any other frequency in the system $\omega \gg \omega_j$, the dielectric function once again becomes simple and has the form

$$\epsilon(\omega) = 1 - \frac{4\pi nze^2}{m\omega^2} \equiv 1 - \frac{\omega_p^2}{\omega^2} \quad (79)$$

where we have introduced the *plasma frequency* of the electron system,

$$\omega_p \equiv \sqrt{\frac{4\pi nze^2}{m}}. \quad (80)$$

For typical values of n in solids, this frequency is of order 10^{16} sec^{-1} which is as large as or larger than the frequency of visible light. Our result is interesting in that the dielectric function is smaller than unity in this regime of frequency, meaning that a point of constant phase in a harmonic wave actually travels faster than the speed of light c . Even more remarkable is the possibility that $\epsilon(\omega) < 0$ in some range of frequency. For this to occur it is necessary to have $\omega < \omega_p$ but at the same time ω must be considerably larger than any resonant frequency ω_j and also larger than the damping parameters γ_j . Such conditions can be attained in some materials; a simple example is a tenuous plasma, or gas of charged particles. Then the resonant frequencies are all zero, the plasma frequency is rather low because the density of charges is not large, and the damping is small. See the following section.

6.1 Dielectric Response of Free Electrons

Some special cases are also worthy of mention. One is the case of free electrons. For these electrons there is no restoring force and so we may set the corresponding ω_j , called ω_0 , to zero. This has a profound effect on the dielectric function at low frequencies. If we extract the free-electron term from the remainder of the dielectric function and regard the latter as some constant ϵ_0 at low frequencies (see Eq. (78)), then we have

$$\epsilon = \epsilon_0 - \frac{4\pi n f_0 e^2}{m\omega(\omega + i\gamma_0)} = \epsilon_0 + i \frac{4\pi n f_0 e^2}{m\omega(\gamma_0 - i\omega)}. \quad (81)$$

This thing is singular as $\omega \rightarrow 0$, reflecting the fact that in the zero-frequency limit, the free electrons will be displaced arbitrarily far from their initial positions by any small electric field, producing a very large polarization. The singular term in ϵ in fact represents the conductivity of the free-electron material. To see how it is related to the conductivity, let us examine Ampère's law using this dielectric function and **no** macroscopic current \mathbf{J} , as this current will be included in the dielectric response (the

polarization produced by the free electrons). From $\nabla \times \mathbf{H} = c^{-1} \partial \mathbf{D} / t$, we find

$$\nabla \times \mathbf{H} = -i \frac{\omega}{c} \left(\epsilon_0 + i \frac{4\pi n f_0 e^2}{m\omega(\gamma_0 - i\omega)} \right) \mathbf{E} = \frac{4\pi}{c} \frac{n f_0 e^2}{m(\gamma_0 - i\omega)} \mathbf{E} - i \frac{\omega}{c} \epsilon_0 \mathbf{E}. \quad (82)$$

By contrast, we may choose **not** to include the free electrons' contribution to the polarization in which case $\epsilon = \epsilon_0$. Then, however, we have to include them as macroscopic current \mathbf{J} ; assuming linear response and isotropy, we may write $\mathbf{J} = \sigma \mathbf{E}$ where σ is the *electrical conductivity*. Using these relations, and Ampère's law, $\nabla \times \mathbf{H} = (4\pi/c) \mathbf{J} + c^{-1} \partial \mathbf{D} / \partial t$, we find

$$\nabla \times \mathbf{H} = \frac{4\pi}{c} \sigma \mathbf{E} - i \frac{\omega}{c} \epsilon_0 \mathbf{E}. \quad (83)$$

Comparison of the two preceding equations shows that by including the contribution of the free electrons in the polarization we have actually derived a simple expression for the conductivity,

$$\sigma = \frac{n f_0 e^2}{m(\gamma_0 - i\omega)} \rightarrow \frac{n f_0 e^2}{m \gamma_0}, \quad (84)$$

the last expression holding in the zero-frequency or static limit.

Comparison of measured conductivities with this result gives one an estimate of the damping constant. In very good metallic conductors such as *Cu* or *Ag*, $\sigma \sim 10^{17} \text{ sec}^{-1}$. The free-electron density is of order 10^{22} cm^{-3} and so one is led to $\gamma_0 \sim 10^{13} \text{ sec}^{-1}$ which is considerably smaller than typical resonant frequencies (for bound electrons, of course).

7 A Model for the Ionosphere

The ionosphere is a region of the upper atmosphere which is ionized by solar radiation (ultraviolet, x-ray, etc.). It may be simply described as a dilute gas of charged particles, composed of electrons and protons or other heavy charged objects. The dielectric properties of this medium are mainly produced by the lighter electrons, so we shall include only them in our description. We then have just one kind of charge

and it has zero resonant frequency. Because the medium is dilute, the damping is small; we shall ignore it. This is the approximation of a *collisionless* plasma and it leaves us with a very simple dielectric function,

$$\epsilon(\omega) = 1 - \frac{\omega_p^2}{\omega^2}. \quad (85)$$

For frequencies smaller than the plasma frequency, $\epsilon(\omega) < 0$, meaning that the wave number is pure imaginary since $k = \omega\sqrt{\epsilon}/c$; the corresponding wave will not propagate because its dependence on position is proportional to $\exp(i\mathbf{k} \cdot \mathbf{x})$ or $\exp(-|\mathbf{k}|z)$ given that $\mathbf{k} \parallel \hat{\mathbf{z}}^{10}$.

In the case of the ionosphere there is an additional complicating factor (which also makes the problem more interesting); the earth has a magnetic field which influences the motions of the electrons and hence the dielectric function. The equation of motion of the charges, including this field \mathbf{B}_0 is

$$m \frac{d^2 \mathbf{x}}{dt^2} = -e \left[\mathbf{E} + \frac{1}{c} \left(\frac{d\mathbf{x}}{dt} \times \mathbf{B}_0 \right) \right]. \quad (86)$$

We ignore the effect of the wave's magnetic induction. We shall also restrict (for simplicity) attention to the case of $\mathbf{k} \parallel \mathbf{B}_0$ and shall ignore the spatial variations of \mathbf{E} . In addition, and without loss of generality, we can let the wave have circular polarization. Hence we write the electric field as $\mathbf{E} = E_0 \boldsymbol{\epsilon}_{\pm} e^{-i\omega t}$.

Under these conditions, \mathbf{x} will be of the form $\mathbf{x} = \mathbf{x}_0 e^{-i\omega t}$; using this relation in the equation of motion, we find

$$-m\omega^2 \mathbf{x}_0 = -e \left[E_0 \boldsymbol{\epsilon}_{\pm} - \frac{i\omega}{c} B_0 (\mathbf{x}_0 \times \hat{\mathbf{z}}) \right]. \quad (87)$$

The solution of this equation is $\mathbf{x}_0 = x_0 \boldsymbol{\epsilon}_{\pm}$; one can see this easily if one realizes that $\boldsymbol{\epsilon}_{\pm} \times \hat{\mathbf{z}} = \pm i \boldsymbol{\epsilon}_{\pm}$:

$$\boldsymbol{\epsilon}_{\pm} \times \hat{\mathbf{z}} = \frac{1}{\sqrt{2}} (\hat{\mathbf{x}} \pm i\hat{\mathbf{y}}) \times \hat{\mathbf{z}} = \frac{1}{\sqrt{2}} (-\hat{\mathbf{y}} \pm i\hat{\mathbf{x}}) = \pm \frac{i}{\sqrt{2}} (\hat{\mathbf{x}} \pm i\hat{\mathbf{y}}) = \pm i \boldsymbol{\epsilon}_{\pm}. \quad (88)$$

¹⁰For most laboratory plasmas, this occurs at microwave frequencies

Hence the equation of motion, using $\mathbf{x}_0 = x_0\boldsymbol{\epsilon}_\pm$, is

$$-m\omega^2 x_0\boldsymbol{\epsilon}_\pm = -e\left[E_0 \pm \frac{\omega B_0 x_0}{c}\right]\boldsymbol{\epsilon}_\pm. \quad (89)$$

or

$$x_0 = \frac{eE_0/m}{\omega(\omega \mp \omega_B)} \quad (90)$$

where $\omega_B \equiv eB_0/mc$ is the *cyclotron frequency*. From this point we may determine the dielectric function by repeating the arguments used in the preceding section and find

$$\epsilon(\omega) = 1 - \frac{\omega_p^2}{\omega(\omega \mp \omega_B)}. \quad (91)$$

Our result tells us that waves with different polarization elicit different dielectric responses from the medium; such a phenomenon is known as *birefringence*. If a wave of general polarization is incident upon the plasma, it is in effect broken into its two circularly polarized components and these propagate independently. It is possible to have a wave with a frequency such that for one component $\epsilon(\omega) < 0$ and for the other, $\epsilon(\omega) > 0$. Hence, one will propagate and the other will not, providing a (not particularly practical) way of producing a circularly polarized wave.

In the specific case of the ionosphere, ω_p , ω_B , and ω can all be quite comparable. The density of electrons, which varies with the time of day and solar activity, is typically $\sim 10^5 - 10^6 \text{ cm}^{-3}$, leading to $\omega_p \sim 10^7 \text{ sec}^{-1}$. The earth's field $B_0 \sim 0.1 - 1.0 \text{ gauss}$, leading to $\omega_B \sim 10^7 \text{ sec}^{-1}$. A wave with $\omega \sim 10^7 \text{ sec}^{-1}$ is in the AM band; short-wave radio frequencies are somewhat higher, and FM radio or television have considerably higher frequencies. This means that FM and television signals are at frequencies so large that $\epsilon \approx 1$ and they propagate right through the ionosphere without significant reflection or attenuation. For this reason, the signals can be received only at locations where there is a direct path through the atmosphere from transmitter to receiver. For the lower frequency signals (short-wave and AM), however, there can be strong reflection from the ionosphere, making it possible to receive them relatively far from the transmitter. The higher the point in the ionosphere where the reflection takes place, the greater the effective range of the signal.

Figure 2: Dielectric constants vs. ω for the ionosphere

Figure 3: Electron density vs. height in the ionosphere

Because the electron density increases with height (and then decreases again), the higher frequencies tend to be reflected at greater heights (if they are reflected at all) than the lower ones, thereby giving greater range. That is why short-wave signals have longer range than AM signals, at least some of the time.

What happens if \mathbf{k} is not parallel to \mathbf{B}_0 ? The medium is still birefringent so that a wave of arbitrary polarization is broken into two components that propagate independently; however, the two components are not simple circularly polarized waves. In addition, the dielectric functions and hence the indices of refraction for these two waves depend on the angle between \mathbf{B}_0 and \mathbf{k} , so the medium is not only birefringent but also anisotropic.

8 Waves in a Dissipative Medium

We have seen in the preceding sections that the dielectric function will in general be complex, reflecting the fact that a wave will be dissipated or damped under many conditions. It therefore behooves us to learn more about the properties of waves when dissipation is present. As we have seen, we can do this by employing a complex dielectric function, and we can also do it, with the same basic results, by letting ϵ be real while introducing a real conductivity and thus a macroscopic current density.

We shall do the latter, for no particular reason.

Suppose that once again we have some linear medium with $\mathbf{D} = \epsilon\mathbf{E}$, $\mathbf{B} = \mu\mathbf{H}$, and $\mathbf{J} = \sigma\mathbf{E}$; ϵ , μ , and σ are taken as real. Then the Maxwell equations become

$$\nabla \cdot \mathbf{B} = 0, \quad \nabla \cdot \mathbf{E} = 0, \quad \nabla \times \mathbf{E} = -\frac{1}{c} \frac{\partial \mathbf{B}}{\partial t}, \quad (92)$$

and

$$\nabla \times \mathbf{B} = \frac{4\pi\mu}{c} \sigma \mathbf{E} + \frac{\epsilon\mu}{c} \frac{\partial \mathbf{E}}{\partial t}. \quad (93)$$

We have set ρ equal to zero in these equations. It may be that there is initially some macroscopic charge density within a conductor. If this is the case, that density will decay to zero with a characteristic time on the order of γ^{-1} where γ is the damping constant introduced in the section on dielectric functions; see Jackson, Problem 7.7.

Let us look for plane wave solutions to the field equations. Set $\mathbf{E}(\mathbf{x}, t) = \mathbf{E}_0 e^{i(\mathbf{k}\cdot\mathbf{x} - \omega t)}$ and $\mathbf{B}(\mathbf{x}, t) = \mathbf{B}_0 e^{i(\mathbf{k}\cdot\mathbf{x} - \omega t)}$. The divergence equations then tell us that $\mathbf{B}_0 \cdot \mathbf{k} = 0$ and $\mathbf{E}_0 \cdot \mathbf{k} = 0$ as in a nondissipative medium. From Faraday's law we find the familiar result

$$\mathbf{B}_0 = \frac{c}{\omega} (\mathbf{k} \times \mathbf{E}_0), \quad (94)$$

and from the Ampère's law we find

$$i(\mathbf{k} \times \mathbf{B}_0) = \frac{4\pi\mu\sigma}{c} \mathbf{E}_0 - i \frac{\omega\mu\epsilon}{c} \mathbf{E}_0. \quad (95)$$

If we take the cross product of \mathbf{k} with Eq. (94) and substitute Eq. (95) into the result where $\mathbf{k} \times \mathbf{B}_0$ appears, we find, after using $\mathbf{k} \times (\mathbf{k} \times \mathbf{E}_0) = -k^2 \mathbf{E}_0$, that

$$-i \frac{4\pi\mu\sigma}{c} \mathbf{E}_0 - \frac{\omega\mu\epsilon}{c} \mathbf{E}_0 = -\frac{ck^2}{\omega} \mathbf{E}_0 \quad (96)$$

or

$$k^2 = \frac{\omega^2\mu\epsilon}{c^2} + i \frac{4\pi\mu\sigma\omega}{c^2}. \quad (97)$$

Taking the point of view that ω is some given real frequency, we can solve this relation for the corresponding wavenumber k , which is complex. If we write $k = k_0 + i\alpha$, then

the real and imaginary parts of Eq. (97) give us two equations which may be solved for k_0 and α :

$$k_0^2 - \alpha^2 = \frac{\omega^2 \epsilon \mu}{c^2} \quad 2k_0 \alpha = \frac{\omega^2 \epsilon \mu}{c^2} \left(\frac{4\pi\sigma}{\epsilon\omega} \right). \quad (98)$$

The solution is

$$\begin{Bmatrix} k_0 \\ \alpha \end{Bmatrix} = \sqrt{\mu\epsilon} \left(\frac{\omega}{c} \right) \left\{ \frac{\sqrt{1 + \left(\frac{4\pi\sigma}{\omega\epsilon} \right)^2} \pm 1}{2} \right\}^{1/2}. \quad (99)$$

where the + sign refers to k_0 and the - sign to α .

This expression appears somewhat impenetrable although it doesn't say anything unexpected or remarkable. It takes on simple forms in the limits of high and low conductivity. The relevant dimensionless parameter is $4\pi\sigma/\epsilon\omega$. If it is much larger than unity, corresponding to a **good conductor**, then

$$k_0 \approx \alpha \approx \frac{\sqrt{2\pi\omega\mu\sigma}}{c} \equiv \frac{1}{\delta} \frac{4\pi\sigma}{\epsilon\omega} \gg 1. \quad (100)$$

where we have introduced the *penetration depth* δ . This is the distance that an electromagnetic wave will penetrate into a good conductor before being attenuated to a fraction $1/e$ of its initial amplitude. Since the wavelength of the wave is $\lambda = 2\pi/k_0$, δ is also a measure of the wavelength in the conductor.

For a **poor conductor**, by which we mean $4\pi\sigma/\omega\epsilon \ll 1$, one has

$$k_0 + i\alpha \approx \sqrt{\mu\epsilon} \frac{\omega}{c} + i \frac{2\pi}{c} \sqrt{\frac{\mu}{\epsilon}} \sigma. \quad (101)$$

Notice that in the latter case, the real part of the wavenumber is the same as in a nonconducting medium and the imaginary part is independent of frequency so that waves of all frequencies are attenuated by equal amounts over a given distance. Also, $\alpha \ll k_0$ which tells us that the wave travels many wavelengths before being attenuated significantly.

For a given σ , α is an increasing function of ω and saturates at high frequencies. Therefore, if one wants a wave to travel as far as possible, one wants to use as low frequency a wave as possible. Then one should be in the good-conductor limit where

the attenuation varies as $\sqrt{\omega}$ and vanishes as $\omega \rightarrow 0$. A well-known application of this rule has to do with radio communication with submarines; sea water is a reasonably good conductor $\sigma \sim 10^{11} \text{ sec}^{-1}$ and so to communicate with a submerged boat, one should send out low frequency signals which will penetrate to greater depths in the ocean than more standard signals.

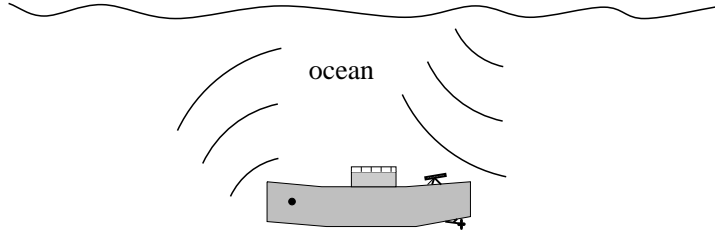


Figure 22: Low frequency waves can propagate through sea water.

Given that we have found the complex wave number, and letting \mathbf{k} point in the z -direction, we have

$$\mathbf{E}(\mathbf{x}, t) = \mathbf{E}_0 e^{i(k_0 z - \omega t)} e^{-\alpha z}; \quad (102)$$

the corresponding magnetic induction is found in the usual way (take $\mathbf{k} \times \mathbf{E}$):

$$\mathbf{B}(\mathbf{x}, t) = \frac{c}{\omega} (k_0 + i\alpha) (\hat{\mathbf{z}} \times \mathbf{E}_0) e^{i(k_0 z - \omega t)} e^{-\alpha z}. \quad (103)$$

Define the complex index of refraction

$$n \equiv \frac{c}{\omega} k = \frac{c}{\omega} (k_0 + i\alpha), \quad (104)$$

so that

$$\mathbf{B} = n(\hat{\mathbf{z}} \times \mathbf{E}). \quad (105)$$

Notice that because n is complex, \mathbf{B} is not in phase with \mathbf{E} ; to make the phase difference explicit, let us write n in polar form:

$$n = |n| e^{i\phi} \quad \text{where} \quad \phi = \arctan\left(\frac{\alpha}{k_0}\right). \quad (106)$$

We can find $|n|$ and ϕ in terms of other parameters; let $\gamma \equiv (4\pi\sigma/\omega\epsilon)^2$. Then

$$\phi = \arctan \left[\frac{\sqrt{1+\gamma}-1}{\sqrt{1+\gamma}+1} \right]^{1/2}. \quad (107)$$

Consider $\tan(2\phi)$:

$$\begin{aligned} \tan(2\phi) &= \frac{2 \tan \phi}{1 - \tan^2 \phi} = 2 \frac{[(\sqrt{1+\gamma}-1)/(\sqrt{1+\gamma}+1)]^{1/2}}{1 - \frac{\sqrt{1+\gamma}-1}{\sqrt{1+\gamma}+1}} \\ &= [(\sqrt{1+\gamma}-1)(\sqrt{1+\gamma}+1)]^{1/2} = \sqrt{\gamma} \end{aligned} \quad (108)$$

Thus,

$$\phi = \frac{1}{2} \arctan \sqrt{\gamma} = \frac{1}{2} \arctan \left(\frac{4\pi\sigma}{\omega\epsilon} \right). \quad (109)$$

Also,

$$|n| = \frac{c}{\omega} \sqrt{k_0^2 + \alpha^2} = \sqrt{\mu\epsilon} \left[1 + \left(\frac{4\pi\sigma}{\omega\epsilon} \right)^2 \right]^{1/4}. \quad (110)$$

Using these results in Eq. (105), we have

$$\mathbf{B}(\mathbf{x}, t) = \sqrt{\mu\epsilon} \left[1 + \left(\frac{4\pi\sigma}{\omega\epsilon} \right)^2 \right]^{1/4} e^{\frac{i}{2} \arctan \left(\frac{4\pi\sigma}{\omega\epsilon} \right)} (\hat{\mathbf{z}} \times \mathbf{E}_0). \quad (111)$$

The amount by which $\mathbf{B}(\mathbf{x}, t)$ is phase-shifted from $\mathbf{E}(\mathbf{x}, t)$ is easily seen from this expression to lie between 0 and $\pi/4$; it is zero in the small σ/ω limit and $\pi/4$ in the large σ/ω limit. Another significant feature of the expression for $\mathbf{B}(\mathbf{x}, t)$ is that in the small σ/ω limit, the amplitude of \mathbf{B} relative to that of \mathbf{E} is just $\sqrt{\mu\epsilon}$ as for insulators. But in the opposite limit, one finds that the relative amplitude is $\sqrt{4\pi\sigma\mu/\omega}$ which is much larger than unity. Here the wave has, relatively speaking, a much larger magnetic induction than electric field.

8.1 Reflection of a Wave Normally Incident on a Conductor

As an example, let us calculate the reflection of a wave normally incident on a conductor from vacuum.

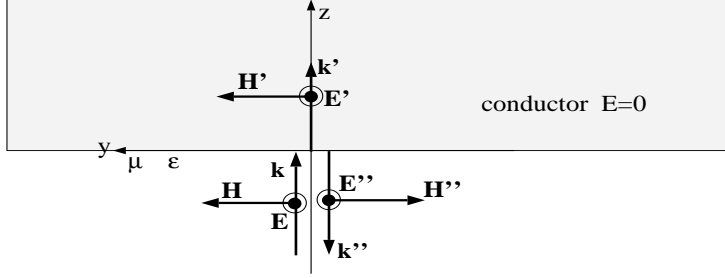


Figure 23: Wave normally incident on a conductor.

Then

$$\mathbf{k} = \frac{\omega}{c} \hat{\mathbf{z}} \quad \mathbf{k}' = \frac{\omega}{c} n \hat{\mathbf{z}}, n = \sqrt{\mu \epsilon} (1 + \gamma)^{1/4} e^{i\phi}. \quad (112)$$

The relevant boundary conditions are \mathbf{H}_t and \mathbf{E}_t continuous. Let $\mathbf{E}_0 = E_0 \hat{\mathbf{x}}$, $\mathbf{E}''_0 = E''_0 \hat{\mathbf{x}}$, and $\mathbf{E}'_0 = E'_0 \hat{\mathbf{x}}$. The corresponding magnetic field amplitudes are $\mathbf{H}_0 = E_0 \hat{\mathbf{y}}$, $\mathbf{H}''_0 = -E''_0 \hat{\mathbf{y}}$, and, for the transmitted wave in the conductor,

$$\mathbf{H}'_0 = \sqrt{\frac{\epsilon}{\mu}} (1 + \gamma)^{1/4} e^{i\phi} E'_0 \hat{\mathbf{y}}. \quad (113)$$

Our boundary conditions give immediately

$$E_0 + E''_0 = E'_0 \quad E_0 - E''_0 = \sqrt{\frac{\epsilon}{\mu}} (1 + \gamma)^{1/4} e^{i\phi} E'_0. \quad (114)$$

These may be combined to yield

$$E'_0 = \frac{2}{1 + \sqrt{\epsilon/\mu} (1 + \gamma)^{1/4} e^{i\phi}} E_0 \quad (115)$$

and

$$E''_0 = \frac{1 - \sqrt{\epsilon/\mu} (1 + \gamma)^{1/4} e^{i\phi}}{1 + \sqrt{\epsilon/\mu} (1 + \gamma)^{1/4} e^{i\phi}} E_0. \quad (116)$$

Let us calculate the Poynting vector in the conductor. Its time average is

$$\langle \mathbf{S}' \rangle = \frac{c}{8\pi} \Re(\mathbf{E}' \times \mathbf{H}'^*) = \frac{c}{8\pi} \Re \left\{ \frac{4|E_0|^2 \sqrt{\epsilon/\mu} (1 + \gamma)^{1/4} e^{-i\phi}}{|1 + \sqrt{\epsilon/\mu} (1 + \gamma)^{1/4} e^{i\phi}|^2} \right\} e^{-2\alpha z} \hat{\mathbf{z}}. \quad (117)$$

Using the interpretation of this vector as the energy current density, we may find the power per unit area transmitted into the conductor by evaluating $\langle \mathbf{S}' \rangle \cdot \hat{\mathbf{z}}$ at $z = 0$,

$$\mathcal{P}' = \frac{c}{2\pi} |E_0|^2 \sqrt{\frac{\epsilon}{\mu}} \frac{(1 + \gamma)^{1/4} \cos \phi}{1 + 2\sqrt{\epsilon/\mu} \cos \phi (1 + \gamma)^{1/4} + (\epsilon/\mu)(1 + \gamma)^{1/2}}. \quad (118)$$

The incident power per unit area is $\mathcal{P} = c|E_0|^2/8\pi$, so the fraction of the incident power which enters the conductor, where it is dissipated as Joule heat, is

$$T = \frac{\mathcal{P}'}{\mathcal{P}} = 4\sqrt{\frac{\epsilon}{\mu}} \frac{(1 + \gamma)^{1/4} \cos \phi}{1 + 2\sqrt{\epsilon/\mu} \cos \phi (1 + \gamma)^{1/4} + (\epsilon/\mu)(1 + \gamma)^{1/2}}. \quad (119)$$

This expression is much simplified in the limit of a good conductor where $\phi = \pi/4$, $\cos \phi = 1/\sqrt{2}$, and $\gamma \gg 1$. Then

$$T \approx 4\sqrt{\frac{\epsilon}{\mu}} \frac{\gamma^{1/4}(1/\sqrt{2})}{\epsilon\gamma^{1/2}/\mu} = 2\sqrt{2}\sqrt{\frac{\mu}{\epsilon}} \sqrt{\frac{\omega\epsilon}{4\pi\sigma}} = \frac{2\mu\omega}{c} \frac{c}{\sqrt{2\pi\sigma\omega\mu}} = \frac{2\mu\omega}{c} \delta. \quad (120)$$

For a good conductor such as Cu , $\sigma \sim 10^{17} \text{ sec}^{-1}$ and so a wave with frequency around 10^{10} sec^{-1} will have $\delta \sim 10^{-4} \text{ cm}$ or $1 \mu\text{m}$. Also, the better the conductor, the smaller the fraction of the incident power lost in the reflection process. For the example just given, $T \approx 10^{-4}$, meaning that the wave can be reflected some ten thousand times before becoming strongly attenuated.

9 Superposition of Waves; Pulses and Packets

No wave is truly monochromatic, although some waves, such as those produced by lasers, are exceedingly close to being so. Fortunately, in the case of linear media, the equations of motion for electromagnetic waves are completely linear and so any sum of harmonic solutions is also a solution. By making use of this superposition “principle” we can construct quite general solutions by superposing solutions of the kind we have already studied.

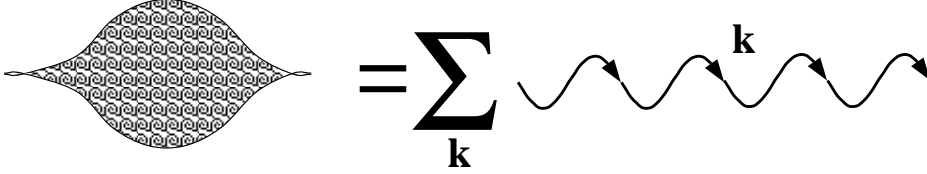


Figure 24: Any pulse in a linear media may be decomposed into a superposition of plane waves.

This procedure amounts to making a Fourier transform of the pulse. For simplicity we shall work in one spatial dimension which simply means that we will superpose waves whose wave vectors are all in the same direction (the z -direction). For the same reason, we shall also employ scalar waves; these could, for example, be the x components of the electric fields of the waves. One such wave has the form $e^{i(kz-\omega(k)t)}$ where we shall not initially restrict $\omega(k)$ to any particular form. Given a set of such waves, we can build a general solution of this kind (wave vector parallel to the z -axis) by integrating over some distribution $A(k)$ of them:

$$u(z, t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dk A(k) e^{i(kz-\omega(k)t)}. \quad (121)$$

At time $t = 0$, this function is simply

$$u(z, 0) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dk A(k) e^{ikz} \quad (122)$$

and the inverse transform gives A in terms of the zero-time wave:

$$A(k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dz u(z, 0) e^{-ikz}. \quad (123)$$

All of the standard rules of Fourier transforms are applicable to the functions $A(k)$ and $u(z, 0)$. For example, if $A(k)$ is a sharply peaked function with width Δk , then the width of $u(z, 0)$ must be of order $1/\Delta k$ or larger, and conversely. One may make this statement more precise by defining

$$(\Delta z)^2 \equiv \langle z^2 \rangle - \langle z \rangle^2 \quad (\Delta k)^2 \equiv \langle k^2 \rangle - \langle k \rangle^2, \quad (124)$$

where

$$\langle f(k) \rangle \equiv \frac{\int_{-\infty}^{\infty} dk f(k) |A(k)|^2}{\int_{-\infty}^{\infty} dk |A(k)|^2} \quad (125)$$

and

$$\langle f(z) \rangle \equiv \frac{\int_{-\infty}^{\infty} dz f(z) |u(z, 0)|^2}{\int_{-\infty}^{\infty} dz |u(z, 0)|^2}. \quad (126)$$

The relation between these widths which must be obeyed is

$$\Delta z \Delta k \geq 1/2. \quad (127)$$

Now, given a “reasonable” initial wave form $u(z, 0)$ ¹¹ with some Δz and a Fourier transform $A(k)$ with some Δk , the question we seek to answer is what will be the nature of $u(z, t)$? The answer is simple in principle because all we have to do is Fourier transform to find $A(k)$ and then do the integral specified by Eq. (121) to find $u(z, t)$. One can always do these integrals numerically if all else fails. Here we shall do some approximate calculations designed to demonstrate a few general points.

Suppose that we have found $A(k)$ and that it is some peaked function centered at k_0 with a width Δk . If $\omega(k)$ is reasonably well approximated by a truncated Taylor’s series expansion for k within Δk of k_0 , then we may write

$$\omega(k) \approx \omega_0 + \left. \frac{d\omega}{dk} \right|_{k_0} (k - k_0) \equiv \omega_0 + v_g (k - k_0) \quad (128)$$

where

$$\omega_0 \equiv \omega(k_0) \quad \text{and} \quad v_g = d\omega/dk|_{k_0}; \quad (129)$$

v_g is called the *group velocity* of the packet; notice that it can depend on the wave number k_0 which characterizes the typical wave numbers in the wave. In this approximation, one finds

$$u(z, t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dk A(k) e^{ik(z-v_g t)} e^{-i\omega_0 t} e^{iv_g k_0 t} = e^{i(v_g k_0 - \omega_0)t} u(z - v_g t, 0). \quad (130)$$

¹¹Its time derivative $\partial u(z, t)/\partial t|_{t=0}$ must also be given to allow a unique solution of the initial value problem; our discussion is therefore incomplete but can be corrected easily.

This result tells us that the wave packet retains its initial form and translates in space at a speed v_g . It does not spread (disperse) or distort in any way. In particular, the energy carried by the wave will move with a speed v_g .

The group velocity is evidently an important quantity. We may write it in terms of the index of refraction by using the defining relation $k = \omega n(\omega)/c$. Take the derivative of this with respect to k :

$$1 = \left(\frac{n}{c} + \frac{\omega}{c} \frac{dn}{d\omega} \right) \frac{d\omega}{dk} \quad (131)$$

or

$$v_g = \frac{c}{n + \omega \frac{dn}{d\omega}}. \quad (132)$$

As an example consider the collisionless plasma relation $n = \sqrt{1 - \omega_p^2/\omega^2}$. One easily finds that

$$v_g = c\sqrt{1 - \omega_p^2/\omega^2}. \quad (133)$$

For $\omega < \omega_p$, the group velocity is imaginary which corresponds to a damped wave; for $\omega > \omega_p$, it is positive and increases from zero to c as ω increases.

Our calculations thus far have not resulted in any spreading or distortion of the wave packet because we did not include higher-order terms in the relation (called a *dispersion relation*) between ω and k . Let's treat a simple example in which $A(k)$ is a gaussian function of $k - k_0$,

$$A(k) = \left(\frac{A_0}{\delta} \right) e^{-(k-k_0)^2/2\delta^2}. \quad (134)$$

Further, let $\omega(k)$ be approximated by

$$\omega(k) = \omega_0 + v_g(k - k_0) + \alpha(k - k_0)^2. \quad (135)$$

The corresponding $u(z, t)$ is

$$u(z, t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dk \frac{A_0}{\delta} e^{-(k-k_0)^2/2\delta^2} e^{ikz - i[\omega_0 + v_g(k-k_0) + \alpha(k-k_0)^2]t}$$

$$\begin{aligned}
&= \frac{A_0}{\delta\sqrt{2\pi}} e^{i(k_0 z - \omega_0 t)} \int_{-\infty}^{\infty} dk e^{i(k-k_0)(z-v_g t)} e^{-(1/2\delta^2 + i\alpha t)(k-k_0)^2} \\
&= \frac{A_0}{\sqrt{1+2i\alpha\delta^2 t}} e^{-(z-v_g t)^2 \delta^2 / [2(1+2i\alpha\delta^2 t)]} e^{i(k_0 z - \omega_0 t)}. \tag{136}
\end{aligned}$$

If $\alpha = 0$, this is a Gaussian-shaped packet which travels at speed v_g with a constant width equal to δ^{-1} . If $\alpha \neq 0$, it is still a Gaussian-shaped packet traveling at speed v_g ; however, it does not have a constant width any longer. To make the development of the width completely clear, consider $|u(z, t)|^2$ which more nearly represents the energy density in the wave:

$$|u(z, t)|^2 = \frac{A_0^2}{\sqrt{1+4\alpha^2\delta^4 t^2}} e^{-(z-v_g t)^2 \delta^2 / (1+4\alpha^2\delta^4 t^2)}. \tag{137}$$

The width of this traveling Gaussian is easily seen to be

$$w(t) = \sqrt{1+4\alpha^2\delta^4 t^2} / \delta. \tag{138}$$

At short times the width increases as the square of the time, while at long times it becomes linear with t .

When the packet spreads, or disperses, in this fashion, to what extent does it make sense to think about the wave as a localized object? One measure is the width of the packet as compared with the distance it has moved. After a long time the width is approximately $2\alpha\delta t$ while the distance the packet has moved is $v_g t$. The ratio of these distances is $2\alpha\delta/v_g$, so our condition for having a localized object is

$$2\alpha\delta/v_g \ll 1. \tag{139}$$

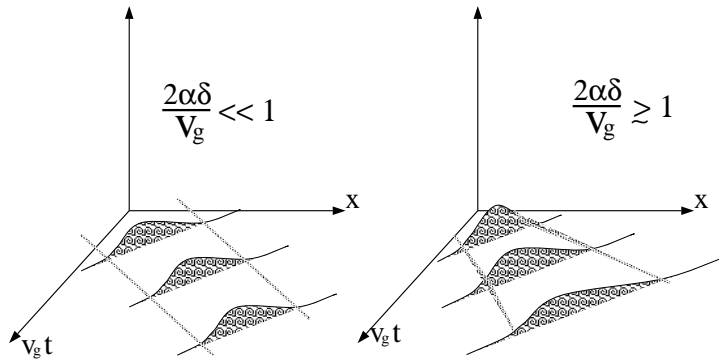


Figure 25: When δ is small, the wave is composed of few wavenumbers.

In addition, of course, the initial width of the packet must be small compared to $v_g t$ which is always possible if one waits long enough. Our inequality clearly puts a limit on the allowable size of α , for a given δ , necessary to have a well-defined pulse. For smaller δ , one can get away with larger α , a simple consequence of the fact that small δ means the width of the packet in k -space is small, leading to less dispersion.

9.1 A Pulse in the Ionosphere

Let's look also at the fate of a wave packet propagating in the ionosphere; we found in an earlier section, treating the ionosphere as a collisionless plasma and with \mathbf{k} parallel to \mathbf{B}_0 , that $\epsilon(\omega) = 1 + \omega_p^2/\omega(\omega_B - \omega)$ for one particular polarization of the wave. If ω is small enough compared to other frequencies, we may approximate in such a way that $n(\omega) = \omega_p/\sqrt{\omega\omega_B}$, which gives rise to anomalous dispersion indeed. Defining $\omega_0 \equiv \omega_p^2/\omega_B$, one finds that the group velocity of a signal is $v_g = 2c\sqrt{\omega/\omega_0}$.

Let us see how a pulse with the same $A(k)$ as in the previous example propagates. We have

$$\begin{aligned} u(z, t) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dk \frac{A_0}{\delta} e^{-(k-k_0)^2/2\delta^2 + ikz - ic^2 k^2 t/\omega_0} \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dk \frac{A_0}{\delta} e^{-(k-k_0)^2/2\delta^2 + i(k-k_0)z + ik_0 z - ic^2 t(k-k_0)^2/\omega_0 - i2c^2 k_0 t(k-k_0)/\omega_0 - ic^2 k_0^2 t/\omega_0} \\ &= \frac{A_0}{1 + 2i\delta^2 c^2 t/\omega_0)^{1/2}} e^{i(k_0 z - c^2 k_0^2 t/\omega_0)} e^{-\frac{(z - 2c^2 k_0 t/\omega_0)^2 \delta^2}{2(1 + 2i\delta^2 c^2 t/\omega_0)}} \quad (140) \end{aligned}$$

This is a traveling, dispersing Gaussian. Its speed is the group velocity $v_g(k_0)$. The width of the Gaussian is

$$w(t) = \sqrt{1 + 4\delta^4 c^4 t^2/\omega_0^2}/\delta \rightarrow 2\delta c^2 t/\omega_0 \quad (141)$$

at long times. The packet spreads at a rate given by $v_w = 2\delta c^2/\omega_0$. The ratio of this spreading rate to the group velocity is δ/k_0 and so we retain a well-defined pulse provided the spread in wavenumber is small compared to the central wavenumber.

Pulses of this general type are generated in the ionosphere by thunderstorms. They have a very broad range of frequencies ranging from very low ones up into at least the AM radio range. The electromagnetic waves tend to be guided along lines of the earth's magnetic induction, and so, if for example the storm is in the southern hemisphere, the waves travel north in the ionosphere along lines of \mathbf{B} and then come back to earth in the northern hemisphere.

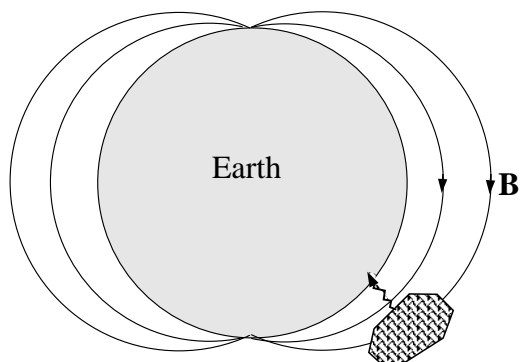


Figure 26: Lightning in the southern hemisphere yields wistlers in the north.

By this time they are much dispersed, with the higher frequency components arriving well before the lower frequency ones since $v_g = 2c\sqrt{\omega/\omega_0}$ for $\omega \ll \omega_0$. Frequencies in the audible range, $\omega \sim 10^2$ or 10^3 sec^{-1} take one or more seconds (a long time for electromagnetic waves) to arrive. If one receives the signal and converts it directly to an audio signal at the same frequency, it sounds like a whistle, starting at high frequencies and continuing down to low ones over a time period of several seconds. This characteristic feature has caused such waves to be known as *wistlers*.

10 Causality and the Dielectric Function

A linear dispersive medium is characterized by a dielectric function $\epsilon(\omega)$ having physical origins that we have just finished exploring. One consequence of having such a relation between $\mathbf{D}(\mathbf{x}, \omega)$ and $\mathbf{E}(\mathbf{x}, \omega)$, that is,

$$\mathbf{D}(\mathbf{x}, \omega) = \epsilon(\omega)\mathbf{E}(\mathbf{x}, \omega), \quad (142)$$

is that the relation between $\mathbf{D}(\mathbf{x}, t)$ and $\mathbf{E}(\mathbf{x}, t)$ is **nonlocal** in time. To see this we have only to look at the Fourier transforms of \mathbf{D} and \mathbf{E} . One has

$$\mathbf{D}(\mathbf{x}, t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d\omega \mathbf{D}(\mathbf{x}, \omega) e^{-i\omega t} \quad (143)$$

and its inverse

$$\mathbf{D}(\mathbf{x}, \omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dt \mathbf{D}(\mathbf{x}, t) e^{-i\omega t}, \quad (144)$$

similar relations hold for $\mathbf{E}(\mathbf{x}, t)$ and $\mathbf{E}(\mathbf{x}, \omega)$. Using the relation $\mathbf{D}(\mathbf{x}, \omega) = \epsilon(\omega)\mathbf{E}(\mathbf{x}, \omega)$, we have

$$\mathbf{D}(\mathbf{x}, t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d\omega \epsilon(\omega) \mathbf{E}(\mathbf{x}, \omega) e^{-i\omega t}. \quad (145)$$

We can write $\mathbf{E}(\mathbf{x}, \omega)$ here as a Fourier integral and so have

$$\begin{aligned} \mathbf{D}(\mathbf{x}, t) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega \epsilon(\omega) e^{-i\omega t} \int_{-\infty}^{\infty} dt' e^{i\omega t'} \mathbf{E}(\mathbf{x}, t') \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} dt d\omega [\epsilon(\omega) - 1 + 1] \mathbf{E}(\mathbf{x}, t') e^{-i\omega(t-t')} = \\ \mathbf{E}(\mathbf{x}, t) + \frac{1}{2\pi} \int_{-\infty}^{\infty} dt d\omega [\epsilon(\omega) - 1] \mathbf{E}(\mathbf{x}, t') e^{-i\omega(t-t')} &\equiv \mathbf{E}(\mathbf{x}, t) + 4\pi \mathbf{P}(\mathbf{x}, t). \end{aligned} \quad (146)$$

The final term, $4\pi \mathbf{P}(\mathbf{x}, t)$, can be written in terms the Fourier transform¹² of $\epsilon(\omega) - 1$; introduce the function

$$G(t) \equiv \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega [\epsilon(\omega) - 1] e^{-i\omega t}. \quad (147)$$

Then we have

$$\mathbf{D}(\mathbf{x}, t) = \mathbf{E}(\mathbf{x}, t) + \int_{-\infty}^{\infty} dt' G(t - t') \mathbf{E}(\mathbf{x}, t') \quad (148)$$

which may also be written as

$$\mathbf{D}(\mathbf{x}, t) = \mathbf{E}(\mathbf{x}, t) + \int_{-\infty}^{\infty} d\tau G(\tau) \mathbf{E}(\mathbf{x}, t - \tau). \quad (149)$$

This equation makes it clear that when the medium has a frequency-dependent dielectric function, as all materials do, then the electric displacement at time t depends on

¹²Provided the order of integration can be reversed and the transform exists.

the electric field not only at time t but also at times other than t . This is somewhat disturbing because one can see that, depending on the character of G , we could get a polarization $\mathbf{P}(\mathbf{x}, t)$ that depends on values of $\mathbf{E}(\mathbf{x}, t')$ for $t' > t$, which means we get an effect arising from a cause that occurs at a time later than the effect. This behavior can be avoided if the function $G(\tau)$ vanishes when $\tau < 0$, and that is what in fact happens.

Let's look at a simple example with

$$\epsilon(\omega) = 1 + \frac{\omega_p^2}{\omega_0^2 - \omega^2 - i\omega\gamma}. \quad (150)$$

Then

$$G(\tau) = \frac{\omega_p^2}{2\pi} \int_{-\infty}^{\infty} d\omega \frac{e^{-i\omega\tau}}{\omega_0^2 - \omega^2 - i\omega\gamma} \quad (151)$$

This integral was made for contour integration techniques. The poles of the integrand are in the lower half-plane in complex frequency space at

$$\omega_{\pm} = \frac{1}{2}[\pm\sqrt{4\omega_0^2 - \gamma^2} - i\gamma]; \quad (152)$$

without producing a contribution to the integral, we can close the contour in the upper (lower) half-plane when τ is smaller (larger) than zero. Because there are poles only in the lower half-plane, we can see immediately that $G(\tau)$ will be zero for $\tau < 0$. That is pleasing since we don't want the displacement (that is, the polarization) to respond at time t to the electric field at times later than t .

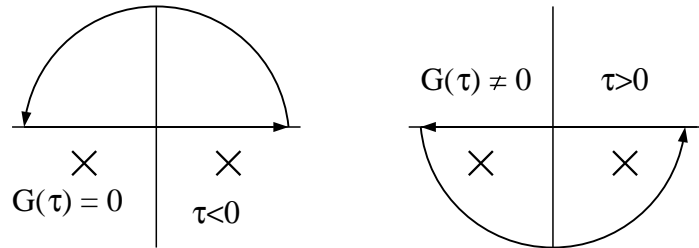


Figure 27: Because there are poles only in the lower half-plane, we can see immediately that $G(\tau)$ will be zero for $\tau < 0$.

Applying Cauchy's theorem to the case of $\tau > 0$, one finds that, for all τ ,

$$G(\tau) = \omega_p^2 e^{-\gamma\tau/2} \frac{\sin(\nu_0\tau)}{\nu_0} \theta(\tau) \quad (153)$$

where $\theta(x)$ is the step function, equal to unity for $x > 0$ and to zero otherwise, and $\nu_0 = \sqrt{\omega_0^2 - \gamma^2/4}$. The characteristic range in time of this function is γ^{-1} and hence the nonlocal (in time) character of the response is not important for frequencies smaller than about γ ; it becomes important for larger ones.

One may naturally wonder whether there should also be nonlocal character of the response in space as well as in time. In fact there should and will be under some conditions. If we look back at our derivation of the model dielectric function, we see that the equation of motion of the particle was solved using $\mathbf{E}(0, t)$ instead of $\mathbf{E}(\mathbf{x}, t)$; the latter is of course the more correct choice. The difference is not important so long as the excursions of the charge from the point on which it is bound are much smaller than the wavelength of the radiation, which is the case for any kind of wave with frequencies up to those of soft X-rays. Hence the response can be expected to be local in space in insulating materials. However, if an electron is free, it can move quite far during a cycle of the field and if it does so, the response will be nonlocal in space as well as time.

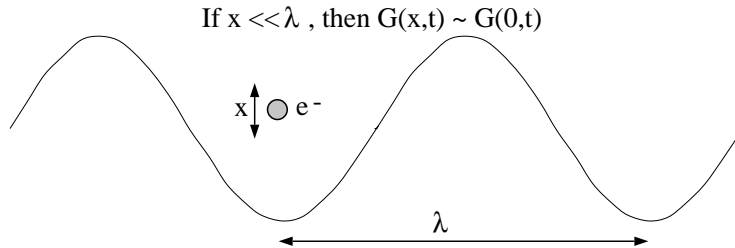


Figure 28: $G(\tau, \mathbf{x})$ will not be \mathbf{x} dependent if the excursions of the charge from the point on which it is bound are much smaller than the wavelength of the radiation.

Returning to the question of causality, we have seen that the simple model dielectric function produces a function $G(t)$ which is zero for $t < 0$, as is necessary if “causality” is to be preserved, by which we mean there is no response in advance of the “cause” of that response. It is easy to see what are the features of the dielectric

function that give rise to the result $G(t) = 0$ for $t < 0$. One is that there are no simple poles of the dielectric function in the upper half of the complex frequency plane. Another is that the dielectric function goes to zero for large ω fast enough that we can do the contour integral as we did it.

More generally, if one wants to have a function $G(t)$ which is consistent with the requirements of causality, this implies certain conditions on *any* $\epsilon(\omega)$. Additional conditions can be extracted from such simple things as the fact that $G(t)$ must be real so that \mathbf{D} is real if \mathbf{E} is. Without going into the details of the matter (see Jackson) let us make some general statements. The reality of G requires that

$$\epsilon(-\omega) = \epsilon^*(\omega^*). \quad (154)$$

That G is zero for negative times requires that $\epsilon(\omega)$ be analytic in the upper half of the frequency plane. Assuming that $G(t) \rightarrow 0$ as $t \rightarrow \infty$, one finds that $\epsilon(\omega)$ is analytic on the real axis. This last statement is actually not true for conductors which give a contribution to $\epsilon \sim i\sigma/\omega$ so that there is a pole at the origin. Finally, from the small-time behavior of $G(t)$, one can infer that at large frequencies the real part of $\epsilon(\omega) - 1$ varies as ω^{-2} while the imaginary part varies as ω^{-3} . This is accomplished by repeatedly integrating by parts

$$\epsilon(\omega) - 1 = \int_0^\infty d\tau G(\tau) e^{i\omega\tau} \approx \frac{iG(0^+)}{\omega} - \frac{G'(0^+)}{\omega^2} + \frac{iG''(0^+)}{\omega^3} + \dots \quad (155)$$

This series is convergent for large ω . The first term vanishes if $G(\tau)$ is continuous across $\tau = 0$. Thus

$$\Re(\epsilon(\omega) - 1) \sim \frac{1}{\omega^2} \quad \Im(\epsilon(\omega) - 1) \sim \frac{1}{\omega^3} \quad (156)$$

From inspection, one may see that the various dielectric functions we have contrived satisfy these conditions.

Given that the dielectric function has the analyticity properties described above, it turns out that by rather standard manipulations making use of Cauchy's integral

theorem, one can write the imaginary part of $\epsilon(\omega)$ in terms of an integral of the real part over real frequencies and conversely. That one can do so is important because it means, for example, that if one succeeds in measuring just the real (imaginary) part, the imaginary (real) part is then known. The downside of this apparent miracle is that one has to know the real or imaginary part for all real frequencies in order to obtain the other part.

To see how this works, notice that as a consequence of the analytic properties of the dielectric function, it obeys the relation

$$\epsilon(z) = 1 + \frac{1}{2\pi i} \oint_C d\omega' \frac{\epsilon(\omega') - 1}{\omega' - z} \quad (157)$$

where the contour does not enter the lower half-plane (where ϵ may have poles) anywhere and where z is inside of the contour. Let C consist of the real axis and a large semicircle which closes the path in the upper half-plane.

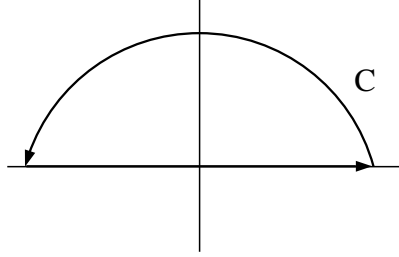


Figure 29: Contour C : $\epsilon(\omega)$ is analytic inside and on C .

Then, given that ϵ falls off fast enough, as described above, at large ω , the semicircular part of the path does not contribute to the integral. Hence we find that

$$\epsilon(z) = 1 + \frac{1}{2\pi i} \int_{-\infty}^{\infty} d\omega' \frac{\epsilon(\omega') - 1}{\omega' - z}. \quad (158)$$

At this juncture, z can be any point in the upper half-plane. Let's use $z = \omega + i\eta$ and take the limit of $\eta \rightarrow 0$, finding

$$\epsilon(\omega + i\eta) = 1 + \frac{1}{2\pi i} \int_{-\infty}^{\infty} d\omega' \frac{\epsilon(\omega') - 1}{\omega' - \omega - i\eta}. \quad (159)$$

The presence of the η in the denominator means that at the integration point $\omega' = \omega$, we must be careful to keep the singularity inside of, or above, the contour. Here we

pick up $2\pi i$ times the residue, and the residue is just $\epsilon(\omega) - 1$. This relation shows identity but is not useful otherwise. However, one can also pull the following trick: If we integrate right across the singularity, taking the principal part (denoted P) of the integral plus an infinitesimal semicircle right below the singularity that amounts to taking $i\pi$ times the residue. Hence we can make the replacement

$$\frac{1}{\omega' - \omega - i\eta} \rightarrow P\left(\frac{1}{\omega' - \omega}\right) + i\pi\delta(\omega' - \omega) \quad (160)$$

where P stands for the principal part; this substitution leads to

$$\epsilon(\omega) = 1 + \frac{1}{\pi i} P \int_{-\infty}^{\infty} d\omega' \frac{\epsilon(\omega') - 1}{\omega' - \omega} \quad (161)$$

Let us write separately the real and imaginary parts of this expression:

$$\begin{aligned} \Re[\epsilon(\omega)] &= 1 + \frac{1}{\pi} P \int_{-\infty}^{\infty} d\omega' \frac{\Im[\epsilon(\omega')]}{\omega' - \omega} \\ \Im[\epsilon(\omega)] &= \frac{1}{\pi} P \int_{-\infty}^{\infty} d\omega' \frac{\Re[\epsilon(\omega') - 1]}{\omega' - \omega} \end{aligned} \quad (162)$$

These equations are known as the *Kramers-Kronig relations* for the dielectric function. They may be written as integrals over only positive frequencies because of the fact that the real part of $\epsilon(\omega)$ is an even function of ω while the imaginary part is odd. It should also be pointed out that we have assumed there is no pole in $\epsilon(\omega)$ at $\omega = 0$; if there is one (conductors have dielectric functions with this property) some modification of these expressions will be necessary.

11 Arrival of a Signal in a Dispersive Medium

Most of the wave trains one receives, such as radio signals, messages from within or without the galaxy (sent by stars, pulsars, neutron stars, etc), and so on, have to traverse dispersive media to get wherever they go. Consequently it is of considerable importance to know how the signals are distorted by the intervening material. The basic idea is this: we have seen how a pulse centered at some particular wave

number or frequency tends to travel with the group velocity of the central frequency and also spreads some as a consequence of the frequency-dependence of the index of refraction or dielectric function. If the dispersion is very large, as in regions of anomalous dispersion, the pulse will not simply spread some but will be distorted beyond recognition. In addition, frequency components in this region will be strongly attenuated and so will disappear from the wave train after awhile. If a signal is initially very broad in frequency, having components ranging from very low ones, where the group velocity is roughly constant and equal to $c/\sqrt{\epsilon(0)}$, to very high ones where $\epsilon(\omega) \approx 1$ and the group velocity is about c , then the signal that arrives after traveling through a significant length of medium will be very different indeed from the initial one. All of the frequency components around the regions of anomalous dispersion will be gone. There will be some high-frequency component which travels at a speed around c and so arrives first; it is generally called the “first precursor.” Then after awhile the remainder of the signal will arrive. The leading edge of this part is called the “second precursor” and it consists of those lower frequency components which have the largest group velocity and which are not appreciably attenuated. These are usually¹³ the very low frequency components.

It is a straightforward matter to determine what the signal will be, using the superposition principle. Consider a pulse in one dimension with an amplitude $u(z, t)$. Given that one knows the form of this pulse and its first space derivatives as functions of time at some initial position in space¹⁴, called z_0 , then one may determine by Fourier analysis the amplitude $A(\omega)$ of the various frequency components in it. Since a frequency component ω propagates according to $\exp[i(\omega n(\omega)z/c - \omega t)]$, it is then easy in principle to find $u(z, t)$:

$$u(z, t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d\omega A(\omega) e^{i(\omega n(\omega)z/c - \omega t)}. \quad (163)$$

¹³But not always; the whistler provides a counter example.

¹⁴Notice that instead of solving an initial value problem in time, we here rephrase it as an initial value problem in space.

If we can do this integral for the index of refraction of our choice, we can find the form of the wave train at all space points at any later time. Among other things, one can show by making use of the analyticity properties of the dielectric function that it is impossible for an electromagnetic signal to travel faster than the speed of light. See Jackson.

As a very simple example, consider a single-resonance dielectric function with no absorption,

$$\epsilon(\omega) = 1 + \frac{\omega_p^2}{\omega_0^2 - \omega^2} = n^2(\omega) \quad (164)$$

or

$$n(\omega) = \left(\frac{\omega_0^2 - \omega^2 + \omega_p^2}{\omega_0^2 - \omega^2} \right)^{1/2}. \quad (165)$$

Then

$$2n \frac{dn}{d\omega} = 2\omega \frac{\omega_p^2}{(\omega_0^2 - \omega^2)^2} \quad (166)$$

so

$$n\omega \frac{dn}{d\omega} + n^2 = \frac{\omega^2 \omega_p^2}{(\omega_0^2 - \omega^2)^2} + 1 + \frac{\omega_p^2}{\omega_0^2 - \omega^2} = \frac{\omega_0^4 - 2\omega_0^2 \omega^2 + \omega^4 + \omega_p^2 \omega_0^2}{(\omega_0^2 - \omega^2)^2}. \quad (167)$$

Hence

$$v_g = \frac{c}{\omega \frac{dn}{d\omega} + n} = c \left(\frac{\omega_0^2 - \omega^2 + \omega_p^2}{\omega_0^2 - \omega^2} \right)^{1/2} \frac{(\omega_0^2 - \omega^2)^2}{(\omega_0^2 - 2\omega_0^2 \omega^2 + \omega^4 + \omega_p^2 \omega_0^2)} \quad (168)$$

The first plot shows the character of v_g and of $n(\omega)$. The group velocity is largest for the largest frequencies; these will combine to provide the first precursor which

may well be weak to the extent that the initial pulse does not contain many high-frequency components. The first precursor continues as lower frequency components (but still larger than $\sqrt{\omega_0^2 + \omega_p^2}$) come through. While this is going on, all of the very low frequency components arrive. This is the second precursor. Finally, if the pulse is actually a long wave train which has one predominant frequency in it, then after some time the received pulse settles down to something more or less harmonic, showing just this frequency.

A Waves in a Conductor

When we discussed the propagation of waves in an ideal dielectric, we showed that the fields were transverse to the direction of propagation. This corresponds to an insulating material, with a vanishing electrical conductivity. When we extend our discussion to include media of finite conductivity σ , there is no a priori reason that the fields will still be transverse to the direction of propagation.

Let's show that we need not worry about any longitudinal fields. Suppose that once again we have some linear medium with $\mathbf{D} = \epsilon\mathbf{E}$, $\mathbf{B} = \mu\mathbf{H}$, and $\mathbf{J} = \sigma\mathbf{E}$; ϵ , μ , and σ are taken as real. Then the Maxwell equations become

$$\nabla \cdot \mathbf{B} = 0, \quad \nabla \cdot \mathbf{E} = 0, \quad \nabla \times \mathbf{E} = -\frac{1}{c} \frac{\partial \mathbf{B}}{\partial t}, \quad (169)$$

and

$$\nabla \times \mathbf{B} = \frac{4\pi\mu}{c} \sigma \mathbf{E} + \frac{\epsilon\mu}{c} \frac{\partial \mathbf{E}}{\partial t}. \quad (170)$$

Let's look for solutions to Maxwell's equations in the form of longitudinal waves,

$$\mathbf{E} = \hat{\mathbf{z}}E(z, t) ; \quad \mathbf{B} = \hat{\mathbf{z}}B(z, t) \quad (171)$$

Since $\nabla \cdot \mathbf{B} = \nabla \cdot \mathbf{E} = 0$, E and B can be functions of time only. Thus $\nabla \times \mathbf{E} = \nabla \times \mathbf{B} = 0$, and the other two Maxwell's equations become

$$\frac{\partial \mathbf{B}}{\partial t} = 0 \quad ; \quad \frac{4\pi\sigma}{c} \mathbf{E} + \frac{\epsilon}{c} \frac{\partial \mathbf{E}}{\partial t} = 0 \quad (172)$$

The first says that \mathbf{B} must be constant. The second says that \mathbf{E} while uniform in space has a time dependence

$$\mathbf{E}(t) = \mathbf{E}(0)e^{-4\pi\sigma t/\epsilon} \quad (173)$$

In a conductor, $\sigma \approx 10^{16} \text{ sec}^{-1}$. Thus $\mathbf{E}(t)$ falls off very rapidly and may be neglected. Thus as worst there is a constant longitudinal \mathbf{B} -field as part of our wave in a conductor. Since Maxwell's equations are linear, we may drop this trivial solution and just consider the transverse fields.

Waveguides and Cavities

John William Strutt

also known as

Lord Rayleigh

(1842 - 1919)

September 17, 2001

Contents

1	Reflection and Transmission at a Conducting Wall	2
1.1	Boundary Conditions	3
1.2	Power and Energy Loss	4
2	Wave Guides	7
2.1	Fundamental Equations	8
2.1.1	Boundary Conditions	10
2.2	Transverse Modes	11
2.2.1	TEM Mode	11
2.2.2	TE and TM Modes	12
2.3	Energy Flow	14
2.3.1	TE Modes in Rectangular and Circular Guides	15
3	Attenuation of Modes in Waveguides	16

In this chapter we continue with the topic of solutions of the Maxwell equations in the form of waves. This time we seek solutions in the presence of bounding surfaces which may take a variety of forms. The basic possibilities are to have boundaries in

1. one dimension only, such as a pair of parallel planes;
2. two dimensions, such as several intersecting planes forming a pipe or channel;
and
3. three dimensions, such as a collection of intersecting planes that completely bound some region of space.

The materials employed to form the boundaries are usually¹ conductors. The mathematical problem is a boundary-value problem for solutions of the Maxwell equations. We shall look at harmonic solutions within the cavity or channel and must match these solutions onto appropriate ones within the walls or bounding materials. If the walls are constructed from a “good” conductor, the boundary conditions become simple and the boundary-value problem itself is not too difficult. This point is explored in the following sections.

1 Reflection and Transmission at a Conducting Wall

We consider the reflection and transmission of a harmonic plane wave incident on a conducting material at a planar surface. We let the incident wave have an arbitrary angle of incidence - which gives a hard problem to solve in the general case - and then imagine that the conductivity is very large - which simplifies the solution by allowing an expansion in a small parameter. Physically, the central point is that if $\sigma \gg \omega$,

¹Dielectric materials are also used, with conditions such that total internal reflection takes place at the surfaces in order to keep the wave within the channel or cavity.

then the skin depth $\delta = c/\sqrt{2\pi\sigma\omega\mu}$ of the wave in the conductor is much smaller than the wavelength λ of the incident wave. The distance over which the fields vary in the conductor depends on the direction. In the direction normal to the surface, this distance is δ ; in directions parallel to the surface, it is λ . Thus by having $\delta \ll \lambda$, we can often ignore variations of the fields parallel to the surface in comparison with variations normal to the surface; in effect, the wave in the conductor travels normal to the surface no matter what the angle of incidence.

1.1 Boundary Conditions

First, let us consider the boundary or continuity conditions at the interface. We can find appropriate conditions by using the Maxwell equations and either Stokes' theorem or Gauss's Law in the usual way. Let us first do this by employing a rectangle or pillbox which has a size t normal to the interface which is much larger than δ . At the same time, the size l of these constructs parallel to the interface must be large compared to t but small compared to λ , so we have the condition

$$\lambda \gg l \gg t \gg \delta \tag{1}$$

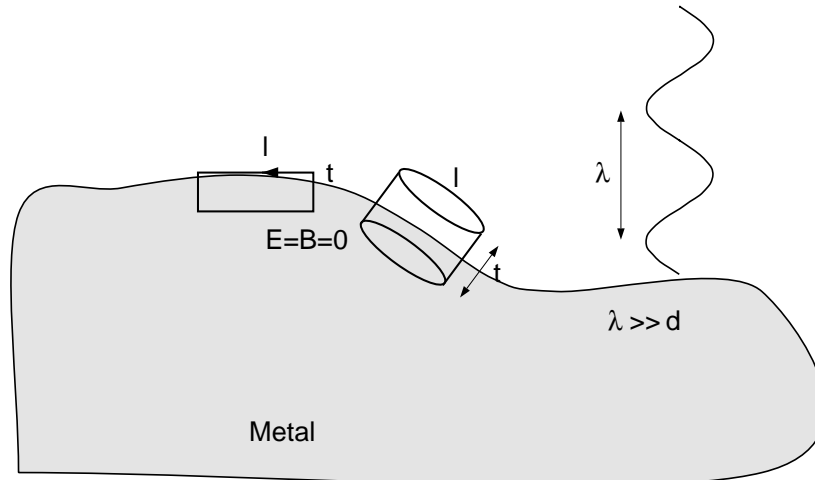


Fig.1: Integration surfaces adjacent to a good conductor.

which can be satisfied by a metal with a large enough conductivity (and an incident

radiation with a small enough frequency). Then, because the side of the rectangle, or face of the pillbox, within the conductor is placed in a region where the transmitted fields have been attenuated to very small values, compared to the incident amplitudes, we can say that these fields are zero. The result is that the continuity conditions become

$$\mathbf{n} \cdot \mathbf{D} = 4\pi\sigma_q \quad \mathbf{n} \times \mathbf{H} = \frac{4\pi}{c}\mathbf{K} \quad \mathbf{n} \cdot \mathbf{B} = 0 \quad \text{and} \quad \mathbf{n} \times \mathbf{E} = 0 \quad (2)$$

where the fields are those just outside of the conductor, and σ_q and \mathbf{K} are the charge and current density on the conductor's surface²; \mathbf{n} is the unit outward normal at the surface of the conductor. These relations are only approximate because we have neglected in particular the term $\partial\mathbf{B}/\partial t$ in Faraday's Law; it gives a correction of order $\omega\delta/c$ times the incident field's amplitude to the statement that the tangential electric field vanishes at the interface. To put it another way, the tangential component of the reflected wave's electric field actually differs from that of the incident wave's electric field by an amount of order $\omega\delta/c$ times the amplitude of the incident wave. In lowest order we ignore this difference. Outside of the conductor, $\partial\mathbf{B}/\partial t$ does not contribute to the integral since we assure the contour has negligible area here.

So far we don't know the surface charge and current densities, but the conditions that the tangential component of \mathbf{E} and the normal component of \mathbf{B} are zero at the interface are already enough to allow us to determine the reflected fields, given the incident ones. Hence we have at this point all of the information we need to obtain, to lowest order in the small parameter, the solution for the waves in the channel or cavity, i.e., the solution to the boundary-value problem posed above.

1.2 Power and Energy Loss

Before going on to look at that problem, however, let's look at the properties of the transmitted wave in the conductor. The reason for doing this is that we want to

²it is only appropriate to talk about surface currents or charges in the limit of a perfect conductor; otherwise, these densities will extend into the conductor to a finite extent

know how much energy is lost in the reflection process. To zero order in $\omega\delta/c$, none is lost, as is evident from the boundary condition which says that the amplitude of the reflected wave is the same as that of the incident wave. We must therefore look at the first-order corrections to this result, and that is most easily done by examining the transmitted wave.

From Faraday's Law and Ampère's Law, assuming a harmonic wave, we find that the fields in the conductor, which are identified by a subscript c , obey the relations

$$\mathbf{B}_c = -i\frac{c}{\omega}(\nabla \times \mathbf{E}_c) \quad \text{and} \quad \mathbf{E}_c = \frac{c}{4\pi\sigma\mu}(\nabla \times \mathbf{B}_c), \quad (3)$$

where we have ignored the displacement current term because it is of order ω/σ relative to the real current term; μ is the permeability of the conductor. Because the fields vary rapidly in the direction normal to the interface (length scale δ) and slowly in directions parallel to the interface (length scale λ), we may ignore spatial derivatives in all directions except the normal one. The conditions³ $\nabla \cdot \mathbf{B}_c = 0$ and $\nabla \cdot \mathbf{D}_c = 0$ tell us, to lowest order, that the fields within the conductor have no components normal to the interface. Taking the curl of the second of Eqs. (3), and using the first of these equations for the curl of \mathbf{E}_c , we find that

$$\left(i + \frac{c^2}{4\pi\sigma\mu\omega} \nabla^2\right) \mathbf{B}_c = 0, \quad (4)$$

or

$$\left(\nabla^2 + i\frac{2}{\delta^2}\right) \mathbf{B}_c = 0, \quad (5)$$

which has the solution

$$\mathbf{B}_c(z, t) = \mathbf{B}_{c0} e^{\kappa z} e^{-i\omega t} \quad (6)$$

where

$$\kappa = \pm(1 - i)/\delta. \quad (7)$$

³Of course, the electric displacement has a non-zero divergence if $\rho \neq 0$; as we saw in Jackson 7.7, any initial non-zero ρ dies out with some characteristic lifetime and so when the steady-state is established, $\rho = 0$.

Because the fields must vanish for $z \rightarrow \infty$, we have to choose the negative root and so find

$$\mathbf{B}_c = \mathbf{B}_{c0} e^{-z/\delta} e^{iz/\delta} e^{-i\omega t}. \quad (8)$$

Also, as is easily shown from this result and one of Eqs. (3),

$$\mathbf{E}_c = -\frac{c}{4\pi\sigma\mu\delta}(1-i)(\hat{\mathbf{z}} \times \mathbf{B}_c). \quad (9)$$

These fields are the same in form as the ones that arise in the case of normal incidence. The amplitudes are somewhat different from that case, however.

The power per unit area entering the conductor is

$$\begin{aligned} \langle \mathbf{S} \cdot \hat{\mathbf{z}} \rangle |_{z=0} &= \frac{c}{8\pi} \Re(\mathbf{E}_c \times \mathbf{H}_c^*) \cdot \hat{\mathbf{z}} |_{z=0} \\ &= -\frac{c}{8\pi} \Re \left\{ \frac{c}{4\pi\sigma\mu\delta} (1-i) [(\hat{\mathbf{z}} \times \mathbf{B}_c) \times (\mathbf{B}_c^*/\mu)] \cdot \hat{\mathbf{z}} |_{z=0} \right\} \\ &= \frac{c^2}{32\pi^2\sigma\mu^2\delta} \Re[(1-i)|\mathbf{B}_c|^2] |_{z=0}. \end{aligned} \quad (10)$$

However, $c^2/2\pi\sigma\omega\mu = \delta^2$, so, writing the power per unit area as \mathcal{P} , we have

$$\mathcal{P} = \frac{\mu\omega\delta}{16\pi} |H_c|^2 |_{z=0}. \quad (11)$$

We can relate H_c at $z = 0$ to the field at the interface on the outside of the conductor by employing an appropriate continuity condition. It is not the one derived above. This time, we use a value of t which is much smaller than δ so that there is only a negligible amount of current passing through the rectangle employed in applying Stokes' Theorem. Then we find that H_c at $z = 0$ is the same as the tangential component of the magnetic field on the outside. For definiteness, let the incident field be polarized perpendicular to the plane of incidence. Then the reflected field has an equal and opposite amplitude (to lowest order) and the sum of the incident and reflected waves' magnetic field amplitudes parallel to the interface is⁴ twice the

⁴We suppose that the exterior medium is vacuum, or at least has $\mu = \epsilon = 1$.

amplitude of the incident electric field times the cosine of the angle of incidence, or $2E_0 \cos \theta$.

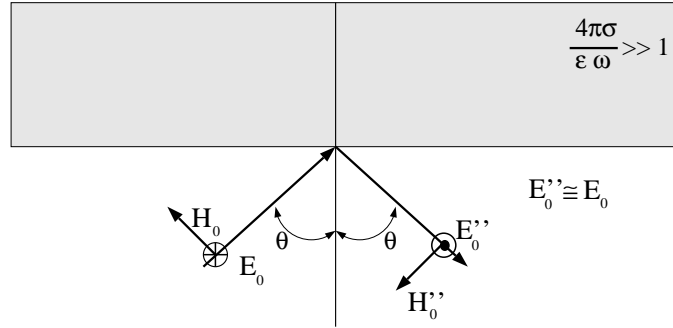


Fig.2: Wave polarized \perp to the plane, between vacuum and a good conductor.

Hence the power loss per unit area in the reflection process, meaning the power per unit area entering the conductor and so not reflected, is

$$\mathcal{P} = \frac{\mu\omega\delta}{4\pi} |E_0|^2 \cos^2 \theta. \quad (12)$$

The ratio of the lost to incident power, which is also the transmission coefficient, is

$$T = \frac{\mu\omega\delta |E_0|^2 \cos^2 \theta / 4\pi}{c |E_0|^2 \cos \theta / 8\pi} = \frac{2\mu\omega\delta}{c} \cos \theta. \quad (13)$$

This agrees with the result we found earlier in the case of a good conductor and for normal incidence, $\theta = 0$.

We will want to use the result for power loss later in connection with the attenuation of waves travelling along a wave guide. First we shall obtain the solution for the electromagnetic field within the waveguide in the limit of perfectly conducting walls.

2 Wave Guides

A waveguide is a hollow conducting pipe, perhaps filled with dielectric. It has a characteristic transverse size on the order of centimeters and is used to transmit electromagnetic energy (waves) from one place to another.

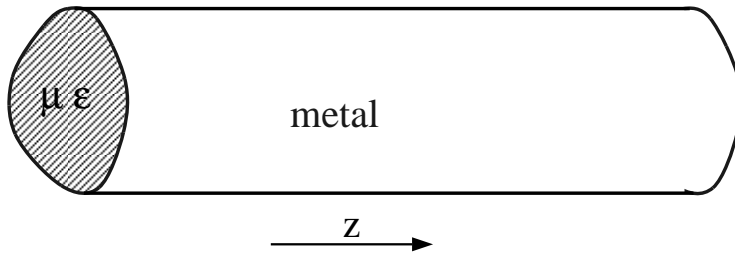


Fig.3: Wave guide of arbitrary cross-section.

The waves typically have frequencies such that the wavelength in vacuum would be comparable to the size of the waveguide. Thus $\omega = 2\pi c/\lambda$ is of order 10^{11} sec^{-1} . If the walls of the guide are constructed of a good conducting material, i.e., one with $\sigma \sim 10^{17} \text{ sec}^{-1}$, then we are in the good conducting limit so that the treatment of the previous section is valid. In particular, $T \sim 10^{-3}$ which means that some 10^3 reflections can take place before the wave is seriously attenuated. Also, we may adopt the boundary conditions that $\mathbf{E}_{tan} = 0 = B_n$ at the conducting surfaces.

2.1 Fundamental Equations

Let the wave guide have its long axis parallel to the z -direction and let its cross-section be invariant under translation along this direction. It is useful to divide operators, such as ∇ and ∇^2 , and also fields into components parallel and perpendicular to the long axis. Thus we write

$$\begin{aligned} \nabla &= \nabla_t + \epsilon_3 \frac{\partial}{\partial z}, & \nabla^2 &= \nabla_t^2 + \frac{\partial^2}{\partial z^2}, \\ \mathbf{E} &= \mathbf{E}_t + \epsilon_3 E_z, & \text{and} & \quad \mathbf{B} = \mathbf{B}_t + \epsilon_3 B_z. \end{aligned} \quad (14)$$

Further we shall assume that the fields' dependence on both z and t is harmonic,

$$\mathbf{B}(\mathbf{x}, t) = \mathbf{B}(x, y)e^{i(kz - \omega t)} \quad \text{and} \quad \mathbf{E}(\mathbf{x}, t) = \mathbf{E}(x, y)e^{i(kz - \omega t)}. \quad (15)$$

Given ω , we need to find k and the amplitudes $\mathbf{B}(x, y)$, $\mathbf{E}(x, y)$. Letting the material within the guide have dielectric constant ϵ and permeability μ , and assuming

no macroscopic sources in this region, we can derive wave equations using familiar methods. Because of the harmonic time dependences, the Maxwell equations read

$$\nabla \times \mathbf{E} = i\frac{\omega}{c}\mathbf{B} \quad \nabla \cdot \mathbf{B} = 0 \quad \nabla \cdot \mathbf{E} = 0 \quad \nabla \times \mathbf{B} = -i\frac{\omega}{c}\mu\epsilon\mathbf{E}. \quad (16)$$

Taking the curl of each curl equation and using the forms of the fields as well as the fact that both fields have zero divergence, we find that the wave equations for all Cartesian components of $\mathbf{E}(x, y)$ and $\mathbf{B}(x, y)$ have the same form; it is

$$\left(\nabla_t^2 - k^2 + \mu\epsilon\frac{\omega^2}{c^2}\right)\psi(x, y) = 0. \quad (17)$$

We can greatly simplify things by noting that if we find E_z and B_z first, then \mathbf{E}_t and \mathbf{B}_t follow. To demonstrate this statement, we shall derive explicit expressions for the latter in terms of the former. Consider the transverse components of the curl of the magnetic induction,

$$[\nabla \times (B_z\boldsymbol{\epsilon}_3 + \mathbf{B}_t)]_t = (\nabla_t B_z) \times \boldsymbol{\epsilon}_3 + \boldsymbol{\epsilon}_3 \times \left(\frac{\partial \mathbf{B}_t}{\partial z}\right) = -i\mu\epsilon\frac{\omega}{c}\mathbf{E}_t. \quad (18)$$

Cross $\boldsymbol{\epsilon}_3$ into this equation to find

$$\nabla_t B_z - \frac{\partial \mathbf{B}_t}{\partial z} = -i\mu\epsilon\frac{\omega}{c}(\boldsymbol{\epsilon}_3 \times \mathbf{E}_t). \quad (19)$$

By similar means, one finds that the transverse part of Faraday's Law can be written as

$$(\nabla_t E_z) \times \boldsymbol{\epsilon}_3 + \boldsymbol{\epsilon}_3 \times \left(\frac{\partial \mathbf{E}_t}{\partial z}\right) = i\frac{\omega}{c}\mathbf{B}_t. \quad (20)$$

Take the derivative with respect to z of the first of these equations and substituted the result into the second equation; the result is

$$\left(\mu\epsilon\frac{\omega^2}{c^2} + \frac{\partial^2}{\partial z^2}\right)\mathbf{B}_t = \nabla_t \left(\frac{\partial B_z}{\partial z}\right) + i\epsilon\mu\frac{\omega}{c}\boldsymbol{\epsilon}_3 \times (\nabla_t E_z). \quad (21)$$

Because the fields' z -dependence is e^{ikz} , we may take the derivatives with respect to z and express the result as an equation for \mathbf{B}_t :

$$\mathbf{B}_t = \frac{1}{\mu\epsilon\omega^2/c^2 - k^2} \left[\nabla_t \left(\frac{\partial B_z}{\partial z}\right) + i\mu\epsilon\frac{\omega}{c}\boldsymbol{\epsilon}_3 \times (\nabla_t E_z) \right]. \quad (22)$$

In the same fashion, we can take the derivative with respect to z of Eq. (20) and substitute into Eq. (19) to find a relation for \mathbf{E}_t ,

$$\mathbf{E}_t = \frac{1}{\mu\epsilon\omega^2/c^2 - k^2} \left[\nabla_t \left(\frac{\partial E_z}{\partial z} \right) - i\frac{\omega}{c} \boldsymbol{\epsilon}_3 \times (\nabla_t B_z) \right]. \quad (23)$$

Now we need only solve for the z components of the fields; from them and the preceding relations, all components are determined.

2.1.1 Boundary Conditions

The boundary condition on $E_z(x, y)$ is that it should vanish at the walls because the tangential component of the electric field is zero there. The other boundary condition, $B_n = 0$, does not put any constraint on B_z ; however, there is a constraint on B_z which can be extracted from the equation for \mathbf{E}_t ; one of the two components of \mathbf{E}_t is tangential to the wall and this one must vanish next to the wall. From Eq. (23) we see that there is a contribution to that component which is proportional to $\partial B_z/\partial n$, so we conclude that the boundary condition on B_z is $\partial B_z/\partial n = 0$. The other contribution to \mathbf{E}_t is proportional to the transverse gradient of E_z at the boundary; since E_z is zero at all points on the boundary, it is clear that this term will not give any tangential component of \mathbf{E}_t at the boundary. Hence the tangential components of \mathbf{E} are zero on the boundary provided E_z vanishes there along with the normal component of the gradient of B_z .

And what of the normal component of \mathbf{B} itself? From Eq. (22) for \mathbf{B}_t , we see that the the normal component of \mathbf{B} at the wall vanishes if, first, the gradient of B_z has zero normal component there, and, second, the component of $\nabla_t E_z$ parallel to the wall vanishes; these conditions are met if $E_z = 0$ and $\partial B_z/\partial n = 0$ everywhere on the boundary. Hence we are left with the following boundary-value, or eigenvalue, problem:

$$\left(\nabla_t^2 - k^2 + \mu\epsilon\frac{\omega^2}{c^2} \right) \begin{Bmatrix} E_z(x, y) \\ B_z(x, y) \end{Bmatrix} = 0, \quad (24)$$

with

$$E_z(x,y) = 0 \quad \text{and} \quad \partial B_z(x,y)/\partial n = 0 \quad (25)$$

on the boundary.

2.2 Transverse Modes

Depending on the geometry, it may or may not be possible to find an eigenvalue k^2 such that the conditions on B_z and E_z are both satisfied. If it is not possible, then either $B_z \equiv 0$ or $E_z \equiv 0$. In the former case, \mathbf{B} is purely transverse and one speaks of a *transverse magnetic mode*, often abbreviated as a TM mode; in the latter case, \mathbf{E} is purely transverse and the mode is called a *transverse electric mode*, or a TE mode. For some geometries it is possible to have both E_z and B_z identically zero although the transverse fields are finite; then we have a *transverse electromagnetic mode* or a TEM mode.

MODE	CHARACTER
TM (Transverse Magnetic)	$B_z = 0$
TE (Transverse Electric)	$E_z = 0$
TEM	$E_z = B_z = 0$

2.2.1 TEM Mode

Let's briefly discuss the TEM modes first. In order to see what are the appropriate equations of motion of the fields, we have to go back to the Maxwell equations. If we look just at the z -component of Faraday's Law and of the generalized Ampère's Law, we find that

$$\boldsymbol{\epsilon}_3 \cdot (\nabla_t \times \mathbf{E}_t) = 0 \quad \text{and} \quad \boldsymbol{\epsilon}_3 \cdot (\nabla_t \times \mathbf{B}_t) = 0. \quad (26)$$

Since ∇_t and \mathbf{E}_t have only x and y components, the curls lie entirely in the z direction, so we can write

$$\nabla_t \times \mathbf{E}_t = 0 \quad \text{and} \quad \nabla_t \times \mathbf{B}_t = 0. \quad (27)$$

From the other two Maxwell equations we find that

$$\nabla_t \cdot \mathbf{E}_t = 0 \quad \text{and} \quad \nabla_t \cdot \mathbf{B}_t = 0. \quad (28)$$

The two transverse fields are solutions of problems identical to statics problems in two dimensions. In particular, the transverse electric field, which must have zero tangential component at the walls of the waveguide, is found by solving an electrostatics problem. If the waveguide is composed of a single conductor, which is an equipotential in the equivalent electrostatics problem, then there is no nontrivial solution. *The conclusion is that a simple single-conductor waveguide cannot have a TEM mode to lowest order in the parameter $\delta\omega/c$.* Where are there TEM modes? These exist in guides which consist of a pair of parallel but electrically unconnected conductors that can be at different potentials. Such things are called *transmission lines*.

2.2.2 TE and TM Modes

Now let's turn to TE and TM modes. We will discuss here only TE modes; TM modes can be treated by making a simple modification of the boundary conditions when solving the eigenvalue problem. For TE modes, $E_z = 0$ and so the transverse fields are given in terms of B_z by

$$\mathbf{B}_t = \frac{1}{\gamma^2} \nabla_t \left(\frac{\partial B_z}{\partial z} \right) = \frac{ik}{\gamma^2} (\nabla_t B_z) \quad (29)$$

and

$$\mathbf{E}_t = -\frac{i}{\gamma^2} \frac{\omega}{c} \boldsymbol{\epsilon}_3 \times (\nabla_t B_z) = -\frac{\omega}{ck} (\boldsymbol{\epsilon}_3 \times \mathbf{B}_t), \quad (30)$$

where $\gamma^2 \equiv \mu\epsilon\omega^2/c^2 - k^2$. This parameter must be determined by solving the eigenvalue equation

$$(\nabla_t^2 + \gamma^2)B_z(x, y) = 0, \quad \text{with } \partial B_z/\partial n = 0 \text{ on } C; \quad (31)$$

C is the boundary of the (cross-section of the) waveguide. This problem will have solutions $B_z(x, y) \rightarrow B_i(x, y)$ with eigenvalues γ_i^2 , $i=1,2,\dots$. In terms of these

eigenvalues, the wavenumber k_i is given by

$$k_i^2(\omega) = \mu\epsilon\omega^2/c^2 - \gamma_i^2, \quad (32)$$

which means that for given ω and i , k is determined.

The eigenvalues γ_i^2 are always positive (otherwise the boundary conditions cannot be satisfied), so we can see from the preceding equation that for a given mode i , ω^2 must be larger than $\omega_i^2 \equiv \gamma_i^2 c^2 / \mu\epsilon$ in order for the squared wavenumber to be positive corresponding to a real wavenumber k . If ω^2 is smaller than this cutoff value, then the wavenumber is imaginary and the wave is attenuated as it moves in the z -direction. As a particular consequence, one can choose ω such that only some fixed number of modes (one, for example) can propagate.

There are two velocities of interest in connection with any mode; these are the phase velocity and the group velocity. The dispersion relation is

$$\omega^2 = \frac{c^2}{\mu\epsilon}(k^2 + \gamma_i^2), \quad (33)$$

which can be expressed also as

$$\frac{\omega^2}{k^2} = \frac{c^2/\mu\epsilon}{1 - \omega_i^2/\omega^2}. \quad (34)$$

From this form, one can see clearly that the phase velocity, ω/k , is always larger than the phase velocity in the absence of walls, $c/\sqrt{\epsilon\mu}$. Further, the phase velocity diverges as ω approaches the cutoff frequency. As for the group velocity, we have $v_g = d\omega/dk$, and

$$\omega \frac{d\omega}{dk} = \frac{c^2}{\mu\epsilon} k \quad \text{or} \quad \frac{\omega}{k} \frac{d\omega}{dk} = \frac{c^2}{\mu\epsilon} = v_p v_g. \quad (35)$$

This equation tells us that the product of the group and phase velocities is a constant, $c^2/\mu\epsilon$; the group velocity itself is

$$v_g = \frac{c^2}{\mu\epsilon} \frac{k}{\omega} = \frac{c}{\sqrt{\epsilon\mu}} \sqrt{1 - \frac{\omega_i^2}{\omega^2}} \quad (36)$$

which is always smaller than $c/\sqrt{\epsilon\mu}$.

2.3 Energy Flow

The significance of the group velocity of the mode becomes clear from a study of the energy flow in the guide. The time-average of the Poynting vector's z component is

$$\langle \mathbf{S} \cdot \boldsymbol{\epsilon}_3 \rangle = \frac{c}{8\pi} \Re(\mathbf{E}_t \times \mathbf{H}_t^*) \cdot \boldsymbol{\epsilon}_3 = \frac{c}{8\pi} \frac{\omega k}{\mu c \gamma^4} |\nabla_t B_z|^2. \quad (37)$$

This must be integrated over the cross-section of the guide to find the power transmitted,

$$\begin{aligned} \mathcal{P} &= \int_S d^2x \langle \mathbf{S} \cdot \boldsymbol{\epsilon}_3 \rangle = \frac{c}{8\pi} \frac{\omega k}{\mu c \gamma^4} \int_S d^2x |\nabla_t B_z|^2 \\ &= \frac{\omega k}{8\pi \mu \gamma^4} \int_S d^2x [\nabla_t \cdot (B_z^* \nabla_t B_z) - B_z^* \nabla_t^2 B_z]. \end{aligned} \quad (38)$$

The first term in the final expression converts to a surface integral which is shown, from the boundary conditions on B_z , to be zero; the second term is made simpler in appearance by using the fact that $\nabla_t^2 B_z = -\gamma^2 B_z$. Thus,

$$\mathcal{P} = \frac{\omega k}{8\pi \mu \gamma^2} \int_S d^2x |B_z|^2. \quad (39)$$

Compare this with the time-averaged energy per unit length in the guide

$$\begin{aligned} U &= \frac{1}{16\pi} \int_S d^2x \left(\epsilon \mathbf{E}_t \cdot \mathbf{E}_t^* + \frac{1}{\mu} [\mathbf{B}_t \cdot \mathbf{B}_t^* + B_z B_z^*] \right) \\ &= \frac{1}{16\pi} \int_S d^2x \left[\epsilon \frac{\omega^2}{c^2 k^2} |\mathbf{B}_t|^2 + \frac{1}{\mu} |\mathbf{B}_t|^2 + \frac{1}{\mu} |B_z|^2 \right] \\ &= \frac{1}{16\pi \mu} \int_S d^2x \left[\left(\epsilon \mu \frac{\omega^2}{c^2 \gamma^4} + \frac{k^2}{\gamma^4} \right) |\nabla_t B_z|^2 + |B_z|^2 \right] \\ &= \frac{1}{16\pi \mu} \int_S d^2x \left[\left(\epsilon \mu \frac{\omega^2}{c^2} + k^2 \right) \frac{1}{\gamma^2} + 1 \right] |B_z|^2 = \frac{1}{8\pi} \left(\epsilon \frac{\omega^2}{c^2} \right) \frac{1}{\gamma^2} \int_S d^2x |B_z|^2. \end{aligned} \quad (40)$$

In arriving at the final result, we've used a whole collection of identities related to the eigenvalue problem.

Comparison of U and \mathcal{P} shows that

$$\frac{\mathcal{P}}{U} = \frac{k}{\omega} \frac{c}{\mu \epsilon} \equiv v_g. \quad (41)$$

The obvious interpretation need not be stated.

2.3.1 TE Modes in Rectangular and Circular Guides

Before going on to other matters, let us look at the solutions of the eigenvalue problem for some standard waveguide shapes, i.e., rectangles and circles, assuming TE modes.

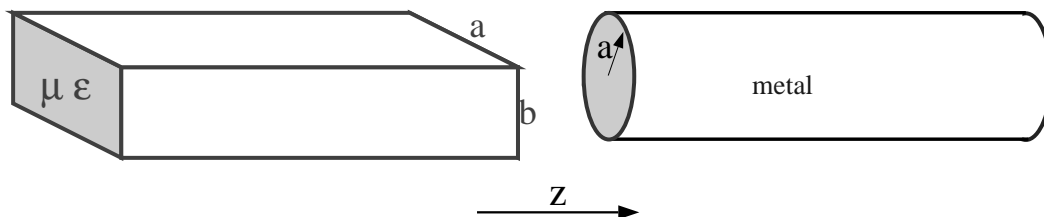


Fig.4: Geometry of Rectangular and Circular Wave Guides.

For the rectangle shown, the solution⁵ for B_z is

$$B_{mn}(x, y) = B_0 \cos\left(\frac{m\pi x}{a}\right) \cos\left(\frac{n\pi y}{b}\right) \quad (42)$$

with

$$\gamma_{mn}^2 = \pi^2 \left(\frac{m^2}{a^2} + \frac{n^2}{b^2} \right). \quad (43)$$

For the circular guide of radius a , on the other hand, B_z becomes

$$B_{mn}(\rho, \phi) = B_0 e^{im\phi} J_m(y_{mn}\rho/a) \quad (44)$$

with

$$\gamma_{mn}^2 = y_{mn}^2/a^2. \quad (45)$$

Here, y_{mn} is the n^{th} zero of the derivative of the Bessel function J_m , $J'_m(y)|_{y_{mn}} = 0$.

The question of which modes will actually be excited in a waveguide for a given source can be worked out (see below); one has to address the question of how the source couples to the eigenfunctions for the different modes. Different configurations

⁵The eigenvalue problem for TE modes is formally equivalent to that of a quantum mechanical particle in a box with somewhat unusual boundary conditions; the case of the TM mode has the usual boundary conditions.

of the source will produce different superpositions of modes. A simple way to guarantee that just one propagating mode will be present is to pick ω smaller than the cutoff frequencies of all modes but one.

3 Attenuation of Modes in Waveguides

Even modes for which the wave number is real will be somewhat damped because of the finite conductivity of the walls. We have seen that the power travelling down the pipe is given by Eq. (39); also, the power lost per unit length is, from Eq. (11),

$$\frac{d\mathcal{P}}{dz} = -\frac{\mu\omega\delta}{16\pi} \oint_C dl |\mathbf{H}_{\parallel}|^2 \quad (46)$$

where \mathbf{H}_{\parallel} is the component of \mathbf{B}/μ which is parallel to the boundary, and the integral is evaluated on the contour formed by the cross-section of the guide. We can evaluate $|\mathbf{H}_{\parallel}|^2$ up to a point. First, for TE modes

$$|\mathbf{H}_{\parallel}|^2 = |\mathbf{n} \times \mathbf{H}|^2 = \frac{1}{\mu^2} (|B_z|^2 + |\mathbf{n} \times \mathbf{B}_t|^2) = \frac{1}{\mu^2} \left[|B_z|^2 + \frac{k^2}{\gamma^4} |\mathbf{n} \times (\nabla_t B_z)|^2 \right]. \quad (47)$$

Further, at the surface $|\mathbf{n} \times (\nabla_t B_z)|^2 = |\nabla_t B_z|^2$ (since $\partial B_z / \partial n = 0$ there). The latter can be expected to be comparable to $|B_z^* \nabla_t^2 B_z| = \gamma^2 |B_z|^2$, so we write

$$|\nabla_t B_z|^2 = \xi \gamma^2 |B_z|^2, \quad (48)$$

where ξ is a mode-dependent dimension-free constant of order unity that is independent of the frequency⁶. Hence,

$$|\mathbf{n} \times \mathbf{H}|^2 = \frac{1}{\mu^2} \left(1 + \xi \frac{k^2}{\gamma^2} \right) |B_z|^2 \quad (49)$$

and

$$\frac{d\mathcal{P}}{dz} = -\frac{\mu\omega\delta}{16\pi} \frac{1}{\mu^2} \left(1 + \xi \frac{k^2}{\gamma^2} \right) \oint_C dl |B_z|^2 \quad (50)$$

⁶For a given geometry of the guide, and a particular mode therein, one may easily calculate ξ .

This is to be compared with the power itself which is obtained by integrating $|B_z|^2$ over a cross-section of the guide. We convert the integral in the preceding equation into one over the cross-section by noticing that $|B_z|$ behaves on the boundary in much the same way as in the interior of the guide, so that

$$\oint_C dl |H_z|^2 = \frac{C}{A} \eta \int_S d^2x |H_z|^2 \quad (51)$$

where A and C are the cross-sectional area and circumference of the guide and η is another dimension-free constant of order unity; it depends on the shape of the guide and on the mode but not on the frequency.

Substituting this expression into Eq. (50) and dividing by the power, given by Eq. (29), we find an attenuation coefficient β ,

$$\beta \equiv - \frac{d\mathcal{P}}{dz} / \mathcal{P} = \frac{\delta}{2k} \eta \frac{C}{A} (\gamma^2 + \xi k^2). \quad (52)$$

Substitute for k using $k^2 = \epsilon\mu\omega^2/c^2 - \gamma^2$, and replace γ using $\gamma^2 = \epsilon\mu\omega_i^2/c^2$; ω_i is the cutoff frequency of the mode. Then we find

$$\beta = \frac{\sqrt{\epsilon\mu}}{2} \sqrt{\frac{\omega_i}{\omega}} \frac{\omega_i}{\sqrt{\omega^2 - \omega_i^2}} \frac{\eta C}{A} \frac{\delta_i \omega_i}{c} \left[1 + \xi \left(\frac{\omega^2 - \omega_i^2}{\omega_i^2} \right) \right]; \quad (53)$$

δ_i is the penetration depth at cutoff. All of the frequency dependence of β is explicit in this result. The damping becomes very large as $\omega \rightarrow \omega_i$; for ω not too close to the cutoff for mode i , and for $\sigma \sim 10^{17} \text{ sec}^{-1}$, we can see that the wave can travel some hundreds of meters without disastrous attenuation. At very high frequencies, $\omega \gg \omega_i$, the attenuation increases once again.

The usefulness of our result is limited; in particular, it breaks down when the frequency approaches the cutoff frequency for the mode. One can do a better calculation by improving the treatment of the boundary conditions. The solution (for the fields) that we have found satisfies boundary conditions only slightly different from the correct ones; it turns out, not surprisingly, that an improved solution can be obtained by looking for small corrections to the fields we already have, with the

corrections determined by demanding that the exact boundary conditions be satisfied to one higher order in ω/σ . This is a straightforward but somewhat technical calculation and we shall not spend time on it.

4 Resonating cavities

The step from a waveguide to a resonating cavity is not a large one. We need only think about what new constraints are placed on the electromagnetic fields if end walls are placed on a waveguide. Suppose that such walls are introduced at $z = 0$ and $z = d$. Then we cannot have traveling waves in the guide but must have instead standing waves; that is, in addition to the fields varying as $e^{i(kz-\omega t)}$, we must have ones that vary as $e^{-i(kz+\omega t)}$ which means that the z and t -dependent parts of the fields can be expressed as some linear combination of $\sin(kz)e^{i\omega t}$ and $\cos(kz)e^{i\omega t}$. Using such a combination, one may proceed as in the previous sections and will find in particular that $E_z(x, y)$ and $B_z(x, y)$ satisfy the same eigenequations as before.

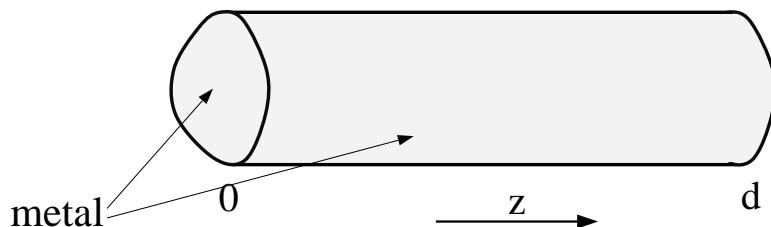


Fig.5: Geometry of a Resonant Cavity.

What is different in a cavity is that there are new boundary conditions or constraints on the fields because of the presence of end walls. If $B_z \neq 0$, then one must add the condition that $B_z = 0$ at $z = 0$ and $z = d$. Also, in order to guarantee that the tangential component of \mathbf{E} , (or \mathbf{E}_t) vanish on the end walls, it must be the case that $\partial E_z/\partial z = 0$ on the end walls; see Eq. (23). These are the additional conditions that must be satisfied for a cavity.

Following tradition established in earlier sections, we examine only TE modes.

Then B_z must vanish at the ends of the cavity, meaning that the solution we want is the one having z -dependence $\sin(kz)$; this is zero at $z = 0$, and we make it zero at $z = d$ by choosing $k = p\pi/d$ with p an integer. It is also true, as for the waveguide, that k must satisfy the equation $k^2 = \mu\epsilon\omega^2/c^2 - \gamma_{mn}^2$. Both of these conditions on k^2 can be satisfied simultaneously only for certain discrete frequencies ω_{mnp} which are given by

$$\omega_{mnp}^2 = \frac{c^2}{\mu\epsilon} \left(\frac{p^2\pi^2}{d^2} + \gamma_{mn}^2 \right). \quad (54)$$

Hence a resonating cavity has a set of discrete natural “resonant” frequencies at which it can support a standing wave electromagnetic field. These frequencies can be tuned by adjusting the size of the cavity, e.g., by changing d .

A resonant cavity is useful in that if excited in a single mode, it will contain monochromatic radiation in the microwave frequency range (i.e. a maser), or would do so if it were perfect. However, the resonance is never perfectly sharp in frequency, meaning that if one Fourier transforms the fields into frequency space (instead of time) the result will not be a delta function at the resonant frequency. There are several contributing factors to the width of the resonance. One important factor is the power loss in the walls. This loss is generally characterized by the “ Q ” of the cavity defined by

$$Q \equiv \omega_0 \left(\frac{\text{stored energy}}{\text{power loss}} \right) \quad (55)$$

where ω_0 is the frequency of the mode in the cavity. In words, Q is 2π times the energy stored in the cavity divided by the energy loss per cycle. From this definition it follows that the connection between Q and the rate at which the energy stored in the cavity decays is $dU/dt = -U\omega_0/Q$ when no new energy is being pumped into the cavity. If Q is independent of the amount of energy in the cavity, then this differential equation has a simple exponential solution,

$$U = U_0 e^{-\omega_0 t/Q}, \quad (56)$$

and, under these circumstances, any field⁷ ψ behaves in time as

$$\psi(t) = \psi_0 e^{-i\omega_0 t} e^{-\omega_0 t/2Q}. \quad (57)$$

Because of the decay, the frequency spectrum of the field contains components in addition to ω_0 . Specifically,

$$\psi(\omega) \sim \int_0^\infty dt e^{i(\omega - \omega_0)t} e^{-\omega_0 t/2Q} = \frac{1}{i(\omega_0 - \omega) + \omega_0/2Q}, \quad (58)$$

or

$$|\psi(\omega)|^2 \sim \frac{1}{(\omega_0 - \omega)^2 + \omega_0^2/4Q^2} \quad (59)$$

which means that the resonance has a width in frequency space which is of order $\omega_0/2Q$.

And what is the value of Q ? The energy (or fields) in the cavity decay in time because of losses in the walls. We learned in section 1 how to compute these losses. For any given mode in a particular cavity it is a straightforward matter to do the calculations, provided one can solve for the fields in that mode. One proceeds in much the same way as in the previous section where we learned how to calculate the power loss in a waveguide. Skipping over the details of the argument, which are much like the calculation of β for a mode of a waveguide, we simply state the conclusion which is that the energy lost per period is of the order of the energy in the cavity times the ratio of the volume of the walls into which the field penetrates to the volume of the cavity. If the area of the walls is A and the volume of the cavity is V , then

$$Q \sim \frac{V}{A\delta}. \quad (60)$$

Thus the relevant parameter for determining the Q of the cavity is the ratio of the penetration depth to the linear size of the cavity. For δ on the order of microns, or 10^{-4} cm , and a cavity having a size on the order of a centimeter, the Q will be on the order of 10^3 to 10^4 .

⁷That is, any component of \mathbf{E} or \mathbf{B} .

Chapter Nine

Radiation

Heinrich Rudolf Hertz
(1857 - 1894)

October 12, 2001

Contents

1	Introduction	1
2	Radiation by a localized source	3
2.1	The Near Zone	6
2.2	The Radiation or Far Zone	6
3	Multipole Expansion of the Radiation Field	9
3.1	Electric Dipole	10
3.1.1	Example: Linear Center-Fed Antenna	13
3.2	Magnetic Dipole	14
3.3	Comparison of Dipoles	16
3.4	Electric Quadrupole	18
3.4.1	Example: Oscillating Charged Spheroid	21
3.5	Large Radiating Systems	23
3.5.1	Example: Linear Array of Dipoles	24

4	Multipole expansion of sources in waveguides	26
4.1	Electric Dipole	26
5	Scattering of Radiation	31
5.1	Scattering of Polarized Light from an Electron	31
5.2	Scattering of Unpolarized Light from an Electron	34
5.3	Elastic Scattering From a Molecule	35
5.3.1	Example: Scattering Off a Hard Sphere	38
5.3.2	Example: A Collection of Molecules	40
6	Diffraction	43
6.1	Scalar Diffraction Theory: Kirchoff Approximation	45
6.2	Babinet's Principle	51
6.3	Fresnel and Fraunhofer Limits	53
7	Example Problems	55
7.1	Example: Diffraction from a Rectangular Aperture	55
7.2	Example: Diffraction from a Circular Aperture	57
7.3	Diffraction from a Cross	60
7.4	Radiation from a Reciprocating Disk	61

1 Introduction

An electromagnetic wave, or electromagnetic **radiation**, has as its sources electric accelerated charges in motion. We have learned a great deal about waves but have not given much thought to the connection between the waves and the sources that produce them. That oversight will be rectified in this chapter.

The **scattering** of electromagnetic waves is produced by bombarding some object (the scatterer) with an electromagnetic wave. Under the influence of the fields in that

wave, charges in the scatterer will be set into some sort of coherent motion¹ and these moving charges will produce radiation, called the scattered wave. Hence scattering phenomena are closely related to radiation phenomena.

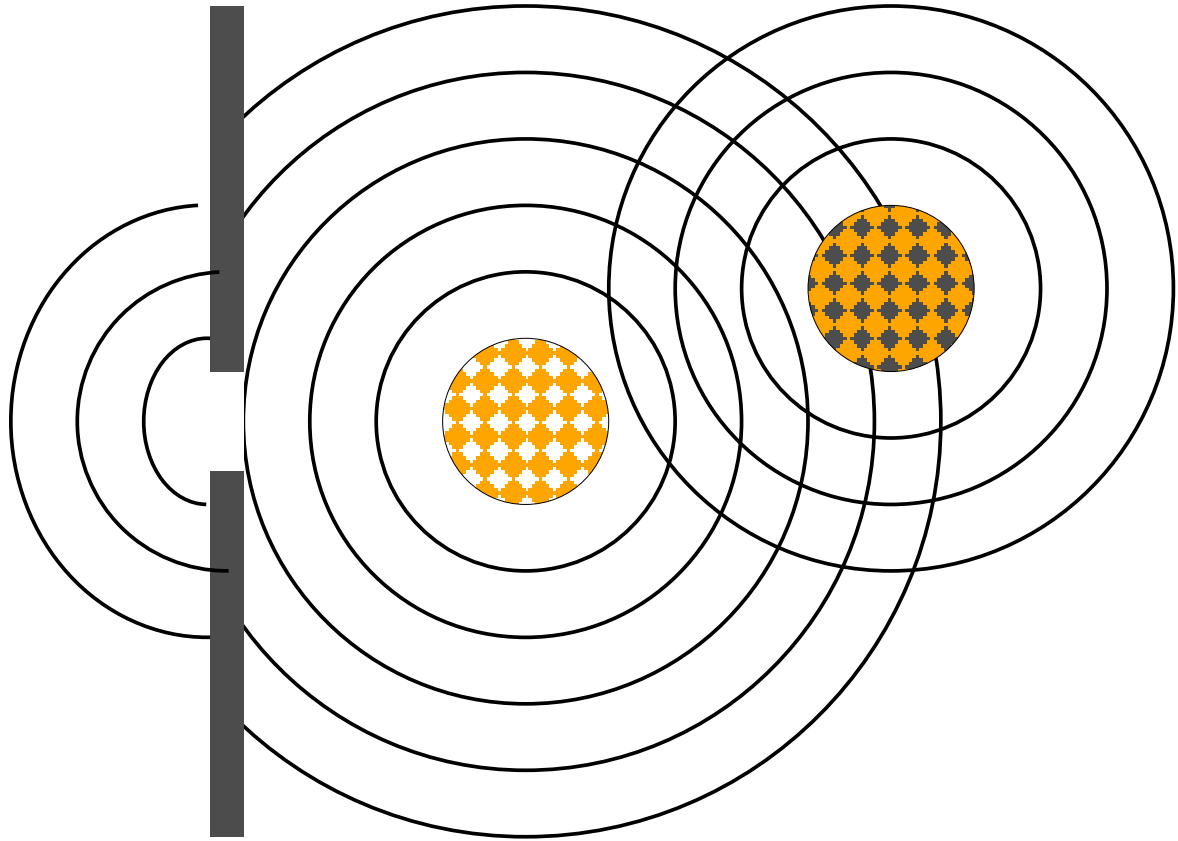
Diffraction of electromagnetic waves is similar. One starts with a wave incident on an opaque screen with holes, or aperture, in it. Charges in the screen, especially around the apertures, are set in motion and produce radiation which in this case is called the diffracted wave.

Thus radiation, scattering, and diffraction are closely related. We shall start our investigation by considering the radiation produced by some specified localized distribution of charges and currents in harmonic motion. We delay until Chap. 14, the discussion of non-harmonic sources.

¹The response to a harmonic excitation is of the same frequency, and thus coherent

Diffraction

Scattering



2 Radiation by a localized source

Suppose that we are given some charge and current densities $\rho(\mathbf{x}, t)$ and $\mathbf{J}(\mathbf{x}, t)$ ². These produce potentials which, in the Lorentz gauge (Chap. 6), can be found immediately using the retarded Green's function $G^{(+)}(\mathbf{x}, t; \mathbf{x}', t')$ which we shall write simply as $G(\mathbf{x}, t; \mathbf{x}', t')$:

$$\mathbf{A}(\mathbf{x}, t) = \frac{1}{c} \int d^3x' dt' G(\mathbf{x}, t; \mathbf{x}', t') \mathbf{J}(\mathbf{x}', t') \quad (1)$$

²In this chapter, we assume $\epsilon = \mu = 1$

$$\Phi(\mathbf{x}, t) = \int d^3x' dt' G(\mathbf{x}, t; \mathbf{x}', t') \rho(\mathbf{x}', t'). \quad (2)$$

The Green's function itself is given by

$$G(\mathbf{x}, t; \mathbf{x}', t') = \frac{\delta(t - t' - |\mathbf{x} - \mathbf{x}'|/c)}{|\mathbf{x} - \mathbf{x}'|}. \quad (3)$$

Because all of the equations we shall use to compute fields are linear in the fields themselves, we may conveniently treat just one Fourier component (frequency) of the field at a time. To this end we write

$$\mathbf{J}(\mathbf{x}, t) = \frac{1}{2} \int_{-\infty}^{\infty} d\omega \mathbf{J}(\mathbf{x}, \omega) e^{-i\omega t} \quad (4)$$

where

$$\mathbf{J}(\mathbf{x}, -\omega) = \mathbf{J}^*(\mathbf{x}, \omega) \quad (5)$$

is required in order that $\mathbf{J}(\mathbf{x}, t)$ be real; Eq. (5) is known as a “reality condition.” We may equally well, and more conveniently, replace Eqs. (4) and (5) by

$$\mathbf{J}(\mathbf{x}, t) = \Re \int_0^{\infty} d\omega \mathbf{J}(\mathbf{x}, \omega) e^{-i\omega t}. \quad (6)$$

We will do this and will in general not bother to write \Re in front of every complex expression whose real part must be taken. We will just have to remember that the real part is the physically meaningful quantity. Similarly, we introduce the Fourier transform of the charge density,

$$\rho(\mathbf{x}, t) = \int_0^{\infty} d\omega \rho(\mathbf{x}, \omega) e^{-i\omega t}. \quad (7)$$

In the following we shall suppose that the sources have just a single frequency component,

$$\mathbf{J}(\mathbf{x}, \omega') = \mathbf{J}(\mathbf{x}) \delta(\omega - \omega'), \quad \omega' > 0 \quad (8)$$

$$\rho(\mathbf{x}, \omega') = \rho(\mathbf{x}) \delta(\omega - \omega'), \quad \omega' > 0. \quad (9)$$

Thus, assuming $\omega > 0$,

$$\mathbf{J}(\mathbf{x}, t) = \mathbf{J}(\mathbf{x}) e^{-i\omega t} \quad \text{and} \quad (10)$$

$$\rho(\mathbf{x}, t) = \rho(\mathbf{x}) e^{-i\omega t}. \quad (11)$$

From the continuity condition on the sources,

$$\frac{\partial \rho(\mathbf{x}, t)}{\partial t} + \nabla \cdot \mathbf{J}(\mathbf{x}, t) = 0, \quad (12)$$

we find that $\rho(\mathbf{x})$ and $\mathbf{J}(\mathbf{x})$ are related by

$$\rho(\mathbf{x}) = -i \frac{\nabla \cdot \mathbf{J}(\mathbf{x})}{\omega}. \quad (13)$$

Using Eq. (10) in Eq. (1) and employing Eq. (3) for the Green's function, we find, upon completing the integration over the time, that

$$\mathbf{A}(\mathbf{x}, t) = \frac{1}{c} \int d^3 x' \frac{\mathbf{J}(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|} e^{ik|\mathbf{x} - \mathbf{x}'|} e^{-i\omega t} \quad (14)$$

where, as usual, $k \equiv \omega/c$. Define $\mathbf{A}(\mathbf{x})$ by

$$\mathbf{A}(\mathbf{x}, t) = \mathbf{A}(\mathbf{x}) e^{-i\omega t}; \quad (15)$$

comparison with Eq. (14) gives

$$\mathbf{A}(\mathbf{x}) = \frac{1}{c} \int d^3 x' \mathbf{J}(\mathbf{x}') \frac{e^{ik|\mathbf{x} - \mathbf{x}'|}}{|\mathbf{x} - \mathbf{x}'|}. \quad (16)$$

From here the recipe is to find $\mathbf{B}(\mathbf{x}, t)$ from the curl of $\mathbf{A}(\mathbf{x}, t)$; then the electric field is found³ from $\nabla \times \mathbf{B}(\mathbf{x}, t) = c^{-1} \partial \mathbf{E}(\mathbf{x}, t) / \partial t$, which holds in regions where the current density vanishes. These fields have the forms

$$\mathbf{B}(\mathbf{x}, t) = \mathbf{B}(\mathbf{x}) e^{-i\omega t} \quad \mathbf{E}(\mathbf{x}, t) = \mathbf{E}(\mathbf{x}) e^{-i\omega t} \quad (17)$$

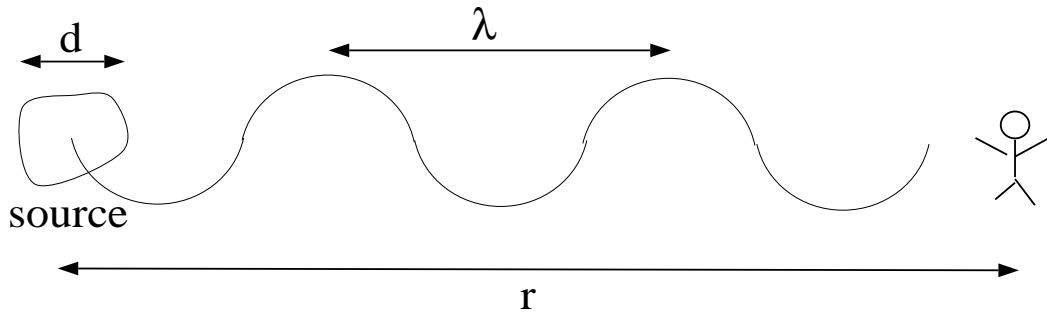
where

$$\mathbf{B}(\mathbf{x}) = \nabla \times \mathbf{A}(\mathbf{x}) \quad \mathbf{E}(\mathbf{x}) = \frac{i}{k} \nabla \times \mathbf{B}(\mathbf{x}). \quad (18)$$

We have reduced everything to a set of straightforward calculations - integrals and derivatives. Doing them exactly can be tedious, so we should spend some time

³Notice that we never have to evaluate the scalar potential.

thinking about whether there are any **simplifying approximations** that may have general validity or at least validity in some cases of interest. There are approximations based on expansions in powers of some small parameter. We can see what may be possible by realizing that there are *three relevant lengths* in any radiating system. Provided the origin of coordinates is taken to be somewhere within the source in the integrals above, these are the size of the source, $r' = |\mathbf{x}'|$; the distance of the observer from the source, represented by $r = |\mathbf{x}|$; and the wavelength of the radiation, $\lambda = 2\pi/k$. The magnitude r' is never larger than d , the size of the source. Focusing on just the relative size of λ and r , we identify the three traditional regimes below.



$d \ll r \ll \lambda$	Near or static zone
$d \ll r \sim \lambda$	Intermediate or induction zone
$d \ll \lambda \ll r$	Far or radiation zone

Here we have specified also that d be much smaller than the other two lengths. That simplifies the discussion of the three regimes and so is a convenience but it is not always met in practice nor is it always necessary. In particular, the fields far away from the source (in the radiation zone) have characteristic forms independent of the relative size of λ and d provided r is large enough. Also, man-made sources such as antennas (and antenna arrays) are often intentionally constructed to have $\lambda \sim d$ and even $\lambda \ll d$ in which case the inequalities above are not always satisfied. On the other hand, natural radiating systems, such as atoms and nuclei, typically do satisfy

the condition $d \ll \lambda$ and $d \ll r$ for any r at which it is practical to detect the radiation.

2.1 The Near Zone

Consider first the **near zone**. Here $d \ll \lambda$ and $r \ll \lambda$, so there is a simple expansion of the exponential factor,

$$e^{ik|\mathbf{x}-\mathbf{x}'|} = 1 + ik|\mathbf{x}-\mathbf{x}'| + \dots \quad (19)$$

which leads to

$$\mathbf{A}(\mathbf{x}, t) = \frac{1}{c} \int d^3x' \frac{\mathbf{J}(\mathbf{x}')}{|\mathbf{x}-\mathbf{x}'|} [1 + ik|\mathbf{x}-\mathbf{x}'| + \dots] e^{-i\omega t}. \quad (20)$$

The leading term in this expansion is

$$\mathbf{A}(\mathbf{x}, t) = \frac{1}{c} \int d^3x' \frac{\mathbf{J}(\mathbf{x}')}{|\mathbf{x}-\mathbf{x}'|} e^{-i\omega t}. \quad (21)$$

Aside from the harmonic time dependence, this is just the vector potential of a static current distribution $\mathbf{J}(\mathbf{x})$, and that is the origin of the name “static” zone; the magnetic induction here has a spatial dependence which is the same as what one would find for a static current distribution with the spatial dependence of the actual oscillating current distribution. We find this result for the simple reason that in the near zone the exponential factor can be approximated as unity.

2.2 The Radiation or Far Zone

In the **radiation or far zone** ($r \gg \lambda \gg d$), the story is completely different because in this regime the behavior of the exponential dominates the integral. We can most easily see what will happen if we first expand the displacement $|\mathbf{x}-\mathbf{x}'|$ in powers of r'/r (d/r):

$$\begin{aligned} |\mathbf{x}-\mathbf{x}'| &= [(\mathbf{x}-\mathbf{x}') \cdot (\mathbf{x}-\mathbf{x}')]^{1/2} = (r^2 - 2\mathbf{x} \cdot \mathbf{x}' + r'^2)^{1/2} \\ &= r \left[1 - \frac{2\mathbf{x} \cdot \mathbf{x}'}{r^2} + \left(\frac{r'}{r}\right)^2 \right]^{1/2} = r \left[1 - 2\frac{\mathbf{n} \cdot \mathbf{x}'}{r} + \left(\frac{r'}{r}\right)^2 \right]^{1/2} \end{aligned} \quad (22)$$

where $\mathbf{n} = \mathbf{x}/r$ is a unit vector in the direction of \mathbf{x} . Carrying the expansion to second order in r'/r , we have

$$|\mathbf{x} - \mathbf{x}'| = r \left[1 - \frac{\mathbf{n} \cdot \mathbf{x}'}{r} + \frac{1}{2} \left(\frac{r'}{r} \right)^2 - \frac{1}{2} \left(\frac{\mathbf{n} \cdot \mathbf{x}'}{r} \right)^2 \right]. \quad (23)$$

Similarly,

$$\frac{1}{|\mathbf{x} - \mathbf{x}'|} = \frac{1}{r} \left\{ 1 + \frac{\mathbf{n} \cdot \mathbf{x}'}{r} + \frac{1}{2} \left(\frac{r'}{r} \right)^2 \left[3 \left(\frac{\mathbf{n} \cdot \mathbf{x}'}{r'} \right)^2 - 1 \right] \right\}. \quad (24)$$

Using the first of these expansions we can write

$$\begin{aligned} e^{ik|\mathbf{x}-\mathbf{x}'|} &= e^{ikr} e^{-ik(\mathbf{n}\cdot\mathbf{x}')} e^{i\frac{kr}{2} \left[\left(\frac{r'}{r} \right)^2 - \left(\frac{\mathbf{n}\cdot\mathbf{x}'}{r} \right)^2 \right]} \\ &= e^{ikr} e^{-ik(\mathbf{n}\cdot\mathbf{x}')} e^{i\frac{kr'^2}{2r} \left[1 - \frac{(\mathbf{n}\cdot\mathbf{x}')^2}{r'^2} \right]}. \end{aligned} \quad (25)$$

Given $r \gg r'$ and $kr'^2/r \ll 1$, it is clear that this exponential function can be approximated by just the first two factors; the third represents a change of phase by an amount small compared to a radian. Further, in the far zone it is sufficient to approximate

$$\frac{1}{|\mathbf{x} - \mathbf{x}'|} = \frac{1}{r}. \quad (26)$$

Putting these pieces together we have

$$\mathbf{A}(\mathbf{x}, t) = \frac{e^{i(kr-\omega t)}}{cr} \int d^3x' \mathbf{J}(\mathbf{x}') e^{-ik(\mathbf{n}\cdot\mathbf{x}')} \quad (r \gg d \text{ and } r \gg d^2/\lambda). \quad (27)$$

This expression is always valid for r “large enough” which means $r \gg r'$ and $r \gg kr'^2$. The relative size of λ and d is unimportant.

The behavior of $\mathbf{A}(\mathbf{x}, t)$ on r and t is explicitly given by the factor in front of the integral; the integral depends on the direction of \mathbf{x} but not on its magnitude. Hence in the far zone the vector potential always takes the form

$$\mathbf{A}(\mathbf{x}, t) = \frac{e^{i(kr-\omega t)}}{r} \mathbf{f}(\theta, \phi) \quad (28)$$

where

$$\mathbf{f}(\theta, \phi) \equiv \frac{1}{c} \int d^3x' \mathbf{J}(\mathbf{x}') e^{-ik(\mathbf{n}\cdot\mathbf{x}')}; \quad (29)$$

θ and ϕ specify, in polar coordinates, the direction of \mathbf{x} (or \mathbf{n}).

Knowing the form of the potential so precisely makes it easy to see what must be the form of the fields \mathbf{E} and \mathbf{B} in the far zone. First,

$$\mathbf{B}(\mathbf{x}, t) = \nabla \times \mathbf{A}(\mathbf{x}, t) = \nabla \times \left(\frac{e^{i(kr - \omega t)}}{r} \mathbf{f}(\theta, \phi) \right) \approx ik \frac{e^{i(kr - \omega t)}}{r} [\mathbf{n} \times \mathbf{f}(\theta, \phi)] \quad (30)$$

where we have discarded terms of relative order λ/r or d/r . Further, from Eqs. (18) and (30), we can find $\mathbf{E}(\mathbf{x}, t)$; to the same order as \mathbf{B} , it is

$$\mathbf{E}(\mathbf{x}, t) = -ik \frac{e^{i(kr - \omega t)}}{r} [\mathbf{n} \times (\mathbf{n} \times \mathbf{f}(\theta, \phi))] = \mathbf{B}(\mathbf{x}, t) \times \mathbf{n}. \quad (31)$$

There are two essential features of these equations.

- First, both \mathbf{E} and \mathbf{B} in the radiation zone is that the field strengths are proportional to r^{-1} ; this is very different from the case for static fields which fall off at least as fast as r^{-2} (consider the static zone).
- Second, the radiation fields are transverse, meaning they are perpendicular to \mathbf{x} or \mathbf{n} ; they are also perpendicular to each other.

The Poynting vector in the far zone also has a simple basic form:

$$\begin{aligned} \langle \mathbf{S} \rangle &= \frac{c}{8\pi} (\mathbf{E} \times \mathbf{B}^*) = \frac{c}{8\pi} \left(\frac{-ik}{r} \right)^2 [\mathbf{n} \times (\mathbf{n} \times \mathbf{f}(\theta, \phi))] \times [\mathbf{n} \times \mathbf{f}^*(\theta, \phi)] \\ &= -\frac{ck^2}{8\pi r^2} \{ [\mathbf{n}(\mathbf{n} \cdot \mathbf{f}(\theta, \phi)) - \mathbf{f}(\theta, \phi)] \times [\mathbf{n} \times \mathbf{f}^*(\theta, \phi)] \} \\ &= -\frac{ck^2}{8\pi r^2} \{ (\mathbf{n} \cdot \mathbf{f}(\theta, \phi)) [\mathbf{n}(\mathbf{n} \cdot \mathbf{f}^*(\theta, \phi)) - \mathbf{f}^*(\theta, \phi)] - \mathbf{n} |\mathbf{f}(\theta, \phi)|^2 + \mathbf{f}^*(\theta, \phi) (\mathbf{n} \cdot \mathbf{f}(\theta, \phi)) \} \\ &= \frac{ck^2}{8\pi r^2} \mathbf{n} [|\mathbf{f}(\theta, \phi)|^2 - |\mathbf{n} \cdot \mathbf{f}(\theta, \phi)|^2]. \end{aligned} \quad (32)$$

This is presumably the time-averaged energy current density. Because it points radially outward⁴, it also gives directly the angular distribution of radiated power:

$$\frac{d\mathcal{P}}{d\Omega} = \frac{ck^2}{8\pi} [|\mathbf{f}(\theta, \phi)|^2 - |\mathbf{n} \cdot \mathbf{f}(\theta, \phi)|^2]. \quad (33)$$

⁴In the near or intermediate zones, there are non-zero components in other directions as well.

If we integrate over the solid angle, we find the total power radiated:

$$\mathcal{P} = \oint_S d^2x \langle \mathbf{S} \rangle \cdot \mathbf{n} = \frac{ck^2}{8\pi} \int d\Omega [|\mathbf{f}(\theta, \phi)|^2 - |\mathbf{n} \cdot \mathbf{f}(\theta, \phi)|^2]. \quad (34)$$

Notice that the radiated power is, appropriately, independent of r .

3 Multipole Expansion of the Radiation Field

Thus far, we have only assumed that $r \gg \lambda$, d . Now we will consider the d/λ .

Consider two limits:

- If $d/\lambda \ll 1$, then all elements of the source will essentially be in phase, and thus an observer at r cannot learn about the structure of the source from the emitted radiation. In this limit, we need only consider the first finite moment in d/λ (if the series is convergent).
- If $d/\lambda \gtrsim 1$, then the elements of the source will not radiate in phase, and an observer at r may learn about the details of the structure of the source by analyzing the interference of the radiation pattern (i.e. Bragg diffraction). In this case, to be discussed in sec. V, we need to retain more terms in the series.

Let us now go back and attempt to evaluate $\mathbf{f}(\theta, \phi)$. If $kd \ll 1$, or $2\pi d/\lambda \ll 1$, then it is not unreasonable to proceed with the evaluation by expanding the exponential function $e^{-ik(\mathbf{n} \cdot \mathbf{x}')}$,

$$\mathbf{f}(\theta, \phi) = \frac{1}{c} \int d^3x' \mathbf{J}(\mathbf{x}') \left[1 - ik(\mathbf{n} \cdot \mathbf{x}') - \frac{1}{2}k^2(\mathbf{n} \cdot \mathbf{x}')^2 + \dots \right]. \quad (35)$$

3.1 Electric Dipole

The first term in the expansion is just the volume integral of $\mathbf{J}(\mathbf{x}')$; one can write it as

$$\frac{1}{c} \int d^3x' \mathbf{J}(\mathbf{x}') = -\frac{1}{c} \int d^3x' [\nabla' \cdot \mathbf{J}(\mathbf{x}')] \mathbf{x}' \quad (36)$$

which one can show by doing integration by parts starting from the right-hand side of this equation. Now employ Eq. (13) to have

$$\frac{1}{c} \int d^3 x' \mathbf{J}(\mathbf{x}') = -ik \int d^3 x' \mathbf{x}' \rho(\mathbf{x}'). \quad (37)$$

The right-hand side can be recognized as the electric dipole moment of the amplitude $\rho(\mathbf{x})$ of the harmonically varying charge distribution. Let us define the electric dipole moment \mathbf{p} in the usual way,

$$\mathbf{p} \equiv \int d^3 x' \mathbf{x}' \rho(\mathbf{x}'). \quad (38)$$

The electric dipole contribution $\mathbf{f}_{ed}(\theta, \phi)$ to $\mathbf{f}(\theta, \phi)$ is thus

$$\mathbf{f}_{ed}(\theta, \phi) = -ik\mathbf{p}; \quad (39)$$

it is in fact independent of θ and ϕ . The corresponding contribution to the vector potential, $\mathbf{A}_{ed}(\mathbf{x}, t)$, is

$$\mathbf{A}_{ed}(\mathbf{x}, t) = -ik\mathbf{p} \frac{e^{i(kr-\omega t)}}{r}. \quad (40)$$

We have used only $d \ll \lambda$ and $d \ll r$; no assumption about the relative size of λ and r has been made. It is somewhat tedious, but nevertheless instructive, to evaluate the fields without making any assumptions so that we can see their form in the near, intermediate, and far zones. First, the magnetic induction is

$$\begin{aligned} \mathbf{B}_{ed}(\mathbf{x}) &= \nabla \times \mathbf{A}_{ed}(\mathbf{x}) = -ik \nabla \left(\frac{e^{ikr}}{r} \right) \times \mathbf{p} \\ &= -ik \left(ik - \frac{1}{r} \right) \frac{e^{ikr}}{r} (\mathbf{n} \times \mathbf{p}) = k^2 \left(1 - \frac{1}{ikr} \right) \frac{e^{ikr}}{r} (\mathbf{n} \times \mathbf{p}). \end{aligned} \quad (41)$$

Further, the electric field is

$$\begin{aligned} \mathbf{E}_{ed}(\mathbf{x}) &= \frac{i}{k} (\nabla \times \mathbf{B}_{ed}(\mathbf{x})) \\ &= ik \nabla \left[\frac{e^{ikr}}{r} \left(1 - \frac{1}{ikr} \right) \right] \times (\mathbf{n} \times \mathbf{p}) + ik \frac{e^{ikr}}{r} \left(1 - \frac{1}{ikr} \right) \nabla \times (\mathbf{n} \times \mathbf{p}) \\ &= ik \left\{ \left(\frac{ik}{r} - \frac{1}{r^2} \right) \left(1 - \frac{1}{ikr} \right) + \frac{1}{ikr^3} \right\} e^{ikr} \mathbf{n} \times (\mathbf{n} \times \mathbf{p}) \end{aligned}$$

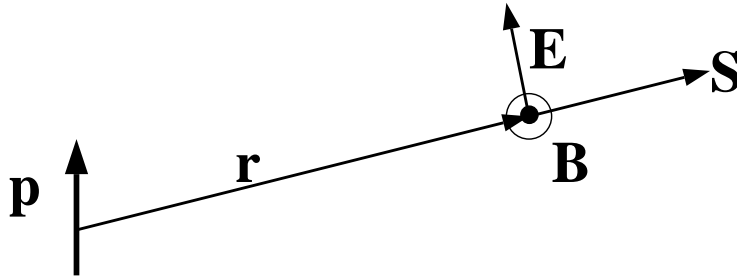
$$\begin{aligned}
& +ik\frac{e^{ikr}}{r}\left(1-\frac{1}{ikr}\right)\left[\left(\mathbf{p}\cdot\nabla\right)\left(\frac{\mathbf{x}}{r}\right)-\mathbf{p}\nabla\cdot\left(\frac{\mathbf{x}}{r}\right)\right] \\
= & \frac{k^2}{r}\left[-1+\frac{2}{ikr}+\frac{2}{k^2r^2}\right]e^{ikr}\mathbf{n}\times\left(\mathbf{n}\times\mathbf{p}\right) \\
& +\frac{ik}{r}e^{ikr}\left(1-\frac{1}{ikr}\right)\left(\frac{\mathbf{p}}{r}-\frac{(\mathbf{p}\cdot\mathbf{n})\mathbf{n}}{r}-\frac{3\mathbf{p}}{r}+\frac{\mathbf{p}}{r}\right) \\
= & \left\{-\frac{k^2}{r}\mathbf{n}\times\left(\mathbf{n}\times\mathbf{p}\right)+\frac{1}{r^3}(1-ikr)[3\mathbf{n}(\mathbf{p}\cdot\mathbf{n})-\mathbf{p}]\right\}e^{ikr}. \tag{42}
\end{aligned}$$

The electric field is divided up in this fashion to bring out, first, the form in the radiation zone which is the first term and, second, the form in the near zone $r \ll \lambda$ which is the second term. The spatial dependence of the latter is the same as the field of a static dipole, $[3\mathbf{n}(\mathbf{p}\cdot\mathbf{n})-\mathbf{p}]/r^3$, but do not forget that it oscillates with angular frequency ω . In the intermediate zone where all contributions are comparable, the field is complex indeed.⁵

In the radiation zone, where the fields become quite simple, they are

$$\mathbf{B}_{ed}(\mathbf{x}) = \frac{k^2}{r}e^{ikr}(\mathbf{n}\times\mathbf{p}) \quad \text{and} \quad \mathbf{E}_{ed} = -\frac{k^2}{r}e^{ikr}\mathbf{n}\times(\mathbf{n}\times\mathbf{p}). \tag{43}$$

The same conclusion may be reached much more simply from Eqs. (30), (31), and (39). As remarked earlier, the fields in the far zone are transverse to \mathbf{n} . If we let \mathbf{p} define the polar axis, then \mathbf{B}_{ed} is azimuthal, i.e., in the direction of ϕ . This is a special feature of electric dipole radiation. Further, $\mathbf{E}_{ed}(\mathbf{x})$ is in the direction of θ .



⁵And this is only the electric dipole part of the field which is, along with the magnetic dipole part of the field, by far the simplest.

From Eqs. (32) and (39) we find that the time-averaged Poynting vector in the radiation zone is

$$\langle \mathbf{S} \rangle = \frac{ck^4}{8\pi r^2} [|\mathbf{p}|^2 - |\mathbf{n} \cdot \mathbf{p}|^2] \mathbf{n} = \frac{ck^4}{8\pi r^2} |\mathbf{p}|^2 \sin^2 \theta \mathbf{n}, \quad (44)$$

where θ is the usual polar angle, i.e., the angle between the dipole moment and the direction at which the radiation is observed. The radiated power per unit solid angle is

$$\frac{d\mathcal{P}}{d\Omega} = r^2 \langle \mathbf{S} \rangle \cdot \mathbf{n} = \frac{ck^4}{8\pi} |\mathbf{p}|^2 \sin^2 \theta, \quad (45)$$

and the total power radiated is

$$\mathcal{P} = \int d\Omega \frac{d\mathcal{P}}{d\Omega} = \frac{ck^4 |\mathbf{p}|^2}{8\pi} \int d\phi d\theta \sin \theta \sin^2 \theta = \frac{ck^4 |\mathbf{p}|^2}{3}. \quad (46)$$

3.1.1 Example: Linear Center-Fed Antenna

Consider the short, linear, “center-fed” antenna shown below.

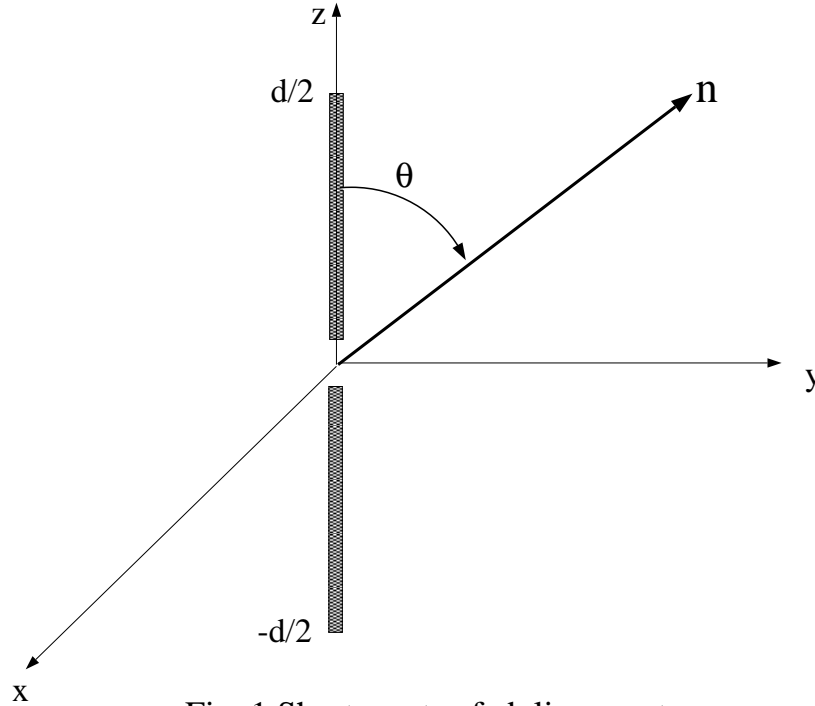


Fig. 1 Short, center-fed, linear antenna.

For such an antenna, the current density can be crudely approximated by

$$\mathbf{J}(\mathbf{x}, t) = \epsilon_3 I_0 \delta(x) \delta(y) \left(1 - 2 \frac{|z|}{d}\right) e^{-i\omega t} \quad (47)$$

for $|z| < d/2$; for $|z| > d/2$, it is zero. Given this current density, we can evaluate the divergence and so find the charge density,

$$\nabla \cdot \mathbf{J}(\mathbf{x}) = -\frac{2I_0}{d} \delta(x) \delta(y) \frac{z}{|z|} = -i\omega \rho(\mathbf{x}) \quad (48)$$

or

$$\rho(\mathbf{x}) = \frac{2iI_0}{\omega d} \delta(x) \delta(y) \frac{z}{|z|}, \quad |z| < d/2. \quad (49)$$

Hence,

$$\mathbf{p} = \int d^3x \mathbf{x} \rho(\mathbf{x}) = \frac{iI_0 d}{2ck} \epsilon_3. \quad (50)$$

Now we have only to plug Eq. (50) into Eqs. (45) and (46) to find

$$\frac{d\mathcal{P}}{d\Omega} = \frac{I_0^2}{32\pi c} (kd)^2 \sin^2 \theta \quad \text{and} \quad \mathcal{P} = \frac{I_0^2 k^2 d^2}{12c}. \quad (51)$$

The calculation in this example may be expected to provide a good approximation to the total radiated power provided $kd \ll 1$ so that the electric dipole term dominates, unless it vanishes. There are many instances where this happens.

3.2 Magnetic Dipole

When this happens it is necessary to look at higher-order terms in the expansion of the phase factor $e^{ik|\mathbf{x}-\mathbf{x}'|}$. Let's look now at the next one. Start from the exact expression for $\mathbf{A}(\mathbf{x})$,

$$\begin{aligned} \mathbf{A}(\mathbf{x}) &= \frac{1}{c} \int d^3x' \mathbf{J}(\mathbf{x}') \frac{e^{ik|\mathbf{x}-\mathbf{x}'|}}{|\mathbf{x}-\mathbf{x}'|} \\ &= \mathbf{A}_{ed}(\mathbf{x}) + \frac{e^{ikr}}{cr} \int d^3x' \mathbf{J}(\mathbf{x}') \left(\frac{\mathbf{n} \cdot \mathbf{x}'}{r} - ik(\mathbf{n} \cdot \mathbf{x}') \right) [1 + \mathcal{O}(d/r, d/\lambda)]. \end{aligned} \quad (52)$$

where the second term in the () comes from the exponential, and the first comes from the corresponding denominator. The integral we wish to examine is

$$\begin{aligned} &\frac{1}{c} \int d^3x' \mathbf{J}(\mathbf{x}') (\mathbf{n} \cdot \mathbf{x}') \\ &= \frac{1}{2c} \int d^3x' \{ [\mathbf{J}(\mathbf{x}') (\mathbf{n} \cdot \mathbf{x}') + \mathbf{x}' (\mathbf{n} \cdot \mathbf{J}(\mathbf{x}'))] + [\mathbf{J}(\mathbf{x}') (\mathbf{n} \cdot \mathbf{x}') - \mathbf{x}' (\mathbf{n} \cdot \mathbf{J}(\mathbf{x}'))] \} \\ &= \frac{1}{2c} \int d^3x' [\mathbf{J}(\mathbf{x}') (\mathbf{n} \cdot \mathbf{x}') + \mathbf{x}' (\mathbf{n} \cdot \mathbf{J}(\mathbf{x}'))] + \frac{1}{2c} \int d^3x' \mathbf{n} \times (\mathbf{J}(\mathbf{x}') \times \mathbf{x}') \end{aligned} \quad (53)$$

What is the point of breaking the integral into two pieces, symmetric and antisymmetric under interchange of \mathbf{x}' and $\mathbf{J}(\mathbf{x}')$? There are several related points. One is that in the near zone the second term on the right-hand side produces a magnetic induction which has the form of the induction of a static magnetic dipole while the first term produces an electric field which has the form of the field of a static electric quadrupole. Hence the radiation from the former is called magnetic dipole radiation while that from the latter is known as electric quadrupole radiation. Also, the

magnetic dipole part produces a purely transverse electric field in all zones while the electric quadrupole part gives a purely transverse magnetic induction. Recall that for electric dipole radiation, $\mathbf{B}(\mathbf{x}, t)$ is purely transverse in all zones as well.

Let us look at the magnetic dipole fields first. From Eqs. (52) and (53) we have

$$\begin{aligned}\mathbf{A}_{md}(\mathbf{x}) &= \frac{e^{ikr}}{r} \left(\frac{1}{r} - ik \right) \mathbf{n} \times \int d^3x' \frac{1}{2c} (\mathbf{J}(\mathbf{x}') \times \mathbf{x}') \\ &\equiv -ik \frac{e^{ikr}}{r} \left(1 - \frac{1}{ikr} \right) (\mathbf{n} \times \mathbf{m})\end{aligned}\quad (54)$$

where

$$\mathbf{m} \equiv \frac{1}{2c} \int d^3x' [\mathbf{x}' \times \mathbf{J}(\mathbf{x}')] \quad (55)$$

is the magnetic dipole moment, familiar from our study of magnetostatics.

Rather than plow ahead with with the evaluation of the curl to find \mathbf{B} , etc., let us recall the electric dipole results

$$\mathbf{A}_{ed}(\mathbf{x}) = -ik \frac{e^{ikr}}{r} \mathbf{p} \quad (56)$$

$$\mathbf{B}_{ed}(\mathbf{x}) = k^2 \frac{e^{ikr}}{r} \left(1 - \frac{1}{ikr} \right) (\mathbf{n} \times \mathbf{p}). \quad (57)$$

We see that \mathbf{B}_{ed} is the same in functional form as \mathbf{A}_{md} ; consequently, \mathbf{E}_{ed} , which is the curl of \mathbf{B}_{ed} , must be the same in form as the curl of \mathbf{A}_{md} , or \mathbf{B}_{md} . Hence we can immediately write, using Eq. (43),

$$\mathbf{B}_{md}(\mathbf{x}) = k^2 \frac{e^{ikr}}{r} \mathbf{n} \times (\mathbf{n} \times \mathbf{m}) + \frac{e^{ikr}}{r^3} (1 - ikr) [3\mathbf{n}(\mathbf{n} \cdot \mathbf{m}) - \mathbf{m}]. \quad (58)$$

Notice that in the near zone $\mathbf{B}_{md}(\mathbf{x})$ is the same as that of a static dipole.

As for the corresponding electric field, we could work through the tedious derivatives of the magnetic induction, but it happens that this is one time when it is much easier to evaluate the the field from the potentials. We know that

$$\mathbf{E}(\mathbf{x}, t) = -\nabla\Phi(\mathbf{x}, t) - \frac{1}{c} \frac{\partial \mathbf{A}(\mathbf{x}, t)}{\partial t}. \quad (59)$$

Also, in the Lorentz gauge

$$\nabla \cdot \mathbf{A}(\mathbf{x}, t) + \frac{1}{c} \frac{\partial \Phi(\mathbf{x}, t)}{\partial t} = 0 \quad \text{or} \quad \nabla \cdot \mathbf{A}(\mathbf{x}) - ik\Phi(\mathbf{x}) = 0. \quad (60)$$

Using this relation for $\Phi(\mathbf{x})$, we find that

$$\mathbf{E}(\mathbf{x}) = -\frac{i}{k} \nabla(\nabla \cdot \mathbf{A}(\mathbf{x})) + ik\mathbf{A}(\mathbf{x}) \quad (61)$$

for any electric field which is harmonic in time. From our expression for $\mathbf{A}_{md}(\mathbf{x})$, one can easily see that $\nabla \cdot \mathbf{A}_{md} = 0$ ⁶ and so \mathbf{E}_{md} is simply proportional to the vector potential,

$$\mathbf{E}_{md}(\mathbf{x}) = -k^2 \frac{e^{ikr}}{r} \left(1 - \frac{1}{ikr}\right) (\mathbf{n} \times \mathbf{m}). \quad (62)$$

3.3 Comparison of Dipoles

To summarize:

- The electric dipole and magnetic dipole fields are the same with \mathbf{E} and \mathbf{B} interchanged, $\mathbf{E}_{md} \Leftrightarrow -\mathbf{B}_{ed}$ and $\mathbf{B}_{md} \Leftrightarrow \mathbf{E}_{ed}$ when $\mathbf{p} \Leftrightarrow \mathbf{m}$.
- In the near zone \mathbf{E}_{ed} and \mathbf{B}_{md} have the form of static dipole fields, while in all zones, \mathbf{B}_{ed} and \mathbf{E}_{md} are purely azimuthal in direction.
- The Poynting vector in the far zone has the same form for both electric dipole and magnetic dipole fields; in the latter case it is

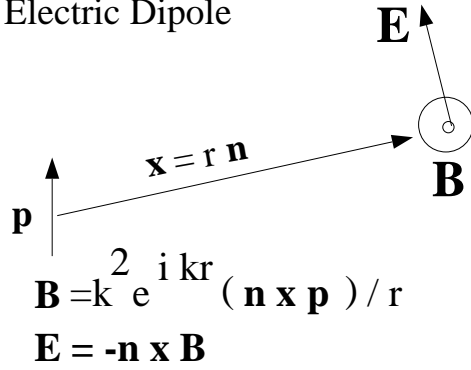
$$\langle \mathbf{S} \rangle = r^2 \mathbf{n} \frac{ck^4}{8\pi} |\mathbf{m}|^2 \sin^2 \theta, \quad (63)$$

leading to results for the power distribution and total power which are the same as for electric dipole radiation, Eqs. (45) and (46), with \mathbf{m} in place of \mathbf{p} . Notice particularly that if one measures $d\mathcal{P}/d\Omega$, and finds the $\sin^2 \theta$ pattern, then one

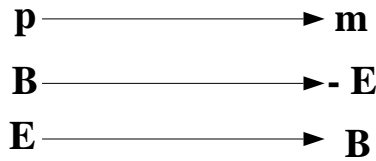
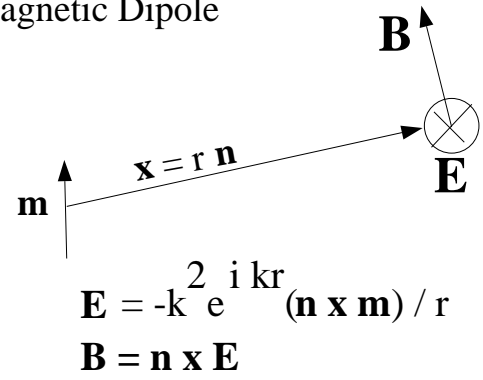
⁶An equivalent statement is that a magnetic dipole is always charge neutral, so that $\Phi = 0$.

knows that the radiation is indeed⁷ dipole radiation; however, it is impossible to distinguish electric dipole radiation from magnetic dipole radiation without examining its polarization.

Electric Dipole



Magnetic Dipole



These formulas *suggest* that for a given set of moving charges, one should get as much power out of an oscillating magnetic dipole as an oscillating electric dipole. This is not so, since the magnetic dipole moment is

$$\mathbf{m} \equiv \frac{1}{2c} \int d^3 x' [\mathbf{x}' \times \mathbf{J}(\mathbf{x}')] \quad (64)$$

thus

$$|\mathbf{m}| \sim \frac{v}{c} Qd \quad (65)$$

where v is the speed of the moving charge, Q is the magnitude of this charge, and d is the size of the current loop. The size of the corresponding electric dipole moment

⁷Higher-order multipoles produce radiation with distinctive angular distributions which are never proportional to $\sin^2 \theta$.

is roughly Qd . Thus,

$$|\mathbf{m}| \sim \frac{v}{c} |\mathbf{p}| \quad (66)$$

We see that

$$\mathcal{P}_{md} \approx \left(\frac{v}{c}\right)^2 \mathcal{P}_{ed}. \quad (67)$$

Thus, in a system with both an electric and magnetic dipole moment, the latter is usually a relativistic correction to the former.

3.4 Electric Quadrupole

Let's go now to the other contribution to the vector potential in Eq. (53). This is called the electric quadrupole term; knowing that, we naturally expect to find that it produces an electric field in the near zone which has the characteristic form of a static electric quadrupole field. The vector potential is

$$\begin{aligned} A_{eq}(\mathbf{x}) &= \frac{e^{ikr}}{r} \left(\frac{1}{r} - ik\right) \frac{1}{2c} \int d^3x' [\mathbf{J}(\mathbf{x}')(\mathbf{n} \cdot \mathbf{x}') + \mathbf{x}'(\mathbf{n} \cdot \mathbf{J}(\mathbf{x}'))] \\ &= -\frac{e^{ikr}}{r} \left(\frac{1}{r} - ik\right) \frac{1}{2c} \int d^3x' (\mathbf{n} \cdot \mathbf{x}') (\nabla' \cdot \mathbf{J}(\mathbf{x}')) \mathbf{x}'. \end{aligned} \quad (68)$$

To demonstrate this algebraic step, consider the i^{th} component of the final expression:

$$\begin{aligned} \frac{1}{2c} \int d^3x' (\mathbf{n} \cdot \mathbf{x}') \left(\sum_j \frac{\partial J_j}{\partial x'_j} \right) x'_i &= -\frac{1}{2c} \sum_j \int d^3x' \frac{\partial}{\partial x'_j} [x'_i (\mathbf{n} \cdot \mathbf{x}')] J_j \\ &= -\frac{1}{2c} \int d^3x' \sum_j [\delta_{ij} (\mathbf{n} \cdot \mathbf{x}') + x'_i n_j] J_j \\ &= -\frac{1}{2c} \int d^3x' [J_i (\mathbf{n} \cdot \mathbf{x}') + x'_i (\mathbf{n} \cdot \mathbf{J})] \end{aligned} \quad (69)$$

which matches the i^{th} component of the original expression in Eq. (68). Making use of Eq. (13) for $\nabla \cdot \mathbf{J}(\mathbf{x})$ and also using $\omega = ck$, we find, from Eq. (68), that

$$\mathbf{A}_{eq}(\mathbf{x}) = -\frac{k^2}{2} \frac{e^{ikr}}{r} \left(1 - \frac{1}{ikr}\right) \int d^3x' \mathbf{x}' (\mathbf{n} \cdot \mathbf{x}') \rho(\mathbf{x}'). \quad (70)$$

There are nine components to the integral since the factor of \mathbf{n} in the integrand can be used to project out three numbers by, for example, letting \mathbf{n} be each of the Cartesian basis vectors. That is, the basic integral over the source, $\rho(\mathbf{x})$, which appears here is a dyadic, $\int d^3x \mathbf{x}\mathbf{x}\rho(\mathbf{x})$. It is symmetric and so has at most six independent components. Notice also that \mathbf{A}_{eq} depends on the direction of \mathbf{n} , which was not the case for either \mathbf{A}_{ed} or \mathbf{A}_{md} ; the evaluation of the fields is further complicated as a consequence.

The electric quadrupole vector potential can be written in terms of the components Q_{ij} of the electric quadrupole moment tensor which we defined long ago. Recall that

$$Q_{ij} \equiv \int d^3x (3x_i x_j - r^2 \delta_{ij}) \rho(\mathbf{x}). \quad (71)$$

Take combinations of these to construct the components Q_i of a vector \mathbf{Q} :

$$Q_i(\mathbf{n}) \equiv \sum_{j=1}^3 Q_{ij} n_j. \quad (72)$$

From these definitions one can show quite easily that

$$\frac{1}{3} \mathbf{n} \times \mathbf{Q}(\mathbf{n}) \equiv \mathbf{n} \times \left(\int d^3x' \mathbf{x}' (\mathbf{n} \cdot \mathbf{x}') \rho(\mathbf{x}') \right). \quad (73)$$

To see this, note that the i^{th} component of $\mathbf{Q}/3$ is

$$\frac{1}{3} Q_i = \int d^3x' [x'_i (\sum_j n_j x'_j) \rho(\mathbf{x}') - r'^2 n_i \rho(\mathbf{x}')/3]. \quad (74)$$

The first term on the right-hand side of this relation produces the i^{th} component of the integral in Eq. (73); the second term is some i -independent quantity multiplied by n_i ; the three terms in \mathbf{Q} of this form give something which is directly proportional to \mathbf{n} and so they do not contribute to $\mathbf{n} \times \mathbf{Q}$.

Thus have we established the validity of Eq. (73). But what good is it? It tells us that we can write $\mathbf{n} \times \mathbf{A}_{eq}$ in terms of $\mathbf{n} \times \mathbf{Q}$, but does not allow us to write \mathbf{A}_{eq} itself in terms of \mathbf{Q} . However, if we restrict our attention to the radiation zone, then

$\mathbf{n} \times \mathbf{A}$ is all we will need, because in this zone, $\mathbf{B} = ik(\mathbf{n} \times \mathbf{A})$ and $\mathbf{E} = -\mathbf{n} \times \mathbf{B}$. Thus, in the far zone,

$$\mathbf{B}_{eq}(x) = -\frac{ik^3 e^{ikr}}{6r} [\mathbf{n} \times \mathbf{Q}(\mathbf{n})] \quad (75)$$

and

$$\mathbf{E}_{eq}(\mathbf{x}) = \frac{ik^3 e^{ikr}}{6r} \mathbf{n} \times [\mathbf{n} \times \mathbf{Q}(\mathbf{n})]. \quad (76)$$

The time-averaged power radiated is

$$\begin{aligned} \frac{d\mathcal{P}}{d\Omega} = r^2 \langle \mathbf{S} \cdot \mathbf{n} \rangle &= \frac{c}{8\pi} r^2 [\mathbf{E}(\mathbf{x}) \times \mathbf{B}^*(\mathbf{x})] \cdot \mathbf{n} = \frac{c}{8\pi} r^2 [\mathbf{B}^*(\mathbf{x}) \times \mathbf{n}] \cdot \mathbf{E}(\mathbf{x}) \\ &= \frac{c}{8\pi} \frac{k^6}{36} |\mathbf{n} \times [\mathbf{n} \times \mathbf{Q}(\mathbf{n})]|^2 = \frac{ck^6}{288\pi} |\mathbf{n} \times \mathbf{Q}(\mathbf{n})|^2. \end{aligned} \quad (77)$$

The right-hand side does not have any single form as a function of θ and ϕ that we can extract because $\mathbf{Q}(\mathbf{n})$ depends in an unknown way (in general) on these angles. We can proceed to a general result only up to a point:

$$\begin{aligned} |\mathbf{n} \times \mathbf{Q}|^2 &= (\mathbf{n} \times \mathbf{Q}) \cdot (\mathbf{n} \times \mathbf{Q}^*) = [(\mathbf{n} \times \mathbf{Q}) \times \mathbf{n}] \cdot \mathbf{Q}^* \\ &= -\mathbf{Q}^* \cdot [\mathbf{n}(\mathbf{Q} \cdot \mathbf{n}) - \mathbf{Q}] = |\mathbf{Q}|^2 - |\mathbf{n} \cdot \mathbf{Q}|^2; \end{aligned} \quad (78)$$

this is a brilliant derivation of the statement that $\sin^2 \theta = 1 - \cos^2 \theta$. Further,

$$\mathbf{Q} \cdot \mathbf{Q}^* - (\mathbf{n} \cdot \mathbf{Q})(\mathbf{n} \cdot \mathbf{Q}^*) = \sum_{ijk} Q_{ij} n_j Q_{ik}^* n_k - \sum_{ijkl} n_i Q_{ij} n_j n_k Q_{kl}^* n_l. \quad (79)$$

The n_i 's are direction cosines and so obey the identities

$$\int d\Omega n_i n_j = \frac{4\pi}{3} \delta_{ij} \quad (80)$$

and

$$\int d\Omega n_i n_j n_k n_l = \frac{4\pi}{15} (\delta_{ij} \delta_{kl} + \delta_{ik} \delta_{jl} + \delta_{il} \delta_{jk}). \quad (81)$$

These enable us at least to get a simple result for the total radiated power:

$$\begin{aligned} \mathcal{P} &= \int d\Omega \frac{d\mathcal{P}}{d\Omega} \\ &= \frac{ck^6}{288\pi} \left\{ \frac{4\pi}{3} \sum_{ij} |Q_{ij}|^2 - \frac{4\pi}{15} \sum_{ik} [Q_{ii} Q_{kk}^* + |Q_{ik}|^2 + Q_{ij} Q_{ji}^*] \right\} \\ &= \frac{ck^6}{360} \sum_{ij} |Q_{ij}|^2 \end{aligned} \quad (82)$$

where we have used the facts that $Q_{ij} = Q_{ji}$ and $\sum_i Q_{ii} = 0$.

By choosing appropriate axes (the principal axes of the quadrupole moment matrix or tensor) one can always put the matrix of quadrupole moments into diagonal form. Further, only two of the diagonal elements can be chosen independently because the trace of the matrix must vanish. Hence any quadrupole moment matrix is a linear combination of two basic ones.

3.4.1 Example: Oscillating Charged Spheroid

A commonly occurring example is an oscillating spheroidal charge distribution leading to

$$Q_{33} = Q_0 \quad \text{and} \quad Q_{22} = Q_{11} = -Q_0/2. \quad (83)$$

Then

$$Q_i = \sum_j Q_{ij} n_j = Q_{ii} n_i \quad (84)$$

or

$$\mathbf{Q} = Q_0 \left[\cos \theta \boldsymbol{\epsilon}_3 - \frac{1}{2} \sin \theta (\cos \phi \boldsymbol{\epsilon}_1 + \sin \phi \boldsymbol{\epsilon}_2) \right]. \quad (85)$$

From this we have

$$\begin{aligned} |\mathbf{Q}|^2 &= Q_0^2 \left(\cos^2 \theta + \frac{1}{4} \sin^2 \theta \right), \\ \mathbf{n} \cdot \mathbf{Q} &= Q_0 \left(\cos^2 \theta - \frac{1}{2} \sin^2 \theta \right), \\ |\mathbf{n} \cdot \mathbf{Q}|^2 &= Q_0^2 \left(\cos^4 \theta - \cos^2 \theta \sin^2 \theta + \frac{1}{4} \sin^4 \theta \right), \end{aligned} \quad (86)$$

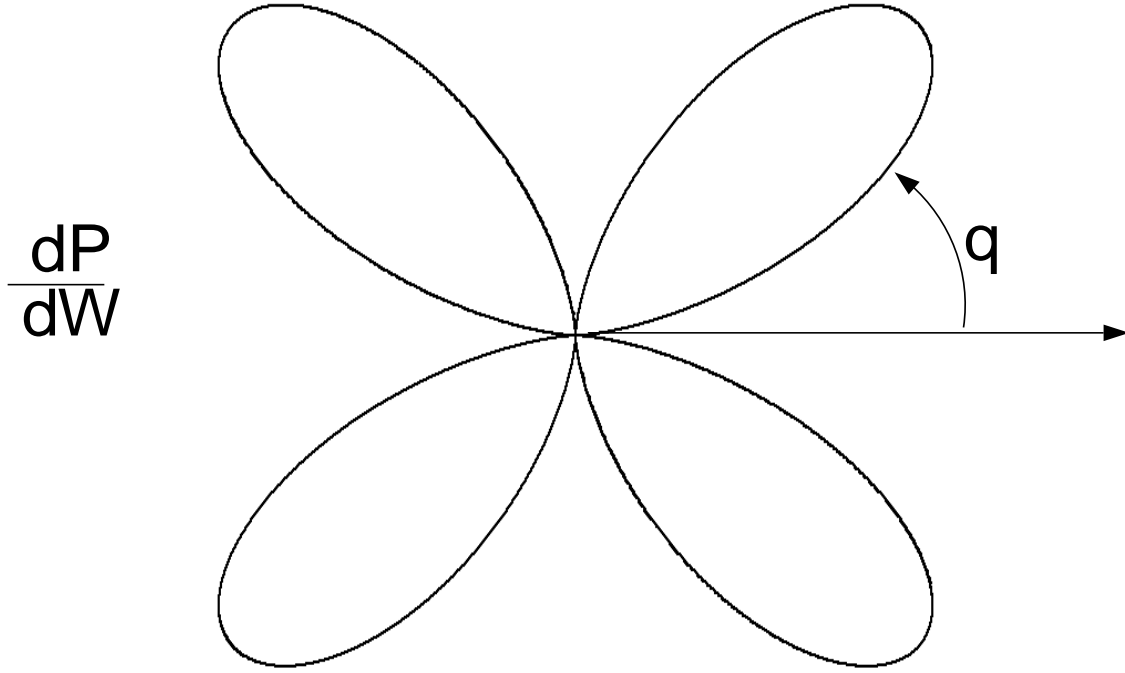
and so

$$\begin{aligned} |\mathbf{Q}|^2 - |\mathbf{n} \cdot \mathbf{Q}|^2 &= Q_0^2 \left[\cos^2 \theta + \frac{1}{4} \sin^2 \theta - \cos^4 \theta + \sin^2 \theta \cos^2 \theta - \frac{1}{4} \sin^4 \theta \right] \\ &= Q_0^2 \left[\cos^2 \theta \sin^2 \theta + \frac{1}{4} \cos^2 \theta \sin^2 \theta + \cos^2 \theta \sin^2 \theta \right] \\ &= \frac{9}{4} Q_0^2 \cos^2 \theta \sin^2 \theta. \end{aligned} \quad (87)$$

Hence,

$$\frac{d\mathcal{P}}{d\Omega} = \frac{ck^6}{128\pi} Q_0^2 \sin^2 \theta \cos^2 \theta. \quad (88)$$

This is a typical, but not uniquely so, electric quadrupole radiation distribution.



Radiation Pattern of an oscillating charged spheriod

Higher-order multipole radiation (including magnetic quadrupole radiation) is found by expanding the factor $e^{-ik(\mathbf{n}\cdot\mathbf{x}')}$ in Eq. (27) to higher order in powers of d/λ . One can do this in a complete and systematic fashion after first developing some appropriate mathematical machinery by generalizing the spherical harmonics to vector fields,⁸ the purpose being to construct an orthonormal set of basis functions for the electromagnetic fields.

⁸Thereby producing the so-called *vector spherical harmonics*. This is the subject matter of Chapter 16.

3.5 Large Radiating Systems

Before abandoning the topic of simple radiating systems, let us look at one example of an antenna which is not small compared the wavelength of the emitted radiation. Our example is an array of antennas, each of which is itself small compared to λ and each of which will be treated as a point dipole. We take the current density of this array to be

$$\mathbf{J}(\mathbf{x}) = I_0 a \sum_j \delta(\mathbf{x} - \mathbf{x}_j) e^{i\phi_j} \boldsymbol{\epsilon}_3. \quad (89)$$

One can easily see that this is an array of point antennas located at positions \mathbf{x}_j ; they all have the same current but are not necessarily in phase, the phase of the j^{th} antenna being given by ϕ_j (the additional phase factor $e^{-i\omega t}$, common to all antennas, has been omitted, as usual).



No matter how large the array, we can specify that \mathbf{x} is large enough that the observation point is in the far zone in which case the vector potential can be taken as

$$\mathbf{A}(\mathbf{x}) = \frac{e^{ikr}}{r} \mathbf{f}(\theta, \phi) \quad (90)$$

where

$$\begin{aligned} \mathbf{f}(\theta, \phi) &= \boldsymbol{\epsilon}_3 \frac{I_0 a}{c} \int d^3x' \sum_j \delta(\mathbf{x}' - \mathbf{x}_j) e^{-ik(\mathbf{n} \cdot \mathbf{x}')} e^{i\phi_j} \\ &= \boldsymbol{\epsilon}_3 \frac{I_0 a}{c} \sum_j e^{i[\phi_j - k(\mathbf{n} \cdot \mathbf{x}_j)]}. \end{aligned} \quad (91)$$

Referring back to Eq. (33) we find that the distribution of radiated power is

$$\frac{d\mathcal{P}}{d\Omega} = \frac{k^2 I_0^2 a^2}{8\pi c} \sin^2 \theta |w|^2 \quad (92)$$

where

$$w \equiv \sum_j e^{i(\phi_j - k\mathbf{n} \cdot \mathbf{x}_j)}. \quad (93)$$

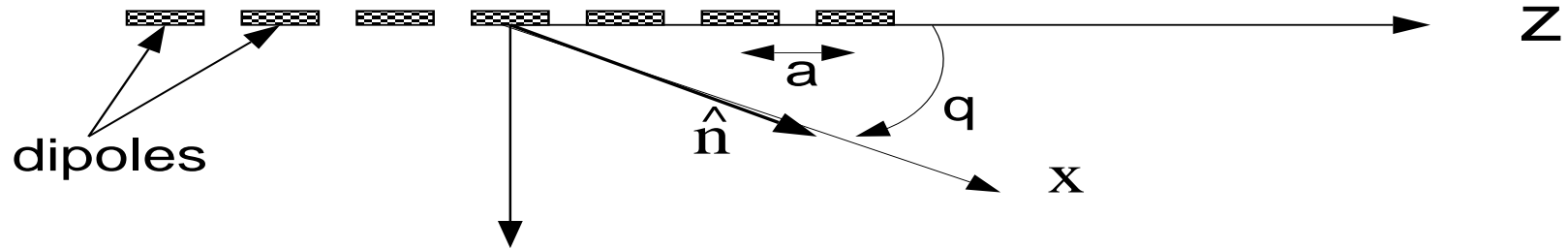
The factor of $\sin^2 \theta$ arises because each element of this array is treated in the dipole approximation. The factor $|w|^2$ contains all of the information about the relative phases of the amplitudes from the different elements, i.e., about the interference of the waves from the different elements.

3.5.1 Example: Linear Array of Dipoles

As a special and explicit example, suppose that

$$\phi_j = j\phi_0 \quad \text{and} \quad \mathbf{x}_j = aj\boldsymbol{\epsilon}_3, \quad (94)$$

meaning that the elements are equally spaced in a line along the z-axis and that they have relative phases that increase linearly along the array.



Then

$$w = \sum_j e^{i(\phi_0 - ka \cos \theta)j} \equiv \sum_j x^j \quad (95)$$

where

$$x = e^{i(\phi_0 - ka \cos \theta)}. \quad (96)$$

The sum is easy to evaluate if we know where j begins and ends:

$$\sum_{j=j_1}^{j_2} x^j = x^{j_1} \frac{1 - x^{j_2 - j_1 + 1}}{1 - x}, \quad (97)$$

so

$$|w|^2 = \left| \frac{1 - e^{i\alpha(\theta)(j_2 - j_1 + 1)}}{1 - e^{i\alpha(\theta)}} \right|^2 \quad (98)$$

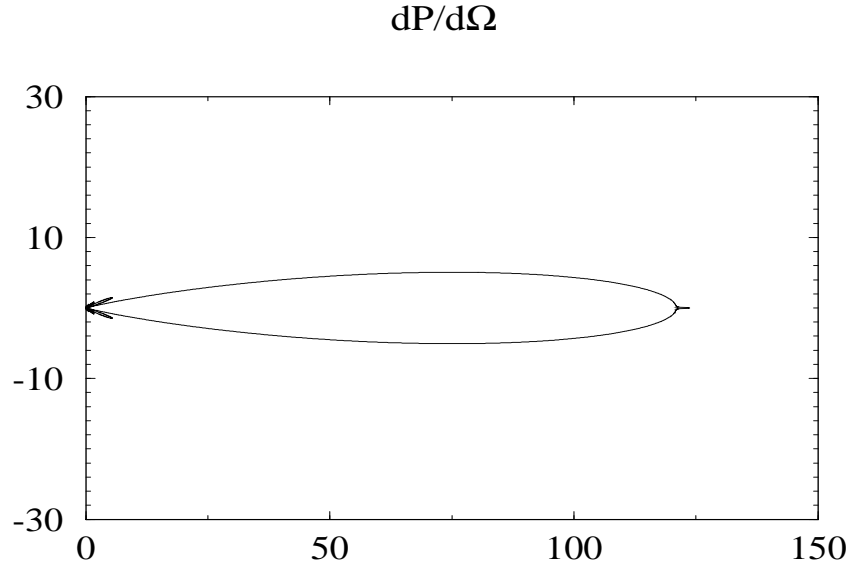
where $\alpha(\theta) \equiv \phi_0 - ka \cos \theta$. Let $j_1 = -n$ and $j_2 = n$, corresponding to an array of $2n + 1$ elements centered at the origin. Then

$$|w|^2 = \left| \frac{1 - e^{i\alpha(2n+1)}}{1 - e^{i\alpha}} \right|^2 = \frac{1 - \cos[(2n + 1)\alpha]}{1 - \cos \alpha}. \quad (99)$$

This is a function which is in general of order unity and which oscillates as a function of θ . It has, however, a large peak of size $(2n + 1)^2$ when $\alpha(\theta)$ is an integral multiple of 2π . The peak occurs at that angle θ_0 where⁹ $\alpha(\theta_0) = 0$ or $\cos \theta_0 = \phi_0/ka$. If we choose $\phi_0 = 0$, the peak is at $\theta_0 = \pi/2$; further, if $ka < 2\pi$ or $a < \lambda$, there is no other such peak. The width of the peak can be determined from the fact that the factor w goes to zero when $(2n + 1)\alpha(\theta) = 2\pi$. Assuming that $n \gg 1$, one finds that the corresponding angle θ differs from θ_0 by η where

$$\eta = \frac{2\pi}{(2n + 1)ka[1 - (\phi_0/ka)^2]} \sim (2n + 1)^{-1}. \quad (100)$$

Hence the antenna becomes increasingly directional with increasing n . The accompanying figure shows the power distribution in units of $k^2 I_0^2 a^2 / 8\pi c$ for 5 elements with $ka = \pi$ and $\phi_0 = 0$.



⁹More generally, $\alpha(\theta_0) = 2m\pi$ where m is an integer; because we control ϕ_0 , we can make it small enough that the peak corresponds to the particular case $m = 0$.

4 Multipole expansion of sources in waveguides

We saw in Chapter 8 that a field in a waveguide could be expanded in the normal modes of the waveguide with an integral over the sources, consisting of a current distribution and apertures, determining the coefficients in the expansion. If the sources are small in size compared to distances over which fields in the normal modes vary, then one can do the integrals in an approximate fashion by making a multipole expansion of the sources.

4.1 Electric Dipole

Consider first the part of the amplitude which is produced by some explicit current distribution. It is

$$A_\lambda^{(\pm)} = -\frac{2\pi Z_\lambda}{c} \int d^3x' \mathbf{J}(\mathbf{x}') \cdot \mathbf{E}_\lambda^{(\mp)}. \quad (101)$$

The field in the mode is given by

$$\mathbf{E}_\lambda^{(\pm)}(\mathbf{x}') = [\mathbf{E}_\lambda(x', y') \pm \epsilon_3 E_{z\lambda}(x', y')] e^{\pm ik_\lambda z'}. \quad (102)$$

The origin of the coordinate \mathbf{x}' is at some appropriately chosen point which is probably not near the center of the source distribution. Let us use a different coordinate \mathbf{x} having as origin a point near the center of the source. Also, let's let the electric field in the mode be called $\mathbf{E}(\mathbf{x})$ as a matter of convenience. Then we have to do an integral of the form

$$\int d^3x \mathbf{J}(\mathbf{x}) \cdot \mathbf{E}(\mathbf{x}) = \int d^3x \mathbf{J}(\mathbf{x}) \cdot \left[\mathbf{E}(0) + \sum_{i,j} \frac{\partial E_i}{\partial x_j} \Big|_0 \epsilon_i x_j + \dots \right] \quad (103)$$

where we are assuming that the field varies little over the size of the source. In this integral the first term can be converted to an integral over the charge density,

$$\int d^3x \mathbf{J}(\mathbf{x}) = -i\omega \int d^3x \mathbf{x} \rho(\mathbf{x}) \quad (104)$$

provided the integration by parts can be done without picking up a contribution from the surface of the integration volume. That is not automatic here because the surface includes some points on the wall of the waveguide including those points where the current is fed into the guide, so one has to exercise some care in applying this formula. Assuming that it is okay, we see that the leading term in the expansion of the integrand produces to a term in the coefficient $A_\lambda^{(\pm)}$ which is proportional to $\mathbf{p} \cdot \mathbf{E}(0)$.

The next term in the expansion can, as we have seen, be divided into symmetric and antisymmetric parts:

$$\begin{aligned} \sum_{i,j} \frac{\partial E_i}{\partial x_j} \Big|_0 \int d^3x J_i(\mathbf{x}) x_j &= \frac{1}{4} \sum_{i,j} \left(\frac{\partial E_i}{\partial x_j} - \frac{\partial E_j}{\partial x_i} \right) \Big|_0 \int d^3x [J_i x_j - J_j x_i] \\ &+ \frac{1}{2} \frac{\partial E_i}{\partial x_j} \Big|_0 \int d^3x [J_i x_j + J_j x_i]. \end{aligned} \quad (105)$$

The first term on the right-hand side is set up in such a way as to display explicitly a component of $\nabla \times \mathbf{E}$, which is $i(\omega/c)\mathbf{B}$, and the same component of the magnetic dipole moment. Hence this term is proportional to $\mathbf{m} \cdot \mathbf{B}(0)$. The remaining term can be handled in the way that we treated the electric quadrupole part of the vector potential earlier; it becomes

$$-\frac{i\omega}{2} \sum_{i,j} \frac{\partial E_i}{\partial x_j} \Big|_0 \int d^3x x_i x_j \rho(\mathbf{x}) = -\frac{i\omega}{6} \sum_{i,j} Q_{ij} \frac{\partial E_i}{\partial x_j} \Big|_0 \quad (106)$$

provided one can throw away the contributions that come from the surface when the integration by parts is done. The final step is achieved by making use of the fact that $\nabla \cdot \mathbf{E} = 0$.

In the next order, not shown in Eq. (103), the antisymmetric part provides the magnetic quadrupole contribution. Without delving into the algebra of the derivation, we state that the result is of the same form as Eq. (106) but with the *magnetic quadrupole moment tensor* Q_{ij}^M in place of the electric quadrupole moment tensor, \mathbf{B} in place of \mathbf{E} , and an overall relative $(-)$. The components of the magnetic quadrupole

moment tensor are defined in the same way as those of the electric quadrupole moment tensor except that in place of $\rho(\mathbf{x})$ there is

$$\rho^M(\mathbf{x}) = -\frac{1}{2c} \nabla \cdot [\mathbf{x} \times \mathbf{J}(\mathbf{x})]. \quad (107)$$

The final result for the amplitude $A_\lambda^{(\pm)}$ with all indices in place is

$$A_\lambda^{(\pm)} = i \frac{2\pi\omega}{c} \left\{ \mathbf{p} \cdot \mathbf{E}_\lambda^{(\mp)}(0) - \mathbf{m} \cdot \mathbf{B}_\lambda^{(\mp)} + \frac{1}{6} \sum_{i,j} \left[Q_{ij} \frac{\partial E_{i\lambda}^{(\mp)}}{\partial x_j} \Big|_0 - Q_{ij}^M \frac{\partial B_{i\lambda}^{(\mp)}}{\partial x_j} \Big|_0 \right] + \dots \right\}. \quad (108)$$

We can see immediately some interesting features of this result. For example, if one wants to produce TM modes, which have z components of \mathbf{E} but not of \mathbf{B} , then this is most efficiently done by designing the source to have a large p_z . At the same time, hardly any TE mode will be generated if there is only a z component of \mathbf{p} because the TE mode has no E_z to couple to \mathbf{p} . Hence this expression gives one a good idea how to design a source to produce, or not produce, modes of a given kind.

Now let's look at the same expansion if the source of radiation is an aperture rather than an explicit current distribution. We derived in chapter 8 that in this case

$$\begin{aligned} A_\lambda^{(\pm)} &= -\frac{Z_\lambda}{2} \int_{S_a} d^2x' \mathbf{n} \cdot [\mathbf{E}(\mathbf{x}') \times \mathbf{H}_\lambda^{(\mp)}(\mathbf{x}')] \\ &= -\frac{Z_\lambda}{2} \int_{S_a} d^2x' [\mathbf{n} \times \mathbf{E}(\mathbf{x}')] \cdot \mathbf{H}_\lambda^{(\mp)}(\mathbf{x}') \end{aligned} \quad (109)$$

where the integral is over the aperture S_a , $\mathbf{E}(\mathbf{x}')$ is the electric field actually present at point \mathbf{x}' , and \mathbf{n} is the inward directed normal at the aperture. Notice that only the tangential component of the electric field contributes to this integral. Assuming that the aperture is small compared to distances over which the fields in the normal modes vary, we can expand the latter and find, after changing the origin to a point in the aperture,

$$\begin{aligned} A_\lambda^{(\pm)} &= -\frac{Z_\lambda}{2} \int_{S_a} d^2x (\mathbf{n} \times \mathbf{E}_{tan}) \cdot \left[\mathbf{H}_\lambda^{(\mp)}(0) + \sum_{i,j} \boldsymbol{\epsilon}_i \frac{\partial H_{i\lambda}^{(\pm)}}{\partial x_j} \Big|_0 x_j + \dots \right] \\ &= -\frac{Z_\lambda}{2} \left\{ \frac{\mathbf{B}_\lambda^{(\mp)}(0)}{\mu} \int_{S_a} d^2x (\mathbf{n} \times \mathbf{E}_{tan}) + \sum_{i,j} \int_{S_a} d^2x (\mathbf{n} \times \mathbf{E}_{tan})_i \frac{\partial H_{i\lambda}^{(\mp)}}{\partial x_j} \Big|_0 x_j + \dots \right\}. \end{aligned} \quad (110)$$

The first term here has already an appropriate form. As for the second one, note that we can break up the integrand into even and odd pieces,

$$\begin{aligned} (\mathbf{n} \times \mathbf{E}_{tan})_i \frac{\partial H_{i\lambda}}{\partial x_j} x_j &= \frac{1}{2} [(\mathbf{n} \times \mathbf{E}_{tan})_i x_j - (\mathbf{n} \times \mathbf{E}_{tan})_j x_i] \frac{\partial H_{i\lambda}^{(\mp)}}{\partial x_j} \Big|_0 \\ &+ \frac{1}{2} [(\mathbf{n} \times \mathbf{E}_{tan})_i x_j + (\mathbf{n} \times \mathbf{E}_{tan})_j x_i] \frac{\partial H_{i\lambda}^{(\mp)}}{\partial x_j} \Big|_0. \end{aligned} \quad (111)$$

Take just the antisymmetric part of this expression and complete the integral over \mathbf{x} . The contribution to the amplitude, aside from a factor of $-Z_\lambda/2$, is

$$\begin{aligned} &\frac{1}{4} \sum_{i,j} \int_{S_a} d^2x [(\mathbf{n} \times \mathbf{E}_{tan})_i x_j - (\mathbf{n} \times \mathbf{E}_{tan})_j x_i] \left(\frac{\partial H_{i\lambda}^{(\mp)}}{\partial x_j} - \frac{\partial H_{j\lambda}^{(\mp)}}{\partial x_i} \right) \Big|_0 \\ &= \frac{1}{4} \sum_{i,j,k} \int_{S_a} d^2x [(\mathbf{n} \times \mathbf{E}_{tan})_i x_j - (\mathbf{n} \times \mathbf{E}_{tan})_j x_i] \epsilon_{kij} \left(-\frac{i\omega\epsilon}{c} E_{k\lambda}^{(\mp)}(0) \right) \\ &= -\frac{i\omega\epsilon}{4c} \int_{S_a} d^2x \mathbf{E}_\lambda^{(\mp)}(0) \cdot [2\mathbf{x} \times (\mathbf{n} \times \mathbf{E}_{tan})] \\ &= -\frac{i\omega\epsilon}{2c} \int_{S_a} d^2x \mathbf{E}_\lambda^{(\mp)}(0) \cdot [\mathbf{x} \times (\mathbf{n} \times \mathbf{E}_{tan})] = -\frac{i\omega\epsilon}{2c} \int_{S_a} d^2x \mathbf{E}_\lambda^{(\mp)}(0) \cdot \mathbf{n}(\mathbf{x} \cdot \mathbf{E}_{tan}). \end{aligned} \quad (112)$$

In this string of algebra we have made use of the fact that $\nabla \times \mathbf{H} = -i(\omega/c)\epsilon\mathbf{E}$ and that \mathbf{E}_{tan} is orthogonal to $\mathbf{E}_\lambda^{(\mp)}$ in the aperture because of the boundary conditions on the latter field.

Putting this piece into the expression for $A_\lambda^{(\pm)}$ along with the leading one, we find

$$A_\lambda^{(\pm)} = -\frac{Z_\lambda}{2} \left\{ \frac{1}{\mu} \mathbf{B}_\lambda^{(\mp)}(0) \cdot \int_{S_a} d^2x \mathbf{n} \times \mathbf{E}_{tan}(\mathbf{x}) - \frac{i\omega\epsilon}{2c} \mathbf{E}_\lambda^{(\mp)}(0) \cdot \mathbf{n} \int_{S_a} d^2x \mathbf{x} \cdot \mathbf{E}_{tan}(\mathbf{x}) \right\} \quad (113)$$

Defining

$$\mathbf{p}_{eff} \equiv \frac{\epsilon}{4\pi} \mathbf{n} \int_{S_a} d^2x \mathbf{x} \cdot \mathbf{E}_{tan}(\mathbf{x}) \quad (114)$$

and

$$\mathbf{m}_{eff} \equiv \frac{c}{2\pi i \mu \omega} \int_{S_a} d^2x \mathbf{n} \times \mathbf{E}_{tan}(\mathbf{x}) \quad (115)$$

we can write

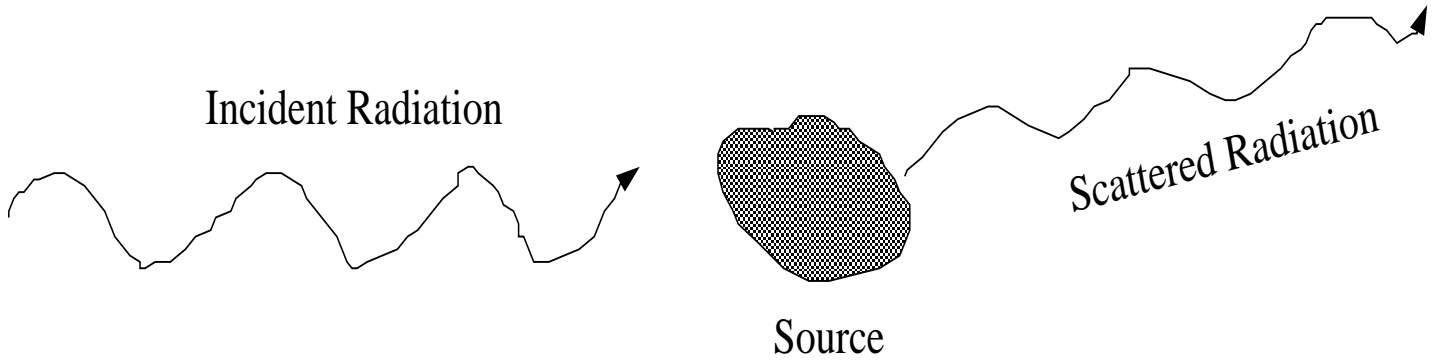
$$A_\lambda^{(\pm)} = \frac{i\pi\omega Z_\lambda}{c} [\mathbf{p}_{eff} \cdot \mathbf{E}_\lambda^{(\mp)}(0) - \mathbf{m}_{eff} \cdot \mathbf{B}_\lambda^{(\mp)}(0) + \dots] \quad (116)$$

Thus we find that in the small wavelength limit, apertures are equivalent to dipole sources. The effective dipole moments are found by solving for, or using some simple approximation for, the fields in the aperture. See Jackson for a description of the particular case of circular apertures.

5 Scattering of Radiation

So far, we have discussed the radiation produced by a harmonically moving source without discussing the origin of the source's motion. In this section we will address the fields which set the source in motion.

Consider the case where the motion is excited by some incident radiation. The incident radiation is absorbed by the source, which then begins to oscillate coherently, and thus generates new radiation. This process is generally called scattering.



5.1 Scattering of Polarized Light from an Electron

For simplicity, let's consider a plane electromagnetic wave incident upon a single electron of charge $-e$. We will assume that the incident field has the form

$$\mathbf{E}_{in}(\mathbf{x}, t) = \epsilon_1 E_0 e^{i(\mathbf{k} \cdot \mathbf{x} - \omega t)} \quad (117)$$

Since the equations which govern the motion of the electron are linear, we expect the electron to respond by oscillating along ϵ_1 at the same frequency (assuming that the magnetic force on the electron is small as long as the electron's motion is nonrelativistic). This will produce a time varying electric dipole moment.

We write the electron position as

$$\mathbf{x}(t) = \mathbf{x}_0 + \delta\mathbf{x}(t) \quad (118)$$

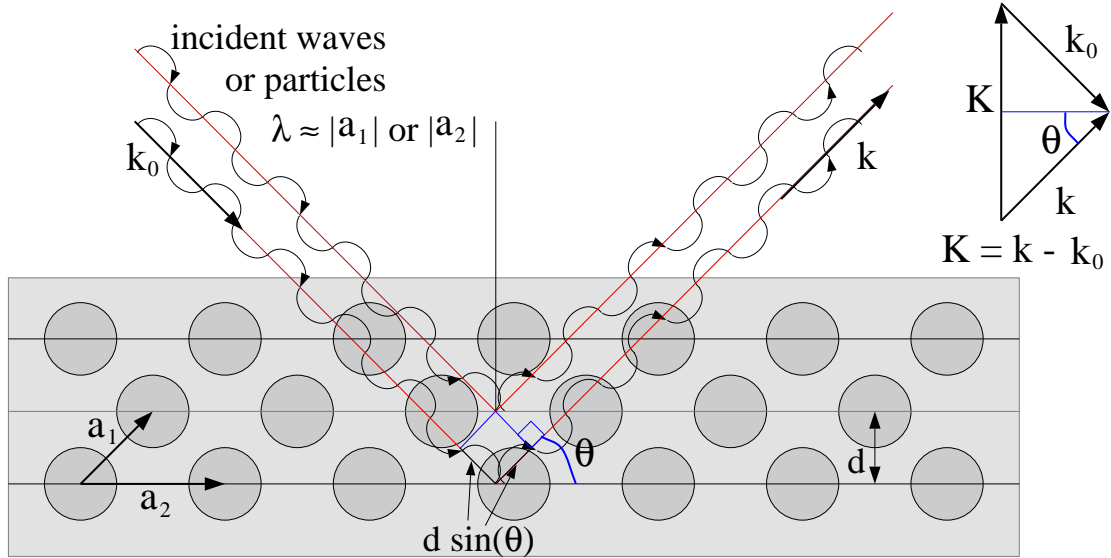
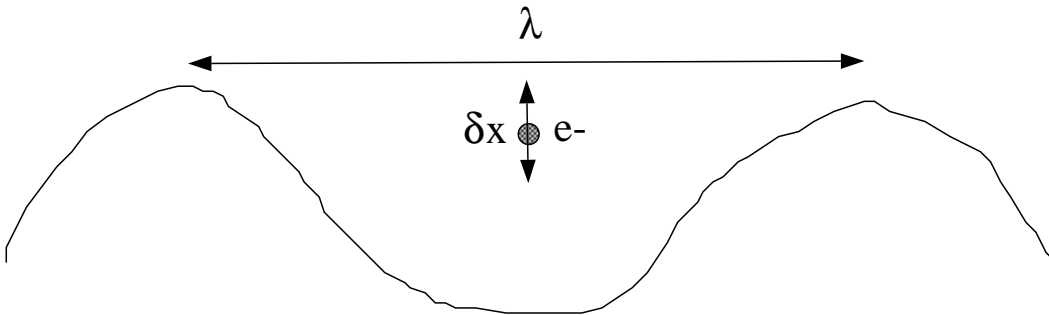


Figure 1: *Scattering of waves or particles with wavelength of roughly the same size as the lattice repeat distance allows us to learn about the lattice structure. We will assume that each electron acts as a dipole scatterer.*

If E_0 is small enough, then $\delta \mathbf{x}$ will be small compared to the wavelength λ of the incident radiation.



This approximation is also consistent with the non-relativistic assumption. When this is case, we can write

$$e^{i(\mathbf{k} \cdot \mathbf{x} - \omega t)} \approx e^{i(\mathbf{k} \cdot \mathbf{x}_0 - \omega t)} \quad (119)$$

so that the field of the incident radiation does not change over the distance traveled by the oscillating electron. If we assume a harmonic form for the electronic motion

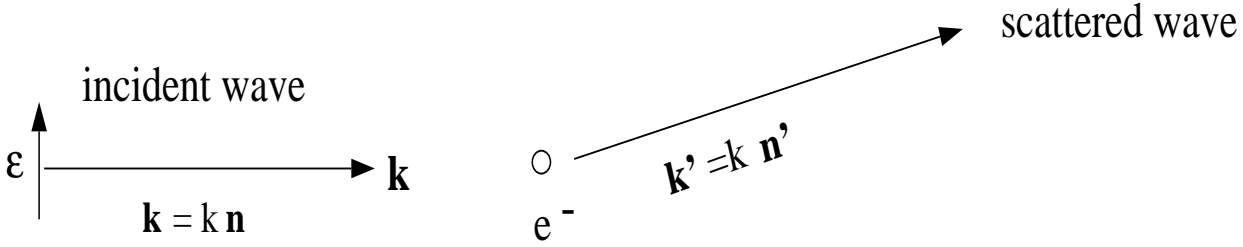
$\delta\mathbf{x}(t) = \delta\mathbf{x}e^{-i\omega t}$, insert this into Newton's law¹⁰ and solve, we get

$$\delta\mathbf{x} = \boldsymbol{\epsilon}_1 \left(\frac{eE_0 e^{i\mathbf{k}\cdot\mathbf{x}_0}}{m\omega^2} \right) \quad (120)$$

The dipole moment of the oscillating electron will then be $\mathbf{p}(t) = \mathbf{p}e^{-i\omega t}$, with

$$\mathbf{p} = -e\delta\mathbf{x} = \boldsymbol{\epsilon}_1 \left(\frac{-e^2 E_0 e^{i\mathbf{k}\cdot\mathbf{x}_0}}{m\omega^2} \right) \quad (121)$$

Now we will solve for the scattered radiation. Let the geometry of the scattering be as shown below.



From our electric dipole formulas, the angular distribution of the scattered power Eq. (45) is

$$\frac{d\mathcal{P}_{scat}}{d\Omega} = r^2 \langle \mathbf{S} \rangle \cdot \mathbf{n}' = \frac{ck^4}{8\pi} [|\mathbf{p}|^2 - |\mathbf{n}' \cdot \mathbf{p}|^2] \quad (122)$$

or

$$\frac{d\mathcal{P}_{scat}}{d\Omega} = \frac{e^4 E_0^2}{8\pi m^2 c^3} [1 - (\mathbf{n}' \cdot \boldsymbol{\epsilon}_1)^2] \quad (123)$$

Note that gives zero radiated power for \mathbf{n}' along $\boldsymbol{\epsilon}_1$. I.e. there is no power radiated in the direction along which the dipole oscillates.

¹⁰ $\mathbf{F} = m\mathbf{a} \rightarrow -e\mathbf{E} = -e\boldsymbol{\epsilon}_1 E_0 e^{i\mathbf{k}\cdot\mathbf{x}_0} = m\delta\ddot{\mathbf{x}}(t) = -m\omega^2 \delta\mathbf{x}$

We may determine the differential cross section for scattering by dividing the radiated power per solid angle obtained above by the incident power per unit area (the incident Poynting vector).

$$\mathbf{S}_{in} = \frac{c}{8\pi} (\mathbf{E} \times \mathbf{B}^*) = \frac{c}{8\pi} E_0^2 \quad (124)$$

thus

$$\frac{d\sigma}{d\Omega} = \frac{1}{|\mathbf{S}_{in}|} \frac{d\mathcal{P}_{scat}}{d\Omega} = \frac{e^4}{m^2 c^4} (1 - (\mathbf{n}' \cdot \boldsymbol{\epsilon}_1)^2) \quad (125)$$

or

$$\frac{d\sigma}{d\Omega} = r_0^2 (1 - (\mathbf{n}' \cdot \boldsymbol{\epsilon}_1)^2) \quad (126)$$

where $r_0 = e^2/mc^2 = 2.8 \times 10^{-13}$ cm is the classical scattering radius of an electron. This is the formula for Thomson scattering of incident light polarized along $\boldsymbol{\epsilon}_1$.

5.2 Scattering of Unpolarized Light from an Electron

Now suppose that the incident light is unpolarized. Then the cross section is the average of the cross sections for the two possible polarizations $\boldsymbol{\epsilon}_1$ and $\boldsymbol{\epsilon}_2$ of the incident wave (why?).

$$\left(\frac{d\sigma}{d\Omega} \right)_{unpol} = \frac{1}{2} \sum_{i=1}^2 r_0^2 (1 - (\mathbf{n}' \cdot \boldsymbol{\epsilon}_i)^2) \quad (127)$$

$$= r_0^2 - \frac{1}{2} \sum_{i=1}^2 r_0^2 ((\mathbf{n}' \cdot \boldsymbol{\epsilon}_i)^2). \quad (128)$$

From the fact that $(\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2, \text{ and } \mathbf{n})$ form an orthonormal triad of unit vectors, one may show that

$$\left(\frac{d\sigma}{d\Omega} \right)_{unpol} = \frac{1}{2} r_0^2 (1 + (\mathbf{n}' \cdot \mathbf{n})^2) \quad (129)$$

This may be rewritten as

$$\left(\frac{d\sigma}{d\Omega} \right)_{unpol} = \frac{1}{2} r_0^2 (1 + \cos^2(\theta)) \quad (130)$$

where θ is the scattering angle of the unpolarized radiation. From this we can see that the scattering is predominantly forward ($\theta = 0$), and backscattering ($\theta = \pi$).

The total cross sections may be obtained by integrating the differential cross sections. Thus

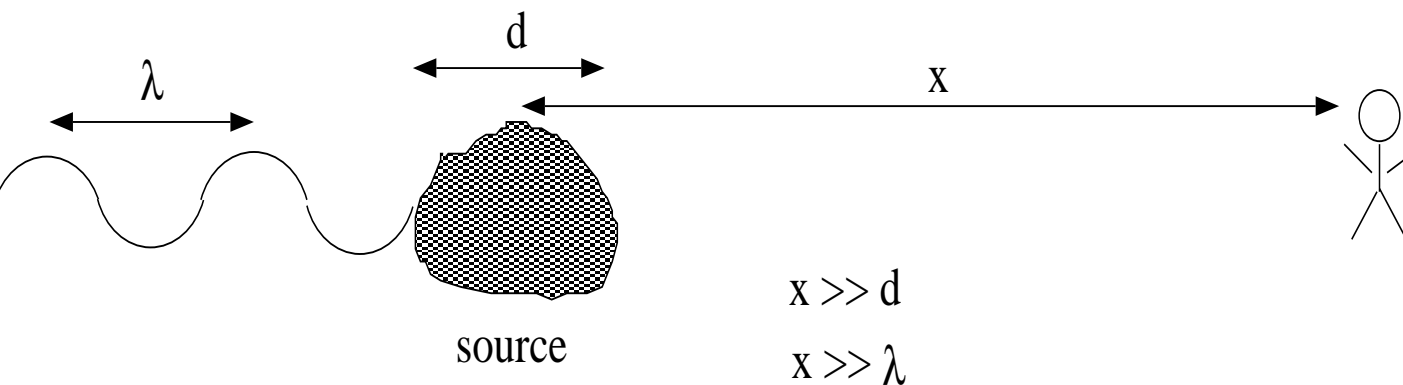
$$\sigma_{pol} = \frac{8\pi}{3}r_0^2 \quad (131)$$

and

$$\sigma_{unpol} = \frac{1}{2}(\sigma_{pol}(\epsilon_1) + \sigma_{pol}(\epsilon_2)) = \sigma_{pol} \quad (132)$$

5.3 Elastic Scattering From a Molecule

Let us now consider the elastic scattering from a molecule¹¹ in which many electrons are generally present. In calculating the field produced by the molecular electrons when light is incident on them, we will again assume that we are making our observation very far away from the molecule, so that



We will *not* assume however, that $\lambda \gg d$, since we want to use the scattering to learn something about the structure of the molecule. This is only possible if our resolution (limited by λ) is smaller than d . Thus we expect our results to include interference effects from different charges in the molecule

¹¹Here we shall use molecule to identify any small collection of charges

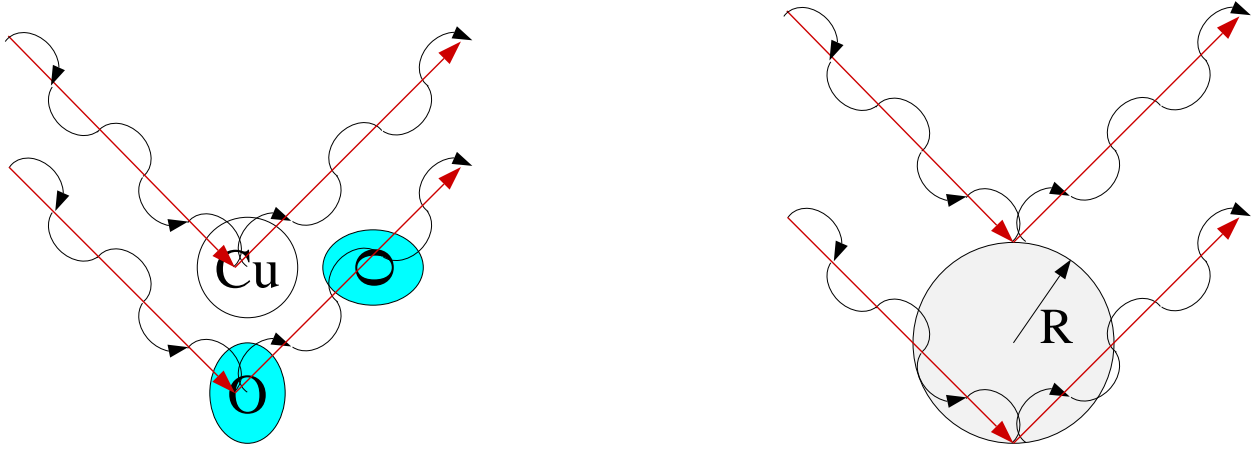


Figure 2: Rays scattered from different elements of the basis, and from different places on the atom, interfere giving the scattered intensity additional structure described by the form factor S and the atomic form factor f , respectively.

From page 9, the vector potential produced by the i 'th electron in the molecule is

$$\mathbf{A}_i(\mathbf{x}) = \frac{e^{ikr}}{cr} \int d^3x' \mathbf{J}_i(\mathbf{x}') e^{-ik(\mathbf{n}\cdot\mathbf{x}')}. \quad (133)$$

This expression includes just the current of the i 'th electron, and assumes $x \gg \lambda$ and $x \gg d$, as well as a harmonic time dependence. To find the current $\mathbf{J}_i(\mathbf{x}')$ for the i 'th electron, we note that the classical current density of an electron is

$$\mathbf{J}_i(\mathbf{x}', t) = -e\mathbf{v}_i(t)\delta(\mathbf{x}' - \mathbf{x}_i) \quad (134)$$

where $(\mathbf{x}_i, \mathbf{v}_i)$ are the location and velocity of the i 'th electron. If the electron is exposed to an incident electromagnetic plane wave, we find that

$$\mathbf{x}_i(t) = \mathbf{x}_{i,0} + \delta\mathbf{x}_i(t) \quad (135)$$

$$\delta\mathbf{x}_i(t) = \boldsymbol{\epsilon}_1 \left(\frac{eE_0}{m\omega^2} e^{i\mathbf{k}\cdot\mathbf{x}_i} e^{-i\omega t} \right) \quad (136)$$

From this we can find the electron velocity, and hence the current

$$\mathbf{J}_i(\mathbf{x}') = \boldsymbol{\epsilon}_1 \left(\frac{ie^2E_0}{m\omega} e^{i\mathbf{k}\cdot\mathbf{x}_i} \delta(\mathbf{x}' - \mathbf{x}_i) \right). \quad (137)$$

In this we have taken out the harmonic time term. Thus the vector potential of the scattered field generated by the i 'th electron in the molecule is

$$\mathbf{A}_i(\mathbf{x}) = \epsilon_1 \frac{ie^2 E_0}{m\omega} \frac{e^{ikr}}{cr} e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{x}_i} \quad (138)$$

where $\mathbf{k}' = k\mathbf{n}'$ is the scattered wave vector and $\mathbf{k} = k\mathbf{n}$ is the incident wave vector. We define the vector $\mathbf{q} = \mathbf{k} - \mathbf{k}'$ which is $(1/\hbar)$ times the momentum transfer from the photon to the electron, then the total vector potential of the scattered field is

$$\mathbf{A}(\mathbf{x}) = \sum_i \mathbf{A}_i(\mathbf{x}) = \epsilon_1 \frac{ie^2 E_0}{m\omega} \frac{e^{ikr}}{cr} \sum_i e^{i\mathbf{q}\cdot\mathbf{x}_i} \quad (139)$$

This differs from the vector potential due to a single scattering electron only through the factor $\sum_i e^{i\mathbf{q}\cdot\mathbf{x}_i}$. Since this factor is independent of the observation point \mathbf{x} , the resulting \mathbf{B} and \mathbf{E} fields are likewise those of a single scattering multiplied by a factor of $\sum_i e^{i\mathbf{q}\cdot\mathbf{x}_i}$. Thus we may immediately write down the differential cross sections. For polarized light

$$\left(\frac{d\sigma}{d\Omega}\right)_{pol} = [r_0^2 (1 - (\mathbf{n}' \cdot \epsilon_1)^2)] \left[\left| \sum_i e^{i\mathbf{q}\cdot\mathbf{x}_i} \right|^2 \right] \quad (140)$$

where the first term in brackets is the single-electron result, and the second term is the structure factor. For unpolarized light

$$\left(\frac{d\sigma}{d\Omega}\right)_{unpol} = \left[\frac{1}{2} r_0^2 (1 + (\mathbf{n}' \cdot \mathbf{n})^2) \right] \left[\left| \sum_i e^{i\mathbf{q}\cdot\mathbf{x}_i} \right|^2 \right] \quad (141)$$

We see that from a measurement of the differential cross section, we may learn about the structure of the object from which we are scattering light. Let's examine the structure factor in more detail.

$$\sum_i e^{i\mathbf{q}\cdot\mathbf{x}_i} = \int_V d^3x \sum_i e^{i\mathbf{q}\cdot\mathbf{x}_i} \delta(\mathbf{x} - \mathbf{x}_i) \quad (142)$$

$$= \int_V d^3x e^{i\mathbf{q}\cdot\mathbf{x}} \sum_i \delta(\mathbf{x} - \mathbf{x}_i) \quad (143)$$

However,

$$\sum_i \delta(\mathbf{x} - \mathbf{x}_i) = n_{el}(\mathbf{x}) \quad (144)$$

which is the electron number density at position \mathbf{x} , thus

$$\sum_i e^{i\mathbf{q}\cdot\mathbf{x}_i} = \int_V d^3x e^{i\mathbf{q}\cdot\mathbf{x}} n_{el}(\mathbf{x}) = n_{el}(-\mathbf{q}) \quad (145)$$

where $n_{el}(-\mathbf{q})$ is the Fourier transform of the electron density. Thus, our differential cross sections are, for polarized light:

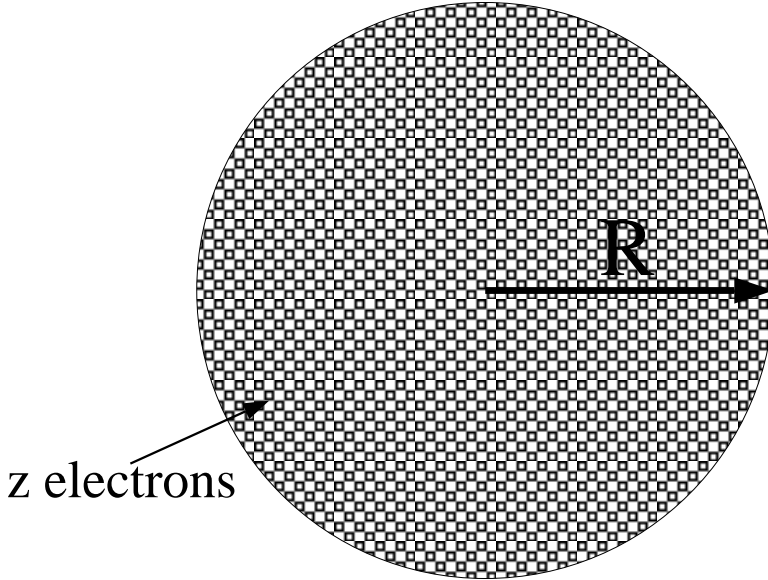
$$\left(\frac{d\sigma}{d\Omega}\right)_{pol} = [r_0^2 (1 - (\mathbf{n}' \cdot \boldsymbol{\epsilon}_1)^2)] [|n_{el}(-\mathbf{q})|^2], \quad (146)$$

and for unpolarized light

$$\left(\frac{d\sigma}{d\Omega}\right)_{unpol} = \left[\frac{1}{2}r_0^2 (1 + (\mathbf{n}' \cdot \mathbf{n})^2)\right] [|n_{el}(-\mathbf{q})|^2] \quad (147)$$

Since $\mathbf{q} = \mathbf{k} - \mathbf{k}'$ depends upon the scattering angle and the wave number of the incident photon, a scan over θ and/or \mathbf{k} gives you information about the electron density on the target.

5.3.1 Example: Scattering Off a Hard Sphere



Single spherical molecule with a sharp cutoff in its electron distribution

Consider scattering from a single molecule with a sharp cutoff in its electron distribution.

$$n_{el}(\mathbf{x}) = \frac{3z}{4\pi R^3} \theta(R - |\mathbf{x}|) \quad (148)$$

where R is the radius of the sphere and z is the number of electrons, then

$$n_{el}(-\mathbf{q}) = \int_V d^3x e^{i\mathbf{q}\cdot\mathbf{x}} n_{el}(\mathbf{x}) = \frac{3z}{q^3 R^3} [\sin(qR) - qR \cos(qR)] , \quad (149)$$

where $j_1(x) = (\sin x/x - \cos x)$ is a spherical Bessel function. Note that this has zeroes at

$$qR = \tan(qR) \quad (150)$$

so that we expect zeroes in the scattering cross section. These are clearly related to the size of the molecule. For forward scattering ($q = 2k \sin(0) = 0$), we have

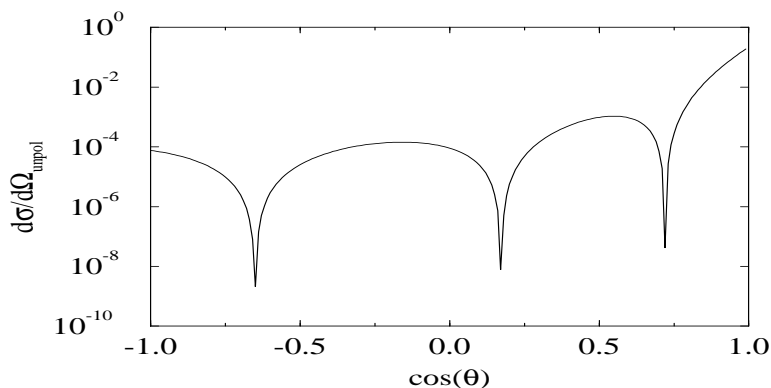
$$n_{el}(-\mathbf{q} = 0) = z \quad (151)$$

This is the maximum value, i.e. all the scattering electrons give constructive interference for forward scattering.

A sketch of the unpolarized cross section

$$\left(\frac{d\sigma}{d\Omega} \right)_{unpol} = \frac{1}{2} \mathbf{r}_0^2 (1 + \cos^2(\theta)) |n_{el}(-\mathbf{q})|^2$$

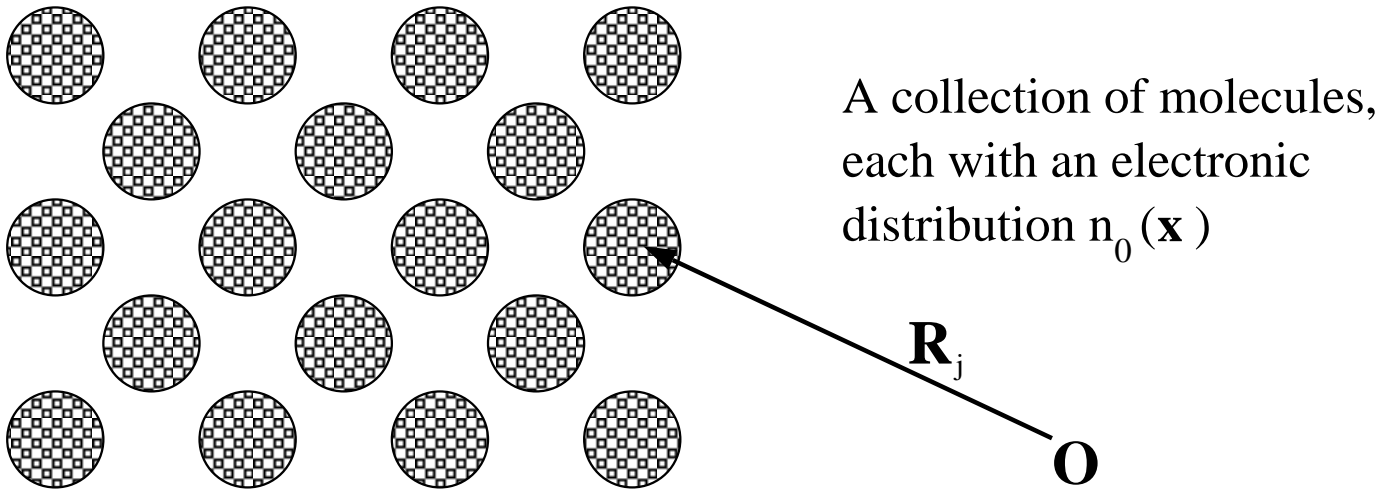
looks like



If we measured this, we could learn the spatial distribution of electrons by inverting the transform. It is important to note that we could not learn the details of the

electron wavefunctions since all the phase information is lost. I.e., the cross section just measures the modulus squared of n_{el} , and so phase information is lost.

5.3.2 Example: A Collection of Molecules



In a macroscopic sample, the total electron density is made up of two parts: the electron distribution within the molecules, and the distribution of the molecules within the sample. For simplicity, let's assume that the sample contains just one type of molecule, each of which has an electron density $n_0(\mathbf{x})$. The total electron number density is then

$$n_{el}(\mathbf{x}) = \sum_j n_0(\mathbf{x} - \mathbf{R}_j) \quad (152)$$

where \mathbf{R}_j is the position of the j 'th molecule. The Fourier transform of this is

$$n_{el}(-\mathbf{q}) = \int_V d^3x e^{i\mathbf{q}\cdot\mathbf{x}} n_{el}(\mathbf{x}) = \sum_j e^{i\mathbf{q}\cdot\mathbf{R}_j} n_0(-\mathbf{q}) \quad (153)$$

which is a weighted sum of the Fourier transforms for each molecule. We see that then the cross section will include the factor

$$|n_0(-\mathbf{q})|^2 \left| \sum_j e^{i\mathbf{q}\cdot\mathbf{R}_j} \right|^2 \quad (154)$$

From above, we know that $n_0(-\mathbf{q} = 0) = z$, the number of electrons in a molecule, so we may write

$$n_0(-\mathbf{q}) = zF(\mathbf{q}) \quad (155)$$

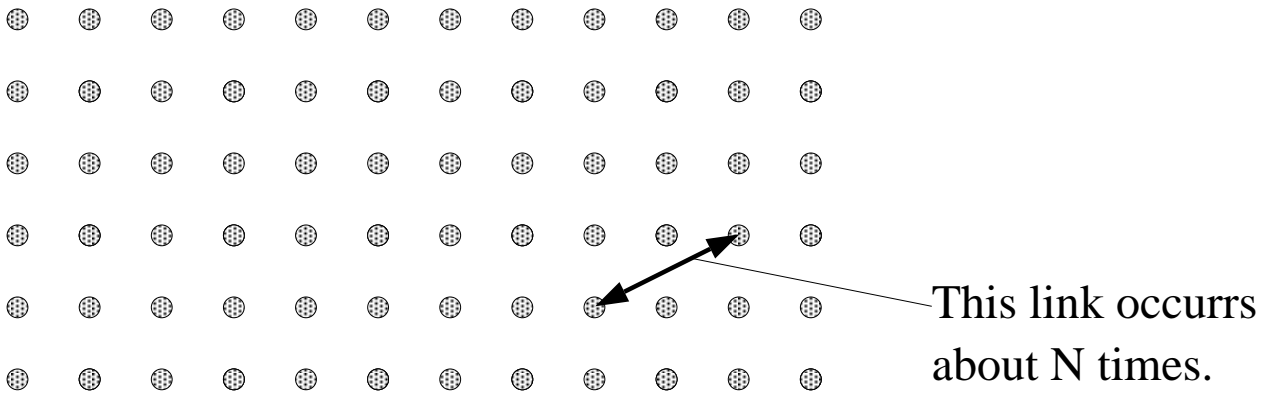
where $F(\mathbf{q})$ is called the form factor of the molecule which is normalized to unity for $\mathbf{q} = 0$. Thus the unpolarized differential cross section becomes

$$\left(\frac{d\sigma}{d\Omega} \right)_{unpol} = \frac{1}{2} z^2 r_0^2 (1 + \cos^2(\theta)) |F(\mathbf{q})|^2 \left| \sum_j e^{i\mathbf{q}\cdot\mathbf{R}_j} \right|^2 \quad (156)$$

In general, we will not know where all the molecules are, nor do we necessarily care. Thus we will look at an average of the term which describes the distribution of the molecules.

$$\left\langle \left| \sum_j e^{i\mathbf{q}\cdot\mathbf{R}_j} \right|^2 \right\rangle = \sum_{jj'} \langle e^{i\mathbf{q}\cdot(\mathbf{R}_j - \mathbf{R}_{j'})} \rangle \approx N \sum_j \langle e^{i\mathbf{q}\cdot(\mathbf{R}_j - \mathbf{R}_0)} \rangle \equiv NS(\mathbf{q}) \quad (157)$$

where N is the number of molecules in the sample, \mathbf{R}_0 is the position of the origin, and $S(\mathbf{q})$ is the structure factor of the sample. The approximation is justified by the fact that (neglecting finite-size effects) each link between sites occurs about N times.



$$S(\mathbf{q}) = \sum_j \langle e^{i\mathbf{q}\cdot(\mathbf{R}_j - \mathbf{R}_0)} \rangle = 1 + \sum_{j \neq 0} \langle e^{i\mathbf{q}\cdot(\mathbf{R}_j - \mathbf{R}_0)} \rangle \quad (158)$$

Note that this goes to 1 as $|\mathbf{q}| \rightarrow \infty$. Then, using the same procedure detailed before, this becomes

$$NS(\mathbf{q}) = V \int_V d^3x e^{i\mathbf{q}\cdot\mathbf{x}} \langle n(\mathbf{x})n(0) \rangle \quad (159)$$

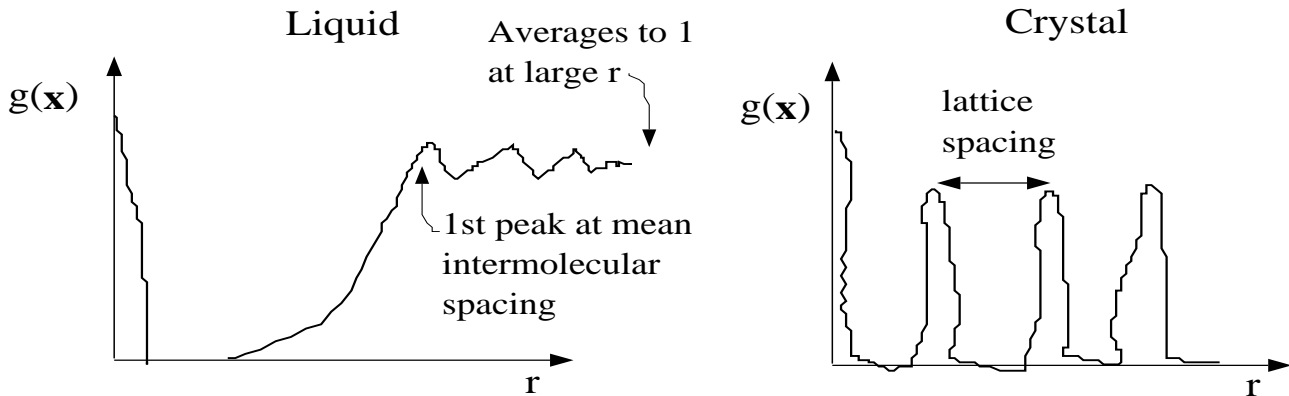
where V is the sample volume, and $n(\mathbf{x})$ is the number density of molecules in the sample.

Here, $\langle n(\mathbf{x})n(0) \rangle$ tell us about correlations between molecular positions.

$$\langle n(\mathbf{x})n(0) \rangle = (n(0))^2 g(\mathbf{x}) \quad (160)$$

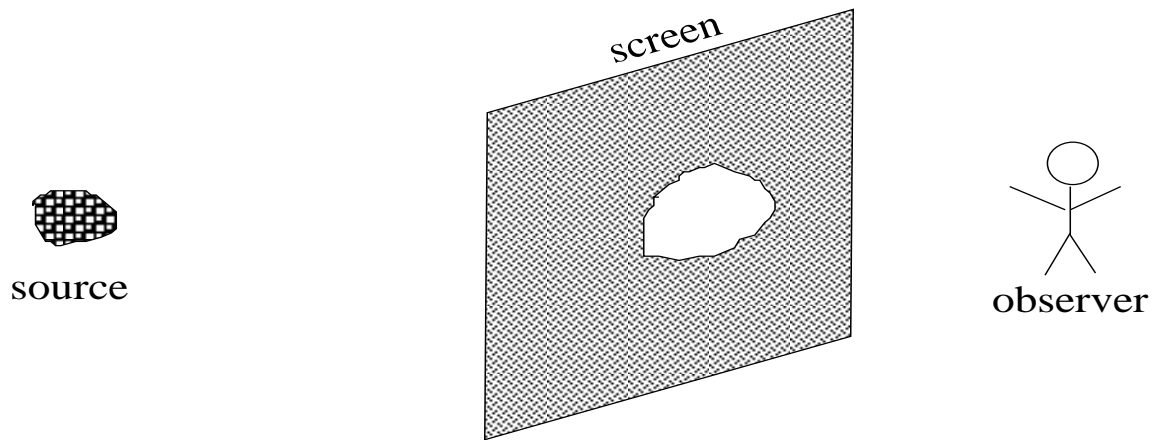
where $g(\mathbf{x})$ is just the probability of finding a molecule at \mathbf{x} if there is one at the origin. This is the normalized two-body correlation function. We see that if the molecular form factor is known, then a measurement of the differential cross section tells us about the Fourier transform of $g(\mathbf{x})$. This is the key to using x-ray (or neutron, etc) scattering to determine the internal structure of materials.

$$\left(\frac{d\sigma}{d\Omega} \right)_{unpol} = \frac{1}{2} z^2 r_0^2 (1 + \cos^2(\theta)) |F(\mathbf{q})|^2 V (n(0))^2 g(\mathbf{q})$$



6 Diffraction

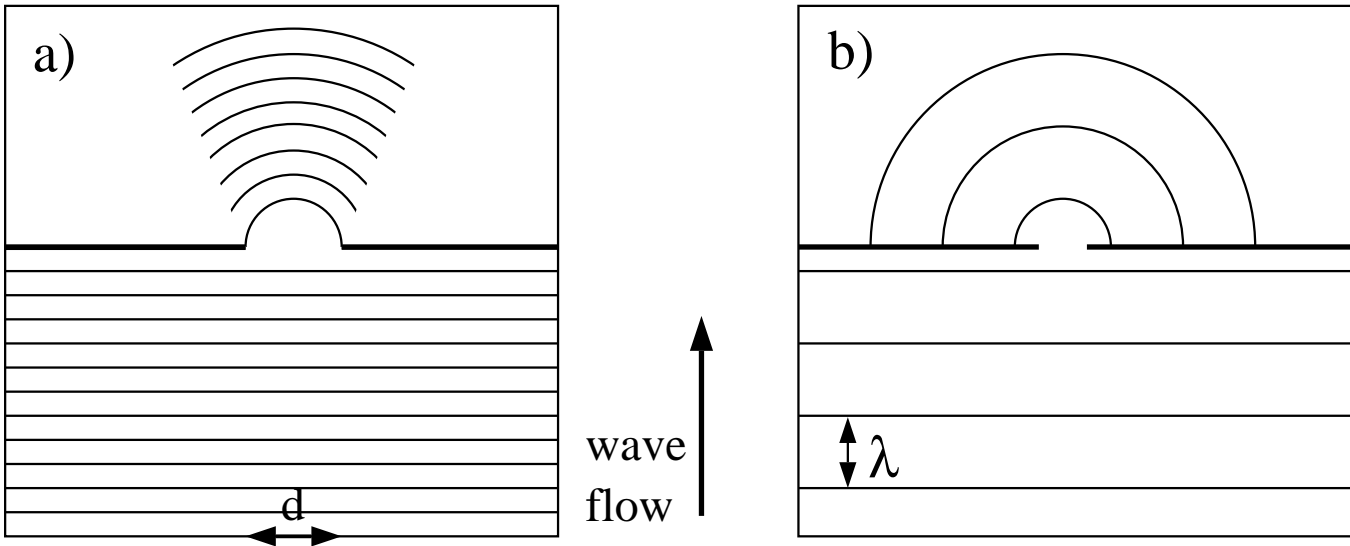
In this section we will discuss diffraction which is related to scattering. The prototypical setup is shown below in which radiation from a source is diffracted through an opaque screen with one or more holes in it.



Clearly, diffraction is the study of the propagation of light or radiation, or rather the deviation of light from rectilinear propagation. As undergraduates we all learned that the propagation of light was governed by **Huygen's Principle** that *every point on a primary wavefront serves as the source of spherical secondary wavelets such that the primary wavefront at some later time is the envelope of these wavelets. Moreover, the wavelets advance with a speed and frequency equal to that of the primary wave at each point in space.*¹². This serves as the paradigm for our study of geometric optics. Diffraction is the first crisis in this paradigm.

To see this consider wave propagation in a ripple tank through an aperture.

¹²Hecht-Zajac, page 60



In the figure, wavefronts are propagating toward the aperture from the bottom of each box. In case a) the wavelength λ is much smaller than the size of the aperture d . In this case, the diffracted waves interfere destructively except immediately in front of the aperture. In case b), $\lambda \gg d$, and no such interference is observed.

Huygen's principle cannot explain the difference between cases a) and b), since it is independent of any wavelength considerations, and thus would predict the same wavefront in each case. The difficulty was resolved by Fresnel. The corresponding **Huygens-Fresnel principle** states that *every unobstructed point of a wavefront, at a given instant of in time, serves as a source of spherical secondary wavelets of the same frequency as the source. The amplitude of the diffracted wave is the sum of the wavelets considering their amplitudes and relative phases*¹³. Applying these ideas clarifies case a). Here what is happening is that the wavelets from the right and left sides of the aperture interfere constructively in front of the aperture (since they travel the same distance and hence remain in phase), whereas these wavelets interfere destructively to the sides of the aperture (since they travel two paths with length differences of order $\lambda/2$). In case b), we approach the limit of a single point source

¹³Hecht-Zajac, page 330

of spherical waves. Beside being rather hypothetical, the Huygens-Fresnel principle involves some approximations which we will discuss later; however, Kirchoff showed that the Huygen's-Fresnel principle is a direct consequence of the wave equation.

6.1 Scalar Diffraction Theory: Kirchoff Approximation

This problem could be solved using the techniques we just developed to treat scattering. I.e. by considering the dynamics of the charged particles in the screen, and then calculating the scattered radiation generated by these particles. However, diffraction is conventionally treated as a boundary value problem in which the presence of the screen is taken into account with boundary conditions on the wave.

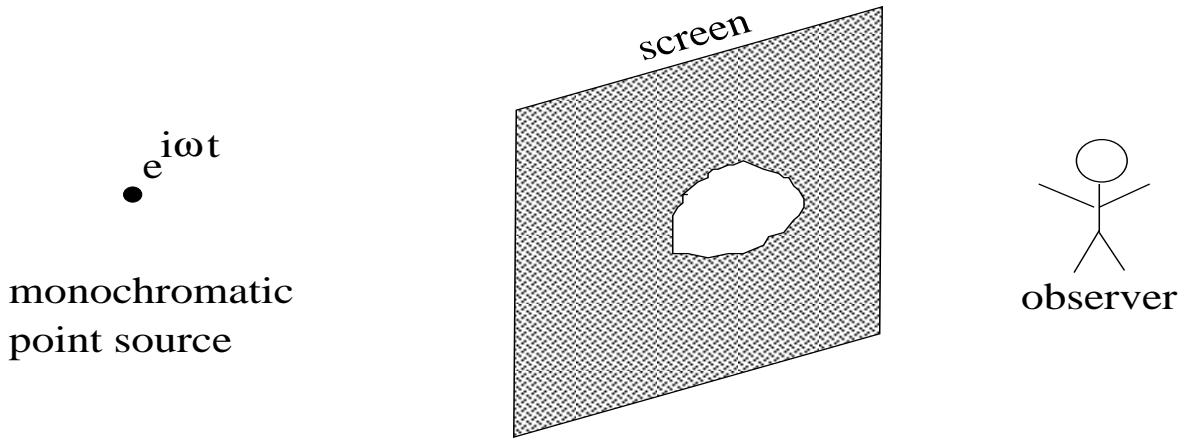
Several excellent references for this problem are worth noting:

1. B.B. Baker and E.T. Copson, **The Mathematical Theory of Huygens Principle**, (Clarendon Press, Oxford, 1950).
2. Landau and Lifshitz, **The Classical Theory of Fields**.
3. L. Eyges, **The Classical Electromagnetic Field**, (Addison-Wesley, Reading, 1972).
4. Hecht-Zajac, **Optics**, (Addison-Wesley, Reading, 1979).

The Baker reference, especially, has a good discussion of the limits of the Kirchoff approximation, of course, Landau and Lifshitz have an excellent discussion of the physics, but Hecht-Zajac's discussion is perhaps the most elementary, and will often be quoted here.

The typical question we ask is, given a strictly monochromatic point source S , what resulting radiation is observed at the point O on the opposite side of the screen. At first this approach may seem to limit us just to point sources. The case of a real extended source which emits non-monochromatic light does not, however, require special treatment. This is because of the linearity of our equations and the complete

independence (incoherence) of the light emitted by different points of the source. The interference terms average to zero. Thus the total diffraction pattern is simply the sum of the intensity distributions obtained from the diffraction of the independent components of the light.



To treat the theory of diffraction, a number of approximations will be necessary.

- First, we assume that we can neglect the vector nature of the electromagnetic fields, and work instead with a scalar complex function ψ (a component of \mathbf{E} or \mathbf{B} , or the single polarization observed in the ripple tank discussed above). In principle, ψ is any of the three components of either \mathbf{E} or \mathbf{B} . In practice, however, the polarization of the radiation is usually ignored and the intensity of the radiation at a point is usually taken as $|\psi(\mathbf{x}, t)|^2$. This first assumption limits the number of geometries we can treat.
- Second, we will generally assume that $\lambda/d \ll 1$, where d is the linear dimension of the aperture or obstacle.
- The third assumption is that we will only look for the first correction to geometric optics due to diffraction. This is often called the Kirchoff approximation, which will be discussed a bit later. (This set of approximations are sometimes also called the Kirchoff approximation scheme.)

We then impose the boundary condition¹⁴

$$\psi(\mathbf{x}, t) = 0 \quad \text{everywhere on the screen} \quad (161)$$

We will assume that ψ obeys the wave equation, and the source is harmonic so that it emits radiation of frequency ω , hence

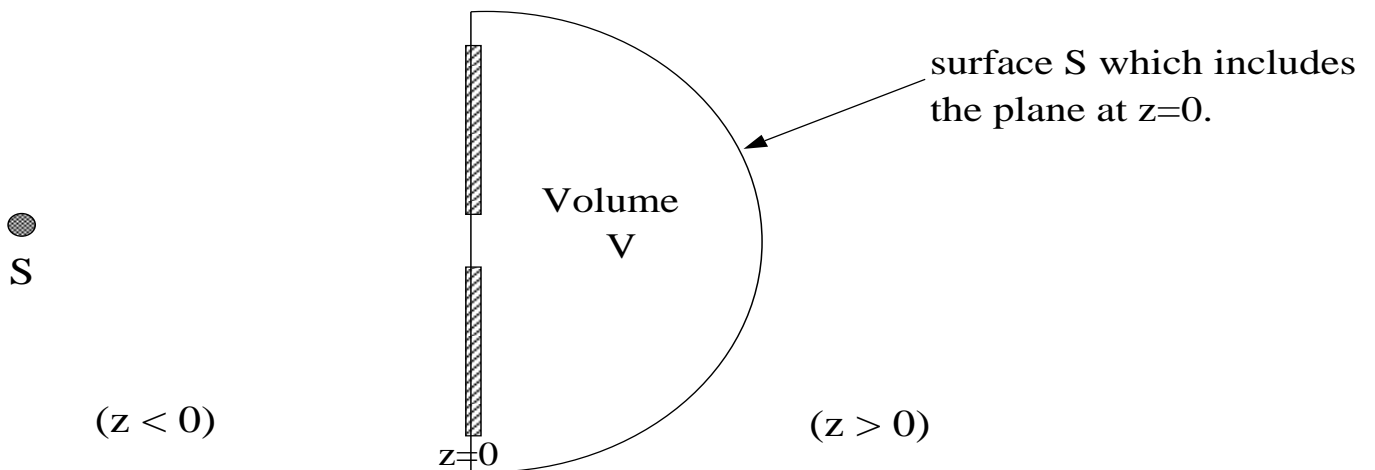
$$\psi(\mathbf{x}, t) = \psi(\mathbf{x})e^{-i\omega t}. \quad (162)$$

Thus the spatial wave function obeys the Helmholtz equation

$$\square^2\psi = (\nabla^2 + k^2)\psi(\mathbf{x}) = 0 \quad , \text{with } k = \omega/c \quad (163)$$

if we assume that the waves propagate in a homogeneous medium and restrict our observations to points away from the source S . As always, the physical quantities will be the *real* amplitude $\Re(\psi(\mathbf{x})e^{-i\omega t})$, and the modulus squared which denotes the time-averaged intensity.

Let's consider just the observation points in the volume V below which is bounded by the screen ($z = 0$ plane), and a hemisphere at infinity.



¹⁴Note that we only impose one boundary condition on the screen. This is to avoid the difficulties when both boundary conditions (Neumann and Dirichlet) are imposed, as discussed in Jackson on page 429

Then, everywhere within V , the Helmholtz equation $\square^2\psi = 0$ is obeyed since the source lies outside of V . To find ψ within V , we use Green's theorem

$$\int_V d^3x' \left(\psi(\mathbf{x}')\nabla'^2\phi(\mathbf{x}') - \phi(\mathbf{x}')\nabla'^2\psi(\mathbf{x}') \right) = \int_S d^2x' \mathbf{n}' \cdot (\psi(\mathbf{x}')\nabla'\phi(\mathbf{x}') - \phi(\mathbf{x}')\nabla'\psi(\mathbf{x}')), \quad (164)$$

or, adding and subtracting $k^2\psi(\mathbf{x}')\phi(\mathbf{x}')$ from the first integrand, we get

$$\int_V d^3x' \left(\psi(\mathbf{x}')(\nabla'^2 + k^2)\phi(\mathbf{x}') - \phi(\mathbf{x}')(\nabla'^2 + k^2)\psi(\mathbf{x}') \right) = \int_S d^2x' \mathbf{n}' \cdot (\psi(\mathbf{x}')\nabla'\phi(\mathbf{x}') - \phi(\mathbf{x}')\nabla'\psi(\mathbf{x}')) \quad (165)$$

This works for any two functions ψ and ϕ . We will take $\psi(\mathbf{x}')$ to be the wave amplitude, and take $\phi(\mathbf{x}') = G(\mathbf{x}, \mathbf{x}')$, where the Dirichlet Green's function satisfies

$$(\nabla^2 + k^2)G(\mathbf{x}, \mathbf{x}') = -4\pi\delta(\mathbf{x} - \mathbf{x}') \text{ in } V \quad (166)$$

i.e., it is the response to a unit point source so that in free space $G(\mathbf{x}, \mathbf{x}') = \frac{e^{ik|\mathbf{x}-\mathbf{x}'|}}{|\mathbf{x}-\mathbf{x}'|}$.

However, need to solve for G with boundary conditions

$$G(\mathbf{x}, \mathbf{x}') = 0 \text{ for } \mathbf{x}' \text{ on } S. \quad (167)$$

since we will be using Dirichlet boundary conditions on ψ : We will specify the value of $\psi(\mathbf{x})$ (as opposed to its derivative) on the boundary S . Now using the facts that

$$(\nabla^2 + k^2)G(\mathbf{x}, \mathbf{x}') = -4\pi\delta(\mathbf{x} - \mathbf{x}') \text{ in } V \quad (168)$$

$$(\nabla^2 + k^2)\psi(\mathbf{x}) = 0 \text{ in } V \quad (169)$$

and that

$$G(\mathbf{x}, \mathbf{x}') = 0 \text{ for } \mathbf{x}' \text{ on } S \quad (170)$$

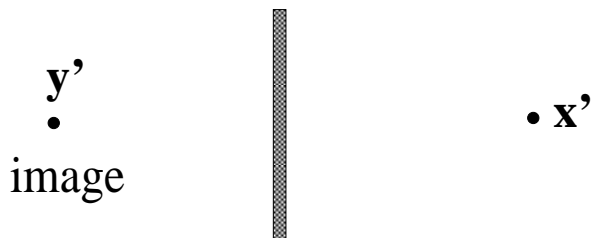
this Green's theorem becomes (the Kirchoff Integral)

$$\psi(\mathbf{x}) = -\frac{1}{4\pi} \int_S d^2x' \mathbf{n}' \cdot (\psi(\mathbf{x}')\nabla'G(\mathbf{x}, \mathbf{x}')) \quad (171)$$

To proceed further we must determine the form of $G(\mathbf{x}, \mathbf{x}')$ and we must now specify $\psi(\mathbf{x}')$ on the surface S . For an infinite planar screen the Green's function is

given by the method of images (analogously to the electrostatic case)

$$G(\mathbf{x}, \mathbf{x}') = \frac{e^{ik|\mathbf{x}-\mathbf{x}'|}}{|\mathbf{x}-\mathbf{x}'|} - \frac{e^{ik|\mathbf{x}-\mathbf{y}'|}}{|\mathbf{x}-\mathbf{y}'|} \quad (172)$$



where $\mathbf{x}' = (x', y', z')$ is in V , but $\mathbf{y}' = (x', y', -z')$ (the image point) is not. This satisfies $(\nabla^2 + k^2)G(\mathbf{x}, \mathbf{x}') = -4\pi\delta(\mathbf{x}-\mathbf{x}')$ in V , and vanishes both on the plane $z' = 0$ and on the hemisphere at infinity. Note that it vanishes as $1/r'^2$ as $r' \rightarrow \infty$. From this it is clear that the integral

$$\psi(\mathbf{x}) = -\frac{1}{4\pi} \int_S d^2x' \mathbf{n}' \cdot (\psi(\mathbf{x}') \nabla' G(\mathbf{x}, \mathbf{x}')) \quad (173)$$

has a vanishing contribution from the hemisphere at infinity, since $\psi(\mathbf{x}')$ vanishes at least as fast as $1/r'$ as $r' \rightarrow \infty$ (since it is a solution of the wave equation for a finite source). Furthermore, from the boundary condition

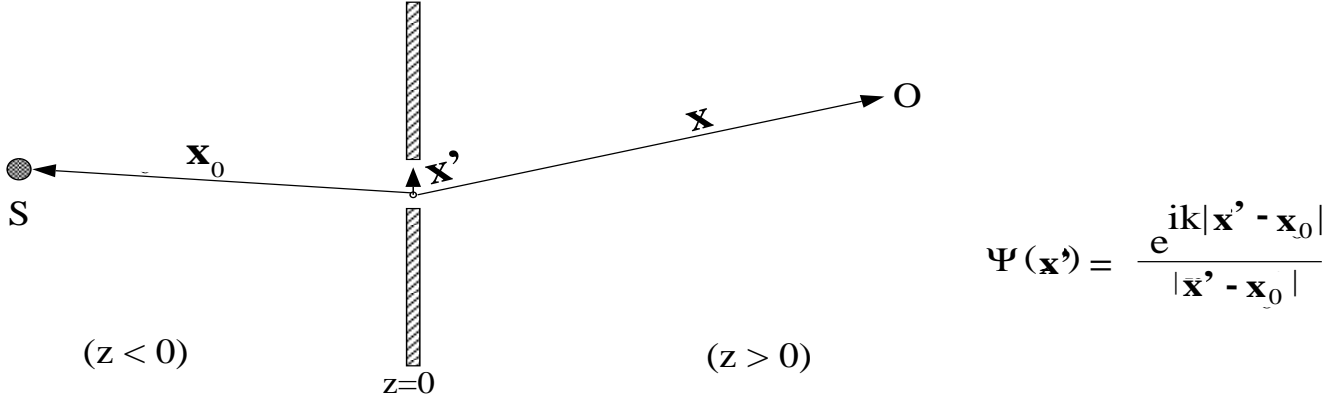
$$\psi(\mathbf{x}') = 0 \quad \text{on the screen} \quad (174)$$

we can see that the integral only gets a nonzero contribution from the opening

$$\psi(\mathbf{x}) = -\frac{1}{4\pi} \int_{\text{opening}} d^2x' \mathbf{n}' \cdot (\psi(\mathbf{x}') \nabla' G(\mathbf{x}, \mathbf{x}')) . \quad (175)$$

So far we have made an exact evaluation of our *scalar* theory. However, to proceed we must make an approximation. We will assume that the value of $\psi(\mathbf{x}')$ in the opening will be the *same* as if the screen was not there at all. This means that the wavelength we are considering must be small compared to the characteristic size of the problem (i.e. the size of the opening). As a result, our formalism will yield just the lowest-order correction due to diffraction to the results of geometrical or ray optics. This approximation is called the **Kirchoff approximation**.

Now to implement this approximation we need two things. First, we need the field strength at the opening. For a source of unit strength at position \mathbf{x}_0 , the field at position \mathbf{x}' in the opening is taken to be



which is just a spherical wave. Second, we need the component of $\nabla'G(\mathbf{x}, \mathbf{x}')$ in the direction of \mathbf{n}' in the opening. Since \mathbf{n}' is the outward normal direction from the volume V , this means we need

$$-\frac{d}{dz'}G(\mathbf{x}, \mathbf{x}') = -\frac{d}{dz'} \left(\frac{e^{ik|\mathbf{x}-\mathbf{x}'|}}{|\mathbf{x}-\mathbf{x}'|} - \frac{e^{ik|\mathbf{x}-\mathbf{y}'|}}{|\mathbf{x}-\mathbf{y}'|} \right). \quad (176)$$

Since we are using a short wavelength approximation, it follows that $k|\mathbf{x}-\mathbf{x}'| \gg 1$ and $k|\mathbf{x}-\mathbf{y}'| \gg 1$. Consequently, in taking the derivative of $G(\mathbf{x}, \mathbf{x}')$, the derivatives coming from the denominators are negligible compared to those from the exponentials.

Thus

$$-\frac{d}{dz'}G(\mathbf{x}, \mathbf{x}') \approx \frac{ik(z-z')}{|\mathbf{x}-\mathbf{x}'|} \frac{e^{ik|\mathbf{x}-\mathbf{x}'|}}{|\mathbf{x}-\mathbf{x}'|} + \frac{ik(z+z')}{|\mathbf{x}-\mathbf{y}'|} \frac{e^{ik|\mathbf{x}-\mathbf{y}'|}}{|\mathbf{x}-\mathbf{y}'|}. \quad (177)$$

We only need to evaluate this in the opening ($z'=0$)

$$-\frac{d}{dz'}G(\mathbf{x}, \mathbf{x}') \Big|_{z'=0} \approx \frac{2ikz}{|\mathbf{x}-\mathbf{x}'|} \frac{e^{ik|\mathbf{x}-\mathbf{x}'|}}{|\mathbf{x}-\mathbf{x}'|} \quad (178)$$

since $\mathbf{x}' = \mathbf{y}'$ here.

Thus, our Kirchoff approximation yields the following expression for the field

observed at \mathbf{x}

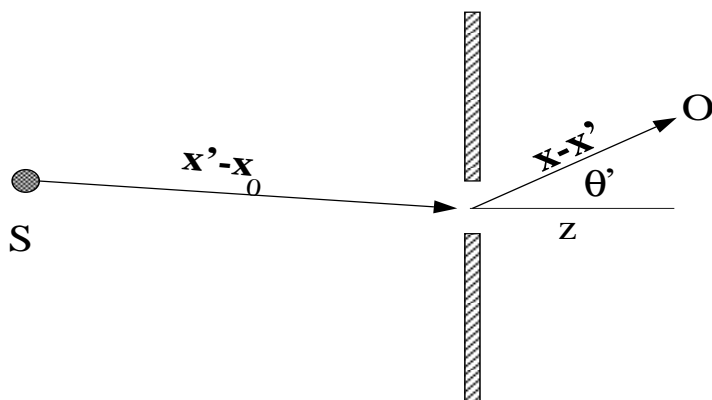
$$\psi(\mathbf{x}) = -\frac{1}{4\pi} \int_{\text{opening}} d^2x' \mathbf{n}' \cdot (\psi(\mathbf{x}') \nabla G(\mathbf{x}, \mathbf{x}')) \quad (179)$$

$$= -\frac{1}{4\pi} \int_{\text{opening}} d^2x' \frac{2ikz}{|\mathbf{x} - \mathbf{x}'|} \frac{e^{ik|\mathbf{x}-\mathbf{x}'|}}{|\mathbf{x} - \mathbf{x}'|} \frac{e^{ik|\mathbf{x}_0-\mathbf{x}'|}}{|\mathbf{x}_0 - \mathbf{x}'|}. \quad (180)$$

We may also write

$$\frac{z}{|\mathbf{x} - \mathbf{x}'|} = \cos(\theta'), \quad (181)$$

where θ' is the angle from the normal to the opening at integration point \mathbf{x}' to the observation point \mathbf{x} .



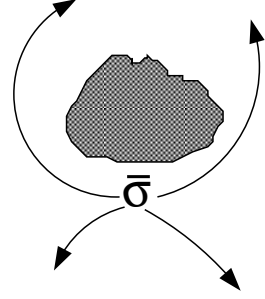
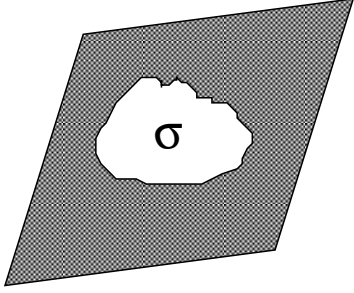
Note that θ' is different for each integration point \mathbf{x}' . Thus our expression becomes

$$\psi(\mathbf{x}) = -\frac{ik}{2\pi} \int_{\text{opening}} d^2x' \frac{e^{ik|\mathbf{x}_0-\mathbf{x}'|}}{|\mathbf{x}_0 - \mathbf{x}'|} \frac{e^{ik|\mathbf{x}-\mathbf{x}'|}}{|\mathbf{x} - \mathbf{x}'|} \cos(\theta') \quad (182)$$

The first term is the field from the source located at \mathbf{x}_0 and received at point \mathbf{x}' in the opening. The second term is the field from the “source” at \mathbf{x}' in the opening received at observation point \mathbf{x} . The third term is the inclination factor. Thus, this integral is just an expression of the Huygens-Fresnel principle.

6.2 Babinet’s Principle

Let’s explore the consequences of this formula. First, consider two different screens which are complimentary.



One has an opening, labeled σ , while the other is just a disk. For the second screen the "opening" is the entire plane $z = 0$ except for the disk. This opening is the complement of σ , so we will call it $\bar{\sigma}$.

We notice something interesting about the amplitudes received at the point O in the two complementary cases, we have

$$\psi(\mathbf{x}) = -\frac{1}{4\pi} \int_{\sigma} d^2x' \mathbf{n}' \cdot (\psi(\mathbf{x}') \nabla G(\mathbf{x}, \mathbf{x}')) \quad (183)$$

$$\overline{\psi(\mathbf{x})} = -\frac{1}{4\pi} \int_{\bar{\sigma}} d^2x' \mathbf{n}' \cdot (\psi(\mathbf{x}') \nabla G(\mathbf{x}, \mathbf{x}')) \quad (184)$$

The *sum* of these two amplitudes is then

$$\psi(\mathbf{x}) + \overline{\psi(\mathbf{x})} = -\frac{1}{4\pi} \int_{\bar{\sigma} + \sigma} d^2x' \mathbf{n}' \cdot (\psi(\mathbf{x}') \nabla G(\mathbf{x}, \mathbf{x}')) \quad (185)$$

However, $\bar{\sigma} + \sigma$ represents the entire plane $z = 0$. Thus $\psi(\mathbf{x}) + \overline{\psi(\mathbf{x})}$ represents the amplitude detected at \mathbf{x} if there was no screen at all. In other words $\psi(\mathbf{x}) + \overline{\psi(\mathbf{x})}$ includes no diffraction.

To interpret this, note that $\psi(\mathbf{x})$ may be written as

$$\psi(\mathbf{x}) = f \frac{e^{ik|\mathbf{x}-\mathbf{x}_0|}}{|\mathbf{x}-\mathbf{x}_0|} + \psi_{diff}(\mathbf{x}) \quad (186)$$

where f is a ray optic amplitude. If $f = 0$, then there is no line-of-sight from the observer to the source, while if $f = 1$, then there is. The second term is the amplitude due to diffraction of the wave. Returning to our amplitudes $\psi(\mathbf{x})$ and $\overline{\psi(\mathbf{x})}$, we see

that one will have $f = 0$ and the other $f = 1$, so that the sum will include a term $e^{ik|\mathbf{x}-\mathbf{x}_0|}/|\mathbf{x}-\mathbf{x}_0|$. In fact, this is all it will include, since $\psi + \bar{\psi}$ includes no diffraction. Thus, we see that the diffraction amplitudes for complimentary screens cancel. This is **Babinet's Principle**. It says that solving one diffraction problem is tantamount to solving its compliment. However, Babinet's principle does *not* say that the intensities will cancel in the two cases! We will see some consequences of Babinet's principle in what follows

6.3 Fresnel and Fraunhofer Limits

Lets return to our expression for $\psi(\mathbf{x})$ in the Kirchoff approximation. Since we assume that the wavelength is small, it follows that

$$ka \gg 1 \quad kr_0 \gg 1 \quad kr \gg 1 \quad (187)$$

where a is the typical aperture size, r_0 is the distance from the screen to the source, and r is the distance to the observer.

In general these approximations should be applied to the expression for $\psi(\mathbf{x})$ only after the integral over the opening has been carried out. This can be done for some simple geometries (as in the homework). What we will do here is to insert the above limits into the Kirchoff approximate expressions for $\psi(\mathbf{x})$. In what follows, we will restrict our attention to apertures, not discs. By doing this it follows that r' will have an upper limit a , so $kr' \gg 1$. In practice, the restriction to apertures is not a real limitation since Babinet's principle allows us to calculate the the diffraction amplitude for a disk from the diffraction amplitude for the complimentary screen.

Once again, our expression for $\psi(\mathbf{x})$ is

$$\psi(\mathbf{x}) = -\frac{ik}{2\pi} \int_{\text{opening}} d^2x' \frac{e^{ik|\mathbf{x}_0-\mathbf{x}'|}}{|\mathbf{x}_0-\mathbf{x}'|} \frac{e^{ik|\mathbf{x}-\mathbf{x}'|}}{|\mathbf{x}-\mathbf{x}'|} \cos(\theta'). \quad (188)$$

Now if a/r_0 and/or a/r are not small, then we are limited in what further approximations are possible since r' is not necessarily small compared to either r or r_0 .

This limit is called **Fresnel Diffraction**, in which the source and/or the observation points are close enough to the aperture that we must worry about the diffraction of spherical waves. This presents a difficult problem, which we will not treat (but please see the homework and Landau and Lifshitz, Classical Theory of Fields, Sec. 60).

In the opposite limit, which we will treat, is called **Fraunhofer Diffraction**. In this limit

$$ka > 1 \quad kr_0 \gg 1 \quad kr \gg 1 \quad a \ll r_0 \quad \text{and} \quad a \ll r \quad (189)$$

Note that we specify $ka > 1$ rather than $ka \gg 1$ since we want to solve for the first finite corrections to the latter inequality, so that

$$\frac{ka^2}{r_0} = ka \left(\frac{a}{r_0} \right) \ll 1 \quad \text{and} \quad \frac{ka^2}{r'} \ll 1 \quad (190)$$

Thus we are looking at plane waves instead of spherical waves in this limit.

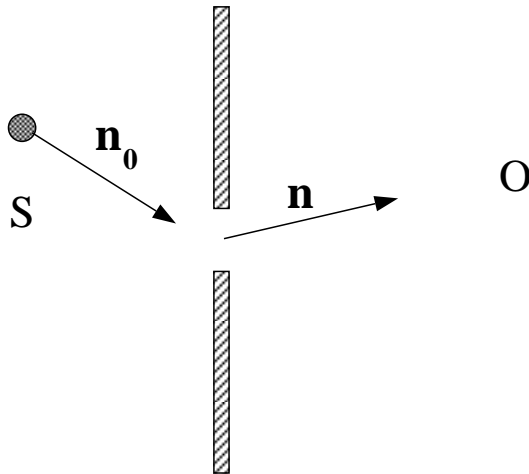
It is possible to simplify our expression for $\psi(\mathbf{x})$ even further in this limit. Then

$$|\mathbf{x}' - \mathbf{x}_0| \approx r_0 - \frac{\mathbf{x}_0 \cdot \mathbf{x}'}{r_0} = r_0 + \mathbf{n}_0 \cdot \mathbf{x}' \quad (191)$$

where \mathbf{n}_0 is a unit vector from the source to the origin (taken to be the center of the aperture), so that $\mathbf{x}_0 = -\mathbf{n}_0 r_0$. Similarly,

$$|\mathbf{x} - \mathbf{x}'| \approx r - \mathbf{n} \cdot \mathbf{x}' \quad (192)$$

where \mathbf{n} is a unit vector from the origin to the observer so that $\mathbf{x} = \mathbf{n}r$.



Thus,

$$\frac{e^{ik|\mathbf{x}'-\mathbf{x}_0|}}{|\mathbf{x}'-\mathbf{x}_0|} \frac{e^{ik|\mathbf{x}-\mathbf{x}'|}}{|\mathbf{x}-\mathbf{x}'|} \approx \frac{e^{ik(r+r_0)}}{rr_0} e^{ik(\mathbf{n}_0-\mathbf{n})\cdot\mathbf{x}'} \quad (193)$$

where we have dropped second-order terms coming from the denominators. Furthermore, since r is large compared to the aperture size, it is appropriate to take

$$\cos(\theta') = \frac{z}{|\mathbf{x}-\mathbf{x}'|} \approx 1 \quad (194)$$

inside the integral. Thus in the Fraunhofer limit

$$\psi(\mathbf{x}) = -\frac{ik}{2\pi} \frac{e^{ik(r+r_0)}}{rr_0} \int_{\text{opening}} d^2x' e^{ik(\mathbf{n}_0-\mathbf{n})\cdot\mathbf{x}'} \quad (195)$$

But the incident wave vector was $k\mathbf{n}$, so

$$k(\mathbf{n}_0-\mathbf{n}) = \mathbf{k}_{in} - \mathbf{k}_{diff} \equiv \mathbf{q} \quad (196)$$

so we finally have

$$\psi(\mathbf{x}) = -\frac{ik}{2\pi} \frac{e^{ik(r+r_0)}}{rr_0} \int_{\text{opening}} d^2x' e^{i\mathbf{q}\cdot\mathbf{x}'} \quad (197)$$

7 Example Problems

Let's consider some examples.

7.1 Example: Diffraction from a Rectangular Aperture

Here the opening is given by $|x'| < a$ and $|y'| < b$, so the integral above is

$$\int_{\text{opening}} d^2x' e^{i\mathbf{q}\cdot\mathbf{x}'} = \int_{-a}^a dx' e^{iq_x x'} \int_{-b}^b dy' e^{iq_y y'} = 2 \frac{\sin(q_x a)}{q_x} 2 \frac{\sin(q_y b)}{q_y} \quad (198)$$

The amplitude at \mathbf{x} is

$$\psi(\mathbf{x}) = -\frac{2ikab}{\pi} \frac{e^{ik(r+r_0)}}{rr_0} \left(\frac{\sin(q_x a)}{aq_x} \right) \left(\frac{\sin(q_y b)}{bq_y} \right) \quad (199)$$

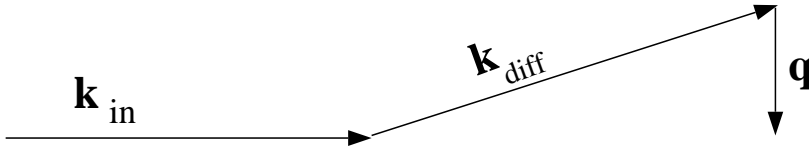
thus, the intensity of the diffracted wave is

$$\mathbf{I}(\mathbf{x}) = |\psi(\mathbf{x})|^2 = I_0 \left(\frac{\sin(q_x a)}{a q_x} \right)^2 \left(\frac{\sin(q_y b)}{b q_y} \right)^2 \quad (200)$$

with

$$I_0 = \frac{4k^2 a^2 b^2}{\pi^2 r_0^2 r^2} \quad (201)$$

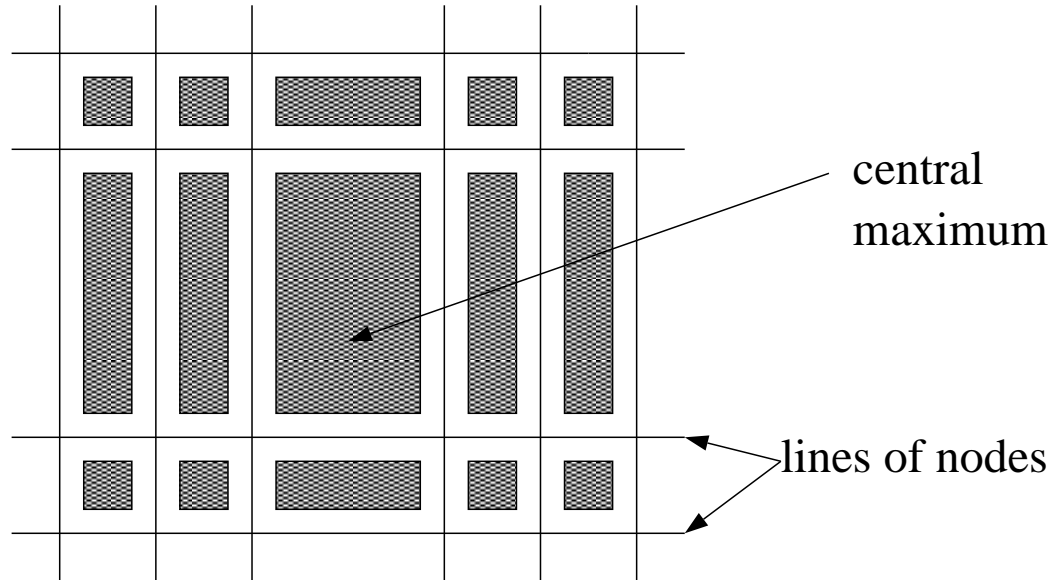
We see that all of the angular dependence of in this is embodied in the dependence on \mathbf{q} , the momentum transfer. Since we are looking at small-angle scattering, we are taking \mathbf{q} to lie essentially in the x-y plane.



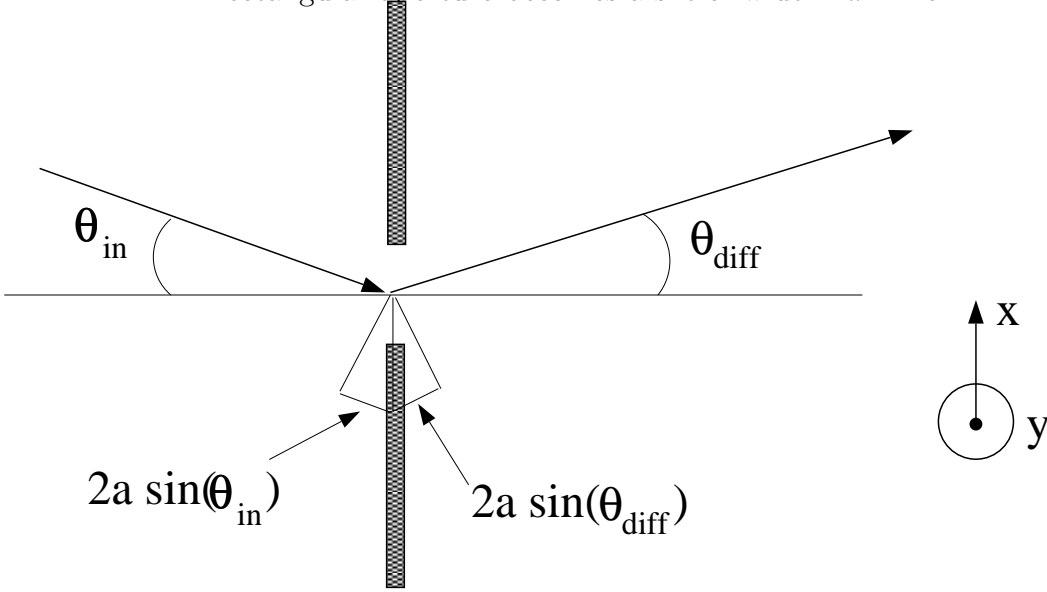
We see that the nodes of $\mathbf{I}(\mathbf{x})$ occur for

$$q_x = \frac{m\pi}{a}, \quad m \neq 0, \quad \text{and/or} \quad q_y = \frac{n\pi}{b}, \quad n \neq 0 \quad (202)$$

The global maximum ($I(\mathbf{x}) = I_0$) occurs for $q_x = q_y = 0$, i.e. for forward scattering. Thus the intensity of the diffracted light looks something like



We can recover the formula for single-slit diffraction by taking $b \gg a$, so that our rectangular aperture becomes a slit of width $2a$. Then



$$q_x = (k_x)_{in} - (k_x)_{diff} = k \sin(\theta_{in}) - k \sin(\theta_{diff}) \quad (203)$$

and the condition that there is a node is

$$q_x = k (\sin(\theta_{in}) - \sin(\theta_{diff})) = \frac{m\pi}{a} \quad m \neq 0. \quad (204)$$

Since $k = 2\pi/\lambda$, this may be written

$$2a (\sin(\theta_{in}) - \sin(\theta_{diff})) = m\lambda \quad m \neq 0 \quad (205)$$

which is the usual expression for a node in single-slit diffraction.

7.2 Example: Diffraction from a Circular Aperture

Here the opening is given by $r' < a$. Our Fraunhofer expression for the amplitude at \mathbf{x} is then

$$\psi(\mathbf{x}) = -\frac{ik}{2\pi} \frac{e^{ik(r+r_0)}}{rr_0} \int_0^a dr' r' \int_0^{2\pi} d\phi' e^{iqr' \cos(\phi')} \quad (206)$$

where we have taken \mathbf{q} to lie entirely in the plane of the aperture, as is appropriate in small-angle diffraction. Now

$$\int_0^{2\pi} d\phi' e^{iqr' \cos(\phi')} = 2\pi J_0(qr') \quad (207)$$

where J_0 is the zeroth-order Bessel function. Our expression for the diffraction amplitude is then

$$\psi(\mathbf{x}) = -\frac{ik}{2\pi} \frac{e^{ik(r+r_0)}}{rr_0} \int_0^a dr' r' 2\pi J_0(qr') = -ika^2 \frac{e^{ik(r+r_0)}}{rr_0} \left(\frac{J_1(qa)}{qa} \right) \quad (208)$$

where we have used $J_1(0) = 0$. The diffracted intensity is then

$$I(\mathbf{x}) = I_0 \left(\frac{J_1(qa)}{qa} \right)^2, \quad I_0 = \frac{k^2 a^4}{r^2 r_0^2} \quad (209)$$

this reaches a maximum at $q = 0$, it then has a minima of zero and maxima which decrease as q increases. For incident light with \mathbf{k} normal to the opening,

$$q = k_{diff} \sin(\theta_{diff}) = k \sin(\theta_{diff}) \quad (210)$$

since the magnitude of the wavevector does not change. Thus,

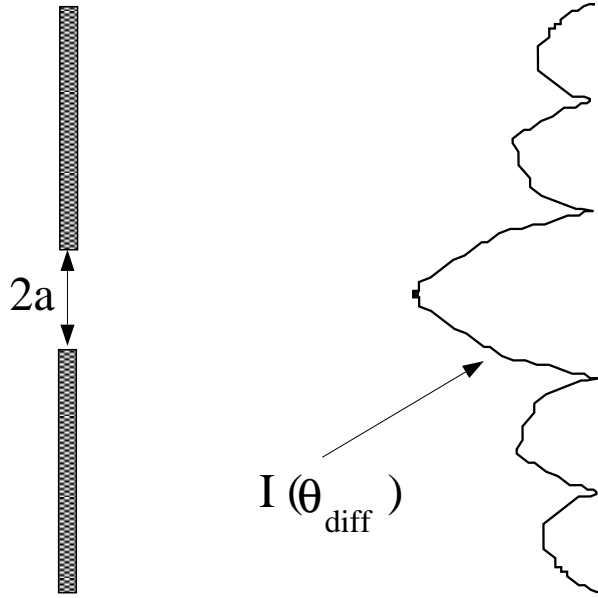
$$I(\theta_{diff}) = I_0 \left(\frac{J_1(ka \sin(\theta_{diff}))}{ka \sin(\theta_{diff})} \right)^2 \quad (211)$$

The first minima of $J_1(x)$ is at $x = 3.83$, hence the first minima of diffraction will occur when

$$ka \sin(\theta_{diff}) = 3.83, \quad \text{or} \quad \sin(\theta_{diff}) = 1.22 \left(\frac{\lambda}{2a} \right). \quad (212)$$

Since, by assumption, λ is small compared to a , $\sin(\theta_{diff}) \approx \theta_{diff}$, and so the angle of the first node is roughly

$$\theta_{diff} = 1.22 \left(\frac{\lambda}{2a} \right). \quad (213)$$



We notice from these examples of Fraunhofer diffraction that all of the light falling on the aperture is deflected. Granted, the intensity for forward scattering ($\mathbf{q} = 0$) is nonzero, but the total amount of light with $\mathbf{q} = 0$ is

$$|\psi(\mathbf{q})|^2 q^2 dq d\Omega \Big|_{\mathbf{q}=0} \rightarrow 0 \quad (214)$$

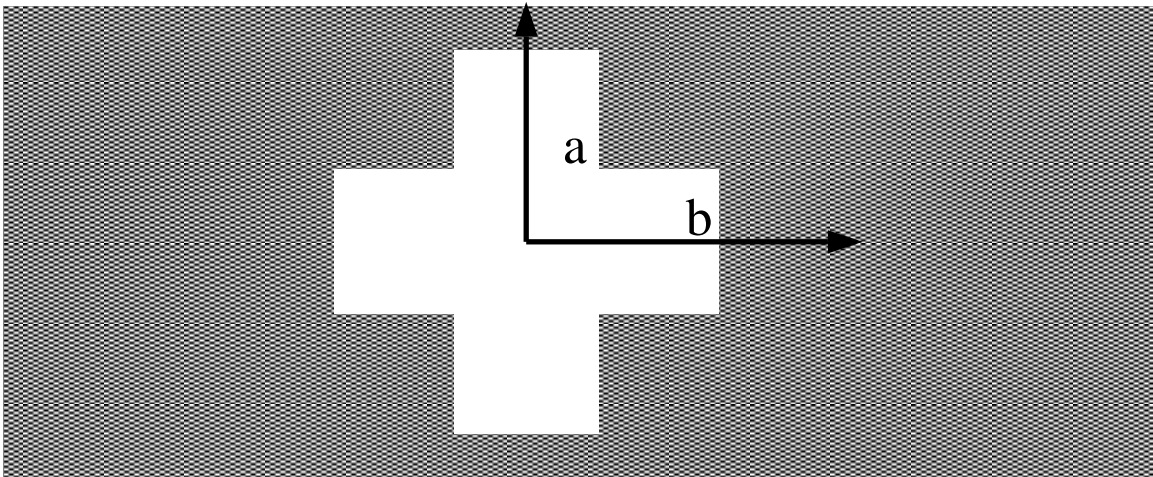
Thus, it is legitimate to say that all of the light is deflected. Let us apply this observation to a disc which is totally absorbing (i.e. a perfect black-body). If plane waves are incident on this disc, then all the light that falls directly upon it is absorbed. If there was no diffraction, then there would be a geometrically perfect shadow cast by the object, and the total cross section of the disc would be its area A . But of course there is diffraction, so in addition to absorbing light, some will also be deflected by the disc. By Babinet's principle, as much light will be diffracted by the disc as by an aperture of the same shape as the disc but in a screen. But we have said that in Fraunhofer diffraction all of the light falling on an aperture is deflected. Thus for our disc, the cross section for deflection of light will be the disc area A as well. Thus the total cross section, including both absorption (inelastic cross section) and diffraction (elastic cross section) is

$$\sigma_{total} = \sigma_{inelastic} + \sigma_{elastic} = A + A = 2A \quad (215)$$

This is twice the area of the disc. Recall that in this argument we assume that $ka \gg 1$, where a is the characteristic size of the disc. In this limit in quantum mechanics one finds that the cross section for scattering from a sphere of radius a is $2\pi a^2$. The factor of two has the same origins here as it does in quantum mechanics.

7.3 Diffraction from a Cross

1. Given a normally incident wave of frequency ω calculate the diffraction pattern produced by an opaque screen with an aperture in the shape symmetric cross of inner dimension a and outer dimension b in the Fraunhofer limit of the scalar Kirchoff approximation. Very roughly tell what the diffraction pattern will look like.
2. The correct answer to the first part of this problem also yields the solution to the diffraction pattern involving a quite different set of apertures. What is this other set of apertures?



Solution. In the Fraunhofer limit, we can use Eq. (197) of the notes.

$$\psi(\mathbf{x}) = -\frac{ik}{2\pi} \frac{e^{ik(r+r_0)}}{rr_0} \int_{\text{opening}} d^2x' e^{i\mathbf{q}\cdot\mathbf{x}'}$$

where r_0 is the distance from the aperture to the source, r is the distance from the aperture to the observer, and \mathbf{q} is the difference between the incident and scattered

wavevector. We will take \mathbf{q} to lie entirely in the plane of the aperture, as is appropriate in the small angle limit.

For the geometry in the figure, the integrals may be performed trivially

$$\psi(\mathbf{x}) = -\frac{ik}{2\pi} \frac{e^{ik(r+r_0)}}{rr_0} \left(\int_{-a}^a dx' \int_{-a}^a dy' + \int_{-a}^a dx' \left(\int_{-b}^{-a} dy' + \int_a^b dy' \right) + \int_{-a}^a dy' \left(\int_{-b}^{-a} dx' + \int_a^b dx' \right) \right) e^{iq_x x'}$$

so that

$$\psi(\mathbf{x}) = \frac{2ik}{\pi q_x q_y} \frac{e^{ik(r+r_0)}}{rr_0} [\sin(aq_x) \sin(bq_y) + \sin(aq_y) \sin(bq_x) - \sin(aq_x) \sin(aq_y)]$$

The corresponding intensity $|\psi(\mathbf{x})|^2$ will have a central maximum with $I \propto (2ab - a^2)^2$, and the whole pattern will be symmetric to rotations modulo $\pi/2$. By Babinet's principle we know that the diffraction from a cross-shaped screen has an amplitude, that when added to $\psi(\mathbf{x})$, is the same as if there was no diffraction at all.

7.4 Radiation from a Reciprocating Disk

A disc of radius a lies in the $z = 0$ plane and is centered at the origin. It is uniformly charged with surface charge density σ , and it rotates around the z -axis with an angular velocity $\Omega \cos(\omega t)$ where Ω and ω are constants. Assuming that the motion is nonrelativistic, find the fields in the radiation zone, the angular distribution of radiated power, and the total radiated power.

Solution. Here we will employ the methods developed to treat the radiation of harmonic current sources. The current of the rotation disc is (in cylindrical coordinates)

$$\mathbf{J}(\mathbf{x}, t) = \delta(z) \sigma \Omega \cos(\omega t) \rho \hat{\phi}$$

or, in complex notation

$$\mathbf{J}(\mathbf{x}, t) = \delta(z) \sigma \Omega e^{-i\omega t} \rho \hat{\phi} = \mathbf{J}(\mathbf{x}) e^{-i\omega t}$$

Then,

$$\mathbf{f}(\theta, \phi) = \frac{1}{c} \int_V d^3x' \mathbf{J}(\mathbf{x}') e^{-ik\mathbf{n}\cdot\mathbf{x}'} = \frac{\sigma \Omega}{c} \int_0^a \rho' d\rho' \int_0^{2\pi} d\phi' \rho' \hat{\phi}' e^{-ik\mathbf{n}\cdot\mathbf{x}'}$$

We can use the azimuthal symmetry of the problem, and assume that \mathbf{n} is in the xz plane. Then $\mathbf{n} = \cos \theta \hat{\mathbf{z}} + \sin \theta \hat{\mathbf{x}}$, $\mathbf{x}' = \rho' (\cos(\phi') \hat{\mathbf{x}} + \sin(\phi') \hat{\mathbf{y}})$, and $\hat{\phi}' = \cos(\phi') \hat{\mathbf{y}} - \sin(\phi') \hat{\mathbf{x}}$. For non-relativistic motion $\Omega a/c \ll 1$, this allows us to keep only the first nonvanishing term in the exponential, then

$$\mathbf{f}(\theta, \phi) = -i \frac{\pi k a^4 \sigma \Omega}{4c} \sin \theta \hat{\phi}'$$

Hence, in the radiation zone,

$$\mathbf{B} = ik \frac{e^{ikr}}{r} (\mathbf{n} \times \mathbf{f}) = -\frac{\pi k^2 a^4 \sigma \Omega}{4c} \sin \theta \frac{e^{ikr}}{r} \hat{\theta}$$

and

$$\mathbf{E} = \mathbf{B} \times \mathbf{n} = \frac{\pi k^2 a^4 \sigma \Omega}{4c} \sin \theta \frac{e^{ikr}}{r} \hat{\phi}'$$

The power distribution is given by

$$\frac{dP}{d\Omega} = \frac{ck^2}{8\pi} [|\mathbf{f}|^2 - |\mathbf{n} \cdot \mathbf{f}|^2] = \frac{ck^2}{8\pi} \frac{\pi^2 k^2 a^8 \sigma^2 \Omega^2}{16c^2} \sin^2(\theta)$$

The total power is obtained by integrating over all solid angles

$$P = \frac{\pi^2 k^4 \sigma^2 \Omega^2 a^8}{48c}$$

The Special Theory of Relativity

Albert Einstein

(1879 - 1955)

November 9, 2001

Contents

1	Einstein's Two Postulates	1
1.1	Galilean Invariance	2
1.2	The difficulty with Galilean Invariance	4
2	Simultaneity, Separation, Causality, and the Light Cone	6
2.1	Simultaneity	6
2.2	Separation and Causality	7
2.3	The Light Cone	8
2.4	The invariance of Separation	10
3	Proper time	11
3.1	Proper Time of an Oscillating Clock	13
4	Lorentz Transformations	14
4.1	Motivation	14
4.2	Derivation	15
4.3	Elapsed Proper Time Revisited	17
4.4	Proper Length and Length Contraction	18

5	Transformation of Velocities	20
5.1	Aberration of Starlight	22
6	Doppler Shift	23
6.1	Stellar Red Shift	26
7	Four-tensors and all that	27
7.1	The Metric Tensor	30
7.2	Differential Operators	33
7.3	Notation	35
8	Representation of the Lorentz transformation	35
9	Covariance of Electrodynamics	40
9.1	Transformations of Source and Fields	40
9.1.1	ρ and \mathbf{J}	40
9.1.2	Potentials	42
9.1.3	Fields, Field-Strength Tensor	43
9.2	Invariance of Maxwell Equations	45
10	Transformation of the electromagnetic field	46
10.1	Fields Due to a Point Charge	48

In this chapter we depart temporarily from the study of electromagnetism to explore Einstein’s special theory of relativity. One reason for doing so is that Maxwell’s field equations are inconsistent with the tenets of “classical” or “Galilean” relativity. After developing the special theory, we will apply it to both particle kinematics and electromagnetism and will find that Maxwell’s equations are completely consistent with the requirements of the special theory.

1 Einstein's Two Postulates

Physical phenomena may be observed and/or described relative to any of an infinite number of “reference frames;” we regard the reference frame as being that one relative to which the measuring apparatus is at rest. The basic claim (or postulate) of relativity, which predates Einstein’s work by many centuries, is that physical phenomena should be unaffected by the choice of the frame from which they are observed. This statement is quite vague. A simple explicit example is a collision of two objects. If they are seen to collide when observed from one frame, then the postulate of relativity says that they will be seen to collide no matter what reference frame is used to make the observation.

1.1 Galilean Invariance

Given that one believes some version of the postulate of relativity, then that person should, when constructing an explanation of the phenomena in question, make a theory which will predict the same phenomena in all reference frames. The original great achievements of this kind were Newton’s theories of mechanics and gravitation. Consider, for example, $\mathbf{F} = m\mathbf{a}$. If the motion of some massive object is observed relative to two different reference frames, the motion will obey this equation in both frames provided the frames themselves are not being accelerated. This qualification leads one to restrict the statement of the relativity principle to unaccelerated or *inertial* reference frames.

In order to test the postulate of relativity, one needs a transformation that makes it possible to translate the values of physical observables from one frame to another. Consider two frames K and K' with K' moving at velocity \mathbf{v} relative to K .

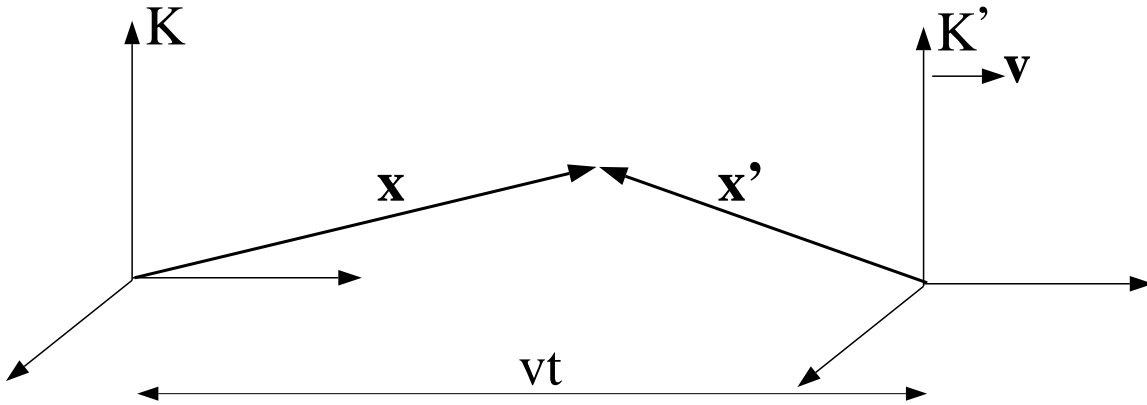


Figure 1: Inertial frames K and K'

Then the (almost obvious) way to relate a space-time point (t, \mathbf{x}) in K to the same point (t', \mathbf{x}') in K' is via the *Galilean transformation*

$$\mathbf{x}' = \mathbf{x} - \mathbf{v}t \quad \text{and} \quad t' = t, \quad (1)$$

or so it was believed up to the time of Einstein. Notice that the transformation is written so that the (space) origins coincide at $t = t' = 0$; we shall say simply that the origins (in space and time) coincide.

In what sense is Newton's law of motion consistent with the Galilean transformation? If his equation satisfies the postulate of relativity, then the motion of a massive object must obey it in both frames; thus

$$\mathbf{F} = m\mathbf{a} \quad \text{and} \quad \mathbf{F}' = m'\mathbf{a}' \quad (2)$$

where primed quantities are measured in K' and unprimed ones in K . Now, experiments demonstrate (not quite correctly) that the force and mass are invariants, meaning that they are the same in all inertial frames, so if Newton's law is to hold in all inertial frames, then it must be the case that $\mathbf{a} = \mathbf{a}'$. The Galilean transformation provides a way of comparing these two quantities. In Eqs. (1), let \mathbf{x} and \mathbf{x}' be the positions of the mass at, respectively, times t and t' in frames K and K' . Then we

have

$$\frac{d\mathbf{x}'}{dt'} = \frac{d\mathbf{x}}{dt} - \mathbf{v} \quad (3)$$

and

$$\mathbf{a}' = \frac{d^2\mathbf{x}'}{dt'^2} = \frac{d^2\mathbf{x}}{dt^2} = \mathbf{a}, \quad (4)$$

assuming \mathbf{v} is a constant. Thus we find that Newton's law, Galileo's transformation, and the observed motions of massive objects are consistent.¹

1.2 The difficulty with Galilean Invariance

Now we come to the dilemma posed by Maxwell's equations. They are not consistent with the postulate of relativity if one uses the Galilean transformation to relate quantities in two different inertial frames. Imagine the quandary of the late-nineteenth-century physicist. He had the Galilean transformation and Newton's equations of motion, backed by enormous experimental evidence, to support the almost self-evident principle of relativity. But he also had the new - and enormously successful - Maxwell theory of light which was not consistent with Galilean relativity. What to do? One possible way out of the morass was easy to find. It was well-known that wavelike phenomena, such as sound, obey wave equations which are not properly "invariant" under Galilean transformations. The reason is simple: These waves are vibrational motions of some medium such as air or water, and this medium will be in motion with different velocities relative to the coordinate axes of different inertial frames. If one understood this, then one could see that although the wave equation takes on different forms relative to different frames, it did correctly describe what goes on in every frame and was not inconsistent with the postulate of relativity.

The appreciation of this fact set off a great search to find the medium, called the "luminiferous ether" or simply the *ether*, whose vibrations constitute electromag-

¹Of course, they aren't consistent at all if one either makes measurements of extraordinary precision or studies particles traveling at an appreciable fraction of the speed of light. Neither of these things was done prior to the twentieth century.

netic waves. The search (i.e. Michelson and Morely) was, as we know, completely unsuccessful², as the ether eluded all seekers.

However, for Einstein, it was the Fizeau experiment (1851) which convinced him that the ether explanation was incorrect. This experiment looked for a change in the phase velocity of light due to its passage through a moving medium, in this case water.

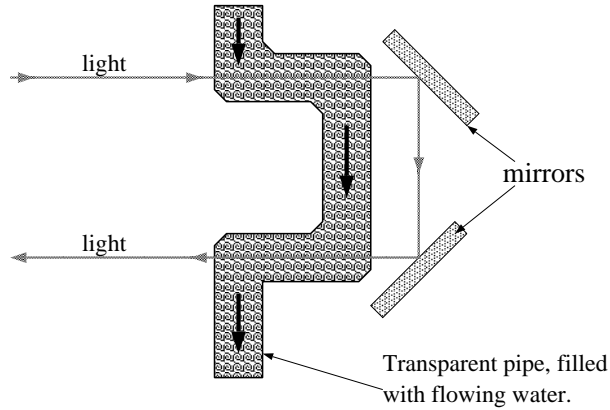


Figure 2: Diagram of Fizeau experiment

Fitzeau found that this phase velocity was given by

$$v_{phase} = \frac{c}{n} \pm v \left(1 - \frac{1}{n^2}\right) \text{ experiment}$$

where n is the index of refraction of the water, and v is its velocity. The plus(minus) sign is taken if the water is moving with(against) the light.

Lets analyze the experiment from a Galilean point of view. The dielectric water is moving in either the same or opposite direction as the light, and so acts as a moving source for the light with is refracted (i.e. reradiated by the water molecules). Nonrelativistically, we just add the velocity v of the source to the wave velocity for the stationary source. Thus Galilean theory says

$$v_{phase} = \frac{c}{n} \pm v \text{ Galilean theory,}$$

²Or completely successful, if we adopt a somewhat different (Einstein's) point of view.

which is clearly inconsistent with experiment.

The stage was now set for Einstein who, in 1905, made the following postulates:

1. Postulate of relativity: The laws of nature and the results of all experiments performed in a given frame of reference are independent of the translational motion of the system as a whole.
2. Postulate of the constancy of the speed of light: The speed of light is independent of the motion of its source.

The first postulate essentially reaffirmed what had long been thought or believed in the specific case of Newton's law, extending it to all phenomena. The second postulate was much more radical. It did away with the ether at a stroke and also with Galilean relativity because it implies that the speed of light is the same in all reference frames which is fundamentally inconsistent with the Galilean transformation.

2 Simultaneity, Separation, Causality, and the Light Cone

2.1 Simultaneity

The second postulate - disturbing in itself - leads to many additional "nonintuitive" predictions. For example, suppose that there are sources of light at points A and C and that they both emit signals that are observed by someone at B which is midway between A and C.

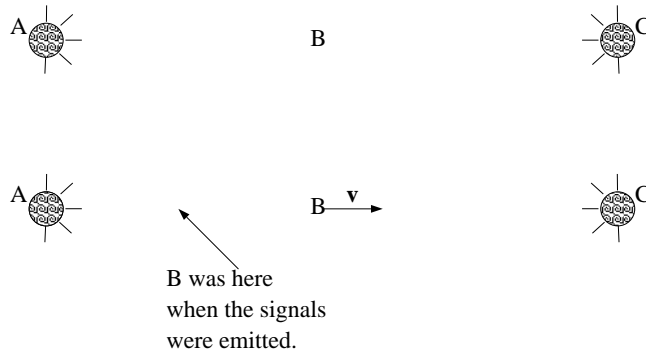


Figure 3: Simultaneity depends upon the rest frame of the observer

If he sees the two signals simultaneously and knows that he is equidistant from the sources, he will conclude quite correctly that the signals were emitted simultaneously. Now suppose that there is a second observer, B', who is moving along the line from A to C and who arrives at B just when the signals do. He will know that the signals were emitted at some earlier time when he was closer to A than to C. Also, since both signals travel with the same speed c in his rest frame (because the speeds of the signals relative to him are independent of the speeds of the sources relative to him), he will conclude that the signal from C was emitted earlier than that from A because it had to travel the greater distance before reaching him. He is as correct as the first observer. Similarly, an observer moving in the opposite direction relative to the first one will conclude from the same reasoning that the signal from A was emitted before that from C. *Hence Einstein's second postulate leads us to the conclusion that events, in this case the emission of light signals, which are simultaneous in one inertial frame are not necessarily simultaneous in other inertial frames.*

2.2 Separation and Causality

If simultaneity is only a relative fact, as opposed to an absolute one, what about causality? Because the order of the members of some pairs of events can be reversed by changing one's reference frame, we must consider whether the events' ability to

influence each other can similarly be affected by a change of reference frame. This question is closely related to a quantity that we shall call the *separation* between the events. Given two events A and B which occur at space-time points (t_1, \mathbf{x}_1) and (t_2, \mathbf{x}_2) , we define the squared separation s_{12}^2 between them to be

$$s_{12}^2 \equiv c^2(t_1 - t_2)^2 - |\mathbf{x}_1 - \mathbf{x}_2|^2. \quad (5)$$

Let the two events be (1) the emission of an electromagnetic signal at some point in vacuum and (2) its reception somewhere else. Then, because the signal travels with the speed c , these events have separation zero, $s_{12}^2 = 0$. This result will be the same in any inertial frame since the signal has the same speed c in all such frames.

Now, if we have two events such that $s_{12}^2 > 0$, then we have a “causal relationship” in the sense that a light signal can get from the first event to the place where the second one occurs before it does occur. Such a separation is called *timelike*. On the other hand, if $s_{12}^2 < 0$, then a light signal cannot get from the first event to the location of the second event before the second event occurs. This separation is called *spacelike*. A separation $s_{12}^2 = 0$ is called *lightlike*.

It is important to ask whether there is some other type of signal that travels faster than c and which could therefore produce a causal relationship between events with a spacelike separation. None has been found and we shall assume that none exists. Consequently, we claim that events with a timelike separation are such that the earlier one can influence the later one, because a signal can get from the first to the location of the second before the latter occurs, but that events with a spacelike separation are such that the earlier one cannot influence the later one because a signal cannot get from the first event to the location of the second one fast enough.

2.3 The Light Cone

The question now is whether the character of the separation between two events, timelike, spacelike, or lightlike, can be changed by changing the frame in which it is

measured. For simplicity, let the “first” event, A, occur (in frame K) at $(t = 0, \mathbf{x} = 0)$ while the second takes place at some general (t, \mathbf{x}) with $t > 0$. Further, let ct be larger than $|\mathbf{x}|$ so that $s^2 > 0$ and A may influence B. Consider these same two events in another frame K' . By an appropriate choice of the origin (in space and time) of this frame, we can make the first event occur here, just as it does in frame K . The second event will be at some (t', \mathbf{x}') .

We can picture the relative positions of the two events in space and time by using a *light cone* as shown.

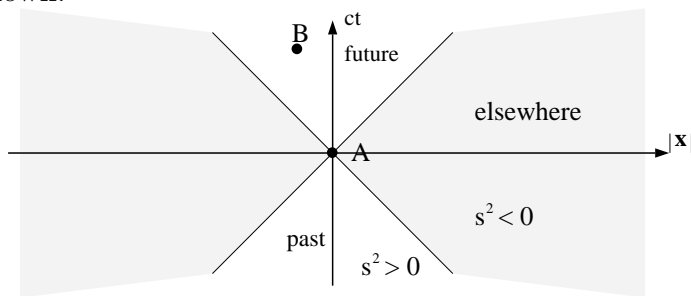


Figure 4: The Light Cone

The vertical axis measures ct ; the horizontal one, separation in space, $|\mathbf{x}|$. The two diagonal lines have slopes ± 1 . The event B is shown within the cone whose axis is the ct axis; any event with a timelike separation relative to the origin will be in here.

The question we wish to ask now is whether, by going to another reference frame, one may cause event B to move across one of the diagonal lines and so wind up in a place where it cannot be influenced by the event at the origin? The point is that A can influence any event inside of the “future” cone; it can be influenced by any event inside of the “past” cone; but it cannot influence, or be influenced by, any event inside of the “elsewhere” region. If an event B and two reference frames K and K' can be found such that the event when expressed in one frame is on the opposite side of a diagonal from where it is in the other frame, then we have made causality a frame-dependent concept.

Suppose that we have two such frames. We can effect a transformation from one to the other by considering a sequence of many frames, each moving at a velocity only

slightly different from the previous one, and such that the first frame in the sequence is K and the final one is K' .

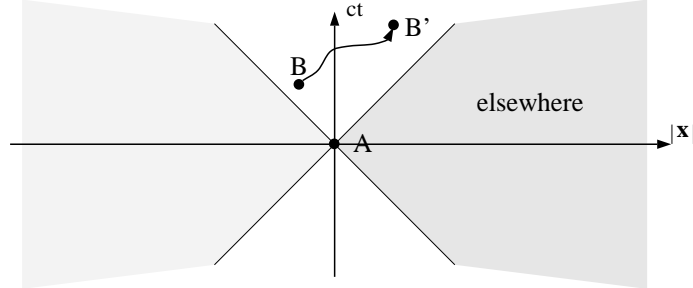


Figure 5: Evolution of the event B as we evolve from frame K to K'

If we let this become an infinite sequence with infinitesimal differences in the velocities of two successive members of the sequence, then the positions of the event B in the sequence of light cones for each of the frames will form a continuous curve when expressed in a single light cone. If B crosses from timelike to spacelike in this sequence, then at some point, B must lie on one of the diagonals. For such an event, the separation from A is lightlike, or zero.

Consider now two events with zero separation $s_{12}^2 = 0$. These events can be coincident with the emission and reception of a light signal. But these events must be coincident with the emission and reception of the light signal in all frames, by the postulate of relativity (the first postulate), and so these events must have zero separation in all inertial frames because of the constancy of the speed of light in all frames. Consequently, what we are trying to do above is in fact impossible; that is, one cannot move an event such as B onto or off of the surface of the light cone by looking at it in a different reference frame, and for this reason, one cannot make it cross the surface of the light cone. An event in the future will be there in all reference frames. One in the past cannot be taken out of the past; and one that is “elsewhere” will be there in any reference frame.

2.4 The invariance of Separation

With a little more thought we can generalize the conclusion of the previous paragraph that two events with zero separation in one frame have zero separation in all frames. In fact, the separation, whatever it may be, between any two events is the same in all frames. We shall call something that is the same in all frames an *invariant*; the separation is an invariant. To argue that this should be the case, suppose that we have two events which are infinitesimally far apart in both space and time so that we may write ds_{12} for s_{12} ,

$$(ds_{12})^2 = c^2(t_1 - t_2)^2 - |\mathbf{x}_1 - \mathbf{x}_2|^2 \quad (6)$$

in frame K . In another inertial frame K' we have separation $(ds'_{12})^2$, and we have argued that this is zero if ds_{12}^2 is zero. If K' is moving with a small speed relative to K , the separations in the two frames must be nearly equal which means that they will be infinitesimal quantities of the same order, or

$$(ds_{12})^2 = A(ds'_{12})^2 \quad (7)$$

where $A = A(v)$ is a finite function of v , the relative speed of the frames. Furthermore, $A(0) = 1$ since the two frames are the same if $v = 0$. Now, if time and space are homogeneous and isotropic, then it must also be true that

$$(ds'_{12})^2 = A(v)(ds_{12})^2. \quad (8)$$

Comparing the preceding equations, we see that the only solutions are $A(v) = \pm 1$; the condition that $A(0) = 1$ means $A(v) = 1$. Hence

$$(ds'_{12})^2 = (ds_{12})^2 \quad (9)$$

which is a relation between differentials that may be integrated to give

$$(s'_{12})^2 = (s_{12})^2, \quad (10)$$

thereby demonstrating that the separation between any two events is an invariant. The locus of all points with a given separation is a hyperbola when drawn on a light cone (or a hyperboloid of revolution if more spatial dimensions are displayed in the light cone).

3 Proper time

Another related invariant is the so-called *proper time*. This is tied to a particular object and is the time that elapses in the rest frame of that object. If the object is accelerated, its rest frame is not an inertial frame.

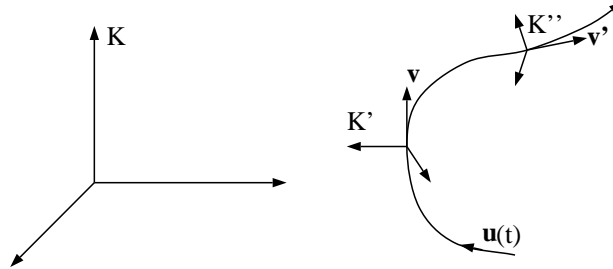


Figure 6: Instantaneous rest frames K' and K'' of an object with velocity $\mathbf{u}(t)$ as measured in K

It is then useful to make use of the “instantaneous” rest frame of the object, meaning an inertial frame relative to which the object is not moving at a particular instant of time. Thus, if in frame K the object has a velocity $\mathbf{u}(t)$, its instantaneous rest frame at time t is a frame K' which moves at velocity $\mathbf{v} = \mathbf{u}(t)$. One may find the object’s proper time by calculating the time that elapses in an infinite sequence of instantaneous rest frames.

Consider an object moving with a trajectory $\mathbf{x}(t)$ relative to frame K . Between t and $t + dt$ it moves a distance $d\mathbf{x}$ as measured in K . Let us ask what time dt' elapses in the frame K' which is the instantaneous rest frame at time t . The one thing we

know is that

$$(ds)^2 \equiv c^2(dt)^2 - (d\mathbf{x})^2 = (ds')^2 = c^2(dt')^2 - (d\mathbf{x}')^2 \quad (11)$$

where, as usual, unprimed quantities are the ones measured relative to K and primed ones are measured in K' . Now, $d\mathbf{x}' = 0$ ³ because the object is at rest in K' at time t . Hence we may drop this contribution to the (infinitesimal) separation and solve for dt' :

$$dt' = \sqrt{(dt)^2 - (d\mathbf{x})^2/c^2} = dt \sqrt{1 - \frac{1}{c^2} \left(\frac{d\mathbf{x}}{dt} \right)^2} = dt \sqrt{1 - u^2/c^2} \quad (12)$$

where $\mathbf{u} \equiv d\mathbf{x}/dt$ is the object's velocity as measured in K . Now we may integrate from some initial time t_1 to a final time t_2 to find the proper time of the object which elapses while time is proceeding from t_1 to t_2 in frame K ; that is, we are adding up all of the time that elapses in an infinite sequence of instantaneous rest frames of the object while time is developing in K from t_1 to t_2 .

$$\tau_2 - \tau_1 = \int_{t_1}^{t_2} dt \sqrt{1 - u^2(t)/c^2}. \quad (13)$$

3.1 Proper Time of an Oscillating Clock

As an example let the object move along a one-dimensional path with $u(t) = c \sin(2\pi t/t_0)$ with $t_1 = 0$ and $t_2 = t_0$. This velocity describes a round trip of a harmonic oscillator with a peak speed of c and a period of t_0 . The corresponding elapsed proper time is

$$\tau_0 = \int_0^{t_0} dt \sqrt{1 - \sin^2(2\pi t/t_0)} = \int_0^{t_0} dt |\cos(2\pi t/t_0)| = 2t_0/\pi. \quad (14)$$

This is smaller than t_0 by a factor of $2/\pi$ which means that a clock carried by the object will show an elapsed time during the trip which is just $2/\pi$ times what a clock which remains in frame K will show, provided the acceleration experienced by the clock which makes the trip doesn't alter the rate at which it runs.

³More correctly $d\mathbf{x}'$ is a second order differential, and hence may be neglected

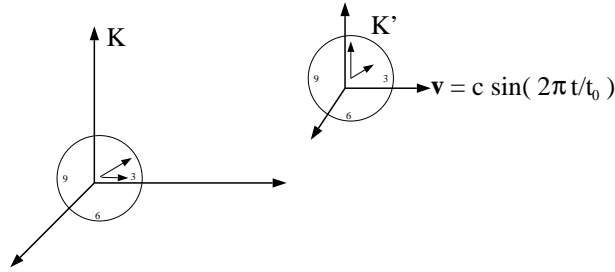


Figure 7: The proper time for the oscillating frame is $2t/\pi$; which is less than the elapsed time in frame K .

If this clock is a traveller, then the traveller ages during the trip by an amount which is only $2/\pi$ of the amount by which someone who stays at rest in K ages. One may wonder whether, from the point of view of the traveller, the one who stayed at home should be the one who ages more “slowly.” If the calculation is done carefully (correctly), one finds that the same conclusion is reached; the traveller has in fact aged less than the stay-at-home.

4 Lorentz Transformations

4.1 Motivation

So far we know the locations (t, \mathbf{x}) and (t', \mathbf{x}') of a space-time point as given in K and K' must be related by

$$c^2t^2 - \mathbf{x} \cdot \mathbf{x} = c^2t'^2 - \mathbf{x}' \cdot \mathbf{x}', \quad (15)$$

given that the origins of the coordinate and time axes of the two frames coincide. This equation looks a lot like the statement that the inner product of a four-dimensional vector, having components ct and $i\mathbf{x}$, with itself is an invariant. It suggests that the transformation relating (t, \mathbf{x}) and (t', \mathbf{x}') is an orthogonal transformation in the four-dimensional space of ct and \mathbf{x} . There is an unusual feature in that the transformation apparently describes an imaginary or complex rotation because the inner product,

or length, that is preserved is $c^2t^2 - \mathbf{x} \cdot \mathbf{x}$ as opposed to $c^2t^2 + \mathbf{x} \cdot \mathbf{x}$. Recall that a rotation in three dimensions around the \mathbf{e}_3 direction by angle ϕ can be represented by a matrix

$$a = \begin{pmatrix} \cos \phi & -\sin \phi & 0 \\ \sin \phi & \cos \phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (16)$$

so that

$$x'_i = \sum_{j=1}^3 a_{ij}x_j; \quad (17)$$

that is,

$$\begin{aligned} x'_1 &= \cos \phi x_1 - \sin \phi x_2 \\ x'_2 &= \sin \phi x_1 + \cos \phi x_2 \\ x'_3 &= x_3. \end{aligned} \quad (18)$$

For an imaginary ϕ , $\phi = i\eta$, $\cos \phi \rightarrow \cosh \eta$, and $\sin \phi \rightarrow -i \sinh \eta$. Further, let us reconstruct the vector as $\mathbf{y} = (x_1, ix_2, ix_3)$ and make the transformation

$$\mathbf{y}'_i = \sum_j a_{ij}y_j. \quad (19)$$

The result, expressed in terms of components of \mathbf{x} , is

$$\begin{aligned} x'_1 &= \cosh \eta x_1 - \sinh \eta x_2 \\ x'_2 &= -\sinh \eta x_1 + \cosh \eta x_2 \\ x'_3 &= x_3; \end{aligned} \quad (20)$$

these are such that

$$x'^2_1 - x'^2_2 - x'^2_3 = x^2_1 - x^2_2 - x^2_3 \quad (21)$$

since $\cosh^2(\eta) - \sinh^2(\eta) = 1$, so we have succeeded in constructing a transformation that produces the right sort of invariant. All we have to do is generalize to four dimensions.

4.2 Derivation

Let's begin by introducing a vector with four components, (x_0, x_1, x_2, x_3) where $x_0 = ct$ and the x_i with $i = 1, 2, 3$ are the usual Cartesian components of the position vector. Then introduce $\vec{y} \equiv (x_0, i\mathbf{x})$ which has the property that

$$\vec{y} \cdot \vec{y} = x_0^2 - \mathbf{x} \cdot \mathbf{x}. \quad (22)$$

This inner product is supposed to be an invariant under the transformation of \mathbf{x} and t that we seek. The transformation in question is a rotation through an imaginary angle $i\eta$ that mixes time and one spatial direction, which we pick to be the first (y_1 or x_1) without loss of generality. The matrix representing this rotation is

$$a = \begin{pmatrix} \cosh \eta & i \sinh \eta & 0 & 0 \\ -i \sinh \eta & \cosh \eta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \quad (23)$$

Now operate with this matrix on \vec{y} to produce \vec{y}' . If we write the components of the latter as $(x'_0, i\mathbf{x}')$, we find the following:

$$\begin{aligned} x'_0 &= \cosh \eta x_0 - \sinh \eta x_1 \\ x'_1 &= -\sinh \eta x_0 + \cosh \eta x_1 \\ x'_2 &= x_2 \\ x'_3 &= x_3. \end{aligned} \quad (24)$$

It is a simple matter to show from these results that $x_0^2 - \mathbf{x} \cdot \mathbf{x}$ is an invariant, i.e.,

$$x_0^2 - \mathbf{x} \cdot \mathbf{x} = x_0'^2 - \mathbf{x}' \cdot \mathbf{x}' \quad (25)$$

which means we have devised an acceptable transformation in the sense that it preserves the separation between two events.

But what is the significance of η ? Let us rewrite $\sinh \eta$ as $\cosh \eta \tanh \eta$. Then we have, in particular,

$$\begin{aligned} x'_0 &= \cosh \eta (x_0 - \tanh \eta x_1) \\ x'_1 &= \cosh \eta (x_1 - \tanh \eta x_0). \end{aligned} \quad (26)$$

The second of these is

$$x'_1 = \cosh \eta (x_1 - \tanh \eta ct). \quad (27)$$

Suppose that we are looking at an object at rest at the origin of K' , and the space-time point (t, \mathbf{x}) is this object's location. Then $x'_1 = 0$ for all t' . As seen from K , the object is at $x_1 - vt$ given that \mathbf{v} , the velocity of the object (and of K') relative to K , is parallel to $\boldsymbol{\epsilon}_1$.

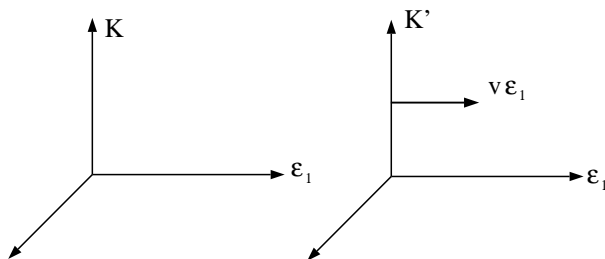


Figure 8: Coordinates for our Lorentz transform.

This is consistent with Eq. (27) provided

$$\tanh \eta \equiv \frac{v}{c} \equiv \beta \quad (28)$$

where β is defined as the speed v relative to the speed of light. From this relation, we find further that

$$\cosh \eta = 1/\sqrt{1 - \beta^2} \equiv \gamma; \quad (29)$$

this expression defines γ , a parameter that comes up repeatedly in the special theory of relativity.

Our determination of the transformation, called the *Lorentz transformation*⁴, is now complete. We find that, given a frame K' moving at velocity $\mathbf{v} = v\boldsymbol{\epsilon}_1$ relative to

⁴H. A. Lorentz devised these transformations prior to Einstein's development of the special theory of relativity; they had in fact been used even earlier by Larmor and perhaps others. Furthermore, it was known that Maxwell's equations were invariant under these transformations, meaning that if these are the right transformations (as opposed to the Galilean transformations), Maxwell's equations are eligible for "law of nature" status.

K , a space-time point (t, \mathbf{x}) in K becomes, in K' , the space-time point (t', \mathbf{x}') with

$$x'_0 = \gamma(x_0 - \beta x_1) \quad x'_1 = \gamma(x_1 - \beta x_0) \quad x'_2 = x_2 \quad x'_3 = x_3. \quad (30)$$

The inverse transformation can be extracted from these equations in a straightforward manner; it may also be inferred from the fact that K is moving at velocity $-\mathbf{v}$ relative to K' which tells us immediately that

$$x_0 = \gamma(x'_0 + \beta x'_1) \quad x_1 = \gamma(x'_1 + \beta x'_0) \quad x_2 = x'_2 \quad x_3 = x'_3. \quad (31)$$

4.3 Elapsed Proper Time Revisited

Let us try to use this transformation to calculate something. First, we revisit the proper time. For an object at rest in K' , \mathbf{x}' does not change with time. Also, from our transformation,

$$ct = \gamma(ct' + \beta x'_1), \quad (32)$$

The differential of this transformation, making use of the fact that the object is instantaneously at rest in K' , gives, $dt = \gamma dt'$ since dx'_1 is second-order in powers of dt' . Stated in another fashion, we are considering the transformation of two events or space-time points. They are the locations of the object at times t' and $t' + dt'$. Because the object is at rest in K' at time t' , its displacement $d\mathbf{x}'$ during the time increment dt' is of order $(dt')^2$ and so may be discarded. The corresponding elapsed time dt in K is thus found to be $dt = \gamma dt'$, using the Lorentz transformations of the two space-time points. This equation may also be written as

$$dt' = dt/\gamma = \sqrt{1 - v^2/c^2} dt. \quad (33)$$

The left-hand side of this equation is the elapsed proper time of the object while dt is the elapsed time measured by observers at rest relative to K . If we introduce \mathbf{u} , the velocity of the object relative to K , and notice that $\mathbf{u} = \mathbf{v}$ at time t , then we can write dt' in terms of $\mathbf{u}(t)$ as

$$dt' = \sqrt{1 - |\mathbf{u}(t)|^2/c^2} dt, \quad (34)$$

where now dt' is the elapsed proper time of the object which moves at velocity \mathbf{u} relative to frame K . We can integrate this relation to find the finite elapsed proper time during an arbitrary time interval (in K),

$$\tau_2 - \tau_1 = \int_{t_1}^{t_2} dt \sqrt{1 - |\mathbf{u}(t)|^2/c^2}. \quad (35)$$

4.4 Proper Length and Length Contraction

Next, we shall examine the *Fitzgerald-Lorentz contraction*. Define the *proper length* of an object as its length, measured in the frame where it is at rest. Let this be L_0 , and let the rest frame be K' , moving at the usual velocity ($\mathbf{v} = v\boldsymbol{\epsilon}_1$) relative to K .

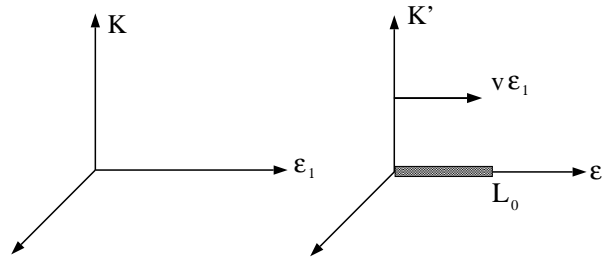


Figure 9: Length contraction occurs along the axis parallel to the velocity.

The relative geometry is shown in the figure. The rod, or object, is positioned in K' so that its ends are at $x'_1 = 0, L_0$. They are there for all t' . In order to find the length of the rod in K , we have to measure the positions of both ends at the **same** time t as measured in K . We can find the results of these measurements from the Lorentz transformation

$$x'_1 = \gamma(x_1 - \beta x_0). \quad (36)$$

Use this relation first with x'_1 equal to 0 and then with $x'_1 = L_0$, using the same time x_0 in both cases, and take the difference of the two equations so obtained. The result is

$$L_0 = \gamma(x_{1R} - x_{1L}) \equiv \gamma L \quad (37)$$

where x_{1R} and x_{1L} are the positions of the right and left ends of the rod at some

particular time, or x_0 . The difference of these is L , the length of the rod as measured in frame K .

Our result for L can be written as

$$L = L_0/\gamma = \sqrt{1 - \beta^2}L_0. \quad (38)$$

This length is smaller than L_0 which means that the object is found (is measured) in K to be shorter than its proper length or its length in the frame where it is at rest. Notice, however, that if we did the same calculation for its length in a direction perpendicular to the direction of \mathbf{v} , we would find that it is the same in K as in K' . Consequently the transformation of the object's volume is

$$V = V_0/\gamma = \sqrt{1 - \beta^2}V_0 \quad (39)$$

where V_0 is the *proper volume* or volume in the rest frame, and V is the volume in a frame moving at speed βc relative to the rest frame.

5 Transformation of Velocities

Because we know how \mathbf{x} and t transform, we can determine how anything that involves functions of these things transforms. For example, velocity. Let an object have velocity \mathbf{u} in K and velocity \mathbf{u}' in K' and let K' move at velocity \mathbf{v} relative to K . We wish to determine how \mathbf{u}' is related to \mathbf{u} . In K' , the object moves a distance $d\mathbf{x}' = \mathbf{u}'dt'$ in time dt' . A similar statement, without any primed quantities, holds in K . The infinitesimal displacements in time and space are related by Lorentz transformations:

$$dt = \gamma(v) \left(dt' + (v/c^2)dx' \right) \quad dx = \gamma(v)(dx' + vdt') \quad dy = dy' \quad dz = dz', \quad (40)$$

where we have let \mathbf{v} be along the direction of x_1 as usual. Taking ratios of the displacements to the time increment, we have

$$u_x = \frac{dx}{dt} = \frac{dx' + vdt'}{dt' + (v/c^2)dx'} = \frac{dx'/dt' + v}{1 + (v/c^2)(dx'/t')} = \frac{u'_x + v}{1 + vu'_x/c^2}, \quad (41)$$

$$u_y = \frac{1}{\gamma(v)} \frac{u'_y}{1 + vu'_x/c^2}, \quad (42)$$

and

$$u_z = \frac{1}{\gamma(v)} \frac{u'_z}{1 + vu'_x/c^2}. \quad (43)$$

These results may be summarized in vectorial form:

$$\mathbf{u}_{\parallel} = \frac{u'_{\parallel} + \mathbf{v}}{1 + \mathbf{v} \cdot \mathbf{u}'/c^2} \quad u_{\perp} = \frac{u'_{\perp}}{\gamma(v)(1 + \mathbf{v} \cdot \mathbf{u}'/c^2)} \quad (44)$$

where the subscripts “ \parallel ” and “ \perp ” refer respectively to the components of the velocities \mathbf{u} and \mathbf{u}' parallel and perpendicular to \mathbf{v} . Notice too that

$$\mathbf{u}_{\parallel} = \left(\frac{\mathbf{u} \cdot \mathbf{v}}{v^2} \right) \mathbf{v} \quad \text{and} \quad \mathbf{u}_{\perp} = \mathbf{u} - \mathbf{u}_{\parallel} = \mathbf{u} - \left(\frac{\mathbf{u} \cdot \mathbf{v}}{v^2} \right) \mathbf{v}. \quad (45)$$

It is sometimes useful to express the transformations for velocity in polar coordinates (u, θ) and (u', θ') such that

$$u_{\parallel} = u \cos \theta \quad \text{and} \quad u_{\perp} = u \sin \theta, \quad (46)$$

etc.; the appropriate expressions are

$$\tan \theta = \frac{u' \sin \theta'}{\gamma(v)(u' \cos \theta' + v)} \quad \text{and} \quad u = \frac{[u'^2 + v^2 + 2u'v \cos \theta' - (vu' \sin \theta'/c)^2]^{1/2}}{1 + u'v \cos \theta'/c^2}. \quad (47)$$

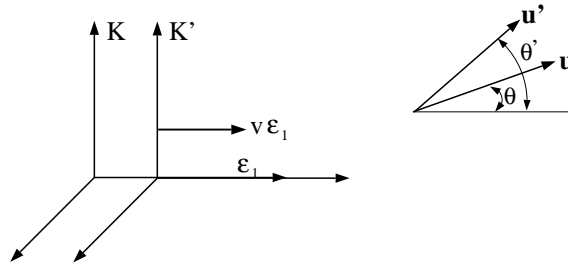


Figure 10: Frames for velocity transform.

The inverses of all of these velocity transformations are easily found by appropriate symmetry arguments based on the fact that the velocity of K relative to K' is just $-\mathbf{v}$.

It is interesting that the velocity transformations are, in contrast to the ones for \mathbf{x} and t , nonlinear. They must be nonlinear because there is a maximum velocity which is the speed of light; combining two velocities, both of which are close to, or equal to, c , cannot give a velocity greater than c . A linear transformation would necessarily allow this to happen, so a nonlinear transformation is required. To see how the transformations rule out finding a frame where an object moves faster than c , let us consider the transformation of a velocity $|\mathbf{u}'| = c$. From the second of Eqs. (47), we see that

$$u = \frac{c^2 + v^2 + 2cv \cos \theta' - v^2(1 - \cos^2 \theta')^{1/2}}{1 + v \cos \theta'/c} = c \frac{[(1 + v \cos \theta'/c)^2]^{1/2}}{1 + v \cos \theta'/c} = c. \quad (48)$$

Thus do we find what we already knew: If something moves at speed c in one frame, then it moves at the same speed in any other frame. More generally, if we had used any $u' \leq c$ and $v \leq c$, we would have recovered a $u \leq c$.

5.1 Aberration of Starlight

An interesting example of the application of the velocity transformation is the observed aberration of starlight. Suppose that an observer is moving with speed v at right angles to the direction of a star that he is watching. If a Galilean transformation is applied to the determination of the apparent direction of the star, one finds that it is seen at an angle ϕ away from its true direction where $\tan \phi = v/c$.

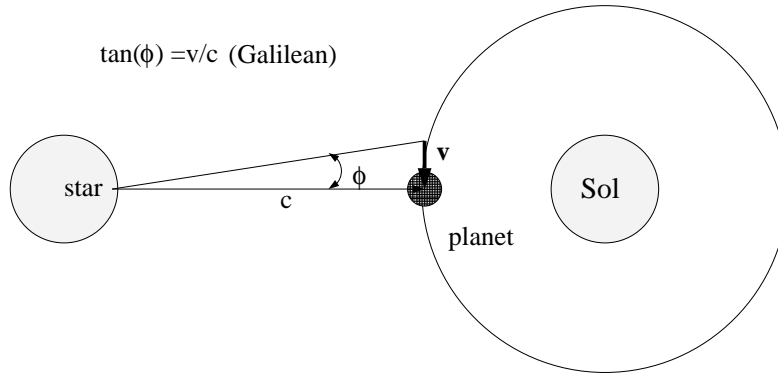


Figure 11: Due to the finite velocity of light, a star is seen an angle ϕ away from its true direction.

One can measure this angle by waiting six months. The velocity \mathbf{v} is provided by the earth's orbital motion; six months later it is reversed and if the observer then looks for the same star, its position will have shifted by 2ϕ , at least according to the Galilean transformation.

But that prediction is not correct. Consider what happens if the Lorentz transformation is used to compute the angle ϕ . Using Eq. (47), we see that the angle θ at which the light from the star appears to be headed in the frame K of the observer is

$$\tan \theta = \frac{c \sin \theta' \sqrt{1 - v^2/c^2}}{c(\cos \theta' + v/c)}. \quad (49)$$

where θ' is its direction in the frame K' which is the rest frame of the sun.

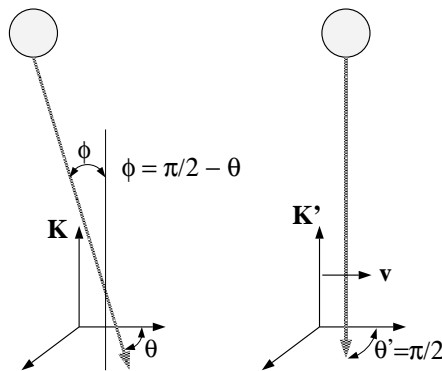


Figure 12: Coordinates for stargazing.

Now suppose that \mathbf{v} is at $\pi/2$ radians to the direction of the light's motion in frame K' so that $\theta' = \pi/2$. Then we find $\tan \theta = c/\gamma(v)v$. To compare with the prediction of the Galilean transformation, we need to find the angle ϕ , which is to say, $\pi/2 - \theta$. From a trigonometric identity, we have

$$\tan \theta = \frac{\tan(\pi/2) - \tan \phi}{1 + \tan(\pi/2) \tan \phi} = \frac{1}{\tan \phi}, \quad (50)$$

and so

$$\tan \phi = v\gamma(v)/c \quad \text{or} \quad \sin \phi = v/c. \quad (51)$$

This is the correct answer; the tangent of the angle ϕ differs from the prediction of the Galilean transformation by a factor of γ which is second-order in powers of v/c .

6 Doppler Shift

The Doppler shift of sound is a well-known and easy to understand phenomenon. It depends on the velocities of the source and observer relative to the medium in which the waves propagate. For electromagnetic waves, this medium does not exist and so the Doppler shift for light takes on its own special - and relatively simple! - form.

Suppose that in frame K there is a plane wave with wave vector \mathbf{k} and frequency ω . Put an observer at some point \mathbf{x} and set him to work counting wave crests as they go past him.

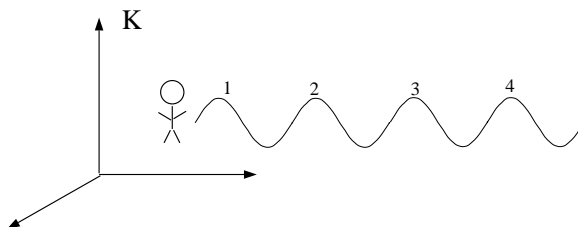


Figure 13: An observer counts wave crests.

Let him begin with the crest which passes the origin at $t = 0$ and continue counting until some later time t . How many crests does he count? We can decide by first determining when he starts. The starting time is $t_0 = (\mathbf{k} \cdot \mathbf{x})/kv_w$ where v_w is the velocity of the wave in frame K . The observer counts from t_0 to t and so counts n crests where

$$n = (t - t_0)/T; \quad (52)$$

T is the period of the wave, $T = 2\pi/\omega$. Hence,

$$n = \frac{1}{2\pi} \left(\omega t - \frac{\omega}{kv_w} \mathbf{k} \cdot \mathbf{x} \right) = \frac{1}{2\pi} (\omega t - \mathbf{k} \cdot \mathbf{x}), \quad (53)$$

since $\omega = v_w k$.

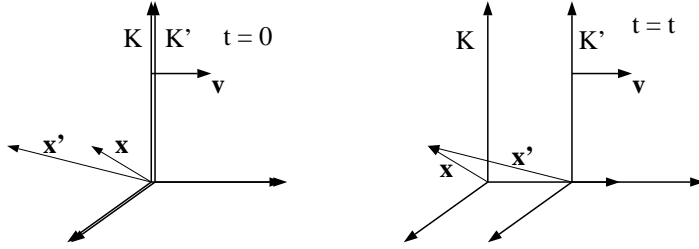


Figure 14: Frame alignments for the Doppler problem.

Now let the same measurement be performed by an observer at rest at a point \mathbf{x}' in frame K' which moves at \mathbf{v} relative to K . We choose \mathbf{x}' in a special way; it must be coincident with \mathbf{x} at time t (measured in K). This observer also counts crests, starting with the one that passed his origin at time $t' = 0$ and stopping with the one that arrives when he (the second observer) is coincident with the first observer. Given the usual transformations, the four-dimensional coordinate origins coincide, and so both observers count the same number of crests. Repeating the argument given for the number counted by the first observer, we find that the number counted by the second observer can be written as

$$n = \frac{1}{2\pi}(\omega't' - \mathbf{k}' \cdot \mathbf{x}') \quad (54)$$

where ω' and \mathbf{k}' are the frequency and wave vector of the wave in K' and (t', \mathbf{x}') is the spacetime point that transforms into (t, \mathbf{x}) . Thus we find

$$\omega t - \mathbf{k} \cdot \mathbf{x} = \omega't' - \mathbf{k}' \cdot \mathbf{x}'. \quad (55)$$

The significance of this relation is that the phase of the wave is an invariant. Further it appears to be the inner product of $(ct, i\mathbf{x})$ and $(\omega/c, i\mathbf{k})$. Because we know how (t, \mathbf{x}) transforms to (t', \mathbf{x}') , we can figure out how $(\omega/c, i\mathbf{k})$ transforms to $(\omega'/c, i\mathbf{k}')$. Let $\omega/c \equiv k_0$ and $\omega'/c \equiv k'_0$ and consider Eq. (55) with the transformations Eq. (30) used for t' and \mathbf{x}' :

$$\omega t - k_1 x_1 - k_2 x_2 - k_3 x_3 = \omega' \gamma (t - \beta x_1/c) - k'_1 \gamma (x_1 - \beta ct) - k'_2 x_2 - k'_3 x_3. \quad (56)$$

Because t and \mathbf{x} are completely arbitrary, we may conclude that

$$\omega = \gamma(\omega' + \beta ck'_1) \quad k_1 = \gamma(k'_1 + \beta\omega'/c) \quad k_2 = k'_2 \quad k_3 = k'_3, \quad (57)$$

or

$$k_0 = \gamma(k'_0 + \beta k'_1) \quad k_1 = \gamma(k'_1 + \beta k'_0) \quad k_2 = k'_2 \quad k_3 = k'_3. \quad (58)$$

We recognize the form of these transformations; they tell us that (k_0, \mathbf{k}) transforms in the same way as $(x_0, i\mathbf{x})$, i.e., via the Lorentz transformation.

Let's spend a few minutes thinking about the conditions under which our result is valid. We assumed when making the argument that we have a plane wave in both K and K' which means, more or less, that we are giving Maxwell's equations Law of Nature status since we assumed that the relevant equation of motion produces plane wave solutions in both frames. In fact, our results are not correct for waves in general, because many types of waves will not have this property (plane waves remain plane waves relative to all reference frames if they are plane waves relative to one frame). But they are correct for electromagnetic waves in vacuum.

Finally, let us look at an alternative form for our transformations. Let

$$\mathbf{k}' = k'(\cos \theta' \boldsymbol{\epsilon}_1 + \sin \theta' \boldsymbol{\epsilon}_2); \quad (59)$$

the component of \mathbf{k}' perpendicular to $\boldsymbol{\epsilon}_1$ is defined to be in the direction of $\boldsymbol{\epsilon}_2$. Further, $k' = \omega'/c$. Then the transformation equations may be used to produce the relations

$$k_1 = \gamma k'(\cos \theta' + \beta) \quad \mathbf{k}_2 = k' \sin \theta' \boldsymbol{\epsilon}_2 \quad \text{and} \quad \omega = \gamma \omega'(1 + \beta \cos \theta') \quad (60)$$

where \mathbf{k}_2 is the component of \mathbf{k} which is perpendicular to $\boldsymbol{\epsilon}_1$. From these results it is easy to show that $\omega - ck$, no surprise, and that

$$\cos \theta = \frac{\cos \theta' + \beta}{1 + \beta \cos \theta'}; \quad (61)$$

θ is the angle that \mathbf{k} makes with the direction of \mathbf{v} , (or $\boldsymbol{\epsilon}_1$); that is,

$$\mathbf{k} = k(\cos \theta \boldsymbol{\epsilon}_1 + \sin \theta \boldsymbol{\epsilon}_2). \quad (62)$$

6.1 Stellar Red Shift

The last of Eqs. (60) in particular may be used to describe the Doppler shift of the frequency of electromagnetic waves in vacuum. A well-known case in point is the “redshift” of light from distant galaxies.

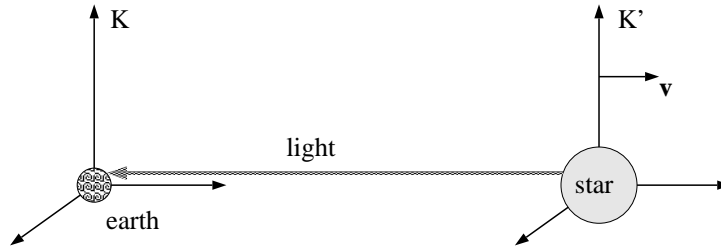


Figure 15: Light from receding stars in K' is redshifted when seen in K .

Given an object receding from the observer in K and emitting light of frequency ω' in its own rest frame, K' , we have $\cos \theta' \approx -1$ and

$$\omega = \gamma \omega' (1 - \beta) = \omega' \sqrt{\frac{1 - \beta}{1 + \beta}}. \quad (63)$$

For, e.g., $\beta = 1/2$, $\omega = \omega'/\sqrt{3}$. The observer sees the light as having much lower frequency than that with which it is emitted; it is “red-shifted.”

7 Four-tensors and all that

⁵ It is no accident that (x_0, \mathbf{x}) and (k_0, \mathbf{k}) transform from K to K' in the same way. They are but two of many sets of four objects or elements that have this property. They are called *four-vectors*. More generally, there are sets of 4^p elements, with $p = 0, 1, 2, \dots$, which have very similar transformation properties and which are called *four-tensors of rank p* . The better to manipulate them when the time comes, let us spend a little time now learning some of the basics of tensor calculus.

⁵The introduction to tensor calculus given in this section is largely drawn from J. L. Synge and A. Schild, *Tensor Calculus*, (University of Toronto Press, Toronto, 1949).

Consider the usual frames K and K' with coordinates \bar{x} and \bar{x}' , respectively; \bar{x} stands for (x_0, \mathbf{x}) and similarly for \bar{x}' . Let there be some transformation from one frame to the other which gives

$$\bar{x}' = x'(\bar{x}), \quad (64)$$

with an inverse,

$$\bar{x} = x(\bar{x}'). \quad (65)$$

These transformations need not in general be linear.

A **contravariant vector** or *rank-one tensor* is defined to be a set of four quantities or elements a^α , $\alpha = 0, 1, 2, 3$, which transform from K to K' according to the rule

$$a'^\alpha = \sum_{\beta=0}^3 \frac{\partial x'^\alpha}{\partial x^\beta} a^\beta \equiv A^\alpha{}_\beta a^\beta. \quad (66)$$

This equation serves to define $A^\alpha{}_\beta$,

$$A^\alpha{}_\beta \equiv \frac{\partial x'^\alpha}{\partial x^\beta}; \quad (67)$$

we have also introduced in the last step the summation convention that a Greek index, which appears in a term as both an upper and a lower index, is summed from zero to three.

For any contravariant vector or tensor, we are going to introduce also a *covariant* vector or tensor whose components will be designated by subscripts. Define a **covariant vector** or *rank-one tensor* as a set of four objects b_α , $\alpha = 0, 1, 2, 3$, which transform according to the rule

$$b'_\alpha = \sum_{\beta=0}^3 \frac{\partial x^\beta}{\partial x'^\alpha} b_\beta \equiv A_\alpha{}^\beta b_\beta \quad (68)$$

where we have defined

$$A_\alpha{}^\beta \equiv \frac{\partial x^\beta}{\partial x'^\alpha}. \quad (69)$$

The generalization to tensors of ranks other than one is straightforward. For example, a rank-two contravariant tensor comprises a set of sixteen objects $T^{\alpha\beta}$ which

transform according to the rule

$$T'^{\alpha\beta} = A^\alpha_\gamma A^\beta_\delta T^{\gamma\delta} \quad (70)$$

and a rank-two covariant tensor has sixteen elements $T_{\alpha\beta}$ which transform according to the rule

$$T'_{\alpha\beta} = A_\alpha^\gamma A_\beta^\delta T_{\gamma\delta}. \quad (71)$$

Mixed tensors can also be of interest. The rank-two mixed tensor \bar{T} is a set of sixteen elements T^α_β which transform according to

$$T'^\alpha_\beta = A^\alpha_\gamma A_\beta^\delta T^\gamma_\delta. \quad (72)$$

Generalizations follow as you would expect.

The *inner product* of \bar{a} and \bar{b} can be⁶ defined as

$$\bar{a} \cdot \bar{b} \equiv b_\alpha a^\alpha. \quad (73)$$

Consider the transformation properties of the inner product:

$$\bar{a}' \cdot \bar{b}' = a'^\alpha b'_\alpha = A_\alpha^\gamma A^\alpha_\delta b_\gamma a^\delta; \quad (74)$$

however,

$$A_\alpha^\gamma A^\alpha_\delta = \frac{\partial x^\gamma}{\partial x'^\alpha} \frac{\partial x'^\alpha}{\partial x^\delta} = \frac{\partial x^\gamma}{\partial x^\delta} = \delta^\gamma_\delta \quad (75)$$

where

$$\delta^\gamma_\delta \equiv \begin{cases} 1 & \gamma = \delta \\ 0 & \gamma \neq \delta \end{cases} \quad (76)$$

Hence

$$\bar{a}' \cdot \bar{b}' = \delta^\gamma_\delta a^\delta b_\gamma = a^\gamma b_\gamma = \bar{a} \cdot \bar{b}. \quad (77)$$

The inner product is an invariant, also known as a *scalar* or *rank-zero tensor*.

Notice that when we wrote the Kronecker delta function, we gave it a superscript and subscript as though it were a rank-two mixed tensor. It in fact is one as we can

⁶We will present a different but equivalent definition later.

show by transforming it from one frame to another. Let δ_α^β be defined as above in the frame K and let it be defined to be a mixed tensor. Then we know how it transforms and so can find it in a different frame K' (where we hope it will turn out to be the same as in frame K):

$$\delta_\beta^{\prime\alpha} = A^\alpha_\gamma A_\beta^\delta \delta_\delta^\gamma = A^\alpha_\gamma A_\beta^\gamma = \frac{\partial x^{\prime\alpha}}{\partial x^\gamma} \frac{\partial x^\gamma}{\partial x^{\prime\beta}} = \frac{\partial x^{\prime\alpha}}{\partial x^{\prime\beta}} = \delta_\beta^\alpha \quad (78)$$

which means that the thing we defined to be a rank-two mixed tensor is remains the same as the Kronecker delta function in all frames.

The operation which enters the definition of the inner product is to set a contravariant and a covariant index equal to each other and then to sum them. This operation is called a *contraction* with respect to the pair of indices in question. It reduces the rank of something by two. That is, the sixteen objects $b_\alpha a^\beta$ form a rank-two tensor, as may be shown easily by checking how it transforms (given the transformation properties of b_α and a^β). After we perform the contraction, we are left with a rank-zero tensor.

7.1 The Metric Tensor

Now think about how we can use these things in relativity. We have a fundamental invariant which is the separation between two events; specifically,

$$(ds)^2 = (dx^0)^2 - (dx^1)^2 - (dx^2)^2 - (dx^3)^2 \quad (79)$$

is an invariant,

$$(ds)^2 = (ds')^2. \quad (80)$$

We would like to write this as an inner product $d\bar{x} \cdot d\bar{x}$, where

$$d\bar{x} \cdot d\bar{x} = dx_\alpha dx^\alpha. \quad (81)$$

However, in order that we can do so, it must be the case that the covariant four-vector $d\bar{x}$ have the components

$$dx_0 = dx^0 \quad \text{and} \quad dx_i = -dx^i \quad \text{for } i = 1, 2, 3. \quad (82)$$

In general, the components of the contravariant and covariant versions of a four-vector are related by the *metric tensor* \bar{g} which is a rank-two tensor that can be expressed in covariant, contravariant, or mixed form (just like any other tensor of rank two or more). In particular, the *covariant metric tensor* is defined for any system by the statement that the separation can be written as

$$(ds)^2 \equiv g_{\alpha,\beta} dx^\alpha dx^\beta, \quad (83)$$

plus the statement that it is a symmetric tensor.

How do we know that this is a tensor? From the fact that its double contraction with the contravariant vector \bar{x} is an invariant and from the fact that it is symmetric, one can prove that it is a rank-two covariant tensor.⁷

In three-dimensional Cartesian coordinates in a Euclidean space such as we are accustomed to thinking about, the covariant metric tensor is just the unit tensor. In curvilinear coordinates (for example, spherical coordinates) it is some other (still simple) thing. For the flat four-dimensional space that one deals with in the special theory of relativity, we can see from Eqs. (79) and (83), and from the condition that \bar{g} is symmetric, that it must be

$$g_{00} = 1, \quad g_{ii} = -1, \quad i = 1, 2, 3, \quad \text{and} \quad g_{\alpha\beta} = 0, \quad \alpha \neq \beta. \quad (84)$$

Next, we introduce the *contravariant metric tensor*. First, we take the determinant of the matrix formed by the covariant metric tensor,

$$g \equiv \det[g_{\alpha\beta}] = -1 \quad (85)$$

Then one introduces the cofactor, written as $\Delta^{\alpha\beta}$, of each element $g_{\alpha\beta}$ in the matrix. The elements of the contravariant metric tensor are defined as

$$g^{\alpha\beta} \equiv \frac{\Delta^{\alpha\beta}}{g}. \quad (86)$$

⁷See, e.g., Synge and Schild.

We need to demonstrate that this thing is indeed a contravariant tensor. From the standard definitions of the determinant and cofactor, we can write

$$g_{\alpha\beta}\Delta^{\alpha\gamma} = g_{\beta\alpha}\Delta^{\gamma\alpha} = \delta_{\beta}^{\gamma}g \quad (87)$$

from which it follows that

$$g_{\alpha\beta}g^{\alpha\gamma} = \delta_{\beta}^{\gamma} = g_{\beta\alpha}g^{\gamma\alpha}. \quad (88)$$

When contracted (as above) with a covariant tensor, the thing we call a contravariant tensor produces a mixed tensor. In addition, it is symmetric which follows from the symmetry of the covariant metric tensor. This is sufficient to prove that the elements $g^{\alpha\beta}$ do form a contravariant tensor.

It is easy to work out the elements of the contravariant metric tensor if one knows the covariant one; for our particular metric tensor they are the same as the elements of the covariant one.

The metric tensor is used to convert contravariant tensors or indices to covariant ones and conversely. Consider for example the elements x_{α} defined by

$$x_{\alpha} = g_{\alpha\beta}x^{\beta}. \quad (89)$$

It is clear that the result is a covariant tensor of rank one. It is the covariant version of the position four-vector \bar{x} and has elements $(x^0, -\mathbf{x})$. Similarly, we may recover the contravariant version of a four-vector or tensor from the covariant version of the same tensor by using the contravariant metric tensor:

$$x^{\alpha} = g^{\alpha\beta}x_{\beta} = g^{\alpha\beta}g_{\beta\gamma}x^{\gamma} = \delta_{\gamma}^{\alpha}x^{\gamma} = x^{\alpha}. \quad (90)$$

More generally, one may raise or lower as many indices as one wishes by using the appropriate metric tensor as many times as needed. Among other things, we can thereby construct a mixed metric tensor,

$$g_{\alpha}^{\beta} = g_{\alpha\gamma}g^{\gamma\beta}; \quad (91)$$

Using the explicit components of the covariant and contravariant metric tensors, one finds that this is precisely the unit mixed tensor, i.e., the Kronecker delta,

$$g_{\alpha}^{\beta} = \delta_{\alpha}^{\beta} \quad (92)$$

Finally, we earlier defined the inner product of two vectors by contracting the covariant version of one with the contravariant version of the other; we can now see that there are numerous other ways to express the inner product:

$$\bar{a} \cdot \bar{b} = a^{\alpha} b_{\alpha} = g_{\alpha\gamma} a^{\alpha} b^{\gamma} = g^{\alpha\gamma} a_{\gamma} b_{\alpha}; \quad (93)$$

In particular, the separation is now seen to be the same as $\bar{x} \cdot \bar{x}$,

$$(s)^2 = g_{\alpha\beta} x^{\alpha} x^{\beta} = x^{\alpha} x_{\alpha}. \quad (94)$$

There is one piece of unfinished business in all of this. We have defined a metric tensor; it was defined so that the separation is an invariant. We still do not know (if we assume we haven't as yet learned about Lorentz transformations) the components A^{α}_{β} and A_{β}^{α} of the transformation matrices. Just any old transformations won't do; it has to be consistent with our metric tensor, i.e., with the condition that the separation is invariant. This implies some conditions on the transformations. We shall return to this point later.

7.2 Differential Operators

Differential operators also have simple transformation properties. Consider the basic example of the four operators $\partial/\partial x^{\alpha}$. The transformation of this from one frame to another is found from the relation

$$\frac{\partial}{\partial x'^{\alpha}} = \frac{\partial x^{\beta}}{\partial x'^{\alpha}} \frac{\partial}{\partial x^{\beta}} \equiv A_{\alpha}^{\beta} \frac{\partial}{\partial x^{\beta}}. \quad (95)$$

The components of this operator transform in the same way as the components of a covariant vector which means that the four differential operators $\partial/\partial x^{\alpha}$ form a

covariant four-vector. That being the case the elements

$$A_\alpha{}^\beta = \frac{\partial x^\beta}{\partial x'^\alpha} \quad (96)$$

are the elements of a rank-two mixed tensor, and that is why we have all along used for them notation which suggests that they are components of such a tensor.

It is equally true that $\partial/\partial x_\alpha$ is a contravariant four-vector operator. Consider

$$\begin{aligned} \frac{\partial}{\partial x'_\alpha} &= \frac{\partial x_\beta}{\partial x'_\alpha} \frac{\partial}{\partial x_\beta} = g^{\beta\gamma} \frac{\partial x^\gamma}{\partial x'^\delta} \frac{\partial x'^\delta}{\partial x'_\alpha} \frac{\partial}{\partial x_\beta} \\ &= g^{\beta\gamma} g^{\delta\alpha} A_\delta{}^\gamma \frac{\partial}{\partial x_\beta} = A^\alpha{}_\beta \frac{\partial}{\partial x_\beta}. \end{aligned} \quad (97)$$

Since the operator transforms in the same way as a contravariant four-vector, it is a contravariant four-vector!

Either of these four-vectors is called the *four-divergence*. Let's introduce some new notation for them:

$$\partial_\alpha \equiv \frac{\partial}{\partial x^\alpha} \quad (98)$$

is the way we shall write a component of the covariant four-divergence, and

$$\partial^\alpha \equiv \frac{\partial}{\partial x_\alpha} \quad (99)$$

is the way we write a component of the contravariant four-divergence.

We can construct some interesting invariants using the four-divergence. For example, the inner product of one of them with an four-vector produces an invariant,

$$\partial^\alpha A_\alpha = \partial_\alpha A^\alpha = \frac{\partial A^0}{\partial x^0} + \nabla \cdot \mathbf{A}. \quad (100)$$

Also, the four-dimensional Laplacian

$$\partial^\alpha \partial_\alpha = \frac{\partial^2}{\partial x^{0^2}} - \nabla \cdot \nabla \equiv \square \quad (101)$$

is an invariant, or scalar, operator.

7.3 Notation

It is natural to present four-vectors using column vectors and rank-two tensors using matrices. Thus a four-vector such as \bar{x} becomes

$$\bar{x} = \begin{pmatrix} x^0 \\ x^1 \\ x^2 \\ x^3 \end{pmatrix}, \quad (102)$$

and its transpose is

$$\tilde{x} = (x^0 \ x^1 \ x^2 \ x^3). \quad (103)$$

Using this notation, we can write, e.g., the inner product of two vectors as

$$\bar{a} \cdot \bar{b} = a_\alpha b^\alpha = g_{\alpha\beta} a^\alpha b^\beta = \tilde{a} \bar{g} \bar{b}. \quad (104)$$

Notice however, that if \tilde{a} were the transpose of the covariant vector, we would write the inner product as $\tilde{a} \bar{b}$. The notation leaves something to be desired. Be that as it may, we can write a transformation as

$$x'^\alpha = A^\alpha_\beta x^\beta \quad \text{or} \quad \bar{x}' = \bar{A} \bar{x}. \quad (105)$$

We will only make use of this abbreviated notation when it is necessary to cause lots of confusion.

8 Representation of the Lorentz transformation

Our next task is to find a general transformation matrix⁸ \bar{A} . As pointed out earlier, the basic fact we have to work with is that the separation is invariant,

$$g_{\alpha\beta} dx^\alpha dx^\beta = g_{\alpha\beta} dx'^\alpha dx'^\beta. \quad (106)$$

⁸This could be the fully contravariant version, the fully covariant version or one of the two mixed versions. If we know one of them, we know the others because we can lower and raise indices with the metric tensor.

Knowing this, and knowing also that

$$x'^{\alpha} = A^{\alpha}_{\beta} x^{\beta}, \quad (107)$$

it is a standard and straightforward exercise in linear algebra to show that

$$\det |A| = \pm 1. \quad (108)$$

Just as in three dimensions, there are proper and improper transformations which satisfy our requirements. The proper ones may be arrived at via a sequence of infinitesimal transformations starting from the identity, $A^{\alpha}_{\beta} = g^{\alpha}_{\beta}$. All transformations generated in this manner have determinant $+1$. The improper ones cannot be constructed in this way, even though some of them can have determinant $+1$. An example is $A^{\alpha}_{\beta} = -g^{\alpha}_{\beta}$; it has determinant $+1$ but is an improper transformation and cannot be arrived at by a sequence of infinitesimal transformations.

In this investigation we shall construct proper Lorentz transformations and shall build them from infinitesimal ones. Let's start by writing

$$A^{\alpha}_{\beta} = \delta^{\alpha}_{\beta} + \Delta\omega^{\alpha}_{\beta}, \quad (109)$$

where $\Delta\omega^{\alpha}_{\beta}$ is an infinitesimal. From the invariance of the interval, one can easily show that of the sixteen components $\Delta\omega^{\alpha}_{\beta}$, the diagonal ones must be zero and the off-diagonal ones must be such that

$$\Delta\omega^{\alpha\beta} = -\Delta\omega^{\beta\alpha}; \quad (110)$$

notice that both indices are now contravariant, in contrast to the previous equation. If we write the preceding relation with one contravariant and one covariant index, we will find the same $-$ sign if the two indices are 1, 2, or 3, and there will be no $-$ if one index is 0 and the other is one of 1, 2, or 3. Evidently, it is simpler to use a completely contravariant form.⁹

⁹The point is, we can use any form for the tensor that we like because all forms can be found from any single one. Therefore, it makes sense to use that form in which the relations are simplest, if there is one.

These results demonstrate that we have just six independent infinitesimals. We may take them to be a set of six numbers without indices if we introduce suitable basis matrices. One such set of matrices is given by

$$(K_1)^\alpha_\beta = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}; \quad (K_2)^\alpha_\beta = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad (111)$$

$$(K_3)^\alpha_\beta = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}; \quad (S_1)^\alpha_\beta = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \quad (112)$$

$$(S_2)^\alpha_\beta = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix}; \quad (S_3)^\alpha_\beta = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \quad (113)$$

The most general infinitesimal transformation can now be written as

$$\bar{A} = \bar{g} - \Delta\boldsymbol{\omega} \cdot \bar{\mathbf{S}} - \Delta\boldsymbol{\zeta} \cdot \bar{\mathbf{K}}; \quad (114)$$

where $\Delta\boldsymbol{\omega}$ contains three independent infinitesimal components as does $\Delta\boldsymbol{\zeta}$; these are, respectively, just infinitesimal coordinate rotations and infinitesimal relative velocities.

Powers of the matrices \bar{K}_i and \bar{S}_i have some very special properties. For example,

$$(\bar{K}_1)^2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad \text{and} \quad (\bar{S}_1)^2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}; \quad (115)$$

consequently, powers of the matrices tend to repeat; the periods of these cycles are two and four for the \bar{K} 's and the \bar{S} 's, respectively so that, for any m and integral n ,

$$(\bar{K}_i)^{m+2n} = (\bar{K}_i)^m \quad \text{and} \quad (\bar{S}_i)^{m+4n} = (\bar{S}_i)^m. \quad (116)$$

We have seen above what is the second power in each case; for the \bar{K} 's, the third power is the same as the first and for the \bar{S} 's, one finds the negative of the first,

$$(\bar{S}_i)^3 = -\bar{S}_i; \quad (117)$$

and finally, the fourth power of one of the \bar{S} 's has two 1's on the diagonal, much like the even powers of the \bar{K} 's.

We can construct the matrix for a finite transformation by making a sequence of many infinitesimal transformations. To this end consider some finite $\boldsymbol{\omega}$ and $\boldsymbol{\zeta}$ and relate them to the infinitesimals by

$$\Delta\boldsymbol{\omega} = \boldsymbol{\omega}/n \quad \text{and} \quad \Delta\boldsymbol{\zeta} = \boldsymbol{\zeta}/n, \quad (118)$$

where n is a very large number. Now apply \bar{A} (given by Eq. (114)) to \bar{x} n times, thereby producing some \bar{x}' :

$$x'^{\alpha} = \left(g - \frac{\boldsymbol{\omega} \cdot \mathbf{S}}{n} - \frac{\boldsymbol{\zeta} \cdot \mathbf{K}}{n} \right)_{\alpha_1}^{\alpha} \left(g - \frac{\boldsymbol{\omega} \cdot \mathbf{S}}{n} - \frac{\boldsymbol{\zeta} \cdot \mathbf{K}}{n} \right)_{\alpha_2}^{\alpha_1} \dots \left(g - \frac{\boldsymbol{\omega} \cdot \mathbf{S}}{n} - \frac{\boldsymbol{\zeta} \cdot \mathbf{K}}{n} \right)_{\alpha_n}^{\alpha_{n-1}} x^{\alpha_n} \quad (119)$$

We want to take the $n \rightarrow \infty$ limit of this expression. In general,

$$\lim_{n \rightarrow \infty} \left(1 + \frac{a}{n} \right)^n = e^a, \quad (120)$$

as one can show by, e.g., considering the logarithm. Applying this fact, we find that

$$x'^{\alpha} = A^{\alpha}_{\beta} x^{\beta} \quad (121)$$

where

$$A^{\alpha}_{\beta} = \left(e^{-\boldsymbol{\omega} \cdot \mathbf{S} - \boldsymbol{\zeta} \cdot \mathbf{K}} \right)_{\beta}^{\alpha}. \quad (122)$$

We can get a little understanding of what this equation is telling us by considering some special cases which are also familiar. For example, let $\boldsymbol{\omega} = 0$ and $\boldsymbol{\zeta} = \zeta \boldsymbol{\epsilon}_1$. Then

$$\begin{aligned}\bar{A} &= e^{-\zeta \bar{K}_1} = 1 - \zeta \bar{K}_1 + \frac{\zeta^2}{2} \bar{K}_1^2 - \frac{\zeta^3}{6} \bar{K}_1^3 + \dots \\ &= 1 - \bar{K}_1 \left(\zeta + \frac{\zeta^3}{6} + \dots \right) + \bar{K}_1^2 \left(1 + \frac{\zeta^2}{2} + \dots \right) - \bar{K}_1^3 \\ &= 1 - \bar{K}_1^2 + (\cosh \zeta) \bar{K}_1^2 - (\sinh \zeta) \bar{K}_1\end{aligned}\tag{123}$$

or

$$A^\alpha_\beta = \begin{pmatrix} \cosh \zeta & -\sinh \zeta & 0 & 0 \\ -\sinh \zeta & \cosh \zeta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}\tag{124}$$

which should be familiar. Similarly, if $\boldsymbol{\omega} = \omega \boldsymbol{\epsilon}_1$ with $\boldsymbol{\zeta} = 0$, one finds

$$A^\alpha_\beta = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos \omega & \sin \omega \\ 0 & 0 & -\sin \omega & \cos \omega \end{pmatrix}\tag{125}$$

which we recognize as a simple rotation around the x -axis.

Our general result for \bar{A} allows us to find the transformation matrix for any combination of $\boldsymbol{\omega}$ and $\boldsymbol{\zeta}$. In particular, one can show that for $\boldsymbol{\omega} = 0$ and general $\boldsymbol{\zeta}$ such that $\boldsymbol{\beta}$ has magnitude $\tanh \zeta$ and is in the direction of $\boldsymbol{\zeta}$. Writing the components of $\boldsymbol{\beta}$ as β_i , $i = 1, 2, 3$, we find that \bar{A} is

$$A^\alpha_\beta = \begin{pmatrix} \gamma & -\gamma\beta_1 & -\gamma\beta_2 & -\gamma\beta_3 \\ -\gamma\beta_1 & 1 + \frac{(\gamma-1)\beta_1^2}{\beta^2} & \frac{(\gamma-1)\beta_1\beta_2}{\beta^2} & \frac{(\gamma-1)\beta_1\beta_3}{\beta^2} \\ -\beta_2 & \frac{(\gamma-1)\beta_1\beta_2}{\beta^2} & 1 + \frac{(\gamma-1)\beta_2^2}{\beta^2} & \frac{(\gamma-1)\beta_2\beta_3}{\beta^2} \\ -\gamma\beta_3 & \frac{(\gamma-1)\beta_1\beta_3}{\beta^2} & \frac{(\gamma-1)\beta_2\beta_3}{\beta^2} & 1 + \frac{(\gamma-1)\beta_3^2}{\beta^2} \end{pmatrix},\tag{126}$$

in case anybody wanted to know.

9 Covariance of Electrodynamics

In this section we are going to demonstrate the consistency of the Maxwell equations with Einstein's first postulate. But first we must decide more precisely what it means for a "law of nature" to be "the same" in all inertial frames. The relevant statement is this: an equation expressing a law of nature must be **invariant in form** under Lorentz transformations. When this is the case, the equation is said to be *Lorentz covariant* or simply *covariant*, which has nothing to do with the definition of covariant as opposed to contravariant tensors. And what is meant by the phrase "invariant in form" which appears above? It means that the quantities in the equation must transform in well-defined ways (as particular components of some four-tensors, for example) and that when terms are grouped in an appropriate manner, each group transforms in the same way as each of the other groups. In order to determine whether the Maxwell equations can have this property, we must first figure out how each of the physical objects in those equations, that is, \mathbf{E} , \mathbf{B} , ρ , and \mathbf{J} , transforms.

9.1 Transformations of Source and Fields

9.1.1 ρ and \mathbf{J}

Let's start with the electric charge. It is an experimental observation that charge is an invariant. If a system has a particular charge q as measured in one frame, then it has the same charge q when the measurements are made in a different frame. From this (experimental) fact and things we already know, we can determine how charge density and current density transform.

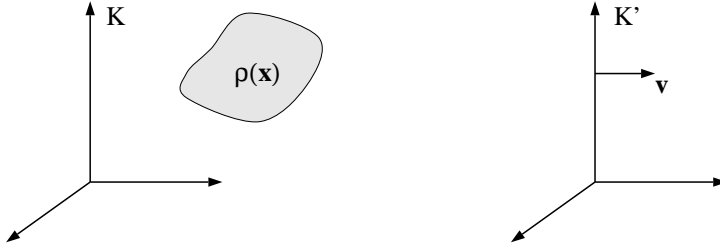


Figure 16: The charge density transforms like time.

Suppose that we have a system with charge density ρ , as measured in K , and ρ' as measured in K' . Then in volume d^3x in K , there is charge dq where

$$dq = \rho d^3x = \rho d^3x dt/dt \quad (127)$$

where we have introduced an infinitesimal time element dt as well. Similarly, in K' , the charge dq' in the volume element d^3x' can be written as

$$dq' = \rho' d^3x' dt'/dt'. \quad (128)$$

Now, if d^3x' is what d^3x transforms into (that is, if it is the same volume element as d^3x), then charge invariance implies that $dq = dq'$. Further, if dt' is what dt transforms into, then we can say that

$$c d^3x' dt' \equiv d^4x' = \left| \frac{\partial(x'^0, x'^1, x'^2, x'^3)}{\partial(x^0, x^1, x^2, x^3)} \right| d^4x \equiv |\det[\bar{A}]| d^4x. \quad (129)$$

But the determinant of \bar{A} is unity, so we have shown that a spacetime volume element is an invariant,

$$d^4x = d^4x'. \quad (130)$$

As applied to the present inquiry, we use this statement along with the equality of dq and dq' (and the invariance of c) to conclude that

$$\rho/dt = \rho'/dt'. \quad (131)$$

This relation can be true only if the charge density transforms in the same way as the time; that is, it must be the 0^{th} component of a four-vector.

Where are the other three components of this four-vector? They are the current density. Since \mathbf{J} is ρ times a velocity, which is in turn the ratio $d\mathbf{x}/dt$, we can write

$$\mathbf{J} = \rho \mathbf{u} = \rho \frac{d\mathbf{x}}{dt}; \quad (132)$$

in view of the fact that ρ/dt is an invariant, \mathbf{J} must transform in the same way as $d\mathbf{x}$, which is to say, as the 1,2,3 components of a (contravariant) four-vector. Hence we have the *contravariant current four-vector*

$$J^\alpha = (c\rho, \mathbf{J}); \quad (133)$$

the covariant current four-vector is

$$J_\alpha = (c\rho, -\mathbf{J}). \quad (134)$$

Knowing this, we are not surprised to find that the charge conservation equation is a four-divergence equation,

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{J} = 0 \quad \text{or} \quad \partial_\alpha J^\alpha = 0. \quad (135)$$

Notice that this equation is “covariant” in the sense introduced earlier; both sides are scalars.

9.1.2 Potentials

Now we shall proceed by demanding that all the relevant equations be Lorentz covariant. We shall apply this requirement to equations that we already have and see what are the implications for the fields \mathbf{E} and \mathbf{B} and also see that no contradictions arise. Let’s start with the equations for the potentials in the Lorentz gauge. The equations of motion are

$$\square \mathbf{A}(\mathbf{x}, t) = \frac{4\pi}{c} \mathbf{J}(\mathbf{x}, t) \quad \text{and} \quad \square \Phi(\mathbf{x}, t) = 4\pi \rho(\mathbf{x}, t); \quad (136)$$

these can all be written in the very brief notation

$$\square A^\alpha(\mathbf{x}, t) = \frac{4\pi}{c} J^\alpha(\mathbf{x}, t) \quad (137)$$

where we have introduced

$$A^\alpha \equiv (\Phi, \mathbf{A}) \quad (138)$$

which must be a contravariant four-vector if the equations of motion above are the correct equations of motion for the potential in the Lorentz gauge in every inertial frame. Notice that the potentials in gauges other than the Lorentz gauge will not form a four-vector.

The Lorentz condition, which is satisfied by potentials in the Lorentz gauge, is

$$\nabla \cdot \mathbf{A}(\mathbf{x}, t) + \frac{1}{c} \frac{\partial \Phi}{\partial t} = 0; \quad (139)$$

this equation may also be written as a four-divergence of a four-vector,

$$\partial_\alpha A^\alpha = 0. \quad (140)$$

9.1.3 Fields, Field-Strength Tensor

Let's look next at \mathbf{E} and \mathbf{B} ; these are given by

$$\mathbf{B}(\mathbf{x}, t) = \nabla \times \mathbf{A}(\mathbf{x}, t) \quad \text{and} \quad \mathbf{E}(\mathbf{x}, t) = -\nabla \Phi(\mathbf{x}, t) - \frac{1}{c} \frac{\partial \mathbf{A}(\mathbf{x}, t)}{\partial t}. \quad (141)$$

Look at just the x -components:

$$E_x = -\frac{1}{c} \frac{\partial A_x}{\partial t} - \frac{\partial \Phi}{\partial x} = -\frac{\partial A^1}{\partial x^0} - \frac{\partial \Phi}{\partial x^1} = -\frac{\partial A^1}{\partial x_0} + \frac{\partial A^0}{\partial x_1} = -\partial^0 A^1 + \partial^1 A^0. \quad (142)$$

Similarly, a component of the magnetic induction turns out to be, e.g.,

$$B_x = -\partial^2 A^3 + \partial^3 A^2. \quad (143)$$

Given the four-vector character of the differential operators and of the potentials, we can see that these particular components of the electric field and magnetic induction are elements of a rank-two tensor which we have expressed here in contravariant form. Let us define the *field-strength tensor* \bar{F} by

$$F^{\alpha\beta} \equiv \partial^\alpha A^\beta - \partial^\beta A^\alpha. \quad (144)$$

This turns out to be

$$F^{\alpha\beta} = \begin{pmatrix} 0 & -E_x & -E_y & -E_z \\ E_x & 0 & -B_z & B_y \\ E_y & B_z & 0 & -B_x \\ E_z & -B_y & B_x & 0 \end{pmatrix}. \quad (145)$$

Because the tensor is antisymmetric, it has just six independent entries, these being the six components of \mathbf{E} and \mathbf{B} .

The corresponding covariant tensor is easily worked out. It is the same as the contravariant one except that the signs of the entries in the first column and the first row are reversed. A somewhat different object which contains the same information is the *dual field-strength tensor* $\bar{\mathcal{F}}$ which is defined by

$$\mathcal{F}^{\alpha\beta} \equiv \epsilon^{\alpha\beta\gamma\delta} \frac{1}{2} F_{\gamma\delta} \quad (146)$$

where the *fully antisymmetric rank-four unit pseudotensor* with components $\epsilon^{\alpha\beta\gamma\delta}$ is in turn defined by specifying (1) that in frame K

$$\epsilon^{\alpha\beta\gamma\delta} \equiv \begin{cases} 1 & \text{if } \alpha\beta\gamma\delta \text{ is an even permutation of } 1234 \\ -1 & \text{if } \alpha\beta\gamma\delta \text{ is an odd permutation of } 1234 \\ 0 & \text{otherwise,} \end{cases} \quad (147)$$

and (2) that it transforms to other frames as a rank-four pseudotensor must,

$$(\epsilon')^{\alpha\beta\gamma\delta} \equiv \det[\bar{A}] A^\alpha_\phi A^\beta_\chi A^\gamma_\psi A^\delta_\omega \epsilon^{\phi\chi\psi\omega}. \quad (148)$$

Applying this definition, one can show that the components of this pseudotensor are given by Eq. (147) not only in frame K but in all inertial frames.

Although $\bar{\epsilon}$ is a pseudotensor as opposed to a true tensor, the distinction will not be important for us so long as we stick to proper Lorentz transformations or to improper ones that have determinant $+1$. In what follows, we will refer to it as a tensor even though we know better; similarly we will refer to the dual tensor as a “tensor” (as was done in the definition) even though it is in fact a pseudotensor.

Returning now to the original point, $\mathcal{F}^{\alpha\beta}$ is, explicitly,

$$\mathcal{F}^{\alpha\beta} = \begin{pmatrix} 0 & -B_x & -B_y & -B_z \\ B_x & 0 & E_z & -E_y \\ B_y & -E_z & 0 & E_x \\ B_z & E_y & -E_x & 0 \end{pmatrix}. \quad (149)$$

9.2 Invariance of Maxwell Equations

Now we know how everything transforms; it remains to be seen whether the Maxwell equations are Lorentz covariant. The inhomogeneous equations are

$$\nabla \cdot \mathbf{E}(\mathbf{x}, t) = 4\pi\rho(\mathbf{x}, t) \quad \text{and} \quad \nabla \times \mathbf{B}(\mathbf{x}, t) - \frac{1}{c} \frac{\partial \mathbf{E}(\mathbf{x}, t)}{\partial t} = \frac{4\pi}{c} \mathbf{J}(\mathbf{x}, t). \quad (150)$$

The first of these is

$$\frac{\partial F^{10}}{\partial x^1} + \frac{\partial F^{20}}{\partial x^2} + \frac{\partial F^{30}}{\partial x^3} = \frac{4\pi}{c} J^0. \quad (151)$$

Because $F^{00} \equiv 0$, we may add a term $\partial F^{00}/\partial x^0$ to the left-hand side of this equation and then find that it reads

$$\partial_\alpha F^{\alpha 0} = \frac{4\pi}{c} J^0. \quad (152)$$

This equation is clearly the 0^{th} component of a four-vector equation in which the left-hand side is obtained by taking the divergence of a rank-two tensor. The other three inhomogeneous Maxwell equations may be analyzed in similar fashion and the four may be concisely written as

$$\partial_\alpha F^{\alpha\beta} = \frac{4\pi}{c} J^\beta \quad (153)$$

where $\beta = 0, 1, 2$, and 3 . These are manifestly Lorentz covariant.

The homogeneous Maxwell equations are

$$\nabla \cdot \mathbf{B}(\mathbf{x}, t) = 0 \quad \text{and} \quad \nabla \times \mathbf{E}(\mathbf{x}, t) = -\frac{1}{c} \frac{\partial \mathbf{B}(\mathbf{x}, t)}{\partial t}. \quad (154)$$

The first one can be written as

$$\frac{\partial \mathcal{F}^{10}}{\partial x^1} + \frac{\partial \mathcal{F}^{20}}{\partial x^2} + \frac{\partial \mathcal{F}^{30}}{\partial x^3} = 0, \quad (155)$$

or, since $\mathcal{F}^{00} = 0$,

$$\partial_\alpha \mathcal{F}^{\alpha 0} = 0. \quad (156)$$

The others can be expressed in similar fashion, and all four are contained in the following equation:

$$\partial_\alpha \mathcal{F}^{\alpha\beta} = 0, \quad (157)$$

where $\beta = 0, 1, 2,$ and 3 . This form is clearly covariant, establishing the covariance of Maxwell's equations. These equations are components of a rank-one pseudotensor. They may also be written as components of a rank-three tensor. Notice that $\nabla \cdot \mathbf{B} = 0$ is, in tensor notation,

$$\frac{\partial F^{31}}{\partial x^2} + \frac{\partial F^{23}}{\partial x^1} + \frac{\partial F^{12}}{\partial x^3} = 0. \quad (158)$$

The remaining three homogeneous Maxwell equations can be expressed in similar fashion, and all four can be written as

$$\partial^\alpha F^{\beta\gamma} + \partial^\gamma F^{\alpha\beta} + \partial^\beta F^{\gamma\alpha} = 0 \quad (159)$$

where $\alpha, \beta,$ and γ are any three of $0,1,2,3$, giving four equations. The other possible choices of the superscripts (involving repetition of two or more values) give nothing (They give $0=0$). Hence we have succeeded in writing each of the homogeneous Maxwell equations in the form of an element of a rank-three tensor and the Lorentz covariant equation we have constructed simply says that this tensor is equal to zero.

10 Transformation of the electromagnetic field

The transformation properties of \mathbf{E} and \mathbf{B} are easily worked out by making use of our knowledge of how a rank-two tensor must transform:

$$(F')^{\alpha\beta} = A^\alpha_\gamma A^\beta_\delta F^{\gamma\delta}, \quad (160)$$

or, in matrix notation,

$$\bar{F}' = \bar{A}\bar{F}\tilde{A} \quad (161)$$

where \tilde{A} is the transpose of the matrix representing \bar{A} . If we pick a frame K' which is moving at velocity $\mathbf{v} = c\beta\mathbf{e}_1$, then

$$\bar{A} = \begin{pmatrix} \gamma & -\beta\gamma & 0 & 0 \\ -\beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \equiv \tilde{A}. \quad (162)$$

Given the field tensor from Eq. (145), we have

$$\bar{F}\tilde{A} = \begin{pmatrix} \beta\gamma E_x & -\gamma E_x & -E_y & -E_z \\ \gamma E_x & -\beta\gamma E_x & -B_z & B_y \\ \gamma E_y - \beta\gamma B_z & -\beta\gamma E_y + \gamma B_z & 0 & -B_x \\ \gamma E_z + \beta\gamma B_y & -\beta\gamma E_z - \gamma B_y & B_x & 0 \end{pmatrix} \quad (163)$$

and

$$\bar{F}' = \begin{pmatrix} 0 & -E_x & -\gamma E_y + \beta\gamma B_y & -\gamma E_z - \beta\gamma B_y \\ E_x & 0 & \beta\gamma E_y - \gamma B_z & \beta\gamma E_z + \gamma B_y \\ \gamma E_y - \beta\gamma B_z & -\beta\gamma E_y + \gamma B_z & 0 & -B_x \\ \gamma E_z + \beta\gamma B_y & -\beta\gamma E_z - \gamma B_y & B_x & 0 \end{pmatrix}. \quad (164)$$

This is an antisymmetric tensor - as it should be - and we can equate individual elements to the appropriate components of \mathbf{B}' and \mathbf{E}' . One finds

$$\begin{aligned} B'_x &= B_x & B'_y &= \gamma(B_y + \beta E_z) & B'_z &= \gamma(B_z - \beta E_y) \\ E'_x &= E_x & E'_y &= \gamma(E_y - \beta B_z) & E'_z &= \gamma(E_z + \beta B_y). \end{aligned} \quad (165)$$

By examining these relations for a bit, one can see that

$$\begin{aligned} \mathbf{E}'_{\parallel} &= \mathbf{E}_{\parallel} & \mathbf{E}'_{\perp} &= \gamma[\mathbf{E}_{\perp} + (\boldsymbol{\beta} \times \mathbf{B})] \\ \mathbf{B}'_{\parallel} &= \mathbf{B}_{\parallel} & \mathbf{B}'_{\perp} &= \gamma[\mathbf{B}_{\perp} - (\boldsymbol{\beta} \times \mathbf{E})] \end{aligned} \quad (166)$$

where the subscripts refer to components of the fields parallel or perpendicular to $\boldsymbol{\beta}$.

From the transformations one may see that when $\mathbf{E} \perp \mathbf{B}$, it is possible to find a frame where one of \mathbf{E}' and \mathbf{B}' (which one?) vanishes. This is achieved by picking $\boldsymbol{\beta} \perp \mathbf{E}$ and $\boldsymbol{\beta} \perp \mathbf{B}$ with an appropriate magnitude. For example, if $|\mathbf{B}| > |\mathbf{E}|$, we take

$$\boldsymbol{\beta} = \beta(\mathbf{E} \times \mathbf{B})/|\mathbf{E}||\mathbf{B}| \quad (167)$$

where β is to be such that $\mathbf{E}'_{\perp} = 0$, or

$$\begin{aligned} 0 &= \mathbf{E} + \boldsymbol{\beta} \times \mathbf{B} = \mathbf{E} + \beta[(\mathbf{E} \times \mathbf{B}) \times \mathbf{B}]/|\mathbf{E}||\mathbf{B}| \\ &= \mathbf{E} - \beta\mathbf{E}|\mathbf{B}|/|\mathbf{E}| = \mathbf{E}(1 - \beta|\mathbf{B}|/|\mathbf{E}|) \end{aligned} \quad (168)$$

so that we find

$$\beta = |\mathbf{E}|/|\mathbf{B}| \quad (169)$$

which is possible if $B > E$.

10.1 Fields Due to a Point Charge

Another example of the use of the transformations is the determination of the fields of a charge moving at constant velocity. Suppose a charge q has velocity $\mathbf{u} = \beta c\boldsymbol{\epsilon}_1$ relative to frame K . Let K' move at this velocity relative to K so that the charge is at rest in the primed frame. Further, choose the coordinates so that the charge is at $\mathbf{x}' = 0$. Then the fields in this frame are

$$\mathbf{B}'(\mathbf{x}', t') = 0 \quad \text{and} \quad \mathbf{E}'(\mathbf{x}', t') = \frac{q}{r'^3}\mathbf{x}'. \quad (170)$$

Let us restrict attention, without loss of generality, to the $z' = 0$ plane. There, the electric field is

$$\mathbf{E}' = \frac{q}{(\sqrt{x'^2 + y'^2})^3}[x'\boldsymbol{\epsilon}_1' + y'\boldsymbol{\epsilon}_2'] \quad (171)$$

Using the transformations of the electromagnetic field, we find that the nonvanishing components of the fields in frame K are

$$E_x = \frac{qx'}{(x'^2 + y'^2)^{3/2}} \quad E_y = \frac{\gamma qy'}{(x'^2 + y'^2)^{3/2}} \quad (172)$$

and

$$B_z = \frac{\gamma\beta qy'}{(x'^2 + y'^2)^{3/2}}. \quad (173)$$

In order for these expressions to be of any use, we should express the fields in terms of t and \mathbf{x} rather than the primed spacetime variables.

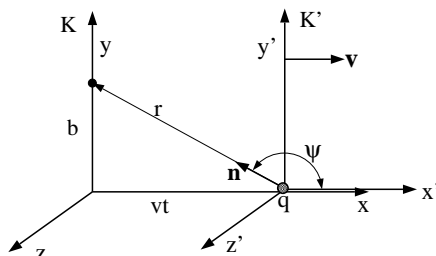


Figure 17: Charge fixed in K' detected by an observer at P in K .

We shall look in particular at the fields at the point $\mathbf{x} = b\mathbf{e}_2$. His position translates, via the Lorentz transformations, into K' as

$$y' = y \quad x' = -\gamma vt \quad \text{and} \quad z' = 0. \quad (174)$$

Using these in the expressions for \mathbf{E} and \mathbf{B} , we find

$$\begin{aligned} E_x &= -\gamma qvt/[b^2 + (\gamma vt)^2]^{3/2} \\ E_y &= \gamma qb/[b^2 + (\gamma vt)^2]^{3/2} \\ B_z &= \gamma\beta qb/[b^2 + (\gamma vt)^2]^{3/2}. \end{aligned} \quad (175)$$

It is instructive to study these results. They tell us the field at a point $(0,b,0)$ in K when a charge q goes along the x axis with speed v , passing the origin at time $t = 0$. The fields are zero at large negative times, then E_x rises and falls to zero at $t = 0$ and repeats this pattern with the opposite sign at positive times. The other two rise to a maximum value at $t = 0$ and then fall to zero at large positive time. The duration in time of the pulse is of order $b/\gamma v$ and becomes very short if $v \rightarrow c$ because then γ becomes arbitrarily large. The maximum field strengths are, for E_y , $\gamma q/b^2$, and, for B_z , $\gamma\beta q/b^2$. Notice that for a highly relativistic particle, $\beta \rightarrow 1$, $E_y \approx B_z$

and also that the maximum pulse strength scales as γ which means it becomes very large (but for a very short time) as the velocity approaches c .

Particle and Field Dynamics

Comte Joseph Louis Lagrange

(1736 - 1813)

November 9, 2001

Contents

1	Lagrangian and Hamiltonian of a Charged Particle in an External Field	2
1.1	Lagrangian of a Free Particle	4
1.1.1	Equations of Motion	5
1.2	Lagrangian of a Charged Particle in Fields	6
1.2.1	Equations of Motion	7
1.3	Hamiltonian of a Charged Particle	8
1.4	Invariant Forms	9
2	Lagrangian for the Electromagnetic Field	11
3	Stress Tensors and Conservation Laws	13
3.1	Free Field Lagrangian and Hamiltonian Densities	14
3.2	Symmetric Stress Tensor	17
3.3	Conservation Laws in the Presence of Sources	19
4	Examples of Relativistic Particle Dynamics	20

4.1	Motion in a Constant Uniform Magnetic Induction	20
4.2	Motion in crossed \mathbf{E} and \mathbf{B} fields, $E < B$	22
4.3	Motion in crossed \mathbf{E} and \mathbf{B} fields, $E > B$	24
4.4	Motion for general uniform \mathbf{E} and \mathbf{B}	26
4.5	Motion in slowly spatially varying $\mathbf{B}(\mathbf{x})$	26

In this chapter we shall study the dynamics of particles and fields. For a particle, the relativistically correct equation of motion is

$$\frac{d\mathbf{p}}{dt} = \mathbf{F} \quad (1)$$

where $\mathbf{p} = m\gamma\mathbf{u}$; the corresponding equation for the time rate of change of the particle's energy is

$$\frac{dE}{dt} = \mathbf{F} \cdot \mathbf{u}. \quad (2)$$

The dynamics of the electromagnetic field is given by the Maxwell equations,

$$\begin{aligned} \nabla \cdot \mathbf{E}(\mathbf{x}, t) &= 4\pi\rho(\mathbf{x}, t) & \nabla \cdot \mathbf{B}(\mathbf{x}, t) &= 0 \\ \nabla \times \mathbf{E}(\mathbf{x}, t) + \frac{1}{c} \frac{\partial \mathbf{B}(\mathbf{x}, t)}{\partial t} &= 0 & \nabla \times \mathbf{B}(\mathbf{x}, t) - \frac{1}{c} \frac{\partial \mathbf{E}(\mathbf{x}, t)}{\partial t} &= \frac{4\pi}{c} \mathbf{J}(\mathbf{x}, t). \end{aligned} \quad (3)$$

These are tied together by the Lorentz force which gives \mathbf{F} in terms of the electromagnetic fields

$$\mathbf{F} = q \left[\mathbf{E} + \frac{1}{c} (\mathbf{u} \times \mathbf{B}) \right] \quad (4)$$

and by the expressions for $\rho(\mathbf{x}, t)$ and $\mathbf{J}(\mathbf{x}, t)$ in terms of the particles' coordinates and velocities

$$\begin{aligned} \rho(\mathbf{x}, t) &= \sum_i q_i \delta(\mathbf{x} - \mathbf{x}_i(t)) \\ \mathbf{J}(\mathbf{x}, t) &= \sum_i q_i \mathbf{u}_i(t) \delta(\mathbf{x} - \mathbf{x}_i(t)). \end{aligned} \quad (5)$$

In view of the fact that we already know all of this, what further do we want to do? Two things: (1) Formulate appropriate covariant Lagrangians and Hamiltonians from which covariant dynamical equations can be derived; and (2) applications.

1 Lagrangian and Hamiltonian of a Charged Particle in an External Field

We want to devise a Lagrangian for a charged particle in the presence of given applied fields which are treated as parameters and not as dynamic variables. This Lagrangian

is to yield the equations of motion Eqs. (1) and (2) with \mathbf{F} given by the Lorentz force. These equations can be written as

$$\frac{dp^\alpha}{dt} = \frac{q}{mc\gamma} F^{\alpha\beta} p_\beta \quad (6)$$

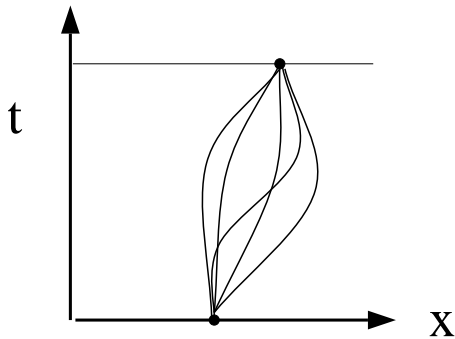
which is almost in Lorentz covariant form. A more obviously Lorentz covariant form can be obtained by using the fact that the infinitesimal time element dt can be related to an infinitesimal proper time element $d\tau$ by $dt = \gamma d\tau$. Then we have

$$\frac{dp^\alpha}{d\tau} = \frac{q}{mc} F^{\alpha\beta} p_\beta. \quad (7)$$

To get a Lagrangian from which these equations follow, we postulate the existence of the action A which may be expressed as an integral,

$$A = \int_a^b dA, \quad (8)$$

over possible “paths” from a to b . The action is an extremum for the actual motion of the system.



Possible paths which contribute to the action.

In this case, the system consists of a single particle. The paths have the constraint that they start at given (\mathbf{x}_a, t_a) and end at (\mathbf{x}_b, t_b) .

Next comes a delicate point. We could say that the first postulate of relativity requires that A be the same in all inertial frames¹, which elevates the action and its

¹This argument is (elegantly) made in *The Classical Theory of Fields*

consequences to “law of nature” status. It seems better to regard the invariance of A as an assumption or postulate in its own right and to see where that leads us.

Rewrite Eq. (8) as follows:

$$A = \int_a^b dA = \int_{t_a}^{t_b} \frac{dA}{dt} dt \equiv \int_{t_a}^{t_b} L dt. \quad (9)$$

This equation expresses nothing more than the parametrization of the integral using the time and the definition of the *Lagrangian* L as the derivative of A with respect to t . Let us further parametrize the integral using the proper time τ of the particle,

$$A = \int_{\tau_a}^{\tau_b} L \gamma d\tau \quad (10)$$

where we use $dt = \gamma d\tau$, $\gamma = 1/\sqrt{1 - u^2/c^2}$, \mathbf{u} being the particle’s velocity as measured in the lab frame or the frame in which the time t is measured. *The proper time is an invariant, so if we believe that A is one also, we have to conclude that $L\gamma$ is an invariant.* This statement of invariance greatly limits the possible forms of L .

1.1 Lagrangian of a Free Particle

Consider first the case of a free particle. What invariants may we construct from the properties of a free particle? We have only the four-vectors \bar{p} and \bar{x} . The presumed translational invariance of space rules out the use of the latter. That leaves only the four-momentum and the single invariant $p^\alpha p_\alpha = m^2 c^2$ which is a constant. Hence we are led to $L\gamma = C$ where C is a constant. Hence, $L = C/\gamma$ and

$$A = C \int_{\tau_a}^{\tau_b} d\tau = C \int_{t_a}^{t_b} dt \sqrt{1 - u^2/c^2}. \quad (11)$$

We may find the constant C by appealing to the nonrelativistic limit and expanding in powers of u^2/c^2 .

$$A \approx C \int_{t_a}^{t_b} dt \left(1 - \frac{u^2}{2c^2} + \dots \right). \quad (12)$$

The term proportional to u^2 should be the usual nonrelativistic Lagrangian of a free particle, $mu^2/2$. This condition leads to

$$C = -mc^2 \quad (13)$$

and so

$$L_f = -mc^2 \sqrt{1 - u^2/c^2} \quad (14)$$

is the free-particle Lagrangian.

1.1.1 Equations of Motion

The equations of motion are found by requiring that A be an extremum,

$$\delta A = \delta \left(\int_{t_a}^{t_b} dt L \right) = 0. \quad (15)$$

The path $\mathbf{x}(t)$ is to be fixed at the end points t_a and t_b , $\delta\mathbf{x}(t_a) = \delta\mathbf{x}(t_b) = 0$. Writing L as a function of the Cartesian components of the position and velocity, we have, allowing for possible position-dependence which will appear if the particle is not free, $L = L(x_i, u_i, t)$, and

$$\delta A = \sum_i \int_{t_a}^{t_b} dt \left[\left(\frac{\partial L}{\partial x_i} \right) \delta x_i + \left(\frac{\partial L}{\partial u_i} \right) \delta u_i \right]. \quad (16)$$

But δu_i is related to δx_i through $u_i = dx_i/dt$, so

$$\begin{aligned} \delta A &= \sum_i \int_{t_a}^{t_b} dt \left[\left(\frac{\partial L}{\partial x_i} \right) \delta x_i + \left(\frac{\partial L}{\partial u_i} \right) \delta \left(\frac{dx_i}{dt} \right) \right] \\ &= \left(\frac{\partial L}{\partial u_i} \right) \delta x_i \Big|_{t_a}^{t_b} + \int_{t_a}^{t_b} dt \left[\frac{\partial L}{\partial x_i} - \frac{d}{dt} \left(\frac{\partial L}{\partial u_i} \right) \right] \delta x_i(t) \end{aligned} \quad (17)$$

where we have integrated by parts to achieve the last step. The first term in the final expression vanishes because $\delta x_i = 0$ at the endpoints of the interval of integration. Arguing that $\delta x_i(t)$ is arbitrary elsewhere, we conclude that the factor [...] in the final expression must vanish everywhere,

$$\frac{d}{dt} \left(\frac{\partial L}{\partial u_i} \right) - \frac{\partial L}{\partial x_i} = 0 \quad (18)$$

for each $i = 1, 2, 3$.

These are the *Euler-Lagrange* equations of motion. Let's apply them to the free-particle Lagrangian L_f ,

$$\frac{\partial L_f}{\partial x_i} = 0 \quad \text{and} \quad \frac{\partial L_f}{\partial u_i} = m\gamma u_i, \quad (19)$$

so

$$\frac{d}{dt}(m\gamma \mathbf{u}) = 0 \quad (20)$$

is the equation of motion. It is the same as

$$\frac{d\mathbf{p}}{dt} = 0 \quad (21)$$

which is correct for a free particle.

1.2 Lagrangian of a Charged Particle in Fields

Next suppose that there are electric and magnetic fields of roughly the same order of magnitude present so that the particle experiences some force and acceleration. Then $L = L_f + L_{int}$ where L_{int} is the “interaction” Lagrangian and contains the information about the fields and forces. For the action to be an invariant, it must be the case that

$$A_{int} \equiv \int_{t_a}^{t_b} L_{int} dt \quad (22)$$

is an invariant which means $L_{int}\gamma$ has to be an invariant. Now, in the nonrelativistic limit one has, to lowest order, $L = T - V$ with $V = q\Phi$, so we have in this limit $L_{int}\gamma = -q\Phi\gamma = -qE\Phi/mc^2 = -(q/mc)p_0A^0$. This is not an invariant but can be made one by including the rest of $\bar{p} \cdot \bar{A}^2$, and we expect that the result is valid not just in the nonrelativistic limit but in general:

$$L_{int}\gamma = -\left(\frac{q}{mc}\right)p_\alpha A^\alpha. \quad (23)$$

²We have little choice other than this form since we only have the p , x and A four-vectors to work with

This choice of L_{int} gives the desired invariant and reduces to the correct static limit. It is the simplest choice of the interaction Lagrangian with the following properties:

1. Translationally invariant (in the sense that it is independent of explicit dependence on \mathbf{x} ; the potentials do depend on \mathbf{x})
2. Linear in the charge (as are the forces on the particle)
3. Linear in the momenta (as are the forces)
4. Linear in the fields (as are the equations of motion of the particle)
5. A function of no time derivatives of p^α (appropriate for the equations of motion)

1.2.1 Equations of Motion

Let us proceed to the Euler-Lagrange equations of motion. The total Lagrangian is

$$L = -mc^2 \sqrt{1 - u^2/c^2} + \frac{q}{c} \mathbf{u} \cdot \mathbf{A} - q\Phi; \quad (24)$$

$$\begin{aligned} \frac{\partial L}{\partial x_i} &= \frac{q}{c} \mathbf{u} \cdot \frac{\partial \mathbf{A}}{\partial x_i} - q \frac{\partial \Phi}{\partial x_i} \\ \frac{\partial L}{\partial u_i} &= \frac{mc^2}{\sqrt{1 - u^2/c^2}} \frac{u_i}{c^2} + \frac{q}{c} A_i, \end{aligned} \quad (25)$$

so

$$\frac{d}{dt} \left(\frac{\partial L}{\partial u_i} \right) = \frac{d}{dt} (m\gamma u_i) + \frac{q}{c} \frac{\partial A_i}{\partial t} + \frac{q}{c} (\mathbf{u} \cdot \nabla) A_i. \quad (26)$$

Notice the last term on the right-hand side of this equation. It is there because when we take the total time derivative, we must remember that the position variable \mathbf{x} on which \mathbf{A} depends is really the position of the particle at time t , so, by application of the chain rule, we pick up a sum of terms, each of which is the derivative of \mathbf{A} with respect to a component of \mathbf{x} times the derivative of that component of \mathbf{x} with respect to t ; the last is a component of the velocity of the particle.

Finally, the equations of motion are

$$\begin{aligned} \frac{d}{dt}(m\gamma u_i) &= -\frac{q}{c} \frac{\partial A_i}{\partial t} - \frac{q}{c} (\mathbf{u} \cdot \nabla) A_i - q \frac{\partial \Phi}{\partial x_i} + \frac{q}{c} \mathbf{u} \cdot \frac{\partial \mathbf{A}}{\partial x_i} \\ &= qE_i + \frac{q}{c} \mathbf{u} \cdot \frac{\partial \mathbf{A}}{\partial x_i} - \frac{q}{c} (\mathbf{u} \cdot \nabla) A_i. \end{aligned} \quad (27)$$

These are supposed to be familiar; consider

$$(\mathbf{u} \times \mathbf{B})_i = [\mathbf{u} \times (\nabla \times \mathbf{A})]_i = [\nabla(\mathbf{u} \cdot \mathbf{A}) - (\mathbf{u} \cdot \nabla)\mathbf{A}]_i = \mathbf{u} \cdot \frac{\partial \mathbf{A}}{\partial x_i} - (\mathbf{u} \cdot \nabla) A_i. \quad (28)$$

Comparison of this expansion with Eq. (27) demonstrates that the latter can be written as

$$\frac{dp_i}{dt} = qE_i + \frac{q}{c} (\mathbf{u} \times \mathbf{B})_i. \quad (29)$$

1.3 Hamiltonian of a Charged Particle

One can also make a Hamiltonian description of the system. Introduce the canonical three-momentum $\boldsymbol{\pi}$ with components

$$\pi_i \equiv \frac{\partial L}{\partial u_i} = \gamma m u_i + \frac{q}{c} A_i = p_i + \frac{q}{c} A_i. \quad (30)$$

Then the Hamiltonian³ is

$$H = \boldsymbol{\pi} \cdot \mathbf{u} - L = \boldsymbol{\pi} \cdot \mathbf{u} + mc^2 \sqrt{1 - u^2/c^2} + q\Phi - \frac{q}{c} (\mathbf{u} \cdot \mathbf{A}). \quad (31)$$

We want H to depend on \mathbf{x} and $\boldsymbol{\pi}$ but not on \mathbf{u} . To this end consider how to write \mathbf{u} in terms of $\boldsymbol{\pi}$,

$$\boldsymbol{\pi} = \frac{m\mathbf{u}}{\sqrt{1 - u^2/c^2}} + \frac{q}{c} \mathbf{A}, \quad (32)$$

or

$$\left(\boldsymbol{\pi} - \frac{q}{c} \mathbf{A} \right)^2 \left(1 - \frac{u^2}{c^2} \right) = m^2 u^2 \quad (33)$$

³The hamiltonian $H(q, p)$ obtained from the Lagrangian through a Legendre transformation $H = \sum_i p_i \dot{q}_i - L(q, \dot{q})$

which may be solved for \mathbf{u} to give

$$\mathbf{u} = c \frac{c\boldsymbol{\pi} - q\mathbf{A}}{\sqrt{m^2c^4 + (c\boldsymbol{\pi} - q\mathbf{A})^2}} \quad (34)$$

Use of this result in the expression for the Hamiltonian leads to

$$H = \sqrt{m^2c^4 + (c\boldsymbol{\pi} - q\mathbf{A})^2} + q\Phi. \quad (35)$$

The development of Hamilton's equations will be left as an exercise.

The Hamiltonian is the 0th component of a four-vector. Notice, from Eq. (35), that

$$(H - q\Phi)^2 - (c\boldsymbol{\pi} - q\mathbf{A})^2 = m^2c^4, \quad (36)$$

is an invariant. This invariant is the inner product of a four-vector with itself. The spacelike components are $c\boldsymbol{\pi} - q\mathbf{A} = c\mathbf{p}$, and the timelike component is $H - q\Phi$. The vector is just the energy-momentum four-vector in the presence of fields,

$$p^\alpha = (E/c, \mathbf{p}) = \left(\frac{1}{c}(H - q\Phi), \boldsymbol{\pi} - \frac{e}{c}\mathbf{A} \right). \quad (37)$$

1.4 Invariant Forms

Next we are going to repeat everything that we have just done, but in a manner that is “manifestly” covariant. That is, we want to rewrite the Lagrangian in terms of invariant 4-vector products. We can write the free-particle Lagrangian as

$$L_f = -mc^2\sqrt{1 - u^2/c^2} = -\frac{1}{\gamma}\sqrt{E^2 - p^2c^2} = -\frac{c}{\gamma}\sqrt{p_\alpha p^\alpha} = -\frac{mc}{\gamma}\sqrt{U_\alpha U^\alpha} \quad (38)$$

where

$$U^\alpha = (E/mc, \mathbf{p}/m) \equiv dx^\alpha/d\tau \quad (39)$$

is a four-vector we shall call the *four-velocity*. The action of the free particle is

$$A = -\int_{t_a}^{t_b} dt \gamma^{-1} mc \sqrt{U^\alpha U_\alpha} = -mc \int_{\tau_a}^{\tau_b} d\tau \sqrt{U^\alpha U_\alpha} \quad (40)$$

Note the manifestly invariant form. However, note that we must also impose the constraint $U^\alpha U_\alpha = c^2$. Thus we may not freely vary this action to find the equations of motion. There are two ways to overcome this. First, we could introduce a Lagrange multiplier to impose the constraint⁴, or we could introduce an additional degree of freedom into our equations, and use it to impose the constraint *a posteriori*. Following Jackson, we will follow the later (less conventional) route. To this end we rewrite the action, and introduce s .

$$A = -mc \int_{\tau_a}^{\tau_b} \sqrt{dx_\alpha dx^\alpha} = -mc \int_{s_a}^{s_b} ds \sqrt{g_{\alpha\beta} \frac{dx^\alpha}{ds} \frac{dx^\beta}{ds}} \quad (41)$$

where the path of integration has been parametrized using some (invariant) s . We shall now **treat** each dx^α/ds as an independent generalized velocity, and the Lagrangian takes on the functional form $L(x^\alpha, dx^\alpha/ds, s)$. This (more general) parametrization of the action integral is just as good as the standard one using the time; the Euler-Lagrange equations of motion, found by demanding that A be an extremum, are familiar in appearance,

$$\frac{d}{ds} \left(\frac{\partial \tilde{L}}{\partial(dx_\alpha/ds)} \right) - \frac{\partial \tilde{L}}{\partial x_\alpha} = 0 \quad (42)$$

In this case, we obtain the equation of motion

$$mc \frac{d}{ds} \left[\frac{dx^\alpha/ds}{\sqrt{\frac{dx^\beta}{ds} \frac{dx_\beta}{ds}}} \right] = 0 \quad (43)$$

These velocities are constrained by the condition

$$\sqrt{g_{\alpha\beta} \frac{dx^\alpha}{ds} \frac{dx^\beta}{ds}} ds = cd\tau \quad (44)$$

because there are really only three independent generalized velocities, so that

$$m \frac{d^2 x^\alpha}{d\tau^2} = 0 \quad (45)$$

⁴This approach is discussed in **Electrodynamics and Classical Theory of Fields and Particles** by A.O. Barut, Dover, page 65

Analyzing and including the interaction Lagrangian in the same manner leads to a total Lagrangian and an action which is

$$A = - \int_{s_a}^{s_b} ds \left[mc \sqrt{g^{\alpha\beta} \frac{dx_\alpha}{ds} \frac{dx_\beta}{ds}} + \frac{e}{c} \frac{dx_\alpha}{ds} A^\alpha(x) \right] \equiv - \int_{s_a}^{s_b} ds \tilde{L}. \quad (46)$$

The equation of motion may be found in the same manner and in the present application these turn out to be

$$m \frac{d^2 x^\alpha}{d\tau^2} = \frac{e}{c} (\partial^\alpha A^\beta - \partial^\beta A^\alpha) \frac{dx_\beta}{d\tau}, \quad (47)$$

and they are correct, as one may show by comparing them with the standard forms. The corresponding canonical momenta are

$$\pi^\alpha = \frac{\partial \tilde{L}}{\partial (dx_\alpha/ds)} = mU^\alpha + \frac{e}{c} A^\alpha. \quad (48)$$

Hence the Hamiltonian is

$$\tilde{H} = \pi_\alpha U^\alpha - \tilde{L} = \frac{1}{2m} \left(\pi_\alpha - \frac{eA_\alpha}{c} \right) \left(\pi^\alpha - \frac{eA^\alpha}{c} \right) - \frac{1}{2} mc^2. \quad (49)$$

Hamilton's equations of motion⁵ are

$$\begin{aligned} \frac{dx^\alpha}{d\tau} &= \frac{\partial \tilde{H}}{\partial \pi_\alpha} = \frac{1}{m} \left(\pi^\alpha - \frac{e}{c} A^\alpha \right) \\ \frac{d\pi^\alpha}{d\tau} &= - \frac{\partial \tilde{H}}{\partial x_\alpha} = \frac{e}{mc} \left(\pi_\beta - \frac{eA_\beta}{c} \right) \partial^\alpha A^\beta \end{aligned} \quad (50)$$

2 Lagrangian for the Electromagnetic Field

The electromagnetic field and fields in general have continuous degrees of freedom. The analog of a generalized coordinate q_i is the value of a field ϕ_k at a point \bar{x} . There are an infinite number of such points and so we have an infinite number of generalized

⁵For $H(p, q)$ Hamilton's equations are $\dot{q}_i = \frac{\partial H}{\partial p_i}$, $\dot{p}_i = -\frac{\partial H}{\partial q_i}$, and $\frac{\partial L}{\partial t} = -\frac{\partial H}{\partial t}$

coordinates. The corresponding “generalized velocities” are derivatives of the field with respect to the variables, $\partial\phi_k(\bar{x})/\partial x^\alpha$ or $\partial\phi_k(\bar{x})\partial x_\alpha$ with $\alpha = 0, 1, 2, 3$.

$$\begin{aligned} q_i &\rightarrow \phi_k(\bar{x}) \\ \dot{q}_i &\rightarrow \frac{\partial\phi_k(\bar{x})}{\partial x^\alpha} \end{aligned} \tag{51}$$

Instead of a Lagrangian L which depends on the coordinates and velocities q_i and \dot{q}_i , one now has a *Lagrangian density* \mathcal{L} , and the Lagrangian is obtained by integrating this density over position space,

$$L = \int d^3x \mathcal{L}(\phi_k(\bar{x}), \partial^\alpha\phi_k(\bar{x})); \tag{52}$$

The action is the integral of this over time, or

$$A = \int d^4x \mathcal{L}(\phi_k(\bar{x}), \partial^\alpha\phi_k(\bar{x})). \tag{53}$$

Given that A and d^4x are invariants, \mathcal{L} must also be an invariant.

The Euler-Lagrange equations of motion are obtained as usual by demanding that A be an extremum with respect to variation of the fields, or

$$\delta A/\delta\phi_k(\bar{x}) = 0 \tag{54}$$

for each field ϕ_k . The resulting equations are, explicitly,

$$\partial^\beta \left(\frac{\partial\mathcal{L}}{\partial(\partial^\beta\phi_k)} \right) - \frac{\partial\mathcal{L}}{\partial\phi_k} = 0. \tag{55}$$

Now let’s turn to the question of an appropriate Lagrangian density for the electromagnetic field. The things we have to work with are $F^{\alpha\beta}$, A^α , and J^α , if we rule out explicit dependence on space and time (a translationally invariant universe). We must make an invariant out of these. One which practically suggests itself is

$$\mathcal{L} = -\frac{1}{16\pi}F_{\alpha\beta}F^{\alpha\beta} - \frac{1}{c}J_\alpha A^\alpha. \tag{56}$$

The various constants are a matter of definition; otherwise we have something which is linear in components of \bar{A} and of \bar{J} , and bilinear derivatives of components of \bar{A} . Let's write it out in detail:

$$\begin{aligned}\mathcal{L} &= -\frac{1}{16\pi}(\partial_\alpha A_\beta - \partial_\beta A_\alpha)(\partial^\alpha A^\beta - \partial^\beta A^\alpha) - \frac{1}{c}J_\alpha A^\alpha \\ &= -\frac{1}{16\pi}g_{\alpha\gamma}g_{\beta\delta}(\partial^\gamma A^\delta - \partial^\delta A^\gamma)(\partial^\alpha A^\beta - \partial^\beta A^\alpha) - \frac{1}{c}J_\alpha A^\alpha\end{aligned}\quad (57)$$

The generalized fields (called ϕ_k above) are the components of \bar{A} . Hence the functional derivatives of \mathcal{L} which enter the Euler-Lagrange equations are

$$\frac{\partial\mathcal{L}}{\partial(\partial^\beta A^\alpha)} = \frac{1}{4\pi}F_{\alpha\beta} \quad \text{and} \quad \frac{\partial\mathcal{L}}{\partial A^\alpha} = -\frac{1}{c}J_\alpha \quad (58)$$

and so the equations of motion are

$$\frac{1}{4\pi}\partial^\beta F_{\beta\alpha} = \frac{1}{c}J_\alpha. \quad (59)$$

These are indeed the four⁶ inhomogeneous Maxwell equations. The homogeneous equations are automatically satisfied because we have constructed the Lagrangian in terms of the potentials. The charge continuity equation follows from taking the contravariant derivative of the equation above,

$$\frac{1}{4\pi}\partial^\alpha\partial^\beta F_{\beta\alpha} = \frac{1}{c}\partial^\alpha J_\alpha; \quad (60)$$

the left-hand side is zero when summed because $F_{\alpha\beta} = -F_{\beta\alpha}$ and so we have

$$\partial^\alpha J_\alpha = 0. \quad (61)$$

3 Stress Tensors and Conservation Laws

Conservation of energy emerges from the usual Lagrangian formulation if L has no explicit dependence on the time; then $dH/dt = 0$ which means that the Hamiltonian

⁶Four equations come from the one scalar and one vector inhomogeneous Maxwell's equations

is a constant of the motion. If we want to carry this sort of thing over to our field theory, we need to construct a *Hamiltonian density* \mathcal{H} whose integral over all position space, H , is interpreted as the energy. If one proceeds in analogy with the particle case, he would take a Lagrangian density

$$\mathcal{L} = \mathcal{L}(\phi_k(\bar{x}), \partial^\alpha \phi_k(\bar{x})); \quad (62)$$

introduce momentum fields

$$\Pi_k(\bar{x}) \equiv \partial \mathcal{L} / \partial (\partial \phi_k / \partial t); \quad (63)$$

and a Hamiltonian density

$$\mathcal{H} = \sum_k \Pi_k(\bar{x}) (\partial \phi_k(\bar{x}) / \partial t) - \mathcal{L}. \quad (64)$$

We are going to generalize this procedure by introducing a rank-two tensor instead of a simple Hamiltonian density. The reason is that if one has a simple density and introduces H as

$$H = \int d^3x \mathcal{H} = \int dx_0 \frac{d^3x}{dx_0} \mathcal{H}, \quad (65)$$

and if one wants this to be an energy, which, as we have seen, transforms as the 0^{th} component of a four-vector, then \mathcal{H} should be the (0,0) component of a rank-two tensor. To this end, let us introduce

$$\psi_k^\alpha(\bar{x}) \equiv \partial \mathcal{L} / \partial (\partial_\alpha \phi_k) \quad (66)$$

and

$$T^{\alpha\beta} \equiv \sum_k \psi_k^\alpha(\bar{x}) \partial^\beta \phi_k - g^{\alpha\beta} \mathcal{L}. \quad (67)$$

This rank-two tensor is called the *canonical stress tensor*.

3.1 Free Field Lagrangian and Hamiltonian Densities

Let's see what form the Lagrangian density and canonical stress tensor take in the absence of any sources \mathbf{J}^α . In this case the Lagrangian density becomes \mathcal{L}_{ff} , the

free-field Lagrangian density.

$$\mathcal{L}_{ff} = -\frac{1}{16\pi} F_{\alpha\beta} F^{\alpha\beta} \quad (68)$$

By carrying out the implied manipulations we find

$$T^{\alpha\beta} = -\frac{1}{4\pi} g^{\alpha\gamma} F_{\gamma\delta} \partial^\beta A^\delta - g^{\alpha\beta} \mathcal{L}_{ff}. \quad (69)$$

Look in particular at T^{00} :

$$\begin{aligned} T^{00} &= -\frac{1}{4\pi} (F_{0\gamma} \partial^0 A^\gamma) - \frac{1}{8\pi} (E^2 - B^2) \\ &= -\frac{1}{4\pi} \left[E_x \frac{1}{c} \frac{\partial A_x}{\partial t} + E_y \frac{1}{c} \frac{\partial A_y}{\partial t} + E_z \frac{1}{c} \frac{\partial A_z}{\partial t} \right] - \frac{1}{8\pi} (E^2 - B^2) \\ &= \frac{1}{4\pi} E^2 + \mathbf{E} \cdot \nabla \Phi - \frac{1}{8\pi} (E^2 - B^2) = \frac{1}{8\pi} (E^2 + B^2) + \frac{\mathbf{E} \cdot \nabla \Phi}{4\pi}. \end{aligned} \quad (70)$$

This contains the expected and desired term $(E^2 + B^2)/8\pi$, which is the feild energy density, but there is an additional term $\mathbf{E} \cdot \nabla \Phi$. Because $\nabla \cdot \mathbf{E} = 0$ for free fields, it is the case that $\mathbf{E} \cdot \nabla \Phi = \nabla \cdot (\mathbf{E}\Phi)$ and so the integral over all space of this part of T^{00} will vanish for a localized field distribution. Hence we find that

$$\int d^3x T^{00} = \frac{1}{8\pi} \int d^3x (E^2 + B^2) \quad (71)$$

is indeed the field energy.

And what of the other components of the stress tensor? These too have some unexpected properties. For example, one can show that

$$T^{0i} = \frac{1}{4\pi} (\mathbf{E} \times \mathbf{B})_i + \frac{1}{4\pi} \nabla \cdot (A_i \mathbf{E}) \quad (72)$$

and

$$T^{i0} = \frac{1}{4\pi} (\mathbf{E} \times \mathbf{B})_i + \frac{1}{4\pi} \left[(\nabla \times (\Phi \mathbf{B}))_i - \frac{\partial}{\partial x_0} (\Phi E_i) \right]. \quad (73)$$

Evidently, this tensor is not symmetric. Also, one would have hoped that these components of the tensor would have turned out to be components of the Poynting

vector, with appropriate scaling, so that we would have found an equation $0 = \partial_\alpha T^{0\alpha}$ which would have been equivalent to the Poynting theorem,

$$\frac{\partial u}{\partial t} + \nabla \cdot \mathbf{S} = 0. \quad (74)$$

Although this is not going to happen, there is some sort of conservation law contained in our stress tensor. One can show that

$$\partial_\alpha T^{\alpha\beta} = 0 \quad (75)$$

which gives not one but four conservation laws. To demonstrate this equation, consider the following:

$$\begin{aligned} \partial_\alpha T^{\alpha\beta} &= \sum_k \partial_\alpha \left[\frac{\partial \mathcal{L}}{\partial(\partial_\alpha \phi_k)} \partial^\beta \phi_k \right] - \partial^\beta \mathcal{L} \\ &= \sum_k \left[\partial_\alpha \left(\frac{\partial \mathcal{L}}{\partial(\partial_\alpha \phi_k)} \right) \partial^\beta \phi_k + \left(\frac{\partial \mathcal{L}}{\partial(\partial_\alpha \phi_k)} \right) \partial^\beta (\partial_\alpha \phi_k) \right] - \partial^\beta \mathcal{L} \\ &= \sum_k \left[\left(\frac{\partial \mathcal{L}}{\partial \phi_k} \right) \partial^\beta \phi_k + \left(\frac{\partial \mathcal{L}}{\partial(\partial_\alpha \phi_k)} \right) \partial^\beta (\partial_\alpha \phi_k) \right] - \partial^\beta \mathcal{L} \end{aligned} \quad (76)$$

where we have used the Euler-Lagrange equations of motion Eq. (55) on the first term in the middle line. Now we can recognize that the terms summed over k in the last line are $\partial^\beta \mathcal{L}$ since \mathcal{L} is a function of the fields ϕ_k and their derivatives $\partial_\alpha \phi_k$. Hence we have demonstrated that

$$\partial_\alpha T^{\alpha\beta} = \partial^\beta \mathcal{L} - \partial^\beta \mathcal{L} = 0 \quad (77)$$

These give familiar global conservation laws when integrated over all space for a localized set of fields. Consider

$$0 = \int d^3x \partial_\alpha T^{\alpha\beta} = \frac{\partial}{\partial x^0} \left(\int d^3x T^{0\beta} \right) + \int d^3x \frac{\partial}{\partial x^i} (T^{i\beta}). \quad (78)$$

The last term on the right-hand side is zero as one shows by integrating over that coordinate with respect to which the derivative is taken and appealing to the fact that

we have localized fields which vanish as $|x^i|$ becomes large. Hence our conclusion is that

$$\frac{d}{dt} \left(\int d^3x T^{0\beta} \right) = 0 \quad (79)$$

If one looks at the explicit components of the tensor, one finds that these simply say the total energy and total momentum are constant, using our identifications (from chapter 6) of u and \mathbf{g} as the energy and momentum density.

$$u = \frac{1}{8\pi} (E^2 + B^2) \quad \mathbf{g} = \frac{1}{4\pi c} (\mathbf{E} \times \mathbf{B}) \quad (80)$$

3.2 Symmetric Stress Tensor

It is troubling that the canonical stress tensor is not symmetric. This becomes a serious problem when one examines the angular momentum. Consider the rank-three tensor

$$M^{\alpha\beta\gamma} \equiv T^{\alpha\beta} x^\gamma - T^{\alpha\gamma} x^\beta. \quad (81)$$

If this is to represent the angular momentum in some way we would like it to provide a conservation law in the form $\partial_\alpha M^{\alpha\beta\gamma} = 0$. But that doesn't happen. Rather,

$$\partial_\alpha M^{\alpha\beta\gamma} = T^{\gamma\beta} - T^{\beta\gamma} + (\partial_\alpha T^{\alpha\beta}) x^\gamma - (\partial_\alpha T^{\alpha\gamma}) x^\beta = T^{\gamma\beta} - T^{\beta\gamma} \quad (82)$$

which doesn't vanish because \bar{T} is not symmetric.

The standard way out of this and other difficulties associated with the asymmetry of the canonical stress tensor is to define a different stress tensor which works. By regrouping terms in the canonical stress tensor one can write

$$T^{\alpha\beta} = \frac{1}{4\pi} \left(g^{\alpha\gamma} F_{\gamma\delta} F^{\delta\beta} + \frac{1}{4} g^{\alpha\beta} F_{\gamma\delta} F^{\gamma\delta} \right) - \frac{1}{4\pi} g^{\alpha\gamma} F_{\gamma\delta} \partial^\delta A^\beta. \quad (83)$$

Now, the second term is

$$\begin{aligned} -\frac{1}{4\pi} g^{\alpha\gamma} F_{\gamma\delta} \partial^\delta A^\beta &= -\frac{1}{4\pi} F^{\alpha\delta} \partial_\delta A^\beta = \frac{1}{4\pi} F^{\delta\alpha} \partial_\delta A^\beta = \\ &= \frac{1}{4\pi} (F^{\delta\alpha} \partial_\delta A^\beta + A^\beta \partial_\delta F^{\delta\alpha}) = \frac{1}{4\pi} \partial_\delta (F^{\delta\alpha} A^\beta). \end{aligned} \quad (84)$$

This is a four-divergence, so for fields of finite extent, it must be the case that

$$\int d^3x \partial_\delta (F^{\delta 0} A^\beta) = 0. \quad (85)$$

Moreover, it has a vanishing four-divergence,

$$\partial_\alpha \partial_\delta (F^{\delta\alpha} A^\beta) = 0, \quad (86)$$

which follows from the antisymmetric character of the field tensor. Hence, if we simply remove this piece from the stress tensor, leaving a new tensor $\bar{\theta}$, known as the *symmetric stress tensor*,

$$\theta^{\alpha\beta} \equiv \frac{1}{4\pi} (g^{\alpha\gamma} F_{\gamma\delta} F^{\delta\beta} + \frac{1}{4} g^{\alpha\beta} F_{\gamma\delta} F^{\gamma\delta}), \quad (87)$$

then this tensor is such that

$$\frac{d}{dt} \left(\int d^3x \theta^{0\beta} \right) = 0 \quad \text{and} \quad \partial_\alpha \theta^{\alpha\beta} = 0. \quad (88)$$

It is easy to work out the components of this tensor; they are ($i, j = 1, 2, 3$)

$$\begin{aligned} \theta^{00} &= \frac{1}{8\pi} (E^2 + B^2) \\ \theta^{i0} &= \theta^{0i} = \frac{1}{4\pi} (\mathbf{E} \times \mathbf{B}) \\ \theta^{ij} &= -\frac{1}{4\pi} [E_i E_j + B_i B_j - \frac{1}{2} \delta_{ij} (E^2 + B^2)]. \end{aligned} \quad (89)$$

Hence in block matrix form,

$$\theta^{\alpha\beta} = \begin{pmatrix} u & c\mathbf{g} \\ c\mathbf{g} & -T_{ij}^{(M)} \end{pmatrix} \quad (90)$$

where $T_{ij}^{(M)}$ is the ij component of the Maxwell stress tensor. The conservation laws

$$\partial_\alpha \theta^{\alpha\beta} = 0 \quad (91)$$

are well-known to us. They are the Poynting theorem, for $\beta = 0$, and the momentum conservation laws

$$\frac{\partial g_i}{\partial t} - \sum_j \frac{\partial T_{ij}^{(M)}}{\partial x_j} = 0. \quad (92)$$

when $\beta = i$

Now consider once again the question of angular momentum. Define

$$M^{\alpha\beta\gamma} \equiv \theta^{\alpha\beta} x^\gamma - \theta^{\alpha\gamma} x^\beta. \quad (93)$$

Then the equations

$$\partial_\alpha M^{\alpha\beta\gamma} = 0 \quad (94)$$

express angular momentum conservation as well as some other things.

3.3 Conservation Laws in the Presence of Sources

Finally, what happens if there are sources? Then we won't find the same form for the conservation laws. Consider

$$\begin{aligned} \partial_\alpha \theta^{\alpha\beta} &= \frac{1}{4\pi} \left[\partial^\gamma (F_{\gamma\delta} F^{\delta\beta}) + \frac{1}{4} \partial^\beta (F_{\gamma\delta} F^{\gamma\delta}) \right] \\ &= \frac{1}{4\pi} \left[(\partial^\gamma F_{\gamma\delta} F^{\delta\beta} + F_{\gamma\delta} (\partial^\gamma F^{\delta\beta})) + \frac{1}{2} F_{\gamma\delta} (\partial^\beta F^{\gamma\delta}) \right]. \end{aligned} \quad (95)$$

Making use of the Maxwell equations $\partial^\gamma F_{\gamma\delta} = \frac{4\pi}{c} J_\delta$, we can rewrite this as

$$\partial_\alpha \theta^{\alpha\beta} + \frac{1}{c} F^{\beta\delta} J_\delta = \frac{1}{8\pi} \left[F_{\gamma\delta} (\partial^\gamma F^{\delta\beta} + \partial^\gamma F^{\delta\beta} + \partial^\beta F^{\gamma\delta}) \right]. \quad (96)$$

Now recall that (these are the homogeneous Maxwell's equations)

$$\partial^\gamma F^{\delta\beta} + \partial^\beta F^{\gamma\delta} + \partial^\delta F^{\beta\gamma} = 0, \quad (97)$$

so Eq. (90) may be written as

$$\partial_\alpha \theta^{\alpha\beta} + \frac{1}{c} F^{\beta\delta} J_\delta = \frac{1}{8\pi} F_{\gamma\delta} (\partial^\gamma F^{\delta\beta} - \partial^\delta F^{\beta\gamma}). \quad (98)$$

However,

$$(\partial^\gamma F^{\delta\beta} - \partial^\delta F^{\beta\gamma}) F_{\gamma\delta} = (\partial^\gamma F^{\delta\beta} + \partial^\delta F^{\gamma\beta}) F_{\gamma\delta} \quad (99)$$

is a contraction of an object symmetric in the indices γ and δ and one which is antisymmetric; therefore it is zero. Hence we conclude that

$$\partial_\alpha \theta^{\alpha\beta} = -\frac{1}{c} F^{\beta\delta} J_\delta. \quad (100)$$

The four equations contained in this conservation law are the familiar ones

$$\frac{\partial u}{\partial t} + \nabla \cdot \mathbf{S} = -\mathbf{J} \cdot \mathbf{E} \quad \text{when } \beta = 0 \quad (101)$$

and

$$\frac{\partial g_i}{\partial t} - \sum_j \frac{\partial}{\partial x^j} T_{ij}^{(M)} = -[\rho E_i + \frac{1}{c}(\mathbf{J} \times \mathbf{B})_i] \quad \text{when } \beta = i. \quad (102)$$

4 Examples of Relativistic Particle Dynamics

4.1 Motion in a Constant Uniform Magnetic Induction

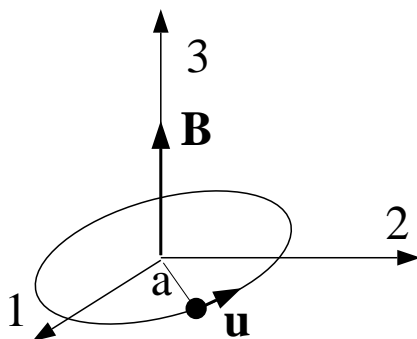
Given an applied constant magnetic induction, the equations of motion for a particle of charge q are

$$\frac{dE}{dt} = \mathbf{F} \cdot \mathbf{u} = 0, \quad \frac{d\mathbf{p}}{dt} = \frac{q}{c}(\mathbf{u} \times \mathbf{B}) = m\gamma \frac{d\mathbf{u}}{dt} \quad (103)$$

where the last step follows from the fact that $\mathbf{p} = m\gamma\mathbf{u}$ and the fact that γmc^2 , the particle's energy, is constant because magnetic forces do no work. Hence the equations reduce to

$$\frac{d\mathbf{u}}{dt} = \mathbf{u} \times \omega_B \quad (104)$$

where $\omega_B = q\mathbf{B}/m\gamma c$. Notice that this frequency depends on the energy of the particle. For definiteness, let $\mathbf{B} = B\epsilon_3$. Also, write $\mathbf{u} = u_{\parallel}\epsilon_3 + \mathbf{u}_{\perp}$ where $\mathbf{u}_{\perp} \cdot \epsilon_3 = 0$.



From the equations of motion, one can see that u_{\parallel} is a constant while \mathbf{u}_{\perp} obeys

$$\frac{d\mathbf{u}_{\perp}}{dt} = \omega_B(\mathbf{u}_{\perp} \times \epsilon_3), \quad (105)$$

or

$$\frac{du_x}{dt} = \omega_B u_y \quad \text{and} \quad \frac{du_y}{dt} = -\omega_B u_x. \quad (106)$$

Combining these we find, e.g.,

$$\frac{d^2 u_x}{dt^2} = -\omega_B^2 u_x \quad (107)$$

with the general solution

$$u_x = u_0 e^{-i\omega_B t} \quad (108)$$

where u_0 is a complex constant. Further,

$$u_y = \frac{1}{\omega_B} \frac{du_x}{dt} = -i u_x, \quad (109)$$

so

$$\mathbf{u}_\perp = u_0 (\boldsymbol{\epsilon}_1 - i\boldsymbol{\epsilon}_2) e^{-i\omega_B t}. \quad (110)$$

We may integrate over time to find the trajectory:

$$\frac{d\mathbf{x}}{dt} = u_\parallel \boldsymbol{\epsilon}_3 + \mathbf{u}_\perp \quad (111)$$

and so

$$\begin{aligned} \mathbf{x}(t) &= \mathbf{x}(0) + \int_0^t dt' [u_\parallel \boldsymbol{\epsilon}_3 + u_0 (\boldsymbol{\epsilon}_1 - i\boldsymbol{\epsilon}_2) e^{-i\omega_B t'}] \\ &= \mathbf{x}(0) + u_\parallel t \boldsymbol{\epsilon}_3 + i \frac{u_0}{\omega_B} (\boldsymbol{\epsilon}_1 - i\boldsymbol{\epsilon}_2) (e^{-i\omega_B t} - 1). \end{aligned} \quad (112)$$

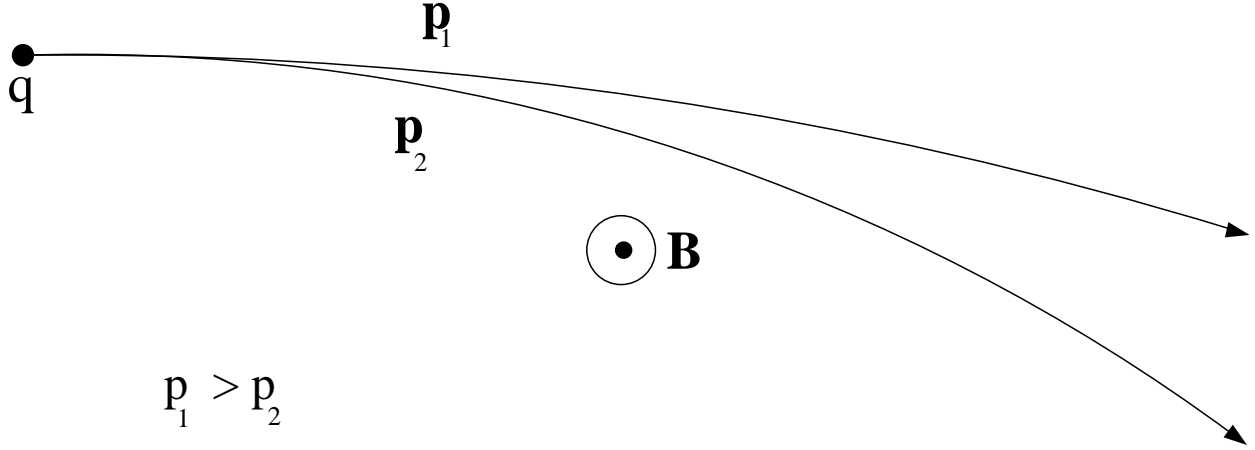
The physical trajectory is the real part of this and is, for real u_0 ,

$$\mathbf{x}(t) = \mathbf{x}(0) + u_\parallel t \boldsymbol{\epsilon}_3 + \frac{u_0}{\omega_B} [\sin(\omega_B t) \boldsymbol{\epsilon}_1 + (\cos(\omega_B t) - 1) \boldsymbol{\epsilon}_2]. \quad (113)$$

This equation describes helical motion with the helix axis parallel to the z -axis. The radius of the axis is a , where $a = u_0/\omega_B$.

It is worthwhile to establish the connection between a and $|\mathbf{p}_\perp|$ where $\mathbf{p}_\perp = m\gamma \mathbf{u}_\perp$ is the momentum in the plane perpendicular to the direction of \mathbf{B} .

$$p_\perp = m\gamma u_0 = m\gamma \omega_B a = m\gamma \frac{qB}{m\gamma c} a = \frac{qBa}{c}. \quad (114)$$



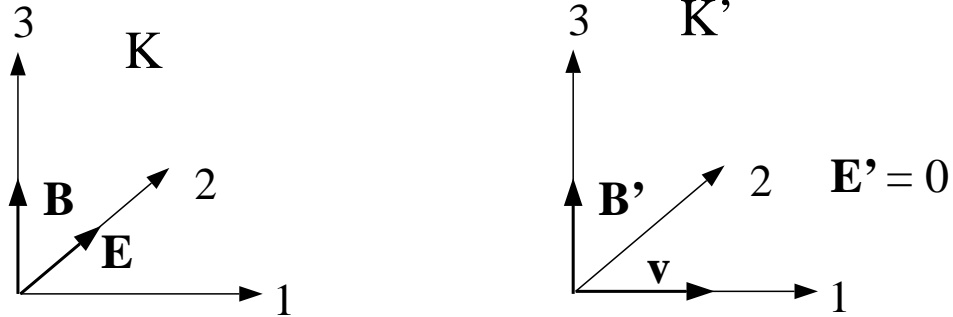
This relation, $p_{\perp} = qBa/c$, tells us the radius of curvature in the plane perpendicular to \mathbf{B} (which is not the same as the radius of curvature of the orbit) is a linear function of p_{\perp} , and it suggests a simple way to select particles of a given momentum out of a beam containing particles with many momenta. One simply passes the beam through a region of space where there is some \mathbf{B} applied transverse to the direction of the beam. The amount by which a particle is deflected will increase with decreasing p_{\perp} and so the beam is spread out much as a prism separates the different frequency components of a beam of light. The device is a momentum selector.

4.2 Motion in crossed \mathbf{E} and \mathbf{B} fields, $E < B$.

For $\mathbf{E} \cdot \mathbf{B} = 0$ in frame K , we can find a frame K' where $\mathbf{E}' = 0$, provided $E < B$. This may be seen from the form of the field transforms.

$$\begin{aligned} \mathbf{E}'_{\parallel} &= \mathbf{E}_{\parallel} & \mathbf{E}'_{\perp} &= \gamma[\mathbf{E}_{\perp} + (\boldsymbol{\beta} \times \mathbf{B})] \\ \mathbf{B}'_{\parallel} &= \mathbf{B}_{\parallel} & \mathbf{B}'_{\perp} &= \gamma[\mathbf{B}_{\perp} - (\boldsymbol{\beta} \times \mathbf{E})] \end{aligned} \quad (115)$$

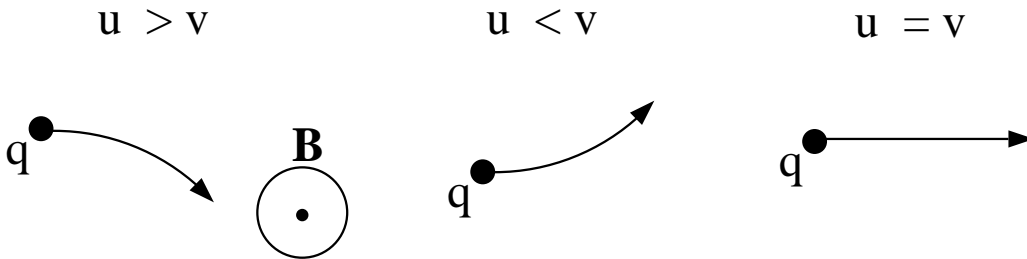
In fact, we have already solved exactly this problem in chapter 11 where we found that K' moves relative to K with a velocity which is $\mathbf{v} = c(\mathbf{E} \times \mathbf{B})/B^2$. If we let $\mathbf{E} = E\boldsymbol{\epsilon}_2$ and $\mathbf{B} = B\boldsymbol{\epsilon}_3$, then $\mathbf{v} = c(E/B)\boldsymbol{\epsilon}_1$, and $\mathbf{B}' = B\sqrt{1 - E^2/B^2}\boldsymbol{\epsilon}_3$.



Now imagine a particle is injected into this system with an initial velocity⁷ $\mathbf{u}(0) = u_0\boldsymbol{\epsilon}_1$. In the frame K' , its initial velocity is

$$\mathbf{u}'(0) = \frac{u_0 - v}{1 + u_0v/c^2}\boldsymbol{\epsilon}_1. \quad (116)$$

From our first example, we know that the particle will proceed to execute circular motion in this frame, always with the same speed $u'(0)$. What then is its motion in frame K ? Superposed on the circular motion will be a drift velocity \mathbf{v} . If $q > 0$ and $u_0 > v$, we get the first motion shown below. But if $u_0 < v$, we get the second motion. For the special case of $u_0 = v$, the particle is at rest in K' which means it moves in K at a constant velocity $\mathbf{u}(t) = \mathbf{v}$.



⁷We could be more general and include a component of \mathbf{u} parallel to \mathbf{B} ; this would not lead to anything significantly different from what we are about to find.

Such a device can be employed as a velocity selector and so it complements the device described in the first example which was a momentum selector. The idea is that a particle coming in with a speed u_0 greater than v will experience a magnetic force greater than the electric force and so it will be deflected accordingly. But one coming in with a speed smaller than v will experience an electric force greater than the magnetic one, and it will be deflected in the other direction.

The picture changes after a while, however, because the particle will speed up and slow down under the influence of the electric field. Suppose that initially $u_0 > v$ ($u_0 < v$). Then the B -field (the E -field) force dominates, and the particle is deflected in such a way that it moves against (with) the electric field. This causes it to slow down (speed up) so that after some time $u_0 < v$ ($u_0 > v$). Then the electric (magnetic) field force dominates, causing the particle to swing around so that it eventually moves with (against) the electric field force. And so on. The end result is a trajectory that produces a time-averaged velocity equal to \mathbf{v} or $c(\mathbf{E} \times \mathbf{B})/B^2$. This is called the *E cross B drift velocity*. It is in the direction of $\mathbf{E} \times \mathbf{B}$ no matter what is the sign of the charge.

4.3 Motion in crossed \mathbf{E} and \mathbf{B} fields, $E > B$

This time we want to consider the motion in a frame K' moving at velocity $\mathbf{v} = c(\mathbf{E} \times \mathbf{B})/E^2 \rightarrow c(B/E)\mathbf{e}_1$, if we keep the same directions of the fields as in the preceding example. In this frame there is only an electric field $\mathbf{E}' = \mathbf{E}\sqrt{1 - B^2/E^2}$ which will cause the particle to move away in the direction of \mathbf{E}' . The equations of motion in K' are

$$mc^2 \frac{d\gamma'}{dt'} = qE' \frac{dy'}{dt'} \quad \text{and} \quad \frac{dp'_y}{dt'} = qE'; \quad (117)$$

the components of the momentum in the other directions are constant. One easily solves to find

$$\mathbf{p}'(t') = \mathbf{p}'(0) + qE't'\mathbf{e}_2 \quad (118)$$

and we can then find γ' directly from the dispersion relation,

$$\gamma' = \frac{1}{mc^2} \sqrt{m^2 c^4 + \mathbf{p}'(t') \cdot \mathbf{p}'(t') c^2}. \quad (119)$$

The speed u'_y is found easily from the equation of motion for γ' which integrates trivially to produce

$$\begin{aligned} y'(t') &= y'(0) + \frac{mc^2}{qE'} (\gamma'(t') - \gamma'(0)) \\ &= y'(0) + \frac{\sqrt{m^2 c^4 + (\mathbf{p}'(t'))^2 c^2} - \sqrt{m^2 c^4 + (\mathbf{p}'(0))^2 c^2}}{qE'}. \end{aligned} \quad (120)$$

Consider also \mathbf{x}'_{\perp} , the component of \mathbf{x}' perpendicular to the electric field. Because $\mathbf{p}'_{\perp}/dt = 0$, it is true that $\gamma' \mathbf{y}'_{\perp} = \gamma'(0) \mathbf{u}'_{\perp}(0)$, a constant. Hence

$$\mathbf{u}'_{\perp}(t') = \mathbf{u}'_{\perp}(0) \sqrt{1 + (\mathbf{p}'(0))^2 / m^2 c^2} / \sqrt{1 + (\mathbf{p}'(t'))^2 / m^2 c^2}. \quad (121)$$

We can integrate the velocity over time to find the displacement of the particle. For the special case that there is no component of $\mathbf{p}'(0)$ in the direction of the field, one finds that

$$\mathbf{x}'_{\perp}(t) - \mathbf{x}'_{\perp}(0) = \frac{\mathbf{p}'_{\perp}(0)}{qE'} \ln \left[\frac{qE't'}{m\gamma(0)} + \sqrt{1 + \left(\frac{qE't'}{m\gamma(0)} \right)^2} \right]. \quad (122)$$

We can combine Eqs. (113) and (115) to remove the time and so have an equation that determines the shape of the trajectory. For simplicity, let $\mathbf{x}'_{\perp}(0) = y'(0) = 0$. Then one finds

$$\frac{x'_{\perp} qE'}{m\gamma'(0)u'(0)} = \ln \left[\sqrt{\left(1 + \frac{qE'y'}{m\gamma'(0)} \right)^2 - 1} + 1 + \frac{qE'y'}{m\gamma'(0)} \right]. \quad (123)$$

For short times satisfying the condition $|qE'y'/m\gamma'(0)| \ll 1$, the trajectory is a parabola,

$$\frac{x'_{\perp} qE'}{m\gamma'(0)u'(0)} \approx \sqrt{\frac{2qE'y'}{m\gamma'(0)}} \quad (124)$$

or

$$y' = \frac{qE'x_{\perp}^2}{2m\gamma'(0)(u'(0))^2}. \quad (125)$$

The long time behavior is displayed for $|qE'y'/m\gamma'(0)| \gg 1$, and it is such that

$$y' = \frac{m\gamma'(0)}{2qE'} \exp\left(\frac{x'_{\perp}qE'}{m\gamma'(0)u'(0)}\right). \quad (126)$$

4.4 Motion for general uniform \mathbf{E} and \mathbf{B} .

Then we cannot find a frame where one of the fields can be made to vanish. But there is a frame where the electric field and magnetic induction are parallel; here the solution of the equations of motion is relatively simple and is left as an exercise.

4.5 Motion in slowly spatially varying $\mathbf{B}(\mathbf{x})$

This problem is greatly simplified by (1) the fact that then energy, or γ , is a constant and by (2) the assumption that $\mathbf{B}(\mathbf{x})$ does not vary much relative to its magnitude over distances on the order of the radius of the particle's orbit.

Chapter Thirteen

Charged Particle Collisions, Energy Loss, Scattering

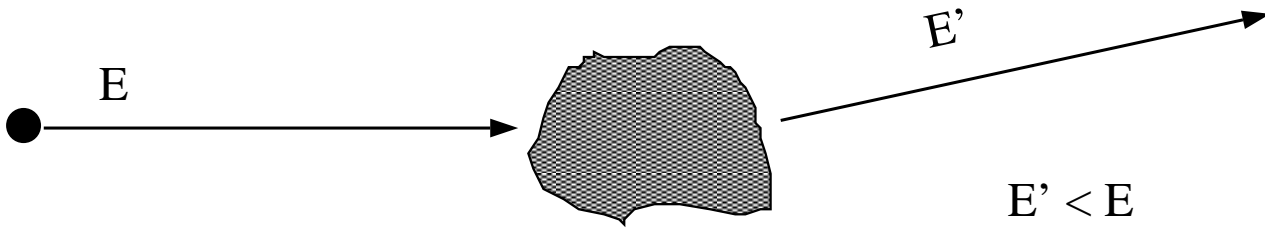
Niels Henrik David Bohr
(1885 - 1962)

September 17, 2001

Contents

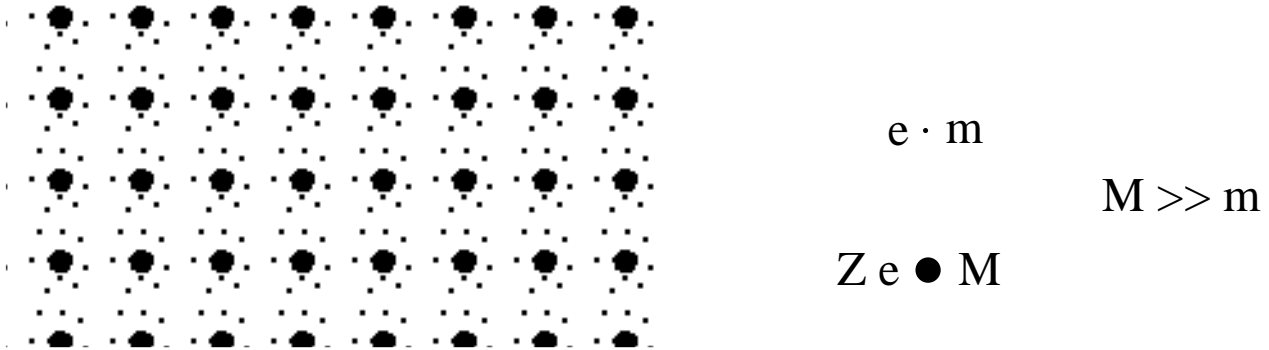
1	Energy Transfer in Coulomb Collisions	3
1.1	Classical Impulse Approximation	3
1.2	Validity of Approximations	6
1.3	Energy Loss	7
2	Collisions with a Harmonically Bound Charge; Energy Loss	10
3	Density Effect in Energy Loss	17
4	Cherenkov Radiation	26
5	Momentum Transfer	29
5.1	Average Angle of Deflection	33
5.1.1	Distribution of Small Angle Scattering	35
5.1.2	The Distribution of Large Angle Scattering	36

The topic of interest is a charged particle traversing a material medium. Such a particle loses energy by scattering from the charged particles, electrons and nuclei, in the material.



This is an interesting system from many points of view. Historically it was extremely important in resolving the question of the structure of matter (the Rutherford atom), and at present energy loss is an important phenomenon in particle physics and is also studied in detail by nuclear engineers and by condensed matter physicists in connection with the properties of materials and radiation damage to materials.

The problem can be studied as a straightforward application of electromagnetism; charged particles scatter from one another with the result that energy and momentum are transferred. The scattering centers in a material are of two distinct types; there are electrons of charge $-e$ and small mass $m \sim 10^{-27} g$, and there are nuclei of charge Ze with Z up to about 10^2 and large mass $M \sim 10^{-22} g$. Thus the nuclear charge is significantly larger than that of an electron, and the nuclear mass is much larger—some 10^5 times larger—than the electronic mass. It is also important to realize that there are Z more electrons than nuclei (Z is the atomic number of the atoms in the material) in a given volume of target material. Consequently the electrons provide Z times as many scattering centers as the nuclei.



As we shall see, it turns out that electrons soak up most of the energy of an incident particle while nuclei are responsible for most of the momentum transfer in the sense that they are more effective than electrons at deflecting the incident particle from its initial direction of motion.

1 Energy Transfer in Coulomb Collisions

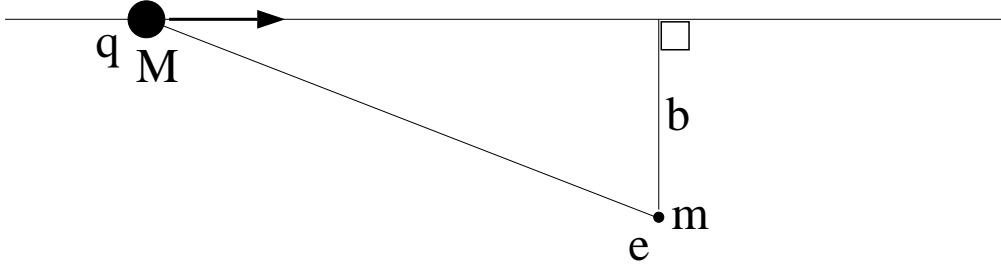
The general problem of energy transfer when a charged particle traverses a material is naturally very complicated. We shall approach it a little at a time starting with the classical impulse approximation applied to a pair of particles.

1.1 Classical Impulse Approximation

Consider a particle (q, M) , where q is the charge and M , the mass, incident with speed v on a second particle $(-e, m)$ at rest in the frame of our calculation. The incident particle has total energy $M\gamma c^2$, where $\gamma = 1/(1 - v^2/c^2)^{1/2}$. In the impulse approximation the incident particle is treated as undeflected by the collision. Further, the target is approximated as stationary during the collision. Then it is easy to calculate the momentum, or impulse, transferred from the incident particle to the target.

Given the approximation that the incident particle's trajectory is unaffected by

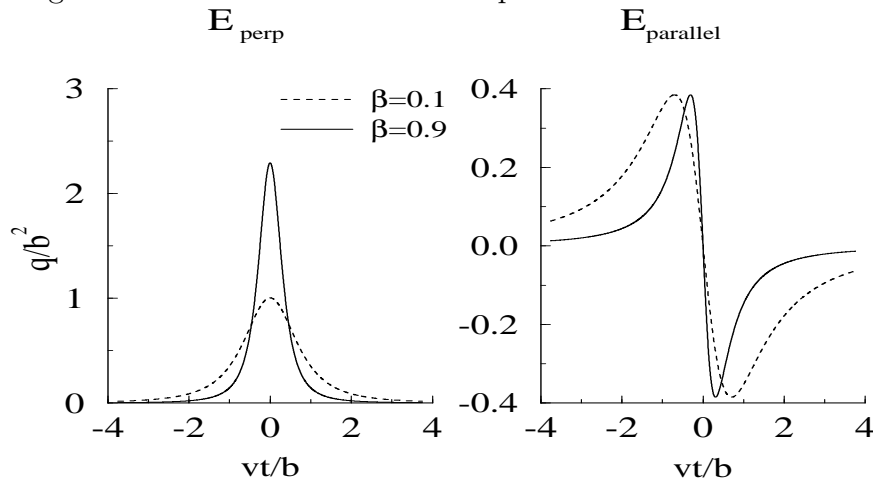
the collision, it travels with constant velocity and passes the target at some distance b called the *impact parameter*.



The momentum transferred to the target can be expressed as the integral over time of the force acting on it, and we can find that from knowledge of the electric field produced by the incident particle at the location of the target. From prior calculations in Chapter 11, we know that this field is $\mathbf{E} = \mathbf{E}_\perp + \mathbf{E}_\parallel$ where the parallel and perpendicular components act parallel and perpendicular to the line of motion of the incident particle. These components are given at the target, by

$$E_\perp(b) = \frac{\gamma qb}{(b^2 + \gamma^2 v^2 t^2)^{3/2}} \quad \text{and} \quad E_\parallel(b) = -\frac{\gamma qvt}{(b^2 + \gamma^2 v^2 t^2)^{3/2}} \quad (1)$$

where the origin of time is chosen so that the particles are closest at $t = 0$.



The integral over time of E_\parallel is zero while that of E_\perp provides the momentum trans-

ferred to the target,

$$\begin{aligned}
 p &= \left| \int_{-\infty}^{\infty} dt (-eE_{\perp}) \right| = \left| \int_{-\infty}^{\infty} dt \frac{e\gamma qb}{(b^2 + \gamma^2 v^2 t^2)^{3/2}} \right| = \left| \frac{eqb}{v} \right| \int_{-\infty}^{\infty} \frac{dx}{(b^2 + x^2)^{3/2}} \\
 &= \left| \frac{qe}{bv} \right| \int_{-\infty}^{\infty} \frac{du}{(1 + u^2)^{3/2}} = \frac{2|qe|}{bv}.
 \end{aligned} \tag{2}$$

Next, we shall assume that $p \ll mc$ so that the energy transfer to the target may be approximated by the non-relativistic formula $p^2/2m$. This gives us a third approximation whose validity we must scrutinize. Adopting it, we find that the energy transfer to the target is¹

$$\Delta E = \frac{p^2}{2m} = \frac{2q^2 e^2}{mb^2 v^2} = \left(\frac{qe}{b} \right) \frac{(qe/b)}{(mv^2/2)} \propto \frac{e^2}{m}. \tag{3}$$

Notice that the energy transfer is proportional to the square of the charge of the target particle and inversely proportional to its mass. Possible targets are electrons and nuclei. A nucleus has a larger charge than an electron by a factor of the atomic number z , giving the nucleus an “advantage” by a factor of z^2 when it comes to extracting energy from the incident particle. However, nuclei are more massive than electrons by a factor of $1836A$ where A is the atomic weight which is as large as or larger than z . Furthermore, there are z more electrons than nuclei to act as targets. *Hence we see that the electrons are more effective than nuclei at taking the energy of the incident particle by a factor of at least 1836.* For this reason, we shall henceforth suppose that the target particle is an electron so long as we are interested in the energy transfer, as opposed to the momentum transfer, from the incident particle to the target.

effect	nucleus	electron
charge	z^2	1
mass	$1/(1836z)$	1
number	1	z
total	$z/(1836)$	z

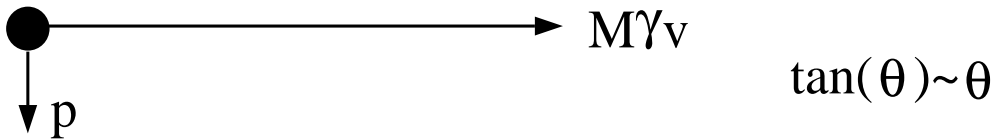
¹We shall suppose $qe > 0$ so that the notation is simplified.

1.2 Validity of Approximations

Our simple calculation of the energy transfer contains three distinct approximations.

1. It is assumed that the incident particle is not deflected from its straight-line path. This assumption is valid so long as the actual angle of deflection θ obeys the inequality $\theta \ll 1$.
2. It is assumed that the target particle is at a particular point during the entire collision. This assumption is valid provided the target recoils a distance d during the collision which is small compared to the impact parameter b , $d \ll b$.
3. We assumed that the recoiling particle is non-relativistic, $p \ll mc$.

We can determine the conditions under which the approximations are valid. **First**, the angle of deflection of the incident particle is of order $p/M\gamma v$,



$$\theta \approx \frac{p}{M\gamma v} \approx \frac{2qe}{\gamma b M v^2} = \frac{2}{\gamma} \left(\frac{(qe/b)}{M v^2} \right). \quad (4)$$

Thus we require that the electrostatic potential energy of interaction at a separation of the impact parameter should be small compared to the energy $M\gamma v^2$ which is something like the energy of the incident particle. This condition is generally met. It is also not independent of the other conditions required for the validity of the impulse approximation.

Second, consider the requirement that the target not recoil far in comparison with b during the collision. From the form of the electric field

$$E_{\perp}(b) = \frac{\gamma q b}{(b^2 + \gamma^2 v^2 t^2)^{3/2}}$$

and hence the force experienced the the particle, we can see that the duration τ of the collision is roughly $b/\gamma v$. During this time the recoil particle moves a distance of order $(p/m)\tau$, assuming it was initially at rest, so the requirement is

$$\frac{b}{\gamma v} \left(\frac{2qe}{mbv} \right) \ll b \quad \text{or} \quad \frac{1}{\gamma} \frac{(qe/b)}{mv^2} \ll 1. \quad (5)$$

This inequality is much like the previous one but note the replacement of M with the mass m . Given that the target is an electron, which has the smallest mass of all charged particles², the present condition is at least as strong as the condition $\theta \ll 1$. Consequently, we can forget about the latter.

Notice that the condition (5) can also be written as

$$\frac{1}{\gamma} \frac{c^2 r_0}{v^2 b} \ll 1 \quad (6)$$

where $r_0 \equiv e^2/mc^2$ is the *classical radius of the electron* which is about $2.82 \times 10^{-13} \text{ cm}$. Thus, provided the factor $c^2/v^2\gamma$ is not much larger than unity, this condition is met for impact parameters (almost) down to r_0 which is also about the size of a baryon or nucleus at which point we would expect the calculation to fail for entirely different reasons. Notice, however, that the condition becomes much more severe if v is not large, i.e., if the incident particle is not relativistic. That is not surprising; the collision will last much longer if the incident particle moves slowly and the target has more time to recoil during the collision.

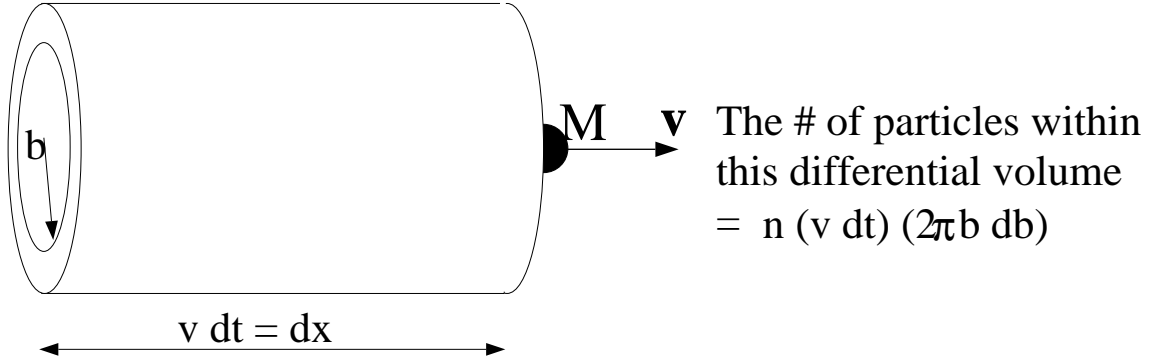
Third, comes the condition that $p \ll mc$, or $2(qe/b)/mvc \ll 1$. This condition is not much different from Eq. (5).

1.3 Energy Loss

We have calculated, in the impulse approximation, the energy absorbed from an incident particle by a single electron. There is never just one electron. We have to figure out how to add up the contributions of many electrons to determine how much

²As far as anyone knows.

energy the incident particle will lose per unit length of its path. Given an electron density n , then an incident particle having speed v will pass $n(vdt)(2\pi b db)$ scatterers per unit time at distances between b and $b + db$.



The consequent energy change of the incident particle in time dt is, from Eq. (3)

$$d^2 E = -dt db v 2\pi b n \left(\frac{2q^2 e^2}{mb^2 v^2} \right). \quad (7)$$

Now integrate over b to get the contributions from scatterers at all distances. This must be done with some care. Let's just integrate b from some minimum to some maximum:

$$\frac{dE}{dt} = -\frac{4\pi n q^2 e^2}{mv} \int_{b_{min}}^{b_{max}} \frac{db}{b} = -\frac{4\pi n q^2 e^2}{mv} \ln(b_{max}/b_{min}). \quad (8)$$

The mathematical necessity of the upper and lower cutoffs on b is clear; the integral would diverge at either end without the cutoff. Physically, what is the reason for them? We have just seen that the impulse approximation breaks down at $b \rightarrow 0$ because the recoiling particle recoils by a distance comparable to or larger than the impact parameter during the collision in that limit. Referring back to the condition that our approximation is valid

$$\frac{b}{\gamma v} \left(\frac{2qe}{mbv} \right) \ll b \quad \text{or} \quad \frac{1}{\gamma} \frac{(qe/b)}{mv^2} \ll 1,$$

we see that a reasonable value for the cutoff is $b_{min} = qe/m\gamma v^2$. This will also make certain that the incident particle's deflection θ is small.

What about the upper cutoff? The physical reason for the breakdown of the impulse approximation (which then necessitates the cutoff) at large b is that when b is large, the collision time $\tau = b/\gamma v$ is long. When this time is long, the natural motions of the target cannot be neglected; that is, the electron or target is not really at rest although we treated it as such when calculating the energy transfer. Most electrons are bound to atoms, molecules, or ions with some binding energy E_e giving them a natural angular frequency of motion $\omega_0 = E_e/\hbar$. The corresponding period is of order $1/\omega_0$. The collision time must be small compared to this time or the impulse approximation, as we have derived it, breaks down. That suggests we choose b_{max} according to $b_{max}/\gamma v = 1/\omega_0$ or $b_{max} = \gamma v/\omega_0$.

Using these cutoffs, we find that the rate of change with time of the incident particle's energy is

$$\frac{dE}{dt} = -\frac{4\pi n q^2 e^2}{mv} \ln\left(\frac{m\gamma^2 v^3}{qe\omega_0}\right). \quad (9)$$

A perhaps more interesting quantity is $dE/dx = v^{-1}dE/dt$,

$$\frac{dE}{dx} = -\frac{4\pi n q^2 e^2}{mv^2} \ln\left(\frac{m\gamma^2 v^3}{qe\omega_0}\right) \quad (10)$$

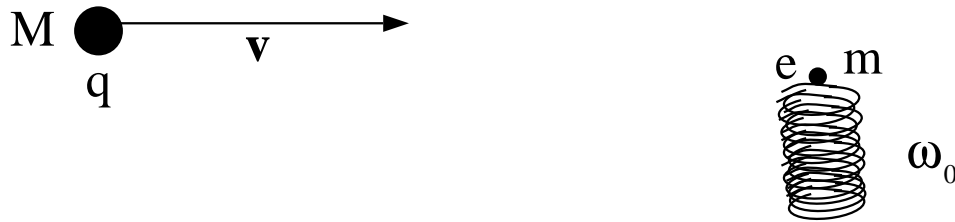
In this derivation, we have determined the lower cutoff on b by looking at the breakdown of the classical impulse approximation. There is also a breakdown associated with quantum effects which implies a somewhat different lower cutoff. The quantum breakdown can be understood by appealing to the uncertainty principle. The value of b is uncertain by an amount related to the momentum of the incident particle. We claim that it has no momentum in the direction in which the impact parameter is measured. We can't really know this to be precisely the case and there has to be an uncertainty in the impact parameter which is of order $\hbar/m\gamma v$. If this uncertainty is comparable to b itself, then our calculation fails. Hence the quantum mechanical cutoff is $b_{min}^{(q)} = \hbar/m\gamma v$. In any given situation, we have to use the larger of the two lower cutoffs. The ratio of the two is

$$\frac{b_{min}^{(q)}}{b_{min}} = \frac{\hbar}{m\gamma v} \frac{\gamma m v^2}{qe} = \frac{\hbar v}{qe} = \frac{1}{\alpha(q/e)} \frac{v}{c} \approx \frac{137}{(q/e)} \frac{v}{c} \quad (11)$$

where $\alpha \equiv e^2/\hbar c \approx 1/137$ is the *fine structure constant*. If this parameter is larger than unity, the quantum cutoff should be employed; if it is smaller than unity, the classical one is appropriate.

2 Collisions with a Harmonically Bound Charge; Energy Loss

One can remove the need for introducing the upper and lower cutoffs on b by doing a more careful treatment of the (classical) energy transfer in the collision. The more careful treatment needed at small b is relegated to the homework. The one needed at large b , which must include the natural motion of the target particle, is given here.



Suppose that the target is bound harmonically at a point, taken as the origin of coordinates, meaning that there is a restoring force $-m\omega_0^2\mathbf{x}$, where \mathbf{x} is the particle's position and ω_0 is the natural frequency of the oscillator, in the absence of damping or perturbing forces. Given that the particle is an electron with mass m and charge $-e$, its equation of motion in the presence of an applied electric field (the one coming from the incident particle) is

$$m\frac{d^2\mathbf{x}}{dt^2} = -m\omega_0^2\mathbf{x} - m\Gamma\frac{d\mathbf{x}}{dt} - e\mathbf{E}(\mathbf{x}, t) \quad (12)$$

where the term $-m\Gamma d\mathbf{x}/dt$ is a damping force proportional to the particle's velocity; Γ is a 'damping constant'. This term is typically small compared to the restoring

force term.

To simplify the solution, we will make several **Approximations**.

1. We have not included the magnetic force which acts on the bound charge. This force is smaller than the electric force by a factor of the recoiling particle's velocity divided by c even if the incident particle is relativistic; since the recoiling particle is not relativistic, we may ignore the magnetic force.
2. We make one more approximation which is to evaluate $\mathbf{E}(\mathbf{x}, t)$ at the origin or point where the target particle is bound; this is reasonable provided $b \gg |\mathbf{x}|$ because then the electric field will vary but little over distances of order $|\mathbf{x}|$.

We solve Eq. (12) by making a Fourier analysis of the motion. Write

$$\mathbf{E}(t) \equiv \mathbf{E}(0, t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d\omega' \mathbf{E}(\omega') e^{-i\omega' t} \quad (13)$$

and

$$\mathbf{x}(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d\omega' \mathbf{x}(\omega') e^{-i\omega' t}. \quad (14)$$

The inverse transforms are

$$\mathbf{E}(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dt \mathbf{E}(t) e^{i\omega t} \quad (15)$$

and

$$\mathbf{x}(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dt \mathbf{x}(t) e^{i\omega t}. \quad (16)$$

Substitute Eqs. (13) and (14) directly into the equation of motion and perform the time derivatives to find

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d\omega' [-\omega'^2 - i\omega'\Gamma + \omega_0^2] \mathbf{x}(\omega') e^{-i\omega' t} = -\frac{(e/m)}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d\omega' \mathbf{E}(\omega') e^{-i\omega' t}. \quad (17)$$

If we multiply by $e^{i\omega t}$ and integrate over t , we obtain a delta-function, $\delta(\omega - \omega')$, and can then integrate trivially over ω' to find a solution for $\mathbf{x}(\omega)$ which is

$$\mathbf{x}(\omega) = -\frac{e\mathbf{E}(\omega)}{m} \frac{1}{\omega_0^2 - i\omega\Gamma - \omega^2}. \quad (18)$$

We could now figure out what is $\mathbf{E}(\omega)$ since we know $\mathbf{E}(t)$ and use it in Eq. (18) to find $\mathbf{x}(\omega)$ and then Fourier transform the latter to find $\mathbf{x}(t)$. But we aren't really interested in $\mathbf{x}(t)$. What we are trying to determine is the energy transferred to the target from the incident charge. That energy can be found as follows:

$$\frac{dE}{dt} = \mathbf{F} \cdot \frac{d\mathbf{x}}{dt} = -e\mathbf{E}(t) \cdot \frac{d\mathbf{x}(t)}{dt} \quad (19)$$

where we again approximate $\mathbf{E}(\mathbf{x}, t)$ with $\mathbf{E}(0, t)$. The total energy transferred in the collision is

$$\begin{aligned} \Delta E &= - \int_{-\infty}^{\infty} dt e\mathbf{E}(t) \cdot \frac{d\mathbf{x}(t)}{dt} \\ &= - \int_{-\infty}^{\infty} dt \frac{e}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d\omega' \mathbf{E}(\omega') e^{-i\omega't} \cdot \frac{d}{dt} \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d\omega \mathbf{x}(\omega) e^{-i\omega t} \right) \\ &= -e \int_{-\infty}^{\infty} d\omega (-i\omega) \mathbf{x}(\omega) \cdot \mathbf{E}(-\omega). \end{aligned} \quad (20)$$

The last step is achieved by, first, taking the time derivative; second, integrating over t to obtain a delta-function $\delta(\omega + \omega')$; and, finally, integrating over ω' .

Because the electric field is real, $\mathbf{E}(-\omega) = \mathbf{E}^*(\omega)$. Similarly, $\mathbf{x}(-\omega) = \mathbf{x}^*(\omega)$; hence

$$\Delta E = ie \int_{-\infty}^{\infty} d\omega \omega \mathbf{x}(\omega) \cdot \mathbf{E}^*(\omega) = \Re \left[2ie \int_0^{\infty} d\omega \omega \mathbf{x}(\omega) \cdot \mathbf{E}^*(\omega) \right]. \quad (21)$$

Using our solution for $\mathbf{x}(\omega)$, we find

$$\begin{aligned} \Delta E &= \Re \left(-2i \frac{e^2}{m} \int_0^{\infty} d\omega \frac{\omega |\mathbf{E}(\omega)|^2}{\omega_0^2 - i\omega\Gamma - \omega^2} \right) \\ &= \Re \left(-2i \frac{e^2}{m} \int_0^{\infty} d\omega \frac{\omega |\mathbf{E}(\omega)|^2 (\omega_0^2 - \omega^2 + i\omega\Gamma)}{(\omega_0^2 - \omega^2)^2 + \omega^2\Gamma^2} \right) \\ &= \frac{e^2}{m} \int_0^{\infty} d\omega \frac{2\omega^2\Gamma |\mathbf{E}(\omega)|^2}{(\omega_0^2 - \omega^2)^2 + \omega^2\Gamma^2} \end{aligned} \quad (22)$$

Finally, consider the limit that Γ is very small (small damping). Then the entire weight in the integrand is at $\omega = \omega_0$ which means that the only part of \mathbf{E} which contributes to the energy transfer is the part whose frequency matches the natural

frequency of the oscillator. In this limit the integral can be done by evaluating $\mathbf{E}(\omega)$ at ω_0 so that

$$\begin{aligned}\Delta E &\approx \frac{2e^2}{m} |\mathbf{E}(\omega_0)|^2 \int_0^\infty d\omega \frac{\omega^2 \Gamma}{(\omega_0^2 - \omega^2)^2 + \omega^2 \Gamma^2} \\ &= \frac{2e^2}{m} |\mathbf{E}(\omega_0)|^2 \int_0^\infty dx \frac{x^2}{[(\omega_0/\Gamma)^2 - x^2]^2 + x^2}\end{aligned}\quad (23)$$

where $x \equiv \omega/\Gamma$. The remaining integral is

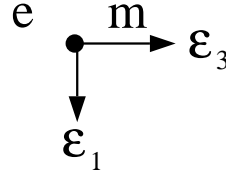
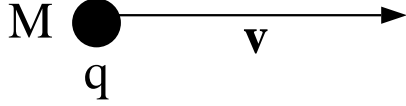
$$\begin{aligned}I &= \int_0^\infty \frac{dx}{[(\omega_0/\Gamma)^2 - x^2]^2/x^2 + 1} \approx \int_{-\omega_0/\Gamma}^\infty \frac{dy}{(\Gamma/\omega_0)^2 [2y\omega_0/\Gamma + y^2]^2 + 1} \\ &\approx \int_{-\infty}^\infty \frac{dy}{4y^2 + 1} = \frac{\pi}{2}.\end{aligned}\quad (24)$$

Hence

$$\Delta E = \frac{\pi e^2}{m} |\mathbf{E}(\omega_0)|^2 \quad (25)$$

in the limit of $\Gamma \ll \omega_0$.

We still need to evaluate $\mathbf{E}(\omega_0)$.



If the incident particle is moving in the z direction and the target lies in the x direction relative to the track of the incident particle, then the components of the electric field are

$$\mathbf{E}_{\parallel}(t) = -\frac{qv\gamma t}{(b^2 + \gamma^2 v^2 t^2)^{3/2}} \boldsymbol{\epsilon}_3 \quad \text{and} \quad \mathbf{E}_{\perp}(t) = \frac{\gamma qb}{(b^2 + \gamma^2 v^2 t^2)^{3/2}} \boldsymbol{\epsilon}_1. \quad (26)$$

Hence

$$\mathbf{E}_{\perp}(\omega) = \frac{qb\gamma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dt \frac{e^{i\omega t}}{(b^2 + \gamma^2 v^2 t^2)^{3/2}} \boldsymbol{\epsilon}_1 = \frac{qb\gamma}{\sqrt{2\pi}} \frac{b}{\gamma v} \frac{1}{b^3} \int_{-\infty}^{\infty} dx \frac{e^{-i(\omega b/\gamma v)x}}{(1+x^2)^{3/2}} \boldsymbol{\epsilon}_1$$

$$= \frac{q}{bv\sqrt{2\pi}} \int_{-\infty}^{\infty} dx \frac{e^{-izx}}{(1+x^2)^{3/2}} \epsilon_1 = \frac{2q}{bv\sqrt{2\pi}} \int_0^{\infty} dx \frac{\cos(zx)}{(1+x^2)^{3/2}} \epsilon_1 \quad (27)$$

where $z = \omega b/\gamma v$. The integral we are contemplating is a Bessel function; that is,

$$K_\nu(z) = \frac{2^\nu \Gamma(\nu + 1/2)}{\sqrt{\pi} z^\nu} \int_0^{\infty} dx \frac{\cos(xz)}{(1+x^2)^{\nu+1/2}}; \quad (28)$$

specifically,

$$K_1(z) = \frac{2\Gamma(3/2)}{\sqrt{\pi} z} \int_0^{\infty} dx \frac{\cos(zx)}{(1+x^2)^{3/2}}. \quad (29)$$

Further, $\Gamma(3/2) = \sqrt{\pi}/2$, so

$$\mathbf{E}_\perp(\omega) = \frac{q}{bv} \sqrt{\frac{2}{\pi}} z K_1(z) \epsilon_1. \quad (30)$$

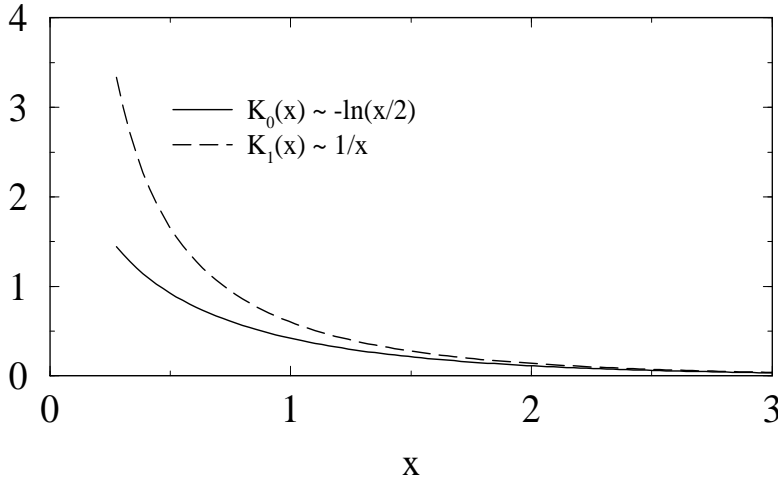
By similar manipulations one finds that

$$\mathbf{E}_\parallel(\omega) = -i \frac{q}{\gamma v b} \sqrt{\frac{2}{\pi}} z K_0(z) \epsilon_3. \quad (31)$$

Hence the energy transfer is, from Eq. (25),

$$\Delta E = \frac{\pi e^2}{m} \frac{q^2}{b^2 v^2} \frac{2}{\pi} \left[z^2 K_1^2(z) + \frac{z^2}{\gamma^2} K_0^2(z) \right] = \frac{2q^2 e^2}{mb^2 v^2} \left[z^2 K_1^2(z) + \frac{z^2}{\gamma^2} K_0^2(z) \right] \quad (32)$$

where $z = \omega_0 b/\gamma v = b/b_{max}$ using $b_{max} = \gamma v/\omega_0$ as per the criterion discussed in the preceding section.



Cutoffs. What is the qualitative behavior of this result as a function of b ? For $b \ll b_{max}$, $z \ll 1$ and $zK_0(z) \rightarrow 0$ as $z \rightarrow 0$ while $zK_1(z) \rightarrow 1$. Making these substitutions in Eq. (32) we find that for small b , the result is the same as what emerged from the impulse approximation.³ Thus we must still insert by hand a cutoff for small b (b_{min}). For large $b \gg b_{max}$, however, $z \gg 1$ and the Bessel functions' behavior is

$$K_0(z) \sim K_1(z) \sim \sqrt{\frac{\pi}{2z}} e^{-z} \quad (33)$$

so that in this regime of b ,

$$\Delta E \approx \frac{2q^2 e^2}{mv^2 b^2} \left[\frac{\pi}{2z} z^2 \left(1 + \frac{1}{\gamma^2} \right) e^{-2z} \right] = \frac{q^2 e^2 \pi z}{mv^2 b^2} \left(1 + \frac{1}{\gamma^2} \right) e^{-2z}. \quad (34)$$

Thus the large b cutoff is automatically included in this formalism

We can find the total energy loss per unit path length by integrating over b as before. Given an electron density n , then an incident particle traversing a distance dx will pass $n(dx)(2\pi b db)$ scatterers with impact parameters between b and $b + db$. The integral for the energy loss by the incident particle can then be put in the form

$$d^2 E = -(2\pi b db)(v dt)n \Delta E,$$

or, since $b = b_{max} z = \frac{v\gamma}{\omega_0} z$ and $dx = v dt$,

$$d^2 E = -2\pi \left(\frac{v\gamma}{\omega_0} \right)^2 z dz dx n \Delta E.$$

Then, integrating on z , we get

$$\frac{dE}{dx} = -2\pi n \frac{2q^2 e^2}{mv^2} \int_{z_{min}}^{\infty} \frac{dz}{z} \left[z^2 K_1^2(z) + \frac{z^2}{\gamma^2} K_0^2(z) \right] \quad (35)$$

where $z_{min} = b_{min}/b_{max} = qe\omega_0/m\gamma^2 v^3$. The integral, which is

$$I \equiv \int_{z_{min}}^{\infty} dz z \left(K_1^2(z) + \frac{1}{\gamma^2} K_0^2(z) \right) = \int_{z_{min}}^{\infty} dz z \left(K_1^2(z) + K_0^2(z) - \frac{v^2}{c^2} K_0^2(z) \right), \quad (36)$$

³Which is almost miraculous because we approximated $\mathbf{E}(\mathbf{x}, t)$ as $\mathbf{E}(0, t)$ which is not good when b is small; evidently, some cancellation of errors takes place.

can be done by making use of certain identities satisfied by the Bessel functions. These identities are

$$K'_\nu(x) = -K_{\nu-1}(x) - \frac{\nu}{x}K_\nu(x) \quad \text{and} \quad K'_\nu(x) = -K_{\nu+1}(x) + \frac{\nu}{x}K_\nu(x); \quad (37)$$

from them, it follows that

$$\begin{aligned} \frac{d}{dx}[xK_0(x)K_1(x)] &= -x[K_0^2(x) + K_1^2(x)] \\ \frac{d}{dx}[x^2(K_1^2(x) - K_0^2(x))] &= -2xK_0^2(x). \end{aligned} \quad (38)$$

It is now easy to do the integral; the result is

$$I = z_{min}K_0(z_{min})K_1(z_{min}) - \frac{v^2}{2c^2}z_{min}^2[K_1^2(z_{min}) - K_0^2(z_{min})]. \quad (39)$$

Now, $z_{min} = qe\omega_0/m\gamma^2v^3 \sim 10^{-7}$ or less for a relativistic particle, so we can expand the Bessel functions in the small argument limit:

$$K_1(x) \approx 1/x \quad \text{and} \quad K_0(x) \approx -[\ln(x/2) + 0.577] = \ln(1.123/x). \quad (40)$$

Thus $I = \ln(1.123/z_{min}) - v^2/2c^2$, and

$$\frac{dE}{dx} = -\frac{4\pi nq^2e^2}{mv^2} \left[\ln \left(\frac{1.123m\gamma^2v^3}{qe\omega_0} \right) - \frac{v^2}{2c^2} \right]. \quad (41)$$

This formula may be easily extended to a (slightly) more realistic form, accounting for different charges with different resonant frequencies. Assume an elemental solid with a density of atoms N , each with Z electrons. The Z electrons will be split into groups of f_j electrons distinguished by the resonant frequency of the group ω_j . The oscillator strengths f_j must satisfy the sum rule $\sum_j f_j = Z$. The groups add linearly so that

$$\frac{dE}{dx} = -4\pi nZ \frac{q^2e^2}{mv^2} \left[\ln B_c - \frac{v^2}{2c^2} \right]. \quad (42)$$

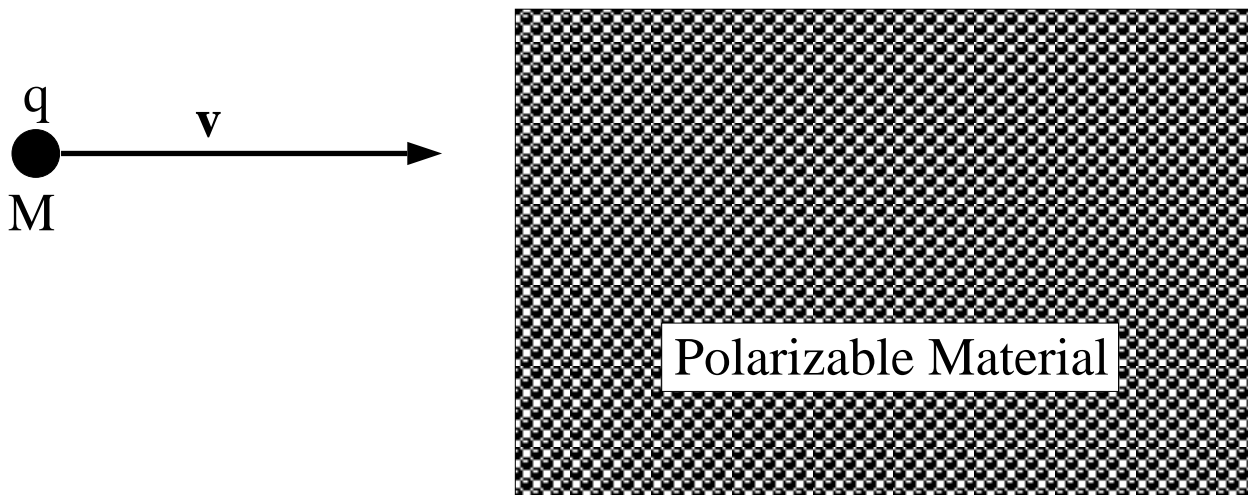
where

$$B_c = \frac{1.123m\gamma^2v^3}{qe \langle \omega \rangle} \quad Z \ln \langle \omega \rangle = \sum_j f_j \ln \omega_j \quad (43)$$

This is the classical energy-loss formula derived by Bohr in 1915. It actually works rather well despite the fact that the effects responsible for the energy loss (scattering of small objects by other small objects) really ought to be treated using quantum theory. The reason why the classical theory works as well as it does is that any macroscopic energy loss is the result of many collisions. The energy loss in each collision is not given very accurately by the classical theory, but Eq. (41) represents the energy loss over a large number of collisions, and that is pretty close to the mark. Thus the usefulness of the classical theory is in part a consequence of statistical effects. Bohr's original formula was eventually superseded by a calculation based on quantum theory and done by Bethe in 1930. Read the appropriate section in Jackson for more details.

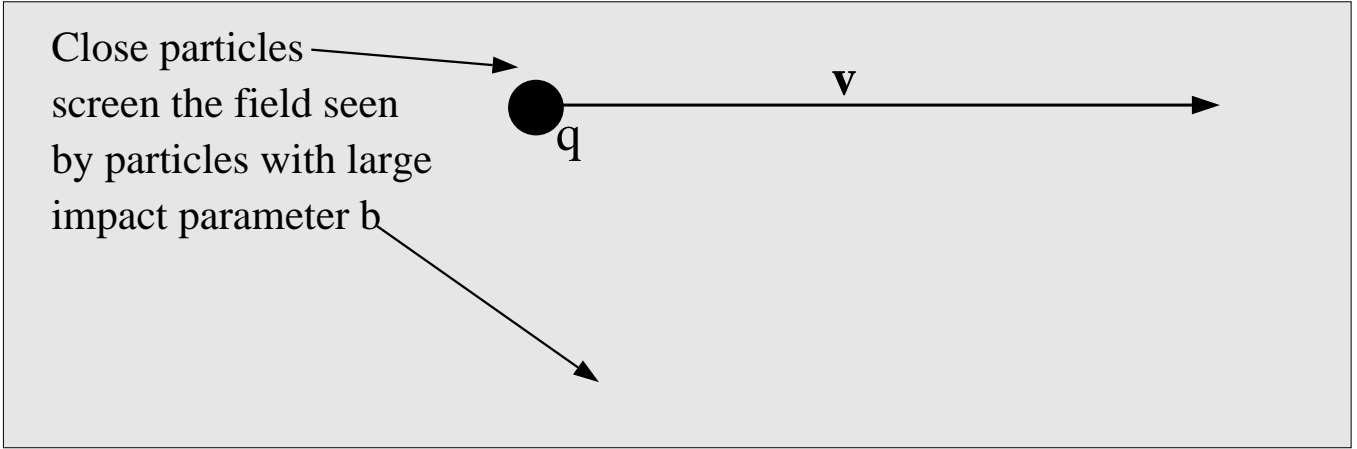
3 Density Effect in Energy Loss

A charged particle traversing a material produces a local electric polarization of that material, as a consequence of which the electric field acting on any given charge in the material is not the electric field that we used in the preceding sections.



This “screening” effect is especially important for collisions of large impact parameter b , since then the field will be screened by the charges closer to the path of the incident

particle.



Thus the energy loss formulas we derived earlier will overestimate the energy loss of a charged particle traversing a polarizable medium. As we will see, this effect is most important for fast or ultra-relativistic particles.

We can produce a calculation of the consequences of this “screening” effect using the familiar formalism of macroscopic electrodynamics. Let the material have a frequency-dependent dielectric function $\epsilon(\omega)$, as discussed in Chapter 7, so that the displacement and macroscopic electric field, expressed as functions of position and frequency, are related by

$$\mathbf{D}(\mathbf{x}, \omega) = \epsilon(\omega)\mathbf{E}(\mathbf{x}, \omega); \quad (44)$$

the connection between any field F as a function of \mathbf{x} and ω and the same field as a function of \mathbf{x} and t is

$$F(\mathbf{x}, t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d\omega F(\mathbf{x}, \omega) e^{-i\omega t} \quad (45)$$

with the inverse transformation

$$F(\mathbf{x}, \omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dt F(\mathbf{x}, t) e^{i\omega t}. \quad (46)$$

Further, let us introduce Fourier transforms in space:

$$F(\mathbf{x}, \omega) = \frac{1}{(\sqrt{2\pi})^3} \int d^3k F(\mathbf{k}, \omega) e^{i\mathbf{k}\cdot\mathbf{x}} \quad (47)$$

with the inverse

$$F(\mathbf{k}, \omega) = \frac{1}{(\sqrt{2\pi})^3} \int d^3x F(\mathbf{x}, \omega) e^{-i\mathbf{k}\cdot\mathbf{x}}. \quad (48)$$

We begin from the macroscopic Maxwell equations with $\mathbf{B} \equiv \mathbf{H}$, i.e., $\mu = 1$; the inhomogeneous equations are

$$\nabla \times \mathbf{B} = \frac{4\pi}{c} \mathbf{J} + \frac{1}{c} \frac{\partial \mathbf{D}}{\partial t} \quad \text{and} \quad \nabla \cdot \mathbf{D} = 4\pi\rho, \quad (49)$$

and the homogeneous field equations may be replaced by

$$\mathbf{B} = \nabla \times \mathbf{A} \quad \text{and} \quad \mathbf{E} = -\nabla\Phi - \frac{1}{c} \frac{\partial \mathbf{A}}{\partial t}. \quad (50)$$

which ensure that $\nabla \cdot \mathbf{B} = 0$ and that $\nabla \times \mathbf{E} = -\frac{1}{c} \partial \mathbf{B} / \partial t$. Fourier transform the Maxwell equations (49) to find

$$i\mathbf{k} \times \mathbf{B} = \frac{4\pi}{c} \mathbf{J} - i\frac{\omega}{c} \mathbf{D} \quad \text{and} \quad i\mathbf{k} \cdot \mathbf{D} = 4\pi\rho. \quad (51)$$

Similarly, from the Fourier transforms of Eqs. (50) one finds

$$\mathbf{B} = i\mathbf{k} \times \mathbf{A} \quad \text{and} \quad \mathbf{E} = -i\mathbf{k}\Phi + i\frac{\omega}{c} \mathbf{A}. \quad (52)$$

Substitution of this pair of equations into the immediately preceding ones and using $\mathbf{D} = \epsilon \mathbf{E}$, we arrive at Fourier transformed wave equations for the potentials:

$$-\mathbf{k}(\mathbf{k} \cdot \mathbf{A}) + k^2 \mathbf{A} = \frac{4\pi}{c} \mathbf{J} - \frac{\omega}{c} \epsilon \left[\mathbf{k}\Phi - \frac{\omega}{c} \mathbf{A} \right] \quad (53)$$

and

$$\epsilon \left[k^2 \Phi - \frac{\omega}{c} \mathbf{k} \cdot \mathbf{A} \right] = 4\pi\rho. \quad (54)$$

We can make the equations for \mathbf{A} and Φ separate by choosing an appropriate gauge; specifically,

$$\mathbf{k} \cdot \mathbf{A}(\mathbf{k}, \omega) = \epsilon \frac{\omega}{c} \Phi(\mathbf{k}, \omega), \quad (55)$$

which is a slightly modified form of the Lorentz gauge. Within this gauge, the equations of motion are

$$(k^2 - \epsilon \omega^2/c^2) \mathbf{A} = \frac{4\pi}{c} \mathbf{J} \quad \text{and} \quad (k^2 - \epsilon \omega^2/c^2) \Phi = 4\pi \frac{\rho}{\epsilon} \quad (56)$$

which are simple familiar⁴ wave equations.

The only macroscopic source is the incident charge⁵, so

$$\rho(\mathbf{x}, t) = q\delta(\mathbf{x} - \mathbf{v}t) \quad \text{and} \quad \mathbf{J}(\mathbf{x}, t) = \mathbf{v}\rho(\mathbf{x}, t) \quad (57)$$

where we approximate \mathbf{v} as a constant, $\mathbf{v} = v\epsilon_3$. The Fourier transforms of these source densities are

$$\begin{aligned} \rho(\mathbf{k}, \omega) &= \frac{1}{(2\pi)^2} \int d^3x dt e^{-i(\mathbf{k}\cdot\mathbf{x} - \omega t)} q\delta(\mathbf{x} - \mathbf{v}t) \\ &= \frac{q}{(2\pi)^2} \int_{-\infty}^{\infty} dt e^{-i(\mathbf{k}\cdot\mathbf{v} - \omega)t} = \frac{q}{2\pi} \delta(\omega - \mathbf{v} \cdot \mathbf{k}) \end{aligned} \quad (58)$$

and, similarly,

$$\mathbf{J}(\mathbf{k}, \omega) = \frac{q\mathbf{v}}{2\pi} \delta(\omega - \mathbf{v} \cdot \mathbf{k}). \quad (59)$$

The solutions for the Fourier-transformed potentials are trivially found:

$$\begin{aligned} \Phi(\mathbf{k}, \omega) &= \left(\frac{2q}{\epsilon}\right) \frac{\delta(\omega - \mathbf{k} \cdot \mathbf{v})}{k^2 - \epsilon \omega^2/c^2} \\ \mathbf{A}(\mathbf{k}, \omega) &= \left(\frac{2q\mathbf{v}}{c}\right) \frac{\delta(\omega - \mathbf{k} \cdot \mathbf{v})}{k^2 - \epsilon \omega^2/c^2} \end{aligned} \quad (60)$$

Now, $\mathbf{E}(\mathbf{k}, \omega) = -i\mathbf{k}\Phi(\mathbf{k}, \omega) + i(\omega/c)\mathbf{A}(\mathbf{k}, \omega)$, so

$$\mathbf{E}(\mathbf{k}, \omega) = 2iq \left(\frac{\omega\mathbf{v}}{c^2} - \frac{\mathbf{k}}{\epsilon} \right) \frac{\delta(\omega - \mathbf{k} \cdot \mathbf{v})}{k^2 - \epsilon \omega^2/c^2} \quad (61)$$

and

$$\mathbf{B}(\mathbf{k}, \omega) = i\mathbf{k} \times \mathbf{A}(\mathbf{k}, \omega) = i \left(\frac{2q}{c} \right) (\mathbf{k} \times \mathbf{v}) \frac{\delta(\omega - \mathbf{k} \cdot \mathbf{v})}{k^2 - \epsilon \omega^2/c^2}. \quad (62)$$

⁴They may not look so familiar because they are in wavenumber and frequency space.

⁵The dielectric function accounts for any sources associated with charges in the material.

Now let us compute the rate at which the incident particle loses energy by finding the flow of electromagnetic energy away from the track of this particle. Let the point at which the fields are to be evaluated be $\mathbf{x} = b\boldsymbol{\epsilon}_1$ and find $\mathbf{E}(\mathbf{x}, \omega)$ and $\mathbf{B}(\mathbf{x}, \omega)$:

$$\begin{aligned}
\mathbf{B}(\mathbf{x}, \omega) &= \frac{1}{(\sqrt{2\pi})^3} \int d^3k e^{i\mathbf{k}\cdot\mathbf{x}} i \left(\frac{2q}{c} \right) (\mathbf{k} \times \mathbf{v}) \frac{\delta(\omega - \mathbf{k} \cdot \mathbf{v})}{k^2 - \epsilon \omega^2/c^2} \\
&= \frac{(2iq/c)}{(2\pi)^{3/2}} \int d^3k e^{ibk_1} (k_2\boldsymbol{\epsilon}_1 - k_1\boldsymbol{\epsilon}_2) v \frac{\delta(\omega - k_3v)}{k^2 - \epsilon \omega^2/c^2} \\
&= -\boldsymbol{\epsilon}_2 \frac{(2iq/c)}{(2\pi)^{3/2}} \int dk_1 dk_2 k_1 e^{ibk_1} \left/ \left[k_1^2 + k_2^2 + \frac{\omega^2}{v^2} \left(1 - \epsilon \frac{v^2}{c^2} \right) \right] \right. \quad (63)
\end{aligned}$$

Set $\lambda^2 = (\omega/v)^2(1 - \epsilon v^2/c^2)$ and $\beta = v/c$. Then

$$\begin{aligned}
\mathbf{B}(\mathbf{x}, \omega) &= -i\boldsymbol{\epsilon}_2 \frac{2q}{c(2\pi)^{3/2}} \int dk_1 dk_2 \frac{k_1 e^{ibk_1}}{k_1^2 + k_2^2 + \lambda^2} = -i\boldsymbol{\epsilon}_2 \frac{q}{c\sqrt{2\pi}} \int dk_1 \frac{k_1 e^{ibk_1}}{\sqrt{k_1^2 + \lambda^2}} \\
&= -\boldsymbol{\epsilon}_2 \frac{q}{c\sqrt{2\pi}} \frac{d}{db} \left(\int dk_1 \frac{e^{ibk_1}}{\sqrt{k_1^2 + \lambda^2}} \right) = -\boldsymbol{\epsilon}_2 \frac{q}{c} \sqrt{\frac{2}{\pi}} \frac{d}{db} \left(\int_0^\infty dx \frac{\cos(b\lambda x)}{\sqrt{1+x^2}} \right) \\
&= -\boldsymbol{\epsilon}_2 \frac{q}{c} \sqrt{\frac{2}{\pi}} \frac{d}{db} [K_0(b\lambda)] = \boldsymbol{\epsilon}_2 \frac{q}{c} \sqrt{\frac{2}{\pi}} \lambda K_1(b\lambda). \quad (64)
\end{aligned}$$

Similarly,

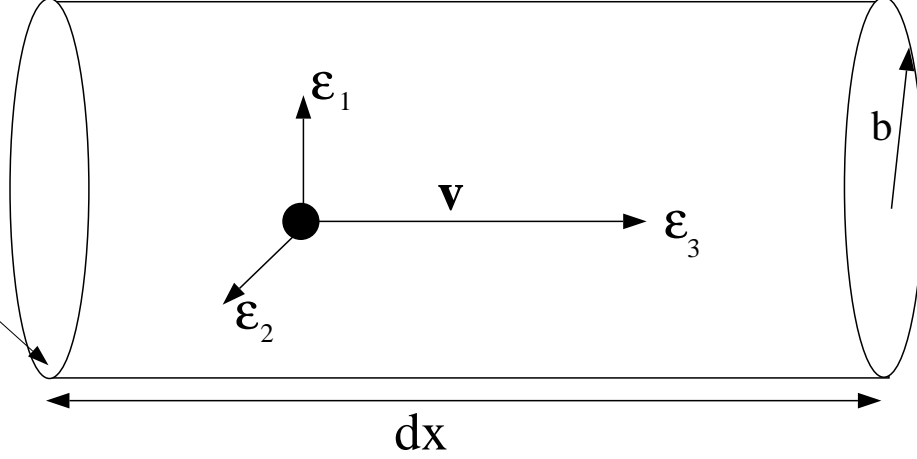
$$\mathbf{E}(\mathbf{x}, \omega) = \boldsymbol{\epsilon}_1 \frac{q}{v} \sqrt{\frac{2}{\pi}} \frac{\lambda}{\epsilon} K_1(b\lambda) - i\boldsymbol{\epsilon}_3 \frac{q\omega}{v^2} \sqrt{\frac{2}{\pi}} \left(\frac{1}{\epsilon} - \beta^2 \right) K_0(b\lambda). \quad (65)$$

Next, for real ϵ , λ^2 may be positive or negative depending on whether the incident particle moves more slowly or more rapidly than the speed of light in the medium, $c' = c/\sqrt{\epsilon}$. For $v < c'$, $\lambda^2 > 0$, λ is real, and \mathbf{E} reduces to our previous result except for the appearance of ϵ here and there. It is then a straightforward matter to calculate $\Delta E(b)$ by the same procedure as before, assuming⁶ the field acting on a target particle is the same as the macroscopic field.

Rather than reproducing the previous calculation, let's look at an alternative: we shall calculate the radial outward part (ρ component) of the Poynting vector at $\mathbf{x} = \rho\boldsymbol{\rho}$.

⁶A risky assumption.

calculate
radiation
on this
cylinder



When this component of \mathbf{S} is integrated over all time and over a closed loop of radius b around the path of the particle, the result is the total electromagnetic field energy which flows away from the particle, per unit length of path, and at distance b from the path. Letting this energy be E_f , to distinguish it from the energy change of anything else (such as the incident particle), we have

$$\left(\frac{dE_f}{dz}\right)_{\rho=b} = \frac{c}{4\pi} 2\pi b \int_{-\infty}^{\infty} dt (\mathbf{E} \times \mathbf{B}) \cdot \mathbf{n} \quad (66)$$

Given the geometry introduced earlier, the quantity $(\mathbf{E} \times \mathbf{B}) \cdot \mathbf{n}$ is just $-E_3 B_2$.

Let's complete the integral:

$$\begin{aligned} \left(\frac{dE_f}{dx}\right)_{\rho=b} &= -\frac{cb}{2} \int_{-\infty}^{\infty} dt B_2(t) E_3(t) = -\frac{cb}{4\pi} \int_{-\infty}^{\infty} dt d\omega d\omega' B_2(\omega') E_3(\omega) e^{-i(\omega+\omega')t} \\ &= -\frac{cb}{2} \int_{-\infty}^{\infty} d\omega B_2(-\omega) E_3(\omega) = -\frac{cb}{2} \int_{-\infty}^{\infty} d\omega E_3(\omega) B_2^*(\omega) \\ &= -cb \Re \left[\int_0^{\infty} d\omega B_2^*(\omega) E_3(\omega) \right] \\ &= -\frac{2cbq^2}{\pi v^2} \Re \left[\int_0^{\infty} d\omega (-i\omega) \left(\frac{1}{\epsilon} - \beta^2\right) K_0(b\lambda) \frac{1}{c} \lambda^* K_1(b\lambda^*) \right] \\ &= \frac{2q^2}{\pi v^2} \Re \left[\int_0^{\infty} d\omega (i\omega \lambda^* b) \left(\frac{1}{\epsilon} - \beta^2\right) K_1(b\lambda^*) K_0(b\lambda) \right], \end{aligned} \quad (67)$$

an expression first derived by Enrico Fermi.

In order for the integral to have a real part, either λ or ϵ must be complex. If ϵ is real, then λ can still be complex if $\epsilon\beta^2 > 1$ meaning that the particle is travelling faster than the speed of light in the material. In this case one finds the phenomenon of Cherenkov radiation which we shall discuss presently.

For now, let us look at the case of complex ϵ . Introduce the frequency-dependent polarization $\mathbf{P}(\mathbf{x}, \omega)$ via the relation

$$\mathbf{D}(\mathbf{x}, \omega) = \mathbf{E}(\mathbf{x}, \omega) + 4\pi\mathbf{P}(\mathbf{x}, \omega); \quad (68)$$

In a linear medium such as we are considering, $\mathbf{P}(\mathbf{x}, \omega) = \chi(\omega)\mathbf{E}(\mathbf{x}, \omega)$ with $\chi(\omega) = (\epsilon(\omega) - 1)/4\pi$. The frequency dependent polarization is just the Fourier transform in time of the usual polarization $\mathbf{P}(\mathbf{x}, t)$. If we calculate it using the damped harmonic oscillator model introduced above and in chapter 7, we find

$$\mathbf{P}(\omega) = \frac{ne^2}{m} \frac{\mathbf{E}(\omega)}{\omega_0^2 - \omega^2 - i\omega\Gamma} \quad (69)$$

where n is the electron density in the material; the corresponding dielectric function is

$$\epsilon(\omega) = 1 + \frac{\omega_p^2}{\omega_0^2 - \omega^2 - i\omega\Gamma} \quad (70)$$

where ω_p is the plasma frequency, $\omega_p^2 = 4\pi ne^2/m$.

Now we have an expression for $\epsilon(\omega)$ based on a simple model. We need to do the integral presented in Eq. (67). Unfortunately that cannot be done in terms of simple functions so we shall approximate the integral in a physically reasonable way. The important range of ω should be $\omega \sim \omega_0$ so that $b\lambda \sim b\omega/v \sim b(\omega_0/v) \ll 1$ for b less than about an atomic size and $v \sim c$; ω_0 is a typical atomic energy. Thus we make the small argument approximations

$$b\lambda^* K_1(b\lambda^*) \approx b\lambda^* \frac{1}{b\lambda^*} = 1 \quad (71)$$

and

$$K_0(b\lambda) \approx \ln(1.123/b\lambda) \quad (72)$$

which leads to

$$\left(\frac{dE_f}{dx}\right)_{\rho=b} = \frac{2q^2}{\pi v^2} \Re \left[\int_0^\infty d\omega i\omega \left(\frac{1}{\epsilon} - \beta^2\right) \ln\left(\frac{1.123}{b\lambda}\right) \right]. \quad (73)$$

Because we just fouled up the integrand in the region $\omega \gg \omega_0$, we had best make sure that no contribution comes from this region of frequency; physically, we believe this should be the case. Since $\epsilon \rightarrow 1$ sufficiently rapidly here (something that should be checked to be sure our belief), we can guarantee convergence of the integral by approximating β^2 with 1. Then

$$\left(\frac{dE_f}{dx}\right)_{\rho=b} = \frac{2q^2}{\pi v^2} \Re(I) \quad (74)$$

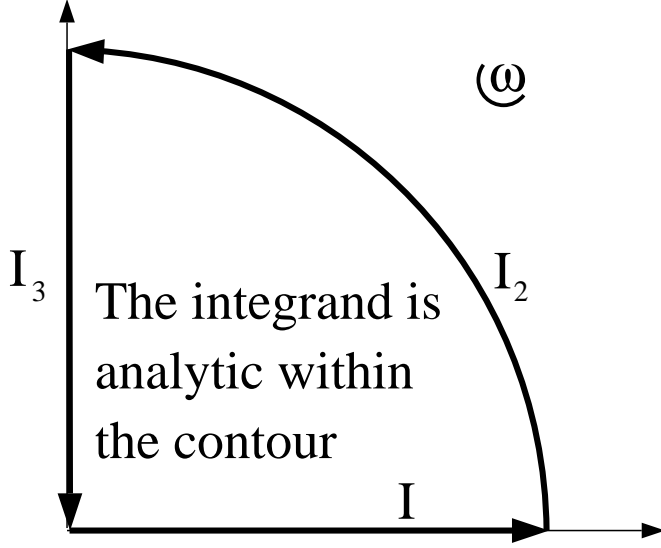
where

$$I = \int_0^\infty d\omega i\omega \left[\ln\left(\frac{1.123c}{\omega b}\right) - \frac{1}{2} \ln(1 - \epsilon) \right] \left(\frac{1 - \epsilon}{\epsilon}\right). \quad (75)$$

Using Eq. (68) for $\epsilon(\omega)$, we have

$$I = i \int_0^\infty d\omega \omega \left(\frac{-\omega_p^2}{\omega_0^2 + \omega_p^2 - \omega^2 - i\omega\Gamma} \right) \left[\ln\left(\frac{1.123c}{\omega_p b}\right) - \ln \omega + \frac{1}{2} \ln(\omega^2 - \omega_0^2 + i\omega\Gamma) \right] \quad (76)$$

We can employ the Cauchy theorem to evaluate this integral by closing the contour around the first quadrant; that is, construct a closed path by adding a quarter-circle from a point where ω is large and real to one where it is large and imaginary and then coming down the positive imaginary- ω axis to the origin.



The total integral around this contour is zero because there are no poles of the integrand within it. This point is clarified by looking for the zeroes of the integrand's denominator and by looking for the zeroes of the logarithm's argument. They are located at points in the lower half plane and so are well away from the interior of the contour.

The integral along the imaginary-frequency axis is, with $\omega = i\Omega$, Ω real,

$$\begin{aligned}
 I_3 &= -i \int_0^\infty id\Omega \frac{-i\Omega\omega_p^2}{\Omega^2 + \omega_0^2 + \omega_p^2 + \Omega\Gamma} \left[\ln\left(\frac{1.123c}{b\omega_p}\right) - \ln(i\Omega) + \frac{1}{2} \ln[-(\Omega^2 + \omega_0^2 + \Omega\Gamma)] \right] \\
 &= i \int_0^\infty d\Omega \frac{-\Omega\omega_p^2}{\Omega^2 + \omega_0^2 + \omega_p^2 + \Omega\Gamma} \left[\ln\left(\frac{1.123c}{b\omega_p}\right) - \ln\Omega + \frac{1}{2} \ln(\Omega^2 + \omega_0^2 + \Omega\Gamma) \right] \quad (77)
 \end{aligned}$$

which is pure imaginary, meaning that $\Re(I_3) = 0$. The integral over the quarter-circle, I_2 , is thus such that $-\Re(I_2) = \Re(I)$, or, letting $\omega = \Omega \exp(i\theta)$ on the quarter-circle,

$$\begin{aligned}
 \Re(I) &= -\Re \int_0^{\pi/2} i\Omega e^{i\theta} i\Omega e^{i\theta} d\theta \left(\frac{-\omega_p^2}{\omega_0^2 + \omega_p^2 - \Omega^2 e^{2i\theta} - i\Omega e^{i\theta}\Gamma} \right) \\
 &\quad \times \left[\ln\left(\frac{1.123c}{b\omega_p}\right) - \ln(\Omega e^{i\theta}) + \frac{1}{2} \ln(\Omega^2 e^{2i\theta} - \omega_0^2 + i\Omega\Gamma e^{i\theta}) \right] \\
 &= \omega_p^2 \Re \int_0^{\pi/2} d\theta \left[\ln\left(\frac{1.123c}{b\omega_p}\right) + \mathcal{O}\left(\frac{\Gamma}{\Omega}\right) \right] = \omega_p^2 \frac{\pi}{2} \ln\left(\frac{1.123c}{b\omega_p}\right). \quad (78)
 \end{aligned}$$

Hence,

$$\left(\frac{dE_f}{dz}\right)_{\rho=b} = \left(\frac{q^2\omega_p^2}{c^2}\right) \ln\left(\frac{1.123c}{b\omega_p}\right); \quad (79)$$

The negative of this quantity is the energy loss of the incident particle per unit distance traveled.

This result is to be compared with the one we found before taking screening into account,

$$\left(\frac{dE}{dz}\right)_{\rho>b} = -\frac{q^2\omega_p^2}{c^2} \left[\ln\left(\frac{1.123\gamma c}{b\omega_0}\right) - \frac{1}{2} \right]. \quad (80)$$

The two differ significantly in principle, if not numerically. In particular, the dependence of our original formula on the specific natural frequency of the target, ω_0 , is gone, replaced by a dependence on ω_p which depends only on the density of the target electrons. Also, a factor of γ has, in our most recent result, disappeared from the argument of the logarithm, meaning that the energy loss by highly relativistic charged particles is much reduced by the screening effect.

4 Cherenkov Radiation

We are also in a position to calculate energy loss by Cherenkov radiation which is something that takes place when the incident particle's speed exceeds the speed of light in the medium. We can avoid the mechanism just discussed and so isolate the Cherenkov radiation mechanism by letting ϵ be real (no damping). In this approximation we will also miss the attenuation of the radiation. Under these conditions, and as discussed in the last section, the only way to get any radiation is if

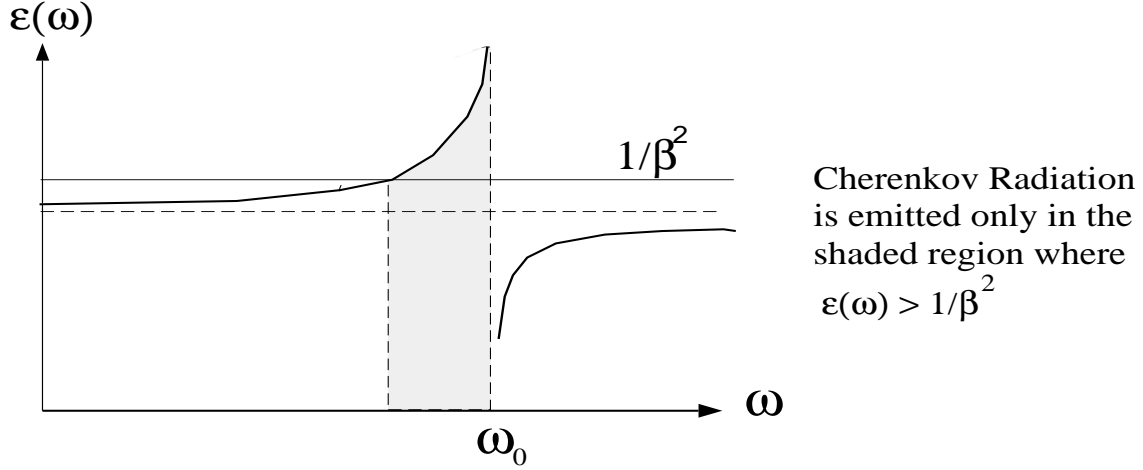
$$\lambda = \frac{\omega}{v} \sqrt{1 - \epsilon(\omega)\beta^2} \in \mathbb{C}, \quad (81)$$

or, more correctly, λ must be imaginary. We must have $v^2 > c^2/\epsilon$ or there will be no radiation. Since $n = \sqrt{\epsilon}$, then $c/\sqrt{\epsilon}$ is the speed of light in the medium, and thus the condition for radiation is that the particle exceed the speed of light in the medium.

This will not happen for all frequencies. By assuming a simple model dielectric function

$$\epsilon(\omega) = 1 + \frac{\omega_p^2}{\omega_0^2 - \omega^2}, \quad (82)$$

and expressing the condition as $\epsilon(\omega) > 1/\beta^2$ we can see that the radiation tends to be emitted near regions of anomalous dispersion.



Under these conditions, we evaluate the fields which are present at distance b from the axis of the incident particle, using b large enough that we can make simple approximations to the Bessel functions, $b|\lambda| \gg 1$. Then

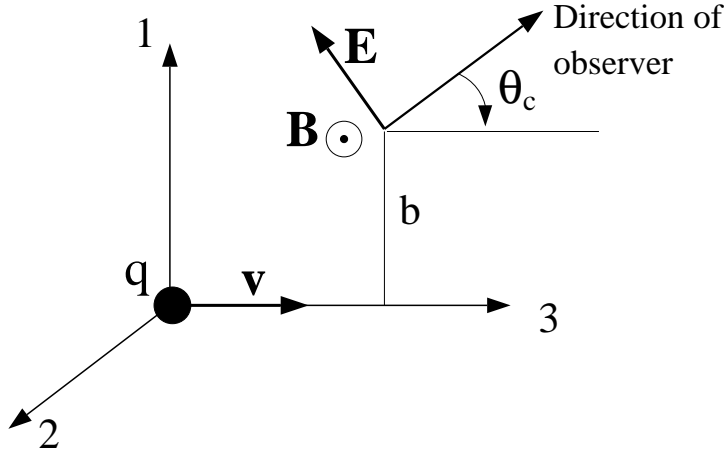
$$K_0(\lambda b) \approx K_1(\lambda b) \approx \sqrt{\frac{\pi}{2\lambda b}} e^{-\lambda b} \quad (83)$$

and so,

$$\mathbf{B}(\mathbf{x}, \omega) = \epsilon_2 \frac{q}{c} \sqrt{\frac{\lambda}{b}} e^{-\lambda b}. \quad (84)$$

Similarly,

$$\mathbf{E}(\mathbf{x}, \omega) = \epsilon_1 \frac{q}{\epsilon v} \sqrt{\frac{\lambda}{b}} e^{-\lambda b} - i\epsilon_3 \frac{q\omega}{v^2} \frac{1}{\sqrt{\lambda b}} \left(\frac{1}{\epsilon} - \beta^2 \right) e^{-\lambda b}. \quad (85)$$



Then from Eq. (67) and the equations above, the field energy passing through the cylinder of radius b per unit length is

$$\left(\frac{dE}{dz}\right)_C = \frac{2q^2}{\pi v^2} \Re \left[\int_0^\infty d\omega i\omega \frac{\pi}{2} \sqrt{\frac{\lambda^*}{\lambda}} e^{-(\lambda+\lambda^*)a} \left(\frac{1}{\epsilon} - \beta^2\right) \right]. \quad (86)$$

The wonderful thing that happens when λ is pure imaginary is that the exponential functions have imaginary arguments and will not become small as b becomes large. Thus we find the energy given off as *Cherenkov radiation* to be

$$\left(\frac{dE}{dz}\right)_C = \frac{q^2}{v^2} \Re \left[\int d\omega i\omega \sqrt{-1} \left(\frac{1 - \epsilon\beta^2}{\epsilon}\right) \right] = \frac{q^2}{c^2} \int d\omega \omega \left(1 - \frac{1}{\epsilon\beta^2}\right) \quad (87)$$

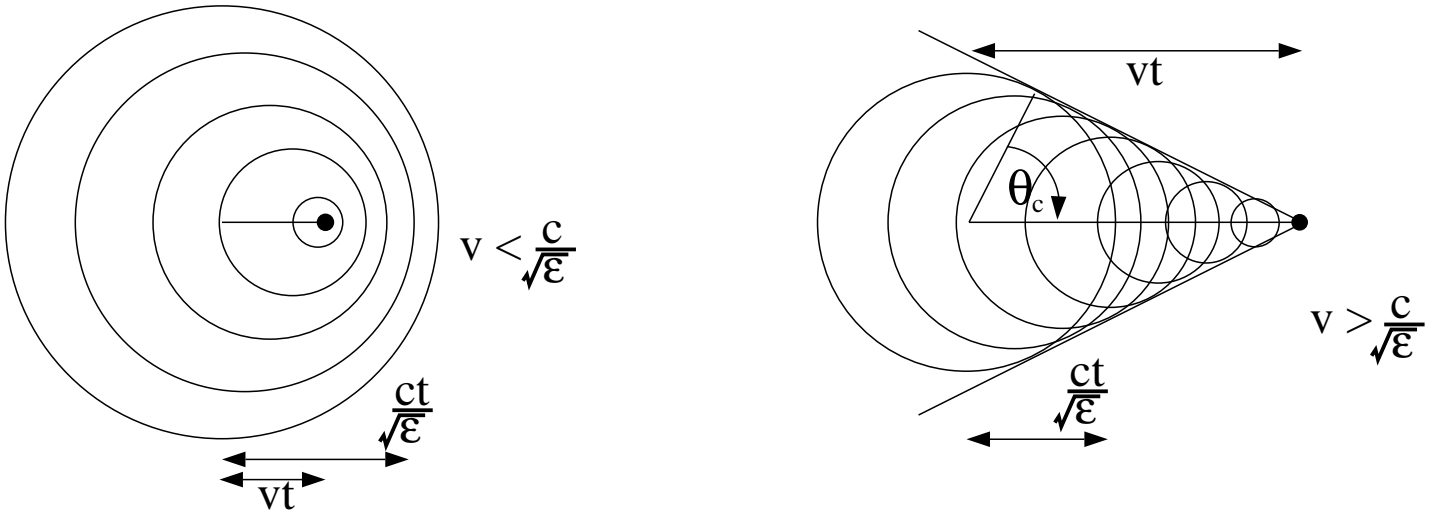
where the integration extends over only those frequencies $\epsilon\beta^2 > 1$. One can see that this is indeed radiative energy loss because it is independent of b provided only b is large enough that the Bessel functions are well-represented by their large-argument forms. In this respect it is quite distinct from the energy loss by transfer of energy to other charged particles that we studied earlier (real as opposed to virtual photons). We were able to treat that energy loss by examining the energy carried by the electromagnetic fields because the mechanism by which the energy is transferred from one particle to another is by means of the fields; in effect, we did that calculation in such a way as to “intercept” the energy that was on its way from one charge to another.

From the picture above it is clear that the radiation is completely linearly polarized in the plane containing the observer and the path of the particle. In addition the angle θ_c of emission of Cherenkov radiation relative to the direction ϵ_3 of the particle's velocity is given by

$$\cos(\theta_c) = \frac{E_1}{\sqrt{E_1^2 + E_3^2}} = \frac{c/n}{v} \quad (88)$$

where $n = \sqrt{\epsilon}$. Thus the condition that λ be complex, and thus that required for Cherenkov radiation, can be rephrased as the requirement that θ_c be a physical angle with a cosine less than unity.

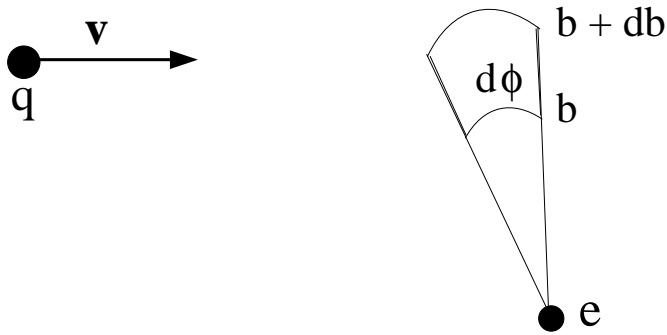
As shown in the picture below, the emission angle θ_c can also be interpreted in terms of a shock wave angle.



5 Momentum Transfer

The final topic we shall study in this chapter is the deflection of the incident particle produced by scattering from the particles in the material through which it moves. The targets mainly responsible for the deflection turn out to be the highly charged ones—the nuclei.

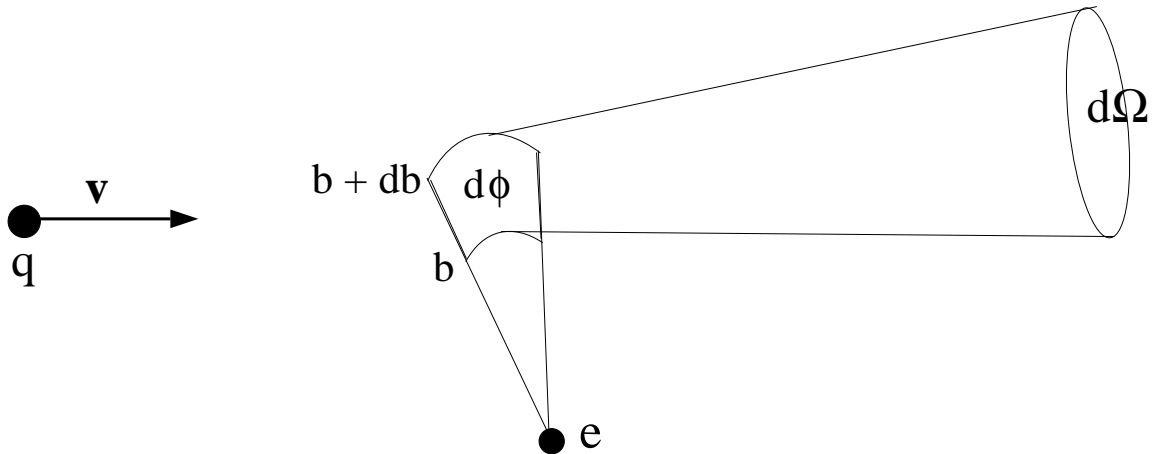
We start by introducing the number of particles incident per unit time on the target with an impact parameter between b and $b + db$ and at an azimuthal angle between ϕ and $\phi + d\phi$.



If the incident beam has a particle number density n and a speed v , then the incident flux is nv particles per unit area per unit time, and the number incident in the area element just described is

$$d^2N = nvbdbd\phi. \quad (89)$$

Now, given a smoothly varying scattering potential, these particles will, after scattering, show up in some element of solid angle $d\Omega$.



Hence we can write that

$$d^2N = N'd\Omega \quad (90)$$

where N' is the number of particles scattered into unit solid angle in unit time and $d\Omega$ is the element of solid angle into which the particular d^2N particles under consideration are scattered. Using Eq. (89), we have

$$nvbd\phi db = N'd\Omega \quad \text{or} \quad bd\phi db = \frac{N'}{nv}d\Omega \quad (91)$$

The quantity N' is proportional to the incident particle flux; that is, the number of particles per unit solid angle that come out in some given direction is directly proportional to the incident flux. Hence a more intrinsic measure of the scattering than N' is provided by the quantity N'/nv , the *differential scattering cross-section* $d\sigma/d\Omega$:

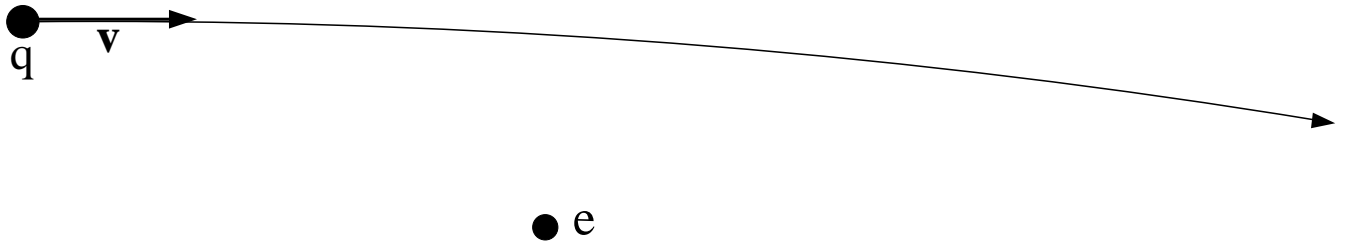
$$\frac{d\sigma}{d\Omega} \equiv \frac{N'}{nv} \quad (92)$$

Making this substitution in Eq. (91), we get

$$bd\phi db = \frac{d\sigma}{d\Omega}d\phi \sin\theta d\theta \quad (93)$$

We will also assume that the potential between the incident particle and the scatterer is central. In this case we have azimuthal symmetry so the particles incident on the target in some increment $d\phi$ of azimuthal angle around ϕ are scattered into the same element of azimuthal angle,

Scattering from a central potential occurs within one plane thus ϕ is unchanged



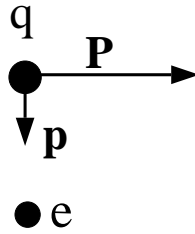
thus we find

$$b db = \frac{d\sigma}{d\Omega} \sin\theta d\theta \quad \text{or} \quad \frac{d\sigma}{d\Omega} = \frac{b}{\sin\theta} \left| \frac{db}{d\theta} \right| \quad (94)$$

where θ is the angle by which the particle is deflected or scattered.

The differential scattering cross-section, by its definition, has dimensions of length squared or area. We can evaluate it if we have an equation relating b and θ . In the impulse approximation, the scattering angle θ is given by ratio of the momentum transfer to the incident momentum; and that is, from Eq. (2),

$$|\theta| = \frac{p}{P} = \left| \frac{2qe}{Pvb} \right| \quad (95)$$



where, in this equation, $\mathbf{P} = \gamma M \mathbf{v}$ is the momentum of the incident particle, and $p = \frac{2|qe|}{bv}$ (Eq. (2)) is the momentum transfer from the incident particle to the target. From this relation we can evaluate $|d\theta/db|$ and find that the cross-section is

$$\frac{d\sigma}{d\Omega} = \frac{b}{\sin \theta} \left| \frac{Pvb^2}{2qe} \right| = \frac{Pv}{2qe \sin \theta} \left(\frac{2qe}{Pv\theta} \right)^3 = \left(\frac{2qe}{Pv} \right)^2 \frac{1}{\theta^4} \quad (96)$$

where we make the small angle approximation $\theta \approx \sin \theta$ which is valid anywhere that the impulse approximation is valid. In this, the small-angle regime, our result matches the Rutherford scattering cross-section.

From Eq. (96) we can see that nuclei are more effective than electrons at producing a given deflection θ . The charge e that appears in the cross-section is the charge of the target, a holdover from when we let the target be an electron. More generally, replace this charge by ze , in case the target is, *e.g.*, a nucleus.

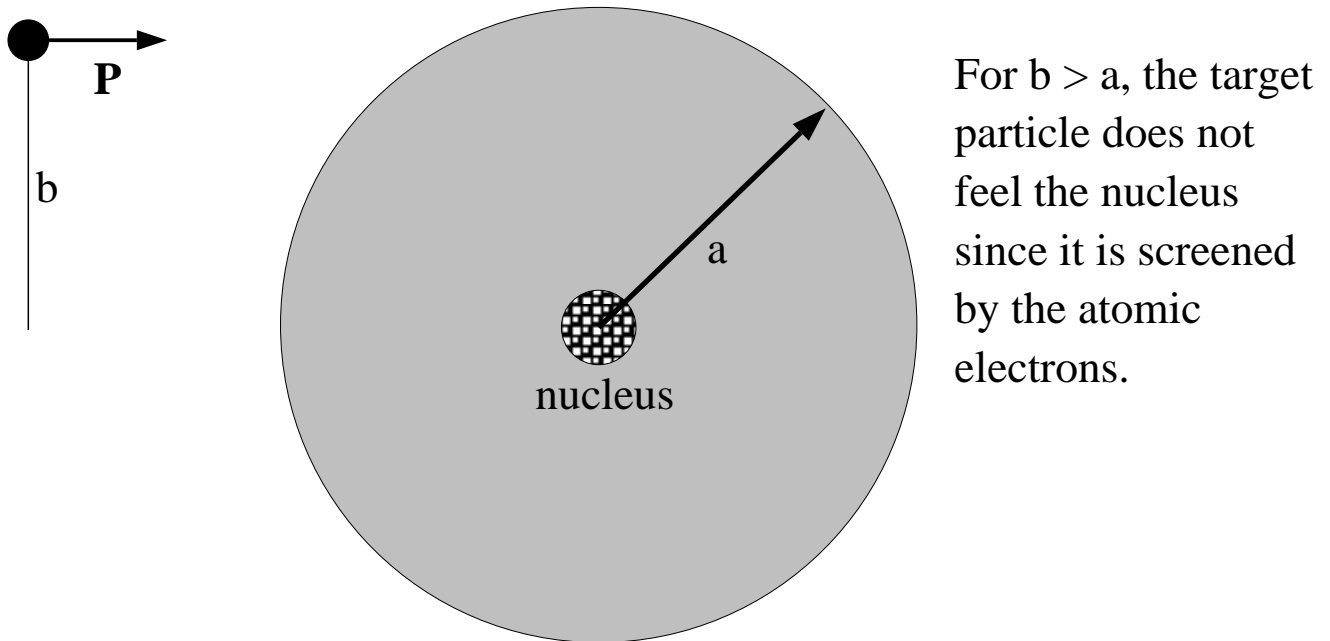
$$\frac{d\sigma}{d\Omega} = \frac{b}{\sin \theta} \left| \frac{Pvb^2}{2qze} \right| = \frac{Pv}{2qze \sin \theta} \left(\frac{2qze}{Pv\theta} \right)^3 = \left(\frac{2qze}{Pv} \right)^2 \frac{1}{\theta^4} \quad (97)$$

One can then see that the cross-section is proportional to z^2 , meaning that a nucleus is more effective by a factor of z^2 at producing a given angle of deflection θ . At the

same time there are z times as many electrons, leading to z times as many scattering events. This is not enough to offset the larger cross-section produced by the nuclei, and therefore they are the dominant scatterers where deflection of the incident particle is concerned.

5.1 Average Angle of Deflection

Of course the target is rarely composed of a single atom. Rather, we generally scatter from a molecular solid, or material. Here, we want to calculate a typical or average angle of deflection produced in a scattering event. That will require integrating over θ using $d\sigma/d\Omega$ as the distribution function. Cutoffs on the integration must be introduced. At small θ , corresponding to large b , the cutoff is determined by the condition $b_{max} \sim a$ where a is an atomic size.



The reason is that for $b > a$, the incident particle passes completely outside of the electronic shell surrounding the nucleus and so the interaction between the incident

particle and nucleus is almost completely screened. Thus

$$\theta_{min} \approx \left| \frac{qze}{Pvb_{max}} \right| \approx \left| \frac{qze}{Pva} \right|. \quad (98)$$

This still leaves a large range of impact parameter b , since the nuclear radius $\sim 10^{-13}$ cm. and that of a typical atomic radius is $\sim 10^{-8}$ cm.. An alternative, quantum-based argument can be made for choosing $\theta_{min} \sim \hbar/pa$. There is also a maximum scattering angle which is not of much significance in the present context; we may suppose that θ_{max} is of order one.

Given appropriate cutoffs, we can determine the mean value of θ^2 in scattering events. Using the small-angle approximation for all trigonometric functions, we have

$$\begin{aligned} \langle \theta^2 \rangle &= \frac{\int d\Omega \theta^2 [d\sigma/d\Omega]}{\int d\Omega [d\sigma/d\Omega]} \approx \frac{\int_{\theta_{min}}^{\theta_{max}} d\theta/\theta}{\int_{\theta_{min}}^{\theta_{max}} d\theta/\theta^3} \\ &= \frac{2 \ln(\theta_{max}/\theta_{min})}{1/\theta_{min}^2 - 1/\theta_{max}^2} \approx 2\theta_{min}^2 \ln(\theta_{max}/\theta_{min}). \end{aligned} \quad (99)$$

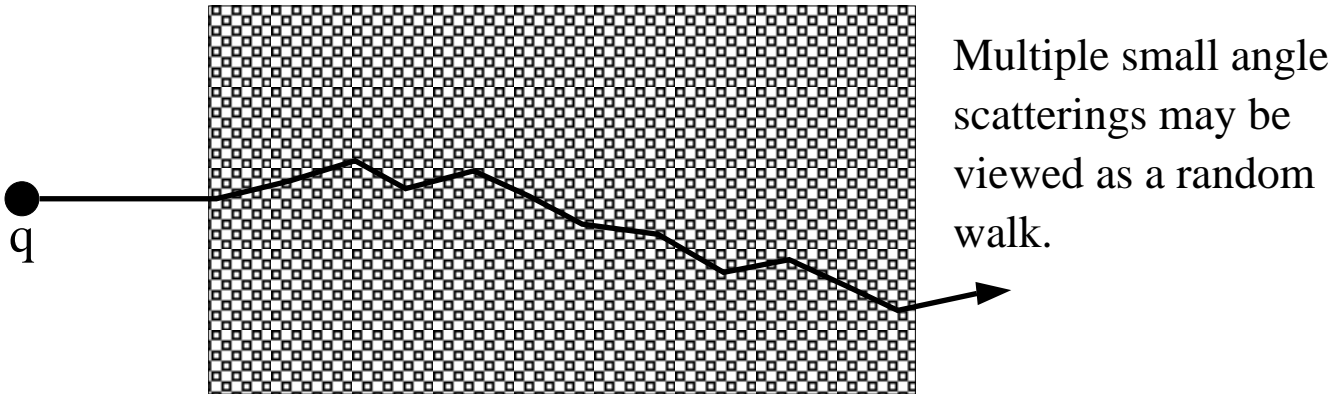
This result is some not-very-large multiple⁷ of θ_{min}^2 . Hence, a single scattering event cannot be expected to deflect the incident particle very much.

A sizable net deflection can be obtained in two quite different ways. One is that a large number of small-angle scatterings can result in a large deflection. The other is that a single large-angle scattering, though rare, can occur. If one bombards a thin slab of target material with a beam of particles, then what one finds is that most of the particles which come through will have experienced a large number of small-angle scatterings and no large-angle scatterings. These will have a distribution of net scattering angles which reflects their experience (many small-angle scatterings). Some particles, however, will have experienced a large-angle scattering in addition to the many small-angle scatterings. They will have a distribution of scattering angles which reflects their experience and which will be quite unlike the distribution of the particles which experience only small-angle scatterings. Let's give each of these possibilities a little further thought.

⁷Because the cross-section is strongly peaked at small angles.

5.1.1 Distribution of Small Angle Scattering

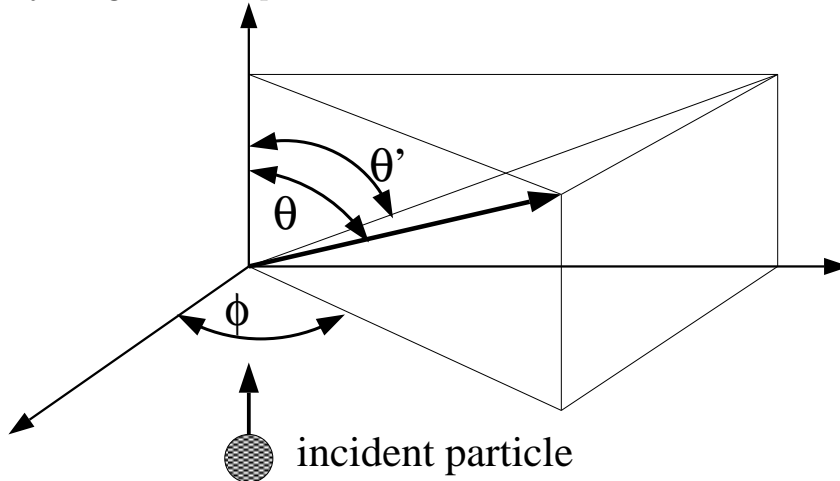
If the particle experiences only a large number of small-angle scattering events, its deflection will resemble a random walk.



A collection of such random walkers will provide a distribution of observed scattering angles which will have approximately a Gaussian form,

$$P(\theta) \sim e^{-\theta^2 / \langle \Theta^2 \rangle}, \quad (100)$$

where $\langle \Theta^2 \rangle$ is the width of the distribution. To carry the analysis further in a quantitative manner, let's make the random walk effectively one-dimensional by projecting it onto a plane.



Consider a particle that is scattered into the direction (θ, ϕ) ; project this direction onto the $y-z$ plane where it becomes θ' with $\theta' = \theta \sin \phi$ for $\theta \ll 1$. Hence $\theta'^2 =$

$\theta^2 \sin^2 \phi$, and the observed mean value of θ'^2 in single scattering events is

$$\langle \theta'^2 \rangle = \frac{\int d\Omega \theta'^2(\theta, \phi) [d\sigma/d\Omega]}{\int d\Omega [d\sigma/d\Omega]} = \frac{1}{2} \langle \theta^2 \rangle. \quad (101)$$

Also, $\langle \theta' \rangle = 0$. Assuming that the scattering directions produced by the different collisions that any one particle suffers are independent, and that there are many such collisions, then, from the theory of the elementary one-dimensional random walk, the normalized distribution of observed net scattering angles θ' is well-approximated by a Gaussian

$$P_m(\theta') = \frac{1}{\sqrt{\pi \langle \Theta^2 \rangle}} e^{-\theta'^2 / \langle \Theta^2 \rangle} \quad (102)$$

with the random walk distribution width

$$\langle \Theta^2 \rangle = N \langle \theta'^2 \rangle, \quad (103)$$

where N is the mean number of collisions experienced by each particle in traversing the material. If the total cross-section is σ , the density of scatterers is n , and the thickness of the slab is a , then $N = n\sigma a$ and so

$$\langle \Theta^2 \rangle = n\sigma a \langle \theta'^2 \rangle. \quad (104)$$

For our particular cross-section Eq. (97), $\sigma = \pi(2qze/Pv)^2/\theta_{min}^2$, so, using also Eq. (99),

$$\langle \Theta^2 \rangle = 2\pi n \left(\frac{2qze}{Pv} \right)^2 a \ln(\theta_{max}/\theta_{min}). \quad (105)$$

5.1.2 The Distribution of Large Angle Scattering

This distribution may be contrasted with the one that arises for particles which undergo a single large-angle scattering and many small-angle ones. If the net effect of the latter is less than the deflection produced by the former, which in some sense defines what we mean by a large-angle scattering, then we need only consider the distribution produced by a single large-angle event. The number of such events is

proportional to the cross-section or, for⁸ $\theta \ll 1$,

$$d\sigma = \frac{d\sigma}{d\Omega} d\Omega = \left(\frac{2qze}{pv} \right)^2 \frac{1}{\theta^4} d\phi \theta d\theta. \quad (106)$$

We may convert θ to θ' using $\theta = \theta' / \sin \phi$,

$$d\sigma = \left(\frac{2qze}{pv} \right)^2 \frac{d\theta'}{\theta'^3} \sin^2 \phi d\phi. \quad (107)$$

Now integrate ϕ from zero to π to pick up all events corresponding to $\theta' > 0$. The result is that⁹

$$d\sigma = \frac{\pi}{2} \left(\frac{2qze}{pv} \right)^2 \frac{d\theta'}{\theta'^3}. \quad (108)$$

For a slab of thickness a with a density n of scatterers, the probability of having a single large-angle scattering in an interval $d\theta'$ around θ' is

$$P_s(\theta') d\theta' = nad\sigma = \frac{\pi}{2} na \left(\frac{2qze}{pv} \right)^2 \frac{d\theta'}{\theta'^3}. \quad (109)$$

Because this distribution falls off only as θ'^{-3} while the multiple-scattering distribution falls off exponentially as θ'^2 , there is some angle θ_0 such that the single-scattering distribution is larger than the multiple scattering one for $\theta' > \theta_0$ and conversely.

Roughly speaking, the total distribution of scattered particles as a function of θ' is just P_m for $\theta' < \theta_0$ and P_s for $\theta' > \theta_0$. In any given system, one can easily compute the two distributions along with θ_0 . It is expected that the description will work quite well for θ' significantly smaller than θ_0 and also for θ' significantly larger. For $\theta' \approx \theta_0$, the actual behavior is complicated considerably by the contribution of particles that have undergone several scatterings through “almost-large” angles. There are not enough such scattering events per particle for them to be properly treated using statistical methods, and they are not easily treated in any other way, except for numerical simulations.

⁸Evidently, “large-angle” means an angle large compared to θ_{min} ; it does not mean an angle so large as to be of order one.

⁹The original $d\sigma$ is a second-order differential; the result of integrating over ϕ , unfortunately still called $d\sigma$, is a first-order differential.

Chapter Fourteen

Radiation by Moving Charges

Sir Joseph Larmor
(1857 - 1942)

September 17, 2001

Contents

1	Liénard-Wiechert potentials	1
2	Radiation from an Accelerated Charge; the Larmor Formula	9
2.1	Relativistic Larmor Formula	12
2.1.1	Example: Synchrotron	14
2.1.2	Example: Linear Acceleration	15
3	Angular distribution of radiation	17
3.1	Example: Parallel acceleration and velocity	17
3.2	Example: Acceleration Perpendicular to Velocity	20
3.3	Comparison of Examples	21
3.4	Radiation of an Ultrarelativistic Charged Particle	22
4	Frequency Distribution of the Radiated Energy	25
4.1	Continuous Frequency Distribution	26
4.2	Discrete Frequency Distribution	30

4.3	Examples	32
4.3.1	A Particle in Instantaneous Circular Motion	32
4.3.2	A Particle in Circular Motion	37
5	Thomson Scattering; Blue Sky	42
6	Cherenkov Radiation Revisited	46
7	Cherenkov Radiation; Transition Radiation	51
7.1	Cherenkov Radiation in a Dilute Collisionless Plasma	55
8	Example Problems	57
8.1	A Relativistic Particle in a Capacitor	57
8.2	Relativistic Electrons at SLAC	58

We have already calculated the radiation produced by some known charge and current distribution. Now we are going to do it again. This time, however, we shall consider that the source is a single charge moving in some fairly arbitrary, possibly relativistic, fashion. Here the methods of chapter 9, e.g., multipole expansions, are impractical and there are better ways to approach the problem.

1 Liénard-Wiechert potentials

The current and charge densities produced by a charge e in motion are

$$\begin{aligned}
 \rho(\mathbf{x}, t) &= e\delta(\mathbf{x} - \mathbf{x}(t)) \\
 \mathbf{J}(\mathbf{x}, t) &= e\mathbf{v}(t)\delta(\mathbf{x} - \mathbf{x}(t))
 \end{aligned}
 \tag{1}$$

if $\mathbf{x}(t)$ is the position of the particle at time t and $\mathbf{v}(t) \equiv d\mathbf{x}(t)/dt \equiv \dot{\mathbf{x}}(t)$ is its velocity. In four-vector notation,

$$J^\mu(\mathbf{x}, t) = ec\beta^\mu\delta(\mathbf{x} - \mathbf{x}(t))
 \tag{2}$$

where $\beta^\mu \equiv (1, \boldsymbol{\beta})$ is **not** a four-vector; $\boldsymbol{\beta} = \mathbf{v}/c$.

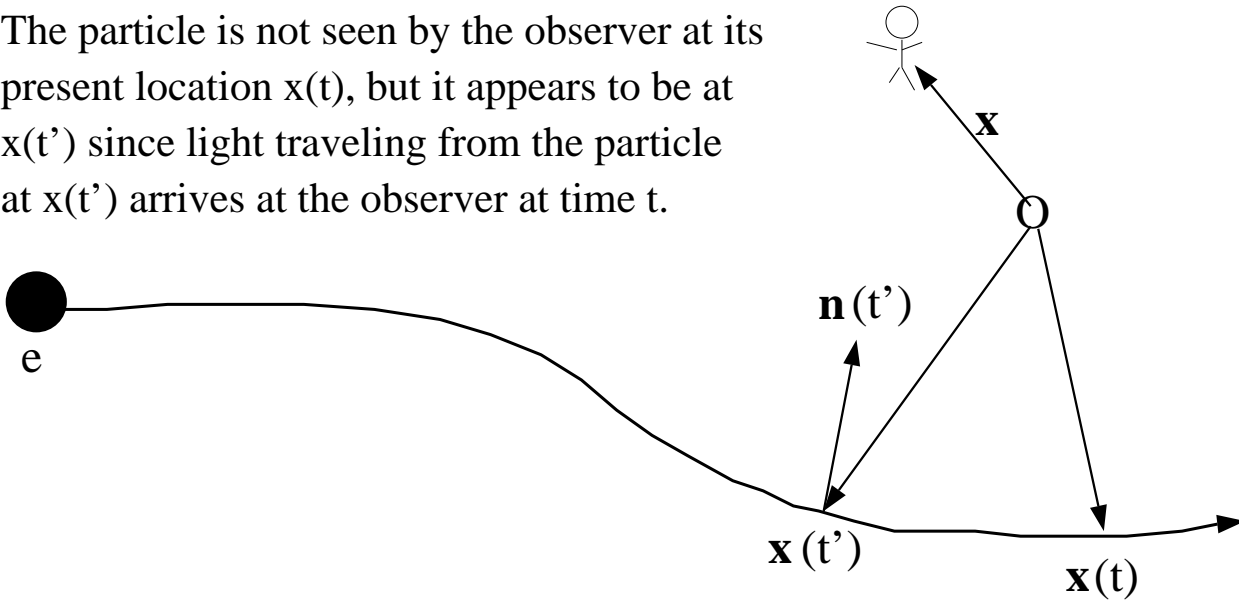
From \mathbf{J} and ρ , one finds \mathbf{A} and Φ . We can do this in an infinite space by making use of the retarded Green's function $G(\mathbf{x}, t; \mathbf{x}', t') = \delta(t - t' - |\mathbf{x} - \mathbf{x}'|/c)/|\mathbf{x} - \mathbf{x}'|$. From here, we can evaluate the electromagnetic field as appropriate derivatives of the potentials. All of these manipulations are straightforward. Furthermore, the integrations are relatively easy because there are many delta functions. The problem becomes interesting and unfamiliar, however, for highly relativistic particles which produce large retardation effects.

Let us start from the integral expression for the potentials:

$$\begin{aligned}
 A^\mu(\mathbf{x}, t) &= \int d^3x' dt' G(\mathbf{x}, t; \mathbf{x}', t') J^\mu(\mathbf{x}', t') \\
 &= \frac{1}{c} \int d^3x' dt' \frac{\delta(t - t' - |\mathbf{x} - \mathbf{x}'|/c) J^\mu(\mathbf{x}', t')}{|\mathbf{x} - \mathbf{x}'|} \\
 &= e \int d^3x' dt' \beta^\mu(t') \frac{\delta(\mathbf{x}' - \mathbf{x}(t')) \delta(t - t' - |\mathbf{x} - \mathbf{x}'|/c)}{|\mathbf{x} - \mathbf{x}'|} \\
 &= e \int dt' \beta^\mu(t') \frac{\delta(t - t' - |\mathbf{x} - \mathbf{x}(t')|/c)}{|\mathbf{x} - \mathbf{x}(t')|} \tag{3}
 \end{aligned}$$

This form reflects the retarded nature of the problem.

The particle is not seen by the observer at its present location $\mathbf{x}(t)$, but it appears to be at $\mathbf{x}(t')$ since light traveling from the particle at $\mathbf{x}(t')$ arrives at the observer at time t .



The evaluation of this integral is not totally simple because the argument of the δ -function is not a simple function of the time t' . In general when one faces an integral of this form, one invokes the rule

$$\int dt' f(t') \delta[g(t')] = f(t_0) \left/ \left| \frac{dg}{dt'} \right|_{t_0} \right. \quad (4)$$

where t_0 is the zero (there may be more than one) of g^1 , i.e., $g(t_0) = 0$. Applying this to the present case, we have

$$g(t') = t' + |\mathbf{x} - \mathbf{x}(t')|/c - t = t' + [(\mathbf{x} - \mathbf{x}(t')) \cdot (\mathbf{x} - \mathbf{x}(t'))]^{1/2}/c - t, \quad (5)$$

so

$$\frac{dg}{dt'} = 1 + \frac{1}{2c} \left(-2 \frac{d\mathbf{x}(t')}{dt'} \right) \cdot \frac{(\mathbf{x} - \mathbf{x}(t'))}{|\mathbf{x} - \mathbf{x}(t')|} = 1 - \frac{\mathbf{v}(t') \cdot (\mathbf{x} - \mathbf{x}(t'))}{c|\mathbf{x} - \mathbf{x}(t')|} \equiv 1 - \boldsymbol{\beta}(t') \cdot \mathbf{n}(t'). \quad (6)$$

The unit vector \mathbf{n} points from the point $\mathbf{x}(t')$ on the particle's path toward the field point \mathbf{x} ; it, and $\boldsymbol{\beta}$, must be evaluated at a time t' which is earlier than t , the time at which the field is evaluated, by some amount which is determined by solving the equation

$$t' + |\mathbf{x} - \mathbf{x}(t')|/c = t. \quad (7)$$

Combining Eqs. (3), (4), and (6) we find that the potentials are given simply by

$$A^\mu(\mathbf{x}, t) = e \left[\frac{\beta^\mu}{|\mathbf{x} - \mathbf{x}(t')|(1 - \boldsymbol{\beta} \cdot \mathbf{n})} \right]_{ret} = \left[\frac{e\beta^\mu}{R\kappa} \right]_{ret} \quad (8)$$

where $R \equiv |\mathbf{x} - \mathbf{x}(t')|$, $\kappa = 1 - \boldsymbol{\beta} \cdot \mathbf{n}$, and the subscript *ret* means that the quantity in brackets [...] must be evaluated at the retarded time t' determined from Eq. (7).

Our potentials, Eq. (8), are known as the *Liénard-Wiechert potentials*. Probably their most significant feature is the fact that they vary inversely as $1 - \boldsymbol{\beta} \cdot \mathbf{n}$ or κ ; this factor can be very close to zero for $\mathbf{n} \parallel \boldsymbol{\beta}$ if β is close to one, i.e., for highly relativistic

¹More generally $\delta[g(x)] = \sum_j \delta(x - x_j)/|g'(x_j)|$ where x_j are the simple zeroes of $g(x)$. If $g'(x_j) = 0$ (a complex zero), then $\delta[g(x)]$ makes no sense.

particles, meaning that there is a strong maximum in the potentials produced by a relativistic particle in the direction of the particle's velocity (at some retarded time).

We wish next to find the electromagnetic field. One may do this in a variety of ways. One is simply **carefully** to take derivatives of the Liénard-Wieckert potentials, a procedure followed in, e.g., Landau and Lifshitz' book, *The Classical Theory of Fields*. Another, much more elegant, is to find the fields in the instantaneous rest frame of the particle and to Lorentz-transform them to the frame interest. Another, not very elegant at all, and about to be employed here, is to go back to the integrals, Eq. (3), for the potentials and take derivatives of these expressions. Consider just the vector potential,

$$\mathbf{A}(\mathbf{x}, t) = e \int dt' \frac{\boldsymbol{\beta}(t') \delta(t' + R(t')/c - t)}{R(t')}. \quad (9)$$

If we wish to find $\mathbf{B}(\mathbf{x}, t)$, we have to take derivatives of R with respect to various components of \mathbf{x} . Consider, for example,

$$\nabla f(R) = \frac{df}{dR} \nabla R = \frac{df}{dR} \frac{\mathbf{R}}{R} = \mathbf{n} \frac{df}{dR}. \quad (10)$$

Application of this simple rule gives (since $\nabla \times (\psi \mathbf{a}) = \nabla \psi \times \mathbf{a} + \psi \nabla \times \mathbf{a}$, and $\nabla \times \boldsymbol{\beta}(t') = 0$ due to the lack of an \mathbf{x} dependence.)

$$\begin{aligned} \mathbf{B}(\mathbf{x}, t) &= \nabla \times \mathbf{A}(\mathbf{x}, t) = e \int dt' \nabla \left(\frac{\delta(t' + R/c - t)}{R} \right) \times \boldsymbol{\beta}(t') \\ &= e \int dt' (\mathbf{n} \times \boldsymbol{\beta}) \left[-\frac{1}{R^2} \delta(t' + R/c - t) + \frac{1}{cR} \delta'(t' + R/c - t) \right] \end{aligned} \quad (11)$$

where the prime on the delta function denotes differentiation with respect to the argument. Hence

$$\mathbf{B}(\mathbf{x}, t) = e \left\{ \left[\frac{\boldsymbol{\beta} \times \mathbf{n}}{\kappa R^2} \right]_{ret} + \int dt' \frac{\kappa}{cR} \left(\frac{\delta'(t' + R/c - t)}{\kappa} \right) (\mathbf{n} \times \boldsymbol{\beta}) \right\} \quad (12)$$

Now, from Eqs. (5) and (6)

$$\frac{d(t' + R/c - t)}{dt'} = \kappa \quad \text{or} \quad \kappa dt' = d(t' + R/c - t) \quad (13)$$

and so

$$\begin{aligned}
\mathbf{B}(\mathbf{x}, t) &= e \left\{ \left[\frac{\boldsymbol{\beta} \times \mathbf{n}}{\kappa R^2} \right]_{ret} + \int d(t' + R/c - t) \left(\frac{\mathbf{n} \times \boldsymbol{\beta}}{c\kappa R} \right) \delta'(t' + R/c - t) \right\} \\
&= e \left\{ \left[\frac{\boldsymbol{\beta} \times \mathbf{n}}{\kappa R^2} \right]_{ret} - \int d(t' + R/c - t) \frac{\partial[(\mathbf{n} \times \boldsymbol{\beta})/c\kappa R]}{\partial(t' + R/c - t)} \delta(t' + R/c - t) \right\} \\
&= e \left\{ \left[\frac{\boldsymbol{\beta} \times \mathbf{n}}{\kappa R^2} \right]_{ret} - \left[\frac{1}{\kappa} \frac{\partial}{\partial t'} \left(\frac{\mathbf{n} \times \boldsymbol{\beta}}{cR\kappa} \right) \right]_{ret} \right\} \\
&= e \left\{ \left[\frac{\boldsymbol{\beta} \times \mathbf{n}}{\kappa R^2} \right]_{ret} + \frac{1}{c} \left[\frac{1}{\kappa} \frac{\partial}{\partial t'} \left(\frac{\boldsymbol{\beta} \times \mathbf{n}}{\kappa R} \right) \right]_{ret} \right\}. \tag{14}
\end{aligned}$$

The electric field can be found by similar manipulations:

$$\begin{aligned}
\mathbf{E}(\mathbf{x}, t) &= -\nabla\Phi(\mathbf{x}, t) - \frac{1}{c} \frac{\partial \mathbf{A}(\mathbf{x}, t)}{\partial t} \\
&= -e \int dt' \mathbf{n} \frac{d}{dR} \left(\frac{\delta(t' + R/c - t)}{R} \right) + \frac{e}{c} \int dt' \frac{\boldsymbol{\beta} \delta'(t' + R/c - t)}{R} \\
&= e \int dt' \left\{ \mathbf{n} \frac{\delta(t' + R/c - t)}{R^2} + \frac{1}{cR} (\boldsymbol{\beta} - \mathbf{n}) \delta'(t' + R/c - t) \right\} \\
&= e \left[\frac{\mathbf{n}}{\kappa R^2} \right]_{ret} - \frac{e}{c} \left[\frac{1}{\kappa} \frac{\partial}{\partial t'} \left(\frac{\boldsymbol{\beta} - \mathbf{n}}{\kappa R} \right) \right]_{ret}. \tag{15}
\end{aligned}$$

We now have expressions for \mathbf{E} and \mathbf{B} , but they involve time derivatives of retarded quantities. We can work out each of these derivatives. First, we consider just the derivative of $\mathbf{n} = \mathbf{R}/R$. One has

$$\frac{\partial \mathbf{R}}{\partial t'} = -\dot{\mathbf{x}}(t') = -\boldsymbol{\beta}c \tag{16}$$

and

$$\frac{\partial R}{\partial t'} = \frac{\partial[(\mathbf{x} - \mathbf{x}(t')) \cdot (\mathbf{x} - \mathbf{x}(t'))]^{1/2}}{\partial t'} = \frac{1}{2R} [-2(\mathbf{x} - \mathbf{x}(t')) \cdot \dot{\mathbf{x}}(t')] = -\mathbf{n} \cdot \dot{\mathbf{x}}(t') = -\mathbf{n} \cdot \boldsymbol{\beta}c \tag{17}$$

Hence,

$$\frac{1}{c} \frac{d\mathbf{n}}{dt'} = \frac{1}{cR} \frac{\partial \mathbf{R}}{\partial t'} - \frac{1}{cR^2} \mathbf{R} \frac{\partial R}{\partial t'} = -\frac{1}{R} [\boldsymbol{\beta} - (\mathbf{n} \cdot \boldsymbol{\beta})\mathbf{n}]. \tag{18}$$

The quantity in brackets is can be written more concisely:

$$(\mathbf{n} \cdot \boldsymbol{\beta})\mathbf{n} - \boldsymbol{\beta} = (\mathbf{n} \cdot \boldsymbol{\beta})\mathbf{n} - (\mathbf{n} \cdot \mathbf{n})\boldsymbol{\beta} = \mathbf{n} \times (\mathbf{n} \times \boldsymbol{\beta}); \tag{19}$$

hence we find

$$\frac{1}{c} \frac{d\mathbf{n}}{dt'} = \frac{1}{R} \mathbf{n} \times (\mathbf{n} \times \boldsymbol{\beta}). \quad (20)$$

Employing this useful result we have, for the electric field,

$$\begin{aligned} \mathbf{E}(\mathbf{x}, t) &= e \left[\frac{\mathbf{n}}{\kappa R^2} + \frac{1}{\kappa^2 R^2} \mathbf{n} \times (\mathbf{n} \times \boldsymbol{\beta}) + \frac{\mathbf{n}}{c\kappa} \frac{\partial}{\partial t'} \left(\frac{1}{\kappa R} \right) - \frac{1}{c\kappa} \frac{\partial}{\partial t'} \left(\frac{\boldsymbol{\beta}}{\kappa R} \right) \right]_{ret} \\ &= e \left[\frac{\mathbf{n}(1 - \mathbf{n} \cdot \boldsymbol{\beta})}{\kappa^2 R^2} + \frac{1}{\kappa^2 R^2} [(\mathbf{n} \cdot \boldsymbol{\beta})\mathbf{n} - \boldsymbol{\beta}] \right]_{ret} \\ &\quad + \left[\frac{\mathbf{n}}{c\kappa} \frac{\partial}{\partial t'} \left(\frac{1}{\kappa R} \right) - \frac{1}{c\kappa} \frac{\partial}{\partial t'} \left(\frac{\boldsymbol{\beta}}{\kappa R} \right) \right]_{ret} \end{aligned} \quad (21)$$

or

$$\mathbf{E}(\mathbf{x}, t) = e \left[\frac{\mathbf{n} - \boldsymbol{\beta}}{\kappa^2 R^2} + \frac{\mathbf{n}}{c\kappa} \frac{\partial}{\partial t'} \left(\frac{1}{\kappa R} \right) - \frac{1}{c\kappa} \frac{\partial}{\partial t'} \left(\frac{\boldsymbol{\beta}}{\kappa R} \right) \right]_{ret}. \quad (22)$$

This expression may be related to that for the magnetic induction,

$$\begin{aligned} \mathbf{B}(\mathbf{x}, t) &= e \left[\frac{\boldsymbol{\beta} \times \mathbf{n}}{\kappa R^2} \right]_{ret} + \frac{e}{c} \left[\frac{1}{\kappa} \frac{\partial}{\partial t'} \left(\frac{\boldsymbol{\beta} \times \mathbf{n}}{\kappa R} \right) \right]_{ret} \\ &= e \left[\frac{\boldsymbol{\beta} \times \mathbf{n}}{\kappa R^2} \right]_{ret} + \frac{e}{c} \left[\frac{1}{\kappa} \frac{\partial}{\partial t'} \left(\frac{\boldsymbol{\beta}}{\kappa R} \right) \times \mathbf{n} \right]_{ret} \\ &\quad + e \left[\frac{\boldsymbol{\beta}}{\kappa^2 R} \times \left(\frac{1}{R} \mathbf{n} \times (\mathbf{n} \times \boldsymbol{\beta}) \right) \right]_{ret} \\ &= e \left[\frac{\boldsymbol{\beta} \times \mathbf{n}}{\kappa R^2} \right]_{ret} + \frac{e}{c} \left[\frac{1}{\kappa} \frac{\partial}{\partial t'} \left(\frac{\boldsymbol{\beta}}{\kappa R} \right) \times \mathbf{n} \right]_{ret} \\ &\quad + e \left[\frac{1}{\kappa^2 R^2} [\mathbf{n}((\mathbf{n} \times \boldsymbol{\beta}) \cdot \boldsymbol{\beta}) - (\mathbf{n} \times \boldsymbol{\beta})(\mathbf{n} \cdot \boldsymbol{\beta})] \right]_{ret} \end{aligned} \quad (23)$$

Next, put all terms proportional to $\mathbf{n} \times \boldsymbol{\beta}$ over the same denominator, which is $\kappa^2 R^2$, and also make use of the fact that $(\mathbf{n} \times \boldsymbol{\beta}) \cdot \boldsymbol{\beta} = 0$, to find

$$\mathbf{B}(\mathbf{x}, t) = e \left[\frac{\boldsymbol{\beta} \times \mathbf{n}}{\kappa^2 R^2} \right]_{ret} + \frac{e}{c} \left[\frac{1}{\kappa} \frac{\partial}{\partial t'} \left(\frac{\boldsymbol{\beta}}{\kappa R} \right) \times \mathbf{n} \right]_{ret} \quad (24)$$

This is our final expression. Comparing it with Eq. (22) for $\mathbf{E}(\mathbf{x}, t)$, we can see that

$$\mathbf{B}(\mathbf{x}, t) = [\mathbf{n}]_{ret} \times \mathbf{E}(\mathbf{x}, t). \quad (25)$$

Thus we can find $\mathbf{B}(\mathbf{x}, t)$ quite easily provided we can find $\mathbf{E}(\mathbf{x}, t)$. Henceforth, we won't spend more time on \mathbf{B} but will examine only \mathbf{E} from which \mathbf{B} then follows trivially.

To have an explicit expression for \mathbf{E} with no time derivatives, we need to evaluate

$$\begin{aligned}
\frac{\partial}{\partial t'} \left(\frac{1}{\kappa R} \right) &= -\frac{1}{\kappa^2 R^2} \left[\kappa(-\mathbf{n} \cdot \boldsymbol{\beta} c) + R \left(-\mathbf{n} \cdot \frac{\partial \boldsymbol{\beta}}{\partial t'} \right) - R \boldsymbol{\beta} \cdot [(\mathbf{n} \cdot \boldsymbol{\beta}) \mathbf{n} - \boldsymbol{\beta}] \frac{c}{R} \right] \\
&= -\frac{c}{\kappa^2 R^2} \left[(1 - \mathbf{n} \cdot \boldsymbol{\beta})(-\mathbf{n} \cdot \boldsymbol{\beta}) - \frac{R}{c} (\mathbf{n} \cdot \dot{\boldsymbol{\beta}}) + \boldsymbol{\beta} \cdot [\boldsymbol{\beta} - (\mathbf{n} \cdot \boldsymbol{\beta}) \mathbf{n}] \right] \\
&= -\frac{c}{\kappa^2 R^2} \left[\beta^2 - \mathbf{n} \cdot \boldsymbol{\beta} - \frac{R}{c} (\mathbf{n} \cdot \dot{\boldsymbol{\beta}}) \right] \tag{26}
\end{aligned}$$

and so

$$\begin{aligned}
\mathbf{E}(\mathbf{x}, t) &= e \left[\frac{\mathbf{n} - \boldsymbol{\beta}}{\kappa^2 R^2} + \frac{\mathbf{n}}{c\kappa} \left(\frac{-c}{\kappa^2 R^2} \right) \left(\beta^2 - \mathbf{n} \cdot \boldsymbol{\beta} - \frac{R}{c} \mathbf{n} \cdot \dot{\boldsymbol{\beta}} \right) - \frac{\dot{\boldsymbol{\beta}}}{c\kappa^2 R} \right]_{ret} \\
&\quad - e \left[\frac{\boldsymbol{\beta}}{c\kappa} \left(\frac{-c}{\kappa^2 R^2} \right) \left(\beta^2 - \mathbf{n} \cdot \boldsymbol{\beta} - \frac{R}{c} \mathbf{n} \cdot \dot{\boldsymbol{\beta}} \right) \right]_{ret} \\
&= e \left[\frac{1}{\kappa^3 R^2} \mathbf{n} \left(1 - \beta^2 + \frac{R}{c} \mathbf{n} \cdot \dot{\boldsymbol{\beta}} \right) + \frac{\boldsymbol{\beta}}{\kappa^3 R^2} \left(-1 + \beta^2 - \frac{R}{c} \mathbf{n} \cdot \dot{\boldsymbol{\beta}} \right) - \frac{1}{c\kappa^2 R} \dot{\boldsymbol{\beta}} \right]_{ret} \\
&= e \left[\frac{(\mathbf{n} - \boldsymbol{\beta})(1 - \beta^2)}{\kappa^3 R^2} \right]_{ret} + e \left[\frac{1}{c\kappa^3 R} \left((\mathbf{n} - \boldsymbol{\beta})(\mathbf{n} \cdot \dot{\boldsymbol{\beta}}) - (1 - \mathbf{n} \cdot \boldsymbol{\beta}) \dot{\boldsymbol{\beta}} \right) \right]_{ret} \tag{27}
\end{aligned}$$

The second bracket in the final expression contains the quantity

$$(\mathbf{n} - \boldsymbol{\beta})(\mathbf{n} \cdot \dot{\boldsymbol{\beta}}) - \mathbf{n} \cdot (\mathbf{n} - \boldsymbol{\beta}) \dot{\boldsymbol{\beta}} = \mathbf{n} \times [(\mathbf{n} - \boldsymbol{\beta}) \times \dot{\boldsymbol{\beta}}], \tag{28}$$

so

$$\mathbf{E}(\mathbf{x}, t) = e \left[\frac{(\mathbf{n} - \boldsymbol{\beta})(1 - \beta^2)}{\kappa^3 R^2} \right]_{ret} + \frac{e}{c} \left[\frac{\mathbf{n} \times [(\mathbf{n} - \boldsymbol{\beta}) \times \dot{\boldsymbol{\beta}}]}{\kappa^3 R} \right]_{ret}. \tag{29}$$

If there is no acceleration, then $\dot{\boldsymbol{\beta}} = 0$ and only the first term in \mathbf{E} and the corresponding term in \mathbf{B} are finite. These terms fall off with distance as $1/R^2$, and hence cannot give rise to a net flux of radiation to infinity. If there is an acceleration, then the second term of \mathbf{E} , and the corresponding term in \mathbf{B} are finite. These fall off as $1/R$, and hence will give rise to radiation, meaning that the charged particle will

emit radiation only if it is accelerated. Hence the two terms are interpreted as the non-radiation and radiation parts respectively.

From these results when $\dot{\boldsymbol{\beta}} = 0$, we should be able to recover the results for the fields of a uniformly moving charge which we derived in chapter 11 (Jackson Eq. 11.152). With the particle moving along the x-direction with a constant velocity v , the fields felt by an observer a distance b away are

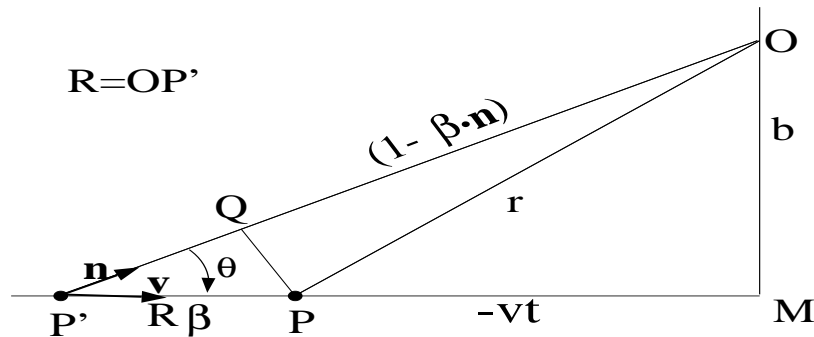
$$\begin{aligned} E_{\parallel} &= -\gamma qvt/[b^2 + (\gamma vt)^2]^{3/2} \\ E_{\perp} &= \gamma qb/[b^2 + (\gamma vt)^2]^{3/2}. \end{aligned} \quad (30)$$

where the observer and particle are closest at time $t = 0$. It is a reasonably simple task to show that this is the same result as Eq. (29) above when $\dot{\boldsymbol{\beta}} = 0$. Consider the diagram below in which the path of a charged particle is along the abscissa. At time t , the retarded and real location of the particle are P' and P respectively, while O is the observation point. The time required for light to travel from P' to the observer O is $t = R/c$, in which time the particle travels $\beta R = PP'$. Thus the distance P'Q is $\beta R \cos \theta = \boldsymbol{\beta} \cdot \mathbf{n}R$. From this it follows that the distance OQ is $R - \boldsymbol{\beta} \cdot \mathbf{n}R$. Then $[(1 - \boldsymbol{\beta} \cdot \mathbf{n})R]^2 = r^2 - (PQ)^2 = r^2 - \beta^2 R^2 \sin^2(\theta)$. Thus, as $R \sin \theta = b$

$$[(1 - \boldsymbol{\beta} \cdot \mathbf{n})R]^2 = b^2 + v^2 t^2 - \beta^2 b^2 = \frac{1}{\gamma^2} (b^2 + \gamma^2 v^2 t^2), \quad (31)$$

and transverse component of Eq. (29) when $\dot{\boldsymbol{\beta}} = 0$ is

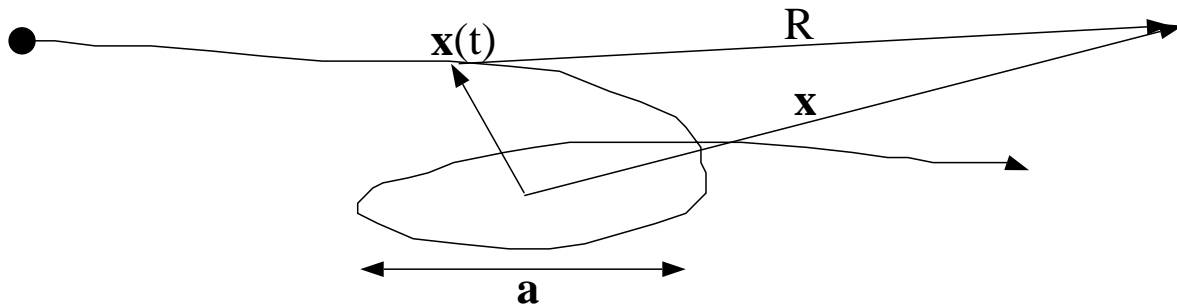
$$e \left[\frac{(R \sin \theta)}{\gamma^2 (1 - \boldsymbol{\beta} \cdot \mathbf{n})^3 R^3} \right]_{ret} = \frac{eb\gamma}{[b^2 + (\gamma vt)^2]^{3/2}} \quad (32)$$



2 Radiation from an Accelerated Charge; the Larmor Formula

Even given Eqs. (25) and (29), the problem of calculating the fields emitted from a charge moving along an arbitrary trajectory is non-trivial. This is largely due to the effect of retardation. The problem is greatly simplified if the particle is not moving too fast.

In order to consider this limit, let's consider the trajectory of a particle shown below.



We will assume that the origin is located in the center of the region of interest of the particles trajectory, and that this region of interest is of linear dimension a (an example would be an electron bound to a classical atom, where a would be the Bohr radius).

There are two ways in which the problem simplifies in the nonrelativistic limit, $\beta \ll 1$. First, we may approximate

$$\kappa \approx 1 \quad \mathbf{n} - \boldsymbol{\beta} \approx \mathbf{n} \quad 1 - \beta^2 \approx 1. \quad (33)$$

Second, and much more importantly, we can approximate functions of the retarded time

$$f(t - |\mathbf{x} - \mathbf{x}(t')|/c) \quad (34)$$

by making a Taylor series expansion around the origin

$$f(t - |\mathbf{x} - \mathbf{x}(t')|/c) \approx f(t - r/c) + \mathbf{x}(t') \cdot \nabla f(t - r/c) + \dots \quad (35)$$

where $r = |\mathbf{x}|$. The ratio of the second term to the first is roughly $f'x(t')/fc$, where the prime means differentiation of f with respect to $t - r/c$. If this is small, as it must be for $\beta \ll 1$, then the second term in the series may be neglected, then

$$f(t - |\mathbf{x} - \mathbf{x}(t')|/c) \approx f(t - r/c). \quad (36)$$

This approximation is sometimes called the dipole approximation (we will see why directly). Here the effects of retardation are simple but not insignificant. They are simply due to the separation of the charge from the observer, not due to the motion of the charge. Thus, the only way they are significant is if the acceleration of the charge changes abruptly and it takes a while for the emitted fields to reach the observer.²

With the approximations described above, the electric field becomes

$$\mathbf{E}_{nr}(\mathbf{x}, t) = e \left[\frac{\mathbf{n}}{R^2} \right]_{ret} + \frac{e}{c} \left[\frac{\mathbf{n} \times (\mathbf{n} \times \dot{\boldsymbol{\beta}})}{R} \right]_{ret}. \quad (37)$$

The Poynting vector is given by

$$\begin{aligned} \mathbf{S} &= \frac{c}{4\pi} (\mathbf{E} \times \mathbf{B}) = \frac{c}{4\pi} \mathbf{E} \times ([\mathbf{n}]_{ret} \times \mathbf{E}) \\ &= \frac{c}{4\pi} \{ [\mathbf{n}]_{ret} \mathbf{E} \cdot \mathbf{E} - \mathbf{E} ([\mathbf{n}]_{ret} \cdot \mathbf{E}) \}. \end{aligned} \quad (38)$$

If we take the limit of large R and keep just the radiation terms, then we find that

$$\mathbf{E}(\mathbf{x}, t) = \frac{e}{c} \left[\frac{\mathbf{n}(\mathbf{n} \cdot \dot{\boldsymbol{\beta}}) - \dot{\boldsymbol{\beta}}}{R} \right]_{ret}; \quad \mathbf{B}(\mathbf{x}, t) = [\mathbf{n}]_{ret} \times \mathbf{E} = \left[-\frac{e}{cR} \mathbf{n} \times \dot{\boldsymbol{\beta}} \right]_{ret} \quad (39)$$

Note that $\mathbf{n} \cdot \mathbf{E} = \mathbf{n} \cdot \mathbf{B} = 0$ for these radiation fields, thus (as usual) the radiation is transverse. The Poynting vector due to the radiation is then given by

$$\mathbf{S} = \frac{ce^2}{4\pi R^2 c^2} \left[\mathbf{n} \left| \mathbf{n} \times (\mathbf{n} \times \dot{\boldsymbol{\beta}}) \right|^2 \right]_{ret} \quad (40)$$

²For example, suppose a particle starts to oscillate back and forth at time $t = 0$, there will be no signal felt at the observer's location \mathbf{x} for what may be a very long time $t = r/c$.

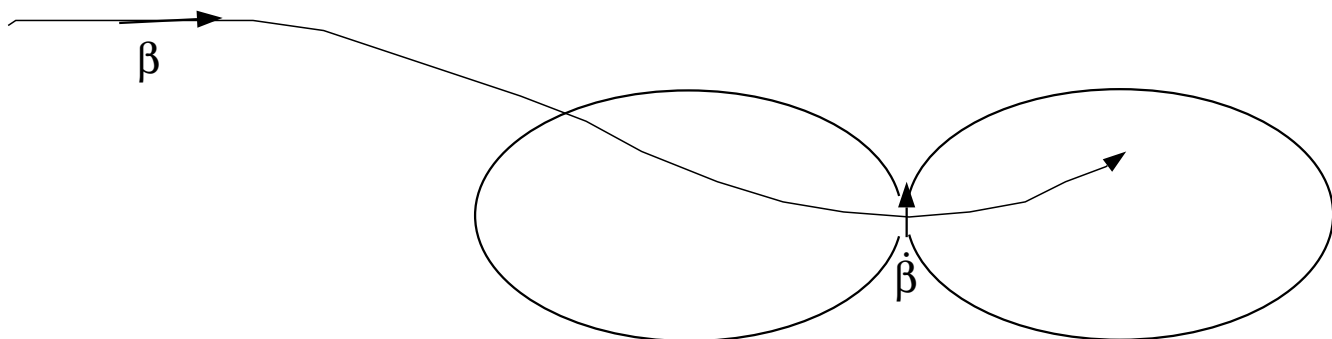
The angular distribution of radiated power is

$$\frac{d\mathcal{P}}{d\Omega} = R^2(\mathbf{S} \cdot \mathbf{n}) \quad (41)$$

which is easily evaluated by expanding the cross products. The result is

$$\frac{d\mathcal{P}}{d\Omega} = \frac{e^2}{4\pi c} (\dot{\beta}^2 \sin^2 \theta) = \frac{e^2}{4\pi c^3} (\dot{\mathbf{v}}^2 \sin^2 \theta) \quad (42)$$

where θ is the angle between \mathbf{n} and $\dot{\boldsymbol{\beta}}$ (at the retarded time).



In these last expressions we have dropped the reminder that \mathbf{n} and R must be evaluated at the retarded time.

The radiation pattern is characteristic of dipole radiation with the reference z axis parallel to the direction of the particle's acceleration. The shape and magnitude of the distribution are independent of the particle's velocity and proportional to the square of the acceleration. The total power radiated is the integral over all directions of the distribution,

$$\mathcal{P} = \int d\Omega \frac{d\mathcal{P}}{d\Omega} = \frac{2}{3} \frac{e^2 \dot{\mathbf{v}}^2}{c^3} \quad (43)$$

which is known as the *Larmor formula* for the power radiated by an accelerated particle.

2.1 Relativistic Larmor Formula

We turn now to the relativistic generalization of the Larmor formula. One can determine this generalization by consideration of how power transforms under Lorentz transformations. There are actually two ways (that I know of) to do this, of which Jackson does one. One can also find it from direct computation and that is what we shall do. The first step is to consider just how we shall define the power radiated by a particle. The point is that the rate at which energy crosses a closed surface surrounding the particle depends on that surface because of retardation. We are going to calculate the power as a function not of the time t at which the fields are measured on the surface but rather as a function of the retarded time t' . Consider a surface S which encloses the particle at all times during which it is radiating. The power crossing unit area at \mathbf{x} on S at time t is $\mathbf{S}(\mathbf{x}, t) \cdot \mathbf{n}$, where \mathbf{n} is a unit outward normal, and so the total energy crossing this unit area is

$$W = \int_{-\infty}^{\infty} dt \mathbf{S}(\mathbf{x}, t) \cdot \mathbf{n}. \quad (44)$$

Now let us transform to the retarded time t' :

$$W = \int_{-\infty}^{\infty} dt' \frac{dt}{dt'} \mathbf{S}(\mathbf{x}, t(t')) \cdot \mathbf{n} = \int_{-\infty}^{\infty} dt' \kappa [\mathbf{S}(\mathbf{x}, t(t')) \cdot \mathbf{n}]. \quad (45)$$

The integrand of this expression we identify as dW/dt' , the rate at which the particle radiates what eventually passes through the unit area on S at \mathbf{x} . This is the *instantaneous* radiated power, and if we multiply it by R^2 we get $d\mathcal{P}(t')/d\Omega$:

$$\frac{d\mathcal{P}(t')}{d\Omega} = \kappa R^2 \mathbf{S} \cdot \mathbf{n}. \quad (46)$$

If we suppose that R is large enough that only the radiation fields need be retained, then we find, from our results for the fields and the definition of the Poynting vector, that

$$\frac{d\mathcal{P}(t')}{d\Omega} = \frac{e^2}{4\pi c} \frac{[\mathbf{n} \times ((\mathbf{n} - \boldsymbol{\beta}) \times \dot{\boldsymbol{\beta}})]^2}{(1 - \mathbf{n} \cdot \boldsymbol{\beta})^5} \Big|_{t'} \quad (47)$$

Of course, if one wants to know the radiated power per unit area at the time that it actually gets there, then he will have to address the calculation of the retardation. In other words, the radiated intensity at time t depends upon the behavior of the particle at t' , and the differential time elements are different as well: $dt = dt'(1 - \mathbf{n} \cdot \boldsymbol{\beta})_{ret}$. For example, if a particle of velocity $\boldsymbol{\beta}$ is impulsively accelerated for a time τ , and then brought to rest, a pulse of radiation will appear at the observer at time $t = r/c$, of duration $\tau(1 - \mathbf{n} \cdot \boldsymbol{\beta})_{ret}$. The total energy lost of the particle, of course, must eventually equal the energy radiated, but the energy lost by the charge per unit time will differ from the energy radiated per unit time by the factor $(1 - \mathbf{n} \cdot \boldsymbol{\beta})_{ret}$

However, by calculating the instantaneous power radiated, we have avoided having to calculate the retardation. Thus we can proceed with a calculation of the energy lost. Our result, Eq. (47), depends in a complicated way on both $\boldsymbol{\beta}$ and $\dot{\boldsymbol{\beta}}$ making it possible for rather remarkable angular distributions to occur. The total power radiated can also be computed by integrating the distribution over directions:

$$\begin{aligned}
\mathcal{P}(t') &= \frac{e^2}{4\pi c} \int \frac{d\Omega}{\kappa^5} [\mathbf{n} \times ((\mathbf{n} - \boldsymbol{\beta}) \times \dot{\boldsymbol{\beta}})]^2 \\
&= \frac{e^2}{4\pi c} \int \frac{d\Omega}{\kappa^5} \left\{ [(\mathbf{n} - \boldsymbol{\beta})(\mathbf{n} \cdot \dot{\boldsymbol{\beta}}) - \dot{\boldsymbol{\beta}}(1 - \boldsymbol{\beta} \cdot \mathbf{n})]^2 \right\} \\
&= \frac{e^2}{4\pi c} \int \frac{d\Omega}{\kappa^5} \left\{ (1 - 2\mathbf{n} \cdot \boldsymbol{\beta} + \beta^2)(\mathbf{n} \cdot \dot{\boldsymbol{\beta}})^2 \right. \\
&\quad \left. - 2(\mathbf{n} \cdot \dot{\boldsymbol{\beta}} - \boldsymbol{\beta} \cdot \dot{\boldsymbol{\beta}})(\dot{\boldsymbol{\beta}} \cdot \mathbf{n})(1 - \mathbf{n} \cdot \boldsymbol{\beta}) + \dot{\boldsymbol{\beta}}^2(1 - 2\boldsymbol{\beta} \cdot \mathbf{n} + (\boldsymbol{\beta} \cdot \mathbf{n})^2) \right\} \quad (48)
\end{aligned}$$

Let the angle between $\boldsymbol{\beta}$ and \mathbf{n} be θ ; further, let $\cos \theta = u$; also, let the angle between $\boldsymbol{\beta}$ and $\dot{\boldsymbol{\beta}}$ be θ_0 . Then

$$\begin{aligned}
\mathcal{P}(t') &= \frac{e^2 \dot{\boldsymbol{\beta}}^2}{2c} \int_{-1}^1 \frac{du}{\kappa^5} \left\{ (\beta^2 - 1) \left[u^2 \cos^2 \theta_0 + \frac{1}{2}(1 - u^2) \sin^2 \theta_0 \right] \right. \\
&\quad \left. + 2 \cos^2 \theta_0 u \beta (1 - \beta u) + (1 - 2\beta u + \beta^2 u^2) \right\} \\
&= \frac{\dot{\boldsymbol{\beta}}^2 e^2}{2c} \int_{-1}^1 \frac{du}{(1 - \beta u)^5} \left\{ -u^2(1 - \beta^2) + 2u\beta(1 - \beta u) + 1 - 2\beta u + \beta^2 u^2 \right. \\
&\quad \left. + \sin^2 \theta_0 \left[(1 - \beta^2)u^2 - (1 - u^2)(1 - \beta^2)/2 - 2\beta u + 2\beta^2 u^2 \right] \right\} \quad (49) \\
&= \frac{\dot{\boldsymbol{\beta}}^2 e^2}{2c} \int_{-1}^1 \frac{du}{(1 - \beta u)^5} \left\{ (1 - u^2) + \sin^2 \theta_0 \left[\left(\frac{3}{2} + \frac{1}{2} \beta^2 \right) u^2 - 2\beta u - \frac{1}{2}(1 - \beta^2) \right] \right\}.
\end{aligned}$$

Now introduce $x \equiv 1 - \beta u$; then

$$\begin{aligned}
\mathcal{P}(t') &= \frac{\dot{\beta}^2 e^2}{2c\beta^3} \int_{1-\beta}^{1+\beta} \frac{dx}{x^5} \left\{ (\beta^2 - 1 + 2x - x^2) \right. \\
&\quad \left. + \sin^2 \theta_0 \left[(3 + \beta^2) (1 - 2x + x^2) / 2 - 2\beta^2 (1 - x) - \beta^2 (1 - \beta^2) / 2 \right] \right\} \\
&= \frac{\dot{\beta}^2 e^2}{2c\beta^3} \int_{1-\beta}^{1+\beta} \frac{dx}{x^5} \left\{ (\beta^2 - 1 + 2x - x^2) \right. \\
&\quad \left. + \sin^2 \theta_0 \left[(3 - 4\beta^2 + \beta^4) / 2 + (-2 + \beta^2)x + (3 + \beta^2) x^2 / 2 \right] \right\} \\
&= \frac{\dot{\beta}^2 e^2}{2c\beta^3} \left\{ \left(\frac{\beta^2 - 1}{4x^4} + \frac{2}{3x^3} - \frac{1}{2x^2} \right) \right. \\
&\quad \left. + \sin^2 \theta_0 \left(\frac{(3/2 - \beta^2/2)(1 - \beta^2)}{4x^4} + \frac{\beta^2 - 3}{3x^3} + \frac{3 + \beta^2}{4x^2} \right) \right\} \\
&= \frac{\dot{\beta}^2 e^2 \gamma^6}{c\beta^3} \left\{ \left(-(\beta + \beta^3) + 2(3\beta + \beta^3)/3 - \beta(1 - \beta^2) \right) \right. \\
&\quad \left. + \sin^2 \theta_0 \left[(3 - \beta^2)(\beta + 3\beta^3)/2 + (-3 + \beta^2)(\beta + \beta^3/3) + (3 + \beta^2)(\beta - \beta^3)/2 \right] \right\} \\
&= \frac{\dot{\beta}^2 e^2 \gamma^6}{c\beta^3} \left\{ \frac{2}{3}\beta^3 - \frac{2}{3}\beta^5 \sin^2 \theta_0 \right\} = \frac{2\dot{\beta}^2 e^2 \gamma^6}{3c} (1 - \beta^2 \sin^2 \theta_0). \tag{50}
\end{aligned}$$

Putting the result back in terms of vectors and their products, we have

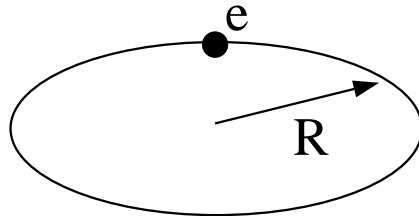
$$\mathcal{P} = \frac{2e^2 \gamma^6}{3c} [\dot{\beta}^2 - (\boldsymbol{\beta} \times \dot{\boldsymbol{\beta}})^2]. \tag{51}$$

This is the relativistic generalization of the Larmor formula.

It is instructive to look at some simple examples.

2.1.1 Example: Synchrotron

An electron moves in a circle of radius R at constant speed.



The acceleration is then entirely centripetal and has magnitude $\dot{\beta} = c\beta^2/R$. Then

$$\dot{\beta}^2 - (\boldsymbol{\beta} \times \dot{\boldsymbol{\beta}})^2 = \frac{c^2\beta^4}{R^2}(1 - \beta^2) = \frac{c^2\beta^4}{R^2\gamma^2}, \quad (52)$$

and so

$$\mathcal{P} = \frac{2e^2c\gamma^4\beta^4}{3R^2} \quad (53)$$

which can be rewritten in terms of the particle's kinetic energy as

$$\mathcal{P} = \frac{2e^2c}{3R^2} \left(\frac{E}{mc^2} \right)^4 \beta^4. \quad (54)$$

The energy emitted per cycle of the motion is $\Delta E = \mathcal{P}\tau$ where τ is the period of the motion, $\tau = 2\pi R/\beta c$. Hence,

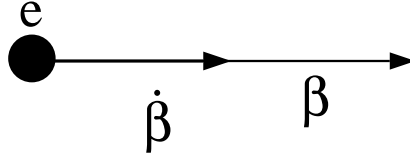
$$\Delta E = \frac{4\pi e^2\beta^3}{3R} \left(\frac{E}{mc^2} \right)^4. \quad (55)$$

An electron of energy $E = 500 \text{ Mev}$ in a synchrotron of radius $R = 10^2 \text{ cm}$ will radiate in each cycle an energy $\Delta E \approx 10^{-8} \text{ erg} \sim 10^4 \text{ ev}$ which is a non-trivial amount if one wants to increase the electron's energy or even to maintain it. Basically, one must apply an accelerating voltage of at least ten thousand volts during each cycle to break even, i.e., to maintain the electron's energy. This radiation is the reason why circular very-high-energy electron accelerators don't exist; however, the synchrotron is a great Xray source.

But if one wants to produce high-energy protons in a circular accelerator, that is a lot easier, especially if one is willing to make R rather large. For example, a 10 Tev proton in a machine of radius $R = 3 \times 10^6 \text{ cm}$, which is about nineteen miles, radiates away considerably less than 10^4 ev in one cycle.

2.1.2 Example: Linear Acceleration

An electron is accelerated in the direction of its velocity, $\boldsymbol{\beta} \parallel \dot{\boldsymbol{\beta}}$.



In this instance one trivially finds from Eq. (51) that

$$\mathcal{P} = \frac{2e^2\gamma^6}{3c}\dot{\beta}^2. \quad (56)$$

This doesn't look very encouraging (for an accelerator design) because of the factor of γ^6 , but against this one has the fact that the only acceleration the particle feels is the one produced by the fields acting to increase the particle's kinetic energy; in the case of the round accelerator, there is a large acceleration even for a particle of constant energy. Thus, in the case of a linear accelerator, there is much less acceleration to produce radiation. To make this point more clearly, let's write the acceleration in terms of the time rate of change of the particle's momentum,

$$\dot{\mathbf{v}} = \frac{d\mathbf{v}}{dt} = \frac{1}{m\gamma^3} \frac{d\mathbf{p}}{dt} \quad (57)$$

and so

$$\mathcal{P} = \frac{2e^2}{3m^2c^3} \left(\frac{d\mathbf{p}}{dt} \right)^2. \quad (58)$$

Now relate the rate of change of momentum to the rate of change of energy of the particle,

$$\frac{dp}{dt} = F = \frac{dE}{dx}, \quad (59)$$

and so

$$\mathcal{P} = \frac{2e^2}{3m^2c^3} \left(\frac{dE}{dx} \right)^2 \quad (60)$$

or

$$\frac{\mathcal{P}}{dE/dt} = \frac{2e^2}{3m^2c^3} \frac{dE/dx}{dx/dt} = \frac{2}{3\beta} \frac{(e^2/mc^2)}{mc^2} \frac{dE}{dx}, \quad (61)$$

which says that the power radiated away is quite negligible in comparison with the rate at which energy is being pumped into the particle unless one pumps the energy in at a rate (in space) dE/dx comparable to the rest energy of the particle, mc^2 , in a distance $r_c = e^2/mc^2$. This distance is rather small, being about $3 \times 10^{-13} \text{ cm}$ for an electron, and it is difficult to put so much energy into the particle in such a small distance (500,000 *volts* in a distance of 10^{-13} cm !). Hence particles in linear accelerators lose an insignificant amount of energy to radiation. The difficulty with such accelerators is that, like all things excepting round ones, they must sooner or later end.

3 Angular distribution of radiation

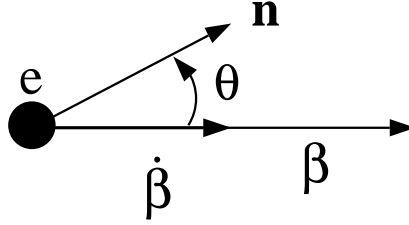
In this section we consider the angular distribution of the radiation emitted by an accelerated charge. On the basis of what we said about the potentials, we expect to find that it is strongly focussed in the forward direction, or parallel to the velocity, in a frame where the particle has a speed close to c . As at the end of the preceding section, we shall consider some particular examples. The basic equation is

$$\frac{d\mathcal{P}}{d\Omega} = \frac{e^2}{4\pi c} \frac{\left\{ \mathbf{n} \times [(\mathbf{n} - \boldsymbol{\beta}) \times \dot{\boldsymbol{\beta}}] \right\}^2}{(1 - \mathbf{n} \cdot \boldsymbol{\beta})^5}; \quad (62)$$

the right-hand side of this expression is to be evaluated at the retarded time t' .

3.1 Example: Parallel acceleration and velocity

Let a particle have parallel velocity and acceleration. If these define the z direction, and \mathbf{n} points at an angle θ to the z axis,

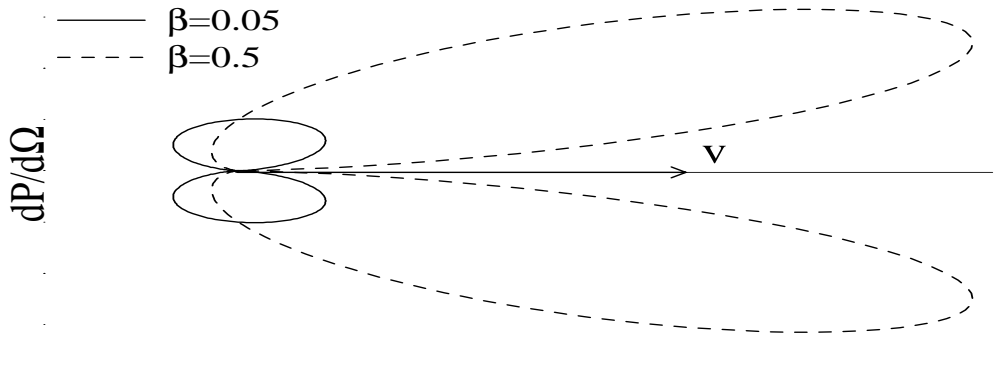


then one finds

$$\frac{d\mathcal{P}}{d\Omega} = \frac{e^2 \dot{\beta}^2}{4\pi c} \frac{\sin^2 \theta}{(1 - \beta \cos \theta)^5} = \frac{e^2 \dot{\mathbf{v}}^2}{4\pi c^3} \frac{\sin^2 \theta}{(1 - \beta \cos \theta)^5}, \quad (63)$$

and the total power is

$$\mathcal{P} = \frac{2e^2 \dot{\mathbf{v}}^2}{3c^3} \gamma^6$$



For β close to one, this expression provides dramatic confirmation of the inadequacy of the multipole expansion for describing radiation from accelerated charges. This radiation pattern is sharply peaked close to the z axis and is actually zero precisely on the axis. One would have to keep many multipole terms to produce such a distribution.

We can find various basic properties which characterize the distribution. It has a

maximum at a value of θ or of $u \equiv \cos \theta$ which we can find by differentiating:

$$0 = \frac{d}{du} \left(\frac{1 - u^2}{(1 - \beta u)^5} \right) = -\frac{2u}{(1 - \beta u)^5} + \frac{5\beta(1 - u^2)}{(1 - \beta u)^6}, \quad (64)$$

or

$$-2u(1 - \beta u) + 5\beta(1 - u^2) = 0. \quad (65)$$

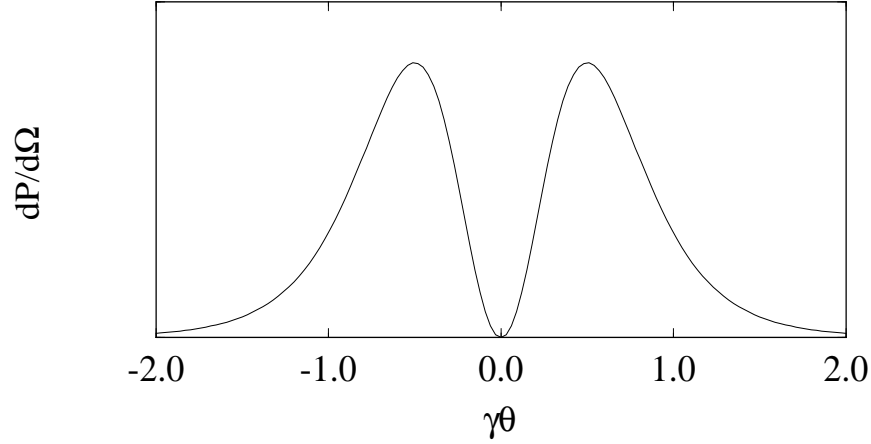
This quadratic equation has the solution

$$u = \frac{1}{3\beta}[\sqrt{1 + 15\beta^2} - 1] \quad \text{or} \quad \theta_m = \arccos \left[\frac{1}{3\beta}(\sqrt{1 + 15\beta^2} - 1) \right]. \quad (66)$$

Further, $\beta^2 = 1 - 1/\gamma^2$, so for β close to one, $\beta \approx 1 - 1/2\gamma^2$ and

$$\begin{aligned} \theta_m &\approx \arccos \left\{ \frac{1}{3} \left(1 + \frac{1}{2\gamma^2} \right) \left[4 \left(1 - \frac{15}{32\gamma^2} \right) - 1 \right] \right\} \\ &\approx \arccos \left(1 - \frac{1}{8\gamma^2} \right) \approx \frac{1}{2\gamma}. \end{aligned} \quad (67)$$

Also, the power distribution becomes, in the relativistic limit,

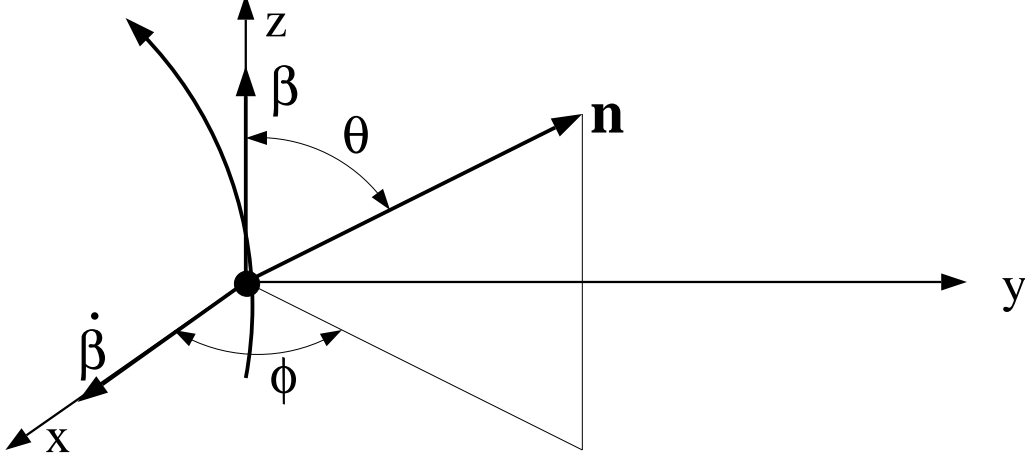


$$\begin{aligned} \frac{dP}{d\Omega} &= \frac{e^2 \dot{v}^2}{4\pi c^3} \frac{\theta^2}{(1 - \beta + \beta\theta^2/2)^5} = \frac{e^2 \dot{v}^2}{4\pi c^3} \frac{32\theta^2}{[2(1 - \beta) + \beta\theta^2]^5} \\ &= \frac{e^2 \dot{v}^2}{4\pi c^3} \frac{32\gamma^5 \theta^2}{(1 + \gamma^2 \theta^2)^5} = \frac{8e^2 \dot{v}^2}{\pi c^3} \gamma^3 \frac{\gamma^2 \theta^2}{(1 + \gamma^2 \theta^2)^5} \end{aligned} \quad (68)$$

This may be integrated to find the rms angle

$$\langle \theta^2 \rangle = \frac{1}{\gamma} = \frac{mc^2}{E} \quad (69)$$

3.2 Example: Acceleration Perpendicular to Velocity



A particle in instantaneously circular motion has its acceleration perpendicular to its velocity. Let $\boldsymbol{\beta} = \beta \boldsymbol{\epsilon}_3$ and $\dot{\boldsymbol{\beta}} = \dot{\beta} \boldsymbol{\epsilon}_1$. Then, letting θ and ϕ specify the direction from the particle (at time t') to the observation point \mathbf{x} , we have

$$\boldsymbol{\epsilon}_3 = \cos \theta \mathbf{n} - \sin \theta \boldsymbol{\theta} \quad \text{and} \quad \boldsymbol{\epsilon}_1 = \sin \theta \cos \phi \mathbf{n} + \cos \theta \cos \phi \boldsymbol{\theta} - \sin \phi \boldsymbol{\phi}. \quad (70)$$

Using these conventions we can work out the relevant vector products:

$$(\mathbf{n} - \boldsymbol{\beta}) \times \dot{\boldsymbol{\beta}} = \dot{\beta} [\cos \theta \cos \phi \boldsymbol{\phi} + \sin \phi \boldsymbol{\theta} - \boldsymbol{\beta} (\sin \theta \sin \phi \mathbf{n} + \cos \theta \sin \phi \boldsymbol{\theta} + \cos \phi \boldsymbol{\phi})] \quad (71)$$

and so

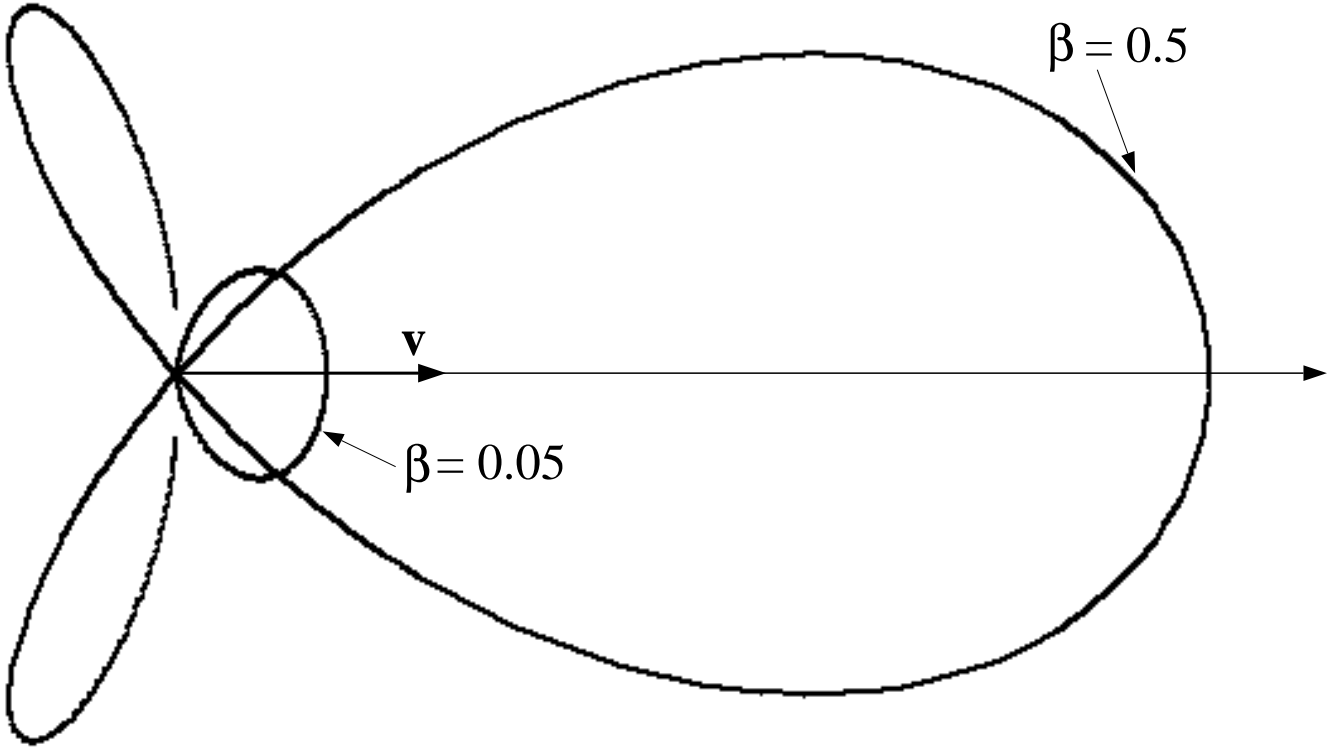
$$\begin{aligned} [\mathbf{n} \times ((\mathbf{n} - \boldsymbol{\beta}) \times \dot{\boldsymbol{\beta}})]^2 &= \dot{\beta}^2 [\cos^2 \phi (\beta - \cos \theta)^2 + \sin^2 \phi (1 - \beta \cos \theta)^2] \\ &= \dot{\beta}^2 [\beta^2 \cos^2 \phi - 2\beta \cos \theta \cos^2 \phi + \cos^2 \theta \cos^2 \phi + \sin^2 \phi \\ &\quad - 2\beta \cos \theta \sin^2 \phi + \beta^2 \cos^2 \theta \sin^2 \phi] \\ &= \dot{\beta}^2 [\cos^2 \phi (\beta^2 + \cos^2 \theta - 1 - \beta^2 \cos^2 \theta) + (1 - \beta \cos \theta)^2] \end{aligned} \quad (72)$$

Hence

$$\begin{aligned} \frac{d\mathcal{P}}{d\Omega} &= \frac{e^2 \dot{v}^2}{4\pi c^3} \left\{ 1 - \frac{\cos^2 \phi (1 - \beta^2) \sin^2 \theta}{(1 - \beta \cos \theta)^2} \right\} \frac{1}{(1 - \beta \cos \theta)^3} \\ &= \frac{e^2 \dot{v}^2}{4\pi c^3 (1 - \beta \cos \theta)^3} \left\{ 1 - \frac{\cos^2 \phi \sin^2 \theta}{\gamma^2 (1 - \beta \cos \theta)^2} \right\}. \end{aligned} \quad (73)$$

In this case we find that there is radiation in the forward direction, or $\theta = 0$. In fact the main peak in the radiation is in this direction. If one makes a small angle expansion of the power distribution in the case of highly relativistic particles, he will find the result

$$\frac{d\mathcal{P}}{d\Omega} = \frac{2e^2\dot{v}^2}{\pi c^3} \frac{\gamma^6}{(1 + \gamma^2\theta^2)^3} \left\{ 1 - \frac{4\gamma^2\theta^2 \cos^2\phi}{(1 + \gamma^2\theta^2)^2} \right\}. \quad (74)$$



Further, the total radiated power is

$$\mathcal{P} = \frac{2e^2\dot{v}^2}{4\pi c^3} \gamma^4. \quad (75)$$

3.3 Comparison of Examples

From the expressions for the total power produced in each of our two examples, we see that for a given magnitude of acceleration, there is a factor of γ^2 more radiation produced when the acceleration is parallel to the velocity than when it is perpendicular. This is a misleading statement in some sense because what is actually applied

to a particle is not an acceleration but a force, and a given force will produce quite different accelerations when applied perpendicular and parallel to the velocity. For a force applied parallel to the velocity,

$$F = \frac{dp}{dt} = m\gamma^3 \frac{dv}{dt} \quad (76)$$

and for a force applied perpendicular to \mathbf{v} ,

$$F = \frac{dp}{dt} = m\gamma \frac{dv}{dt}. \quad (77)$$

Hence the powers produced in the two cases, expressed in terms of the forces or dp/dt , are

$$\mathcal{P}_\perp = \frac{2e^2}{3m^2c^3} \gamma^2 \left(\frac{dp}{dt} \right)^2 \quad \text{and} \quad \mathcal{P}_\parallel = \frac{2e^2}{3m^2c^3} \left(\frac{dp}{dt} \right)^2. \quad (78)$$

Thus, for a given applied force, γ^2 more radiation is produced if it is perpendicular to \mathbf{v} than if it is parallel. Of course, if the force comes from a magnetic field, then it has to be perpendicular to the velocity.

3.4 Radiation of an Ultrarelativistic Charged Particle

The radiation emitted at any instant from an accelerating charged particle may be decomposed into components coming from the parallel and perpendicular accelerations of the particle. From the discussion above, it is clear that the radiation of an ultrarelativistic $\gamma \gg 1$ particle is dominated by the perpendicular acceleration component. Thus the radiation is approximately the same as that emitted by a particle moving instantaneously in a circle of radius

$$\rho = \frac{v^2}{\dot{v}_\perp} \quad (79)$$

Since the angular width of the pulse is $\sim 1/\gamma$, the particle will travel a distance

$$d \sim \rho \Delta\theta \sim \frac{\rho}{\gamma} \quad (80)$$

while illuminating the observer for a time

$$\Delta t \sim \frac{d}{v} \sim \frac{\rho}{\gamma v}. \quad (81)$$

If we assume that pattern of radiation is roughly that of a coiled beam of rectangular cross-section then the front edge of the beam will travel

$$D = c\Delta t \sim \frac{\rho c}{\gamma v} = \frac{\rho}{\gamma\beta} \quad (82)$$

in time Δt ; whereas, the trailing edge will be a distance

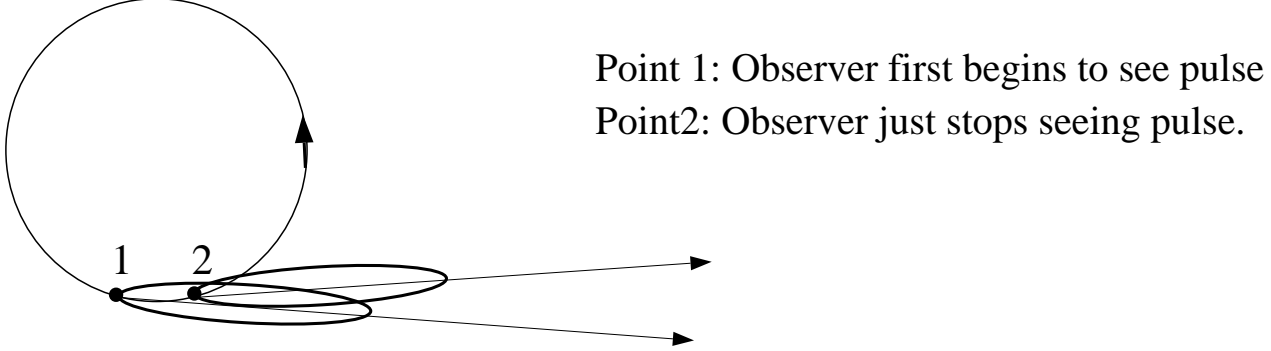
$$L = D - d \approx \left(\frac{\beta - \beta^2}{\beta^2} \right) \frac{\rho}{\gamma} \approx \left(1 - \frac{1}{2\gamma^2} - 1 + \frac{1}{\gamma^2} \right) \frac{\rho}{\gamma} = \frac{\rho}{2\gamma^3} \quad (83)$$

behind the front edge since the charge moves the distance d in the same time interval. Thus the pulse width is roughly L in space or L/c in time.

This derivation (straight out of Jackson) raises about as many questions as it answers. Perhaps the most fundamental is this: why is Δt different than the observed width of the pulse L/c . The difference is that Δt is really a difference of retarded times; whereas, L/c is in the time frame of the observer. Lets repeat this calculation more carefully, distinguishing between the observer's time and the particles (retarded) time.

The pulse width for a distant observer would be

$$\frac{L}{c} = t_2 - t_1 = (t'_2 + R_2/c) - (t'_1 + R_1/c) = \Delta t - (R_1 - R_2)/c. \quad (84)$$

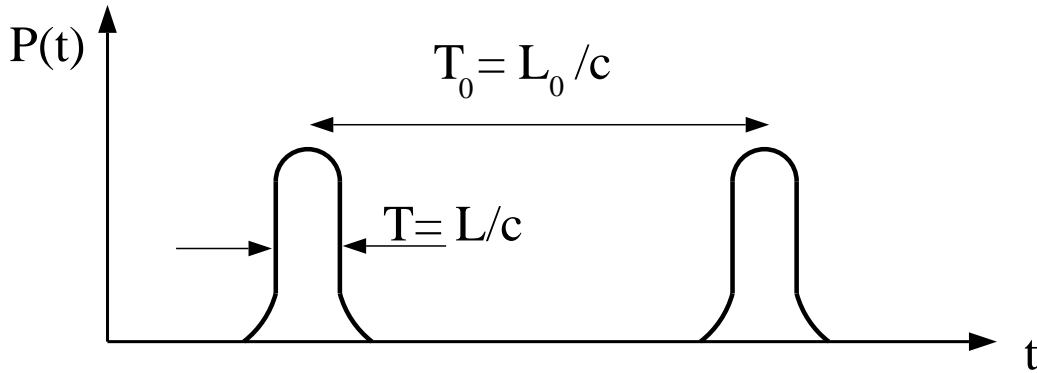


Since $R_1 - R_2 \approx v' \Delta t$, we have

$$\frac{L}{c} = t_2 - t_1 = \Delta t(1 - v/c) \approx \frac{\Delta t}{2\gamma^2} = \frac{\rho}{2\gamma^3}, \quad (85)$$

where the right hand side is to be evaluated in the retarded time.

An observer would receive periodic pulses of width L/c in time.



From the principles of Fourier decomposition, we can also estimate the type of radiation the observer would receive. From the uncertainty principle $\Delta t \Delta \omega \sim 1$, the pulse would contain components up to a cutoff³ of roughly

$$\omega_c \sim \frac{c}{L} \sim \frac{c}{\rho} \gamma^3 \quad (86)$$

³Higher frequency components would yield a narrower pulse

However, the frequency ω_0 of the circular motion is $v/\rho \approx c/\rho$. Thus, a broad spectrum of radiation is emitted up to γ^3 times the fundamental frequency of the rotation. This kind of radiation, called *synchrotron radiation* provides a good, more or less continuous, source of radiation in the range from visible and ultraviolet to soft X-rays.

Thus in a 200 MeV electron synchrotron where $\gamma \approx 400$ and the fundamental frequency is $\omega_0 \approx 3 \times 10^8$, the frequency of the emitted radiation extends to 2×10^{16} . In a 10 GeV synchrotron, x-rays can be produced. Thus synchrotrons can be used as high intensity radiation sources. Synchrotrons are now even being used in industry as radiation sources for x-ray lithography⁴.

4 Frequency Distribution of the Radiated Energy

The radiation produced by rotating charges is predominantly at the fundamental frequency ω_0 of the motion with much smaller amounts at integral multiples, or harmonics, of this frequency. The expansion parameter in the problem is $k_0 a = \omega_0 a/c \sim v/c = \beta$, and the energy emitted at higher frequencies than the fundamental is proportional to some power of this parameter. For $\beta \ll 1$, this energy will be relatively small. But if $\beta \sim 1$, there will be a significant fraction of the total radiated energy appearing at higher frequencies.

There is a simple and instructive way to see roughly how many harmonics will contribute to the radiation. For a relativistic particle, $d\mathcal{P}/d\Omega$ is a peaked distribution which has a width in angle $\delta\theta \sim 1/\gamma$. If the particle is, e.g., travelling in a circle with a frequency of motion ω_0 , then the particle sweeps through an angle $\delta\theta$ in a time $\delta t' \sim \delta\theta/\omega_0 = 1/\gamma\omega_0$. This is of the order of the duration of the pulse observed at some fixed point in space, but measured in units of the time at the source. The time

⁴**Physics Today**, October 1991.

which passes at the location of the observer is

$$\delta t = \delta t' \frac{dt}{dt'} = \kappa \delta t'. \quad (87)$$

Further, during the pulse, \mathbf{n} is parallel to $\boldsymbol{\beta}$, $\mathbf{n} \cdot \boldsymbol{\beta} = \beta$ with corrections of order $1/\gamma^2$ and so $\kappa \sim 1 - \beta[1 + \mathcal{O}(1/\gamma^2)] \sim 1/\gamma^2$. Hence $\delta t \sim 1/\gamma^3 \omega_0$. Now, a pulse which lasts a time δt at a point must contain in it frequencies of order⁵ $1/\delta t$. Thus the pulse which our observer sees must have in particular frequencies of order $\omega \sim \gamma^3 \omega_0$. If the particle is highly relativistic, $\gamma \gg 1$, then the typical frequencies in the pulse will be much larger than the frequency of the particle's motion, or ω_0 ; they will be γ^3 times as large, meaning that many harmonics must contribute to the pulse.

4.1 Continuous Frequency Distribution

To make our analysis of the radiation more quantitative we are going to consider the Fourier transforms in time of the fields at an observation point \mathbf{x} . Start from the expression for the angular distribution of radiated power:

$$\frac{d\mathcal{P}}{d\Omega} = \frac{c}{4\pi} (\mathbf{E}(\mathbf{x}, t) \times \mathbf{B}(\mathbf{x}, t)) \cdot [\mathbf{n}R^2]_{ret}. \quad (88)$$

Using just the radiation fields, noting that they are transverse to \mathbf{n} far away from the source, and making use the fact that $\mathbf{B}(\mathbf{x}, t) = [\mathbf{n}]_{ret} \times \mathbf{E}(\mathbf{x}, t)$, we can write

$$\frac{d\mathcal{P}}{d\Omega} = \frac{c}{4\pi} \{[R]_{ret} \mathbf{E}(\mathbf{x}, t)\}^2. \quad (89)$$

For simplicity of notation, define

$$\mathbf{a}(t) \equiv \sqrt{\frac{c}{4\pi}} [R]_{ret} \mathbf{E}(\mathbf{x}, t). \quad (90)$$

Then, in terms of $\mathbf{a}(t)$ the angular distribution of radiated power is

$$\frac{d\mathcal{P}}{d\Omega} = [\mathbf{a}(t)]^2. \quad (91)$$

⁵In quantum theory this statement would be called the uncertainty principle.

Further, from Eq. (29),

$$\mathbf{a}(t) = \sqrt{\frac{e^2}{4\pi c}} \left[\frac{\mathbf{n} \times [(\mathbf{n} - \boldsymbol{\beta}) \times \dot{\boldsymbol{\beta}}]}{\kappa^3} \right]_{ret}. \quad (92)$$

Introduce also the Fourier transform⁶ $\mathbf{a}(\omega)$ of $\mathbf{a}(t)$:

$$\mathbf{a}(t) \equiv \frac{1}{\sqrt{2\pi}} \int d\omega e^{-i\omega t} \mathbf{a}(\omega); \quad \mathbf{a}(\omega) = \frac{1}{\sqrt{2\pi}} \int dt e^{i\omega t} \mathbf{a}(t). \quad (93)$$

Using the Fourier integral (and its complex conjugate) for $\mathbf{a}(t)$ in Eq. (91), we can write the power distribution as

$$\frac{d\mathcal{P}(t)}{d\Omega} = \frac{1}{2\pi} \int d\omega d\omega' \mathbf{a}(\omega) e^{-i\omega t} \cdot \mathbf{a}^*(\omega') e^{i\omega' t}. \quad (94)$$

Integrating over all t , and thereby generating an expression for the angular distribution of the total radiated energy, we find

$$\begin{aligned} \frac{dW}{d\Omega} &\equiv \int_{-\infty}^{\infty} dt \frac{d\mathcal{P}(t)}{d\Omega} = \int d\omega d\omega' \delta(\omega - \omega') \mathbf{a}(\omega) \cdot \mathbf{a}^*(\omega') \\ &= \int_{-\infty}^{\infty} d\omega |\mathbf{a}(\omega)|^2 = \int_0^{\infty} d\omega \left\{ |\mathbf{a}(\omega)|^2 + |\mathbf{a}(-\omega)|^2 \right\}. \end{aligned} \quad (95)$$

The reality of $\mathbf{a}(t)$ demands that $\mathbf{a}^*(-\omega) = \mathbf{a}(\omega)$, and so the two terms in the integrand are identical:

$$\frac{dW}{d\Omega} = 2 \int_0^{\infty} d\omega |\mathbf{a}(\omega)|^2 \equiv \int_0^{\infty} d\omega \frac{dI(\omega)}{d\Omega}. \quad (96)$$

The integrand, $dI/d\Omega$, is interpreted as the total radiation received per unit frequency per unit solid angle during the entire pulse of radiation. It is simply the square of $\mathbf{a}(\omega)$; further $\mathbf{a}(\omega)$ is

$$\begin{aligned} \mathbf{a}(\omega) &= \sqrt{\frac{e^2}{4\pi c}} \frac{1}{\sqrt{2\pi}} \int dt e^{i\omega t} \left[\frac{\mathbf{n} \times [(\mathbf{n} - \boldsymbol{\beta}) \times \dot{\boldsymbol{\beta}}]}{\kappa^3} \right]_{ret} \\ &= \sqrt{\frac{e^2}{8\pi^2 c}} \int dt' \left[\frac{\mathbf{n} \times [(\mathbf{n} - \boldsymbol{\beta}) \times \dot{\boldsymbol{\beta}}]}{\kappa^2} \right] e^{i\omega(t'+R(t')/c)} \end{aligned} \quad (97)$$

⁶The field $\mathbf{a}(t)$ depends on \mathbf{x} as well as on t ; we suppress the former dependence as it is not of interest at present.

It's *déjà vu* all over again. Recalling the typical equations generated in Chapter 9, we can see that this one probably can be obtained without great difficulty from what we did there. Proceeding in familiar fashion, then, let us look at this integral in the far zone which means approximate R in the exponent by

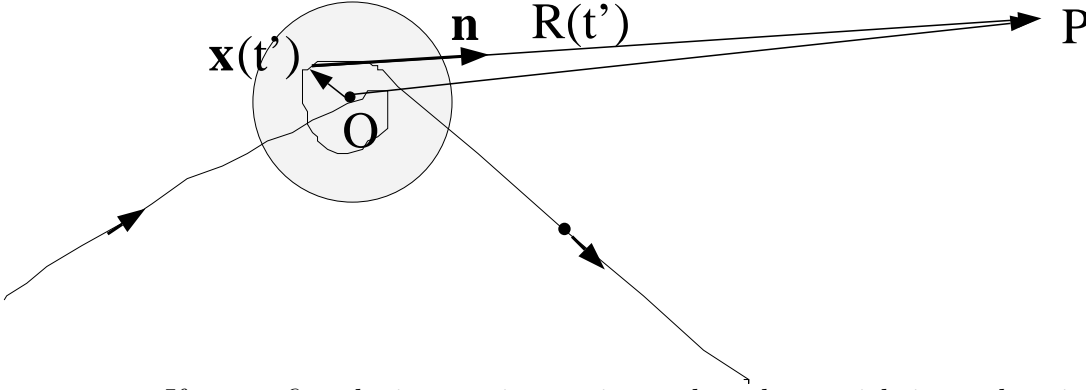
$$R = |\mathbf{x} - \mathbf{x}(t')| \approx r - \mathbf{n} \cdot \mathbf{x}(t') \quad (98)$$

and so

$$\mathbf{a}(\omega) = \sqrt{\frac{e^2}{8\pi^2 c}} \int dt' e^{i\omega t'} e^{i\omega r/c} e^{-i\omega \mathbf{n} \cdot \mathbf{x}(t')/c} \left[\frac{\mathbf{n} \times [(\mathbf{n} - \boldsymbol{\beta}) \times \dot{\boldsymbol{\beta}}]}{\kappa^2} \right] \quad (99)$$

or

$$\mathbf{a}(\omega) = e^{i\omega r/c} \sqrt{\frac{e^2}{8\pi^2 c}} \int dt' e^{i\omega(t' - \mathbf{n} \cdot \mathbf{x}(t')/c)} \left[\frac{\mathbf{n} \times [(\mathbf{n} - \boldsymbol{\beta}) \times \dot{\boldsymbol{\beta}}]}{\kappa^2} \right]. \quad (100)$$



If we confine the integration to times when the particle is accelerating, i.e. confine it to the shaded region above, then we can put this integral into a form such that the integrand involves $\boldsymbol{\beta}$ but not $\dot{\boldsymbol{\beta}}$; that can be done by, in essence, a parts integration.

First, consider

$$\frac{d}{dt'} \left(\frac{\mathbf{n} \times (\mathbf{n} \times \boldsymbol{\beta})}{\kappa} \right) = -\frac{\mathbf{n} \times (\mathbf{n} \times \boldsymbol{\beta})}{\kappa^2} (-\mathbf{n} \cdot \dot{\boldsymbol{\beta}}) + \frac{\mathbf{n} \times (\mathbf{n} \times \dot{\boldsymbol{\beta}})}{\kappa^2} (1 - \mathbf{n} \cdot \boldsymbol{\beta}) \quad (101)$$

where we have not kept derivatives of \mathbf{n} because they give corrections of relative order $|\mathbf{x}(t')/R|$. Now group terms as follows:

$$\frac{d}{dt'} \left(\frac{\mathbf{n} \times (\mathbf{n} \times \boldsymbol{\beta})}{\kappa} \right) = \frac{\mathbf{n} \times (\mathbf{n} \times \dot{\boldsymbol{\beta}})}{\kappa^2} + \frac{\mathbf{n} \times \left\{ \mathbf{n} \times [(\mathbf{n} \cdot \dot{\boldsymbol{\beta}})\boldsymbol{\beta} - (\mathbf{n} \cdot \boldsymbol{\beta})\dot{\boldsymbol{\beta}}] \right\}}{\kappa^2}$$

$$\begin{aligned}
&= \frac{\mathbf{n} \times (\mathbf{n} \times \dot{\boldsymbol{\beta}})}{\kappa^2} + \frac{\mathbf{n} \times \left\{ \mathbf{n} \times [\mathbf{n} \times (\boldsymbol{\beta} \times \dot{\boldsymbol{\beta}})] \right\}}{\kappa^2} \\
&= \frac{\mathbf{n} \times [(\mathbf{n} - \boldsymbol{\beta}) \times \dot{\boldsymbol{\beta}}]}{\kappa^2}
\end{aligned} \tag{102}$$

where we have used the fact that

$$\mathbf{n} \times \left\{ \mathbf{n} \times [\mathbf{n} \times (\boldsymbol{\beta} \times \dot{\boldsymbol{\beta}})] \right\} = -\mathbf{n} \times (\boldsymbol{\beta} \times \dot{\boldsymbol{\beta}}). \tag{103}$$

Using this identity to do a parts integration of the expression for $\mathbf{a}(\omega)$ we find (again, drop terms proportional to powers of $1/R$)

$$\begin{aligned}
\mathbf{a}(\omega) &= -e^{i\omega r/c} \sqrt{\frac{e^2}{8\pi^2 c}} \int dt' i\omega (1 - \mathbf{n} \cdot \boldsymbol{\beta}) \left(\frac{\mathbf{n} \times (\mathbf{n} \times \boldsymbol{\beta})}{\kappa} \right) e^{i\omega(t' - \mathbf{n} \cdot \mathbf{x}(t')/c)} \\
&= -\sqrt{\frac{e^2}{8\pi^2 c}} i\omega e^{i\omega r/c} \int dt' [\mathbf{n} \times (\mathbf{n} \times \boldsymbol{\beta})] e^{i\omega(t' - \mathbf{n} \cdot \mathbf{x}(t')/c)}.
\end{aligned} \tag{104}$$

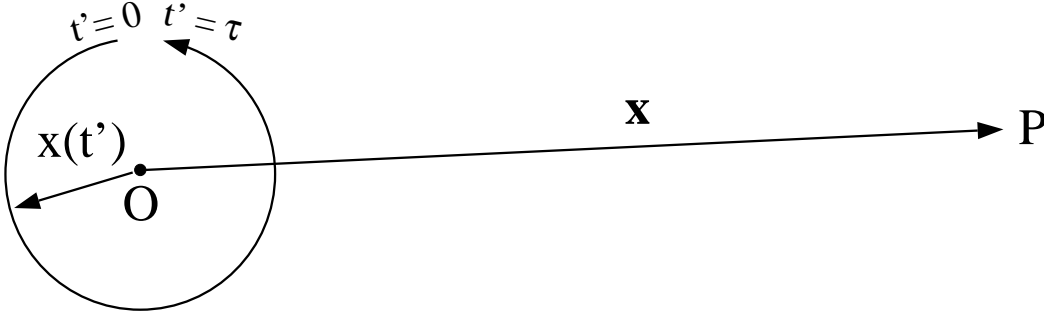
Combining this result and Eq. (96), we find for the radiated energy per unit frequency per unit solid angle

$$\frac{dI(\omega)}{d\Omega} = \frac{e^2 \omega^2}{4\pi^2 c} \left| \int dt' [\mathbf{n} \times (\mathbf{n} \times \boldsymbol{\beta})] e^{i\omega(t' - \mathbf{n} \cdot \mathbf{x}(t')/c)} \right|^2. \tag{105}$$

In this derivation we were not careful when doing the parts integration, meaning that we did not worry about whether the terms involving the integrand evaluated at the endpoints of the interval of integration contribute to the result. Consequently, in any application of Eq. (105), one should check to see whether this assumption is justified; there are occasions when it is not and then, naturally, the contributions from the endpoint(s) must be included. Our formulation of $dI/d\Omega$ is most appropriate for a source with an open orbit in which case the natural limits on the integration are $\pm\infty$, and there is usually no difficulty in ignoring the contributions from the endpoints where the particle is far away and unaccelerated; however, that is not enough to guarantee that the endpoints contribute nothing.

4.2 Discrete Frequency Distribution

For truly cyclic motion, it is more convenient and perhaps more sensible from a physical point of view to set up a Fourier series to express the frequency distribution of the radiation.



The point is that for non-cyclic motion, the distribution of radiation in frequency will be continuous and the integral formula we derived above is appropriate for expressing this distribution. But for cyclic motion, the radiation will be distributed in frequency space only at harmonics or multiples of the fundamental frequency of the motion and so a sum or series expansion of $dI(\omega)/d\Omega$ is more appropriate for expressing the distribution. We shall now set up this sum. To get started, suppose that the period of the motion is $\tau' = 2\pi/\omega_0$. Then, letting the period measured by an observer at a point \mathbf{x} be τ , one can show that $\tau = \tau'$. That is, a time t for the observer and the corresponding retarded time t' are related by

$$t = t' + R(t')/c. \quad (106)$$

One period later in the life of the source, its time has increased to $t' + \tau'$ and the signal emitted by the source at this time will reach the observer at a time $t + \tau$ which is

$$t + \tau = t' + \tau' + R(t' + \tau')/c. \quad (107)$$

But $R(t' + \tau') = R(t')$ for the cyclic motion so

$$t + \tau = t' + \tau' + R(t')/c. \quad (108)$$

Comparing Eqs. (106) and (108), we see that $\tau = \tau'$.

The radiation fields produced by this cyclic motion of a charged particle will also be periodic. Hence where we formerly had a Fourier integral for $\mathbf{a}(t)$, we now have a Fourier series, a sum over frequencies which are integral multiples of the sources's frequency,

$$\mathbf{a}(t) = \sqrt{\frac{c}{4\pi}} [R\mathbf{E}(t)] \equiv \sum_{n=-\infty}^{\infty} \mathbf{a}_n e^{-in\omega_0 t} \quad (109)$$

with

$$\mathbf{a}_n = \frac{\omega_0}{2\pi} \int_0^{2\pi/\omega_0} dt \mathbf{a}(t) e^{in\omega_0 t}. \quad (110)$$

The energy received during one cycle, per unit solid angle, at some point \mathbf{x} is the integral of $d\mathcal{P}/d\Omega$ over one period. We shall write it as $dW/d\Omega$,

$$\begin{aligned} \frac{dW}{d\Omega} &\equiv \int_0^{2\pi/\omega_0} dt \frac{d\mathcal{P}(t)}{d\Omega} = \int_0^{2\pi/\omega_0} dt |\mathbf{a}(t)|^2 \\ &= \sum_{n,m} \int_0^{2\pi/\omega_0} dt (\mathbf{a}_n \cdot \mathbf{a}_m^*) e^{-i(n-m)\omega_0 t} = \frac{2\pi}{\omega_0} \sum_{n=-\infty}^{\infty} |\mathbf{a}_n|^2 \\ &= \frac{4\pi}{\omega_0} \sum_{n=1}^{\infty} |\mathbf{a}_n|^2. \end{aligned} \quad (111)$$

Notice that the $n = 0$ term has been discarded in the final expression; that is okay because there is no radiation at zero frequency. The radiated energy at the frequency $n\omega_0$ is determined by \mathbf{a}_n , which is

$$\begin{aligned} \mathbf{a}_n &= \frac{\omega_0}{2\pi} \int_0^{2\pi/\omega_0} dt \left[\frac{\mathbf{n} \times [(\mathbf{n} - \boldsymbol{\beta}) \times \dot{\boldsymbol{\beta}}]}{\kappa^3} \right]_{ret} \sqrt{\frac{e^2}{4\pi c}} e^{in\omega_0 t} \\ &= \frac{\omega_0}{2\pi} \sqrt{\frac{e^2}{4\pi c}} \int_0^{2\pi/\omega_0} dt' \frac{\mathbf{n} \times [(\mathbf{n} - \boldsymbol{\beta}) \times \dot{\boldsymbol{\beta}}]}{\kappa^2} e^{in\omega_0(t'+R(t')/c)}. \end{aligned} \quad (112)$$

We can now do the same integration by parts that led to Eq. (104) and find

$$\mathbf{a}_n = -\frac{\omega_0}{2\pi} \sqrt{\frac{e^2}{4\pi c}} in\omega_0 \int_0^{2\pi/\omega_0} dt' [\mathbf{n} \times (\mathbf{n} \times \boldsymbol{\beta})] e^{in\omega_0(t'+R(t')/c)}, \quad (113)$$

with no contribution ever from the end points of the interval as they correspond to the same point on the periodic orbit of the particle. From Eq. (113), $|\mathbf{a}_n|^2$ follows

easily and hence $dW/d\Omega$ from Eq. (111). Notice also that the time-averaged power distribution is just

$$\left(\frac{d\mathcal{P}}{d\Omega}\right)_{ave} = \frac{dW}{d\Omega} / \left(\frac{2\pi}{\omega_0}\right) \quad (114)$$

4.3 Examples

4.3.1 A Particle in Instantaneous Circular Motion

Suppose we have a particle in instantaneous circular motion, meaning that its acceleration is, at least temporarily, perpendicular to its velocity. Any particle subjected to a magnetic field but no electric field will satisfy this condition. We might do the calculation by supposing that the motion is truly periodic and circular and using the relatively easily applied Fourier series approach just developed. It is more difficult and therefore more challenging to use the Fourier integral approach. Let's try the latter.

First, we want to characterize the orbit of the particle. A circular orbit at constant speed can be described as

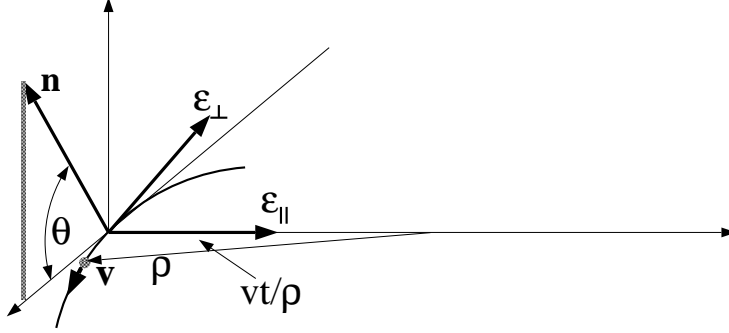
$$\boldsymbol{\beta} = \beta[(\cos \omega_0 t')\boldsymbol{\epsilon}_1 + (\sin \omega_0 t')\boldsymbol{\epsilon}_2] \quad (115)$$

and, for sufficiently small times t' , meaning $\omega_0 t' \ll 1$, we have

$$\boldsymbol{\beta} \approx \beta(\boldsymbol{\epsilon}_1 + \omega_0 t' \boldsymbol{\epsilon}_2) \quad (116)$$

with corrections of order $(\omega_0 t')^2$. Let the observer be located in the x - z plane; then $\mathbf{n} = \cos \theta \boldsymbol{\epsilon}_1 + \sin \theta \boldsymbol{\epsilon}_3$ where $\theta \ll 1$ if the observer is to experience the strong pulse of radiation that the particle emits in the forward direction. We know that the times of importance at the source for this pulse at the position of the observer are of order $t' \sim 1/\omega_0 \gamma$. Consequently, in our approximation for $\boldsymbol{\beta}$, we lose corrections of relative order $(\omega_0 t')^2 \sim 1/\gamma^2$ in each of the components.

Let's work out $\mathbf{n} \times (\mathbf{n} \times \boldsymbol{\beta})$. Define a unit vector $\boldsymbol{\epsilon}_\perp \equiv \mathbf{n} \times \boldsymbol{\epsilon}_2 = \cos \theta \boldsymbol{\epsilon}_3 - \sin \theta \boldsymbol{\epsilon}_1$; it will prove to be useful.



$$\mathbf{n} \times \boldsymbol{\beta} = \beta[(\mathbf{n} \times \boldsymbol{\epsilon}_1) + \omega_0 t'(\mathbf{n} \times \boldsymbol{\epsilon}_2)] = \beta(\sin \theta \boldsymbol{\epsilon}_2 + \omega_0 t' \boldsymbol{\epsilon}_\perp); \quad (117)$$

and

$$\mathbf{n} \times (\mathbf{n} \times \boldsymbol{\beta}) = \beta(\sin \theta \boldsymbol{\epsilon}_\perp - \omega_0 t' \boldsymbol{\epsilon}_2). \quad (118)$$

Further, the interesting range of θ is of order $1/\gamma \ll 1$, so let us replace $\sin \theta$ by θ . Further, let $\boldsymbol{\epsilon}_2$ be designated $\boldsymbol{\epsilon}_\parallel$. Then, using this equation in Eq. (105) and setting $n\omega_0 = \omega$, we find

$$\frac{dI(\omega)}{d\Omega} = \frac{e^2 \omega^2}{4\pi^2 c} \left| \int dt' (\beta \theta \boldsymbol{\epsilon}_\perp - \omega_0 t' \boldsymbol{\epsilon}_\parallel) e^{i\omega(t' - \mathbf{n} \cdot \mathbf{x}(t')/c)} \right|^2. \quad (119)$$

We need to evaluate the exponent in order to complete the integral. Consider

$$\begin{aligned} \mathbf{n} \cdot \mathbf{x}(t') &= \cos \theta (\boldsymbol{\epsilon}_1 \cdot \mathbf{x}(t')) = \frac{c\beta}{\omega_0} \cos \theta \sin \omega_0 t' \\ &\approx \frac{c\beta}{\omega_0} \left(1 - \frac{\theta^2}{2}\right) \left(\omega_0 t' - \frac{1}{6}(\omega_0 t')^3\right); \end{aligned} \quad (120)$$

Notice that we have kept the leading term and corrections to it of order $1/\gamma^2$. Basically, we have kept all phases of order unity when $\omega_0 t' \sim 1/\gamma$, $\theta \sim 1/\gamma$, and $\omega \sim \omega_0 \gamma^3$, these being what we believe to be the important ranges of t' , ω , and θ . Hence the total phase is, to the order indicated,

$$\begin{aligned} \omega[t' - \mathbf{n} \cdot \mathbf{x}(t')/c] &= \omega t' - \frac{\omega\beta}{\omega_0} \omega_0 t' \left(1 - \frac{\theta^2}{2}\right) + \frac{\omega\beta}{\omega_0} \frac{1}{6} (\omega_0 t')^3 \left(1 - \frac{\theta^2}{2}\right) \\ &= \omega t' \left(1 - \beta + \frac{\beta\theta^2}{2}\right) + \frac{\omega\beta}{6\omega_0} (\omega_0 t')^3 \left(1 - \frac{\theta^2}{2}\right) \\ &\approx \omega t' \left(\frac{1}{2\gamma^2} + \frac{\theta^2}{2}\right) + \frac{\omega}{6\omega_0} (\omega_0 t')^3 \end{aligned}$$

$$= \left(\frac{\omega}{2\omega_0} \right) \left[\omega_0 t' \left(\frac{1}{\gamma^2} + \theta^2 \right) + \frac{1}{3} (\omega_0 t')^3 \right]. \quad (121)$$

With the foregoing expression for the phase, we are able to write the frequency and angle distribution of the intensity as

$$\frac{dI(\omega)}{d\Omega} = \frac{e^2 \omega^2}{4\pi^2 c} \left| \int_{-\infty}^{\infty} dt' (\theta \epsilon_{\perp} - \omega_0 t' \epsilon_{\parallel}) e^{i \frac{\omega}{2\omega_0} \left[\omega_0 t' \left(\frac{1}{\gamma^2} + \theta^2 \right) + \frac{1}{3} (\omega_0 t')^3 \right]} \right|^2. \quad (122)$$

The integration has been extended to $\pm\infty$ even though the integrand is accurate only for $|\omega_0 t'| \sim 1/\gamma$ (or less). The extended range of integration is sensible only if there is no important contribution coming from other regimes of t' . That is the case because the term in the phase proportional to t'^3 produces rapid oscillations of the integrand at larger $|t'|$ which yield a very small net contribution to the integral.

Introduce x such that $\omega_0 t' \equiv x(1/\gamma^2 + \theta^2)^{1/2}$. The important range of t' corresponds to $|x| \sim 1$ and the frequency distribution of the intensity is given by

$$\frac{dI(\omega)}{d\Omega} = \frac{e^2 \omega^2}{4\pi^2 c} \left| \frac{1}{\omega_0} \sqrt{\frac{1}{\gamma^2} + \theta^2} \int_{-\infty}^{\infty} dx \left[\theta \epsilon_{\perp} - \sqrt{\frac{1}{\gamma^2} + \theta^2} x \epsilon_{\parallel} \right] e^{i \frac{\omega}{2\omega_0} \left(\frac{1}{\gamma^2} + \theta^2 \right)^{3/2} \left(x + \frac{1}{3} x^3 \right)} \right|^2. \quad (123)$$

Let

$$\eta = \frac{\omega}{3\omega_0} \left(\frac{1}{\gamma^2} + \theta^2 \right)^{3/2}. \quad (124)$$

Then

$$\begin{aligned} \frac{dI(\omega)}{d\Omega} &= \frac{e^2 \omega^2}{4\pi^2 c \omega_0^2} \left| \int_{-\infty}^{\infty} dx \left[\sqrt{\frac{1}{\gamma^2} + \theta^2} \theta \epsilon_{\perp} + \left(\frac{1}{\gamma^2} + \theta^2 \right) x \epsilon_{\parallel} \right] e^{i \frac{3}{2} \eta (x + x^3/3)} \right|^2 \\ &\equiv \frac{e^2 \omega^2}{4\pi^2 c \omega_0^2} \left| \sqrt{\frac{1}{\gamma^2} + \theta^2} \theta I_{\perp} \epsilon_{\perp} + \left(\frac{1}{\gamma^2} + \theta^2 \right) I_{\parallel} \epsilon_{\parallel} \right|^2. \end{aligned} \quad (125)$$

where

$$I_{\perp} = \int_{-\infty}^{\infty} dt e^{i(3\eta t + \eta t^3)/2} \quad (126)$$

and

$$I_{\parallel} = \int_{-\infty}^{\infty} dt t e^{i(3\eta t + \eta t^3)/2}. \quad (127)$$

These integrals may be expressed in terms of *Airy functions* which are modified Bessel functions of order 1/3 and 2/3. An integral representation⁷ of the function Ai is

$$\frac{\pi}{(3a)^{1/3}} Ai[x/(3a)^{1/3}] = \int_0^\infty dt \cos(xt + at^3) = \frac{1}{2} \int_{-\infty}^\infty dt e^{i(xt+at^3)}. \quad (128)$$

From this representation one can see that

$$I_\perp = \frac{2\pi}{(3\eta/2)^{1/3}} Ai[(3\eta/2)^{2/3}]. \quad (129)$$

As for I_\parallel , it is

$$\begin{aligned} I_\parallel &= \int_{-\infty}^\infty dt te^{i(3\eta t + \eta t^3)/2} = \int_{-\infty}^\infty dt te^{i(xt + \eta t^3/2)} \Big|_{x=3\eta/2} \\ &= \frac{1}{i} \frac{d}{dx} \left(\int_{-\infty}^\infty dt e^{i(xt + \eta t^3/2)} \right) \Big|_{x=3\eta/2} = \frac{1}{i} \frac{d}{dx} \left(\frac{2\pi}{(3\eta/2)^{1/3}} Ai[x/(3\eta/2)^{1/3}] \right) \Big|_{x=3\eta/2} \\ &= \frac{2\pi}{i(3\eta/2)^{2/3}} Ai'[(3\eta/2)^{2/3}]. \end{aligned} \quad (130)$$

The prime on the Airy function denotes differentiation with respect to the argument.

The connection between Airy functions and modified Bessel functions is⁸

$$Ai[(3\eta/2)^{2/3}] = \frac{1}{\pi} \left[\frac{(3\eta/2)^{2/3}}{3} \right]^{1/2} K_{1/3}(\eta). \quad (131)$$

Also⁹,

$$-Ai'[(3\eta/2)^{2/3}] = \frac{1}{\pi} \frac{(3\eta/2)^{2/3}}{\sqrt{3}} K_{2/3}(\eta). \quad (132)$$

Thus we may express the result for the frequency distribution of intensity in terms of modified Bessel functions as

$$\begin{aligned} \frac{dI(\omega)}{d\Omega} &= \frac{e^2}{4\pi^2 c} \left(\frac{\omega}{\omega_0} \right)^2 \left| \sqrt{\frac{1}{\gamma^2} + \theta^2} \theta \frac{2}{\sqrt{3}} K_{1/3}(\eta) \epsilon_\perp + \left(\frac{1}{\gamma^2} + \theta^2 \right) \frac{2}{i\sqrt{3}} K_{2/3}(\eta) \epsilon_\parallel \right|^2 \\ &= \frac{e^2}{3\pi^2 c} \left(\frac{\omega}{\omega_0} \right)^2 \left(\frac{1}{\gamma^2} + \theta^2 \right)^2 \left[K_{2/3}^2(\eta) + \frac{\theta^2}{1/\gamma^2 + \theta^2} K_{1/3}^2(\eta) \right] \\ &= \frac{3e^2 \gamma^2}{\pi^2 c} \left(\frac{\omega}{\omega_c} \right)^2 (1 + \gamma^2 \theta^2)^2 \left[K_{2/3}^2(\eta) + \frac{\gamma^2 \theta^2}{1 + \gamma^2 \theta^2} K_{1/3}^2(\eta) \right] \end{aligned} \quad (133)$$

⁷Abramowitz and Stegun, 10.4.32.

⁸Abramowitz and Stegun, 10.4.14.

⁹Abramowitz and Stegun, 10.4.16.

with

$$\omega_c = 3\gamma^3\omega_0 \quad \text{and} \quad \eta = \frac{\omega}{3\omega_0} \left(\frac{1}{\gamma^2} + \theta^2 \right)^{3/2} = \frac{\omega}{\omega_c} (1 + \theta^2\gamma^2)^{3/2}. \quad (134)$$

This is a fairly transparent, if not simple, result. The variable η is proportional to ω and is scaled by $3\gamma^3\omega_0 \equiv \omega_c$ which we believe, on the basis of arguments given earlier, to be the appropriate or characteristic scale of frequency in this radiating system. For $\eta \ll 1$,

$$K_\nu(\eta) \sim \frac{\Gamma(\nu)}{2} \left(\frac{2}{\eta} \right)^\nu \quad (135)$$

and for $\eta \gg 1$,

$$K_\nu(\eta) \sim \sqrt{\frac{\pi}{2\eta}} e^{-\eta}. \quad (136)$$

Thus, for $\eta \ll 1$ and at $\theta = 0$,

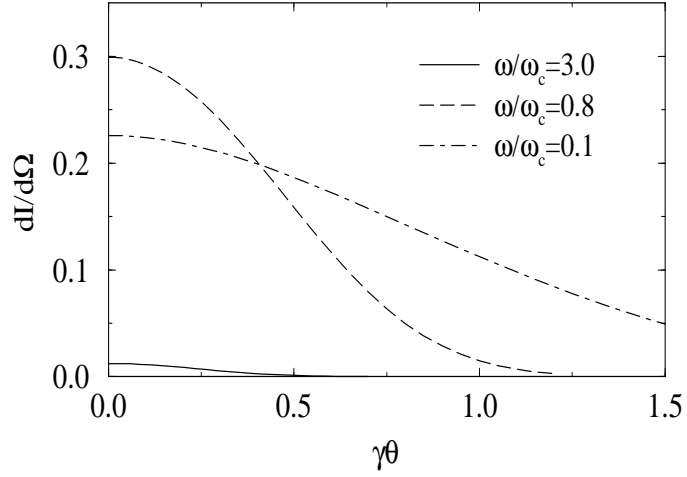
$$\frac{dI(\omega)}{d\Omega} = \frac{3e^2}{\pi^2c} \gamma^2 \left(\frac{\omega}{\omega_c} \right)^2 \left(\frac{\Gamma(2/3)}{2} \right)^2 \left(\frac{2}{\omega/\omega_c} \right)^{4/3} = \frac{e^2}{c} \left(\frac{\Gamma(2/3)}{\pi} \right)^2 \left(\frac{3}{4} \right)^{1/3} \left(\frac{\omega}{\omega_0} \right)^{2/3}. \quad (137)$$

If $\theta \neq 0$, we pick up a contribution proportional to $K_{1/3}^2$, leading to an additional term which is proportional to $\omega^{4/3}$. Note also that the term we do have is produced by waves with the electric field polarized in the plane of the particle's orbit.

In the large frequency regime, $\omega \gg \omega_c$ and $\theta = 0$, we find

$$\frac{dI(\omega)}{d\Omega} = \frac{e^2}{\pi^2c} 3\gamma^2 \left(\frac{\omega}{\omega_c} \right)^2 \frac{\pi}{2\omega/\omega_c} e^{-2\omega/\omega_c} = \frac{e^2}{2\pi c \gamma} \frac{\omega}{\omega_c} e^{-2\omega/3\gamma^3\omega_0}. \quad (138)$$

The accompanying figure shows the angular distribution of radiation at several values



of ω/ω_c .

This figure also very clearly makes the point that at a given frequency, except for the very low ones, the intensity is greatest at the smallest angles.

4.3.2 A Particle in Circular Motion

Let's do the same sort of calculation for truly circular motion. Then, from Eqs. (111) and (113), we find that the energy emitted per cycle per unit solid angle is

$$\frac{dW}{d\Omega} = \sum_{n=0}^{\infty} \frac{dW_n}{d\Omega} = \frac{4\pi}{\omega_0} \sum_{n=1}^{\infty} |\mathbf{a}_n|^2 \quad (139)$$

with

$$|\mathbf{a}_n| = \sqrt{\frac{e^2}{16\pi^3 c} n \omega_0^2} \left| \int_0^{2\pi/\omega_0} dt' (\mathbf{n} \times \boldsymbol{\beta}) e^{in\omega_0(t' - \mathbf{n} \cdot \mathbf{x}(t')/c)} \right|. \quad (140)$$

An individual term in the sum, $dW_n/d\Omega$, is the energy per cycle per unit solid angle radiated at frequency $\omega_n = n\omega_0$. It can be written as

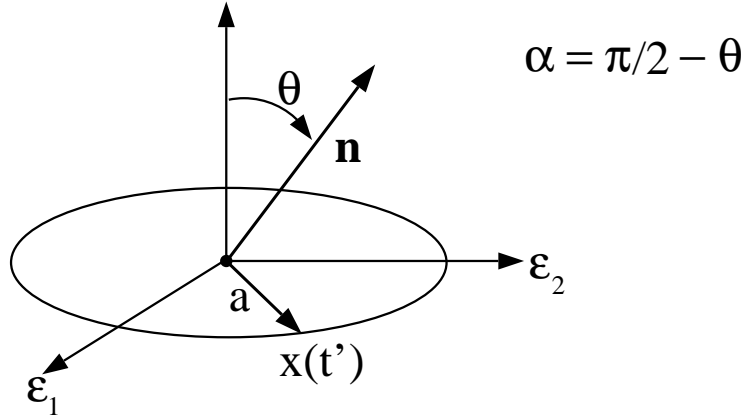
$$\frac{dW_n}{d\Omega} = \frac{4\pi}{\omega_0} \frac{e^2 n^2 \omega_0^4}{16\pi^3 c} \left| \int_0^{2\pi/\omega_0} dt' (\mathbf{n} \times \boldsymbol{\beta}) e^{in\omega_0(t' - \mathbf{n} \cdot \mathbf{x}(t')/c)} \right|^2. \quad (141)$$

If we divide by the period, $2\pi/\omega_0$, we find the time-averaged power at frequency $n\omega_0$ per unit solid angle,

$$\frac{d\mathcal{P}_n}{d\Omega} = \frac{e^2 n^2 \omega_0^4}{8\pi^3 c} \left| \int_0^{2\pi/\omega_0} dt' (\mathbf{n} \times \boldsymbol{\beta}) e^{in\omega_0(t' - \mathbf{n} \cdot \mathbf{x}(t')/c)} \right|^2. \quad (142)$$

Let the motion be in the x - y plane as before,

$$\mathbf{x}(t') = a[\boldsymbol{\epsilon}_1 \cos(\omega_0 t') + \boldsymbol{\epsilon}_2 \sin(\omega_0 t')] \quad (143)$$



so that

$$\boldsymbol{\beta}(t') = \frac{\omega_0 a}{c} [-\boldsymbol{\epsilon}_1 \sin(\omega_0 t') + \boldsymbol{\epsilon}_2 \cos(\omega_0 t')]. \quad (144)$$

Also, write \mathbf{n} in terms of the usual¹⁰ spherical coordinates,

$$\mathbf{n} = \boldsymbol{\epsilon}_3 \cos \theta + \boldsymbol{\epsilon}_1 \sin \theta \cos \phi + \boldsymbol{\epsilon}_2 \sin \theta \sin \phi. \quad (145)$$

Then

$$\mathbf{n} \cdot \mathbf{x}(t') = a[\sin \theta \cos \phi \cos(\omega_0 t') + \sin \theta \sin \phi \sin(\omega_0 t')] = a \sin \theta \cos(\omega_0 t' - \phi), \quad (146)$$

and

$$\begin{aligned} \mathbf{n} \times \boldsymbol{\beta}(t') &= \frac{a\omega_0}{c} [-\boldsymbol{\epsilon}_2 \cos \theta \sin(\omega_0 t') - \boldsymbol{\epsilon}_1 \cos \theta \cos(\omega_0 t') \\ &\quad + \boldsymbol{\epsilon}_3 \sin \theta (\cos \phi \cos \omega_0 t' + \sin \phi \sin \omega_0 t')] \\ &= \frac{a\omega_0}{c} [-\boldsymbol{\epsilon}_2 \cos \theta \sin(\omega_0 t') - \boldsymbol{\epsilon}_1 \cos \theta \cos(\omega_0 t') + \boldsymbol{\epsilon}_3 \sin \theta \cos(\omega_0 t' - \phi)] \end{aligned} \quad (147)$$

¹⁰But be aware that the angle θ in this example is the polar angle and not the latitude as in the previous example; the angle α introduced below is the latitude, i.e., the same as the θ of the previous example.

We must do the integral

$$\mathbf{I} = \int_0^{2\pi/\omega_0} dt' (\mathbf{n} \times \boldsymbol{\beta}) e^{in\omega_0(t' - \mathbf{n} \cdot \mathbf{x}(t')/c)} \equiv \frac{\beta}{\omega_0} [-\boldsymbol{\epsilon}_1 K \cos \theta - \boldsymbol{\epsilon}_2 J \cos \theta + \boldsymbol{\epsilon}_3 L' \sin \theta] \quad (148)$$

where $\omega_0 t' = y$, and

$$\begin{aligned} K &= \int_0^{2\pi} dy \cos(y + \phi) e^{i(ny - n\beta \sin \theta \cos y)} e^{in\phi} \\ J &= \int_0^{2\pi} dy \sin(y + \phi) e^{i(ny - n\beta \sin \theta \cos y)} e^{in\phi} \\ L' &= \int_0^{2\pi} dy \cos y e^{i(ny - n\beta \sin \theta \cos y)} e^{in\phi} \end{aligned} \quad (149)$$

with $\beta = a\omega_0/c$. Thus

$$\begin{aligned} \begin{Bmatrix} J \\ K \end{Bmatrix} &= \int_0^{2\pi} dy \begin{Bmatrix} \sin(y + \phi) \\ \cos(y + \phi) \end{Bmatrix} e^{in\phi} e^{iny} e^{-in\beta \sin \theta \cos y} \\ &= e^{in\phi} \int_0^{2\pi} dy \left[\sin \phi \begin{Bmatrix} \cos y \\ -\sin y \end{Bmatrix} + \cos \phi \begin{Bmatrix} \sin y \\ \cos y \end{Bmatrix} \right] e^{i(ny - n\beta \sin \theta \cos y)} \\ &\equiv e^{in\phi} \left[\sin \phi \begin{Bmatrix} L \\ M \end{Bmatrix} + \cos \phi \begin{Bmatrix} M \\ L \end{Bmatrix} \right] \end{aligned} \quad (150)$$

where

$$\begin{aligned} M &= \int_0^{2\pi} dy \sin y e^{iny - in\beta \sin \theta \cos y} \\ &= \frac{1}{in\beta \sin \theta} \int_0^{2\pi} dy e^{iny} \frac{d}{dy} (e^{-in\beta \sin \theta \cos y}) \\ &= -\frac{1}{\beta \sin \theta} \int_0^{2\pi} dy e^{iny} e^{-in\beta \sin \theta \cos y} = -\frac{2\pi}{\beta \sin \theta} \frac{J_n(n\beta \sin \theta)}{i^n}, \end{aligned} \quad (151)$$

and

$$\begin{aligned} L &= \int_0^{2\pi} dy \cos y e^{iny} e^{-in\beta \sin \theta \cos y} \\ &= \frac{1}{-i d(n\beta \sin \theta)} \int_0^{2\pi} dy e^{i(ny - n\beta \sin \theta \cos y)} \\ &= -\frac{2\pi}{i} \frac{1}{i^n} \frac{dJ_n(n\beta \sin \theta)}{d(n\beta \sin \theta)} \end{aligned} \quad (152)$$

Hence

$$\begin{aligned}
|\mathbf{I}|^2 &= \frac{\beta^2}{\omega_0^2} \left[\cos^2 \theta (|K|^2 + |J|^2) + \sin^2 \theta |L|^2 \right] \\
&= \frac{4\pi^2 \beta^2}{\omega_0^2} \left[\cos^2 \theta \left(\frac{J_n(n\beta \sin \theta)}{\beta \sin \theta} \right)^2 + (\cos^2 \theta + \sin^2 \theta) \left(\frac{dJ_n(n\beta \sin \theta)}{d(n\beta \sin \theta)} \right)^2 \right] \quad (153)
\end{aligned}$$

Also,

$$\frac{d\mathcal{P}_n}{d\Omega} = \frac{e^2 n^2 \beta^2 \omega_0^2}{2\pi c} \left[\left(\frac{dJ_n(n\beta \sin \theta)}{d(n\beta \sin \theta)} \right)^2 + \frac{\cot^2 \theta}{\beta^2} (J_n(n\beta \sin \theta))^2 \right]. \quad (154)$$

For better comparison of this result with things we already know, let us look at some limiting cases. First, in the nonrelativistic limit, $\beta \ll 1$, we expect that only the $n = 1$ term will contribute appreciably and, in addition, we can use the small argument approximation to the Bessel function,

$$J_1(x) \approx x/2 \quad \text{and} \quad J_1'(x) \approx 1/2 \quad (155)$$

so that

$$\frac{d\mathcal{P}_1}{d\Omega} = \frac{e^2 \beta^2 \omega_0^2}{2\pi c} \left(\frac{1}{4} + \frac{\cot^2 \theta \sin^2 \theta}{4} \right) = \frac{e^2 \omega_0^4 a^2}{8\pi c^3} (1 + \cos^2 \theta) \quad (156)$$

which may be compared with the result of a completely nonrelativistic calculation as in, e.g., Jackson, Problem 14.2(b). In the highly relativistic limit, on the other hand, we have $\beta \approx 1$ and we also know that most of the radiation is close to the equatorial plane or $\theta \approx \pi/2$. Let us introduce $\alpha \equiv \pi/2 - \theta \ll 1$. Then

$$\frac{d\mathcal{P}_n}{d\Omega} = \frac{e^2 n^2 \beta^2 \omega_0^2}{2\pi c} \left[\left(\frac{dJ_n(x)}{dx} \right)^2 + \frac{\tan^2 \alpha}{\beta^2} (J_n(x))^2 \right] \Bigg|_{x=n\beta \cos \alpha}. \quad (157)$$

The argument of the Bessel function J_n is comparable to n but is always less than n . For this particular range of argument, it is the case¹¹ that

$$J_n(x) = \frac{1}{\pi} \sqrt{\frac{2(n-x)}{3x}} K_{1/3} \left(\frac{2\sqrt{2}(n-x)^{3/2}}{3\sqrt{x}} \right), \quad (158)$$

¹¹See, e.g., Watson, *Bessel Functions*, p. 249.

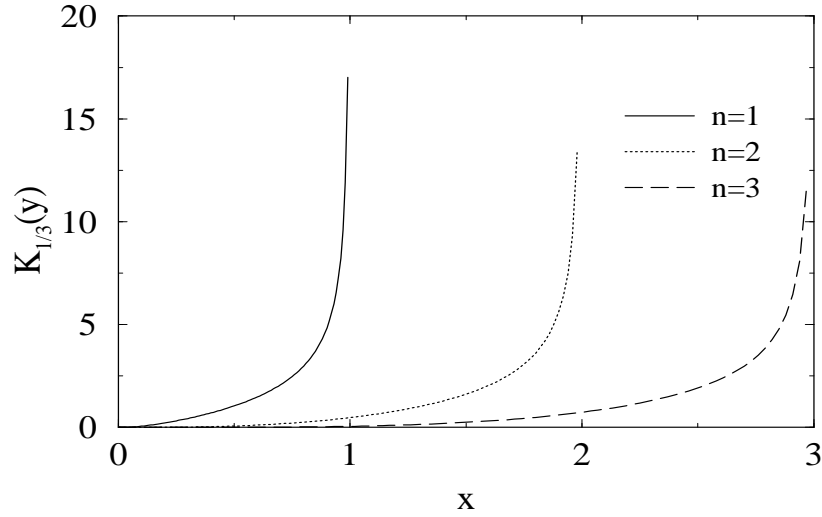
or

$$J_n(x) = \frac{1}{\pi} \frac{y^{1/3}}{3^{1/6} x^{1/3}} K_{1/3}(y) \quad \text{where} \quad y = \frac{2\sqrt{2}(n-x)^{3/2}}{3\sqrt{x}}. \quad (159)$$

Also,

$$\frac{dJ_n(x)}{dx} = \frac{1}{\pi 3^{1/6}} \frac{d}{dx} \left(\frac{y^{1/3} K_{1/3}(y)}{x^{1/3}} \right). \quad (160)$$

As shown in the figure below, for the interesting values of x , $K_{1/3}(y)$ varies much more than x .



Thus, the preceding derivative may be approximated by

$$\begin{aligned} \frac{dJ_n(x)}{dx} &\approx \frac{1}{\pi 3^{1/6} x^{1/3}} \frac{d[y^{1/3} K_{1/3}(y)]}{dy} \frac{dy}{dx} \\ &= -\frac{y^{1/3} K_{2/3}(y)}{\pi 3^{1/6} x^{1/3}} \frac{dy}{dx} = \frac{y^{1/3} K_{2/3}(y)}{3^{1/6} x^{1/3}} \frac{\sqrt{2}(n-x)^{1/2}}{3x^{3/2}} (2x+n). \end{aligned} \quad (161)$$

Hence

$$\frac{d\mathcal{P}_n}{d\Omega} = \frac{e^2 \beta^2 n^2 \omega_0^2 y^{2/3}}{2\pi c \pi^2 3^{1/3} x^{2/3}} \left[\frac{n^2 \sin^2 \alpha}{x^2} K_{1/3}^2(y) + \frac{2(n-x)(2x+n)^2}{9x^3} K_{2/3}^2(y) \right]. \quad (162)$$

Now let's see what can be said about y and x . Expanding in powers of α and $1/\gamma$, we have

$$y \approx \frac{n}{3} \left(\frac{1}{\gamma^2} + \alpha^2 \right)^{3/2} \quad \text{and} \quad x \approx n \left(1 - \frac{1}{2\gamma^2} - \frac{\alpha^2}{2} \right), \quad (163)$$

so

$$\frac{d\mathcal{P}_n}{d\Omega} = \frac{e^2\omega_0^2 n^2}{6\pi^3 c} \left(\frac{1}{\gamma^2} + \alpha^2 \right)^2 \left[K_{2/3}^2(y) + \frac{\alpha^2 \gamma^2}{1 + \alpha^2 \gamma^2} K_{1/3}^2(y) \right]. \quad (164)$$

This term describes radiation at the particular frequency $\omega = n\omega_0$, so we can also write it as

$$\frac{d\mathcal{P}_n}{d\Omega} = \frac{\omega_0^2}{2\pi} \frac{e^2(\omega^2/\omega_0^2)}{3\pi^2 c \gamma^6} (1 + \alpha^2 \gamma^2)^2 \gamma^2 \left[K_{2/3}^2(y) + \frac{\alpha^2 \gamma^2}{1 + \alpha^2 \gamma^2} K_{1/3}^2(y) \right] \quad (165)$$

with

$$y = \frac{\omega}{3\omega_0 \gamma^3} (1 + \alpha^2 \gamma^2)^{3/2} \equiv \frac{\omega}{\omega_c} (1 + \alpha^2 \gamma^2)^{3/2}. \quad (166)$$

By grouping terms appropriately in Eq. (165) we can write

$$\frac{d\mathcal{P}_n}{d\Omega} = \frac{\omega_0^2}{2\pi} \frac{3e^2}{\pi^2 c} \left(\frac{\omega}{\omega_c} \right)^2 (1 + \alpha^2 \gamma^2)^2 \gamma^2 \left[K_{2/3}^2(y) + \frac{\alpha^2 \gamma^2}{1 + \alpha^2 \gamma^2} K_{1/3}^2(y) \right]. \quad (167)$$

This is the time-averaged power received at latitude α , at frequency $\omega = n\omega_0$, per unit solid angle. The energy/cycle received at this frequency is obtained if we multiply by $2\pi/\omega_0$, and, finally, the energy per unit frequency is found if we multiply by a factor of the inverse spacing of the harmonics, or $1/\omega_0$. Hence we conclude that for a single pulse, meaning one cycle of the particle,

$$\frac{dI(\omega)}{d\Omega} = \frac{2\pi}{\omega_0^2} \frac{d\mathcal{P}_n}{d\Omega} = \frac{3e^2}{\pi^2 c} \left(\frac{\omega}{\omega_0} \right)^2 (1 + \alpha^2 \gamma^2)^2 \gamma^2 \left[K_{2/3}^2(y) + \frac{\alpha^2 \gamma^2}{1 + \alpha^2 \gamma^2} K_{1/3}^2(y) \right] \quad (168)$$

which is precisely what we found when we calculated the radiated intensity from a particle in instantaneous circular motion.

5 Thomson Scattering; Blue Sky

It's time for a change of pace to something with less nineteenth-century analysis. *Thomson scattering* provides just such a respite. It is the scattering of radiation by a free charge. The mechanism involved, classically, is the coupling of the charge to the incident electric field \mathbf{E}_0 ; this produces acceleration of the charge and consequent

radiation by the accelerated charge which becomes the “scattered” radiation. We saw in Chapter 9 how one can describe scattering of electromagnetic radiation in this way. We didn’t actually do Thomson scattering at that time, so we’ll do it now.

Thomson scattering is to be distinguished from *Compton scattering* in which the same phenomenon is treated using quantum theory. One must use quantum theory, when λ , the wavelength of the radiation, is comparable to the *Compton wavelength* of the scatterer, $\lambda = 2\pi c/\omega \sim \hbar/mc \equiv \lambda_c$; λ_c is the Compton wavelength. This condition may also be written as $\hbar\omega \sim 2\pi mc^2$ which says that if the photon energy $\hbar\omega$ (\sim eV) is comparable to the particle’s rest energy mc^2 (\sim MeV), then the calculation must be done using quantum theory. For, e.g., visible light the photon energy is much smaller than an electron’s rest energy so the Thomson scattering, or classical, calculation is quite adequate.

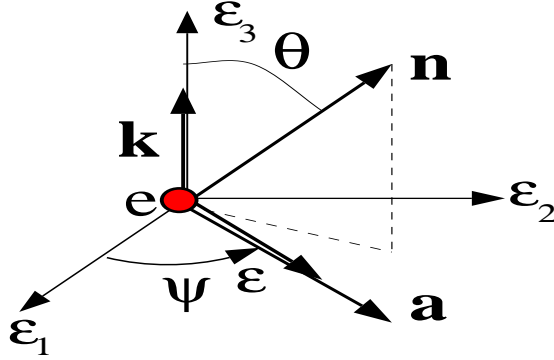
Assuming a nonrelativistic particle, we can make use of the Larmor formula for the scattered radiation,

$$\frac{d\mathcal{P}}{d\Omega} = \frac{e^2}{4\pi c^3} a^2 \sin^2 \Theta \quad \text{where} \quad \Theta = \angle(\mathbf{n}, \mathbf{a}) \quad (169)$$

for the instantaneous radiated power. To find the acceleration \mathbf{a} of the particle, we must solve the equation of motion of the scatterer. The force on it is provided by the incident electric field,

$$\mathbf{E}_0 = E_0 \boldsymbol{\epsilon} e^{i(\mathbf{k}\cdot\mathbf{x} - \omega t)} \quad (170)$$

where for convenience we shall let $\mathbf{k} = k\boldsymbol{\epsilon}_3$ and assume linear polarization of the incident plane wave $\boldsymbol{\epsilon} = \cos \psi \boldsymbol{\epsilon}_1 + \sin \psi \boldsymbol{\epsilon}_2$.



Similarly, we shall suppose for simplicity that the charged particle has no velocity component along the direction of \mathbf{k} and shall ignore the magnetic force which is reasonable so long as the particle is nonrelativistic. Letting the particle be located in the $z = 0$ plane, we find that it experiences an incident field

$$\mathbf{E}_0 = \epsilon E_0 e^{-i\omega t}; \quad (171)$$

given that it has a mass m and charge e , the force acting on it is $\mathbf{F} = e\mathbf{E}_0 = m\mathbf{a}$, so

$$\mathbf{a} = \epsilon \frac{eE_0}{m} e^{-i\omega t}, \quad (172)$$

and

$$\mathbf{a} \cdot \mathbf{n} = \frac{eE_0}{m} \sin \theta (\cos \psi \cos \phi + \sin \psi \sin \phi) e^{-i\omega t}. \quad (173)$$

where θ and ϕ specify, in spherical coordinates, the direction to the observation point.

The time average of the power emitted in some particular direction will be proportional to

$$\langle a^2 \sin^2 \Theta \rangle = \langle a^2 (1 - \cos^2 \Theta) \rangle = \langle a^2 - (\mathbf{a} \cdot \mathbf{n})^2 \rangle = \frac{e^2 E_0^2}{2m^2} [1 - \sin^2 \theta \cos^2(\psi - \phi)] \quad (174)$$

where the brackets $\langle \dots \rangle$ denote a time average. If the incident radiation is not polarized, then we must average over ψ with the result that

$$\langle a^2 \sin^2 \Theta \rangle = \frac{e^2 E_0^2}{2m^2} \left(1 - \frac{1}{2} \sin^2 \theta \right) = \frac{e^2 E_0^2}{4m^2} (1 + \cos^2 \theta) \quad (175)$$

and

$$\frac{d\mathcal{P}}{d\Omega} = \frac{e^4}{16\pi c^3 m^2} |E_0|^2 (1 + \cos^2 \theta) \quad (176)$$

is the time-averaged power distribution when the incident wave is unpolarized.

We define the scattering cross-section in the usual way, i.e., the time-averaged power per unit solid angle divided by the time-averaged incident power per unit area,

$$\frac{d\sigma}{d\Omega} = \frac{d\mathcal{P}/d\Omega}{(c/8\pi)|E_0|^2} = \frac{1}{2} \left(\frac{e^2}{mc^2} \right)^2 (1 + \cos^2 \theta). \quad (177)$$

This is J. J. Thomson's formula for the scattering of light by a charged particle. The total cross-section is

$$\sigma = \int d\Omega \left(\frac{d\sigma}{d\Omega} \right) = \frac{1}{2} \left(\frac{e^2}{mc^2} \right)^2 4\pi(1 + 1/3) = \frac{8\pi}{3} \left(\frac{e^2}{mc^2} \right)^2 \equiv \frac{8\pi}{3} r_c^2 \quad (178)$$

where r_c is the classical radius of the particle. For an electron it is $\sim 3 \times 10^{-13} \text{ cm}$ and σ is about $0.7 \times 10^{-24} \text{ cm}^2$.

We may also calculate the scattering of radiation by a bound charge using the model of a damped harmonic oscillator. Let a charge e with mass m be bound at the origin of coordinates with a natural frequency of oscillation ω_0 and damping constant Γ . Then, given an applied electric field

$$\mathbf{E}_0(\mathbf{x}, t) = \epsilon E_0 e^{i(\mathbf{k}\cdot\mathbf{x} - \omega t)} \quad (179)$$

with $\mathbf{k} = k\epsilon_3$, we know from earlier calculations that the charge will respond with a displacement

$$\mathbf{x}(t) = \mathbf{x}_0 e^{-i\omega t} \quad \text{where} \quad \mathbf{x}_0 = \frac{eE_0/m}{\omega_0^2 - \omega^2 - i\omega\Gamma} \epsilon \quad (180)$$

if we approximate $\mathbf{E}_0(\mathbf{x}, t)$ by $\mathbf{E}_0(0, t)$ as is reasonable when the particle's displacement from the origin is small compared to the wavelength of the incident radiation. That is certainly true for visible light and an atomic electron.

From $\mathbf{x}(t)$ it is a simple matter to compute the acceleration,

$$\mathbf{a} = -\epsilon \frac{eE_0\omega^2/m}{\omega_0^2 - \omega^2 - i\omega\Gamma} e^{-i\omega t}, \quad (181)$$

and then to find the radiated, or scattered, time- and incident-polarization-averaged power per unit solid angle,

$$\frac{d\mathcal{P}}{d\Omega} = \frac{e^2}{16\pi c^3} \left(\frac{eE_0}{m} \right)^2 \left(\frac{1 + \cos^2 \theta}{(1 - \omega_0^2/\omega^2)^2 + \Gamma^2/\omega^2} \right); \quad (182)$$

the scattering cross-section follows:

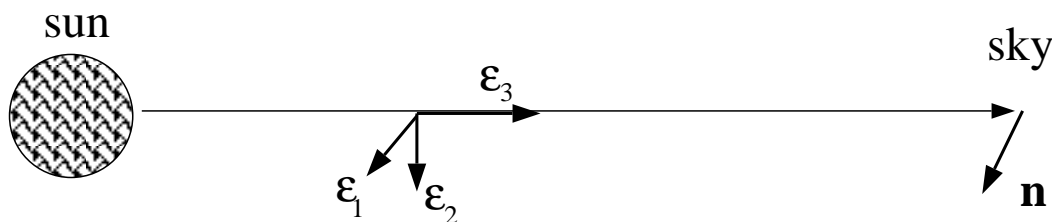
$$\frac{d\sigma}{d\Omega} = \frac{1}{2} r_c^2 \left(\frac{1 + \cos^2 \theta}{(1 - \omega_0^2/\omega^2)^2 + \Gamma^2/\omega^2} \right). \quad (183)$$

For sufficiently large ω , $\omega \gg \omega_0, \Gamma$, the cross-section reduces to the Thomson result as it should since for very large ω the particle will respond to the field as though it were a free particle. Also, for $\omega \ll \omega_0$ and ω_0^2/Γ , we find the Rayleigh scattering result,

$$\frac{d\sigma}{d\Omega} = \frac{1}{2} r_c^2 \left(\frac{\omega}{\omega_0} \right)^4 (1 + \cos^2 \theta) \quad (184)$$

with the characteristic ω^4 behavior indicating dipole scattering.

Why is the sky blue? Why is the light polarized when we look at the sky perpendicular to the line of sight from us to the sun?



P

6 Cherenkov Radiation Revisited

While studying the energy loss of a charged particle traversing a material, we derived an expression for the rate of energy loss through Cherenkov radiation. Specifically,

for a charge q moving at speed β through a medium with a real dielectric function $\epsilon(\omega)$, we found that the energy loss per unit path length is

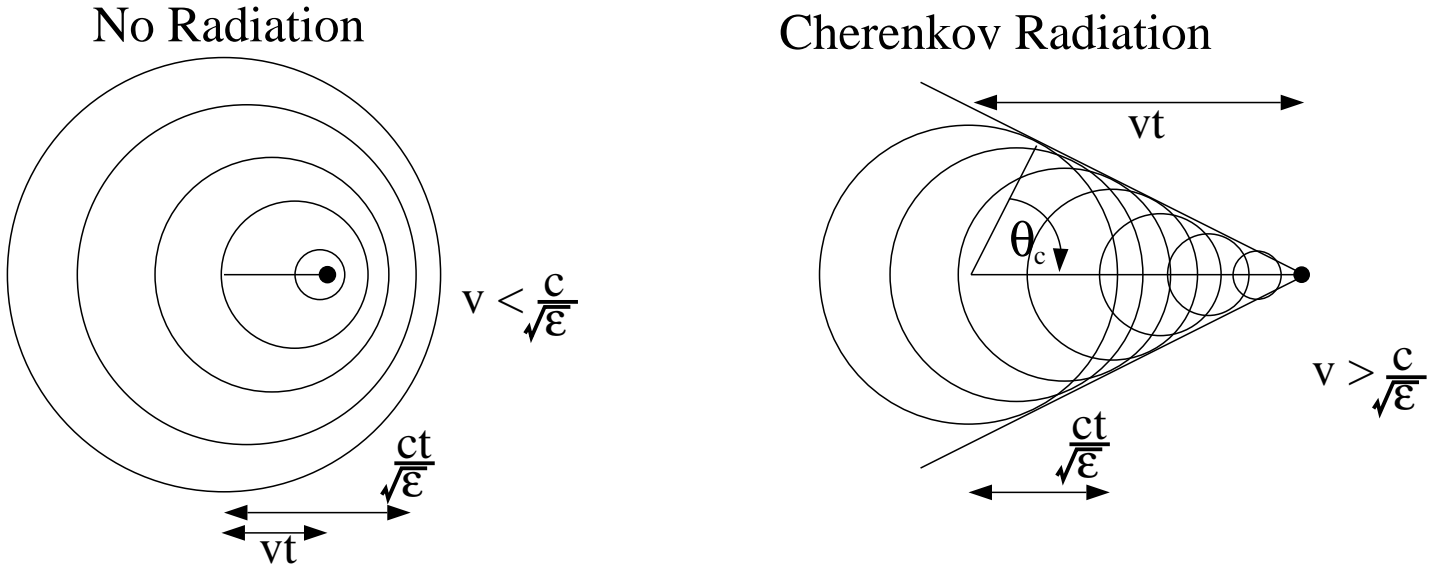
$$\frac{dE}{dx} = \frac{q^2}{c^2} \int_{\epsilon\beta^2 > 1} d\omega \omega \left(1 - \frac{1}{\epsilon(\omega)\beta^2}\right). \quad (185)$$

The form of this expression suggests that the radiation per unit length of path per unit frequency is given by the integrand,

$$\frac{dI(\omega)}{dx} = \omega \frac{q^2}{c^2} \left(1 - \frac{1}{\epsilon(\omega)\beta^2}\right). \quad (186)$$

What we did not do is look at the form of the potentials and fields in real space and time. That is interesting and revealing, so we are going to do it now using retarded potentials.

Before starting, let's be sure that we understand the physical mechanism giving rise to the radiation. It is not the incident particle, which may be reasonably described as having constant velocity, that is doing the radiating. Rather, the incident particle produces fields which act on the particles in the medium, causing them to be accelerated in various ways. They then produce radiation fields which, when the incident particle moves more rapidly than the speed of light in the medium, but not when it moves more slowly, add in a coherent fashion to give Cherenkov radiation.



It is not easy to see why the fields must cancel when the particle is moving slower than the speed of light, but it *is* easy to see why they must not when it is moving faster than light in the medium. Consider the right hand side of the figure above. Before the wake of the radiation hits a particle in the medium, it does not feel the incident particle. Once the wake hits a particular particle in the medium, not only it, but all of its neighbors accelerate in a direction perpendicular to the wake (depending upon the relative charge). Clearly these particles will radiate coherently.

If we are going to produce a calculation of Cherenkov radiation, then, we have to find the total field produced by all of the particles and not just the incident particle. We in fact did that in Chapter 13 in the space of \mathbf{k} and ω . Here we want to determine appropriate Liénard-Wieckert potentials for this field so as to find it in real space and time. That turns out not to be very hard if we ignore the frequency-dependence of the dielectric function.

Consider the Fourier-transformed potentials $\mathbf{A}(\mathbf{k}, \omega)$ and $\Phi(\mathbf{k}, \omega)$ produced by the incident charge in the medium where it is taken to have constant velocity. As we have seen, these obey the equations

$$\begin{aligned} \left(k^2 - \frac{\omega^2}{c^2}\epsilon\right) \Phi(\mathbf{k}, \omega) &= 4\pi \frac{\rho(\mathbf{k}, \omega)}{\epsilon} \\ \left(k^2 - \frac{\omega^2}{c^2}\epsilon\right) \mathbf{A}(\mathbf{k}, \omega) &= \frac{4\pi}{c} \mathbf{J}(\mathbf{k}, \omega) \end{aligned} \quad (187)$$

where ρ and \mathbf{J} are the macroscopic sources, i.e., the charge and current density of the incident particle. Given that ϵ is independent of ω , then these are the same as the equations obeyed by the potentials of a fictitious system consisting of a point particle with charge $q/\sqrt{\epsilon}$ moving at constant velocity \mathbf{v} in a ‘vacuum’ where the speed of light is $c' = c/\sqrt{\epsilon}$. Let’s rewrite them in such a way as to see this correspondence more clearly:

$$\begin{aligned} \left(k^2 - \frac{\omega^2}{c'^2}\right) [\sqrt{\epsilon}\Phi(\mathbf{k}, \omega)] &= 4\pi \frac{\rho(\mathbf{k}, \omega)}{\sqrt{\epsilon}} \\ \left(k^2 - \frac{\omega^2}{c'^2}\right) \mathbf{A}(\mathbf{k}, \omega) &= \frac{4\pi}{c'} \frac{\mathbf{J}(\mathbf{k}, \omega)}{\sqrt{\epsilon}}. \end{aligned} \quad (188)$$

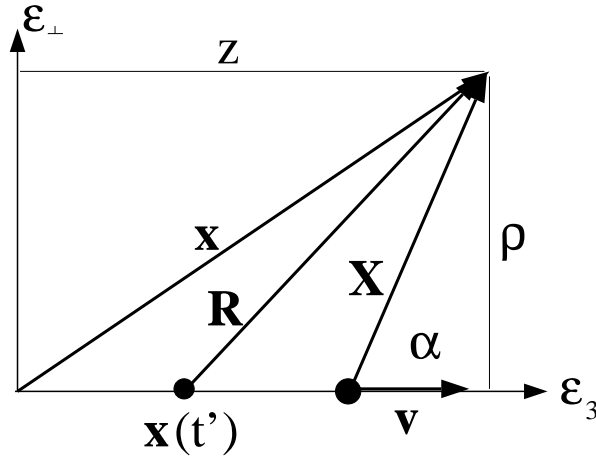
Now let's think about these potentials in real space and time. Because they are the same as a particle with a renormalized charge $q/\sqrt{\epsilon}$ moving in a vacuum with a renormalized speed of light is c' , we can write them as Liénard-Wieckert potentials using the renormalized charge and velocity of signal propagation,

$$\sqrt{\epsilon}\Phi(\mathbf{x}, t) = \frac{q}{\sqrt{\epsilon}} \left[\frac{1}{\kappa R} \right]_{ret} \quad (189)$$

and similarly for $\mathbf{A}(\mathbf{x}, t)$. In this case the retardation means that κR should be evaluated at the time $t' = t - R(t')/c'$. Further, if $\mathbf{x}(t') = \mathbf{v}t'$, $\mathbf{v} = v\boldsymbol{\epsilon}_3$, and $\mathbf{x} = z\boldsymbol{\epsilon}_3 + \rho\boldsymbol{\epsilon}_\perp$, then

$$R(t') = (z - vt')\boldsymbol{\epsilon}_3 + \rho\boldsymbol{\epsilon}_\perp. \quad (190)$$

Also, for this system, $\kappa = 1 - \mathbf{n} \cdot \mathbf{v}/c'$. Notice that κ can be negative; the absolute value of κ should be employed in evaluating the potential because the potential really involves $|\kappa R|$ as one may see by going back to its derivation, especially Eq. (4).



It is useful to introduce a vector $\mathbf{X} \equiv \mathbf{x} - \mathbf{v}t$ which is the relative displacement of the observation point and the particle at time t . Then

$$\mathbf{R} = \mathbf{x} - \mathbf{v}t' = \mathbf{x} - \mathbf{v}t + \mathbf{v}(t - t') = \mathbf{X} + \mathbf{v}(t - t'), \quad (191)$$

and

$$t - t' = R/c' = |\mathbf{X} + \mathbf{v}(t - t')|/c'. \quad (192)$$

Square this relation to find

$$(t - t')^2 = \frac{X^2}{c'^2} + \frac{v^2}{c'^2}(t - t')^2 + \left(\frac{2\mathbf{X} \cdot \mathbf{v}}{c'^2} \right) (t - t'). \quad (193)$$

This is a quadratic equation that can be solved for $t - t'$; there are two solutions which are

$$t - t' = \frac{-\mathbf{X} \cdot \mathbf{v} \pm \sqrt{(\mathbf{X} \cdot \mathbf{v})^2 - (v^2 - c'^2)X^2}}{v^2 - c'^2}. \quad (194)$$

Acceptable solutions must be real and positive. Given that $v^2 > c'^2$, which we know to be the regime where there is Cherenkov radiation, we find that there are either no such solutions or there are two of them. The conditions under which there are two are

$$\mathbf{X} \cdot \mathbf{v} < 0 \quad \text{and} \quad (\mathbf{X} \cdot \mathbf{v})^2 > (v^2 - c'^2)X^2. \quad (195)$$

Let the angle between \mathbf{X} and \mathbf{v} be α . Then we require, first, that α be larger than $\pi/2$ and, second, that $X^2 v^2 \cos^2 \alpha > (v^2 - c'^2)X^2$ or $\cos^2 \alpha > 1 - c'^2/v^2$. Hence there is a cutoff angle α_0 given by

$$\alpha_0 = \arccos(-\sqrt{1 - c'^2/v^2}) \quad (196)$$

such that for $\alpha < \alpha_0$ there is no potential. There can thus be potentials and fields at time t only within a cone whose apex is the current position of the particle and which has an apex angle of $\pi - \alpha_0$. Within this cone the potential is the sum of two terms, $\Phi = \Phi_1 + \Phi_2$, corresponding to the two allowed values of $t - t'$. Making use of Eq. (191), and the fact that $\mathbf{R} \parallel \mathbf{n}$, we see that we can write, for either case,

$$[\kappa R]_{ret} = |(1 - \mathbf{n} \cdot \mathbf{v}/c) \cdot \mathbf{R}| = R - \mathbf{v} \cdot \mathbf{R}/c = |\mathbf{X} + \mathbf{v}(t - t')| - \mathbf{X} \cdot \mathbf{v}/c' - v^2(t - t')/c'. \quad (197)$$

Using Eqs. (189) and (197) with Eq. (194) for $t - t'$, one finds that $\Phi_1 \equiv \Phi_2$ and that

$$\Phi(\mathbf{x}, t) = \left(\frac{2q}{\epsilon} \right) \frac{1}{X \sqrt{1 - (v^2/c'^2) \sin^2 \alpha}}. \quad (198)$$

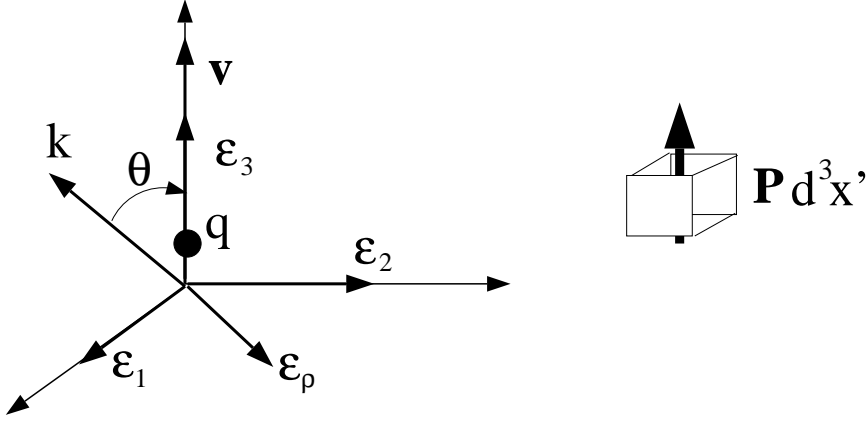
A similar expression can be found for $\mathbf{A}(\mathbf{x}, t)$ and with a bit more work one can compute the radiated power, recovering the same equations as found in chapter 13

for the particular case that $\epsilon(\omega)$ is a constant. Such a dielectric function never exists, of course, and so our conclusions are flawed in some respects. One of them has to do with the form of Φ close to $\alpha = \alpha_0$; Eq. (198) indicates that it in fact diverges here. That would indeed happen if signals with all frequencies traveled at speed c' so that such a singular non-dispersing wave front could be built by superposing waves with many different wavelengths including ones approaching zero. In reality, there is no such singularity although the amplitude of the wave does have a strong maximum at the leading edge.

7 Cherenkov Radiation; Transition Radiation

This time we will do a calculation using perturbation theory much the way we did scattering *via* perturbation theory. We will learn a little bit more about the character of Cherenkov radiation and will also derive a new (to us) phenomenon. The basic requirement for validity of the calculation is to have $|\epsilon(\omega) - 1| \ll 1$ so that we can get away with calculating the true macroscopic fields as a correction to the fields produced by the incident particle in vacuum. We have shown that, when expressed as a function of \mathbf{x} and ω , the electric field of a particle with charge q moving at constant velocity $\mathbf{v} = v\mathbf{e}_3$ on a trajectory $\mathbf{x}(t) = \mathbf{v}t$ is, in vacuum,

$$\mathbf{E}_i(\mathbf{x}', \omega) = \sqrt{\frac{2}{\pi}} \frac{q\omega}{\gamma v^2} e^{i\omega z'/v} [K_1(\omega\rho'/\gamma v)\mathbf{e}_{\rho'} - (i/\gamma)K_0(\omega\rho'/\gamma v)\mathbf{e}_3]. \quad (199)$$



This field produces a polarization in the medium which is

$$\mathbf{P}(\mathbf{x}'\omega) = \frac{\epsilon(\omega) - 1}{4\pi} \mathbf{E}_i(\mathbf{x}', \omega) \quad (200)$$

and a dipole moment $\mathbf{P}(\mathbf{x}', \omega)d^3x'$ in a volume element d^3x' . From chapter 9 (Eq. (43)), we know that such a harmonic dipole moment gives rise to a radiation field

$$d\mathbf{E}_{rad}(\mathbf{x}, \omega) = \frac{e^{ikR}}{R} k^2 [\mathbf{n} \times \{\mathbf{P}(\mathbf{x}', \omega)d^3x'\}] \times \mathbf{n}. \quad (201)$$

In the radiation zone we can expand $R = |\mathbf{x} - \mathbf{x}'|$ as $R = r - \mathbf{n} \cdot \mathbf{x}'$, leading to

$$\begin{aligned} \mathbf{E}_{rad}(\mathbf{x}, \omega) &= \frac{e^{ikr}}{r} \left(\frac{\epsilon(\omega) - 1}{4\pi} \right) k^2 \int_V d^3x' e^{-ik\mathbf{n}\cdot\mathbf{x}'} \{[\mathbf{n} \times \mathbf{E}_i(\mathbf{x}', \omega)] \times \mathbf{n}\} \quad (202) \\ &= \frac{e^{ikr}}{r} \left(\frac{\epsilon(\omega) - 1}{4\pi} \right) k^2 \int d^2x'_\perp \{[\mathbf{n} \times \mathbf{E}_i(\mathbf{x}'_\perp, 0, \omega)] \times \mathbf{n}\} e^{-ikx' \sin \theta} \int dz' e^{i(\omega z'/v - kz' \cos \theta)} \quad (203) \end{aligned}$$

where we have specified $\mathbf{k} = k(\epsilon_3 \cos \theta + \epsilon_1 \sin \theta)$ without loss of generality since the fields are invariant under rotation around the direction of \mathbf{v} . Also, $k = \omega \sqrt{\epsilon(\omega)}/c = \omega/v_p$ where v_p is the phase velocity of a wave with frequency ω . If the medium is infinite in the z direction, we can complete the integration over z' with ease and find

$$\mathbf{E}_{rad}(\mathbf{x}, \omega) = \frac{e^{ikr}}{r} \left(\frac{\epsilon(\omega) - 1}{4\pi} \right) k^2 \delta(\omega/v - k \cos \theta) \int d^2x'_\perp \{[\mathbf{n} \times \mathbf{E}_i(\mathbf{x}'_\perp, \omega)] \times \mathbf{n}\} e^{ikx' \sin \theta}. \quad (204)$$

Notice that k and ω are related by $k = \omega\sqrt{\epsilon(\omega)}/c = \omega/v_p$ so that we find no radiation field unless $\cos\theta = v_p/v$ which can only happen if $v > v_p$. We seem to be on the right track. The implication is that there is radiation coming out of this system in a direction \mathbf{n} which makes an angle α_0 with the z axis where $\alpha_0 = \arccos(v_p/v)$. What we have shown in chapter 13 is that this is the direction of the outgoing radiation, which is Cherenkov radiation, is perpendicular to the “bow wave” shown in the figure earlier.

We can get something new out of this calculation also. Suppose that $v < v_p$ and that the medium is finite in extent in the z direction. If it begins at $z = 0$ and ends at some arbitrary, large, value of z , then the integral over z' is not from $-\infty$ to $+\infty$ but rather starts from zero and will be non-zero even though $v < v_p$.



The corresponding radiation, known as *transition radiation*, comes about because of the presence of the boundary, in this case between vacuum and dielectric. The integral we must do can be evaluated with the use of a convergence factor (which emulates the damping present in most materials):

$$\begin{aligned} \int_0^\infty dz' e^{iz'(\omega/v - k \cos\theta)} &= \lim_{\eta \rightarrow 0} \int_0^\infty dz' e^{[-\eta + i(\omega/v - k \cos\theta)]z'} \\ &= \lim_{\eta \rightarrow 0} \left(\frac{-1}{-\eta + i(\omega/v - k \cos\theta)} \right) = \frac{i}{\omega/v - k \cos\theta}, \end{aligned} \quad (205)$$

and so

$$\mathbf{E}_{rad}(\mathbf{x}, \omega) = \frac{e^{ikr}}{r} \frac{\epsilon(\omega) - 1}{4\pi} k^2 \left(\frac{i}{\omega/v - k \cos \theta} \right) \int d^2 x'_\perp \{ [\mathbf{n} \times \mathbf{E}_i(\mathbf{x}'_\perp, \omega)] \times \mathbf{n} \} e^{-ikx' \sin \theta}. \quad (206)$$

More generally, a slab of material of finite thickness d produces an integral over z' which is

$$\int_0^d dz' e^{i(\omega/v - k \cos \theta)z'} = \frac{i(1 - e^{i(\omega/v - k \cos \theta)d})}{\omega/v - k \cos \theta}. \quad (207)$$

Whether we treat a finite slab or no, we have to do the integral over the other two coordinates. The integrand involves $(\mathbf{n} \times \mathbf{E}_i) \times \mathbf{n}$. For a given $\mathbf{n} = \mathbf{k}/k$, \mathbf{E}_i , and \mathbf{x}' , and noting that $\boldsymbol{\epsilon}_{\rho'} = \cos \phi' \boldsymbol{\epsilon}_1 + \sin \phi' \boldsymbol{\epsilon}_2$, we have

$$(\mathbf{n} \times \mathbf{E}_i) \times \mathbf{n} = (E_\rho \cos \phi' \cos \theta - E_z \sin \theta)(\boldsymbol{\epsilon}_2 \times \mathbf{n}) + E_\rho \sin \phi' \boldsymbol{\epsilon}_2. \quad (208)$$

The second term is an odd function of ϕ' and will give zero when integrated over \mathbf{x}'_\perp . Thus we must do the integral

$$\begin{aligned} \mathbf{I} &= \int d^2 x'_\perp (\mathbf{n} \times \mathbf{E}_i) \times \mathbf{n} e^{-ikx' \sin \theta} \\ &= \sqrt{\frac{2}{\pi}} \frac{q\omega}{\gamma v^2} (\boldsymbol{\epsilon}_2 \times \mathbf{n}) \int dx' dy' [K_1(\omega\rho'/\gamma v) \cos \theta \cos \phi' + (i/\gamma)K_0(\omega\rho'/\gamma v) \sin \theta] e^{-ikx' \sin \theta} \end{aligned} \quad (209)$$

Now, $\cos \phi' = x'/\rho'$ and

$$\cos \phi' K_1(\omega\rho'/\gamma v) = \frac{x'}{\rho'} K_1(\omega\rho'/\gamma v) = -\frac{\gamma v}{\omega} \frac{\partial K_0(\omega\rho'/\gamma v)}{\partial x'}. \quad (211)$$

Using this in the expression for \mathbf{I} , we can do an integration by parts and find

$$\begin{aligned} \mathbf{I} &= \sqrt{\frac{2}{\pi}} \frac{q\omega}{\gamma v^2} (\boldsymbol{\epsilon}_2 \times \mathbf{n}) \int dx' dy' \left[-\frac{ik\gamma v \sin \theta}{\omega} K_0(\omega\rho'/\gamma v) \cos \theta + \frac{i}{\gamma} K_0(\omega\rho'/\gamma v) \sin \theta \right] \\ &= \sqrt{\frac{2}{\pi}} \frac{q}{v} (-i \sin \theta) \left(k \cos \theta - \frac{\omega}{\gamma^2 v} \right) (\boldsymbol{\epsilon}_2 \times \mathbf{n}) \int dx' dy' K_0(\omega\rho'/\gamma v) e^{-ikx' \sin \theta}. \end{aligned} \quad (212)$$

The integral over x' can be done using the identity

$$\int_0^\infty dz K_0(\beta\sqrt{z^2 + t^2}) \cos(\alpha z) = \frac{\pi}{2\sqrt{\alpha^2 + \beta^2}} e^{-|t|\sqrt{\alpha^2 + \beta^2}}. \quad (213)$$

Thus,

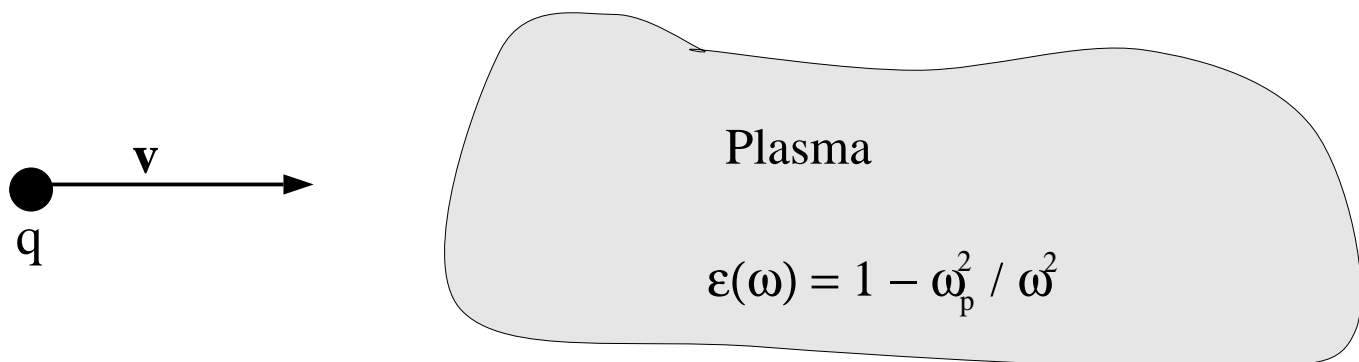
$$\begin{aligned}
\mathbf{I} &= \sqrt{\frac{2}{\pi}} \frac{q}{v} \boldsymbol{\epsilon}_2 \times \mathbf{n}(-i \sin \theta) \left(k \cos \theta - \frac{\omega}{\gamma^2 v} \right) \int_{-\infty}^{\infty} \frac{\pi dy' \exp(-|y'| \sqrt{k^2 \sin^2 \theta + \omega^2 / \gamma^2 v^2})}{\sqrt{k^2 \sin^2 \theta + \omega^2 / \gamma^2 v^2}} \\
&= \sqrt{\frac{2}{\pi}} \frac{q}{v} \boldsymbol{\epsilon}_2 \times \mathbf{n}(-i \sin \theta) \left(k \cos \theta - \frac{\omega}{\gamma^2 v} \right) \frac{2\pi}{k^2 \sin^2 \theta + \omega^2 / \gamma^2 v^2}. \tag{214}
\end{aligned}$$

Hence

$$\mathbf{E}_{rad} = e^{ikr} \frac{\epsilon(\omega) - 1}{4\pi r} \frac{k^2 \sin \theta}{\omega/v - k \cos \theta} \frac{2\sqrt{2\pi}(q/v) \boldsymbol{\epsilon}_2 \times \mathbf{n}}{k^2 \sin^2 \theta + \omega^2 / \gamma^2 v^2} \left(k \cos \theta - \frac{\omega}{\gamma^2 v} \right). \tag{215}$$

7.1 Cherenkov Radiation in a Dilute Collisionless Plasma

This is a general result. Let us look at some specific example in a simple limit in order to extract the salient qualitative features of transition radiation. First, we shall choose a dielectric function. A very simple one, which has the added virtue that for some frequencies $\epsilon < 1$ so that it is impossible to have Cherenkov radiation, is that for a dilute collisionless plasma, $\epsilon(\omega) = 1 - \omega_p^2 / \omega^2$.



For ω large enough, this function is less than, but close to, unity so that it satisfies the criteria for validity of the perturbation theory. Then

$$(\epsilon(\omega) - 1)k^2 / 4\pi = -\omega_p^2 k^2 / 4\pi \omega^2 = -\omega_p^2 / 4\pi c^2 \tag{216}$$

where we approximate ω by ck wherever it is not important.¹² Suppose also that

¹²This can be done in some places because $\epsilon(\omega) \approx 1$; in other places the difference $\epsilon(\omega) - 1$ is needed and here we cannot set $\epsilon(\omega)$ equal to 1.

$\gamma \gg 1$, i.e., that the particle is highly relativistic. These conditions allow us to approximate as follows:

$$\frac{\omega}{v} - k \cos \theta = \frac{\omega}{v} - \frac{\omega}{v_p} \cos \theta = \frac{\omega}{\beta c} - \frac{\omega}{c} \sqrt{\epsilon} \cos \theta \approx \frac{\omega}{c(1 - 1/2\gamma^2)} - \frac{\omega}{c} \left(1 - \frac{\omega_p^2}{2\omega^2}\right) \left(1 - \frac{\theta^2}{2}\right) \quad (217)$$

$$\approx \frac{\omega}{c} \left(\frac{1}{2\gamma^2} + \frac{\omega_p^2}{2\omega^2} + \frac{\theta^2}{2}\right) = \frac{\omega}{2\gamma^2 c} \left(1 + \frac{\gamma^2 \omega_p^2}{\omega^2} + \gamma^2 \theta^2\right) \equiv \frac{\omega}{2\gamma^2 c} \left(1 + \frac{1}{\nu^2} + \eta\right) \quad (218)$$

where

$$\nu \equiv \omega/\gamma\omega_p \quad \eta \equiv \gamma^2 \theta^2 \quad \beta = \sqrt{1 - \gamma^{-2}} \approx 1 - 1/2\gamma^2 \quad \text{and} \quad \sqrt{\epsilon} = \sqrt{1 - \frac{\omega_p^2}{\omega^2}} \approx 1 - \frac{\omega_p^2}{2\omega^2}. \quad (219)$$

Also,

$$\frac{\omega^2}{\gamma^2 v^2} + k^2 \sin^2 \theta = \frac{\omega^2}{\gamma^2 \beta^2 c^2} + \frac{\omega^2 \epsilon \sin^2 \theta}{c^2} = \frac{\omega^2}{\gamma^2 \beta^2 c^2} [1 + \gamma^2 \beta^2 \epsilon \sin^2 \theta] \approx \frac{\omega^2}{\gamma^2 c^2} (1 + \eta) \quad (220)$$

while

$$k \cos \theta - \frac{\omega}{v\gamma^2} = \frac{\omega}{c} \sqrt{\epsilon(\omega)} \cos \theta - \frac{\omega}{\beta c \gamma^2} = \frac{\omega}{c} \left(\sqrt{\epsilon(\omega)} \cos \theta - \frac{1}{\beta \gamma^2} \right) \approx \frac{\omega}{c}. \quad (221)$$

Hence,

$$\begin{aligned} \mathbf{E}_{rad} &= \frac{e^{ikr}}{r} \left(\frac{-\omega_p^2}{4\pi c^2} \right) \frac{\theta(2\sqrt{2\pi})(q/c)(\omega/c) \boldsymbol{\epsilon}_2 \times \mathbf{n}}{(\omega/2\gamma^2 c)(1 + \eta + 1/\nu^2)(\omega^2/\gamma^2 c^2)(1 + \eta)} \\ &= -\frac{e^{ikr}}{r} \frac{\omega_p^2}{4\pi} \frac{2\gamma^4 \theta 2\sqrt{2\pi} q \boldsymbol{\epsilon}_2 \times \mathbf{n} \omega}{c\omega(1 + \eta + 1/\nu^2)\omega^2(1 + \eta)} \\ &= -\frac{e^{ikr}}{r} \sqrt{\frac{2}{\pi}} \frac{\gamma\sqrt{\eta}}{\nu^2} \frac{q \boldsymbol{\epsilon}_2 \times \mathbf{n}}{c(1 + \eta + 1/\nu^2)(1 + \eta)}. \end{aligned} \quad (222)$$

The radiated energy per unit frequency per unit solid angle is¹³

$$\frac{d^2 I(\omega)}{d\Omega d\omega} = 2 \frac{c}{4\pi} |r \mathbf{E}_{rad}(\mathbf{x}, \omega)|^2 = \frac{q^2 \gamma^2}{\pi^2 c} \frac{\eta}{\nu^4 (1 + \eta + 1/\nu^2)^2 (1 + \eta)^2}. \quad (223)$$

¹³Previously, we just called this $dI/d\Omega$.

We can write $d\Omega = \sin\theta d\theta d\phi \approx \theta d\theta d\phi = d\eta d\phi/2\gamma^2$ and integrate over ϕ to find the distribution per unit η . Also, let's write the frequency in terms of ν , $d\nu = d\omega/\gamma\omega_p$, to find

$$\frac{d^2I}{d\eta d\nu} = \int d\phi \frac{d^2I}{d\Omega d\omega} \frac{d\omega}{d\nu} \frac{d\Omega}{d\eta} = \frac{d^2I}{d\Omega d\omega} \frac{\gamma\omega_p\pi}{\gamma^2} = \frac{q^2\gamma\omega_p}{\pi c} \frac{\eta}{\nu^4(1+\eta+1/\nu^2)^2(1+\eta)^2}. \quad (224)$$

This expression fails at small ν ($\omega \sim \omega_p$). It falls off as η^{-3} at large η . It peaks as a function of η around $\eta = 1$ for small ν and at $\eta = 1/3$ for large ν . If one integrates over η the result is

$$\frac{dI}{d\nu} = \frac{q^2\gamma\omega_p}{\pi c} [(1+2\nu^2)\ln(1+1/\nu^2) - 2]. \quad (225)$$

Further, the total energy radiated is

$$I = \int_0^\infty d\nu \frac{dI}{d\nu} = \frac{q^2\gamma\omega_p}{3c} = \frac{(q/e)^2\gamma\hbar\omega_p}{3} \frac{e^2}{\hbar c} = \frac{(q/e)^2}{3(137)}(\gamma\hbar\omega_p). \quad (226)$$

A typical photon energy is $\gamma\hbar\omega_p/3$ ($\nu = 1/3$), so the number of photons emitted on average is $(q/e)^2/137$ which is quite a bit smaller than one. However, a sizable amount of transition radiation can be obtained by employing a stack of thin slabs of material with adjacent slabs having significantly different dielectric constants. Then there is some radiation produced at each interface between different materials.

8 Example Problems

8.1 A Relativistic Particle in a Capacitor

A particle of charge e and mass m initially at rest is accelerated across a parallel plate capacitor held at (stat) voltage V ; the distance between the plates is d . Assuming nonrelativistic motion, find the total energy radiated by the particle during this process. Then, *without calculation, answer or estimate the following*:

1. The angular distribution of the radiated energy of the particle,

2. the order of magnitude of typical frequencies emitted from the charge, and
3. the angular distribution of radiated energy from the charge if it were highly relativistic.

Solution. The field in a parallel plate capacitor is constant, thus so is the acceleration of the charged particle and its radiation.

$$P = \frac{2}{3} \frac{e^2}{c^3} a^2 \quad W = \int P dt = \frac{2}{3} \frac{e^2}{c^3} t$$

where t is the duration of the pulse.

$$a = F/m = eE/m = eV/md \quad \text{and} \quad t = \sqrt{2d/a}$$

Hence, solving for W

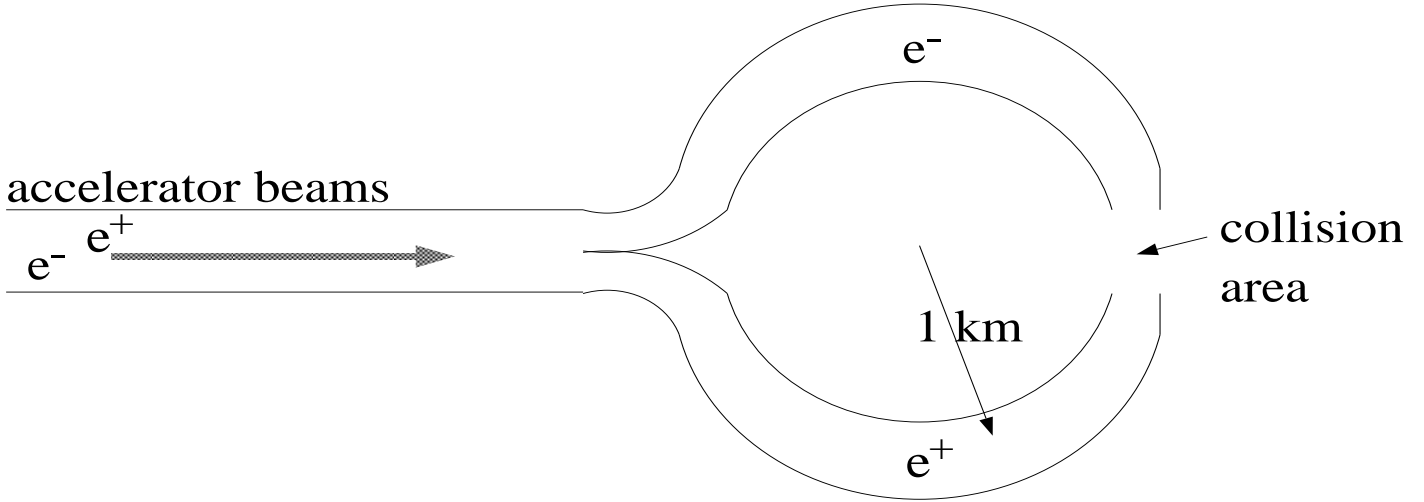
$$W = \frac{2\sqrt{2}e^7/2V^{3/2}}{3c^3m^{3/2}d}$$

In the non relativistic limit, the angular distribution of the radiation is given by $\sin^2(\theta)$ where the angle is measured relative to the velocity vector. The typical frequencies can be found using the Fourier uncertainty principle. Since the retarded duration of the pulse is the same as the observer's duration for a nonrelativistic particle, we have $\omega \sim 1/t = \sqrt{eV/md^2}$. In the relativistic limit the angular distribution of the radiation will be strongly pitched in the direction of the velocity/acceleration, but will be zero along the axis. The maximum of the pulse of radiation will be at an angle of $\theta \sim 1/\gamma$ away from the axis defined by the velocity vector.

8.2 Relativistic Electrons at SLAC

At the Stanford linear accelerator, devices have been added at the end of the accelerator to guide electrons and positrons around roughly semicircular paths until they collide head-on as shown in the sketch below. If each particle has a total energy of 50GeV and rest energy of 0.5 MeV, while the circular paths have radii of about

1km, roughly what is the fraction of the particles energy lost to radiation before the collision takes place?



Solution. To determine the power radiated, we must use the relativistic Larmor formula. For $\dot{\boldsymbol{\beta}} \perp \boldsymbol{\beta}$, $|\boldsymbol{\beta} \times \dot{\boldsymbol{\beta}}| = |\beta\dot{\beta}|$, so the power radiated is

$$P = \frac{2e^2}{3c} \gamma^6 \left[\dot{\boldsymbol{\beta}}^2 - (\boldsymbol{\beta} \times \dot{\boldsymbol{\beta}})^2 \right] = \frac{2e^2}{3c} \gamma^6 \beta^2 (1 - \beta^2) = \frac{2e^2}{3c} \gamma^4 \dot{\beta}^2$$

The acceleration is centripetal, so $\dot{\beta} = v^2/cr \approx c/r$, so

$$P = \frac{2e^2}{3c} \gamma^4 \frac{c^2}{r^2}$$

Counting the initial deflection of the particles into the circular region, both the electron and the positron travel about 3/4 of a circle. The time it takes to do this is roughly $\tau = 3\pi r/2c$ (assuming that the particles travel roughly at velocity c), so the energy radiated is

$$\Delta E = \frac{\pi e^2}{r} \gamma^4$$

Since $\gamma = E/mc^2$, the relative energy loss is

$$\frac{\Delta E}{E} = \frac{\pi e^2/r}{mc^2} \gamma^3$$

For $E = 50$ GeV and $mc^2 = 0.511$ MeV, $\gamma \approx 10^5$, and the relative energy loss is

$$\frac{\Delta E}{E} = \frac{\pi 23 (10^{-20}) (10^{15})}{0.911 (10^{-27}) (9) (10^{20}) (10^5)} \approx 0.9 \times 10^{-2}$$

or roughly only one percent of the energy is lost.