

5. Electromagnetism and Relativity

We've seen that Maxwell's equations have wave solutions which travel at the speed of light. But there's another place in physics where the speed of light plays a prominent role: the theory of special relativity. How does electromagnetism fit with special relativity?

Historically, the Maxwell equations were discovered before the theory of special relativity. It was thought that the light waves we derived above must be oscillations of some substance which fills all of space. This was dubbed the *aether*. The idea was that Maxwell's equations only hold in the frame in which the aether is at rest; light should then travel at speed c relative to the aether.

We now know that the concept of the aether is unnecessary baggage. Instead, Maxwell's equations hold in all inertial frames and are the first equations of physics which are consistent with the laws of special relativity. Ultimately, it was by studying the Maxwell equations that Lorentz was able to determine the form of the Lorentz transformations which subsequently laid the foundation for Einstein's vision of space and time.

Our goal in this section is to view electromagnetism through the lens of relativity. We will find that observers in different frames will disagree on what they call electric fields and what they call magnetic fields. They will observe different charge densities and different currents. But all will agree that these quantities are related by the same Maxwell equations. Moreover, there is a pay-off to this. It's only when we formulate the Maxwell equations in a way which is manifestly consistent with relativity that we see their true beauty. The slightly cumbersome vector calculus equations that we've been playing with throughout these lectures will be replaced by a much more elegant and simple-looking set of equations.

5.1 A Review of Special Relativity

We start with a very quick review of the relevant concepts of special relativity. (For more details see the lecture notes on [Dynamics and Relativity](#)). The basic postulate of relativity is that the laws of physics are the same in all inertial reference frames. The guts of the theory tell us how things look to observers who are moving relative to each other.

The first observer sits in an inertial frame \mathcal{S} with spacetime coordinates (ct, x, y, z) the second observer sits in an inertial frame \mathcal{S}' with spacetime coordinates (ct', x', y', z') .

If we take \mathcal{S}' to be moving with speed v in the x -direction relative to \mathcal{S} then the coordinate systems are related by the Lorentz boost

$$x' = \gamma \left(x - \frac{v}{c} ct \right) \quad \text{and} \quad ct' = \gamma \left(ct - \frac{v}{c} x \right) \quad (5.1)$$

while $y' = y$ and $z' = z$. Here c is the speed of light which has the value,

$$c = 299792458 \text{ m s}^{-1}$$

Meanwhile γ is the ubiquitous factor

$$\gamma = \sqrt{\frac{1}{1 - v^2/c^2}} \quad (5.2)$$

The Lorentz transformation (5.1) encodes within it all of the fun ideas of time dilation and length contraction that we saw in our first course on relativity.

5.1.1 Four-Vectors

It's extremely useful to package these spacetime coordinates in 4-vectors, with indices running from $\mu = 0$ to $\mu = 3$

$$X^\mu = (ct, x, y, z) \quad \mu = 0, 1, 2, 3$$

Note that the index is a superscript rather than subscript. This will be important shortly. A general Lorentz transformation is a linear map from X to X' of the form

$$(X')^\mu = \Lambda^\mu_\nu X^\nu$$

Here Λ is a 4×4 matrix which obeys the matrix equation

$$\Lambda^T \eta \Lambda = \eta \quad \Leftrightarrow \quad \Lambda^\rho_\mu \eta_{\rho\sigma} \Lambda^\sigma_\nu = \eta_{\mu\nu} \quad (5.3)$$

with $\eta_{\mu\nu}$ the *Minkowski metric*

$$\eta_{\mu\nu} = \text{diag}(+1, -1, -1, -1)$$

The solutions to (5.3) fall into two classes. The first class is simply rotations. Given a 3×3 rotation matrix R obeying $R^T R = 1$, we can construct a Lorentz transformation Λ obeying (5.3) by embedding R in the spatial part,

$$\Lambda^\mu_\nu = \left(\begin{array}{c|ccc} 1 & 0 & 0 & 0 \\ \hline 0 & & & \\ 0 & & R & \\ 0 & & & \end{array} \right) \quad (5.4)$$

These transformations describe how to relate the coordinates of two observers who are rotated with respect to each other.

The other class of solutions to (5.3) are the Lorentz boosts. These are the transformations appropriate for observers moving relative to each other. The Lorentz transformation (5.1) is equivalent to

$$\Lambda^\mu{}_\nu = \left(\begin{array}{cc|cc} \gamma & -\gamma v/c & 0 & 0 \\ -\gamma v/c & \gamma & 0 & 0 \\ \hline 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array} \right) \quad (5.5)$$

There are similar solutions associated to boosts along the y and z axes.

The beauty of 4-vectors is that it's extremely easy to write down invariant quantities. These are things which all observers, no matter which their reference frame, can agree on. To construct these we take the inner product of two 4-vectors. The trick is that this inner product uses the Minkowski metric and so comes with some minus signs. For example, the square of the distance from the origin to some point in spacetime labelled by X is

$$X \cdot X = X^\mu \eta_{\mu\nu} X^\nu = c^2 t^2 - x^2 - y^2 - z^2$$

which is the invariant interval. Similarly, if we're given two four-vectors X and Y then the inner product $X \cdot Y = X^\mu \eta_{\mu\nu} Y^\nu$ is also a Lorentz invariant.

5.1.2 Proper Time

The key to building relativistic theories of Nature is to find the variables that have nice properties under Lorentz transformations. The 4-vectors X , labelling spacetime points, are a good start. But we need more. Here we review how the other kinematical variables of velocity, momentum and acceleration fit into 4-vectors.

Suppose that, in some frame, the particle traces out a worldline. The clever trick is to find a way to parameterise this path in a way that all observers agree upon. The natural choice is the *proper time* τ , the duration of time experienced by the particle itself. If you're sitting in some frame, watching some particle move with an old-fashioned Newtonian 3-velocity $\mathbf{u}(t)$, then it's simple to show that the relationship between your time t and the proper time of the particle τ is given by

$$\frac{dt}{d\tau} = \gamma(\mathbf{u})$$

The proper time allows us to define the 4-velocity and the 4-momentum. Suppose that the particle traces out a path $X(\tau)$ in some frame. Then the 4-velocity is

$$U = \frac{dX}{d\tau} = \gamma \begin{pmatrix} c \\ \mathbf{u} \end{pmatrix}$$

Similarly, the 4-momentum is $P = mU$ where m is the rest mass of the particle. We write

$$P = \begin{pmatrix} E/c \\ \mathbf{p} \end{pmatrix} \tag{5.6}$$

where $E = m\gamma c^2$ is the energy of the particle and $\mathbf{p} = \gamma m\mathbf{u}$ is the 3-momentum in special relativity.

The importance of U and P is that they too are 4-vectors. Because all observers agree on τ , the transformation law of U and P are inherited from X . This means that under a Lorentz transformation, they too change as $U \rightarrow \Lambda U$ and $P \rightarrow \Lambda P$. And it means that inner products of U and P are guaranteed to be Lorentz invariant.

5.1.3 Indices Up, Indices Down

Before we move on, we do need to introduce one extra notational novelty. The minus signs in the Minkowski metric η means that it's useful to introduce a slight twist to the usual summation convention of repeated indices. For all the 4-vectors that we introduced above, we always place the spacetime index $\mu = 0, 1, 2, 3$ as a superscript (i.e. up) rather than a subscript.

$$X^\mu = \begin{pmatrix} ct \\ \mathbf{x} \end{pmatrix}$$

This is because the same object with an index down, X_μ , will mean something subtly different. We define

$$X_\mu = \begin{pmatrix} ct \\ -\mathbf{x} \end{pmatrix}$$

With this convention, the Minkowski inner product can be written using the usual convention of summing over repeated indices as

$$X^\mu X_\mu = c^2 t^2 - \mathbf{x} \cdot \mathbf{x}$$

In contrast, $X^\mu X^\mu = c^2 t^2 + \mathbf{x}^2$ is a dumb thing to write in the context of special relativity since it looks very different to observers in different inertial frames. In fact, we will shortly declare it illegal to write things like $X^\mu X^\mu$.

There is a natural way to think of X_μ in terms of X^μ using the Minkowski metric $\eta_{\mu\nu} = \text{diag}(+1, -1, -1, -1)$. The following equation is trivially true:

$$X_\mu = \eta_{\mu\nu} X^\nu$$

This means that we can think of the Minkowski metric as allowing us to lower indices. To raise indices back up, we need the inverse of $\eta_{\mu\nu}$ which, fortunately, is the same matrix: $\eta^{\mu\nu} = \text{diag}(+1, -1, -1, -1)$ which means we have $\eta^{\mu\rho}\eta_{\rho\nu} = \delta^\mu_\nu$ and we can write

$$X^\nu = \eta^{\nu\mu} X_\mu$$

From now on, we're going to retain this distinction between all upper and lower indices. All the four-vectors that we've met so far have upper indices. But all can be lowered in the same way. For example, we have

$$U_\mu = \gamma \begin{pmatrix} c \\ -\mathbf{u} \end{pmatrix} \quad (5.7)$$

This trick of distinguishing between indices up and indices down provides a simple formalism to ensure that all objects have nice transformation properties under the Lorentz group. We insist that, just as in the usual summation convention, repeated indices only ever appear in pairs. But now we further insist that pairs always appear with one index up and the other down. The result will be an object which is invariant under Lorentz transformations.

5.1.4 Vectors, Covectors and Tensors

In future courses, you will learn that there is somewhat deeper mathematics lying behind distinguishing X^μ and X_μ : formally, these objects live in different spaces (sometimes called dual spaces). We'll continue to refer to X^μ as vectors, but to distinguish them, we'll call X_μ *covectors*. (In slightly fancier language, the components of the vector X^μ are sometimes said to be *contravariant* while the components of the covector X_μ are said to be *covariant*).

For now, the primary difference between a vector and covector is how they transform under rotations and boosts. We know that, under a Lorentz transformation, any 4-vector changes as

$$X^\mu \rightarrow X'^\mu = \Lambda^\mu_\nu X^\nu$$

From this, we see that a covector should transform as

$$\begin{aligned} X_\mu &\rightarrow X'_\mu = \eta_{\mu\rho} X'^\rho \\ &= \eta_{\mu\rho} \Lambda^\rho_\sigma X^\sigma \\ &= \eta_{\mu\rho} \Lambda^\rho_\sigma \eta^{\sigma\nu} X_\nu \end{aligned}$$

Using our rule for raising and lowering indices, now applied to the Lorentz transformation Λ , we can also write this as

$$X_\mu \rightarrow \Lambda_\mu^\nu X_\nu$$

where our notation is now getting dangerously subtle: you have to stare to see whether the upper or lower index on the Lorentz transformation comes first.

There is a sense in which Λ_μ^ν can be thought of as the components of the inverse matrix Λ^{-1} . To see this, we go back to the definition of the Lorentz transformation (5.3), and start to use our new rules for raising and lowering indices

$$\begin{aligned} \Lambda^\rho_\mu \eta_{\rho\sigma} \Lambda^\sigma_\nu = \eta_{\mu\nu} &\Rightarrow \Lambda^\rho_\mu \Lambda_{\rho\nu} = \eta_{\mu\nu} \\ &\Rightarrow \Lambda^\rho_\mu \Lambda_\rho^\sigma = \delta_\mu^\sigma \\ &\Rightarrow \Lambda_\rho^\sigma \Lambda^\rho_\mu = \delta_\mu^\sigma \end{aligned}$$

In the last line above, we've simply reversed the order of the two terms on the left. (When written in index notation, these are just the entries of the matrix so there's no problem with commuting them). Now we compare this to the formula for the inverse of a matrix,

$$(\Lambda^{-1})^\sigma_\rho \Lambda^\rho_\mu = \delta_\mu^\sigma \quad \Rightarrow \quad (\Lambda^{-1})^\sigma_\rho = \Lambda_\rho^\sigma \quad (5.8)$$

Note that you need to be careful where you place the indices in equations like this. The result (5.8) is analogous to the statement that the inverse of a rotation matrix is the transpose matrix. For general Lorentz transformations, we learn that the inverse is sort of the transpose where “sort of” means that there are minus signs from raising and lowering. The placement of indices in (5.8) tells us where those minus signs go.

The upshot of (5.8) is that if we want to abandon index notation all together then vectors transform as $X \rightarrow \Lambda X$ while covectors – which, for the purpose of this sentence, we'll call \tilde{X} – transform as $\tilde{X} \rightarrow \Lambda^{-1} \tilde{X}$. However, in what follows, we have no intention of abandoning index notation. Instead, we will embrace it. It will be our friend and our guide in showing that the Maxwell equations are consistent with special relativity.

A particularly useful example of a covector is the *four-derivative*. This is the relativistic generalisation of ∇ , defined by

$$\partial_\mu = \frac{\partial}{\partial X^\mu} = \left(\frac{1}{c} \frac{\partial}{\partial t}, \nabla \right)$$

Notice that the superscript on the spacetime 4-vector X^μ has migrated to a subscript on the derivative ∂_μ . For this to make notational sense, we should check that ∂_μ does indeed transform as covector. This is a simple application of the chain rule. Under a Lorentz transformation, $X^\mu \rightarrow X'^\mu = \Lambda^\mu_\nu X^\nu$, so we have

$$\partial_\mu = \frac{\partial}{\partial X^\mu} \rightarrow \frac{\partial}{\partial X'^\mu} = \frac{\partial X^\nu}{\partial X'^\mu} \frac{\partial}{\partial X^\nu} = (\Lambda^{-1})^\nu_\mu \partial_\nu = \Lambda_\mu^\nu \partial_\nu$$

which is indeed the transformation of a co-vector.

Tensors

Vectors and covectors are the simplest examples of objects which have nice transformation properties under the Lorentz group. But there are many more examples. The most general object can have a bunch of upper indices and a bunch of lower indices, $T^{\mu_1 \dots \mu_n}_{\nu_1 \dots \nu_m}$. These objects are also called *tensors* of type (n, m) . In order to qualify as a tensor, they must transform under a Lorentz transformation as

$$T'^{\mu_1 \dots \mu_n}_{\nu_1 \dots \nu_m} = \Lambda^{\mu_1}_{\rho_1} \dots \Lambda^{\mu_n}_{\rho_n} \Lambda^{\sigma_1}_{\nu_1} \dots \Lambda^{\sigma_m}_{\nu_m} T^{\rho_1 \dots \rho_n}_{\sigma_1 \dots \sigma_m} \quad (5.9)$$

You can always use the Minkowski metric to raise and lower indices on tensors, changing the type of tensor but keeping the total number of indices $n + m$ fixed.

Tensors of this kind are the building blocks of all our theories. This is because if you build equations only out of tensors which transform in this manner then, as long as the μ, ν, \dots indices match up on both sides of the equation, you're guaranteed to have an equation that looks the same in all inertial frames. Such equations are said to be *covariant*. You'll see more of this kind of thing in courses on *General Relativity* and *Differential Geometry*.

In some sense, this index notation is too good. Remember all those wonderful things that you first learned about in special relativity: time dilation and length contraction and twins and spaceships so on. You'll never have to worry about those again. From now on, you can guarantee that you're working with a theory consistent with relativity by ensuring two simple things

- That you only deal with tensors.
- That the indices match up on both sides of the equation.

It's sad, but true. It's all part of growing up and not having fun anymore.

5.2 Conserved Currents

We started these lectures by discussing the charge density $\rho(\mathbf{x}, t)$, the current density $\mathbf{J}(\mathbf{x}, t)$ and their relation through the continuity equation,

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{J} = 0$$

which tells us that charge is locally conserved.

The continuity equation is already fully consistent with relativity. To see this, we first need to appreciate that the charge and current densities sit nicely together in a 4-vector,

$$J^\mu = \begin{pmatrix} \rho c \\ \mathbf{J} \end{pmatrix}$$

Of course, placing objects in a four-vector has consequence: it tells us how these objects look to different observers. Let's quickly convince ourselves that it makes sense that charge density and current do indeed transform in this way. We can start by considering a situation where there are only static charges with density ρ_0 and no current. So $J^\mu = (\rho_0, 0)$. Now, in a frame that is boosted by velocity \mathbf{v} , the current will appear as $J'^\mu = \Lambda^\mu{}_\nu J^\nu$ with the Lorentz transformation given by (5.5). The new charge density and current are then

$$\rho' = \gamma \rho_0 \quad , \quad \mathbf{J}' = -\gamma \rho_0 \mathbf{v}$$

The first of these equations tells us that different observers see different charge densities. This is because of Lorentz contraction: charge *density* means charge per unit volume. And the volume gets squeezed because lengths parallel to the motion undergo Lorentz contraction. That's the reason for the factor of γ in the observed charge density. Meanwhile, the second of these equations is just the relativistic extension of the formula $\mathbf{J} = \rho \mathbf{v}$ that we first saw in the introduction. (The extra minus sign is because \mathbf{v} here denotes the velocity of the boosted observer; the charge is therefore moving with relative velocity $-\mathbf{v}$).

In our new, relativistic, notation, the continuity equation takes the particularly simple form

$$\partial_\mu J^\mu = 0 \tag{5.10}$$

This equation is Lorentz invariant. This follows simply because the indices are contracted in the right way: one up, and one down.

5.2.1 Magnetism and Relativity

We've learned something unsurprising: boosted charge gives rise to a current. But, combined with our previous knowledge, this tells us something new and important: boosted electric fields must give rise to magnetic fields. The rest of this chapter will be devoted to understanding the details of how this happens. But first, we're going to look at a simple example where we can re-derive the magnetic force purely from the Coulomb force and a dose of Lorentz contraction.

To start, consider a bunch of positive charges $+q$ moving along a line with speed $+v$ and a bunch of negative charges $-q$ moving in the opposite direction with speed $-v$ as shown in the figure. If there is equal density, n , of positive and negative charges then the charge density vanishes while the current is

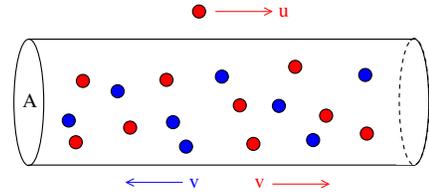


Figure 47:

$$I = 2nAqv$$

where A is the cross-sectional area of the wire. Now consider a test particle, also carrying charge q , which is moving parallel to the wire with some speed u . It doesn't feel any electric force because the wire is neutral, but we know it experiences a magnetic force. Here we will show how to find an expression for this force without ever invoking the phenomenon of magnetism.

The trick is to move to the rest frame of the test particle. This means we have to boost by speed u . The usual addition formula tells us that the velocities of the positive and negative charges now differ, given by

$$v_{\pm} = \frac{v \mp u}{1 \mp uv/c^2}$$

But with the boost comes a Lorentz contraction which means that the charge density changes. Moreover, because the velocities of positive and negative charges are now different, this will mean that, viewed from the rest frame of our particle, the wire is no longer neutral. Let's see how this works. First, we'll introduce n_0 , the density of charges when the particles in the wire are at rest. Then the density of the $+q$ charges in the original frame is

$$\rho = qn = \gamma(v)qn_0$$

The charge density of the $-q$ particles is the same, but with opposite sign, so that in the original frame the wire is neutral. However, in our new frame, the charge densities

are

$$\rho_{\pm} = qn_{\pm} = q\gamma(v_{\pm})n_0 = \left(1 \mp \frac{uv}{c^2}\right) \gamma(u)\gamma(v) qn_0$$

where you've got to do a little bit of algebra to get to the last result. Since $v_- > v_+$, we have $n_- > n_+$ and the wire carries negative charge. The overall net charge density in the new frame is

$$\rho' = qn' = q(n_+ - n_-) = -\frac{2uv}{c^2} \gamma(u) qn$$

But we know that a line of electric charge creates an electric field; we calculated it in (2.6); it is

$$E(r) = -\frac{2uv}{c^2} \frac{\gamma(u) qnA}{2\pi\epsilon_0 r} \hat{\mathbf{r}}$$

where r is the radial direction away from the wire. This means that, in its rest frame, the particle experiences a force

$$F' = -u\gamma(u) \frac{nAq^2v}{\pi\epsilon_0 c^2 r}$$

where the minus sign tells us that the force is towards the wire for $u > 0$. But if there's a force in one frame, there must also be a force in another. Transforming back to where we came from, we conclude that even when the wire is neutral there has to be a force

$$F = \frac{F'}{\gamma(u)} = -u \frac{nq^2Av}{\pi\epsilon_0 c^2 r} = -uq \frac{\mu_0 I}{2\pi r} \quad (5.11)$$

But this precisely agrees with the Lorentz force law, with the magnetic field given by the expression (3.5) that we computed for a straight wire. Notice that if $u > 0$ then the test particle – which has charge q – is moving in the same direction as the particles in the wire which have charge q and the force is attractive. If $u < 0$ then it moves in the opposite direction and the force is repulsive.

This analysis provides an explicit demonstration of how an electric force in one frame of reference is interpreted as a magnetic force in another. There's also something rather surprising about the result. We're used to thinking of length contraction as an exotic result which is only important when we approach the speed of light. Yet the electrons in a wire crawl along. They take around an hour to travel a meter! Nonetheless, we can easily detect the magnetic force between two wires which, as we've seen above, can be directly attributed to the length contraction in the electron density.

The discussion above needs a minor alteration for actual wires. In the rest frame of the wire the positive charges – which are ions, atoms stripped of some of their electrons – are stationary while the electrons move. Following the explanation above, you might think that there is an imbalance of charge density already in this frame. But that’s not correct. The current is due to some battery feeding electrons into the wire and taking them out the other end. And this is done in such a way that the wire is neutral in the rest frame, with the electron density exactly compensating the ion density. In contrast, if we moved to a frame in which the ions and electrons had equal and opposite speeds, the wire would appear charged. Although the starting point is slightly different, the end result remains.

5.3 Gauge Potentials and the Electromagnetic Tensor

Under Lorentz transformations, electric and magnetic fields will transform into each other. In this section, we want to understand more precisely how this happens. At first sight, it looks as if it’s going to be tricky. So far the objects which have nice transformation properties under Lorentz transformations are 4-vectors. But here we’ve got two 3-vectors, \mathbf{E} and \mathbf{B} . How do we make those transform into each other?

5.3.1 Gauge Invariance and Relativity

To get an idea for how this happens, we first turn to some objects that we met previously: the scalar and vector potentials ϕ and \mathbf{A} . Recall that we introduced these to solve some of the equations of electrostatics and magnetostatics,

$$\begin{aligned}\nabla \times \mathbf{E} = 0 &\quad \Rightarrow \quad \mathbf{E} = -\nabla\phi \\ \nabla \cdot \mathbf{B} = 0 &\quad \Rightarrow \quad \mathbf{B} = \nabla \times \mathbf{A}\end{aligned}$$

However, in general these expressions can’t be correct. We know that when \mathbf{B} and \mathbf{E} change with time, the two source-free Maxwell equations are

$$\nabla \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = 0 \quad \text{and} \quad \nabla \cdot \mathbf{B} = 0$$

Nonetheless, it’s still possible to use the scalar and vector potentials to solve both of these equations. The solutions are

$$\mathbf{E} = -\nabla\phi - \frac{\partial \mathbf{A}}{\partial t} \quad \text{and} \quad \mathbf{B} = \nabla \times \mathbf{A}$$

where now $\phi = \phi(\mathbf{x}, t)$ and $\mathbf{A} = \mathbf{A}(\mathbf{x}, t)$.

Just as we saw before, there is no unique choice of ϕ and \mathbf{A} . We can always shift $\mathbf{A} \rightarrow \mathbf{A} + \nabla\chi$ and \mathbf{B} remains unchanged. However, now this requires a compensating shift of ϕ .

$$\phi \rightarrow \phi - \frac{\partial\chi}{\partial t} \quad \text{and} \quad \mathbf{A} \rightarrow \mathbf{A} + \nabla\chi \quad (5.12)$$

with $\chi = \chi(\mathbf{x}, t)$. These are *gauge transformations*. They reproduce our earlier gauge transformation for \mathbf{A} , while also encompassing constant shifts in ϕ .

How does this help with our attempt to reformulate electromagnetism in a way compatible with special relativity? Well, now we have a scalar, and a 3-vector: these are ripe to place in a 4-vector. We define

$$A^\mu = \begin{pmatrix} \phi/c \\ \mathbf{A} \end{pmatrix}$$

Or, equivalently, $A_\mu = (\phi/c, -\mathbf{A})$. In this language, the gauge transformations (5.12) take a particularly nice form,

$$A_\mu \rightarrow A_\mu - \partial_\mu\chi \quad (5.13)$$

where χ is any function of space and time

5.3.2 The Electromagnetic Tensor

We now have all the ingredients necessary to determine how the electric and magnetic fields transform. From the 4-derivative $\partial_\mu = (\partial/\partial(ct), \nabla)$ and the 4-vector $A_\mu = (\phi/c, -\mathbf{A})$, we can form the anti-symmetric tensor

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$$

This is constructed to be invariant under gauge transformations (5.13). We have

$$F_{\mu\nu} \rightarrow F_{\mu\nu} + \partial_\mu\partial_\nu\chi - \partial_\nu\partial_\mu\chi = F_{\mu\nu}$$

This already suggests that the components involve the \mathbf{E} and \mathbf{B} fields. To check that this is indeed the case, we can do a few small computations,

$$F_{01} = \frac{1}{c} \frac{\partial(-A_x)}{\partial t} - \frac{\partial(\phi/c)}{\partial x} = \frac{E_x}{c}$$

and

$$F_{12} = \frac{\partial(-A_y)}{\partial x} - \frac{\partial(-A_x)}{\partial y} = -B_z$$

Similar computations for all other entries give us a matrix of electric and magnetic fields,

$$F_{\mu\nu} = \begin{pmatrix} 0 & E_x/c & E_y/c & E_z/c \\ -E_x/c & 0 & -B_z & B_y \\ -E_y/c & B_z & 0 & -B_x \\ -E_z/c & -B_y & B_x & 0 \end{pmatrix} \quad (5.14)$$

This, then, is the answer to our original question. You can make a Lorentz covariant object consisting of two 3-vectors by arranging them in an anti-symmetric tensor. $F_{\mu\nu}$ is called the *electromagnetic tensor*. Equivalently, we can raise both indices using the Minkowski metric to get

$$F^{\mu\nu} = \eta^{\mu\rho}\eta^{\nu\sigma}F_{\rho\sigma} = \begin{pmatrix} 0 & -E_x/c & -E_y/c & -E_z/c \\ E_x/c & 0 & -B_z & B_y \\ E_y/c & B_z & 0 & -B_x \\ E_z/c & -B_y & B_x & 0 \end{pmatrix}$$

Both $F_{\mu\nu}$ and $F^{\mu\nu}$ are tensors. They are tensors because they're constructed out of objects, A_μ , ∂_μ and $\eta_{\mu\nu}$, which themselves transform nicely under the Lorentz group. This means that the field strength must transform as

$$F'^{\mu\nu} = \Lambda^\mu{}_\rho \Lambda^\nu{}_\sigma F^{\rho\sigma} \quad (5.15)$$

Alternatively, if you want to get rid of the indices, this reads $F' = \Lambda F \Lambda^T$. The observer in a new frame sees electric and magnetic fields \mathbf{E}' and \mathbf{B}' that differ from the original observer. The two are related by (5.15). Let's look at what this means in a couple of illustrative examples.

Rotations

To compute the transformation (5.15), it's probably simplest to just do the sums that are implicit in the repeated ρ and σ labels. Alternatively, if you want to revert to matrix multiplication then this is the same as $F' = \Lambda F \Lambda^T$. Either way, we get the same result. For a rotation, the 3×3 matrix R is embedded in the lower-right hand block of Λ as shown in (5.4). A quick calculation shows that the transformation of the electric and magnetic fields in (5.15) is as expected,

$$\mathbf{E}' = R\mathbf{E} \quad \text{and} \quad \mathbf{B}' = R\mathbf{B}$$

Boosts

Things are more interesting for boosts. Let's consider a boost v in the x -direction, with Λ given by (5.5). Again, you need to do a few short calculations. For example, we have

$$\begin{aligned} -\frac{E'_x}{c} &= F'^{01} = \Lambda^0_\rho \Lambda^1_\sigma F^{\rho\sigma} \\ &= \Lambda^0_0 \Lambda^1_1 F^{01} + \Lambda^0_1 \Lambda^1_0 F^{10} \\ &= \frac{\gamma^2 v^2}{c^2} \frac{E_x}{c} - \gamma^2 \frac{E_x}{c} = -\frac{E_x}{c} \end{aligned}$$

and

$$\begin{aligned} -\frac{E'_y}{c} &= F'^{02} = \Lambda^0_\rho \Lambda^2_\sigma F^{\rho\sigma} \\ &= \Lambda^0_0 \Lambda^2_2 F^{02} + \Lambda^0_1 \Lambda^2_2 F^{12} \\ &= -\gamma \frac{E_y}{c} + \frac{\gamma v}{c} B_z = -\frac{\gamma}{c} (E_y - v B_z) \end{aligned}$$

and

$$\begin{aligned} -B'_z &= F'^{12} = \Lambda^1_\rho \Lambda^2_\sigma F^{\rho\sigma} \\ &= \Lambda^1_0 \Lambda^2_2 F^{02} + \Lambda^1_1 \Lambda^2_2 F^{12} \\ &= \frac{\gamma v}{c^2} E_y - \gamma B_z = -\gamma (B_z - v E_y / c^2) \end{aligned}$$

The final result for the transformation of the electric field after a boost in the x -direction is

$$\begin{aligned} E'_x &= E_x \\ E'_y &= \gamma (E_y - v B_z) \\ E'_z &= \gamma (E_z + v B_y) \end{aligned} \tag{5.16}$$

and, for the magnetic field,

$$\begin{aligned} B'_x &= B_x \\ B'_y &= \gamma \left(B_y + \frac{v}{c^2} E_z \right) \\ B'_z &= \gamma \left(B_z - \frac{v}{c^2} E_y \right) \end{aligned} \tag{5.17}$$

As we anticipated above, what appears to be a magnetic field to one observer looks like an electric field to another, and vice versa.

Note that in the limit $v \ll c$, we have $\mathbf{E}' = \mathbf{E} + \mathbf{v} \times \mathbf{B}$ and $\mathbf{B}' = \mathbf{B}$. This can be thought of as the Galilean boost of electric and magnetic fields. We recognise $\mathbf{E} + \mathbf{v} \times \mathbf{B}$ as the combination that appears in the Lorentz force law. We'll return to this force in Section 5.4.1 where we'll see how it's compatible with special relativity.

5.3.3 An Example: A Boosted Line Charge

In Section 2.1.3, we computed the electric field due to a line with uniform charge density η per unit length. If we take the line to lie along the x -axis, we have (2.6)

$$\mathbf{E} = \frac{\eta}{2\pi\epsilon_0(y^2 + z^2)} \begin{pmatrix} 0 \\ y \\ z \end{pmatrix} \quad (5.18)$$

Meanwhile, the magnetic field vanishes for static electric charges: $\mathbf{B} = 0$. Let's see what this looks like from the perspective of an observer moving with speed v in the x -direction, parallel to the wire. In the moving frame the electric and magnetic fields are given by (5.16) and (5.17). These read

$$\begin{aligned} \mathbf{E}' &= \frac{\eta\gamma}{2\pi\epsilon_0(y^2 + z^2)} \begin{pmatrix} 0 \\ y \\ z \end{pmatrix} = \frac{\eta\gamma}{2\pi\epsilon_0(y'^2 + z'^2)} \begin{pmatrix} 0 \\ y' \\ z' \end{pmatrix} \\ \mathbf{B}' &= \frac{\eta\gamma v}{2\pi\epsilon_0 c^2(y^2 + z^2)} \begin{pmatrix} 0 \\ z \\ -y \end{pmatrix} = \frac{\eta\gamma v}{2\pi\epsilon_0 c^2(y'^2 + z'^2)} \begin{pmatrix} 0 \\ z' \\ -y' \end{pmatrix} \end{aligned} \quad (5.19)$$

In the second equality, we've rewritten the expression in terms of the coordinates of \mathcal{S}' which, because the boost is in the x -direction, are trivial: $y = y'$ and $z = z'$.

From the perspective of an observer in frame \mathcal{S}' , the charge density in the wire is $\eta' = \gamma\eta$, where the factor of γ comes from Lorentz contraction. This can be seen in the expression above for the electric field. Since the charge density is now moving, the observer in frame \mathcal{S}' sees a current $I' = -\gamma\eta v$. Then we can rewrite (5.19) as

$$\mathbf{B}' = \frac{\mu_0 I'}{2\pi\sqrt{y'^2 + z'^2}} \hat{\phi}' \quad (5.20)$$

But this is something that we've seen before. It's the magnetic field due to a current in a wire (3.5). We computed this in Section 3.1.1 using Ampère's law. But here we've re-derived the same result without ever mentioning Ampère's law! Instead, our starting point (5.18) needed Gauss' law and we then used only the Lorentz transformation of electric and magnetic fields. We can only conclude that, under a Lorentz transformation, Gauss' law must be related to Ampère's law. Indeed, we'll shortly see explicitly that this is the case. For now, it's worth repeating the lesson that we learned in Section 5.2.1: the magnetic field can be viewed as a relativistic effect.

5.3.4 Another Example: A Boosted Point Charge

Consider a point charge Q , stationary in an inertial frame \mathcal{S} . We know that its electric field is given by

$$\mathbf{E} = \frac{Q}{4\pi\epsilon_0 r^2} \hat{\mathbf{r}} = \frac{Q}{4\pi\epsilon_0 [x^2 + y^2 + z^2]^{3/2}} \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

while its magnetic field vanishes. Now let's look at this same particle from the frame \mathcal{S}' , moving with velocity $\mathbf{v} = (v, 0, 0)$ with respect to \mathcal{S} . The Lorentz boost which relates the two is given by (5.5) and so the new electric field are given by (5.16),

$$\mathbf{E}' = \frac{Q}{4\pi\epsilon_0 [x^2 + y^2 + z^2]^{3/2}} \begin{pmatrix} x \\ \gamma y \\ \gamma z \end{pmatrix}$$

But this is still expressed in terms of the original coordinates. We should now rewrite this in terms of the coordinates of \mathcal{S}' , which are $x' = \gamma(x - vt)$ and $y' = y$ and $z' = z$. Inverting these, we have

$$\mathbf{E}' = \frac{Q\gamma}{4\pi\epsilon_0 [\gamma^2(x' + vt')^2 + y'^2 + z'^2]^{3/2}} \begin{pmatrix} x' + vt' \\ y' \\ z' \end{pmatrix} \quad (5.21)$$

In the frame \mathcal{S}' , the particle sits at $\mathbf{x}' = (-vt', 0, 0)$, so we see that the electric field emanates from the position of the charge, as it should. For now, let's look at the electric field when $t' = 0$ so that the particle sits at the origin in the new frame. The electric field points outwards radially, along the direction

$$\mathbf{r}' = \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix}$$

However, the electric field is not isotropic. This arises from the denominator of (5.21) which is not proportional to r'^3 because there's an extra factor of γ^2 in front of the x' component. Instead, at $t' = 0$, the denominator involves the combination

$$\begin{aligned} \gamma^2 x'^2 + y'^2 + z'^2 &= (\gamma^2 - 1)x'^2 + \mathbf{r}'^2 \\ &= \frac{v^2 \gamma^2}{c^2} x'^2 + \mathbf{r}'^2 \\ &= \left(\frac{v^2 \gamma^2}{c^2} \cos^2 \theta + 1 \right) \mathbf{r}'^2 \\ &= \gamma^2 \left(1 - \frac{v^2}{c^2} \sin^2 \theta \right) \mathbf{r}'^2 \end{aligned}$$

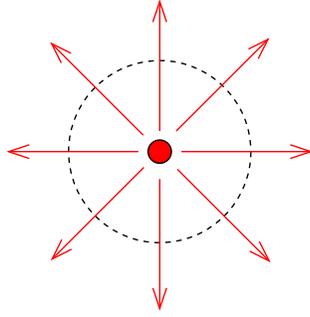


Figure 48: The isotropic field lines of a static charge

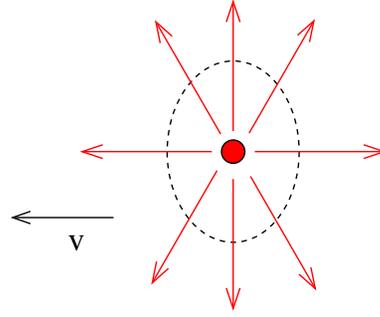


Figure 49: The squeezed field lines of a moving charge

where the θ is the angle between \mathbf{r}' and the x' -axis and, in the last line, we've just used some simple trig and the definition of $\gamma^2 = 1/(1 - v^2/c^2)$. This means that we can write the electric field in frame S' as

$$\mathbf{E}' = \frac{1}{\gamma^2(1 - v^2 \sin^2 \theta/c^2)^{3/2}} \frac{Q}{4\pi\epsilon_0 r'^2} \hat{\mathbf{r}}'$$

The pre-factor is responsible for the fact that the electric field is not isotropic. We see that it reduces the electric field along the x' -axis (i.e. when $\theta = 0$) and increases the field along the perpendicular y' and z' axes (i.e. when $\theta = \pi/2$). This can be thought of as a consequence of Lorentz contraction, squeezing the electric field lines in the direction of travel.

The moving particle also gives rise to a magnetic field. This is easily computed using the Lorentz transformations (5.17). It is

$$\mathbf{B} = \frac{\mu_0 Q \gamma v}{4\pi[\gamma^2(x' + vt')^2 + y'^2 + z'^2]^{3/2}} \begin{pmatrix} 0 \\ z' \\ -y' \end{pmatrix}$$

5.3.5 Lorentz Scalars

We can now ask a familiar question: is there any combination of the electric and magnetic fields that all observers agree upon? Now we have the power of index notation at our disposal, this is easy to answer. We just need to write down an object that doesn't have any floating μ or ν indices. Unfortunately, we don't get to use the obvious choice of $\eta_{\mu\nu} F^{\mu\nu}$ because this vanishes on account of the anti-symmetry of $F^{\mu\nu}$. The simplest thing we can write down is

$$\frac{1}{2} F_{\mu\nu} F^{\mu\nu} = -\frac{\mathbf{E}^2}{c^2} + \mathbf{B}^2$$

Note the relative minus sign between \mathbf{E} and \mathbf{B} , mirroring a similar minus sign in the spacetime interval.

However, this isn't the only Lorentz scalar that we can construct from \mathbf{E} and \mathbf{B} . There is another, somewhat more subtle, object. To build this, we need to appreciate that Minkowski spacetime comes equipped with another natural tensor object, beyond the familiar metric $\eta_{\mu\nu}$. This is the fully anti-symmetric object known as the *alternating tensor*,

$$\epsilon^{\mu\nu\rho\sigma} = \begin{cases} +1 & \text{if } \mu\nu\rho\sigma \text{ is an even permutation of } 0123 \\ -1 & \text{if } \mu\nu\rho\sigma \text{ is an odd permutation of } 0123 \end{cases}$$

while $\epsilon^{\mu\nu\rho\sigma} = 0$ if there are any repeated indices.

To see why this is a natural object in Minkowski space, let's look at how it changes under Lorentz transformations. The usual tensor transformation is

$$\epsilon'^{\mu\nu\rho\sigma} = \Lambda^\mu{}_\kappa \Lambda^\nu{}_\lambda \Lambda^\rho{}_\alpha \Lambda^\sigma{}_\beta \epsilon^{\kappa\lambda\alpha\beta}$$

It's simple to check that $\epsilon'^{\mu\nu\rho\sigma}$ is also full anti-symmetric; it inherits this property from $\epsilon^{\kappa\lambda\alpha\beta}$ on the right-hand side. But this means that $\epsilon'^{\mu\nu\rho\sigma}$ must be proportional to $\epsilon^{\mu\nu\rho\sigma}$. We only need to determine the constant of proportionality. To do this, we can look at

$$\epsilon'^{0123} = \Lambda^0{}_\kappa \Lambda^1{}_\lambda \Lambda^2{}_\alpha \Lambda^3{}_\beta \epsilon^{\kappa\lambda\alpha\beta} = \det(\Lambda)$$

Now any Lorentz transformations have $\det(\Lambda) = \pm 1$. Those with $\det(\Lambda) = 1$ make up the “proper Lorentz group” $SO(1,3)$. (This was covered in the *Dynamics and Relativity* notes). These proper Lorentz transformations do not include reflections or time reversal. We learn that the alternating tensor $\epsilon^{\mu\nu\rho\sigma}$ is invariant under proper Lorentz transformations. What it's really telling us is that Minkowski space comes with an oriented orthonormal basis. By lowering indices with the Minkowski metric, we can also construct the tensor $\epsilon_{\mu\nu\rho\sigma}$ which has $\epsilon_{0123} = -1$.

The alternating tensor allows us to construct a second tensor field, sometimes called the *dual electromagnetic tensor* (although “dual” is perhaps the most overused word in physics),

$$\tilde{F}^{\mu\nu} = \frac{1}{2} \epsilon^{\mu\nu\rho\sigma} F_{\rho\sigma} = \begin{pmatrix} 0 & -B_x & -B_y & -B_z \\ B_x & 0 & E_z/c & -E_y/c \\ B_y & -E_z/c & 0 & E_x/c \\ B_z & E_y/c & -E_x/c & 0 \end{pmatrix} \quad (5.22)$$

$\tilde{F}^{\mu\nu}$ is sometimes also written as $*F^{\mu\nu}$. We see that this looks just like $F^{\mu\nu}$ but with the electric and magnetic fields swapped around. Actually, looking closely you'll see that there's a minus sign difference as well: $\tilde{F}^{\mu\nu}$ arises from $F^{\mu\nu}$ by the substitution $\mathbf{E} \rightarrow c\mathbf{B}$ and $\mathbf{B} \rightarrow -\mathbf{E}/c$.

The statement that $\tilde{F}^{\mu\nu}$ is a tensor means that it too has nice properties under Lorentz transformations,

$$\tilde{F}'^{\mu\nu} = \Lambda^\mu{}_\rho \Lambda^\nu{}_\sigma \tilde{F}^{\rho\sigma}$$

and we can use this to build new Lorentz invariant quantities. Taking the obvious square of \tilde{F} doesn't give us anything new, since

$$\tilde{F}^{\mu\nu} \tilde{F}_{\mu\nu} = -F^{\mu\nu} F_{\mu\nu}$$

But by contracting \tilde{F} with the original F we do find a new Lorentz invariant

$$\frac{1}{4} \tilde{F}^{\mu\nu} F_{\mu\nu} = \frac{1}{c} \mathbf{E} \cdot \mathbf{B}$$

This tells us that the inner-product of \mathbf{E} and \mathbf{B} is the same viewed in all frames.

5.4 Maxwell Equations

We now have the machinery to write the Maxwell equations in a way which is manifestly compatible with special relativity. They take a particularly simple form:

$$\partial_\mu F^{\mu\nu} = \mu_0 J^\nu \quad \text{and} \quad \partial_\mu \tilde{F}^{\mu\nu} = 0 \quad (5.23)$$

Pretty aren't they!

The Maxwell equations are not *invariant* under Lorentz transformations. This is because there is the dangling ν index on both sides. However, because the equations are built out of objects which transform nicely – $F^{\mu\nu}$, $\tilde{F}^{\mu\nu}$, J^μ and ∂_μ – the equations themselves also transform nicely. For example, we will see shortly that Gauss' law transforms into Ampère's law under a Lorentz boost, something we anticipated in Section 5.3.3. We say that the equations are *covariant* under Lorentz transformations.

This means that an observer in a different frame will mix everything up: space and time, charges and currents, and electric and magnetic fields. Although observers disagree on what these things are, they all agree on how they fit together. This is what it means for an equation to be covariant: the ingredients change, but the relationship between them stays the same. All observers agree that, in their frame, the electric and magnetic fields are governed by the same Maxwell equations.

Given the objects $F^{\mu\nu}$, $\tilde{F}^{\mu\nu}$, J^μ and ∂_μ , the Maxwell equations are not the only thing you could write down compatible with Lorentz invariance. But they are by far the simplest. Any other equation would be non-linear in F or \tilde{F} or contain more derivative terms or some such thing. Of course, simplicity is no guarantee that equations are correct. For this we need experiment. But surprisingly often in physics we find that the simplest equations are also the right ones.

Unpacking the Maxwell Equations

Let's now check that the Maxwell equations (5.23) in relativistic form do indeed coincide with the vector calculus equations that we've been studying in this course. We just need to expand the different parts of the equation. The components of the first Maxwell equation give

$$\begin{aligned}\partial_i F^{i0} = \mu_0 J^0 &\Rightarrow \nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0} \\ \partial_\mu F^{\mu i} = \mu_0 J^i &\Rightarrow -\frac{1}{c^2} \frac{\partial \mathbf{E}}{\partial t} + \nabla \times \mathbf{B} = \mu_0 \mathbf{J}\end{aligned}$$

In the first equation, which arises from $\nu = 0$, we sum only over spatial indices $i = 1, 2, 3$ because $F^{00} = 0$. Meanwhile the components of the second Maxwell equation give

$$\begin{aligned}\partial_i \tilde{F}^{i0} = 0 &\Rightarrow \nabla \cdot \mathbf{B} = 0 \\ \partial_\mu \tilde{F}^{\mu i} = 0 &\Rightarrow \frac{\partial \mathbf{B}}{\partial t} + \nabla \times \mathbf{E} = 0\end{aligned}$$

These, of course, are the familiar equations that we've all grown to love over this course.

Here a few further, simple comments about the advantages of writing the Maxwell equations in relativistic form. First, the Maxwell equations imply that current is conserved. This follows because $F^{\mu\nu}$ is anti-symmetric, so $\partial_\mu \partial_\nu F^{\mu\nu} = 0$ automatically, simply because $\partial_\mu \partial_\nu$ is symmetric. The first of the Maxwell equations (5.23) then requires that the continuity equation holds

$$\partial_\mu J^\mu = 0$$

This is the same calculation that we did in vector notation in Section 4.2.1. Note that it's marginally easier in the relativistic framework.

The second Maxwell equation can be written in a number of different ways. It is equivalent to:

$$\partial_\mu \tilde{F}^{\mu\nu} = 0 \quad \Leftrightarrow \quad \epsilon^{\mu\nu\rho\sigma} \partial_\nu F_{\rho\sigma} = 0 \quad \Leftrightarrow \quad \partial_\rho F_{\mu\nu} + \partial_\nu F_{\rho\mu} + \partial_\mu F_{\nu\rho} = 0$$

where the last of these equalities follows because the equation is constructed so that it is fully anti-symmetric with respect to exchanging any of the indices ρ, μ and ν . (Just expand out for a few examples to see this).

The gauge potential A_μ was originally introduced to solve the two Maxwell equations which are contained in $\partial_\mu \tilde{F}^{\mu\nu} = 0$. Again, this is marginally easier to see in relativistic notation. If we write $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ then

$$\partial_\mu \tilde{F}^{\mu\nu} = \frac{1}{2} \epsilon^{\mu\nu\rho\sigma} \partial_\mu F_{\rho\sigma} = \frac{1}{2} \epsilon^{\mu\nu\rho\sigma} \partial_\mu (\partial_\rho A_\sigma - \partial_\sigma A_\rho) = 0$$

where the final equality holds because of the symmetry of the two derivatives, combined with the anti-symmetry of the ϵ -tensor. This means that we could equally well write the Maxwell equations as

$$\partial_\mu F^{\mu\nu} = \mu_0 J^\nu \quad \text{where} \quad F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$$

The first of these coincides with the first equation in (5.23); the second is an alternative way of writing the second equation in (5.23). In more advanced formulations of electromagnetism (for example, in the Lagrangian formulation), this is the form in which the Maxwell equations arise.

5.4.1 The Lorentz Force Law

There's one last aspect of electromagnetism that we need to show is compatible with relativity: the Lorentz force law. In the Newtonian world, the equation of motion for a particle moving with velocity \mathbf{u} and momentum $\mathbf{p} = m\mathbf{u}$ is

$$\frac{d\mathbf{p}}{dt} = q(\mathbf{E} + \mathbf{u} \times \mathbf{B}) \tag{5.24}$$

We want to write this equation in 4-vector notation in a way that makes it clear how all the objects change under Lorentz transformations.

By now it should be intuitively clear how this is going to work. A moving particle experiences the magnetic force. But if we boost to its rest frame, there is no magnetic force. Instead, the magnetic field transforms into an electric field and we find the same force, now interpreted as an electric force.

The relativistic version of (5.24) involves the 4-momentum P^μ , defined in (5.6), the proper time τ , reviewed in Section 5.1.2, and our new friend the electromagnetic tensor $F^{\mu\nu}$. The electromagnetic force acting on a point particle of charge q can then be written as

$$\frac{dP^\mu}{d\tau} = q F^{\mu\nu} U_\nu \tag{5.25}$$

where the 4-velocity is

$$U^\mu = \frac{dX^\mu}{d\tau} = \gamma \begin{pmatrix} c \\ \mathbf{u} \end{pmatrix} \quad (5.26)$$

and the 4-momentum is $P = mU$. Again, we see that the relativistic form of the equation (5.25) is somewhat prettier than the original equation (5.24).

Unpacking the Lorentz Force Law

Let's check to see that the relativistic equation (5.25) is giving us the right physics. It is, of course, four equations: one for each $\mu = 0, 1, 2, 3$. It's simple to multiply out the right-hand side, remembering that U_μ comes with an extra minus sign in the spatial components relative to (5.26). We find that the $\mu = 1, 2, 3$ components of (5.25) arrange themselves into a familiar vector equation,

$$\frac{d\mathbf{p}}{d\tau} = q\gamma(\mathbf{E} + \mathbf{u} \times \mathbf{B}) \quad \Rightarrow \quad \frac{d\mathbf{p}}{dt} = q(\mathbf{E} + \mathbf{u} \times \mathbf{B}) \quad (5.27)$$

where we've used the relationship $dt/d\tau = \gamma$. We find that we recover the Lorentz force law. Actually, there's a slight difference from the usual Newtonian force law (5.24), although the difference is buried in our notation. In the Newtonian setting, the momentum is $\mathbf{p} = m\mathbf{u}$. However, in the relativistic setting above, the momentum is $\mathbf{p} = m\gamma\mathbf{u}$. Needless to say, the relativistic version is correct, although the difference only shows up at high speeds.

The relativistic formulation of the Lorentz force (5.25) also contains an extra equation coming from $\mu = 0$. This reads

$$\frac{dP^0}{d\tau} = \frac{q}{c} \gamma \mathbf{E} \cdot \mathbf{u} \quad (5.28)$$

Recall that the temporal component of the four-momentum is the energy $P^0 = E/c$. Here the energy is $E = m\gamma c^2$ which includes both the rest-mass of the particle and its kinetic energy. The extra equation in (5.25) is simply telling us that the kinetic energy increases when work is done by an electric field

$$\frac{d(\text{Energy})}{dt} = q\mathbf{E} \cdot \mathbf{u}$$

where I've written energy as a word rather than as E to avoid confusing it with the electric field \mathbf{E} .

5.4.2 Motion in Constant Fields

We already know how electric and magnetic fields act on particles in a Newtonian world. Electric fields accelerate particles in straight lines; magnetic fields make particles go in circles. Here we're going to redo this analysis in the relativistic framework. The Lorentz force law remains the same. The only difference is that momentum is now $\mathbf{p} = m\gamma\mathbf{u}$. We'll see how this changes things.

Constant Electric Field

Consider a vanishing magnetic field and constant electric field $\mathbf{E} = (E, 0, 0)$. (Note that E here denotes electric field, not energy!). The equation of motion (5.27) for a charged particle with velocity $\mathbf{u} = (u, 0, 0)$ is

$$m \frac{d(\gamma u)}{dt} = qE \quad \Rightarrow \quad m\gamma u = qEt$$

where we've implicitly assumed that the particle starts from rest at $t = 0$. Rearranging, we get

$$u = \frac{dx}{dt} = \frac{qEt}{\sqrt{m^2 + q^2 E^2 t^2 / c^2}}$$

Reassuringly, the speed never exceeds the speed of light. Instead, $u \rightarrow c$ as $t \rightarrow \infty$ as one would expect. It's simple to integrate this once more. If the particle starts from the origin, we have

$$x = \frac{mc^2}{qE} \left(\sqrt{1 + \frac{q^2 E^2 t^2}{m^2 c^2}} - 1 \right)$$

For early times, when the speeds are not too high, this reduces to

$$mx \approx \frac{1}{2} qEt^2 + \dots$$

which is the usual non-relativistic result for particles undergoing constant acceleration in a straight line.

Constant Magnetic Field

Now let's turn the electric field off and look at the case of constant magnetic field $\mathbf{B} = (0, 0, B)$. In the non-relativistic world, we know that particles turn circles with frequency $\omega = qB/m$. Let's see how relativity changes things.

We start by looking at the zeroth component of the force equation (5.28) which, in the absence of an electric field, reads

$$\frac{dP^0}{d\tau} = 0$$

This tells us that magnetic fields do no work. We knew this from our course on Newtonian physics, but it remains true in the relativistic context. So we know that energy, $E = m\gamma c^2$, is constant. But this tells us that the speed (i.e. the magnitude of the velocity) remains constant. In other words, the velocity, and hence the position, once again turn circles. The equation of motion is now

$$m \frac{d(\gamma \mathbf{u})}{dt} = q \mathbf{u} \times \mathbf{B}$$

Since γ is constant, the equation takes the same form as in the non-relativistic case and the solutions are circles (or helices if the particle also moves in the z -direction). The only difference is that the frequency with which the particle moves in a circle now depends on how fast the particle is moving,

$$\omega = \frac{qB}{m\gamma}$$

If you wanted, you could interpret this as due to the relativistic increase in the mass of a moving particle. Naturally, for small speeds $\gamma \approx 1$ and we reproduce the more familiar cyclotron frequency $\omega \approx qB/m$.

So far we have looked at situations in which $\mathbf{E} = 0$ and in which $\mathbf{B} = 0$. But we've seen that $\mathbf{E} \cdot \mathbf{B} = 0$ and $\mathbf{E}^2 - \mathbf{B}^2$ are both Lorentz invariant quantities. This means that the solutions we've described above can be boosted to apply to any situation where $\mathbf{E} \cdot \mathbf{B} = 0$ and $\mathbf{E}^2 - \mathbf{B}^2$ is either > 0 or < 0 . In the general situation, both electric and magnetic fields are turned on so $\mathbf{E} \cdot \mathbf{B} \neq 0$ and we have three possibilities to consider depending on whether $\mathbf{E}^2 - \mathbf{B}^2$ is > 0 or < 0 or $= 0$.

6. Electromagnetic Radiation

We've seen that Maxwell's equations allow for wave solutions. This is light. Or, more generally, electromagnetic radiation. But how do you generate these waves from a collection of electric charges? In other words, how do you make light?

We know that a stationary electric charge produce a stationary electric field. If we boost this charge so it moves at a constant speed, it produces a stationary magnetic field. In this section, we will see that propagating electromagnetic waves are created by *accelerating* charges.

6.1 Retarded Potentials

We start by simply solving the Maxwell equations for a given current distribution $J^\mu = (\rho c, \mathbf{J})$. We did this in Section 2 and Section 3 for situations where both charges and currents are independent of time. Here we're going to solve the Maxwell equations in full generality where the charges and currents are time dependent.

We know that we can solve half of Maxwell's equations by introducing the gauge potential $A_\mu = (\phi/c, -\mathbf{A})$ and writing $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$. Then the remaining equations become

$$\partial_\nu F^{\nu\mu} = \mu_0 J^\mu \quad \Rightarrow \quad \square A^\mu - \partial^\mu(\partial_\nu A^\nu) = \mu_0 J^\mu \quad (6.1)$$

where \square is the wave operator: $\square = \partial_\mu \partial^\mu = (1/c^2)\partial^2/\partial t^2 - \nabla^2$.

This equation is invariant under gauge transformations

$$A^\mu \rightarrow A^\mu + \partial^\mu \chi \quad (6.2)$$

Any two gauge potentials related by the transformation (6.2) are considered physically equivalent. We will use this symmetry to help us solve (6.1). To do this we make a gauge choice:

Claim: We can use the gauge symmetry (6.2) to choose A^μ to satisfy

$$\partial_\mu A^\mu = 0 \quad (6.3)$$

This is known as *Lorentz Gauge*. It was actually discovered by a guy named Lorenz who had the misfortune to discover a gauge choice that is Lorentz invariant: all observers will agree on the gauge condition (6.3).

Proof: Suppose you are handed a gauge potential A_μ which doesn't obey (6.3) but, instead, satisfies $\partial_\mu A^\mu = f$ for some function f . Then do a gauge transformation of the form (6.2). Your new gauge potential will obey $\partial_\mu A^\mu + \square\chi = f$. This means that if you can find a gauge transformation χ which satisfies $\square\chi = f$ then your new gauge potential will be in Lorentz gauge. Such a χ can always be found. This follows from general facts about differential equations. (Note that this proof is essentially the same as we used in Section 3.2.2 when proving that we could always choose Coulomb gauge $\nabla \cdot \mathbf{A} = 0$). \square

If we are in Lorentz gauge then the Maxwell equations (6.1) become particularly simple; they reduce to the sourced wave equation

$$\square A^\mu = \left(\frac{1}{c^2} \frac{\partial^2}{\partial t^2} - \nabla^2 \right) A^\mu = \mu_0 J^\mu \quad (6.4)$$

Our goal is to solve this equation, subject to the condition (6.3). We'll assume that J has compact spatial support, meaning that the charges and currents are restricted to some finite region of space. As an aside, notice that this is the same kind of equation as $\square\chi = -f$ which we needed to solve to go Lorentz gauge in the first place. This means that the methods we develop below will allow us to figure out both how to go to Lorentz gauge, and also how to solve for A_μ once we're there.

In the following, we'll solve (6.4) in two (marginally) different ways. The first way is quicker; the second way gives us a deeper understanding of what's going on.

6.1.1 Green's Function for the Helmholtz Equation

For our first method, we will Fourier transform A_μ and J_μ in time, but not in space. We write

$$A_\mu(\mathbf{x}, t) = \int_{-\infty}^{+\infty} \frac{d\omega}{2\pi} \tilde{A}_\mu(\mathbf{x}, \omega) e^{-i\omega t} \quad \text{and} \quad J_\mu(\mathbf{x}, t) = \int_{-\infty}^{+\infty} \frac{d\omega}{2\pi} \tilde{J}_\mu(\mathbf{x}, \omega) e^{-i\omega t}$$

Now the Fourier components $\tilde{A}_\mu(\mathbf{x}, \omega)$ obey the equation

$$\left(\nabla^2 + \frac{\omega^2}{c^2} \right) \tilde{A}_\mu = -\mu_0 \tilde{J}_\mu \quad (6.5)$$

This is the *Helmholtz equation* with source given by the current \tilde{J} .

When $\omega = 0$, the Helmholtz equation reduces to the Poisson equation that we needed in our discussion of electrostatics. We solved the Poisson equation using the method of Green's functions when discussing electrostatics in Section 2.2.3. Here we'll do the same for the Helmholtz equation. The Green's function for the Helmholtz equation obeys

$$\left(\nabla^2 + \frac{\omega^2}{c^2}\right) G_\omega(\mathbf{x}; \mathbf{x}') = \delta^3(\mathbf{x} - \mathbf{x}')$$

Translational and rotational invariance ensure that the solutions to this equation are of the form $G_\omega(\mathbf{x}; \mathbf{x}') = G_\omega(r)$ with $r = |\mathbf{x} - \mathbf{x}'|$. We can then write this as the ordinary differential equation,

$$\frac{1}{r^2} \frac{d}{dr} \left(r^2 \frac{dG_\omega}{dr} \right) + \frac{\omega^2}{c^2} G_\omega = \delta^3(r) \quad (6.6)$$

We want solutions that vanish as $r \rightarrow \infty$. However, even with this restriction, there are still two such solutions. Away from the origin, they take the form

$$G_\omega \sim \frac{e^{\pm i\omega r/c}}{r}$$

We will see shortly that there is a nice physical interpretation of these two Green's functions. First, let's figure out the coefficient that sits in front of the Green's function. This is determined by the delta-function. We integrate both sides of (6.6) over a ball of radius R . We get

$$4\pi \int_0^R dr r^2 \left[\frac{1}{r^2} \frac{d}{dr} \left(r^2 \frac{dG_\omega}{dr} \right) + \frac{\omega^2}{c^2} G_\omega \right] = 1$$

Now, taking the limit $R \rightarrow 0$, only the first term on the left-hand side survives. Moreover, only the first term of $dG_\omega/dr \sim (-1/r^2 \pm i\omega/cr)e^{\pm i\omega r/c}$ survives. We find that the two Green's functions for the Helmholtz equation are

$$G_\omega(r) = -\frac{1}{4\pi} \frac{e^{\pm i\omega r/c}}{r}$$

Note that this agrees with the Green's function for the Poisson equation when $\omega = 0$.

Retarded Potentials

So which \pm sign should we take? The answer depends on what we want to do with the Green's function. For our purposes, we'll nearly always need $G_\omega \sim e^{+i\omega r/c}/r$. Let's see

why. The Green's function G_ω allows us to write the Fourier components \tilde{A}_μ in (6.5) as

$$\tilde{A}_\mu(\mathbf{x}, \omega) = \frac{\mu_0}{4\pi} \int d^3x' \frac{e^{+i\omega|\mathbf{x}-\mathbf{x}'|/c}}{|\mathbf{x}-\mathbf{x}'|} \tilde{J}_\mu(\mathbf{x}', \omega)$$

which, in turn, means that the time-dependent gauge potential becomes

$$A_\mu(\mathbf{x}, t) = \frac{\mu_0}{4\pi} \int \frac{d\omega}{2\pi} \int d^3x' \frac{e^{-i\omega(t-|\mathbf{x}-\mathbf{x}'|/c)}}{|\mathbf{x}-\mathbf{x}'|} \tilde{J}_\mu(\mathbf{x}')$$

But now the integral over ω is just the inverse Fourier transform. With one difference: what was the time variable t has become the *retarded time*, t_{ret} , with

$$ct_{\text{ret}} = ct - |\mathbf{x} - \mathbf{x}'|$$

We have our final result,

$$A_\mu(\mathbf{x}, t) = \frac{\mu_0}{4\pi} \int d^3x' \frac{J_\mu(\mathbf{x}', t_{\text{ret}})}{|\mathbf{x} - \mathbf{x}'|} \quad (6.7)$$

This is called the *retarded potential*. To determine the contribution at point \mathbf{x} and time t , we integrate the current over all of space, weighted with the Green's function factor $1/|\mathbf{x} - \mathbf{x}'|$ which captures the fact that points further away contribute more weakly.

After all this work, we've arrived at something rather nice. The general form of the answer is very similar to the result for electrostatic potential and magnetostatic vector potential that we derived in Sections 2 and 3. Recall that when the charge density and current were independent of time, we found

$$\phi(\mathbf{x}) = \frac{1}{4\pi\epsilon_0} \int d^3x' \frac{\rho(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|} \quad \text{and} \quad \mathbf{A}(\mathbf{x}) = \frac{\mu_0}{4\pi} \int d^3x' \frac{\mathbf{J}(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|}$$

But when the charge density and current do depend on time, we see from (6.7) that something new happens: the gauge field at point \mathbf{x} and time t depends on the current configuration at point \mathbf{x}' and the *earlier* time $t_{\text{ret}} = t - |\mathbf{x} - \mathbf{x}'|/c$. This, of course, is due to causality. The difference $t - t_{\text{ret}}$ is just the time it took the signal to propagate from \mathbf{x}' to \mathbf{x} , travelling at the speed of light. Of course, we know that Maxwell's equations are consistent with relativity so something like this had to happen; we couldn't have signals travelling instantaneously. Nonetheless, it's pleasing to see how this drops out of our Green's functionology.

Finally, we can see what would happen were we to choose the other Green's function, $G_\omega \sim e^{-i\omega r/c}/r$. Following through the steps above, we see that the retarded time t_{ret} is replaced by the advanced time $t_{\text{adv}} = t + |\mathbf{x} - \mathbf{x}'|/c$. Such a solution would mean that the gauge field depends on what the current is doing in the future, rather than in the past. These solutions are typically thrown out as being unphysical. We'll have (a little) more to say about them at the end of the next section.

6.1.2 Green's Function for the Wave Equation

The expression for the retarded potential (6.7) is important. In this section, we provide a slightly different derivation. This will give us more insight into the origin of the retarded and advanced solutions. Moreover, the techniques below will also be useful in later courses⁴.

We started our previous derivation by Fourier transforming only the time coordinate, to change the wave equation into the Helmholtz equation. Here we'll treat time and space on more equal footing and solve the wave equation directly. We again make use of Green's functions. The Green's function for the wave equation obeys

$$\left(\nabla^2 - \frac{1}{c^2} \frac{\partial^2}{\partial t^2}\right) G(\mathbf{x}, t; \mathbf{x}', t') = \delta^3(\mathbf{x} - \mathbf{x}') \delta(t - t') \quad (6.8)$$

Translational invariance in space and time means that the Green's function takes the form $G(\mathbf{x}, t; \mathbf{x}', t) = G(\mathbf{x} - \mathbf{x}', t - t')$. To determine this function $G(\mathbf{r}, t)$, with $\mathbf{r} = \mathbf{x} - \mathbf{x}'$, we Fourier transform both space and time coordinates,

$$G(\mathbf{x}, t) = \int \frac{d\omega d^3k}{(2\pi)^4} \tilde{G}(\mathbf{k}, \omega) e^{i(\mathbf{k}\cdot\mathbf{r} - \omega t)} \quad (6.9)$$

Choosing $\mathbf{x}' = 0$ and $t' = 0$, the wave equation (6.8) then becomes

$$\begin{aligned} \left(\nabla^2 - \frac{1}{c^2} \frac{\partial^2}{\partial t^2}\right) G(\mathbf{r}, t) &= \int \frac{d\omega d^3k}{(2\pi)^4} \tilde{G}(\omega, \mathbf{k}) \left(\nabla^2 - \frac{1}{c^2} \frac{\partial^2}{\partial t^2}\right) e^{i(\mathbf{k}\cdot\mathbf{r} - \omega t)} \\ &= \int \frac{d\omega d^3k}{(2\pi)^4} \tilde{G}(\mathbf{k}, \omega) \left(-k^2 + \frac{\omega^2}{c^2}\right) e^{i(\mathbf{k}\cdot\mathbf{r} - \omega t)} \\ &= \delta^3(\mathbf{r}) \delta(t) = \int \frac{d\omega d^3k}{(2\pi)^4} e^{i(\mathbf{k}\cdot\mathbf{r} - \omega t)} \end{aligned}$$

Equating the terms inside the integral, we see that the Fourier transform of the Green's function takes the simple form

$$\tilde{G}(\mathbf{k}, \omega) = -\frac{1}{k^2 - \omega^2/c^2}$$

⁴A very similar discussion can be found in the lecture notes on *Quantum Field Theory*.

But notice that this diverges when $\omega^2 = c^2 k^2$. This pole results in an ambiguity in the Green's function in real space which, from (6.9), is given by

$$G(\mathbf{r}, t) = - \int \frac{d\omega d^3k}{(2\pi)^4} \frac{1}{k^2 - \omega^2/c^2} e^{i(\mathbf{k}\cdot\mathbf{r} - \omega t)}$$

We need some way of dealing with that pole in the integral. To see what's going on, it's useful to change to polar coordinates for the momentum integrals over \mathbf{k} . This will allow us to deal with that $e^{i\mathbf{k}\cdot\mathbf{r}}$ factor. The best way to do this is to think of fixing \mathbf{r} and then choosing the k_z axis to point along x . We then write $\mathbf{k}\cdot\mathbf{r} = kr \cos\theta$, and the Green's function becomes

$$G(\mathbf{r}, t) = - \frac{1}{(2\pi)^4} \int_0^{2\pi} d\phi \int_0^\pi d\theta \sin\theta \int_0^\infty dk k^2 \int_{-\infty}^{+\infty} d\omega \frac{1}{k^2 - \omega^2/c^2} e^{i(kr \cos\theta - \omega t)}$$

Now the $d\phi$ integral is trivial, while the $d\theta$ integral is

$$\int_0^\pi d\theta \sin\theta e^{ikr \cos\theta} = - \frac{1}{ikr} \int_0^\pi d\theta \left[\frac{d}{d\theta} e^{ikr \cos\theta} \right] = - \frac{1}{ikr} [e^{-ikr} - e^{+ikr}] = 2 \frac{\sin kr}{kr}$$

After performing these angular integrals, the real space Green's function becomes

$$G(\mathbf{r}, t) = \frac{1}{4\pi^3} \int_0^\infty dk c^2 k^2 \frac{\sin kr}{kr} \int_{-\infty}^{+\infty} d\omega \frac{e^{-i\omega t}}{(\omega - ck)(\omega + ck)}$$

Now we have to face up to those poles. We'll work by fixing k and doing the ω integral first. (Afterwards, we'll then have to do the k integral). It's clear that we run into two poles at $\omega = \pm ck$ when we do the ω integral and we need a prescription for dealing with these. To do this, we need to pick a contour C in the complex ω plane which runs along the real axis but skips around the poles. There are different choices for C . Each of them provides a Green's function which obeys (6.8) but, as we will now see, these Green's functions are different. What's more, this difference has a nice physical interpretation.

Retarded Green's Function

To proceed, let's just pick a particular C and see what happens. We choose a contour which skips above the poles at $\omega = \pm ck$ as shown in the diagram. This results in what's called the *retarded Greens function*; we denote it as $G_{\text{ret}}(\mathbf{r}, t)$. As we now show, it depends crucially on whether $t < 0$ or $t > 0$.

Let's first look at the case with $t < 0$. Here, $e^{-i\omega t} \rightarrow 0$ when $\omega \rightarrow +i\infty$. This means that, for $t < 0$, we can close the contour C in the upper-half plane as shown in the figure and the extra semi-circle doesn't give rise to any further contribution. But there are no poles in the upper-half plane. This means that, by the Cauchy residue theorem, $G_{\text{ret}}(\mathbf{r}, t) = 0$ when $t < 0$.

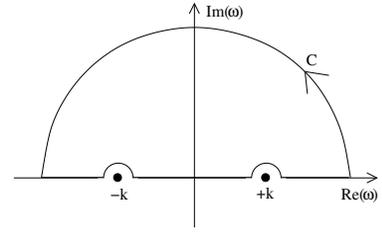


Figure 50:

In contrast, when $t > 0$ we have $e^{-i\omega t} \rightarrow 0$ when $\omega \rightarrow -i\infty$, which means that we get to close the contour in the lower-half plane. Now we do pick up contributions to the integral from the two poles at $\omega = \pm ck$. This time the Cauchy residue theorem gives

$$\begin{aligned} \int_C d\omega \frac{e^{-i\omega t}}{(\omega - ck)(\omega + ck)} &= -2\pi i \left[\frac{e^{-ickt}}{2ck} - \frac{e^{+ickt}}{2ck} \right] \\ &= -\frac{2\pi}{ck} \sin ckt \end{aligned} \quad (t > 0)$$

So, for $t > 0$, the Green's function becomes

$$\begin{aligned} G_{\text{ret}}(\mathbf{r}, t) &= -\frac{1}{2\pi^2} \frac{1}{r} \int_0^\infty dk c \sin kr \sin ckt \\ &= \frac{1}{4\pi^2} \frac{1}{r} \int_{-\infty}^\infty dk \frac{c}{4} (e^{ikr} - e^{-ikr})(e^{ickt} - e^{-ickt}) \\ &= \frac{1}{4\pi^2} \frac{1}{r} \int_{-\infty}^\infty dk \frac{c}{4} (e^{ik(r+ct)} + e^{-ik(r+ct)} - e^{ik(r-ct)} - e^{-ik(r-ct)}) \end{aligned}$$

Each of these final integrals is a delta-function of the form $\delta(r \pm ct)$. But, obviously, $r > 0$ while this form of the Green's function is only valid for $t > 0$. So the $\delta(r + ct)$ terms don't contribute and we're left with

$$G_{\text{ret}}(\mathbf{x}, t) = -\frac{1}{4\pi} \frac{c}{r} \delta(r - ct) \quad t > 0$$

We can absorb the factor of c into the delta-function. (Recall that $\delta(x/a) = |a|\delta(x)$ for any constant a). So we finally get the answer for the *retarded Green's function*

$$G_{\text{ret}}(\mathbf{r}, t) = \begin{cases} 0 & t < 0 \\ -\frac{1}{4\pi r} \delta(t_{\text{ret}}) & t > 0 \end{cases}$$

where t_{ret} is the retarded time that we met earlier,

$$t_{\text{ret}} = t - \frac{r}{c}$$

The delta-function ensures that the Green's function is only non-vanishing on the light-cone emanating from the origin.

Finally, with the retarded Green's function in hand, we can construct what we really want: solutions to the wave equation (6.4). These solutions are given by

$$\begin{aligned}
 A_\mu(\mathbf{x}, t) &= -\mu_0 \int d^3x' dt' G_{\text{ret}}(\mathbf{x}, t; \mathbf{x}', t') J_\mu(\mathbf{x}', t') \\
 &= \frac{\mu_0}{4\pi} \int d^3x' dt' \frac{\delta(t_{\text{ret}})}{|\mathbf{x} - \mathbf{x}'|} J_\mu(\mathbf{x}', t') \\
 &= \frac{\mu_0}{4\pi} \int d^3x' \frac{J_\mu(\mathbf{x}', t_{\text{ret}})}{|\mathbf{x} - \mathbf{x}'|}
 \end{aligned} \tag{6.10}$$

Happily, we find the same expression for the retarded potential that we derived previously in (6.7).

Advanced Green's Function

Let us briefly look at other Green's functions. We can pick the contour C in the complex ω -plane to skip below the two poles on the real axis. This results in what's called the *advanced Green's function*. Now, when $t > 0$, we complete the contour in the lower-half plane, as shown in the figure, where the lack of poles means that the advanced Green's function vanishes. Meanwhile, for $t < 0$, we complete the contour in the upper-half plane and get

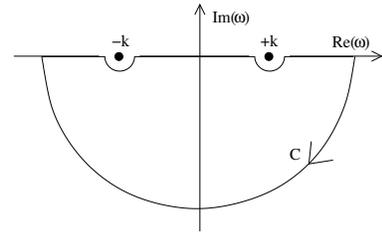


Figure 51:

$$G_{\text{adv}}(\mathbf{r}, t) = \begin{cases} -\frac{1}{4\pi r} \delta(t_{\text{adv}}) & t < 0 \\ 0 & t > 0 \end{cases}$$

where

$$t_{\text{adv}} = t + \frac{r}{c}$$

The resulting solution gives a solution known as the advanced potential,

$$A_\mu(\mathbf{x}, t) = \frac{\mu_0}{4\pi} \int d^3x' \frac{J_\mu(\mathbf{x}', t_{\text{adv}})}{|\mathbf{x} - \mathbf{x}'|}$$

It's hard to think of this solution as anything other than unphysical. Taken at face value, the effect of the current and charges now propagates backwards in time to determine the gauge potential A_μ . The sensible thing is clearly to throw these solutions away.

However, it's worth pointing out that the choice of the retarded propagator G_{ret} rather than the advanced propagator G_{adv} is an extra ingredient that we should add to the theory of electromagnetism. The Maxwell equations themselves are time symmetric; the choice of which solutions are physical is not.

There is some interesting history attached to this. A number of physicists have felt uncomfortable at imposing this time asymmetry only at the level of solutions, and attempted to rescue the advanced propagator in some way. The most well-known of these is the Feynman-Wheeler absorber theory, which uses a time symmetric propagator, with the time asymmetry arising from boundary conditions. However, I think it's fair to say that these ideas have not resulted in any deeper understanding of how time emerges in physics.

Finally, there is yet another propagator that we can use. This comes from picking a contour C that skips under the first pole and over the second. It is known as the *Feynman propagator* and plays an important role in quantum field theory.

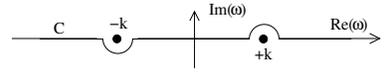


Figure 52:

6.1.3 Checking Lorentz Gauge

There is a loose end hanging over from our previous discussion. We have derived the general solution to the wave equation (6.4) for A_μ . This is given by the retarded potential

$$A_\mu(\mathbf{x}, t) = \frac{\mu_0}{4\pi} \int d^3x' \frac{J_\mu(\mathbf{x}', t_{\text{ret}})}{|\mathbf{x} - \mathbf{x}'|} \quad (6.11)$$

But the wave equation is only equivalent to the Maxwell equations if it obeys the Lorentz gauge fixing condition, $\partial_\mu A^\mu = 0$. We still need to check that this holds. In fact, this follows from the conservation of the current: $\partial_\mu J^\mu = 0$. To show this, it's actually simplest to return to a slightly earlier form of this expression (6.10)

$$A_\mu(\mathbf{x}, t) = -\mu_0 \int d^3x' dt' G_{\text{ret}}(\mathbf{x}, t; \mathbf{x}', t') J_\mu(\mathbf{x}', t')$$

The advantage of this is that both time and space remain on an equal footing. We have

$$\partial_\mu A^\mu(\mathbf{x}, t) = -\mu_0 \int d^3x' dt' \partial_\mu G_{\text{ret}}(\mathbf{x}, t; \mathbf{x}', t') J^\mu(\mathbf{x}', t')$$

But now we use the fact that $G_{\text{ret}}(\mathbf{x}, t; \mathbf{x}', t')$ depends on $\mathbf{x} - \mathbf{x}'$ and $t - t'$ to change the derivative ∂_μ acting on x into a derivative ∂'_μ acting on x' . We pick up a minus sign for

our troubles. We then integrate by parts to find,

$$\begin{aligned}\partial_\mu A^\mu(\mathbf{x}, t) &= +\mu_0 \int d^3x' dt' \partial'_\mu G_{\text{ret}}(\mathbf{x}, t; \mathbf{x}', t') J^\mu(\mathbf{x}', t') \\ &= -\mu_0 \int d^3x' dt' G_{\text{ret}}(\mathbf{x}, t; \mathbf{x}', t') \partial'_\mu J^\mu(\mathbf{x}', t') \\ &= 0\end{aligned}$$

as required. If you prefer, you can also run through the same basic steps with the form of the solution (6.11). You have to be a little careful because t_{ret} now also depends on \mathbf{x} and \mathbf{x}' so you get extra terms at various stages when you differentiate. But it all drops out in the wash and you again find that Lorentz gauge is satisfied courtesy of current conservation.

6.2 Dipole Radiation

Let's now use our retarded potential to understand something new. This is the setup: there's some localised region V in which there is a time-dependent distribution of charges and currents. But we're a long way from this region. We want to know what the resulting electromagnetic field looks like.

Our basic formula is the retarded potential,

$$A_\mu(\mathbf{x}, t) = \frac{\mu_0}{4\pi} \int_V d^3x' \frac{J_\mu(\mathbf{x}', t_{\text{ret}})}{|\mathbf{x} - \mathbf{x}'|} \quad (6.12)$$

The current $J^\mu(\mathbf{x}', t)$ is non-zero only for $\mathbf{x}' \in V$. We denote the size of the region V as d and we're interested in what's happening at a point \mathbf{x} which is a distance $r = |\mathbf{x}|$ away. (A word of warning: in this section we're using $r = |\mathbf{x}|$ which differs from our notation in Section 6.1 where we used $r = |\mathbf{x} - \mathbf{x}'|$). If $|\mathbf{x} - \mathbf{x}'| \gg d$ for all $\mathbf{x}' \in V$ then we can approximate $|\mathbf{x} - \mathbf{x}'| \approx |\mathbf{x}| = r$. In fact, we will keep the leading order correction to this which we get by Taylor expansion. (This is the same Taylor expansion that we needed when deriving the multipole expansion for electrostatics in Section 2.2.3). We have

$$\begin{aligned}|\mathbf{x} - \mathbf{x}'| &= r - \frac{\mathbf{x} \cdot \mathbf{x}'}{r} + \dots \\ \Rightarrow \frac{1}{|\mathbf{x} - \mathbf{x}'|} &= \frac{1}{r} + \frac{\mathbf{x} \cdot \mathbf{x}'}{r^3} + \dots\end{aligned} \quad (6.13)$$

There is a new ingredient compared to the electrostatic case: we have a factor of $|\mathbf{x} - \mathbf{x}'|$ that sits inside $t_{\text{ret}} = t - |\mathbf{x} - \mathbf{x}'|/c$ as well, so that

$$J_\mu(\mathbf{x}', t_{\text{ret}}) = J_\mu(\mathbf{x}', t - r/c + \mathbf{x} \cdot \mathbf{x}'/rc + \dots)$$

Now we'd like to further expand out this argument. But, to do that, we need to know something about what the current is doing. We will assume that the motion of the charges and current are *non-relativistic* so that the current doesn't change very much over the time $\tau \sim d/c$ that it takes light to cross the region V . For example, if the current varies with characteristic frequency ω (so that $J \sim e^{-i\omega t}$), then this requirement becomes $d/c \ll 1/\omega$. Then we can further Taylor expand the current to write

$$J_\mu(\mathbf{x}', t_{\text{ret}}) = J_\mu(\mathbf{x}', t - r/c) + \dot{J}_\mu(\mathbf{x}', t - r/c) \frac{\mathbf{x} \cdot \mathbf{x}'}{rc} + \dots \quad (6.14)$$

We start by looking at the leading order terms in both these Taylor expansions.

6.2.1 Electric Dipole Radiation

At leading order in d/r , the retarded potential becomes simply

$$A_\mu(\mathbf{x}, t) \approx \frac{\mu_0}{4\pi r} \int_V d^3x' J_\mu(\mathbf{x}', t - r/c)$$

This is known as the *electric dipole approximation*. (We'll see why very shortly). We want to use this to compute the electric and magnetic fields far from the localised source. It turns out to be simplest to first compute the magnetic field using the 3-vector form of the above equation,

$$\mathbf{A}(\mathbf{x}, t) \approx \frac{\mu_0}{4\pi r} \int_V d^3x' \mathbf{J}(\mathbf{x}', t - r/c)$$

We can manipulate the integral of the current using the conservation formula $\dot{\rho} + \nabla \cdot \mathbf{J} = 0$. (The argument is basically a repeat of the kind of arguments we used in the magnetostatics section 3.3.2). We do this by first noting the identity

$$\partial_j(J_j x_i) = (\partial_j J_j) x_i + J_i = -\dot{\rho} x_i + J_i$$

We integrate this over all of space and discard the total derivative to find

$$\int d^3x' \mathbf{J}(\mathbf{x}') = \frac{d}{dt} \int d^3x' \rho(\mathbf{x}') \mathbf{x}' = \dot{\mathbf{p}}$$

where we recognise \mathbf{p} as the electric dipole moment of the configuration. We learn that the vector potential is determined by the change of the electric dipole moment,

$$\mathbf{A}(\mathbf{x}, t) \approx \frac{\mu_0}{4\pi r} \dot{\mathbf{p}}(t - r/c)$$

This, of course, is where the *electric dipole* approximation gets its name.

We now use this to compute the magnetic field $\mathbf{B} = \nabla \times \mathbf{A}$. There are two contributions: one when ∇ acts on the $1/r$ term, and another when ∇ acts on the r in the argument of $\dot{\mathbf{p}}$. These give, respectively,

$$\mathbf{B} \approx -\frac{\mu_0}{4\pi r^2} \hat{\mathbf{x}} \times \dot{\mathbf{p}}(t - r/c) - \frac{\mu_0}{4\pi r c} \hat{\mathbf{x}} \times \ddot{\mathbf{p}}(t - r/c)$$

where we've used the fact that $\nabla r = \hat{\mathbf{x}}$. Which of these two terms is bigger? As we get further from the source, we would expect that the second, $1/r$, term dominates over the first, $1/r^2$ term. We can make this more precise. Suppose that the source is oscillating at some frequency ω , so that $\ddot{\mathbf{p}} \sim \omega \dot{\mathbf{p}}$. We expect that it will make waves at the characteristic wavelength $\lambda = c/\omega$. Then, as long we're at distances $r \gg \lambda$, the second term dominates and we have

$$\mathbf{B}(t, \mathbf{x}) \approx -\frac{\mu_0}{4\pi r c} \hat{\mathbf{x}} \times \ddot{\mathbf{p}}(t - r/c) \quad (6.15)$$

The region $r \gg \lambda$ is called the *far-field zone* or, sometimes, the *radiation zone*. We've now made two successive approximations, valid if we have a hierarchy of scales in our problem: $r \gg \lambda \gg d$.

To get the corresponding electric field, it's actually simpler to use the Maxwell equation $\dot{\mathbf{E}} = c^2 \nabla \times \mathbf{B}$. Again, if we care only about large distances, $r \gg \lambda$, the curl of \mathbf{B} is dominated by ∇ acting on the argument of $\ddot{\mathbf{p}}$. We get

$$\begin{aligned} \nabla \times \mathbf{B} &\approx \frac{\mu_0}{4\pi r c^2} \hat{\mathbf{x}} \times (\hat{\mathbf{x}} \times \ddot{\mathbf{p}}(t - r/c)) \\ \Rightarrow \quad \mathbf{E} &\approx \frac{\mu_0}{4\pi r} \hat{\mathbf{x}} \times (\hat{\mathbf{x}} \times \ddot{\mathbf{p}}(t - r/c)) \end{aligned} \quad (6.16)$$

Notice that the electric and magnetic field are related in the same way that we saw for plane waves, namely

$$\mathbf{E} = -c \hat{\mathbf{x}} \times \mathbf{B}$$

although, now, this only holds when we're suitably far from the source, $r \gg \lambda$. What's happening here is that oscillating dipole is emitting spherical waves. At radius $r \gg \lambda$ these can be thought of as essentially planar.

Notice, also, that the electric field is dropping off slowly as $1/r$. This, of course, is even slower than the usual Coulomb force fall-off.

6.2.2 Power Radiated: Larmor Formula

We can look at the power radiated away by the source. This is computed by the Poynting vector which we first met in Section 4.4. It is given by

$$\mathbf{S} = \frac{1}{\mu_0} \mathbf{E} \times \mathbf{B} = \frac{c}{\mu_0} |\mathbf{B}|^2 \hat{\mathbf{x}} = \frac{\mu_0}{16\pi^2 r^2 c} |\hat{\mathbf{x}} \times \ddot{\mathbf{p}}|^2 \hat{\mathbf{x}}$$

The fact that \mathbf{S} lies in the direction $\hat{\mathbf{x}}$ means that the power is emitted radially. The fact that it drops off as $1/r^2$ follows from the conservation of energy. It means that the total energy flux, computed by integrating \mathbf{S} over a large surface, is constant, independent of r .

Although the radiation is radial, it is not uniform. Suppose that the dipole oscillates in the $\hat{\mathbf{z}}$ direction. Then we have

$$\mathbf{S} = \frac{\mu_0}{16\pi^2 r^2 c} |\ddot{\mathbf{p}}|^2 \sin^2 \theta \hat{\mathbf{z}} \quad (6.17)$$

where θ is the angle between $\hat{\mathbf{x}}$ and the z -axis. The emitted power is largest in the plane perpendicular to the dipole. A sketch of this is shown in the figure.

A device which converts currents into electromagnetic waves (typically in the radio spectrum) is called an *antenna*. We see that it's not possible to create a dipole antenna which emits radiation uniformly. There's actually some nice topology underlying this observation. Look at a sphere which surrounds the antenna at large distance. The radiation is emitted radially, which means that the magnetic field \mathbf{B} lies tangent to the sphere. But there's an intuitive result in topology called the *hairy ball theorem* which says that you can't smoothly comb the hair on a sphere. Or, more precisely, there does not exist a nowhere vanishing vector field on a sphere. Instead, any vector field like \mathbf{B} must vanish at two or more points. In this present context, that ensures that \mathbf{S} too vanishes at two points.

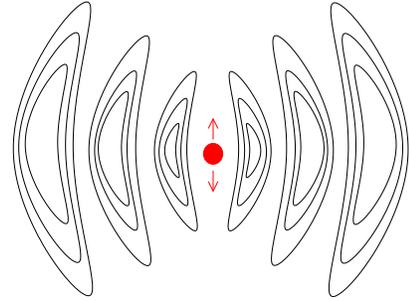


Figure 53:

The total radiated power, \mathcal{P} , is computed by integrating over a sphere,

$$\mathcal{P} = \int_{S^2} d^2\mathbf{r} \cdot \mathbf{S} = \frac{\mu_0}{16\pi^2 c} |\ddot{\mathbf{p}}|^2 \int_0^{2\pi} d\phi \int_0^\pi d\theta \sin^3 \theta$$

where one of the factors of $\sin \theta$ comes from the Jacobian. The integral is easily performed, to get

$$\mathcal{P} = \frac{\mu_0}{6\pi c} |\ddot{\mathbf{p}}|^2 \quad (6.18)$$

Finally, the dipole term $\ddot{\mathbf{p}}$ is still time dependent. It's common practice to compute the time averaged power. The most common example is when the dipole oscillates with frequency ω , so that $|\ddot{\mathbf{p}}|^2 \sim \cos^2(\omega t)$. (Recall that we're only allowed to work with complex expressions when we have linear equations). Then, integrating over a period, $T = 2\pi/\omega$, just gives an extra factor of 1/2.

Let's look at a simple example. Take a particle of charge Q , oscillating in the \hat{z} direction with frequency ω and amplitude d . Then we have $\mathbf{p} = p\hat{z}e^{i\omega t}$ with the dipole moment $p = Qd$. Similarly, $\ddot{\mathbf{p}} = -\omega^2 p\hat{z}e^{i\omega t}$. The end result for the time averaged power $\bar{\mathcal{P}}$ is

$$\bar{\mathcal{P}} = \frac{\mu_0 p^2 \omega^4}{12\pi c} \quad (6.19)$$

This is the *Larmor formula* for the time-averaged power radiated by an oscillating charge. The formula is often described in terms of the acceleration, $a = d\omega^2$. Then it reads

$$\bar{\mathcal{P}} = \frac{Q^2 a^2}{12\pi\epsilon_0 c^3} \quad (6.20)$$

where we've also swapped the μ_0 in the numerator for $\epsilon_0 c^2$ in the denominator.

6.2.3 An Application: Instability of Classical Matter

The popular picture of an atom consists of a bunch of electrons orbiting a nucleus, like planets around a star. But this isn't what an atom looks like. Let's see why.

We'll consider a Hydrogen atom, with an electron orbiting around a proton, fixed at the origin. (The two really orbit each other around their common centre of mass, but the mass of the electron is $m_e \approx 9 \times 10^{-31} \text{ Kg}$, while the mass of the proton is about 1800 bigger, so this is a good approximation). The equation of motion for the electron is

$$m_e \ddot{\mathbf{r}} = -\frac{e^2}{4\pi\epsilon_0} \frac{\hat{\mathbf{r}}}{r^2}$$

The dipole moment of the atom is $\mathbf{p} = e\mathbf{r}$ so the equation of motion tells us $\ddot{\mathbf{p}}$. Plugging this into (6.18), we can get an expression for the amount of energy emitted by the electron,

$$\mathcal{P} = \frac{\mu_0}{6\pi c} \left(\frac{e^3}{4\pi\epsilon_0 m_e r^2} \right)^2$$

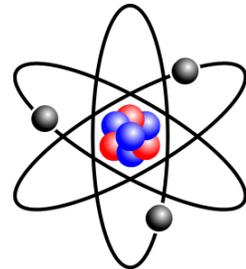


Figure 54: This is not what an atom looks like.

As the electron emits radiation, it loses energy and must, therefore, spiral towards the nucleus. We know from classical mechanics that the energy of the orbit depends on its eccentricity. For simplicity, let's assume that the orbit is circular with energy

$$E = -\frac{e^2}{4\pi\epsilon_0} \frac{1}{2r}$$

Then we can equate the change in energy with the emitted power to get

$$\dot{E} = \frac{e^2}{8\pi\epsilon_0 r^2} \dot{r} = -\mathcal{P} = -\frac{\mu_0}{6\pi c} \left(\frac{e^3}{4\pi\epsilon_0 m_e r^2} \right)^2$$

which gives us an equation that tells us how the radius of the orbit changes,

$$\dot{r} = -\frac{\mu_0 e^4}{12\pi^2 c \epsilon_0 m_e^2 r^2}$$

Suppose that we start at some time, $t = 0$, with a classical orbit with radius r_0 . Then we can calculate how long it takes for the electron to spiral down to the origin at $r = 0$. It is

$$T = \int_0^T dt = \int_{r_0}^0 \frac{1}{\dot{r}} dr = \frac{4\pi^2 c \epsilon_0 m_e^2 r_0^3}{\mu_0 e^4}$$

Now let's plug in some small numbers. We can take the size of the atom to be $r_0 \approx 5 \times 10^{-11} m$. (This is roughly the Bohr radius that can be derived theoretically using quantum mechanics). Then we find that the lifetime of the hydrogen atom is

$$T \approx 10^{-11} \text{ s}$$

That's a little on the small size. The Universe is 14 billion years old and hydrogen atoms seem in no danger of decaying.

Of course, what we're learning here is something dramatic: the whole framework of classical physics breaks down when we look at the atomic scale and has to be replaced with quantum mechanics. And, although we talk about electron orbits in quantum mechanics, they are very different objects than the classical orbits drawn in the picture. In particular, an electron in the ground state of the hydrogen atom emits no radiation. (Electrons in higher states do emit radiation with some probability, ultimately decaying down to the ground state).

6.2.4 Magnetic Dipole and Electric Quadrupole Radiation

The electric dipole approximation to radiation is sufficient for most applications. Obvious exceptions are when the dipole \mathbf{p} vanishes or, for some reason, doesn't change in time. For completeness, we describe here the leading order corrections to the electric dipole approximations.

The Taylor expansion of the retarded potential was given in (6.13) and (6.14). Putting them together, we get

$$\begin{aligned} A_\mu(\mathbf{x}, t) &= \frac{\mu_0}{4\pi} \int_A d^3x' \frac{J_\mu(\mathbf{x}', t_{\text{ret}})}{|\mathbf{x} - \mathbf{x}'|} \\ &= \frac{\mu_0}{4\pi r} \int_A d^3x' \left(J_\mu(\mathbf{x}', t - r/c) + \dot{J}_\mu(\mathbf{x}', t - r/c) \frac{\mathbf{x} \cdot \mathbf{x}'}{rc} \right) \left(1 + \frac{\mathbf{x} \cdot \mathbf{x}'}{r^2} \right) + \dots \end{aligned}$$

The first term is the electric dipole approximation that we discussed in above. We will refer to this as A_μ^{ED} . Corrections to this arise as two Taylor series. Ultimately we will only be interested in the far-field region. At far enough distance, the terms in the first bracket will always dominate the terms in the second bracket, which are suppressed by $1/r$. We therefore have

$$A_\mu(\mathbf{x}, t) \approx A_\mu^{\text{ED}}(\mathbf{x}, t) + \frac{\mu_0}{4\pi r^2 c} \int_A d^3x' (\mathbf{x} \cdot \mathbf{x}') \dot{J}_\mu(\mathbf{x}', t - r/c)$$

As in the electric dipole case, it's simplest if we focus on the vector potential

$$\mathbf{A}(\mathbf{x}, t) \approx \mathbf{A}^{\text{ED}}(\mathbf{x}, t) + \frac{\mu_0}{4\pi r^2 c} \int d^3x' (\mathbf{x} \cdot \mathbf{x}') \dot{\mathbf{J}}(\mathbf{x}', t - r/c) \quad (6.21)$$

The integral involves the kind of expression that we met first when we discussed magnetic dipoles in Section 3.3.2. We use the slightly odd expression,

$$\partial_j(J_j x_i x_k) = (\partial_j J_j) x_i x_k + J_i x_k + J_k x_i = -\dot{\rho} x_i x_k + J_i x_k + J_k x_i$$

Because \mathbf{J} in (6.21) is a function of \mathbf{x}' , we apply this identity to the $J_i x'_j$ terms in the expression. We drop the boundary term at infinity, remembering that we're actually dealing with \dot{J} rather than J , write the integral above as

$$\int d^3x' x_j x'_j \dot{J}_i = \frac{x_j}{2} \int d^3x' (x'_j \dot{J}_i - x'_i \dot{J}_j + \ddot{\rho} x'_i x'_j)$$

Then, using the appropriate vector product identity, we have

$$\int d^3x' (\mathbf{x} \cdot \mathbf{x}') \dot{\mathbf{J}} = \frac{1}{2} \mathbf{x} \times \int d^3x' \dot{\mathbf{J}} \times \mathbf{x}' + \frac{1}{2} \int d^3x' (\mathbf{x} \cdot \mathbf{x}') \mathbf{x}' \ddot{\rho}$$

Using this, we may write (6.21) as

$$\mathbf{A}(\mathbf{x}, t) \approx \mathbf{A}^{\text{ED}}(\mathbf{x}, t) + \mathbf{A}^{\text{MD}}(\mathbf{x}, t) + \mathbf{A}^{\text{EQ}}(\mathbf{x}, t)$$

where \mathbf{A}^{MD} is the *magnetic dipole* contribution and is given by

$$\mathbf{A}^{\text{MD}}(\mathbf{x}, t) = -\frac{\mu_0}{8\pi r^2 c} \mathbf{x} \times \int d^3x' \mathbf{x}' \times \dot{\mathbf{J}}(\mathbf{x}', t - r/c) \quad (6.22)$$

and \mathbf{A}^{EQ} is the *electric quadrupole* contribution and is given by

$$\mathbf{A}^{\text{EQ}}(\mathbf{x}, t) = \frac{\mu_0}{8\pi r^2 c} \int d^3x' (\mathbf{x} \cdot \mathbf{x}') \mathbf{x}' \ddot{\rho}(\mathbf{x}', t - r/c) \quad (6.23)$$

The names we have given to each of these contributions will become clearer as we look at their properties in more detail.

Magnetic Dipole Radiation

Recall that, for a general current distribution, the magnetic dipole \mathbf{m} is defined by

$$\mathbf{m} = \frac{1}{2} \int d^3x' \mathbf{x}' \times \mathbf{J}(\mathbf{x}')$$

The magnetic dipole contribution to radiation (6.22) can then be written as

$$\mathbf{A}^{\text{MD}}(\mathbf{x}, t) = -\frac{\mu_0}{4\pi r c} \hat{\mathbf{x}} \times \dot{\mathbf{m}}(t - r/c)$$

This means that varying loops of current will also emit radiation. Once again, the leading order contribution to the magnetic field, $\mathbf{B} = \nabla \times \mathbf{A}$, arises when the curl hits the argument of \mathbf{m} . We have

$$\mathbf{B}^{\text{MD}}(\mathbf{x}, t) \approx \frac{\mu_0}{4\pi r c^2} \hat{\mathbf{x}} \times (\hat{\mathbf{x}} \times \ddot{\mathbf{m}}(t - r/c))$$

Using the Maxwell equation $\dot{\mathbf{E}}^{\text{MD}} = c^2 \nabla \times \mathbf{B}^{\text{MD}}$ to compute the electric field, we have

$$\mathbf{E}^{\text{MD}}(\mathbf{x}, t) \approx \frac{\mu_0}{4\pi r c} \hat{\mathbf{x}} \times \ddot{\mathbf{m}}(t - r/c)$$

The end result is very similar to the expression for \mathbf{B} and \mathbf{E} that we saw in (6.15) and (6.16) for the electric dipole radiation. This means that the radiated power has the same angular form, with the Poynting vector now given by

$$\mathbf{S}^{\text{MD}} = \frac{\mu_0}{16\pi^2 r^2 c^3} |\ddot{\mathbf{m}}|^2 \sin^2 \theta \hat{\mathbf{z}} \quad (6.24)$$

Integrating over all space gives us the power emitted,

$$\mathcal{P}^{\text{MD}} = \frac{\mu_0}{6\pi c^3} |\ddot{\mathbf{m}}|^2 \quad (6.25)$$

This takes the same form as the electric dipole result (6.18), but with the electric dipole replaced by the magnetic dipole. Notice, however, that for non-relativistic particles, the magnetic dipole radiation is substantially smaller than the electric dipole contribution. For a particle of charge Q , oscillating a distance d with frequency ω , we have $p \sim Qd$ and $m \sim Qd^2\omega$. This means that the ratio of radiated powers is

$$\frac{\mathcal{P}^{\text{MD}}}{\mathcal{P}^{\text{ED}}} \sim \frac{d^2\omega^2}{c^2} \sim \frac{v^2}{c^2}$$

where v is the speed of the particle.

Electric Quadrupole Radiation

The electric quadrupole tensor \mathbb{Q}_{ij} arises as the $1/r^4$ term in the expansion of the electric field for a general, static charge distribution. It is defined by

$$\mathbb{Q}_{ij} = \int d^3x' \rho(\mathbf{x}') (3x'_i x'_j - \delta_{ij} x'^2)$$

This is not quite of the right form to account for the contribution to the potential (6.23). Instead, we have

$$A_i^{\text{EQ}}(\mathbf{x}, t) = -\frac{\mu_0}{24\pi r^2 c} \left(x_j \ddot{\mathbb{Q}}_{ij}(t - r/c) + x_i \int d^3x' x'^2 \ddot{\rho}(x', t - r/c) \right)$$

The second term looks like a mess, but it doesn't do anything. This is because it's radial and so vanishes when we take the curl to compute the magnetic field. Neither does it contribute to the electric field which, in our case, we will again determine from the Maxwell equation. This means we are entitled to write

$$\mathbf{A}^{\text{EQ}}(\mathbf{x}, t) = -\frac{\mu_0}{24\pi r^2 c} \mathbf{x} \cdot \ddot{\mathbb{Q}}(t - r/c)$$

where $(\mathbf{x} \cdot \mathbb{Q})_i = x_j \mathbb{Q}_{ij}$. Correspondingly, the magnetic and electric fields at large distance are

$$\begin{aligned} \mathbf{B}^{\text{EQ}}(\mathbf{x}, t) &\approx \frac{\mu_0}{24\pi r c^2} \hat{\mathbf{x}} \times (\hat{\mathbf{x}} \cdot \ddot{\mathbb{Q}}) \\ \mathbf{E}^{\text{EQ}}(\mathbf{x}, t) &\approx \frac{\mu_0}{24\pi r c} \left((\hat{\mathbf{x}} \cdot \ddot{\mathbb{Q}} \cdot \hat{\mathbf{x}}) \hat{\mathbf{x}} - (\hat{\mathbf{x}} \cdot \ddot{\mathbb{Q}}) \right) \end{aligned}$$

We may again compute the Poynting vector and radiated power. The details depend on the exact structure of \mathbb{Q} , but the angular dependence of the radiation is now different from that seen in the dipole cases.

Finally, you may wonder about the cross terms between the ED, MD and EQ components of the field strengths when computing the quadratic Poynting vector. It turns out that, courtesy of their different spatial structures, these cross-term vanish when computing the total integrated power.

6.2.5 An Application: Pulsars

Pulsars are lighthouses in the sky, spinning neutron stars continuously beaming out radiation which sweeps past our line of sight once every rotation. They have been observed with periods between 10^{-3} seconds and 8 seconds.

Neutron stars typically carry a very large magnetic field. This arises from the parent star which, as it collapses, reduces in size by a factor of about 10^5 . This squeezes the magnetic flux lines, which gets multiplied by a factor of 10^{10} . The resulting magnetic field is typically around 10^8 Tesla, but can be as high as 10^{11} Tesla. For comparison, the highest magnetic field that we have succeeded in creating in a laboratory is a paltry 100 Tesla or so.

The simplest model of a pulsar has the resulting magnetic dipole moment \mathbf{m} of the neutron star misaligned with the angular velocity. This resulting magnetic dipole radiation creates the desired lighthouse effect. Consider the set-up shown in the picture. We take the pulsar to rotate about the z -axis with frequency Ω . The magnetic moment sits at an angle α relative to the z -axis, so rotates as

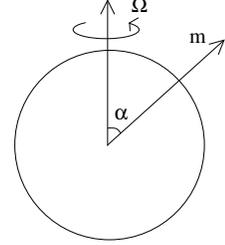


Figure 55:

$$\mathbf{m} = m_0 (\sin(\alpha) \cos(\Omega t) \hat{\mathbf{x}} + \sin(\alpha) \sin(\Omega t) \hat{\mathbf{y}} + \cos \alpha \hat{\mathbf{z}})$$

The power emitted (6.25) is then

$$\mathcal{P} = \frac{\mu_0}{6\pi c} m_0^2 \Omega^2 \sin^2 \alpha$$

At the surface of the neutron star, it's reasonable to assume that the magnetic field is given by the dipole moment. In Section 3.3, we computed the magnetic field due to a dipole moment: it is

$$\mathbf{B}(\mathbf{r}) = \frac{\mu_0}{4\pi} \left(\frac{3(\mathbf{m} \cdot \hat{\mathbf{r}}) \hat{\mathbf{r}} - \mathbf{m}}{R^3} \right)$$

where R is the radius of the star. This means that $B_{\max} = \mu_0 m_0 / 2\pi R^3$ and the power emitted is

$$\mathcal{P} = \frac{2\pi R^6 B_{\max}^2}{3c\mu_0} \Omega^2 \sin^2 \alpha \quad (6.26)$$

Because the pulsar is emitting radiation, it must lose energy. And this means it slows down. The rotational energy of a the pulsar is given by

$$E = \frac{1}{2} I \Omega^2$$

where $I = \frac{2}{5} M R^2$ is the moment of inertia of a sphere of mass M and radius R . Equating the power emitted with the loss of rotational kinetic energy gives

$$\mathcal{P} = -\dot{E} = -I\Omega\dot{\Omega} \quad (6.27)$$

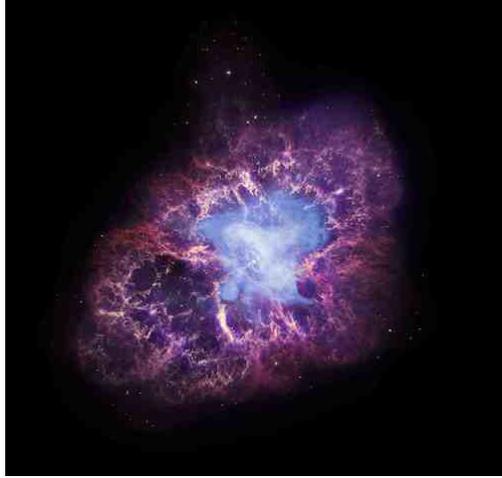


Figure 56: A composite image of the Crab Nebula, taken by the Hubble, Chandra and Spitzer space telescopes.

Let's put some big numbers into these equations. In 1054, Chinese astronomers saw a new star appear in the sky. 6500 light years away, a star had gone supernova. It left behind a pulsar which, today, emits large quantities of radiation, illuminating the part of the sky we call the Crab nebula. This is shown in the picture.

The Crab pulsar has mass $M \approx 1.4M_{\text{Sun}} \approx 3 \times 10^{30} \text{ Kg}$ and radius $R \approx 15 \text{ km}$. It spins about 30 times a second, so $\Omega \approx 60\pi \text{ s}^{-1}$. It's also seen to be slowing down with $\dot{\Omega} = -2 \times 10^{-9} \text{ s}^{-2}$. From this information alone, we can calculate that it loses energy at a rate of $\dot{E} = I\Omega\dot{\Omega} \approx -10^{32} \text{ Js}^{-1}$. That's a whopping amount of energy to be losing every second. In fact, it's enough energy to light up the entire Crab nebula. Which, of course, it has to be! Moreover, we can use (6.26) and (6.27) to estimate the magnetic field on the surface of the pulsar. Plugging in the numbers give $B_{\text{max}} \sin \alpha \approx 10^8 \text{ Tesla}$.

6.3 Scattering

In this short section, we describe the application of our radiation formulae to the phenomenon of *scattering*. Here's the set-up: an electromagnetic wave comes in and hits a particle. In response, the particle oscillates and, in doing so, radiates. This new radiation moves out in different directions from the incoming wave. This is the way that light is scattered.

6.3.1 Thomson Scattering

We start by considering free, charged particles where the process is known as Thomson

scattering. The particles respond to an electric field by accelerating, as dictated by Newton's law

$$m\ddot{\mathbf{x}} = q\mathbf{E}$$

The incoming radiation takes the form $\mathbf{E} = \mathbf{E}_0 e^{i(\mathbf{k}\cdot\mathbf{r} - \omega t)}$. To solve for the motion of the particle, we're going to assume that it doesn't move very far from its central position, which we can take to be the origin $\mathbf{r} = 0$. Here, "not very far" means small compared to the wavelength of the electric field. In this case, we can replace the electric field by $\mathbf{E} \approx \mathbf{E}_0 e^{-i\omega t}$, and the particle undergoes simple harmonic motion

$$\mathbf{x}(t) = -\frac{q\mathbf{E}_0}{m\omega^2} \sin(\omega t)$$

We should now check that the motion of the particle is indeed small compared to the wavelength of light. The maximum distance that the particle gets is $x_{\max} = qE_0/m\omega^2$, so our analysis will only be valid if we satisfy

$$\frac{qE_0}{m\omega^2} \ll \frac{c}{\omega} \quad \Rightarrow \quad \frac{qE_0}{m\omega c} \ll 1 \quad (6.28)$$

This requirement has a happy corollary, since it also ensures that the maximum speed of the particle $v_{\max} = qE_0/m\omega \ll c$, so the particle motion is non-relativistic. This means that we can use the dipole approximation to radiation that we developed in the previous section. We computed the time-averaged radiated power in (6.20): it is given by

$$\bar{\mathcal{P}}_{\text{radiated}} = \frac{\mu_0 q^4 E_0^2}{12\pi m^2 c}$$

It's often useful to compare the strength of the emitted radiation to that of the incoming radiation. The relevant quantity to describe the incoming radiation is the time-averaged magnitude of the Poynting vector. Recall from Section 4.4 that the Poynting vector for a wave with wavevector \mathbf{k} is

$$\mathbf{S} = \frac{1}{\mu_0} \mathbf{E} \times \mathbf{B} = \frac{cE_0^2}{\mu_0} \hat{\mathbf{k}} \sin^2(\mathbf{k} \cdot \mathbf{x} - \omega t)$$

Taking the time average over a single period, $T = 2\pi/\omega$, gives us the average energy flux of the incoming radiation,

$$\bar{S}_{\text{incident}} = \frac{cE_0^2}{2\mu_0}$$

with the factor of two coming from the averaging. The ratio of the outgoing to incoming powers is called the *cross-section* for scattering. It is given by

$$\sigma = \frac{\bar{P}_{\text{radiated}}}{\bar{S}_{\text{incident}}} = \frac{\mu_0^2 q^4}{6\pi m^2 c^2}$$

The cross-section has the dimensions of area. To highlight this, it's useful to write it as

$$\sigma = \frac{8\pi}{3} r_q^2 \tag{6.29}$$

where the length scale r_q is known as the *classical radius* of the particle and is given by

$$\frac{q^2}{4\pi\epsilon_0 r_q} = mc^2$$

This equation tells us how to think of r_q . Up to some numerical factors, it equates the Coulomb energy of a particle in a ball of size r_q with its relativistic rest mass. Ultimately, this is not the right way to think of the size of point particles. (The right way involves quantum mechanics). But it is a useful concept in the classical world. For the electron, $r_e \approx 2.8 \times 10^{-15} \text{ m}$.

The Thomson cross-section (6.29) is slightly smaller than the (classical) geometric cross-section of the particle (which would be the area of the disc, $4\pi r_q^2$). For us, the most important point is that the cross-section does not depend on the frequency of the incident light. It means that all wavelengths of light are scattered equally by free, charged particles, at least within the regime of validity (6.28). For electrons, the Thomson cross-section is $\sigma \approx 6 \times 10^{-30} \text{ m}^2$.

6.3.2 Rayleigh Scattering

Rayleigh scattering describes the scattering of light off a neutral atom or molecule. Unlike in the case of Thomson scattering, the centre of mass of the atom does not accelerate. Instead, as we will see in Section 7.1.1, the atom undergoes polarisation

$$\mathbf{p} = \alpha \mathbf{E}$$

We will present a simple atomic model to compute the proportionality constant in Section 7.5.1, where we will show that it takes the form (7.29),

$$\alpha = \frac{q^2/m}{-\omega^2 + \omega_0^2 - i\gamma\omega}$$

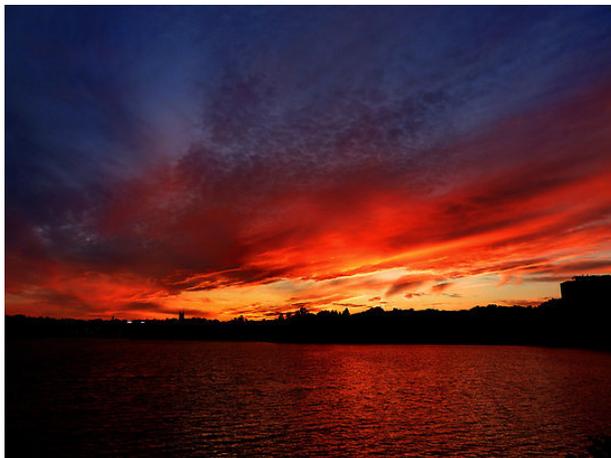


Figure 57: Now you know why.

Here ω_0 is the natural oscillation frequency of the atom while ω is the frequency of incoming light. For many cases of interest (such as visible light scattering off molecules in the atmosphere), we have $\omega_0 \gg \omega$, and we can approximate α as a constant,

$$\alpha \approx \frac{q^2}{\omega_0^2 m}$$

We can now compute the time-average power radiated in this case. It's best to use the version of Larmor's formula involving the electric dipole (6.19), since we can just substitute in the results above. We have

$$\bar{\mathcal{P}}_{\text{radiated}} = \frac{\mu_0 \alpha^2 E_0^2 \omega^4}{12\pi c}$$

In this case, the cross-section for Rayleigh scattering is given by

$$\sigma = \frac{\bar{\mathcal{P}}_{\text{radiated}}}{\mathcal{S}_{\text{incident}}} = \frac{\mu_0^2 q^4}{6\pi m^2 c^2} \left(\frac{\omega}{\omega_0}\right)^4 = \frac{8\pi r_q^2}{3} \left(\frac{\omega}{\omega_0}\right)^4$$

We see that the cross-section now has more structure. It increases for high frequencies, $\sigma \sim \omega^4$ or, equivalently, for short wavelengths $\sigma \sim 1/\lambda^4$. This is important. The most famous example is the colour of the sky. Nitrogen and oxygen in the atmosphere scatter short-wavelength blue light more than the long-wavelength red light. This means that the blue light from the Sun gets scattered many times and so appears to come from all regions of the sky. In contrast, the longer wavelength red and yellow light gets scattered less, which is why the Sun appears to be yellow. (In the absence of an atmosphere, the light from the Sun would be more or less white). This effect is particularly apparent at

sunset, when the light from the Sun passes through a much larger slice of atmosphere and, correspondingly, much more of the blue light is scattered, leaving behind only red.

6.4 Radiation From a Single Particle

In the previous section, we have developed the multipole expansion for radiation emitted from a source. We needed to invoke a couple of approximations. First, we assumed that we were far from the source. Second, we assumed that the motion of charges and currents within the source was non-relativistic.

In this section, we're going to develop a formalism which does not rely on these approximations. We will determine the field generated by a particle with charge q , moving on an arbitrary trajectory $\mathbf{r}(t)$, with velocity $\mathbf{v}(t)$ and acceleration $\mathbf{a}(t)$. It won't matter how far we are from the particle; it won't matter how fast the particle is moving. The particle has charge density

$$\rho(\mathbf{x}, t) = q\delta^3(\mathbf{x} - \mathbf{r}(t)) \quad (6.30)$$

and current

$$\mathbf{J}(\mathbf{x}, t) = q\mathbf{v}(t)\delta^3(\mathbf{x} - \mathbf{r}(t)) \quad (6.31)$$

Our goal is find the general solution to the Maxwell equations by substituting these expressions into the solution (6.7) for the retarded potential,

$$A_\mu(\mathbf{x}, t) = \frac{\mu_0}{4\pi} \int d^3x' \frac{J_\mu(\mathbf{x}', t_{\text{ret}})}{|\mathbf{x} - \mathbf{x}'|} \quad (6.32)$$

The result is known as Liénard-Wierchert potentials.

6.4.1 Liénard-Wierchert Potentials

If we simply plug (6.30) into the expression for the retarded electric potential (6.32), we get

$$\phi(\mathbf{x}, t) = \frac{q}{4\pi\epsilon_0} \int d^3x' \frac{1}{|\mathbf{x} - \mathbf{x}'|} \delta^3(\mathbf{x}' - \mathbf{r}(t_{\text{ret}}))$$

Here we're denoting the position of the particle as $\mathbf{r}(t)$, while we're interested in the value of the electric potential at some different point \mathbf{x} which does not lie on the trajectory $\mathbf{r}(t)$. We can use the delta-function to do the spatial integral, but it's a little cumbersome because the \mathbf{x}' appears in the argument of the delta-function both in the

obvious place, and also in $t_{\text{ret}} = t - |\mathbf{x} - \mathbf{x}'|/c$. It turns out to be useful to shift this awkwardness into a slightly different delta-function over time. We write,

$$\begin{aligned}\phi(\mathbf{x}, t) &= \frac{q}{4\pi\epsilon_0} \int dt' \int d^3x' \frac{1}{|\mathbf{x} - \mathbf{x}'|} \delta^3(\mathbf{x}' - \mathbf{r}(t')) \delta(t' - t_{\text{ret}}) \\ &= \frac{q}{4\pi\epsilon_0} \int dt' \frac{1}{|\mathbf{x} - \mathbf{r}(t')|} \delta(t - t' - |\mathbf{x} - \mathbf{r}(t')|/c)\end{aligned}\quad (6.33)$$

We still have the same issue in doing the $\int dt'$ integral, with t' appearing in two places in the argument. But it's more straightforward to see how to deal with it. We introduce the separation vector

$$\mathbf{R}(t) = \mathbf{x} - \mathbf{r}(t)$$

Then, if we define $f(t') = t' + R(t')/c$, we can write

$$\begin{aligned}\phi(\mathbf{x}, t) &= \frac{q}{4\pi\epsilon_0} \int dt' \frac{1}{R(t')} \delta(t - f(t')) \\ &= \frac{q}{4\pi\epsilon_0} \int df \frac{dt'}{df} \frac{1}{R(t')} \delta(t - f(t')) \\ &= \frac{q}{4\pi\epsilon_0} \left[\frac{dt'}{df} \frac{1}{R(t')} \right]_{f(t')=t}\end{aligned}$$

A quick calculation gives

$$\frac{df}{dt'} = 1 - \frac{\hat{\mathbf{R}}(t') \cdot \mathbf{v}(t')}{c}$$

with $\mathbf{v}(t) = \dot{\mathbf{r}}(t) = -\dot{\mathbf{R}}(t)$. This leaves us with our final expression for the scalar potential

$$\phi(\mathbf{x}, t) = \frac{q}{4\pi\epsilon_0} \left[\frac{c}{c - \hat{\mathbf{R}}(t') \cdot \mathbf{v}(t')} \frac{1}{R(t')} \right]_{\text{ret}}\quad (6.34)$$

Exactly the same set of manipulations will give us a similar expression for the vector potential,

$$\mathbf{A}(\mathbf{x}, t) = \frac{q\mu_0}{4\pi} \left[\frac{c}{c - \hat{\mathbf{R}}(t') \cdot \mathbf{v}(t')} \frac{\mathbf{v}(t')}{R(t')} \right]_{\text{ret}}\quad (6.35)$$

Equations (6.34) and (6.35) are the *Liénard-Wiechert potentials*. In both expressions “ret” stands for “retarded” and means that they should be evaluated at time t' determined by the requirement that

$$t' + R(t')/c = t\quad (6.36)$$

This equation has an intuitive explanation. If you trace back light-sheets from the point \mathbf{x} , they intersect the trajectory of the particle at time t' , as shown in the figure. The Liénard-Wierchert potentials are telling us that the field at point \mathbf{x} is determined by what the particle was doing at this time t' .

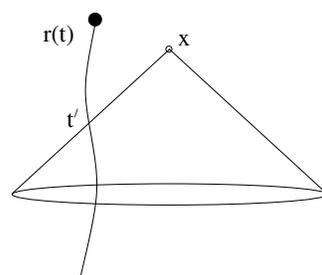


Figure 58:

6.4.2 A Simple Example: A Particle Moving with Constant Velocity

The Liénard-Wierchert potentials (6.34) and (6.35) have the same basic structure that we for the Coulomb law in electrostatics and the Biot-Savart law in magnetostatics. The difference lies in the need to evaluate the potentials at time t' . But there is also the extra factor $1/(1 - \hat{\mathbf{R}} \cdot \mathbf{v}/c)$. To get a feel for this, let's look at a simple example. We'll take a particle which moves at constant speed in the $\hat{\mathbf{z}}$ direction, so that

$$\mathbf{r}(t) = vt\hat{\mathbf{z}} \quad \Rightarrow \quad \mathbf{v}(t) = v\hat{\mathbf{z}}$$

To simplify life even further, we'll compute the potentials at a point in the $z = 0$ plane, so that $\mathbf{x} = (x, y, 0)$. We'll ask how the fields change as the particle passes through. The equation (6.36) to determine the retarded time becomes

$$t' + \sqrt{x^2 + y^2 + v^2 t'^2}/c = t$$

Squaring this equation (after first making the right-hand side $t - t'$) gives us a quadratic in t' ,

$$t'^2 - 2\gamma^2 t t' + \gamma^2 (t^2 - r^2/c^2) = 0$$

where we see the factor $\gamma = (1 - v^2/c^2)^{-1/2}$, familiar from special relativity naturally emerging. The quadratic has two roots. We're interested in the one with the minus sign, corresponding to the retarded time. This is

$$t' = \gamma^2 t - \frac{\gamma}{c} \sqrt{v^2 t^2 + r^2/\gamma^2} \quad (6.37)$$

We now need to deal with the various factors in the numerator of the Liénard-Wierchert potential (6.34). Pleasingly, they combine together nicely. We have $R(t') = c(t - t')$. Meanwhile, $\mathbf{R}(t') \cdot \mathbf{v}(t') = (\mathbf{x} - \mathbf{r}(t')) \cdot \mathbf{v} = -\mathbf{r}(t') \cdot \mathbf{v} = -v^2 t'$ since we've taken \mathbf{x} to lie

perpendicular to \mathbf{v} . Put together, this gives us

$$\begin{aligned}\phi(\mathbf{x}, t) &= \frac{q}{4\pi\epsilon_0} \frac{1}{[1 + v^2 t'/c(t - t')]} \frac{1}{c(t - t')} \\ &= \frac{q}{4\pi\epsilon_0} \frac{1}{c(t - t') + v^2 t'} \\ &= \frac{1}{4\pi\epsilon_0} \frac{1}{c(t - t'/\gamma^2)}\end{aligned}$$

But, using our solution (6.37), this becomes

$$\phi(\mathbf{x}, t) = \frac{q}{4\pi\epsilon_0} \frac{1}{[v^2 t^2 + (x^2 + y^2)/\gamma^2]}$$

Similarly, the vector potential is

$$\mathbf{A}(\mathbf{x}, t) = \frac{q\mu_0}{4\pi} \frac{\mathbf{v}}{[v^2 t^2 + (x^2 + y^2)/\gamma^2]}$$

How should we interpret these results? The distance from the particle to the point \mathbf{x} is $r^2 = x^2 + y^2 + v^2 t^2$. The potentials look very close to those due to a particle a distance r away, but with one difference: there is a contraction in the x and y directions. Of course, we know very well what this means: it is the usual Lorentz contraction in special relativity.

In fact, we previously derived the expression for the electric and magnetic field of a moving particle in Section 5.3.4, simply by acting with a Lorentz boost on the static fields. The calculation here was somewhat more involved, but it didn't assume any relativity. Instead, the Lorentz contraction follows only by solving the Maxwell equations. Historically, this kind of calculation is how Lorentz first encountered his contractions.

6.4.3 Computing the Electric and Magnetic Fields

We now compute the electric and magnetic fields due to a particle undergoing arbitrary motion. In principle this is straightforward: we just need to take our equations (6.34) and (6.35)

$$\begin{aligned}\phi(\mathbf{x}, t) &= \frac{q}{4\pi\epsilon_0} \left[\frac{c}{c - \hat{\mathbf{R}}(t') \cdot \mathbf{v}(t')} \frac{1}{R(t')} \right]_{\text{ret}} \\ \mathbf{A}(\mathbf{x}, t) &= \frac{q\mu_0}{4\pi} \left[\frac{c}{c - \hat{\mathbf{R}}(t') \cdot \mathbf{v}(t')} \frac{\mathbf{v}(t')}{R(t')} \right]_{\text{ret}}\end{aligned}$$

where $\mathbf{R}(t') = \mathbf{x} - \mathbf{r}(t')$. We then plug these into the standard expressions for the electric field $\mathbf{E} = -\nabla\phi - \partial\mathbf{A}/\partial t$ and the magnetic field $\mathbf{B} = \nabla \times \mathbf{A}$. However, in

practice, this is a little fiddly. It's because the terms in these equations are evaluated at the retarded time t' determined by the equation $t' + R(t')/c = t$. This means that when we differentiate (either by $\partial/\partial t$ or by ∇), the retarded time also changes and so gives a contribution. It turns out that it's actually simpler to return to our earlier expression (6.33),

$$\phi(\mathbf{x}, t) = \frac{q}{4\pi\epsilon_0} \int dt' \frac{1}{R(t')} \delta(t - t' - R(t')/c)$$

and a similar expression for the vector potential,

$$\mathbf{A}(\mathbf{x}, t) = \frac{q\mu_0}{4\pi} \int dt' \frac{\mathbf{v}(t')}{R(t')} \delta(t - t' - R(t')/c) \quad (6.38)$$

This will turn out to be marginally easier to deal with.

The Electric Field

We start with the electric field $\mathbf{E} = -\nabla\phi - \partial\mathbf{A}/\partial t$. We call the argument of the delta-function

$$s = t - t' - R(t')$$

We then have

$$\begin{aligned} \nabla\phi &= \frac{q}{4\pi\epsilon_0} \int dt' \left[-\frac{\nabla R}{R^2} \delta(s) - \frac{1}{R} \delta'(s) \frac{\nabla R}{c} \right] \\ &= \frac{q}{4\pi\epsilon_0} \int ds \left| \frac{\partial t'}{\partial s} \right| \left[-\frac{\nabla R}{R^2} \delta(s) - \frac{\nabla R}{Rc} \delta'(s) \right] \end{aligned} \quad (6.39)$$

The Jacobian factor from changing the integral variable is the given by

$$\frac{\partial s}{\partial t'} = -1 + \hat{\mathbf{R}}(t') \cdot \mathbf{v}(t')/c$$

This quantity will appear a lot in what follows, so we give it a new name. We define

$$\kappa = 1 - \hat{\mathbf{R}}(t') \cdot \mathbf{v}(t')/c$$

so that $\partial t'/\partial s = -1/\kappa$. Integrating the second term in (6.39) by parts, we can then write

$$\begin{aligned} \nabla\phi &= \frac{q}{4\pi\epsilon_0} \int ds \left[-\frac{\nabla R}{\kappa R^2} + \frac{d}{ds} \left(\frac{\nabla R}{\kappa R c} \right) \right] \delta(s) \\ &= \frac{q}{4\pi\epsilon_0} \int ds \left[-\frac{\nabla R}{\kappa R^2} - \frac{1}{\kappa} \frac{d}{dt'} \left(\frac{\nabla R}{\kappa R c} \right) \right] \delta(s) \end{aligned}$$

Meanwhile, the vector potential term gives

$$\frac{\partial \mathbf{A}}{\partial t} = \frac{q\mu_0}{4\pi} \int dt' \frac{\mathbf{v}}{R} \delta'(s) \frac{\partial s}{\partial t}$$

But $\partial s / \partial t = 1$. Moving forward, we have

$$\begin{aligned} \frac{\partial \mathbf{A}}{\partial t} &= \frac{q\mu_0}{4\pi} \int ds \left| \frac{\partial t'}{\partial s} \right| \frac{\mathbf{v}}{R} \delta'(s) \\ &= -\frac{q\mu_0}{4\pi} \int ds \left[\frac{d}{ds} \left(\frac{\mathbf{v}}{\kappa R} \right) \right] \delta(s) \\ &= \frac{q\mu_0}{4\pi} \int ds \frac{1}{\kappa} \left[\frac{d}{dt'} \left(\frac{\mathbf{v}}{\kappa R} \right) \right] \delta(s) \end{aligned}$$

Putting this together, we get

$$\begin{aligned} \mathbf{E} &= \frac{q}{4\pi\epsilon_0} \int ds \left[\frac{\nabla R}{\kappa R^2} + \frac{1}{\kappa c} \frac{d}{dt'} \left(\frac{\nabla R - \mathbf{v}/c}{\kappa R} \right) \right] \delta(s) \\ &= \frac{q}{4\pi\epsilon_0} \left[\frac{\hat{\mathbf{R}}}{\kappa R^2} + \frac{1}{\kappa c} \frac{d}{dt'} \left(\frac{\hat{\mathbf{R}} - \mathbf{v}/c}{\kappa R} \right) \right]_{\text{ret}} \end{aligned} \quad (6.40)$$

We're still left with some calculations to do. Specifically, we need to take the derivative d/dt' . This involves a couple of small steps. First,

$$\frac{d\hat{\mathbf{R}}}{dt'} = \frac{d}{dt'} \left(\frac{\mathbf{R}}{R} \right) = -\frac{\mathbf{v}}{R} + \frac{\mathbf{R}}{R^2} (\hat{\mathbf{R}} \cdot \mathbf{v}) = -\frac{1}{R} \left(\mathbf{v} - (\mathbf{v} \cdot \hat{\mathbf{R}}) \hat{\mathbf{R}} \right)$$

Also,

$$\frac{d}{dt'} (\kappa R) = \frac{d}{dt'} (R - \mathbf{R} \cdot \mathbf{v}/c) = -\mathbf{v} \cdot \hat{\mathbf{R}} + v^2/c - \mathbf{R} \cdot \mathbf{a}/c$$

Putting these together, we get

$$\frac{d}{dt'} \left(\frac{\hat{\mathbf{R}} - \mathbf{v}/c}{\kappa R} \right) = -\frac{1}{\kappa R^2} \left(\mathbf{v} - (\mathbf{v} \cdot \hat{\mathbf{R}}) \hat{\mathbf{R}} \right) - \frac{\mathbf{a}}{\kappa R c} + \frac{\hat{\mathbf{R}} - \mathbf{v}/c}{\kappa^2 R^2} \left(\mathbf{v} \cdot \hat{\mathbf{R}} - v^2/c + \mathbf{R} \cdot \mathbf{a}/c \right)$$

We write the $\mathbf{v} \cdot \hat{\mathbf{R}}$ terms as $\mathbf{v} \cdot \hat{\mathbf{R}} = c(1 - \kappa)$. Then, expanding this out, we find that a bunch of terms cancel, until we're left with

$$\begin{aligned} \frac{d}{dt'} \left(\frac{\hat{\mathbf{R}} - \mathbf{v}/c}{\kappa R} \right) &= -\frac{c\hat{\mathbf{R}}}{R^2} + \frac{c(\hat{\mathbf{R}} - \mathbf{v}/c)}{\kappa^2 R^2} (1 - v^2/c^2) + \frac{1}{\kappa^2 R c} \left[(\hat{\mathbf{R}} - \mathbf{v}/c) \hat{\mathbf{R}} \cdot \mathbf{a} - \kappa a \right] \\ &= -\frac{c\hat{\mathbf{R}}}{R^2} + \frac{c(\hat{\mathbf{R}} - \mathbf{v}/c)}{\gamma^2 \kappa^2 R^2} + \frac{\hat{\mathbf{R}} \times [(\hat{\mathbf{R}} - \mathbf{v}/c) \times \mathbf{a}]}{\kappa^2 R c} \end{aligned} \quad (6.41)$$

where we've introduced the usual γ factor from special relativity: $\gamma^2 = 1/(1 - v^2/c^2)$. Now we can plug this into (6.40) to find our ultimate expression for the electric field,

$$\mathbf{E}(\mathbf{x}, t) = \frac{q}{4\pi\epsilon_0} \left[\frac{\hat{\mathbf{R}} - \mathbf{v}/c}{\gamma^2 \kappa^3 R^2} + \frac{\hat{\mathbf{R}} \times [(\hat{\mathbf{R}} - \mathbf{v}/c) \times \mathbf{a}]}{\kappa^3 R c^2} \right]_{\text{ret}} \quad (6.42)$$

Since it's been a long journey, let's recall what everything in this expression means. The particle traces out a trajectory $\mathbf{r}(t)$, while we sit at some position \mathbf{x} which is where the electric field is evaluated. The vector $\mathbf{R}(t)$ is the difference: $\mathbf{R} = \mathbf{x} - \mathbf{r}$. The *ret* subscript means that we evaluate everything in the square brackets at time t' , determined by the condition $t' + R(t')/c = t$. Finally,

$$\kappa = 1 - \frac{\hat{\mathbf{R}} \cdot \mathbf{v}}{c} \quad \text{and} \quad \gamma^2 = \frac{1}{1 - v^2/c^2}$$

The electric field (6.42) has two terms.

- The first term drops off as $1/R^2$. This is what becomes of the usual Coulomb field. It can be thought of as the part of the electric field that remains bound to the particle. The fact that it is proportional to $\hat{\mathbf{R}}$, with a slight off-set from the velocity, means that it is roughly isotropic.
- The second term drops off as $1/R$ and is proportional to the acceleration. This describes the radiation emitted by the particle. Its dependence on the acceleration means that it's highly directional.

The Magnetic Field

To compute the magnetic field, we start with the expression (6.38),

$$\mathbf{A}(\mathbf{x}, t) = \frac{q\mu_0}{4\pi} \int dt' \frac{\mathbf{v}(t')}{R(t')} \delta(s)$$

with $s = t - t' - R(t')/c$. Then, using similar manipulations to those above, we have

$$\begin{aligned} \mathbf{B} = \nabla \times \mathbf{A} &= \frac{q\mu_0}{4\pi} \int dt' \left[-\frac{\nabla R}{R^2} \times \mathbf{v} \delta(s) + \frac{\nabla s \times \mathbf{v}}{R} \delta'(s) \right] \\ &= \frac{q\mu_0}{4\pi} \int ds \left[-\frac{\nabla R}{\kappa R^2} \times \mathbf{v} - \frac{1}{\kappa} \frac{d}{dt'} \left(\frac{\nabla R \times \mathbf{v}}{\kappa R c} \right) \right] \delta(s) \end{aligned} \quad (6.43)$$

We've already done the hard work necessary to compute this time derivative. We can write,

$$\begin{aligned} \frac{d}{dt'} \left(\frac{\nabla R \times \mathbf{v}}{\kappa R} \right) &= \frac{d}{dt'} \left(\frac{(\hat{\mathbf{R}} - \mathbf{v}/c) \times \mathbf{v}}{\kappa R} \right) \\ &= \frac{d}{dt'} \left(\frac{\hat{\mathbf{R}} - \mathbf{v}/c}{\kappa R} \right) \times \mathbf{v} + \frac{\hat{\mathbf{R}} - \mathbf{v}/c}{\kappa R} \times \mathbf{a} \end{aligned}$$

Now we can use (6.41). A little algebra shows that terms of the form $\mathbf{v} \times \mathbf{a}$ cancel, and we're left with

$$\frac{d}{dt'} \left(\frac{\hat{\mathbf{R}} \times \mathbf{v}}{\kappa R} \right) = -\frac{c\hat{\mathbf{R}} \times \mathbf{v}}{R^2} + \frac{c\hat{\mathbf{R}} \times \mathbf{v}}{\gamma^2 \kappa^2 R^2} + \frac{(\mathbf{R} \cdot \mathbf{a}) \hat{\mathbf{R}} \times \mathbf{v}}{c\kappa^2 R^2} + \frac{\hat{\mathbf{R}} \times \mathbf{a}}{\kappa R}$$

Substituting this into (6.43), a little re-arranging of the terms gives us our final expression for the magnetic field,

$$\mathbf{B} = -\frac{q\mu_0}{4\pi} \left[\frac{\hat{\mathbf{R}} \times \mathbf{v}}{\gamma^2 \kappa^3 R^2} + \frac{(\hat{\mathbf{R}} \cdot \mathbf{a})(\hat{\mathbf{R}} \times \mathbf{v}/c) + \kappa \hat{\mathbf{R}} \times \mathbf{a}}{c\kappa^3 R} \right]_{\text{ret}} \quad (6.44)$$

We see that this has a similar form to the electric field (6.42). The first term falls off as $1/R^2$ and is bound to the particle. It vanishes when $\mathbf{v} = 0$ which tells us that a charged particle only gives rise to a magnetic field when it moves. The second term falls off as $1/R$. This is generated by the acceleration and describes the radiation emitted by the particle. You can check that \mathbf{E} in (6.42) and \mathbf{B} in (6.44) are related through

$$\mathbf{B} = \frac{1}{c} [\hat{\mathbf{R}}]_{\text{ret}} \times \mathbf{E} \quad (6.45)$$

as you might expect.

6.4.4 A Covariant Formalism for Radiation

Before we make use of the Liénard-Wiechert potentials, we're going to do something a little odd: we're going to derive them again. This time, however, we'll make use of the Lorentz invariant notation of electromagnetism. This won't teach us anything new about physics and the results of this section aren't needed for what follows. But it will give us some practice on manipulating these covariant quantities. Moreover, the final result will be pleasingly concise.

A Covariant Retarded Potential

We start with our expression for the retarded potential (6.32) in terms of the current,

$$A_\mu(\mathbf{x}, t) = \frac{\mu_0}{4\pi} \int d^3x' \frac{J_\mu(\mathbf{x}', t_{\text{ret}})}{|\mathbf{x} - \mathbf{x}'|} \quad (6.46)$$

with $t_{\text{ret}} = t - |\mathbf{x} - \mathbf{x}'|/c$. This has been the key formula that we've used throughout this section. Because it was derived from the Maxwell equations, this formula should be Lorentz covariant, meaning that someone in a different inertial frame will write down the same equation. Although this *should* be true, it's not at all obvious from the way that (6.46) is written that it actually is true. The equation involves only integration over space, and the denominator depends only on the spatial distance between two points. Neither of these are concepts that different observers agree upon.

So our first task is to rewrite (6.46) in a way which is manifestly Lorentz covariant. To do this, we work with four-vectors $X^\mu = (ct, \mathbf{x})$ and take a quantity which everyone agrees upon: the spacetime distance between two points

$$(X - X')^2 = \eta_{\mu\nu}(X^\mu - X'^\mu)(X^\nu - X'^\nu) = c^2(t - t')^2 - |\mathbf{x} - \mathbf{x}'|^2$$

Consider the delta-function $\delta((X - X')^2)$, which is non-vanishing only when X and X' are null-separated. This is a Lorentz-invariant object. Let's see what it looks like when written in terms of the time coordinate t . We will need the general result for delta-functions

$$\delta(f(x)) = \sum_{x_i} \frac{\delta(x - x_i)}{|f'(x_i)|} \quad (6.47)$$

where the sum is over all roots $f(x_i) = 0$. Using this, we can write

$$\begin{aligned} \delta((X - X')^2) &= \delta([c(t' - t) + |\mathbf{x} - \mathbf{x}'|][c(t' - t) - |\mathbf{x} - \mathbf{x}'|]) \\ &= \frac{\delta(ct' - ct + |\mathbf{x} - \mathbf{x}'|)}{2c|t - t'|} + \frac{\delta(ct' - ct - |\mathbf{x} - \mathbf{x}'|)}{2c|t - t'|} \\ &= \frac{\delta(ct' - ct + |\mathbf{x} - \mathbf{x}'|)}{2|\mathbf{x} - \mathbf{x}'|} + \frac{\delta(ct' - ct - |\mathbf{x} - \mathbf{x}'|)}{2|\mathbf{x} - \mathbf{x}'|} \end{aligned}$$

The argument of the first delta-function is $ct' - ct_{\text{ret}}$ and so this term contributes only if $t' < t$. The argument of the second delta-function is $ct' - ct_{\text{adv}}$ and so this term can contribute only contribute if $t' > t$. But the temporal ordering of two spacetime points is also something all observers agree upon, as long as those points are either timelike or null separated. And here the delta-function requires the points to be null separated.

This means that if we picked just one of these terms, that choice would be Lorentz invariant. Mathematically, we do this using the Heaviside step-function

$$\Theta(t - t') = \begin{cases} 1 & t > t' \\ 0 & t < t' \end{cases}$$

We have

$$\delta((X - X')^2) \Theta(t - t') = \frac{\delta(ct' - ct_{\text{ret}})}{2|\mathbf{x} - \mathbf{x}'|} \quad (6.48)$$

The left-hand side is manifestly Lorentz invariant. The right-hand side doesn't look Lorentz invariant, but this formula tells us that it must be! Now we can make use of this to rewrite (6.46) in a way that the Lorentz covariance is obvious. It is

$$A_\mu(X) = \frac{\mu_0}{2\pi} \int d^4 X' J_\mu(X') \delta((X - X')^2) \Theta(t - t') \quad (6.49)$$

where the integration is now over spacetime, $d^4 X' = c dt' d^3 x'$. The combination of the delta-function and step-functions ensure that this integration is limited to the past light-cone of a point.

A Covariant Current

Next, we want a covariant expression for the current formed by a moving charged particle. We saw earlier that a particle tracing out a trajectory $\mathbf{y}(t)$ gives rise to a charge density (6.30) and current (6.31) given by

$$\rho(\mathbf{x}, t) = q \delta^3(\mathbf{x} - \mathbf{y}(t)) \quad \text{and} \quad \mathbf{J}(\mathbf{x}, t) = q \mathbf{v}(t) \delta^3(\mathbf{x} - \mathbf{y}(t)) \quad (6.50)$$

(We've changed notation from $\mathbf{r}(t)$ to $\mathbf{y}(t)$ to denote the trajectory of the particle). How can we write this in a manifestly covariant form?

We know from our course on Special Relativity that the best way to parametrise the worldline of a particle is by using its proper time τ . We'll take the particle to have trajectory $Y^\mu(\tau) = (ct(\tau), \mathbf{y}(\tau))$. Then the covariant form of the current is

$$J^\mu(X) = qc \int d\tau \dot{Y}^\mu(\tau) \delta^4(X^\nu - Y^\nu(\tau)) \quad (6.51)$$

It's not obvious that (6.51) is the same as (6.50). To see that it is, we can decompose the delta-function as

$$\delta^4(X^\nu - Y^\nu(\tau)) = \delta(ct - Y^0(\tau)) \delta^3(\mathbf{x} - \mathbf{y}(\tau))$$

The first factor allows us to do the integral over $d\tau$, but at the expense of picking up a Jacobian-like factor $1/\dot{Y}^0$ from (6.47). We have

$$J^\mu = \frac{qc\dot{Y}^\mu}{\dot{Y}^0} \delta^3(\mathbf{x} - \mathbf{y}(\tau))$$

which does give us back the same expressions (6.50).

Covariant Liènard-Wierchert Potentials

We can now combine (6.49) and (6.51) to get the retarded potential,

$$\begin{aligned} A^\mu(X) &= \frac{\mu_0 qc}{4\pi} \int d^4 X' \int d\tau \dot{Y}^\mu(\tau) \delta^4(X'^\nu - Y^\nu(\tau)) \frac{\delta(ct' - ct_{\text{ret}})}{|\mathbf{x} - \mathbf{x}'|} \\ &= \frac{\mu_0 qc}{4\pi} \int d\tau \dot{Y}^\mu(\tau) \frac{\delta(ct - Y^0(\tau) - |\mathbf{x} - \mathbf{y}(\tau)|)}{|\mathbf{x} - \mathbf{y}(\tau)|} \end{aligned}$$

This remaining delta-function implicitly allows us to do the integral over proper time. Using (6.48) we can rewrite it as

$$\frac{\delta(ct - Y^0(\tau) - |\mathbf{x} - \mathbf{y}(\tau)|)}{2|\mathbf{x} - \mathbf{y}(\tau)|} = \delta(R(\tau) \cdot R(\tau)) \Theta(R^0(\tau)) \quad (6.52)$$

where we've introduced the separation 4-vector

$$R^\mu = X^\mu - Y^\mu(\tau)$$

The delta-function and step-function in (6.52) pick out a unique value of the proper time that contributes to the gauge potential at point X . We call this proper time τ_\star . It is the retarded time lying along a null direction, $R(\tau_\star) \cdot R(\tau_\star) = 0$. This should be thought of as the proper time version of our previous formula (6.36).

The form (6.52) allows us to do the integral over τ . But we still pick up a Jacobian-like factor from (6.47). This gives

$$\delta(R(\tau) \cdot R(\tau)) \Theta(R^0(\tau)) = \frac{\delta(\tau - \tau_\star)}{2|R^\mu(\tau_\star)\dot{Y}^\mu(\tau_\star)|}$$

Putting all of this together gives our covariant form for the Liènard-Wierchert potential,

$$A^\mu(X) = \frac{\mu_0 qc}{4\pi} \frac{\dot{Y}^\mu(\tau_\star)}{|R^\nu(\tau_\star)\dot{Y}_\nu(\tau_\star)|}$$

This is our promised, compact expression. Expanding it out will give the previous results for the scalar (6.34) and vector (6.35) potentials. (To see this, you'll need to first show that $|R^\nu(\tau_\star)\dot{Y}^\nu(\tau_\star)| = c\gamma(\tau_\star)R(\tau_\star)(1 - \hat{\mathbf{R}}(\tau_\star) \cdot \mathbf{v}(\tau_\star)/c)$.)

The next step is to compute the field strength $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$. This is what took us some time in Section 6.4.3. It turns out to be somewhat easier in the covariant approach. We need to remember that τ_\star is a function of X^μ . Then, we get

$$F_{\mu\nu} = \frac{\mu_0 qc}{4\pi} \left(\frac{\ddot{Y}_\nu(\tau_\star)}{|R^\rho(\tau_\star)\dot{Y}_\rho(\tau_\star)|} \frac{\partial\tau_\star}{\partial X^\mu} - \frac{\dot{Y}_\nu(\tau_\star)}{|R^\rho(\tau_\star)\dot{Y}_\rho(\tau_\star)|^2} \frac{\partial|R^\sigma(\tau_\star)\dot{Y}_\sigma(\tau_\star)|}{\partial X^\mu} \right) - (\mu \leftrightarrow \nu) \quad (6.53)$$

The simplest way to compute $\partial\tau_\star/\partial X^\mu$ is to start with $\eta_{\rho\sigma}R^\rho(\tau_\star)R^\sigma(\tau_\star) = 0$. Differentiating gives

$$\eta_{\rho\sigma}R^\rho(\tau_\star)\partial_\mu R^\sigma(\tau_\star) = \eta_{\rho\sigma}R^\rho(\tau_\star) \left(\delta_\mu^\sigma - \dot{Y}^\sigma(\tau_\star)\partial_\mu\tau_\star \right) = 0$$

Rearranging gives

$$\frac{\partial\tau_\star}{\partial X^\mu} = \frac{R_\mu(\tau_\star)}{R^\nu(\tau_\star)\dot{Y}_\nu(\tau_\star)}$$

For the other term, we have

$$\begin{aligned} \frac{\partial|R^\sigma(\tau_\star)\dot{Y}_\sigma(\tau_\star)|}{\partial X^\mu} &= \left(\delta_\mu^\sigma - \dot{Y}^\sigma(\tau_\star)\partial_\mu\tau_\star \right) \dot{Y}_\sigma(\tau_\star) + R^\sigma(\tau_\star)\ddot{Y}_\sigma(\tau_\star)\partial_\mu\tau_\star \\ &= \left(R^\sigma(\tau_\star)\ddot{Y}_\sigma(\tau_\star) + c^2 \right) \partial_\mu\tau_\star + \dot{Y}_\mu(\tau_\star) \end{aligned}$$

where we've used $\dot{Y}^\mu\dot{Y}_\mu = c^2$. Using these in (6.53), we get our final expression for the field strength,

$$F_{\mu\nu}(X) = \frac{\mu_0 qc}{4\pi} \frac{1}{R^\rho\dot{Y}_\rho} \left[(-c^2 + R^\lambda\ddot{Y}_\lambda) \frac{R_\mu\dot{Y}_\nu - R_\nu\dot{Y}_\mu}{(R^\sigma\dot{Y}_\sigma)^2} + \frac{\ddot{Y}_\mu R_\nu - \ddot{Y}_\nu R_\mu}{R^\sigma\dot{Y}_\sigma} \right] \quad (6.54)$$

This is the covariant field strength. It takes a little work to write this in terms of the component \mathbf{E} and \mathbf{B} fields but the final answer is, of course, given by (6.42) and (6.44) that we derived previously. Indeed, you can see the general structure in (6.54). The first term is proportional to velocity and goes as $1/R^2$; the second term is proportional to acceleration and goes as $1/R$.

6.4.5 Bremsstrahlung, Cyclotron and Synchrotron Radiation

To end our discussion, we derive the radiation due to some simple relativistic motion.

Power Radiated Again: Relativistic Larmor Formula

In Section 6.2.2, we derived the Larmor formula for the emitted power in the electric dipole approximation to radiation. In this section, we present the full, relativistic version of this formula.

We'll work with the expressions for the radiation fields \mathbf{E} (6.42) and \mathbf{B} (6.44). As previously, we consider only the radiative part of the electric and magnetic fields which drops off as $1/R$. The Poynting vector is

$$\mathbf{S} = \frac{1}{\mu_0} \mathbf{E} \times \mathbf{B} = \frac{1}{\mu_0 c} \mathbf{E} \times (\hat{\mathbf{R}} \times \mathbf{E}) = \frac{1}{\mu_0 c} |\mathbf{E}|^2 \hat{\mathbf{R}}$$

where all of these expressions are to be computed at the retarded time. The second equality follows from the relation (6.45), while the final equality follows because the radiative part of the electric field (6.42) is perpendicular to $\hat{\mathbf{R}}$. Using the expression (6.42), we have

$$\mathbf{S} = \frac{q^2}{16\pi^2 \epsilon_0 c^3} \frac{|\hat{\mathbf{R}} \times [(\hat{\mathbf{R}} - \mathbf{v}/c) \times \mathbf{a}]|^2}{\kappa^6 R^2} \hat{\mathbf{R}}$$

with $\kappa = 1 - \hat{\mathbf{R}} \cdot \mathbf{v}/c$.

Recall that everything in the formula above is evaluated at the retarded time t' , defined by $t' + R(t')/c = t$. This means, that the coordinates are set up so that we can integrate \mathbf{S} over a sphere of radius R that surrounds the particle at its retarded time. However, there is a subtlety in computing the emitted power, associated to the Doppler effect. The energy emitted per unit time t is not the same as the energy emitted per unit time t' . They differ by the factor $dt/dt' = \kappa$. The power emitted per unit time t' , per solid angle $d\Omega$, is

$$\frac{d\mathcal{P}}{d\Omega} = \kappa R^2 \mathbf{S} \cdot \hat{\mathbf{R}} = \frac{q^2}{16\pi^2 \epsilon_0 c^3} \frac{|\hat{\mathbf{R}} \times [(\hat{\mathbf{R}} - \mathbf{v}/c) \times \mathbf{a}]|^2}{\kappa^5} \quad (6.55)$$

To compute the emitted power, we must integrate this expression over the sphere. This is somewhat tedious. The result is given by

$$\mathcal{P} = \frac{q^2}{6\pi \epsilon_0 c^3} \gamma^4 \left(a^2 + \frac{\gamma^2}{c^2} (\mathbf{v} \cdot \mathbf{a})^2 \right) \quad (6.56)$$

This is the relativistic version of the Larmor formula (6.18). (There is a factor of 2 difference when compared to (6.20) because the former equation was time averaged). We now apply this to some simple examples.

Bremsstrahlung

Suppose a particle is travelling in a straight line, with velocity \mathbf{v} parallel to acceleration \mathbf{a} . The most common situation of this type occurs when a particle decelerates. In this case, the emitted radiation is called *bremsstrahlung*, German for “braking radiation”.

We'll sit at some point \mathbf{x} , at which the radiation reaches us from the retarded point on the particle's trajectory $\mathbf{r}(t')$. As before, we define $\mathbf{R}(t') = \mathbf{x} - \mathbf{r}(t')$. We introduce the angle θ , defined by

$$\hat{\mathbf{R}} \cdot \mathbf{v} = v \cos \theta$$

Because the $\mathbf{v} \times \mathbf{a}$ term in (6.55) vanishes, the angular dependence of the radiation is rather simple in this case. It is given by

$$\frac{d\mathcal{P}}{d\Omega} = \frac{q^2 a^2}{16\pi\epsilon_0 c^3} \frac{\sin^2 \theta}{(1 - (v/c) \cos \theta)^5}$$

For $v \ll c$, the radiation is largest in the direction $\theta \approx \pi/2$, perpendicular to the direction of travel. But, at relativistic speeds, $v \rightarrow c$, the radiation is beamed in the forward direction in two lobes, one on either side of the particle's trajectory. The total power emitted is (6.56) which, in this case, simplifies to

$$\mathcal{P} = \frac{q^2 \gamma^6 a^2}{6\pi\epsilon_0 c^3}$$

Cyclotron and Synchrotron Radiation

Suppose that the particle travels in a circle, with $\mathbf{v} \cdot \mathbf{a} = 0$. We'll pick axes so that \mathbf{a} is aligned with the x -axis and \mathbf{v} is aligned with the z -axis. Then we write

$$\hat{\mathbf{R}} = \sin \theta \cos \phi \hat{\mathbf{x}} + \sin \theta \sin \phi \hat{\mathbf{y}} + \cos \theta \hat{\mathbf{z}}$$

After a little algebra, we find that the angular dependence of the emitted radiation is

$$\frac{d\mathcal{P}}{d\Omega} = \frac{q^2 a^2}{16\pi\epsilon_0 c^3} \frac{1}{(1 - (v/c) \cos \theta)^3} \left(1 - \frac{\sin^2 \theta \cos^2 \phi}{\gamma^2 (1 - (v/c) \cos \theta)^2} \right)$$

At non-relativistic speeds, $v \ll c$, the angular dependence takes the somewhat simpler form $(1 - \sin^2 \theta \cos^2 \phi)$. In this limit, the radiation is referred to as *cyclotron radiation*. In contrast, in the relativistic limit $v \rightarrow c$, the radiation is again beamed mostly in the forwards direction. This limit is referred to as *synchrotron radiation*. The total emitted power (6.56) is this time given by

$$\mathcal{P} = \frac{q^2 \gamma^4 a^2}{6\pi\epsilon_0 c^3}$$

Note that the factors of γ differ from the case of linear acceleration.

7. Electromagnetism in Matter

Until now, we've focussed exclusively on electric and magnetic fields in vacuum. We end this course by describing the behaviour of electric and magnetic fields inside materials, whether solids, liquids or gases.

The materials that we would like to discuss are insulators which, in this context, are usually called *dielectrics*. These materials are the opposite of conductors: they don't have any charges that are free to move around. Moreover, they are typically neutral so that – at least when averaged – the charge density vanishes: $\rho = 0$. You might think that such neutral materials can't have too much effect on electric and magnetic fields. But, as we will see, things are more subtle and interesting.

7.1 Electric Fields in Matter

The fate of electric fields inside a dielectric depends on the microscopic make-up of the material. We going to work only with the simplest models. We'll consider our material to be constructed from a lattice of neutral atoms. Each of these atoms consists of a positively charged nuclei, surrounded by a negatively charged cloud of electrons. A cartoon of this is shown in the figure; the nucleus is drawn in red, the cloud of electrons in yellow.

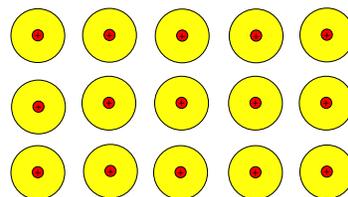


Figure 59: A simple model of a neutral material

Suppose that electric field \mathbf{E} is applied to this material. What happens? Although each atom is neutral, its individual parts are not. This results in an effect called *polarisation*: the positively charged nucleus gets pushed a little in the direction of \mathbf{E} ; the negatively charged cloud gets pushed a little in the opposite direction. (This is not to be confused with the orientation of the electromagnetic wave which also has the name “polarisation”).

The net effect is that the neutral atom gains an electric dipole moment. Recall from Section 2 that two equal and opposite charges, $+q$ and $-q$, separated by a distance \mathbf{d} , have an electric dipole $\mathbf{p} = q\mathbf{d}$. By convention, \mathbf{p} points from the negative charge to the positive charge.

It turns out that in most materials, the induced electric dipole is proportional to the electric field,

$$\mathbf{p} = \alpha\mathbf{E} \tag{7.1}$$

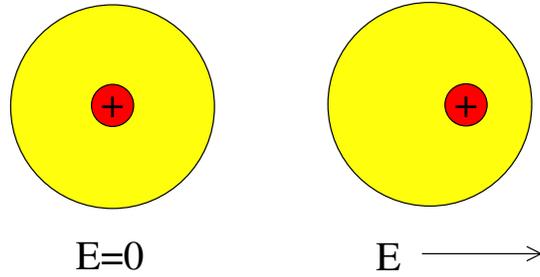


Figure 60: The polarisation of an atom

The proportionality factor α is called the *atomic polarisability*. Because \mathbf{p} points from negative to positive charge, it points in the same direction as \mathbf{E} . The electric field will also result in higher multipole moments of the atoms. (For example, the cloud of electrons will be distorted). We will ignore these effects.

A Simple Model for Atomic Polarizability

Here's a simple model which illustrates how the relationship (7.1) arises. It also gives a ball-park figure for the value of the atomic polarizability α . Consider a nucleus of charge $+q$, surrounded by a spherical cloud of electrons of radius a . We'll take this cloud to have uniform charge density. If we just focus on the electron cloud for now, the electric field it produces was computed in Section 2: it rises linearly inside the cloud, before dropping off as $1/r^2$ outside the cloud. Here we're interested in the linearly increasing behaviour inside

$$\mathbf{E}_{\text{cloud}} = \frac{1}{4\pi\epsilon_0} \frac{qr}{a^3} \hat{\mathbf{r}} \quad (r < a) \quad (7.2)$$

In the absence of an external field, the nucleus feels the field due to the cloud and sits at $r = 0$. Now apply an external electric field \mathbf{E} . The nucleus will be displaced to sit at a point where $\mathbf{E} + \mathbf{E}_{\text{cloud}} = 0$. In other words, it will be displaced by

$$\mathbf{r} = \frac{4\pi\epsilon_0 a^3}{q} \mathbf{E} \quad \Rightarrow \quad \mathbf{p} = q\mathbf{r} = 4\pi\epsilon_0 a^3 \mathbf{E}$$

This gives the simple expression $\alpha = 4\pi\epsilon_0 a^3$. This isn't too far off the experimentally measured values. For example, for hydrogen $\alpha/4\pi\epsilon_0 \approx 0.7 \times 10^{-30} \text{ m}^3$ which, from the above formula, suggests that the size of the cloud is around $a \sim 10^{-10} \text{ m}$.

7.1.1 Polarisation

We've learnt that applying an electric field to a material causes each atom to pick up a dipole moment. We say that the material is *polarised*. The *polarisation* \mathbf{P} is defined

to be the average dipole moment per unit volume. If n is the density of atoms, each with dipole moment \mathbf{p} , then we can write

$$\mathbf{P} = n\mathbf{p} \tag{7.3}$$

We've actually dodged a bullet in writing this simple equation and evaded a subtle, but important, point. Let me try to explain. Viewed as a function of spatial position, the dipole moment $\mathbf{p}(\mathbf{r})$ is ridiculously complicated, varying wildly on distances comparable to the atomic scale. We really couldn't care less about any of this. We just want the average dipole moment, and that's what the equation above captures. But we do care if the average dipole moment varies over large, macroscopic distances. For example, the density n may be larger in some parts of the solid than others. And, as we'll see, this is going to give important, physical effects. This means that we don't want to take the average of $\mathbf{p}(\mathbf{r})$ over the whole solid since this would wash out all variations. Instead, we just want to average over small distances, blurring out any atomic messiness, but still allowing \mathbf{P} to depend on \mathbf{r} over large scales. The equation $\mathbf{P} = n\mathbf{p}$ is supposed to be shorthand for all of this. Needless to say, we could do a better job of defining \mathbf{P} if forced to, but it won't be necessary in what follows.

The polarisation of neutral atoms is not the only way that materials can become polarised. One simple example is water. Each H_2O molecule already carries a dipole moment. (The oxygen atom carries a net negative charge, with each hydrogen carrying a positive charge). However, usually these molecules are jumbled up in water, each pointing in a different direction so that the dipole moments cancel out and the polarisation is $\mathbf{P} = 0$. This changes if we apply an electric field. Now the dipoles all want to align with the electric field, again leading to a polarisation.

In general, the polarisation \mathbf{P} can be a complicated function of the electric field \mathbf{E} . However, most materials it turns out that \mathbf{P} is proportional to \mathbf{E} . Such materials are called *linear dielectrics*. They have

$$\mathbf{P} = \epsilon_0\chi_e\mathbf{E} \tag{7.4}$$

where χ_e is called the *electric susceptibility*. It is always positive: $\chi_e > 0$. Our simple minded computation of atomic polarisability above gave such a linear relationship, with $\epsilon_0\chi_e = n\alpha$.

The reason why most materials are linear dielectrics follows from some simple dimensional analysis. Any function that has $\mathbf{P}(\mathbf{E} = 0) = 0$ can be Taylor expanded as a linear term + quadratic + cubic and so on. For suitably small electric fields, the linear

term always dominates. But how small is small? To determine when the quadratic and higher order terms become important, we need to know the relevant scale in the problem. For us, this is the scale of electric fields inside the atom. But these are huge. In most situations, the applied electric field leading to the polarisation is a tiny perturbation and the linear term dominates. Nonetheless, from this discussion it should be clear that we do expect the linearity to fail for suitably high electric fields.

There are other exceptions to linear dielectrics. Perhaps the most striking exception are materials for which $\mathbf{P} \neq 0$ even in the absence of an electric field. Such materials – which are not particularly common – are called *ferroelectric*. For what it's worth, an example is $BaTiO_3$.

Bound Charge

Whatever the cause, when a material is polarised there will be regions in which there is a build up of electric charge. This is called *bound charge* to emphasise the fact that it's not allowed to move and is arising from polarisation effects. Let's illustrate this with a simple example before we describe the general case. Let's go back to our lattice of neutral atoms. As we've seen, in the presence of an electric field they become polarised, as shown in the figure. However, as long as the polarisation is uniform, so \mathbf{P} is constant, there is no net charge in the middle of the material: averaged over many atoms, the total charge remains the same. The only place that there is a net build up of charge is on the surface. In contrast, if $\mathbf{P}(\mathbf{r})$ is not constant, there will also be regions in the middle that have excess electric charge.

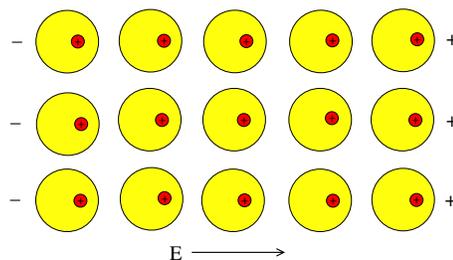


Figure 61: A polarised material

To describe this, recall that the electric potential due to each dipole \mathbf{p} is

$$\phi(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \frac{\mathbf{p} \cdot \mathbf{r}}{r^3}$$

(We computed this in Section 2). Integrating over all these dipoles, we can write the potential in terms of the polarisation,

$$\phi(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \int_V d^3r' \frac{\mathbf{P}(\mathbf{r}') \cdot (\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^3}$$

We then have the following manipulations.,

$$\begin{aligned}\phi(\mathbf{r}) &= \frac{1}{4\pi\epsilon_0} \int_V d^3r' \mathbf{P}(\mathbf{r}') \cdot \nabla' \left(\frac{1}{|\mathbf{r} - \mathbf{r}'|} \right) \\ &= \frac{1}{4\pi\epsilon_0} \int_S d\mathbf{S} \cdot \frac{\mathbf{P}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} - \frac{1}{4\pi\epsilon_0} \int_V d^3r' \frac{\nabla' \cdot \mathbf{P}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}\end{aligned}$$

where S is the boundary of V . But both of these terms have a very natural interpretation. The first is the kind of potential that we would get from a surface charge,

$$\sigma_{\text{bound}} = \mathbf{P} \cdot \hat{\mathbf{n}}$$

where $\hat{\mathbf{n}}$ is the normal to the surface S . The second term is the kind of potential that we would get from a charge density of the form,

$$\rho_{\text{bound}}(\mathbf{r}) = -\nabla \cdot \mathbf{P}(\mathbf{r}) \quad (7.5)$$

This matches our intuition above. If the polarisation \mathbf{P} is constant then we only find a surface charge. But if \mathbf{P} varies throughout the material then this can lead to non-vanishing charge density sitting inside the material.

7.1.2 Electric Displacement

We learned in our first course that the electric field obeys Gauss' law

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0}$$

This is a fundamental law of Nature. It doesn't change just because we're inside a material. But, from our discussion above, we see that there's a natural way to separate the electric charge into two different types. There is the bound charge ρ_{bound} that arises due to polarisation. And then there is anything else. This could be some electric impurities that are stuck in the dielectric, or it could be charge that is free to move because our insulator wasn't quite as good an insulator as we originally assumed. The only important thing is that this other charge does not arise due to polarisation. We call this extra charge *free charge*, ρ_{free} . Gauss' law reads

$$\begin{aligned}\nabla \cdot \mathbf{E} &= \frac{1}{\epsilon_0} (\rho_{\text{free}} + \rho_{\text{bound}}) \\ &= \frac{1}{\epsilon_0} (\rho_{\text{free}} - \nabla \cdot \mathbf{P})\end{aligned}$$

We define the *electric displacement*,

$$\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P} \quad (7.6)$$

This obeys

$$\nabla \cdot \mathbf{D} = \rho_{\text{free}} \quad (7.7)$$

That's quite nice. Gauss' law for the displacement involves only the free charge; any bound charge arising from polarisation has been absorbed into the definition of \mathbf{D} .

For linear dielectrics, the polarisation is given by (7.4) and the displacement is proportional to the electric field. We write

$$\mathbf{D} = \epsilon \mathbf{E}$$

where $\epsilon = \epsilon_0(1 + \chi_e)$ is called the *permittivity* of the material. We see that, for linear dielectrics, things are rather simple: all we have to do is replace ϵ_0 with ϵ everywhere. Because $\epsilon > \epsilon_0$, it means that the electric field will be decreased. We say that it is *screened* by the bound charge. The amount by which the electric field is reduced is given by the dimensionless *relative permittivity* or *dielectric constant*,

$$\epsilon_r = \frac{\epsilon}{\epsilon_0} = 1 + \chi_e$$

For gases, ϵ_r is very close to 1. (It differs at one part in 10^{-3} or less). For water, $\epsilon_r \approx 80$.

An Example: A Dielectric Sphere

As a simple example, consider a sphere of dielectric material of radius R . We'll place a charge Q at the centre. This gives rise to an electric field which polarises the sphere and creates bound charge. We want to understand the resulting electric field \mathbf{E} and electric displacement \mathbf{D} .

The modified Gauss' law (7.7) allows us to easily compute \mathbf{D} using the same kind of methods that we used in Section 2. We have

$$\mathbf{D} = \frac{Q}{4\pi r^2} \hat{\mathbf{r}} \quad (r < R)$$

where the condition $r < R$ means that this holds inside the dielectric. The electric field is then given by

$$\mathbf{E} = \frac{Q}{4\pi \epsilon r^2} \hat{\mathbf{r}} = \frac{Q/\epsilon_r}{4\pi \epsilon_0 r^2} \hat{\mathbf{r}} \quad (r < R) \quad (7.8)$$

This is what we'd expect from a charge Q/ϵ_r placed at the origin. The interpretation of this is that there is the bound charge gathers at the origin, screening the original charge Q . This bound charge is shown as the yellow ring in the figure surrounding the original charge in red. The amount of bound charge is simply the difference

$$Q_{\text{bound}} = \frac{Q}{\epsilon_r} - Q = \frac{1 - \epsilon_r}{\epsilon_r} Q = -\frac{\chi_e}{\epsilon_r} Q$$

This bound charge came from the polarisation of the sphere. But the sphere is a neutral object which means that total charge on it has to be zero. To accomplish this, there must be an equal, but opposite, charge on the surface of the sphere. This is shown as the red rim in the figure. This surface charge is given by

$$4\pi R^2 \sigma_{\text{bound}} = -Q_{\text{bound}} = \frac{\epsilon_r - 1}{\epsilon_r} Q$$

We know from our first course that such a surface charge will lead to a discontinuity in the electric field. And that's exactly what happens. Inside the sphere, the electric field is given by (7.8). Meanwhile outside the sphere, Gauss' law knows nothing about the intricacies of polarisation and we get the usual electric field due to a charge Q ,

$$\mathbf{E} = \frac{Q}{4\pi\epsilon_0 r^2} \hat{\mathbf{r}} \quad (r > R)$$

At the surface $r = R$ there is a discontinuity,

$$\mathbf{E} \cdot \hat{\mathbf{r}}|_+ - \mathbf{E} \cdot \hat{\mathbf{r}}|_- = \frac{Q}{4\pi\epsilon_0 R^2} - \frac{Q}{4\pi\epsilon R^2} = \frac{\sigma_{\text{bound}}}{\epsilon_0}$$

which is precisely the expected discontinuity due to surface charge.

7.2 Magnetic Fields in Matter

Electric fields are created by charges; magnetic fields are created by currents. We learned in our first course that the simplest way to characterise any localised current distribution is through a *magnetic dipole moment* \mathbf{m} . For example, a current I moving in a planar loop of area A with normal $\hat{\mathbf{n}}$ has magnetic dipole moment,

$$\mathbf{m} = IA\hat{\mathbf{n}}$$

The resulting long-distance gauge field and magnetic field are

$$\mathbf{A}(\mathbf{r}) = \frac{\mu_0}{4\pi} \frac{\mathbf{m} \times \mathbf{r}}{r^3} \quad \Rightarrow \quad \mathbf{B}(\mathbf{r}) = \frac{\mu_0}{4\pi} \left(\frac{3(\mathbf{m} \cdot \hat{\mathbf{r}})\hat{\mathbf{r}} - \mathbf{m}}{r^3} \right)$$

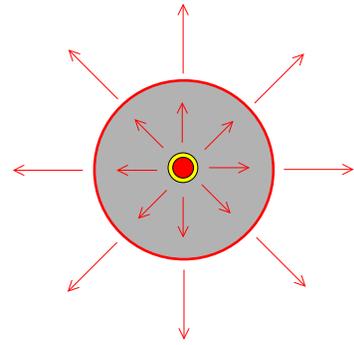


Figure 62: A polarised material

The basic idea of this section is that current loops, and their associated dipole moments, already exist inside materials. They arise through two mechanisms:

- Electrons orbiting the nucleus carry angular momentum and act as magnetic dipole moments.
- Electrons carry an intrinsic spin. This is purely a quantum mechanical effect. This too contributes to the magnetic dipole moment.

In the last section, we defined the polarisation \mathbf{P} to be the average dipole moment per unit volume. In analogy, we define the *magnetisation* \mathbf{M} to be the average magnetic dipole moment per unit volume. Just as in the polarisation case, here “average” means averaging over atomic distances, but keeping any macroscopic variations of the polarisation $\mathbf{M}(\mathbf{r})$. It’s annoyingly difficult to come up with simple yet concise notation for this. I’ll choose to write,

$$\mathbf{M}(\mathbf{r}) = n\langle\mathbf{m}(\mathbf{r})\rangle$$

where n is the density of magnetic dipoles (which can, in principle, also depend on position) and the notation $\langle\cdot\rangle$ means averaging over atomic distance scales. In most (but not all) materials, if there is no applied magnetic field then the different atomic dipoles all point in random directions. This means that, after averaging, $\langle\mathbf{m}\rangle = 0$ when $\mathbf{B} = 0$. However, when a magnetic field is applied, the dipoles line up. The magnetisation typically takes the form $\mathbf{M} \propto \mathbf{B}$. We’re going to use a slightly strange notation for the proportionality constant. (It’s historical but, as we’ll see, it turns out to simplify a later equation)

$$\mathbf{M} = \frac{1}{\mu_0} \frac{\chi_m}{1 + \chi_m} \mathbf{B} \quad (7.9)$$

where χ_m is the *magnetic susceptibility*. The magnetic properties of materials fall into three different categories. The first two are dictated by the sign of χ_m :

- *Diamagnetism*: $-1 < \chi_m < 0$. The magnetisation of diamagnetic materials points in the opposite direction to the applied magnetic field. Most metals are diamagnetic, including copper and gold. Most non-metallic materials are also diamagnetic including, importantly, water with $\chi_m \approx -10^{-5}$. This means, famously, that frogs are also diamagnetic. Superconductors can be thought of as “perfect” diamagnets with $\chi_m = -1$.
- *Paramagnetism*: $\chi_m > 0$. In paramagnets, the magnetisation points in the same direction as the field. There are a number of paramagnetic metals, including Tungsten, Cesium and Aluminium.

We see that the situation is already richer than what we saw in the previous section. There, the polarisation takes the form $\mathbf{P} = \epsilon_0 \chi_e \mathbf{E}$ with $\chi_e > 0$. In contrast, χ_m can have either sign. On top of this, there is another important class of material that don't obey (7.9). These are ferromagnets:

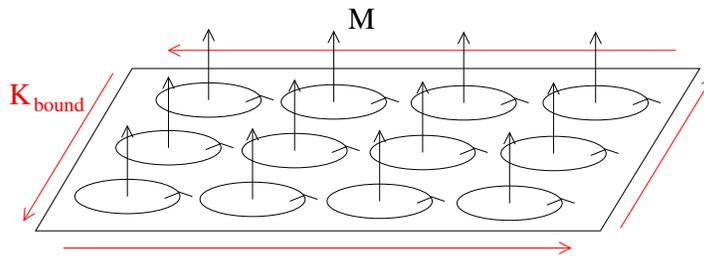
- *Ferromagnetism*: $\mathbf{M} \neq 0$ when $\mathbf{B} = 0$. Materials with this property are what you usually call “magnets”. They're the things stuck to your fridge. The direction of \mathbf{B} is from the south pole to the north. Only a few elements are ferromagnetic. The most familiar is Iron. Nickel and Cobalt are other examples.

In this course, we won't describe the microscopic effects that cause these different magnetic properties. They all involve quantum mechanics. (Indeed, the Bohr-van Leeuwen theorem says magnetism *can't* happen in a classical world — see the lecture notes on *Classical Dynamics*). A number of mechanisms for paramagnetism and diamagnetism in metals are described in the lecture notes on *Statistical Physics*.

7.2.1 Bound Currents

In the previous section, we saw that when a material is polarised, it results in bound charge. There is a similar story here. When a material becomes magnetised (at least in an anisotropic way), there will necessarily be regions in which there is a current. This is called the *bound current*.

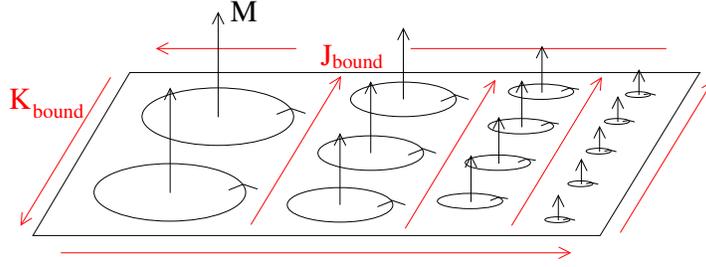
Let's first give an intuitive picture for where these bound currents appear from. Consider a bunch of equal magnetic dipoles arranged uniformly on a plane like this:



The currents in the interior region cancel out and we're left only with a surface current around the edge. In Section 3, we denoted a surface current as \mathbf{K} . We'll follow this notation and call the surface current arising from a constant, internal magnetisation $\mathbf{K}_{\text{bound}}$.

Now consider instead a situation where the dipoles are arranged on a plane, but have different sizes. We'll put the big ones to the left and the small ones to the right, like

this:



In this case, the currents in the interior no longer cancel. As we can see from the picture, they go into the page. Since \mathbf{M} is out of the page, and we've arranged things so that \mathbf{M} varies from left to right, this suggests that $\mathbf{J}_{\text{bound}} \sim \nabla \times \mathbf{M}$.

Let's now put some equations on this intuition. We know that the gauge potential due to a magnetic dipole is

$$\mathbf{A}(\mathbf{r}) = \frac{\mu_0}{4\pi} \frac{\mathbf{m} \times \mathbf{r}}{r^3}$$

Integrating over all dipoles, and doing the same kinds of manipulations that we saw for the polarisations, we have

$$\begin{aligned} \mathbf{A}(\mathbf{r}) &= \frac{\mu_0}{4\pi} \int_V d^3r' \frac{\mathbf{M}(\mathbf{r}') \times (\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^3} \\ &= \frac{\mu_0}{4\pi} \int_V d^3r' \mathbf{M}(\mathbf{r}') \times \nabla' \left(\frac{1}{|\mathbf{r} - \mathbf{r}'|} \right) \\ &= -\frac{\mu_0}{4\pi} \int_S d\mathbf{S}' \times \frac{\mathbf{M}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + \frac{\mu_0}{4\pi} \int_V d^3r' \frac{\nabla' \times \mathbf{M}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \end{aligned}$$

Again, both of these terms have natural interpretation. The first can be thought of as due to a surface current

$$\mathbf{K}_{\text{bound}} = \mathbf{M} \times \hat{\mathbf{n}}$$

where $\hat{\mathbf{n}}$ is normal to the surface. The second term is the bound current in the bulk of the material. We can compare its form to the general expression for the Biot-Savart law that we derived in Section 3,

$$\mathbf{A}(\mathbf{r}) = \frac{\mu_0}{4\pi} \int d^3r' \frac{\mathbf{J}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}$$

We see that the bound current is given by

$$\mathbf{J}_{\text{bound}} = \nabla \times \mathbf{M} \tag{7.10}$$

as expected from our intuitive description above. Note that the bound current is a steady current, in the sense that it obeys $\nabla \cdot \mathbf{J}_{\text{bound}} = 0$.

7.2.2 Ampère's Law Revisited

Recall that Ampère's law describes the magnetic field generated by static currents. We've now learned that, in a material, there can be two contributions to a current: the bound current $\mathbf{J}_{\text{bound}}$ that we've discussed above, and the current \mathbf{J}_{free} from freely flowing electrons that we were implicitly talking. In Section 3, we were implicitly talking about \mathbf{J}_{free} when we discussed currents. Ampère's law does not distinguish between these two currents; the magnetic field receives contributions from both.

$$\begin{aligned}\nabla \times \mathbf{B} &= \mu_0(\mathbf{J}_{\text{free}} + \mathbf{J}_{\text{bound}}) \\ &= \mu_0\mathbf{J}_{\text{free}} + \mu_0\nabla \times \mathbf{M}\end{aligned}$$

We define the *magnetising field*, \mathbf{H} as

$$\mathbf{H} = \frac{1}{\mu_0}\mathbf{B} - \mathbf{M} \quad (7.11)$$

This obeys

$$\nabla \times \mathbf{H} = \mathbf{J}_{\text{free}} \quad (7.12)$$

We see that the field \mathbf{H} plays a similar role to the electric displacement \mathbf{D} ; the effect of the bound currents have been absorbed into \mathbf{H} , so that only the free currents contribute. Note, however, that we can't quite forget about \mathbf{B} entirely, since it obeys $\nabla \cdot \mathbf{B} = 0$. In contrast, we don't necessarily have " $\nabla \cdot \mathbf{H} = 0$ ". Rather annoyingly, in a number of books \mathbf{H} is called the magnetic field and \mathbf{B} is called the magnetic induction. But this is stupid terminology so we won't use it.

For diamagnets or paramagnets, the magnetisation is linear in the applied magnetic field \mathbf{B} and we can write

$$\mathbf{B} = \mu\mathbf{H}$$

A little algebra shows that $\mu = \mu_0(1 + \chi_m)$. It is called the *permeability*. For most materials, μ differs from μ_0 only by 1 part in 10^5 or so. Finally, note that the somewhat strange definition (7.9) leaves us with the more sensible relationship between \mathbf{M} and \mathbf{H} ,

$$\mathbf{M} = \chi_m\mathbf{H}$$

7.3 Macroscopic Maxwell Equations

We've seen that the presence of bound charge and bound currents in matter can be absorbed into the definitions of \mathbf{D} and \mathbf{H} . This allowed us to present versions of Gauss' law (7.7) and Ampère's law (7.12) which feature only the free charges and free currents. These equations hold for electrostatic and magnetostatic situations respectively. In this section we explain how to reformulate Maxwell's equations in matter in more general, time dependent, situations.

Famously, when fields depend on time there is an extra term required in Ampère's law. However, there is also an extra term in the expression (7.10) for the bound current. This arises because the bound charge, ρ_{bound} , no longer sits still. It moves. But although it moves, it must still be locally conserved which means that it should satisfy a continuity equation

$$\nabla \cdot \mathbf{J}_{\text{bound}} = -\frac{\partial \rho_{\text{bound}}}{\partial t}$$

From our earlier analysis (7.5), we can express the bound charge in terms of the polarisation: $\rho_{\text{bound}} = -\nabla \cdot \mathbf{P}$. Including both this contribution and the contribution (7.10) from the magnetisation, we have the more general expression for the bound current

$$\mathbf{J}_{\text{bound}} = \nabla \times \mathbf{M} + \frac{\partial \mathbf{P}}{\partial t}$$

Let's see how we can package the Maxwell equation using this notation. We're interested in the extension to Ampère's law which reads

$$\begin{aligned} \nabla \times \mathbf{B} - \frac{1}{c^2} \frac{\partial \mathbf{E}}{\partial t} &= \mu_0 \mathbf{J}_{\text{free}} + \mu_0 \mathbf{J}_{\text{bound}} \\ &= \mu_0 \mathbf{J}_{\text{free}} + \mu_0 \nabla \times \mathbf{M} + \mu_0 \frac{\partial \mathbf{P}}{\partial t} \end{aligned}$$

As before, we can use the definition of \mathbf{H} in (7.11) to absorb the magnetisation term. But we can also use the definition of \mathbf{D} to absorb the polarisation term. We're left with the Maxwell equation

$$\nabla \times \mathbf{H} - \frac{\partial \mathbf{D}}{\partial t} = \mathbf{J}_{\text{free}}$$

The Macroscopic Maxwell Equations

Let's gather together everything we've learned. Inside matter, the four Maxwell equations become

$$\begin{aligned} \nabla \cdot \mathbf{D} = \rho_{\text{free}} \quad \text{and} \quad \nabla \times \mathbf{H} - \frac{\partial \mathbf{D}}{\partial t} = \mathbf{J}_{\text{free}} \\ \nabla \cdot \mathbf{B} = 0 \quad \text{and} \quad \nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \end{aligned} \tag{7.13}$$

There are the *macroscopic Maxwell equations*. Note that half of them are written in terms of the original \mathbf{E} and \mathbf{B} while the other half are written in terms of \mathbf{D} and \mathbf{H} . Before we solve them, we need to know the relationships between these quantities. In the simplest, linear materials, this can be written as

$$\mathbf{D} = \epsilon\mathbf{E} \quad \text{and} \quad \mathbf{B} = \mu\mathbf{H}$$

Doesn't all this look simple! The atomic mess that accompanies most materials can simply be absorbed into two constants, the permittivity ϵ and the permeability μ . Be warned, however: things are not always as simple as they seem. In particular, we'll see in Section 7.5 that the permittivity ϵ is not as constant as we're pretending.

7.3.1 A First Look at Waves in Matter

We saw earlier how the Maxwell equations give rise to propagating waves, travelling with speed c . We call these waves "light". Much of our interest in this section will be on what becomes of these waves when we work with the macroscopic Maxwell equations. What happens when they bounce off different materials? What really happens when they propagate through materials?

Let's start by looking at the basics. In the absence of any free charge or currents, the macroscopic Maxwell equations (7.13) become

$$\begin{aligned} \nabla \cdot \mathbf{D} = 0 \quad \text{and} \quad \nabla \times \mathbf{H} &= \frac{\partial \mathbf{D}}{\partial t} \\ \nabla \cdot \mathbf{B} = 0 \quad \text{and} \quad \nabla \times \mathbf{E} &= -\frac{\partial \mathbf{B}}{\partial t} \end{aligned} \tag{7.14}$$

which should be viewed together with the relationships $\mathbf{D} = \epsilon\mathbf{E}$ and $\mathbf{B} = \mu\mathbf{H}$. But these are of exactly the same form as the Maxwell equations in vacuum. Which means that, at first glance, the propagation of waves through a medium works just like in vacuum. All we have to do is replace $\epsilon_0 \rightarrow \epsilon$ and $\mu_0 \rightarrow \mu$. By the same sort of manipulations that we used in Section 4.3, we can derive the wave equations

$$\frac{1}{v^2} \frac{\partial^2 \mathbf{E}}{\partial t^2} - \nabla^2 \mathbf{E} = 0 \quad \text{and} \quad \frac{1}{v^2} \frac{\partial^2 \mathbf{H}}{\partial t^2} - \nabla^2 \mathbf{H} = 0$$

The only difference from what we saw before is that the speed of propagation is now given by

$$v^2 = \frac{1}{\epsilon\mu}$$

This is less than the speed in vacuum: $v^2 \leq c^2$. It's common to define the *index of refraction*, n , as

$$n = \frac{c}{v} \geq 1 \quad (7.15)$$

In most materials, $\mu \approx \mu_0$. In this case, the index of refraction is given in terms of the dielectric constant as

$$n \approx \sqrt{\epsilon_r}$$

The monochromatic, plane wave solutions to the macroscopic wave equations take the familiar form

$$\mathbf{E} = \mathbf{E}_0 e^{i(\mathbf{k} \cdot \mathbf{x} + \omega t)} \quad \text{and} \quad \mathbf{B} = \mathbf{B}_0 e^{i(\mathbf{k} \cdot \mathbf{x} + \omega t)}$$

where the dispersion relation is now given by

$$\omega^2 = v^2 k^2$$

The polarisation vectors must obey $\mathbf{E}_0 \cdot \mathbf{k} = \mathbf{B}_0 \cdot \mathbf{k} = 0$ and

$$\mathbf{B}_0 = \frac{\hat{\mathbf{k}} \times \mathbf{E}_0}{v}$$

Boundary Conditions

In what follows, we're going to spend a lot of time bouncing waves off various surfaces. We'll typically consider an interface between two dielectric materials with different permittivities, ϵ_1 and ϵ_2 . In this situation, we need to know how to patch together the fields on either side.

Let's first recall the boundary conditions that we derived in Sections 2 and 3. In the presence of surface charge, the electric field normal to the surface is discontinuous, while the electric field tangent to the surface is continuous. For magnetic fields, it's the other way around: in the presence of a surface current, the magnetic field normal to the surface is continuous while the magnetic field tangent to the surface is discontinuous.

What happens with dielectrics? Now we have two options of the electric field, \mathbf{E} and \mathbf{D} , and two options for the magnetic field, \mathbf{B} and \mathbf{H} . They can't both be continuous because they're related by $\mathbf{D} = \epsilon \mathbf{E}$ and $\mathbf{B} = \mu \mathbf{H}$ and we'll be interested in situation where ϵ (and possibly μ) are different on either side. Nonetheless, we can use the same kind of computations that we saw previously to derive the boundary conditions. Roughly, we get one boundary condition from each of the Maxwell equations.

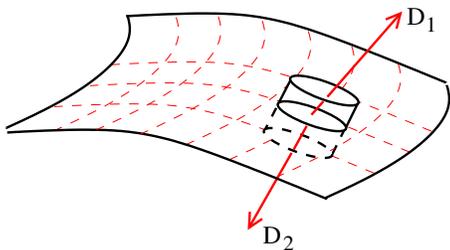


Figure 63: The normal component of the electric field is discontinuous

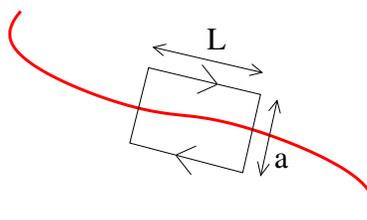


Figure 64: The tangential component of the electric field is continuous.

For example, consider the Gaussian pillbox shown in the left-hand figure above. Integrating the Maxwell equation $\nabla \cdot \mathbf{D} = \rho_{\text{free}}$ tells us that the normal component of \mathbf{D} is discontinuous in the presence of surface charge,

$$\hat{\mathbf{n}} \cdot (\mathbf{D}_2 - \mathbf{D}_1) = \sigma \quad (7.16)$$

where $\hat{\mathbf{n}}$ is the normal component pointing from 1 into 2. Here σ refers only to the free surface charge. It does not include any bound charges. Similarly, integrating $\nabla \cdot \mathbf{B} = 0$ over the same Gaussian pillbox tells us that the normal component of the magnetic field is continuous,

$$\hat{\mathbf{n}} \cdot (\mathbf{B}_2 - \mathbf{B}_1) = 0 \quad (7.17)$$

To determine the tangential components, we integrate the appropriate field around the loop shown in the right-hand figure above. By Stoke's theorem, this is going to be equal to the integral of the curl of the field over the bounding surface. This tells us what the appropriate field is: it's whatever appears in the Maxwell equations with a curl. So if we integrate \mathbf{E} around the loop, we get the result

$$\hat{\mathbf{n}} \times (\mathbf{E}_2 - \mathbf{E}_1) = 0 \quad (7.18)$$

Meanwhile, integrating \mathbf{H} around the loop tells us the discontinuity condition for the magnetic field

$$\hat{\mathbf{n}} \times (\mathbf{H}_2 - \mathbf{H}_1) = \mathbf{K} \quad (7.19)$$

where \mathbf{K} is the surface current.

7.4 Reflection and Refraction

We're now going to shine light on something and watch how it bounces off. We did something very similar in Section 4.3, where the light reflected off a conductor. Here, we're going to shine the light from one dielectric material into another. These two materials will be characterised by the parameters ϵ_1, μ_1 and ϵ_2, μ_2 . We'll place the interface at $x = 0$, with "region one" to the left and "region two" to the right.

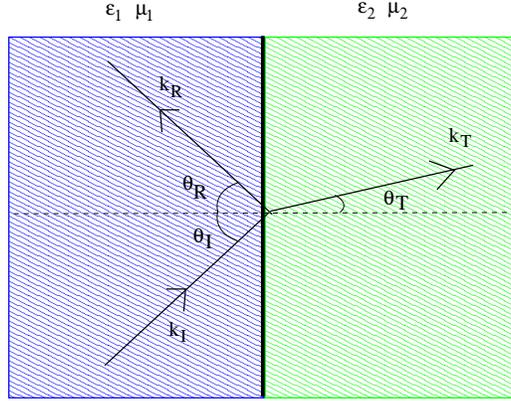


Figure 65: Incident, reflected and transmitted waves in a dielectric interface.

We send in an incident wave from region one towards the interface with a frequency ω_I and wavevector \mathbf{k}_I ,

$$\mathbf{E}_{\text{inc}} = \mathbf{E}_I e^{i(\mathbf{k}_I \cdot \mathbf{x} - \omega_I t)}$$

where

$$\mathbf{k}_I = k_I \cos \theta_I \hat{\mathbf{x}} + k_I \sin \theta_I \hat{\mathbf{z}}$$

When the wave hits the interface, two things can happen. It can be reflected, or it can pass through to the other region. In fact, in general, both of these things will happen. The reflected wave takes the general form,

$$\mathbf{E}_{\text{ref}} = \mathbf{E}_R e^{i(\mathbf{k}_R \cdot \mathbf{x} - \omega_R t)}$$

where we've allowed for the possibility that the amplitude, frequency, wavevector and polarisation all may change. We will write the reflected wavevector as

$$\mathbf{k}_R = -k_R \cos \theta_R \hat{\mathbf{x}} + k_R \sin \theta_R \hat{\mathbf{z}}$$

Meanwhile, the part of the wave that passes through the interface and into the second region is the transmitted wave which takes the form,

$$\mathbf{E}_{\text{trans}} = \mathbf{E}_T e^{i(\mathbf{k}_T \cdot \mathbf{x} - \omega_T t)}$$

with

$$\mathbf{k}_T = k_T \cos \theta_T \hat{\mathbf{x}} + k_T \sin \theta_T \hat{\mathbf{z}} \quad (7.20)$$

Again, we've allowed for the possibility that all the different properties of the wave could differ from the incoming wave. The electric field then takes the general form,

$$\mathbf{E} = \begin{cases} \mathbf{E}_{\text{inc}} + \mathbf{E}_{\text{ref}} & x < 0 \\ \mathbf{E}_{\text{trans}} & x > 0 \end{cases}$$

All of this is summarised in the figure.

We want to impose the matching conditions (7.16), (7.18), (7.19) and (7.17), with no surface charges and no surface currents. To start, we need the phase factors to be equal for all time. This means that we must have

$$\omega_I = \omega_R = \omega_T \quad (7.21)$$

and

$$\mathbf{k}_I \cdot \mathbf{x} = \mathbf{k}_R \cdot \mathbf{x} = \mathbf{k}_T \cdot \mathbf{x} \quad \text{at } x = 0 \quad (7.22)$$

This latter condition tells us that all of the wavevectors lie in the (x, z) -plane because \mathbf{k}_I originally lay in this plane. It further imposes the equality of the $\hat{\mathbf{z}}$ components of the wavevectors:

$$k_I \sin \theta_I = k_R \sin \theta_R = k_T \sin \theta_T \quad (7.23)$$

But, in each region, the frequency and wavenumbers are related, through the dispersion relation, to the speed of the wave. In region 1, we have $\omega_I = v_1 k_I$ and $\omega_R = v_1 k_R$ which, using (7.21) and (7.23), tells us that

$$\theta_I = \theta_R$$

This is the familiar law of reflection.

Meanwhile, in region 2 we have $\omega_T = v_2 k_T$. Now (7.21) and (7.23) tell us that

$$\frac{\sin \theta_I}{v_1} = \frac{\sin \theta_T}{v_2}$$

In terms of the refractive index $n = c/v$, this reads

$$n_1 \sin \theta_I = n_2 \sin \theta_T \quad (7.24)$$

This is the law of refraction, known as *Snell's law*.

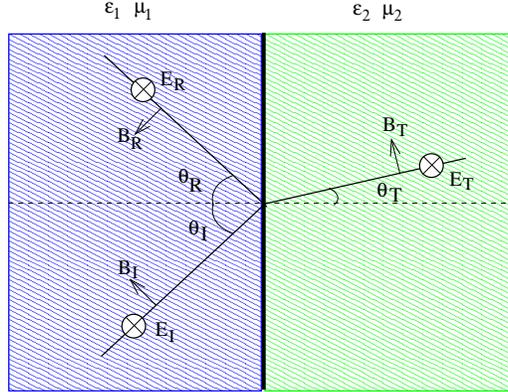


Figure 66: Incident, reflected and transmitted waves with normal polarisation.

7.4.1 Fresnel Equations

There's more information to be extracted from this calculation: we can look at the amplitudes of the reflected and transmitted waves. As we now show, this depends on the polarisation of the incident wave. There are two cases:

Normal Polarisation:

When the direction of $\mathbf{E}_I = E_I \hat{\mathbf{y}}$ is normal to the (x, z) -plane of incidence, it's simple to check that the electric polarisation of the other waves must lie in the same direction: $\mathbf{E}_R = E_T \hat{\mathbf{y}}$ and $\mathbf{E}_T = E_T \hat{\mathbf{y}}$. This situation, shown in Figure 66, is sometimes referred to as *s-polarised* (because the German word for normal begins with s).

The matching condition (7.18) requires

$$E_I + E_R = E_T$$

Meanwhile, as we saw in (7.16), the magnetic fields are given by $\mathbf{B} = (\hat{\mathbf{k}} \times \mathbf{E})/v$. The matching condition (7.19) then tells us that

$$B_I \cos \theta_I - B_R \cos \theta_R = B_T \cos \theta_T \quad \Rightarrow \quad \frac{E_I - E_R}{v_1} \cos \theta_I = \frac{E_T}{v_2} \cos \theta_T$$

With a little algebra, we can massage these conditions into the expressions,

$$\frac{E_R}{E_I} = \frac{n_1 \cos \theta_I - n_2 \cos \theta_T}{n_1 \cos \theta_I + n_2 \cos \theta_T} \quad \text{and} \quad \frac{E_T}{E_I} = \frac{2n_1 \cos \theta_I}{n_1 \cos \theta_I + n_2 \cos \theta_T} \quad (7.25)$$

These are the *Fresnel equations* for normal polarised light. We can then use Snell's law (7.24) to get the amplitudes in terms of the refractive indices and the incident angle θ_I .

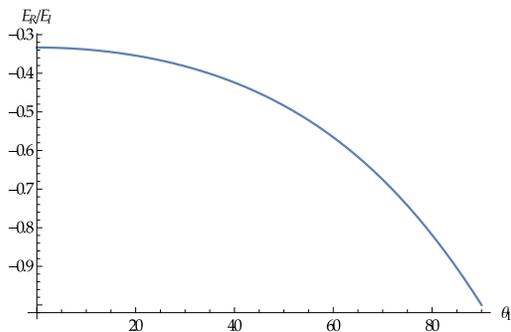


Figure 67: The reflected field with normal polarisation

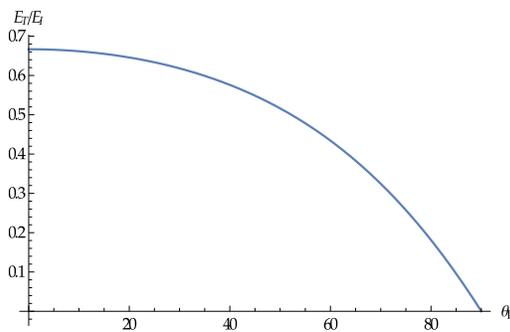


Figure 68: The transmitted field with normal polarisation

The most common example is if region 1 contains only air, with $n_1 \approx 1$, and region 2 consists of some transparent material. (For example, glass which has $n_2 \approx 1.5$). The normalised reflected and transmitted fields are plotted in the figures above for $n_1 = 1$ and $n_2 = 2$, with θ_I plotted in degrees along the horizontal axis). Note that the vertical axes are different; negative for the reflected wave, positive for the transmitted wave. In particular, when $\theta = 90^\circ$, the whole wave is reflected and nothing is transmitted.

Parallel Polarisation:

The case in which the electric field lies within the (x, z) -plane of incidence is sometimes referred to as *p-polarised* (because the English word for parallel begins with p). It is shown in Figure 69. Of course, we still require $\mathbf{E}_I \cdot \mathbf{k} = 0$, which means that

$$\mathbf{E}_I = -E_I \sin \theta_I \hat{\mathbf{x}} + E_I \cos \theta_I \hat{\mathbf{z}}$$

with similar expressions for \mathbf{E}_R and \mathbf{E}_T . The magnetic field now lies in the $\pm \hat{\mathbf{y}}$ direction. The matching condition (7.18) equates the components of the electric field tangential to the surface. This means

$$E_I \cos \theta_I + E_R \cos \theta_R = E_T \cos \theta_T$$

while the matching condition (7.19) for the components of magnetic field tangent to the surface gives

$$B_I - B_R = B_T \quad \Rightarrow \quad \frac{E_I - E_R}{v_1} = \frac{E_T}{v_2}$$

where the minus sign for B_R can be traced to the fact that the direction of the \mathbf{B} field (relative to \mathbf{k}) points in the opposite direction after a reflection. These two conditions can be written as

$$\frac{E_R}{E_I} = \frac{n_1 \cos \theta_T - n_2 \cos \theta_I}{n_1 \cos \theta_T + n_2 \cos \theta_I} \quad \text{and} \quad \frac{E_T}{E_I} = \frac{2n_1 \cos \theta_I}{n_1 \cos \theta_T + n_2 \cos \theta_I} \quad (7.26)$$

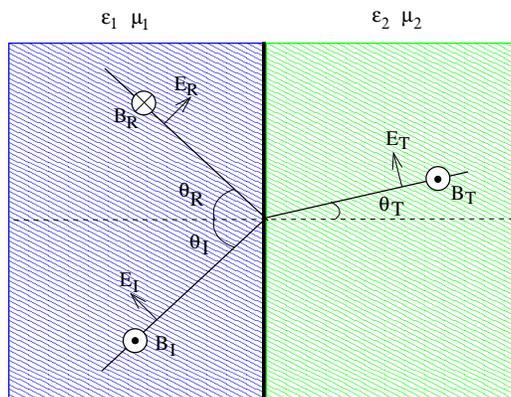


Figure 69: Incident, reflected and transmitted waves with parallel polarisation.

These are the *Fresnel equations* for parallel polarised light. Note that when the incident wave is normal to the surface, so both $\theta_I = \theta_T = 0$, the amplitudes for the normal (7.25) and parallel (7.26) polarisations coincide. But in general, they are different.

We can again plot the reflected and transmitted amplitudes in the case $n_1 = 1$ and $n_2 = 2$, shown in the figure.

Brewster's Angle

We can see from the left-hand figure that something interesting happens in the case of parallel polarisation. There is an angle for which there is no reflected wave. Everything gets transmitted. This is called the *Brewster Angle*, θ_B . It occurs when $n_1 \cos \theta_T = n_2 \cos \theta_I$. Of course, we also need to obey Snell's law (7.24). These two conditions are only satisfied when $\theta_I + \theta_T = \pi/2$. The Brewster angle is given by

$$\tan \theta_B = \frac{n_2}{n_1}$$

For the transmission of waves from air to glass, $\theta_B \approx 56^\circ$.

Brewster's angle gives a simple way to create polarised light: shine unpolarised light on a dielectric at angle θ_B and the only thing that bounces back has normal polarisation. This is the way sunglasses work to block out polarised light from the Sun. It is also the way polarising filters work.

7.4.2 Total Internal Reflection

Let's return to Snell's law (7.24) that tells us the angle of refraction,

$$\sin \theta_T = \frac{n_1}{n_2} \sin \theta_I$$

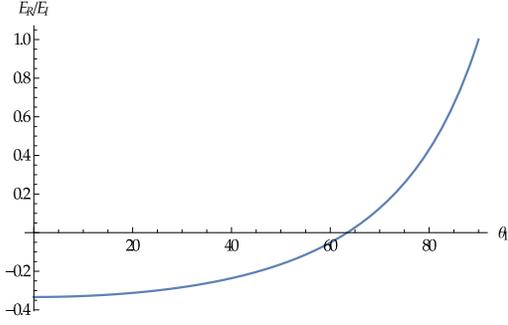


Figure 70: The reflected field with parallel polarisation

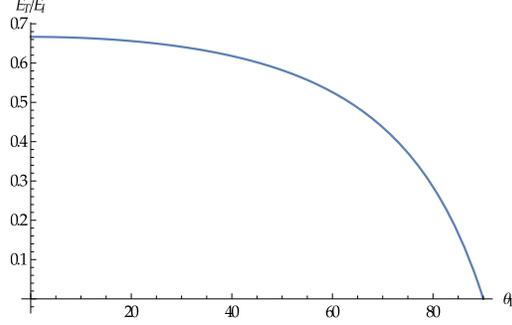


Figure 71: The transmitted field with parallel polarisation

But there's a problem with this equation: if $n_2 > n_1$ then the right-hand side can be greater than one, in which case there are no solutions. This happens at the critical angle of incidence, θ_C , defined by

$$\sin \theta_C = \frac{n_2}{n_1}$$

For example, if light is moving from glass, into air, then $\theta_C \approx 42^\circ$. At this angle, and beyond, there is no transmitted wave. Everything is reflected. This is called *total internal reflection*. It's what makes diamonds sparkle and makes optical fibres to work.

Here our interest is not in jewellery, but rather in a theoretical puzzle about how total internal reflection can be consistent. After all, we've computed the amplitude of the transmitted electric field in (7.25) and (7.26) and it's simple to check that it doesn't vanish when $\theta_I = \theta_C$. What's going on?

The answer lies back in our expression for the transmitted wavevector \mathbf{k}_T which we decomposed in (7.20) using geometry. The matching condition (7.22) tells us that $\mathbf{k}_T \cdot \hat{\mathbf{y}} = 0$ and

$$\mathbf{k}_T \cdot \hat{\mathbf{z}} = \mathbf{k}_I \cdot \hat{\mathbf{z}} = \frac{\omega_I}{v_1} \sin \theta_I$$

But, from the matching of frequencies (7.21), we know that $\omega_I = \omega_T \equiv \omega$. We also know that the magnitude of the transmitted wavevector is given by $|\mathbf{k}_T|^2 = \omega^2/v_2^2$. But this means that the component of the wavevector in the $\hat{\mathbf{x}}$ direction of propagation must be

$$\mathbf{k}_T \cdot \hat{\mathbf{x}} = \pm \sqrt{|\mathbf{k}_T|^2 - (\mathbf{k}_T \cdot \hat{\mathbf{z}})^2} = \pm \frac{\omega}{v_2} \sqrt{1 - \frac{v_2^2 \sin^2 \theta_I}{v_1^2}} = \pm \frac{\omega}{v_2} \sqrt{1 - \frac{n_1^2 \sin^2 \theta_I}{n_2^2}}$$

We see that when $n_1 \sin \theta_I / n_2 > 1$, the $\hat{\mathbf{x}}$ component of the wavevector is imaginary! We'll write $\mathbf{k}_T \cdot \hat{\mathbf{x}} = \pm i\omega\alpha/v_2$. An imaginary wavevector sounds strange, but it's very simple to interpret: we simply substitute it into our wave solution to find

$$\mathbf{E}_{\text{trans}} = \mathbf{E}_T e^{(i\mathbf{k}_T \cdot \hat{\mathbf{z}} - \omega t)} e^{\mp \omega\alpha x/v_2} \quad x > 0$$

Picking the minus sign in the exponent gives the physically sensible solution which decays as we move into region 2. We see that beyond the critical angle θ_C , there is no propagating wave in region 2. Instead it is replaced by a decaying solution. This is called an *evanescent wave*.

As we'll now see, the idea that the wavevector can be imaginary is very useful in a many other circumstances.

7.5 Dispersion

The dielectric constant $\epsilon_r = \epsilon/\epsilon_0$ is poorly named. It is not constant. This is because, in the presence of time-dependent electric fields, the permittivity typically depends on the frequency: $\epsilon = \epsilon(\omega)$. In this section, we will first provide a simple model to understand why this is the case and what form of $\epsilon(\omega)$ we should expect. We'll then move on to see the consequences of this frequency dependence.

7.5.1 Atomic Polarizability Revisited

In Section 7.1, we introduced a simple model for electric polarizability. This treats the atom as a point-like nucleus with charge q , surrounded by a cloud of electrons which we treat as a solid ball of radius a with uniform charge density. It's obviously a daft model for the atom, but it will do for our purposes.

Suppose that the centre of the electron cloud is displaced by a distance r . (You can equivalently think of the nucleus as displaced by the same distance in the opposite direction). We previously computed the restoring force (7.2) which acts on cloud,

$$\mathbf{F}_{\text{cloud}} = -\frac{q^2}{4\pi\epsilon_0 a^3} \mathbf{r} = -m\omega_0^2 \mathbf{r}$$

In the final equality, we've introduced the mass m of the cloud and defined the quantity ω_0 which we will call the *resonant frequency*.

In Section 7.1, we just looked at the equilibrium configuration of the electron cloud. Here, instead, we want to subject the atom to a time-dependent electric field $\mathbf{E}(t)$. In this situation, the electron cloud also feels a damping force

$$\mathbf{F}_{\text{damping}} = -m\gamma \dot{\mathbf{r}} \tag{7.27}$$

for some constant coefficient γ . You might find it strange to see such a friction term occurring for an atomic system. After all, we usually learn that friction is the effect of averaging over many many atoms. The purpose of this term is to capture the fact that the atom can lose energy, either to surrounding atoms or emitted electromagnetic radiation. If we now apply a time dependent electric field $\mathbf{E}(t)$ to this atom, the equation of motion for the displacement is

$$m\ddot{\mathbf{r}} = -q\mathbf{E}(t) - m\omega_0^2\mathbf{r} + m\gamma\dot{\mathbf{r}} \quad (7.28)$$

Solutions to this describe the atomic cloud oscillating about the nucleus.

The time dependent electric field will be of the wave form that we've seen throughout these lectures: $\mathbf{E} = \mathbf{E}_0 e^{i(\mathbf{k}\cdot\mathbf{r} - \omega t)}$. However, the atom is tiny. In particular, it is small compared to the wavelength of (at least) visible light, meaning $ka \ll 1$. For this reason, we can ignore the fact that the phase oscillates in space and work with an electric field of the form $\mathbf{E}(t) = \mathbf{E}_0 e^{-i\omega t}$. Then (7.28) is the equation for a forced, damped harmonic oscillator. We search for solutions to (7.28) of the form $\mathbf{r}(t) = \mathbf{r}_0 e^{-i\omega t}$. (In the end we will take the real part). The solution is

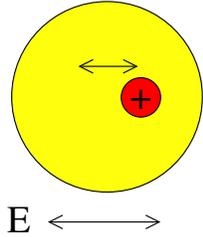


Figure 72:

$$\mathbf{r}_0 = -\frac{q\mathbf{E}_0}{m} \frac{1}{-\omega^2 + \omega_0^2 - i\gamma\omega}$$

This gives the atomic polarisability $\mathbf{p} = \alpha\mathbf{E}$, where

$$\alpha = \frac{q^2/m}{-\omega^2 + \omega_0^2 - i\gamma\omega} \quad (7.29)$$

As promised, the polarisability depends on the frequency. Moreover, it is also complex. This has the effect that the polarisation of the atom is not in phase with the oscillating electric field.

Because the polarisability is both frequency dependent and complex, the permittivity $\epsilon(\omega)$ will also be both frequency dependent and complex. (In the simplest settings, they are related by $\epsilon(\omega) = \epsilon_0 + n\alpha(\omega)$ where n is the density of atoms). We'll now see the effect this has on the propagation of electromagnetic waves through materials.

7.5.2 Electromagnetic Waves Revisited

To start, we'll consider a general form of the permittivity $\epsilon(\omega)$ which is both frequency dependent and complex; we'll return to the specific form arising from the polarisability

(7.29) later. In contrast, we will assume that the magnetic thing μ is both constant and real, which turns out to be a good approximation for most materials. This means that we have

$$\mathbf{D} = \epsilon(\omega)\mathbf{E} \quad \text{and} \quad \mathbf{B} = \mu\mathbf{H}$$

We'll look for plane wave solutions, so that the electric and magnetic fields takes the form

$$\mathbf{E}(\mathbf{x}, t) = \mathbf{E}(\omega) e^{i(\mathbf{k}\cdot\mathbf{x} - \omega t)} \quad \text{and} \quad \mathbf{B}(\mathbf{x}, t) = \mathbf{B}(\omega) e^{i(\mathbf{k}\cdot\mathbf{x} - \omega t)}$$

Maxwell's equations in matter were given in (7.14). The first two simply tell us

$$\begin{aligned} \nabla \cdot \mathbf{D} = 0 &\quad \Rightarrow \quad \epsilon(\omega)\mathbf{k} \cdot \mathbf{E}(\omega) = 0 \\ \nabla \cdot \mathbf{B} = 0 &\quad \Rightarrow \quad \mathbf{k} \cdot \mathbf{B}(\omega) = 0 \end{aligned}$$

These are the statements that the electric and magnetic fields remain transverse to the direction of propagation. (In fact there's a caveat here: if $\epsilon(\omega) = 0$ for some frequency ω , then the electric field need not be transverse. This won't affect our discussion here, but we will see an example of this when we turn to conductors in Section 7.6). Meanwhile, the other two equations are

$$\begin{aligned} \nabla \times \mathbf{H} = \frac{\partial \mathbf{D}}{\partial t} &\quad \Rightarrow \quad \mathbf{k} \times \mathbf{B}(\omega) = -\mu\epsilon(\omega)\omega\mathbf{E}(\omega) \\ \nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} &\quad \Rightarrow \quad \mathbf{k} \times \mathbf{E}(\omega) = \omega\mathbf{B}(\omega) \end{aligned} \quad (7.30)$$

We do the same manipulation that we've seen before: look at $\mathbf{k} \times (\mathbf{k} \times \mathbf{E})$ and use the fact that $\mathbf{k} \cdot \mathbf{E} = 0$. This gives us the *dispersion* relation

$$\mathbf{k} \cdot \mathbf{k} = \mu\epsilon(\omega)\omega^2 \quad (7.31)$$

We need to understand what this equation is telling us. In particular, $\epsilon(\omega)$ is typically complex. This, in turn, means that the wavevector \mathbf{k} will also be complex. To be specific, we'll look at waves propagating in the z -direction and write $\mathbf{k} = k\hat{\mathbf{z}}$. We'll write the real and imaginary parts as

$$\epsilon(\omega) = \epsilon_1(\omega) + i\epsilon_2(\omega) \quad \text{and} \quad k = k_1 + ik_2$$

Then the dispersion relation reads

$$k_1 + ik_2 = \omega\sqrt{\mu}\sqrt{\epsilon_1 + i\epsilon_2} \quad (7.32)$$

and the electric field takes the form

$$\mathbf{E}(\mathbf{x}, t) = \mathbf{E}(\omega) e^{-k_2 z} e^{i(k_1 z - \omega t)} \quad (7.33)$$

We now see the consequence of the imaginary part of $\epsilon(\omega)$; it causes the amplitude of the wave to decay as it extends in the z -direction. This is also called *attenuation*. The real part, k_1 , determines the oscillating part of the wave. The fact that ϵ depends on ω means that waves of different frequencies travel with different speeds. We'll discuss shortly the ways of characterising these speeds.

The magnetic field is

$$\mathbf{B}(\omega) = \frac{k}{\omega} \hat{\mathbf{z}} \times \mathbf{E}(\omega) = \frac{|k| e^{i\phi}}{\omega} \hat{\mathbf{z}} \times \mathbf{E}(\omega)$$

where $\phi = \tan^{-1}(k_2/k_1)$ is the phase of the complex wavenumber k . This is the second consequence of a complex permittivity $\epsilon(\omega)$; it results in the electric and magnetic fields oscillating out of phase. The profile of the magnetic field is

$$\mathbf{B}(\mathbf{x}, t) = \frac{|k|}{\omega} (\hat{\mathbf{z}} \times \mathbf{E}(\omega)) e^{-k_2 z} e^{i(k_1 z - \omega t + \phi)} \quad (7.34)$$

As always, the physical fields are simply the real parts of (7.33) and (7.34), namely

$$\begin{aligned} \mathbf{E}(\mathbf{x}, t) &= \mathbf{E}(\omega) e^{-k_2 z} \cos(k_1 z - \omega t) \\ \mathbf{B}(\mathbf{x}, t) &= \frac{|k|}{\omega} (\hat{\mathbf{z}} \times \mathbf{E}(\omega)) e^{-k_2 z} \cos(k_1 z - \omega t + \phi) \end{aligned}$$

To recap: the imaginary part of ϵ means that $k_2 \neq 0$. This has two effects: it leads to the damping of the fields, and to the phase shift between \mathbf{E} and \mathbf{B} .

Measures of Velocity

The other new feature of $\epsilon(\omega)$ is that it depends on the frequency ω . The dispersion relation (7.31) then immediately tells us that waves of different frequencies travel at different speeds. There are two, useful characterisations of these speeds. The *phase velocity* is defined as

$$v_p = \frac{\omega}{k_1}$$

As we can see from (7.33) and (7.34), a wave of a fixed frequency ω propagates with phase velocity $v_p(\omega)$.

Waves of different frequency will travel with different phase velocities v_p . This means that for wave pulses, which consist of many different frequencies, different parts of the wave will travel with different speeds. This will typically result in a change of shape of the pulse as it moves along. We'd like to find a way to characterise the speed of the whole pulse. The usual measure is the *group velocity*, defined as

$$v_g = \frac{d\omega}{dk_1}$$

where we've inverted (7.31) so that we're now viewing frequency as a function of (real) wavenumber: $\omega(k_1)$.

To see why the group velocity is a good measure of the speed, let's build a pulse by superposing lots of waves of different frequencies. To make life simple, we'll briefly set $\epsilon_2 = 0$ and $k_1 = k$ for now so that we don't have to think about damping effects. Then, focussing on the electric field, we can build a pulse by writing

$$\mathbf{E}(\mathbf{x}, t) = \int \frac{dk}{2\pi} \mathbf{E}(k) e^{i(kz - \omega t)}$$

Suppose that our choice of wavepacket $\mathbf{E}(k)$ is heavily peaked around some fixed wavenumber k_0 . Then we can expand the exponent as

$$\begin{aligned} kz - \omega(k)t &\approx kz - \omega(k_0)t - \left. \frac{d\omega}{dk} \right|_{k_0} (k - k_0)t \\ &= -[\omega(k_0) + v_g(k_0)]t + k[z - v_g(k_0)t] \end{aligned}$$

The first term is just a constant oscillation in time; the second, k -dependent term is the one of interest. It tells us that the peak of the wave pulse is moving to the right with approximate speed $v_g(k_0)$.

Following (7.15), we also define the index of refraction

$$n(\omega) = \frac{c}{v_p(\omega)}$$

This allows us to write a relation between the group and phase velocities:

$$\frac{1}{v_g} = \frac{dk_1}{d\omega} = \frac{d}{d\omega} \left(\frac{n\omega}{c} \right) = \frac{1}{v_p} + \frac{\omega}{c} \frac{dn}{d\omega}$$

Materials with $dn/d\omega > 0$ have $v_g < v_p$; this is called *normal dispersion*. Materials with $dn/d\omega < 0$ have $v_g > v_p$; this is called *anomalous dispersion*.

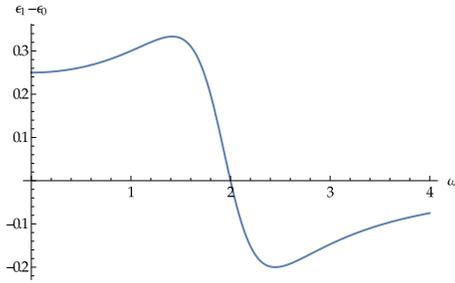


Figure 73: The real part of the permittivity, $\epsilon_1 - \epsilon_0$

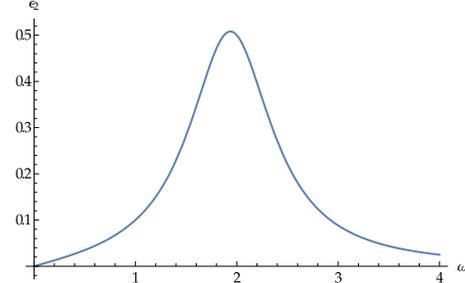


Figure 74: The imaginary part of the permittivity, ϵ_2

7.5.3 A Model for Dispersion

Let's see how this story works for our simple model of atomic polarisability $\alpha(\omega)$ given in (7.29). The permittivity is $\epsilon(\omega) = \epsilon_0 + n\alpha(\omega)$ where n is the density of atoms. The real and imaginary parts $\epsilon = \epsilon_1 + i\epsilon_2$ are

$$\epsilon_1 = \epsilon_0 - \frac{nq^2}{m} \frac{\omega^2 - \omega_0^2}{(\omega^2 - \omega_0^2)^2 + \gamma^2\omega^2}$$

$$\epsilon_2 = \frac{nq^2}{m} \frac{\gamma\omega}{(\omega^2 - \omega_0^2)^2 + \gamma^2\omega^2}$$

These functions look like this: (These particular plots are made with $\gamma = 1$ and $\omega_0 = 2$ and $nq^2/m = 1$).

The real part is an even function: it has a maximum at $\omega = \omega_0 - \gamma/2$ and a minimum at $\omega = \omega_0 + \gamma/2$, each offset from the resonant frequency by an amount proportional to the damping γ . The imaginary part is an odd function; it has a maximum at $\omega = \omega_0$, the resonant frequency of the atom. The width of the imaginary part is roughly $\gamma/2$.

A quantity that will prove important later is the *plasma frequency*, ω_p . This is defined as

$$\omega_p^2 = \frac{nq^2}{m\epsilon_0} \tag{7.35}$$

We'll see the relevance of this quantity in Section 7.6. But for now it will simply be a useful combination that appears in some formulae below.

The dispersion relation (7.32) tells us

$$k_1^2 - k_2^2 + 2ik_1k_2 = \omega^2\mu(\epsilon_1 + i\epsilon_2)$$

Equating real and imaginary parts, we have

$$\begin{aligned} k_1 &= \pm\omega\sqrt{\mu}\left(\frac{1}{2}\sqrt{\epsilon_1^2 + \epsilon_2^2} + \frac{1}{2}\epsilon_1\right)^{1/2} \\ k_2 &= \pm\omega\sqrt{\mu}\left(\frac{1}{2}\sqrt{\epsilon_1^2 + \epsilon_2^2} - \frac{1}{2}\epsilon_1\right)^{1/2} \end{aligned} \quad (7.36)$$

To understand how light propagates through the material, we need to look at the values of k_1 and k_2 for different values of the frequency. There are three different types of behaviour.

Transparent Propagation: Very high or very low frequencies

The most straightforward physics happens when $\epsilon_1 > 0$ and $\epsilon_1 \gg \epsilon_2$. For our simple model, this occurs when $\omega < \omega_0 - \gamma/2$ or when $\omega > \omega_*$, the value at which $\epsilon_1(\omega_*) = 0$.

Expanding to leading order, we have

$$k_1 \approx \pm\omega\sqrt{\mu\epsilon_1} \quad \text{and} \quad k_2 \approx \pm\omega\sqrt{\frac{\mu\epsilon_2^2}{4\epsilon_1}} = \left(\frac{\epsilon_2}{2\epsilon_1}\right)k_1 \ll k_1$$

Because $k_2 \ll k_1$, the damping is small. This means that the material is transparent at these frequencies.

There's more to this story. For the low frequencies, $\epsilon_1 > \epsilon_0 + nq^2/m\omega_0^2$. This is the same kind of situation that we dealt with in Section 7.3. The phase velocity $v_p < c$ in this regime. For high frequencies, however, $\epsilon_1 < \epsilon_0$; in fact, $\epsilon_1(\omega) \rightarrow \epsilon_0$ from below as $\omega \rightarrow \infty$. This means that $v_p > c$ in this region. This is nothing to be scared of! The plane wave is already spread throughout space; it's not communicating any information faster than light. Instead, pulses propagate at the group velocity, v_g . This is less than the speed of light, $v_g < c$, in both high and low frequency regimes.

Resonant Absorption: $\omega \approx \omega_0$

Resonant absorption occurs when $\epsilon_2 \gg |\epsilon_1|$. In our model, this phenomenon is most pronounced when $\omega_0 \gg \gamma$ so that the resonant peak of ϵ_2 is sharp. Then for frequencies close to the resonance, $\omega \approx \omega_0 \pm \gamma/2$, we have

$$\epsilon_1 \approx \epsilon_0 \quad \text{and} \quad \epsilon_2 \approx \frac{nq^2}{m} \frac{1}{\omega_0\gamma} = \epsilon_0 \left(\frac{\omega_p}{\omega_0}\right)^2 \frac{\omega_0}{\gamma}$$

We see that we meet the requirement for resonant absorption if we also have $\omega_p \gtrsim \omega_0$. When $\epsilon_2 \gg |\epsilon_1|$, we can expand (7.36) to find

$$k_1 \approx k_2 \approx \pm\omega\sqrt{\frac{\mu\epsilon_2}{2}}$$

The fact that $k_2 \approx k_1$ means that the wave decays very rapidly: it has effectively disappeared within just a few wavelengths of propagation. This is because the frequency of the wave is tuned to coincide with the natural frequency of the atoms, which easily become excited, absorbing energy from the wave.

Total Reflection:

The third region of interest occurs when $\epsilon_1 < 0$ and $|\epsilon_1| \gg \epsilon_2$. In our model, it is roughly for frequencies $\omega_0 + \gamma/2 < \omega < \omega_*$. Now, the expansion of (7.36) gives

$$k_1 \approx \pm\omega\sqrt{\mu}\left(\frac{1}{2}|\epsilon_1| + \frac{1}{4}\frac{\epsilon_2^2}{|\epsilon_1|} + \frac{1}{2}\epsilon_1 + \dots\right)^{1/2} \approx \pm\omega\frac{\epsilon_2}{2}\sqrt{\frac{\mu}{|\epsilon_1|}}$$

and

$$k_2 \approx \pm\omega\sqrt{\mu|\epsilon_1|} = \frac{|\epsilon_1|}{2\epsilon_2}k_1 \gg k_1$$

Now the wavenumber is almost pure imaginary. The wave doesn't even manage to get a few wavelengths before it decays. It's almost all gone before it even travels a single wavelength.

We're not tuned to the resonant frequency, so this time the wave isn't being absorbed by the atoms. Instead, the applied electromagnetic field is almost entirely cancelled out by the induced electric and magnetic fields due to polarisation.

7.5.4 Causality and the Kramers-Kronig Relation

Throughout this section, we used the relationship between the polarisation \mathbf{p} and applied electric field \mathbf{E} . In frequency space, this reads

$$\mathbf{p}(\omega) = \alpha(\omega)\mathbf{E}(\omega) \tag{7.37}$$

Relationships of this kind appear in many places in physics. The polarisability $\alpha(\omega)$ is an example of a *response function*. As their name suggests, such functions tell us how some object – in this case \mathbf{p} – respond to a change in circumstance – in this case, the application of an electric field.

There is a general theory around the properties of response functions⁵. The most important fact follows from causality. The basic idea is that if we start off with a vanishing electric field and turn it on only at some fixed time, t_* , then the polarisation shouldn't respond to this until after t_* . This sounds obvious. But how is it encoded in the mathematics?

The causality properties are somewhat hidden in (7.37) because we're thinking of the electric field as oscillating at some fixed frequency, which implicitly means that it oscillates for all time. If we want to turn the electric field on and off in time then we need to think about superposing fields of lots of different frequencies. This, of course, is the essence of the Fourier transform. If we shake the electric field at lots of different frequencies, its time dependence is given by

$$\mathbf{E}(t) = \int_{-\infty}^{+\infty} \frac{d\omega}{2\pi} \mathbf{E}(\omega) e^{-i\omega t}$$

where, if we want $\mathbf{E}(t)$ to be real, we should take $\mathbf{E}(-\omega) = \mathbf{E}(\omega)^*$. Conversely, for a given time dependence of the electric field, the component at some frequency ω is given by the inverse Fourier transform,

$$\mathbf{E}(\omega) = \int_{-\infty}^{+\infty} dt \mathbf{E}(t) e^{i\omega t}$$

Let's now see what this tells us about the time dependence of the polarisation \mathbf{p} . Using (7.37), we have

$$\begin{aligned} \mathbf{p}(t) &= \int_{-\infty}^{+\infty} \frac{d\omega}{2\pi} \mathbf{p}(\omega) e^{-i\omega t} \\ &= \int_{-\infty}^{+\infty} \frac{d\omega}{2\pi} \alpha(\omega) \int_{-\infty}^{+\infty} dt' \mathbf{E}(t') e^{-i\omega(t-t')} \\ &= \int_{-\infty}^{+\infty} dt' \tilde{\alpha}(t-t') \mathbf{E}(t') \end{aligned} \tag{7.38}$$

where, in the final line, we've introduced the Fourier transform of the polarisability,

$$\tilde{\alpha}(t) = \int_{-\infty}^{+\infty} \frac{d\omega}{2\pi} \alpha(\omega) e^{-i\omega t} \tag{7.39}$$

(Note that I've been marginally inconsistent in my notation here. I've added the tilde above $\tilde{\alpha}$ to stress that this is the Fourier transform of $\alpha(\omega)$ even though I didn't do the same to \mathbf{p} and \mathbf{E}).

⁵You can learn more about this in the *Response Functions* section of the lectures on *Kinetic Theory*.

Equation (7.38) relates the time dependence of \mathbf{p} to the time dependence of the electric field \mathbf{E} . It's telling us that the effect isn't immediate; the polarisation at time t depends on what the electric field was doing at all times t' . But now we can state the requirement of causality: the response function must obey

$$\tilde{\alpha}(t) = 0 \quad \text{for } t < 0$$

Using (7.39), we can translate this back into a statement about the response function in frequency space. When $t < 0$, we can perform the integral over ω by completing the contour in the upper-half plane as shown in the figure. Along the extra semi-circle, the exponent is $-i\omega t \rightarrow -\infty$ for $t < 0$, ensuring that this part of the integral vanishes. By the residue theorem, the integral is just given by the sum of residues inside the contour. If we want $\alpha(t) = 0$ for $t < 0$, we need there to be no poles. In other words, we learn that

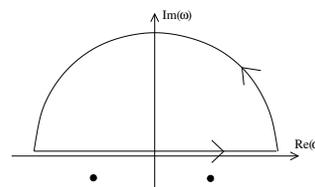


Figure 75:

$$\alpha(\omega) \text{ is analytic for } \text{Im } \omega > 0$$

In contrast, $\alpha(\omega)$ can have poles in the lower-half imaginary plane. For example, if you look at our expression for the polarisability in (7.29), you can see that there are two poles at $\omega = -i\gamma/2 \pm \sqrt{\omega_0^2 - \gamma^2/4}$. Both lie in the lower-half of the complex ω plane.

The fact that α is analytic in the upper-half plane means that there is a relationship between its real and imaginary parts. This is called the Kramers-Kronig relation. Our task in this section is to derive it. We start by providing a few general mathematical statements about complex integrals.

A Discontinuous Function

First, consider a general function $\rho(\omega)$. We'll ask that $\rho(\omega)$ is meromorphic, meaning that it is analytic apart from at isolated poles. But, for now, we won't place any restrictions on the position of these poles. (We will shortly replace $\rho(\omega)$ by $\alpha(\omega)$ which, as we've just seen, has no poles in the upper half plane). We can define a new function $f(\omega)$ by the integral,

$$f(\omega) = \frac{1}{i\pi} \int_a^b \frac{\rho(\omega')}{\omega' - \omega} d\omega' \quad (7.40)$$

Here the integral is taken along the interval $\omega' \in [a, b]$ of the real line. However, when ω also lies in this interval, we have a problem because the integral diverges at $\omega' = \omega$.

To avoid this, we can simply deform the contour of the integral into the complex plane, either running just above the singularity along $\omega' + i\epsilon$ or just below the singularity along $\omega' - i\epsilon$. Alternatively (in fact, equivalently) we could just shift the position of the singularity to $\omega \rightarrow \omega \mp \epsilon$. In both cases we just skim by the singularity and the integral is well defined. The only problem is that we get different answers depending on which way we do things. Indeed, the difference between the two answers is given by Cauchy's residue theorem,

$$\frac{1}{2}[f(\omega + i\epsilon) - f(\omega - i\epsilon)] = \rho(\omega) \quad (7.41)$$

The difference between $f(\omega + i\epsilon)$ and $f(\omega - i\epsilon)$ means that the function $f(\omega)$ is discontinuous across the real axis for $\omega \in [a, b]$. If $\rho(\omega)$ is everywhere analytic, this discontinuity is a branch cut.

We can also define the average of the two functions either side of the discontinuity. This is usually called the *principal value*, and is denoted by adding the symbol \mathcal{P} before the integral,

$$\frac{1}{2}[f(\omega + i\epsilon) + f(\omega - i\epsilon)] \equiv \frac{1}{i\pi} \mathcal{P} \int_a^b \frac{\rho(\omega')}{\omega' - \omega} d\omega' \quad (7.42)$$

We can get a better handle on the meaning of this principal part if we look at the real and imaginary pieces of the denominator in the integrand $1/[\omega' - (\omega \pm i\epsilon)]$,

$$\frac{1}{\omega' - (\omega \pm i\epsilon)} = \frac{\omega' - \omega}{(\omega' - \omega)^2 + \epsilon^2} \pm \frac{i\epsilon}{(\omega' - \omega)^2 + \epsilon^2} \quad (7.43)$$

The real and imaginary parts of this function are shown in the figures.

We can isolate the real part by taking the sum of $f(\omega + i\epsilon)$ and $f(\omega - i\epsilon)$ in (7.42). It can be thought of as a suitably cut-off version of $1/(\omega' - \omega)$. It's as if we have deleted a small segment of this function lying symmetrically about divergent point ω and replaced it with a smooth function going through zero. This is the usual definition of the principal part of an integral.

Similarly, the imaginary part can be thought of as a regularised delta-function. As $\epsilon \rightarrow 0$, it tends towards a delta function, as expected from (7.41).

Kramers-Kronig

Let's now apply this discussion to our polarisability response function $\alpha(\omega)$. We'll be interested in the integral

$$\frac{1}{i\pi} \oint_C d\omega' \frac{\alpha(\omega')}{\omega' - \omega} \quad \omega \in \mathbf{R} \quad (7.44)$$

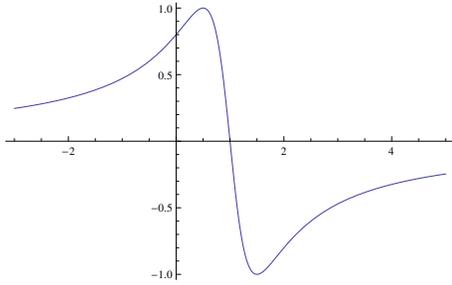


Figure 76: The real part of the function plotted with $\omega' = 1$ and $\epsilon = 0.5$.

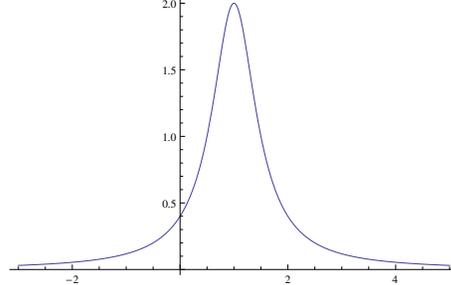


Figure 77: The imaginary part of the function plotted with $\omega' = 1$ and $\epsilon = 0.5$.

where the contour C skims just above the real axis, before closing at infinity in the upper-half plane. We'll need to make one additional assumption: that $\alpha(\omega)$ falls off faster than $1/|\omega|$ at infinity. If this holds, the integral is the same as we considered in (7.40) with $[a, b] \rightarrow [-\infty, +\infty]$. Indeed, in the language of the previous discussion, the integral is $f(\omega - i\epsilon)$, with $\rho = \alpha$.

We apply the formulae (7.41) and (7.42). It gives

$$f(\omega - i\epsilon) = \frac{1}{i\pi} \mathcal{P} \left[\int_{-\infty}^{+\infty} d\omega' \frac{\alpha(\omega')}{\omega' - \omega} \right] - \alpha(\omega)$$

But we know the integral in (7.44) has to be zero since $\alpha(\omega)$ has no poles in the upper-half plane. This means that $f(\omega - i\epsilon) = 0$, or

$$\alpha(\omega) = \frac{1}{i\pi} \mathcal{P} \int_{-\infty}^{+\infty} d\omega' \frac{\alpha(\omega')}{\omega' - \omega}$$

The important part for us is that factor of “ i ” sitting in the denominator. Taking real and imaginary parts, we learn that

$$\text{Re } \alpha(\omega) = \mathcal{P} \int_{-\infty}^{+\infty} \frac{d\omega'}{\pi} \frac{\text{Im } \alpha(\omega')}{\omega' - \omega}$$

and

$$\text{Im } \alpha(\omega) = -\mathcal{P} \int_{-\infty}^{+\infty} \frac{d\omega'}{\pi} \frac{\text{Re } \alpha(\omega')}{\omega' - \omega}$$

These are the *Kramers-Kronig* relations. They follow from causality alone and tell us that the imaginary part of the response function is determined in terms of the real part, and vice-versa. However, the relationship is not local in frequency space: you need to know $\text{Re } \alpha(\omega)$ for all frequencies in order to reconstruct $\text{Im } \alpha(\omega)$ for any single frequency.

7.6 Conductors Revisited

Until now, we've only discussed electromagnetic waves propagating through insulators. (Or, dielectrics to give them their fancy name). What happens in conductors where electric charges are free to move? We met a cheap model of a conductor in Section 2.4, where we described them as objects which screen electric fields. Here we'll do a slightly better job and understand how this happens dynamically.

7.6.1 The Drude Model

The Drude model is simple. Really simple. It describes the electrons moving in a conductor as billiard-balls, bouncing off things. The electrons have mass m , charge q and velocity $\mathbf{v} = \dot{\mathbf{r}}$. We treat them classically using $F = ma$; the equation of motion is

$$m \frac{d\mathbf{v}}{dt} = q\mathbf{E} - \frac{m}{\tau} \mathbf{v} \quad (7.45)$$

The force is due to an applied electric field \mathbf{E} , together with a linear friction term. This friction term captures the effect of electrons hitting things, whether the background lattice of fixed ions, impurities, or each other. (Really, these latter processes should be treated in the quantum theory but we'll stick with a classical treatment here). The coefficient τ is called the *scattering time*. It should be thought of as the average time that the electron travels before it bounces off something. For reference, in a good metal, $\tau \approx 10^{-14} s$. (Note that this friction term is the same as (7.27) that we wrote for the atomic polarisability, although the mechanisms behind it may be different in the two cases).

We start by applying an electric field which is constant in space but oscillating in time

$$\mathbf{E} = \mathbf{E}(\omega) e^{-i\omega t}$$

This can be thought of as applying an AC voltage to a conductor. We look for solutions of the form

$$\mathbf{v} = \mathbf{v}(\omega) e^{-i\omega t}$$

Plugging this into (7.45) gives

$$\left(-i\omega + \frac{1}{\tau}\right) \mathbf{v}(\omega) = \frac{q}{m} \mathbf{E}(\omega)$$

The current density is $\mathbf{J} = nq\mathbf{v}$, where n is the density of charge carriers, so the solution tells us that

$$\mathbf{J}(\omega) = \sigma(\omega) \mathbf{E}(\omega) \quad (7.46)$$

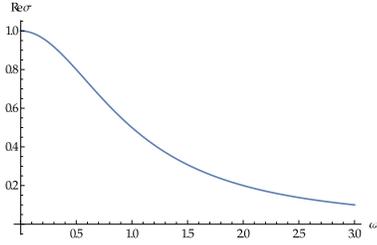


Figure 78: The real, dissipative part of the conductivity

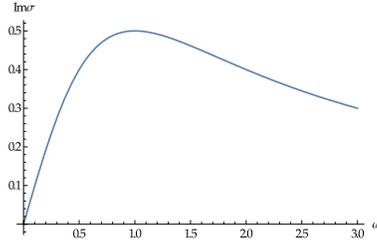


Figure 79: The imaginary, reactive part of the conductivity

This, of course, is Ohm’s law. The proportionality constant $\sigma(\omega)$ depends on the frequency and is given by

$$\sigma(\omega) = \frac{\sigma_{\text{DC}}}{1 - i\omega\tau} \quad (7.47)$$

It is usually referred to as the *optical conductivity*. In the limit of vanishing frequency, $\omega = 0$, it reduces to the *DC conductivity*,

$$\sigma_{\text{DC}} = \frac{nq^2\tau}{m}$$

The DC conductivity is real and is inversely related to the *resistivity* $\rho = 1/\sigma_{\text{DC}}$. In contrast, the optical conductivity is complex. Its real and imaginary parts are given by

$$\text{Re } \sigma(\omega) = \frac{\sigma_{\text{DC}}}{1 + \omega^2\tau^2} \quad \text{and} \quad \text{Im } \sigma(\omega) = \frac{\sigma_{\text{DC}} \omega\tau}{1 + \omega^2\tau^2}$$

These are plotted for $\sigma_{\text{DC}} = 1$ and $\tau = 1$:

The conductivity is complex simply because we’re working in Fourier space. The real part tells us about the dissipation of energy in the system. The bump at low frequencies, $\omega \sim 1/\tau$, is referred to as the *Drude peak*. The imaginary part of the conductivity tells us about the response of the system. (To see how this is relevant note that, in the Fourier ansatz, the velocity is related to the position by $\mathbf{v} = \dot{\mathbf{r}} = -i\omega\mathbf{r}$). At very large frequencies, $\omega\tau \gg 1$, the conductivity becomes almost purely imaginary, $\sigma(\omega) \sim i/\omega\tau$. This should be thought of as the conductivity of a free particle; you’re shaking it so fast that it turns around and goes the other way before it’s had the chance to hit something.

Although we derived our result (7.47) using a simple, Newtonian model of free electrons, the expression for the conductivity itself is surprisingly robust. In fact, it survives just about every subsequent revolution in physics; the development of quantum mechanics and Fermi surfaces, the presence of lattices and Bloch waves, even interactions between electrons in a framework known as Landau’s Fermi liquid model. In all of

these, the optical conductivity (7.47) remains the correct answer⁶. (This is true, at least, at low frequencies, At very high frequencies other effects can come in and change the story).

7.6.2 Electromagnetic Waves in Conductors

Let’s now ask our favourite question: how do electromagnetic waves move through a material? The macroscopic Maxwell equations (7.14) that we wrote before assumed that there were no free charges or currents around. Now we’re in a conductor, we need to include the charge density and current terms on the right-hand side:

$$\begin{aligned} \nabla \cdot \mathbf{D} = \rho \quad \text{and} \quad \nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} \\ \nabla \cdot \mathbf{B} = 0 \quad \text{and} \quad \nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \end{aligned} \quad (7.48)$$

It’s important to remember that here ρ refers only to the free charge. (We called it ρ_{free} in Section 7.1). We can still have bound charge in conductors, trapped around the ions of the lattice, but this has already been absorbed in the definition of \mathbf{D} which is given by

$$\mathbf{D} = \epsilon(\omega)\mathbf{E}$$

Similarly, the current \mathbf{J} is due only to the free charge.

We now apply a spatially varying, oscillating electromagnetic field, using the familiar ansatz,

$$\mathbf{E}(\mathbf{x}, t) = \mathbf{E}(\omega)e^{i(\mathbf{k}\cdot\mathbf{x}-\omega t)} \quad \text{and} \quad \mathbf{B}(\mathbf{x}, t) = \mathbf{B}(\omega)e^{i(\mathbf{k}\cdot\mathbf{x}-\omega t)} \quad (7.49)$$

At this point, we need to do something that isn’t obviously allowed: we will continue to use Ohm’s law (7.46), even in the presence of a varying electric field, so that

$$\mathbf{J}(\mathbf{x}, t) = \sigma(\omega)\mathbf{E}(\omega)e^{i(\mathbf{k}\cdot\mathbf{x}-\omega t)} \quad (7.50)$$

This looks dubious; we derived Ohm’s law by assuming that the electric field was the same everywhere in space. Why do we now get to use it when the electric field varies? For this to be valid, we need to assume that over the time scales τ , relevant in the

⁶As an extreme example, the conductivity of the horizon of certain black holes can be computed in general relativity. Even here, the result at low frequency is given by the simple Drude formula (7.47)! Details can be found in Gary Horowitz, Jorge Santos and David Tong, “*Optical Conductivity with Holographic Lattices*”, arXiv:1204.0519.

derivation of Ohm's law, the electric field is more or less constant. This will be true if the wavelength of the electric field, $\lambda = 2\pi/|\mathbf{k}|$ is greater than the distance travelled by the electrons between collisions. This distance, known as the *mean free path*, is given by $l = \langle v \rangle \tau$, where v is the average speed. In most metals, $l \approx 10^{-7} m$. (This is around 1000 lattice spacings; to understand how it can be so large requires a quantum treatment of the electrons). This means that we should be able to trust (7.50) for wavelengths $\lambda \gtrsim l \approx 10^{-7} m$, which is roughly around the visible spectrum.

The continuity equation $\nabla \cdot \mathbf{J} + d\rho/dt = 0$ tells us that if the current oscillates, then the charge density must as well. In Fourier space, the continuity equation becomes

$$\rho = \frac{\mathbf{k} \cdot \mathbf{J}}{\omega} = \frac{\sigma(\omega)}{\omega} \mathbf{k} \cdot \mathbf{E}(\omega) e^{i(\mathbf{k} \cdot \mathbf{x} - \omega t)} \quad (7.51)$$

We can now plug these ansätze into the Maxwell equations (7.48). We also need $\mathbf{B} = \mu \mathbf{H}$ where, as previously, we'll take μ to be independent of frequency. We have

$$\begin{aligned} \nabla \cdot \mathbf{D} = \rho &\Rightarrow i \left(\epsilon(\omega) + i \frac{\sigma(\omega)}{\omega} \right) \mathbf{k} \cdot \mathbf{E}(\omega) = 0 \\ \nabla \cdot \mathbf{B} = 0 &\Rightarrow \mathbf{k} \cdot \mathbf{B}(\omega) = 0 \end{aligned} \quad (7.52)$$

As before, these tell us that the electric and magnetic fields are transverse to the direction of propagation. Although, as we mentioned previously, there is a caveat to this statement: if we can find a frequency for which $\epsilon(\omega) + i\sigma(\omega)/\omega = 0$ then longitudinal waves are allowed for the electric field. We will discuss this possibility in Section 7.6.3. For now focus on the transverse fields $\mathbf{k} \cdot \mathbf{E} = \mathbf{k} \cdot \mathbf{B} = 0$.

The other two equations are

$$\begin{aligned} \nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} &\Rightarrow i\mathbf{k} \times \mathbf{B}(\omega) = -i\mu\omega \left(\epsilon(\omega) + i \frac{\sigma(\omega)}{\omega} \right) \mathbf{E}(\omega) \\ \nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} &\Rightarrow \mathbf{k} \times \mathbf{E}(\omega) = \omega \mathbf{B}(\omega) \end{aligned}$$

The end result is that the equations governing waves in a conductor take exactly the same form as those derived in (7.30) governing waves in an insulator. The only difference is that we have to make the substitution

$$\epsilon(\omega) \longrightarrow \epsilon^{\text{eff}}(\omega) = \epsilon(\omega) + i \frac{\sigma(\omega)}{\omega}$$

This means that we can happily import our results from Section 7.5. In particular, the dispersion relation is given by

$$\mathbf{k} \cdot \mathbf{k} = \mu \epsilon^{\text{eff}}(\omega) \omega^2 \quad (7.53)$$

Let's now see how this extra term affects the physics, assuming that the optical conductivity takes the Drude form

$$\sigma(\omega) = \frac{\sigma_{\text{DC}}}{1 - i\omega\tau}$$

Low Frequencies

At frequencies that are low compared to the scattering time, $\omega\tau \ll 1$, we have $\sigma(\omega) \approx \sigma_{\text{DC}}$. This means that the real and imaginary parts of ϵ^{eff} are

$$\epsilon^{\text{eff}} = \epsilon_1^{\text{eff}} + i\epsilon_2^{\text{eff}} \approx \epsilon_1 + i\left(\epsilon_2 + \frac{\sigma_{\text{DC}}}{\omega}\right) \quad (7.54)$$

For sufficiently small ω , we always have $\epsilon_2^{\text{eff}} \gg \epsilon_1^{\text{eff}}$. This is the regime that we called *resonant absorption* in Section 7.5. The physics here is the same; no waves can propagate through the conductor; all are absorbed by the mobile electrons.

In this regime, the effective dielectric constant is totally dominated by the contribution from the conductivity and is almost pure imaginary: $\epsilon^{\text{eff}} \approx i\sigma_{\text{DC}}/\omega$. The dispersion relation (7.53) then tells us that the wavenumber is

$$k = k_1 + ik_2 = \sqrt{i\mu\omega\sigma_{\text{DC}}} = \sqrt{\frac{\mu\omega\sigma_{\text{DC}}}{2}}(1 + i)$$

So $k_1 = k_2$. This means that, for a wave travelling in the z -direction, so $\mathbf{k} = k\hat{\mathbf{z}}$, the electric field takes the form

$$\mathbf{E}(z, t) = \mathbf{E}(\omega)e^{-z/\delta} e^{i(k_1z - \omega t)}$$

where

$$\delta = \frac{1}{k_2} = \sqrt{\frac{2}{\mu\omega\sigma_{\text{DC}}}}$$

The distance δ is called the *skin depth*. It is the distance that electromagnetic waves will penetrate into a conductor. Note that as $\omega \rightarrow 0$, the waves get further and further into the conductor.

The fact that $k_1 = k_2$ also tells us, through (7.34), that the electric and magnetic fields oscillate $\pi/4$ out of phase. (The phase difference is given by $\tan\phi = k_2/k_1$). Finally, the magnitudes of the ratio of the electric and magnetic field amplitudes are given by

$$\frac{|\mathbf{B}(\omega)|}{|\mathbf{E}(\omega)|} = \frac{k}{\omega} = \sqrt{\frac{\mu\sigma_{\text{DC}}}{\omega}}$$

As $\omega \rightarrow 0$, we see that more and more of the energy lies in the magnetic, rather than electric, field.

High Frequencies

Let's now look at what happens for high frequencies. By this, we mean both $\omega\tau \gg 1$, so that $\sigma(\omega) \approx i\sigma_{\text{DC}}/\omega\tau$ and $\omega \gg \omega_0$ so that $\epsilon(\omega) \approx \epsilon_0$. Now the effective permittivity is more or less real,

$$\epsilon^{\text{eff}}(\omega) \approx \epsilon_0 - \frac{\sigma_{\text{DC}}}{\omega^2\tau} = \epsilon_0 \left(1 - \frac{\omega_p^2}{\omega^2}\right) \quad (7.55)$$

where we are using the notation of the plasma frequency $\omega_p^2 = nq^2/m\epsilon_0$ that we introduced in (7.35). What happens next depends on the sign of ϵ^{eff} :

- $\omega > \omega_p$: At these high frequencies, $\epsilon^{\text{eff}} > 0$ and k is real. This is the regime of transparent propagation. We see that, at suitably high frequencies, conductors become transparent. The dispersion relation is $\omega^2 = \omega_p^2 + c^2k^2$.
- $\omega < \omega_p$: This regime only exists if $\omega_p > \omega_0, 1/\tau$. (This is usually the case). Now $\epsilon^{\text{eff}} < 0$ so k is purely imaginary. This is the regime of total reflection; no wave can propagate inside the conductor.

We see that the plasma frequency ω_p sets the lower-limit for when waves can propagate through a conductor. For most metals, $\omega_p^{-1} \approx 10^{-16}\text{s}$ with a corresponding wavelength of $\lambda_p \approx 3 \times 10^{-10}\text{ m}$. This lies firmly in the ultraviolet, meaning that visible light is reflected. This is why most metals are shiny. (Note, however, that this is smaller than the wavelength that we needed to really trust (7.50); you would have to work harder to get a more robust derivation of this effect).

There's a cute application of this effect. In the upper atmosphere of the Earth, many atoms are ionised and the gas acts like a plasma with $\omega_p \approx 2\pi \times 9\text{ MHz}$. Only electromagnetic waves above this frequency can make it through. This includes FM radio waves. But, in contrast, AM radio waves are below this frequency and bounce back to Earth. This is why you can hear AM radio far away. And why aliens can't.

7.6.3 Plasma Oscillations

We noted in (7.52) that there's a get out clause in the requirement that the electric field is transverse to the propagating wave. The Maxwell equation reads

$$\nabla \cdot \mathbf{D} = \rho \quad \Rightarrow \quad i \left(\epsilon(\omega) + i \frac{\sigma(\omega)}{\omega} \right) \mathbf{k} \cdot \mathbf{E}(\omega) = 0$$

Which means that we can have $\mathbf{k} \cdot \mathbf{E} \neq 0$ as long as $\epsilon^{\text{eff}}(\omega) = \epsilon(\omega) + i\sigma(\omega)/\omega = 0$.

We could try to satisfy this requirement at low frequencies where the effective permittivity is given by (7.54). Since we typically have $\epsilon_1 \gg \epsilon_2$ in this regime, this is approximately

$$\epsilon^{\text{eff}}(\omega) \approx \epsilon_1 + i \frac{\sigma_{\text{DC}}}{\omega}$$

Which can only vanish if we take the frequency to be purely imaginary,

$$\omega = -i \frac{\sigma_{\text{DC}}}{\epsilon_1}$$

This is easy to interpret. Plugging it into the ansatz (7.49), we have

$$\mathbf{E}(\mathbf{x}, t) = \mathbf{E}(\omega) e^{i\mathbf{k}\cdot\mathbf{x}} e^{-\sigma_{\text{DC}}t/\epsilon_1}$$

which is telling us that if you try to put such a low frequency longitudinal field in a conductor then it will decay in time $\sim \epsilon_1/\sigma_{\text{DC}}$. This is not the solution we're looking for.

More interesting is what happens at high frequencies, $\omega \gg 1/\tau, \omega_0$, where the effective permittivity is given by (7.55). It vanishes at $\omega = \omega_p$:

$$\epsilon^{\text{eff}}(\omega_p) \approx 0$$

Now we can have a new, propagating solution in which $\mathbf{B} = 0$, while \mathbf{E} is parallel to \mathbf{k} . This is a longitudinal wave. It is given by

$$\mathbf{E}(\mathbf{x}, t) = \mathbf{E}(\omega_p) e^{i(\mathbf{k}\cdot\mathbf{x} - \omega_p t)}$$

By the relation (7.51), we see that for these longitudinal waves the charge density is also oscillating,

$$\rho(\mathbf{x}, t) = \mathbf{k} \cdot \mathbf{E}(\omega_p) e^{i(\mathbf{k}\cdot\mathbf{x} - \omega_p t)}$$

These are called *plasma oscillations*.

Note that, while the frequency of oscillation is always ω_p , the wavenumber k can be anything. This slightly strange state of affairs is changed if you take into account thermal motion of the electrons. This results in an electron pressure which acts as a restoring force on the plasma, inducing a non-trivial dispersion relation. When quantised, the resulting particles are called *plasmons*.

7.6.4 Dispersion Relations in Quantum Mechanics

So far we've derived a number of dispersion relations for various wave excitations. In all cases, these become particle excitations when we include quantum mechanics.

The paradigmatic example is the way light waves are comprised of *photons*. These are massless particles with energy E and momentum \mathbf{p} given by

$$E = \hbar\omega \quad \text{and} \quad \mathbf{p} = \hbar\mathbf{k} \quad (7.56)$$

With this dictionary, the wave dispersion relation becomes the familiar energy-momentum relation for massless particles that we met in our special relativity course,

$$\omega = kc \quad \Rightarrow \quad E = pc$$

The relationships (7.56) continue to hold when we quantise any other dispersion relation. However, one of the main lessons of this section is that both the wavevector and frequency can be complex. These too have interpretations after we quantise. A complex k means that the wave dies away quickly, typically after some boundary. In the quantum world, this just means that the particle excitations are confined close to the boundary. Meanwhile, an imaginary ω means that the wave dies down over time. In the quantum world, the imaginary part of ω has the interpretation as the lifetime of the particle.

7.7 Charge Screening

Take a system in which charges are free to move around. To be specific, we'll talk about a metal but everything we say could apply to any plasma. Then take another charge and place it at a fixed location in the middle of the system. This could be, for example, an impurity in the metal. What happens?

The mobile charges will be either attracted or repelled by the impurity. If the impurity has positive charge, the mobile, negatively charged electrons will want to cluster around it. The charge of these electrons acts to cancel out the charge of the impurity so that, viewed from afar, the region around the impurity will appear to have greatly reduced charge. There is a similar story if the charge of the impurity is negative; now the electrons are repelled, exposing the lattice of positively charged ions that lies underneath. Once again, the total charge of a region around the impurity will be greatly reduced. This is the phenomenon of *charge screening*.

Our goal here is to understand more quantitatively how this happens and, in particular, how the effective charge of the impurity changes as we move away from it. As we'll see, ultimately quantum effects will result in some rather surprising behaviour. I should mention that, unlike other parts of these notes, this section will need results from both quantum mechanics and statistical mechanics.

7.7.1 Classical Screening: The Debye-Hückel model

We'll start by looking at a simple classical model for charge screening which will give us some intuition for what's going on. Our metal consists of a mobile gas of electrons, each of charge q . These are described by a charge density $\rho(\mathbf{r})$. In the absence of any impurity, we would have $\rho(\mathbf{r}) = \rho_0$, some constant.

The entire metal is neutral. The charges of the mobile electrons are cancelled by the charges of the ions that they leave behind, fixed in position in the crystal lattice. Instead of trying to model this lattice with any accuracy, we'll simply pretend that it has a uniform, constant charge density $-\rho_0$, ensuring that the total system is neutral. This very simple toy model sometimes goes by the toy name of *jellium*.

Now we introduce the impurity by placing a fixed charge Q at the origin. We want to know how the electron density $\rho(\mathbf{r})$ responds. The presence of the impurity sets up an electric field, with the electrostatic potential $\phi(\mathbf{r})$ fixed by Gauss' law

$$\nabla^2\phi = -\frac{1}{\epsilon_0} (Q\delta^3(\mathbf{r}) - \rho_0 + \rho(\mathbf{r})) \quad (7.57)$$

Here the $-\rho_0$ term is due to the uniform background charge, while $\rho(\mathbf{r})$ is due to the electron density. It should be clear that this equation alone is not enough to solve for both $\rho(\mathbf{r})$ and $\phi(\mathbf{r})$. To make progress, we need to understand more about the forces governing the charge distribution $\rho(\mathbf{r})$. This sounds like it might be a difficult problem. However, rather than approach it as a problem in classical mechanics, we do something clever: we import some tools from statistical mechanics⁷.

We place our system at temperature T . The charge density $\rho(\mathbf{r})$ will be proportional to the probability of finding a charge q at position \mathbf{r} . If we assume that there are no correlations between the electrons, this is just given by the Boltzmann distribution. The potential energy needed to put a charge q at position \mathbf{r} is simply $q\phi(\mathbf{r})$ so we have

$$\rho(\mathbf{r}) = \rho_0 e^{-q\phi(\mathbf{r})/k_B T} \quad (7.58)$$

where the normalisation ρ_0 is fixed by assuming that far from the impurity $\phi(\mathbf{r}) \rightarrow 0$ and the system settles down to its original state.

⁷See the lecture notes on *Statistical Physics*. The Debye-Hückel model was described in Section 2.6 of these notes.

The result (7.58) is a very simple solution to what looks like a complicated problem. Of course, in part this is the beauty of statistical mechanics. But there is also an important approximation that has gone into this result: we assume that a given electron feels the average potential produced by all the others. We neglect any fluctuations around this average. This is an example of the *mean field approximation*, sometimes called the *Hartree approximation*. (We used the same kind of trick in the *Statistical Physics* notes when we first introduced the Ising model).

For suitably large temperatures, we can expand the Boltzmann distribution and write

$$\rho(\mathbf{r}) \approx \rho_0 \left(1 - \frac{q\phi(\mathbf{r})}{k_B T} + \dots \right)$$

Substituting this into Gauss' law (7.57) then gives

$$\left(\nabla^2 - \frac{1}{\lambda_D^2} \right) \phi(\mathbf{r}) = -\frac{Q}{\epsilon_0} \delta^3(\mathbf{r})$$

where λ_D is called the *Debye screening length* (we'll see why shortly) and is given by

$$\lambda_D^2 = \frac{k_B T \epsilon_0}{q^2 n_0} \quad (7.59)$$

We've written this in terms of the number density n_0 of electrons instead of the charge density $\rho_0 = qn_0$. The solution to this equation is

$$\phi(\mathbf{r}) = \frac{Q e^{-r/\lambda_D}}{4\pi\epsilon_0 r} \quad (7.60)$$

This equation clearly shows the screening phenomenon that we're interested in. At short distances $r \ll \lambda_D$, the electric field due to the impurity doesn't look very much different from the familiar Coulomb field. But at larger distances $r \gg \lambda_D$, the screening changes the potential dramatically and it now dies off exponentially quickly rather than as a power-law. Note that the electrons become less efficient at screening the impurity as the temperature increases. In contrast, if we take this result at face value, it looks as if they can screen the impurity arbitrarily well at low temperatures. But, of course, the classical description of electrons is not valid at low temperatures. Instead we need to turn to quantum mechanics.

7.7.2 The Dielectric Function

Before we look at quantum versions of screening, it's useful to first introduce some new terminology. Let's again consider introducing an impurity into the system, this

time with some fixed charge distribution $\rho^{\text{ext}}(\mathbf{r})$, where “ext” stands for “external”. We know that, taken on its own, this will induce a background electric field with potential

$$\nabla^2 \phi^{\text{ext}} = -\frac{\rho^{\text{ext}}}{\epsilon_0}$$

But we also know that the presence of the impurity will affect the charge distribution of the mobile electrons. We’ll call $\rho^{\text{ind}}(\mathbf{r}) = \rho(\mathbf{r}) - \rho_0$ the “induced charge”. We know that the actual electric field will be given by the sum of ρ^{ext} and ρ^{ind} ,

$$\nabla^2 \phi = -\frac{1}{\epsilon_0} (\rho^{\text{ext}}(\mathbf{r}) + \rho^{\text{ind}}(\mathbf{r}))$$

This set-up is very similar to our discussion in Section 7.1 when we first introduced the idea of polarisation \mathbf{P} and the electric displacement \mathbf{D} . In that case, we were interested in insulators and the polarisation described the response of bound charge to an applied electric field. Now we’re discussing conductors and the polarisation should be thought of as the response of the mobile electrons to an external electric field. In other words, $\nabla \cdot \mathbf{P} = -\rho^{\text{ind}}$. (Compare this to (7.5) for an insulator). Meanwhile, the electric displacement \mathbf{D} is the electric field that you apply to the material, as opposed to \mathbf{E} which is the actual electric field inside the material. In the present context, that means

$$\mathbf{E} = -\nabla \phi \quad \text{and} \quad \mathbf{D} = -\epsilon_0 \nabla \phi^{\text{ext}}$$

When we first introduced \mathbf{E} and \mathbf{D} , we defined the relationship between them to be simply $\mathbf{D} = \epsilon \mathbf{E}$, where ϵ is the permittivity. Later, in Section 7.5, we realised that ϵ could depend on the frequency of the applied electric field. Now we’re interested in static situations, so there’s no frequency, but the electric fields vary in space. Therefore we shouldn’t be surprised to learn that ϵ now depends on the wavelength, or wavevector, of the electric fields.

It’s worth explaining a little more how this arises. The first thing we could try is to relate $\mathbf{E}(\mathbf{r})$ and $\mathbf{D}(\mathbf{r})$. The problem is that this relationship is not local in space. An applied electric field $\mathbf{D}(\mathbf{r})$ will move charges far away which, in turn, will affect the electric field $\mathbf{E}(\mathbf{r})$ far away. This means that, in real space, the relationship between \mathbf{D} and \mathbf{E} takes the form,

$$\mathbf{D}(\mathbf{r}) = \int d^3 r' \epsilon(\mathbf{r} - \mathbf{r}') \mathbf{E}(\mathbf{r}') \tag{7.61}$$

The quantity $\epsilon(\mathbf{r} - \mathbf{r}')$ is known as the dielectric response function. It depends only on the difference $\mathbf{r} - \mathbf{r}'$ because the underlying system is translationally invariant. This

relationship looks somewhat simpler if we Fourier transform and work in momentum space. We write

$$\mathbf{D}(\mathbf{k}) = \int d^3r e^{-i\mathbf{k}\cdot\mathbf{r}} \mathbf{D}(\mathbf{r}) \quad \Leftrightarrow \quad \mathbf{D}(\mathbf{r}) = \int \frac{d^3k}{(2\pi)^3} e^{i\mathbf{k}\cdot\mathbf{r}} \mathbf{D}(\mathbf{k})$$

and similar expressions for other quantities. (Note that we're using the notation in which the function and its Fourier transform are distinguished only by their argument). Taking the Fourier transform of both sides of (7.61), we have

$$\mathbf{D}(\mathbf{k}) = \int d^3r e^{-i\mathbf{k}\cdot\mathbf{r}} \mathbf{D}(\mathbf{r}) = \int d^3r \int d^3r' e^{-i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')} \epsilon(\mathbf{r}-\mathbf{r}') e^{-i\mathbf{k}\cdot\mathbf{r}'} \mathbf{E}(\mathbf{r}')$$

But this final expression is just the product of two Fourier transforms. This tells us that we have the promised expression

$$\mathbf{D}(\mathbf{k}) = \epsilon(\mathbf{k}) \mathbf{E}(\mathbf{k})$$

The quantity $\epsilon(\mathbf{k})$ is called the *dielectric function*. The constant permittivity that we first met in Section 7.1 is simply given by $\epsilon(\mathbf{k} \rightarrow 0)$.

In what follows, we'll work with the potentials ϕ and charge densities ρ , rather than \mathbf{D} and \mathbf{E} . The dielectric function is then defined as

$$\phi^{\text{ext}}(\mathbf{k}) = \epsilon(\mathbf{k}) \phi(\mathbf{k}) \tag{7.62}$$

We write $\phi = \phi^{\text{ext}} + \phi^{\text{ind}}$, where

$$-\nabla^2 \phi^{\text{ind}} = \frac{\rho^{\text{ind}}}{\epsilon_0} \quad \Rightarrow \quad k^2 \phi^{\text{ind}}(\mathbf{k}) = \frac{\rho^{\text{ind}}(\mathbf{k})}{\epsilon_0}$$

Rearranging (7.62) then gives us an expression for the dielectric function in terms of the induced charge ρ^{ind} and the total electrostatic potential ϕ .

$$\epsilon(\mathbf{k}) = 1 - \frac{1}{\epsilon_0 k^2} \frac{\rho^{\text{ind}}(\mathbf{k})}{\phi(\mathbf{k})} \tag{7.63}$$

This will turn out to be the most useful form in what follows.

Debye-Hückel Revisited

So far, we've just given a bunch of definitions. They'll be useful moving forward, but first let's see how we can recover the results of the Debye-Hückel model using

this machinery. We know from (7.58) how the induced charge ρ^{ind} is related to the electrostatic potential,

$$\rho^{\text{ind}}(\mathbf{r}) = \rho_0 \left(e^{-q\phi(\mathbf{r})/k_B T} - 1 \right) \approx -\frac{q\rho_0\phi(\mathbf{r})}{k_B T} + \dots \quad (7.64)$$

To leading order, we then also get a linear relationship between the Fourier components,

$$\rho^{\text{ind}}(\mathbf{k}) \approx -\frac{q\rho_0}{k_B T} \phi(\mathbf{k})$$

Substituting this into (7.63) gives us an expression for the dielectric function,

$$\epsilon(\mathbf{k}) = 1 + \frac{k_D^2}{k^2} \quad (7.65)$$

where $k_D^2 = q\rho_0/\epsilon_0 k_B T = 1/\lambda_D^2$, with λ_D the Debye screening length that we introduced in (7.59).

Let's now see the physics that's encoded in the dielectric function. Suppose that we place a point charge at the origin. We have

$$\phi^{\text{ext}}(\mathbf{r}) = \frac{Q}{4\pi\epsilon_0 r} \quad \Rightarrow \quad \phi^{\text{ext}}(\mathbf{k}) = \frac{Q}{\epsilon_0 k^2}$$

Then, using the form of the dielectric function (7.65), the resulting electrostatic potential ϕ is given by

$$\phi(\mathbf{k}) = \frac{\phi^{\text{ext}}(\mathbf{k})}{\epsilon(\mathbf{k})} = \frac{Q}{\epsilon_0(k^2 + k_D^2)}$$

We need to do the inverse Fourier transform of $\phi(\mathbf{k})$ to find $\phi(\mathbf{r})$. Let's see how to do it; we have

$$\phi(\mathbf{r}) = \int \frac{d^3 k}{(2\pi)^3} e^{i\mathbf{k}\cdot\mathbf{r}} \phi(\mathbf{k}) = \frac{Q}{(2\pi)^3 \epsilon_0} \int_0^{2\pi} d\phi \int_0^\pi d\theta \sin\theta \int_0^\infty dk \frac{k^2}{k^2 + k_D^2} e^{ikr \cos\theta}$$

where, in the second equality, we've chosen to work in spherical polar coordinates in which the k_z axis is aligned with \mathbf{r} , so that $\mathbf{k} \cdot \mathbf{r} = kr \cos\theta$. We do the integrals over the two angular variables, to get

$$\begin{aligned} \phi(\mathbf{r}) &= \frac{Q}{(2\pi)^2 \epsilon_0} \int_0^\infty dk \frac{k^2}{k^2 + k_D^2} \frac{2 \sin kr}{kr} \\ &= \frac{Q}{(2\pi)^2 \epsilon_0 r} \int_{-\infty}^\infty dk \frac{k \sin kr}{k^2 + k_D^2} \\ &= \frac{Q}{2\pi \epsilon_0 r} \operatorname{Re} \left[\int_{-\infty}^{+\infty} \frac{dk}{2\pi i} \frac{k e^{ikr}}{k^2 + k_D^2} \right] \end{aligned}$$

We compute this last integral by closing the contour in the upper-half plane with $k \rightarrow +i\infty$, picking up the pole at $k = +ik_D$. This gives our final answer for the electrostatic potential,

$$\phi(\mathbf{r}) = \frac{Q e^{-r/\lambda_D}}{4\pi\epsilon_0 r}$$

That's quite nice: we see that the dielectric function (7.65) contains the same physics (7.60) that we saw earlier in the direct computation of classical electrostatic screening. We could also compute the induced charge density to find

$$\rho^{\text{ind}}(\mathbf{r}) = -\frac{Q e^{-r/\lambda_D}}{4\pi\lambda_D^2 r}$$

which agrees with (7.64).

But the dielectric function $\epsilon(\mathbf{k})$ contains more information: it tells us how the system responds to each Fourier mode of an externally placed charge density. This means that we can use it to compute the response to any shape $\rho^{\text{ext}}(\mathbf{r})$.

Here, for example, is one very simple bit of physics contained in $\epsilon(\mathbf{k})$. In the limit $k \rightarrow 0$, we have $\epsilon(k) \rightarrow \infty$. This means that, in the presence of any constant, applied electric field \mathbf{D} , the electric field inside the material will be $\mathbf{E} = \mathbf{D}/\epsilon = 0$. But you knew this already: it's the statement that you can't have electric fields inside conductors because the charges will always move to cancel it. More generally, classical conductors will effectively screen any applied electric field which doesn't vary much on distances smaller than λ_D .

7.7.3 Thomas-Fermi Theory

The Debye-Hückel result describes screening by classical particles. But, as we lower the temperature, we know that quantum effects become important. Our first pass at this is called the *Thomas-Fermi approximation*. It's basically the same idea that we used in the Debye-Hückel approach, but with the probability determined by the Fermi-Dirac distribution rather than the classical Boltzmann distribution.

We work in the grand canonical ensemble, with temperature T and chemical potential μ . Recall that the probability of finding a fermion in a state $|k\rangle$ with energy E_k is given by the Fermi-Dirac distribution

$$f(\mathbf{k}) = \frac{1}{e^{(E_k - \mu)/k_B T} + 1} \tag{7.66}$$

The chemical potential μ is determined by the requirement that the equilibrium charge density is $\rho(\mu) = \rho_0$, where

$$\rho(\mu) = g_s \int \frac{d^3k}{(2\pi)^3} \frac{q}{e^{(E_k - \mu)/k_B T} + 1} \quad (7.67)$$

Here g_s is the spin degeneracy factor which we usually take to be $g_s = 2$.

Let's now place the external charge density $\rho^{\text{ext}}(\mathbf{r})$ in the system. The story is the same as we saw before: the mobile charges move, resulting in an induced charge density $\rho^{\text{ind}}(\mathbf{r})$, and a total electrostatic potential $\phi(\mathbf{r})$. The Thomas-Fermi approximation involves working with the new probability distribution

$$f(\mathbf{k}, \mathbf{r}) = \frac{1}{e^{(E_k + q\phi(\mathbf{r}) - \mu)/k_B T} + 1} \quad (7.68)$$

This can be thought of as either changing the energy to $E = E_k + q\phi(\mathbf{r})$ or, alternatively, allowing for a spatially varying chemical potential $\mu \rightarrow \mu - q\phi(\mathbf{r})$.

The first thing to say about the probability distribution (7.68) is that it doesn't make any sense! It claims to be the probability for a state with momentum \mathbf{k} and position \mathbf{r} , yet states in quantum mechanics are, famously, not labelled by both momentum and position at the same time! So what's going on? We should think of (7.68) as an approximation that is valid when $\phi(\mathbf{r})$ is very slowly varying compared to any microscopic length scales. Then we can look in a patch of space where $\phi(\mathbf{r})$ is roughly constant and apply (7.68). In a neighbouring patch of space we again apply (7.68), now with a slightly different value of $\phi(\mathbf{r})$. This idea of *local equilibrium* underlies the Thomas-Fermi (and, indeed, the Debye-Hückel) approximations.

Let's see how this works in practice. The spatially dependent charge density is now given by

$$\rho(\mathbf{r}; \mu) = g_s \int \frac{d^3k}{(2\pi)^3} \frac{q}{e^{(E_k + q\phi(\mathbf{r}) - \mu)/k_B T} + 1} \quad (7.69)$$

We're interested in computing the induced charge density $\rho^{\text{ind}}(\mathbf{r}) = \rho(\mathbf{r}) - \rho_0$. Combining (7.69) and (7.67), we have

$$\rho^{\text{ind}}(\mathbf{r}) = g_s \int \frac{d^3k}{(2\pi)^3} \left[\frac{q}{e^{(E_k + q\phi(\mathbf{r}) - \mu)/k_B T} + 1} - \frac{q}{e^{(E_k - \mu)/k_B T} + 1} \right]$$

But we can rewrite this using the notation of (7.67) simply as

$$\rho^{\text{ind}}(\mathbf{r}) = \rho(\mu - q\phi(\mathbf{r})) - \rho(\mu) \approx -q\phi(\mathbf{r}) \frac{\partial \rho(\mu)}{\partial \mu}$$

where, in the last step, we have Taylor expanded the function which is valid under the assumption that $q\phi(\mathbf{r}) \ll \mu$. But this immediately gives us an expression for the dielectric function using (7.63),

$$\epsilon(\mathbf{k}) = 1 + \frac{\partial \rho}{\partial \mu} \frac{q}{\epsilon_0 k^2}$$

We're almost there. We still need to figure out what $\partial \rho / \partial \mu$ is. This is particularly easy if we work at $T = 0$, where we can identify the chemical potential μ with the Fermi energy: $\mu = E_F$. In this case, the Fermi-Dirac distribution is a step function and the total charge density is simply given by

$$\rho(E_F) = q \int_0^{E_F} dE g(E)$$

where $g(E)$ is the density of states (we'll remind ourselves what form the density of states takes below). We learn that $\partial \rho / \partial E_F = g(E_F)$ and the dielectric function is given by

$$\epsilon(\mathbf{k}) = 1 + \frac{q^2 g(E_F)}{\epsilon_0 k^2} \tag{7.70}$$

Note that the functional form of $\epsilon(\mathbf{k})$ is exactly the same as we saw in the classical case (7.65). The only thing that's changed is the coefficient of the $1/k^2$ term which, as we saw before, determines the screening length. Let's look at a simple example.

A Simple Example

For non-relativistic particles, the energy is given by $E = \hbar^2 k^2 / 2m$. In three spatial dimensions, the density of states is given by⁸

$$g(E) = g_s \frac{1}{4\pi^2} \left(\frac{2m}{\hbar^2} \right)^{3/2} E^{1/2}$$

This is kind of a mess, but there's a neater way to write $g(E_F)$. (This neater way will also allow for a simple comparison to the Debye screening length as well). At zero temperature, the total charge density is

$$\rho_0 = q \int_0^{E_F} dE g(E)$$

⁸See the lecture notes on *Statistical Physics* for details on how to compute the density of states. The $g(E)$ we use here differs slightly from that presented in the Statistical Physics lectures because it does not include an overall volume factor. This is because we want to compute the number density of particles rather than the total number of particles.

Using this, we have

$$g(E_F) = \frac{3}{2q} \frac{\rho_0}{E_F}$$

and we can write the dielectric function as

$$\epsilon(\mathbf{k}) = 1 + \frac{k_{TF}^2}{k^2}$$

where $k_{TF}^2 = 3q\rho_0/2\epsilon_0 E_F$. This is our expression for the Thomas-Fermi screening length $\lambda_{TF} = 1/k_{TF}$.

It's instructive to compare this screening length with the classical Debye length λ_D . We have

$$\frac{\lambda_D^2}{\lambda_{TF}^2} = \frac{2}{3} \frac{T}{T_F}$$

where $T_F = k_B E_F$ is the Fermi temperature. The classical analysis can only be trusted at temperature $T \gg T_F$ where $\lambda_D \gg \lambda_{TF}$. But, for metals, the Fermi temperature is hot; something like $10^4, K$. This means that, at room temperature, $T \ll T_F$ and our quantum result above (which, strictly speaking, was only valid at $T = 0$) is a good approximation. Here $\lambda_D \ll \lambda_{TF}$. The upshot is that quantum mechanics acts to increase the screening length beyond that suggested by classical physics.

7.7.4 Lindhard Theory

The Thomas-Fermi approximation is straightforward, but it relies crucially on the potential $\phi(\mathbf{r})$ varying only over large scales. However, as we will now see, the most interesting physics arises due to variations of $\phi(\mathbf{r})$ over small scales (or, equivalently, large \mathbf{k}). For this we need to work harder.

The key idea is to go back to basics where, here, basics means quantum mechanics. Before we add the impurity, the energy eigenstates are plane waves $|\mathbf{k}\rangle$ with energy $E(k) = \hbar^2 k^2 / 2m$. To determine the dielectric function (7.63), we only need to know how the mobile charge density $\rho(\mathbf{r})$ changes in the presence of a potential $\phi(\mathbf{r})$. We can do this by considering a small perturbation to the Hamiltonian of the form

$$\Delta H = q\phi(\mathbf{r})$$

The energy eigenstate that is labelled by \mathbf{k} now shifts. We call the new state $|\psi(\mathbf{k})\rangle$. Ultimately, our goal is to compute the induced charge density. For an electron in state $|\psi(\mathbf{k})\rangle$, the probability of finding it at position \mathbf{r} is simply $|\langle \mathbf{r} | \psi(\mathbf{k}) \rangle|^2$. Which means

that, for this state, the change in the density is $|\langle \mathbf{r} | \psi(\mathbf{k}) \rangle|^2 - |\langle \mathbf{r} | \mathbf{k} \rangle|^2$. The induced charge density $\rho^{\text{ind}}(\mathbf{r})$ is obtained by summing over all such states, weighted with the Fermi-Dirac distribution function. We have

$$\rho^{\text{ind}}(\mathbf{r}) = qg_s \int \frac{d^3k}{(2\pi)^3} f(k) [|\langle \mathbf{r} | \psi(\mathbf{k}) \rangle|^2 - |\langle \mathbf{r} | \mathbf{k} \rangle|^2]$$

where $f(k)$ is the Fermi-Dirac distribution (7.66) and we've remembered to include the spin degeneracy factor $g_s = 2$. To make progress, we need to get to work computing the overlap of states.

To first order in perturbation theory, the new energy eigenstate is given by

$$|\psi(\mathbf{k})\rangle = |\mathbf{k}\rangle + \int \frac{d^3k'}{(2\pi)^3} \frac{\langle \mathbf{k}' | \Delta H | \mathbf{k} \rangle}{E(k) - E(k')} |\mathbf{k}'\rangle$$

Keeping only terms linear in ΔH , we can expand this out to read

$$|\langle \mathbf{r} | \psi(\mathbf{k}) \rangle|^2 - |\langle \mathbf{r} | \mathbf{k} \rangle|^2 = \int \frac{d^3k'}{(2\pi)^3} \left[\langle \mathbf{r} | \mathbf{k} \rangle \frac{\langle \mathbf{k} | \Delta H | \mathbf{k}' \rangle}{E(k) - E(k')} \langle \mathbf{k}' | \mathbf{r} \rangle + \langle \mathbf{k} | \mathbf{r} \rangle \frac{\langle \mathbf{k}' | \Delta H | \mathbf{k} \rangle}{E(k) - E(k')} \langle \mathbf{r} | \mathbf{k}' \rangle \right]$$

But we have expressions for each of these matrix elements. Of course, the plane waves take the form $\langle \mathbf{r} | \mathbf{k} \rangle = e^{i\mathbf{k}\cdot\mathbf{r}}$, while the matrix elements of the perturbed Hamiltonian are

$$\langle \mathbf{k}' | q\phi(\mathbf{r}) | \mathbf{k} \rangle = \int d^3r d^3r' e^{i(\mathbf{k}\cdot\mathbf{r} - \mathbf{k}'\cdot\mathbf{r}')} \langle \mathbf{r}' | q\phi(\mathbf{r}) | \mathbf{r} \rangle = q\phi(\mathbf{k} - \mathbf{k}')$$

In other words, it gives the Fourier transform of the electrostatic potential. Putting this together, we arrive at an integral expression for the induced charge,

$$\rho^{\text{ind}}(\mathbf{r}) = q^2 g_s \int \frac{d^3k}{(2\pi)^3} \frac{d^3k'}{(2\pi)^3} f(k) \left[\frac{e^{-i(\mathbf{k}' - \mathbf{k})\cdot\mathbf{r}} \phi(\mathbf{k} - \mathbf{k}')}{E(k) - E(k')} + \frac{e^{-i(\mathbf{k} - \mathbf{k}')\cdot\mathbf{r}} \phi(\mathbf{k}' - \mathbf{k})}{E(k) - E(k')} \right]$$

Of course, what we really want for the dielectric function (7.63) is the Fourier transform of the induced charge,

$$\rho^{\text{ind}}(\mathbf{k}) = \int d^3r e^{-i\mathbf{k}\cdot\mathbf{r}} \rho^{\text{ind}}(\mathbf{r})$$

Thankfully, doing the $\int d^3r$ integral gives rise to a delta-function which simplifies our life rather than complicating it. Performing some relabelling of dummy integration variables, we have

$$\frac{\rho^{\text{ind}}(\mathbf{k})}{\phi(\mathbf{k})} = q^2 g_s \int \frac{d^3k'}{(2\pi)^3} f(k') \left[\frac{1}{E(k') - E(|\mathbf{k}' - \mathbf{k}|)} + \frac{1}{E(k') - E(|\mathbf{k} + \mathbf{k}'|)} \right] \quad (7.71)$$

These two terms are more similar than they look. If we change the dummy integration variable in the first term to $\mathbf{k}' \rightarrow \mathbf{k}' + \mathbf{k}$ then we can write

$$\frac{\rho^{\text{ind}}(\mathbf{k})}{\phi(\mathbf{k})} = q^2 g_s \int \frac{d^3 k'}{(2\pi)^3} \frac{f(|\mathbf{k} + \mathbf{k}'|) - f(k')}{E(|\mathbf{k} + \mathbf{k}'|) - E(k')} \quad (7.72)$$

The left-hand side is exactly what we want. The right-hand side is an integral. It's not too hard to do this integral, but let's first check that this result gives something sensible.

Thomas-Fermi Revisited

Let's first see how we can recover the Thomas-Fermi result for the dielectric function. Recall that the Thomas-Fermi approximation was only valid when the potential $\phi(\mathbf{r})$, and hence the induced charge $\rho^{\text{ind}}(\mathbf{r})$, vary slowly over large distances. In the present context, this means it is valid at small k . But here we can simply Taylor expand the numerator and denominator of (7.72).

$$E(|\mathbf{k} + \mathbf{k}'|) - E(k') \approx \frac{\partial E}{\partial \mathbf{k}'} \cdot \mathbf{k}$$

and

$$f(|\mathbf{k} + \mathbf{k}'|) - f(k') \approx \frac{\partial f}{\partial E} \frac{\partial E}{\partial \mathbf{k}'} \cdot \mathbf{k}$$

So we have

$$\frac{\rho^{\text{ind}}(\mathbf{k})}{\phi(\mathbf{k})} = q^2 g_s \int \frac{d^3 k'}{(2\pi)^3} \frac{\partial f}{\partial E} = q^2 \int dE g(E) \frac{\partial f}{\partial E}$$

where the last step is essentially the definition of the density of states $g(E)$. But at $T = 0$, the Fermi-Dirac distribution $f(E)$ is just a step function, and $\partial f / \partial E = -\delta(E - E_F)$. So at $T = 0$, we get

$$\frac{\rho^{\text{ind}}(\mathbf{k})}{\phi(\mathbf{k})} = q^2 g(E_F) \quad \Rightarrow \quad \epsilon(\mathbf{k}) = 1 + \frac{q^2 g(E_F)}{\epsilon_0 k^2}$$

which we recognise as the Thomas-Fermi result (7.70) that we derived previously.

The Lindhard Function

While the Thomas-Fermi approximation suffices for variations over large scales and small k , our real interest here is in what happens at large k . As we will now show, quantum mechanics gives rise to some interesting features in the screening when impurities have structure on scales of order $\sim 1/k_F$ where k_F is the Fermi-wavevector. For this, we need to go back to the Lindhard result

$$\frac{\rho^{\text{ind}}(\mathbf{k})}{\phi(\mathbf{k})} = q^2 g_s \int \frac{d^3 k'}{(2\pi)^3} \frac{f(|\mathbf{k} + \mathbf{k}'|) - f(k')}{E(|\mathbf{k} + \mathbf{k}'|) - E(k')}$$

Our task is to do this integral properly.

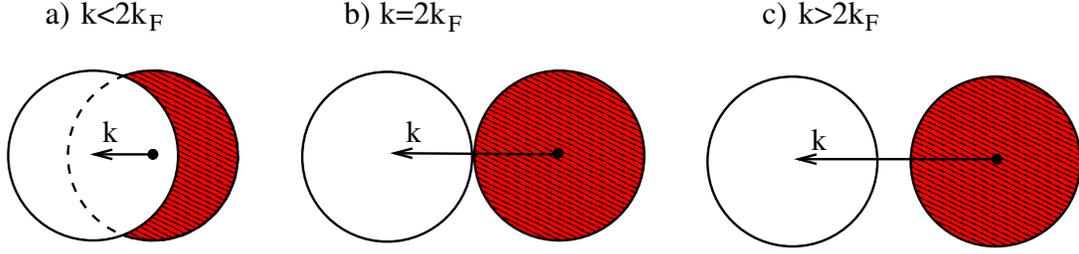


Figure 80: The two Fermi surfaces in momentum space. The integration region Σ is shown shaded in red for a) $k < 2k_F$, b) $k = 2k_F$ and c) $k > 2k_F$.

Let's firstly get a sense for what the integrand looks like. We'll work at $T = 0$, so the Fermi-Dirac distribution function $f(k)$ is a step function with

$$f(k) = \begin{cases} 1 & k < k_F \\ 0 & k > k_F \end{cases}$$

This makes the integral much easier. All the subtleties now come from figuring out which region in momentum space gives a non-vanishing contribution. The filled states associated to $f(k')$ form a ball in momentum space of radius k_F , centered at the origin. Meanwhile, the filled states associated to $f(|\mathbf{k}' + \mathbf{k}|)$ form a ball in momentum space of radius k_F centered at $\mathbf{k}' = -\mathbf{k}$. These are shown in a number of cases in Figure 80. Because the integral comes with a factor of $f(|\mathbf{k} + \mathbf{k}'|) - f(k')$, it gets contributions only from states that are empty in one ball but filled in the other. We call this region Σ ; it is the shaded red region shown in the figures. There is also a mirror region in the other ball that also contributes to the integral, but this simply gives an overall factor of 2. So we have

$$\frac{\rho^{\text{ind}}(\mathbf{k})}{\phi(\mathbf{k})} = 2q^2 g_s \int_{\Sigma} \frac{d^3 k'}{(2\pi)^3} \frac{1}{E(|\mathbf{k} + \mathbf{k}'|) - E(k')}$$

The important physics lies in the fact that the nature of Σ changes as we vary k . For $k < 2k_F$, Σ is a crescent-shaped region as shown in Figure 80a. But for $k \geq 2k_F$, Σ is the whole Fermi ball as shown in Figures 80b and 80c.

We'll work with non-relativistic fermions with $E = \hbar^2 k^2 / 2m$. While the graphical picture above will be useful to get intuition for the physics, to do the integral it's

actually simpler to return to the form (7.71). At zero temperature, we have

$$\begin{aligned}\frac{\rho^{\text{ind}}(\mathbf{k})}{\phi(\mathbf{k})} &= q^2 g_s \int_{k \leq k_F} \frac{d^3 k'}{(2\pi)^3} \left[\frac{1}{-k^2 + 2\mathbf{k} \cdot \mathbf{k}'} + \frac{1}{-k^2 - 2\mathbf{k} \cdot \mathbf{k}'} \right] \\ &= -q^2 g_s \frac{2m}{\hbar^2} \int_{k' \leq k_F} \frac{d^3 k'}{(2\pi)^3} \frac{2}{k^2 - 2\mathbf{k}' \cdot \mathbf{k}}\end{aligned}$$

where the two terms double-up because rotational symmetry ensures that the physics is invariant under $\mathbf{k} \rightarrow -\mathbf{k}$. Now the integration domain remains fixed as we vary \mathbf{k} , with the graphical change of topology that we saw above buried in the integrand. For $k \leq 2k_F$, the denominator in the integrand can vanish. This reflects the fact that transitions between an occupied and unoccupied state with the same energy are possible. It corresponds to the situation depicted in Figure 80a. But for $k > 2k_F$, the denominator is always positive. This corresponds to the situation shown in Figure 80c.

To proceed, we work in polar coordinates for \mathbf{k}' with the z-axis aligned with \mathbf{k} . We have

$$\begin{aligned}\frac{\rho^{\text{ind}}(\mathbf{k})}{\phi(\mathbf{k})} &= -\frac{4mq^2 g_s}{(2\pi)^2 \hbar^2} \int_0^\pi d\theta \sin \theta \int_0^{k_F} dk' \frac{k'^2}{k^2 - 2kk' \cos \theta} \\ &= \frac{2mq^2 g_s}{(2\pi)^2 \hbar^2} \frac{1}{k} \int_0^{k_F} dk' k' \log \left| \frac{k^2 + 2kk'}{k^2 - 2kk'} \right|\end{aligned}$$

But this is now an integral that we can do; the general form is

$$\int dy y \log \left(\frac{ay + b}{-ay + b} \right) = \frac{by}{a} + \frac{1}{2} \left(y^2 - \frac{b^2}{a^2} \right) \log \left(\frac{ay + b}{-ay + b} \right)$$

We then have

$$\frac{\rho^{\text{ind}}(\mathbf{k})}{\phi(\mathbf{k})} = -\frac{2mq^2 g_s}{(2\pi)^2 \hbar^2} \frac{1}{k} \left[\frac{kk_F}{2} + \frac{1}{2} \left(k_F^2 - \frac{k^2}{4} \right) \log \left| \frac{2kk_F + k^2}{-2kk_F + k^2} \right| \right]$$

This gives our final expression, known as the *Lindhard dielectric function*,

$$\epsilon(k) = 1 + \frac{k_{TF}^2}{k^2} F \left(\frac{k}{2k_F} \right)$$

where all the constants that we gathered along our journey sit in $k_{TF}^2 = q^2 g(E_F)/\epsilon_0 = g_s q^2 m k_F / 2\pi^2 \hbar^2 \epsilon_0$. This is the Thomas-Fermi wave result that we saw previously, but now it is dressed by the function

$$F(x) = \frac{1}{2} + \frac{1-x^2}{4x} \log \left| \frac{x+1}{x-1} \right|$$

At small k we have $F(x \rightarrow 0) = 1$ and we recover the Thomas-Fermi result.

For variations on very small scales, we're interested in the large k regime where $x \rightarrow \infty$ and $F(x) \rightarrow 1/3x^2$. (You have to go to third order in the Taylor expansion of the log to see this!). This means that on small scales we have

$$\epsilon(k) \rightarrow 1 + \frac{4k_{TF}^2 k_F^2}{3k^4}$$

However, the most interesting physics occurs at near $k = 2k_F$.

7.7.5 Friedel Oscillations

We saw above that there's a qualitative difference in the accessible states when $k < 2k_F$ and $k > 2k_F$. Our goal is to understand what this means for the physics. The dielectric function itself is nice and continuous at $k = 2k_F$, with $F(x = 1) = 1/2$. However, it is not smooth: the derivative of the dielectric function suffers a logarithmic singularity,

$$F'(x \rightarrow 1^+) \rightarrow \frac{1}{2} \log \left(\frac{x-1}{2} \right)$$

This has an important consequence for the screening of a point charge.

As we saw in Section 7.7.2, a point charge gives rise to the external potential

$$\phi^{\text{ext}}(\mathbf{k}) = \frac{Q}{\epsilon_0 k^2}$$

and, after screening, the true potential is $\phi(\mathbf{k}) = \phi^{\text{ext}}(\mathbf{k})/\epsilon(\mathbf{k})$. However, the Fourier transform back to real space is now somewhat complicated. It turns out that it's easier to work directly with the induced charge density $\rho^{\text{ind}}(\mathbf{r})$. From the definition of the dielectric function (7.63), the induced charge density in the presence of a point charge $\phi^{\text{ext}}(\mathbf{k}) = Q/\epsilon_0 k^2$ is given by,

$$\rho^{\text{ind}}(\mathbf{k}) = -Q \frac{\epsilon(\mathbf{k}) - 1}{\epsilon(\mathbf{k})}$$

where, for $k \approx 2k_F$, we have

$$\frac{\epsilon(\mathbf{k}) - 1}{\epsilon(\mathbf{k})} = \frac{k_{TF}^2}{8k_F^2} \left(1 + \frac{k - 2k_F}{2k_F} \log \left(\frac{k - 2k_F}{4k_F} \right) + \dots \right) \quad (7.73)$$

Now we want to Fourier transform this back to real space. We repeat the steps that we took in Section 7.7.2 for the Debye-Hückel model to get

$$\rho^{\text{ind}}(\mathbf{r}) = -Q \int \frac{d^3k}{(2\pi)^3} e^{i\mathbf{k}\cdot\mathbf{r}} \left(\frac{\epsilon(\mathbf{k}) - 1}{\epsilon(\mathbf{k})} \right) = -\frac{Q}{2\pi^2} \frac{1}{r} \int_0^\infty dk \left(\frac{k\epsilon(\mathbf{k}) - k}{\epsilon(\mathbf{k})} \right) \sin kr$$

At this stage, it's useful if we integrate by parts twice. We have

$$\rho^{\text{ind}}(\mathbf{r}) = \frac{Q}{2\pi^2} \frac{1}{r^3} \int_0^\infty dk \frac{d^2}{dk^2} \left(\frac{k\epsilon(\mathbf{k}) - k}{\epsilon(\mathbf{k})} \right) \sin kr$$

Of course, the Fourier integral requires us to know $\epsilon(\mathbf{k})$ at all values of k , rather than just around $k = 2k_F$. Suppose, however, that we're interested in the behaviour a long way from the point charge. At large r , the $\sin kr$ factor oscillates very rapidly with k , ensuring that the induced charge at large distances is essentially vanishing. This was responsible for the exponential behaviour of the screening that we saw in both the Debye-Hückel and Thomas-Fermi models. However, at $k = 2k_F$ the other factor in the integrand diverges,

$$\frac{d^2}{dk^2} \left(\frac{k\epsilon(\mathbf{k}) - k}{\epsilon(\mathbf{k})} \right) \approx \frac{k_{TF}^2}{4k_F^2} \frac{1}{k - 2k_F}$$

This will now give rise to a long-range contribution. Therefore, if we only care about this long-distance behaviour, we need only integrate over some small interval I about $k = 2k_F$,

$$\begin{aligned} \rho^{\text{ind}}(\mathbf{r}) &\approx \frac{Qk_{TF}^2}{8\pi^2k_F^2} \frac{1}{r^3} \int_I dk \frac{\sin kr}{k - 2k_F} \\ &= \frac{Qk_{TF}^2}{8\pi^2k_F^2} \frac{1}{r^3} \int_I dk \left[\frac{\cos(2k_F r) \sin((q - 2k_F)r)}{k - 2k_F} + \frac{\sin(2k_F r) \cos((k - 2k_F)r)}{k - 2k_F} \right] \end{aligned}$$

where we've used a little trigonometry. The second term above vanishes on parity grounds (contributions from either side of $k = k_F$ cancel). We can approximate the first term by extending the range of the integral to all k (because, as we've just argued, the main contribution comes from the interval I anyway). Using $\int_{-\infty}^{+\infty} dx \sin x/x = \pi$, we get our final expression for the long-distance charge density induced by a point charge,

$$\rho^{\text{ind}}(\mathbf{r}) \approx \frac{Qk_{TF}^2}{8\pi k_F^2} \frac{\cos(2k_F r)}{r^3} \quad (7.74)$$

We learn that the effect of the Fermi surface is to dramatically change the screening of electric charge. Instead of the usual exponential screening, we instead find a power-law fall off, albeit weaker than the Coulomb force in vacuum (i.e. $1/r^3$ instead of $1/r$). Moreover, the sign of the induced charge oscillates. These are called *Friedel oscillations*. They provide a very visual way to see the edge of the Fermi surface. This figure shows some Friedel oscillations on a two-dimensional surface⁹. You can make out a bright

⁹The figure is taken from *Direct Observation of Friedel Oscillations around Incorporated Si_{Ga} Dopants in GaAs by Low-Temperature Scanning Tunneling Microscopy* by M van der Wielen, A van Roij and H. van Kempen, Physical Review Letters 76, 7 (1996).

central region, surrounded by a black ring, surrounded by another white ring. This corresponds to a Fermi wavelength of around $\lambda_F \sim 10^{-8} m$.

Heuristically, what's going on here is that the wavefunction of the electrons has a finite size.. At zero temperature, the states with lowest energy have wavelength $\lambda = 1/k_F$. These modes enthusiastically cluster around the impurity, keen to reduce its charge but, unaware of their own cumbersome nature, end up overscreening. Other electrons have to then respond to undo the damage and the story is then repeated, over exuberance piled upon over exuberance. The end result is a highly inefficient screening mechanism and the wonderful rippling patterns of charge that are seen in scanning tunnelling microscopes.

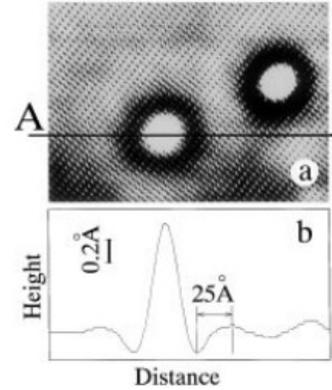


Figure 81: Friedel oscillations in GaAs doped with Silicon.