Electromagnetism Lecture Notes

I took over this course from the late, dear Sandro De Vita in 2018; but I re-wrote it entirely so this is a new second year course in Electromagnetism. It actually makes a neat continuation from my previous first year course, Fields and Waves (an inspired concept due, I think, to the late Alan Michette). Together the two make an introductory course on Electromagnetism, but of course are no substitute for the many brilliant textbooks. However there may be some value in the approach I have taken, the order in which I present the material, and possibly one or two insights.

I have drawn heavily on the textbooks of Griffiths, Panofsky and Phillips, Lorrain and Corson, and of course the sublime Arnold Sommerfeld. I am grateful to Edmund Adam Paxton for telling me about Helmholtz decomposition. Griffiths now has it in an appendix to his fourth edition; and curiously Panofsky and Phillips open their textbook with it, but they do not call it Helmholtz decomposition. Perhaps this is because it was first discovered by Stokes.

This course was divided into nine sections as follows.

1. Revision of vector calculus
2. Revision of electrodynamics
3. Lorentz force
4. Maxwell's equations
5. Scalar and vector potential
6. Auxiliary fields
7. Static electric fields in matter
8. Magnetism in matter
9. Propagation and absorption

Tony Paxton, King's College London, 30 June 2021

Contents

1. Revision of vector calculus

1.1 Vectors

A vector is an object that describes a quantity having both magnitude and direction. We can write it as a product of the two,

$$\mathbf{v} = v\hat{\mathbf{v}}$$

$\hat{\mathbf{v}}$ is the unit vector and is dimensionless. v is sometimes written $|\mathbf{v}|$. Once we agree a Cartesian coordinate system, we can illustrate $\mathbf{v}$ as an arrow emanating from the origin.

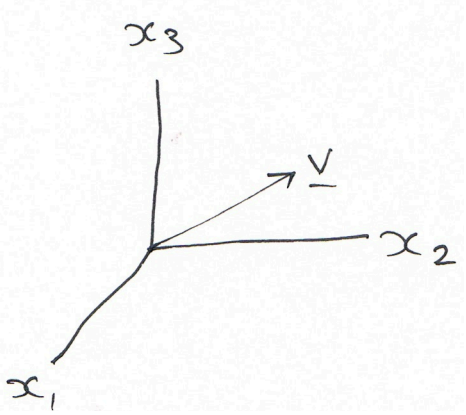


FIGURE 1–1

x_1, x_2, x_3 are sometimes written x, y, z. Now we can put $\mathbf{v}$ into its *components*,

$$\mathbf{v} = \hat{\mathbf{i}}\, v_1 + \hat{\mathbf{j}}\, v_2 + \hat{\mathbf{k}}\, v_3$$

$\hat{\mathbf{i}}$, $\hat{\mathbf{j}}$ and $\hat{\mathbf{k}}$, which we will write as $\hat{\mathbf{e}}_1$, $\hat{\mathbf{e}}_2$ and $\hat{\mathbf{e}}_3$, are *unit vectors* along the x_1, x_2 and x_3 axes.

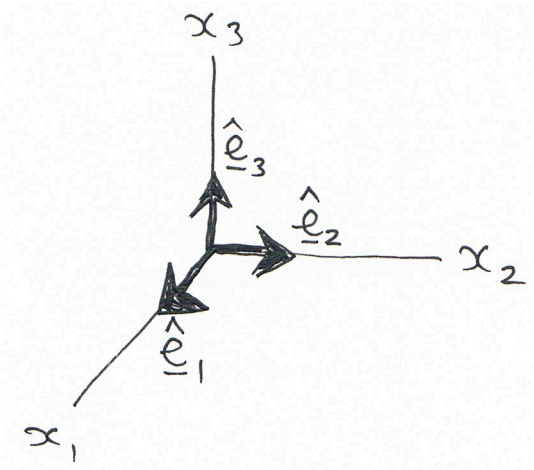


FIGURE 1–2

There are two ways to multiply vectors.

1. The scalar, or dot-product between vectors $\mathbf{u}$ and $\mathbf{v}$ is

$$\mathbf{u}\cdot\mathbf{v} = u_1v_1 + u_2v_2 + u_3v_3 = \sum_{i=1}^{3} u_iv_i$$

$$= uv\cos\theta$$

and θ is the angle between $\mathbf{v}$ and $\mathbf{u}$. We will adopt the *Einstein summation convention* and omit the summation sign. It is understood that if in a product, or in a single term, the same index appears twice, then that index is summed over from one to three. Hence

$$\mathbf{u}\cdot\mathbf{v} = u_i v_i$$

It is very important to realise that i here is a *dummy index*, so I can equally well write†

$$\mathbf{u}\cdot\mathbf{v} = u_k v_k = u_j v_j = u_m v_m = \ldots$$

2. There is a vector whose magnitude is the product of the magnitudes of the two vectors being multiplied times the sine of the angle between them, and whose direction is perpendicular in a right-handed sense to both of them. This is the cross product,

$$\begin{aligned}\mathbf{u}\times\mathbf{v} = \hat{\mathbf{e}}_1\,(u_2v_3 - u_3v_2)\\ + \hat{\mathbf{e}}_2\,(u_3v_1 - u_1v_3)\\ + \hat{\mathbf{e}}_3\,(u_1v_2 - u_2v_1)\end{aligned} \tag{1.1}$$

An easy way to write this down *without even thinking* is to use these two sequences.

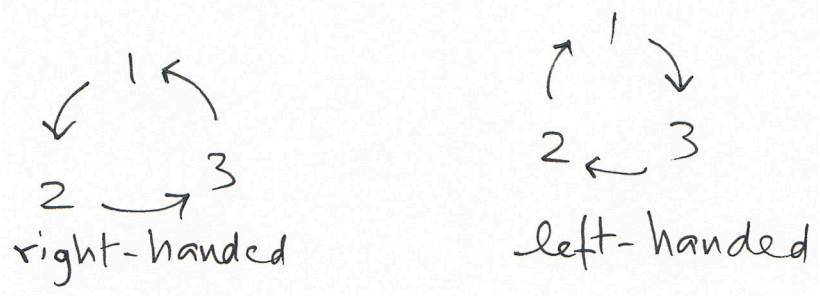


FIGURE 1–3

First write the cross product leaving the indices blank,

$$\begin{aligned}\mathbf{u}\times\mathbf{v} = \hat{\mathbf{e}}_1\,(u\_v\_ - u\_v\_)\\ + \hat{\mathbf{e}}_2\,(u\_v\_ - u\_v\_)\\ + \hat{\mathbf{e}}_3\,(u\_v\_ - u\_v\_)\end{aligned}$$

then fill them in using the following rule. In the first line the prefix is $\hat{\mathbf{e}}_1 = \hat{\mathbf{i}}$ so the "1" index *won't* appear in the following parentheses; then fill in the blanks, in the first product starting with "1" and going around the sequence right-handed and in the second product going round the sequence left-handed: hence the indices to fill in in the first line are 2, 3 then 3, 2. In the second line the prefix is $\hat{\mathbf{e}}_2$ so "2" won't appear in the following indices which are in the order 3, 1 and then 1, 3 (right-handed then

† There is a close analogy with the variable in an integration. You know that,

$$\int f(x)\mathrm{d}x = \int f(y)\mathrm{d}y = \int f(z)\mathrm{d}z\ldots$$

left-handed). In the third row, do the same: the indices to use are 1 and 2 and they go in right and then left handed sequence, 1, 2 then 2, 1. You will recover equation (1.1).

There is an object called the alternating tensor, or Levi-Civita symbol which has these properties,

$$\epsilon_{ijk} = \begin{cases} 0 & \text{if any of the } i,\ j \text{ or } k \text{ are the same} \\ 1 & \text{if } i \to j \to k \text{ is in right-handed sequence} \\ -1 & \text{if } i \to j \to k \text{ is in left-handed sequence} \end{cases}$$

In terms of this symbol if

$$\mathbf{w} = \mathbf{u} \times \mathbf{v}$$

then

$$w_i = \epsilon_{ijk} u_j v_k \tag{1.2}$$

Note that

$$\mathbf{u} \times \mathbf{v} = -\mathbf{v} \times \mathbf{u}$$

You can convince yourself that (1.2) is equivalent to (1.1). For example

$$w_1 = \epsilon_{1jk} u_j v_k = u_2 v_3 - u_3 v_2$$

Another useful symbol is the *Kronecker delta*

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

A useful identity is

$$\epsilon_{ijk}\epsilon_{i\ell m} = \delta_{j\ell}\delta_{km} - \delta_{jm}\delta_{k\ell}$$

Again, there's an easy way to remember this. First write the equation without indices on the right-hand side,

$$\epsilon_{ijk}\epsilon_{i\ell m} = \delta_{\_\_}\delta_{\_\_} - \delta_{\_\_}\delta_{\_\_}$$

then fill in the blanks: i won't appear because it's a dummy index and is summed over on the left-hand side; we have from the left-hand side these indices in this order $jk\ell m$. Fill in the j and k

$$\epsilon_{ijk}\epsilon_{i\ell m} = \delta_{j\_}\delta_{k\_} - \delta_{j\_}\delta_{k\_}$$

then fill in the ℓ and m and you just have to remember that in the second term, they go in in reverse:

$$\epsilon_{ijk}\epsilon_{i\ell m} = \delta_{j\ell}\delta_{km} - \delta_{jm}\delta_{k\ell}$$

all done.

An example of the use of the Kronecker delta is

$$\hat{\mathbf{e}}_i \cdot \hat{\mathbf{e}}_j = \delta_{ij}$$

Another example is

$$v_i u_j \delta_{ij} = v_i u_i \,(= v_1 u_1 + v_2 u_2 + v_3 u_3)$$

In the sum

$$\sum_{j=1}^{3} u_i \delta_{ij} = u_j$$

δ_{ij} *picks out* the term for which $j = i$ and "rejects" the other two by multiplying them by zero.

1.2 Vector calculus

1.2.1 Fields and the "del" operator

A field is a function that describes a physical quantity at all points in space. For a scalar field that quantity is specified by a single number, for example, temperature, density, pressure, gravitational or electric potential, $V(\mathbf{r})$. For a vector field such as electric field, $\mathbf{E}(\mathbf{r})$ or magnetic field, $\mathbf{B}(\mathbf{r})$, the quantity is a vector.

We want to know how the field quantities, in particular V, $\mathbf{E}$ and $\mathbf{B}$, vary with position, $\mathbf{r}$: that is, how they change as I move about in space. In the first instance then I want to know how rapidly they vary close to a point $\mathbf{r}$. For example, if I'm on a mountain, then I am interested in how the height above sea level $h(x_1, x_2)$ changes as I walk east (x_1, say) or north (x_2, say). Then I know whether I ascend or descend as I move and how steeply. The mathematical object that measures the slope is the vector called "grad" or "del",

$$\begin{aligned}\boldsymbol{\nabla} &= \hat{\mathbf{e}}_1 \frac{\partial}{\partial x_1} + \hat{\mathbf{e}}_2 \frac{\partial}{\partial x_2} + \hat{\mathbf{e}}_3 \frac{\partial}{\partial x_3} \\ &= \hat{\mathbf{e}}_i \frac{\partial}{\partial x_i}\end{aligned}$$

and the second line expresses the first using the Einstein convention. $\boldsymbol{\nabla}$ is an operator and as such is *naked* until it has a function on which to operate. I will introduce the notation

$$\partial_i = \frac{\partial}{\partial x_i}$$

so that I can write in compact form

$$\boldsymbol{\nabla} = \hat{\mathbf{e}}_i \partial_i$$

The vector $\boldsymbol{\nabla}$ can do three things.

1. It can act on a scalar field. $\boldsymbol{\nabla} V(\mathbf{r}) = \boldsymbol{\nabla} V(x_1, x_2, x_3)$ is a vector that points in the direction of most rapid increase of V in space and whose magnitude is the slope in that direction. For obvious reasons $\boldsymbol{\nabla} V$ is called the gradient of V,

$$\begin{aligned}\text{grad}V = \boldsymbol{\nabla} V &= \hat{\mathbf{e}}_1 \frac{\partial V}{\partial x_1} + \hat{\mathbf{e}}_2 \frac{\partial V}{\partial x_2} + \hat{\mathbf{e}}_3 \frac{\partial V}{\partial x_3} \\ &= \hat{\mathbf{e}}_i \partial_i V\end{aligned}$$

   You already know that in electrostatics,

$$\mathbf{E} = -\boldsymbol{\nabla} V$$

2. It can act on a vector field as a dot-product. If

$$\mathbf{E} = \hat{\mathbf{e}}_1 E_1 + \hat{\mathbf{e}}_2 E_2 + \hat{\mathbf{e}}_3 E_3 = \hat{\mathbf{e}}_i E_i$$

then "grad dot $\mathbf{E}$" is the divergence of $\mathbf{E}$,

$$\begin{aligned}\text{div}\mathbf{E} = \boldsymbol{\nabla}\cdot\mathbf{E} &= \frac{\partial E_1}{\partial x_1} + \frac{\partial E_2}{\partial x_2} + \frac{\partial E_3}{\partial x_3} \\ &= \partial_i E_i\end{aligned}$$

3. The cross product of $\boldsymbol{\nabla}$ with $\mathbf{E}$ is the circulation or curl of $\mathbf{E}$,

$$\begin{aligned}\text{curl}\mathbf{E} = \boldsymbol{\nabla}\times\mathbf{E} &= \hat{\mathbf{i}}\left(\frac{\partial E_3}{\partial x_2} - \frac{\partial E_2}{\partial x_3}\right) + \hat{\mathbf{j}}\left(\frac{\partial E_1}{\partial x_3} - \frac{\partial E_3}{\partial x_1}\right) + \hat{\mathbf{k}}\left(\frac{\partial E_2}{\partial x_1} - \frac{\partial E_1}{\partial x_2}\right) \\ &= \hat{\mathbf{e}}_1\left(\partial_2 E_3 - \partial_3 E_2\right) + \hat{\mathbf{e}}_2\left(\partial_3 E_1 - \partial_1 E_3\right) + \hat{\mathbf{e}}_3\left(\partial_1 E_2 - \partial_2 E_1\right)\end{aligned}$$

In the first line is the conventional way to write the curl, the second line is neater and illustrates how using the rules for constructing equation (1.1) you can never make a mitsake writing out the curl of a vector. There is an even more compact way to write the curl: if $\mathbf{u} = \boldsymbol{\nabla}\times\mathbf{v}$, then its i^{th} component is

$$u_i = \epsilon_{ijk}\partial_j v_k$$

1.2.2 Laplace operator

In cartesian coordinates,†

$$\begin{aligned}\boldsymbol{\nabla}\cdot\boldsymbol{\nabla} = \nabla^2 &= \frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} + \frac{\partial^2}{\partial x_3^2} \\ &= \partial_i\partial_i\end{aligned}$$

This is called the Laplace operator. It can act on a scalar function (famously the kinetic energy in quantum mechanics is $-(\hbar^2/2m)\nabla^2\psi$) or a vector function. Acting on a scalar field, $f(\mathbf{r})$,

$$\nabla^2 f = \frac{\partial^2 f}{\partial x_1^2} + \frac{\partial^2 f}{\partial x_2^2} + \frac{\partial^2 f}{\partial x_3^2} = \partial_i\partial_i f$$

Acting on a vector field, $\mathbf{u}(\mathbf{r})$, the Laplacian produces a vector whose 1-component is the Laplacian of u_1 and so on...

$$\begin{aligned}\nabla^2\mathbf{u} &= \left(\nabla^2 u_1\right)\hat{\mathbf{i}} + \left(\nabla^2 u_2\right)\hat{\mathbf{j}} + \left(\nabla^2 u_3\right)\hat{\mathbf{k}} \\ &= \left(\nabla^2 u_i\right)\hat{\mathbf{e}}_i\end{aligned}$$

† Mathematical physicists use Δ as a symbol for the Laplace operator; and use the prefixes "grad", "div" and "curl", rather than using the symbol $\boldsymbol{\nabla}$ as if it were a vector. This is mathematically more general than what I'm doing here, which is valid only in a Cartesian coordinate system.

1.2.3 Product rules

There is a number of *product rules* that you need to know (but not, of course, to remember). For any scalar functions, $f(\mathbf{r})$ and $g(\mathbf{r})$, and vector functions, $\mathbf{u}(\mathbf{r})$ and $\mathbf{v}(\mathbf{r})$,

$$\boldsymbol{\nabla}(fg) = f(\boldsymbol{\nabla}g) + g(\boldsymbol{\nabla}f)$$

$$\boldsymbol{\nabla}(\mathbf{u}\cdot\mathbf{v}) = \mathbf{u}\times(\boldsymbol{\nabla}\times\mathbf{v}) + \mathbf{v}\times(\boldsymbol{\nabla}\times\mathbf{u}) + (\mathbf{u}\cdot\boldsymbol{\nabla})\mathbf{v} + (\mathbf{v}\cdot\boldsymbol{\nabla})\mathbf{u}$$

$$\boldsymbol{\nabla}(f\mathbf{u}) = f(\boldsymbol{\nabla}\cdot\mathbf{u}) + \mathbf{u}(\boldsymbol{\nabla}f)$$

$$\boldsymbol{\nabla}\cdot(\mathbf{u}\times\mathbf{v}) = \mathbf{v}(\boldsymbol{\nabla}\times\mathbf{u}) + \mathbf{u}(\boldsymbol{\nabla}\times\mathbf{v})$$

$$\boldsymbol{\nabla}\times(f\mathbf{u}) = f(\boldsymbol{\nabla}\times\mathbf{u}) - \mathbf{u}\times(\boldsymbol{\nabla}f)$$

$$\boldsymbol{\nabla}\times(\mathbf{u}\times\mathbf{v}) = (\mathbf{v}\cdot\boldsymbol{\nabla})\mathbf{u} - (\mathbf{u}\cdot\boldsymbol{\nabla})\mathbf{v} + \mathbf{u}(\boldsymbol{\nabla}\cdot\mathbf{v}) - \mathbf{v}(\boldsymbol{\nabla}\cdot\mathbf{u})$$

As an exercise, write these all out using index notation and the Einstein summation convention. Then prove them if you like.

1.2.4 Second derivative rules

The second derivative rules are,

$$\begin{aligned}\boldsymbol{\nabla}\cdot(\boldsymbol{\nabla}\times\mathbf{u}) &= 0\\ \boldsymbol{\nabla}\times(\boldsymbol{\nabla}f) &= \mathbf{0}\\ \boldsymbol{\nabla}\times(\boldsymbol{\nabla}\times\mathbf{u}) &= \boldsymbol{\nabla}(\boldsymbol{\nabla}\cdot\mathbf{u}) - \nabla^2\mathbf{u}\end{aligned} \tag{1.3}$$

The first two are very important and are almost intuitively obvious. In words,

The divergence of a curl is always zero
The curl of a gradient is always zero

It's easiest to prove these using index notation. For example, to prove that $\boldsymbol{\nabla}\times(\boldsymbol{\nabla}f) = \mathbf{0}$, let $\boldsymbol{\nabla}\times(\boldsymbol{\nabla}f) = \mathbf{w}$ and let $\mathbf{u} = \boldsymbol{\nabla}f$. Then,

$$u_i = \partial_i f$$

and

$$w_i = \epsilon_{ijk}\partial_j u_k = \epsilon_{ijk}\partial_j\partial_k f$$

For the case $i = 1$,

$$\begin{aligned}w_1 &= \epsilon_{1jk}\partial_j\partial_k f\\ &= \epsilon_{123}\partial_2\partial_3 f + \epsilon_{132}\partial_3\partial_2 f\\ &= \partial_2\partial_3 f - \partial_3\partial_2 f = 0\end{aligned}$$

as long as f is "well-behaved" in the sense the order of differentation may be exchanged. The same is true for w_2 and w_3, hence

$$\mathbf{w} = \mathbf{0} \qquad \text{Q.E.D.}$$

Now let us show that

$$\boldsymbol{\nabla} \cdot (\boldsymbol{\nabla} \times \mathbf{u}) = 0$$

This is a dot product so the result is a scalar; call it f. Then

$$f = \boldsymbol{\nabla} \cdot (\boldsymbol{\nabla} \times \mathbf{u})$$

Let $\mathbf{v} = \boldsymbol{\nabla} \times \mathbf{u}$, then $f = \boldsymbol{\nabla} \cdot \mathbf{v} = \partial_i v_i$ and $v_i = \epsilon_{ijk}\partial_j u_k$, so that

$$f = \partial_i \epsilon_{ijk} \partial_j u_k = \epsilon_{ijk} \partial_i \partial_j u_k$$

You can now take separately the three terms, $i = 1, 2, 3$, and show they are all zero for a well-behaved function $\mathbf{u}$. But there's a way to see in your head that $f = 0$. Consider

$$\epsilon_{ijk} \partial_i \partial_j u_k$$

All three indices are summed-over, dummy indices; so I can swap them without affecting the outcome. So

$$\epsilon_{ijk}\, \partial_i \partial_j u_k = \epsilon_{jik}\, \partial_j \partial_i u_k$$

The first term, ϵ_{ijk} is an <u>odd</u> function in that it changes sign when I interchange i and j; whereas the second term is an *even* function since I just interchange the order of differentiation when I interchange i and j. The product of an odd and even function is zero; or put another way, the left and right hand sides are equal but also the one is equal to minus the other, so both sides must be zero. Hence

$$f = \epsilon_{ijk}\, \partial_i \partial_j u_k = \epsilon_{jik}\, \partial_j \partial_i u_k = 0 \qquad \text{Q.E.D}$$

Finally let us show that

$$\boldsymbol{\nabla} \times (\boldsymbol{\nabla} \times \mathbf{u}) = \boldsymbol{\nabla} (\boldsymbol{\nabla} \cdot \mathbf{u}) - \nabla^2 \mathbf{u}$$

Let $\mathbf{w} = \boldsymbol{\nabla} \times \mathbf{u}$, that is, $w_i = \epsilon_{ijk}\partial_j u_k$. Then the left hand side is, say,

$$\mathbf{v} = \boldsymbol{\nabla} \times (\boldsymbol{\nabla} \times \mathbf{u}) = \boldsymbol{\nabla} \times \mathbf{w}$$

So

$$\begin{aligned} v_\ell &= \epsilon_{\ell mi} \partial_m w_i \\ &= \epsilon_{\ell mi} \partial_m \epsilon_{ijk} \partial_j u_k \\ &= \epsilon_{\ell mi} \epsilon_{ijk} \partial_m \partial_j u_k \\ &= \epsilon_{i\ell m} \epsilon_{ijk} \partial_m \partial_j u_k \\ &= (\delta_{\ell j}\delta_{mk} - \delta_{\ell k}\delta_{mj})\, \partial_m \partial_j u_k \\ &= \partial_k \partial_\ell u_k - \partial_j \partial_j u_\ell \end{aligned}$$

Take the 1-component,

$$\begin{aligned} v_1 &= \partial_1 \partial_k u_k - \partial_j \partial_j u_1 \\ &= \frac{\partial}{\partial x_1}\left(\frac{\partial u_1}{\partial x_1} + \frac{\partial u_2}{\partial x_2} + \frac{\partial u_3}{\partial x_3}\right) - \left(\frac{\partial^2 u_1}{\partial x_1^2} + \frac{\partial^2 u_1}{\partial x_2^2} + \frac{\partial^2 u_1}{\partial x_3^2}\right) \end{aligned}$$

We recognise this as the 1-component of

$$\mathbf{v} = \boldsymbol{\nabla} (\boldsymbol{\nabla} \cdot \mathbf{u}) - \nabla^2 \mathbf{u} \qquad \text{Q.E.D}$$

1.2.5 Gauss's theorem

Gauss's theorem is also called the *divergence theorem*. It converts a volume integral into a surface integral. In words it may be stated that the flux of any vector field integrated over a closed surface is proportional to the sum or integral of the strengths of all sources enclosed within the surface. For a vector field $\mathbf{u}(\mathbf{r})$,

$$\int_\Omega (\boldsymbol{\nabla} \cdot \mathbf{u})\, \mathrm{d}\tau = \oint_S \mathbf{u} \cdot \mathrm{d}\mathbf{a}$$

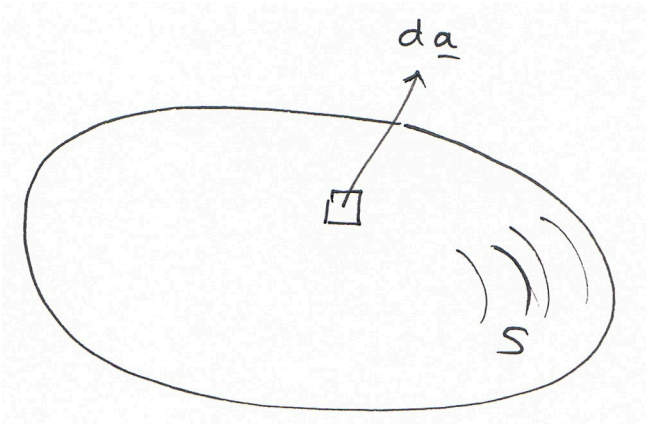


FIGURE 1–4

Here Ω is the volume enclosed by the surface S; $\mathrm{d}\tau$ is shorthand for $\mathrm{d}x_1\mathrm{d}x_2\mathrm{d}x_3$[†] and $\mathrm{d}\mathbf{a}$ is shorthand for $\hat{\mathbf{n}}\mathrm{d}a$ in which $\mathrm{d}a$ is an infinitesimal increment of the surface, having an unit normal vector $\hat{\mathbf{n}}$. In index notation, Gauss's law reads,

$$\int_\Omega \partial_i u_i \mathrm{d}\tau = \oint_S u_i \mathrm{d}a_i$$

You have already encountered this theorem in Gauss's law in electrostatics.

† $\int \mathrm{d}\tau = \int\int\int \mathrm{d}x_1\mathrm{d}x_2\mathrm{d}x_3$

1.2.6 Stokes's theorem

Stokes's theorem converts a surface integral into a line integral. If we consider a closed line, Γ, and any associated open surface (that is any surface whose boundary is Γ) then for any vector field, $\mathbf{u}(\mathbf{r})$, the curl of $\mathbf{u}$ integrated over the surface, S, is equal to the integral of $\mathbf{u}$ taken around Γ.

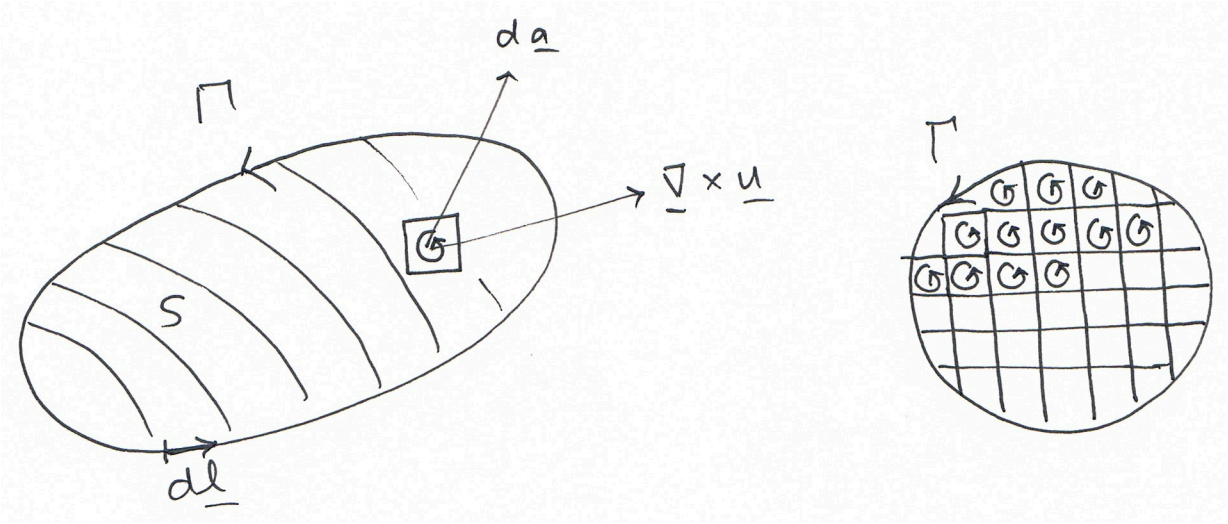


FIGURE 1–5

This may be evident to you from the figure. If I divide up the surface into its infinitesimal areas, d$\mathbf{a}$, and I draw the circulation at each location then you see that at each divide between areas there is an up (or left) arrow that cancels a down (or right) arrow so the only surviving arrows are the circulation around the boundary.

In symbols,

$$\int_S (\boldsymbol{\nabla} \times \mathbf{u}) \cdot \mathrm{d}\mathbf{a} = \oint_\Gamma \mathbf{u} \cdot \mathrm{d}\boldsymbol{\ell}$$

In index notation, Stokes's law reads,

$$\int_S \epsilon_{ijk} \partial_j u_k \mathrm{d}a_i = \oint_\Gamma u_i \mathrm{d}\ell_i$$

1.2.7 Green's theorem

We shan't need this, but for completeness here is Green's theorem. For any scalar fields, $f(\mathbf{r})$ and $g(\mathbf{r})$,

$$\int_\Omega \left(f\nabla^2 g - g\nabla^2 f \right) \mathrm{d}\tau = \oint_S \left(f\boldsymbol{\nabla} g - g\boldsymbol{\nabla} f \right) \cdot \mathrm{d}\mathbf{a}$$

where the domains of integration are as in Gauss's theorem.

1.3 The Helmholtz decomposition

You will need to read section 2, Revision of Electrostatics and Magnetostatics, and complete Problems Class 1, before proceeding further.

Let us consider a few facts.

1. In electrostatics the electric field is *irrotational*: $\boldsymbol{\nabla} \times \mathbf{E} = \mathbf{0}$.
2. In consequence the static electric field may be written as (minus) the gradient of a scalar field, the *electric potential*: $\mathbf{E} = -\boldsymbol{\nabla} V$.
3. According to Gauss's law, the divergence of the electric field is proportional to the quantity of *source* of the field: $\boldsymbol{\nabla} \cdot \mathbf{E} = \rho/\epsilon_0$.
4. Take a look at figure 2–1 in section 2 (reproduced below) and recall the convention that a *source point* is denoted $\mathbf{r}_1$ and a *field point* $\mathbf{r}_0$. Their vector difference, $\mathbf{R} = \mathbf{r}_0 - \mathbf{r}_1$ is called the *separation vector*. By reference to figure 2–1, the electric potential at the *field* point in terms of the *source* charge density is, equation (2.1.1),

$$V(\mathbf{r}_0) = \frac{1}{4\pi\epsilon_0} \int_\Omega \frac{\rho(\mathbf{r}_1)}{R}\, \mathrm{d}\tau_1 \tag{2.1.1}$$

5. The magnetic field, $\mathbf{B}$, by contrast has zero divergence but non zero curl.
6. There is no magnetic scalar potential, but we will learn much later in section 5, that there is a magnetic *vector* potential.
7. By analogy with (2.1.1) we will learn in section 5 that the vector potential is

$$\mathbf{A}(\mathbf{r}_0) = \frac{\mu_0}{4\pi} \int_\Omega \frac{\mathbf{J}(\mathbf{r_1})}{R}\, \mathrm{d}\tau_1 \tag{5.2.5}$$
$$\frac{\mu_0}{4\pi} = \frac{1}{c^2}\,\frac{1}{4\pi\epsilon_0}$$

in which the *source* of the vector potential is the current density, $\mathbf{J}$, see figure 2–6 (reproduced below)

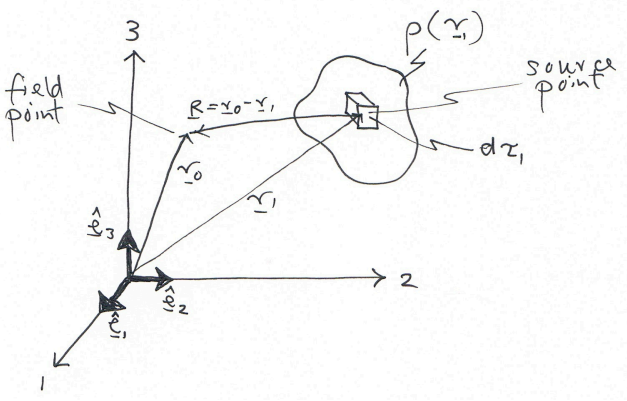


FIGURE 2–1

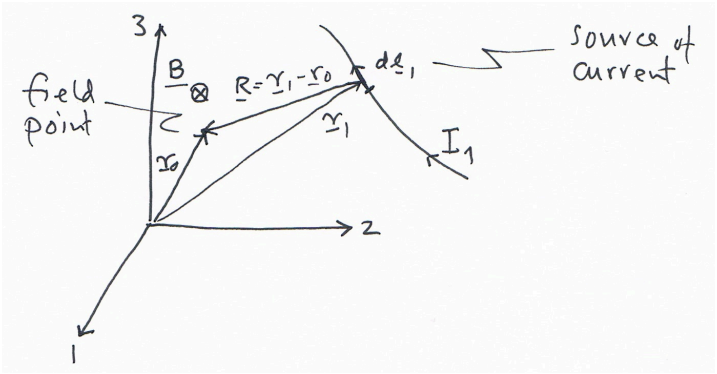


FIGURE 2–6

We don't need any of this in what follows, which is purely mathematical, but it does suggest to us that a vector field, $\mathbf{F}$, may arise from a source of divergence and a source of curl, each of which will act as sources to a *potential* that is developed at a field point.

Irrespective of that, the *mathematical* question is, can any vector field, $\mathbf{F}$, be *decomposed* into a part that is pure curl and a part that is pure divergence; and is the decomposition unique? We will now prove that the answer to both these is "yes". A corollary to this is that the field can be decomposed into a part that is divergence-free and a part that is curl-free. A further corollary of this is that any vector field is fully specified if both its curl and its divergence are given. An important *caveat* is that there must be no sources at infinity and the potentials must decay with distance from the sources at least as fast as $\sim 1/R$—which they do in (2.1.1) and (5.2.5).

Before we start the proof of Helmholtz's theorem, let me make a point about the domain, Ω of integration in (2.1.1) and (1.3). In accord with figure 2–1 we use a volume that encloses only the charge that is of interest to us. There may be other charges in the Universe, but we are interested in the field at $\mathbf{r}_0$ *due only to the charge within* Ω. By the principle of superposition, the remaining charges in the Universe will not affect the answer; and if we seek the total field we will need to include them. The same goes for figure 2–6 although the problem is more vexed since we cannot conceive of a finite length of current carrying wire. But again the principle of superposition applies to the Biot–Savart law so we can notionally divide up a current carrying wire into segments and ask what is the magnetic field due to each segment (see Problems Class 2, question 1). For *mathematicians* the use of finite integration domains is problematic[†] So we assert that the charge or current contained within Ω is *all the charge or current in the Universe* and we extend the domain to "all space", $\mathbb{R}^3$.[‡]

To begin, we write that the divergence of a general vector field is

$$\boldsymbol{\nabla}\cdot\mathbf{F} = s \tag{1.4a}$$

so that s is the source of a scalar potential, that we will call V. The curl of the field is

$$\boldsymbol{\nabla}\times\mathbf{F} = \mathbf{c} \tag{1.4b}$$

so that $\mathbf{c}$ is the source of a vector potential we will call $\mathbf{A}$. The *Helmholtz decomposition* into a part that is curl-free plus a part that is divergence-free is,

$$\mathbf{F} = -\boldsymbol{\nabla}V + \boldsymbol{\nabla}\times\mathbf{A} \tag{1.5a}$$

and the first term is curl-free since the curl of a div is zero, and the second term is divergence-free because the div of a curl is zero. We now prove[§] that if

$$V(\mathbf{r}_0) = \frac{1}{4\pi}\int_{\mathbb{R}^3}\frac{s(\mathbf{r}_1)}{R}\,\mathrm{d}\tau_1 \tag{1.5b}$$

$$\mathbf{A}(\mathbf{r}_0) = \frac{1}{4\pi}\int_{\mathbb{R}^3}\frac{\mathbf{c}(\mathbf{r}_1)}{R}\,\mathrm{d}\tau_1 \tag{1.5c}$$

---

† see for example `https://math.stackexchange.com/questions/2426369/problem-of-integrating-dirac-delta-function-singular-point-at-integration-limit`

‡ See Griffiths, *Introduction to Electrodynamics*, 4th ed., section 2.2.2.

§ For other proofs, see Griffiths, *loc. cit.* Appendix B; and your 5CCP3000 notes, section 3.6.4.3.

then (1.5) is consistent with (1.4).

First let us show that with V expressed by (1.5b), (1.5a) leads to (1.4a). I take the divergence of (1.5a) at a field point,†

$$\begin{aligned}\boldsymbol{\nabla}\cdot\mathbf{F}(\mathbf{r}_0) &= -\nabla^2 V(\mathbf{r}_0)\\ &= -\frac{1}{4\pi}\nabla^2\int_{\mathbb{R}^3}\frac{s(\mathbf{r}_1)}{R}\,\mathrm{d}\tau_1\\ &= -\frac{1}{4\pi}\int_{\mathbb{R}^3} s(\mathbf{r}_1)\nabla^2\frac{1}{R}\,\mathrm{d}\tau_1\\ &= \int_{\mathbb{R}^3} s(\mathbf{r}_1)\delta(\mathbf{r}_0-\mathbf{r}_1)\,\mathrm{d}\tau_1\\ &= s(\mathbf{r}_0)\end{aligned} \tag{1.6}$$

Let me explain the steps. In the first line, the second term in (1.5a) disappears because the div of a curl is zero. In the second line I inserted (1.5b). s is a function of only the *source point* coordinates, while ∇^2 acts on the *field point* coordinates so in the third line the Laplace operator can be moved forward to act on $1/R$ which depends on both source and field coordinates. In the fourth line use is made of the identity (see Problems Class 1) $\nabla^2(1/R) = -4\pi\delta(\mathbf{R})$ and by the properties of the Dirac "delta function" this picks out from the integral just the integrand, s, evaluated at the *field point*, $\mathbf{r}_0$.

Next we need to ask whether the curl of $\mathbf{F}$ in (1.5a) is consistent with (1.4b). Taking the curl of (1.5a) only the second term survives because the curl of a div is zero,

$$\begin{aligned}\boldsymbol{\nabla}\times\mathbf{F} &= \boldsymbol{\nabla}\times(\boldsymbol{\nabla}\times\mathbf{A})\\ &= \boldsymbol{\nabla}(\boldsymbol{\nabla}\cdot\mathbf{A}) - \nabla^2\mathbf{A} \quad 3^{\text{rd}}\text{ second derivative rule}\\ &= \boldsymbol{\nabla}\left(\boldsymbol{\nabla}\cdot\frac{1}{4\pi}\int_{\mathbb{R}^3}\frac{\mathbf{c}(\mathbf{r}_1)}{R}\,\mathrm{d}\tau_1\right) - \nabla^2\frac{1}{4\pi}\int_{\mathbb{R}^3}\frac{\mathbf{c}(\mathbf{r}_1)}{R}\,\mathrm{d}\tau_1\\ &= \frac{1}{4\pi}\left(\int_{\mathbb{R}^3}(\mathbf{c}(\mathbf{r}_1)\cdot\boldsymbol{\nabla})\,\boldsymbol{\nabla}\frac{1}{R}\,\mathrm{d}\tau_1 - \int_{\mathbb{R}^3}\mathbf{c}(\mathbf{r}_1)\nabla^2\frac{1}{R}\,\mathrm{d}\tau_1\right)\end{aligned} \tag{1.7}$$

Here, again, because $\boldsymbol{\nabla}$ and ∇^2 act on the field point coordinates of $\mathbf{r}_0$, and not on $\mathbf{r}_1$, they can be moved up against the function $1/R$. If we could show that the first term is zero, then by the same argument as above we can use the identity $\nabla^2(1/R) = -4\pi\delta(\mathbf{R})$ and we will arrive at the result,

$$\boldsymbol{\nabla}\times\mathbf{F}(\mathbf{r}_0) = \int_{\mathbb{R}^3}\mathbf{c}(\mathbf{r}_1)\,\delta(\mathbf{r}_0-\mathbf{r}_1)\,\mathrm{d}\tau_1 = \mathbf{c}(\mathbf{r}_0) \tag{1.8}$$

which is (1.4b) as required for the proof.

---

† Be aware that when we write equations like (1.4) it is implicit that the left and right hand sides are evaluated at the same point in space. For example the differential form of Gauss's law, $\boldsymbol{\nabla}\cdot\mathbf{E} = \rho/\epsilon_0$ means that the divergence of $\mathbf{E}$ is equal to the source density *at that field point*. If there is no charge density at the field point then $\boldsymbol{\nabla}\cdot\mathbf{E} = 0$.

Now let's examine the first term in (1.7), which we hope to show is zero. First to be as clear as possible, recall Problems Class 1 and note that what we've been calling $\boldsymbol{\nabla}$ is actually $\boldsymbol{\nabla}_0$, namely the operator acting on the field coordinates. When it's to act on the source coordinates we use the symbol $\boldsymbol{\nabla}_1$. You remember that $\boldsymbol{\nabla}_1 = -\boldsymbol{\nabla}_0$. The integral that we hope will vanish is,

$$\begin{aligned}\mathbf{I} &= \int_{\mathbb{R}^3} (\mathbf{c}(\mathbf{r}_1)\cdot\boldsymbol{\nabla}_0)\,\boldsymbol{\nabla}_0 \frac{1}{R}\,\mathrm{d}\tau_1 \\ &= \int_{\mathbb{R}^3} (\mathbf{c}(\mathbf{r}_1)\cdot\boldsymbol{\nabla}_1)\,\boldsymbol{\nabla}_1 \frac{1}{R}\,\mathrm{d}\tau_1\end{aligned}$$

Note that $\mathbf{I}$ is a *vector*. It has components I_1, I_2 and I_3. The 3−component, for example, is

$$I_3 = \int_{\mathbb{R}^3} \mathbf{c}\cdot\boldsymbol{\nabla}_1 \frac{\partial}{\partial z_1}\frac{1}{R}\,\mathrm{d}\tau_1 \tag{1.9}$$

Now, by the rule for the differentiation of a product,

$$\boldsymbol{\nabla}_1\cdot\left(\mathbf{c}\,\frac{\partial}{\partial z_1}\frac{1}{R}\right) = \mathbf{c}\cdot\left(\boldsymbol{\nabla}_1\frac{\partial}{\partial z_1}\frac{1}{R}\right) + \boldsymbol{\nabla}_1\cdot\mathbf{c}\,\frac{\partial}{\partial z_1}\frac{1}{R}$$

which means that (1.9) is,

$$I_3 = \int_{\mathbb{R}^3} \boldsymbol{\nabla}_1\cdot\left(\mathbf{c}\,\frac{\partial}{\partial z_1}\frac{1}{R}\right)\,\mathrm{d}\tau_1 - \int_{\mathbb{R}^3} \boldsymbol{\nabla}_1\cdot\mathbf{c}\,\frac{\partial}{\partial z_1}\frac{1}{R}\,\mathrm{d}\tau_1$$

Both these terms are zero: the second because the divergence of $\mathbf{c}$ is zero as follows from (1.4b), the div of a curl is zero; the first term by virtue of the divergence theorem,

$$\int_{\mathbb{R}^3} \boldsymbol{\nabla}_1\cdot\left(\mathbf{c}\,\frac{\partial}{\partial z_1}\frac{1}{R}\right)\,\mathrm{d}\tau_1 = \int_S \left(\mathbf{c}\,\frac{\partial}{\partial z_1}\frac{1}{R}\right)\cdot\mathrm{d}\mathbf{a}$$

The surface S is the boundary of $\mathbb{R}^3$ which is at infinity where the source, $\mathbf{c}$, vanishes. It is this last piece of the proof that sets the condition on the Helmholtz decomposition that it only applies to fields whose sources vanish at infinity. We have proved that $I_3 = 0$; the same will go for the two other components so we have proved that $\mathbf{I} = \mathbf{0}$ and so (1.8) is proved.

We may draw the following conclusions (which you may check against the facts we listed on page 10) from the Helmholtz decomposition of a vector field, $\mathbf{F}$, that is created by sources which vanish at infinity.

1. Once the sources $s(\mathbf{r})$ and $\mathbf{c}(\mathbf{r})$ are given then the field arising from them is uniquely given by (1.5). It can always be derived from a combination of scalar and vector potentials.
2. If there is only the source density s then $\mathbf{F}$ can be expressed as minus the gradient of a scalar potential. If there is only circulation density, then the field can be expressed as the curl of a vector potential. If both s and $\mathbf{c}$ are zero everywhere then $\mathbf{F}$ is zero everywhere.

3. If at the field point there are no sources, but there are sources elsewhere, then if these remote sources are "charges" s then the field satisfies $\nabla^2 V = 0$; and if the remote sources are "currents" $\mathbf{c}$ then the field is a solution of $\boldsymbol{\nabla} \times \boldsymbol{\nabla} \times \mathbf{A} = \mathbf{0}$. These last two, are so called *harmonic* differential equations.

The last point is important to grasp. You are familiar with the fact that at a field point where there is no charge, the electrostatic field is given by Gauss's law,

$$\boldsymbol{\nabla} \cdot \mathbf{E} = 0$$

The matter is rather subtle in the case of magnetostatics. Consider the magnetic field due to a long straight current carrying wire. At any point not actually *in* the wire the current density is zero and so (1.4b) implies that

$$\boldsymbol{\nabla} \times \mathbf{B} = \mathbf{0}$$

Since it is always true that $\boldsymbol{\nabla} \cdot \mathbf{B} = 0$, we have a magnetic field that has zero curl *and* zero divergence. This does not mean that the field itself is zero. In fact you know there is a field and you know how to sketch it and to calculate it; but contrary to what your sketch might look like its curl is actually zero—see Problems Class 1, C1.4e. In this example there is also a vector potential remote from the wire. It points in the direction of the current.

In point 1 just now I used the word *uniquely*. So if you're still with me and have courage for a final push, then let's prove that the Helmholtz decomposition is unique; again if there are no sources at infinity. Imagine there were two functions, $\mathbf{F}_1$ and $\mathbf{F}_2$ both arising from the same set of sources and both of which satisfied (1.4) simultaneously. Consider the difference function, $\mathbf{W} = \mathbf{F}_2 - \mathbf{F}_1$. Because (1.4) are linear equations, we would have,

$$\boldsymbol{\nabla} \cdot \mathbf{W} = 0$$
$$\boldsymbol{\nabla} \times \mathbf{W} = \mathbf{0}$$

everywhere, and $\mathbf{W} = \mathbf{0}$ at infinity. To prove that if we have found a field $\mathbf{F}_1$ that derives from a set of sources then there can be no other field that derives from the same sources, we need to show $\mathbf{W} = \mathbf{0}$ so that any other field we find is actually identical to the first. In view of the second equation, $\boldsymbol{\nabla} \times \mathbf{W} = \mathbf{0}$, we must have

$$\mathbf{W} = -\boldsymbol{\nabla}\psi \tag{1.10}$$

that is, since $\mathbf{W}$ is irrotational it must be minus the gradient of some scalar field, which we will call ψ. Similarly from the first equation, $\boldsymbol{\nabla} \cdot \mathbf{W} = 0$, if I apply the $\boldsymbol{\nabla}$ operator to both sides of (1.10) I get,

$$\nabla^2 \psi = 0 \tag{1.11}$$

Consider the vector field, $\psi \boldsymbol{\nabla}\psi$ and apply the divergence theorem to it,

$$\begin{aligned}
\int_{\mathbb{R}^3} \boldsymbol{\nabla} \cdot (\psi \boldsymbol{\nabla}\psi)\, \mathrm{d}\tau &= \int_{\mathbb{R}^3} \left(\psi \nabla^2 \psi + (\boldsymbol{\nabla}\psi)^2\right) \mathrm{d}\tau && \text{by the product rule for derivatives} \\
&= \int_S (\psi \boldsymbol{\nabla}\psi) \cdot \mathrm{d}\mathbf{a} && \text{by the divergence theorem} \\
&= 0 && \text{since the boundary is at infinity}
\end{aligned}$$

Since the first term in the parentheses on the right hand side of the first line is zero by virtue of (1.11) it must also be true that

$$\int_{\mathbb{R}^3} (\boldsymbol{\nabla}\psi)^2 = 0$$

and since $(\boldsymbol{\nabla}\psi)^2 = \mathbf{W}^2 \geq 0$ then $\mathbf{W} = \mathbf{0}$ and the uniqueness is established. Note that for $\psi\boldsymbol{\nabla}\psi$ to vanish at infinity it is not enough that there are no sources at infinity; the sources must also fall off with distance at least as fast as $\sim 1/r$ so that $\psi\boldsymbol{\nabla}\psi$ falls off at least as fast as $\sim 1/r^3$ to guarantee that the surface integral is zero.

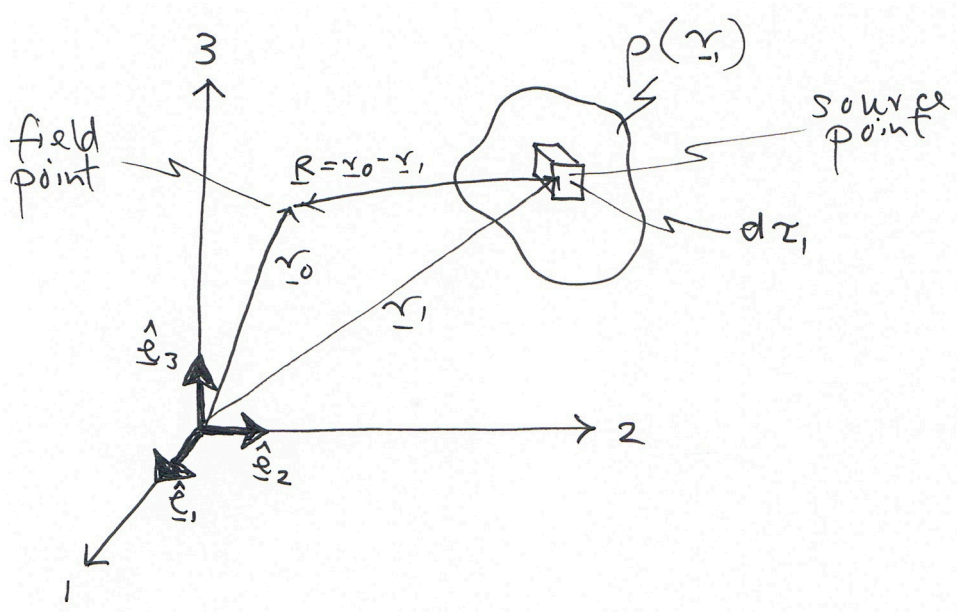
5CCP2380 Problems 1

Scalar and vector fields, for example, $V(\mathbf{r})$, $\mathbf{E}(\mathbf{r})$ and $\mathbf{B}(\mathbf{r})$, exist at points $\mathbf{r}$ in a three dimensional space which is spanned by the (dimensionless) unit vectors, $\hat{\mathbf{e}}_\mathbf{1}$, $\hat{\mathbf{e}}_\mathbf{2}$ and $\hat{\mathbf{e}}_\mathbf{3}$. The specification of these vectors is equivalent to fixing a choice of cartesian coordinate system $(x,\, y,\, z)$. Hence $\mathbf{r} = x\hat{\mathbf{e}}_\mathbf{1} + y\hat{\mathbf{e}}_\mathbf{2} + z\hat{\mathbf{e}}_\mathbf{3}$ means that x, y and z are the *components* of the vector $\mathbf{r}$. In vector calculus we employ the "grad" or "del" operator which takes partial derivatives of the field at the point $\mathbf{r}$, with respect to the components of $\mathbf{r}$:

$$\boldsymbol{\nabla} = \hat{\mathbf{e}}_\mathbf{1}\frac{\partial}{\partial x} + \hat{\mathbf{e}}_\mathbf{2}\frac{\partial}{\partial y} + \hat{\mathbf{e}}_\mathbf{3}\frac{\partial}{\partial z}$$

We will (almost) always denote the cartesian coordinates of a source point by $\mathbf{r}_1 = x_1\hat{\mathbf{e}}_\mathbf{1} + y_1\hat{\mathbf{e}}_\mathbf{2} + z_1\hat{\mathbf{e}}_\mathbf{3}$ and those of a field point by $\mathbf{r}_0 = x_0\hat{\mathbf{e}}_\mathbf{1} + y_0\hat{\mathbf{e}}_\mathbf{2} + z_0\hat{\mathbf{e}}_\mathbf{3}$. The vector pointing *from* the source *to* the field point is called the "separation vector"

$$\mathbf{R} = \mathbf{r}_0 - \mathbf{r}_1 = X\hat{\mathbf{e}}_\mathbf{1} + Y\hat{\mathbf{e}}_\mathbf{2} + Z\hat{\mathbf{e}}_\mathbf{3} = (x_0 - x_1)\hat{\mathbf{e}}_\mathbf{1} + (y_0 - y_1)\hat{\mathbf{e}}_\mathbf{2} + (z_0 - z_1)\hat{\mathbf{e}}_\mathbf{3}$$



The figure illustrates the ubiquitous geometry of electrostatics. There is a charge distribution, or distribution of point charges, $\rho(\mathbf{r})$. This is divided in the sense of the differential calculus into infinitesimal cubic elements, $\mathrm{d}\tau_1 = \mathrm{d}x_1\mathrm{d}y_1\mathrm{d}z_1$, each at a movable *source point*, $\mathbf{r}_1$. We are interested in the electric field (or the electric potential) that the distribution produces at the *field point*, $\mathbf{r}_0$. If the infinitesimal amount of charge within $\mathrm{d}\tau_1$ is $\mathrm{d}q$, then the electric field at the field point, according to Coulomb's law is

$$\mathrm{d}\mathbf{E}(\mathbf{r}_0) = \frac{1}{4\pi\epsilon_0}\,\frac{\mathrm{d}q}{R^2}\,\hat{\mathbf{R}}$$

The total electric field at $\mathbf{r}_0$ is the sum of these increments of field as I reposition the box at all values of $\mathbf{r}_1$ that range over the volume Ω containing the charge distribution. Summing the field over all these contributions in accord with the principle of superposition

leads to

$$\begin{aligned}\mathbf{E}(\mathbf{r}_0) &= \int \mathrm{d}x_1 \int \mathrm{d}y_1 \int \mathrm{d}z_1 \, \mathrm{d}\mathbf{E}(x_0, y_0, z_0) \\ &= \frac{1}{4\pi\epsilon_0} \int \frac{\rho(\mathbf{r}_1)}{R^2} \, \hat{\mathbf{R}} \, \mathrm{d}\tau_1\end{aligned}$$

This statement encapsulates the entirety of electrostatics, all of which is constructed on just the two postulates—Coulomb's law, and the principle of superposition. Generally this is a difficult intrgral because the connecting vector, $\mathbf{R}$, is a function of the source point vector, $\mathbf{r}_1$, which is the variable of integration. However the integral can be done in certain symmetric situations, such as the field due to a point charge, a line charge and a circular disk of charge (if the field point is above the centre of the disk). Some such problems can also be done using Gauss's law.

Griffiths uses a different notation: for $\mathbf{r}_0$ he uses $\mathbf{r}$, and for $\mathbf{r}_1$ he uses $\mathbf{r}'$. Then he uses a curious curly "r" for $\mathbf{R}$ which I can't even typeset, let alone write on the blackboard. He also uses for the dimensionless unit vectors, $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, $\hat{\mathbf{z}}$, instead of $\hat{\mathbf{e}}_\mathbf{1}$, $\hat{\mathbf{e}}_\mathbf{2}$, $\hat{\mathbf{e}}_\mathbf{3}$. In index notation which I will teach you briefly we use $\hat{\mathbf{e}}_i$ which is most economical and easiest to manipulate using the summation convention.

You can easily convince yourself that $\boldsymbol{\nabla} r = \hat{\mathbf{r}}$. But what do I mean by $\boldsymbol{\nabla} R$? The separation vector is a function of six coordinates, $\mathbf{R} = \mathbf{R}(x_0, y_0, z_0, x_1, y_1, z_1)$ so I cannot mean differentiate with respect to x, y and z because these variables to not appear in $\mathbf{R}$. So I must mean differentiate with respect to the components of *either* $\mathbf{r}_0$ *or* of $\mathbf{r}_1$. Physically what that means is, am I interested in how the field at the field point varies as I waggle the test charge, or as I waggle the source charge? I will distinguish these two instances by a subscript on $\boldsymbol{\nabla}$ in the following definitions,

$$\boldsymbol{\nabla}_0 = \hat{\mathbf{e}}_\mathbf{1} \frac{\partial}{\partial x_0} + \hat{\mathbf{e}}_\mathbf{2} \frac{\partial}{\partial y_0} + \hat{\mathbf{e}}_\mathbf{3} \frac{\partial}{\partial z_0}$$

and

$$\boldsymbol{\nabla}_1 = \hat{\mathbf{e}}_\mathbf{1} \frac{\partial}{\partial x_1} + \hat{\mathbf{e}}_\mathbf{2} \frac{\partial}{\partial y_1} + \hat{\mathbf{e}}_\mathbf{3} \frac{\partial}{\partial z_1}$$

When acting on functions $F = F(\mathbf{R})$ of the separation vector $\mathbf{R} = \mathbf{r}_0 - \mathbf{r}_1$, we have the identity $\boldsymbol{\nabla}_1 = -\boldsymbol{\nabla}_0$, which follows from the chain rule. See Exercise C1.1c.

In my lectures notes I may well get sloppy and use $\mathbf{r}$ for a separation vector. So be sure to check whether a vector is intended as a position vector or a separation vector. Also I may leave off the subscript to $\boldsymbol{\nabla}$ when it acts on a separation vector, in which case it is implied that $\boldsymbol{\nabla}$ operates on the field coordinates.

Do some or all of these problems in class

C1.1 For the separation vector,

$$\mathbf{R} = X\hat{\mathbf{e}}_\mathbf{1} + Y\hat{\mathbf{e}}_\mathbf{2} + Z\hat{\mathbf{e}}_\mathbf{3} = (x_0 - x_1)\hat{\mathbf{e}}_\mathbf{1} + (y_0 - y_1)\hat{\mathbf{e}}_\mathbf{2} + (z_0 - z_1)\hat{\mathbf{e}}_\mathbf{3}$$

let

$$R = \sqrt{(x_0 - x_1)^2 + (y_0 - y_1)^2 + (z_0 - z_1)^2}$$

(**a**) What is $\boldsymbol{\nabla}_0(R^n)$?

(**b**) By explicit differentiation show that

$$\boldsymbol{\nabla}_0 \left(\frac{1}{R}\right) = -\frac{\mathbf{R}}{R^3}$$

Is this consistent with the answer to C1.1a? Write down the electric field due to a point charge in terms of $\boldsymbol{\nabla}_0(1/R)$ and comment on the result.

(**c**) Show that

$$\boldsymbol{\nabla}_0 \left(\frac{1}{R}\right) = -\boldsymbol{\nabla}_1 \left(\frac{1}{R}\right)$$

(**d**) Show that

$$\nabla^2 \left(\frac{1}{R}\right) = -4\pi\, \delta(\mathbf{R})$$

where $\nabla^2 = \boldsymbol{\nabla}_0 \cdot \boldsymbol{\nabla}_0$ and $\delta(\mathbf{R})$ is the Dirac "delta function".

C1.2 Using Gauss's law and the divergence theorem, calculate the electric field due to a point charge and hence deduce Coulomb's law.

C1.3 A point charge is located at the origin of a Cartesian coordinate system. For any point $\mathbf{r}$ *not at the origin* show that the divergence of the electric field due to this charge is zero. Discuss this result in relation to Gauss's law.

C1.4 Consider the vector field in three dimensional space, with p a real number:

$$\mathbf{v}(\mathbf{r}) = \frac{1}{(x^2 + y^2)^p}\,(-y\hat{\mathbf{e}}_\mathbf{1} + x\hat{\mathbf{e}}_\mathbf{2})$$

(**a**) Sketch the vector field in the case $p = 0$ and guess what its curl might be. Give an example of a physical realisation of the field.

(**b**) Find the curl of $\mathbf{v}$.

(**c**) Discuss the cases: $p = 0$, $p = 1$. Comment on your finding when $p = 1$.

(**d**) In the case $p = 0$, find the line integral of $\mathbf{v}$ around a circular loop centred on the origin and verify that the field is consistent with Stokes's theorem.

(**e**) How does the case $p = 1$ relate to the magnetic field due to a long straight current-carrying wire?

5CCP2380 Problems 1—Solutions

C1.1a Do the 1-component:

$$\begin{aligned}(\boldsymbol{\nabla}_0 R^n)_1 &= \hat{\mathbf{i}}\,\frac{\partial}{\partial x_0}\left((x_0-x_1)^2+(y_0-y_1)^2+(z_0-z_1)^2\right)^{n/2}\\ &= \hat{\mathbf{i}}\,\frac{1}{2}n\left(R^2\right)^{\frac{1}{2}n-1}2\,(x_0-x_1)\\ &= \hat{\mathbf{i}}\,nR^{n-2}X\end{aligned}$$

Similarly for 2- and 3-components:

$$\begin{aligned}\boldsymbol{\nabla}_0 R^n &= nR^{n-2}\left(X\hat{\mathbf{i}}+Y\hat{\mathbf{j}}+Z\hat{\mathbf{k}}\right)\\ &= n\mathbf{R}R^{n-2}\\ &= n\hat{\mathbf{R}}R^{n-1}\end{aligned}$$

□

C1.1b Do the 1-component:

$$\begin{aligned}\left(\boldsymbol{\nabla}_0 R^{-1}\right)_1 &= \hat{\mathbf{i}}\,\frac{\partial}{\partial x_0}\left((x_0-x_1)^2+(y_0-y_1)^2+(z_0-z_1)^2\right)^{-\frac{1}{2}}\\ &= \hat{\mathbf{i}}\cdot-\frac{1}{2}\left(R^2\right)^{-\frac{3}{2}}2\,(x_0-x_1)\\ &= -\hat{\mathbf{i}}\,\frac{X}{R^3}\end{aligned}$$

Similarly for 2- and 3-components:

$$\boldsymbol{\nabla}_0 R^{-1} = -\frac{\mathbf{R}}{R^3} = -\frac{\hat{\mathbf{R}}}{R^2}$$

□

This is consistent with C1.1a. The electric field due to a point charge is

$$\begin{aligned}\mathbf{E} &= \frac{1}{4\pi\epsilon_0}\,q\,\frac{\hat{\mathbf{r}}}{r^2}\\ &= -\frac{1}{4\pi\epsilon_0}\,q\boldsymbol{\nabla}\left(\frac{1}{r}\right)\end{aligned}$$

This makes sense because we know that the electric field is equal to minus the gradient of a potential that goes like $1/r$.

C1.1c This is easy because the term like

$$\frac{\partial}{\partial x_0}\,(x_0-x_1)^2 = 2\,(x_0-x_1)$$

becomes

$$\frac{\partial}{\partial x_1}\,(x_0-x_1)^2 = -2\,(x_0-x_1)$$

so this picks up a minus sign.

C1.1d The Dirac "delta function" is not really a function; it's a distribution defined by its properties when associated with functions,

$$\delta(\mathbf{R}) = 0 \qquad \text{if } \mathbf{R} \neq \mathbf{0} \quad \text{that is, } \mathbf{r}_0 \neq \mathbf{r}_1$$
$$\int_{\mathbb{R}^3} \delta(\mathbf{R}) \mathrm{d}\tau_1 = 1$$

For any arbitrary function, $f(\mathbf{r}_1)$,

$$\int_{\mathbb{R}^3} f(\mathbf{r}_1)\delta(\mathbf{R})\mathrm{d}\tau_1 = \int_{\mathbb{R}^3} f(\mathbf{r}_1)\delta(\mathbf{r}_0 - \mathbf{r}_1)\mathrm{d}\tau_1 = f(\mathbf{r}_0)$$

We first work out Laplace of $1/R$. We use the result of C1.1b,

$$\begin{aligned}\nabla^2 \frac{1}{R} = \boldsymbol{\nabla}_0 \cdot \boldsymbol{\nabla}_0 \frac{1}{R} &= -\boldsymbol{\nabla}_0 \cdot \frac{\mathbf{R}}{R^3} \\ &= -\frac{\partial}{\partial x_0}\frac{X}{R^3} - \frac{\partial}{\partial y_0}\frac{Y}{R^3} - \frac{\partial}{\partial z_0}\frac{Z}{R^3}\end{aligned}$$

Working out the first term,

$$\frac{\partial}{\partial x_0}\frac{x_0 - x_1}{\left((x_0 - x_1)^2 + (y_0 - y_1)^2 + (z_0 - z_1)^2\right)^{\frac{3}{2}}} = \frac{1}{R^3} - 3\frac{X^2}{R^5}$$

and adding the next two, I get,

$$\nabla^2 \frac{1}{R} = -\boldsymbol{\nabla}_0 \cdot \frac{\mathbf{R}}{R^3} = -\frac{3}{R^3} + 3\frac{X^2 + Y^2 + Z^2}{R^5} = -\frac{3}{R^3} + \frac{3}{R^3}$$

This is zero *unless* $R = 0$ in which case it blows up. To show that this is equivalent to minus 4π times the "delta function" we need to prove that

$$\int_{\mathbb{R}^3} \nabla^2 \frac{1}{R} \mathrm{d}\tau_1 = -4\pi$$

Now according to the divergence theorem,

$$\begin{aligned}\int_{\mathbb{R}^3} \nabla^2 \frac{1}{R}\, \mathrm{d}\tau_1 &= \int_S \boldsymbol{\nabla}_0 \frac{1}{R} \cdot \mathrm{d}\mathbf{a}' \\ &= -\int_S \frac{1}{R^2}\hat{\mathbf{R}} \cdot \mathrm{d}\mathbf{a}' \\ &= -\int_S \mathrm{d}\Omega \\ &= -4\pi\end{aligned}$$

□

The second line follows from the result of C1.1b and the last line follows because $\int \mathrm{d}\Omega$ is the integral over the solid angle subtended at $\mathbf{r}_0$ over the integration surface S. This completes the proof that

$$\nabla^2 \left(\frac{1}{R}\right) = -4\pi\,\delta(\mathbf{R})$$

C1.2 We place the point charge, q, at the centre of a sphere of radius r. If we denote the volume domain by Ω and the boundary of the domain by S, then,

$$\int_\Omega \boldsymbol{\nabla}\cdot\mathbf{E}\,\mathrm{d}\tau = \int_\Omega \frac{1}{\epsilon_0}\rho\,\mathrm{d}\tau$$

is Gauss's law in integral form. The right hand side represents the total charge within the sphere, which is q. The left hand side is transformed into a surface integral using the divergence theorem,

$$\int_\Omega \boldsymbol{\nabla}\cdot\mathbf{E} = \oint_S \mathbf{E}\cdot\mathrm{d}\mathbf{a}$$

Now, the electric field is constant over the surface of the sphere by symmetry so we have,

$$\begin{aligned} E\oint_S \mathrm{d}a &= \frac{1}{\epsilon_0}\,q \\ E\cdot 4\pi r^2 &= \frac{1}{\epsilon_0}\,q \\ E &= \frac{1}{4\pi\epsilon_0}\,\frac{q}{r^2} \end{aligned}$$

which is the force per unit test charge at a distance r from q. Hence if the test charge is of q_0 Coulomb then,

$$F = \frac{1}{4\pi\epsilon_0}\,\frac{qq_0}{r^2}$$ □

which is Coulomb's law.

C1.3 The electric field at $\mathbf{r}$ is,

$$\mathbf{E}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\,q\,\frac{\hat{\mathbf{r}}}{r^2} = \frac{q}{4\pi\epsilon_0}\,\frac{\mathbf{r}}{r^3}$$

We calculate,

$$\boldsymbol{\nabla}\cdot\mathbf{E} = \frac{1}{4\pi\epsilon_0}\left(\frac{\partial}{\partial x}\frac{x}{r^3} + \frac{\partial}{\partial y}\frac{y}{r^3} + \frac{\partial}{\partial z}\frac{z}{r^3}\right)$$

Now,

$$\begin{aligned} \frac{\partial}{\partial x}\,x\,\left(x^2+y^2+z^2\right)^{-\frac{3}{2}} &= \frac{1}{r^3} - \frac{3}{2}\left(r^2\right)^{-\frac{3}{2}-1}\cdot 2x \\ &= \frac{1}{r^3} - \frac{3x^2}{r^5} \end{aligned}$$

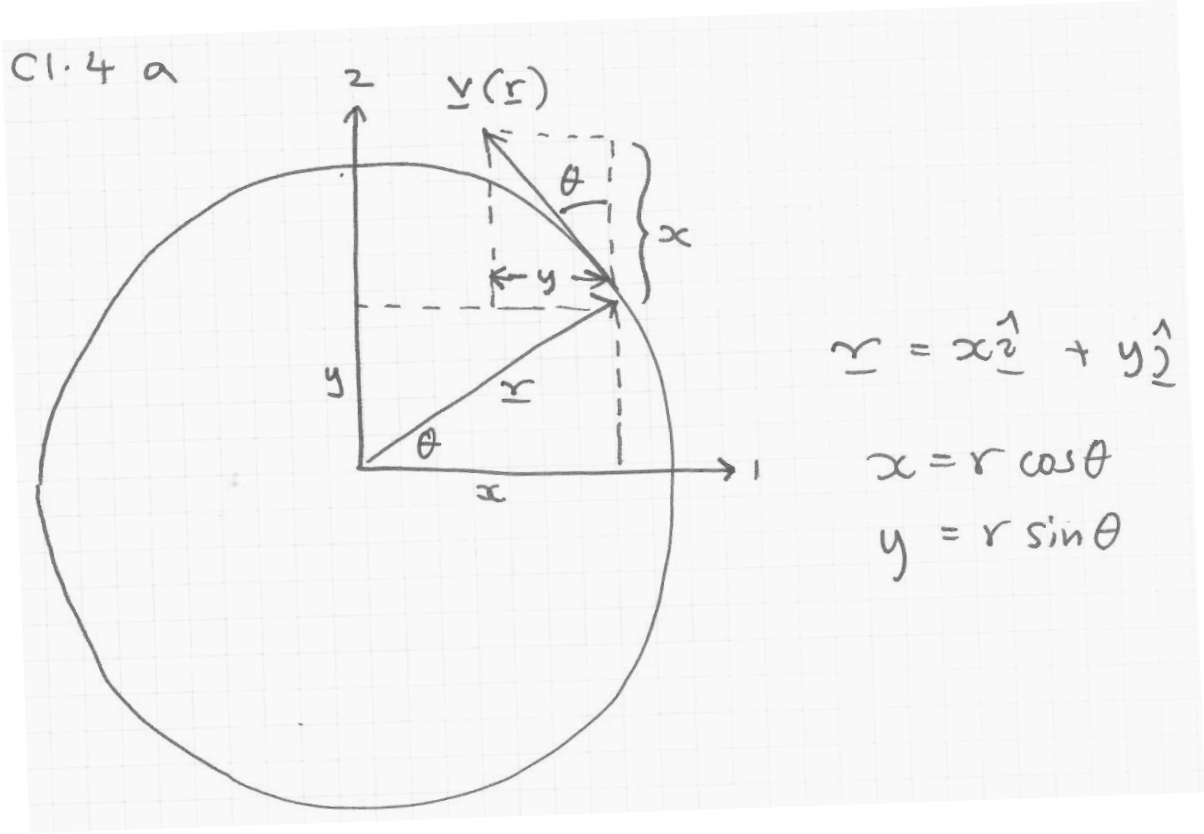
and similarly for the other two terms; adding them gives

$$\begin{aligned}\boldsymbol{\nabla}\cdot\mathbf{E} &= \frac{1}{4\pi\epsilon_0}\left(\frac{3}{r^3}-\frac{1}{r^5}\left(x^2+y^2+z^2\right)\right)\\ &= \frac{1}{4\pi\epsilon_0}\left(\frac{3}{r^3}-\frac{3r^2}{r^5}\right)\\ &= 0 \qquad r\neq 0\end{aligned}$$

At the origin, $r = 0$, this diverges. This is consistent with Gauss's law because,

$$\boldsymbol{\nabla}\cdot\mathbf{E} = \frac{1}{\epsilon_0}\rho = \frac{1}{\epsilon_0}\delta(\mathbf{r})$$

means the divergence is zero everywhere the charge density is zero and since the only charge is a point charge at the origin then the charge density is indeed zero everywhere else. This is neatly captured by use of the Dirac "delta function."



In the case, $p = 0$, this is the velocity field of a point on a spinning solid disc. If the angular velocity is ω radians per second then

$$\omega = \frac{\mathrm{d}\theta}{\mathrm{d}t}$$

and the velocity vector field, $\mathbf{v}(\mathbf{r})$ is,

$$\begin{aligned}\mathbf{v} = \frac{\mathrm{d}}{\mathrm{d}t}\mathbf{r} &= \frac{\mathrm{d}}{\mathrm{d}t}x\hat{\mathbf{i}} + \frac{\mathrm{d}}{\mathrm{d}t}y\hat{\mathbf{j}} \\ &= \frac{\mathrm{d}x}{\mathrm{d}\theta}\frac{\mathrm{d}\theta}{\mathrm{d}t}\hat{\mathbf{i}} + \frac{\mathrm{d}y}{\mathrm{d}\theta}\frac{\mathrm{d}\theta}{\mathrm{d}t}\hat{\mathbf{j}} \\ &= -r\sin\theta\frac{\mathrm{d}\theta}{\mathrm{d}t}\hat{\mathbf{i}} + r\cos\theta\frac{\mathrm{d}\theta}{\mathrm{d}t}\hat{\mathbf{j}} \\ &= -y\omega\hat{\mathbf{i}} + x\omega\hat{\mathbf{j}}\end{aligned}$$

so our field with $p = 0$ corresponds to $\omega = 1$.

C1.4b The vector field is,

$$\mathbf{v}(\mathbf{r}) = \frac{1}{(x^2+y^2)^p}\left(-y\hat{\mathbf{i}} + x\hat{\mathbf{j}}\right)$$

The curl is, (see "Revision of vector calculus")

$$\begin{aligned}\boldsymbol{\nabla}\times\mathbf{v} = \hat{\mathbf{i}}&\left(\frac{\partial v_z}{\partial y} - \frac{\partial v_y}{\partial z}\right) \\ +\hat{\mathbf{j}}&\left(\frac{\partial v_x}{\partial z} - \frac{\partial v_z}{\partial x}\right) \\ +\hat{\mathbf{k}}&\left(\frac{\partial v_y}{\partial x} - \frac{\partial v_x}{\partial y}\right)\end{aligned}$$

The only derivatives that are not zero are,

$$\frac{\partial v_y}{\partial x} \quad \text{and} \quad \frac{\partial v_x}{\partial y}$$

They are,

$$\begin{aligned}\frac{\partial}{\partial x}v_y = \frac{\partial}{\partial x}\,x\left(x^2+y^2\right)^{-p} &= \left(x^2+y^2\right)^{-p} - xp\left(x^2+y^2\right)^{-p-1}\cdot 2x \\ &= \frac{1}{(r^2)^p} - \frac{2px^2}{(r^2)^{p+1}} \\ \frac{\partial}{\partial y}v_x = \frac{\partial}{\partial x}\,-y\left(x^2+y^2\right)^{-p} &= -\frac{1}{(r^2)^p} + \frac{2py^2}{(r^2)^{p+1}}\end{aligned}$$

Therefore the curl of the field is,

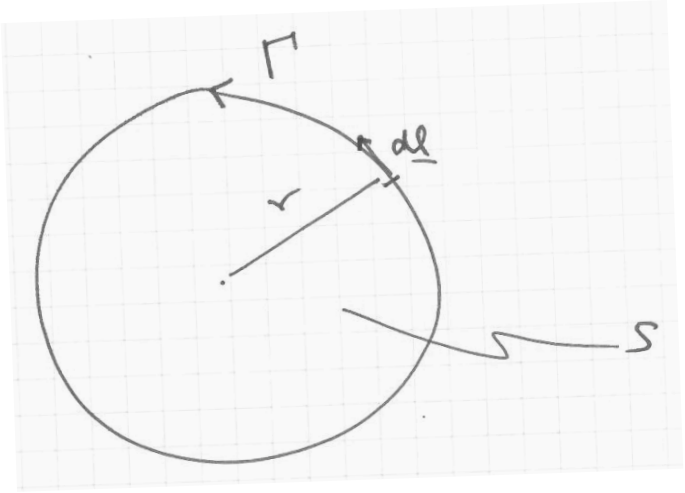
$$\begin{aligned}\boldsymbol{\nabla}\times\mathbf{v} &= \left(\frac{2}{(r^2)^p} - \frac{2px^2}{(r^2)^{p+1}} - \frac{2py^2}{(r^2)^{p+1}}\right)\hat{\mathbf{k}} \\ &= \left(\frac{2}{r^{2p}} - \frac{2pr^2}{(r^2)^{p+1}}\right)\hat{\mathbf{k}} \\ &= \left(\frac{2}{r^{2p}} - \frac{2p}{r^{2p}}\right)\hat{\mathbf{k}}\end{aligned}$$

C1.4c In the case $p = 0$ we get

$$\boldsymbol{\nabla} \times \mathbf{v} = 2\hat{\mathbf{k}}$$

which is consistent with the answer to C1.4a. For $p > 0$ there is a singularity at the origin. For $p = 1$ apart from at the origin where the curl is infinite, the curl is everywhere else zero.

C1.4d We further examine the case $p = 0$ and verify that Stokes's law applies.



Consider a circular line of radius r. The line is denoted Γ and it bounds a flat surface, S. Stokes's theorem is,

$$\oint_\Gamma \mathbf{v} \cdot \mathrm{d}\boldsymbol{\ell} = \int_S \boldsymbol{\nabla} \times \mathbf{v} \cdot \mathrm{d}\mathbf{a}$$

For the left hand side, $\mathbf{v}$ is parallel to $\mathrm{d}\boldsymbol{\ell}$ and so $\mathbf{v} \cdot \mathrm{d}\boldsymbol{\ell} = v\mathrm{d}\ell$. And the magnitude of $\mathbf{v}$ is

$$v = \sqrt{x^2 + y^2} = r$$

So we have,

$$\begin{aligned}\oint_\Gamma \mathbf{v} \cdot \mathrm{d}\boldsymbol{\ell} = v \oint_\Gamma \mathrm{d}\ell &= r \cdot 2\pi r \\ &= 2\pi r^2\end{aligned}$$

On the right hand side, the curl of $\mathbf{v}$ is $2\hat{\mathbf{k}}$ and $\hat{\mathbf{k}}$ is parallel to $\mathrm{d}\mathbf{a}$. So,

$$\begin{aligned}\int_S \boldsymbol{\nabla} \times \mathbf{v} \cdot \mathrm{d}\mathbf{a} = 2\int_S \hat{\mathbf{k}} \cdot \mathrm{d}\mathbf{a} &= 2\int_S \mathrm{d}a \\ &= 2\pi r^2\end{aligned}$$

and Stokes's theorem is verified.

C1.4e The magnetic field due to a long straight current carrying wire falls off like $1/r$ and so corresponds to the case $p = 1$. Here there is an apparent paradox: Stokes's law appears to be contradicted since while the line integral around a circular line as in C1.4d clearly is not zero the curl has vanished everywhere except at the origin. The resolution of the paradox is to properly insert a Dirac "delta function" into the curl. Alternatively, *don't expect* Stokes's theorem to apply since the domain of integration, S, is not simply connected.

2. Revision of Electrodynamics

2.1 Revision of electrostatics

You know that an assembly of point charges, $\{q\}$, or a distribution of charge, $\rho(\mathbf{r})$, will produce a force on a small test charge in accordance with Coulomb's law and the principle of superposition. The force thus measured is a vector quantity and the force per unit test charge is the vector $\mathbf{E}$ which varies with the position $\mathbf{r}_0$ of the field point such as to define the *electric field*, $\mathbf{E}(\mathbf{r}_0)$ [N C^{-1}].

The notation we use is to refer all vectors to a Cartesian coordinate system so that any vector $\mathbf{r}$ is expressed as a linear combination of the unit vectors $\hat{\mathbf{e}}_1$, $\hat{\mathbf{e}}_2$ and $\hat{\mathbf{e}}_3$,

$$\mathbf{r} = x\hat{\mathbf{e}}_1 + y\hat{\mathbf{e}}_2 + z\hat{\mathbf{e}}_3$$

We denote the source points by $\mathbf{r}_1, \mathbf{r}_2 \cdots$ if there are more than one source charge; and a distribution of charge as $\rho(\mathbf{r}_1)$. ρ is a *scalar* field and $\mathbf{E}$ is a *vector* field.

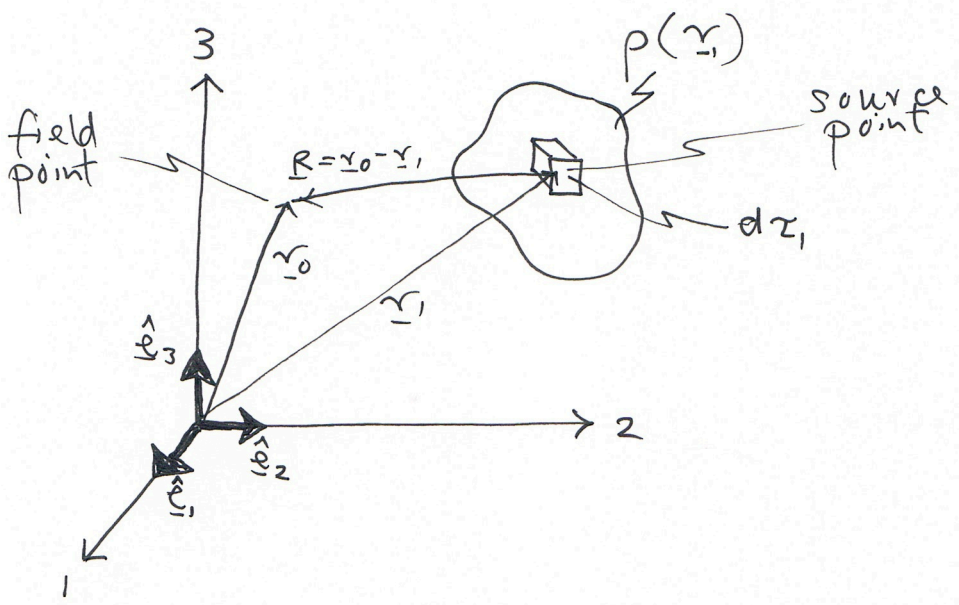


FIGURE 2–1

Figure 2–1 shows a distribution of charge, $\rho(\mathbf{r}_1)$, which exists only within the the volume Ω outlined by a closed curve (see my comments on page 11 of Section 1). In the little cube, there is an infinitesimal amount of charge, dq, so the field at the field point, according to Coulomb's law, is

$$\mathrm{d}\mathbf{E}(\mathbf{r}_0) = \frac{1}{4\pi\epsilon_0}\,\frac{\mathrm{d}q}{R^2}\,\hat{\mathbf{R}}$$

The vector $\mathbf{R} = \mathbf{r_0} - \mathbf{r_1}$ is called the *separation vector*.

The total electric field at $\mathbf{r}_0$ is the sum of these increments of field as I reposition the box at all values of $\mathbf{r}_1$ that range over the volume Ω containing the charge distribution. Summing the field over all these contributions in accord with the principle of superposition

leads to

$$\begin{aligned}\mathbf{E}(\mathbf{r}_0) &= \int \mathrm{d}x_1 \int \mathrm{d}y_1 \int \mathrm{d}z_1 \, \mathrm{d}\mathbf{E}(x_0, y_0, z_0) \\ &= \frac{1}{4\pi\epsilon_0} \int \frac{\rho(\mathbf{r}_1)}{R^2} \, \hat{\mathbf{R}} \, \mathrm{d}\tau_1\end{aligned}$$

where $\mathrm{d}\tau_1 = \mathrm{d}x_1 \mathrm{d}y_1 \mathrm{d}z_1$.

If all the charges are stationary with respect to each other and with respect to the field point the electric field has some very special properties. We call the study of such fields *electrostatics*.

1. The field obeys Gauss's law†

$$\boldsymbol{\nabla} \cdot \mathbf{E} = \frac{1}{\epsilon_0} \rho(\mathbf{r})$$

In integral form, we say that the divergence of the field through the surface of some volume, Ω, is equal to $1/\epsilon_0$ times the charge contained within that volume. The constant ϵ_0 is exactly equal to

$$\frac{10^7}{4\pi c^2}$$

where c is the speed of light in a vacuum. In maths, the integral form of Gauss's law reads,

$$\int_\Omega \boldsymbol{\nabla} \cdot \mathbf{E} \, \mathrm{d}\tau = \oint_S \mathbf{E} \cdot \mathrm{d}\mathbf{a} = \frac{1}{\epsilon_0} \int_\Omega \rho(\mathbf{r}) \, \mathrm{d}\tau$$

The surface S here is that closed surface which bounds the volume Ω. The first equality is true for any well-behaved vector field and is called the *divergence theorem*.

If S is a (not necessarily) closed surface, then $\mathrm{d}\mathbf{a} = \hat{\mathbf{n}} \mathrm{d}a$ is an infinitesimal element of that surface with $\hat{\mathbf{n}}$ being the unit vector normal to the surface element (figure 2–2).

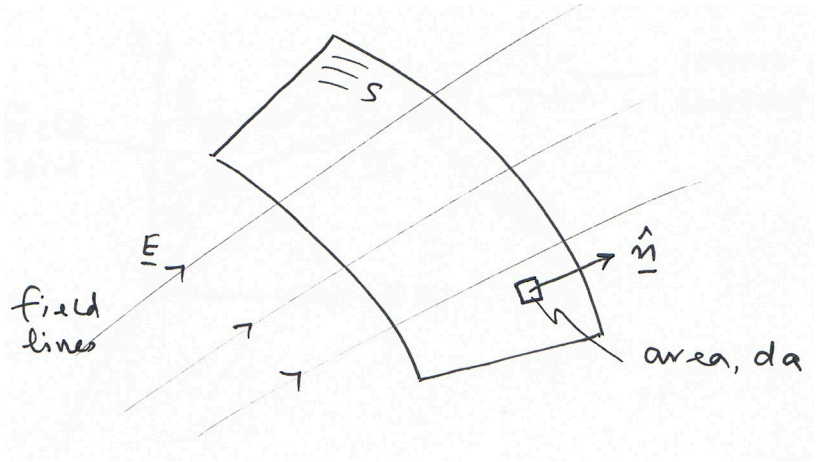


FIGURE 2–2

Of course we have

$$\int_S \mathrm{d}a = A \ , \qquad \text{the area}$$

† This is also true for moving charges.

You recall that the flux is defined as

$$\Phi_E = \int_S \mathbf{E} \cdot \mathrm{d}\mathbf{a} \qquad [\mathrm{N\ C^{-1}m^2}]$$

so in simple words, Gauss's law states that

$$\text{the flux penetrating a closed surface} = \frac{1}{\epsilon_0} \times \text{ the charge inside}$$

2. In electrostatics, the electric field is irrotational,

$$\boldsymbol{\nabla} \times \mathbf{E} = \mathbf{0}$$

In integral form

$$\oint_\Gamma \mathbf{E} \cdot \mathrm{d}\boldsymbol{\ell} = 0$$

If you integrate the field around any closed loop you get zero (figure 2–3).

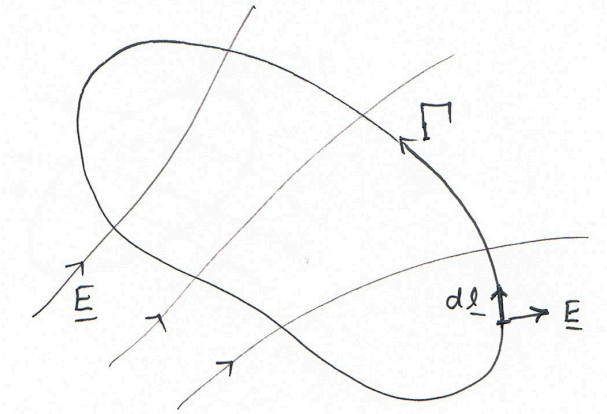


FIGURE 2–3

For any field, the integral around any closed loop is equal to the curl of the field integrated over *any* surface that is bounded by the loop (figure 2–4),

$$\oint_\Gamma \mathbf{E} \cdot \mathrm{d}\boldsymbol{\ell} = \int_S \boldsymbol{\nabla} \times \mathbf{E} \cdot \mathrm{d}\mathbf{a}$$

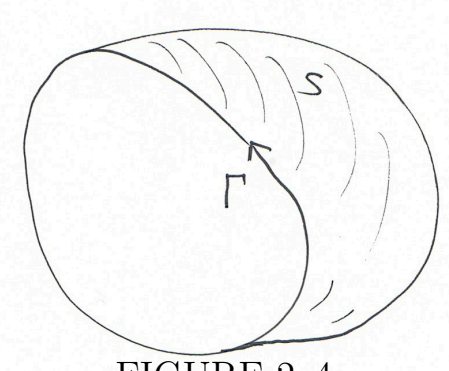


FIGURE 2–4

This is Stokes's theorem (see section 1.2.6).

As an exercise, consider the following two possible forms for an electrostatic field.

a. $\mathbf{E} = C\left(y^2\hat{\mathbf{e}}_1 + (2xy + z^2)\,\hat{\mathbf{e}}_2 + 2yz\hat{\mathbf{e}}_3\right)$

b. $\mathbf{E} = Cy\hat{\mathbf{e}}_1$

Which, if any, of these is a legitimate electric field?

3. Because

$$\oint_\Gamma \mathbf{E}\cdot \mathrm{d}\boldsymbol{\ell} = 0$$

this implies that the work done in carrying a test charge around a closed path is zero. This is not surprising since the electrostatic force is *conservative*.

Also in view of this equation, the three components of $\mathbf{E}$, (E_1, E_2 and E_3) are not independent. This fact is equivalent to the conditions imposed by $\boldsymbol{\nabla}\times\mathbf{E} = \mathbf{0}$ (see the exercise above). In fact, to describe an electric field, a scalar field called the *electric potential*, $V(\mathbf{r})$ will suffice. It is related to $\mathbf{E}$ through the equation,

$$\mathbf{E} = -\boldsymbol{\nabla}V$$

If we combine this with Gauss's law we get Poisson's equation,

$$\nabla^2 V = -\frac{1}{\epsilon_0}\,\rho$$

which relates the electric potential to the charge distribution. The geometry is the same as for the electric field, figure 2–1,

$$V(\mathbf{r}_0) = \frac{1}{4\pi\epsilon_0}\int_\Omega \frac{\rho(\mathbf{r}_1)}{R}\,\mathrm{d}\tau_1 \tag{2.1.1}$$

Example. An empty cavity of arbitrary shape is excavated inside a conducting body, again of arbitrary shape. Show that the electric field inside the cavity is zero.

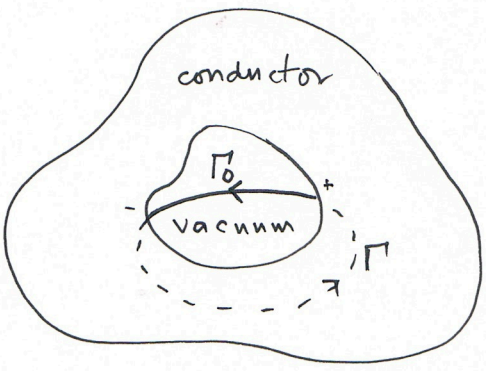


FIGURE 2–5

We consider a line, Γ_0, terminating at two points on the cavity wall. If there were an electric field in the cavity then there would be charges on the surface of the cavity.

However if I extend Γ_0 so as to close it into a loop via the conductor where $\mathbf{E} = \mathbf{0}$ then since

$$\oint_\Gamma \mathbf{E} \cdot \mathrm{d}\boldsymbol{\ell} = 0$$

then the integral along the line Γ_0 must be zero so there can be no charge on the cavity walls and no electric field inside the cavity.

This is the principle behind the *Faraday cage*.

2.2 Revision of magnetostatics

Magnetostatics is the study of forces between *steady currents*. There is a strong analogy with electrostatics: the law of Biot and Savart is the equivalent of Coulomb's law, and Ampère's law plays a role similar to Gauss's law—especially in that each can find the fields due to idealised symmetric objects, such as an infinite line of charge or an infinite line of current (be sure that you can do the *revision problem sheet* at the "Revision" section on KEATS). We talk of *current*, **I**, [C s^{-1}(or amp, A)] and *current density*, **J**, [C s^{-1} m^{-2}] which are vectors. Just as in electrostatics *forces* are the response to a *field*. A point charge sets up an electric field in accord with Coulomb's law; a steady current, I, sets up a magnetic field in accord with the law of Biot and Savart. See figure 2–6 and note the "deliberate mistake"—I've put a cross rather than a dot in the circle representing the **B**-field.

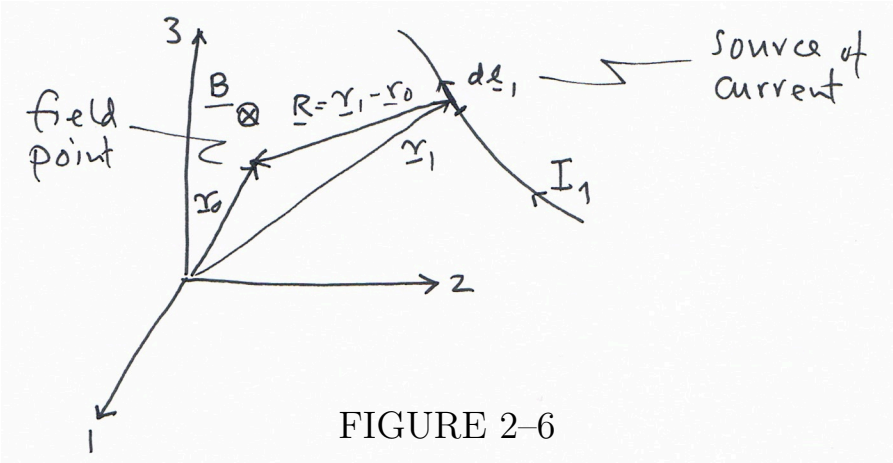


FIGURE 2–6

The Biot–Savart law reads,

$$\mathbf{B}(\mathbf{r}_0) = \frac{\mu_0 I_1}{4\pi} \int_{\Gamma_1} \frac{\mathrm{d}\boldsymbol{\ell}_1 \times \hat{\mathbf{R}}}{R^2} \tag{2.2.1}$$

Again, as in figure 2–1, we have a *source*, here of current, at source points $\mathbf{r}_1$ and the source gives rise to a *field* which we measure at a field point, $\mathbf{r}_0$. Equation (2.2.1) says that to get the total field due to the distribtion of sources, this time along the path Γ_1, we have to integrate over all the infinitesimal increments of current.

The constant $\mu_0 = 4\pi \times 10^{-7}$ [N A^{-1}] and **B** is in Tesla [T] in SI units. (1 T = 1 kg m C^{-1} s^{-1}).

The magnetic force on a second segment of current carrying wire is

$$\mathbf{F}_{\text{mag}} = I_0 \int_{\Gamma_0} \mathrm{d}\boldsymbol{\ell}_0 \times \mathbf{B}$$

So whereas in electrostatics we place a small test charge at the field point in order to detect the force, or field strength; in magnetostatics we place a second current carrying wire at the field point and measure the force and hence the magnetic field, $\mathbf{B}$, due to the source current, I_1. So, in summary, I_1 is the current in the wire that *generates* the magnetic field, $\mathbf{B}$; I_0 is the current in the wire that experiences the force due to $\mathbf{B}$.

Magnetic forces do no work. Prove this from the Lorentz force.

We know that in electrostatics, the force on a point charge, q, in an electric field, $\mathbf{E}$, is $q\mathbf{E}$. There is also a magnetic force on a point charge which is

$$\mathbf{F}_{\text{mag}} = q\mathbf{v} \times \mathbf{B}$$

where $\mathbf{v}$ is the velocity of the point charge. These combine to give the Lorentz force,

$$\mathbf{F} = q\left(\mathbf{E} + \mathbf{v} \times \mathbf{B}\right)$$

You may regard both Coulomb's law and the law of Biot and Savart as empirical laws based on laboratory evidence. Be aware that the *whole of electrostatics* can be built on the two postulates: Coulomb's law and the principle of superposition. The *whole of electrodynamics* rests on just these principles plus the experimental law of Biot and Savart. However at a more fundamental level, all of electrodynamics follows from just Coulomb's law and the principle of superposition *plus* Einstein's two postulates of special relativity which are, that the laws of physics (not just the laws of mechanics) are the same in all inertial frames and the speed of light is always the same measured from within any inertial frame (see section 3).

There are no magnetic charges (monopoles) so the equivalent of Gauss's law in magnetism is

$$\boldsymbol{\nabla} \cdot \mathbf{B} = 0$$

There *are* magnetic dipoles and these are the simplest magnetic objects. The magnetic dipole may be thought of as arising from a small current loop (figure 2–7),

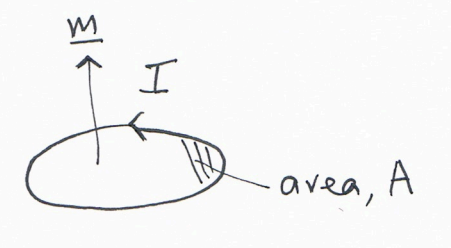


FIGURE 2–7

The magnetic dipole moment is,

$$\mathbf{m} = \hat{\mathbf{n}}\, I A$$

where $\hat{\mathbf{n}}$ is the unit normal to the loop, A is the area and I is the current.

Magnetic flux is defined as for electrostatics,

$$\Phi_B = \int_S \mathbf{B} \cdot \mathrm{d}\mathbf{a}$$

1 Tm^2 is called a Weber.

Ampère's law is a consequence of the Biot–Savart law but was first discovered by experiment. In integral form and in words: if I add all the infinitesimal components of magnetic field around a closed loop the result is equal to μ_0 times the total current that penetrates ("links") the loop. In maths,

$$\oint_\Gamma \mathbf{B} \cdot \mathrm{d}\boldsymbol{\ell} = \mu_0 I_{\text{enclosed}}$$

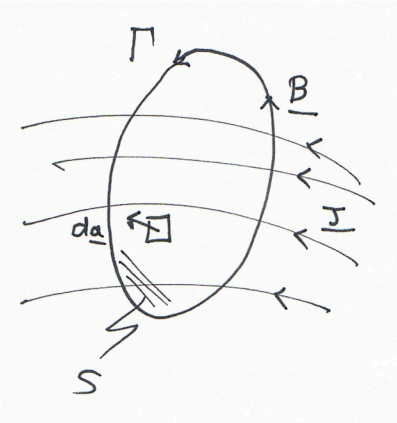


FIGURE 2–8

By applying Stokes's theorem we have

$$\oint_\Gamma \mathbf{B} \cdot \mathrm{d}\boldsymbol{\ell} = \int_S \boldsymbol{\nabla} \times \mathbf{B} \cdot \mathrm{d}\mathbf{a} = \mu_0 \int_S \mathbf{J} \cdot \mathrm{d}\mathbf{a}$$

and this is true for *any* surface that is bounded by Γ. Hence in differential form, Ampère's law is,

$$\boldsymbol{\nabla} \times \mathbf{B} = \mu_0 \mathbf{J}$$

2.3 Revision of Faraday's law

Faraday's law goes beyond electrostatics and magnetostatics. It describes how a changing magnetic field "induces" an electric field. We will start with the *flux rule*,

$$\mathcal{E} = -\frac{\partial \Phi_B}{\partial t}$$

If the magnetic flux changes through a loop† then this change sets up an electromotive force, $\mathcal{E}$, (or voltage) in the loop. The minus sign reflects *Lenz's law*: "nature abhors a change in flux". This means that if the loop can carry a current so that the e.m.f. produces current in the loop, then that current will produce a magnetic field in accord with Ampère's law that will will act to *oppose* the change in flux—that means the magnetic flux produced by the current will have opposite direction to the growing flux. You may have seen this effect as a spark in a light switch when you turn the light *off*.

Because the "induced" electromotance (e.m.f.) is,

$$\mathcal{E} = \oint \mathbf{E} \cdot \mathrm{d}\boldsymbol{\ell}$$

we have from the flux rule

$$\oint \mathbf{E} \cdot \mathrm{d}\boldsymbol{\ell} = -\int \frac{\partial \mathbf{B}}{\partial t} \cdot \mathrm{d}\mathbf{a}$$

which is Faraday's law in integral form. In differential form, as a result of Stokes's law,

$$\boldsymbol{\nabla} \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}$$

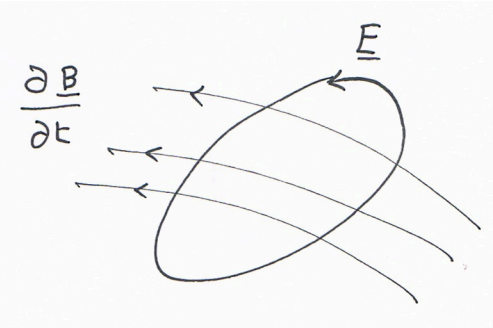


FIGURE 2–9

Of course if $\mathbf{B}$ is unchanging we recover the result in electrostatics

$$\boldsymbol{\nabla} \times \mathbf{E} = \mathbf{0}$$

In summary of section 2, we have these equations,

$$\boldsymbol{\nabla} \cdot \mathbf{E} = \frac{1}{\epsilon_0}\rho \quad ; \qquad \boldsymbol{\nabla} \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}$$
$$\boldsymbol{\nabla} \cdot \mathbf{B} = 0 \quad ; \qquad \boldsymbol{\nabla} \times \mathbf{B} = \mu_0 \mathbf{J}$$

It would be well if you can remember these; and more importantly have a picture in your head of what each one means physically.

† This "loop" may be a completely notional object, just as is an "ampèrian loop", that is, a closed line drawn in space; or it may be an actual loop of wire in which case, if the wire is conducting, the changing magnetic field will induce a current in the loop.

3. Lorentz force

3.1 Relation between electric and magnetic field (after Richard Feynman)

You may wish to follow Faraday and think of electric and magnetic effects as separate phenomena. But you may wonder how current carrying wires attract or repel each other; or why a compass needle is deflected near a current carrying wire as observed by Ampère. In fact as these great scientists suspected, electricity and magnetism are connected as explained mathematically by Heaviside and Maxwell. Indeed they are "two sides of the same coin"; or if you prefer, *there is no magnetism*—it is a relativistic effect of electricity, as Einstein discovered.

Imagine an electron that is travelling a distance r away from and parallel to a straight current carrying wire at a speed u.

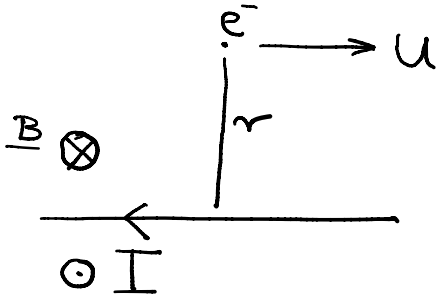


FIGURE 3–1

The current in the wire sets up a magnetic field circulating around the wire, pointing into the page above and out of the page below (figure 3–1). The magnitude of the magnetic field according to Ampère's law is,

$$B = \frac{\mu_0}{2\pi}\frac{I}{r} = \frac{1}{4\pi\epsilon_0 c^2}\frac{2I}{r}$$

The electron experiences a Lorentz force. Since there is no electric field as the wire is neutral,

$$\mathbf{F} = (-e)\mathbf{u}\times\mathbf{B}$$

and this points towards the wire, so the electron will follow a curved trajectory heading towards the wire.† Combining the two equations, we have,

$$\begin{aligned} F &= \frac{1}{4\pi\epsilon_0 c^2}\frac{2Iu(-e)}{r} \\ &= \frac{1}{4\pi\epsilon_0 c^2}\frac{2qIu}{r} \end{aligned}$$

where $q = -e$ is the charge on the electron.

† This is the origin of the fact that parallel current carrying wires attract, or repel, if the currents are in the same, or the opposite direction.

Now we look closely into the metal wire. We find that there are stationary ions which are positively charged and negatively charged conduction electrons moving in the opposite direction to I.

FIGURE 3–2

If the wire contains a charge density of electrons $\rho_{\rm e}(<0)$ [C m^{-3}] Coulomb per cubic metre and the electrons move at a speed $v_{\rm e}$† the amount of charge that flows across a cross sectional area, A, per unit time is $-\rho_{\rm e}v_{\rm e}A$ and so the current density is,

$$J = -\rho_{\rm e}v_{\rm e} \qquad [\text{C m}^{-2}\ \text{s}^{-1}]$$

and the current in the wire is

$$I = -\rho_{\rm e}v_{\rm e}A \qquad [\text{amp}]$$

To make this discussion easier I will assume the case in which the speed, u, of the electron outside is the same as the average drift velocity of the electrons in the wire. Then I have that the force on the outside electron is,

$$\begin{aligned} F &= \frac{1}{4\pi\epsilon_0 c^2}\,\frac{2uq}{r}\,(-\rho_{\rm e}uA) \\ &= \frac{1}{4\pi\epsilon_0}\,\frac{2q\rho_{\rm i}A}{r}\,\frac{u^2}{c^2} \end{aligned} \tag{3.1.1}$$

The situation is as illustrated in figure 3–3.

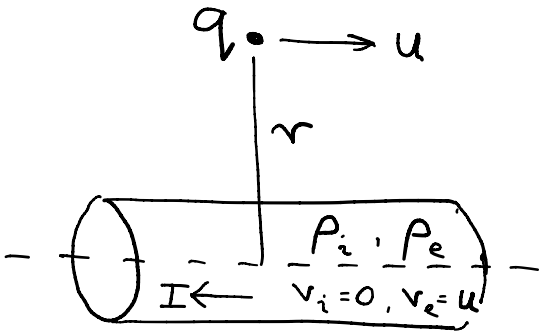


FIGURE 3–3

Here $\rho_{\rm i}$ [C m^{-3}] and $v_{\rm i}$ are the density and speed (zero) of the stationary ions. In this view we are observing the situation from within an inertial laboratory frame, S, in which the wire is stationary. The wire is also neutral so that $\rho_{\rm i} + \rho_{\rm e} = 0$.

† As a matter of interest do you know how fast that is? If you turn on a bulb at a switch in a D.C. circuit the light comes on instantly, but how long does it take a particular electron to get from the switch to the bulb? $v_{\rm e}$ is called the "drift velocity" in solid state physics. (The answer is typically about one foot per hour.)

Now view it from within an inertial frame S' in which the electron is stationary. Now the wire is moving at a speed u to the left (figure 3–4).

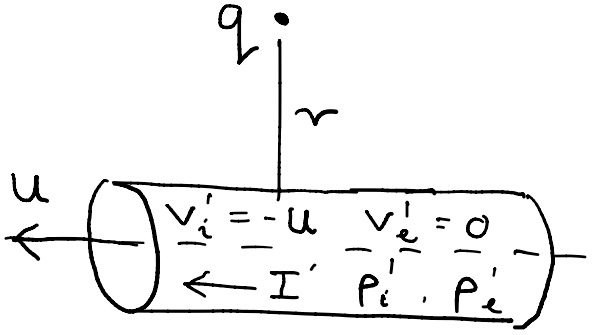


FIGURE 3–4

In this frame, the densities of the ions and electrons are ρ_{i}' and ρ_{e}' and their speeds are, $v_{\mathrm{i}}' = -u$ and $v_{\mathrm{e}}' = 0$.

By Einstein's principle, the physics must be the same; the electron must move towards the wire, but since its speed is zero it cannot experience a Lorentz force due to the magnetic field of the wire.

What is happening here is that the electron now feels an *electric field* due to the fact that in this inertial frame, *the wire is charged*. It is no longer neutral because the Lorentz contraction of the moving ions increases their density compared to the electrons in the wire which are stationary in S'.

Firstly, recall that in special relativity, certain quantities are *conserved* and certain quantities are *invariant* with respect to a change of inertial reference frame. For example, total relativistic energy and momentum are conserved, but they are not invariant; mass is invariant but it is not conserved because it can be converted into energy, $E = \gamma mc^2$. Charge on the other hand is rather special—it is both conserved and invariant. As far as we know there is no way to create or destroy charge so it is conserved. It is also invariant: it is the same whatever inertial frame the observer is in. There is ample evidence for this. For example, when I heat up a solid and its electrons move faster, it does not change its charge. The fundamental unit of charge, e, is the same in all frames.

Secondly, let's calculate ρ_{i}, the density of ions which are stationary in the wire, but seen from a frame that is moving relative to the wire. In a length L of wire of cross section A there are $\rho_{\mathrm{i}} LA$ ions, and this is the total positive charge, Q, in the length, L, of wire. If the wire is moving with speed u relative to an observer then that observer sees the length reduced to

$$L' = L\sqrt{1-\beta^2} = \frac{L}{\gamma}$$

Here, as usual,

$$\beta = \frac{u}{c}$$

and

$$\gamma = \frac{1}{\sqrt{1-\frac{u^2}{c^2}}} = \frac{1}{\sqrt{1-\beta^2}} \quad ; \qquad 1 \leq \gamma < \infty$$

and so to the observer in S' the positive density is $\rho'_{\rm i} L' A$ which also equals Q because charge is invariant. So we have $\rho'_{\rm i} L' A = \rho_{\rm i} L A$ and so the positive density as seen by the observer in S' is

$$\rho'_{\rm i} = \frac{\rho_{\rm i}}{\sqrt{1-\beta^2}} = \gamma \rho_{\rm i} \tag{3.1.2}$$

(Note that A is the same in both frames as its dimensions are in y and z which are perpendicular to the direction of travel.) This argument leads us to the following result. In frame S' in which the wire is moving and the outside electron is stationary, the density of *ions* as seen by the outside electron is given by (3.1.2); and this is *larger* than $\rho_{\rm i}$ since $\gamma > 1$.

What is the charge density of *electrons* in the wire seen by the outside electron? We may use the same argument as above, but we return to the frame S in which an observer sees the electrons in motion exactly as the observer in S' saw the ions in motion. So we have to reverse the primes in (3.1.2) and we assert that electron charge density as seen from S is

$$\rho_{\rm e} = \frac{\rho'_{\rm e}}{\sqrt{1-\beta^2}} = \gamma \rho'_{\rm e}$$

So whereas $\rho'_{\rm i}$ is greater than $\rho_{\rm i}$, it's the other way around for the electron charge density. *The outside electron sees the wire as having a positive charge* since whereas in frame S,

$$\rho = \rho_{\rm i} + \rho_{\rm e} = 0 \tag{3.1.3}$$

the wire is neutral; in frame S', the wire is charged,

$$\begin{aligned} \rho' = \rho'_{\rm i} + \rho'_{\rm e} &= \gamma \rho_{\rm i} + \frac{1}{\gamma}\rho_{\rm e} \\ &= \rho_{\rm i} \frac{\beta^2}{\sqrt{1-\beta^2}} \end{aligned} \tag{3.1.4}$$

and the second line follows from (3.1.3). The electric field due to an infinitely long charged cylinder of charge density ρ' is easily worked out using Gauss's law† and (3.1.4),

$$\begin{aligned} E' = \frac{1}{4\pi\epsilon_0}\,\frac{2\rho' A}{r} &= \frac{1}{4\pi\epsilon_0}\,\frac{2\rho_{\rm i} A \beta^2}{r\sqrt{1-\beta^2}} \\ &= \frac{1}{4\pi\epsilon_0}\,\gamma\,\frac{2\rho_{\rm i} A \beta^2}{r} \end{aligned}$$

Therefore the force on the outside electron in frame S' is

$$F' = \frac{1}{4\pi\epsilon_0}\, q\, \frac{2\rho_{\rm i} A}{r}\, \gamma\, \frac{u^2}{c^2}$$

† It had better be easy for you!!

compared to the force in frame S which we have just shown in equation (3.1.1) to be

$$F = \frac{1}{4\pi\epsilon_0}\, q\, \frac{2\rho_{\rm e} A}{r}\, \frac{u^2}{c^2}$$

Seeing that $\rho_{\rm i}$ and $-\rho_{\rm e}$ are the same (the wire is neutral in its rest frame) we have

$$F' = \gamma F$$

which is what is expected, given the Lorentz transformation of force (see the Appendix to this section). So the *physics* is the same in either frame but the *fields* look different.

In frame S the force is "magnetic" (figure 3–5)

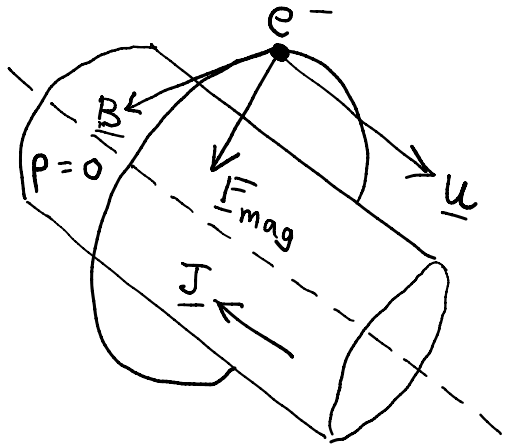


FIGURE 3–5

and in frame S' the force is "electric" (figure 3–6). There is still a magnetic field, $\mathbf{B}'$, in this frame, but the electron does not feel it because it is stationary.

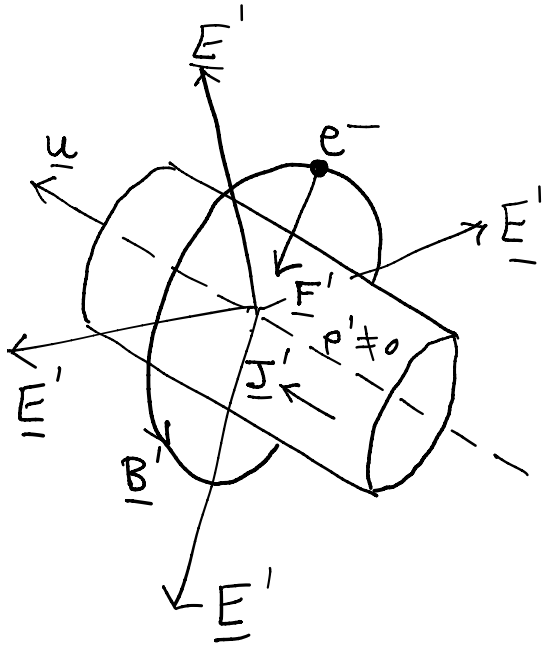


FIGURE 3–6

3.2 Lorentz force and magnetic field (after Lorrain and Corson)

3.2.1 Two charges travelling at the same velocity

Consider two point charges, q_1 and q_2 [C], which are travelling on parallel, straight trajectories as observed in a laboratory frame, S.

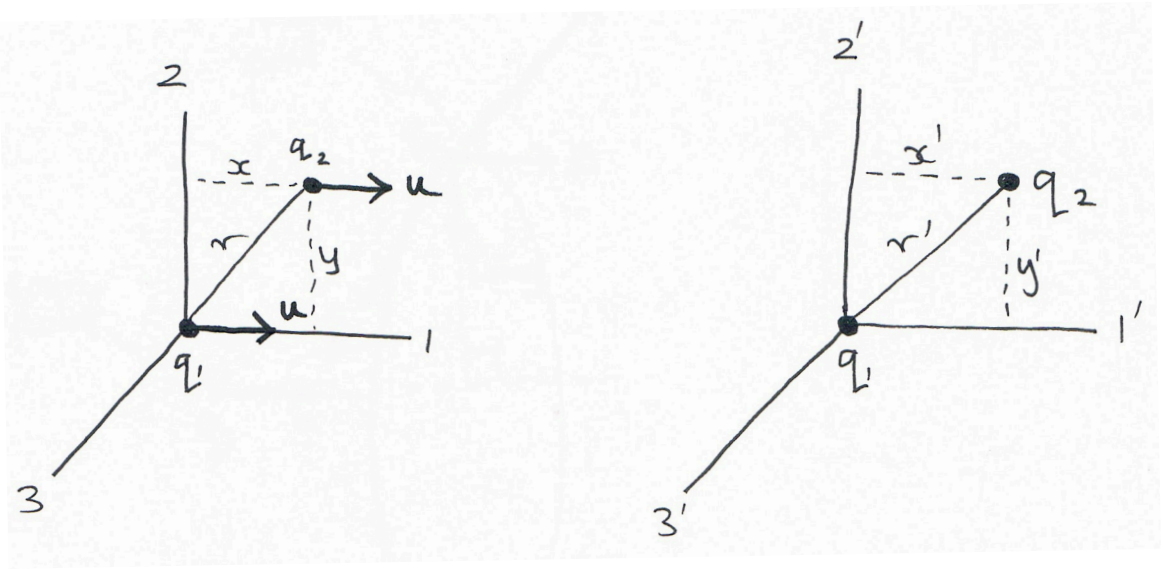


FIGURE 3–7 Two charges, q_1 and q_2 at constant velocity $\mathbf{u} = u\,\hat{\mathbf{i}}$ in the laboratory frame, S, (left) and in frame S' (right) in which the two charges are stationary.

But first, imagine that you are an observer in frame S' which is travelling with the two charges so that in S' both charges are stationary.

The force on q_2 *due to* q_1 is

$$\mathbf{F}' = F'_x\,\hat{\mathbf{i}} + F'_y\,\hat{\mathbf{j}}$$

with (see figure 3–7, right)

$$\begin{aligned} F'_x &= \frac{1}{4\pi\epsilon_0}\, q_1 q_2\, \frac{x'}{r'^3} \\ F'_y &= \frac{1}{4\pi\epsilon_0}\, q_1 q_2\, \frac{y'}{r'^3} \\ F'_z &= 0 \end{aligned} \tag{3.2.1}$$

Now examine the problem from the point of view of an observer in the laboratory frame S (see figure 3–7, left).

I now need to make a Lorentz transformation of the forces (3.2.1) to the same forces observed from S (see the Appendix to this section). The rule in the case that the

charges are stationary in S' is: forces along the direction of travel are unaltered, but forces perpendicular to the relative direction of travel of the two frames are transformed by the factor γ^{-1}, where

$$\gamma = \frac{1}{\sqrt{1-\dfrac{u^2}{c^2}}} = \frac{1}{\sqrt{1-\beta^2}} \quad ; \qquad 1 \leq \gamma < \infty$$

Hence in S,

$$F_x = F'_x \qquad \text{and} \qquad F_y = \frac{F'_y}{\gamma}$$

This gives us, using (3.2.1),

$$\begin{aligned} F_x &= \frac{1}{4\pi\epsilon_0}\, q_1 q_2\, \frac{x'}{r'^3} \\ F_y &= \frac{1}{4\pi\epsilon_0}\, \frac{1}{\gamma}\, q_1 q_2\, \frac{y'}{r'^3} \\ F_z &= 0 \end{aligned} \tag{3.2.2}$$

But these are written in terms of the distances x' and y' as measured in S'. These are Lorentz contracted in x-direction when observed from S and so we have,

$$\begin{aligned} x &= \frac{x'}{\gamma} \\ y &= y' \end{aligned}$$

So the transformed forces (3.2.2) in terms of x and y are,

$$\begin{aligned} F_x &= \frac{1}{4\pi\epsilon_0}\, q_1 q_2\, \frac{\gamma x}{\left(\gamma^2 x^2 + y^2\right)^{3/2}} \\ F_y &= \frac{1}{4\pi\epsilon_0}\, q_1 q_2\, \frac{y}{\gamma\left(\gamma^2 x^2 + y^2\right)^{3/2}} \\ &= \frac{1}{4\pi\epsilon_0}\, q_1 q_2\, \frac{\gamma y}{\left(\gamma^2 x^2 + y^2\right)^{3/2}} \left(1-\beta^2\right) \\ F_z &= 0 \end{aligned}$$

I can combine these three equations for the three components of $\mathbf{F}$ into a vector formula,

$$\mathbf{F} = \frac{1}{4\pi\epsilon_0}\, q_1 q_2\, \frac{\gamma}{\left(\gamma^2 x^2 + y^2\right)^{3/2}} \left(x\hat{\mathbf{i}} + y\hat{\mathbf{j}} - \frac{u^2}{c^2} y\hat{\mathbf{j}} \right)$$

and, recalling that $\mathbf{r} = x\hat{\mathbf{i}} + y\hat{\mathbf{j}}$, this is the same as

$$\mathbf{F} = \frac{1}{4\pi\epsilon_0}\, q_1 q_2\, \frac{\gamma}{\left(\gamma^2 x^2 + y^2\right)^{3/2}} \left(\mathbf{r} - \frac{u^2}{c^2} y\hat{\mathbf{j}} \right) \tag{3.2.3a}$$

Now I cunningly notice that since $\mathbf{u}$ is in the x-direction

$$u\hat{\mathbf{j}} = -\mathbf{u} \times \hat{\mathbf{k}}$$

because $\hat{\mathbf{i}} \times \hat{\mathbf{k}} = -\hat{\mathbf{j}}$ (the minus sign coming from this being a left handed sequence, see section 1, figure 1–3). Just to see how this next step works if you can't see it at once, let us define

$$\chi = \frac{\gamma}{\left(\gamma^2 x^2 + y^2\right)^{3/2}}$$

then

$$\mathbf{F} = \frac{1}{4\pi\epsilon_0}\, q_1 q_2\, \chi \left(\mathbf{r} - \frac{u^2}{c^2} y \hat{\mathbf{j}} \right)$$

and the second term is (omitting the $q_1 q_2 / 4\pi\epsilon_0$)

$$-\chi \frac{u^2}{c^2} y \hat{\mathbf{j}} = \chi \frac{u}{c^2} y \mathbf{u} \times \hat{\mathbf{k}} = \mathbf{u} \times \chi \frac{u}{c^2} y \hat{\mathbf{k}}$$

so that I can re-write the force (3.2.3a) as a force on q_2 *due to* q_1,

$$\begin{aligned} \mathbf{F} &= \frac{q_2}{4\pi\epsilon_0} \left\{ \frac{\gamma q_1 \mathbf{r}}{\left(\gamma^2 x^2 + y^2\right)^{3/2}} + \left(\mathbf{u} \times \frac{u\gamma q_1 y \hat{\mathbf{k}}}{c^2 \left(\gamma^2 x^2 + y^2\right)^{3/2}} \right) \right\} \\ &= \mathbf{F}_{\text{electric}} + \mathbf{F}_{\text{mag}} \end{aligned} \tag{3.2.3b}$$

Let's examine these two terms in the force.

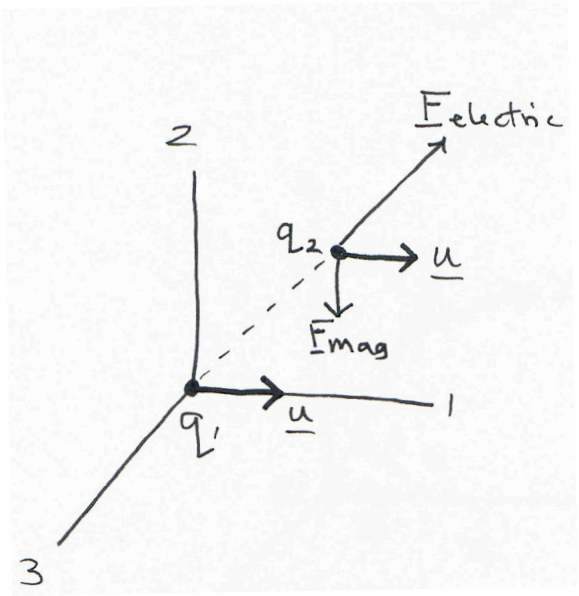


FIGURE 3–8

$\mathbf{F}_{\text{electric}}$ is the electrostatic Coulomb force as you can verify by supposing $u = 0$ and hence $\gamma = 1$. The second term represents a force acting in a direction perpendicular to $\mathbf{u}$ and which is zero if $u = 0$. From the mathematical form we see that this force acts in a direction that is perpendicular to $\mathbf{u}$ and perpendicular to a field that we might call the *magnetic field*, having the property,

$$\mathbf{B} = \frac{\mu_0}{4\pi} \frac{u\gamma q_1 y}{\left(\gamma^2 x^2 + y^2\right)^{3/2}} \hat{\mathbf{k}} \quad , \qquad \frac{\mu_0}{4\pi} = \frac{1}{4\pi\epsilon_0} \frac{1}{c^2} \tag{3.2.4}$$

Since this points in the $\hat{\mathbf{k}}$-direction and it doesn't matter how we choose the $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$-directions, this vector must circulate in a right handed sense about the trajectory of q_1 (figure 3–9).

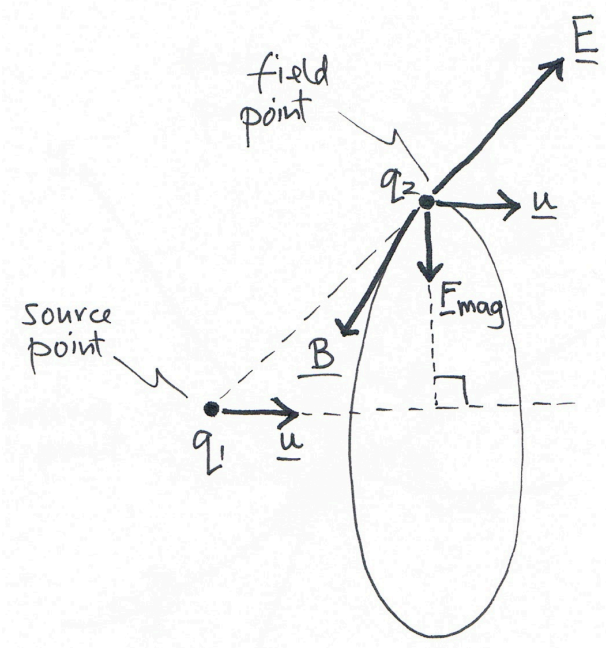


FIGURE 3–9

If we now combine (3.2.3b) with (3.2.4) and we express the electric field as the electric force per unit test charge, then we may rewrite (3.2.3b) as,

$$\mathbf{F} = q_2 \left(\mathbf{E} + \mathbf{u} \times \mathbf{B} \right)$$

3.2.2 Two charges travelling at different velocities

Now I want to make the problem a bit harder, but a bit more general. Suppose the two charges are not travelling with the same velocity. Look at figure 3–10.

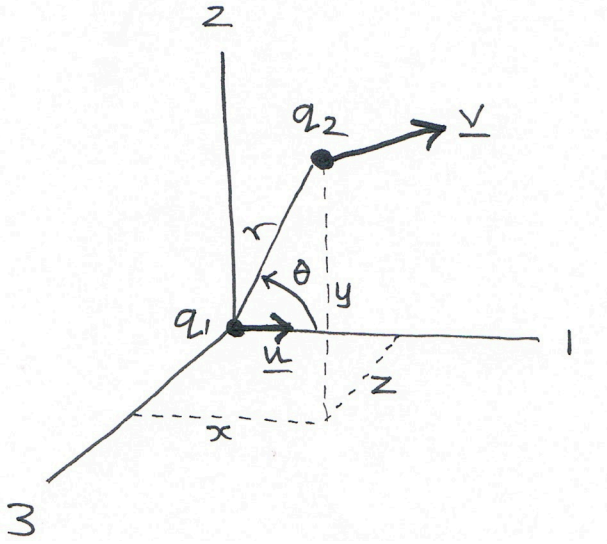


FIGURE 3–10

Now q_1 is moving in the laboratory frame, S, with a velocity $\mathbf{u}$; and q_1 serves as the source charge which sets up a magnetic field which is felt by the test charge q_2 which is travelling in the same frame at a different velocity, $\mathbf{v}$. We focus on a time, $t = 0$, at which the origin of the laboratory frame, S, coincides with the origin of the frame S' in which q_1 is stationary. In frame S' the force on q_2 due to q_1 is *electric*,

$$\begin{aligned} F'_x &= \frac{1}{4\pi\epsilon_0}\, q_1 q_2 \frac{x'}{r'^3} \\ F'_y &= \frac{1}{4\pi\epsilon_0}\, q_1 q_2 \frac{y'}{r'^3} \\ F'_z &= \frac{1}{4\pi\epsilon_0}\, q_1 q_2 \frac{z'}{r'^3} \end{aligned}$$

We transform these forces into the laboratory frame, S, (see Appendix to this section),

$$\begin{aligned} F_x &= \frac{1}{4\pi\epsilon_0}\, q_1 q_2\, \frac{1}{r'^3} \left(x' + \frac{v'_y\, u}{c^2 + v'_x\, u}\, y' + \frac{v'_z\, u}{c^2 + v'_x\, u}\, z' \right) \\ &= \frac{1}{4\pi\epsilon_0}\, q_1 q_2\, \frac{1}{r'^3} \left(\gamma\,(x - ut) + \gamma \frac{v_y\, u}{c^2}\, y + \gamma \frac{v_z\, u}{c^2}\, z \right) \\ F_y &= \frac{1}{4\pi\epsilon_0}\, q_1 q_2\, \frac{1}{r'^3} \frac{1}{\gamma} \frac{1}{1 + \frac{v'_x\, u}{c^2}}\, y' \\ &= \frac{1}{4\pi\epsilon_0}\, q_1 q_2\, \frac{1}{r'^3} \gamma^2 \left(1 - \frac{v_x\, u}{c^2} \right) y \\ F_z &= \frac{1}{4\pi\epsilon_0}\, q_1 q_2\, \frac{1}{r'^3} \frac{1}{\gamma} \frac{1}{1 + \frac{v'_x\, u}{c^2}}\, z' \\ &= \frac{1}{4\pi\epsilon_0}\, q_1 q_2\, \frac{1}{r'^3} \gamma^2 \left(1 - \frac{v_x\, u}{c^2} \right) z \end{aligned}$$

For each component of the force, in the first line is the transformation of the force, and in the second the transformation of the coordinate. In reference to figure 3–10 you see that,

$$\mathbf{r} = (x - ut)\, \hat{\mathbf{i}} + y\, \hat{\mathbf{j}} + z\, \hat{\mathbf{k}}$$

and rather as we saw in (3.2.3a) we can gather up terms in the x, y and z component of the force and after a little algebra we recognise a vector product,†

$$\begin{aligned} \mathbf{F} &= \frac{1}{4\pi\epsilon_0}\, \gamma q_1 q_2\, \frac{1}{r'^3} \left(\mathbf{r} + \frac{u}{c^2} \left((v_y\, y + v_z\, z)\, \hat{\mathbf{i}} - v_x\, y\, \hat{\mathbf{j}} - v_x\, z\, \hat{\mathbf{k}} \right) \right) \\ &= \frac{1}{4\pi\epsilon_0}\, \gamma q_1 q_2\, \frac{1}{r'^3} \left(\mathbf{r} + \frac{1}{c^2}\, \mathbf{v} \times (\mathbf{u} \times \mathbf{r}) \right) \end{aligned}$$

We've not yet transformed the factor $1/r'^3$. To this end, we write,

$$\begin{aligned} r'^2 &= \gamma^2\, (x - ut)^2 + y^2 + z^2 \\ &= \gamma^2 \left((x - ut)^2 + \left(y^2 + z^2 \right) \left(1 - \beta^2 \right) \right) \\ &= \gamma^2 r^2 \left(1 - \beta^2 \sin^2 \theta \right) \end{aligned}$$

† Expand out the second line, noting $\mathbf{u} = u_x \hat{\mathbf{i}}$, $u_y = u_z = 0$.

where θ is the angle between $\mathbf{r}$ and the x-axis at $t = 0$ (figure 3–10). Finally,

$$\begin{aligned}\mathbf{F} &= \frac{1}{4\pi\epsilon_0}\, q_1 q_2 \frac{1}{\gamma^2 r^2 \left(1-\beta^2\sin^2\theta\right)^{\frac{3}{2}}}\left(\hat{\mathbf{r}} + \frac{1}{c^2}\mathbf{v}\times\mathbf{u}\times\hat{\mathbf{r}}\right)\\ &= q_2\left(\mathbf{E} + \mathbf{v}\times\mathbf{B}\right)\end{aligned} \tag{3.2.5}$$

with

$$\mathbf{E} = \frac{1}{4\pi\epsilon_0}\, q_1 \frac{1}{\gamma^2 r^2 \left(1-\beta^2\sin^2\theta\right)^{\frac{3}{2}}}\,\hat{\mathbf{r}} \tag{3.2.6a}$$

$$\mathbf{B} = \frac{\mu_0}{4\pi}\, q_1 \frac{u\sin\theta}{\gamma^2 r^2 \left(1-\beta^2\sin^2\theta\right)^{\frac{3}{2}}}\,\hat{\boldsymbol{\phi}} \tag{3.2.6b}$$

where ϕ is the azimuthal angle about $\mathbf{u}$. In cartesian coordinates,

$$\mathbf{E} = \frac{1}{4\pi\epsilon_0}\,\gamma q_1 \frac{(x-ut)\hat{\mathbf{i}} + y\,\hat{\mathbf{j}} + z\,\hat{\mathbf{k}}}{\left(\gamma^2(x-ut)^2 + y^2 + z^2\right)^{\frac{3}{2}}} \tag{3.2.7a}$$

$$\mathbf{B} = \frac{\mu_0}{4\pi}\,\gamma q_1 \frac{u(-z\,\hat{\mathbf{j}} + y\,\hat{\mathbf{k}})}{\left(\gamma^2(x-ut)^2 + y^2 + z^2\right)^{\frac{3}{2}}} \tag{3.2.7b}$$

and you can see that, as you should expect, these reduce to (3.2.3b) in the case that $u = v$, $t = 0$ and $z = 0$.

The source charge and the fields that it generates (3.2.6) are illustrated in figure 3–11.

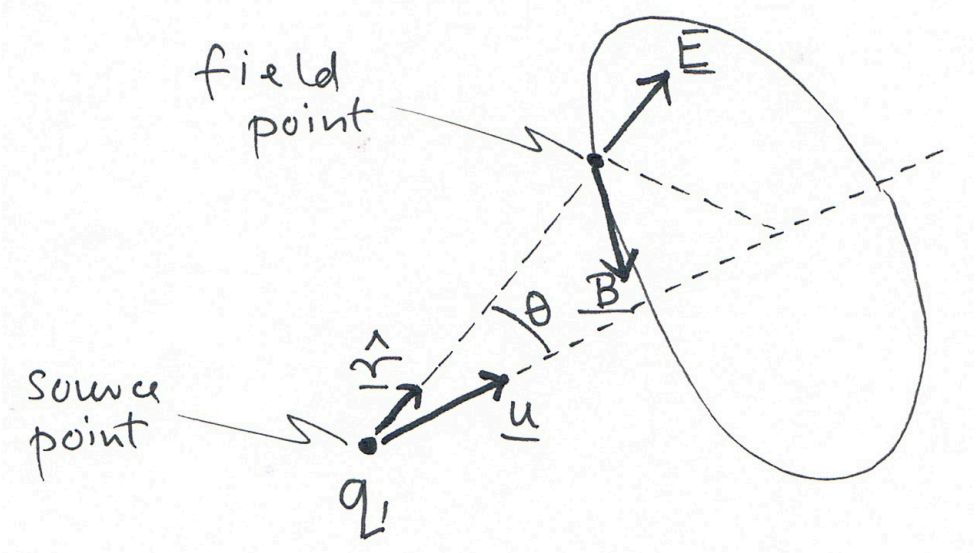


FIGURE 3–11, note that $\mathbf{B}\|\mathbf{u}\times\hat{\mathbf{r}}$, so it is perpendicular to both $\mathbf{u}$ and $\hat{\mathbf{r}}$ in a right handed sense. The cross product brings in the factor $\sin\theta$ in (3.2.6b).

Just write down (3.2.5) again,

$$\mathbf{F} = q\,(\mathbf{E} + \mathbf{v} \times \mathbf{B})$$

where I've dropped the subscript "2". A moving source charge has generated a magnetic field and an electric field and (3.2.5) says that a test charge, q, moving with a velocity $\mathbf{v}$ will experience the force $\mathbf{F}$.

We have derived the Lorentz force from just Coulomb's law and Einstein's principle in special relativity. Are you shocked? Impressed? Bored and indifferent? Magnetism is truly just a relativistic consequence of electricity.

Conventionally, we write,

$$\frac{1}{\epsilon_0 c^2} = \mu_0 = 4\pi \times 10^{-7} \qquad \text{henry m}^{-1}$$

In essence,

$$4\pi\epsilon_0 c^2 = 10^7 \qquad \text{in SI units}$$

defines the Coulomb as unit of charge, or if you prefer, the Amp as unit of current, since all other variables are designated units in the SI system from mechanics.

3.2.3 Summary so far...

Let us briefly recap section *3.2.1*. We have two charges, q_1 and q_2 that are travelling at equal speed, u, along parallel straight line trajectories in the x-direction. This is the laboratory frame S. We imagine a second inertial frame, S', which is moving at the same speed as the two charges in the x-direction relative to frame S. This is the classic two frames for studying special relativity. In frame S' (figure 3–7, right) the problem is trivial: there is a repulsive force on q_2 due to q_1 which is given by Coulomb's law (3.2.1),

$$\begin{aligned}\mathbf{F}' = q_2\mathbf{E}' &= \frac{1}{4\pi\epsilon_0}\; q_1\,\frac{1}{r'^3}\;(x'\hat{\mathbf{i}} + y'\hat{\mathbf{j}}\,)\\ &= \frac{1}{4\pi\epsilon_0}\; q_1\,\frac{1}{r'^3}\,\mathbf{r}'\end{aligned}$$

We are treating q_1 as the *source charge*, so it has set up an electric field, $\mathbf{E}$, which is detected by the *test charge* q_2. In the laboratory frame (figure 3–7, left), the force is (3.2.2),

$$\mathbf{F} = q_2\mathbf{E} = \frac{1}{4\pi\epsilon_0}\; q_1\,\frac{1}{r'^3}\;\left(x'\hat{\mathbf{i}} + \gamma^{-1}y'\hat{\mathbf{j}}\,\right)$$

Clearly, $\mathbf{E}$ and $\mathbf{E}'$ are not the same. As I change my frame of reference I get a different electric field (see the last table in the Appendix to this section). Mind you, we have expressed the electric field in S in terms of the coordinates as measured in S', so we make the appropriate Lorentz transformations of x' and y' and we get (3.2.3b),

$$\mathbf{F} = q_2\,(\mathbf{E} + \mathbf{u} \times \mathbf{B})$$

in which

$$\mathbf{E} = \frac{1}{4\pi\epsilon_0}\ q_1\ \frac{\gamma}{(\gamma^2 x^2 + y^2)^{3/2}}\,\mathbf{r}$$

and (3.2.4)

$$\mathbf{B} = \frac{1}{4\pi\epsilon_0}\ q_1\ \frac{\gamma u y}{(\gamma^2 x^2 + y^2)^{3/2}}\,\hat{\mathbf{k}}$$

An apparently *completely new* field has emerged which we call **B**, the magnetic field. But the message is that we can express the force acting between q_2 and q_1 as wholly electric, or electric and magnetic, depending on the choice of inertial frame that we choose from which to observe the force. **E** and **B** are intricately connected.

I have placed a box about (3.2.7)—the electric and magnetic fields due to a charge q_1 moving at a speed u along the x-axis. They are detected by a test charge q_2, moving at a velocity **v**, via the Lorentz force experience by that test charge (3.2.5). Equations (3.2.7) are *fundamental*—the starting point for demonstrating the law of Biot and Savart (Section 3.3) and hence Ampère's law (Section 3.4), Faraday's law (Section 3.5), and ultimately to showing that Maxwell's equations are *invariant under Lorentz transformations*.

Here are a few questions and problems to test your understanding.

1. Do the charges in section 3.2 attract or repel? Or are there frames from which they are observed to attract and frames from which they are observed to repel? What is the essential difference between this example and the example in section 3.1 in which the electron is *attracted* to the wire?
2. In the example in section 3.1, the drift velocity is typically twelve orders of magnitude smaller than the speed of light. $\gamma = 1$ is correct to 25 decimal places. Why are we talking of relativistic effects here?
3. Can you rework the problem in section 3.1 in which the outside electron is travelling at a different speed to the electrons in the wire in the laboratory frame?† (We rather artificially set them to be equal for simplicity in the algebra.)
4. In the example in section 3.2, is there an inertial frame from which the measured force is purely magnetic, that is, in which the electric field is zero? Just as the magnetic field is zero in frame S'.
5. This whole treatment rests solely on an inverse square force law (Coulomb's law) and the principles of special relativity. So why is there no "magnetism" in gravitation? Or is there…?

† The key to answering this question is that the components of **J**, and ρ, make up a relativistic four-vector, $J^\mu = (c\rho, J_x, J_y, J_z)$ in the same way that the components of momentum and energy do. You need to know how J^μ transforms under a Lorentz transformation. The fact is that it transforms exactly as does the four-vector $x^\mu = (ct, x, y, z)$.

3.3 Law of Biot and Savart

I can generalise the illustration in figure 3–11, so that the source charge becomes representative of an infinitesimal element of the current flowing along a wire. This is shown in figure 3–12.

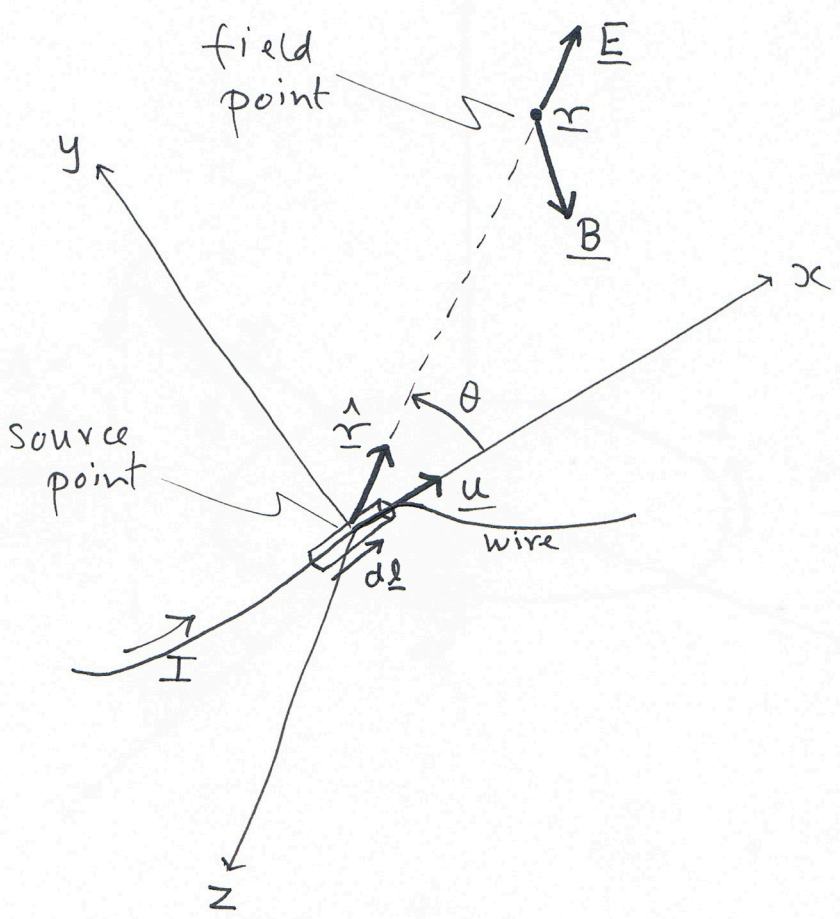


FIGURE 3–12

The element of wire of length $\mathrm{d}\ell$ has ion and electron charge densities per unit length, λ_i and λ_e. The electrons are moving along the x-direction at a speed $-u$. The total charge in the laboratory frame is zero,

$$\lambda_\mathrm{i}\mathrm{d}\ell + \lambda_\mathrm{e}\mathrm{d}\ell = 0$$

At the field point $\mathbf{r} = x\,\hat{\mathbf{i}} + y\hat{\mathbf{j}} + z\,\hat{\mathbf{k}}$ the electric field due to the positive charges is,

$$\mathbf{E}_\mathrm{i} = \frac{1}{4\pi\epsilon_0}\,\frac{\lambda_\mathrm{i}\,\mathrm{d}\ell}{r^2}\,\hat{\mathbf{r}}$$

and they produce no magnetic field. Treat the electronic charge contained within the element of wire as a single source charge, $q_1 = \lambda_\mathrm{e}\,\mathrm{d}\ell$. Then from (3.2.6),

$$\mathbf{E}_\mathrm{e} = \frac{1}{4\pi\epsilon_0}\,\frac{(1-\beta^2)\lambda_\mathrm{e}\,\mathrm{d}\ell}{r^2\left(1-\beta^2\sin^2\theta\right)^{\frac{3}{2}}}\,\hat{\mathbf{r}}$$

$$\mathbf{B}_\mathrm{e} = -\frac{\mu_0}{4\pi}\,\frac{(1-\beta^2)u\,\lambda_\mathrm{e}\,\mathrm{d}\ell\,\sin\theta}{r^2\left(1-\beta^2\sin^2\theta\right)^{\frac{3}{2}}}\,\hat{\boldsymbol{\phi}}$$

The total fields are,

$$\mathbf{E} = \frac{1}{4\pi\epsilon_0}\,\frac{\lambda_\mathrm{i}\,\mathrm{d}\ell}{r^2}\left(1 - \frac{1-\beta^2}{\left(1-\beta^2\sin^2\theta\right)^{\frac{3}{2}}}\right)\,\hat{\mathbf{r}}$$
$$\mathbf{B} = \frac{\mu_0}{4\pi}\,\frac{\lambda_\mathrm{i}\,\mathrm{d}\ell}{r^2}\,\frac{(1-\beta^2)u\,\sin\theta}{\left(1-\beta^2\sin^2\theta\right)^{\frac{3}{2}}}\,\hat{\boldsymbol{\phi}}$$

Because of the truly tiny drift velocity, we will have $\beta^2 \approx 10^{-25}$ so we may happily set $\beta^2 = 0$. Then we have, using $I = -u\,\lambda_\mathrm{e} = u\,\lambda_\mathrm{i}$,

$$\begin{aligned}\mathbf{B} &= \frac{\mu_0}{4\pi}\,\frac{I\,\mathrm{d}\ell}{r^2}\,\sin\theta\,\hat{\boldsymbol{\phi}}\\ &= \frac{\mu_0}{4\pi}\,I\,\frac{\mathrm{d}\boldsymbol{\ell}\times\hat{\mathbf{r}}}{r^2}\end{aligned}$$

If I integrate this over an entire loop of wire, I get,

$$\mathbf{B} = \frac{\mu_0}{4\pi}\;I\;\oint\frac{\mathrm{d}\boldsymbol{\ell}\times\hat{\mathbf{r}}}{r^2}$$

which is the *Law of Biot and Savart*. Because we have neglected β^2 we see that the Biot–Savart law applies to *slow* currents in wires. When dealing with fast charges the relativistic form must be used.

3.4 Ampère's law

At the risk of becoming boring, I will repeat the observation that I hope you find revealing, if not revolutionary; namely, that initially we have learned that *electrostatics* is based entirely on two principles: Coulomb's law and the principle of superposition. Now by adding Einstein's two postulates of special relativity we find that we can obtain the law of Biot and Savart which is the foundation of magnetostatics. Again, at the risk of repeating myself, *magnetism is a relativistic consequence of electricity*. To complete this section, I will briefly show that you can also obtain Ampère's law from the Biot–Savart law.

Start with the law of Biot and Savart in terms of current density,

$$\begin{aligned}\mathbf{B}(\mathbf{r}_0) &= \frac{\mu_0}{4\pi}\int\frac{\mathbf{J}(\mathbf{r}_1)\times\mathbf{r}}{r^2}\mathrm{d}\tau_1\\ &= -\frac{\mu_0}{4\pi}\int\mathbf{J}(\mathbf{r}_1)\times\boldsymbol{\nabla}\frac{1}{r}\mathrm{d}\tau_1\end{aligned}$$

See section 1 for the notation; $\mathbf{r} = \mathbf{r}_0 - \mathbf{r}_1$ is the connecting vector and I use $\boldsymbol{\nabla} = \boldsymbol{\nabla}_0$. We take the curl of $\mathbf{B}$, with respect to the field point coordinates,

$$\boldsymbol{\nabla}\times\mathbf{B}(\mathbf{r}_0) = -\frac{\mu_0}{4\pi}\int\boldsymbol{\nabla}\times\mathbf{J}(\mathbf{r}_1)\times\boldsymbol{\nabla}\frac{1}{r}\mathrm{d}\tau_1$$

By the product rule, section 1.2.3,

$$\boldsymbol{\nabla}\times\mathbf{J}\times\boldsymbol{\nabla}\frac{1}{r}=\left(\boldsymbol{\nabla}\frac{1}{r}\cdot\boldsymbol{\nabla}\right)\mathbf{J}-(\mathbf{J}\cdot\boldsymbol{\nabla})\,\boldsymbol{\nabla}\frac{1}{r}+\mathbf{J}\left(\boldsymbol{\nabla}\cdot\boldsymbol{\nabla}\frac{1}{r}\right)-\boldsymbol{\nabla}\frac{1}{r}\,(\boldsymbol{\nabla}\cdot\mathbf{J})$$

The first and last terms on the right hand side are zero because derivatives of $\mathbf{J}(\mathbf{r}_1)$ with respect to $\mathbf{r}_0$ vanish, and I am left with

$$\begin{aligned}\boldsymbol{\nabla}\times\mathbf{B}&=-\frac{\mu_0}{4\pi}\int\mathbf{J}\,\nabla^2\frac{1}{r}\mathrm{d}\tau_1+\frac{\mu_0}{4\pi}\int(\mathbf{J}\cdot\boldsymbol{\nabla})\,\boldsymbol{\nabla}\frac{1}{r}\mathrm{d}\tau_1\\&=\mu_0\int\mathbf{J}\,\delta(\mathbf{r})\mathrm{d}\tau_1\\&=\mu_0\mathbf{J}\end{aligned}$$

which is Ampère's law. The second integral in the first line is zero—see section 1.3, page 13.

3.5 Faraday's Law

The algebra is not very revealing, but you are welcome to try it; it's not hard. Equations (3.2.7) are consistent with the statement, "the curl of **E** is equal to minus the time derivative of **B**." Hence, again from only Coulomb's law and Einstein's principles, we can deduce not only Ampère's law, but also Faraday's law.

3.6 Gauss's Law

Again, I will leave the maths to you, but you may have asked yourself, is Gauss's law true in the case of a *moving* charge? So if I place a spherical Gaussian surface, stationary in my laboratory frame, so that it encloses a moving charge; is the flux of electric field still equal to $1/\epsilon_0$ times the charge? And the answer is yes. You can derive the result by using the formula (3.2.7a) for the electric field and performing an integration over the sphere. You can also show, using (3.2.7b) that $\boldsymbol{\nabla}\cdot\mathbf{B}=0$ is true also in this case. If I combine this with the result of Section 3.5 I have shown that three of the Maxwell equations are invariant under a Lorentz transformation. Full details are in Lorrain and Corson, where you see the fourth Maxwell equation proved to be Lorentz invariant. This, you may agree, is a staggering result: Maxwell's equations were discovered more than 40 years before Einstein's paper in 1905, "Zur Elektrodynamik bewegter Körper" ("On the Electrodynamics of Moving Bodies")—yes his first special relativity paper was not about mechanics, it was about electrodynamics. In fact it was the three experiments of Faraday (the moving coil, the moving magnet and the changing magnetic field) all pointing to the flux rule that made him ponder the problem. Maxwell's equations also survived the quantum mechanical revolution without modification, which is a great deal more than can be said about Newton's laws, which survived neither.

Appendix—transformation tables (after Lorrain and Corson)

For future reference and in case you want to take through the transformations in detail in the text, here are standard formulas for transformations from frame S to frame S' which is moving with respect to S in the x-direction with velocity $\mathbf{u} = u\,\hat{\mathbf{i}}$. Tables are taken from Lorrain and Corson. First, is the famous Lorentz transformation.

$t = \gamma\left(t' + \frac{u}{c^2}x'\right)$	$t' = \gamma\left(t - \frac{u}{c^2}x\right)$
$x = \gamma\left(x' + ut'\right)$	$x' = \gamma\left(x - ut\right)$
$y = y'$	$y' = y$
$z = z'$	$z' = z$

The transformation of a velocity follows.

$v_x = \dfrac{v'_x + u}{1 + \frac{v'_x u}{c^2}}$	$v'_x = \dfrac{v_x - u}{1 - \frac{v_x u}{c^2}}$
$v_y = \dfrac{v'_y}{\gamma\left(1 + \frac{v'_x u}{c^2}\right)}$	$v'_y = \dfrac{v_y}{\gamma\left(1 - \frac{v_x u}{c^2}\right)}$
$v_z = \dfrac{v'_z}{\gamma\left(1 + \frac{v'_x u}{c^2}\right)}$	$v'_z = \dfrac{v_z}{\gamma\left(1 - \frac{v_x u}{c^2}\right)}$

Here is the transformation of force that you need for section 3.2.

$F_x = F'_x + \dfrac{u}{c^2 + v'_x u}\left(v'_y F'_y + v'_z F'_z\right)$	$F'_x = F_x - \dfrac{u}{c^2 - v_x u}\left(v_y F_y + v_z F_z\right)$
$F_y = \dfrac{F'_y}{\gamma\left(1 + \frac{v'_x u}{c^2}\right)}$	$F'_y = \dfrac{F_y}{\gamma\left(1 - \frac{v_x u}{c^2}\right)}$
$F_z = \dfrac{F'_z}{\gamma\left(1 + \frac{v'_x u}{c^2}\right)}$	$F'_z = \dfrac{F'_z}{\gamma\left(1 - \frac{v_x u}{c^2}\right)}$

In part answer to question 3.2 on page 13, here are the how the fields transform under a Lorentz transformation.

$E_x = E'_x$	$E'_x = E_x$
$E_y = \gamma\left(E'_y + uB'_z\right)$	$E'_y = \gamma\left(E_y - uB_z\right)$
$E_z = \gamma\left(E'_z - uB'_y\right)$	$E'_z = \gamma\left(E_z + uB_y\right)$
$B_x = B'_x$	$B'_x = B_x$
$B_y = \gamma\left(B'_y - \frac{u}{c^2}E'_z\right)$	$B'_y = \gamma\left(B_y + \frac{u}{c^2}E_z\right)$
$B_z = \gamma\left(B'_z + \frac{u}{c^2}E'_y\right)$	$B'_z = \gamma\left(B_z - \frac{u}{c^2}E_y\right)$

There follows an *outline*, but not a detailed proof, of how these tables are obtained. For a full development, you must consult your lecture notes on relativity and textbooks such as Griffiths, and Lorrain and Corson.

Lorentz transformation and 4-vectors

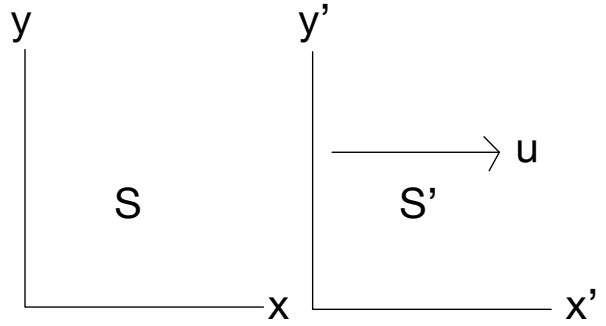


FIGURE 3–13

As is usual (figure 3–13), we imagine two inertial frames, S and S' which are travelling at a speed u relative to each other along a common x-axis. The standard notation is

$$\beta = \frac{u}{c}$$

and

$$\gamma = \frac{1}{\sqrt{1-\frac{u^2}{c^2}}} = \frac{1}{\sqrt{1-\beta^2}} \quad ; \qquad 1 \leq \gamma < \infty$$

Quantities as measured by an observer in S' will be denoted using the same symbols as in S but with a prime " $'$ " attached. Thus x is the position of an object along the x-axis and x' is its position as observed by an observer in S' along the x'-axis. The time indicated on a clock in S is denoted t and the time shown on a clock in S' is t'. It is usual to set $t = t' = 0$ when the origins of S and S' coincide, that is when $x = x' = 0$.

As you probably know, if you combine the three components of position, x, y and z, which we call x^1, x^2 and x^3 with $x^0 = ct$ we can make up a 4-vector

$$x^\mu = \left(x^0, x^1, x^2, x^3\right)$$

then the transformation of any 4-vector takes the same structure as the transformation of x^μ, namely,

$$\begin{aligned} x'^0 &= \gamma\left(x^0 - \beta x^1\right) \\ x'^1 &= \gamma\left(x^1 - \beta x^0\right) \\ x'^2 &= x^2 \\ x'^3 &= x^3 \end{aligned}$$

which you can easily verify is identical to the Lorentz transformation in the first table on page 17. Every 4-vector with index in the upper position, for example, x^μ is so called contravariant; it has a covariant counterpart, with index in the lower position and differing only by a minus sign at the zeroth component in the $(-+++)$ metric signature,†

$$x_\mu = \left(-x^0, x^1, x^2, x^3\right)$$

Transformation of velocity

The velocity of an object as observed from S is $\mathbf{v} = v_x\hat{\mathbf{i}} + v_y\hat{\mathbf{j}} + v_z\hat{\mathbf{k}}$. If the object is observed from S' then its velocity is $\mathbf{v}' = v'_x\hat{\mathbf{i}}\,' + v'_y\hat{\mathbf{j}}\,' + v'_z\hat{\mathbf{k}}\,'$.

An observer in S' measures the x'-component of the velocity of an object to be,

$$v'_x = \frac{\mathrm{d}x'}{\mathrm{d}t'}$$

What velocity along the x-axis does an observer in S measure? The x-component of the velocity of the object observed from S is,

$$v_x = \frac{\mathrm{d}x}{\mathrm{d}t} = \frac{\mathrm{d}x}{\mathrm{d}t'}\frac{\mathrm{d}t'}{\mathrm{d}t} \tag{1}$$

using the chain rule. We write,

$$\frac{\mathrm{d}x}{\mathrm{d}t'} = \gamma\frac{\mathrm{d}}{\mathrm{d}t'}\left(x' + ut'\right) = \gamma\left(v'_x + u\right) \tag{2}$$

where I have used the Lorentz transformation of the distance x. That disposes of the first term in (1). The second term is,

$$\frac{\mathrm{d}t'}{\mathrm{d}t} = \gamma\frac{\mathrm{d}}{\mathrm{d}t}\left(t - \frac{u}{c^2}x\right) = \gamma\left(1 - \frac{u}{c^2}v_x\right) \tag{3}$$

using the Lorentz transformation of the time. Combining the two terms, we have,

$$v_x = \gamma^2\left(v'_x + u\right)\left(1 - \frac{u}{c^2}v_x\right)$$

† Many people use the $(+---)$ metric signature in which case the sign of the three *spatial* components change when the index is "lowered". I am using the same convention as Griffiths, but you will find that some authors place the non-spatial component of the 4-vector at the end and label it number 4 rather than 0. Some 4-vectors may also be found multiplied by some power of c; for example, the 4-velocity may be our η^μ/c to make it dimensionless. Following Sommerfeld, Lorrain and Corson place a $\sqrt{-1}$ to multiply the temporal component so that the minus sign turns up in the dot product without having to bother with upper and lower index notation. I don't know what you will be using in other courses, so be aware there is no single, "right" way to construct a 4-vector.

If I make v_x the subject of this equation, I get,

$$v_x = \frac{v'_x + u}{1 + \frac{v'_x u}{c^2}} \tag{4}$$

which is the first entry in the transformation table on page 17 for velocity. I get the inverse transformation simply by reversing the direction of relative travel of the two frames, namely by changing the sign of u,

$$v'_x = \frac{v_x - u}{1 - \frac{v'_x u}{c^2}}$$

You can compare (4) with the Galilean transformation of speed. If I'm on a train and I throw a tennis ball in the forward travelling direction of the train with a speed v'_x, then an observer by the track measures the speed of the tennis ball to be $v'_x + u$ if u is the speed of the train. Equation (4) is the relativistic correction of this. For if instead of throwing a ball I shine a torch then the photons travel forward at the speed $v'_x = c$. The observer on the track does not see them travelling a speed $c + u$, but also c as you see from (4). *The speed of light is the same in all frames.*

Now we do the y-component of the velocity of the object observed from S. By symmetry, the z-component will be equivalent. We start with,

$$v'_y = \frac{\mathrm{d}y'}{\mathrm{d}t'}$$

and

$$v_y = \frac{\mathrm{d}y}{\mathrm{d}t} = \frac{\mathrm{d}y}{\mathrm{d}t'}\frac{\mathrm{d}t'}{\mathrm{d}t}$$

But for this case $y' = y$ and so we can also write, using (3),

$$v_y = \frac{\mathrm{d}y'}{\mathrm{d}t'}\frac{\mathrm{d}t'}{\mathrm{d}t} = v'_y\frac{\mathrm{d}t'}{\mathrm{d}t} = v'_y\gamma\left(1 - \frac{u}{c^2}v_x\right)$$

We can't use this as our transformation of the velocity from S' to S because on the right hand side we have a mixture of quantities in both frames (*viz.*, v'_y and v_x). But we use this to write the transformation from S into S',

$$v'_y = \frac{v_y}{\gamma\left(1 - \frac{v_x u}{c^2}\right)}$$

And the reverse transformation is, as before, obtained simply by changing the sign of u,

$$v_y = \frac{v'_y}{\gamma\left(1 + \frac{v'_x u}{c^2}\right)}$$

Similar formulas apply to the z-component and we get the second transformation table as shown on page 17. I shan't do it here, but it's fairly easy to use the same method to find how the acceleration transforms.

Suppose an aeroplane is travelling in a straight line along the x-direction as observed from the ground, S; then, as Griffiths writes, the pilot may announce that we are travelling at a speed of $v_x = 500$ m.p.h. and by that she means the distance travelled on the ground divided by the time advanced by a clock on the ground. That is,

$$v_x = \frac{\mathrm{d}x}{\mathrm{d}t} \tag{5}$$

This is what matters if a passenger wished to know whether he will arrive in time. On the other hand if he cares whether he will be hungry on arrival he may be more interested in the distance covered during an interval of time, $\mathrm{d}\tau$, as shown on his watch. What he sees on his watch is called "proper time" (a corruption of the French *propre*) meaning his own time. We give this a special symbol τ because this is an invariant. The time that he perceives is the same in all intertial frames, because you can't alter what a person sees by being in a different frame. All the same the passenger's watch is running slowly compared to a watch on the ground which is why the journey as he experiences it will take longer than the actual time on the ground from take off to landing. By time dilation the interval $\mathrm{d}\tau$ is shorter than the interval $\mathrm{d}t$, †

$$\mathrm{d}t = \mathrm{d}\tau \, \frac{1}{\sqrt{1 - v_x^2/c^2}} \tag{6}$$

So *the passenger* is interested in the speed,

$$\eta_x = \frac{\mathrm{d}x}{\mathrm{d}\tau} \tag{7}$$

which is a strange mixed quantity called his *proper velocity*. A car speedometer measures proper velocity. Interestingly the proper velocity has a more fundamental place in special relativity because it transforms like a four-vector under Lorentz transformations. In fact, $\boldsymbol{\eta}$ is the spatial part of the four-vector

$$\eta^\mu = \frac{\mathrm{d}x^\mu}{\mathrm{d}\tau}$$

and the zeroth component is

$$\eta^0 = \frac{\mathrm{d}x^0}{\mathrm{d}\tau} = c\frac{\mathrm{d}t}{\mathrm{d}\tau} = \frac{c}{\sqrt{1 - v_x^2/c^2}} \tag{8}$$

† So even though he arrives at lunchtime, he may not actually be hungry. Muons are created by cosmic ray events in the upper atmosphere, and though their proper lifetime is only 2 μs, they do actually arrive on earth because to us their clocks are running slow. If the muon's speed is, say, $0.9c$ you can work out how far it travels before it decays.

using (6) and since $x^0 = ct$. The Lorentz transformation of the proper velocity four-vector is much simpler than that of the velocity that we have just derived because the Lorentz transformation only affects the numerator, since the denominator is invariant. In fact you can show using the Lorentz transformation that,

$$\begin{aligned} \eta'^0 &= \gamma\left(\eta^0 - \beta\eta^1\right) \\ \eta'^1 &= \gamma\left(\eta^1 - \beta\eta^0\right) \\ \eta'^2 &= \eta^2 \\ \eta'^3 &= \eta^3 \end{aligned} \tag{9}$$

although you could have written this down immediately since you know that all 4-vectors transform like x^μ (see page 18).

Transformation of volume

In a moment you will need the transformation of a volume, namely the little element of volume, $\mathrm{d}\mathcal{V} = \mathrm{d}x\mathrm{d}y\mathrm{d}z$, which we normally denote $\mathrm{d}\tau$ but in this appendix will denote $\mathrm{d}\mathcal{V}$ to avoid confusion with proper time. If the element of volume is a cube of side a, with its sides parallel to the x, y and z directions, and if it is at rest in S' then when observed from S only its length in the x-direction is contracted. Therefore its volume as observed in S' is $\mathrm{d}\mathcal{V}'$ and its volume observed in S is $\mathrm{d}\mathcal{V} = \mathrm{d}\mathcal{V}'/\gamma$. That's quite easy, but suppose the volume element is not at rest in either frame, but still the frames are in relative motion along the x-direction only. Then the velocity of the element in S' is v'_x and so to the observer who is stationary in S' its length is contracted in the x-direction and so in S',

$$\mathrm{d}\mathcal{V}' = a^3\sqrt{1-(v'_x/c)^2}$$

while to an observer in S in which the x-component of its velocity is v_x, its volume is,

$$\mathrm{d}\mathcal{V} = a^3\sqrt{1-(v_x/c)^2}$$

Now I know how to transform the velocity using equation (4). So inserting (4) in place of v_x, I get, after a fair amount of algebra,

$$\begin{aligned} \mathrm{d}\mathcal{V} &= a^3\,\frac{\sqrt{\left(1-u^2/c^2\right)\left(1-(v'_x/c)^2\right)}}{1+\dfrac{v'_x u}{c^2}} \\ &= \mathrm{d}\mathcal{V}'\,\frac{1}{\gamma\left(1-\dfrac{v'_x u}{c^2}\right)} \end{aligned}$$

The reverse transformation is, in S is,

$$\mathrm{d}\mathcal{V}' = \mathrm{d}\mathcal{V}\,\frac{1}{\gamma\left(1-\dfrac{v_x u}{c^2}\right)}$$

Even if the volume element is not in motion only along the x-axis in S or S' the result is unchanged. This is because the relative motion of S and S' is along the x-axis (that is the *definition* of S and S') and so the y and z-dimensions of $\mathrm{d}\mathcal{V}$ are not Lorentz contracted.

Momentum

Momentum is mass times velocity. But do we mean velocity as in equation (5) or proper velocity as in equation (7)? If relativistic momentum is to be conserved it is necessary to define it in terms of proper velocity, thus,

$$\mathbf{p} = m\boldsymbol{\eta}$$

In this way, the x-component of relativistic momentum is,

$$p_x = m\eta_x = \frac{1}{\sqrt{1 - v_x^2/c^2}} m v_x$$

By m we mean the invariant mass, or *rest mass*. In an older textbook, you may find that the author uses a subscript, say m_0, for the rest mass and then uses the symbol m to mean "relativistic mass". In the case of two inertial frames S and S' with a relative speed u along the common x-direction (as usual) the two are related (in the older notation) by

$$m = \gamma m_0$$

This means that Einstein's famous equation is $E = mc^2$, or $E = \gamma m_0 c^2$ (in the older notation). There is some sense in using relativistic mass—for example a physicist really will notice that a high speed particle requires a much bigger force to deflect it by a certain amount than the same particle at low speed, so that laboratory experience is that faster particles have a larger inertia. But on the other hand E is the relativistic energy and so it actually doesn't make sense to use both relativistic energy *and* relativistic mass, since they are the same quantity just related by multiplying by the constant c^2. So this has gone out of use now: we treat mass as invariant—the same in all inertial frames—just as charge is invariant. It saves us carrying around a redundant quantity, and the only price we pay is that Einstein's famous equation is modified to $E = \gamma mc^2$. If you find that too large a price then by all means carry on using relativistic mass: many physicists do. By the way, mass is not conserved, as we know from the destructive force of a nuclear bomb or the power generated by a nuclear reactor. By contrast, charge is conserved as well as invariant. Energy and momentum are conserved, but not invariant.

As an aside, the three components of momentum are part of the momentum 4-vector in special relativity. The zero component is $p^0 = \gamma mc = E/c$ which is the relativistic energy divided by c. You can show by taking the inner product of the momentum 4-vector with itself the famous formula for relativistic energy, namely,

$$E^2 = p^2c^2 + m^2c^4$$

Force

Force is rate of change of momentum,

$$\mathbf{F} = \frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t}$$

The transformation of force goes like this. In frame S',

$$\begin{aligned} F'_y &= \frac{\mathrm{d}p'_y}{\mathrm{d}t'} = \frac{\mathrm{d}p_y}{\mathrm{d}t'} = \frac{\mathrm{d}p_y}{\mathrm{d}t}\frac{\mathrm{d}t}{\mathrm{d}t'} \\ &= \frac{\mathrm{d}p_y}{\mathrm{d}t}\frac{1}{\gamma\left(1-\frac{v_x u}{c^2}\right)} \\ &= \frac{F_y}{\gamma\left(1-\frac{v_x u}{c^2}\right)} \end{aligned}$$

In the first line I used $p_y = p'_y$ since momentum is only transformed in the direction of relative motion; and in the second line I used (3); and finally I used

$$F_y = \frac{\mathrm{d}p_y}{\mathrm{d}t}$$

By symmetry the same goes for the z-direction and we get the second and third lines in the transformation table for force on page 17. It's messier to get the x-component. We need to know how a 4-vector transforms, see page 18. The algebra goes as follows,

$$\begin{aligned} F'_x &= \frac{\mathrm{d}p'_x}{\mathrm{d}t'} \\ &= \frac{\gamma\mathrm{d}p_x - \gamma\beta\mathrm{d}p^0}{\gamma\mathrm{d}t - \frac{\gamma\beta}{c}\mathrm{d}x} \\ &= \frac{1}{1-\frac{\beta}{c}\frac{\mathrm{d}x}{\mathrm{d}t}}\left(\frac{\mathrm{d}p_x}{\mathrm{d}t} - \beta\frac{\mathrm{d}p^0}{\mathrm{d}t}\right) \\ &= \frac{1}{1-\frac{u}{c^2}v_x}\left(F_x - \frac{u}{c^2}\frac{\mathrm{d}E}{\mathrm{d}t}\right) \end{aligned}$$

and we have used the fact that the zero component of the momentum 4-vector is E/c. So what is $\mathrm{d}E/\mathrm{d}t$? This is the rate at which the force is doing work on the particle which is travelling at speed v_x; in fact,

$$\frac{\mathrm{d}E}{\mathrm{d}t} = \mathbf{F}\cdot\mathbf{v} = F_x v_x + F_y v_y + F_z v_z$$

so that

$$F'_x = \frac{1}{1-\frac{u}{c^2}v_x}\left(F_x - \frac{u}{c^2}\left(F_x v_x + F_y v_y + F_z v_z\right)\right)$$

and finally we get the first line in the transformation of force table on page 17,

$$F'_x = F_x - \frac{u}{c^2 - u v_x}\left(v_y F_y + v_z F_z\right)$$

We get the reverse transformation, as usual, just by reversing the sign of u.

Out of curiosity you might ask why we don't define the force in relativity as the rate of change of momentum with respect to a particle's proper time,

$$\mathbf{K} = \frac{\mathrm{d}\mathbf{p}}{\mathrm{d}\tau}$$

In fact, $\mathbf{K}$ is called the *Minkowski force* and relativistic dynamics can be formulated in terms of $\mathbf{K}$ rather than $\mathbf{F}$. An advantage is that K^μ is a 4-vector, which is rather obvious since p^μ is a 4-vector and τ is an invariant. The zeroth component of K^μ is,

$$K^0 = \frac{1}{c}\frac{\mathrm{d}E}{\mathrm{d}\tau}$$

which is the power delivered into the particle per unit of its proper time. The transformation of $\mathbf{K}$ would have been much easier just as the transformation of proper velocity was easier, but the important fact is that it turns out that *the Lorentz force is an ordinary force,* not a Minkowski force, so we need to use it.

The current density 4-vector

Think of some charge distribution which is moving at uniform velocity. Because it's moving there is a *current*. If this is now observed from a different inertial frame, what charge density and what current is observed? This, of course, relates to the problem we worked out in Section 3.1.

Imagine a collection of $\mathrm{d}n$ charges of amount q. If these are contained within a infinitesimal volume element $\mathrm{d}\mathsf{v}$ then the charge density,

$$\rho = \frac{q\mathrm{d}n}{\mathrm{d}\mathsf{v}}$$

is uniform. If this is viewed from frame S' then

$$\rho' = \frac{q\mathrm{d}n}{\mathrm{d}\mathsf{v}'}$$

Here $\mathrm{d}\mathsf{v} = \mathrm{d}x\mathrm{d}y\mathrm{d}z$ and $\mathrm{d}\mathsf{v}' = \mathrm{d}x'\mathrm{d}y'\mathrm{d}z'$ are infinitesimal elements of volume; $q\mathrm{d}n$, the amount of charge, is invariant. Therefore,

$$\rho' = \rho\frac{\mathrm{d}\mathsf{v}}{\mathrm{d}\mathsf{v}'}$$

The collection of charges is travelling with velocity $\mathbf{v}$ in S and $\mathbf{v}'$ in S'. The transformation of a volume element as given on page 22 is,

$$\mathrm{d}\mathsf{v}' = \mathrm{d}\mathsf{v}\,\frac{1}{\gamma\left(1-\frac{v_x u}{c^2}\right)}$$

and so the transformation of the charge density is,

$$\rho' = \rho\,\gamma\left(1 - \frac{v_x u}{c^2}\right)$$

and in reverse,

$$\rho = \rho'\,\gamma\left(1 + \frac{v'_x u}{c^2}\right)$$

If the charge is stationary in the laboratory frame, S, then we recover equation (3.1.2), $\rho' = \gamma\rho$. However, if the collection of charges has a speed v_x along the x-axis in S and a speed v'_x along x' in S', then there are current densities observed in either frame,

$$J_x = \rho v_x \;\; ; \quad \text{and} \quad J'_x = \rho' v'_x$$

I insert these into the previous two equations and I get,

$$\rho = \gamma\left(\rho' + \frac{u}{c^2}J'_x\right)$$

and

$$\rho' = \gamma\left(\rho - \frac{u}{c^2}J_x\right)$$

This is how charge density transforms between two inertial frames. I can now deduce the transformation of the current density as follows,

$$\begin{aligned} J_x &= \rho v_x \\ &= \rho'\,\gamma\left(1 + \frac{v'_x u}{c^2}\right)\frac{v'_x + u}{1 + \frac{v'_x u}{c^2}} \\ &= \gamma\,(J'_x + u\rho') \end{aligned}$$

In the second line I transformed ρ using the formula just above, and multiplied that with v_x transformed using (4). Similarly,

$$J'_x = \gamma\,(J_x - u\rho)$$

Now I need the transformation of J_y and J_z. Again I write the current density as the product of a charge density and a velocity and I transform each,

$$\begin{aligned} J_y &= \rho v_y \\ &= \rho'\,\gamma\left(1 + \frac{v'_x u}{c^2}\right)\frac{v'_y}{\gamma\left(1 + \frac{v'_x u}{c^2}\right)} \\ &= \rho' v'_y = J'_y \end{aligned}$$

since everything cancels; this is rather obvious since the current density is not expected to be altered except in its component in the x direction. Similarly, $J_z = J'_z$. I can now summarise what we have discovered in a table as on page 17,

$\rho = \gamma\left(\rho' + \frac{u}{c^2}J'_x\right)$	$\rho' = \gamma\left(\rho - \frac{u}{c^2}J_x\right)$
$J_x = \gamma\left(J'_x + u\rho'\right)$	$J'_x = \gamma\left(J_x - u\rho\right)$
$J_y = J'_y$	$J'_y = J_y$
$J_z = J'_z$	$J'_z = J_z$

and if you compare this with the first table on page 17 you see it is identical as long as I replace x with J_x, y with J_y, z with J_z and t with ρ. This means that charge and current density transform as a 4-vector,

$$J^\mu = (c\rho, J_x, J_y, J_z)$$

which transforms like all other 4-vectors—for example equation (9)—so I don't need to include a table on page 17. It is by manipulating and Lorentz transformation of this that you can solve Feynman's problem of Section 3.1 in the case that the outside electron is travelling at a different speed to the drift velocity in the wire.

You may recall that while the distance and time difference between two events appear to be different when observed from different frames, there is a quantity called the *invariant interval* which is the same in all frames. If two events occur at x^μ_A and x^μ_B then the *displacement 4-vector* is $d^\mu = x^\mu_B - x^\mu_A$, and the invariant interval between events A and B is $d^\mu d_\mu$. You may recall furthermore that if this interval is positive then the displacement 4-vector is *spacelike*, if it's negative it is *timelike*, and if it's zero it's *lightlike*. In the same way, while, as we have seen in Section 3.1, the charge and current are different when viewed from different frames, the scalar product of J^μ with itself is an invariant.

$$J^\mu J_\mu = J^2 - c^2\rho^2$$

with Einstein summation over $\mu = 0 \ldots 3$: the same in all frames. So, for example in the problem of Section 3.1—the current carrying wire—the charge density in S is zero and the current is carried by the electrons; while in S' the wire is not neutral and the current is carried by the ions. Clearly ρ and $\mathbf{J}$ are not the same in each frame, but the invariant $J^2 - c^2\rho^2$ *is* the same in all frames.

The electromagnetic field

We have seen that the amount of charge and current in a given situation, as indeed the length and time difference between events, will depend on the inertial frame belonging to the observer. In fact by changing the frame of the observer, current can become charge and vice-versa: only the combination $J^\mu J_\mu$ is invariant. Furthermore we have seen in Section 3.2 that electric and magnetic fields possess a similar property—what in one frame appears as a magnetic field, in another frame is observed to be an electric field.

So we now need to derive the final table on page 17, showing how an electromagnetic field as observed in frame S looks if observed from frame S'. We start with the Lorentz force.

We imagine a point charge, q, that is moving through electric and magnetic fields with velocity $\mathbf{v}$ in S. The force it experiences is,

$$\mathbf{F} = q\left(\mathbf{E} + \mathbf{v} \times \mathbf{B}\right)$$

When viewed in S' the force is (as always we place primes on quantities observed in S'),

$$\mathbf{F}' = q\left(\mathbf{E}' + \mathbf{v}' \times \mathbf{B}'\right) \tag{10}$$

I don't put a prime on q because charge is invariant (the same in all inertial frames). We write $\mathbf{F}$ in its components,

$$\begin{aligned} F_x &= q\left(E_x + v_y B_z - v_z B_y\right) \\ F_y &= q\left(E_y + v_z B_x - v_x B_z\right) \\ F_z &= q\left(E_z + v_x B_y - v_y B_x\right) \end{aligned} \tag{11}$$

We know that these forces when observed from S' will have been transformed and I can transform these components of force using the table on page 17,

$$\begin{aligned} F'_x &= F_x - \frac{u}{c^2 - v_x u}\left(v_y F_y + v_z F_z\right) \\ F'_y &= \frac{F_y}{\gamma\left(1 - \frac{v_x u}{c^2}\right)} \\ F'_z &= \frac{F'_z}{\gamma\left(1 - \frac{v_x u}{c^2}\right)} \end{aligned} \tag{12}$$

My task is to take the force components, F_x, F_y and F_z in (12) and replace them with the Lorentz forces in (11). The idea is to examine the components of the force in S' and identify how they fit into (10). Then I will have the forces as observed in S' in terms of the electromagnetic field components in S. However (12) contains velocities in S and I want them in S' so I will also need to substitute them with the velocities in S' using the transformation table on page 17,

$$\begin{aligned} v_x &= \frac{v'_x + u}{1 + \frac{v'_x u}{c^2}} \\ v_y &= \frac{v'_y}{\gamma\left(1 + \frac{v'_x u}{c^2}\right)} \\ v_z &= \frac{v'_z}{\gamma\left(1 + \frac{v'_x u}{c^2}\right)} \end{aligned} \tag{13}$$

So I need to put (11) and (13) into (12) and see what I get. As always in special relativity the maths is not particularly hard but the algebra is ugly. In this case it is particularly messy. The algebra goes like this. It's best first to define, say, λ and λ', such that,

$$\lambda = 1 - \frac{v_x u}{c^2} \qquad ; \quad \lambda' = 1 + \frac{v'_x u}{c^2}$$

Then, to get first the y-component of $\mathbf{F}'$, from (12) and (13),

$$F'_y = \frac{F_y}{\gamma\lambda} \quad ; \quad v_x = \frac{1}{\lambda'}\left(v'_x + u\right) \quad ; \quad v_z = \frac{v'_z}{\gamma\lambda'}$$

You will also see, if you expand it out, that,

$$\lambda\lambda' = 1 - \frac{u^2}{c^2} = \frac{1}{\gamma^2}$$

Now, starting with (11) and transforming F_y,

$$\begin{aligned} F'_y &= \frac{q}{\gamma\lambda}\left(E_y + \frac{1}{\gamma\lambda'}v'_z B_x - \frac{1}{\lambda'}\left(v'_x + u\right)B_z\right) \\ &= \frac{q}{\gamma\lambda\lambda'}\left(\lambda' E_y + \frac{1}{\gamma}v'_z B_x - \left(v'_x + u\right)B_z\right) \\ &= q\left(\gamma\left(1 + \frac{v'_x u}{c^2}\right)E_y + v'_z B_x - \gamma\left(v'_x + u\right)B_z\right) \\ &= q\left(\gamma\left(E_y - uB_z\right) + v'_z B_x - v'_x\gamma\left(B_z - \frac{u}{c^2}E_y\right)\right) \\ &= q\left(E'_y + v'_z B'_x - v'_x B'_z\right) \end{aligned}$$

This is the y'-component of the Lorentz force in S' (10) and by a comparison of terms, I find,

$$\begin{aligned} E'_y &= \gamma\left(E_y - uB_z\right) \\ B'_x &= B_x \\ B'_z &= \gamma\left(B_z - \frac{u}{c^2}E_y\right) \end{aligned} \tag{14}$$

We make progress! I have three out of the six relations I need. A repeat of the above for F'_z furnishes us with two further relations, namely,

$$\begin{aligned} E'_z &= \gamma\left(E_z + uB_y\right) \\ B'_y &= \gamma\left(B_y + \frac{u}{c^2}E_z\right) \end{aligned} \tag{15}$$

One more is needed. Simply work out F'_x for the case that $v_x = v_z = 0$ and obtain,

$$E'_x = E_x \tag{16}$$

Equations (14), (15) and (16) are the relations gathered in the last table on page 17.

If you want a completely different way to obtain these results, which is more physically motivated and hence very insightful, then consult Griffiths. He demonstrates (16) by imagining a capacitor with charged plates arranged perpendicular to the common x–x' axis. The area of the plates is the same in both frames so the charge density is the same in S and S'; and the gap between the plates is Lorentz contracted. However, the electric field does not depend on the plate separation and hence is the same in both frames.

He also has a lovely way of showing that $B'_x = B_x$. Imagine an electric coil oriented along the common x–x' axis, having n turns per metre, carrying a current, I, and stationary in S in which the magnetic field is $\mu_0 nI$. When viewed from S' the observer sees the coil shortened by the Lorenz contraction so the number of turns per metre is more by a factor γ. But the clock in S is running slowly as far as the observer is S' is concerned so the current is accordingly smaller in S' by a factor $1/\gamma$. So the two effects cancel and the magnetic field is the same in both frames!

The electromagnetic potentials

It appears that **B** and **E** do not transform each like the spatial part of a 4-vector, since they get mixed up together. Of course you can't make a 4-vector out of six components. In fact, as Griffiths explains, the six components of **E** and **B** may be combined into an antisymmetric, second rank, so called *field tensor*,

$$F^{\mu\nu} = \begin{pmatrix} 0 & \frac{1}{c}E_x & \frac{1}{c}E_y & \frac{1}{c}E_z \\ -\frac{1}{c}E_x & 0 & B_z & -B_y \\ -\frac{1}{c}E_y & -B_z & 0 & B_x \\ -\frac{1}{c}E_z & B_y & -B_x & 0 \end{pmatrix}$$

If we look for a 4-vector that represents the electromagnetic field then we come upon the notion that it is the *potentials*, not the *fields* that are most fundamental in relativity. Indeed this also the case in quantum mechanics—we need to express the Hamiltonian operator in terms of the scalar and vector potentials, V and **A** (see Section 1), not the fields.

Possibly the most fundamental object in electromagnetism is the *4-potential*,

$$A^\mu = \left(\frac{1}{c}V, A_x, A_y, A_z\right)$$

Using the transformation of a 4-vector you can easily make a Lorentz transformation of this. Of course what happens is that a magnetic potential in one frame may look like an electric potential in another frame. You may check, or look up in Griffiths, that,

$$F^{\mu\nu} = \frac{\partial A^\nu}{\partial x_\mu} - \frac{\partial A^\mu}{\partial x_\nu}$$

which is indeed antisymmetric.

I have already said that our central equations are (3.2.7) because these describe the electric and magnetic fields due to a moving charge, specifically with a velocity $\mathbf{u} = u\hat{\mathbf{i}}$; and by the principle of superposition they may be used to describe collections of charges, or a charge density as in Section 3.3. Let us now use equations (3.2.7) to deduce the scalar and vector potentials due to a moving charge. You can do the maths at home, or look it up in Lorrain and Corson. I will state the result,

$$\mathbf{A} = \hat{\mathbf{i}}\ \frac{\mu_0}{4\pi}\ q\,\gamma\ \frac{u}{\sqrt{\gamma^2\,(x-ut)^2 + y^2 + z^2}} \tag{17}$$

(It is generally the case that the vector potential in a problem points in the direction of the current.) You can show using (3.2.7b) that (17) is consistent with,

$$\mathbf{B} = \boldsymbol{\nabla} \times \mathbf{A}$$

which is the definition of the vector potential.

You already know that in electrostatics, the scalar electric potential, V, is defined using,

$$\mathbf{E} = -\boldsymbol{\nabla} V$$

When we come to it in Section 5, you will learn in electrodynamics,

$$\mathbf{E} = -\boldsymbol{\nabla} V - \frac{\partial \mathbf{A}}{\partial t} \tag{18}$$

However, we can show this now. By taking spatial derivatives of the following equation (19) you will arrive at (3.2.7a). In this way,

$$V = \frac{1}{4\pi\epsilon_0}\, q\,\gamma\, \frac{1}{\sqrt{\gamma^2 (x-ut)^2 + y^2 + z^2}} \tag{19}$$

which, of course, reduces to the potential due to a stationary point charge in the case $u = 0$, $\gamma = 1$. If you take both spatial and time derivatives of (19) and (17) you will be able to confirm (18).

Finally, you can confirm that the 4-potential transforms like all 4-vectors by using the transformations of $\mathbf{E}$ and $\mathbf{B}$ from the last table on page 17.

In Section 5, we will come across the formula,

$$\frac{1}{c^2}\frac{\partial V}{\partial t} + \boldsymbol{\nabla} \cdot \mathbf{A} = 0$$

which expresses the "Lorenz gauge". If I define a differential operator 4-vector,

$$\begin{aligned} \partial_\mu &= \left(\frac{\partial}{\partial x^0}, \frac{\partial}{\partial x^1}, \frac{\partial}{\partial x^2}, \frac{\partial}{\partial x^3} \right) \\ &= \left(\frac{1}{c}\frac{\partial}{\partial t}, \frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right) \end{aligned}$$

then I have the scalar product with the 4-potential,

$$\partial_\mu\, A^\mu = \frac{\partial A^\mu}{\partial x^\mu} = 0$$

which neatly sums up the Lorenz gauge. In fact this is the *4-divergence* of A^μ which turns out to vanish in the Lorenz guage.

We will come to the equation of continuity in Section 4; it states,

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

which we can now also write as,

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

In words, "the divergence of the current density 4-vector is zero." This is the law of charge conservation (not to be confused with charge invariance).

If I make the inner product of ∂_μ with

$$\partial^\mu = \left(-\frac{1}{c}\frac{\partial}{\partial t}, \frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right)$$

I get

$$\begin{aligned} \partial_\mu \partial^\mu &= \left(\frac{1}{c}\frac{\partial}{\partial t}, \boldsymbol{\nabla} \right) \cdot \left(-\frac{1}{c}\frac{\partial}{\partial t}, \boldsymbol{\nabla} \right) \\ &= -\frac{1}{c^2}\frac{\partial}{\partial t^2} + \nabla^2 \end{aligned} \tag{20}$$

This differential operator is called the D'Alembertian and is the equivalent of the Laplace operator in Minkowski spacetime. It is usually given the symbol $\Box$ in electromagnetism (or sometimes $\Box^2$ in analogy with ∇^2).†

The beautiful thing is that we can use this to express equations (5.3.10) in Section 5, which are the Maxwell equations in terms of the potentials, in the wonderfully elegant and simple,

$$\Box A^\mu = -\mu_0 \, J^\mu$$

Of course this again shows that Maxwell's equations are Lorentz invariant and have survived the revolution of Einstein's special relativity intact. In the absence of sources, "box A is zero". How simple is that?

† Authors who use the $(+ - - -)$ metric signature define the D'Alembertian as $-\partial_\mu \partial^\mu$, so it comes out the same as (20).

4. Maxwell's equations

4.1 All of classical physics

According to Feynman, this is all of classical physics.

Maxwell's equations

I. $$\boldsymbol{\nabla}\cdot\mathbf{E}=\frac{\rho}{\epsilon_0}\ ,\ \text{(flux of }\mathbf{E}\text{ through a closed surface)}=\frac{1}{\epsilon_0}\text{(charge inside)}$$

II. $$\boldsymbol{\nabla}\times\mathbf{E}=-\frac{\partial\mathbf{B}}{\partial t}\ ,\ \begin{matrix}\text{(line integral of}\\ \mathbf{E}\text{ around a loop)}\end{matrix}=-\frac{\partial}{\partial t}\text{(flux of }\mathbf{B}\text{ through the loop)}$$

III. $$\boldsymbol{\nabla}\cdot\mathbf{B}=0\ ,\ \text{(flux of }\mathbf{B}\text{ through a closed surface)}=\text{zero}$$

IV. $$c^2\boldsymbol{\nabla}\times\mathbf{B}=\frac{1}{\epsilon_0}\mathbf{J}+\frac{\partial\mathbf{E}}{\partial t}\ ,\ \begin{matrix}c^2\text{(integral of }\mathbf{B}\\ \text{around a loop)}\end{matrix}=\begin{matrix}\frac{1}{\epsilon_0}\text{(current through the loop)}+\\ \frac{\partial}{\partial t}\text{(flux of }\mathbf{E}\text{ through the loop)}\end{matrix}$$

Conservation of charge

$$\boldsymbol{\nabla}\cdot\mathbf{J}=-\frac{\partial\rho}{\partial t}\ ,\ \begin{matrix}\text{(flux of current density}\\ \text{through a closed surface)}\end{matrix}=\frac{\partial}{\partial t}\text{(charge inside)}$$

Lorentz force law

$$\mathbf{F}=q\left(\mathbf{E}+\mathbf{v}\times\mathbf{B}\right)$$

Law of motion

$$\frac{\partial}{\partial t}\mathbf{p}=\mathbf{F}\ ,\ \text{where }\mathbf{p}=\frac{m\mathbf{v}}{\sqrt{1-\frac{v^2}{c^2}}}\ ,\ \text{(Newton's law with Einstein's modification)}$$

Gravitation

$$\mathbf{F}=-G\frac{m_1m_2}{r^2}\,\mathbf{e}_r$$

4.2 Maxwell's equations

Most of the above you know already. The conservation law reflects the fact that charge cannot be created or destroyed. It can only move from one place to another. So considering a volume Ω the charge inside that volume at some time, t, must be

$$Q(t) = \int_\Omega \rho(\mathbf{r}, t)\, \mathrm{d}\tau$$

If charge enters or leaves the volume then there must be current flowing through the surface, S, that bounds the volume. Hence the rate of change of the charge contained in Ω is[†]

$$\frac{\mathrm{d}Q}{\mathrm{d}t} = -\oint_S \mathbf{J} \cdot \mathrm{d}\mathbf{a}$$

Combining these two equations,

$$\int_\Omega \frac{\partial \rho}{\partial t} \mathrm{d}\tau = -\oint_S \mathbf{J} \cdot \mathrm{d}\mathbf{a} = -\int_\Omega \boldsymbol{\nabla} \cdot \mathbf{J}\, \mathrm{d}\tau$$

where I have used the divergence theorem. This is true for any volume, so the integrands must be equal. It follows that,

$$\boxed{\frac{\partial \rho}{\partial t} = -\boldsymbol{\nabla} \cdot \mathbf{J}}$$

This powerful conservation law, or continuity equation, applies in many branches of physics and engineering when we deal with the flow of a conserved quantity—for example the mass density in fluid dynamics.

The first three equations in section 4.1 we recognise as,

I. : Gauss's law
II. : Faraday's law of induction
III. : non existence of magnetic charges

The fourth,

IV. $$c^2 \boldsymbol{\nabla} \times \mathbf{B} = \frac{1}{\epsilon_0}\mathbf{J} + \frac{\partial \mathbf{E}}{\partial t}$$

is Ampère's law,

$$\boldsymbol{\nabla} \times \mathbf{B} = \mu_0 \mathbf{J}$$

[†] The minus sign is important: the vector $\mathrm{d}\mathbf{a}$ points to the outside of the closed surface; if charge enters then $\dot{Q} > 0$ and so $\mathbf{J}$ points inwards since charge is flowing *into* Ω.

with

$$c^2 = \frac{1}{\epsilon_0 \mu_0}$$

and an extra term! Where does that come from? Ampère was not able to detect this in his experiments and Maxwell deduced it purely by theorising. A modern explanation is this.

We know that the divergence of a curl is always zero (section 1). Faraday's law is consistent with this because,

$$\boldsymbol{\nabla} \cdot (\boldsymbol{\nabla} \times \mathbf{E}) = \boldsymbol{\nabla} \cdot \left(-\frac{\partial \mathbf{B}}{\partial t} \right) = -\frac{\partial}{\partial t} (\boldsymbol{\nabla} \cdot \mathbf{B}) = 0$$

But applied to Ampère's law we get,

$$\boldsymbol{\nabla} \cdot (\boldsymbol{\nabla} \times \mathbf{B}) = \mu_0 \boldsymbol{\nabla} \cdot \mathbf{J} \tag{4.2.1}$$

and this is only zero for *steady currents*, that is, in magnetostatics. In fact you can see that Ampère's law will fail in the following example. In the circuit in figure 4–1 a capacitor is being charged.

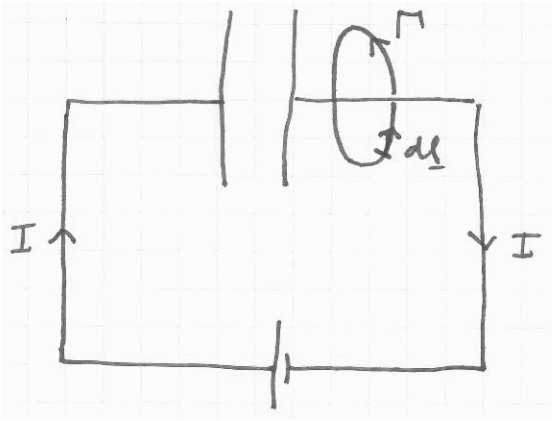


FIGURE 4–1

In integral form Ampère's law is

$$\oint_\Gamma \mathbf{B} \cdot \mathrm{d}\boldsymbol{\ell} = \mu_0 I_{\text{enclosed}}$$

around an amperian loop such as the one shown in figure 4–1. What we mean by the current enclosed, I_{enclosed}, or "linked" by the loop is the current density integrated over any surface bounded by the loop. If I use the obvious flat surface bounded by Γ as in figure 4–2 then

$$I_{\text{enclosed}} = I$$

as it should be.

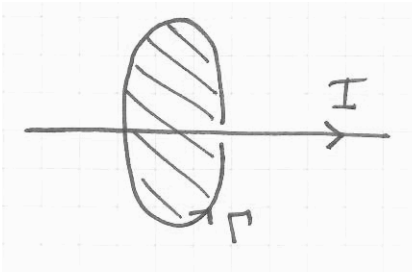


FIGURE 4–2

But suppose I use a surface that encloses one plate of the capacitor as shown in figure 4–3?

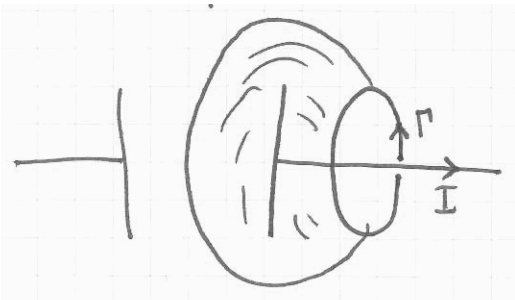


FIGURE 4–3

No current links (penetrates) this surface and

$$I_{\text{enclosed}} = 0$$

The issue is with the form of the right hand side of (4.2.1) namely, $\boldsymbol{\nabla} \cdot \mathbf{J}$ which should be zero, but isn't.

From the continuity law and Gauss's law,

$$\begin{aligned} \boldsymbol{\nabla} \cdot \mathbf{J} &= -\frac{\partial \rho}{\partial t} = -\frac{\partial}{\partial t}\left(\epsilon_0 \boldsymbol{\nabla} \cdot \mathbf{E}\right) \\ &= -\boldsymbol{\nabla} \cdot \epsilon_0 \frac{\partial \mathbf{E}}{\partial t} \end{aligned}$$

This suggests to us that since, from this,

$$\boldsymbol{\nabla} \cdot \left(\mathbf{J} + \epsilon_0 \frac{\partial \mathbf{E}}{\partial t}\right) = 0$$

we replace $\mathbf{J}$ in Ampère's law with the two terms in the parentheses. Ampère's law becomes modified to,

$$\boldsymbol{\nabla} \times \mathbf{B} = \mu_0 \mathbf{J} + \mu_0 \epsilon_0 \frac{\partial \mathbf{E}}{\partial t} \qquad \text{IV.}$$

This changes nothing in magnetostatics since in that case nothing depends on time (by definition). Note that there is now a parallel with Faraday's law since here we see that a changing *electric* field will "induce" a *magnetic* field.

We call $\epsilon_0 \partial \mathbf{E}/\partial t$ the "vacuum displacement current density". You can see that this now fixes the capacitor dilemma. In the second choice of surface bounded by Γ, figure 4–3, there is no linked current, but there is a changing electric flux penetrating the surface. Assuming we can neglect edge effects (large surface area of capacitor plates, small separation between plates) we have

$$E = \frac{\sigma}{\epsilon_0} = \frac{1}{\epsilon_0} \frac{\text{charge on plate}}{\text{area of plate}} = \frac{1}{\epsilon_0} \frac{Q}{A}$$

Between the plates,

$$\frac{\partial E}{\partial t} = \frac{1}{\epsilon_0 A} \frac{\partial Q}{\partial t} = \frac{1}{\epsilon_0 A} I \tag{4.2.2}$$

In integral form our modified Ampère's law is,

$$\oint_\Gamma \mathbf{B} \cdot \mathrm{d}\boldsymbol{\ell} = \mu_0 I_{\text{enclosed}} + \mu_0 \epsilon_0 \int_S \frac{\partial \mathbf{E}}{\partial t} \cdot \mathrm{d}\mathbf{a}$$

and while for this choice of surface $I_{\text{enclosed}} = 0$,

$$\mu_0 \epsilon_0 \int_S \frac{\partial \mathbf{E}}{\partial t} \cdot \mathrm{d}\mathbf{a} = \mu_0 I$$

using (4.2.2) for a flat surface and constant $\mathbf{E} \| \mathrm{d}\mathbf{a}$, and so for either surface

$$\oint_\Gamma \mathbf{B} \cdot \mathrm{d}\boldsymbol{\ell} = \mu_0 I$$

and Ampère's law is rescued!

Actually, as Lorrain and Corson explain, Maxwell's correction to Ampère's law can be deduced by insisting that Gauss's law is invariant under a Lorentz transformation. We want to calculate the curl of $\mathbf{B}$ that arises from charges that are *moving* in the laboratory frame, S. We can ignore charges that are stationary in that frame. To start with to make things simple, imagine that there is a density of charge, ρ, moving in the x-direction in S with a speed u. Then,

$$\mathbf{J} = \rho\, u\, \hat{\mathbf{i}}$$

In frame S' which moves at a velocity $u\,\hat{\mathbf{i}}$ with respect to S, $\mathbf{J}' = \mathbf{0}$ and owing to the Lorentz contraction of the volume, the charge density in S' is

$$\rho' = \frac{\rho}{\gamma}$$

where,

$$\gamma = \frac{1}{\sqrt{1 - \dfrac{u^2}{c^2}}} = \frac{1}{\sqrt{1 - \beta^2}} \quad ; \qquad 1 \leq \gamma < \infty$$

In S', $\mathbf{B}' = \mathbf{0}$ because the charges we are interested in are stationary; and Gauss's law asserts that,

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

I now transform this into frame S using derivatives of the Lorentz transformation of coordinates and the rules for transforming the $\mathbf{B}$ and $\mathbf{E}$ fields (see section 3, appendix),

$$\gamma \left(\frac{\partial}{\partial x} + \frac{u}{c^2} \frac{\partial}{\partial t} \right) E_x + \gamma \frac{\partial}{\partial y} \left(E_y - u B_z \right) + \gamma \frac{\partial}{\partial z} \left(E_z + u B_y \right) = \frac{\rho}{\gamma \epsilon_0}$$

I gather up some terms and this becomes the same as,

$$\boldsymbol{\nabla} \cdot \mathbf{E} + \frac{u}{c^2} \frac{\partial E_x}{\partial t} - u \left(\frac{\partial B_z}{\partial y} - \frac{\partial B_y}{\partial z} \right) = \left(1 - \beta^2 \right) \frac{\rho}{\epsilon_0}$$

(See how the curl of $\mathbf{B}$ is turning up?) Because Gauss's law applies of course in frame S, then $\boldsymbol{\nabla} \cdot \mathbf{E} = \rho / \epsilon_0$. So dividing the above equation through by u, I get, remembering that $\beta = u/c$,

$$\frac{1}{c^2} \frac{\partial E_x}{\partial t} - \left(\frac{\partial B_z}{\partial y} - \frac{\partial B_y}{\partial z} \right) = - \frac{\rho u}{\epsilon_0 c^2} = - \frac{J_x}{\epsilon_0 c^2} = - \mu_0 \, J_x$$

This is the result for the case of a current density along x. If we generalise to three components of $\mathbf{J}$, the result is,

$$\boldsymbol{\nabla} \times \mathbf{B} - \frac{1}{c^2} \frac{\partial \mathbf{E}}{\partial t} = \mu_0 \, \mathbf{J}$$

which is the Ampère–Maxwell law, here *derived* from Gauss's law and the Lorentz transformation.

Appendix—The Story of the "Maxwell Equations" (copied from Lorrain and Corson)

The story of the Maxwell equations is fascinating in many ways.

First, as we shall see, the equations should be called, more appropriately, the Heaviside equations.

Second, the story covers less than the fourth quarter of the nineteenth century.

Third, there were few *Maxwellians*, the main actors being Oliver Heaviside (1850–1925), who was self-taught, George Francis FitzGerald (1851–1901), who worked at Trinity College Dublin, Oliver Lodge (1851–1940), who was at University College in Liverpool, and Heinrich Hertz (1857–1894), who was an experimental physicist at the Karlsruhe Physical Institute. John-Henry Poynting (1852–1914) contributed a major concept when he proposed his theorem in 1884, but Heaviside discovered it practically at the same time. James Clerk Maxwell (1831–1879) was an actor through the publication in 1873 of his *Treatise on Electricity and Magnetism*. At the time of his death, he was working on a second edition of the *Treatise*.

Fourth, all the actors were British, except for Hertz, who was German. Maxwell, who was Scottish, became Professor of Experimental Physics at Cambridge but, after his death, there were no Maxwellians at Cambridge and none at Oxford.

Fifth, the main actor was Heaviside, who had to leave school when he was sixteen. He started life as a telegrapher in Newcastle, where he stayed for eight years, but had no real job for the rest of his life. He was a prolific writer, despite the fact that he had poor health and lived as a recluse, partly by temperament, partly because of his deafness, and partly because of his great poverty. For many years he published his theories in *The Electrician*, which was a trade journal. For this he was paid 40 pounds per year, which was less than the salary of a laborer. Near the end of his life he received a modest government pension of 120 pounds per year, but remained exceedingly poor and miserable, despite the many honors that were bestowed upon him. Although he was self-taught, he had become one of the best, if not *the* best, physicist-mathematician of his time.

Maxwell's *Treatise* is unreadable. [Bruce] Hunt ["The Maxwellians", Cornell University Press, 1994], who clearly regrets having spent so much time on it, says it is rambling, obscure, inconsistent, contradictory, awkward, confusing, and on some points simply wrong; that it has no clear focus, no orderly presentation, and that the mathematics is clumsy and involved! It is a wonder that the Maxwellians managed to master it. Maxwell had great admiration for Faraday, and he was attempting to put Faraday's ideas in mathematical form.

The *Treatise* does *not* state the Maxwell equations because Maxwell reasoned mostly in terms of the potentials V (then called ψ) and $\mathbf{A}$. Shifting from V and $\mathbf{A}$ to $\mathbf{E}$

and $\mathbf{B}$ caused much debate about the "murder of ψ." The potentials then came to be considered as mathematical fictions.

Heaviside focused on the electromagnetic *field* [...]. His main motivation was to improve signaling on submarine cables. He finally recast the Maxwell theory in the form of the four vector "Maxwell" equations in 1884, in the same year that Hertz stated them in Cartesian form. Hertz admitted that Heaviside had the priority, in view of the fact that his own "proof" was "very shaky." At the time, vector notation seemed to everyone to be highly esoteric.†

Modern Physics owes much to Heaviside. He was the first to apply vector analysis as we know it today to electromagnetic fields, and the first to use *rationalized* units. (With unrationalized units, factors of 4π appear in most of the important equations.) He discovered the *Poynting* theorem the same year as Poynting. He proposed the existence of the Heaviside layer and developed Operational Calculus. He also discovered the "Lorentz" force $Q\mathbf{v} \times \mathbf{B}$ fifteen years before Hendrick Lorentz (1853–1928). That was a most important result because it showed how the field contracts lengthwise and that *the contraction involves the factor γ of relativity!"* This phenomenon was later called the *FitzGerald contraction* and served to explain the negative result of the Michelson-Morley experiment.

† Vector calculus was invented independently and simultaneously by Josiah Willard Gibbs (1839–1903) at Yale University. The notation as we use it today is due to Gibbs. [A. T. P.]

5. Scalar and vector potential

5.1 Electric scalar potential

We have not been given an absolute zero of potential energy in the Universe. Generally in physics we are concerned only with *differences* or *changes* in potential energy. To establish a zero of energy we have to adopt a convention. For example, in electrostatics, we usually assert that,

$$V(\mathbf{r}) = 0 \qquad \text{as} \qquad r \to \infty$$

In fact the potential is deduced from the electric field which does have an absolute value. You know that,

$$\mathbf{E} = -\boldsymbol{\nabla} V$$

To V I can always add an arbitrary constant, C, since $\boldsymbol{\nabla}(V + C) = \boldsymbol{\nabla} V$ since C is *constant*. This emphasises that potential is not uniquely defined for a given electric field.

However potential has many uses as a device and as a concept. In particular, because it's a scalar field, V is much easier to manipulate than $\mathbf{E}$. Recall how much easier it is to find the electric field due to an ideal dipole by calculating first the electric potential and differentiating it, rather than making a direct solution of Coulomb's law.

5.2 Multipole expansion of the electric scalar potential

You are aware from excercises this year and last that the *details* of a charge distribution cannot necessarily be discerned from a distance. For example, in spherical symmetry the electric potential due to a charge, Q, distributed about a ball is the same whether the charge is, say, only on the surface if the ball is a conductor; evenly distributed; or distributed in any other way, for example as a power law from the centre to the surface (see for example, Problems Class 2, C2.2). In fact from outside the ball one sees the potential due to a point charge of amount Q placed at the centre of the ball. The electric potential due to a dipole of strength $p = qd$, from a distance cannot be used to deduce either q or d uniquely. Any combination producing the same p (say if I double the charge and halve the distance) will have the same electric potential *seen from a distance*. Here, we find that unlike in the case of the ball with spherical symmetry for which the assertion is *exact*, in the case of the dipole, the result is only exact if we assume an ideal dipole. A related approximate statement is that any localised distribution of charge, from a distance, is indistiguishable from a point charge. If the total charge is zero, then to a first approximation the distribution looks like an ideal dipole when "viewed" from afar. This short sightedness appears as a drawback, but it can also be exploited. We saw in Problems Class 3, C3.2, that there are ideal assemblies of point charges, successively monopoles, dipoles, quadrupoles, octupoles, *etc.* and I gave you a nifty formula for generating these using successive derivatives of the reciprocal separation distance $1/R$.

We can formalise these statements into a theorem which asserts that any charge distribution which is confined within a finite volume Ω *looks from outside* like a superposition

of multipoles; and the relative strengths decrease as the field point is moved further away from the source. I now want to prove this.

A matter of notation: up to now I have been happy to deviate from Griffiths and others and use $\mathbf{r}_0$ for a field point and $\mathbf{r}_1$ for a source point and either $\mathbf{r}$ or $\mathbf{R}$ for the separation vector $\mathbf{r}_0 - \mathbf{r}_1$. *In this section*, 5, I will revert to Griffiths's notation and use $\mathbf{r}$ for the field point and $\mathbf{r}'$ for the source point(s). $\mathrm{d}\tau_1$ becomes $\mathrm{d}\tau'$. I will not use Griffiths's curly 'r' for the separation vector, I will use $\mathbf{R}$.†

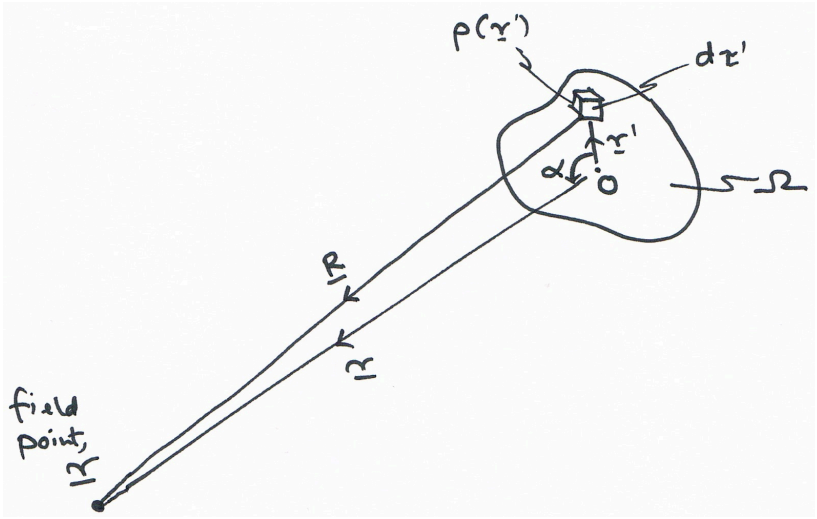


FIGURE 5–1

Figure 5–1 shows an arbitary distribution of charge, localised within a volume, Ω. We want the electric potential at a *distant* field point, $\mathbf{r}$,

$$V(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\int_\Omega \frac{\rho(\mathbf{r}')}{R}\mathrm{d}\tau' \tag{5.2.1}$$

It makes sense to place the origin, O, of the cartesian coordinate system inside Ω so that, firstly, $r' \ll r$ or more strictly $d/r \ll 1$ where d represents a typical dimension of the volume, say $d \approx \Omega^{1/3}$. Then we also have $R \approx r$. In view of this and since the potential is a function of $1/R$, this suggests that we expand $1/R$ in powers of $1/r$. Let us put the vector $\mathbf{r}' = (r'_1, r'_2, r'_3)$ into its components,

$$\mathbf{r}' = r'_i\hat{\mathbf{e}}_i$$

Then the Taylor expansion of $1/R$ about $1/r$ is, using the Einstein summation convention

† You may well ask why didn't I adopt this from the outset! Maybe if I'd written these notes first I would have done. However I don't like to reserve the undecorated $\mathbf{r}$ for a particular vector, the field point, as I like to be free to use it as a general vector. Also I can't typeset Griffiths's curly 'r' and I can't write it on the blackboard. Many textbooks, for example the standard text by Jackson make no extra symbol for the separation vector, so whereas we can write, say, 1/R, he has to write $1/\left|\mathbf{r}-\mathbf{r}'\right|$.

(see Section 1, and the Appendix to this section),

$$\frac{1}{R} = \frac{1}{r} - r'_i \left(\frac{\partial}{\partial r_i} \frac{1}{r} \right)_{r'_i=0} + \frac{1}{2!} r'_i r'_j \left(\frac{\partial^2}{\partial r_i \partial r_j} \frac{1}{r} \right)_{r'_i=r'_j=0} + \dots \tag{5.2.2}$$

This is now substituted into (5.2.1),

$$\begin{aligned} V(\mathbf{r}) &= \frac{1}{4\pi\epsilon_0} \int_\Omega \frac{\rho(\mathbf{r}')}{R} \mathrm{d}\tau' \\ &= \frac{1}{4\pi\epsilon_0} \left(\frac{1}{r} \int \rho \, \mathrm{d}\tau' - \left(\frac{\partial}{\partial r_i} \frac{1}{r} \right) \int r'_i \rho \, \mathrm{d}\tau' + \frac{1}{2!} \left(\frac{\partial^2}{\partial r_i \partial r_j} \frac{1}{r} \right) \int r'_i r'_j \rho \, \mathrm{d}\tau' + \dots \right) \end{aligned}$$

Again, the derivatives are evaluated at $R = r$, the integrals are over the volume Ω that confines ρ and the Einstein summation convention is implied. This helps to expose this as a *multipole expansion* because the first moment of the charge distribution is just the total charge,

$$Q = \int \rho \, \mathrm{d}\tau'$$

The i-component of the dipole moment vector $\mathbf{p}$ is

$$p_i = \int r'_i \rho \, \mathrm{d}\tau'$$

and the ij element of the quadrupole moment tensor is

$$\mathcal{Q}_{ij} = \int r'_i r'_j \rho \, \mathrm{d}\tau'$$

You have already seen the coefficients to the multipole moments in Problems Class 3, equation (7). Indeed once the summations are made, the first three terms in the multipole expansion of the electric potential are (see the Appendix to this section),

$$V(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \left(\frac{Q}{r} + \frac{1}{r^2} \int (\cos\alpha) \, (r'\rho) \, \mathrm{d}\tau' + \frac{1}{r^3} \int \frac{1}{2} \left(3\cos^2\alpha - 1 \right) \left(r'^2 \rho \right) \, \mathrm{d}\tau' \dots \right)$$

where α is the angle between $\mathbf{r}$ and $\mathbf{r}'$. (See also Griffiths, section 3.4.) You expect terms to be progressively smaller at large r (distant field point). These take the form that we already saw in the case of linear multipoles in Problems Class 3, namely, multipole moment times Legendre polynomial. We cannot take the coefficients outside the integral signs because the angle α depends on the variable of integration, r'.

Do the multipole moments depend on the choice of origin? This is, after all, arbitrary, although we chose it to be within the volume Ω so as to make the most of the Taylor expansion. The answer is, in general *yes*. But clearly Q, the total charge, is independent of origin. You can show that if $Q = 0$, the body is uncharged, then the dipole moment is independent of origin. In fact if the first n multipole moments are zero, you can show that the $(n+1)^{\text{th}}$ is independent of the choice of origin.

5.3 Magnetic vector potential

You natually ask whether there is a magnetic potential associated with the **B**-field. If there are no currents or electric fields, then Ampère's law is,

$$\boldsymbol{\nabla} \times \mathbf{B} = \mathbf{0} \qquad \longleftarrow \text{ no currents or electric fields}$$

and we know that $\boldsymbol{\nabla} \cdot \mathbf{B} = 0$ always, so in this case there is a scalar magnetic potential, because **B** is now irrotational (see Section 8.2).

But in general there is no scalar magnetic potential, but there is a *magnetic vector potential* (see section 1.3).

We know that for any vector field, $\mathbf{u}(\mathbf{r})$,

$$\boldsymbol{\nabla} \cdot (\boldsymbol{\nabla} \times \mathbf{u}) = 0$$

Also, because,

$$\boldsymbol{\nabla} \cdot \mathbf{B} = 0$$

then according to the Helmholtz theorem (Section 1.3) **B** has to be the curl of some vector field. We call this the magnetic vector potential, **A**. So always,

$$\mathbf{B} = \boldsymbol{\nabla} \times \mathbf{A}$$

However just as in the case of the electric potential this is not unambiguous. I can always add to **A** any vector function whose curl is zero, that is, any irrotational vector field; or equivalently I can add the gradient of any scalar field since we know that

$$\boldsymbol{\nabla} \times (\boldsymbol{\nabla} f) = \mathbf{0} \quad \forall f(\mathbf{r})$$

so if I replace **A** with $\mathbf{A} + \boldsymbol{\nabla} f$ then,

$$\mathbf{B} = \boldsymbol{\nabla} \times (\mathbf{A} + \boldsymbol{\nabla} f) = \boldsymbol{\nabla} \times \mathbf{A}$$

and I get the same magnetic field.

It's not usually a huge advantage to manipulate the electric potential by adding a constant; but choosing the gradient of some function to add to the magnetic vector potential, **A**, has advantages. It is called "choice of gauge." The fact that **A** is physically unchanged by adding a vector field $\boldsymbol{\nabla} f$ is called *gauge invariance*.

Suppose I write Ampère's law in terms of **A**,

$$\begin{aligned} \boldsymbol{\nabla} \times \mathbf{B} &= \mu_0 \mathbf{J} \\ &= \boldsymbol{\nabla} \times (\boldsymbol{\nabla} \times \mathbf{A}) \\ &= \boldsymbol{\nabla} (\boldsymbol{\nabla} \cdot \mathbf{A}) - \nabla^2 A \end{aligned} \tag{5.3.1}$$

using the third second derivative rule (section 1.2.4, equation (1.3)). We can always choose, if we wish, a gauge in which

$$\boldsymbol{\nabla}\cdot\mathbf{A}=0 \qquad \longleftarrow \text{Coulomb gauge: } \mathbf{A} \text{ is solenoidal}$$

How can we be sure that we can do that? Well, suppose our original vector potential, $\mathbf{A}_0$, say, did not have zero divergence. I can add to $\mathbf{A}_0$ the gradient of some scalar function $f(\mathbf{r})$ to make a new vector potential which will give the same magnetic field,

$$\mathbf{A}=\mathbf{A}_0+\boldsymbol{\nabla}f$$

from which

$$\boldsymbol{\nabla}\cdot\mathbf{A}=\boldsymbol{\nabla}\cdot\mathbf{A}_0+\nabla^2 f$$

This will be zero, as required, if a function $f(\mathbf{r})$ can be found such that,

$$\nabla^2 f=-\boldsymbol{\nabla}\cdot\mathbf{A}_0 \tag{5.3.2}$$

Now recall Poisson's equation in electrostatics,

$$\nabla^2 V=-\frac{\rho}{\epsilon_0} \tag{5.3.3}$$

whose solution is the electric potential due to a bounded charge density, $\rho(\mathbf{r}')$,

$$V(\mathbf{r})=\frac{1}{4\pi\epsilon_0}\int\frac{\rho}{R}\mathrm{d}\tau' \tag{5.3.4}$$

where $\mathbf{r}$ is a field point, $\mathbf{r}'$ is a source point and $\mathbf{R}=\mathbf{r}'-\mathbf{r}$ is a separation vector. By analogy with this, then it follows by comparison of (5.3.3) with (5.3.4) that (5.3.2) will be true if I choose my function $f(\mathbf{r})$ to be,

$$f(\mathbf{r})=\frac{1}{4\pi}\int\frac{\boldsymbol{\nabla}\cdot\mathbf{A}_0}{R}\mathrm{d}\tau'$$

You don't normally have to work out what this function is—it's enough that it exists so that you can apply the condition $\boldsymbol{\nabla}\cdot\mathbf{A}=0$ to the vector potential, in other words you may treat it as a *solenoidal* field. You are then working within the so called *Coulomb gauge*. In that case the first term on the right hand side of Ampère's law (5.3.1) is zero and Ampère's law in the Coulomb gauge reads,

$$\nabla^2\mathbf{A}=-\mu_0\mathbf{J}$$

(Note that this is the Laplace of a *vector* field, see section 1.2.2) Again I make the analogy with (5.3.3) and (5.3.4), and I am able immediately to write down the vector potential in the Coulomb gauge in terms of the current density,

$$\mathbf{A}(\mathbf{r})=\frac{\mu_0}{4\pi}\int\frac{\mathbf{J}(\mathbf{r}')}{R}\mathrm{d}\tau' \tag{5.3.5}$$

This is extremely useful as we will see in some Problems, because once we have the current density in a problem, we can work out the vector potential and hence the magnetic field that is produced by the current density. Note that the same techniques, well known in electrostatics, that apply to solving (5.3.4) can be used to solve a problem in magnetostatics using (5.3.5). It doesn't matter that we've made a choice of gauge to help us, because the resulting magnetic field is independent of the choice.

When we work with electromagnetic waves, or in quantum mechanics when we wish to include the magnetic vector potential in the Hamiltonian, it is more convenient, instead of $\boldsymbol{\nabla} \cdot \mathbf{A} = 0$, to use a vector potential whose divergence is,

$$\boldsymbol{\nabla} \cdot \mathbf{A} = -\mu_0 \epsilon_0 \frac{\partial V}{\partial t} \qquad \longleftarrow \text{Lorenz gauge}$$

This is called the *Lorenz gauge* (without a "t" as in Venezuela).

The point is that we can choose anything we like for $\boldsymbol{\nabla} \cdot \mathbf{A}$, not just zero: the definition $\mathbf{B} = \boldsymbol{\nabla} \times \mathbf{A}$ specifies the curl of $\mathbf{A}$ but not its divergence—we have the freedom to choose what we like for that, thanks to the Helmholtz theorem (section 1.3).

Let us put

$$\mathbf{B} = \boldsymbol{\nabla} \times \mathbf{A} \tag{5.3.6}$$

into Faraday's law,

$$\boldsymbol{\nabla} \times \mathbf{E} = -\frac{\partial}{\partial t} (\boldsymbol{\nabla} \times \mathbf{A})$$

which is the same as,

$$\boldsymbol{\nabla} \times \left(\mathbf{E} + \frac{\partial \mathbf{A}}{\partial t} \right) = \mathbf{0} \tag{5.3.7}$$

We know that in *electrostatics*, in which $\partial(\text{anything})/\partial t = 0$, $\boldsymbol{\nabla} \times \mathbf{E} = \mathbf{0}$, so (5.3.7) is evidently its generalisation into *electrodynamics*. If the curl of a vector field is zero, then that field must be the gradient of a scalar field, so guided by the electrostatic case in which

$$\boldsymbol{\nabla} \times \mathbf{E} = \mathbf{0} \quad \text{leads to} \quad \mathbf{E} = -\boldsymbol{\nabla} V$$

we infer that,

$$\mathbf{E} + \frac{\partial \mathbf{A}}{\partial t} = -\boldsymbol{\nabla} V \tag{5.3.8}$$

Equations (5.3.6) and (5.3.8) serve to express the fields in terms of potentials.

We can now find the *electrodynamic* extension of Poisson's equation by substituting (5.3.8) in Gauss's law,

$$\nabla^2 V + \frac{\partial}{\partial t} (\boldsymbol{\nabla} \cdot \mathbf{A}) = -\frac{1}{\epsilon_0} \rho \tag{5.3.9}$$

while putting (5.3.6) into the Ampère–Maxwell equation yields us,

$$\boldsymbol{\nabla}\times(\boldsymbol{\nabla}\times\mathbf{A})=\mu_0\mathbf{J}-\mu_0\epsilon_0\boldsymbol{\nabla}\left(\frac{\partial V}{\partial t}\right)-\mu_0\epsilon_0\frac{\partial^2\mathbf{A}}{\partial t^2}$$

which, using $\boldsymbol{\nabla}\times\boldsymbol{\nabla}\times\mathbf{A}=\boldsymbol{\nabla}(\boldsymbol{\nabla}\cdot\mathbf{A})-\nabla^2\mathbf{A}$, can be rearranged into this,

$$\left(\nabla^2\mathbf{A}-\mu_0\epsilon_0\frac{\partial^2\mathbf{A}}{\partial t^2}\right)-\boldsymbol{\nabla}\left(\boldsymbol{\nabla}\cdot\mathbf{A}+\mu_0\epsilon_0\boldsymbol{\nabla}\frac{\partial V}{\partial t}\right)=-\mu_0\mathbf{J}$$

This is where the Lorenz gauge comes in handy, since the second parenthesis vanishes. In that gauge, Maxwell's equations are combined into these two equations,

$$\begin{aligned}\nabla^2V-\frac{1}{c^2}\frac{\partial^2V}{\partial t^2}&=-\frac{1}{\epsilon_0}\rho\\ \nabla^2\mathbf{A}-\frac{1}{c^2}\frac{\partial^2\mathbf{A}}{\partial t^2}&=-\mu_0\mathbf{J}\end{aligned}\tag{5.3.10}$$

the first coming from inserting the Lorenz gauge into (5.3.9). In the absence of sources, when the right hand sides are zero, these are the electromagnetic wave equations in terms of potentials (see Problems Class 4). They come into their own in relativistic electrodynamics.

5.4 Digression into angular momentum and magnetic moment

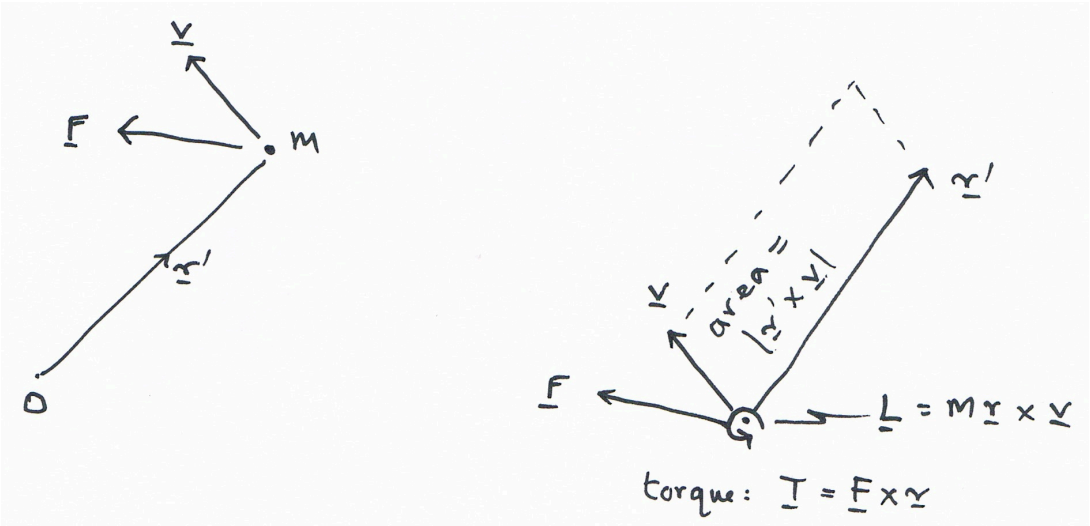


FIGURE 5–2

I wish to make a digression. According to Newton's first law, a mass remains in linear motion unless acted upon by a force. In this way linear momentum, $\mathbf{p}=m\mathbf{v}=m\dot{\mathbf{r}}$, is *conserved*. If we are to change the momentum, that is, to cause the mass to accelerate, then the necessary force, $\mathbf{F}$, is proportional to the amount by which we wish to change the momentum, $\mathbf{F}=m\ddot{\mathbf{r}}=m\dot{\mathbf{p}}$, which is Newton's second law—in words, "force is the

rate of change of momentum". In addition to linear momentum a mass may also possess angular momentum, which is also conserved. The equivalent to Newton's second law for non rectilinear motion is in words, "torque is the rate of change of angular momentum". Unlike the linear case, angular momentum and torque are only defined with respect to a choice of origin. Everything is fine as long as you choose the same origin for both. Figure 5–2 shows a point mass whose position with respect to a chosen origin is $\mathbf{r}'$. It is travelling at an instantaneous velocity $\mathbf{v}$ and is acted on by an instantaneous force $\mathbf{F}$. If I draw all three vectors from the origin then figure 5–2 shows the vector relations, $\mathbf{L} = m\mathbf{r}' \times \mathbf{v} = \mathbf{r}' \times \mathbf{p}$ and $\mathbf{T} = \mathbf{F} \times \mathbf{r}'$. In a simple case such as a point mass moving in a circle about another stationary mass, then the obvious choice of origin is the centre of the orbit, the force acts inwards between the two masses (gravitation) the velocity is tangent to the orbit and the angular momentum is constant (figure 5–3). It acts as an axial vector in a direction perpendicular to the plane of the orbit and has magnitude $L = mr'v$. I can extend this to the case of some distribution of mass that is in motion. This might be a body of fluid, say, that has a mass density ρ_m which is constant if the fluid is incompressible or if we are dealing with a solid. The velocity of an infinitesimal increment of the mass, $\mathrm{d}\tau'$, at position $\mathbf{r}'$ is $\mathbf{v}(\mathbf{r}')$. The total angular momentum of the body is their vector product integrated over the volume, Ω, which it occupies,

$$\mathbf{L} = \int_{\Omega} \rho_m \left(\mathbf{r}' \times \mathbf{v}\right) \mathrm{d}\tau' \tag{5.4.1}$$

which depends, as above, on the choice of origin from which to draw the vector $\mathbf{r}'$ to the mass source point at $\mathrm{d}\tau'$. In solid mechanics, this is often expressed simply as $\mathbf{L} = I\boldsymbol{\omega}$ which *defines* if you like the "moment of inertia", I, of the body about the origin; $\boldsymbol{\omega} = \dot{\theta}\hat{\mathbf{k}}$, say, is the angular velocity about the same origin if we choose the 3-direction to be along the axis of rotation.†

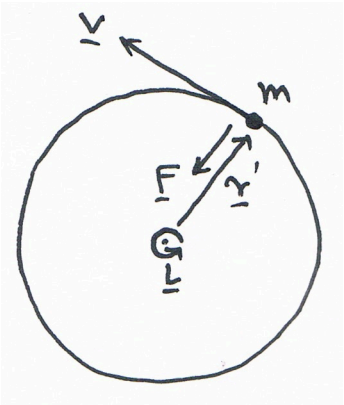


FIGURE 5–3

† Vectors such as $\mathbf{L}$, $\boldsymbol{\omega}$, and also magnetic field and magnetic moment are *axial vectors* or *pseudovectors*. They behave differently under reflection in a mirror compared to *polar vectors* such as velocity, linear momentum, electric field. Axial vectors are usually cross products and hence governed by a *convention*, namely the "right hand rule". Really they shouldn't be drawn as arrows but as directed circles about an axis.
See `https://en.wikipedia.org/wiki/Pseudovector`

Now suppose instead of a distribution of mass in motion I have a distribution, $\rho(\mathbf{r}')$, of charge in motion. I construct the (axial) vector,

$$\boldsymbol{\mu} = \frac{1}{2}\int_{\Omega} \rho\,(\mathbf{r}' \times \mathbf{v})\ \mathrm{d}\tau' \tag{5.4.2}$$

I will have that $\rho\mathbf{v} = \mathbf{J}(\mathbf{r}')$ which is a *current density*, so this becomes,

$$\boldsymbol{\mu} = \frac{1}{2}\int_{\Omega} (\mathbf{r}' \times \mathbf{J})\ \mathrm{d}\tau' \tag{5.4.3}$$

and this is called the *magnetic moment vector*. If the body possesses mass and charge distributions and these are comprised of particles of mass m and charge q and if the angular momentum and magnetic moments are parallel, then the ratio of the two is always,

$$\gamma_c = \frac{\mu}{L} = \frac{1}{2}\frac{q}{m} \tag{5.4.4}$$

irrespective of the complexity of the body since however tricky are the integrals in (5.4.1) and (5.4.2), they cancel (see, for an example, Section 8.6.2). The quantity γ_c is the classical *gyromagnetic ratio*. We observe that the gyromagnetic ratio belonging to a rotating charged body of uniform mass and charge density is independent of its shape and its angular velocity. Gyromagnetic ratio plays a big role in atomic physics and quantum mechanics, as you will find in other courses.

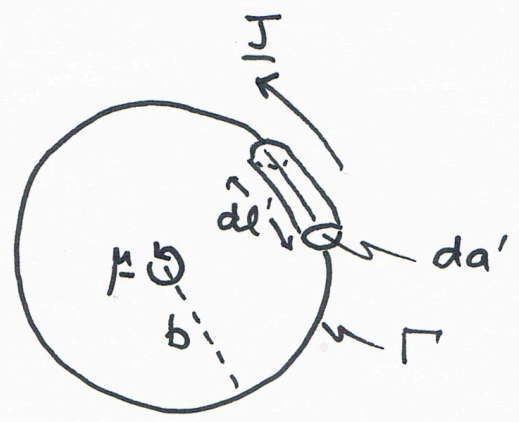


FIGURE 5–4

You will wonder what the "half" is doing in (5.4.2) and (5.4.3) since this seems to disturb the cleanness of (5.4.4). But consider a circular loop of current, I, having a radius b (figure 5–4). We may construct the current density by thinking of an increment of the loop having length $\mathrm{d}\ell'$ as a tube of cross sectional area $\mathrm{d}a'$, figure 5–4. Then the element of current $J\mathrm{d}\tau' = J\mathrm{d}a'\mathrm{d}\ell' = I\mathrm{d}\ell'$ since $J\mathrm{d}a' = I$, the current. We then evaluate the integral (5.4.3) which has become, since $\mathbf{J}$ is perpendicular to $\mathbf{r}'$, and $r' = b$,

$$\begin{aligned}\mu &= \frac{1}{2}\int_{\Gamma} r' J\,\mathrm{d}a'\,\mathrm{d}\ell' \\ &= \frac{1}{2} Ib \int_{\Gamma} \mathrm{d}\ell' \\ &= I\pi b^2\end{aligned}$$

This is "current times area" which you are familiar with as the magnetic moment about an origin at the centre of the circular loop (see question 6 in your revision problem sheet at KEATS). Note that in this section, I have to use $\boldsymbol{\mu}$ for magnetic moment, rather than $\mathbf{m}$ as elsewhere in the notes, so as not to confuse it with mass.

5.5 Multipole expansion of the magnetic vector potential

Compare (5.2.1) with (5.3.5). Since we expanded the electric potential in powers of $1/r$ this suggests we try the same with the magnetic vector potential. We start with the expansion,

$$\int \frac{\mathbf{J}}{R}\mathrm{d}\tau' = \frac{1}{r}\int \mathbf{J}\,\mathrm{d}\tau' - \left(\frac{\partial}{\partial r_i}\frac{1}{r}\right)\int r_i'\,\mathbf{J}\,\mathrm{d}\tau' + \frac{1}{2}\left(\frac{\partial^2}{\partial r_i \partial r_j}\frac{1}{r}\right)\int r_i' r_j'\,\mathbf{J}\,\mathrm{d}\tau' + \ldots$$

Integrals are over the volume Ω that confines the current density and the derivatives are evaluated at $\mathbf{r}' = \mathbf{0}$. You may be able to see without proof that the first term is zero. This is because we are dealing with magnetostatics so that the currents are not varying in time—so called *stationary currents*. Therefore the continuity equation is (see section 4, page 1),

$$\boldsymbol{\nabla}' \cdot \mathbf{J} = 0 \tag{5.5.1}$$

No current is entering or leaving the volume Ω and since current cannot end abruptly within Ω the current must be constituted of one or more loops. We then must have that,†

$$\int \mathbf{J}\,\mathrm{d}\tau' = \sum_{\text{loops}} J \oint \mathrm{d}\boldsymbol{\ell} = 0$$

A key point in making the expansion of the vector potential due to a bounded distribution of current is that we are content to obtain only the lowest order term in the expansion. The reason for permitting this is that we are to use the expansion for the study of magnetism in matter, and here the currents arise from atomic scale orbits of

† If you want a mathematical proof, then, starting with (5.5.1),

$$\begin{aligned} 0 = \int \boldsymbol{\nabla}' \cdot \mathbf{J}\, r_i'\,\mathrm{d}\tau' &= \int \boldsymbol{\nabla}' \cdot (r_i'\,\mathbf{J})\,\mathrm{d}\tau' - \int (\mathbf{J}\cdot\boldsymbol{\nabla}' r_i')\,\mathrm{d}\tau' \\ &= \int_S (r_i'\,\mathbf{J})\cdot \mathrm{d}S - \int J_i \frac{\partial r_i'}{\partial r_j'}\,\mathrm{d}\tau' \\ &= -\int J_i\,\mathrm{d}\tau' \end{aligned}$$

In the second line we used the divergence theorem and sending the boundary to infinity, got zero. We also used

$$\frac{\partial r_i'}{\partial r_j'} = \delta_{ij}$$

electrons and the size of these are at least ten orders of magnitude smaller than the distance from these current sources to the field point. So we take only the second term in the expansion having shown that the first term is zero. Focus therefore on the expression,

$$-\left(\frac{\partial}{\partial r_i}\frac{1}{r}\right)\int r'_i\,J_j\,\mathrm{d}\tau' = \frac{r_i}{r^3}\int r'_i\,J_j\,\mathrm{d}\tau' \tag{5.5.2}$$

which is the j-component of a vector, and in which we have done the differentiation, see Appendix (A2). The first thing to do is to add and subtract $r'_j\,J_i$ to $r'_i\,J_j$,

$$r'_i\,J_j = \frac{1}{2}\left(r'_i\,J_j + r'_j\,J_i\right) + \frac{1}{2}\left(r'_i\,J_j - r'_j\,J_i\right) \tag{5.5.3}$$

This will appear under an integral sign in (5.5.2), so let's show that the first term will vanish when integrated over the volume Ω. In view of (5.5.1),

$$r'_i\,r'_j\,\boldsymbol{\nabla}'\cdot\mathbf{J} = 0 = \boldsymbol{\nabla}'\cdot\left(r'_i\,r'_j\,\mathbf{J}\right) - \mathbf{J}\cdot\boldsymbol{\nabla}'\left(r'_i\,r'_j\right)$$

by the rule for differentiating a product. When the first term is integrated over Ω we will use the divergence theorem to convert this to an integral over a surface that we will take to infinity where the currents vanish. This leaves,

$$\begin{aligned}
0 &= \int \mathbf{J}\cdot\boldsymbol{\nabla}'\left(r'_i\,r'_j\right)\mathrm{d}\tau' \\
&= \int J_k\,\frac{\partial}{\partial r'_k}\left(r'_i\,r'_j\right)\mathrm{d}\tau' \\
&= \int J_k\left(r'_i\,\frac{\partial r'_j}{\partial r'_k} + r'_j\,\frac{\partial r'_i}{\partial r'_k}\right)\mathrm{d}\tau' \\
&= \int \left(J_j\,r'_i + J_i\,r'_j\right)\mathrm{d}\tau'
\end{aligned}$$

Here, k is a dummy index that is summed over using the Einstein convention and we have used,

$$\frac{\partial r'_i}{\partial r'_j} = \delta_{ij}$$

This means that using the substitution (5.5.3) in (5.5.2), we are left with,

$$\frac{1}{r^3}\int r_i\,r'_j\,J_j\,\mathrm{d}\tau' = \frac{1}{2}\frac{1}{r^3}\int r_i\left(r'_i\,J_j - r'_j\,J_i\right)\mathrm{d}\tau' \tag{5.5.4}$$

This is where the "half" turns up that appears in the definition (5.4.2). I have put r_i under the integral sign for clarity in what follows although it could be put before the integral sign since, as a *field point* coordinate, it is independent of the *source point* variables of integration, $\mathrm{d}\tau'$.

Now consider the vector,

$$(\mathbf{r}'\times\mathbf{J})\times\mathbf{r}$$

I will show that the integrand in (5.5.4) is the j-component of this vector. For the purposes of manipulation, define a vector $\mathbf{v} = \mathbf{r}' \times \mathbf{J}$ so that $(\mathbf{r}' \times \mathbf{J}) \times \mathbf{r} = \mathbf{v} \times \mathbf{r}$. Using the alternating tensor (see Section 1), the m-component of $\mathbf{v}$ is,

$$v_m = \epsilon_{mpq}\, r'_p\, J_q$$

and the j-component of $(\mathbf{r}' \times \mathbf{J}) \times \mathbf{r}$ is

$$\begin{aligned}
\left[(\mathbf{r}' \times \mathbf{J}) \times \mathbf{r}\right]_j = \left[\mathbf{v} \times \mathbf{r}\right]_j &= \epsilon_{jmi}\, v_m\, r_i \\
&= \epsilon_{jmi}\epsilon_{mpq}\, r_i\, r'_p\, J_q \\
&= \epsilon_{mij}\epsilon_{mpq}\, r_i\, r'_p\, J_q \\
&= (\delta_{ip}\delta_{jq} - \delta_{iq}\delta_{jp})\, r_i\, r'_p\, J_q \\
&= r_i\, r'_i\, J_j - r_i\, r'_j\, J_i \\
&= r_i \left(r'_i\, J_j - r'_j\, J_i\right)
\end{aligned}$$

This is the integrand in (5.5.4). In the third line I have interchanged two indices in the Levi-Civita symbol so there is no change of sign. In the fourth, I've used the nifty identity from Section 1.1.

To summarise, we have proved that to lowest order, the j-component of the vector potential in the Coulomb gauge, as due to the j-component of a current density is,

$$\begin{aligned}
A_j &= \frac{\mu_0}{4\pi} \int \frac{J_j}{R}\, \mathrm{d}\tau' \\
&\approx -\frac{\mu_0}{4\pi} \left(\frac{\partial}{\partial r_i}\frac{1}{r}\right) \int r'_i\, J_j\, \mathrm{d}\tau' \\
&= \frac{\mu_0}{4\pi}\, \frac{1}{r^3} \int r_i\, r'_i\, J_j\, \mathrm{d}\tau' \\
&= \frac{\mu_0}{4\pi}\, \frac{1}{2}\frac{1}{r^3} \int r_i \left(r'_i\, J_j - r'_j\, J_i\right) \mathrm{d}\tau' \\
&= \frac{\mu_0}{4\pi}\, \frac{1}{r^3} \int \frac{1}{2} \left[(\mathbf{r}' \times \mathbf{J}) \times \mathbf{r}\right]_j\, \mathrm{d}\tau'
\end{aligned}$$

This is true for all three components of the vectors $\mathbf{A}$ and $\mathbf{J}$, hence, finally,

$$\begin{aligned}
\mathbf{A} &= \frac{\mu_0}{4\pi}\, \frac{1}{r^3} \int \frac{1}{2}\, \mathbf{r}' \times \mathbf{J} \times \mathbf{r}\, \mathrm{d}\tau' \\
&= \frac{\mu_0}{4\pi}\, \frac{1}{r^3}\, \boldsymbol{\mu} \times \mathbf{r} \\
&= -\frac{\mu_0}{4\pi}\, \boldsymbol{\mu} \times \boldsymbol{\nabla}\frac{1}{r} \\
&= -\frac{\mu_0}{4\pi}\, \mathbf{m} \times \boldsymbol{\nabla}\frac{1}{r}
\end{aligned} \tag{5.5.5}$$

In the second line I used our definition of the magnetic moment (5.4.3) and in the third line our standard result from Problems Class 1. In the last line I've changed the notation

back to using **m** for magnetic moment that is used in the rest of the notes. It's been a long bout of maths, but worth it since we have discovered this pretty formula relating the vector potential at a field point **r** due to a magnetic moment vector at a source point $\mathbf{r}'$. The connecting vector doesn't appear because we are working to lowest order.

If the source of the vector field is made up of many current loops so that there is a distribution of magnetic moments (figure 5–5), we then rewrite (5.5.5) in terms of the *magnetisation*, **M**, the magnetic moment per unit volume, in analogy with *polarisation*, **P**, the dipole moment per unit volume. These are central quantities when we come to electrostatics and magnetostatics in matter. In this way, if a body is magnetised in a way that the magnetic moment, $\mathbf{M}(\mathbf{r}')$, per unit volume varies from place to place then the vector potential as seen from outside the magnet is,

$$\mathbf{A}(\mathbf{r}) = \frac{\mu_0}{4\pi} \int_\Omega \mathbf{M}(\mathbf{r}') \times \boldsymbol{\nabla}' \frac{1}{R} \, \mathrm{d}\tau' \tag{5.5.6}$$

This is the integral of (5.5.5) over the source magnetisation. I have changed the sign by switching the derivatives to be with respect to the source coordinates and replaced the field point vector with the connecting vector since to lowest order these are not distinguishable.

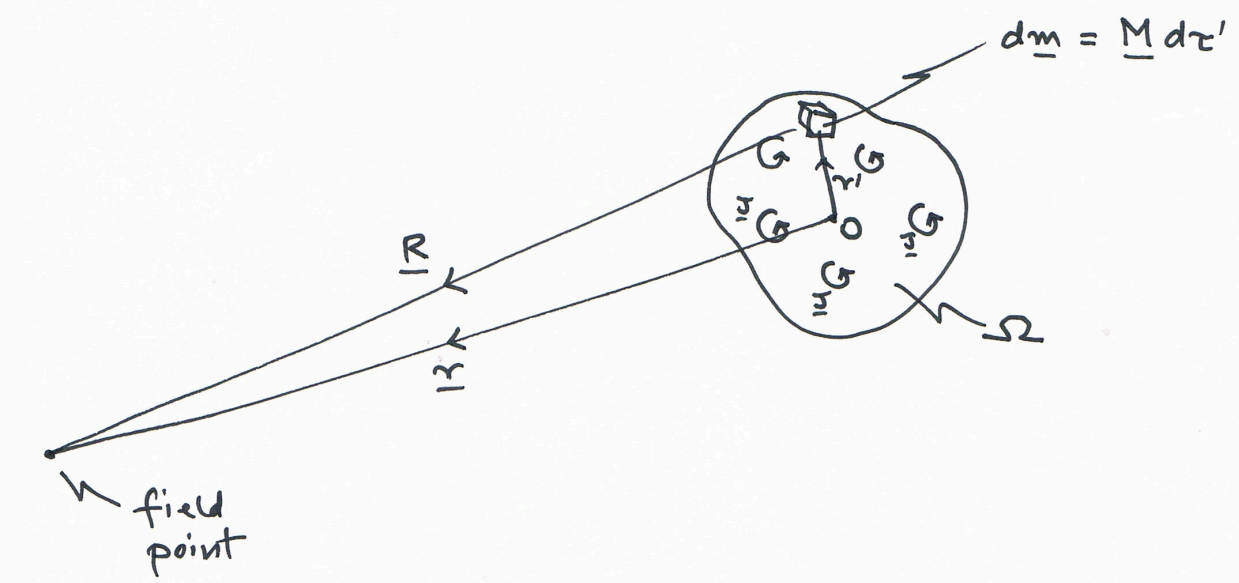


FIGURE 5–5

5.6 Exact result

I know that quite a few students find it very unsatisfactory to be presented with "approximate" results. You may even be disgusted that old hands seem to be quite content to deal with an approach that is not properly correct—that we are happy to be sloppy. You are paying over £9000 per year to be given the result (5.5.5) which is only correct *to lowest order*. For you, I will now give the exact result for the magnetic vector potential due to a circular loop of current. On the one hand it's only practically useful if you have a computer and are a careful programmer as it involves an integral that can't be done. On the other hand it is very general because any planar loop of current can be decomposed into a number of circular loops and the principle of superposition applies.

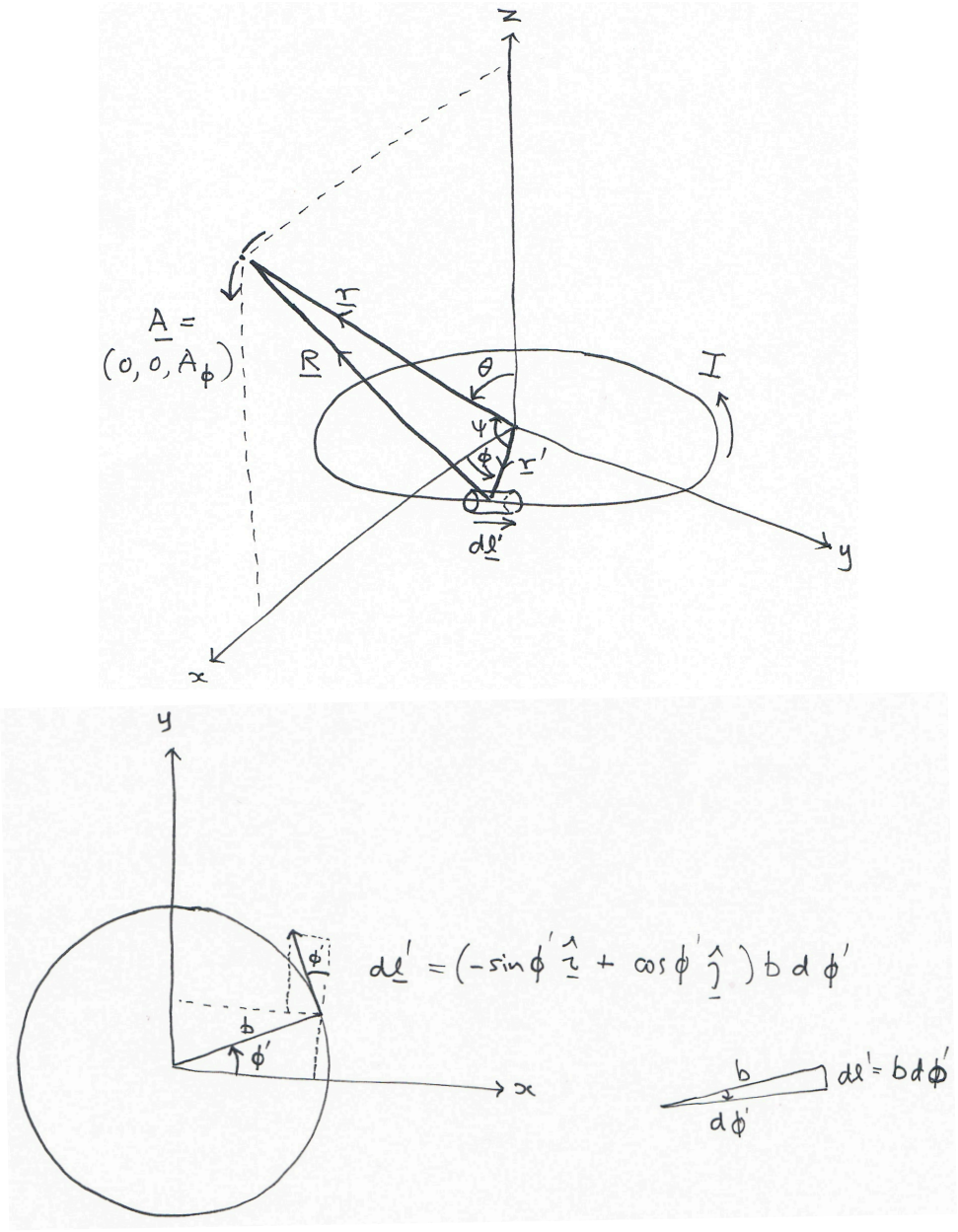


FIGURE 5–6

Figure 5–6 shows the geometry. We have a loop of current, I, of radius b and we place the origin of a cartesian system at the centre of the loop. A field point, $\mathbf{r}$, is placed at an arbitrary position, not necessarily "distant" from the origin. The problem has circular symmetry and it makes sense to use spherical coordinates. Because of the symmetry it doesn't matter where we put our x and y axes so we may as well put the x axis such that the angle ϕ' is zero when the element of current intersects the x-axis, as in figure 5–6. As we have already seen in figure 5–4, we can identify $\mathbf{J}\mathrm{d}\tau'$ with $I\mathrm{d}\boldsymbol{\ell}'$. θ is the polar angle of the *field point*; ϕ' is the azimuthal angle of the *source point*—hence the prime. Geometry dictates that, (see solutions to problem 1.4a),

$$\mathrm{d}\boldsymbol{\ell} = (-\sin\phi'\hat{\mathbf{i}} + \cos\phi'\hat{\mathbf{j}})\, b\,\mathrm{d}\phi'$$

By the cosine rule,

$$R^2 = b^2 + r^2 - 2br\cos\psi$$

where ψ is the angle between $\mathbf{r}$ and $\mathbf{r}'$. You can probably convince yourself that $\cos\psi = \cos\phi'\,\sin\theta$, since when $\phi' = 0$ then $\psi = 90^\circ - \theta$ and when $\phi' = 90^\circ$ then $\mathbf{r}$ and $\mathbf{r}'$ are perpendicular. This all adds up to the formula, starting from (5.3.5),

$$\begin{aligned}\mathbf{A}(\mathbf{r}) &= \frac{\mu_0}{4\pi}\, I \int \frac{\mathrm{d}\boldsymbol{\ell}}{R} \\ &= \frac{\mu_0}{4\pi}\, I \int \frac{(-\sin\phi'\hat{\mathbf{i}} + \cos\phi'\hat{\mathbf{j}})\, b\mathrm{d}\phi'}{\sqrt{b^2 + r^2 - 2br\cos\phi'\sin\theta}}\end{aligned}$$

As Lorrain and Corson remark, "A little thought will show that for any given value of r', we have two symmetrical $\mathrm{d}\boldsymbol{\ell}'$ vectors whose y-components add and whose x-components cancel. Then we need only calculate the y, or *azimuthal* component of $\mathbf{A}$." Hence

$$A_\phi(r,\theta) = \frac{\mu_0}{4\pi}\, Ib \int_0^{2\pi} \frac{\cos\phi'\mathrm{d}\phi'}{\sqrt{b^2 + r^2 - 2br\cos\phi'\sin\theta}} \tag{5.6.1}$$

We cannot do this integral, but we can treat some special cases and limits. For example, if $\theta = 0$ then the field point is on the axis of the circle. Interestingly the vector potential is zero along the z-axis. You can see this because the denominator is now independent of ϕ' and you are left with an integral of $\cos\phi'$ from zero to 2π which is zero. Looked at another way, this amounts to an integral of $\mathrm{d}\boldsymbol{\ell}$ over a circle and of course all the $\mathrm{d}\boldsymbol{\ell}$ add to zero if you start and end at the same point on the circle (or indeed any path). If my field point is close to the z-axis so that $\sin\theta$ is small, or indeed if alternatively $b \ll r$ so that the field point is distant, then in either case I can expand the square root and to lowest order,

$$\begin{aligned}A_\phi(r,\theta) &= \frac{\mu_0}{4\pi}\, Ib\, \frac{1}{\sqrt{b^2 + r^2}} \int_0^{2\pi} \left(\cos\phi' + \frac{br}{b^2 + r^2}\sin\theta\cos^2\phi'\right)\mathrm{d}\phi' \\ &= \frac{\mu_0 I\, b^2\, r\sin\theta}{4\,(b^2 + r^2)^{\frac{3}{2}}}\end{aligned}$$

because this integral *can* be done.

Now, of course, in this limit we can find the magnetic field, $\mathbf{B} = \boldsymbol{\nabla} \times \mathbf{A}$. It's easily done if we stay with spherical coordinates, because there is only the one component of $\mathbf{A}$ and the curl has only two components (see Griffiths, section 1.4.1),

$$B_r = \frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\, A_\phi\right) = \frac{\mu_0\, I\, b^2\, \cos\theta}{2\left(b^2+r^2\right)^{\frac{3}{2}}}$$
$$B_\theta = -\frac{1}{r}\frac{\partial}{\partial r}\left(r\, A_\phi\right) = -\frac{\mu_0\, I\, b^2\, \sin\theta}{4\left(b^2+r^2\right)^{\frac{5}{2}}}\left(2b^2 - r^2\right)$$

The first formula is the most interesting because if $\theta = 0$ then this is the z-component of the magnetic field down the centreline of a circular current loop, which you have already calculated using the law of Biot and Savart,

$$B_z = \frac{1}{2}\mu_0\, I \frac{b^2}{\left(b^2+z^2\right)^{\frac{3}{2}}}$$

which is an exact result, because we started with the limit of $\theta \to 0$ and then set $\theta = 0$ in the final step.

5.7 Magnetic dipole field

In the solutions to the Revision Problems at `KEATS` I derived the electric field due an ideal electric dipole, of strength $\mathbf{p}$,

$$\mathbf{E}_{\text{dip}}(\mathbf{r})\frac{1}{4\pi\epsilon_0}\,\frac{1}{r^3}\left(3\left(\mathbf{p}\cdot\hat{\mathbf{r}}\right)\hat{\mathbf{r}} - \mathbf{p}\right)$$

and without proof I stated the equivalent formula for the magnetic field due to a pure dipole of moment $\mathbf{m}$,

$$\mathbf{B}_{\text{dip}}(\mathbf{r}) = \frac{\mu_0}{4\pi}\,\frac{1}{r^3}\left(3\left(\mathbf{m}\cdot\hat{\mathbf{r}}\right)\hat{\mathbf{r}} - \mathbf{m}\right)$$

and I pointed out the similarity.

Now you can derive this second formula from (5.5.5) in the form,

$$\mathbf{A}(\mathbf{r}) = \frac{\mu_0}{4\pi}\,\frac{1}{r^3}\,\mathbf{m}\times\mathbf{r}$$

The magnetic field is,

$$\begin{aligned}\mathbf{B} &= \boldsymbol{\nabla}\times\mathbf{A}\\ &= \frac{\mu_0}{4\pi}\,\boldsymbol{\nabla}\times\left(\mathbf{m}\times\frac{\mathbf{r}}{r^3}\right)\\ &= \frac{\mu_0}{4\pi}\,\left(\left(\frac{\mathbf{r}}{r^3}\cdot\boldsymbol{\nabla}\right)\mathbf{m} - \left(\mathbf{m}\cdot\boldsymbol{\nabla}\right)\frac{\mathbf{r}}{r^3} + \mathbf{m}\left(\boldsymbol{\nabla}\cdot\frac{\mathbf{r}}{r^3}\right) - \frac{\mathbf{r}}{r^3}\left(\boldsymbol{\nabla}\cdot\mathbf{m}\right)\right)\end{aligned}$$

using the product rule (Section 1.2.3),

$$\boldsymbol{\nabla}\times\left(\mathbf{u}\times\mathbf{v}\right) = \left(\mathbf{v}\cdot\boldsymbol{\nabla}\right)\mathbf{u} - \left(\mathbf{u}\cdot\boldsymbol{\nabla}\right)\mathbf{v} + \mathbf{u}\left(\boldsymbol{\nabla}\cdot\mathbf{v}\right) - \mathbf{v}\left(\boldsymbol{\nabla}\cdot\mathbf{u}\right)$$

By combining the results of Problems Class 1.1(b) and (d) you find that the third term is zero except at the origin. Also $\mathbf{m}$ is constant and independent of $\mathbf{r}$ so its derivatives vanish. This just leaves us with,

$$\mathbf{B} = -\left(\mathbf{m}\cdot\boldsymbol{\nabla}\right)\frac{\mathbf{r}}{r^3}$$

Consider the following derivative,

$$m_x\frac{\partial}{\partial x}\frac{\mathbf{r}}{r^3} = \frac{m_x}{r^3}\hat{\mathbf{i}} - 3m_x x\frac{\mathbf{r}}{r^5}$$

and this must apply also to the y and z components of $\mathbf{m}$. This means that,

$$\left(\mathbf{m}\cdot\boldsymbol{\nabla}\right)\frac{\mathbf{r}}{r^3} = \frac{\mathbf{m}}{r^3} - \frac{3\left(\mathbf{m}\cdot\mathbf{r}\right)\mathbf{r}}{r^5}$$

Therefore,

$$\mathbf{B} = \frac{\mu_0}{4\pi}\,\frac{1}{r^3}\left(3\left(\mathbf{m}\cdot\hat{\mathbf{r}}\right)\hat{\mathbf{r}} - \mathbf{m}\right)$$

which completes our proof.

Appendix: Taylor expansion of $1/R$

The Taylor expansion of a function $f(x_1+\delta x_1, x_2+\delta x_2, x_3+\delta x_3)$ about (x_1,x_2,x_3) is

$$f(x_1+\delta x_1, x_2+\delta x_2, x_3+\delta x_3) = f(x_1,x_2,x_3) + \delta x_1\frac{\partial f}{\partial x_1} + \delta x_2\frac{\partial f}{\partial x_2} + \delta x_3\frac{\partial f}{\partial x_3} + \frac{1}{2!}\sum_{i=1}^{3}\delta x_i\delta x_j\frac{\partial^2 f}{\partial x_i\partial x_j} + \ldots$$

the derivatives being evaluated at $\delta x_i = 0$. In our case we have vectors which are functions of three variables, namely,

$$\mathbf{r} = r_i\hat{\mathbf{e}}_i \quad , \qquad \mathbf{r}' = r_i'\hat{\mathbf{e}}_i$$

and

$$\mathbf{R} = \mathbf{r} - \mathbf{r}'$$

We let

$$f(r_1,r_2,r_3) = \frac{1}{r} = \left(r_1^2 + r_2^2 + r_3^2\right)^{-\frac{1}{2}} \tag{A1}$$

Then

$$\frac{\partial f}{\partial r_i} = -\frac{1}{r^3}\,r_i \tag{A2}$$

and

$$\frac{\partial^2 f}{\partial x_i\partial x_j} = \frac{3}{r^5}\,r_i r_j - \frac{1}{r^3}\delta_{ij} \tag{A3}$$

which you can confirm by differentiating (A1) and noting that

$$\frac{\partial x_i}{\partial x_j} = \delta_{ij}$$

We want the expansion of

$$\frac{1}{R} = \frac{1}{|\mathbf{r} - \mathbf{r}'|}$$

about $1/r$, that is about the point $\mathbf{r}' = \mathbf{0}$, the origin in figure 5–1. This means we want, for $r_i = \{r_1, r_2, r_3\}$ and $r'_i = \{r'_1, r'_2, r'_3\}$,

$$f(r_i - r'_i) = f(r_i) + (-r'_i)\left.\frac{\partial f}{\partial r_i}\right|_{r'_i=0} + \frac{1}{2!}(-r_i)(-r_j)\left.\frac{\partial^2 f}{\partial r_i \partial r_j}\right|_{r'_i=r'_j=0} + \ldots$$

with implicit summations over i and j. For f given by (A1) this is,

$$\begin{aligned}\frac{1}{R} &= \frac{1}{r} - r'_i \frac{\partial}{\partial r_i}\frac{1}{r} + r'_i r'_j \frac{\partial^2}{\partial r_i \partial r_j}\frac{1}{r} + \ldots \\ &= \frac{1}{r} + \frac{1}{r^3} r'_i r_i + \frac{1}{2} r'_i r'_j \left(\frac{3 r_i r_j}{r^5} - \frac{1}{r^3}\delta_{ij}\right) + \ldots\end{aligned}$$

having put in the derivatives (A2) and (A3). Now, we use,

$$r'_i r'_j \, \delta_{ij} = r'_i r'_i = r'^2$$

and $r_i r_i = r^2$. Also if the angle between $\mathbf{r}$ and $\mathbf{r}'$ is α (figure 5–1), we have,

$$\cos\alpha = \frac{\mathbf{r}\cdot\mathbf{r}'}{r r'} = \frac{r_i r'_i}{r r'}$$

The expansion of $1/R$ becomes

$$\begin{aligned}\frac{1}{|\mathbf{r} - \mathbf{r}'|} &= \frac{1}{r} + \frac{1}{r^3} r r' \cos\alpha + \frac{1}{2}\left(\frac{3 r^2 r'^2 \cos^2\alpha}{r^5} - \frac{r'^2}{r^3}\right) + \ldots \\ &= \frac{1}{r} + \frac{1}{r^2}(\cos\alpha)\, r' + \frac{1}{2}\frac{1}{r^3}\left(3\cos^2\alpha - 1\right) r'^2 + \ldots \\ &= \frac{1}{r}\sum_{n=0}^{\infty}\left(\frac{r'}{r}\right)^n P_n(\cos\alpha)\end{aligned}$$

in which $P_n(x)$ are the Legendre polynomials, see equation (11), Problems Class 3. Also see Griffiths, section 3.4 for an alternative derivation.

5CCP2380 Problems Section 5
Electric point multipoles (after Panofsky and Phillips)

C5.1 For a point charge, $+q$ at a source point $\mathbf{r}'$ of a Cartesian coordinate system, the electric potential at a field point $\mathbf{r}$ is

$$V(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\frac{q}{R} = \frac{1}{4\pi\epsilon_0}\frac{q}{\sqrt{(x-x')^2+(y-y')^2+(z-z')^2}} \tag{1}$$

where, here, $\mathbf{R} = \mathbf{r} - \mathbf{r}'$ is a separation vector.

(**a**) Here is a rule for making a dipole. (*i*) Make a copy of the point charge and translate it an amount $\Delta\mathbf{r}'$ so it is at $\mathbf{r}' + \Delta\mathbf{r}'$. (*ii*) Leave the original charge where it is at $\mathbf{r}'$ but change its sign. As shown in figure 1, I have chosen to translate along the x-axis so,

$$\Delta\mathbf{r}' = \Delta x'\hat{\mathbf{i}}$$

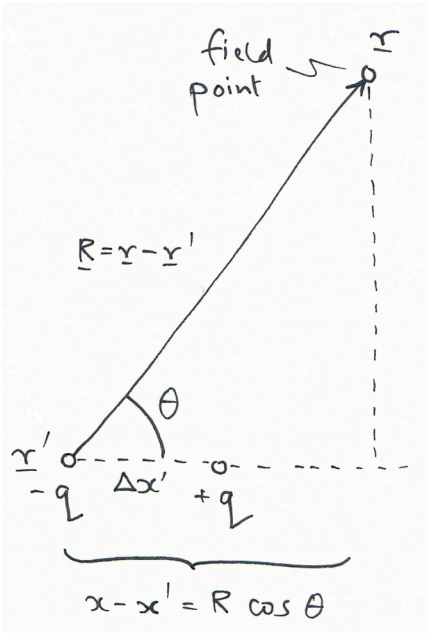


FIGURE 1

Recall what is meant by the derivative of a function $f(x)$ in differential calculus,

$$\frac{\mathrm{d}f}{\mathrm{d}x} = \lim_{\Delta x\to 0}\frac{f(x+\Delta x)-f(x)}{\Delta x}$$

Hence show that the electric potential of a dipole located at a field point $\mathbf{r}'$ oriented in the x-direction is

$$\begin{aligned} V^{\text{dipole}}(\mathbf{r}) &= \Delta x'\frac{\partial V}{\partial x'} \\ &= \frac{1}{4\pi\epsilon_0}\Delta x'\frac{\partial}{\partial x'}\frac{1}{R} && (2)\\ &= \frac{1}{4\pi\epsilon_0}\frac{q}{R^3}\Delta x'\,(x-x') \quad \text{by differentiating (1)} && (3)\\ &= \frac{1}{4\pi\epsilon_0}\,(q\Delta x')\,\frac{\cos\theta}{R^2} = \frac{1}{4\pi\epsilon_0}\frac{p\,\cos\theta}{R^2} && (4)\end{aligned}$$

where

$$p = q\,\Delta x'$$

is the dipole moment and θ is the angle between the x-axis and the vector $\mathbf{r}$. Check figure 1 to convince yourself of the correctness of the last line. This is, of course, the standard formula for the potential due to a dipole oriented along the x-axis.

For an arbitrary orientation of the dipole, I can generalise (2) as follows,

$$\begin{aligned} V^{\text{dipole}}(\mathbf{r}) = \Delta\mathbf{r}' \cdot \boldsymbol{\nabla}' V &= \frac{1}{4\pi\epsilon_0}\, q\,\Delta\mathbf{r}' \cdot \boldsymbol{\nabla}' \left(\frac{1}{R}\right) \\ &= \frac{1}{4\pi\epsilon_0}\, \mathbf{p} \cdot \boldsymbol{\nabla}' \left(\frac{1}{R}\right) \\ &= -\frac{1}{4\pi\epsilon_0}\, \mathbf{p} \cdot \boldsymbol{\nabla} \left(\frac{1}{R}\right) \qquad \text{by the result of C1.2} \\ &= \frac{1}{4\pi\epsilon_0}\, \frac{\mathbf{p}\cdot\mathbf{r}}{R^3} \qquad \text{by the result of C1.1b} \\ &= \frac{1}{4\pi\epsilon_0}\, \frac{\mathbf{p}\cdot\hat{\mathbf{r}}}{R^2} \end{aligned}$$

This is the well-known electric potential due to the dipole, but this nifty derivation is much cleaner and neater than the standard method using Taylor's expansion in, for example, Griffiths.

(**b**) Now we use the same rule to make an *electric quadrupole*. (*i*) Make a copy of the dipole and shift it along the y-axis by an amount $\Delta y'$. (*ii*) Leave the original dipole in place but the change the sign of both its charges. Show that the electric potential at $\mathbf{r}$ due to the quadrupole is,

$$V^{\text{quadrupole}}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\, 3q\,\Delta y'\,\Delta x' \cos\theta \sin\theta \frac{1}{R^3} \tag{5}$$

You will need to use the Taylor expansion in the Appendix to the Section 5 notes.

(**c**) Now create an octupole using the same rule. (*i*) Make a copy of the quadrupole and shift it along the z-axis by an amount $\Delta z'$. (*ii*) Leave the original quadrupole in place but the change the sign of all its charges. You now have a cuboid of point charges as illustrated in figure 2. Show that the octupole electric potential is,

$$V^{\text{octupole}}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\, 15q\,\Delta z'\,\Delta y'\,\Delta x' \cos\alpha \cos\beta \cos\gamma \frac{1}{R^4} \tag{6}$$

where α, β and γ are the angles as shown in figure 2.

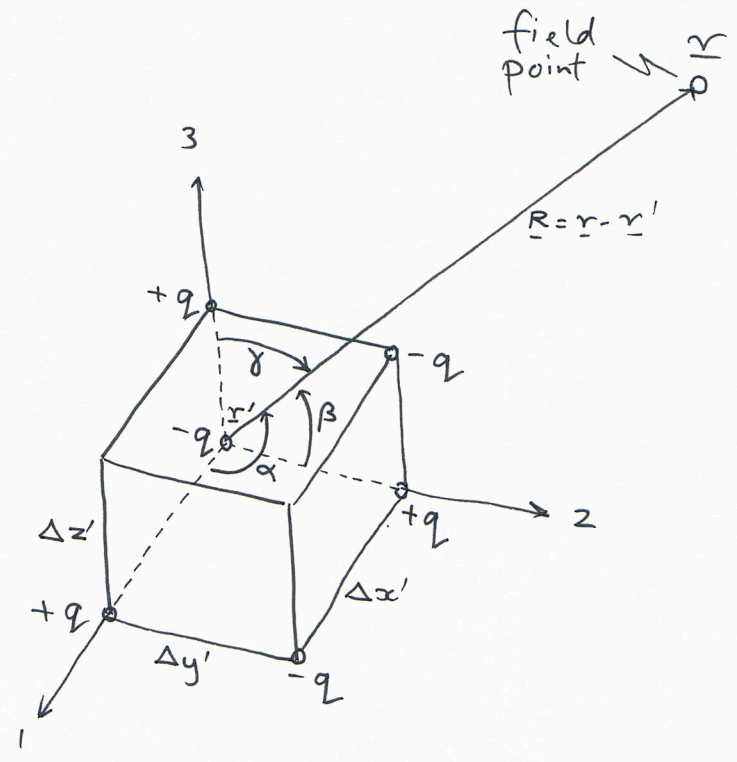


FIGURE 2

(**d**) Your results are consistent with the general formula for a 2^n-pole of point charges as the result of n successive derivatives of $1/R$[†],

$$V^{(2^n)}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\, p^{(2^n)}_{ij\ldots}\, \frac{1}{n!}\, \frac{\partial^n}{\partial r'_i \partial r'_j \ldots}\left(\frac{1}{R}\right) \tag{7}$$
$$= \frac{1}{4\pi\epsilon_0}\, p^{(2^n)}_{ij\ldots}\, \frac{(-1)^n}{n!}\, \frac{\partial^n}{\partial r_i \partial r_j \ldots}\left(\frac{1}{R}\right)$$

Here we are putting the source point vector $\mathbf{r}' = (r'_1, r'_2, r'_3)$ into its components,

$$\mathbf{r}' = r'_i \hat{\mathbf{e}}_i \qquad \text{also} \qquad \mathbf{r} = r_i \hat{\mathbf{e}}_i$$

using the summation convention (see Section 1). The second line follows from the fact that, in our notation, $\boldsymbol{\nabla}' = -\boldsymbol{\nabla}$ when applied to $1/R$, by the result of problem C1.2. Check this consistency in the case of the formulas (4), (5) and (6) above and then in each case identify what is the value of the multipole moment tensor components, $p^{(2)}_x$, $p^{(4)}_{xy}$ and $p^{(8)}_{xyz}$.

[†] There is no Einstein summation implied in this formula—it is taken to apply to any pre-chosen set of indices $\{i, j, \ldots\}$ taken from $\{x, y, z\}$; for example, we will set them all to x in equation (8) below

(e) **Linear multipoles**. We began with a point charge, $+q$, at the source point, $\mathbf{r}'$, and we made a copy, which we shifted by an amount $\Delta x'$ along the x-axis. Then we changed the sign of the original charge to $-q$. Thereby we created a dipole, which is *ideal* or *pure* as we let $\Delta x'$ tend to zero; which is why we could use our trick of differentiating as in (2) to find the electric potential at the field point $\mathbf{r}$. We then made a quadrupole by the same process of copy, shift and change sign. In this question we focus on *linear multipoles* in which all the shifts are along the x-axis.

1. Create a linear quadrupole by making a copy of the dipole, shifting the copy by $\Delta x'$, and changing the sign of each charge of the original dipole. You now have three charges in a row: $+q$, $-2q$, $+q$ since the middle charge is now two charges superposed. Continue this process and sketch the first few linear multipoles. Make an illustration of the rows listing the charges, one under the other to make a triangle. What do you notice? What is the sequence of charges of the linear 2^{12}-pole?

2. Show that in the case that all displacements are along the x-direction, equation (7) becomes,

$$\begin{aligned} V^{(2^n)}(\mathbf{r}) &= \frac{1}{4\pi\epsilon_0}\, p^{(2^n)}\, \frac{1}{n!}\, \frac{\partial^n}{\partial x'^n}\left(\frac{1}{R}\right) \\ &= \frac{1}{4\pi\epsilon_0}\, p^{(2^n)}\, \frac{(-1)^n}{n!}\, \frac{\partial^n}{\partial x^n}\left(\frac{1}{R}\right) \end{aligned} \tag{8}$$

3. Show by differentiation, using (8), that the electric potential at $\mathbf{r}$ due to the linear quadrupole is

$$V^{(2^2)}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\, q\,(\Delta x')^2\, \frac{1}{R^3}\left(3\cos^2\theta - 1\right) \tag{9}$$

where θ is the angle between $\mathbf{R}$ and the x-axis. What is the quadrupole moment?

4. Show by differentiation, using (8), that the electric potential at $\mathbf{r}$ due to the linear octupole is

$$V^{(2^3)}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\, q\,(\Delta x')^3\, \frac{1}{R^4}\left(15\cos^3\theta - 9\cos\theta\right) \tag{10}$$

What is the octupole moment?

5. Equation (8) contains the *generating function* for the Legendre polynomials, $P_n(\cos\theta)$,

$$P_n(\cos\theta) = (-1)^n\, \frac{1}{n!}\, R^{n+1} \frac{\partial^n}{\partial x^n}\left(\frac{1}{R}\right) \tag{11}$$

You recall that if we change the variable of differentiation from x' to x we pick up a minus sign for each such substitution—hence the factor $(-1)^n$. Work out the first few Legendre polynomials and show that your answers are consistent with (9) and (10) and your choice of definition of the multipole moment. To save doing a lot of differentiation, you may use Bonnet's recurrence formula (see Legendre Polynomials at `wikipedia`).

5CCP2380 Problems 5—Solutions

C5.1 (a) The electric potential at a field point $\mathbf{r}$, due a point charge q at a source point $\mathbf{r}'$ is

$$\begin{aligned} V^{\text{point}}(\mathbf{r}) &= \frac{1}{4\pi\epsilon_0}\,\frac{q}{R} = \frac{1}{4\pi\epsilon_0}\,q\,\left((x-x')^2+(y-y')^2+(z-z')^2\right)^{-\frac{1}{2}} \\ &= f(x') \end{aligned} \tag{S1}$$

Here, $\mathbf{R}=\mathbf{r}-\mathbf{r}'$ is the separation vector and I have *defined* the right hand side as a function f of x' only as I can treat x, y, y' z and z' as constant. This is because, at least in the first demonstration, I am interested in how the electric potential at a fixed field point varies as I move the source charge along the x-axis.

I then can form the first derivative,

$$\begin{aligned} \frac{\partial f}{\partial x'} &= \frac{1}{4\pi\epsilon_0}\,q\left(-\frac{1}{2}\right)\left(R^2\right)^{-\frac{3}{2}}\cdot 2\,(x-x')\cdot(-1) \\ &= \frac{1}{4\pi\epsilon_0}\,q\frac{1}{R^3}\,(x-x') \end{aligned}$$

Now consider a charge $+q$ placed a small distance $\Delta x'$ from a charge $-q$ along the x-axis (see FIGURE 1 in the problem sheet for Section 5).

The electric potential at the field point $\mathbf{r}$ is the sum of the potentials, V_-^{point} and V_+^{point} due to the two charges. So using your definition (S1) you can write,

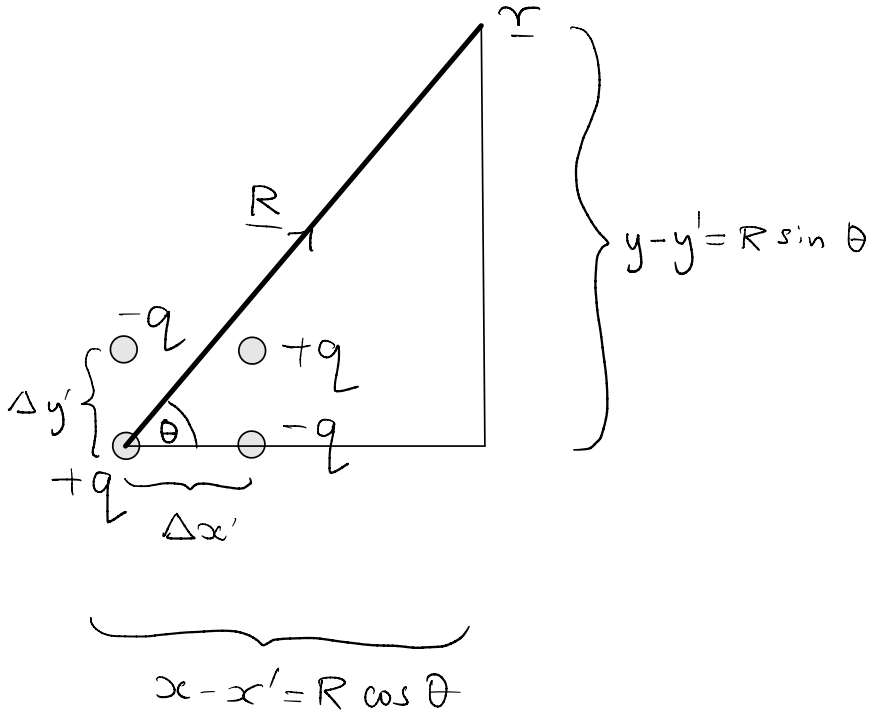
$$V_-^{\text{point}} = -f\,(x') \text{ and } V_+^{\text{point}} = f\,(x'+\Delta x')$$

and this suggests to you that we want to employ the definition of a first derivative, or alternatively that we are to make a Taylor expansion to first order in the small distance $\Delta x'$. Indeed to this order, the *dipole electric potential* is,

$$\begin{aligned} V^{\text{dip}} &= V_-^{\text{point}} + V_+^{\text{point}} \qquad \text{(S2)} \\ &= f\,(x'+\Delta x') - f\,(x') \\ &= \Delta x'\,\frac{\partial f}{\partial x'} \text{ to first order} \\ &= \frac{1}{4\pi\epsilon_0}\,q\Delta x'\frac{1}{R^3}\,(x-x') \qquad \text{(S3)} \\ &= \frac{1}{4\pi\epsilon_0}\,\frac{1}{R^2}p\cos\theta \end{aligned}$$

where we define the dipole moment to be $p=q\Delta x'$, and $x-x'=R\cos\theta$ as you see in FIGURE 1 in the problem sheet for Section 5. Of course you've done this problem already after using the cosine rule and expanding $1/r_+ - 1/r_-$ to first order. Students sometimes ask, if we've done a problem one way why do have to learn another way to do the same problem? Well, if often offers deeper insight; and in this case this second way of doing it opens up the next parts of the problem for which the first approach would be extremely cumbersome. The game we are playing is as explained in the question. We place a charge $+q$ at a source point $\mathbf{r}'$. We then make a copy and shift the copy a small distance along an axis of the cartesian system. Then we change the sign of the original point charge to $-q$. In that way, to begin with we have made a dipole.

(b) Now we do the quadrupole problem by placing two additional charges to form a rectangle, as shown here in the figure below. In fact we are pursuing the same game! You take your object (a dipole now), you make a copy and shift the copy a small distance—this time along the y-axis. The you change the signs of the charges in the original dipole. You have made a *quadrupole* as shown here.



Now the electric potential due to the quadrupole is the same as the sum of that due to the two dipoles it is constructed from. By analogy with (S2) we write,

$$\begin{aligned} V^{\text{quad}} &= V_+^{\text{dip}} + V_-^{\text{dip}} \\ &= f(x',y') - f(x'+\Delta x',y') - f(x',y'+\Delta y') + f(x'+\Delta x',y'+\Delta y') \\ &= \frac{1}{4\pi\epsilon_0}\, q\Delta y'\Delta x' \,\frac{\partial}{\partial y'}\,\frac{\partial}{\partial x'}\,\frac{1}{R} \qquad \text{to lowest order} \\ &= \frac{1}{4\pi\epsilon_0}\, q\Delta y'\Delta x' \,\frac{\partial}{\partial y'}\,\frac{1}{R^3}\,(x-x') \quad \text{using (S3)} \\ &= \frac{1}{4\pi\epsilon_0}\, q\Delta y'\Delta x' \;(x-x')\;\frac{\partial}{\partial y'}\frac{1}{R^3} \end{aligned} \tag{S4}$$

After the second line you will use the Taylor expansion as in the Appendix on page 17 of the Section 5 Notes. You will need these expansions,

$$\begin{aligned} f(x'+\Delta x',y') &= f(x',y') + \Delta x'\frac{\partial f}{\partial x'} + \frac{1}{2}\Delta x'^2\frac{\partial^2 f}{\partial x'^2} \\ f(x',y'+\Delta y') &= f(x',y') + \Delta y'\frac{\partial f}{\partial y'} + \frac{1}{2}\Delta y'^2\frac{\partial^2 f}{\partial y'^2} \\ f(x'+\Delta x',y'+\Delta y') &= f(x',y') + \Delta x'\frac{\partial f}{\partial x'} + \Delta y'\frac{\partial f}{\partial y'} \\ &\quad + \frac{1}{2}\Delta x'^2\frac{\partial^2 f}{\partial x'^2} + \frac{1}{2}\Delta y'^2\frac{\partial^2 f}{\partial y'^2} + \Delta y'\Delta x'\frac{\partial^2 f}{\partial y'\partial x'} \end{aligned}$$

when you add these up, you find the first order terms cancel and you are left with (S4) to lowest order.

Now,

$$\frac{\partial}{\partial y'}\left((x-x')^2+(y-y')^2+(z-z')^2\right)^{-\frac{3}{2}} = \left(-\frac{3}{2}\right)\left(R^2\right)^{-\frac{5}{2}}\cdot 2\,(y-y')\cdot(-1)$$
$$= 3\,(y-y')\,\frac{1}{R^5}$$

So,

$$V^{\text{quad}} = \frac{1}{4\pi\epsilon_0}\,3q\Delta y'\Delta x'\,(x-x')\,(y-y')\,\frac{1}{R^5}$$
$$= q\frac{1}{4\pi\epsilon_0}\,\Delta y'\Delta x'\,\frac{3\cos\theta\sin\theta}{R^3}$$

with $\cos\theta$ and $\sin\theta$ as shown in the figure, above.

(c) We continue our game... we take the quadrupole, make a copy, and shift the copy a small amount along the z-direction. Then we change the sign of the charges in the original quadrupole. The octupole is illustrated in FIGURE 2 in the Section 5 problems. (I omitted some of the charges by mistake). You hardly need to think—inspired by (S4) you write down,

$$\begin{aligned} V^{\text{oct}} &= V_-^{\text{quad}} + V_+^{\text{quad}} \\ &= \frac{1}{4\pi\epsilon_0}\,q\Delta y'\Delta x'\Delta z'\,\frac{\partial^3}{\partial x'\partial y'\partial z'}\,\frac{1}{R} \qquad \text{to lowest order} \\ &= \frac{1}{4\pi\epsilon_0}\,q\Delta y'\Delta x'\Delta z'\,\frac{\partial}{\partial z'}\,(x-x')\,(y-y')\,\frac{3}{R^5} \\ &= \frac{1}{4\pi\epsilon_0}\,q\Delta y'\Delta x'\Delta z'\,\,(x-x')\,(y-y')\,(z-z')\,\frac{15}{R^7} \\ &= \frac{1}{4\pi\epsilon_0}\,15q\,\Delta y'\Delta x'\Delta z'\,\frac{\cos\alpha\cos\beta\cos\gamma}{R^4} \end{aligned} \tag{S5}$$

where the angles α, β and γ are as shown in in FIGURE 2 in the problem sheet for Section 5. Note that these constructions are all made by taking successive derivatives of the reciprocal of the separation vector which you learned to do in the problems to Section 1.

(d) For the electric potential due to the dipole, equation (7) in the problem sheet is,

$$V^{\text{dip}} = V^{(2^1)} = \frac{1}{4\pi\epsilon_0}\,p_x^{(2^1)}\,\frac{1}{1!}\,\frac{\partial}{\partial x'}\,\frac{1}{R}$$

This means that by comparison with (S3) we identify the dipole moment as,

$$p_x^{(2^1)} = 1!\,q\Delta x' \tag{S6}$$

This is the "2^1-pole moment". The notation arises from the usage that a multipole moment is designated 2^ℓ. So $\ell = 0$ corrresponds to the monopole (point charge), $\ell = 1$ corresponds to the dipole, $\ell = 2$ to the quadrupole, $\ell = 3$ to the octupole and so on. We use a similar usage in atomic physics: the $\ell = 0$ orbital is an s-orbital which is spherically symmetrical, the $\ell = 1$ orbital is a p-orbital with two lobes of positive and negative wavefunction like the two charges in the dipole; $\ell = 2$ is the quantum number of the d-orbitals and so on.

In the same vein, for the quadrupole, equation (7) in the problem sheet reads,

$$V^{\text{quad}} = V^{(2^2)} = \frac{1}{4\pi\epsilon_0}\, p_{xy}^{(2^2)} \frac{1}{2!} \frac{\partial}{\partial x'} \frac{\partial}{\partial y'} \frac{1}{R}$$

so by comparison with (S4) it is clear that

$$p_{xy}^{(2^2)} = -2!\, q\Delta y' \Delta x' \tag{S7}$$

You know that the dipole moment $\mathbf{p}$ is a *vector* (or rank 1 tensor) and $p_x^{(2^1)}$ in (S6) is its x-component. Well, $p_{xy}^{(2^2)}$ in (S7) is the xy-element of the rank 2 quadrupole moment tensor. Of course the factorial signs in (S6) and (S7) are unnecessary, but they are there to follow the pattern as we come now to the octupole.

Equation (7) in the problem sheet is, in the case of the octupole,

$$V^{(2^3)} = \frac{1}{4\pi\epsilon_0}\, p_{xyz}^{(2^3)} \frac{1}{3!} \frac{\partial^3}{\partial x' \partial y' \partial z'} \frac{1}{R} = \frac{1}{4\pi\epsilon_0}\, p_{xyz}^{(2^3)} \frac{1}{3!} \frac{\partial}{\partial z'} \frac{\partial}{\partial y'} \frac{\partial}{\partial x'} \frac{1}{R}$$

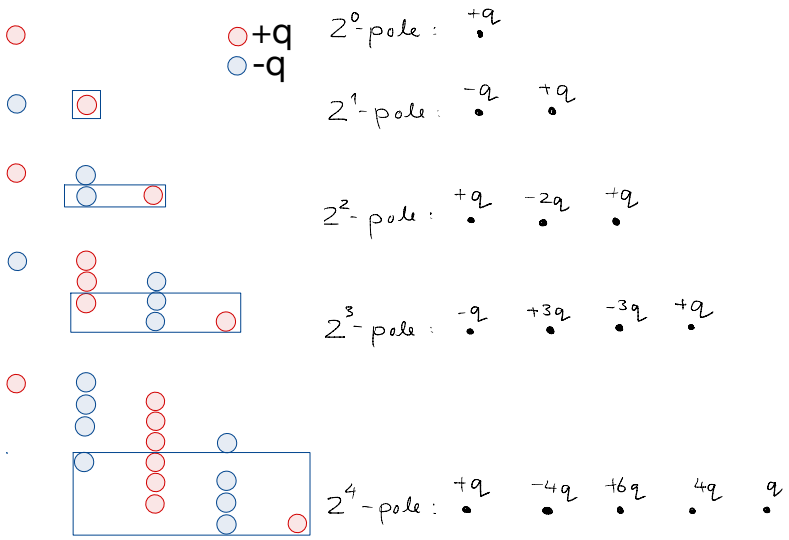
so by comparison with (S5) you can identify,

$$p_{xyz}^{(2^3)} = 3!\, q\Delta z' \Delta y' \Delta x'$$

and this is the xyz-element of the rank 3 *octupole moment tensor.*

You have now completed the game and built multipole moments up to $\ell = 3$. You can't really go any further in three dimensions, because if you make a copy of the octupole, change the sign of its charges and shift it along either x, y or z you've run out of dimensions! You can do it of course, but you will then have more than two charges placed along any one of the axes and so the trick of summing over the dipoles won't work. This thought takes us to the next game in which we can make *linear* multipoles of arbitrary ℓ by placing charges only on, say, the x-axis.

(e) To illustrate the new game, I have sketched below the first few placements of charge. For clarity I have displaced each copy (in a box) slightly below the x-axis but actually these are meant to represent linear arrays of point charges. If you add up the total charge in each column then this is the amount of point charge in each component of the multipole. Note that each blue box is a shifted copy of the previous multipole, while the charges not in the box are those of the previous multipole unshifted but with the sign of all its charges changed.



1. If you now write down this succession of point charges in the form of a triangle you find something rather surprising. Apart from the alternating signs, you have Pascal's triangle:

$$\begin{array}{lccccccccccc}
\ell = 0 & & & & & & +1 & & & & & \\
\ell = 1 & & & & & -1 & & +1 & & & & \\
\ell = 2 & & & & +1 & & -2 & & +1 & & & \\
\ell = 3 & & & -1 & & +3 & & -3 & & +1 & & \\
\ell = 4 & & +1 & & -4 & & +6 & & -4 & & +1 &
\end{array}$$

   This means it's easy to write down the sequence of charges in any linear multipole. For example, you are asked for the $\ell = 12$ linear 2^{12}-pole. The sequence is the set of binomial coefficients,

$$\binom{12}{0}, -\binom{12}{1}, \binom{12}{2}, -\binom{12}{3}, \binom{12}{4}, -\binom{12}{5}, \binom{12}{6}, -\binom{12}{7}, \binom{12}{8}, -\binom{12}{9}, \binom{12}{10}, -\binom{12}{11}, \binom{12}{12}$$

2. In equation (7) in the problem sheet, set all the components, i, $j \ldots$ to x. Then,

$$\begin{aligned} V^{(2^n)} &= \frac{1}{4\pi\epsilon_0} p^{(2^n)} \frac{1}{n!} \frac{\partial^n}{\partial x'^n} \frac{1}{R} \\ &= \frac{1}{4\pi\epsilon_0} p^{(2^n)} \frac{(-1)^n}{n!} \frac{\partial^n}{\partial x^n} \frac{1}{R} \end{aligned} \tag{S8}$$

   since each substitution of x' with x brings a change of sign.

3. Let's find successive derivatives of $1/R$.

   <u>First derivative</u>

$$\begin{aligned} \frac{\partial}{\partial x} \frac{1}{R} &= -(x - x') \frac{1}{R^3} \\ &= -\frac{\cos\theta}{R^2} \end{aligned}$$

Second derivative

$$\begin{aligned}\frac{\partial^2}{\partial x^2}\frac{1}{R} &= -\frac{\partial}{\partial x}\left(x-x'\right)\frac{1}{R^3}\\ &= -\left(x-x'\right)\frac{\partial}{\partial x}\frac{1}{R^3} - \frac{1}{R^3}\frac{\partial}{\partial x}\left(x-x'\right)\\ &= 3\left(x-x'\right)\left(x-x'\right)\frac{1}{R^5} - \frac{1}{R^3}\\ &= \frac{1}{R^3}\left(3\cos^2\theta - 1\right)\end{aligned}$$

Third derivative

$$\begin{aligned}\frac{\partial^3}{\partial x^3}\frac{1}{R} &= \frac{\partial}{\partial x}\left(3\left(x-x'\right)^2\frac{1}{R^5} - \frac{1}{R^3}\right)\\ &= 3\frac{\partial}{\partial x}\left(\left(x-x'\right)^2\frac{1}{R^5}\right) - \frac{\partial}{\partial x}\frac{1}{R^3}\\ &= 3\frac{1}{R^5}\frac{\partial}{\partial x}\left(x-x'\right)^2 + 3\left(x-x'\right)^2\frac{\partial}{\partial x}\frac{1}{R^5} + 3\left(x-x'\right)\frac{1}{R^5}\\ &= \frac{3}{R^5}\cdot 2\left(x-x'\right) + 3\left(x-x'\right)^2\cdot\left(-5\frac{x-x'}{R^7}\right) + 3\frac{x-x'}{R^5}\\ &= \frac{6\cos\theta}{R^4} - 15\frac{\cos^3\theta}{R^4} + \frac{3\cos\theta}{R^4}\\ &= \frac{1}{R^4}\left(9\cos\theta - 15\cos^3\theta\right)\end{aligned}$$

remembering that $x - x' = R\cos\theta$.

Now we can use (S8) and our derivatives of $1/R$ to get the electric potential due to the linear qudrupole,

$$\begin{aligned}V^{(2^2)} &= \frac{1}{4\pi\epsilon_0}p^{(2^2)}\frac{(-1)^2}{2!}\frac{1}{R^3}\left(3\cos^2\theta - 1\right)\\ &= \frac{1}{4\pi\epsilon_0}q\left(\Delta x'\right)^2\frac{1}{R^3}\left(3\cos^2\theta - 1\right)\end{aligned}$$

since, as you found earlier (S7), the xx-component of the quadrupole tensor is,

$$p_{xx}^{(2^2)} = 2!\,q\left(\Delta x'\right)^2$$

4. We can do the same for the octupole.

$$\begin{aligned}V^{(2^3)} &= \frac{1}{4\pi\epsilon_0}p^{(2^3)}\frac{(-1)^3}{3!}\frac{1}{R^4}\left(9\cos\theta - 15\cos^3\theta\right)\\ &= \frac{1}{4\pi\epsilon_0}p^{(2^3)}\frac{1}{3!}\frac{1}{R^4}\left(15\cos^3\theta - 9\cos\theta\right)\\ &= \frac{1}{4\pi\epsilon_0}q\left(\Delta x'\right)^3\frac{1}{R^4}\left(15\cos^3\theta - 9\cos\theta\right)\end{aligned}$$

again identifying the xxx-component of the octupole tensor,

$$p_{xxx}^{(2^3)} = 3!\, q\, (\Delta x')^3$$

5. I'll let you take it from here...

6. Auxiliary fields

6.1 Dimensional analysis (after Arnold Sommerfeld)

All of electrodynamics can be formulated as we have done thus far in terms of just two *fields*, **E** and **B**. This means that if we move to electromagnetism in matter, then *in principle* we can use Maxwell's equations as we have them. However due to the intricate detail in the atomic and sub-atomic structure of matter and the hugely varying electric fields as a field point moves between the atoms in matter, we are forced to make some *averaging*—for example in section 7 we will describe the macroscopic electric field. This needn't worry you. You are happy that water possesses a density. This is the mass per unit volume *averaged* over a large enough number of atoms to smooth out that huge variation in density from within an oxygen nucleus to that in the space between molecules, but over a small enough number of atoms that the density is uniform over a given region. In the sense of such an averaging, it is useful and sometimes essential to invent two further, auxiliary fields in electrodynamics.

The electric field is, by definition, the electric force experienced per unit test charge.

$$\mathbf{E} \doteq \frac{\text{force}}{\text{charge}} \doteq \frac{\text{newton}}{\text{coulomb}}$$

Here, I am principally making a dimensional argument so the "$\doteq$" sign for the purposes of this section implies an equality in dimension only. Maxwell introduced a field that is auxiliary to **E** which he called the "electric displacement", although it would have been better to call it "electric excitation." We are asked to regard the electric field as a realisation of "action at a distance" in the sense that a source charge at $\mathbf{r}_1$ sets up a field that produces the force on a test charge at the field point $\mathbf{r}_0$. An alternative interpretation, favoured by Maxwell, is that the source charge produces an excitation of its surrounding medium which leads to field lines of a field, **D**, whose divergence, similar to Gauss's law is

$$\boldsymbol{\nabla} \cdot \mathbf{D} = \rho_{\text{free}}$$

Clearly, since by analogy with the electric flux over a spherical surface of radius r,

$$4\pi r^2 D = Q_{\text{free}}$$

the dimensions of D are

$$\mathbf{D} \doteq \frac{\text{charge}}{\text{area}} \doteq \frac{\text{coulomb}}{\text{m}^2}$$

I put a subscript "free" on the charge to indicate this is a source charge that *I* have placed there and which *I* control; to distinguish it from charge that may be moved around, outside of my control, within the medium which is excited by the source charge. If that medium is a vacuum then there is no such inaccessible, so called, "bound charge" and the action on a distant test charge is equally well described in terms of **E** or **D**. We will see in section 7 that the effect of free charge on matter is to *polarise* the medium and **D** is used to express the superposition of the electric field (multiplied by ϵ_0) produced

by the accessible, free charge and the dipole moment per unit volume, $\mathbf{P}$, that arises from excitation of the atomic structure of matter. In this way,

$$\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P}$$

Note that dipole moment per unit volume has the same dimensions as $\mathbf{D}$:

$$\mathbf{P} \doteq \frac{\text{charge} \times \text{distance}}{\text{volume}} \doteq \frac{\text{coulomb}}{\text{m}^2}$$

Current density, $\mathbf{J}$, has dimensions

$$\mathbf{J} \doteq \frac{\text{charge}}{\text{area} \times \text{time}} \doteq \frac{\text{coulomb}}{\text{m}^2\text{s}}$$

which is the same as the units of $\partial\mathbf{D}/\partial t$, which quantity we will later call the "displacement current density".

In summary, $\mathbf{D}$ is the field whose source is the free charge, while $\mathbf{E}$ is the field whose source is the total (free plus bound charge). However these two fields have different dimensions, related by the constant ϵ_0.

Let us set up a similar argument in regard to $\mathbf{B}$. The magnetic field exerts a force on a magnetic pole having a certain strength, Q_m. Then the force per unit pole strength is the magnetic field,

$$\mathbf{B} \doteq \frac{\text{force}}{\text{pole strength}} \doteq \frac{\text{newton}}{Q_m}$$

Of course we know there are no poles, except at the ends of bar magnets—but there's always an opposite pole at the other end. In fact the elementary magnetic object is a dipole; and rather than using a bar magnet as analogy, remember that a loop of area, A, carrying a current, I, is the model for an ideal magnetic dipole having a magnetic moment,

$$\mu = IA$$

Since the strength of a dipole is

$$\mu = \text{pole strength} \times \text{pole separation}$$

we infer that the dimensions of pole strength are

$$Q_m \doteq \frac{\text{magnetic moment}}{\text{pole separation}} \doteq \frac{\text{current} \times \text{area}}{\text{length}} \doteq \text{charge} \times \frac{\text{length}}{\text{time}}$$

so

$$\mathbf{B} \doteq \frac{\text{newton}}{\text{coulomb}} \frac{\text{s}}{\text{m}}$$

As in electrostatics, a magnetic field may serve to excite, in this case to "magnetise", matter—that is to introduce a magnetic moment per unit volume, $\mathbf{M}$. Hence we introduce a magnetic excitation vector field, designated the symbol $\mathbf{H}$. By analogy with the electric excitation which has dimensions charge/area, the dimensions of $\mathbf{H}$ are

$$\mathbf{H} \doteq \frac{\text{pole strength}}{\text{area}} \doteq \frac{Q_m}{\text{m}^2} \doteq \frac{\text{coulomb}}{\text{ms}} \doteq \frac{\text{ampere}}{\text{m}}$$

You may recognise these units as "ampere turns per unit length" which describes the strength of a laboratory solenoid. When we adjust the current in a coil we are controlling the so called "free current density", $\mathbf{J}_{\text{free}}$. Magnetising the medium amounts to setting up bound currents that are inaccessible to us, just as are bound charges in matter. We will see in Section 8 that the bound current density is

$$\mathbf{J}_{\text{b}} = \boldsymbol{\nabla} \times \mathbf{M}$$

and the Ampère–Maxwell law in terms of the magnetic excitation is

$$\boldsymbol{\nabla} \times \mathbf{H} = \mathbf{J}_{\text{free}} + \frac{\partial \mathbf{D}}{\partial t} \tag{6.1.1}$$

The names given to the **B** and **H** fields are problematic. If you are not dealing with magnetism in matter, then there is no need to use **H** since $\mu_0\mathbf{H} = \mathbf{B}$. In the laboratory one deals with **H** because this is what can be controlled by varying the current in a solenoid; in fact $H = nI$ if a current I is passed through a solenoid of n turns per meter. If the solenoid is empty of matter then the magnetic field is just $B = \mu_0 nI$. But if you are dealing with magnetism in matter you may need to use both **B** and **H** fields. The problem is what to call them. Lorrain and Corson call **H** the "magnetic field intensity," and **B** the "magnetic induction." Let us rather follow Griffiths and the sublime Arnold Sommerfeld, and call **B** the magnetic field or magnetic field strength and when we need to use **H** we will call it the *magnetic excitation*, following Sommerfeld and by analogy with the *electric excitation*, **D**. Sommerfeld writes, "The unhappy term 'magnetic field' for **H**, ... has led into error none less than Maxwell himself, who puts the force exerted by the field on a magnetic pole, m, as equal to $m\mathbf{H}$." It is best to refer to **B** as "magnetic field" or "magnetic induction" and only use the words "**H**-field" without ever giving a name to **H**.

6.2 Comparisons of E and B and their auxiliary fields

We can separate out the total current density into accessible and inaccessible contributions. If we leave out convective currents arising from relative motion between the medium and an observer, then we can identify three sources of current: (*i*) free current, that is, current that we control in circuits, particle beams and so on; (*ii*) polarisation currents due to movements of bound charge within matter if the polarisation **P** is time dependent; (*iii*) bound magnetisation currents which are stationary currents, inaccessible to observation, that are due to quantum mechanical effects that lead to atomic scale magnetism. The sum of these amount to a *total* current density

$$\mathbf{J}_{\text{total}} = \mathbf{J}_{\text{free}} + \frac{\partial \mathbf{P}}{\partial t} + (\boldsymbol{\nabla} \times \mathbf{M})$$

As we saw in Maxwell's correction to Ampère's law (Section 4.2), Maxwell added to $\mathbf{J}_{\text{total}}$ the "vacuum displacement current density", $\epsilon_0 \partial\mathbf{E}/\partial t$. This gives us the complete

current density, $\mathbf{C}$, that appears in the Ampère–Maxwell law,

$$\begin{aligned}\boldsymbol{\nabla}\times\mathbf{B} &= \mu_0\mathbf{C}\\ &= \mu_0\mathbf{J}_{\text{total}} + \epsilon_0\mu_0\frac{\partial\mathbf{E}}{\partial t}\end{aligned}$$

It is worth noting that the divergence of the total current density is not necessarily zero and so $\mathbf{J}_{\text{total}}$ is not *solenoidal*. However $\boldsymbol{\nabla}\cdot\mathbf{C}=0$, and so the "complete" current density, the sum of the free, the bound and the displacement current densities,

$$\mathbf{C} = \mathbf{J}_{\text{free}} + (\boldsymbol{\nabla}\times\mathbf{M}) + \frac{\partial\mathbf{D}}{\partial t}$$

is solenoidal. By solenoidal we mean a field that has curl but no divergence—this is essential so that when I take $\boldsymbol{\nabla}\cdot\boldsymbol{\nabla}\times\mathbf{B} = \mu_0\boldsymbol{\nabla}\cdot\mathbf{C}$, I get zero as required by vector calculus.

So let me repeat: $\mathbf{D}$ is the electric "displacement" field whose source is the free charge density only, while $\mathbf{E}$ is the electric field whose source is the total (free plus bound) charge density. Similarly, $\mathbf{B}$ is the magnetic field whose source is the complete current density, $\mathbf{C}$, while $\mathbf{H}$ is the magnetic "excitation" field whose source (6.1.1) is the free current density, $\mathbf{J}_{\text{free}}$, plus the displacement current density which is

$$\frac{\partial\mathbf{D}}{\partial t} = \frac{\partial\mathbf{P}}{\partial t} + \epsilon_0\frac{\partial\mathbf{E}}{\partial t}$$

The auxiliary fields are, thereby,

$$\boxed{\mathbf{D} = \epsilon_0\mathbf{E} + \mathbf{P} \qquad ; \quad \mathbf{H} = \frac{1}{\mu_0}\mathbf{B} - \mathbf{M}} \tag{6.2.1}$$

These are not quite symmetric. (i) the constant equivalent to ϵ_0 has historically been assigned a reciprocal $\mu_0^{-1} = \epsilon_0 c^2$. (ii) In the first equation there is a plus rather than a minus sign because whereas bound current is $\mathbf{J}_{\text{b}} = +\boldsymbol{\nabla}\times\mathbf{M}$, bound charge density is $\rho_b = -\boldsymbol{\nabla}\cdot\mathbf{P}$. Take care! Because $\mathbf{J}_{\text{free}}$ is not solenoidal and $\mathbf{D}$ is not irrotational,

$$\boldsymbol{\nabla}\times\mathbf{D} \neq \mathbf{0} \qquad ; \quad \boldsymbol{\nabla}\cdot\mathbf{H} \neq 0$$

This also means that $\mathbf{D}$ is not the gradient of any scalar potential.

Sommerfeld points out another distinction between $\mathbf{E}$ and $\mathbf{D}$, and between $\mathbf{B}$ and $\mathbf{H}$. $\mathbf{E}$ and $\mathbf{B}$ are *intensive*, and $\mathbf{D}$ and $\mathbf{H}$ are *extensive*. $\mathbf{E}$ and $\mathbf{B}$ are the *causes* and $\mathbf{D}$ and $\mathbf{H}$ are the *effects*. In thermodynamics we talk of intensive and extensive properties; that is to say, answers to the question "how intense", as opposed to "how much". Good examples are the pairs of conjugate variables in thermodynamics, $-p$ and V, and T and S; or in mechanics, stress and strain. The combined first and second laws state that

the change in internal energy when the only work done is against a constant external pressure, at constant temperature is

$$\mathrm{d}U = T\mathrm{d}S - p\,\mathrm{d}V$$

Dividing through by the volume, this becomes the change in energy *density* and conventionally we use lower case symbols s and u and u for extensive quantities per unit volume,

$$\mathrm{d}u = T\mathrm{d}s - p\,\mathrm{d}v$$

If, in addition there is a uniform, constant $\mathbf{E}$ field influencing a dielectric substance with temperature-independent dielectric constant then the change in internal energy density is

$$\mathrm{d}u = T\mathrm{d}s - p\,\mathrm{d}v + \mathbf{E}\cdot\mathrm{d}\mathbf{D}$$

Actually the analogy can be misleading in the case of $\mathbf{B}$ and $\mathbf{H}$. I'll show you why and introduce at the same time, Poynting's theorem.

6.3 Poynting's theorem

We have these two Maxwell equations,†

$$\begin{aligned}\frac{\partial\mathbf{B}}{\partial t} &= -\boldsymbol{\nabla}\times\mathbf{E}\\ \frac{\partial\mathbf{D}}{\partial t} + \mathbf{J}_{\text{free}} &= \boldsymbol{\nabla}\times\mathbf{H}\end{aligned}$$

If I make a dot product of $\mathbf{H}$ with the first and a dot product of $\mathbf{E}$ with second and add the results, I get

$$\mathbf{H}\cdot\frac{\partial\mathbf{B}}{\partial t} + \mathbf{E}\cdot\frac{\partial\mathbf{D}}{\partial t} + \mathbf{E}\cdot\mathbf{J}_{\text{free}} = \mathbf{E}\cdot(\boldsymbol{\nabla}\times\mathbf{H}) - \mathbf{H}\cdot(\boldsymbol{\nabla}\times\mathbf{E})$$

I use the vector identity for two fields $\mathbf{u}$ and $\mathbf{v}$,

$$\mathbf{v}\cdot(\boldsymbol{\nabla}\times\mathbf{u}) - \mathbf{u}\cdot(\boldsymbol{\nabla}\times\mathbf{v}) = \boldsymbol{\nabla}\cdot(\mathbf{u}\times\mathbf{v})$$

I introduce the *Poynting vector*,

$$\mathbf{S} = \mathbf{E}\times\mathbf{H}$$

and I get *Poynting's theorem*

$$\mathbf{H}\cdot\frac{\partial\mathbf{B}}{\partial t} + \mathbf{E}\cdot\frac{\partial\mathbf{D}}{\partial t} + \mathbf{E}\cdot\mathbf{J}_{\text{free}} + \boldsymbol{\nabla}\cdot\mathbf{S} = 0 \tag{6.3.1}$$

The term

$$\mathcal{W}_{\text{joule}} = \mathbf{E}\cdot\mathbf{J}_{\text{free}} \qquad [\text{Joule s}^{-1}\ \text{m}^{-3}]$$

† Boltzmann asked, "Was it a god who wrote these lines…?"

is the rate of so called *Joule heating*; it is the dissipation of energy into heat (Watts per metre-cubed) as a result of the electric field acting on a volume density of moving charge.‡ We recall that the energy density contained within an electric field in vacuum is (I use lower case u as usual in thermodynamics to mean internal energy per unit volume—or energy density),

$$u_{\mathrm{e}} = \frac{1}{2}\epsilon_0 E^2 \qquad [\text{Joule m}^{-3}]$$

More generally (but not completely generally—see below) this should be

$$u_{\mathrm{e}} = \frac{1}{2}\mathbf{E}\cdot\mathbf{D}$$

because as we saw just now, the change in internal energy is $\mathbf{E}\cdot\mathrm{d}\mathbf{D}$. The "half" comes about just as when you calculate the energy stored in a stretched spring. Counterintuitively, Poynting's theorem requires us to assert that the energy stored in the magnetic field is to be obtained from

$$u_{\mathrm{mag}} = \int \mathbf{H}\cdot\frac{\partial\mathbf{B}}{\partial t}\,\mathrm{d}t = \int \mathbf{H}\cdot\mathrm{d}\mathbf{B}$$

The rate of change of electric and magnetic energy density is (see Appendix),

$$\frac{\partial u}{\partial t} = \mathbf{E}\cdot\frac{\partial\mathbf{D}}{\partial t} + \mathbf{H}\cdot\frac{\partial\mathbf{B}}{\partial t} \tag{6.3.2}$$

and in terms of this Poynting's theorem is

$$\boxed{\frac{\partial u}{\partial t} + \boldsymbol{\nabla}\cdot\mathbf{S} = -\mathcal{W}_{\mathrm{joule}}}$$

This expresses the *energy balance* of the electromagnetic field. The Joule heat is a dissipative loss, and the left hand side corresponds to energy transfer between different regions of space. You can see this more clearly by integrating over a given volume and using the divergence theorem,

$$\oint_{\mathcal{S}} \mathbf{S}\cdot\mathrm{d}\mathbf{a} = -\frac{\partial}{\partial t}\int_{\Omega} u\,\mathrm{d}\tau - \int_{\Omega}\mathcal{W}_{\mathrm{joule}}\,\mathrm{d}\tau$$

You see at once that this is a conservation law: the flux of energy leaving the surface, $\mathcal{S}$ that bounds a volume, Ω, is equal to the rate of loss of electromagnetic field energy, minus any heat dissipated by Joule heating. This last term, on the right hand side, is zero unless there are conductors present within Ω.

Using the dimensional analysis in subsection 6.1, you will soon see that the dimensions of the terms in Poynting's theorem (6.3.1) are

$$\frac{\text{newton}}{\text{m}^2\text{s}} = \frac{\text{joule}}{\text{m}^3\text{s}}$$

that is, energy per unit volume per unit time. This includes the term $\boldsymbol{\nabla}\cdot\mathbf{S}$ so the dimensions of the Poynting vector are energy per unit area per unit time; exactly what you expect of an energy flux: the *power* entering or leaving through the surface, $\mathcal{S}$, per unit area.

‡ $\mathbf{J}_{\mathrm{free}} = \sigma\mathbf{E}$ is Ohm's law, then $\mathbf{E}\cdot\mathbf{J}_{\mathrm{free}} = J^2_{\mathrm{free}}/\sigma$. This is I^2R per unit volume.

6.4 Electromagnetic energy

Equation (6.3.2) demonstrates the strange anomaly I mentioned at the end of subsection 6.2 in the magnetic energy. Whereas in the case of electrostatics an increment of internal energy is

$$\mathrm{d}U_\mathrm{e} = (\text{intensive quantity}) \times \mathrm{d}(\text{extensive quantity})$$

an increment of magnetic energy is the other way around

$$\mathrm{d}U_\mathrm{mag} = (\text{extensive quantity}) \times \mathrm{d}(\text{intensive quantity})$$

I am as puzzled as you by this. However, except in materials such as ferromagnets or non linear dielectrics, $\mathbf{D}$ is proportional to $\mathbf{E}$; or $\mathbf{H}$ is proportional to $\mathbf{B}$ and we can integrate (6.3.2) to give

$$U_\mathrm{e} = \frac{1}{2}\int_\Omega \mathbf{E}\cdot\mathbf{D}\,\mathrm{d}\tau \qquad \text{and} \qquad U_\mathrm{mag} = \frac{1}{2}\int_\Omega \mathbf{H}\cdot\mathbf{B}\,\mathrm{d}\tau$$

We can re-write these in terms of sources and potentials, rather than fields. You already know that[†]

$$U_\mathrm{e} = \frac{1}{2}\int_\Omega \rho_\mathrm{free}\, V\,\mathrm{d}\tau$$

where $\rho_\mathrm{free}(\mathbf{r})$ is the free charge density and $V(\mathbf{r})$ is the electric potential. I can first write

$$U_\mathrm{mag} = \frac{1}{2}\int_\Omega \mathbf{H}\cdot(\boldsymbol{\nabla}\times\mathbf{A})\,\mathrm{d}\tau$$

I use the vector identity $\boldsymbol{\nabla}\cdot(\mathbf{u}\times\mathbf{v}) = \mathbf{v}\cdot(\boldsymbol{\nabla}\times\mathbf{u}) - \mathbf{u}\cdot(\boldsymbol{\nabla}\times\mathbf{v})$ and $\boldsymbol{\nabla}\times\mathbf{H} = \mathbf{J}_\mathrm{free}$;[‡] I apply the divergence theorem and drop the surface term by taking the boundary of Ω to infinity where $\mathbf{H} = \mathbf{0}$, and I get

$$U_\mathrm{mag} = \frac{1}{2}\int_\Omega \mathbf{J}_\mathrm{free}\cdot\mathbf{A}\,\mathrm{d}\tau$$

now in terms of the source $\mathbf{J}_\mathrm{free}$ and vector potential $\mathbf{A}$.

---

† First write $\mathbf{E} = -\boldsymbol{\nabla}V$, transfer the $\boldsymbol{\nabla}$ to $\mathbf{D}$ at the expense of the a minus sign, throw away the surface term and use $\boldsymbol{\nabla}\cdot\mathbf{D} = \rho_\mathrm{free}$.

‡ So this is valid in the absence of vacuum displacement current—magnetostatics. See equation (6.1.1). But it will be fine in the case of slowly varying fields.

6.5 Maxwell's equations in terms of auxiliary fields

As you know, Maxwell's equations are,

$$\boldsymbol{\nabla}\cdot\mathbf{E} = \frac{\rho}{\epsilon_0} \tag{6.5.1a}$$

$$\boldsymbol{\nabla}\cdot\mathbf{B} = 0 \tag{6.5.1b}$$

$$\boldsymbol{\nabla}\times\mathbf{E} = -\frac{\partial\mathbf{B}}{\partial t} \tag{6.5.1c}$$

$$\boldsymbol{\nabla}\times\mathbf{B} = \mu_0\mathbf{J}_{\text{total}} + \mu_0\epsilon_0\frac{\partial\mathbf{E}}{\partial t} \tag{6.5.1d}$$

They are often written instead in terms of $\mathbf{E}$ and $\mathbf{B}$ *and* their auxiliary fields,

$$\boldsymbol{\nabla}\cdot\mathbf{D} = \rho_{\text{free}} \tag{6.5.2a}$$

$$\boldsymbol{\nabla}\cdot\mathbf{B} = 0 \tag{6.5.2b}$$

$$\boldsymbol{\nabla}\times\mathbf{E} = -\frac{\partial\mathbf{B}}{\partial t} \tag{6.5.2c}$$

$$\boldsymbol{\nabla}\times\mathbf{H} = \mathbf{J}_{\text{free}} + \frac{\partial\mathbf{D}}{\partial t} \tag{6.5.2d}$$

This second way of writing them is equivalent and identical to the first in the vacuum, in which $\mathbf{H} = \mathbf{B}/\mu_0$ and $\mathbf{D} = \epsilon_0\mathbf{E}$. There may be reasons for prefering the second set if you are working with the free charge and the free current density. This is especially significant if the medium is not vacuum, but some ponderable medium. In that case the $\mathbf{E}$ and $\mathbf{B}$ will cause excitations in the medium which are manifested as bound charges and bound currents. We cannot control these and so we would like equations that work directly with only those *free sources* of field that we control in the laboratory. In that case, however we need to know something of the properties of matter which comes from outside the purview of electrodynamics. We require a *constitutive relation* which is a generic term meaning an empirical connection between a cause and an effect. Examples of constitutive relations are Hooke's law, which states by how much a solid will stretch when it is pulled by a given force; or Ohm's law which says how much current will flow under a given potential difference. Constitutive laws turn up in diffusion, heat conduction, piezoelectricity, viscosity, thermoelectric effects, and so on. For our purpose we require constitutive equations that tell us how much "excitation" in a particular piece of matter is produced by a field. Remember Sommerfeld's remark at the last paragraph of page 4 in this Section: the intensive quantities, $\mathbf{E}$ and $\mathbf{B}$ are the *causes* of an amount of extensive quantity, $\mathbf{D}$ and $\mathbf{H}$. Conventionally we replace

$$\mathbf{D} = \epsilon_0\mathbf{E} \qquad \text{and} \qquad \mathbf{H} = \mu_0^{-1}\mathbf{B}$$

in the vacuum, with

$$\mathbf{D} = \epsilon\mathbf{E} \qquad \text{and} \qquad \mathbf{H} = \mu^{-1}\mathbf{B} \tag{6.5.3}$$

It's actually beautifully simple, as we just do all of electromagnetism with just the simple removal of a subscript "zero" on the constants. We call ϵ the *permittivity* and

μ the *permeability* of a particular substance. I will have a lot more say about these in Sections 7 and 8. But for now note the two most important observations.

1. Constitutive laws are *outside* the discipline of electrodynamics; they cannot be predicted using Maxwell's equations. They amount to *additional information* that are needed to apply Maxwell's equations in matter. They need to be obtained chiefly from experiment, but in some instances they may be predicted by theories such as quantum mechanics.
2. Equations (6.5.3) are linear equations: the response is in linear proportion to the cause. The *exact* equations are (6.2.1). Famous equations like Hooke's law and Ohm's law *are* linear equations, but really they are just the lowest order term in a Taylor expansion. There is no fundamental theory that says current does not also depend on the square, and higher powers, of the potential difference; or that strain does not depend on higher powers of the stress. And in fact they do—once voltage or stress gets large. Or the material is weird, like a semiconductor *pn*-junction, or a stretch of rubber. Similarly the electric polarisation is only linear in the electric field at "small" fields. The property of an optical material which has observable higher order terms in the polarisation as a function of electric field is the subject of the vast field of *non linear optics* which you will learn elsewhere. Everywhere in these notes we will stick to the linear response.

Appendix—work and energy in the electromagnetic field

Electromotance

A charge in an electric field experiences a Coulomb force and if that charge moves, then the field does work on the charge. It is easy to see that if moving charges amount to a current density, $\mathbf{J}$, then in the presence of an electric field, $\mathbf{E}$, the quantity $\mathbf{J}\cdot\mathbf{E}$ is the rate at which work is done per unit volume. In other words this is the change in potential energy of the current system per unit time, per unit volume. So in a volume Ω the rate of energy production is

$$\int_\Omega \mathbf{J}\cdot\mathbf{E}\,\mathrm{d}\tau = -\int_\Omega \mathbf{J}\cdot\boldsymbol{\nabla}V\,\mathrm{d}\tau$$

where I have used $\mathbf{E}=-\boldsymbol{\nabla}V$ to write the energy production in terms of the electric potential. I now prove that this is zero! By partial integration,

$$-\int_\Omega \mathbf{J}\cdot\boldsymbol{\nabla}V\,\mathrm{d}\tau = \int_\Omega (\boldsymbol{\nabla}\cdot\mathbf{J})\,V\,\mathrm{d}\tau + \int_S V\mathbf{J}\cdot\mathrm{d}\mathbf{a}$$

If we are dealing in steady currents, such as in electric circuits and particle beams, then $\boldsymbol{\nabla}\cdot\mathbf{J}=0$, and the right hand surface term vanishes since I can make my volume region extend to outside any current density.

What does this mean? It means that my assertion that $\mathbf{E}=-\boldsymbol{\nabla}V$ is wrong! The sort of electric field that is capable of producing electric current is *not* irrotational. In fact

you can see this if you look at Faraday's law. But I want to approach Faraday's law from this point.

Let us then divide the electric field into a part, $\mathbf{E}$, which is irrotational, or *conservative*, and which thereby may be obtained from an electric potential, and the so called electromotive field, $\mathbf{E}'$, which is responsible for driving currents—it is *this* field that is non zero in the wires of an electric circuit, even though you have been taught that a metal cannot sustain an electric field. In fact it can, and it does: these are the fields that exist throughout the wires of a circuit, that lurk in the bends and kinks, and which act to ensure that the wire remains neutral and the current is constant so that no charge accumulates.

Ohm's law now reads

$$\mathbf{J} = \sigma\left(\mathbf{E} + \mathbf{E}'\right)$$

and the *electromotance* (or emf) is defined by an integral around the circuit,

$$\mathcal{E} = \oint \left(\mathbf{E} + \mathbf{E}'\right) \cdot \mathrm{d}\boldsymbol{\ell} = \oint \mathbf{E}' \cdot \mathrm{d}\boldsymbol{\ell} = \frac{1}{\sigma} \oint \mathbf{J} \cdot \mathrm{d}\boldsymbol{\ell}$$

where the part involving the conservative field drops out since it integrates to zero around a closed path. If the circuit consists of wire of cross sectional area, A, length, L, and the current, I, is constant around the circuit, then we find the well known formula,

$$\mathcal{E} = I \oint \frac{\mathrm{d}\boldsymbol{\ell}}{\sigma A} = I \frac{L}{\sigma A} = IR$$

or $V = IR$, where R is the resistance.

Faraday's law of induction

Now Michael Faraday discovered in his laboratory, that if he connected a loop of wire of resistance, R, to a source of emf (battery at $\mathcal{E}$ volts) and an ammeter; and if he grew a magnetic field such that the magnetic flux through the wire changed, then a change in current, I, would be measured. The magnetic flux is the product of the magnetic field and the projected surface, $\mathcal{S}$, that bounds the loop of wire,

$$\Phi_m = \int_{\mathcal{S}} \mathbf{B} \cdot \mathrm{d}\mathbf{a}$$

What Faraday found is that the current in the circuit is not that expected from Ohm's law, $IR - \mathcal{E} = 0$; instead,

$$IR - \mathcal{E} = -\frac{\mathrm{d}\Phi_m}{\mathrm{d}t}$$

so the current differs what what is expected by minus the rate of change of magnetic flux. Faraday discovered in three versions of the experiment that the same result is obtained however the change in flux is achieved: by moving the loop in and out of a

fixed field; by moving the magnet; or by keeping the equipment stationary and changing the field by varying the current in a solenoid.

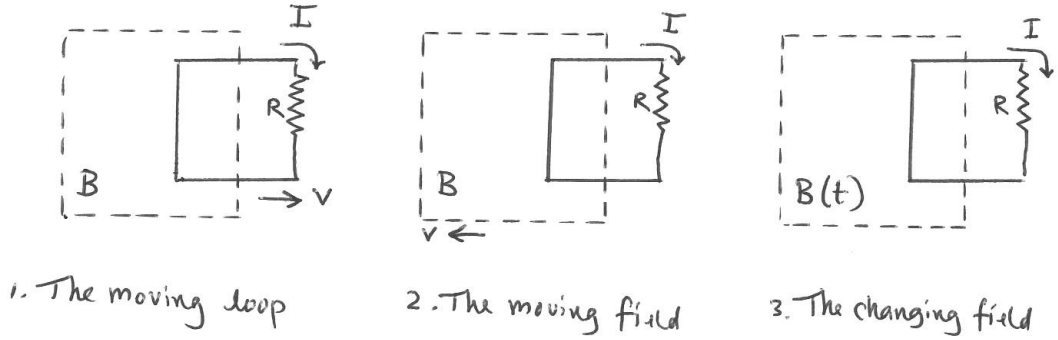


FIGURE 6–1: The three Faraday experiments. It is the observation that all three are consistent with the flux rule that led Albert Einstein to discover the special relativity.

James Clerk Maxwell realised that the Faraday experiments could be interpreted in a more general way. The additional current observed by Faraday has to be the result of the generation ("induction") of a non conservative electromotive field, $\mathbf{E}$, (we will drop the prime superscript) so that he wrote,

$$\oint \mathbf{E} \cdot \mathrm{d}\boldsymbol{\ell} = -\frac{\mathrm{d}\Phi_m}{\mathrm{d}t}$$

This leads to

$$\oint \mathbf{E} \cdot \mathrm{d}\boldsymbol{\ell} = -\int_{\mathcal{S}} \frac{\partial \mathbf{B}}{\partial t} \cdot \mathrm{d}\mathbf{a}$$

and using Stokes's theorem we get Faraday's law of induction (the third Maxwell equation),

$$\boldsymbol{\nabla} \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}$$

and since the curl of this electric field is not zero, it must be a non conservative field, as stated at the outset.

Magnetic energy

You are aware that potential energy is stored in the electric field, and we'll come back to derive this in the next section; but what about energy stored in the magnetic field?

Imagine an electric circuit in which a battery which produces an electromotive field, $\mathbf{E}'$, is producing Joule heating in a resistor and is also pumping energy into a magnetic field, say by increasing the current through the coils of a choke (inductor). We will have, as above,

$$\mathbf{J} = \sigma \left(\mathbf{E} + \mathbf{E}' \right)$$

and if I take the scalar product of both sides with $\mathbf{J}$, I get,

$$\mathbf{E}' \cdot \mathbf{J} = \frac{J^2}{\sigma} - \mathbf{E} \cdot \mathbf{J}$$

On the left hand side is the rate of doing work by the battery. On the right hand side is the Joule heating, and the remaining term, $-\mathbf{E} \cdot \mathbf{J}$, must be the rate of feeding energy into the magnetic field, since that is the only remaining process in our thought experiment. Let us allow the rate of change in the system to be sufficiently slow that we may neglect any displacement currents. Then I may use Ampère's law in the form, $\boldsymbol{\nabla} \times \mathbf{H} = \mathbf{J}$—since I am controlling all the current, it is free current. I make this substitution and integrate over a volume enclosing the system to find the total power delivered into the magnetic field; that is, the rate of doing work is,

$$\frac{\mathrm{d}U_{\mathrm{mag}}}{\mathrm{d}t} = -\int_\Omega \mathbf{E} \cdot (\boldsymbol{\nabla} \times \mathbf{H}) \, \mathrm{d}\tau$$

Now, by a vector identity,

$$\mathbf{E} \cdot (\boldsymbol{\nabla} \times \mathbf{H}) = \mathbf{H} \cdot (\boldsymbol{\nabla} \times \mathbf{E}) - \boldsymbol{\nabla} \cdot (\mathbf{E} \times \mathbf{H})$$

and if I integrate over the volume, apply Faraday's law to the first term on the right hand side, and the divergence theorem to the second, I find

$$-\int_\Omega \mathbf{E} \cdot (\boldsymbol{\nabla} \times \mathbf{H}) \, \mathrm{d}\tau = \int_\Omega \mathbf{H} \cdot \frac{\partial \mathbf{B}}{\partial t} \, \mathrm{d}\tau + \int_\mathcal{S} (\mathbf{E} \times \mathbf{H}) \cdot \mathrm{d}\mathbf{a}$$

I am permitted to drop the surface integral in the last term if I can take the volume of integration large enough so that the $\mathbf{E}$ and $\mathbf{H}$ fields have vanished at the boundary.

This formulation addresses only the case of steady, or slowly varying currents. I may not do this if the system is *radiating*, for example through dipole radiation. Given that we are not dealing with radiation so that the Poynting vector falls away at least like r^{-5}, we are left with,

$$\frac{\mathrm{d}U_{\mathrm{mag}}}{\mathrm{d}t} = \int_\Omega \mathbf{H} \cdot \frac{\partial \mathbf{B}}{\partial t} \, \mathrm{d}\tau$$

which is the result we are seeking, and which we also found in (6.3.2) on page 6 of this section.

Electric energy

To derive the result,

$$\mathrm{d}U_{\mathrm{e}} = \int_\Omega \mathbf{E} \cdot \mathrm{d}\mathbf{D} \, \mathrm{d}\tau$$

I imagine a thought experiment in which there exists an electric field, $\mathbf{E}$, and I move a small increment of free charge density, $\delta\rho$, from infinity into the field. The amount of work done is,

$$\delta W = \int_\Omega V \delta\rho \, \mathrm{d}\tau$$

where V is the electric potential at the position where I place the charge, and I have integrated over a volume enclosing the increment of charge density. I can use Gauss's law in terms of the electric excitation field, $\mathbf{D}$, and the free charge, to write,

$$\delta W = \int_\Omega V\,\delta\left(\boldsymbol{\nabla}\cdot\mathbf{D}\right)\,\mathrm{d}\tau = \int_\Omega V\,\boldsymbol{\nabla}\cdot\delta\mathbf{D}\,\mathrm{d}\tau$$

I can use a vector identity,

$$\boldsymbol{\nabla}\cdot(\mathbf{v}f) = f\,\boldsymbol{\nabla}\cdot\mathbf{v} + \mathbf{v}\cdot\boldsymbol{\nabla}f$$

to obtain,

$$\begin{aligned}\delta W &= \int_\Omega \boldsymbol{\nabla}\cdot(V\delta\mathbf{D})\,\mathrm{d}\tau - \int_\Omega \delta\mathbf{D}\cdot\boldsymbol{\nabla}V\,\mathrm{d}\tau \\ &= \int_\mathcal{S}(V\delta\mathbf{D})\cdot\mathrm{d}\mathbf{a} + \int_\Omega \delta\mathbf{D}\cdot\mathbf{E}\,\mathrm{d}\tau\end{aligned}$$

I will drop the surface term for the usual reasons, and I am left with,

$$\delta W = \int_\Omega \mathbf{E}\cdot\delta\mathbf{D}\,\mathrm{d}\tau$$

This leads me to the analogue of the magnetic case, above, for the rate of change of electrostatic energy, namely,

$$\frac{\mathrm{d}U_\mathrm{e}}{\mathrm{d}t} = \int_\Omega \mathbf{E}\cdot\frac{\partial\mathbf{D}}{\partial t}\,\mathrm{d}\tau$$

to compare with (6.3.2).

This states that as I shovel charge into a region, not necessarily vacuum, where there is a field, $\mathbf{E}$, I am essentially increasing the excitation, $\mathbf{D}$. So I would like to integrate from $\mathbf{D}=\mathbf{0}$ to $\mathbf{D}$ to obtain the stored energy in the electric field. However as I change $\mathbf{D}$, I am also changing $\mathbf{E}$, so to integrate I need the *constitutive relation*, $\mathbf{D}=\epsilon\mathbf{E}$, where ϵ is the permittivity of the medium. I can then obtain the energy,

$$\begin{aligned}U_\mathrm{e} = \int_0^D \delta W &= \frac{1}{2}\int_\Omega \epsilon E^2\,\mathrm{d}\tau \\ &= \frac{1}{2}\int_\Omega \mathbf{E}\cdot\mathbf{D}\,\mathrm{d}\tau\end{aligned}$$

5CCP2380 Problems Section 6

C6.1 Poynting's theorem is described in the notes to Section 6, Auxiliary Fields. Now work through the problem from the point of view of Griffiths. Note that he evokes the Lorentz force as a starting point. Identify the quantities used in either statement of the theorem and convince yourself that they amount to the same physics.

Griffiths imagines a volume, $\mathcal{V}$, in space containing some moving charges so that within $\mathcal{V}$ there is a current density, $\mathbf{J}(\mathbf{r})$. As these charges are sloshing around, he asks: in a time interval between t and $t + \mathrm{d}t$ what is the *work done* by the electromagnetic forces acting on those charges? The work done on each charge, q, as it moves along an element of path $\mathrm{d}\boldsymbol{\ell}$ at velocity $\mathbf{v}$ is

$$\begin{aligned}\mathbf{F} \cdot \mathrm{d}\boldsymbol{\ell} &= q\,(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \cdot \mathbf{v}\,\mathrm{d}t \\ &= q\,\mathbf{E} \cdot \mathbf{v}\,\mathrm{d}t\end{aligned}$$

Note the *magnetic force does no work* because $\mathbf{v}\times\mathbf{B}$ is perpendicular to $\mathbf{v}$ so $(\mathbf{v}\times\mathbf{B})\cdot\mathbf{v}$ is zero. For simplicity the charges are all of the same magnitude and sign, so I can replace q with $\rho\mathrm{d}\tau$ and then $\rho\mathbf{v} = \mathbf{J}$. Show that the rate of doing work (the power in Watts) *on* all the charges, *by* the electric field is

$$\frac{\mathrm{d}W}{\mathrm{d}t} = \int_{\mathcal{V}} \mathbf{E} \cdot \mathbf{J}\,\mathrm{d}\tau$$

Use the Ampére–Maxwell law

$$\boldsymbol{\nabla} \times \mathbf{B} = \mu_0 \mathbf{J} + \mu_0 \epsilon_0 \frac{\partial \mathbf{E}}{\partial t}$$

to find $\mathbf{E} \cdot \mathbf{J}$ in terms of $\mathbf{E}$ and $\mathbf{B}$. Then use Faraday's law to show that[†]

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

Hence show that

$$\mathbf{E} \cdot \mathbf{J} = -\frac{1}{2}\frac{\partial}{\partial t}\left(\epsilon_0 E^2 + \frac{1}{\mu_0} B^2\right) - \boldsymbol{\nabla} \cdot \mathbf{S}$$

where

$$\mathbf{S} = \frac{1}{\mu_0}\,(\mathbf{E} \times \mathbf{B})$$

[†] You will need the identity

$$\boldsymbol{\nabla} \cdot (\mathbf{u} \times \mathbf{v}) = \mathbf{v} \cdot (\boldsymbol{\nabla} \times \mathbf{u}) - \mathbf{u} \cdot (\boldsymbol{\nabla} \times \mathbf{v})$$

is Poynting's vector. Show that, integrating over the volume $\mathcal{V}$ and converting a volume integral to an integral over the bounding surface, $\mathcal{S}$, gives the continuity equation,

$$\frac{\mathrm{d}W}{\mathrm{d}t} = -\frac{\partial}{\partial t}\int_{\mathcal{V}} u_{\mathrm{vac}}\,\mathrm{d}\tau - \oint_{\mathcal{S}} \mathbf{S}\cdot\mathrm{d}\mathbf{a}$$

where

$$u_{\mathrm{vac}} = \frac{1}{2}\left(\epsilon_0 E^2 + \frac{1}{\mu_0}B^2\right)$$

is the energy stored in the electromagnetic field. Under what circumstances can we write this as

$$\frac{\partial u}{\partial t} = -\boldsymbol{\nabla}\cdot\mathbf{S}$$

and how do you interpret this equation?

C6.2 An infinitely long, tightly wound solenoid of radius a carries a current, I, and has N turns per metre.

(i) Find the vector potential $\mathbf{A}$ inside and outside the solenoid as a function of distance from the centre line. Give the magnitude and direction of $\mathbf{A}$.

(ii) The current is now increased at a uniform rate $\mathrm{d}I/\mathrm{d}t$. Find the Poynting vector $\mathbf{S} = \mu_0^{-1}\mathbf{E}\times\mathbf{B}$ over the outer surface of the coil. Sketch the solenoid end-on and indicate the directions of the current and $\mathbf{S}$.

C6.3 Consider a section of length L of a long cylindrical wire of resistivity, $\mathcal{R}$, and radius a that is connected to a battery that supplies an e.m.f. which drives a current density, $\mathbf{J}$, along the wire. You are interested in the energy flow across the surface of this segment of wire. Show that the magnitude of the Poynting vector at the suface is

$$S = \frac{1}{2}\mathcal{R}J^2 a$$

If Ω is the volume domain of the segment and $\mathcal{S}$ is its cylindrical surface, show that

$$\int_{\mathcal{S}} \mathbf{S}\cdot\mathrm{d}\mathbf{a} = -\int_{\Omega} \mathcal{R}J^2\mathrm{d}\tau$$

is minus the rate of Joule heating. In which direction is $\mathbf{S}$ poynting? (pun intended)

C6.4 Draw two short parallel lines as a cartoon of a plate capacitor. Place positive charge on the left and negative on the right and draw the electric field lines. Note that from *outside* and a long way off the capacitor looks like an electric dipole. Now draw a circular section of wire that causes the capacitor to discharge through the wire. The electric field is now collapsing: draw the lines of $\dot{\mathbf{E}}$, and use dot and cross symbols to indicate the $\mathbf{H}$-field, recalling that $\boldsymbol{\nabla}\times\mathbf{H} = \epsilon_0\dot{\mathbf{E}}$. Now, maybe on another diagram, draw the fields in the vacuum just outside the wire and indicate the direction of $\mathbf{S} = \mathbf{E}\times\mathbf{H}$. How is electromagnetic energy being carried along the wire?

5CCP2380 Problems 6—Solutions

C6.1 If a point charge, q, is in motion with a velocity, $\mathbf{v}$, and experiences an electric field, $\mathbf{E}$, then the increment of work done *on* the charge *by* the field in an infinitesimal time interval $\mathrm{d}t$ is,

$$\mathrm{d}W = q\,\mathbf{E}\cdot\mathbf{v}\,\mathrm{d}t \tag{1}$$

I now want to sum, or to be more precise to integrate, over all the charges in a volume $\mathcal{V}$. If I treat each integrating volume, $\mathrm{d}\tau$, as a charge: $q = \rho\mathrm{d}\tau$, where ρ is the charge density—constant within the infinitesimal volume, $\mathrm{d}\tau$—then the rate of doing work from (1) is,

$$\frac{\mathrm{d}W}{\mathrm{d}t} = \int_{\mathcal{V}} \mathbf{E}\cdot\mathbf{v}\,\rho\,\mathrm{d}\tau$$

But if the charge distribution is flowing through the volume element with a velocity $\mathbf{v}$ then $\rho\mathbf{v} = \mathbf{J}$ is the current density at the position of the volume element, $\mathrm{d}\tau$. So I now have,

$$\frac{\mathrm{d}W}{\mathrm{d}t} = \int_{\mathcal{V}} \mathbf{E}\cdot\mathbf{J}\,\mathrm{d}\tau \tag{2}$$

I now apply the scalar product of $\mathbf{E}$ to the Ampère-Maxwell law,

$$\begin{aligned}\mathbf{E}\cdot(\boldsymbol{\nabla}\times\mathbf{B}) &= \mu_0\,\mathbf{E}\cdot\mathbf{J} + \mu_0\epsilon_0\,\mathbf{E}\cdot\frac{\partial\mathbf{E}}{\partial t} &\text{(3a)}\\ &= \mathbf{B}\cdot(\boldsymbol{\nabla}\times\mathbf{E}) - \boldsymbol{\nabla}\cdot(\mathbf{E}\times\mathbf{B})\\ &= -\mathbf{B}\cdot\frac{\partial\mathbf{B}}{\partial t} - \boldsymbol{\nabla}\cdot(\mathbf{E}\times\mathbf{B}) &\text{(3b)}\end{aligned}$$

using the vector identity $\boldsymbol{\nabla}\cdot(\mathbf{w}\times\mathbf{v}) = \mathbf{v}\,(\boldsymbol{\nabla}\times\mathbf{w}) + \mathbf{w}\,(\boldsymbol{\nabla}\times\mathbf{v})$ from page 1 of the notes to Section 1 in the second line, and Faraday's law in the last line. Now,

$$\frac{\partial}{\partial t}(\mathbf{E}\cdot\mathbf{E}) = \mathbf{E}\cdot\frac{\partial\mathbf{E}}{\partial t} + \frac{\partial\mathbf{E}}{\partial t}\cdot\mathbf{E} = 2\mathbf{E}\cdot\frac{\partial\mathbf{E}}{\partial t}$$

so,

$$\mathbf{E}\cdot\frac{\partial\mathbf{E}}{\partial t} = \frac{1}{2}\frac{\partial}{\partial t}E^2 \tag{4}$$

and similarly,

$$\mathbf{B}\cdot\frac{\partial\mathbf{B}}{\partial t} = \frac{1}{2}\frac{\partial}{\partial t}B^2 \tag{5}$$

So equating the right hand sides of (3a) and (3b) and using (4) and (5) I can write,

$$\begin{aligned}\mathbf{E}\cdot\mathbf{J} &= -\epsilon_0\,\mathbf{E}\cdot\frac{\partial\mathbf{E}}{\partial t} - \frac{1}{\mu_0}\mathbf{B}\cdot\frac{\partial\mathbf{B}}{\partial t} - \frac{1}{\mu_0}\boldsymbol{\nabla}\cdot(\mathbf{E}\times\mathbf{B})\\ &= -\frac{1}{2}\epsilon_0\frac{\partial}{\partial t}E^2 - \frac{1}{2}\frac{1}{\mu_0}\frac{\partial}{\partial t}B^2 - \boldsymbol{\nabla}\cdot\mathbf{S}\\ &= -\frac{1}{2}\frac{\partial}{\partial t}\left(\epsilon_0 E^2 + \frac{1}{\mu_0}B^2\right) - \boldsymbol{\nabla}\cdot\mathbf{S}\\ &= -\frac{\partial u_{\mathrm{vac}}}{\partial t} - \boldsymbol{\nabla}\cdot\mathbf{S} &\text{(6)}\end{aligned}$$

and the energy per unit volume of the electromagnetic field inside $\mathcal{V}$ is,

$$u_{\text{vac}} = \frac{1}{2}\left(\epsilon_0 E^2 + \frac{1}{\mu_0}B^2\right)$$

I combine this with (2) and I get,

$$\begin{aligned}\frac{\mathrm{d}W}{\mathrm{d}t} &= \int_{\mathcal{V}} \mathbf{E}\cdot\mathbf{J} \\ &= -\frac{\partial}{\partial t}\int_{\mathcal{V}} u_{\text{vac}}\,\mathrm{d}\tau - \int_{\mathcal{V}} \boldsymbol{\nabla}\cdot\mathbf{S}\,\mathrm{d}\tau\end{aligned}$$

and using the divergence theorem to convert the volume integral to an integral over the surface $\mathcal{S}$ bounding $\mathcal{V}$,

$$\frac{\mathrm{d}W}{\mathrm{d}t} = -\frac{\partial}{\partial t}\int_{\mathcal{V}} u_{\text{vac}}\,\mathrm{d}\tau - \oint_{\mathcal{S}} \mathbf{S}\cdot\mathrm{d}\mathbf{a}$$

This whole development is still valid in empty space—that is, space in which there are electromagnetic fields but no charges. An important example is space in which electromagnetic waves are propagating. In that case, $\mathbf{J} = \mathbf{0}$ and from (6) we deduce that,

$$\frac{\partial u_{\text{vac}}}{\partial t} = -\boldsymbol{\nabla}\cdot\mathbf{S} \quad \text{where no charge exists} \tag{7}$$

This is a conservation law, or continuity equation for electromagnetic energy (compare with the continuity equation for charge density in your Section 4 notes).

In the case of electromagnetic waves, the Poynting vector tells you how much power is being delivered into space from an oscillating (Hertzian) dipole or dipole antenna which is radiating electromagnetic waves.

Note that (6) infers that Griffiths defines the Poynting vector as,

$$\mathbf{S} = \frac{1}{\mu_0}\mathbf{E}\times\mathbf{B}$$

whereas we define in the Section 6 lecture notes,

$$\mathbf{S} = \mathbf{E}\times\mathbf{H}$$

These are equivalent unless there are bound currents in which case we will write Poynting's theorem as,

$$\frac{\partial u}{\partial t} + \mathbf{E}\cdot\mathbf{J}_{\text{f}} + \boldsymbol{\nabla}\cdot(\mathbf{E}\times\mathbf{H}) = 0$$

where, if you like,

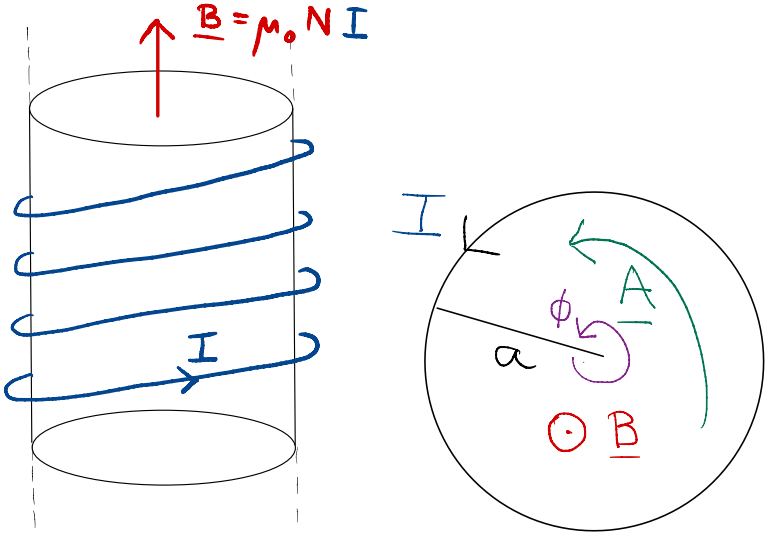
$$u = \frac{1}{2}\left(\mathbf{E}\cdot\mathbf{D} + \mathbf{H}\cdot\mathbf{B}\right)$$

is the "thermodynamic energy density" while u_{vac} is the "vacuum" energy density; whereas Griffiths will write Poynting's theorem as,

$$\left(\frac{\partial u_{\text{vac}}}{\partial t} + \mathbf{E}\cdot\mathbf{J}_{\text{b}}\right) + \mathbf{E}\cdot\mathbf{J}_{\text{f}} + \frac{1}{\mu_0}\boldsymbol{\nabla}\cdot(\mathbf{E}\times\mathbf{B}) = 0$$

In this way Griffiths modifies the energy density of the electromagnetic field to include the work done by the electric field on the bound charge. This means that for us, equation (7) is true in the absence of *conductors* while for Griffiths it's only true in the absence of *matter*.

C6.2 (i) Here is a sketch of the experiment. The coil, or solenoid, has a radius a, and has N windings, or turns, per metre. The current is I. You know from your first year classical physics that the magnetic field is zero outside (if the coil is "infinitely long") and is uniform inside with magnitude $\mu_0 NI$ and its direction is right handed, counter clockwise to the current.

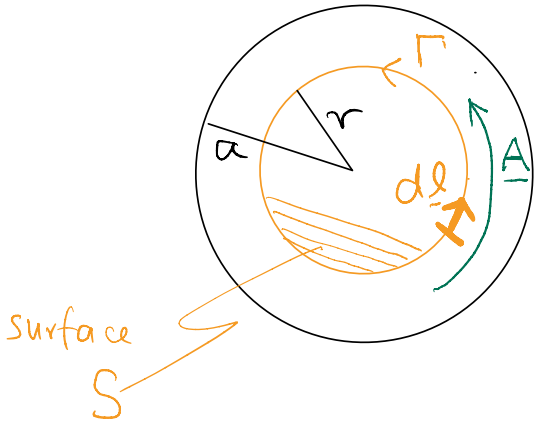


The second figure shows the view from above the coil. How do I know to indicate the vector potential as I have done? Well, firstly if you look at equation (5.3.5) in the Section 5 notes you see that in the Coulomb gauge,

$$\mathbf{A}(\mathbf{r}_0) = \frac{\mu_0}{4\pi}\int \frac{\mathbf{J}(\mathbf{r}_1)}{r}\mathrm{d}\tau_1$$

Since there is a vector on both sides and there are no cross or scalar products, the vector potential must be parallel to the current density. This is generally a good starting point for guessing the vector potential in a problem. Secondly since $\mathbf{B} = \boldsymbol{\nabla}\times\mathbf{A}$, the vector potential as I have sketched it is consistent with the magnetic field (see Problem C1.4, Section 1).

Now, by symmetry we expect the vector potential to depend only on the distance from the centre of the coil and not upon the value of the azimuthal angle, ϕ. However the vector potential may depend on the distance, r, from the centre of the coil. The way to calculate it is to construct an imaginary integration circuit, Γ, as show in the next diagram.



We will then integrate $\mathbf{B}$ over the flat surface, $\mathcal{S}$, that is bounded by the circular loop Γ. Indeed the magnetic field integrated over $\mathcal{S}$ is the magnetic flux penetrating that fraction of the cross section of the coil which is enclosed by a circle of radius r,

$$\int_{\mathcal{S}} \mathbf{B} \cdot \mathrm{d}\mathbf{a} = B \int_{\mathcal{S}} \mathrm{d}a = B \times \text{ area} = \pi r^2 B \tag{8}$$

As usual the integral is very easy—$\mathbf{B}$ is constant and parallel to the normal to each increment of surface $\mathrm{d}\mathbf{a}$ so we just get the field times the area enclosed by Γ and this is proportional to r^2. Now, I know that $\mathbf{B} = \boldsymbol{\nabla} \times \mathbf{A}$ so let's integrate the curl of $\mathbf{A}$ over the same surface Γ and then use Stokes's theorem.

$$\int_{\mathcal{S}} \boldsymbol{\nabla} \times \mathbf{A} \cdot \mathrm{d}\mathbf{a} = \oint_{\Gamma} \mathbf{A} \cdot \mathrm{d}\boldsymbol{\ell} \tag{9}$$

In words, the integral over a surface of the curl of a vector field is equal to the integral of the field itself over the line that bounds that surface. Now, again the integral on the right hand side is easy—$\mathbf{A}$ is parallel to $\mathrm{d}\boldsymbol{\ell}$, so $\mathbf{A} \cdot \mathrm{d}\boldsymbol{\ell} = A\mathrm{d}\ell$, A is constant and comes outside the integral which then becomes just the line integral of $\mathrm{d}\ell$ around the curve which is the just the length of that curve, $2\pi r$. This means that when I combine (8) and (9) I get,

$$\pi r^2 B = \pi r^2 \, \mu_0 N I = 2\pi r A \tag{10}$$

The final result that we are seeking is,

$$\mathbf{A} = \frac{1}{2}\mu_0 I N r \hat{\boldsymbol{\phi}} \quad , \qquad r < a$$

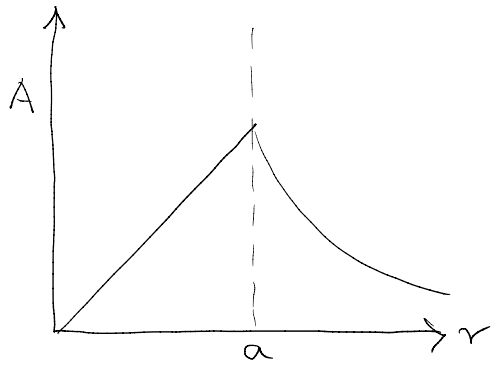
where $\hat{\boldsymbol{\phi}}$ is the unit vector in the direction of increasing azimuthal angle, ϕ, as in polar coordinates.

That was the case that $r < a$, as drawn in the figure. What about the vector potential *outside* the coil? Surely that's zero you say since there is no magnetic field outside the coil. Wrong! If I place my integration path outside the coil where $r > a$ there is a still magnetic flux because Γ encloses the whole width of the coil.

So for any $r > a$ the magnetic flux is constant and equal to $2\pi a^2 B$. So I replace r with a on the left hand side of (10) but leave r on the right hand side. Then I get,

$$\mathbf{A} = \frac{1}{2}\mu_0 I N \frac{a^2}{r}\hat{\boldsymbol{\phi}} \quad , \qquad r > a$$

As always in problems of this kind, you make a quick sketch of the field as a function of distance from the centre. You confirm that for $r = a$ the two expressions coincide and you illustrate that the vector potential increases linearly up to $r = a$ and then decays like $1/r$ outside the coil



Are you surprised that there is vector potential in a region of space where there is no magnetic field? Since the vector potential is anyway undefined to within a choice of gauge and since it is the magnetic field that is measurable, you may take the view that it doesn't really matter. But for future reference, let me tell you that in quantum mechanics what matters are the *potentials* not the *fields*. As you know the Schrödinger equation is,

$$i\hbar\dot{\psi} = H\psi$$

and the hamiltonian is the kinetic plus the potential energy,

$$H = T + U = \frac{p^2}{2m} + U$$

The momentum is the operator formed from the canonical substitution,

$$\mathbf{p} \to -i\hbar\boldsymbol{\nabla}$$

so the kinetic energy is $-(\hbar^2/2m)\nabla^2$. The equation of motion for a particle of charge q in an electric field is the Schrödinger equation with $U = qV$ where V is the electric potential. If there is also a magnetic field then so called *canonical momentum* is $\mathbf{p} - q\mathbf{A}$ and so the hamiltonian for a charged particle in an electromagnetic field is,

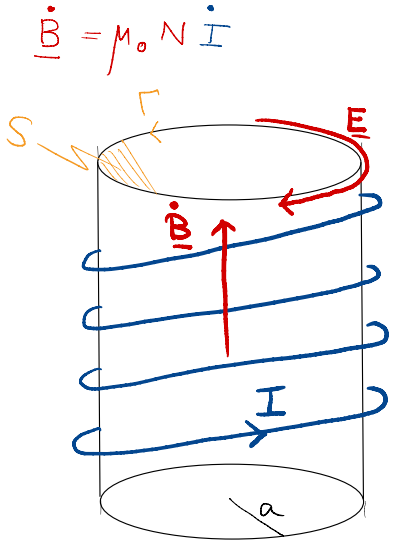
$$H = \frac{1}{2m}\left(-i\hbar\boldsymbol{\nabla} - q\mathbf{A}\right)^2 + qV$$

The point is that a quantum particle may detect vector potential even in a region of space in which there is no magnetic field and the effect is observable. You will learn of this when you study the Aharanov-Bohm effect.

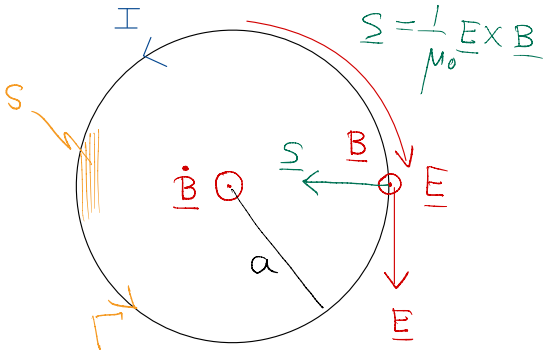
(*ii*) Now you are asked, what if the current in the coil is increased at a steady rate? If the magnetic field is $B = \mu_0 N I$ then by differentiating with respect to time, we have the time rate of change of the magnetic field being,

$$\dot{B} = \mu_0 N \dot{I}$$

The situation is as shown in the next figure.



By Faraday's law the changing magnetic field induces an electric field and by Lenz's law the direction of the induced electromotance is such as to *oppose* the change in magnetic flux. By this argument or just by respecting the minus sign in Faraday's law you deduce that the direction of the electric field is clockwise from above as shown in the figure. Now consider the surface $\mathcal{S}$ which is the cross section of the coil bounded by the line, Γ. You can see from the next figure that at the edge of the coil the Poynting vector is pointing inwards. We want to find its magnitude.



Well, again appealing to Faradays's law,

$$-\dot{\mathbf{B}} = \boldsymbol{\nabla} \times \mathbf{E}$$

and again as in part (i) you integrate both sides of this equation over the surface, $\mathcal{S}$, and again use Stokes's theorem to convert the surface integral of the curl of $\mathbf{E}$ to a line integral of $\mathbf{E}$ itself along the boundary Γ,

$$-\frac{\partial}{\partial t}\int_{\mathcal{S}} \mathbf{B} \cdot \mathrm{d}\mathbf{a} = \int_{\mathcal{S}} \boldsymbol{\nabla} \times \mathbf{E} \cdot \mathrm{d}\mathbf{a} = \oint \mathbf{E} \cdot \mathrm{d}\boldsymbol{\ell}$$

Now these integrals are just as easy to do. **B** is constant and parallel to d**a**, **E** is constant and parallel to d$\boldsymbol{\ell}$; so we get,

$$-\dot{B} \times \text{area} = E \times \text{circumference}$$

leading to,

$$2\pi a E = \pi a^2 \dot{B} = \pi a^2 \mu_0 N \dot{I}$$

and hence the induced electric field is,

$$E = \frac{1}{2}\, a \mu_0 N \dot{I}$$

Now I can construct the Poynting vector at any particular time during the ramp-up of the current, when the instantaneous current is I and the magnetic field is $B = \mu_0 N I$. Its magnitude is,

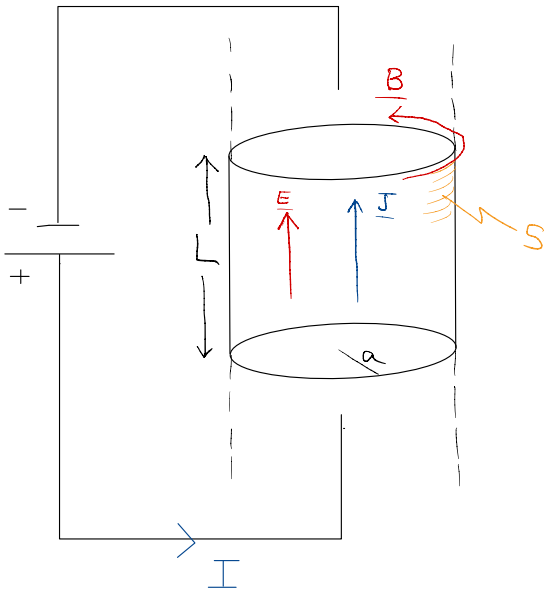
$$S = \frac{1}{\mu_0}\, E\, B = \frac{1}{\mu_0}\, \frac{1}{2}\, a \mu_0 N \dot{I}\, \mu_0 N I$$

so finally,

$$S = \frac{1}{2} \mu_0 a N^2 I \dot{I}$$

and **S** is pointing radially inwards.

C6.2 Imagine a blown up image of part of the the wire. Or you can think of this as the resistor as in the cartoon I showed in the video, Auxiliary Fields—2. Indeed I promised that we would work through this problem in detail.



So consider a resistor of length L. The current, and hence the current density flows up the wire as shown and this is driven by an electromotive field, **E**, also upwards since the electric potential is more positive at the lower end than the top end. Ohm's law is,

$$\mathbf{E} = \mathcal{R}\mathbf{J}$$

This calls for some comment. Firstly $\mathcal{R}$ is the *resistivity* and is the reciprocal of the *conductivity*, σ. (Up to now I have written $\mathbf{J} = \sigma\mathbf{E}$.) Secondly you may be more familiar with Ohm's law in the form $V = IR$, in which R is *resistance*. Please think about it for a bit. $\mathbf{J} = \sigma\mathbf{E}$ is the more fundamental statement; it asserts that the linear response of a conductor to an electric field is a current density. If I hand you a piece of copper, you can tell me its resistivity because you can look that up in data tables (it's 1.7×10^{-8} Ohm-metre at room temperature—more than thirty orders of magnitude smaller than the resistivity of teflon). It is a property of the substance copper. But its resistance (in Ohms), if it's, say, a wire, depends on its thickness; so to tell me that you'd have to measure it and if it's an odd shape you may have to do a calculation.

Back to the problem. From Ampère's law the magnetic field circulating the cylindrical wire or resistor at its cylindrical surface is,

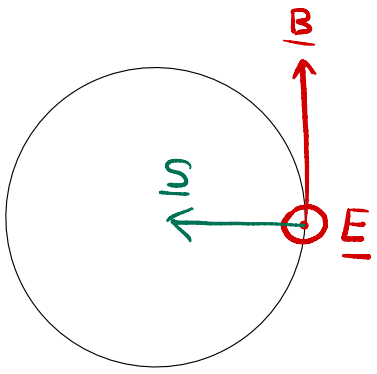
$$\mathbf{B} = \frac{\mu_0}{2\pi}\frac{I}{a}\hat{\boldsymbol{\phi}}$$

But $I = J\pi a^2$ (current density times area), so the magnitude of $\mathbf{B}$ is,

$$B = \frac{1}{2}\mu_0 J a$$

The Poynting vector is directed inwards (see the next figure) as you can deduce from the right handed cross product,

$$\mathbf{S} = \frac{1}{\mu_0}\mathbf{E} \times \mathbf{B}$$



The magnitude of the Poynting vector is,

$$S = \frac{1}{\mu_0}EB = \frac{1}{\mu_0}\mathcal{R}J\,\frac{1}{2}\mu_0 J a$$

So that our result is,

$$S = \frac{1}{2}J^2 a\mathcal{R}$$

If I now integrate $\mathbf{S}$ over the external cylindrical surface, $\mathcal{S}$ of the resistor of length L I get,

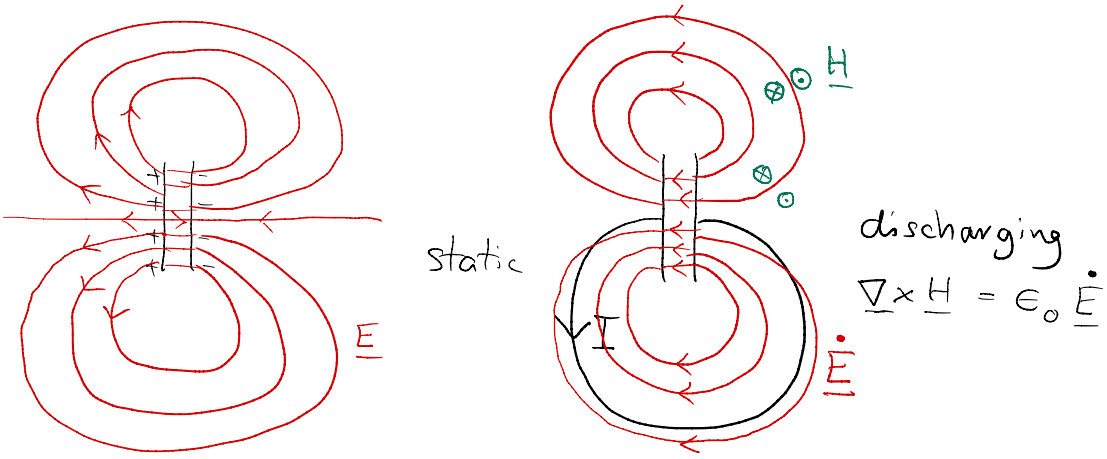
$$\begin{aligned}\int_{\mathcal{S}} \mathbf{S}\cdot \mathrm{d}\mathbf{a} &= -2\pi aL\,\frac{1}{2}J^2 a\mathcal{R}\\ &= -J^2\mathcal{R}\,\pi a^2 L\\ &= -J^2\mathcal{R}\times \text{ volume}\end{aligned}$$

and this is, as expected, the *Joule heating.* (See the video, Auxiliary Fields—2.)

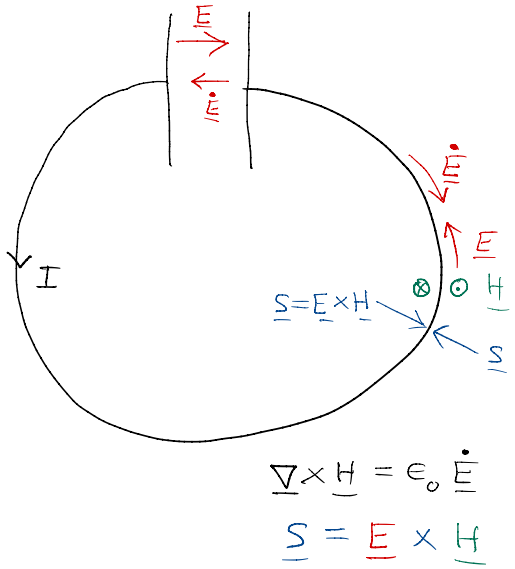
In both of the previous two problems, we have a battery e.m.f. feeding energy into a volume of interest. Because in both cases the magnetic field of interest is either in the vacuum within the coil or in the vacuum just outside the surface of the resistor, it doesn't matter whether we use the form $\mu_0^{-1}\mathbf{E}\times\mathbf{B}$ or $\mathbf{E}\times\mathbf{H}$ as our Poynting vector. In the first case energy is fed into the coil in order to support an increasing magnetic field against the back e.m.f. of Lenz's law; in the second energy from the battery is fed into the resistor to provide the Joule heat. The viewpoint of the Poynting vector is that you can calculate the energy fed in or out of electromagnetic fields inside a volume of interest without reference to the battery e.m.f. In both examples, the direction of $\mathbf{S}$ indicates energy fed *into* the volume.

Finally, try and think a bit about how electromagnetic energy is stored within the vacuum outside a charged or magnetised object. It is hard to visualise. You can easily accept that the energy in a stretched spring is stored in the elastic distortion of the metal, but how does the vacuum store energy? (It also stores angular momentum.) This puzzled the Victorians, including Maxwell, greatly; and they were convinced that the vacuum was in fact permeated by an elastic substance called ether that was capable of supporting the vibrations of the fields of an electromagnetic wave.

C6.4 This is a problem first proposed and solved by Poynting himself in 1884. It resolved a heated controversy over the transmission in cable telegraphy. Heaviside and others were locked in a conflict with telegraph engineers over the best way to transmit signals over the transatlantic cable and through telegraph wires. One point at issue was the role of self inductance as proved by Heaviside in his "telegraph equation" which predicts the ratios of self inductance, L, resistance, R capacitance, C, and "leakage" conductance through the insulation, $G = I/V$ (all per unit length of cable), needed to minimise distortion of the signal: namely, $C/G = L/R$. The pig-headed telegraph engineers believed that self-inductance and leakage should always be as small as possible and in fact telegraphy was set back by over 20 years because of the refusal of W. H. Preece, head engineer of the Post Office Telegraph Department, in particular, to pay heed to the scientists. Another point of contention was how is energy transmitted down a wire. Thomson (later Lord Kelvin) and others took the rather obvious view that the current of electricity carried the energy inside the wire. Heaviside argued that the current was conducted principally along the surface of the wire only. This problem indicates the correct and maybe surprising view that, again, in fact energy is stored in the vacuum (the Maxwellians believed this was the "ether") and transmitted in the region *outside the wire.* Hence the importance of not just the wire itself but the insulating and sheathing substances (for example gutta-percha rubber and steel protection).



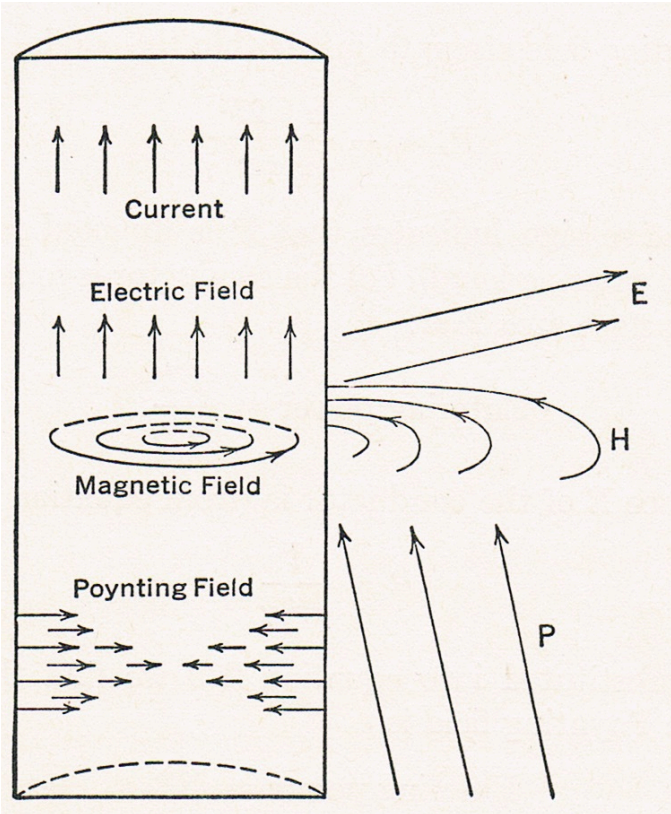
In the static case the **E**-field is as shown on the left. When the wire (black) is connected the capacitor discharges and as the **E**-field collapses, field lines of $\dot{\mathbf{E}}$ appear having the same pattern as **E** but opposite sense. The changing **E**-field induces an **H**-field according to $\boldsymbol{\nabla} \times \mathbf{H} = \epsilon_0 \dot{\mathbf{E}}$ which is the Ampère–Maxwell law in which only the displacement current survives as there is no free current outside the wire.



Draw just the wire and indicate the directions of $\dot{\mathbf{E}}$, **E** and **H** and you see that as in previous examples the Poynting vector is feeding electromagnetic energy into the wire. So it is *not* as the early physicists supposed the "charged fluid" (what we now know as electrons) that are carrying the energy, it is the electromagnetic field outside the wire. Poynting wrote, "..prevailing and somewhat vague opinion [that energy is] in some way... carried down along the conductor by the current." Indeed, as Hunt writes, "... recognise the surrounding air or insulation as the true conveyor of electromagnetic energy."

I will leave you with a perplexing issue. In Section 3, I wrote of a "neutral wire" carrying a current; and indeed the crux of that argument rested on the wire being *neutral* in the laboratory frame. But *is* the wire neutral? In fact *not*, but the

reasons are subtle. You can work through problem 7.43 in Griffiths or for a deeper treatment see Assis, *et al.*, *Foundations of Physics*, **29**, 729 (1999) and references therein.



I found the above sketch in H. H. Skilling, *Fundamentals of Electric Waves*, (New York, John Wiley, 1942). Griffiths and also Sommerfeld and others work out the electric fields outside and inside the wire. Inside it is directed along the wire in the direction of the current; outside the field is subtly dependent on the boundary conditions but you notice that it has both tangential (as inside the wire) and a radial component. Because there is a discontinuity in the radial component of **D** there *must* be free charge on the surface of the wire on account of Gauss's law. You note that that Poynting field is feeding energy into the wire to account for the Joule heating but in addition there is a component of **S** carrying energy along the wire (Skilling calls it **P**). This is very significant, as Heaviside first explained (although he wasn't listened to by the Post Office engineers) because the energy flow along a telegraph wire or a transatlantic cable is actually carried down the *insulating* dielectric, *not* the metal cable.

7. Electric fields in matter

7.1 Conductors

You know that if I place a conductor, such as a metal object, in a *static* electric field, the electric field inside the conductor is always zero.† This is because there are mobile charges, almost always *electrons*, that arrange themselves on the surface of the conductor in such a way as to exactly cancel the external electric field.

Suppose I have a parallel plate capacitor and arrange a surface charge density, $\pm\sigma_{\text{free}}$ on the two plates. The field inside is $E_0 = \sigma_{\text{free}}/\epsilon_0$ and if the area of the plates is A, the charge is $Q = \sigma_{\text{free}}\, A$ and the capacitance is

$$C = \frac{Q}{V} = \frac{\epsilon_0\, A}{d}$$

where V is the potential difference

$$V = E_0\, d = \frac{\sigma_{\text{free}}\, d}{\epsilon_0}$$

and d is the separation between the plates.

Now I insert a piece of metal of width b into the space between the plates, figure 7–1, and adjust the instrumentation to maintain the surface charge density to be σ_{free}.‡

Charges *equal and opposite* to those on the plates are "induced" on the metal surfaces such as to produce an equal and opposite field to $\mathbf{E}_0$ inside the metal.

What is now the capacitance? Well, Q is the same but the *potential* is reduced because it's the integral of the field between the plates and now the field is restricted to the vacuum of width $(d - b)$. Hence

$$V = \frac{\sigma_{\text{free}}}{\epsilon_0}\,(d - b)$$

and the capacitance is

$$C = \frac{\epsilon_0\, A}{d\left(1 - \dfrac{b}{d}\right)}$$

---

† I exclude here the non irrotational electromotive field that is reponsible for generating currents in metals.

‡ You will see that we talk of the "free" charge density as that charge that we *control* by suitable instrumentation. You will also learn that we adopt the same concept in magnetism of "free" current that we control. Actually it's more obvious in magnetism that we control current, say in a solenoid and that immediately establishes a magnetic excitation, $\mathbf{H}$, inside the coil. It's less clear that we can do this in electrostatics because typical instrumentation controls *current* and *voltage*, but not *charge*. So in that sense we have to accept that *at least in principle* we can control the charge density, say on a capacitor plate although the instrumentation may not be trivial.

the capacitance has been *increased* by the factor

$$\frac{d}{d-b}$$

Note, I have written the charge on the plates with a subscript, 'free', to denote *free charge*, so called. This is the charge that I control by suitable instrumentation. The "induced" charge is out of my control, unmeasurable, and sometimes called "inaccessible" charge. We will call this "polarisation charge," or "bound charge," σ_{b}. In the present example, the bound charge density is numerically equal to the free charge density.

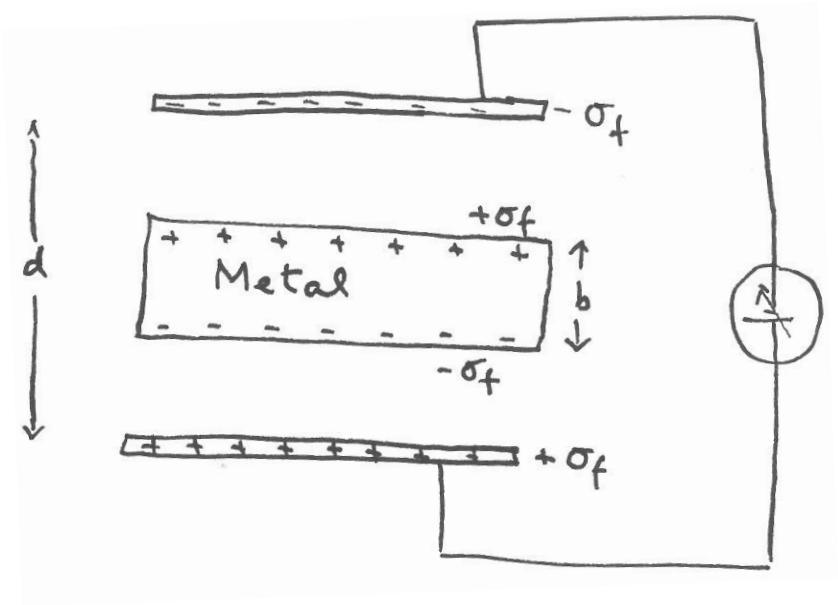


FIGURE 7–1

7.2 Insulators

Of course, well all know that capacitance is increased by inserting an insulating material, or *dielectric*, in between the plates. Can we guess how this might be happening? There are no fully mobile charges in a dielectric (unless it breaks down under high voltage), but maybe enough charge can move about a bit—enough for some charge to accumulate on the surface. Then the picture would look like figure 7–2.

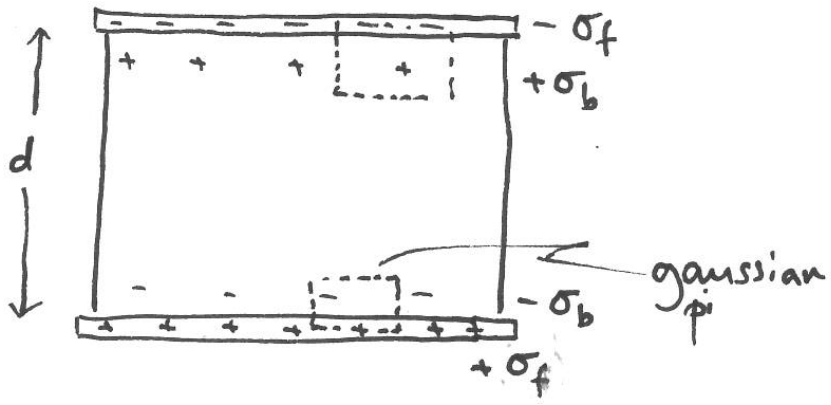


FIGURE 7–2

Using a gaussian "pillbox", in which the charge per unit area inside is $\sigma_{\text{free}} - \sigma_{\text{b}}$, we see that the field in the dielectric is

$$E = \frac{1}{\epsilon_0}\left(\sigma_{\text{free}} - \sigma_{\text{b}}\right) \tag{7.2.1}$$

and the capacitance is

$$C = \frac{\epsilon_0 A}{d}\,\frac{\sigma_{\text{free}}}{\sigma_{\text{free}} - \sigma_{\text{b}}} = \frac{\epsilon_r \epsilon_0 A}{d}$$

which increases the vacuum capacitance by a factor

$$\epsilon_r = \frac{\sigma_{\text{free}}}{\sigma_{\text{free}} - \sigma_{\text{b}}}$$

called relative permittivity, or dielectric constant. Note, again that σ_{free} is the free charge (per unit area) that I put on the plates. This causes an externally applied electric field, $\mathbf{E}_0$, to appear in the dielectric which responds with a distribution of bound charge; this changes the field, $\mathbf{E}$, in the dielectric, causing the bound charge to change, and so on... We cannot control, or even measure, the bound charge and how much you get depends on the applied field, $\mathbf{E}_0$, and on the nature of the dielectric material.

7.3 Polarisation

What is happening is that the field between the capacitor plates is causing *polarisation* in the dielectric. That is, dipoles are lining up with the field to produce bound charges on the surface. As long as the field is uniform and the dielectric is homogeneous, then no bound charge will accumulate inside the material. In simple terms, the tail of any one dipole is cancelled by the head of the one behind it, figure 7–3.

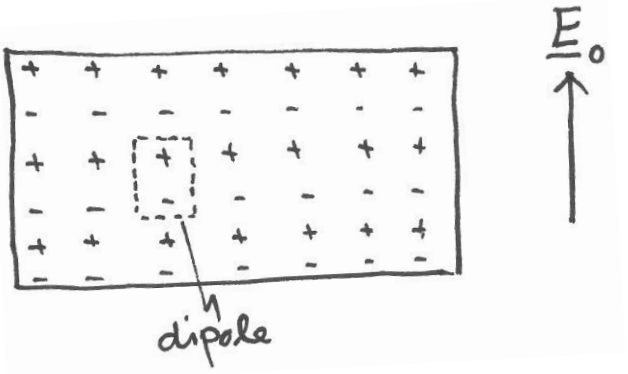


FIGURE 7–3

Polarisation can come about in two ways.

1. The dipoles already exist in the material of the dielectric and are caused to line up by the torque acting on them due to the electric field. This phenomenon occurs in gases and liquids. The *dipole moment per unit volume* is

$$\mathbf{P} = \frac{Np^2}{3kT}\,\mathbf{E} \tag{7.3.1}$$

where $\mathbf{p}$ is the permanent dipole moment of the molecule, N is the number of molecules per unit volume, k is the Boltzmann constant and T the absolute temperature. I give a proof of this in the Appendix to this section; but you won't be able to follow it until you've studied statistical mechanics.

2. The electric field itself polarises the molecules of the dielectric. This can happen in two ways.

   2.1 Ionic polarisation, in which ions of different charge are displaced slightly with respect to each other in the electric field. Think of the NaCl crystal made up of two interpenetrating face centred cubic lattices of Na^+ and Cl^- ions respectively. They are bound into exact cubic symmetry by the electric forces (Madelung energy); but imagine if placed in between the plates of a capacitor: the two sublattices will be attracted to opposite plates causing a rigid displacement between positive and negative sublattices and hence a permanent dipole moment. Indeed the layer of Na^+ ions closest to the negative plate amounts to a surface bound charge density and likewise for the anions at the other end of the crystal. Ionic polarisation is the principal mechanism of polarisation of dielectrics such as barium titanate, in solid state capacitors.

   2.2 Electronic polarisation, in which the nucleus of an atom is displaced slightly from its electron cloud as a result of the electric field, $\mathbf{E}_0$. The ease with which an atom can be polarised is characterised by its *electronic polarisability*, α, defined such that the dipole moment as a function of the electric field is

$$\mathbf{p} = \alpha\,\mathbf{E} \tag{7.3.2}$$

   It is not always true that the polarisation is *linearly* proportional to the field, but it must be if the field is small enough. We do not treat *any* non linear effects in these lectures.

   To obtain some insight, and an estimate of α, consider an atom in an oscillating electric field. The atom may be in a vacuum or form part of a solid or liquid. The electrons will oscillate about the nucleus and their centre of charge, x, will be in accord with Newton's second law,

$$m\ddot{x} + m\omega_0^2 x = -eE\cos\omega t$$

   where m is some effective mass of the electrons (or just the electron mass in the case of a hydrogen atom) and $-e$ the total charge. Actually, for simplicity let's suppose we are dealing with a hydrogen atom and $-e$ is then the electron charge. This equation describes, as you know, the motion of an undamped oscillator with resonant frequency ω_0 being driven at a frequency ω. The amplitude is (see additional notes "simple harmonic motion" on `KEATS`)

$$x = \frac{-eE}{m\,(\omega_0^2 - \omega^2)}$$

   so the displacement is huge at resonance, at which ω is usually in the visible or UV absorption bands.

But *we* want to consider just a static field, so we set $\omega = 0$ and get

$$x = \frac{-eE}{m\omega_0^2}$$

Since this is the displacement of the centre of charge of the electron with respect to the nucleus, then the dipole moment must be

$$p = -ex = \frac{e^2 E}{m\omega_0^2}$$

which is linear in the field. Together with $p = \alpha E$ we find

$$\alpha = \frac{e^2}{m\omega_0^2}$$

So the polarisability depends on the fundamental constants e^2 and m and the squared inverse of the resonant frequency—the "floppiness" of the electrons—it makes sense. As an estimate of ω_0 for hydrogen we identify the resonant energy $\hbar\omega_0$ with the ionisation energy which is one Rydberg unit of energy,

$$\hbar\omega_0 = 1\text{Ry} = \frac{1}{2}\frac{1}{(4\pi\epsilon_0)^2}\frac{me^4}{\hbar^2} = 13.6\text{eV}$$

It follows that

$$\begin{aligned}\alpha &= 4\,(4\pi\epsilon_0)^4\,\frac{\hbar^6}{m^2e^8}\frac{e^2}{m} \\ &= 4\,(4\pi\epsilon_0)\left(4\pi\epsilon_0\frac{\hbar^2}{me^2}\right)^3 \\ &= 16\pi\,\epsilon_0\,a_0^3\end{aligned}$$

where a_0 is the Bohr radius, or "radius of the hydrogen atom" (0.529×10^{-10} m). So, a rule of thumb is that the polarisability of an atom is proportional to its "radius cubed" or atomic "size". This is a pretty good estimate: for hydrogen

$$\frac{\alpha}{4\pi\epsilon_0} = 4a_0^3 = 0.592\times10^{-30}\ \text{m}^3$$

while the observed value is 0.667×10^{-30} m^3.†

† Be careful! because some textbooks define α through $p = \epsilon_0\alpha E$ so that α has units of m^3.

Be careful to distinguish $\mathbf{p}$ [Cm], the dipole moment of an object; and $\mathbf{P}$ [Cm^{-2}] (having the same units as surface bound charge), the dipole moment per unit volume of a dielectric.

If, now, we *add* the electric polarisation to the polarisation due to orientating of fixed dipoles, we have for a gas or a liquid,

$$\mathbf{P} = \left(N\alpha + \frac{Np^2}{3kT} \right) \mathbf{E}$$

In a solid the problem is much harder and even the solid state physics textbooks sometimes gloss over the difficulty that in a periodic crystal the dipole moment is not well-defined.

All the same it is always true that *all three* mechanisms of polarisation of a dielectric:

- alignment of fixed dipoles
- electronic polarisation of atoms or molecules
- ionic polarisation

result in dipoles aligning *along* the direction of the applied electric field, $\mathbf{E}_0$ so as to partially cancel it. Polarisation always reduces the field in a dielectric.† A metal is a "perfect dielectric"—it cancels the field totally. The field generated by polarisation is sometimes called the *depolarising field*, for rather obvious reasons.

† We will see later the contrast with magnetic materials which can be either *diamagnets*, which reduce the magnetic field within the material compared to the applied field; and *paramagnets* which enhance the field. There is no dielectric equivalent of a paramagnet. While we're on this comparison, the equivalent of a permanent magnet in electrostatics is the "electret"—see Problems Set 6, problem C6.5.

7.4 The case of an inhomogeneous dielectric in a non uniform field

Unlike in figure 7–3, section 7.3, the density of dipoles may vary from place to place either due to variations in the material, or variations in the electric field, $\mathbf{E}(\mathbf{r})$. The geometry, as shown in figure 7–4 will be familiar to you.

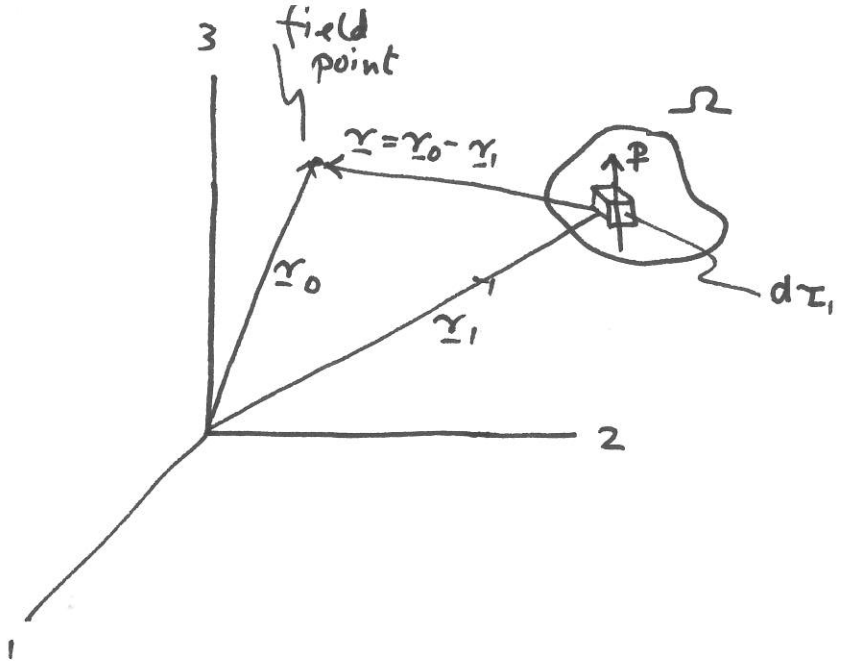


FIGURE 7–4

How do we find the electric field *caused* by this distribution of dipoles? Well, the dipole moment per unit volume, $\mathbf{P}$, may vary from place to place, but we regard it as made up of a collection of little dipoles of moment $\mathbf{p}$; in any volume element, $\mathrm{d}\tau_1$, the dipole moment is

$$\mathbf{p} = \mathbf{P}(\mathbf{r}_1)\,\mathrm{d}\tau_1$$

For a single dipole, $\mathbf{p}(\mathbf{r}_1)$, at a source point $\mathbf{r}_1$, the electric potential at $\mathbf{r}_0$, the field point, is (see Problems Set 3, C3.1(**a**))

$$V(\mathbf{r}_0) = \frac{1}{4\pi\epsilon_0}\,\mathbf{p}(\mathbf{r}_1)\cdot\boldsymbol{\nabla}_1\left(\frac{1}{r}\right)$$

where r is the magnitude of the connecting vector, $\mathbf{r} = \mathbf{r}_0 - \mathbf{r}_1$. If there is a distribution of dipoles in some volume Ω, amounting to a dipole moment per unit volume, $\mathbf{P}$, then in a volume element $\mathrm{d}\tau_1$ there is a dipole moment $\mathbf{p} = \mathbf{P}\,\mathrm{d}\tau_1$ so

$$V(\mathbf{r}_0) = \frac{1}{4\pi\epsilon_0}\int_\Omega \mathbf{P}\cdot\boldsymbol{\nabla}_1\left(\frac{1}{r}\right)\mathrm{d}\tau_1 \tag{7.4.1}$$

We use the product rule for any vector field, $\mathbf{u}(\mathbf{r})$, and any scalar field, $f(\mathbf{r})$,

$$\boldsymbol{\nabla}\cdot(f\mathbf{u}) = f\,(\boldsymbol{\nabla}\cdot\mathbf{u}) + \mathbf{u}\cdot(\boldsymbol{\nabla}f)$$

and find that

$$V(\mathbf{r}_0) = \frac{1}{4\pi\epsilon_0}\left(\int_\Omega \boldsymbol{\nabla}_1 \cdot \left(\frac{\mathbf{P}}{r}\right) \mathrm{d}\tau_1 - \int_\Omega \frac{1}{r}\left(\boldsymbol{\nabla}_1 \cdot \mathbf{P}\right) \mathrm{d}\tau_1\right)$$

then using the divergence theorem

$$V(\mathbf{r}_0) = \frac{1}{4\pi\epsilon_0} \oint_S \frac{\mathbf{P} \cdot \mathrm{d}\mathbf{a}_1}{r} - \frac{1}{4\pi\epsilon_0} \int_\Omega \frac{1}{r} \boldsymbol{\nabla} \cdot \mathbf{P} \, \mathrm{d}\tau_1 \tag{7.4.2}$$

Note two mathematical points.

1. You could have got straight here from (7.4.1) by partial integration: transfer the $\boldsymbol{\nabla}_1$ operator to $\mathbf{P}$, change the sign and add the integral of $\mathbf{P}/r$ over the boundary.
2. Having transferred $\boldsymbol{\nabla}_1$ to $\mathbf{P}$ it no longer acts on a function of the connecting vector so I can drop the subscript.

What! a revealing formula is (7.4.2). The first term is the potential due to a surface charge, $\mathbf{P} \cdot \mathrm{d}\mathbf{a}_1$, call it σ_b, plastered over the surface bounding the volume containing the dielectric if we define

$$\boxed{\sigma_\mathrm{b} = \mathbf{P} \cdot \hat{\mathbf{n}}} \tag{7.4.3}$$

where $\hat{\mathbf{n}}$ is the normal to the surface at the volume element $\mathrm{d}\mathbf{a}_1$ The second term in (7.4.2) is the potential due to a charge distribution within the volume of the dielectric if I define

$$\boxed{\rho_\mathrm{b} = -\boldsymbol{\nabla} \cdot \mathbf{P}} \tag{7.4.4}$$

as the *bound charge density*. So to find the electric field caused by a polarised dielectric, if you know the dipole moment per unit volume (the *polarisation*) everywhere then you calculate the surface bound charge density as $\mathbf{P} \cdot \hat{\mathbf{n}}$ and the volume bound charge density as $-\boldsymbol{\nabla} \cdot \mathbf{P}$ (see Problems Class 6.2). (Note that if the polarisation is uniform throughout the dielectric then all its derivatives are zero and $\rho_\mathrm{b} = 0$ as we found for the capacitor in the uniform field $\mathbf{E}_0$.) Then the potential is that due to these two,

$$\boxed{V(\mathbf{r}_0) = \frac{1}{4\pi\epsilon_0} \oint_S \frac{\sigma_\mathrm{b}}{r} \, \mathrm{d}a_1 + \frac{1}{4\pi\epsilon_0} \int_\Omega \frac{\rho_\mathrm{b}}{r} \, \mathrm{d}\tau_1} \tag{7.4.5}$$

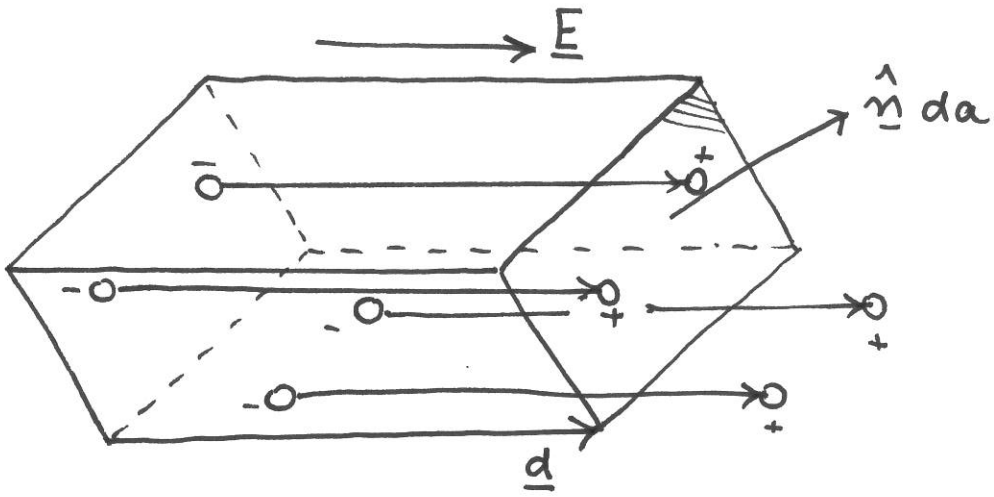


FIGURE 7–5

If you don't like the mathematical derivation of (7.4.3) and (7.4.4), then here is a pictorial demonstration. Consider a small surface element within or on the outer surface of the dielectric material having an area da and unit normal vector $\hat{\mathbf{n}}$, so that as usual $\mathrm{d}\mathbf{a} = \hat{\mathbf{n}}\,\mathrm{d}a$. If the dielectric comprises unpolarised molecules then when an electric field, $\mathbf{E}$, is turned on inside the dielectric positive and negative charges, $\pm q$ will separate by a vector $\mathbf{d}$ to create a dipole moment of $q\mathbf{d}$ per molecule. It's relative movement so let's suppose the negative charges stay put and the positive charges all move an amount $\mathbf{d}$. Some, n_+, say, will cross through the surface; the total charge crossing our surface in the outward direction is $\mathrm{d}Q = n_+ q$. Suppose I now construct a parallelepiped as illustrated in figure 7–5. whose front and back surfaces are d$\mathbf{a}$ and $-\mathrm{d}\mathbf{a}$ and whose sides are the vector $\mathbf{d}$, and whose volume therefore is $\mathrm{d}\Omega = \mathbf{d}\cdot\mathrm{d}\mathbf{a}$. Clearly the number of positive charges that move through the end-surface must equal the number of charges within the parallelepiped before they separated from their negative ends (because the length of the box is d). The number of molecules per unit volume is therefore $N = n_+/\mathrm{d}\Omega$; the dipole moment of each molecule is $\mathbf{p} = q\mathbf{d}$ and hence $\mathbf{P} = N\mathbf{p}$, and so we have,

$$\begin{aligned}\mathrm{d}Q &= n_+ q\\ &= (N\,\mathbf{d}\cdot\hat{\mathbf{n}}\,\mathrm{d}a)\;q\\ &= N\,\mathbf{p}\cdot\hat{\mathbf{n}}\,\mathrm{d}a\\ &= \mathbf{P}\cdot\hat{\mathbf{n}}\,\mathrm{d}a\end{aligned}$$

If our chosen plane, d$\mathbf{a}$, coincides with an outer surface of the material then we recover the result (7.4.3) that the the bound surface charge density is

$$\frac{\mathrm{d}Q}{\mathrm{d}a} = \mathbf{P}\cdot\hat{\mathbf{n}} = \sigma_{\mathrm{b}}$$

So we have the result that on application of the field $\mathbf{E}$ the amount of charge that crosses an arbitrarily orientated surface element, d$\mathbf{a}$, is $\mathbf{P}\cdot\mathrm{d}\mathbf{a}$. Let us now consider an arbitrary volume, Ω contained entirely within the matter of the dielectric then on application of the field the amount of charge that leaves the volume is

$$Q = \oint_S \mathbf{P}\cdot\mathrm{d}\mathbf{a}$$

Since the volume was originally neutral then the amount of charge that remains within Ω is $-Q$. So this amount of charge is the volume bound charge density, $\rho_{\rm b}$, integrated over the volume, Ω. This means that,

$$\int_\Omega \rho_{\rm b}\,{\rm d}\tau = -Q = -\oint_S \mathbf{P}\cdot{\rm d}\mathbf{a} = -\int_\Omega \boldsymbol{\nabla}\cdot\mathbf{P}\,{\rm d}\tau$$

after using the divergence theorem. Since the volume Ω is arbitrary it follows that the integrands are equal and we recover equation (7.4.4),

$$\rho_{\rm b} = -\boldsymbol{\nabla}\cdot\mathbf{P}$$

7.5 The electric field due to a uniformly polarised ball

This is an important problem in electrostatics. You will tackle it in a different way in Problems Class 6. It turns out that the field inside is constant and the field outside is that of a physical dipole.

First we do a rather tricky problem: two balls of opposite charge, slightly displaced. I have a ball of radius R carrying a charge Q, uniformly distributed; and another ball of radius R carrying a uniformly distributed charge $-Q$. These are superimposed (so there is no charge anywhere) and then the negative ball is slightly displaced by an amount d along the x-axis. This produces positive and negative surface charges rather like the bound surface charge on a slab of dielectric. What is the electric field in the overlapping region?

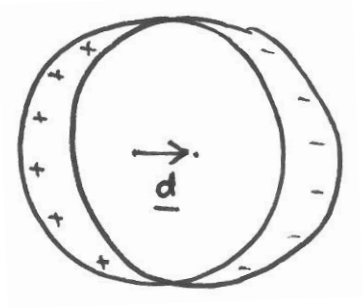


FIGURE 7–6

The charge density in each ball is

$$\rho = \pm\frac{Q}{\frac{4}{3}\pi R^3}$$

and the electric field inside a ball carrying a uniform density is found using Gauss's law. The field is radial and the flux over a gaussian sphere of radius r is

$$E\times 4\pi r^2 = \frac{1}{\epsilon_0}\,Q_{\rm enclosed} = \frac{1}{\epsilon_0}\,\rho\,\frac{4}{3}\pi r^3$$

and so

$$E = \frac{1}{\epsilon_0}\frac{1}{3}\rho r \tag{7.5.1}$$

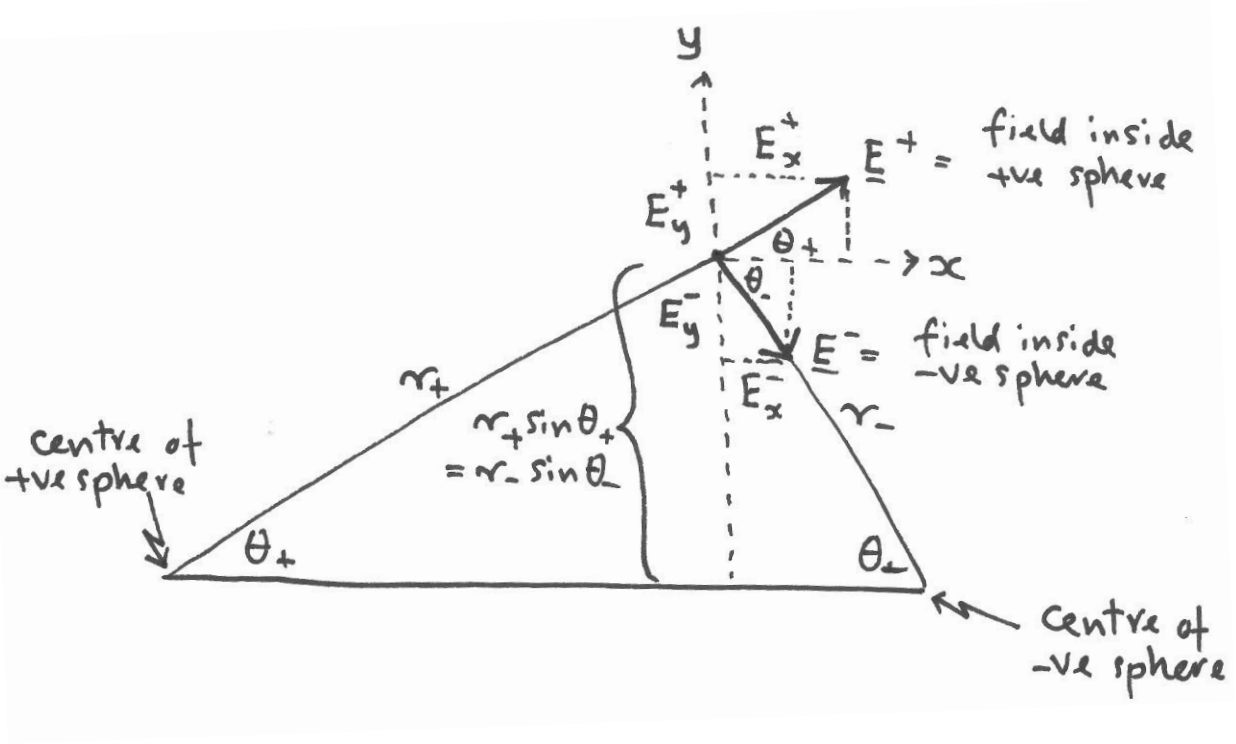


FIGURE 7–7

Now consider a field point in the overlapping region between the two balls. We have, by reference to figure 7–7, and using (7.5.1)

$$\begin{aligned}
E_x^+ &= \left|\mathbf{E}^+\right| \cos\theta_+ = \frac{1}{\epsilon_0}\frac{1}{3}\rho\, r_+ \cos\theta_+ \\
E_y^+ &= \left|\mathbf{E}^+\right| \sin\theta_+ = \frac{1}{\epsilon_0}\frac{1}{3}\rho\, r_+ \sin\theta_+ \\
E_x^- &= \left|\mathbf{E}^-\right| \cos\theta_- = \frac{1}{\epsilon_0}\frac{1}{3}\rho\, r_- \cos\theta_- \\
E_y^- &= -\left|\mathbf{E}^-\right| \sin\theta_- = -\frac{1}{\epsilon_0}\frac{1}{3}\rho\, r_- \sin\theta_-
\end{aligned}$$

By the cosine rule,

$$\begin{aligned}
r_-^2 &= r_+^2 + d^2 - 2r_+ d\cos\theta_+ \\
r_+^2 &= r_-^2 + d^2 - 2r_- d\cos\theta_-
\end{aligned}$$

and adding these gives,

$$r_+\cos\theta_+ + r_-\cos\theta_- = d$$

and so the x-component of the field is

$$E_x = E_x^+ + E_x^- = \frac{1}{\epsilon_0}\frac{1}{3}\rho\, d$$

The y- and z-components are zero. Outside the ball, as you will confirm in Problems Class 6, the electric field is that of a physical dipole of strength Qd. The electric field lines are drawn in figure 7–8.

The *main result* is that the electric field inside is uniform and is

$$\mathbf{E} = \frac{1}{3\epsilon_0}\rho\,\mathbf{d}$$

since $\mathbf{d}$ points in the x-direction. But because by convention the dipole moment points from the negative to the positive ends,

$$\mathbf{p} = -Q\mathbf{d} = -\frac{4}{3}\pi R^3\rho\,\mathbf{d} = \frac{4}{3}\pi R^3\,\mathbf{P} \qquad \text{(polarisation}\times\text{volume)}$$

so, *finally* the field inside a uniformly polarised ball is

$$\mathbf{E} = -\frac{1}{3\epsilon_0}\mathbf{P} \tag{7.5.2}$$

where $\mathbf{P}$ is the dipole moment per unit volume—the polarisation.

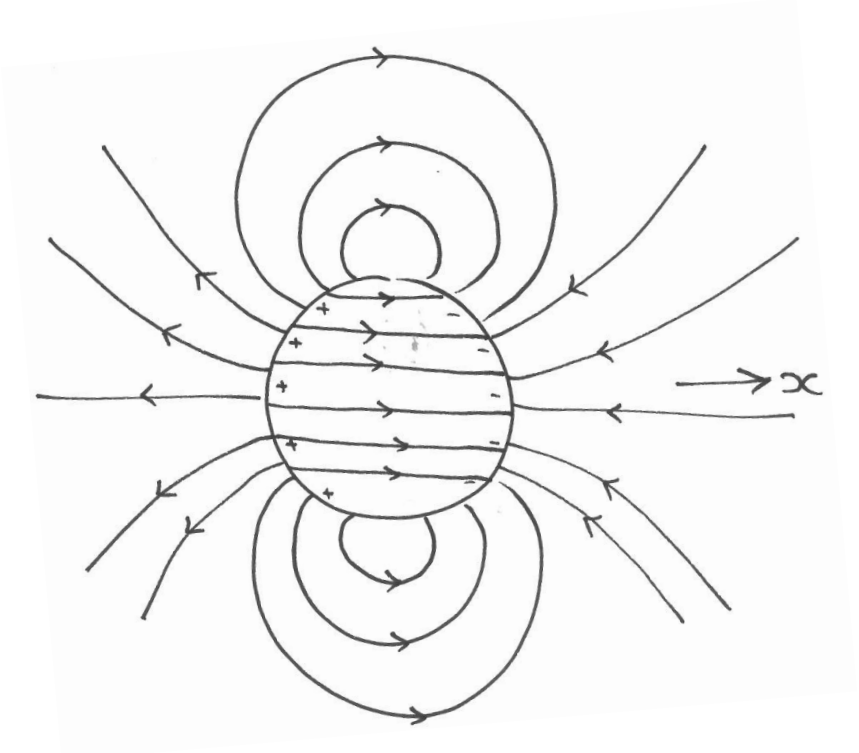


FIGURE 7–8

7.6 The macroscopic and local fields

Now we ask, what is really the electric field inside a dielectric? This is a tricky question for three reasons.

1. I can apply a field $\mathbf{E}_0$ by placing fixed charges outside a dielectric, for example on the plates of a capacitor. This field induces bound charges which create a depolarising field, which in turn modifies the bound charges, which changes the field inside the dielectric, and so on... How do I break out of this loop? In a sense this is resolved by introducing the *electric displacement field*, $\mathbf{D}$. (See section 6 on Auxiliary Fields.) Figure 7–9 shows the various components involved in the process of polarising the dielectric and establishing the "macroscopic field".

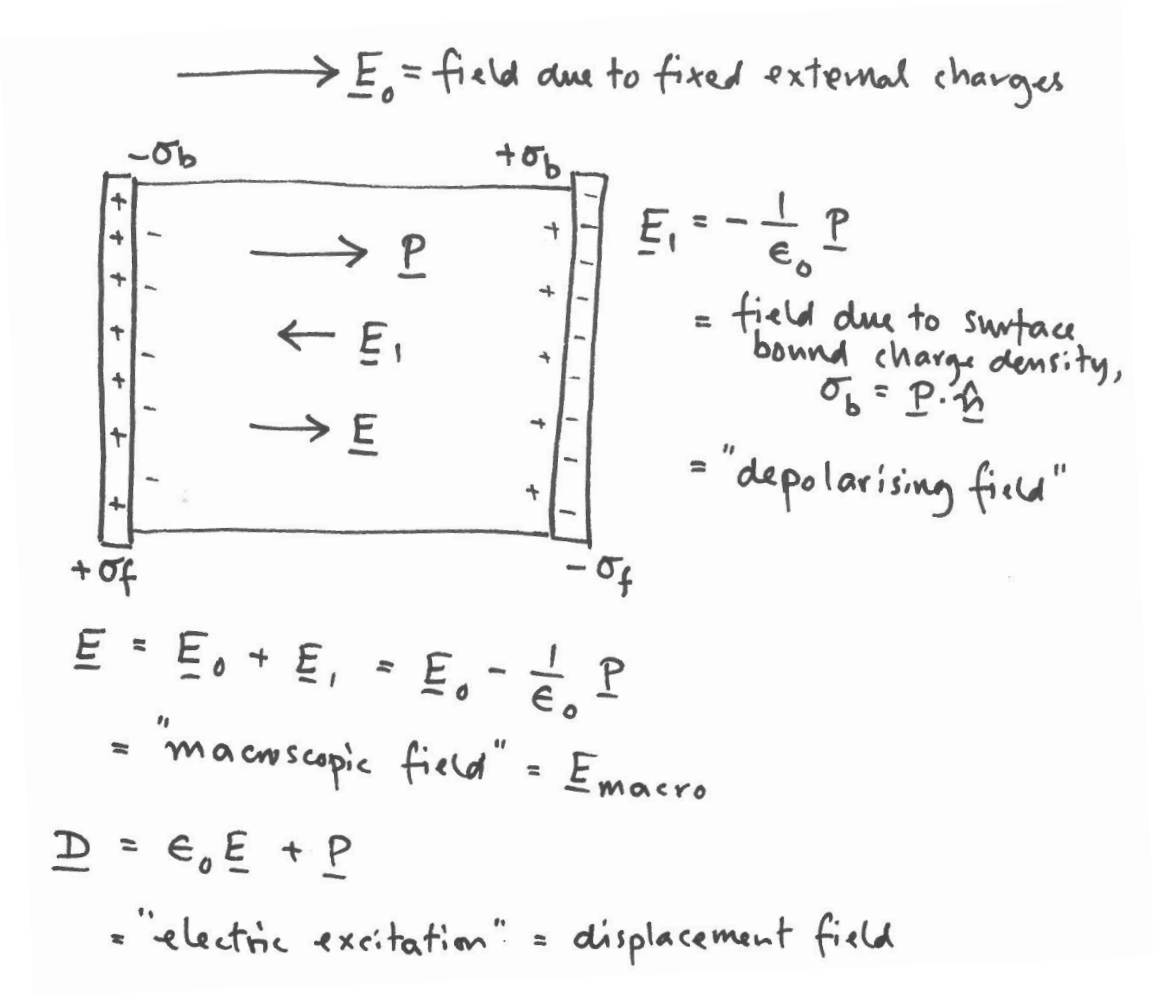


FIGURE 7–9. In this figure, $\mathbf{E}_1$ is the *depolarising field.* Do not confuse it with $\mathbf{E}_1$ in subsection 7.7, below, in which successive approximations are labelled $\mathbf{E}_0$, $\mathbf{E}_1$, $\mathbf{E}_2$...

2. The actual field inside matter is obviously very complicated and rapidly varying as you would find if you wandered about with a test charge between the atoms. Really we want some kind of *average* field or *macroscopic field*, $\mathbf{E}_{\text{macro}}$, averaged over maybe some thousands of atoms so it still varies, but smoothly, from place to place. Now,

$\mathbf{E}_{\text{macro}}$ is the total field inside the dielectric, that is, the field due to external fixed charges plus the field due to any bound charge caused by polarisation (see figure 7–9). We will just write $\mathbf{E} = \mathbf{E}_{\text{macro}}$ for this total field. This is the field as I have written it in equation (7.2.1).

Often we find ourselves dealing with *linear dielectrics* in which the polarisation (dipole moment per unit volume) is linearly related to the macroscopic field,

$$\mathbf{P} = \epsilon_0 \chi_e \mathbf{E} \tag{7.6.1}$$

and the constant of proportionality, χ_e, is the dimensionless *electric susceptibility*. It is also common to define the *permittivity* of a linear dielectric as[†]

$$\epsilon = \epsilon_0 \left(1 + \chi_e\right) \tag{7.6.2}$$

and the *relative permittivity* or *dielectric constant* (often given the symbol κ),

$$\epsilon_r = 1 + \chi_e = \frac{\epsilon}{\epsilon_0} \tag{7.6.3}$$

3. A third complication is that the field that an atom actually experiences, and which causes it to polarise is <u>not</u> the average, macroscopic field $\mathbf{E}$ inside the dielectric. The atom actually sees a *local field*, $\mathbf{E}_{\text{local}}$, not necessarily equal to $\mathbf{E}_{\text{macro}}$. So two questions arise.

   3.1 How does the *local* field depend on the macroscopic field, and how does it relate to the polarisation $\mathbf{P}$? (The "Lorentz relation".)

   3.2 How does the permittivity of a dielectric depend on the polarisabilty, α, of its atoms? (The Clausius–Mossotti relation.)

   These are very significant questions because they address how what is going on at the atomic scale influences our continuum field picture of electromagnetism.

We now answer these two questions. Let us suppose that we think of an atom as occupying a spherical hole that it has carved out of the dielectric. So the first question is, what is the electric field within a cavity in a polarised dielectric? Actually it depends on the shape of the cavity. We consider three cases.

1. If it is a narrow slot, or coin-shape, lying parallel to the field, then looking end on as in figure 7–10, we know that the integral around the dotted path must be zero since $\boldsymbol{\nabla} \times \mathbf{E} = \mathbf{0}$, or

$$\oint \mathbf{E} \cdot \mathrm{d}\boldsymbol{\ell} = 0$$

[†] In vacuum even if $\mathbf{E} \neq \mathbf{0}$, $\mathbf{P} = \mathbf{0}$ and hence χ_e must be zero in the vacuum. This is why you sometimes read that ϵ_0 is the "permittivity of the vacuum", or of free space. But it's not really a very useful statement. $1/4\pi\epsilon_0$ is really there just to get the SI units right.

The field along the downward leg must equal minus the field along the upward leg so considering that the line senses are opposite the two contributions will sum to zero. So the field inside the slot is the same as outside,

$$\mathbf{E}_{\text{slot}} = \mathbf{E} \qquad \longleftarrow \text{ long slot parallel to } \mathbf{E}$$

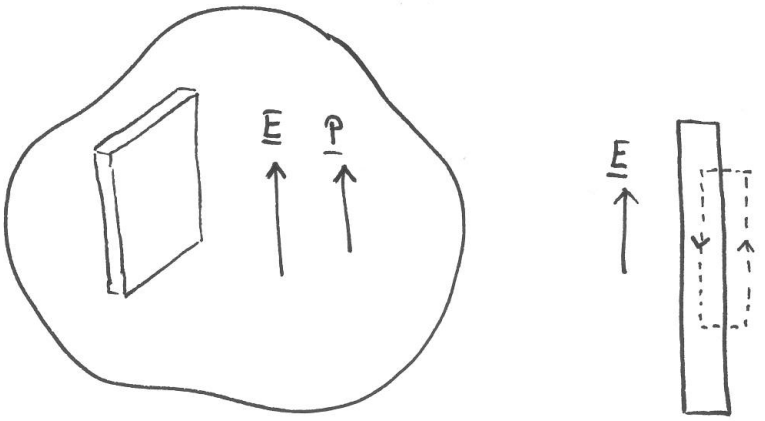


FIGURE 7–10

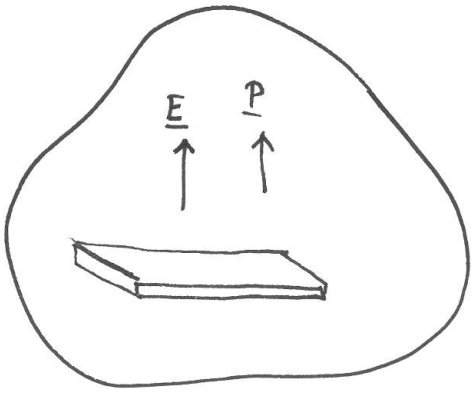


FIGURE 7–11

2. If we have a slot orientated so that its short edge is parallel to the field as in figure 7–11 then we make the following observations. Effectively excavating such a cavity divides the dielectric into two: creating two new surfaces with normal parallel to **E**. Therefore bound charges now appear in the same amounts that are already on the external faces of the dielectric as illustrated in figure 7–12. To find the field in the slot we construct a gaussian pillbox as shown and we sum the two contributions of the flux: $\mathbf{E}_{\text{slot}}$ *entering* from below and thus contributing to the flux *leaving* the pillbox with a minus sign; and **E** leaving the top surface. We get,

$$A\,(E - E_{\text{slot}}) = -\frac{1}{\epsilon_0}\,\sigma_{\text{b}} A$$

if A is the area of the pillbox surfaces, and the bound charge is negative as our gaussian pillbox surrounds the top internal surface. Therefore

$$E - E_{\text{slot}} = -\frac{1}{\epsilon_0}\,\sigma_{\text{b}} = -\frac{1}{\epsilon_0}\,P$$

because

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

since $\mathbf{P}$ is parallel to $\hat{\mathbf{n}}$. This means that the electric field inside this cavity is *enhanced* by polarisation to

$$\mathbf{E}_{\text{slot}} = \mathbf{E} + \frac{1}{\epsilon_0}\mathbf{P} = \frac{1}{\epsilon_0}\mathbf{D} \qquad \longleftarrow \text{ long slot perpendicular to } \mathbf{E}$$

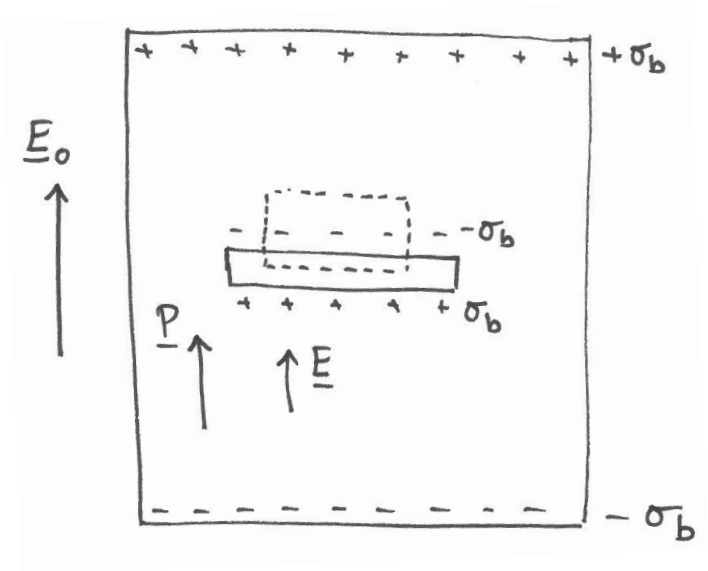


FIGURE 7–12

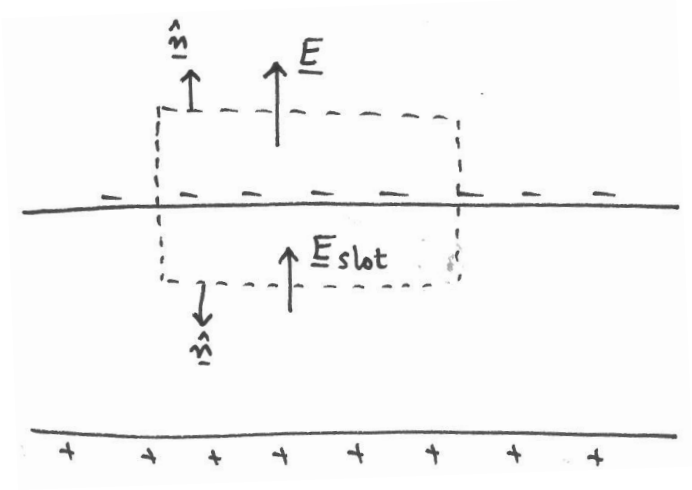


FIGURE 7–13

This is a useful way of visualising what is meant by the displacement field, or electric excitation, $\mathbf{D}$, within a dielectric. *It is ϵ_0 times the electric field in a coin-shaped cavity whose flat faces are perpendicular to the polarisation vector.* We will see below that the source of $\mathbf{D}$ is the free charge, while as you know the source of $\mathbf{E}$ is the total charge. Hence looking at figure 7–13 you can see that there is a flux of $\mathbf{E}$ through the gaussian pillbox by virtue of the bound charge distributed over the exposed surfaces of the dielectric. However there is no *source* of $\mathbf{D}$ and hence the $\mathbf{D}$ field lines must be continuous across the exposed surfaces. Hence the $\mathbf{D}$ field is the the same inside the cavity as in the bulk of the dielectric, $\mathbf{D} = \epsilon_0\mathbf{E} + \mathbf{P} = \epsilon_0\mathbf{E}_{\text{slot}}$.

3. Now we do the spherical cavity, which approximates best to the environment of a roughly spherical atom or molecule. Put another way, the local field inside a spherical

cavity is the appropriate field that causes an atom to polarise and so by reference to equation (7.3.2) we should write

$$\mathbf{p} = \alpha\, \mathbf{E}_{\text{local}}$$

and $\mathbf{E}_{\text{local}} = \mathbf{E}_{\text{hole}}$, the field inside a spherical hole. It turns out to be in between the two values for the slots. We imagine a thought experiment in which we remove a spherical plug of polarised dielectric leaving behind a hole, but without allowing any bound charge to move. What's the field in the hole left behind? Well, the field in the plug before we removed it is the macroscopic field, $\mathbf{E}$. And by the principle of superposition this must be the the field in the hole left behind plus the field in the plug once we've removed it and carried it into the vacuum away from the dielectric. In this way,

$$\mathbf{E} = \mathbf{E}_{\text{hole}} + \mathbf{E}_{\text{plug}}$$

But we know the field inside a ball of uniformly polarised dielectric, it's constant and equal to $-\mathbf{P}/3\epsilon_0$ according to equation (7.5.2). So we have

$$\begin{aligned} \mathbf{E}_{\text{hole}} &= \mathbf{E} - \mathbf{E}_{\text{plug}} \\ &= \mathbf{E} + \frac{1}{3\epsilon_0}\mathbf{P} \qquad \longleftarrow \text{spherical cavity} \end{aligned}$$

So an atom in a dielectric doesn't see the field due to external fixed charges, $\mathbf{E}_0$, and it doesn't see the macroscopic field, $\mathbf{E}$; it sees a *local field* equal to

$$\mathbf{E}_{\text{local}} = \mathbf{E} + \frac{1}{3\epsilon_0}\mathbf{P}$$

This is called the Lorentz relation, and is true for liquid dielectrics and for crystalline dielectrics that have cubic crystal point symmetry. (You'll find a good discussion of this in Charles Kittel's *Introduction to Solid State Physics*, 8th edition, John Wiley and Sons, 2004, chapter 16.)

That's answered the first question. The second is how the dielectric constant relates to the polarisability. We expect the dipole moment per atom to be $\alpha\mathbf{E}_{\text{local}}$. If there are N molecules per unit volume, the dipole moment per unit volume—the polarisation—is

$$\mathbf{P} = N\alpha\mathbf{E}_{\text{local}} = N\alpha\left(\mathbf{E} + \frac{1}{3\epsilon_0}\mathbf{P}\right)$$

This formula reflects the dependence of $\mathbf{P}$ on itself. It's how we break the infinite loop I mentioned at point 1 in the beginning of this section 7.6—the field depends on the polarisation and the polarisation depends on the field. The problem is one of *self consistency* often encountered in field theories† and this one is trivially easy to resolve: I just make $\mathbf{P}$ the subject of the above equation,

$$\mathbf{P} = \frac{N\alpha}{1 - \frac{1}{3\epsilon_0}}\,\mathbf{E} = (\epsilon_r - 1)\,\epsilon_0\mathbf{E}$$

† For example, John C. Slater's "self consistent field" in quantum mechanics.

where, using (7.6.1–3), $\mathbf{P} = \epsilon_0 \chi_e \mathbf{E}$ and $\chi_e = \epsilon_r - 1$. Solving for α we find

$$\alpha = \frac{3\epsilon_0}{N} \frac{\epsilon_r - 1}{\epsilon_r + 2}$$

which is the famous† Clausius–Mossotti equation. It's significant because its a relation between an atomic property, the atomic polarisability, and a continuum property, the dielectric constant. It's only an approximation because an atom isn't really embedded in a spherical hole in a dielectric. To test it people have measured the electronic polarisability of a gas of atoms or molecules and compared that with the dielectric constant of the liquid. For example,

substance	predicted ϵ_r	observed ϵ_r
CS_2	2.76	2.64
O_2	1.51	1.51
CCl_4	2.54	2.24
Ar	1.52	1.54

7.7 Ball of dielectric in a uniform electric field

Let's step back half a page to the *self consistency problem*. There's a nice example of this in Griffiths's textbook. Suppose I place a ball of a linear dielectric substance into a uniform electric field, $\mathbf{E}_0$. What is the macroscopic field and polarisation inside the ball? Well, as a naive first approximation, we might suppose the polarisation is just,

$$\mathbf{P}_0 = \epsilon_0 \chi_e \mathbf{E}_0$$

But the polarisation $\mathbf{P}_0$ gives rise to an electric field,

$$\mathbf{E}_1 = -\frac{1}{3\epsilon_0}\mathbf{P}_0 = -\frac{1}{3}\chi_e \mathbf{E}_0$$

according to equation (7.5.2). So *now* the field is $\mathbf{E}_0 + \mathbf{E}_1$; and as I pointed out in point 1 on page 13 (subsection 7.6), this process needs to be continued indefinitely—the field produces a polarisation which gives rise to a new field, which modifies the polarisation for ever. Of course this sequence of events occurs very rapdily and the substance soon solves the self consistency problem itself. But how do we solve it? In this case we can. $\mathbf{E}_1$ produces a polarisation,

$$\mathbf{P}_1 = \epsilon_0 \chi_e \mathbf{E}_1 = -\frac{1}{3}\epsilon_0 \chi_e^2 \mathbf{E}_0$$

† although not so famous that lots of textbooks spell Mossotti wrongly. Also in some books it appears differently if the authors use the definition of α using $p = \epsilon_0 \alpha E$ as I mentioned in a footnote on page 5.

At the next iteration,

$$\mathbf{E}_2 = -\frac{1}{3\epsilon_0}\mathbf{P}_1 = \frac{1}{9}\chi_e^2\mathbf{E}_0$$

You can see the pattern emerging. It looks as if at the n^{th} iteration of this procedure,

$$\mathbf{E}_n = \left(-\frac{1}{3}\chi_e\right)^n \mathbf{E}_0$$

so the self consistent electric field within the dielectric ball due to the external electric field, $\mathbf{E}_0$ is,

$$\mathbf{E} = \mathbf{E}_0 + \mathbf{E}_1 + \mathbf{E}_2 + \ldots = \mathbf{E}_0 \sum_{n=0}^{\infty} \left(-\frac{1}{3}\chi_e\right)^n$$

The geometric series can be summed exactly,

$$\sum_{n=0}^{\infty} x^n = \frac{1}{1-x}$$

Eventually, the electric field inside the dielectric is,

$$\mathbf{E} = \frac{1}{1+\frac{1}{3}\chi_e}\mathbf{E}_0$$

If I use the formula (7.6.3), $\epsilon_r = 1 + \chi_e$, then this leads to,

$$\mathbf{E} = \frac{3}{\epsilon_r + 2}\mathbf{E}_0$$

Note that the electric field *inside the ball* is uniform. So is the polarisation. Using (7.6.1),

$$\begin{aligned}\mathbf{P} &= \epsilon_0\chi_e\frac{3}{\epsilon_r+2}\mathbf{E}_0 \\ &= \frac{3}{2}\frac{\epsilon-\epsilon_0}{1+\epsilon_r}\mathbf{E}_0\end{aligned}$$

using (7.6.2). The fields inside and outside the ball are sketched in figure 7–14.

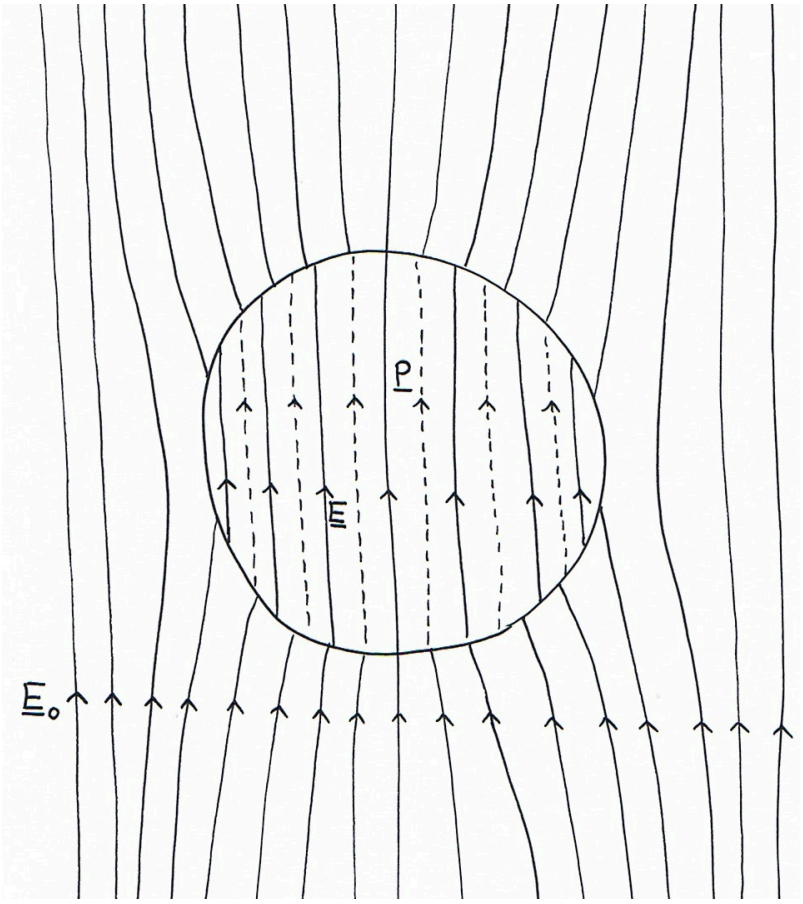


FIGURE 7–14

7.8 The electric displacement—Gauss's law in matter‡

We have seen that the total charge density is the free charge density plus the bound charge density,

$$\rho = \rho_{\text{free}} + \rho_{\text{b}}$$

ρ_{free} is the charge density that we put, say, on the plates of a capacitor to create a field $\mathbf{E}_0$. ρ_{b} is the volume charge density resulting from redistribution of slightly mobile charges in the dielectric material as a response to the field, $\mathbf{E}_0$. The integral of all the bound charge, volume and surface, over the dielectric is zero as no charge has been created or destroyed—just moved about a bit (see Problems Set 6, C6.4)

Now Gauss's law is true even in matter. The electric field within matter is taken to be the *averaged* or *macroscopic* field, $\mathbf{E}$ and

$$\epsilon_0 \boldsymbol{\nabla} \cdot \mathbf{E} = \rho = \rho_{\text{free}} + \rho_{\text{b}} = \rho_{\text{free}} - \boldsymbol{\nabla} \cdot \mathbf{P}$$

using (7.4.4). I can re-write this as

$$\boldsymbol{\nabla} \cdot (\epsilon_0 \mathbf{E} + \mathbf{P}) = \rho_{\text{free}}$$

‡ In addition to this section, please read section 6, Auxiliary Fields.

and define the *electric displacement*

$$\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P} \tag{7.7.1}$$

Now, Gauss's law looks like

$$\boldsymbol{\nabla} \cdot \mathbf{D} = \rho_{\text{free}}$$

in terms only of the *free* charge density, which is the charge that we control. Note that this is valid in *all matter* and also the vacuum, where $\mathbf{P} = \mathbf{0}$, $\rho_{\text{b}} = 0$ and $\mathbf{D} = \epsilon_0 \mathbf{E}$. In a linear dielectric only we also have the additional constitutive relation,

$$\mathbf{P} = (\epsilon_r - 1)\ \epsilon_0\, \mathbf{E}$$

where the dielectric constant (relative permittivity, ϵ_r) is a property of the dielectric material. So in linear dielectrics we have

$$\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P} = \epsilon_0 \mathbf{E} + (\epsilon_r - 1)\ \epsilon_0\, \mathbf{E} = \epsilon_r \epsilon_0 \mathbf{E}$$

Commonly you only know the free charge and not the bound charge. But you can use all the tricks you have learned for spheres, sheets, lines to solve Gauss's law

$$\boldsymbol{\nabla} \cdot \mathbf{D} = \rho_{\text{free}}$$

and then if you know your medium is linear you get the macroscopic electric field,

$$\mathbf{E} = \frac{1}{\epsilon_r \epsilon_0}\, \mathbf{D}$$

Beware! though that you cannot take the analogy with $\mathbf{E}$ any further: you may not assume that $\mathbf{D}$ is irrotational, nor that it is the derivative of any scalar field. Generally $\boldsymbol{\nabla} \times \mathbf{D} \neq \mathbf{0}$. To see that, just run a little loop from the vacuum outside through the surface of a dielectric and back into the vacuum to close it.

Appendix—Langevin equation (after Grant and Phillips)

Imagine a unit volume of gas or liquid polarisable dielectric. It contains N molecular dipoles randomly orientated. If a number $\mathrm{d}N$ are oriented in a range of angles between θ and $\theta + \mathrm{d}\theta$ with respect to some fixed direction, then,

$$\frac{\mathrm{d}N}{N} = \frac{1}{2} \sin\theta\, \mathrm{d}\theta$$

If now a uniform electric field $\mathbf{E}$ is applied, the dipoles will experience a torque that tends to orientate them along the field. Each dipole then possesses a potential energy

$$U = -\mathbf{p} \cdot \mathbf{E}$$

If we now regard θ as the angle of the dipole with respect to the applied field then the number of dipoles per unit volume orientated within the range between θ and $\theta + \mathrm{d}\theta$ is

$$\mathrm{d}N = \text{constant} \times \exp\left(\mathbf{p} \cdot \mathbf{E}/kT\right)$$

where T is absolute temperature and k is the Boltzmann constant. If we set

$$x = \frac{pE}{kT}$$

then

$$N = \text{constant} \times \int_0^\pi e^{x\cos\theta} \sin\theta \, \mathrm{d}\theta$$

which gives you the constant. All those dipoles orientated within the range between θ and $\theta + \mathrm{d}\theta$ amount to a dipole moment per unit volume of

$$\mathrm{d}P = p \, \mathrm{d}N \, \cos\theta = pN \frac{e^{x\cos\theta} \sin\theta \cos\theta \, \mathrm{d}\theta}{\int_0^\pi e^{x\cos\theta} \sin\theta \, \mathrm{d}\theta}$$

so integrating over all angles from $\theta = 0$ to $\theta = \pi$,

$$P = pN \frac{\int_0^\pi e^{x\cos\theta} \sin\theta \cos\theta \, \mathrm{d}\theta}{\int_0^\pi e^{x\cos\theta} \sin\theta \, \mathrm{d}\theta}$$

If you do the integrals you'll find,

$$P = pN \left(\coth x - \frac{1}{x} \right) = \left(\coth \frac{pE}{kT} - \frac{kT}{pE} \right)$$

This is the Langevin equation. If you think about some typical values of fields, dipole moments and temperatures you'll convince yourself that under realistic circumstances it is reasonable to assert that

$$P = \frac{Np^2}{3kT} \, E$$

which is equation (7.3.1) on page 3. So $P \propto E$ and the polarisability of a single molecular dipole is,

$$\alpha = \frac{p^2}{3kT}$$

8. Magnetic fields in matter

8.1. Magnetisation

If you place matter in a magnetic field you may *magnetise* it. That is, it will acquire a magnetic dipole moment per unit volume, $\mathbf{M}$, called its magnetisation. In contrast to electric polarisation the magnetisation may point in the direction of the applied field or against it. Or, the matter may become permanently magnetised. Hence we distinguish three classes of behaviour (for further details, see the excellent section 11.1 in Edward Purcell and David Morin, *Electricity and Magnetism*, third edition, 2013, Cambridge University Press).

1. Diamagnetism: magnetisation opposes the applied field,
$$\hat{\mathbf{M}} = -\hat{\mathbf{B}}$$
This is a *weak* effect but all matter is diamagnetic. Superconductors are *perfect* diamagnets: they *reject* magnetic field.
2. Paramagnetism: magnetisation is parallel to the applied magnetic field,
$$\hat{\mathbf{M}} = \hat{\mathbf{B}}$$
This is a *strong* effect, but does not occur in all matter, but usually only in substances whose atoms have an odd number of electrons.
3. Ferromagnetism: these substances retain a magnetic field after the external field is removed. In some, the internal field may be switched by varying the applied field.

All magnetic fields are caused by electric currents, even in matter. In that case the currents are provided by orbiting electrons, or their intrinsic spin. The origins of magnetism fall outside the scope of classical electromagnetism and cannot be plausibly discussed without quantum mechanics.

8.2. Ferromagnets

In the case of ferromagnets we take it that the magnetic moment per unit volume, or magnetisation, $\mathbf{M}$, is fixed in the material and independent of external fields. The auxiliary field, the "magnetic excitation", $\mathbf{H}$, is defined by

$$\mathbf{H} = \frac{1}{\mu_0}\mathbf{B} - \mathbf{M} \tag{8.2.1}$$

and it has the property,

$$\boldsymbol{\nabla} \times \mathbf{H} = \mathbf{J}_{\text{free}}$$

where $\mathbf{J}_{\text{free}}$ is the free current density. In the case of a permanently magnetised material, there is no free current and so,

$$\begin{aligned} \boldsymbol{\nabla} \times \mathbf{H} &= \mathbf{0} \\ \boldsymbol{\nabla} \times \mathbf{B} &= \mu_0 \boldsymbol{\nabla} \times \mathbf{M} \neq \mathbf{0} \end{aligned} \tag{8.2.2}$$

Because $\boldsymbol{\nabla}\cdot\mathbf{B}=0$ it follows from (8.2.1) that

$$\boldsymbol{\nabla}\cdot\mathbf{H}=-\boldsymbol{\nabla}\cdot\mathbf{M} \tag{8.2.3}$$

and because $\mathbf{H}$ in this case is irrotational (8.2.2), $\mathbf{H}$ must be derivable from the gradient of a scalar potential. In analogy with

$$\mathbf{E}=-\boldsymbol{\nabla}V$$

in electrostatics, in magnetostatics, and *in a ferromagnet*, we can write,

$$\mathbf{H}=-\boldsymbol{\nabla}V_m\ ;\qquad \text{iff } \boldsymbol{\nabla}\times\mathbf{H}=\mathbf{0}$$

and in analogy with Gauss's law, we will have,

$$\boldsymbol{\nabla}\cdot\mathbf{H}=-\boldsymbol{\nabla}\cdot\mathbf{M}=\rho_m$$

where ρ_m is the *density of magnetic sources*—north and south poles, if you like. Please take note of the remarkable parallel of the following development with the equivalent arguments in electrostatics in matter (subsection 7.4). In exact analogy with electrostatics, the scalar potential can be expressed in terms of *volume* and *surface* pole densities, see section 7, equation (7.4.2),

$$V_m(\mathbf{r}_0)=\frac{1}{4\pi}\oint_S\frac{\mathbf{M}\cdot\mathrm{d}\mathbf{a}_1}{r}-\frac{1}{4\pi}\int_\Omega\frac{\boldsymbol{\nabla}\cdot\mathbf{M}}{r}\mathrm{d}\tau_1$$

This formula is only useful in the present case (8.2.2) in which the $\mathbf{H}$-field is irrotational so that a magnetic *scalar* potential exists. We will see below, however, that *in the general case*, we can develop the vector potential in terms of surface and volume bound *current* densities in the same way that we expressed the electric potential in equation (7.4.5) in terms of surface and volume bound *charge* densities. The vector potential is, in view of (5.5.6),†

$$\begin{aligned}\mathbf{A}(\mathbf{r}_0)&=\frac{\mu_0}{4\pi}\int_\Omega\mathbf{M}\times\boldsymbol{\nabla}_1\left(\frac{1}{r}\right)\mathrm{d}\tau_1\\&=\frac{\mu_0}{4\pi}\int_\Omega\frac{\boldsymbol{\nabla}_1\times\mathbf{M}}{r}\mathrm{d}\tau_1+\frac{\mu_0}{4\pi}\oint_S\frac{\mathbf{M}\times\hat{\mathbf{n}}}{r}\mathrm{d}a_1\end{aligned}$$

We can therefore identify

$$\boldsymbol{\nabla}\times\mathbf{M}=\mathbf{J}_\mathrm{b}$$

† To go from the first to the second line you need two identities, the fifth product rule on page 6 of section 1 (see p. 4, below) and,

$$\int_\Omega\boldsymbol{\nabla}\times\mathbf{u}\,\mathrm{d}\tau=-\oint_S\mathbf{u}\times\mathrm{d}\mathbf{a}$$

as a *bound volume current density*, and

$$\mathbf{M} \times \hat{\mathbf{n}} = \mathbf{K}_\mathrm{b}$$

as a *bound surface current density*. If the permanent magnet is uniformly magnetised, then $\mathbf{M}$ is constant and independent of $\mathbf{r}$,

$$\mathbf{J}_\mathrm{b} = \boldsymbol{\nabla} \times \mathbf{M} = \mathbf{0}$$

and there are only surface currents. If a cylinder of ferromagnet is uniformly magnetised along its axis, then there are surface currents, or equivalently a density of poles at either end. These are the *sources* of $\mathbf{H}$.

Outside the ferromagnet, the lines of $\mathbf{B}$ and $\mathbf{H}$ coincide, since,

$$\mathbf{B} = \mu_0 \mathbf{H}$$

Inside the ferromagnet $\mathbf{B}/\mu_0$ and $\mathbf{H}$ differ by $\mathbf{M}$, and $\mathbf{B} = \mu_0(\mathbf{H} + \mathbf{M})$ is opposite in direction to $\mathbf{H}$. This is possible because $\mathbf{M}$ is opposite to $\mathbf{H}$: you can see this from (8.2.3) which implies that $\mathbf{H}$-field lines point away from a north pole, whereas $\mathbf{M}$-field line point into a north pole. You can see that $\mathbf{H}$ must be opposite to $\mathbf{B}$ *inside* the ferromagnet, because they are parallel *outside* the magnet, yet if you integrate $\mathbf{H}$ around any loop, including one that enters and leaves the magnet along its side, the result must be zero, since $\boldsymbol{\nabla} \times \mathbf{H} = \mathbf{0}$.

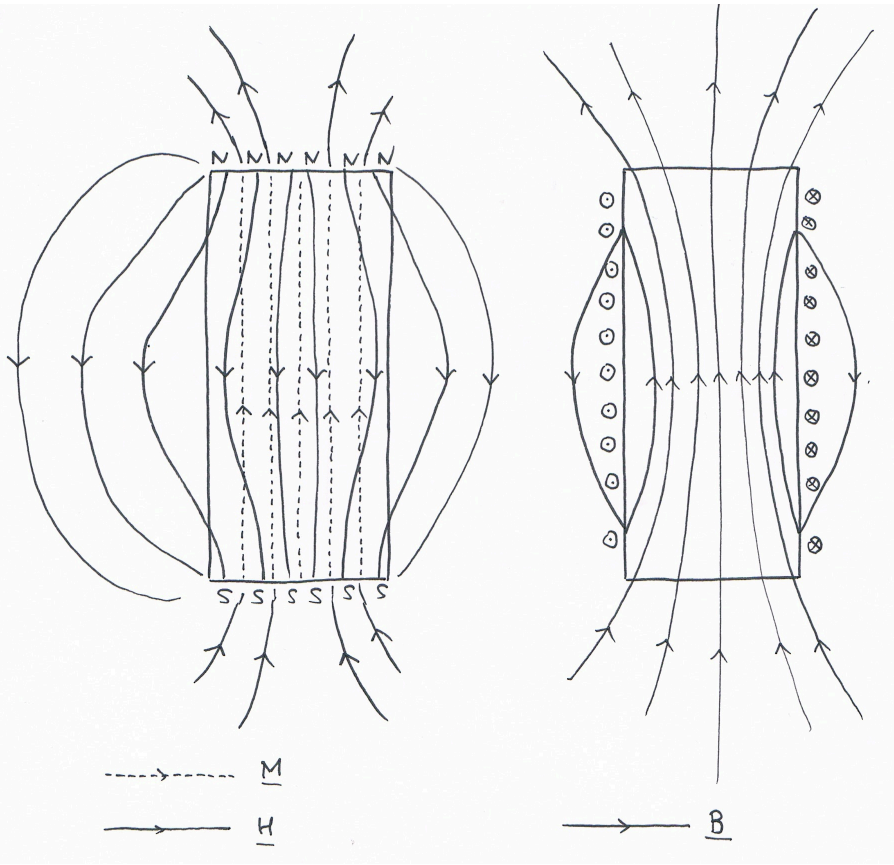


FIGURE 8–1

Field lines inside and outside a permanent bar magnet are sketched in figure 8–1. Note that the **B**-field is *identical* to that of a solenoid whose current in the coil amounts to an equivalent surface bound current density, $\mathbf{K}_\mathrm{b}$. The symbols to the left and to the right in the right hand image are supposed to represent the current in the coils of a solenoid that would produce the equivalent magnetic field. Note that **H**-field lines emerge *out* of north poles and *into* south poles in the way that **E**-field lines emerge from positive charges and converge on negative charges—compare $\boldsymbol{\nabla}\cdot\mathbf{H}=\rho_m$ with $\boldsymbol{\nabla}\cdot\mathbf{E}=\rho/\epsilon_0$ (see Problems Class 6.5). Similarly magnetisation points from south to north as polarisation points from negative to positive charge. Outside the magnet field lines flow from north to south. The earth behaves in some respects like a bar magnet; but the North Pole was so called because it attracts the north pole of a compass. The earth's North Pole is actually a south pole.

8.3. Bound currents

If a body is magnetised then there must be little circulating currents. These are characterised by a distribution of magnetic moments. If the magnetic moment per unit volume, or *magnetisation* at a source point, $\mathbf{r}_1$, is $\mathbf{M}(\mathbf{r}_1)$, then in the element of volume, $\mathrm{d}\tau_1$, there exists a magnetic moment,

$$\mathrm{d}\mathbf{m}=\mathbf{M}\,\mathrm{d}\tau_1$$

Then according to equation (5.5.6), the vector potential at a field point, $\mathbf{r}_0$, due to the complete distribution of magnetic moments is,

$$\mathbf{A}(\mathbf{r}_0)=\frac{\mu_0}{4\pi}\int\mathbf{M}(\mathbf{r}_1)\times\boldsymbol{\nabla}_1\frac{1}{r}\mathrm{d}\tau_1$$

where $\mathbf{r}$ is the connecting vector. I will use the fifth product rule on page 6 of section 1, namely,

$$\boldsymbol{\nabla}\times(f\mathbf{u})=f\,(\boldsymbol{\nabla}\times\mathbf{u})-\mathbf{u}\times(\boldsymbol{\nabla}f)$$

to write

$$\mathbf{M}\times\boldsymbol{\nabla}\frac{1}{r}=\frac{1}{r}\boldsymbol{\nabla}\times\mathbf{M}-\boldsymbol{\nabla}\times\frac{\mathbf{M}}{r} \tag{8.3.1}$$

Now I need to use a corollary of the divergence theorem (`https://en.wikipedia.org/wiki/Divergence_theorem`),

$$\int_\Omega\boldsymbol{\nabla}\times\mathbf{u}\,\mathrm{d}\tau=-\oint_S\mathbf{u}\times\hat{\mathbf{n}}\mathrm{d}a$$

Therefore when I integrate the last term in (8.3.1) and take the boundary out to infinity where there is no magnetisation, I will get zero. This leaves,

$$\mathbf{A}(\mathbf{r}_0)=\frac{\mu_0}{4\pi}\int\frac{\boldsymbol{\nabla}_1\times\mathbf{M}(\mathbf{r}_1)}{r}\mathrm{d}\tau_1 \tag{8.3.2}$$

Compare this with equation (5.3.5)

$$\mathbf{A}(\mathbf{r}_0)=\frac{\mu_0}{4\pi}\int\frac{\mathbf{J}(\mathbf{r}_1)}{r}\mathrm{d}\tau_1 \tag{5.3.5}$$

It is clear that there is a *formal equivalence* between $\boldsymbol{\nabla}\times\mathbf{M}$ and a current density. These are not *free currents* in the sense that we control them in electric circuits or particle beams; instead in exact analogy with the bound charges in electrostatics we define

$$\boxed{\mathbf{J}_\mathrm{b} = \boldsymbol{\nabla}\times\mathbf{M}} \tag{8.3.3}$$

as the *bound current density* in magnetism.

Just as in electrostatics, there are also *surface bound currents*. Think of a slab of magnetised matter of thickness h.

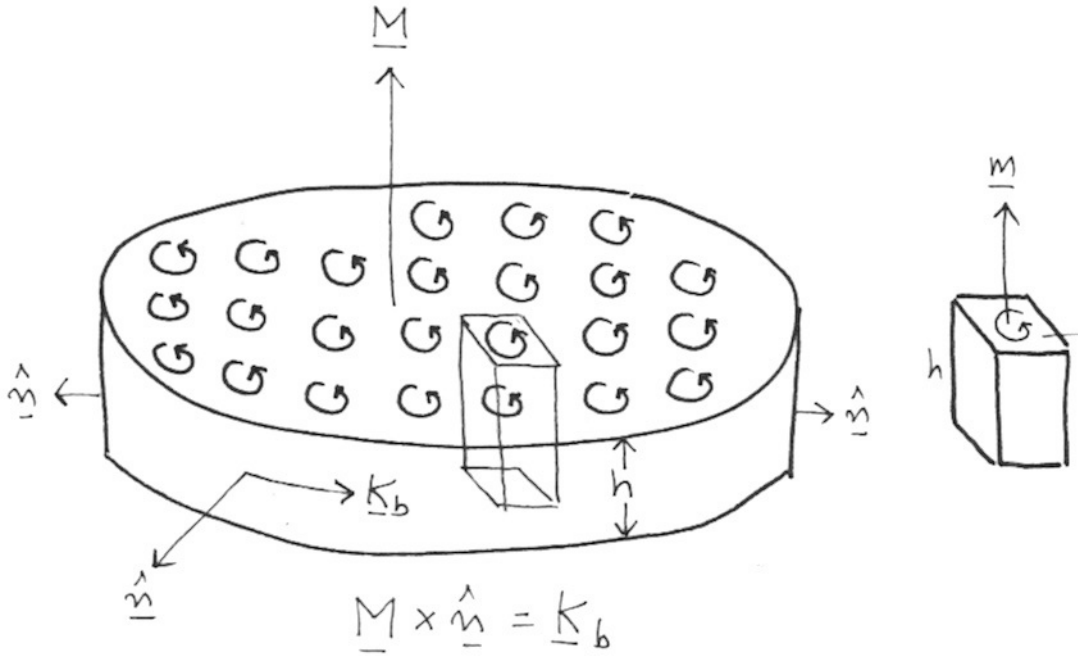


FIGURE 8–2

If $\mathbf{M}$ is uniform then within the slab all the currents cancel and we are left with a ribbon of current around the edge (it's the same as the argument leading to Stokes's theorem—see figure 1–5), figure 8–2. This is a surface current density (amp per unit thickness),

$$K_\mathrm{b} = \frac{I}{h}$$

If we regard each little loop of bound current, I, as contained within an oblong box with through-thickness, h, and area a, then its magnetic moment is $m = Ia$ and its magnetisation is its moment divided by the volume of the box. Since the magnetisation is uniform it's the same throughout the coin-shaped volume,

$$M = \frac{m}{ah} = \frac{Ia}{ah} = \frac{I}{h}$$

Hence,

$$K_\mathrm{b} = M$$

It is clear from figure 8–2 that $\mathbf{M}$, the normal to the edge of the "coin", $\hat{\mathbf{n}}$ and $\mathbf{K}_\mathrm{b}$ are mutually perpendicular and form a right-handed set in that order. It follows that the surface bound current vector must be,

$$\boxed{\mathbf{K}_\mathrm{b} = \mathbf{M} \times \hat{\mathbf{n}}} \tag{8.3.4}$$

We call this a *bound* current density because it's associated with little localised loops of current. No electron makes the trip right around the surface of the piece of magnetised matter. You can't get an electric shock from a fridge magnet.

The currents inside *don't* cancel if $\mathbf{M}$ is non uniform. In that case we consider two contiguous loops (let's make them square) in figure 8–3.

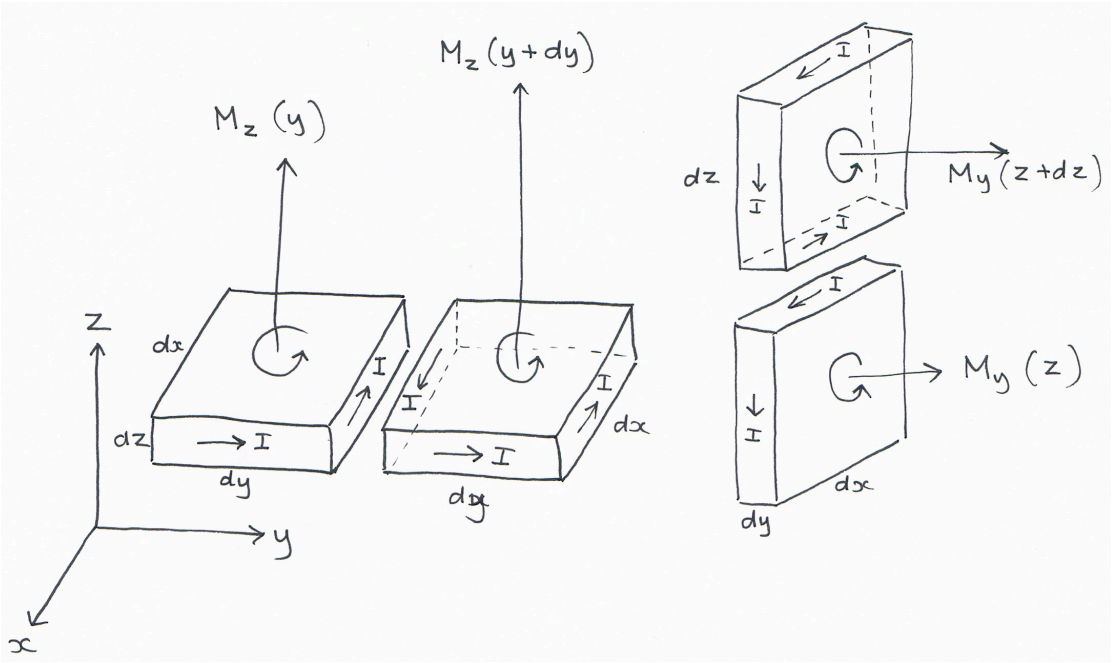


FIGURE 8–3

Look at figure 8–3. At the join between the two slabs on the left the currents do not exactly balance because the magnetic moment, m, of the left slab is smaller than the magnetic moment of the right slab. The magnetisation of one of the slabs is,

$$M = \frac{1}{\text{volume}}\, m = \frac{1}{\text{volume}}\, \text{area} \times I = \frac{1}{\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z}\mathrm{d}x\,\mathrm{d}y\,I$$

therefore $I = M\mathrm{d}z$. The net current in the x-direction at the join is,

$$\begin{aligned} I_x &= I(y+\mathrm{d}y) - I(y) \\ &= \big(M_z(y+\mathrm{d}y) - M_z(y)\big)\mathrm{d}z \\ &= \left(M_z(y) + \mathrm{d}y\frac{\mathrm{d}M_z}{\mathrm{d}y} - M_z(y)\right)\mathrm{d}z \\ &= \frac{\mathrm{d}M_z}{\mathrm{d}y}\,\mathrm{d}y\,\mathrm{d}z \end{aligned}$$

using a Taylor expansion on the third line. Dividing this by the area of a side face of the slab, we get the the x-component of the volume current density,

$$J_x = \frac{I_x}{\mathrm{d}y\,\mathrm{d}z} = \frac{\partial M_z}{\partial y}$$

If the magnetisation were also non uniform in the z-direction then looking at the two slabs to the right of figure 8–3, the magnetisation of the upper slab is larger, say, than the magnetisation of the lower slab; and we go through the same argument and get this additional contribution to the x-component of the current density,

$$-\frac{\partial M_y}{\partial z}$$

So in total,

$$J_x = \frac{\partial M_z}{\partial y} - \frac{\partial M_y}{\partial z}$$

You can construct J_y and J_z similarly, and you see that we have the components of a curl. We write the *bound volume current density* as,

$$\mathbf{J}_\mathrm{b} = \boldsymbol{\nabla} \times \mathbf{M}$$

This is equation (8.3.3) which we obtained mathematically earlier in this subsection.

8.4. The H-field

If we separate the total current density everywhere into free and bound, then we write,

$$\mathbf{J} = \mathbf{J}_\mathrm{free} + \mathbf{J}_\mathrm{b}$$

and $\mathbf{J}_\mathrm{free}$ is the current that we control in circuits, coils, electron beams and so on. Then Ampère's law is

$$\begin{aligned}\frac{1}{\mu_0}\,\boldsymbol{\nabla} \times \mathbf{B} &= \mathbf{J} \\ &= \mathbf{J}_\mathrm{free} + \mathbf{J}_\mathrm{b} \\ &= \mathbf{J}_\mathrm{free} + \boldsymbol{\nabla} \times \mathbf{M}\end{aligned}$$

using (8.3.3). Then,

$$\boldsymbol{\nabla} \times \left(\frac{1}{\mu_0}\,\mathbf{B} - \mathbf{M}\right) = \mathbf{J}_\mathrm{free}$$

and just as we wrote Gauss's law in terms of free charge and the electric displacement, $\mathbf{D}$; we define,†

$$\mathbf{H} = \frac{1}{\mu_0}\,\mathbf{B} - \mathbf{M} \tag{8.4.1}$$

† Compare this with

$$\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P} \tag{7.7.1}$$

and Ampère's law is,

$$\boldsymbol{\nabla} \times \mathbf{H} = \mathbf{J}_{\text{free}} \tag{8.4.2}$$

or in integral form,

$$\oint_{\Gamma} \mathbf{H} \cdot \mathrm{d}\boldsymbol{\ell} = I_{\text{f,enclosed}}$$

where $I_{\text{f,enclosed}}$ is the free current that links the loop Γ. Of course in a vacuum,

$$\mathbf{H} = \frac{1}{\mu_0} \mathbf{B}$$

H is a much more useful quantity in experiment than **D**. This is because we easily control the current in a solenoid, say; we really don't control free charge—we apply a voltage and that produces an **E**-field. So while **H** is usually the "magnetic field" of an experiment, *do not think* that (8.4.2) means "**H** is the field whose source is $\mathbf{J}_{\text{free}}$." Ampère's law completely determines **B** in magnetostatics because we know that $\boldsymbol{\nabla} \cdot \mathbf{B} = 0$ and therefore we know both the curl and the divergence of **B**: this is sufficient to determine **B** uniquely (see section 1.3). The divergence of **H** is not zero. In fact

$$\boldsymbol{\nabla} \cdot \mathbf{H} = -\boldsymbol{\nabla} \cdot \mathbf{M} \tag{8.2.3),}$$

and this only vanishes if **M** is uniform, or has some other special properties. So **H** is the solution to the coupled differential equations (8.4.2) and (8.2.3).

8.5. Linear magnetic substances

For diamagnetic and paramagnetic substances it is usually found that the magnetisation is proportional to the applied magnetic field. Conventionally the *magnetic susceptibility* is defined through **H**, not **B**, since it is **H** that is controlled in experiment,

$$\mathbf{M} = \chi_m \mathbf{H} \tag{8.5.1}$$

Compare this with the *electric susceptibility* that we defined in section 7.6. Because it is **E** that we control in the laboratory and not **D**, we relate the linear response of the polarisation to the electric field,

$$\mathbf{P} = \epsilon_0 \chi_e \mathbf{E} \tag{7.6.1}$$

and the electric susceptibility is kept dimensionless by the insertion of the electric permittivity of free space, ϵ_0. This would not have been necessary if we were defining the proportionality constant between **P** and **D** since these have the same dimensions (Coulomb per square metre). Therefore there is no need for a μ_0 to appear in (8.5.1) because **M** and **H** have the *same dimensions* (amp per metre). Then in analogy with equations (7.6.2–3) we use,

$$\begin{aligned} \mathbf{B} = \mu_0 (\mathbf{H} + \mathbf{M}) &= \mu_0 (1 + \chi_m) \mathbf{H} \\ &= \mu_0 \mu_r \mathbf{H} \\ &= \mu \mathbf{H} \end{aligned}$$

where $\mu = \mu_0\mu_r$ is called the *permeability* of a substance.

$$\mu_r = 1 + \chi_m = \frac{\mu}{\mu_0}$$

is the *relative permeability.* Compare this with the relative permittivity in (6.6.3), †

$$\epsilon_r = 1 + \chi_e = \frac{\epsilon}{\epsilon_0} \tag{7.6.3}$$

So, if we apply a magnetic field, $\mathbf{B}_0$, and we place in it a linear medium, we can find its magnetisation,

$$\mathbf{M} = \frac{\chi_m}{\mu}\,\mathbf{B}$$

The magnetic susceptibility may be either negative (diamagnetism) or positive (paramagnetism). This is illustrated in figure 8–4 and contrasted with the electrostatic case that we discussed in section 7.6. There is no analogue of paramagnetism in electrostatics, despite the term "paraelectric" which is just a particular form of dielectric. Paraelectrics and ferroelectrics belong in a course of solid state physics and will not concern us here.

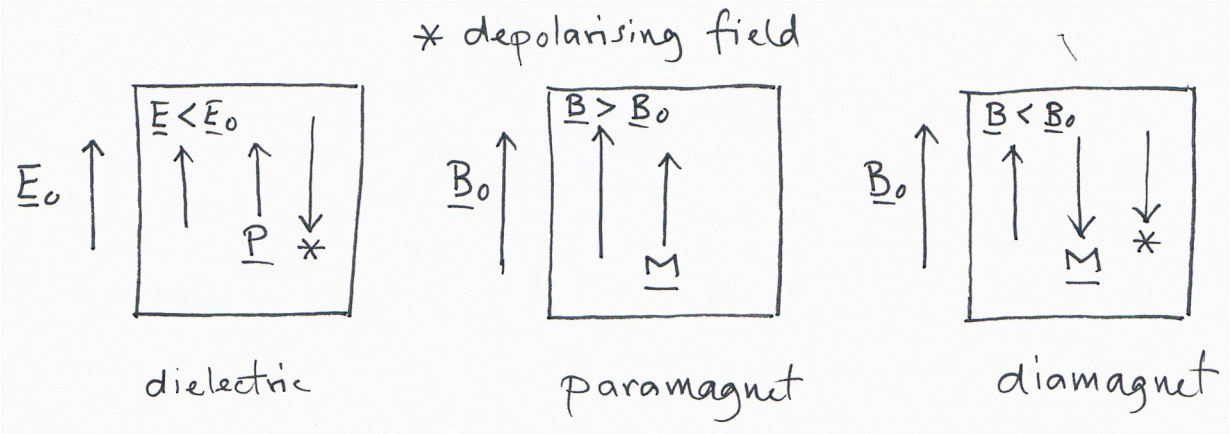


FIGURE 8–4

8.6 Origins of magnetism in matter

This is a quantum mechanical phenomenon, so we need to take great care if we apply ideas from classical physics. We can get some insight, but our conclusions might also be utterly wrong.

8.6.1 Orbital magnetism

Imagine an electron in orbit about the proton in a hydrogen atom. You could regard this as a little loop of current having an associated magnetic moment. Of course a single moving point charge doesn't amount to a current, but suspend your disbelief (you'll have to do that a lot in this section) and consider that

$$\text{current} \times \text{period} = \text{charge}$$

† It's a mess: ϵ and μ are "permittivity" and "permeability" (relative permittivity is also called "dielectric constant."). χ_e and χ_m are "susceptibilities."

So the "current" is,

$$I = \frac{v}{2\pi r}\,(-e) = \frac{1}{2\pi}\,\omega_0\,(-e)$$

in which $\omega_0 = v/r$ is the angular velocity. The magnetic moment is current times area (see figure 8–5),†

$$\boldsymbol{\mu} = I \times \pi r^2 \hat{\mathbf{k}} = -\frac{1}{2}\omega_0\, e\, r^2\, \hat{\mathbf{k}} \tag{8.6.1}$$

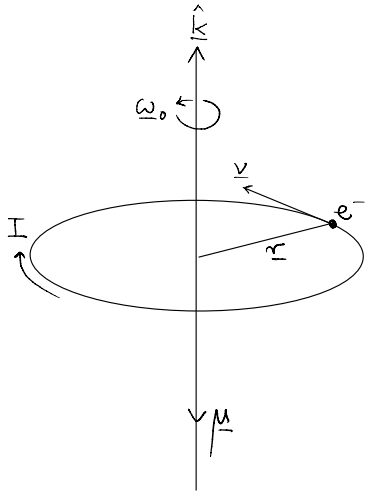


FIGURE 8–5

Recall that the angular momentum is the cross product of the radius vector and the linear momentum vector,

$$\begin{aligned}\boldsymbol{\ell} &= \mathbf{r} \times \mathbf{p} \\ &= m\mathbf{r} \times \mathbf{v} \\ &= m\boldsymbol{\omega}_0 r^2\end{aligned}$$

and

$$\boldsymbol{\omega}_0 = \frac{1}{r^2}\mathbf{r} \times \mathbf{v}$$

is the axial angular velocity vector. So the magnitude of the angular momentum is,

$$\ell = m\,\omega_0\,r^2 \tag{8.6.2}$$

if m is the electron mas. Then the magnitude of the magnetic moment is,

$$\mu = \frac{e}{2m}\,\ell$$

We encountered the *classical gyromagnetic ratio* in subsection 5.4. It is, by definition,

$$\gamma_c = \frac{\text{magnetic moment magnitude}}{\text{angular momentum magnitude}} = \frac{\mu}{\ell} = \frac{e}{2m}$$

† In this section, I will use $\boldsymbol{\mu}$ rather than $\mathbf{m}$ for magnetic moment to avoid confusion with mass.

This is a property of any point charge in circular orbit and we noted already in section 5 that this ratio does not depend either on the radius of the orbit or the angular velocity.

In quantum mechanics, the expectation value of orbital angular momentum is,

$$\ell = m_\ell \, \hbar$$

in which m_ℓ is the *magnetic quantum number* and $\hbar = h/2\pi$ is the "reduced Planck constant." So the magnetic moment is a few Bohr magneton in size, the value of the Bohr magneton being an electronic scale measure of magnetic moment,

$$\mu_B = \frac{e\hbar}{2m} \tag{8.6.3}$$

Now, the centripetal acceleration of the electron is due to the Coulomb force of the proton. Then combining Newton's law with Coulomb's law,

$$m\omega_0^2 r = \frac{1}{4\pi\epsilon_0}\,\frac{e^2}{r^2} \tag{8.6.4}$$

Suppose I apply a uniform magnetic field, figure 8–6,

$$\mathbf{B}_0 = B_0 \, \hat{\mathbf{k}}$$

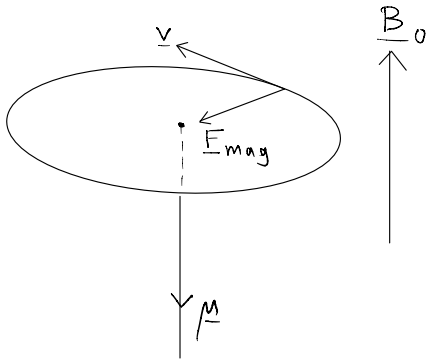


FIGURE 8–6

I get an extra force acting inwards,

$$\mathbf{F}_{\text{mag}} = (-e)\mathbf{v} \times \mathbf{B}_0$$

This will alter the angular velocity because (8.6.4) becomes modified to,

$$\begin{aligned} m\omega^2 r &= \frac{1}{4\pi\epsilon_0}\,\frac{e^2}{r^2} + e\omega r B_0 \\ &= m\omega_0^2 r + e\omega r B_0 \end{aligned}$$

If I solve this quadratic equation under the assumption of a small applied field, strictly: $B_0^2 \ll m/\pi\epsilon_0 r^3$, I find,

$$\omega = \omega_0 + \omega_L$$

The additional term is called the *Larmor frequency*,

$$\omega_L = \frac{eB_0}{2m}$$

The effect of the applied magnetic field is to *increase* the angular velocity of the orbiting electron. So the magnetic moment is now changed by an amount (8.6.1),

$$\Delta\boldsymbol{\mu} = -\frac{1}{2}\omega_L e r^2 \hat{\mathbf{k}}$$

which is opposite to B_0. If I had applied B_0 in the other direction,

$$\mathbf{B}_0 = B_0(-\hat{\mathbf{k}}\,)$$

this would have *slowed down* the orbit and the change in magnetic moment would have been,

$$\Delta\boldsymbol{\mu} = +\frac{1}{2}\omega_L e r^2 \hat{\mathbf{k}}$$

also opposed to $\mathbf{B}_0$. The Larmor frequency may be either positive or negative, remembering that it is the *magnitude* of an *axial* vector.

So, the magnetic field always changes the orbital magnetic moment such as to produce an opposing magnetic field. This is *diamagnetism*, and is analogous to electric polarisation in the sense that the matter responds by causing a *depolarising field*; such that the field inside is *reduced* by polarisation of the atoms from which the substance is made. Mostly, diamagnetism is very weak; an exception is a normal superconductor that expells all magnetic flux so that $\mathbf{B} = \mathbf{0}$ within the superconductor in the superconducting state.

In addition, the magnetic field will exert a torque on the orbital magnetic moment attempting to line up the moment with the applied field. This would *add to*, rather than *opppose* the applied field giving a paramagnetic response to the applied field. But the orbiting electron behaves like a gyroscope: it *precesses* in the magnetic field, actually with an angular frequency equal to the Larmor frequency, ω_L. So orbital paramagnetism is an even weaker effect.

The electron itself has an intrinsic magnetic moment, of magnitude $\sim \mu_B$ according to quantum mechanics, and this will align with an applied magnetic field. But, again according to quantum mechanics, each electron of angular momentum, or "spin", $+\frac{1}{2}\hbar$, is accompanied by one of $-\frac{1}{2}\hbar$, and so only atoms or molecules with an odd number of electrons will in general be paramagnetic.

For a full discussion, see the *Feynman Lectures on Physics*, Vol. 2, chapter 34.

8.6.2 Spin magnetism

Applying classical mechanics, as we have just done, to the orbiting electron actually gives the right quantum mechanical and observed result for the orbital gyromagnetic ratio, namely,

$$\gamma_o = \frac{\mu}{\ell} = -\frac{1}{2}\frac{e}{m} = g\left(\frac{-e}{2m}\right) \tag{8.6.5}$$

and g is called the *g-factor*: $g = 1$ for orbital moments.

So let's be really crazy and try it on an electron! We *pretend* that an electron is a spinning, charged ball of radius R, charge q and mass m. This ball then has a charge density,

$$\rho_q = \frac{q}{\frac{4}{3}\pi R^3}$$

and a mass density

$$\rho_m = \frac{m}{\frac{4}{3}\pi R^3}$$

We will find the magnetic moment and angular momentum simultaneously. Consider the ball to be made up of infinitesimal rings of spinning charge and mass, figure 8–7.

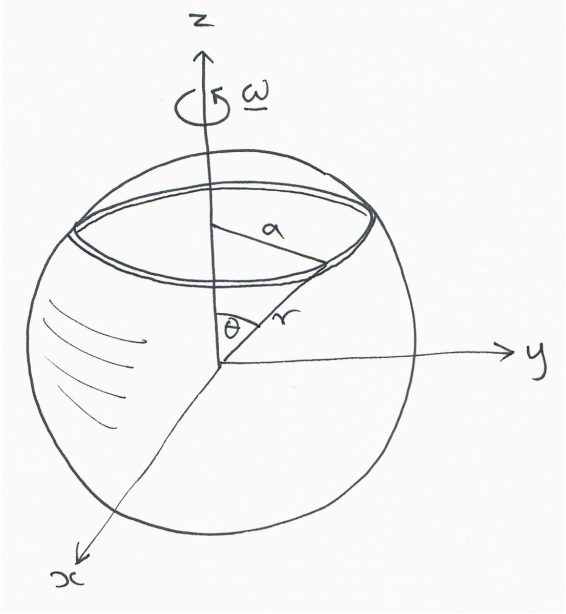


FIGURE 8–7

In cross section in the y-z plane, the area of cross section is as in figure 8–8.

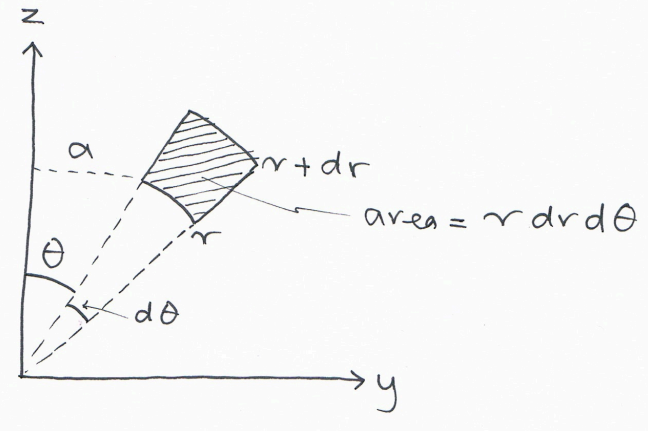


FIGURE 8–8

The radius of the ring is

$$a = r\sin\theta$$

The volume of the ring is

$$2\pi a\, r\mathrm{d}r\, d\theta = 2\pi r^2 \mathrm{d}r\sin\theta\,\mathrm{d}\theta$$

so the charge carried by the ring is

$$\mathrm{d}q = 2\pi\,\rho_q\, r^2\mathrm{d}r\ \sin\theta\,\mathrm{d}\theta$$

and its mass is

$$\mathrm{d}m = 2\pi\,\rho_m\, r^2\mathrm{d}r\ \sin\theta\,\mathrm{d}\theta$$

Now, angular momentum is (8.6.2),

$$\text{mass} \times \text{angular velocity} \times \text{radius}^2$$

so the little rings contribute to the angular momentum, ℓ, of the ball an amount,

$$\begin{aligned}\mathrm{d}\ell &= \mathrm{d}m\,\omega\, a^2\\ &= 2\pi\,\omega\,\rho_m\, r^4\mathrm{d}r\ \sin^3\theta\,\mathrm{d}\theta\end{aligned}$$

and comparing with (8.6.1) we see that the contribution to the magnetic moment is,

$$\begin{aligned}\mathrm{d}\mu &= \frac{1}{2}\omega\,\mathrm{d}q\, a^2\\ &= \frac{1}{2}\,2\pi\,\omega\,\rho_q\, r^4\mathrm{d}r\ \sin^3\theta\,\mathrm{d}\theta\end{aligned}$$

Now we need to integrate. Look at figure 8.7. The ring indicated is at the outer radius of a disc so we'd have to integrate over all values of θ and r to pick up all the mass and charge in the ball. That means that we integrate over r from 0 to R and over θ from 0 to π,

$$\begin{aligned}\ell &= 2\pi\,\omega\frac{m}{V}\int_0^R r^4\mathrm{d}r\int_0^\pi \sin^3\theta\,\mathrm{d}\theta\\ \mu &= \pi\,\omega\frac{q}{V}\int_0^R r^4\mathrm{d}r\int_0^\pi \sin^3\theta\,\mathrm{d}\theta\end{aligned}$$

where $V = 4\pi R^3/3$ is the volume of the ball. We only want the gyromagnetic ratio, so in the ratio ℓ/μ the integrals will cancel and we don't even need to do them (see section 5.2). But for completeness, integrating over r is easy and

$$\int \sin^3\,\mathrm{d}\theta = \frac{1}{3}\cos^3\theta - \cos\theta + \text{constant}$$

as you can check by differentiation. So we get

$$\mu = \frac{1}{5}\omega\, q\, R^5$$

and[†]

$$\ell = \frac{2}{5}\omega\, m\, R^5$$

Finally, the gyromagnetic ratio for the spinning charged ball is,

$$\gamma = \frac{\mu}{\ell} = \frac{1}{2}\frac{q}{m}$$

This is the same as the gyromagnetic ratio for a point charge in orbit. Maybe this is not surprising since the solid ball is made up of infinitesimal pieces of charge and mass, and we know that the gyromagnetic ratio actually does not depend on the radius of its orbit so they all just contribute to the total mass and charge.

If the ball is an electron, $q = -e$ and the *spin* angular momentum, according to quantum mechanics, is $\ell_s = \frac{1}{2}\hbar$, so a *classical* treatment results in the magnetic moment of an electron as

$$\mu = -\frac{1}{4}\frac{e\hbar}{m} \qquad \longleftarrow \text{wrong!} \tag{8.6.6}$$

which unfortunately is wrong!! It's too small by a tiny bit more that a half. This is not really surprising. (What *is* surprising is that the classical treatment is correct for the orbital magnetic moment of an electron in a hydrogen atom.) You cannot for a moment expect that the electron can be described as a spinning solid ball. If it were, then as Pauli pointed out the surface of the ball would be travelling faster than the speed of light.[‡] Indeed as you see in the additional note on `KEATS` about point charges, we don't really know what the radius of an electron is—it may indeed be point-like. Nonetheless, quantum mechanics does strange things: the electron has a measurable gyromagnetic ratio [*] and a measurable magnetic moment.

We saw at equation (8.6.3) that an appropriate unit of atomic magnetic moments is the Bohr magneton,[‖]

$$\mu_B = \frac{1}{2}\frac{e\hbar}{m} = 9.274009994 \times 10^{-24}\ \text{J T}^{-1}$$

and then if we accept that (8.6.6) is out by almost exactly one half, the intrinsic spin magnetic moment of the electron is,

$$\mu \approx -\mu_B$$

We can also adopt the approach of equation (8.6.5) and write for the intrinsic gyromagnetic ratio of the electron

$$\gamma_{\text{e}} = g_e \left(\frac{-e}{2m}\right)$$

[†] You may have known this: $\ell = I\omega$ where $I = \frac{2}{5}mR^2$ is the moment of inertia of a uniform ball.

[‡] See `http://www.lorentz.leidenuniv.nl/history/spin/spin.html`

[*] So does the neutron although clearly the ratio of charge to angular momentum *ought* to be zero.

[‖] The permament magnetic moment per atom in pure iron is 2.2 μ_B. This means that the difference between the numbers of spin-up and spin-down electrons on each atom is 2.2—not an integer!

in which g_e is the *g-factor* for electron spin. If g_e were exactly equal to two, then the magnetic moment of the electron would be exactly minus one Bohr magneton, as you can see by writing,

$$\mu = \gamma_e \, \ell_s = g_e \left(\frac{-e}{2m} \right) \frac{1}{2}\hbar = -\frac{1}{2} g_e \mu_B$$

In fact one of the triumphs of Dirac's relativistic theory of the electron is that g_e comes out to be exactly two. However measurements find that $g_e = 2.0023\ldots$. Quantum electrodynamics which was invented by Hans Bethe, Shin'ichirō Tomonaga, Julian Schwinger, Richard Feynman and Freeman Dyson, predicts that,

$$g_e = 2\left(1 + \frac{\alpha}{2\pi} + \ldots\right)$$

where α is the fine structure constant (see the additional note on `KEATS` about point charges). And this prediction is in very close agreement, to many decimal places, with the measurement; this is regarded as one of the great triumphs of physics.†

Here is a table of measured gyromagnetic ratio, magnetic moment and g-factor for four particles.

	neutron	proton	electron	muon
charge	0	1	–1	–1
spin	$\frac{1}{2}\hbar$	$\frac{1}{2}\hbar$	$\frac{1}{2}\hbar$	$\frac{1}{2}\hbar$
γ (rad s^{-1} T^{-1})	-1.832×10^{8}	2.765×10^{8}	-1.761×10^{11}	-8.516×10^{8}
μ (J T^{-1})	-0.966×10^{-26}	1.411×10^{-26}	-928.5×10^{-26}	-4.490×10^{-26}
g	-3.826085	5.585694	-2.002319	-2.002332

We define electron and neutron Bohr magneton in terms of the electron and proton masses, m_e and m_p as,

$$\mu_B = \frac{e\hbar}{2m_e} = 927.401 \times 10^{-26} \text{ J T}^{-1}$$
$$\mu_N = \frac{e\hbar}{2m_p} = 0.505078 \times 10^{-26} \text{ J T}^{-1}$$

and for the neutron and proton, we use,

$$\gamma = g\gamma_c$$

with

$$\gamma_c = \frac{1}{2}\frac{e}{m_p}$$

Similarly, for the muon we use the muon mass, $m_\mu = 206.768 m_e$.

† See `https://en.wikipedia.org/wiki/Precision_tests_of_QED`

These are important considerations in areas such as nuclear magnetic resonance and MRI because the measured Larmor frequency $\omega_L = \gamma_c B$ is proportional to the magnetic field, and the proportionality constant is the gyromagnetic ratio. In MRI you are imaging the protons in water by causing them to precess in an **H**-field and putting them into spin-flip resonance by application of a perpendicular RF field at the Larmor frequency.

Let us stop NOW! You cannot understand magnetism in matter using classical physics. In fact the great Neils Bohr in his 1911 PhD thesis showed using statistical mechanics that the magnetisation of any classical system is identically zero. That must have puzzled him greatly. But I get the impression that physicists of the first three decades of the twentieth century were in a state of more or less permanent puzzlement. It's much less exciting now: we know everything; and the things we don't know like how many leptons are there, or string theory don't really matter very much...

9. Propagation and absorption

9.1. Introduction

You saw in Problems Set 4 that a solution to Maxwell's equations in vacuum is a travelling electromagnetic wave of **E** and **B** fields. I urge you to read chapter 18 in the Feynman Lectures. For me, as a student, section 18.4 came as something of an "epiphany". Feynman imagines a sheet of charge of infinite extent in the x-y plane and infinitely thin. He imagines that this sheet is set into motion in the x-direction for a period of time, T, and then brought to rest again. The two dimensional current density induces a magnetic field; at first, infinitely close to the current sheet, but as the **B** field expands outward in the z-direction it, in turn, induces an **E** field and so on. This is a consequence of the wonderful symmetry of Maxwell's equations that we have already discussed—a result of Maxwell's discovery of the "displacement current". So the consequence of the brief motion of the charge sheet is to launch an electromagnetic wave into the vacuum. This travels off into space as a plane wave in the z-direction: each oscillation of the **B** field inducing an **E** field and each oscillation of the **E** field inducing a **B** field feeding off itself forever.

9.2 Maxwell's equations for propagating and evanescent waves

In the absence of any free charge Maxwell's equations are (Section 6, (6.5.2)),

$$\boldsymbol{\nabla}\cdot\mathbf{D}=0 \qquad (i)$$

$$\boldsymbol{\nabla}\cdot\mathbf{B}=0 \qquad (ii)$$

$$\boldsymbol{\nabla}\times\mathbf{E}=-\frac{\partial\mathbf{B}}{\partial t} \qquad (iii)$$

$$\boldsymbol{\nabla}\times\mathbf{H}=\mathbf{J}_{\text{free}}+\frac{\partial\mathbf{D}}{\partial t} \qquad (iv)$$

Assuming *linear* media, I will use $\mathbf{B}=\mu\mathbf{H}$, where μ is the magnetic permeability of the medium; and in addition I will have $\mathbf{D}=\epsilon\mathbf{E}$ where ϵ is the permittivity of the medium. I assume that the material properties, ϵ, μ and conductivity, σ, are time and space independent. I now take the curl of (iii),

$$\begin{aligned}\boldsymbol{\nabla}\times(\boldsymbol{\nabla}\times\mathbf{E})&=-\mu\,\frac{\partial}{\partial t}\,(\boldsymbol{\nabla}\times\mathbf{H})\\&=-\mu\,\frac{\partial}{\partial t}\left(\mathbf{J}_{\text{free}}+\epsilon\,\frac{\partial\mathbf{E}}{\partial t}\right)\end{aligned}$$

using (iv). Then comes the identity, for any vector field, $\mathbf{u}(\mathbf{r})$

$$\boldsymbol{\nabla}\times(\boldsymbol{\nabla}\times\mathbf{u})=\boldsymbol{\nabla}\,(\boldsymbol{\nabla}\cdot\mathbf{u})-\nabla^2\mathbf{u}$$

and in view of the fact that $\boldsymbol{\nabla}\cdot\mathbf{E}=0$ in a region absent of free charge as long a ϵ is uniform,† and in view of the fact that according to Ohm's law, in the absence of any non

† $\boldsymbol{\nabla}\cdot\mathbf{E}=\boldsymbol{\nabla}\cdot(\epsilon^{-1}\mathbf{D})=\epsilon^{-1}\boldsymbol{\nabla}\cdot\mathbf{D}+\mathbf{D}\cdot\boldsymbol{\nabla}\,(\epsilon^{-1})=\epsilon^{-1}\rho_{\text{free}}$.

irrotational, electromotive electric field, the free current density is $\mathbf{J}_{\text{free}} = \sigma\mathbf{E}$, where σ is the conductivity of the medium, I arrive at

$$\nabla^2\mathbf{E} - \epsilon\,\mu\,\frac{\partial^2\mathbf{E}}{\partial t^2} - \mu\,\sigma\,\frac{\partial\mathbf{E}}{\partial t} = 0 \tag{9.2.1}$$

If I take the curl of (iv), then

$$\boldsymbol{\nabla}\times(\boldsymbol{\nabla}\times\mathbf{H}) = \boldsymbol{\nabla}\times\mathbf{J}_{\text{free}} + \epsilon\,\frac{\partial}{\partial t}\,(\boldsymbol{\nabla}\times\mathbf{E})$$

Again, I use the vector identity; I allow that $\boldsymbol{\nabla}\cdot\mathbf{H} = 0$;† and I use Faraday's law (iii) and $\mathbf{J}_{\text{free}} = \sigma\mathbf{E}$ and I get,

$$\nabla^2\mathbf{H} - \epsilon\,\mu\,\frac{\partial^2\mathbf{H}}{\partial t^2} - \mu\,\sigma\,\frac{\partial\mathbf{H}}{\partial t} = 0 \tag{9.2.2}$$

You can see again the wonderful symmetry in (9.2.1) and (9.2.2). What we have achieved is to combine the four first order partial differential Maxwell equations, which are mixed in $\mathbf{E}$ and $\mathbf{H}$, into two second order partial differential equations which are unmixed—one for $\mathbf{H}$ and one for $\mathbf{E}$. It is enough just to consider the solutions to one. The other field has to come along for the ride. Experiments with photographic plates, and others, point to the $\mathbf{E}$-field as that which carries properties most familiar to us as "light" so let's analyse (9.2.1). Do not be concerned because this is a vector equation—just think of it as three equations, one for each of the three components, $E_i(\mathbf{r})$, of $\mathbf{E}$, each a scalar field,

$$\nabla^2 E_i - \epsilon\,\mu\,\frac{\partial^2 E_i}{\partial t^2} - \mu\,\sigma\,\frac{\partial E_i}{\partial t} = 0 \tag{9.2.3}$$

9.3 The Wave Equation

Before going further, let's digress a bit on the subject of linear second order differential equations concerning a function $f(z,t)$. You already know from Problems Set 4, that

$$\frac{\partial^2 f}{\partial z^2} = \frac{1}{v^2}\frac{\partial^2 f}{\partial t^2} \tag{9.3.1}$$

is the *classical wave equation* that describes the propagation of the disturbance f according to

$$f(z,t) = A\cos(kz \pm \omega t) \tag{9.3.2}$$

with amplitude A. This situation corresponds to (9.2.3) in the event that the third term is neglected or zero (this would be the case for electromagnetic waves propagating in a medium with zero conductivity, $\sigma = 0$) and if we are considering a plane wave

† As long as the permeability is uniform everywhere in the medium then $\boldsymbol{\nabla}\cdot\mathbf{H} = \mu^{-1}\boldsymbol{\nabla}\cdot\mathbf{B} + \mathbf{B}\cdot\boldsymbol{\nabla}(\mu^{-1}) = 0$.

propagating in the z-direction and whose $\mathbf{E}$ vector is polarised in the x-direction. The wave equation is

$$\frac{1}{v^2}\frac{\partial^2 E_x}{\partial t^2} = \frac{\partial^2 E_x}{\partial z^2} \tag{9.3.3}$$

A solution is

$$E_x = A\cos(kz - \omega t) \tag{9.3.4}$$

If you differentiate (9.3.4) twice both with respect to z and with respect to t and put the results into (9.3.3) you see that

$$k = \frac{\omega}{v} \tag{9.3.5}$$

The phase velocity of the wave would be given by

$$\frac{1}{v^2} = \epsilon\mu = \epsilon_r\epsilon_0\,\mu_r\mu_0 = \frac{\epsilon_r\mu_r}{c^2}$$

because $c^2 = 1/\epsilon_0\mu_0$, so in vacuum in which $\epsilon_r = \mu_r = 1$ the wave propagates at the speed of light, c.† In the case the medium is not vacuum but some matter then the phase velocity would be

$$v = \frac{1}{\sqrt{\epsilon\mu}} = \frac{c}{\sqrt{\epsilon_r\mu_r}} \tag{9.3.6}$$

and this is a certain fraction of the speed of light. The ratio, $n = c/v$, of the speed of light to the phase velocity is called the refractive index of the medium. So the refractive index is related to the permittivity and permeability of the medium according to,

$$n = \sqrt{\frac{\epsilon\mu}{\epsilon_0\mu_0}} = \sqrt{\epsilon_r\mu_r} \approx \sqrt{\epsilon_r} \tag{9.3.7}$$

We allow the approximation because in paramagnetic and diamagnetic materials μ and μ_0 differ by only about 10^{-5}. In paramagnets, $\mu_r > 1$ and in diamagnets, $\mu_r < 1$; but in both cases $\mu_r \approx 1$ to within about 10^{-5}. Of course the refractive index of the vacuum is one.

There are two principal reasons why assuming that $\sigma = 0$ does not furnish us with a complete description of the propagation of electromagnetic waves.

1. In neglecting the third term in (9.2.3) the only *current* that we allow is the displacement current. I pointed out in section 6.2 that, in addition to the vacuum displacement current, if we leave out convective currents arising from relative motion between the medium and an observer, then we can identify three sources of current: (i) free current, that is, current that we control in circuits, particle beams

† Of course we have to recall here the remarkable truth as first intrigued Maxwell and others in the mid nineteenth century that the product of two quantities that can be measured by the forces between charged objects and steady currents turns out to be related to speed of propagation of light. Maxwell wrote, "we can scarcely avoid the inference that light consists in the transverse undulations of the same medium which is the cause of electric and magnetic phenomena."

and so on; (*ii*) polarisation currents due to movements of bound charge within matter if the polarisation $\mathbf{P}$ is time dependent; (*iii*) bound magnetisation currents, $\boldsymbol{\nabla} \times \mathbf{M}$. If polarisation or magnetisation current sources exist in the medium as a result of the presence of the oscillating electric or magnetic fields then they must be accounted for. The bound current is included implicitly in the Ampère–Maxwell law. The polarisation current is built in to the displacement current,

$$\frac{\partial \mathbf{D}}{\partial t} = \frac{\partial \mathbf{P}}{\partial t} + \epsilon_0 \frac{\partial \mathbf{E}}{\partial t} = \begin{matrix}\text{polarisation} \\ \text{density current}\end{matrix} + \begin{matrix}\text{vacuum displacement} \\ \text{density current}\end{matrix}$$

This leaves the free current which is not accounted for thus far.

2. Equation (9.3.6) asserts that the phase velocity, v, is *independent* of the frequency, ω. This is of course true in the vacuum, but you all know that it's not true in so called *dispersive media*. The obvious example is the splitting of white light into its colours by a prism. Another well known piece of evidence is that the measured *static* dielectric constant of water is about 81, so the refractive index is predicted by (9.3.7) to be about 9 which is far from what is observed in the deflection of a visible ray of light in water—namely that the refractive index is about 1.3. The point is that the constants, ϵ_r, μ_r and σ have been introduced from electrostatics and magnetostatics as determined in experiments using *static* electric and magnetic fields. But if an applied electric field is oscillating like $\mathbf{E} = \mathbf{E}_0 \cos \omega t$ then it will drive the polarisation in the manner of any driven oscillator, like $\mathbf{P} = (\epsilon_r - 1)\, \epsilon_0 \mathbf{E}_0 \cos(\omega t - \phi)$ (see section 7, subsection 7); so the response may be out of phase with the driving field and there may be resonances which produce large responses and damping which attenuate the response and so you expect the dielectric "constant", ϵ_r, to be *frequency dependent*. Most of the textbooks at this point state that in fact ϵ_r, μ_r and σ are all frequency dependent without detailed comment and to me, at least, that seems unsatisfactory as we need to revise what we mean by these material properties.

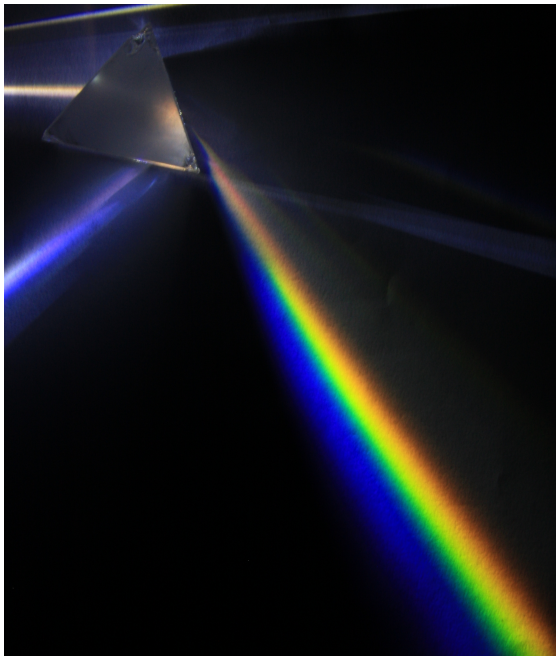


FIGURE 9–1: A prism.

D Kuru, CC BY-SA 3.0 at, https://commons.wikimedia.org/w/index.php?curid=7082370

What this means is that the assumptions we have made so far are valid in the static limit—that is, in the limit of long wavelength electromagnetic waves, or slowly varying fields. Let's just simply assert that ϵ_r is frequency dependent and solve in detail for electromagnetic waves in insulators. Then we'll see in subsection 8, *et seq.*, how frequency dependence arises naturally once we allow that $\sigma \neq 0$.

9.4 Electromagnetic Plane Wave (after Panofsky and Phillips)

Let us begin our analysis of equation (9.2.3) by assuming a medium that has zero conductivity. Then the second term in (9.2.3) vanishes, since $\sigma = 0$ and we're left with,

$$\nabla^2 E_i = \epsilon\,\mu\,\frac{\partial^2 E_i}{\partial t^2}$$

which is the classical wave equation describing the propagation of the components of the electric field vector with a phase velocity

$$v = \frac{1}{\sqrt{\epsilon\mu}}$$

The plane wave is a one-dimensional solution to Maxwell's equations. Exactly as described by Feynman in chapter 18 there is a wave front that propagates in, say, the z-direction, with the **E** and **B** fields oscillating in perpendicular directions in the x-y plane. A good way to describe this mathematically and to obtain exact insight from Maxwell's equations is as follows.

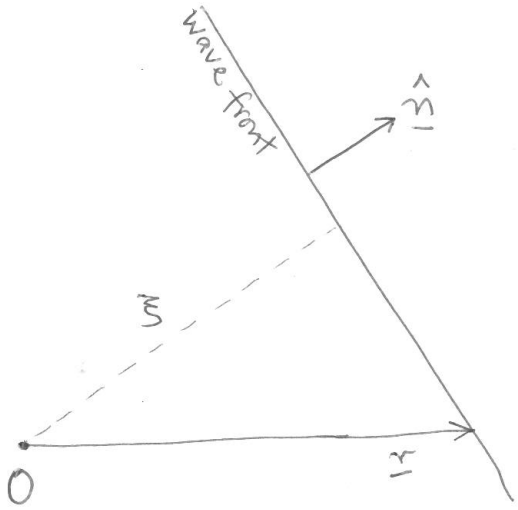


FIGURE 9–2: A plane wavefront travelling away from the origin. The unit normal to the wavefront is $\hat{\mathbf{n}}$; ξ is the perpendicular distance from origin to the instantaneous wavefront; $\mathbf{r}$ is a vector connecting the origin to a point on the wavefront.

All fields are functions only of the distance of a given plane from an origin as shown in figure 9–2. The plane is the wavefront and has a unit normal $\hat{\mathbf{n}}$ which is thereby the direction of propagation of the wave. We will call the perpendicular distance from the plane to the origin ξ. This means that if the plane wave set out from the origin with a wavefront perpendicular to $\hat{\mathbf{n}}$, then at some time t it has travelled a distance ξ. This simplifies all the spacial derivatives in the Maxwell equations because the fields are only functions of ξ. Therefore the "grad" operator reduces to

$$\boldsymbol{\nabla} = \hat{\mathbf{n}} \frac{\partial}{\partial \xi}$$

With this expression for $\boldsymbol{\nabla}$ Maxwell's equations, in a medium free of charge and electromotive fields, are

$$\hat{\mathbf{n}} \cdot \frac{\partial \mathbf{D}}{\partial \xi} = 0 \qquad (i)$$

$$\hat{\mathbf{n}} \cdot \frac{\partial \mathbf{B}}{\partial \xi} = 0 \qquad (ii)$$

$$\hat{\mathbf{n}} \times \frac{\partial \mathbf{E}}{\partial \xi} = -\frac{\partial \mathbf{B}}{\partial t} \qquad (iii)$$

$$\begin{aligned} \hat{\mathbf{n}} \times \frac{\partial \mathbf{H}}{\partial \xi} &= \mathbf{J}_{\text{free}} + \frac{\partial \mathbf{D}}{\partial t} \\ &= \sigma \mathbf{E} + \frac{\partial \mathbf{D}}{\partial t} \end{aligned} \qquad (iv)$$

We take the scalar product of $\hat{\mathbf{n}}$ with (iv) which gives us,

$$\hat{\mathbf{n}} \cdot \left(\frac{\sigma}{\epsilon} + \frac{\partial}{\partial t} \right) \mathbf{D} = 0 \qquad (9.4.1)$$

Since the scalar product of $\hat{\mathbf{n}}$ with $\partial \mathbf{D}/\partial \xi$ (i) is zero this means components of $\mathbf{D}$ and $\mathbf{E}$ do not vary with position in the direction of $\hat{\mathbf{n}}$. The time dependence of these normal components is given by solutions of (9.4.1) which are *evanescent*. So the normal component of $\mathbf{E}$, E_n, must decay like

$$E_n = A\, e^{-t/\tau} \ , \qquad \tau = \frac{\epsilon}{\sigma}$$

Furthermore, taking the scalar product of $\hat{\mathbf{n}}$ and (iii), gives us

$$\begin{aligned} \hat{\mathbf{n}} \cdot \frac{\partial \mathbf{B}}{\partial t} &= 0 \\ \hat{\mathbf{n}} \cdot \frac{\partial \mathbf{B}}{\partial \xi} &= 0 \end{aligned} \qquad (ii)$$

so that exactly as for $\mathbf{E}$ there is no component of the $\mathbf{B}$ field in the direction of propagation that survives without dissipating. If the medium is non conducting then these normal components of $\mathbf{E}$ and $\mathbf{B}$ vanish identically, because $\sigma = 0$ implies that $\tau \to \infty$. This expresses the well known fact the the electric field is oscillating *in the plane of the*

wavefront. In other words the plane electromagnetic wave is *transverse*. Look ahead to figure 9–3 which is probably already familiar to you. The arrows show that $\mathbf{E}$ and $\mathbf{B}$ are *perpendicular* to $\hat{\mathbf{n}}$ (in this case the z-direction). If there *is* a component parallel to z, then it must be constant as a function of z; and it cannot be oscillating in time in that direction: it must be decaying exponentially.

Let's now find the time dependence of the transverse components of $\mathbf{E}$ and $\mathbf{B}$. In this geometry, equation (9.2.1) in a medium having zero conductivity is

$$\frac{\partial^2 \mathbf{E}}{\partial \xi^2} = \mu\,\epsilon\,\frac{\partial^2 \mathbf{E}}{\partial t^2} \tag{9.4.2}$$

The general solution will be waves as in equation (9.3.2), and Problems Set 4, equation (2), $f(\xi, t) = g(\xi - vt)$, which are one-dimensional waves travelling in the plus ξ-direction with a speed

$$v = \frac{1}{\sqrt{\epsilon\mu}} = \frac{c}{\sqrt{\epsilon_r \mu_r}}$$

which is the *phase velocity*. The phase velocity is only equal to the speed of light, c, if the wave is propagating in the vacuum. Otherwise its speed is necessarily less than c and, as you know, we define the *refractive index* of the medium as the ratio $n = c/v \geq 1$. Hence, as we saw in section 9.3,

$$n = \sqrt{\frac{\epsilon\mu}{\epsilon_0\mu_0}} = \sqrt{\epsilon_r \mu_r} \tag{9.3.7}$$

We can write the solution to (9.4.2) as,

$$\mathbf{E} = \mathbf{E}_0\, e^{ik\xi - i\omega t} \tag{9.4.4a}$$

in which

$$k = \frac{\omega}{v} = \frac{2\pi}{\lambda}$$

is the *wavenumber* and λ is the wavelength. We can express the wavenumber as a *wavevector* which points in the direction of propagation,

$$\mathbf{k} = \omega\frac{\mathbf{v}}{v^2} = \frac{\omega}{v}\hat{\mathbf{n}} \tag{9.4.3}$$

where $\mathbf{v}$ is the phase velocity vector which points in the direction $\hat{\mathbf{n}}$, the normal to the wavefront. With these definitions, the wavefunction for a forward travelling wave is

$$\mathbf{E} = \mathbf{E}_0\, e^{i(\mathbf{k}\cdot\mathbf{r} - \omega t)} \tag{9.4.4b}$$

and, of course, the equivalent wavefunction for $\mathbf{B}$ is

$$\mathbf{B} = \mathbf{B}_0\, e^{i(\mathbf{k}\cdot\mathbf{r} - \omega t)} \tag{9.4.4c}$$

I now use (*iii*), differentiate (9.4.4c) with respect to t, and get

$$\hat{\mathbf{n}} \times \frac{\partial \mathbf{E}}{\partial \xi} = i\omega\mathbf{B} = ik\hat{\mathbf{n}} \times \mathbf{E}$$

where the last step follows by differentiating (9.4.4a) with respect to ξ. This becomes, using (9.4.3),

$$\mathbf{B} = \sqrt{\epsilon\mu}\,\hat{\mathbf{n}} \times \mathbf{E} = \frac{1}{\omega}\,\mathbf{k} \times \mathbf{E} \tag{9.4.5}$$

Since (*iii*) is Faraday's law you may assert that the oscillating *magnetic field* has "induced" an *electric field*† and that as clearly indicated by (9.4.5) these two fields are at all times mutually perpendicular. In fact the system $(\mathbf{E}, \mathbf{B}, \mathbf{k})$ forms a right-handed coordinate system, in that order, see figure 9–3. Numerically, in the SI system, it is clear from (9.4.5) that the magnitude of the $\mathbf{E}$ field is $k/\omega = v$ times the magnitude of the $\mathbf{B}$ field and hence in conjuction with (9.4.4b) and (9.4.4c) we have these two equations for the $\mathbf{E}$ and $\mathbf{B}$ fields,

$$\mathbf{E} = E_0\, e^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)}\,\hat{\mathbf{i}} \tag{9.4.6a}$$

$$\mathbf{B} = \frac{E_0}{v}\, e^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)}\,\hat{\mathbf{j}} \tag{9.4.6b}$$

$$= \frac{k}{\omega}\, E_0\, e^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)}\,\hat{\mathbf{j}} \tag{9.4.6c}$$

if we choose the z-axis as the direction of propagation, $\hat{\mathbf{n}} = \hat{\mathbf{k}}$ (this is the unit vector along z, not the wavevector). In vacuum, $v = c$. The two fields are clearly oscillating in phase, see figure 9–3.

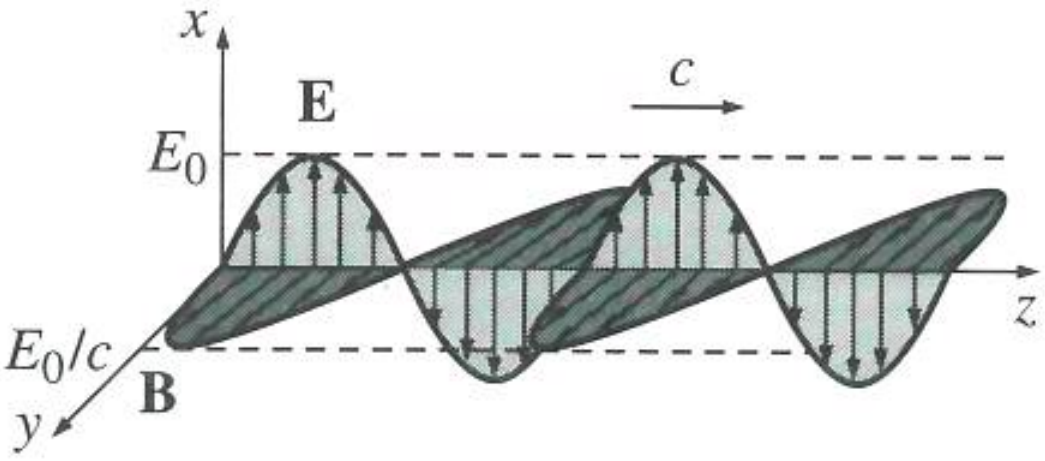


FIGURE 9–3: The electric and magnetic fields in an electromagnetic wave propagating in the vacuum. $\mathbf{E}$ and $\mathbf{B}$ are in phase and the phase velocity is c. Taken from Griffiths. Note, especially for section 9.11, below, that this is a *plane polarised* wave. The $\mathbf{E}$ vector is oscillating in the xz-plane and conventionally we use the $\mathbf{E}$ vector to thus establish xz as the plane of polarisation. We could equally well have chosen the $\mathbf{B}$ vector.

† Conversely, by the symmetry of Maxwell's equations you may also infer that the $\mathbf{E}$ field has "induced" the $\mathbf{B}$ field in accord with the Ampère–Maxwell law (*iv*).

9.5 Electromagnetic plane wave energy density

We enquire, what is the amount of energy that is carried away from the source of the electromagnetic wave? Because there is no dissipation in solutions to the classical wave equation (9.4.2) energy is conserved and is carried, as we know, within the oscillating electric and magnetic fields (see Section 6.3). The energy contained in the electric field is the square of $|\mathbf{E}|$ averaged over a cycle. You know that when we average, say, $E^2 = E_0^2 \cos^2(k\xi - \omega t)$ over a cycle we get $E_0^2/2$. In the case of complex waves, we can use the following identity.† If $\mathbf{u}$ and $\mathbf{v}$ are *complex* vectors, then

$$\overline{(\mathrm{Re}\ \mathbf{u}) \cdot (\mathrm{Re}\ \mathbf{v})} = \frac{1}{2}\mathrm{Re}\ (\mathbf{u} \cdot \mathbf{v}^*) = \frac{1}{2}\mathrm{Re}\ (\mathbf{u}^* \cdot \mathbf{v})$$

where the '*' implies a complex conjugate. In words, "the time average of the product of the real parts of two vectors both of which vary as $e^{-i\omega t}$ is one half the real part of the product of one of the vectors and the complex conjugate of the other." Since the products on the right are independent of time ($e^{-i\omega t}e^{i\omega t} = 1$) there's no need for the "averaging" bar over them. If we apply this formula to the fields (9.4.6) we get an average energy density (see Section 6.4),

$$\overline{U} = \frac{1}{2}\overline{(\mathbf{E} \cdot \mathbf{D} + \mathbf{H} \cdot \mathbf{B})} = \frac{1}{2}\epsilon E_0^2 \qquad [\text{Joules metre}^{-3}]$$

using $v = 1/\sqrt{\epsilon\mu}$ and the "constitutive relations" $\mathbf{D} = \epsilon\mathbf{E}$ and $\mathbf{B} = \mu\mathbf{H}$.

The Poynting vector is $\mathbf{S} = \mathbf{E} \times \mathbf{H}$, pointing in the direction of propagation, $\hat{\mathbf{n}}$, and its average over time is

$$\overline{\mathbf{E} \times \mathbf{H}} = \frac{1}{2}\sqrt{\frac{\epsilon}{\mu}}\, E_0^2\, \hat{\mathbf{n}}$$

using again the constitutive relations. If we use equation (9.4.3), then we have, finally,

$$\mathbf{S} = \overline{U}\, \mathbf{v} \qquad [\text{Watts metre}^{-2}]$$

It makes sense that the energy density is propagated at the same phase velocity as the electromagnetic wave.

---

† I'll leave you to prove this for yourself by writing it all out in terms of real and imaginary parts.

9.6 The Diffusion Equation

To continue our digression into differential equations, suppose the second term in (9.2.3) can be neglected; then instead of (9.3.1) we're looking at an equation like,

$$\frac{\partial^2 f}{\partial x^2} = \text{constant} \times \frac{\partial f}{\partial t} \tag{9.6.1}$$

This is called the *classical diffusion equation*. It describes, for example, the propagation of ink that has been dropped into water. If $f(\mathbf{r}, t)$ is the concentration of ink then (9.6.1) says that as time goes on the concentration flattens out so that if at the beginning there's a very high concentration where the ink entered the water and zero concentration far away in the liquid then eventually the ink will distribute itself evenly so that $f(\mathbf{r}, \infty)$ is constant everywhere. You can see that the equation of motion is flattening out the concentration profile because the rate of change of concentration (the R.H.S.) is proportional to the *curvature* of the concentration profile (the L.H.S). In diffusion (9.6.1) is called "Fick's second law" and the constant is the diffusivity. It is very interesting and significant to note that the Schrödinger equation,

$$\begin{aligned}\frac{\partial^2 \psi}{\partial x^2} &= -i\hbar \frac{\partial \psi}{\partial t} \\ &= \hbar \frac{\partial \psi}{\partial (it)}\end{aligned}$$

is a diffusion equation for the probability amplitude $\psi(x)$ in *imaginary* time, and $\hbar$ is the "diffusivity".

9.7 Relaxation time

We're not quite ready to tackle (9.2.3) as a combined wave and diffusion equation. First we have to enquire into timescales. And we need some insight into the relative magnitudes of the propagating (second order in time) and diffusive (first order in time) terms. If there are non stationary (that is, time dependent) free currents set up in the medium by the electromagnetic fields, leading to a current density $\mathbf{J}_{\text{free}}(\mathbf{r}, t)$, we want to know how long it takes for these transients to die away in the medium, leaving just a stationary, time independent, current density, $\mathbf{J}_s$. If there are non stationary currents then the equation of continuity is

$$\boldsymbol{\nabla} \cdot \mathbf{J}_{\text{free}} + \frac{\partial \rho_{\text{free}}}{\partial t} = 0$$

According to Gauss's law for the free charge,

$$\boldsymbol{\nabla} \cdot \mathbf{D} = \rho_{\text{free}}, \quad \text{hence} \qquad \frac{\partial}{\partial t} \boldsymbol{\nabla} \cdot \mathbf{D} = \frac{\partial \rho_{\text{free}}}{\partial t}, \quad \text{hence} \qquad \boldsymbol{\nabla} \cdot \frac{\partial \mathbf{D}}{\partial t} = \frac{\partial \rho_{\text{free}}}{\partial t}$$

This gives us

$$\boldsymbol{\nabla} \cdot \left(\mathbf{J}_{\text{free}} + \frac{\partial \mathbf{D}}{\partial t} \right) = 0$$

We also know that the continuity equation for stationary currents reduces to

$$\boldsymbol{\nabla} \cdot \mathbf{J}_s = 0$$

because the time derivative is zero if the current is time independent, and this implies that since the divergence of both sides is zero, we must have this equation,

$$\mathbf{J}_{\text{free}} + \frac{\partial \mathbf{D}}{\partial t} = \mathbf{J}_s$$

to describe the decay of the non stationary currents generated by electromagnetic fields into stationary currents. By using Ohm's law and assuming that the medium is linear we have,

$$\mathbf{D} = \epsilon \mathbf{E} = \frac{\epsilon}{\sigma} \mathbf{J}_{\text{free}}$$

so that

$$\left(1 + \frac{\epsilon}{\sigma} \frac{\partial}{\partial t}\right) \mathbf{J}_{\text{free}} = \mathbf{J}_s$$

The solution to this differential equation is

$$\mathbf{J}_{\text{free}} = \mathbf{J}_s + \mathbf{J}_0 \, e^{-t/\tau}$$

where $\mathbf{J}_0$ is the initial value of the non stationary current. There's only one important point here—and that is that the transient non stationary current dies down with a time constant

$$\tau = \frac{\epsilon}{\sigma} \tag{9.7.1}$$

the ratio of the permittivity of the medium to its conductivity. A medium such as a metal with a very large conductivity cannot support a flow of charge without rapidly re-establishing stationary conditions. In fact, in a metal the relaxation time is typically about 10^{-14} s which implies that the diffusive equation is appropriate for all frequencies lower than those in the visible spectrum. This amounts to saying that the *displacement* current is negligible compared to the *conduction* current in metals at all frequencies below about 10^{14} Hz, see figure 9–4. This is why metals are *shiny* and *opaque*.

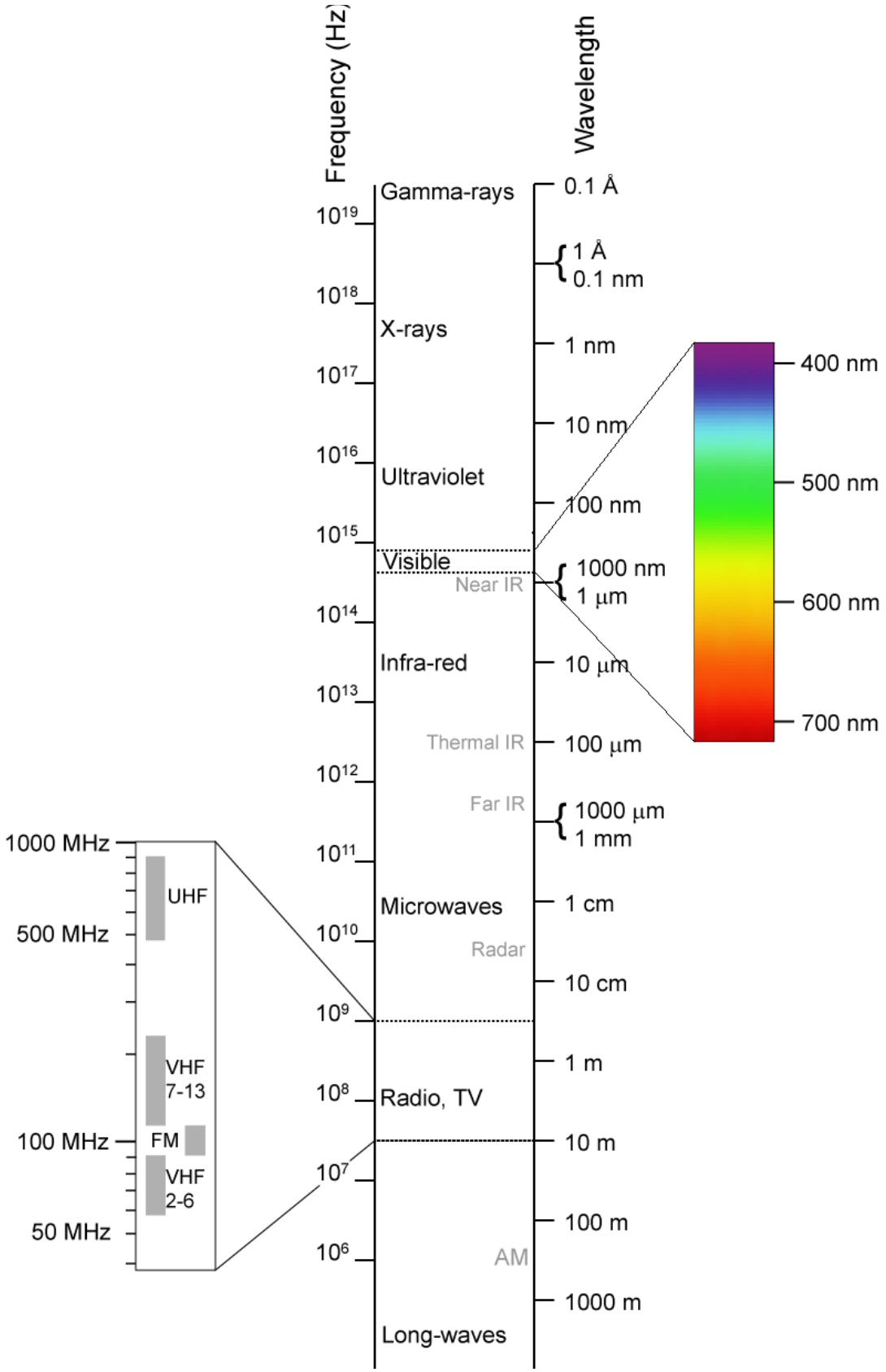


FIGURE 9–4: The electromagnetic spectrum

Victor Blacus, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=22428451

9.8 Complex Wave Equation—dispersion

Now we embark on the general solution to the coupled differential equations (9.2.1) and (9.2.2). I will follow Griffiths, section 9.4, here and first multiply (9.2.2) through by μ. The two equations are

$$\nabla^2\mathbf{E} = \epsilon\,\mu\,\frac{\partial^2\mathbf{E}}{\partial t^2} + \mu\,\sigma\,\frac{\partial\mathbf{E}}{\partial t}\ ; \qquad \nabla^2\mathbf{B} = \epsilon\,\mu\,\frac{\partial^2\mathbf{B}}{\partial t^2} + \mu\,\sigma\,\frac{\partial\mathbf{B}}{\partial t} \tag{9.8.1}$$

The solutions are the exponential wavefunctions, as (9.4.4b) and (9.4.4c) for the classical wave equation, but now the amplitude is complex,

$$\widetilde{\mathbf{E}} = \widetilde{\mathbf{E}}_0\,e^{i\tilde{\mathbf{k}}\cdot\mathbf{r}-i\omega t}\ ; \qquad \widetilde{\mathbf{B}} = \widetilde{\mathbf{B}}_0\,e^{i\tilde{\mathbf{k}}\cdot\mathbf{r}-i\omega t}$$

Since we are interested in *plane waves* let's take the propagation direction as the z-direction, $\hat{\mathbf{n}} = \hat{\mathbf{k}}$, so we can simplify this to

$$\widetilde{\mathbf{E}} = \widetilde{\mathbf{E}}_0\,e^{i\tilde{k}z-i\omega t}\ ; \qquad \widetilde{\mathbf{B}} = \widetilde{\mathbf{B}}_0\,e^{i\tilde{k}z-i\omega t} \tag{9.8.2}$$

As a matter of notation, I will use a "tilde" over a symbol to indicate that this is a complex number and I will use a "prime" and a "double prime" to denote the real and imaginary parts of these complex numbers. In this way I write,

$$\tilde{k} = k' + i\,k'' \tag{9.8.3}$$

You can check that (9.8.2) is a solution of (9.8.1) by taking partial derivatives with respect to z and t and substituting. You should do this as an exercise and you'll find that the wavenumber *is* indeed complex,

$$\tilde{k}^2 = \epsilon\mu\omega^2 + i\,\sigma\mu\omega \tag{9.8.4}$$

This reduces to (9.3.5), $k = \sqrt{\epsilon\mu}\omega$, in the case that $\sigma = 0$. Now if I put (9.8.3) into (9.8.2) I get

$$\widetilde{\mathbf{E}} = \widetilde{\mathbf{E}}_0\,e^{ik'z-i\omega t}\,e^{-k''z}\ ; \qquad \widetilde{\mathbf{B}} = \widetilde{\mathbf{B}}_0\,e^{ik'z-i\omega t}\,e^{-k''z} \tag{9.8.5}$$

The wave is now *evanescent*—the amplitude is reduced by a factor $1/e$ over a distance $1/k''$, called the "skin depth" of the medium. Just as we found in subsection 4, above, there is no propagating $\mathbf{E}$ or $\mathbf{B}$ field in the z-direction and so we must have that $\mathbf{E}$ is polarised in, say, the x-direction and in view of (9.4.5) the $\mathbf{B}$ field must be polarised in the y-direction. In this case, by comparison with (9.4.6)

$$\widetilde{\mathbf{E}} = \widetilde{E}_0\,e^{ik'z-i\omega t}\,e^{-k''z}\,\hat{\mathbf{i}}\ ; \qquad \widetilde{\mathbf{B}} = \frac{\tilde{k}}{\omega}\widetilde{E}_0\,e^{ik'z-i\omega t}\,e^{-k''z}\,\hat{\mathbf{j}} \tag{9.8.6}$$

If I express the complex wavevector in the form†

$$\tilde{k} = |\tilde{k}|\,e^{i\phi}$$

† For any complex number $z = a + ib$ I can cast this into "polar form" $z = |z|e^{i\phi}$, where $\tan\phi = b/a$, and $|z| = \sqrt{(a+ib)(a-ib)} = \sqrt{a^2+b^2}$, recalling the Argand diagram.

then

$$\phi = \arctan \kappa$$

where†

$$\kappa = \frac{k''}{k'} = \frac{\text{Im } \tilde{k}}{\text{Re } \tilde{k}} \tag{9.8.7}$$

You can see from (9.8.6) that the complex amplitude of the **B** field is

$$\widetilde{B}_0 = \frac{\tilde{k}}{\omega}\, \widetilde{E}_0$$

Therefore separating into polar form (I won't bother to put a "prime" on the real parts of $\widetilde{E}_0$ and $\widetilde{B}_0$),

$$B_0\, e^{i\phi_B} = \frac{1}{\omega}\, |\tilde{k}|\, e^{i\phi}\, E_0 e^{i\phi_E}$$

The imaginary parts on either side must be equal, and so the phases of the **E** and **B** fields are related to the argument of the complex wavenumber,

$$\phi_B - \phi_E = \phi > 0$$

and you see that the magnetic field *lags behind* the electric field as long as the wavenumber is complex. Now if I take the real parts of $\widetilde{\mathbf{E}}$ and $\widetilde{\mathbf{B}}$ (9.8.6) and put in the phase difference I get for the oscillating electromagnetic wave in a conducting medium,

$$\mathbf{E}(z,t) = E_0\, e^{-k''z}\, \cos\left(k'z - \omega t + \phi_E\right) \hat{\mathbf{i}} \tag{9.8.8a}$$

$$\mathbf{B}(z,t) = \frac{|\tilde{k}|}{\omega}\, E_0\, e^{-k''z}\, \cos\left(k'z - \omega t + \phi_E + \phi\right) \hat{\mathbf{j}} \tag{9.8.8b}$$

Compare these with (9.4.6). The above equations reduce to these in the case that $\sigma = 0$.

In Problems Set 8 you will calculate the energy density for the evanescent wave as we did above in subsection 5. You will also calculate the intensity, which is always proportional to the *square* of the amplitude and you will find,

$$I = \frac{k'}{2\mu\omega}\, E_0^2\, e^{-2k''z}$$

Often the quantity $2k''$ is called the absorption coefficient, $\alpha(\omega)$, written as a function of frequency.

---

† Sorry, I'm using k'' where Griffiths uses κ and I'm using κ for the ratio k''/k' of the imaginary and real parts of the wavenumber.

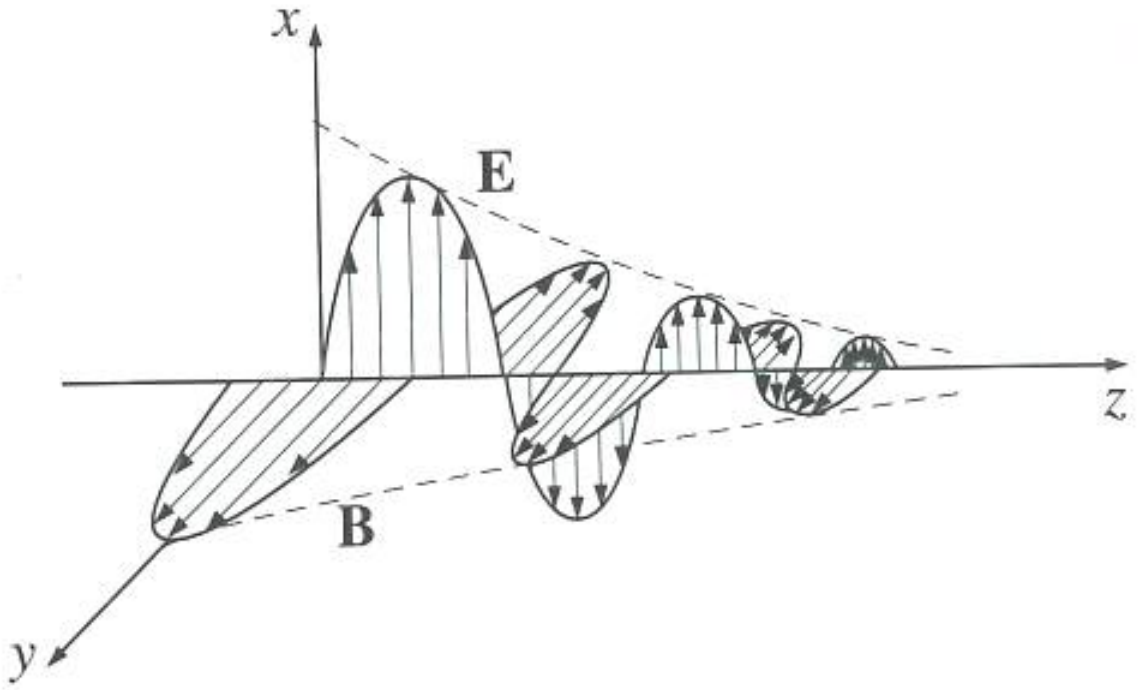


FIGURE 9–5: The electric and magnetic fields in an electromagnetic wave propagating in dispersive medium. **E** leads **B** by ϕ and the phase velocity is v. Taken from Griffiths.

9.9 Index of refraction and coefficient of absorption

Please be sure that you understand that nothing observable is ever imaginary. The use of a complex wavefunction is merely a mathematical device. This is why at the end of the last subsection we took the real part of the fields to get the actual physical electric and magnetic fields. On the other hand we do see very clear physics behind the notion of a complex wavenumber. The real part, k', is 2π divided by the wavelength of the wave. Moreover, the phase velocity is ω/k' and so the refractive index is $n(\omega) = c/v = ck'/\omega$. The imaginary part, k'', is a measure of the distance over which the amplitude decays in space; $\alpha = 2k''$ is a measure of the distance over which energy is absorbed by the medium. The difference is that now k' and k'' *are functions of the frequency*, ω. (Use the footnote on the next page and equation (9.8.4) to find formulas for k' and k'' in terms of ω, ϵ, μ and σ.)

Another important point that I want to get across is that by treating ϵ and σ as complex numbers takes us outside the realm of classical electromagnetism. Maxwell's equations do not admit complex ϵ and σ; we make them complex as a device for solving (9.8.1) which equations contain only real numbers. In this way, as Sommerfeld emphasises, optics is an extension of formal electrodynamics. In fact Sommerfeld writes,† "In Maxwell's

† *Optics*, translated by O. Laporte and P Moldauer, (New York, Academic, 1954) p. 34.

theory metals are characterised by the conductivity σ. However actual conduction of electricity has to be thought of as a phenomenon consisting of the interactions between free electrons and metallic ions at fixed positions and as being brought about by an averaging of many elementary processes. Only in stationary or slowly varying fields does this averaging lead to a constant which is independent of frequency. One cannot expect that the phenomenological Maxwell theory will suffice in the visible region of the spectrum. We have already encountered the failure of the theory in the optics of transparent media (the failure of water, for instance, to satisfy the Maxwell relation $n^2 = \epsilon_r$.)" See point 2 on page 4 of these notes. I have tried to keep to a consistent notation to reflect this by reserving ϵ and σ for the Maxwell equation permittivity and conductivity, necessarily real and in association with constant or slowly varying fields; and then using ϵ' and σ' as the real parts of the *complex* permittivity and conductivity. I hope this hasn't confused you.

You will find in your reading that very generally in physics and especially in solid state physics and optics, the notion of complex wavenumber is extended to complex dielectric constant and complex refractive index (see also Problems Set 8). Generally speaking the imaginary part of any such physical quantity gives a measure of the amount of dissipation—if you like, the relative weight of the diffusive term in the complex wave equation. Let us start with equation (9.8.4) which gives the real and imaginary parts of the square of the complex wavenumber. If I write

$$\tilde{k} = \sqrt{\tilde{\epsilon}\mu}\omega \tag{9.9.1}$$

Then I have defined a *complex permittivity*

$$\tilde{\epsilon} = \epsilon' + \frac{i\sigma}{\omega} \tag{9.9.2}$$

in which σ is the Ohm's law, or DC, conductivity. You can see this by putting (9.9.2) into (9.9.1) and squaring† both sides to recover (9.8.4).

† When I "square" a complex number $z = a + ib$, I mean $a^2 + b^2 + 2iab$. If I take the square root of this complex number, z, I get rather horrible formulas. If

$$\sqrt{a + ib} = c + id$$

then c and d are the real and imaginary parts of the square root of z. They are

$$c = \frac{1}{\sqrt{2}}\sqrt{\sqrt{a^2 + b^2} + a}$$
$$d = \frac{\text{sign } b}{\sqrt{2}}\sqrt{\sqrt{a^2 + b^2} - a}$$

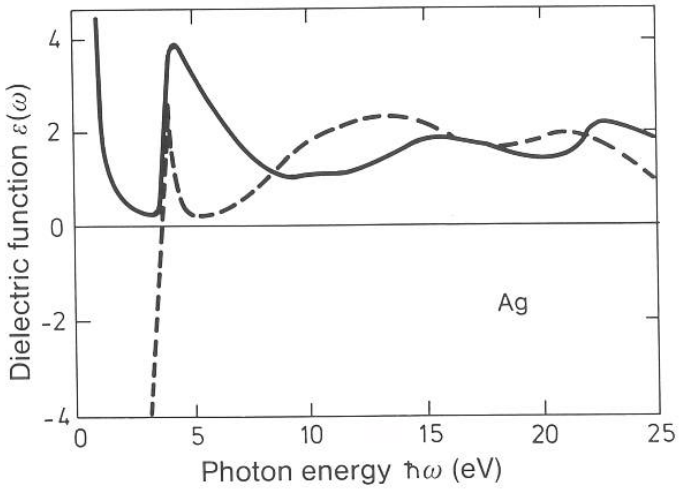


FIGURE 9–6: Real (broken line) and imaginary parts of the complex dielectric constant, $\tilde{\epsilon}_r$, as a function of photon energy (you can convert this to frequency) in silver.

Figure 9–6 shows real and imaginary parts of the relative permittivity of silver metal. These are clearly complicated functions of the frequency (photon energy). You can see from figure 9–4 that since $\hbar\omega = 5$ corresponds to a frequency of 7.6×10^{15} Hz while $\hbar\omega = 25$ corresponds to a frequency of 3.8×10^{16} Hz, these are *ultraviolet* frequencies. These are sufficiently high frequencies that only the electrons are excited by the electromagnetic wave; the nuclei are too massive to respond. Note the striking "anomalous dispersion": below about 5 eV the real part of the dielectric constant is negative. This means that the instantaneous response, $\mathbf{D}$, to the applied field, $\mathbf{E}$, is in the opposite direction. This reflects that in this regime the electrons are reacting about 180° out of phase with the electromagnetic field. I'll describe a model that demonstrates this in the next subsection.

It is also common to define a complex refractive index; by comparison with (9.3.7) and (9.8.7),

$$\tilde{n} = \sqrt{\frac{\tilde{\epsilon}\mu}{\epsilon_0\mu_0}} \approx \sqrt{\tilde{\epsilon}_r} \tag{9.9.3a}$$

$$\begin{aligned}\tilde{n} &= n' + in'' \\ &= n'\,(1 + i\kappa)\end{aligned} \tag{9.9.3b}$$

from which it follows that

$$\begin{aligned}\tilde{k} &= \frac{\omega}{c}\,\tilde{n} \\ &= \frac{\omega}{c}\,n'\,(1 + i\kappa)\end{aligned} \tag{9.9.4}$$

Also, if we use the approximation (9.9.3a) then

$$\frac{1}{\epsilon_0}\,(\epsilon' + i\epsilon'') = (n' + in'')^2$$

and by equating real and imaginary parts, it follows that,

$$\epsilon' = \epsilon_0 \left(n'^2 + n''^2\right) \tag{9.9.5a}$$
$$\epsilon'' = 2\epsilon_0 n' n'' \tag{9.9.5b}$$
$$= \frac{\sigma}{\omega} \tag{9.9.5c}$$

The last line follows from (9.9.2). The *absorption coefficient* is then

$$\alpha(\omega) = 2\omega \,\mathrm{Im}\sqrt{\tilde{\epsilon}\mu} \tag{9.9.6a}$$
$$= 2k'' = 2\frac{\omega}{c}\,\kappa\, n' \tag{9.9.6b}$$

using (9.9.4). The *index of refraction* or refractive index is, from (9.9.3a) and (9.9.1),

$$n(\omega) = \mathrm{Re}\sqrt{\frac{\tilde{\epsilon}\mu}{\epsilon_0\mu_0}} \tag{9.9.7a}$$
$$= \frac{c}{\omega}k' \tag{9.9.7b}$$
$$= n' \tag{9.9.7c}$$

and is now *frequency dependent* in contrast to (9.3.7). This means that electromagnetic waves of different frequency and light of different colours travel at different speeds in a dispersive medium. Now we can understand the way a prism splits white light into colours, which we could not do with the classical wave equation without simply asserting that the relative permittivity was frequency dependent.

From a practical point of view the absorption coefficent and refractive index as functions of frequency are the most relevant material properties in optics. In (9.9.6) and (9.9.7) I have given equivalent representations to make it easier to follow the experimental literature. In fact, as you will have noticed, there is no real reason for me to introduce the quantity κ in (9.8.7). I have done so because the sublime Arnold Sommerfeld uses the forms (9.9.3b) and (9.9.4) and indeed he calls κ the "absorption coefficent", not $2k''$. Since he invented most of this theory it is fitting to include this, and it makes some sense to use a dimensionless quantity for this rather than one with dimensions $[\mathrm{L}]^{-1}$; moreover κ serves as a *ratio* of the amount of absorption and the the amount of propagation.

As you know the conductivity of matter varies hugely between about 10^7 $\mathrm{Ohm}^{-1}\mathrm{m}^{-1}$ in silver to as little as 10^{-25} $\mathrm{Ohm}^{-1}\mathrm{m}^{-1}$ in Teflon. In figure 9–7 I show data from a recent paper on the conductivity of a complex glass which is doped with silver ions. It is notable that this is not conductivity by *electrons* but by Ag^{++} *ions*. In contrast to figure 9–6, these frequencies are low enough for the ions to be mobile. Since this is a log–log plot it is striking how strong is the frequency dependence of both real and imaginary parts of the conductivity.

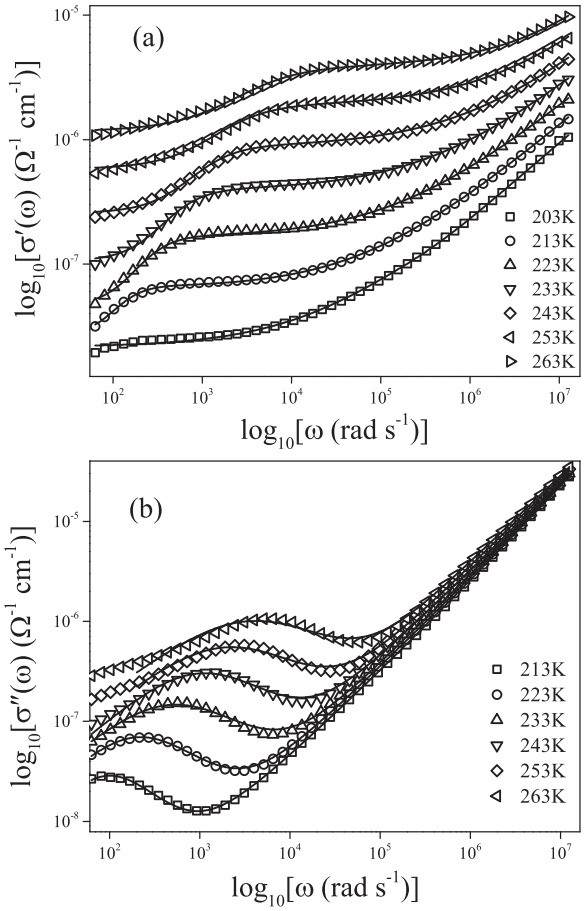


FIGURE 9–7: Real (a) and imaginary (b) parts of the complex conductivity as a function of frequency in a complex glass doped with silver iodide. (*J. Applied Phys.*, **121**, 125104 (2017))

9.10 A model for the complex dielectric constant

In this section we enquire into how a frequency dependent and complex dielectric response might come about. If an oscillating electric field is attempting to polarise a dielectric substance then the field will be causing the charged electrons and ions to respond by oscillating in position and this will cause an oscillating polarisation. Since the complex wave equation (9.8.1) contains propagating and absorbing (diffusive) contributions a good model is to pretend that the electrons and ions are anchored in place by springs of stiffness k_s and that their motion is damped by a velocity dependent force. Therefore the equation of motion of an electron of charge, $-e$ in the field is

$$m\frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2} + m\gamma\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} + m\omega_0^2\mathbf{r} = -e\mathbf{E} = -e\,\mathbf{E}_0 e^{-i\omega t}$$

It should be understood that in this context $\mathbf{E}$ is the *local field*, $\mathbf{E}_\mathrm{local}$, (see section 7, page 13). Here $\omega_0 = \sqrt{k_\mathrm{s}/m}$ is the natural frequency of the charge having mass m and $m\gamma$ is the damping coefficient. The damping ratio is $\zeta = \gamma/2\omega_0$, and γ has units of s^{-1}.

The amplitude of the oscillation is (see additional notes "simple harmonic motion" on KEATS),

$$\mathbf{r} = \frac{-e\mathbf{E}/m}{(\omega_0^2 - \omega^2) - i\gamma\omega}$$

If we recall that the polarisation is $\mathbf{P} = -e\mathbf{r}N$ where N is the number of oscillators per unit volume then

$$\mathbf{P} = \frac{(e^2N/m)}{(\omega_0^2 - \omega^2) - i\gamma\omega}\,\mathbf{E} = \epsilon_0\left(\frac{\tilde{\epsilon}}{\epsilon_0} - 1\right)\mathbf{E}$$

using (7.6.1) from section 7, page 13. This leads us to a *complex dielectric constant*,

$$\tilde{\epsilon}_r = \frac{\tilde{\epsilon}}{\epsilon_0} = 1 + \frac{(e^2N/\epsilon_0 m)}{(\omega_0^2 - \omega^2) - i\gamma\omega}$$

I can take real and imaginary parts of this to get, in comparison to (9.9.5),

$$\epsilon'_r = \left(n'^2 - n''^2\right) = 1 + \sum \frac{e^2N}{\epsilon_0 m}\,\frac{\omega_0^2 - \omega^2}{\left(\omega_0^2 - \omega^2\right)^2 + \gamma^2\omega^2} \tag{9.10.1a}$$

$$\epsilon''_r = 2n'n'' = \frac{\sigma}{\omega} = \sum \frac{e^2N}{\epsilon_0 m}\,\frac{\gamma\omega}{\left(\omega_0^2 - \omega^2\right)^2 + \gamma^2\omega^2} \tag{9.10.1b}$$

I have put a "summation sign" in a symbolic way to indicate that in any material there will be a very large number of oscillators with different "masses" and charges, since they may be electrons or even ions or nuclei with charges that are multiples of e. The remarkable fact is that equations (9.10.1)—the "Drude-Lorentz" oscillators—are very general and have been found to describe most phenomena in matter that give rise to propagation and absorption. The same formulas even emerge from a quantum mechanical analysis in which the numerators are replaced by the so called *oscillator strengths*, f_i, which obey the "f-sum rule" discovered by Hans Bethe in 1930. Each term in (9.10.1) corresponds to some oscillation of the medium that has a resonant frequency ω_0. Clearly then if the electromagnetic wave has a frequency close to the resonant frequency of a particular process then there will be characteristic behaviour of the optical constants at that frequency which will be rather general in shape. This is illustrated in figure 9–8 which shows the real and imaginary parts of $\tilde{\epsilon}_r$ and $\tilde{n}$ as functions of frequency. The figure uses an example of $\hbar\omega_0 = 4$ eV and a damping of γ/ω_0=0.25. Note what happens at the resonance, $\hbar\omega_0 = 4$ eV (which is in the ultra violet). If there were no damping then $\gamma = 0$ and the polarisation would diverge. Hence the necessity to include damping and admit a complex dielectric constant. At resonance there is a peak in ϵ''_r; and ϵ'_r goes with negative slope through zero: this is called "anomalous absorption". At frequencies far from ω_0, ϵ''_r goes to zero and the dielectric constant is real. The graph of the refractive index is divided into regions corresponding to transmission (T), absorption, (A) and reflection, (R). The greatest absorption is near resonance where the oscillator can be driven with the greatest amplitude and absorb the energy of the electromagnetic wave which thereby becomes evanescent (9.9.6). Far from resonance, there is no opportunity to set the matter's particles into oscillation and the wave is transmitted. It's very instructive to compare figure 9–8 with figure 9–6. There you see that the real part of the dielectric function is rising from the left out of a

resonance at low frequency and that the real part is falling from an associated absorption peak. This must be at visible light frequency where we know that silver is reflecting and absorbing. The structure in figure 9–6 at higher frequencies is much more involved than the model in figure 9–8 for a single oscillator and this reflects all the collective and single particle oscillations of the metallic electron gas.

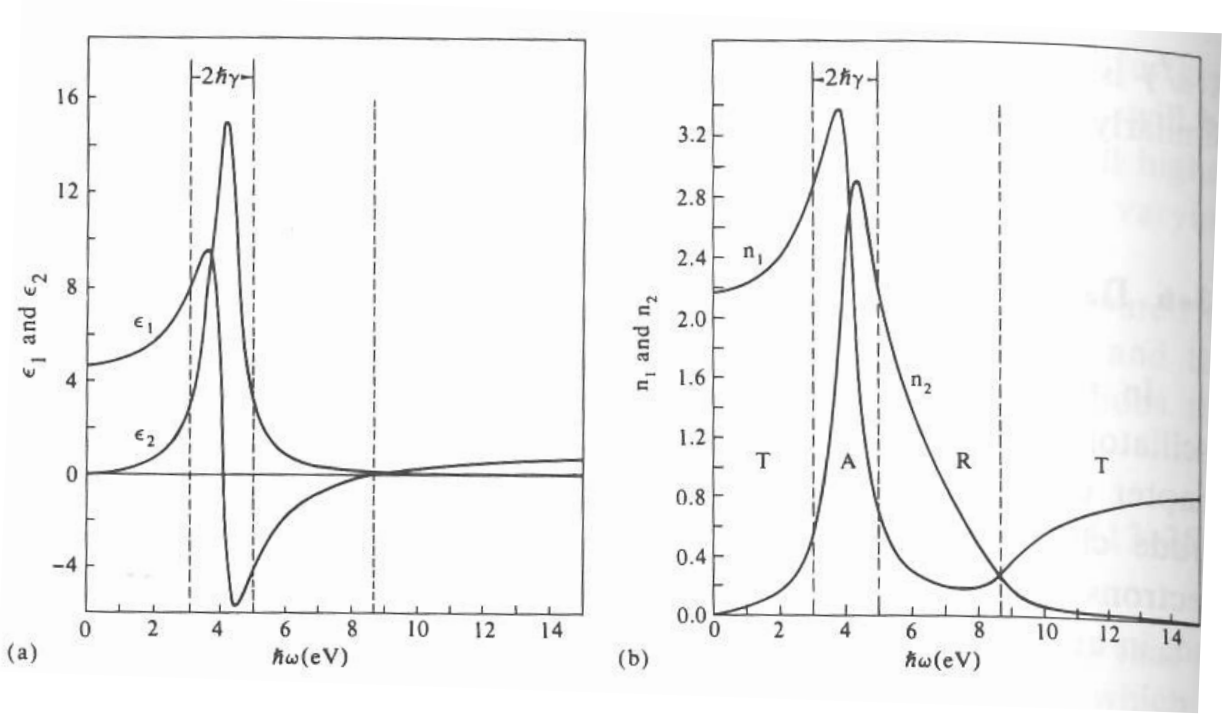


FIGURE 9–8: Real and imaginary parts of the dielectric constant and refractive index for a single Drude–Lorentz oscillator (9.10.1). After F. Wooten, *Optical Properties of Solids*, (New York, Academic Press, 1972) who uses subscripts "1" and "2" in place of our prime and double prime for real and imaginary parts.

9.11 Reflection and refraction

We turn now to *reflection* and *refraction*—with which you are already familiar. We shall *derive* the rules of geometric optics.

9.11.1 Boundary conditions

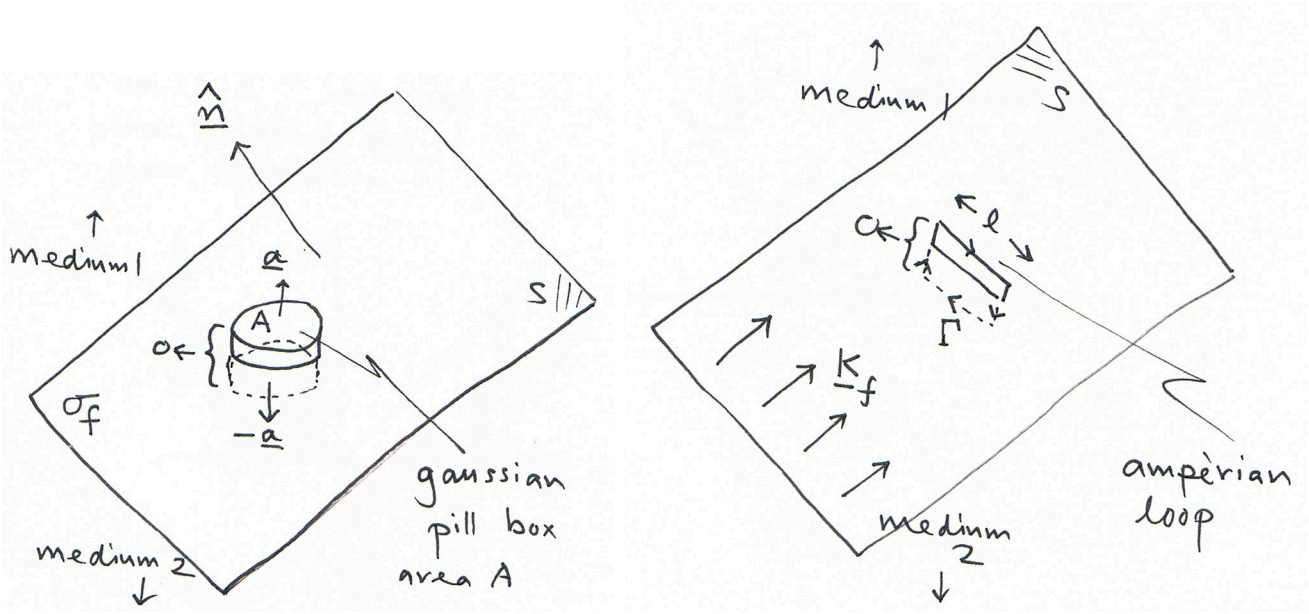


FIGURE 9–9

First we focus on the interface, or boundary, between two media of different optical constants. We ask, what conditions are imposed on the fields across the boundary? Look at figure 9–9. A boundary is shown between medium 1 and medium 2. The fields immediately at either side of the boundary will be denoted with subscripts 1 and 2. Consider, on the left, a gaussian pill box which straddles the boundary with areas, A, above and below, but with vanishingly small width. According to Gauss's law,

$$\boldsymbol{\nabla}\cdot\mathbf{D} = Q_{\mathrm{f,enclosed}} = \sigma_{\mathrm{free}}A$$

It follows that

$$\mathbf{D}_1\cdot\mathbf{a} - \mathbf{D}_2\cdot\mathbf{a} = \sigma_{\mathrm{free}}A$$

where $\mathbf{a}$ is the vector of length A directed normal to the upper surface. The scalar products represent the magnitude of the component of $\mathbf{D}$ that is projected normal to the boundary times A. Therefore, at the boundary,

$$D_1^{\perp} - D_2^{\perp} = \sigma_{\mathrm{free}}$$

and in linear media,

$$\epsilon_1 E_1^{\perp} - \epsilon_2 E_2^{\perp} = \sigma_{\mathrm{free}}$$

If there is no free charge at the boundary then,

$$D_1^{\perp} = D_2^{\perp}$$

and in linear media,

$$\epsilon_1 E_1^\perp = \epsilon_2 E_2^\perp$$

By the same argument applied to $\boldsymbol{\nabla} \cdot \mathbf{B} = 0$ we must have,

$$B_1^\perp = B_2^\perp$$

and in linear media,

$$\mu_1 H_1^\perp = \mu_2 H_2^\perp$$

This means that in the absence of free surface charge, *the normal component* of $\mathbf{D}$ is continous across the boundary; and always the *the normal component* of $\mathbf{B}$ is continous across the boundary.

What about the parallel components? Look at the right hand cartoon in figure 9–9. I've drawn an ampèrian loop whose long arms are parallel to the boundary and whose short arms are vanishingly small. We consider line integrals around the path marked with arrows. According to Faraday's law,

$$\oint_\Gamma \mathbf{E} \cdot \mathrm{d}\boldsymbol{\ell} = -\frac{\partial}{\partial t} \int_S \mathbf{B} \cdot \mathrm{d}\mathbf{a}$$

But if the width is zero then the area bounded by Γ is zero and so the right hand side is zero. This leaves,

$$\mathbf{E}_1 \cdot \boldsymbol{\ell} - \mathbf{E}_2 \cdot \boldsymbol{\ell} = 0$$

Therefore what remains of the vector $\mathbf{E}$ after taking away the perpendicular component lies in the boundary and this vector, $\mathbf{E}^\parallel$, is therefore continuous across the boundary,

$$\mathbf{E}_1^\parallel = \mathbf{E}_2^\parallel$$

By a similar arguments based on Ampère's law, we must have,

$$\mathbf{H}_1 \cdot \boldsymbol{\ell} - \mathbf{H}_2 \cdot \boldsymbol{\ell} = I_{\mathrm{f,enclosed}}$$

Again, because the loop is infinitesimally narrow, the only enclosed current can arise from a free surface current, $\mathbf{K}_\mathrm{f}$, and the enclosed current must be the component of $\mathbf{K}_\mathrm{f}$ normal to the loop,

$$I_{\mathrm{f,enclosed}} = \mathbf{K}_\mathrm{f} \cdot (\hat{\mathbf{n}} \times \boldsymbol{\ell}) = (\mathbf{K}_\mathrm{f} \times \hat{\mathbf{n}}) \cdot \boldsymbol{\ell}$$

where $\hat{\mathbf{n}}$ is the unit vector normal to the boundary plane. So the discontinuity in the parallel components of $\mathbf{H}$ is

$$\mathbf{H}_1 \cdot \boldsymbol{\ell} - \mathbf{H}_2 \cdot \boldsymbol{\ell} = (\mathbf{K}_\mathrm{f} \times \hat{\mathbf{n}}) \cdot \boldsymbol{\ell}$$

We particularly want to apply these rules to the field vectors of an electromagnetic wave as it strikes the interface between two media—for example monochromatic light reaching the surface of a glass plate in air. We do not expect there to be free charge or

current in these circumstances so we may apply the following boundary conditions on the fields infinitesimally close the interface on either side,

$$\epsilon_1 E_1^\perp = \epsilon_2 E_2^\perp \tag{9.11.1a}$$
$$\mu_1 H_1^\perp = \mu_2 H_2^\perp \tag{9.11.1b}$$
$$\mathbf{E}_1^\parallel = \mathbf{E}_2^\parallel \tag{9.11.1c}$$
$$\mathbf{H}_1^\parallel = \mathbf{H}_2^\parallel \tag{9.11.1d}$$

Note that the perpendicular components are scalar because we know that the direction is along $\hat{\mathbf{n}}$; but having subtracted that off, what is left *lies in the boundary* but it's direction in the boundary is arbitrary. Hence the parallel components are vectors. Mathematically, what we're saying is that, for example,

$$\mathbf{E} = E^\perp \hat{\mathbf{n}} + \mathbf{E}^\parallel$$

9.11.2 Laws of geometric optics

Figure 9–10 shows the well known geometry associated with a "ray" of light striking a surface; in general a part of the light is reflected and a part is transmitted, or "refracted," into the second medium. Our notation is to use prime and double-prime to indicate the refracted and reflected components.

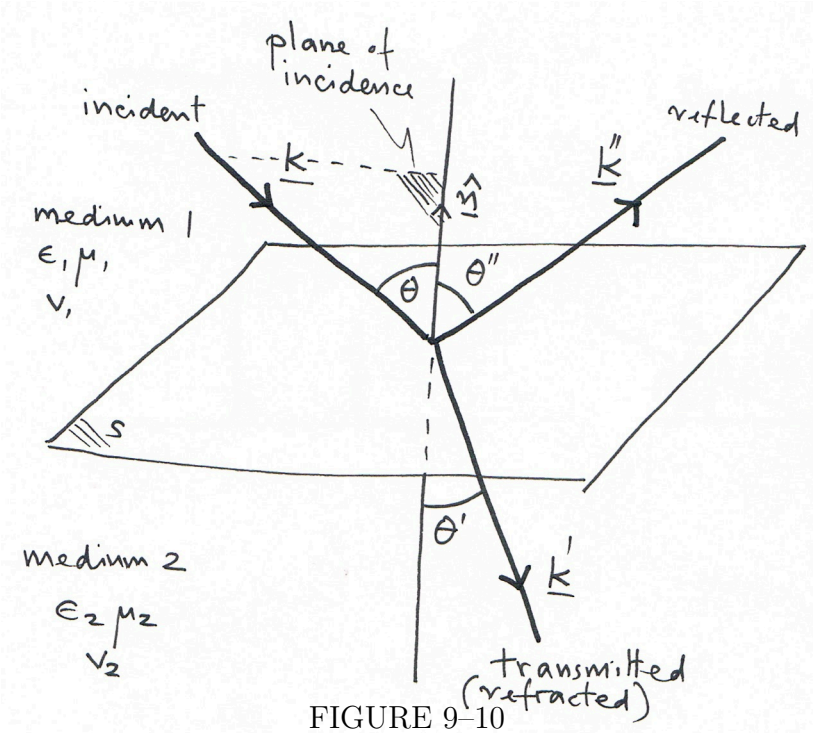


FIGURE 9–10

According to (9.4.4b) and (9.4.5) the wavefunction of the incident EM wave is,

$$\mathbf{E} = \mathbf{E}_0\, e^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)} \tag{9.11.2a}$$
$$\mathbf{H} = \frac{1}{\omega\mu_1}\,\mathbf{k}\times\mathbf{E} \tag{9.11.2b}$$

We now write equivalent expressions for the transmitted and reflected wavefunctions,

$$\mathbf{E}' = \mathbf{E}'_0\, e^{i(\mathbf{k}'\cdot\mathbf{r}-\omega t)} \qquad \mathbf{H}' = \frac{1}{\omega\mu_2}\,\mathbf{k}'\times\mathbf{E}' \tag{9.11.2c}$$

$$\mathbf{E}'' = \mathbf{E}''_0\, e^{i(\mathbf{k}''\cdot\mathbf{r}-\omega t)} \qquad \mathbf{H}'' = \frac{1}{\omega\mu_1}\,\mathbf{k}''\times\mathbf{E}'' \tag{9.11.2d}$$

At the interface, in view of (9.11.1) the tangential components of $\mathbf{E}$ and $\mathbf{H}$ are continuous. This means that the *exponents* in (9.11.2) must all be equal. We already know that the frequency must be same as only the speed and wavelength of light is affected by a change of medium. We now also see that,

$$\mathbf{k}\cdot\mathbf{r} = \mathbf{k}'\cdot\mathbf{r} = \mathbf{k}''\cdot\mathbf{r} \tag{9.11.3}$$

If we place the origin of the vector $\mathbf{r}$ somewhere within the boundary then everywhere on the boundary,

$$\hat{\mathbf{n}}\cdot\mathbf{r} = 0$$

and

$$\hat{\mathbf{n}}\times(\hat{\mathbf{n}}\times\mathbf{r}) = (\hat{\mathbf{n}}\cdot\mathbf{r})\,\hat{\mathbf{n}} - \mathbf{r} = -\mathbf{r}$$

which follows from the triple vector product rule. I put this into (9.11.3). Firstly,

$$-\mathbf{k}\cdot\mathbf{r} = \mathbf{k}\cdot(\hat{\mathbf{n}}\times(\hat{\mathbf{n}}\times\mathbf{r})) = (\mathbf{k}\times\hat{\mathbf{n}})\cdot(\hat{\mathbf{n}}\times\mathbf{r})$$

using the vector product rule, $\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c}) = (\mathbf{a}\times\mathbf{b})\cdot\mathbf{c}$. Then (9.11.3) is the same as,

$$(\mathbf{k}\times\hat{\mathbf{n}})\cdot(\hat{\mathbf{n}}\times\mathbf{r}) = (\mathbf{k}'\times\hat{\mathbf{n}})\cdot(\hat{\mathbf{n}}\times\mathbf{r}) = (\mathbf{k}''\times\hat{\mathbf{n}})\cdot(\hat{\mathbf{n}}\times\mathbf{r})$$

which is true within the boundary for any arbitrary vector $\mathbf{r}$. Hence it must be true that,

$$(\mathbf{k}\times\hat{\mathbf{n}}) = (\mathbf{k}'\times\hat{\mathbf{n}}) = (\mathbf{k}''\times\hat{\mathbf{n}})$$

This means that all three $\mathbf{k}$-vectors are *coplanar*—they all lie in the *plane of incidence*, which is the plane defined by the direction of the incoming ray and the normal to the boundary. This is the well known first rule of geometric optics. By the definition of the vector cross product,

$$k\sin\theta = k''\sin\theta'' = k'\sin\theta'$$

Recalling that $k = \omega/v$ where v is the phase velocity in the medium, it's clear that $k = k''$ since the incident and reflected wave are in the same medium, and

$$\frac{k}{k'} = \frac{v_2}{v_1}$$

We therefore must have,

$$\theta'' = \theta$$

the *angle of reflection is equal to the angle of incidence* which is the second rule of geometric optics. It also follows clearly that

$$\frac{v_1}{v_2} = \frac{k'}{k} = \frac{\sin\theta}{\sin\theta'} = \frac{n_2}{n_1} \overset{\text{def}}{=} n \tag{9.11.4}$$

which is the third rule of geometric optics, also known as Snell's law. *For this subsection, 9.11, only* we will use n to denote the ratio of the refractive indices of medium 2 and medium 1, as indicated in (9.11.4). In the case that medium 1 is vacuum, or air, then n is the refractive index of medium 2.

Please note that we have obtained all three laws of geometric optics simply from the conditions of continuity of the field in the absence of surface sources. It's not suprising then that these rules apply to many situations of waves transferring from one medium to another in which they travel at different speeds, for example, sound waves.

9.11.3 Fresnel's equations

The purpose of this section is to outline exactly how you would go about, for a particular pair of media, wavelength of light and angle of incidence, finding the intensities of reflected and transmitted waves. You will also be able to find the Brewster angle and find the conditions for total internal reflection as found, say, in an optic fibre cable.

The total field just at the side of the boundary lying in medium 1 is the sum of the field vectors of the incident and reflected waves. This must match exactly the field in medium 2 immediately on the other side of the boundary, in accord with our conditions (9.11.1). Because the *tangential* $\mathbf{E}$ and $\mathbf{H}$ fields must be continuous, we express this by equating the cross products of the total fields with the normal to the boundary,

$$\begin{aligned} \hat{\mathbf{n}} \times (\mathbf{E} + \mathbf{E}'') &= \hat{\mathbf{n}} \times \mathbf{E}' \\ \hat{\mathbf{n}} \times (\mathbf{H} + \mathbf{H}'') &= \hat{\mathbf{n}} \times \mathbf{H}' \end{aligned} \tag{9.11.5a}$$

Because

$$\mathbf{H} = \frac{1}{\omega\mu}\mathbf{k} \times \mathbf{E} \tag{9.4.6}$$

and ω is the same for all waves, the second of the pair of equations above reads,

$$\frac{1}{\mu_1}\hat{\mathbf{n}} \times (\mathbf{k} \times \mathbf{E} + \mathbf{k}'' \times \mathbf{E}'') = \frac{1}{\mu_2}\hat{\mathbf{n}} \times (\mathbf{k}' \times \mathbf{E}') \tag{9.11.5b}$$

In view of the remarks in this section, after equation (9.3.7) we will take $\mu_1 = \mu_2$ in all that follows.

We are ready to ask, what are the relations between $\mathbf{E}$, $\mathbf{E}''$ and $\mathbf{E}'$? Since the intensity of radiation is proportional to the electric field vector *squared*, we will also have access to the intensities of reflected and transmitted waves. We could also calculate Poynting vectors; although we will not carry out a full programme of this, but leave it to your upcoming courses in optics.

Now, $\mathbf{E}$ and $\mathbf{H}$ are *perpendicular* to $\mathbf{k}$ because the wave is *transverse*—the field vectors are perpendicular to the propagation direction. Otherwise the orientation of $\mathbf{E}$ with respect to the boundary plane is arbitrary. In order to proceed, we we consider two

orthogonal directions for **E** so that a result can be obtained for any orientation of **E** by making linear combinations of the two. These two cases are, **1.** **E** *perpendicular to* the plane of incidence, and **2.** **E** *lying in* the plane of incidence, see figure 9–11.

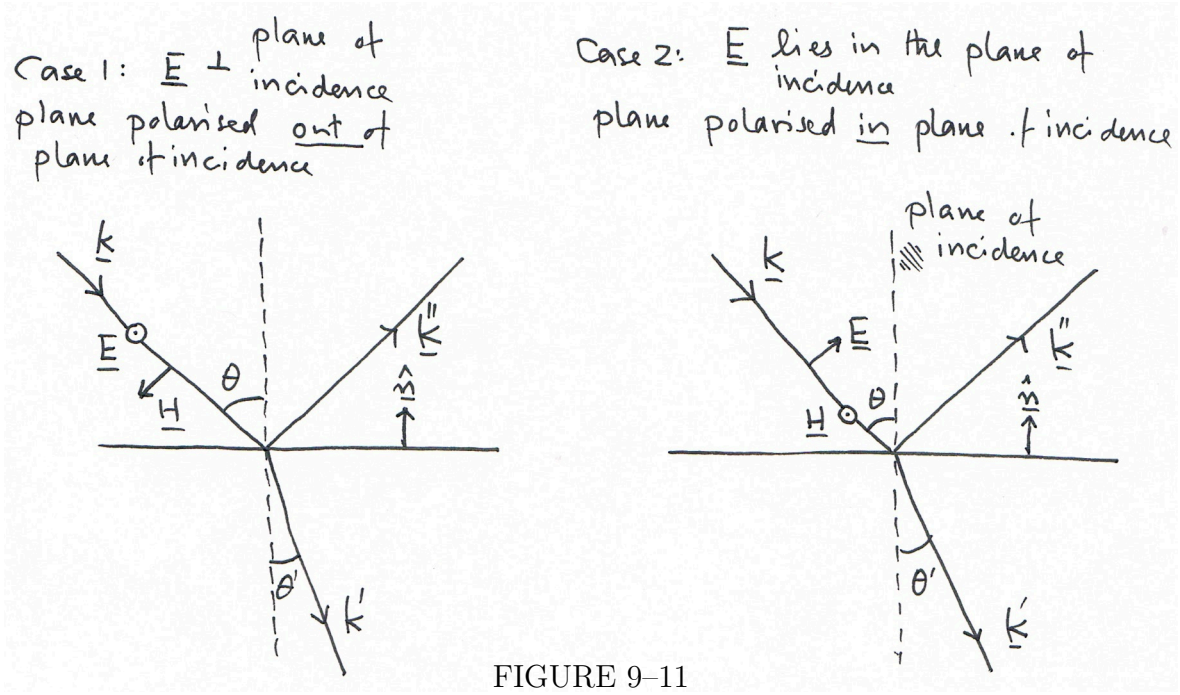


FIGURE 9–11

Case 1: If **E** is perpendicular to the plane of incidence, figure 9–11 left, then all three **E** vectors are tangential to the boundary plane. So the amplitudes just above and below the plane must be equal,

$$E + E'' = E' \tag{9.11.6a}$$

I apply the triple product rule, $\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = \mathbf{b}(\mathbf{a} \cdot \mathbf{c}) - \mathbf{c}(\mathbf{a} \cdot \mathbf{b})$, to the three terms in (9.11.5b),

$$\begin{aligned}
\hat{\mathbf{n}} \times (\mathbf{k} \times \mathbf{E}) &= \mathbf{k}\,(\hat{\mathbf{n}} \cdot \mathbf{E}) - \mathbf{E}\,(\hat{\mathbf{n}} \cdot \mathbf{k}) \\
\hat{\mathbf{n}} \times (\mathbf{k} \times \mathbf{E}'') &= \mathbf{k}''\,(\hat{\mathbf{n}} \cdot \mathbf{E}'') - \mathbf{E}''\,(\hat{\mathbf{n}} \cdot \mathbf{k}'') \\
\hat{\mathbf{n}} \times (\mathbf{k} \times \mathbf{E}') &= \mathbf{k}'\,(\hat{\mathbf{n}} \cdot \mathbf{E}') - \mathbf{E}'\,(\hat{\mathbf{n}} \cdot \mathbf{k}')
\end{aligned}$$

and since the **E** vectors are all lying in the plane of the interface,

$$\hat{\mathbf{n}} \cdot \mathbf{E} = \hat{\mathbf{n}} \cdot \mathbf{E}'' = \hat{\mathbf{n}} \cdot \mathbf{E}' = 0$$

When I put this back into (9.11.5b), allowing that $\mu_1 = \mu_2$, I get,

$$\mathbf{E}\,(\hat{\mathbf{n}} \cdot \mathbf{k}) + \mathbf{E}''\,(\hat{\mathbf{n}} \cdot \mathbf{k}'') = \mathbf{E}'\,(\hat{\mathbf{n}} \cdot \mathbf{k}')$$

But we have agreed that the exponential factors in the wavefunctions must be equal at the boundary so this implies that,

$$E\,(\hat{\mathbf{n}} \cdot \mathbf{k}) + E''\,(\hat{\mathbf{n}} \cdot \mathbf{k}'') = E'\,(\hat{\mathbf{n}} \cdot \mathbf{k}')$$

which is the same as,

$$E\cos\theta - E''\cos\theta'' = nE'\cos\theta' \tag{9.11.6b}$$

where $n = k'/k$ (9.11.4). Check figure 9–11 to see where the minus sign comes from (the opposite directions of $\mathbf{k}$ and $\mathbf{k}''$). Now I combine (9.11.6a) and (9.11.6b) and do some simple but tedious algebra to find these formulas for the ratios of the amplitudes of the transmitted and reflected waves to the amplitude of the incident wave.

$$\frac{E'}{E} = \frac{2\cos\theta}{\cos\theta + n\cos\theta'} = \frac{2\cos\theta\sin\theta'}{\sin(\theta+\theta')} \tag{9.11.7a}$$

$$\frac{E''}{E} = \frac{\cos\theta - n\cos\theta'}{\cos\theta + n\cos\theta'} = \frac{\sin(\theta'-\theta)}{\sin(\theta'+\theta)} \tag{9.11.7b}$$

In each case I have used Snell's law (9.11.4) to replace n with $\sin\theta/\sin\theta'$.

We shall return and analyse equations (9.11.7) later, once we have treated the second case, in which the $\mathbf{E}$ vector *lies in* the plane of incidence, figure 9–11 right. Before coming to that case, and briefly, I digress to point out that the Maxwell–Ampère's law in the absence of free current is,

$$\boldsymbol{\nabla} \times \mathbf{H} = \epsilon \frac{\partial \mathbf{E}}{\partial t}$$

and if $\mathbf{H}$ and $\mathbf{E}$ are plane waves in the form, amplitude$\times e^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)}$ then you can easily see that,

$$i\mathbf{k} \times \mathbf{H} = -i\omega\epsilon\mathbf{E}$$

Using $k = \omega/v$ and $v^2 = 1/\epsilon\mu$ it follows that,

$$\mathbf{E} = -\frac{\mu\omega}{k^2}\mathbf{k} \times \mathbf{H} \tag{9.11.8}$$

so the magnitudes of the $\mathbf{E}$ and $\mathbf{H}$ field vectors are related through,

$$E = -\frac{\mu\omega}{k}H$$

This means that since ω are μ are taken the same in either medium, the relations between the incident, transmitted and reflected amplitudes are,

$$\frac{E'}{E} = \frac{1}{n}\frac{H'}{H} \qquad \text{and} \qquad \frac{E''}{E} = \frac{H''}{H} \tag{9.11.9}$$

where $n = k'/k$ (9.11.4).

Now, we come to **Case 2**. $\mathbf{E}$ lies in the plane of incidence, figure 9–11 right. Therefore, now, $\mathbf{H}$ is parallel to the boundary; so we can use practically the same argument as for case 1, but in terms of $\mathbf{H}$ rather than $\mathbf{E}$. We will use the boundary condition (9.11.5a),

$$\hat{\mathbf{n}} \times (\mathbf{E} + \mathbf{E}'') = \hat{\mathbf{n}} \times \mathbf{E}' \tag{9.11.5a}$$

and insert (9.11.8); after cancelling ω and μ we have, similar to before,

$$\frac{1}{k^2}H\,(\hat{\mathbf{n}}\cdot\mathbf{k}) + \frac{1}{k''^2}H''\,(\hat{\mathbf{n}}\cdot\mathbf{k}'') = \frac{1}{k'^2}H'\,(\hat{\mathbf{n}}\cdot\mathbf{k}')$$

which leaves us an equation just like (9.11.6b),

$$H\cos\theta - H''\cos\theta'' = n^{-1}H'\cos\theta'$$

having remembered that $k'' = k$ and $n = k'/k$ (9.11.4). I combine this with the equivalent to (9.11.6a) which expresses continuity of the tangential component of **H** across the boundary,

$$H + H'' = H'$$

and exactly as before I will get these equations, the same as (9.11.7) except n is replaced with n^{-1},

$$\frac{H'}{H} = \frac{2\cos\theta}{\cos\theta + n^{-1}\cos\theta'} \tag{9.11.10a}$$

$$\frac{H''}{H} = \frac{\cos\theta - n^{-1}\cos\theta'}{\cos\theta + n^{-1}\cos\theta'} \tag{9.11.10b}$$

We now use (9.11.9) to cast these in terms of E, and again apply Snell's law to replace n with $\sin\theta/\sin\theta'$. We get,

$$\frac{E'}{E} = \frac{2n\cos\theta}{\cos\theta + n^{-1}\cos\theta'} = \frac{2\sin\theta'\cos\theta}{\sin(\theta'+\theta)\cos(\theta'-\theta)} \tag{9.11.7c}$$

$$\frac{E''}{E} = \frac{\cos\theta - n^{-1}\cos\theta'}{\cos\theta + n^{-1}\cos\theta'} = \frac{\tan(\theta-\theta')}{\tan(\theta+\theta')} \tag{9.11.7d}$$

(I confess I used `WolframAlpha` to get the last one.) Equations (9.11.7a,b) are the *first* and (9.11.7c,d) the *second* Fresnel formulas. Let's display them properly.

	refracted amplitude		**reflected amplitude**
E out of plane:	$\dfrac{E'}{E} = \dfrac{2\cos\theta\sin\theta'}{\sin(\theta+\theta')}$	;	$\dfrac{E''}{E} = \dfrac{\sin(\theta'-\theta)}{\sin(\theta'+\theta)}$
E in plane:	$\dfrac{E'}{E} = \dfrac{2\sin\theta'\cos\theta}{\sin(\theta'+\theta)\cos(\theta'-\theta)}$	;	$\dfrac{E''}{E} = \dfrac{\tan(\theta-\theta')}{\tan(\theta+\theta')}$

9.11.4 Analysis of Fresnel's equations

I'm sorry it's been a lot of algebra, and maybe in other courses you are given Fresnel's equations without proof. You may also see them in different but equivalent mathematical forms. They are the starting points for many analyses in optics. I'll mention just one or two briefly.

1. The two formulas for the amplitudes of the reflected waves are particularly instructive to analyse. In figures 9–10 and 11 I have drawn the case of medium 2 having greater refractive index than medium 1, typically of light in air falling on, say, a glass sheet. In that case, $\theta' < \theta$ and so you see that for light polarised perpendicular to the plane of incidence E'' is *negative*, which doesn't make sense—all amplitudes are real and positive. The way around this is to multiply the reflected wavefunction by $-1 = e^{i\pi}$ using Euler's beautiful equation.† (You may always multiply the solution to a linear differential equation by a constant.) This means that if the incoming wavefunction is $E_0 e^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)}$ then the reflected wave is $E_0 e^{i(\mathbf{k}\cdot\mathbf{r}-\omega t+\pi)}$, *exactly* 180° *out of phase.* Any component of the wave that is plane polarised *in* the plane of incidence will not be phase shifted on reflection because if $\theta' < \theta$ then $\tan(\theta - \theta') > 0$. Therefore in general light that is polarised in an arbitrary plane will be phase shifted by some angle on reflection, depending on the plane of polarisation, the angle of incidence and the refractive indices of the media. The resulting combination of the two plane polarised waves, out of phase by 180° amounts to *circular polarised* light. No doubt you will be given the full details in your future courses in optics.

   For light impinging at *normal incidence*, figure 9–12, left, there is no plane of incidence because you cannot define a plane using just one vector, and both **E** and **H** vectors are parallel to the boundary. In principle then both Fresnel equations apply and also $\theta = \theta' = 0$. I'm not a good mathematician and according to `https://math.stackexchange.com/questions/177067/lhospital-rule-for-two-variable` there is no rule of L'Hôpital for functions of more that one variable. However after playing a bit with `WolframAlpha` I get the impression that you may assert,

$$\lim_{\theta \to 0,\, \theta' \to 0} \left(\frac{\sin(\theta' - \theta)}{\sin(\theta' + \theta)} \right) = -1 \quad \text{and} \quad \lim_{\theta \to 0,\, \theta' \to 0} \left(\frac{\tan(\theta - \theta')}{\tan(\theta' + \theta)} \right) = -1$$

   In addition, since $\theta = \theta' = 0$, according to the Fresnel equations for refraction the transmitted amplitudes are zero, at normal incidence there is total reflection and the reflected wave is exactly 180° out of phase. In your discussions of phenomena such as the colours at a film of oil, the fringes called Newton's rings and lens coatings, you use the fact that light reflected by a more optically dense medium is shifted in phase by 180°. Now you see why. Of course there's always some transmission through glass because you cannot get light travelling exactly perpendicular to the surface.

† Euler's equation, $e^{i\pi} + 1 = 0$ *is* truly remarkable—combining zero, the first integer, the imaginary number and the two most prominent transcendental numbers into one formula!

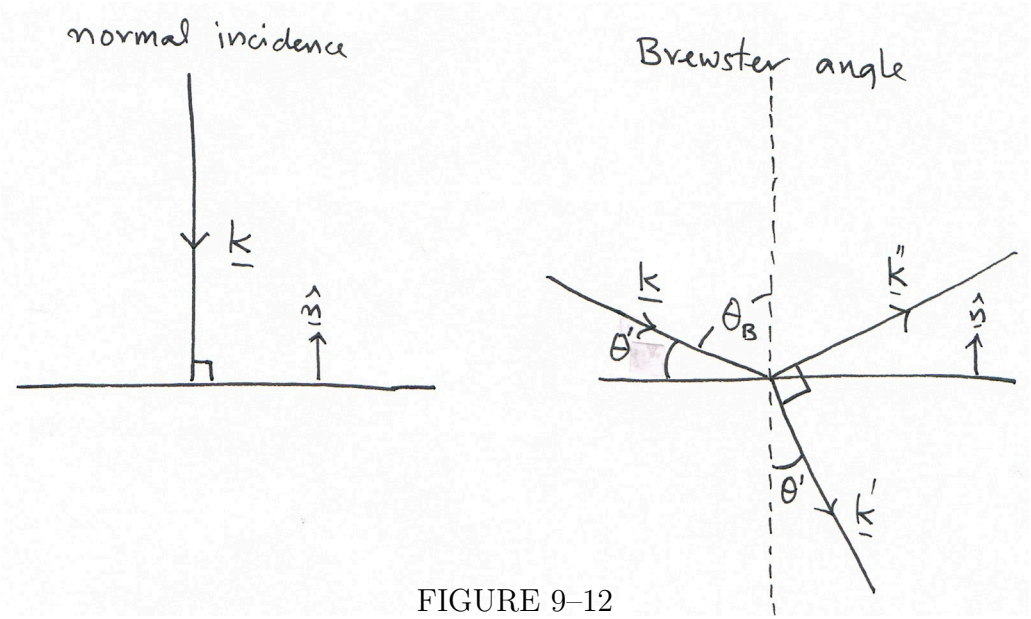


FIGURE 9–12

2. If $\theta + \theta' = 90°$ then $\mathbf{k}'$ and $\mathbf{k}''$ are perpendicular to each other, figure 9–12, right. Then according to Fresnel's equations, the amplitude of reflected light that is plane polarised in the plane of incidence is zero since the denominator, $\tan \pi/2 = \infty$. This means that unpolarised light impinging at a certain angle of incidence such that $\theta + \theta' = 90°$ will reflect light that is plane polarised since the $\mathbf{E}$ vector will be pointing in the direction normal to the plane of incidence. This is exploited by a stack of dielectric plates to make a *polariser*. The angle of incidence that leads to this condition is called the "Brewster angle," θ_B. We can easily find what it is. We have,

$$\theta_\mathrm{B} + \theta' = \frac{1}{2}\pi$$

and according to Snell's law,

$$n_1 \sin\theta_\mathrm{B} = n_2 \sin\theta' = n_2 \sin\left(\frac{1}{2}\pi - \theta_\mathrm{B}\right) = n_2 \cos\theta_\mathrm{B}$$

Hence

$$\tan\theta_\mathrm{B} = \frac{n_2}{n_1}$$

You can understand Brewster's law in the following way. The oscillation of the electric field along the direction of the refracted wave is perpendicular to $\mathbf{k}'$ and lies in the plane of incidence. When light strikes the surface it sets electrons into oscillation and these act as Hertzian dipoles radiating light. However if the Brewster condition is realised then the electrons are oscillating *back and forth in the direction of the reflected wave*, $\mathbf{k}''$. A dipole radiator does not radiate in the direction of its motion: this is true of an antenna as well as of an electron. Hence there is no radiation in the direction of $\mathbf{k}''$ and hence no reflected wave in the Brewster geometry.

My hope is that this will give you a good start and some structured notes to refer back to when you come to do optics properly. There, of course, you are faced with very much less ideal situations of media with curved surfaces, which are transmitting and absorbing; non linear media; and even media with *negative* permittivities or permeabilities—so called meta materials. I'm sure that in real research and industry Maxwell's and Fresnel's equations are solved by computer for complex geometries and multiple scattering events. I hope that this will also set up the background for you when to come to study the interaction of radiation with matter in solid state physics. Colin Latimer made a point to me once that when a student of history has studied, say, the Civil War and taken the exam, then that's the end of it. It's not really needed for the next course on Disraeli and Gladstone. But in physics it's very different and much more challenging—most advanced courses are predicated on earlier ones. You can't just take the exam and move on. Sorry.

5CCP2380 Problems 9

C9.1 Compare typical wavelengths of electromagnetic waves with the sizes of unit cells in crystals of typical metals and metal oxides. Is it reasonable to adopt the notion of "macroscopic" electric field? What sort of radiation has a wavelength of similar dimension to the spacing between atoms in solids? How should we then think about the interaction of electromagnetic waves with matter?

C9.2 (After Griffiths)

(a) Find expressions for k' and k'' in terms of ω, ϵ, μ and σ (see the footnote on page 16 of the section 9 lecture notes).

(b) Show that the time-averaged energy density of an electromagnetic wave in a conducting medium, using equations (9.8.8) in the lecture notes, is,

$$\overline{U} = \frac{1}{2\mu}\left(\frac{k'}{\omega}\right)^2 E_0^2\, e^{-2k''z}$$

Show that the magnetic contribution always dominates.

(c) Find the Poynting vector and show that its magnitude, the *intensity* is

$$I = \frac{k'}{2\mu\omega}\, E_0^2 e^{-2k''z}$$

C9.3 In *electrostatics* we assert that

$$\mathbf{P} = \epsilon_0 \chi_e \mathbf{E} \tag{1}$$

(See equation (7.6.1) in the lecture notes.) If the electric field is the real part of

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{E}_0\, e^{i\mathbf{k}\cdot\mathbf{r} - i\omega t}$$

then as discussed in the lecture notes, section 9, subsection 9.3, the polarisation will be forced to oscillate at the same frequency, but it may oscillate *out of phase* with $\mathbf{E}$,

$$\mathbf{P}(\mathbf{r}, t) = \mathbf{P}_0\, e^{i\mathbf{k}\cdot\mathbf{r} - i\omega t - i\phi}$$

These are the forced, damped oscillations described in "Simple Harmonic Motion"—additional notes on `KEATS`.

(a) Show that the relationship between $\mathbf{P}$ and $\mathbf{E}$ is consistent with a complex electric susceptibility,

$$\tilde{\chi}_e = \chi_e' + i\chi_e''$$

so that (1) becomes

$$\mathbf{P} = \epsilon_0\, (\chi_e' + i\chi_e'')\, \mathbf{E}$$

Show that

$$\epsilon_0\, \chi_e'\, \mathbf{E}_0 = \mathbf{P}_0 \cos\phi \qquad \text{and} \qquad \epsilon_0\, \chi_e''\, \mathbf{E}_0 = \mathbf{P}_0 \sin\phi$$

Explain that the polarisation comprises a component that is in phase with the electric field and one that is 90° out of phase and that the ratio of the amplitudes of the two components is χ_e'/χ_e''.

(b) In the interaction of electromagnetic radiation with matter we do not expect free currents to appear. If the medium is non magnetic then the current density is as given on page 4 of the section 6 notes,

$$\frac{\partial \mathbf{D}}{\partial t} = \frac{\partial \mathbf{P}}{\partial t} + \epsilon_0 \frac{\partial \mathbf{E}}{\partial t}$$

The vacuum current density is not an actual current—it does not correspond to the movement of charge, as you see in the example on page 4 of the section 4 notes, and cannot be excited by an electric field. This means that we are left with only the polarisation current, corresponding to the oscillations of the bound charge due to excitation by the oscillating **E**-field of the EM wave (this is why Maxwell called this *displacement current*: he was thinking of the displacment of otherwise fixed "charges" in dielectrics; of course at that time the actual nature of charge was a mystery). We have,

$$\mathbf{J} = \frac{\partial \mathbf{P}}{\partial t}$$

Show that the current is 90° out of phase with the polarisation.

(c) The displacement field is related to the electric field, for a linear dielectric as,

$$\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P} = \tilde{\epsilon}\, \mathbf{E}$$

where $\tilde{\epsilon} = \epsilon' + i\epsilon''$ is the *complex permittivity*. Show that

$$\frac{\epsilon'}{\epsilon_0} = 1 + \chi_e' \qquad \text{and} \qquad \frac{\epsilon''}{\epsilon_0} = \chi_e'' \tag{2}$$

(d) The *complex conductivity* is

$$\tilde{\sigma} = \sigma' + i\sigma''$$

Show that again if the medium is non magnetic, or strictly if the diamagnetic response is negligible,

$$\mathbf{J} = -i\omega \mathbf{P} = -i\omega\epsilon_0 \tilde{\chi}_e \mathbf{E} = \epsilon_0 \omega \left(\chi_e'' - i\chi_e'\right) \mathbf{E}$$

Hence show that,

$$\begin{aligned} \sigma' &= \epsilon_0\, \omega \chi_e'' \qquad \text{and} \qquad \sigma'' = -\epsilon_0\, \omega \chi_e' \\ \sigma' &= \omega \epsilon'' \qquad \text{and} \qquad \sigma'' = -\omega\left(\epsilon' - \epsilon_0\right) = -\epsilon_0\, \omega \left(\epsilon_r' - 1\right) \end{aligned} \tag{3}$$

You can see that from this point of view, everything stems fundamentally from the complex susceptibility arising from the phase difference between **P** and **E**. This depends on the frequency because it depends on the damping coefficient, $m\gamma$ (section 9.2), of the oscillating charges which in turn depends on the mass of the charges. Electrons can respond to high frequency optical and UV radiation whereas ions can only respond to low frequencies such as in alternating current circuits. Secondly note that the real and imaginary parts of the dielectric constant and the conductivity then follow as functions of the real and imaginary parts of the electric susceptibility through equations (2) and (3). I should repeat that complex conductivity and permittivity fall *outside the purview* of Maxwell's equations and the classical electrodynamics; see section 9.9.

5CCP2380 Problems 9—Solutions

C9.2(a) The answers, after some algebra are,

$$k' = \omega\sqrt{\frac{1}{2}\epsilon\mu}\left(\sqrt{1+\left(\frac{\sigma}{\epsilon\omega}\right)^2}+1\right)^{\frac{1}{2}} \tag{1a}$$

$$k'' = \omega\sqrt{\frac{1}{2}\epsilon\mu}\left(\sqrt{1+\left(\frac{\sigma}{\epsilon\omega}\right)^2}-1\right)^{\frac{1}{2}} \tag{1b}$$

These equations, first discovered by Heaviside in 1888, furnish us with the connection between the material constants of Maxwell's equations and the complex wavenumbers in optics, see equation (9.9.1) in Section 9.9 of the notes.

(b) The energy density is,

$$\begin{aligned} u &= \frac{1}{2}\epsilon E^2 + \frac{1}{2}\frac{1}{\mu}B^2 \\ &= \frac{1}{2}e^{-2k''z}\left(\epsilon E_0^2\cos^2\left(k'z-\omega t+\phi_E\right)+\frac{1}{\mu}B_0^2\cos^2\left(k'z-\omega t+\phi_E+\phi\right)\right) \end{aligned}$$

where,

$$B_0 = \frac{1}{\omega}|\tilde{k}|E_0 \qquad ; \qquad B_0^2 = \frac{1}{\omega^2}|\tilde{k}|E_0^2$$

and,

$$|\tilde{k}|^2 = \omega^2\epsilon\mu\sqrt{1+\left(\frac{\sigma}{\epsilon\omega}\right)^2} \qquad \text{using (1)}$$

The average over a cycle of $\cos^2$ any angle, is one half (just integrate $\cos^2 x$ from zero to 2π), so the average energy density is,

$$\begin{aligned} \bar{u} &= \frac{1}{2}e^{-2k''z}\cdot\frac{1}{2}\left(\epsilon E_0^2+\frac{1}{\mu}E_0^2\epsilon\mu\sqrt{1+\left(\frac{\sigma}{\epsilon\omega}\right)^2}\right) \\ &= \frac{1}{4}e^{-2k''z}\epsilon E_0^2\left(1+\sqrt{1+\left(\frac{\sigma}{\epsilon\omega}\right)^2}\right) \end{aligned}$$

Now use equations (1) to show that,

$$1+\sqrt{1+\left(\frac{\sigma}{\epsilon\omega}\right)^2} = \frac{2}{\epsilon\mu}\left(\frac{k'}{\omega}\right)^2$$

so that,

$$\begin{aligned} \bar{u} &= \frac{1}{4}e^{-2k''z}\,\epsilon E_0^2\frac{2}{\epsilon\mu}\left(\frac{k'}{\omega}\right)^2 \\ &= \frac{1}{2\mu}\left(\frac{k'}{\omega}\right)^2 E_0^2\,e^{-2k''z} \end{aligned}$$

□

The ratio of the magnetic to electric contributions is

$$\begin{aligned}\frac{\bar{u}_{\text{mag}}}{\bar{u}_{\text{el}}} &= \frac{\frac{1}{\mu}B_0^2}{\epsilon E_0^2} = \frac{1}{\mu\epsilon}\mu\epsilon\sqrt{1+\left(\frac{\sigma}{\epsilon\omega}\right)^2} \\ &= \sqrt{1+\left(\frac{\sigma}{\epsilon\omega}\right)^2}\end{aligned}$$ □

which is *greater than one*.

(c) We need to find the Poynting vector,

$$\begin{aligned}\mathbf{S} &= \frac{1}{\mu}\mathbf{E}\times\mathbf{B} \\ &= \frac{1}{\mu}E_0B_0\,e^{-2k''z}\,\cos\left(k'z-\omega t+\phi_E\right)\cos\left(k'z-\omega t+\phi_E+\phi\right)\hat{\mathbf{k}}\end{aligned}$$

Again we require the average over one whole cycle, and we find,

$$\frac{1}{2\pi}\int_0^{2\pi}\cos\theta\,\cos\left(\theta+\phi\right)\mathrm{d}\theta = \frac{1}{2}\cos\phi$$

so the average Poynting vector is,

$$\bar{\mathbf{S}} = \frac{1}{2\mu}E_0B_0\,e^{-2k''z}\,\cos\phi\,\hat{\mathbf{k}}$$

The *intensity* is the magnitude of $\bar{\mathbf{S}}$, namely,

$$\begin{aligned}I &= \frac{1}{2\mu}E_0B_0\,e^{-2k''z}\,\cos\phi \\ &= \frac{1}{2\mu}E_0^2\,e^{-2k''z}\,\frac{|\tilde{k}|}{\omega}\,\cos\phi\end{aligned}$$

If I use the form of $\tilde{k}$ and the end of page 13, Section 9 notes,

$$\tilde{k} = |\tilde{k}|\,e^{i\phi}$$

then I find that the real part of $\tilde{k}$ is,

$$k' = |\tilde{k}|\,\cos\phi$$

So finally

$$I = \frac{k'}{2\mu\omega}\,E_0^2\,e^{-2k''z}$$ □

In the vacuum, from (9.8.4) $k = k' = \sqrt{\epsilon\mu}\omega$, $k''=0$, and this reduces to,

$$I = \frac{1}{2}\sqrt{\frac{\epsilon}{\mu}}E_0^2$$

which is the result we obtained on page 9 of the Section 9 notes.

C9.3(a) We start with,

$$\mathbf{E} = \mathbf{E}_0\, e^{i\mathbf{k}\cdot\mathbf{r}-i\omega t}$$
$$\mathbf{P} = \mathbf{P}_0\, e^{i\mathbf{k}\cdot\mathbf{r}-i\omega t-i\phi}$$

The relation for *linear matter*, $\mathbf{P} = \epsilon_0\tilde{\chi}_e\mathbf{E}$, in the case of complex susceptibility is,

$$\mathbf{P} = \epsilon_0\left(\chi_e' + i\chi_e''\right)\mathbf{E}$$

Then,

$$\mathbf{P}_0 e^{i\mathbf{k}\cdot\mathbf{r}-i\omega t-i\phi} = \epsilon_0\mathbf{E}_0\left(\chi_e' + i\chi_e''\right)e^{i\mathbf{k}\cdot\mathbf{r}-i\omega t}$$

that is,

$$\mathbf{P}_0 e^{i\phi} = \epsilon_0\mathbf{E}_0\left(\chi_e' + i\chi_e''\right)$$

or

$$\mathbf{P}_0\cos\phi - i\mathbf{P}_0\sin\phi = \epsilon_0\mathbf{E}_0\chi_e' + i\epsilon_0\chi_e''$$

Now you equate real and imaginary parts,†

$$\epsilon_0\chi_e'\mathbf{E}_0 = \mathbf{P}_0\cos\phi \qquad ; \qquad \epsilon_0\chi_e''\mathbf{E}_0 = \mathbf{P}_0\sin\phi$$

If $\phi = 0$ then $\mathbf{E}$ and $\mathbf{P}$ are in phase and,

$$\mathbf{P} = \epsilon_0\chi_e'\mathbf{E} \;\; ; \;\; \chi_e'' = 0 \;\; ; \;\; \tilde{\chi}_e \text{ is real, wave is unattenuated}$$

If $\phi = \frac{1}{2}\pi$ then $\mathbf{E}$ and $\mathbf{P}$ are 90° out of phase and,

$$\mathbf{P} = \epsilon_0\chi_e''\mathbf{E} \;\; ; \;\; \chi_e' = 0 \;\; ; \;\; \tilde{\chi}_e \text{ is imaginary, wave is evanescent}$$

In the first case the wave is a pure undamped oscillation which propagates unattenuated through the medium, in the second case there is no oscillatory wave and the fields decay exponentially in the medium. These are the two extreme cases; in general $0 < \phi < \frac{1}{2}\pi$ and χ_e'/χ_e'' is the *ratio of amplitudes* of the in-phase and out-of-phase components of $\mathbf{P}$.

(b) For the time dependence of the polarisation we will write,

$$\mathbf{P} = \mathbf{P}_0 e^{-i\omega t} = \mathbf{P}_0\left(\cos\omega t - i\sin\omega t\right)$$

In that case,

$$\begin{aligned}\mathbf{J} = \frac{\partial\mathbf{P}}{\partial t} &= -i\omega\mathbf{P}_0 e^{-i\omega t}\\ &= -\omega\mathbf{P}_0 i\left(\cos\omega t - i\sin\omega t\right)\\ &= -\omega\mathbf{P}_0\left(\sin\omega t + i\cos\omega t\right)\end{aligned}$$

Of course we are interested in the *real parts* of $\mathbf{J}$ and $\mathbf{P}$ (using complex representation of the time dependence is just a mathematical device) and you see at once that these are 90° out of phase.

† I hope you're OK that you can always do this: if $x + iy = p + iq$ then $x = p$ and $y = q$.

(c) It is always true that,

$$\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P} \tag{1}$$

and in linear matter only, there is the constitutive relation,

$$\mathbf{D} = \tilde{\epsilon}\mathbf{E} \tag{2}$$

Remember that in Maxwell's equation $\mathbf{D} = \epsilon\mathbf{E}$, ϵ is real and this is only intended to describe *static* or *very slowly varying* fields (slowly varying means that we may neglect the displacement current). The extension to a *complex* permittivity, $\tilde{\epsilon}$, is forced on us by the observation that optical constants depend on the frequency of a changing electric field (see Section 9.9 of the lecture notes). Now use, from (a),

$$\mathbf{P} = \epsilon_0 \left(\chi_e' + i\chi_e''\right) \mathbf{E} \tag{3}$$

and insert (3) into (1) to get,

$$\begin{aligned} \mathbf{D} &= \epsilon_0 \mathbf{E} + \epsilon_0 \left(\chi_e' + i\chi_e''\right) \mathbf{E} && \text{(4a)} \\ &= \left(\epsilon' + i\epsilon''\right) \mathbf{E} && \text{(4b)} \end{aligned}$$

which serves to *define* the complex permittivity, $\tilde{\epsilon} = \epsilon' + i\epsilon''$, by comparison with (2). I can now equate real and imaginary parts of (4a) and (4b),

$$\epsilon' = \epsilon_0 \left(1 + \chi_e'\right) \quad ; \qquad \epsilon'' = \epsilon_0 \chi_e'' \tag{5}$$

which gives me equation (2) in the problems.

(d) As in (b) we start with,

$$\mathbf{J} = \frac{\partial \mathbf{P}}{\partial t} = -i\omega \mathbf{P}$$

Then using (2),

$$\begin{aligned} \mathbf{J} &= -i\omega \mathbf{P} = -i\omega\epsilon_0 \tilde{\chi}_e \mathbf{E} \\ &= \epsilon_0 \omega \left(\chi_e'' - i\chi_e'\right) \mathbf{E} \\ &= \tilde{\sigma}\mathbf{E} \end{aligned}$$

and this serves to *define* the complex conductivity, $\tilde{\sigma} = \sigma' + i\sigma''$, in an extension to the Ohm's law of Maxwell's equations which applies to *direct current* or *very slowly varying* current only. The real and imaginary parts of σ' are found by equating real and imaginary parts, Using also (5) we get the results we are asked to find in the problem,

$$\begin{aligned} \sigma' &= \omega\epsilon_0 \chi_e'' = \omega\epsilon_0 \epsilon_r'' = \omega\epsilon'' \\ \sigma'' &= -\omega\epsilon_0 \chi_e' = -\omega\epsilon_0 \left(\epsilon_r' - 1\right) = -\omega \left(\epsilon' - \epsilon_0\right) \end{aligned}$$

You may find these relations useful when you come to study optix. It is interesting that the *real* part of $\tilde{\sigma}$ relates to the *imaginary* part of $\tilde{\epsilon}$, and *vice versa*. I wonder why that is?