Energy in electromagnetism

Andrew M. Steane

Exeter College and Department of Atomic and Laser Physics, University of Oxford.

June 5, 2024

Abstract

This note presents the main ideas concerning energy and energy flow in classical electromagnetism, at 2nd year undergraduate level.

Abandon point charges all ye who enter here.

Contents

1 Basic concepts

There are three fundamental concepts:

$$\text{field energy density} \qquad u = \frac{1}{2}(\mathbf{E}\cdot\mathbf{D} + \mathbf{B}\cdot\mathbf{H}) \tag{1}$$

$$\text{Poynting vector (field energy flux)} \qquad \mathbf{N} = \mathbf{E}\times\mathbf{H} \tag{2}$$

$$\text{power density} \qquad p = \mathbf{E}\cdot\mathbf{j}_{\rm c} \tag{3}$$

where (as usual)

$$\mathbf{D} = \epsilon_0\mathbf{E} + \mathbf{P}, \qquad \mathbf{H} = \frac{1}{\mu_0}\mathbf{B} - \mathbf{M} \tag{4}$$

and (note well), $\mathbf{j}_{\rm c}$ in (3) is the conduction current, not the total current. The total current is

$$\mathbf{j} = \mathbf{j}_{\rm c} + \boldsymbol{\nabla}\times\mathbf{M} + \frac{\partial\mathbf{P}}{\partial t}. \tag{5}$$

The power density p is the rate at which energy is transferred from the fields to the free charges, per unit volume. To derive this, observe that the work done per unit time on a charge q moving at velocity $\mathbf{v}$ is $\mathbf{f}\cdot\mathbf{v} = q(\mathbf{E}+\mathbf{v}\times\mathbf{B})\cdot\mathbf{v} = q\mathbf{E}\cdot\mathbf{v}$ and therefore the rate per unit volume, if there are n such particles per unit volume, is $nq\mathbf{E}\cdot\mathbf{v} = \mathbf{E}\cdot\mathbf{j}_{\rm c}$ by using $\mathbf{j}_{\rm c} = nq\mathbf{v}$.

I have begun with the above quantities because I think they offer the simplest route to thinking about energy in electromagnetism. But there are some related quantities which are not the same as the above, although they go by the same names. They are:

$$\text{free field energy density} \qquad u_0 = \frac{1}{2}\epsilon_0(E^2 + c^2B^2) \tag{6}$$

$$\text{Poynting-vector-like quantity} \qquad \mathbf{N}_0 = \epsilon_0 c^2\mathbf{E}\times\mathbf{B} \tag{7}$$

$$\text{total power density} \qquad p_0 = \mathbf{E}\cdot\mathbf{j} \tag{8}$$

The difference between u and u_0 is one thing we will seek to clarify in the following (they are the same when $\mathbf{P} = \mathbf{M} = 0$). For the moment, simply note that u_0 is energy associated just with the fields $\mathbf{E}$ and $\mathbf{B}$, whereas u includes energy associated with polarization and magnetization as well. Polarization and magnetization are properties of matter, so u is not a property of the electromagnetic fields alone. It is a property of fields and matter together. The name 'field energy density' remains legitimate, if a little misleading, since one can apply the mathematical term 'field' to $\mathbf{P}$ and $\mathbf{M}$, and in consequence it is also applied to $\mathbf{D}$ and $\mathbf{H}$.

The conservation of energy is expressed by

$$-\frac{\partial}{\partial t}\int_{\mathcal{R}} u \,{\rm d}\tau = \oint_{(\mathcal{R})} \mathbf{N}\cdot{\rm d}\mathbf{s} + \int_{\mathcal{R}} \mathbf{E}\cdot\mathbf{j}_{\rm c}\,{\rm d}\tau \tag{9}$$

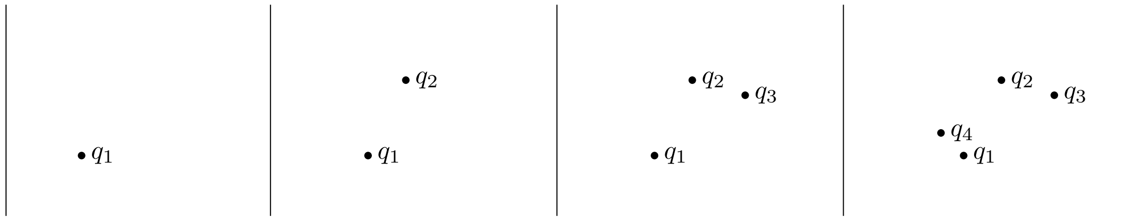

Figure 1: Building up an assembly of charges by introducing the charges one by one.

where $\mathcal{R}$ is a region of space, $(\mathcal{R})$ is the surface of the region, $\mathrm{d}\tau$ is a volume element and $\mathrm{d}\mathbf{s}$ is a surface element. The left hand side of this equation is the rate of loss of field energy in the region. The right hand side gives first the rate at which field energy is flowing out, and then the rate at which energy is going to the particles. Clearly these are flows which reduce u, hence the minus sign on the left. By using Gauss's theorem we express the surface integral as a volume integral of $\boldsymbol{\nabla} \cdot \mathbf{N}$:

$$\oint_{(\mathcal{R})} \mathbf{N} \cdot \mathrm{d}\mathbf{s} = \int_{\mathcal{R}} \boldsymbol{\nabla} \cdot \mathbf{N} \, \mathrm{d}\tau. \tag{10}$$

Hence we obtain

$$\int_R \frac{\partial u}{\partial t} + \boldsymbol{\nabla} \cdot \mathbf{N} + \mathbf{E} \cdot \mathbf{j}_{\mathrm{c}} \, \mathrm{d}\tau = 0. \tag{11}$$

(It is legitimate to bring the $\partial/\mathrm{d}\partial\mathrm{t}$ inside the integral, because the integral is just a sum and the limits of integration to not depend on t.) Now in general if an integral is zero it does not necessarily follow that the integrand is zero. However, if the integral is zero *for all regions of integration*, as here, then it must be that the integrand is zero. Therefore

Conservation of energy (continuity equation)

$$\frac{\partial u}{\partial t} + \boldsymbol{\nabla} \cdot \mathbf{N} + \mathbf{E} \cdot \mathbf{j}_{\mathrm{c}} = 0. \tag{12}$$

This is an important equation. It goes by the name *continuity equation.*

1.1 Static set of charges

Finally, another basic concept is the work required to assemble a static system of charges. To bring the first charge no work is required: $W_1 = 0$. To bring the second requires work $W_2 = q_2 V_{12}$ where V_{12} is electric potential *owing to* 1, *at the location of* 2 (see Fig 1). To bring the third charge requires $W_3 = q_3 V_{13} + q_3 V_{23}$. And so on:

$$W_i = q_i \sum_{j=1}^{i-1} V_{ji}. \tag{13}$$

Hence the total work is

$$W = \sum_i W_i = \sum_i \sum_{j=1}^{i-1} q_i V_{ji}. \tag{14}$$

Now it is a property of the electric potential of point charges that $V_{ji} = V_{ij}$. Hence we can write

$$W = \frac{1}{2} \sum_i \sum_{j=1}^{i-1} q_i (V_{ji} + V_{ij}). \tag{15}$$

Now suppose we set out all those V_{ij} values in a table:

$$\begin{matrix} V_{11} & V_{12} & V_{13} & \cdots & V_{1n} \\ V_{21} & V_{22} & V_{23} & \cdots & V_{2n} \\ \vdots & \vdots & \vdots & & \vdots \\ V_{n1} & V_{n2} & V_{n3} & \cdots & V_{nn} \end{matrix} \tag{16}$$

The sum in (15) includes all the off-diagonal entries, and only those, so we can write it as:

$$W = \frac{1}{2} \sum_{i=1}^{n} q_i \left(\sum_{j=1, j \neq i}^{n} V_{ij} \right). \tag{17}$$

The quantity in the bracket here is the potential at the location of the i'th charge owing to all the other charges once the assembly of all the charges has been completed. Hence we have

$$W = \frac{1}{2} \sum_i q_i V_i \tag{18}$$

where V_i is the potential at the i'th charge in the completed assembly. Note the factor half! That factor appears repeatedly in energy formulae in electromagnetism. Also note the usage of the phrase 'potential at the i'th charge'. This refers to the potential at some location owing to all the *other* charges. If you forget this, you will find yourself dealing with infinite quantities and you will get in a muddle.

Going over now to a continuous distribution of charge, we put $q_i \to \rho \mathrm{d}\tau$ for charge density ρ, and therefore

$$W = \frac{1}{2} \int V \rho \, \mathrm{d}\tau \tag{19}$$

This seems straightforward, but for completeness we ought to check that we correctly handled the omission of the V_{ii} parts from the integral. To do this one must first abandon point charges, replacing them with small balls of charge of some finite radius and charge density. For such balls, as the radius tends to zero, so does the charge, in such a way that infinite quantities are avoided. It all works out fine but we shall omit these details here. (See box for further comments.)

Finally, a bit of fun. For static problems we have (by definition of V), $\mathbf{E} = -\boldsymbol{\nabla} V$ and the first Maxwell equation gives $\rho = \boldsymbol{\nabla} \cdot \mathbf{D}$. The combination $V\rho$ therefore contains a factor whose derivative we know, and a factor whose integral we know. With this in mind, we can perform a form of integration by parts, as follows. First observe that

$$\boldsymbol{\nabla} \cdot (V\mathbf{D}) = (\boldsymbol{\nabla} V) \cdot \mathbf{D} + V \boldsymbol{\nabla} \cdot \mathbf{D} \;=\; -\mathbf{E} \cdot \mathbf{D} + V\rho. \tag{20}$$

Therefore

$$\begin{aligned} \int V\rho \mathrm{d}\tau &= \int \mathbf{E} \cdot \mathbf{D} \mathrm{d}\tau + \int \boldsymbol{\nabla} \cdot (\mathrm{V}\mathbf{D}) \mathrm{d}\tau \\ &= \int \mathbf{E} \cdot \mathbf{D} \mathrm{d}\tau + \int (\mathrm{V}\mathbf{D}) \cdot \mathrm{d}\mathbf{s} \end{aligned} \tag{21}$$

where we used Gauss's theorem to convert a volume integral to a surface integral. Now we let the region of integration be a huge sphere around the set of charges. In the limit as the radius of this sphere tends to infinity, the dominant term in both V and $\mathbf{D}$ will be that owing to the total charge Q, where V goes as Q/r and D as Q/r^2. If the total charge is zero then the fields will fall off quicker than this. It follows that $V\mathbf{D}$ falls off at least as fast as $1/r^3$. Therefore

$$\int (V\mathbf{D}) \cdot \mathrm{d}\mathbf{s} = \mathrm{O}(1/\mathrm{r}) \to 0. \tag{22}$$

(Note: the integral is over the spherical surface; in this integral r is constant and can be brought out the front.) We thus obtain

$$\int V\rho \, \mathrm{d}\tau \;=\; \int \mathbf{E} \cdot \mathbf{D} \, \mathrm{d}\tau \qquad \text{(static case)} \tag{23}$$

You can now see that our work formula (19) is consistent with our field energy formula (1). (This part of the discussion is limited to the static case because we employed $\mathbf{E} = -\boldsymbol{\nabla} V$ which is not true more generally.)

1.2 Simple examples

Let's start with our old friends the idealized circuit elements: resistor, capacitor and inductor.

We begin with energy flow (Poynting vector). The electrical power supplied is $P = VI$. For a resistor in the shape of a long cylinder of length d and radius r, $V = Ed$ where E is the electric field along the resistor, and the magnetic field is $B = \mu_0 I / 2\pi r$ in loops around the surface. Hence

$$VI = Ed\, 2\pi r B/\mu_0 = (2\pi r d) E H. \tag{25}$$

The quantity in the bracket is the area of the curved surface of the resistor, so we have that the energy flux (power per unit area) agrees with the Poynting vector, (2). The direction of $\mathbf{N}$ in this example is *into* the resistor. Field energy is flowing in, and it is being given up to the work required to push the

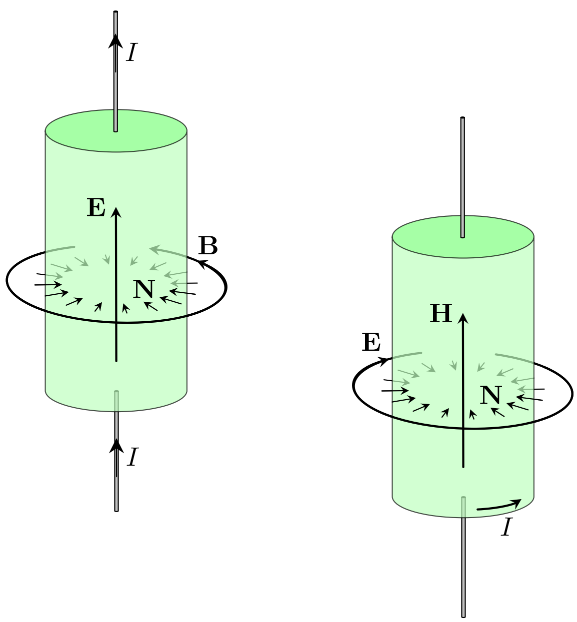

Figure 2: Fields and energy flow. Left: a resistor, or a capacitor in which the charge is increasing. Right: an inductor in which the current is increasing. Note: these diagrams show the fields in or very near the device in question (resistor, capacitor, inductor). Near the ends and further away the field lines change their directions, such that energy is conserved and flows towards the circuit element (or away if the capacitor or inductor is discharging).

current against the resistance of the material. (Inside the material this energy is quickly converted into random vibrations; the energy eventually emerges via convection and heat conduction and radiant heat; we have not attempted to model those processes even though they are largely electromagnetic too.)

Now the capacitor. We treat a cylindrical parallel-plate capacitor with a small gap d and large radius r. We have $P = VI$ and $V = Ed$. In a static case there is no energy flow and no magnetic field. If the capacitor is being charged, on the other hand, then the electric field is increasing and there is a magnetic field in loops around. At the radius r the magnetic field is $B = \mu_0 I/2\pi r$ (to obtain this one may use Ampère's law applied to a surface which extends to the wire supplying the current, thus avoiding the $\partial\mathbf{E}/\partial t$ term, or one may use some other surface or method). Hence for the capacitor we obtain (25) again.

Now the inductor. The magnetic field is $H = nI$ where n is the number of turns per unit length and I is the current. For a constant current there is no energy flow and no electric field. If the current is changing, there is an electric field E in loops around, such that the net voltage (or e.m.f. if you prefer) across the whole length of the wire is

$$V = N2\pi rE \tag{26}$$

where N is the number of turns. Hence

$$VI = (N2\pi rE)(H/n) = (2\pi rd)EH \tag{27}$$

where d is the length of the solenoid, and we have (25) once again.

In the case of the capacitor and the solenoid, the field energy flowing in or out goes to increase and decrease the field energy density inside the device. For the capacitor we have capacitance $C = \epsilon A/d$ (where we write $\epsilon = \epsilon_0\epsilon_r$) with $A = \pi r^2$ the area of the plates. To charge a capacitor from zero the total work required is

$$W = \int VI\mathrm{dt} = \int \mathrm{VdQ} = \int \mathrm{CVdV} = \frac{1}{2}\mathrm{CV}^2. \tag{28}$$

The volume is Ad so the energy density is

$$\frac{W}{Ad} = \frac{1}{2}\frac{\epsilon A}{d}(Ed)^2\frac{1}{Ad} = \frac{1}{2}\epsilon E^2 \tag{29}$$

in agreement with (1).

To make a current flow in a solenoid, starting from zero current, the work required is

$$W = \int VI\mathrm{dt} = \int \mathrm{LIdI} = \frac{1}{2}\mathrm{LI}^2. \tag{30}$$

The inductance is $L = \mu N^2 A/d$ so the energy density is

$$\frac{W}{Ad} = \frac{1}{2}\mu\frac{N^2 A}{d}\left(\frac{H}{n}\right)^2\frac{1}{Ad} = \frac{1}{2}\mu H^2 \tag{31}$$

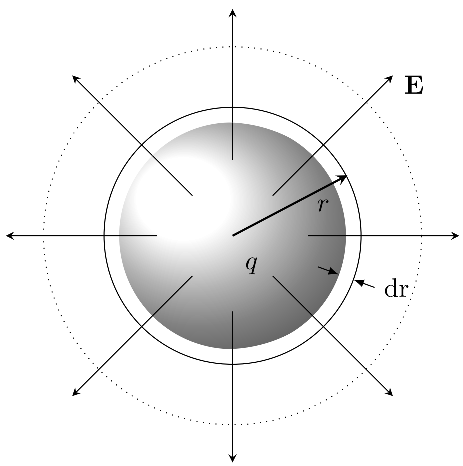

Figure 3: Adding a layer of thickness dr to a charged sphere. At the illustrated stage of construction, the sphere has reached some radius $r < R$ and it has a charge $q = Qr^3/R^3$. It will eventually grow to size R, at which point the charge will arrive at the total Q.

in agreement with (1) once again. (Note, by employing the concepts of relative permittivity and relative permeability we have here assumed the simplest kind of dielectric material whose response is linear and isotropic, meaning the polarization is along **E** and proportion to it, and the the magnetization is along **H** and proportional to it.)

1.2.1 The charged sphere

The uniformly charged sphere is a very useful example for learning purposes. We shall treat a sphere with no polarization ($\epsilon_r = 1$) first, and then a sphere with polarization. In each case we shall calculate the energy by two methods: work to bring the charges to together, and field energy.

1. A uniform charged sphere of non-polarized matter.
We suppose that the physical situation begins with an empty region of space, with, a very long way away and surrounding it, all the charge which will eventually be brought in to make the sphere. We calculate the work required to bring the charge together. At some stage of partial construction, the situation is as shown in Fig. 3. The sphere has attained radius r and the charge in it is $q = Qr^3/R^3$ where R is the radius it will have when its total charge reaches Q. Therefore the work required to bring in a new piece of charge dq is

$$\mathrm{dW} = \frac{\mathrm{q\,dq}}{4\pi\epsilon_0 \mathrm{r}} = \frac{\mathrm{Qr}^2}{4\pi\epsilon_0 \mathrm{R}^3}\mathrm{dq} \tag{32}$$

where we used that the whole situation is spherically symmetric so the simple Coulomb formula for the potential applies. With the arrival of charge dq, the radius grows in such a way as to keep the charge density ρ uniform. This is achieved if[1]

$$\mathrm{dq} = 4\pi \mathrm{r}^2 \rho \mathrm{dr} = 4\pi \mathrm{r}^2 \frac{\mathrm{Q}}{(4/3)\pi \mathrm{R}^3} \mathrm{dr} = 3\mathrm{Q}\frac{\mathrm{r}^2}{\mathrm{R}^3}\mathrm{dr}. \tag{33}$$

Hence we find

$$\begin{aligned} W &= \int \mathrm{dW} = \frac{3\mathrm{Q}^2}{4\pi\epsilon_0 \mathrm{R}^6}\int_0^{\mathrm{R}} \mathrm{r}^4 \mathrm{dr} \\ &= \frac{3}{5}\frac{Q^2}{4\pi\epsilon_0 R} \end{aligned} \tag{34}$$

Now let's examine the field energy. Inside the sphere the elecric field is radially outwards (for positive ρ) and equal to the amount given by the Coulomb law for the charge q. Hence

$$E_{(r<R)} = \frac{q}{4\pi\epsilon_0 r^2} = \frac{Q}{4\pi\epsilon_0 R^3} r \tag{35}$$

The total field energy inside the sphere is therefore

$$U_{\mathrm{inside}} = \int \frac{1}{2}\epsilon_0 E^2 \mathrm{d}\tau = \frac{1}{2}\epsilon_0 \left(\frac{\mathrm{Q}}{4\pi\epsilon_0 \mathrm{R}^3}\right)^2 \int_0^{\mathrm{R}} \mathrm{r}^2 4\pi \mathrm{r}^2 \mathrm{dr} = \frac{1}{10}\frac{\mathrm{Q}^2}{4\pi\epsilon_0 \mathrm{R}}. \tag{36}$$

This does *not* match the work required to assemble the sphere. Why not? It is because the work required to assemble the sphere has also produced all the other field energy: the energy which is situated *outside* the sphere. It must do this, because the work is supplied, at each stage of the process, *at exactly the location where the force acts.* The force acts on the charge at each radius r on its entire journey from infinity, pushing energy into the field at each place. The field outside the completed sphere is $E = Q/(4\pi\epsilon_0 r^2)$ (Coulomb law again, by using the Gauss theorem) so the field energy outside the sphere is

$$U_{\mathrm{outside}} = \int \frac{1}{2}\epsilon_0 E^2 \mathrm{d}\tau = \frac{1}{2}\epsilon_0 \left(\frac{\mathrm{Q}}{4\pi\epsilon_0}\right)^2 \int_{\mathrm{R}}^{\infty} \frac{1}{\mathrm{r}^4} 4\pi \mathrm{r}^2 \mathrm{dr} = \frac{1}{2}\frac{\mathrm{Q}^2}{4\pi\epsilon_0 \mathrm{R}}. \tag{37}$$

The total $U_{\mathrm{inside}} + U_{\mathrm{outside}}$ now agrees with W from (34).

2. A uniformly charged dielectric sphere.

Now let's suppose our charged sphere is made of polarizable matter. We assume the simple case $\mathbf{D} = \epsilon_0\epsilon_r\mathbf{E}$. The calculation of the field energy goes just as before, except that now we should use $\mathbf{E}\cdot\mathbf{D} = \epsilon E^2$ in (36) so

$$U_{\mathrm{inside}} = \frac{1}{10}\frac{Q^2}{4\pi\epsilon_0\epsilon_r R}. \tag{38}$$

The field outside is the same as before, so the result for the total field energy is

$$U = U_{\mathrm{inside}} + U_{\mathrm{outside}} = \frac{1}{2}\frac{Q^2}{4\pi\epsilon_0 R}\left(\frac{1}{5\epsilon_r} + 1\right). \tag{39}$$

[1] To get this you could equally well just differentiate $q = Qr^3/R^3$.

This is the correct result. The next part of the discussion is considerably more subtle and should be skipped (to the end of this section) on first reading, but you will need it later on if you ever want to understand polarization in other contexts, such as in plasma physics.

The work required to bring the charges to the sphere is calculated exactly as before, so we shall obtain (34) again, with the result that now we have a discrepancy between W and the field energy. What happened? Has the edifice of classical field theory crumbled? No: the problem is that we have not yet accounted for all the charge distribution in the sphere. The problem is hiding in the difference between $\mathbf{j}$ and $\mathbf{j}_c$, and the difference between ρ_f (the free charge density) and ρ_{tot}. The field energy calculation is correct and complete, but we have not yet finished the work calculation. We need to account for all the little dipoles associated with polarization. To do this, we can imagine first supplying the free charge density, which requires the work we have calculated, leading to (34). Next, we imagine bringing in, from a long way away, lots of little electric dipoles, and planting them in the sphere in just the right way so as to form the polarization of the final situation. We do this using dipoles all of fixed magnitude (the varying polarization can be produced by a varying density of such dipoles). It is important to treat fixed magnitude in the following to as to avoid the complication of tracking energy when a dipole is stretched.

As each dipole is brought in, we can arrange it so that the dipole is always oriented perpendicular to the local electric field, so there is no force on it and no work is done. Eventually the dipole arrives at its final location. Once there, we have to rotate it to alignment with the local field. In so doing the potential energy of the dipole in the field changes from zero to

$$-\mathbf{p} \cdot \mathbf{E} \tag{40}$$

where $\mathbf{p}$ is the dipole moment. Once the dipole is aligned with the field, therefore, we *receive back* an amount of energy $\mathbf{p} \cdot \mathbf{E}$, in which $\mathbf{E}$ refers to the field *before* it has been adjusted by the newly arrived dipole. By arguing thus for the whole sphere we find that the work required to place the dipoles is negative and given by

$$W_{\rm pol} = -\frac{1}{2} \int \mathbf{E}_1 \cdot \mathbf{P} \, {\rm d}\tau \tag{41}$$

where $\mathbf{E}_1$ refers to the field of the unpolarized sphere, and the factor $1/2$ is to avoid double-counting, just as in the argument leading to (19). Using now $\mathbf{P} = \epsilon_0 \mathbf{E}_2 (\epsilon_r - 1)$ where $\mathbf{E}_2$ is the final field (the one in the completed, polarized, sphere) we have

$$\begin{aligned} W_{\rm pol} &= -\frac{1}{2}\epsilon_0 \int \mathbf{E}_1 \cdot \mathbf{E}_2 (\epsilon_r - 1) \, {\rm d}\tau \\ &= -\frac{1}{2}\epsilon_0 \int \mathbf{E}_1^2 (1 - 1/\epsilon_r) \, {\rm d}\tau \end{aligned} \tag{42}$$

The sum $W + W_{\rm pol}$ now agrees with the field energy (39).

The above amounts to a rigorous proof of (41) but for added comfort, let's derive it another way. Consider the field energy in the case where there is no magnetic field:

$$u = \frac{1}{2}\mathbf{E} \cdot \mathbf{D} = \frac{1}{2}\mathbf{E} \cdot (\epsilon_0 \mathbf{E} + \mathbf{P}) \, . \tag{43}$$

To find the work required to introduce polarization, we want to compare this with what the energy would be without polarization, being careful to note that this will change the electric field too. So the comparison we want is

$$\Delta u = \frac{1}{2}\mathbf{E}_2 \cdot (\epsilon_0 \mathbf{E}_2 + \mathbf{P}) - \frac{1}{2}\epsilon_0 \mathbf{E}_1^2 \tag{44}$$

where $\mathbf{E}_1 = \epsilon_r \mathbf{E}_2$. We find

$$\Delta u = \frac{1}{2}\epsilon_0 \mathbf{E}_2 \cdot \mathbf{E}_2 (1 - \epsilon_r^2) + \frac{1}{2}\mathbf{E}_2 \cdot \mathbf{P}. \tag{45}$$

Now $\mathbf{P} = \epsilon_0(\epsilon_r - 1)\mathbf{E}_2$, so the first term can be written in terms of $\mathbf{P}$:

$$\Delta u = -\frac{1}{2}\mathbf{E}_2 \cdot \mathbf{P}(1 + \epsilon_r) + \frac{1}{2}\mathbf{E}_2 \cdot \mathbf{P} \;=\; -\frac{\epsilon_r}{2}\mathbf{E}_2 \cdot \mathbf{P} \;=\; -\frac{1}{2}\mathbf{E}_1 \cdot \mathbf{P}, \tag{46}$$

in agreement with (41). It might seem as if the second calculation assumed the answer because it employed the field energy formulae. But those formulae are themselves derivable by energy conservation, starting from the work required to move a charge against a field (Poynting's argument), so we have not argued in a circle.

2 Electromagnetic waves

Coming now to electromagnetic waves, we shall restrict to the case where $\mathbf{D}$ is parallel to $\mathbf{E}$ and $\mathbf{H}$ is parallel to $\mathbf{B}$. This describes isotropic dielectrics, conductors and plasmas, for example.

It will be useful to adopt complex notation, where we can write a plane wave, for example, in the form

$$\mathbf{E} = \mathbf{E}_0 e^{i(kz-\omega t)}, \qquad \mathbf{B} = \mathbf{B}_0 e^{i((kz-\omega t)} \tag{47}$$

$$\mathbf{D} = \mathbf{D}_0 e^{i(kz-\omega t)}, \qquad \mathbf{H} = \mathbf{H}_0 e^{i((kz-\omega t)} \tag{48}$$

where all the constants may be complex numbers, with the exception of $\mathbf{E}_0$ and ω. By allowing the wave amplitudes $\mathbf{B}_0$, $\mathbf{D}_0, \mathbf{H}_0$ to be possibly complex we can account for phase differences between the oscillations (and we choose the origin of t such that $\mathbf{E}_0$ is real). By allowing k to be complex we can account for a wider set of wave motions, including decaying waves such as occur in conducting media.

It is understood, in this notation, that the physical fields are given by the real part of the complex quantities. So for example if k is real then

$$\mathbf{E}_{\rm phys} = \mathrm{Re}\,(\mathbf{E}) = \mathbf{E}_0 \cos(kz - \omega t) \tag{49}$$

and if $k = \alpha + i\beta$ then

$$\mathbf{E}_{\rm phys} = \mathrm{Re}\,(\mathbf{E}) = \mathbf{E}_0 e^{-\beta z} \cos(\alpha z - \omega t). \tag{50}$$

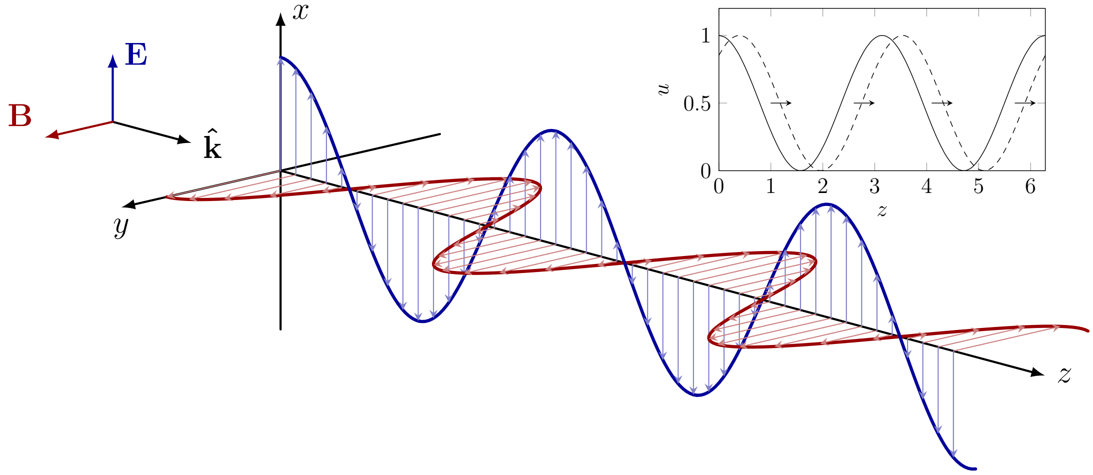

Figure 4: A travelling wave in free space. The main diagram shows the fields at some instant of time. The inset shows u at two successive instants of time (the later curve is dashed). The inset arrows show the direction of $\mathbf{N}$ (its amplitude matches u).

For plane waves of every kind, the third Maxwell equation gives

$$\mathbf{k} \times \mathbf{E}_0 \quad = \quad \omega \mathbf{B}_0. \tag{51}$$

Therefore

1. The magnetic waves are transverse (i.e. $\mathbf{B}$ is perpendicular to $\mathbf{k}$).
2. For a travelling wave, $\mathbf{B}$ oscillates in phase with $\mathbf{E}$ if and only if k is real.
3. The amplitudes are related by

$$B_0 = \frac{\omega}{|k|} E_0 \sin\theta \tag{52}$$

where θ is the angle between $\mathbf{E}$ and $\mathbf{k}$. This angle is 90^0 in ordinary circumstances.

For wave motion, the quantities u and $\mathbf{N}$ are usually neither uniform in space nor constant in time. They may oscillate at the wave frequency or at twice the wave frequency, for example. But when averaged over time the result may be independent of both time and position. Here are some examples.

1. **Travelling waves in empty space**

$$
\begin{aligned}
\mathbf{E} &= \mathbf{E}_0 \cos(kz - \omega t) && (53)\\
\mathbf{B} &= \mathbf{B}_0 \cos(kz - sin^{\omega} t) && (54)\\
& k \text{ is real} && (55)\\
\omega &= kc && (56)\\
E_0 &= cB_0 && (57)\\
u &= \frac{1}{2}\epsilon_0 (E_0^2 + c^2 B_0^2)\cos^2(kz - \omega t) \\
&= \epsilon_0 E_0^2 \cos^2(kz - \omega t) && (58)\\
\langle u \rangle &= \frac{1}{2}\epsilon_0 E_0^2 && (59)\\
\mathbf{N} &= \epsilon_0 c E_0^2 \cos^2(kz - \omega t)\hat{\mathbf{z}} && (60)\\
\langle \mathbf{N} \rangle &= \frac{1}{2}\epsilon_0 c E_0^2 \;=\; \langle u \rangle c\, \hat{\mathbf{z}} && (61)
\end{aligned}
$$

The notation $\langle \cdots \rangle$ here indicates the time average. Notice that for these waves the electric and magnetic fields contribute equally to the energy. Also, the final result ($\langle N \rangle = \langle u \rangle c$) is easily derived by considering the energy in a cylinder of length ct and cross-section A: this energy is $\langle u \rangle ctA$ and it is transported across a plane in a time t, so the flux is $\langle u \rangle c$.

2. **Standing waves in empty space**

$$
\begin{aligned}
\mathbf{E} &= \mathbf{E}_0\, 2\sin(kz)\sin(\omega t) && (62)\\
\mathbf{B} &= \mathbf{B}_0\, 2\cos(kz)\cos(\omega t) && (63)\\
E_0 &= cB_0 && (64)\\
u &= 2\epsilon_0 E_0^2 \left(\sin^2(kz)\sin^2(\omega t) + \cos^2(kz)\cos^2(\omega t)\right) && (65)\\
\langle u \rangle &= \epsilon_0 E_0^2 && (66)\\
\mathbf{N} &= \epsilon_0 c E_0^2 \sin(2kz)\sin(2\omega t)\hat{\mathbf{z}} && (67)\\
\langle \mathbf{N} \rangle &= 0. && (68)
\end{aligned}
$$

We have defined the amplitudes such that these standing waves are the sum of two travelling waves, each of which has amplitude $\mathbf{E}_0$. In a standing wave the electric and magnetic contributions are 90° out of phase, and the energy oscillates between them. Nodes of $\mathbf{E}$ are located at anti-nodes of $\mathbf{B}$. The Poynting vector indicates the movement of the energy to and fro between the nodes of $\mathbf{E}$ and nodes of $\mathbf{B}$.

For waves in a conductor and a plasma, see the note *Electromagnetic waves in plasmas and conductors.*

Notice that in the above we were careful to use real-number quantities when calculating u and $\mathbf{N}$. One should be careful with the complex notation when dealing with quantities that are not linear in the fields, because for two complex numbers z_1 and z_2,

$$\mathrm{Re}\,(z_1 z_2) \neq \mathrm{Re}\,(z_1)\,\mathrm{Re}\,(z_2)\,. \tag{69}$$

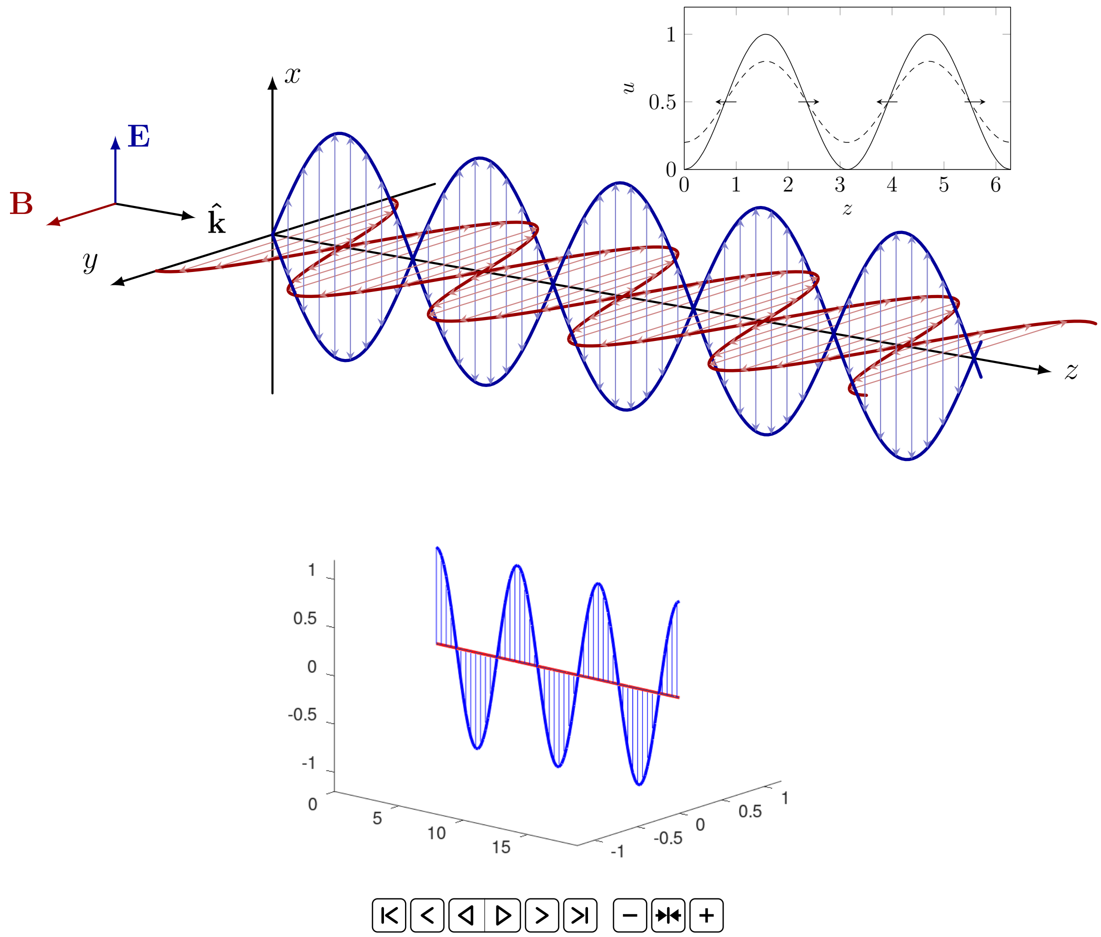

Figure 5: A standing wave in free space. The main diagram shows the extreme values of the fields. They oscillate in place, such that the result alternates between purely electric and purely magnetic. The inset shows u at two successive instants of time (the later curve is dashed). The inset arrows indicate the direction of $\mathbf{N}$ (it oscillates over time as well as position). The animation shows the fields evolving over time.

It follows that if $\mathbf{E}$ and $\mathbf{B}$ are complex valued, then

$$\begin{aligned} u &= \frac{1}{2}\left(\mathrm{Re}\,(\mathbf{E})\cdot \mathrm{Re}\,(\mathbf{D}) + \mathrm{Re}\,(\mathbf{B})\cdot \mathrm{Re}\,(\mathbf{H})\right), && (70)\\ \mathbf{N} &= \mathrm{Re}\,(\mathbf{E})\times \mathrm{Re}\,(\mathbf{H}) && (71)\end{aligned}$$

but

$$\begin{aligned} u &\neq \mathrm{Re}\left(\frac{1}{2}(\mathbf{E}\cdot\mathbf{D} + \mathbf{B}\cdot\mathbf{H})\right) && (72)\\ \mathbf{N} &\neq \mathrm{Re}\,(\mathbf{E}\times\mathbf{H}). && (73)\end{aligned}$$

It follows that one should **not** write $\mathbf{N} = \mathbf{E}\times\mathbf{H}$ if in fact $\mathbf{E}$ and $\mathbf{H}$ are complex.

Having made this cautionary observation, we can also note that, fortunately, there is a useful simplification if we just want to know the time-averaged quantities. If $z_1(t)$ and $z_2(t)$ are complex quantities of the form

$$z_1 = A_1 e^{i(-\omega t+\phi_1)}, \qquad z_2 = A_2 e^{i(-\omega t+\phi_2)} \tag{74}$$

where A_1, A_2, ϕ_1, ϕ_2 are all real and independent of t, then

$$\begin{aligned} \mathrm{Re}\,(z_1)\,\mathrm{Re}\,(z_2) &= A_1A_2\cos(\omega t+\phi_1)\cos(\omega t+\phi_2)\\ &= A_1A_2\cos(\theta)\cos(\theta+\Delta\phi) && (75)\end{aligned}$$

where $\theta = \omega t + \phi_1$ and $\Delta\phi = \phi_2 - \phi_1$. Therefore

$$\begin{aligned} \mathrm{Re}\,(z_1)\,\mathrm{Re}\,(z_2) &= A_1A_2\cos(\theta)\left[\cos(\theta)\cos(\Delta\phi) - \sin(\theta)\sin(\Delta\phi)\right]\\ &= A_1A_2\left[\cos^2(\theta)\cos(\Delta\phi) - (1/2)\sin(2\theta)\sin(\Delta\phi)\right] && (76)\end{aligned}$$

Taking now a time average, one finds $\langle\cos^2(\theta)\rangle = 1/2$ and $\langle\sin(2\theta)\rangle = 0$. Hence

$$\langle \mathrm{Re}\,(z_1)\,\mathrm{Re}\,(z_2)\rangle = \frac{1}{2}A_1A_2\cos(\Delta\phi). \tag{77}$$

Now observe that this same quantity can also be obtained from

$$z_1^* z_2 = A_1A_2 e^{i(\phi_2-\phi_1)} \tag{78}$$

by taking half real part. The overall conclusion is, for quantities oscillating as in (74),

$$\langle \mathrm{Re}\,(z_1)\,\mathrm{Re}\,(z_2)\rangle = \frac{1}{2}\mathrm{Re}\,(z_1^* z_2). \tag{79}$$

This purely mathematical observation can be useful in simplifying the algebra when we want to find the time average of the energy density or the Poynting vector. For example, for $\mathbf{E} = \mathbf{E}_0 e^{i(kz-\omega t)}$ and $\mathbf{H} = \mathbf{H}_0 e^{ikz-\omega t}$, with $\mathbf{E}_0$ and $\mathbf{H}_0$ both real, we find immediately $\langle\mathbf{N}\rangle = (1/2)\mathbf{E}_0\times\mathbf{H}_0$.

3 Derivation of u and $\mathbf{N}$: Poynting's theorem

The conservation of energy is a remarkable feature of classical electromagnetism. When the Maxwell equations were first discovered it was not self-evident whether or not the theory would respect energy conservation. That it does (and momentum conservation too) is a remarkable outcome. It means there is an intimate connection between the field equations and the Lorentz force equation: for a given force, not all field equations would respect energy conservation. With modern methods we can connect this to a Lagrangian density for both charge and fields together which is independent of time and position. We shall not explore that treatment here however. In the present discussion we shall take the Maxwell and Lorentz force equations as a starting point, and see what we can discover.

The following derivation is owing to John Henry Poynting (1852-1914).

We consider the following set of equations:

$$\nabla \cdot \mathbf{D} = \rho_{\mathrm{f}}, \qquad \nabla \cdot \mathbf{B} = 0 \tag{80}$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \qquad \nabla \times \mathbf{H} = \mathbf{j}_{\mathrm{c}} + \frac{\partial \mathbf{D}}{\partial t} \tag{81}$$

$$p = \mathbf{E} \cdot \mathbf{j}_{\mathrm{c}} \tag{82}$$

where the last equation is that part of the power density (rate of doing work per unit volume) which delivers energy to the motion of free charges. We want to find expressions for quantities u and $\mathbf{N}$, such that the energy conservation equation (continuity equation) (12) shall be satisfied. In the derivation we do not claim to know anything else about u and $\mathbf{N}$ in the first instance, but we suspect they can be connected to the fields in some way. The starting point of the derivation is to note that the term $\mathbf{E} \cdot \mathbf{j}_{\mathrm{c}}$ in (12) involves only quantities that appear in the Maxwell equations. We can find out about it by dotting $\mathbf{E}$ onto the fourth Maxwell equation, giving

$$\mathbf{E} \cdot (\nabla \times \mathbf{H}) = \mathbf{E} \cdot \mathbf{j}_{\mathrm{c}} + \mathbf{E} \cdot \frac{\partial \mathbf{D}}{\partial t} \tag{83}$$

Next, employ the vector identity

$$\nabla \cdot (\mathbf{E} \times \mathbf{H}) = \mathbf{H} \cdot (\nabla \times \mathbf{E}) - \mathbf{E} \cdot (\nabla \times \mathbf{H}) \tag{84}$$

(This is valid for any pair of vector fields, but we shall be applying it to $\mathbf{E}$ and $\mathbf{H}$ in particular). By employing this in (83) we find

$$\mathbf{H} \cdot (\nabla \times \mathbf{E}) - \nabla \cdot (\mathbf{E} \times \mathbf{H}) = \mathbf{E} \cdot \mathbf{j}_{\mathrm{c}} + \mathbf{E} \cdot \frac{\partial \mathbf{D}}{\partial t} \tag{85}$$

The term involving $\nabla \times \mathbf{E}$ can now be replaced by employing the third Maxwell equation, and we have

$$\mathbf{E} \cdot \frac{\partial \mathbf{D}}{\partial t} + \mathbf{H} \cdot \frac{\partial \mathbf{B}}{\partial t} + \nabla \cdot (\mathbf{E} \times \mathbf{H}) + \mathbf{E} \cdot \mathbf{j}_{\mathrm{c}} = 0. \tag{86}$$

This is looking a lot like the continuity equation! We will have that equation if we assign

$$\mathbf{N} = \mathbf{E} \times \mathbf{H} \tag{87}$$

as long as we can also find a u such that

$$\frac{\partial u}{\partial t} = \mathbf{E} \cdot \frac{\partial \mathbf{D}}{\partial t} + \mathbf{H} \cdot \frac{\partial \mathbf{B}}{\partial t}. \tag{88}$$

The simplest case is when $\mathbf{D} = \epsilon \mathbf{E}$ and $\mathbf{B} = \mu \mathbf{H}$ (for time-independent ϵ, μ) for then one has

$$\frac{\partial}{\partial t}(\mathbf{E} \cdot \mathbf{D}) = \epsilon \frac{\partial}{\partial t}(\mathbf{E} \cdot \mathbf{E}) = 2\epsilon \mathbf{E} \cdot \frac{\partial \mathbf{E}}{\partial t} = 2\mathbf{E} \cdot \frac{\partial \mathbf{D}}{\partial t} \tag{89}$$

and similarly,

$$\frac{\partial}{\partial t}(\mathbf{B} \cdot \mathbf{H}) = 2\mathbf{H} \cdot \frac{\partial \mathbf{B}}{\partial t}. \tag{90}$$

After substituting these results into (88) one finds that the expression we need is

$$u = \frac{1}{2}(\mathbf{E} \cdot \mathbf{D} + \mathbf{B} \cdot \mathbf{H}). \tag{91}$$

Thus we have derived expressions for energy density and energy flux in electromagnetic fields which give the conservation of energy.

There are two further aspects to clear up. First, we only presented the derivation in the simple case where $\mathbf{D} = \epsilon \mathbf{E}$ and $\mathbf{B} = \mu \mathbf{H}$. However it is not hard to generalize further. We write

$$\mathbf{P} = \epsilon_0 \chi_{\rm e} \mathbf{E}, \qquad \mathbf{M} = \chi_{\rm m} \mathbf{H} \tag{92}$$

where $\chi_{\rm e}$ is the electric susceptibility and $\chi_{\rm m}$ is the magnetic susceptibility. We can allow that these quantities are tensors, and we can allow that they may depend on the fields (a non-linear response) *as long as they remain time-independent*. Under this restriction we shall find $\mathbf{P} \cdot \dot{\mathbf{E}} = \dot{\mathbf{P}} \cdot \mathbf{E}$, and $\mathbf{M} \cdot \dot{\mathbf{H}} = \dot{\mathbf{M}} \cdot \mathbf{H}$, which leads to $\mathbf{D} \cdot \dot{\mathbf{E}} = \dot{\mathbf{D}} \cdot \mathbf{E}$, and $\mathbf{B} \cdot \dot{\mathbf{H}} = \dot{\mathbf{B}} \cdot \mathbf{H}$, which is all we require in order to obtain (91).

Finally, a further detail. The solution we have found for u and $\mathbf{N}$ is not unique. We can add to u any time-independent function and still get the same $\partial u / \partial t$, and we can add to $\mathbf{N}$ the curl of any scalar function and still get the same $\boldsymbol{\nabla} \cdot \mathbf{N}$. It follows that energy conservation alone will not pin down the expressions uniquely. If the energy density had a further contribution then it will make a contribution to the rest energy and therefore the rest mass of a system with internal fields. This will influence the inertia and the gravitational effects and therefore could in principle be observed. No such effects have been found and it is believed that our formulae for u and $\mathbf{N}$ are the right ones.

4 Relationship between u and u_0

The reader should note that in the above derivation of Poynting's theorem the power density term includes only the work done on free charges. The $\mathbf{E} \cdot \mathbf{j}_{\rm c}$ term does not include the work done to change the distribution of polarization and magnetization, because this work is taken care of elsewhere in the equations: it has been incorporated into $\mathbf{N}$ via the $\mathbf{H}$ field, and into u via both $\mathbf{D}$ and $\mathbf{H}$. This can be quite a puzzling idea when you first meet it!

To get some more insight, let's consider now the following set of equations:

$$\nabla \cdot \mathbf{E} = \rho_{\rm tot}/\epsilon_0, \qquad \nabla \cdot \mathbf{B} = 0 \tag{93}$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \qquad \nabla \times \mathbf{B} = \mu_0 \mathbf{j} + \epsilon_0 \mu_0 \frac{\partial \mathbf{E}}{\partial t} \tag{94}$$

$$p_0 = \mathbf{E} \cdot \mathbf{j} \tag{95}$$

These are called Maxwell's equations, and (80)–(82) are also called Maxwell's equations. Why the same name? It is because the two sets are entirely equivalent to one another. $\mathbf{D}$, $\mathbf{H}$, ρ and $\mathbf{j}_{\rm c}$ are defined expressly so as to bring about this equivalence.

If we now set out from (93)–(95) and apply Poynting's argument, we shall find

$$\frac{\partial u_0}{\partial t} + \nabla \cdot \mathbf{N}_0 + \mathbf{E} \cdot \mathbf{j} = 0. \tag{96}$$

This expresses conservation of energy just as (12) does, and it is *just as general* as (12). It can be used to describe fields in dielectric media, for example. The two results are, indeed, equivalent: each can be derived from the other (as you may like to show as an exercise). But there is a difference in the way we read the equations. In (12) the power density term is $\mathbf{E} \cdot \mathbf{j}_{\rm c}$ whereas in (96) the power density term is $\mathbf{E} \cdot \mathbf{j}$. The first refers to energy used up in pushing just the free charge; the second refers to energy used up in pushing the total charge (including the part owing to time-variation of $\mathbf{P}$ and spatial variation of $\mathbf{M}$). This idea is sufficiently confusing that one may wonder why $\mathbf{D}$ and $\mathbf{H}$ were ever introduced. However, if we simply trust the mathematics then $\mathbf{D}$ and $\mathbf{H}$ prove to be very useful.

One can elucidate the difference between the two approaches by observing that

$$u = u_0 + \frac{1}{2}(\mathbf{E} \cdot \mathbf{P} - \mathbf{B} \cdot \mathbf{M}) \tag{97}$$

$$\mathbf{N} = \mathbf{N}_0 - \mathbf{E} \times \mathbf{M} \tag{98}$$

$$p = p_0 + \mathbf{E} \cdot (\nabla \times \mathbf{M}) + \mathbf{E} \cdot \frac{\partial \mathbf{P}}{\partial t} \tag{99}$$

An example of (97) occurred in our treatment of the charged dielectric sphere, eqns (44)–(46).

The myth of the point charge.

In electromagnetism we often find it useful to talk about physical entities called 'point charges' whose charge is non-zero, mass is finite, and whose physical extension (diameter) is zero. No such entities exist in fact, and indeed they do not make sense even from a theoretical point of view! So what is going on?

First, on the physical non-existence. The closest approximation to the notional 'point charge' is the electron. These are entities described by quantum theory, in which there is a fixed amount of rest-mass and charge, and which can carry energy and momentum in the usual way. However no electron has ever been localized in an infinitesimally small region of space, because if it were then the position uncertainty would tend to zero, with the consequence that the momentum uncertainty tends to infinity, which implies that the kinetic energy would also increase without limit, and the situation is impossible. (What happens in fact when one tightly confines electrons is that multiple electron-positron pairs are produced). It follows that, to get a good physical intuition about electrons in ordinary circumstances you should think of them as spread out a little. A diameter of order a few fm (10^{-15} m) gives a reasonable picture. This is the diameter where the field energy is of the order of the rest mass energy if we model the electron as a little ball of smeared-out charge.

So much for the physical myth. What about the mathematical possibility? Is there anything wrong with introducing mathematical point charges into our discussion of electromagnetism? The answer is that for discussions of motion in response to a force, the notion of a point-like charge is often ok, but there are cases where one must be more careful. The difficulties arise when energy and radiation reaction are considered. The total field energy of a point charge is infinite (see (34)) so really it is hopeless dealing with point charges (of finite charge) if we want to be careful about energy and momentum. To the rescue comes the following idea: *let the charge tend to zero as the volume does.* What this means is, for small entities adopt a model where the charge density is everywhere finite. A sphere of charge density ρ and radius r has charge

$$q = \frac{4}{3}\pi r^3 \rho \tag{24}$$

The point particle limit $r \to 0$ is now also a zero charge limit, $q \to 0$, and furthermore the field energy tends to zero too. If a given discussion is not concerned with the actual magnitude of the charge then one can choose some value of r small compared to all other lengths under discussion, and then the charge will be small but non-zero, and one can apply formulae such as $\mathbf{f} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B})$ without encountering non-physical predictions.