Electromagnetism

University of Oxford
Second Year, Part A2

Caroline Terquem
Department of Physics

caroline.terquem@physics.ox.ac.uk

Michaelmas Term 2018

Contents

1 Potentials **9**

1.1 Static scalar and vector potentials . . . 9

1.1.1 Definitions . . . 9

1.1.2 Poisson's and Laplace's equations . . . 10

1.1.3 Solutions to Laplace's equation: separation of variables . . . 13

1.1.4 Multipole expansion and dipoles . . . 18

1.2 Time–dependent scalar and vector potentials . . . 22

1.2.1 Gauge transformations . . . 22

1.2.2 Lorenz gauge and Maxwell's equations . . . 23

1.2.3 Retarded potentials . . . 24

2 Electric fields in matter **25**

2.1 Polarization in dielectrics . . . 25

2.1.1 Atomic dipoles . . . 25

2.1.2 Molecular dipoles . . . 26

2.1.3 Polarization . . . 27

2.2 Polarization charges and current . . . 28

2.2.1 Polarization charges . . . 28

2.2.2 Polarization current . . . 29

2.3 The electric field caused by polarized matter . . . 30

2.3.1 Field outside the dielectric . . . 30

2.3.2 Field inside the dielectric . . . 31

2.4 The electric displacement vector **D** . . . 33

2.4.1 Gauss's law . . . 33

2.4.2 Ampère's law . . . 34

2.4.3 Boundary conditions on **E** and **D** . . . 35

2.5 Linear dielectrics . . . 36

2.6 Energy in the presence of dielectrics . . . 37

3 Magnetic fields in matter **41**

3.1 Magnetic materials . . . 41

3.1.1 Review: Torque and force on a magnetic dipole . . . 42

3.1.2 Diamagnetism . . . 43

3.1.3 Paramagnetism . . . 45
3.1.4 Magnetization . . . 46
3.2 Magnetization currents . . . 46
3.2.1 Surface currents . . . 46
3.2.2 Volume currents . . . 48
3.3 The magnetic field caused by magnetized matter . . . 49
3.3.1 Field outside the material . . . 49
3.3.2 Field inside the material . . . 50
3.4 The auxiliary field **H** . . . 53
3.4.1 Ampère's law . . . 53
3.4.2 Boundary conditions on **B** and **H** . . . 54
3.5 Linear magnetic materials . . . 55
3.6 Energy in the presence of magnetic materials . . . 56
3.7 Ferromagnetism . . . 59
3.7.1 Microscopic description . . . 59
3.7.2 Hysteresis loop . . . 60
3.7.3 Hysteresis loss . . . 61

4 Electromagnetic waves in vacuum 63
4.1 Maxwell's equations and boundary conditions . . . 63
4.1.1 Local form of Maxwell's equations . . . 63
4.1.2 Integral form of Maxwell's equations . . . 63
4.1.3 Boundary conditions . . . 64
4.2 Electromagnetic waves in vacuum . . . 65
4.2.1 Monochromatic plane waves . . . 66
4.2.2 Electromagnetic waves . . . 68
4.2.3 Polarization . . . 69
4.3 Energy and momentum transport by electromagnetic waves . . . 71
4.3.1 The Poynting vector . . . 71
4.3.2 Energy conservation for a system of charges and electromagnetic fields 73
4.3.3 Application to monochromatic plane waves . . . 74
4.3.4 Momentum transport and radiation pressure . . . 75

5 Electromagnetic waves in matter 77
5.1 Maxwell's equations in matter and boundary conditions . . . 77
5.1.1 Maxwell's equations . . . 77
5.1.2 Boundary conditions . . . 78
5.2 Electromagnetic waves in non–conducting linear media . . . 78
5.2.1 Ewald–Oseen extinction theorem . . . 79
5.2.2 Monochromatic plane waves . . . 79
5.2.3 Energy transport . . . 81
5.3 Reflection and transmission at the boundary between two linear media . . . 82

5.3.1 Normal incidence . . . 82
5.3.2 Oblique incidence and Brewster's angle . . . 85
5.4 Electromagnetic waves in conductors . . . 86
5.4.1 Skin depth . . . 86
5.4.2 Impedance . . . 89
5.4.3 Energy transport . . . 90
5.5 Reflection and transmission at a conducting surface . . . 90
5.6 Electromagnetic waves in media with frequency dependent permittivity . . . 92
5.6.1 Dispersion of glass . . . 93
5.6.2 A simple model to explain dispersion . . . 93
5.6.3 Absorption and anomalous dispersion . . . 96
5.6.4 Group velocity . . . 100
5.6.5 Plasma frequency . . . 100

6 Electromagnetism and special relativity 103
6.1 Einstein's postulates . . . 103
6.2 Review: Lorentz transformations . . . 104
6.2.1 Transformation of coordinates . . . 104
6.2.2 Time dilatation . . . 105
6.2.3 Lorentz contraction . . . 105
6.2.4 Velocity addition . . . 105
6.3 Relativistic electrodynamics . . . 106
6.3.1 Invariance of electric charge . . . 106
6.3.2 Transformation of the electric field . . . 106
6.3.3 Transformation of the electric and magnetic fields . . . 108

7 Transmission lines and waveguides 113
7.1 Transmission lines . . . 114
7.1.1 Equivalent circuit of a transmission line . . . 114
7.1.2 The telegrapher's equations and characteristic impedance . . . 114
7.1.3 Parallel wire transmission line . . . 117
7.1.4 Reflection by the load . . . 119
7.1.5 Impedance matching and quarter–wave transformer . . . 120
7.2 Waveguides and resonant cavities . . . 122
7.2.1 Generalities about waveguides . . . 122
7.2.2 TE waves in a rectangular wave guide . . . 125
7.2.3 Resonant cavities . . . 128

A Radiation 131
A.1 Power radiated by an accelerated point charge . . . 131
A.2 Electric dipole radiation . . . 135
A.2.1 Oscillating dipole . . . 135
A.2.2 Antennas . . . 136

A.3 Thomson scattering 137
A.3.1 Thomson scattering cross section 137
A.3.2 Thomson, Compton, resonant and Rayleigh scattering 138

These notes borrow from the following books:

David J. Griffiths, *Introduction to Electrodynamics*, 4th edition (Cambridge University Press)

Edward M. Purcell & David J. Morin, *Electricity and Magnetism*, 3rd edition (Cambridge University Press)

John D. Jackson, *Classical Electrodynamics*, 3rd edition (Wiley)

Richard P. Feynman, Robert B. Leighton & Matthew Sands *The Feynman Lectures on Physics*, Volume II (Basic Books)

These notes are meant to be a support for the course, but they should not replace textbooks. It is strongly advised that at least one of the books listed above is used regularly, as they provide much more details about the subject and lots of examples and problems. To quote William Faulkner: *Read! You'll absorb it.*

Chapter 1

Potentials

1.1 Static scalar and vector potentials

1.1.1 Definitions

As seen in the fist year course, for static fields in the presence of electric charges of density ρ and electric currents of density $\mathbf{J}$, Maxwell's equations are:

$$\boldsymbol{\nabla}\cdot\mathbf{E} = \frac{\rho}{\epsilon_0}, \tag{1.1}$$

$$\boldsymbol{\nabla}\cdot\mathbf{B} = 0, \tag{1.2}$$

$$\boldsymbol{\nabla}\times\mathbf{E} = \mathbf{0}, \tag{1.3}$$

$$\boldsymbol{\nabla}\times\mathbf{B} = \mu_0\mathbf{J}. \tag{1.4}$$

Equation (1.3) implies that there exists a **scalar potential** V such that:

$$\boxed{\mathbf{E} = -\boldsymbol{\nabla}V.} \tag{1.5}$$

The scalar potential is not uniquely defined, as any function $V' = V + K$, where K is a constant, also satisfies $\mathbf{E} = -\boldsymbol{\nabla}V'$. The physical interpretation of equation (1.5) is that the electric potential V evaluated at some position $\mathbf{r}$ is the work required to bring a unit positive charge from some reference point to the position $\mathbf{r}$ in the presence of the field $\mathbf{E}$. In other words, V is the potential energy per unit charge.

Similarly, equation (1.2) implies that there exists a **vector potential A** such that:

$$\boxed{\mathbf{B} = \boldsymbol{\nabla}\times\mathbf{A}.} \tag{1.6}$$

Here again, the vector potential $\mathbf{A}$ is not uniquely defined, as we can have different potentials that give the same magnetic field. If both $\mathbf{A}$ and $\mathbf{A}'$ are associated with the same field, then: $\mathbf{B} = \boldsymbol{\nabla}\times\mathbf{A} = \boldsymbol{\nabla}\times\mathbf{A}'$, which implies: $\boldsymbol{\nabla}\times(\mathbf{A}' - \mathbf{A}) = \mathbf{0}$. The vector $\mathbf{A}' - \mathbf{A}$ is curl free, and therefore can be written as the gradient of a scalar ϕ: $\mathbf{A}' - \mathbf{A} = \boldsymbol{\nabla}\phi$, so that $\mathbf{A}' = \mathbf{A} + \boldsymbol{\nabla}\phi$.

From the definition of the vector potential given by equation (1.6), it is not straightforward to assign a physical meaning to $\mathbf{A}$. However, it will be seen in third year that, in the same way that the momentum $\mathbf{p}$ and the energy E combine to form the four–momentum $(E/c, \mathbf{p})$ in relativity, $\mathbf{A}$ and V combine to form the electromagnetic four–potential $(V/c, \mathbf{A})$. Also, in quantum electrodynamics, $\mathbf{A}$ and V, and not $\mathbf{B}$ and $\mathbf{E}$, are the fundamental quantities entering the equations that replace Maxwell equations (see Feynman, sections 15.4 and 15.5).

1.1.2 Poisson's and Laplace's equations

With $\mathbf{E}$ and $\mathbf{B}$ given by equations (1.5) and (1.6), Maxwell's equations (1.2) and (1.3) are satisfied. We now insert equation (1.5) into Gauss's law (1.1) to obtain **Poisson's equation** for V:

$$\boxed{\nabla^2 V = -\frac{\rho}{\epsilon_0}.} \tag{1.7}$$

Finally, we insert equation (1.6) into Ampère's law (1.4) to obtain:

$$\boldsymbol{\nabla} \times (\boldsymbol{\nabla} \times \mathbf{A}) = \mu_0 \mathbf{J}.$$

Using the identity $\boldsymbol{\nabla} \times (\boldsymbol{\nabla} \times \mathbf{A}) = \boldsymbol{\nabla} (\boldsymbol{\nabla} \cdot \mathbf{A}) - \nabla^2 \mathbf{A}$, the above equation can be written under the form:

$$\nabla^2 \mathbf{A} - \boldsymbol{\nabla} (\boldsymbol{\nabla} \cdot \mathbf{A}) = -\mu_0 \mathbf{J}. \tag{1.8}$$

The curl of the vector potential is specified by equation (1.6), but its divergence can be chosen freely. We have indeed $\boldsymbol{\nabla} \cdot \mathbf{A}' = \boldsymbol{\nabla} \cdot \mathbf{A} + \nabla^2 \phi$. By choosing ϕ appropriately, and given $\mathbf{A}$, we can obtain a vector potential $\mathbf{A}'$ which has whatever divergence we want. To simplify equation (1.8), **we chose** $\boldsymbol{\nabla} \cdot \mathbf{A} = 0$. This equation then becomes:

$$\boxed{\nabla^2 \mathbf{A} = -\mu_0 \mathbf{J},} \tag{1.9}$$

which is equivalent to three Poisson's equations for the *cartesian* components of $\mathbf{A}$: $\nabla^2 A_x = -\mu_0 J_x$ and similarly for the y– and z–components. Note, however, that the cylindrical and spherical components of $\mathbf{A}$ do *not* satisfy Poisson's equation[1].

Using the potentials therefore enables us to replace Maxwell's equations by Poisson's equations (1.7) and (1.9). Once they are solved, the fields $\mathbf{E}$ and $\mathbf{B}$ can be calculated using equations (1.5) and (1.6).

[1] $\nabla^2 \mathbf{A}$, *defined* through $\nabla^2 \mathbf{A} = \boldsymbol{\nabla} (\boldsymbol{\nabla} \cdot \mathbf{A}) - \boldsymbol{\nabla} \times (\boldsymbol{\nabla} \times \mathbf{A})$, is a vector which x, y and z components are $\nabla^2 A_x$, $\nabla^2 A_y$ and $\nabla^2 A_z$. That is to say, the definition of $\nabla^2 \mathbf{A}$ makes *explicit* reference to *cartesian* coordinates. It follows that, in cylindrical coordinates for example, as the unit vectors depend on the coordinates, the components of $\nabla^2 \mathbf{A}$ are *not* $\nabla^2 A_{r,\theta,z}$. Therefore, A_r, for example, does *not* satisfy Poisson's equation $\nabla^2 A_r = -\mu_0 J_r$.

In a region of space where there are no charges, that is to say away from the charges which create the potential, Poisson's equation for V reduces to **Laplace's equation**:

$$\boxed{\nabla^2 V = 0.} \tag{1.10}$$

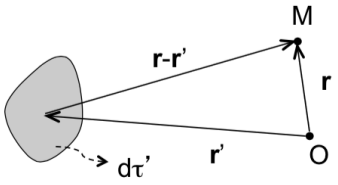
Similarly, in a region of space where there are no currents, $\mathbf{A}$ satisfies:

$$\boxed{\nabla^2 \mathbf{A} = \mathbf{0},} \tag{1.11}$$

which is equivalent to three Laplace's equations for the *cartesian* components of $\mathbf{A}$.

Solutions to Poisson's equation:

The potential of a point charge q at a distance d from the charge is $q/(4\pi\epsilon_0 d)$, taking the reference point (zero of potential) at infinity. Using the superposition principle, we then can write the scalar potential V created at a point M by a charge distribution as:



$$\boxed{V(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \iiint_{\mathcal{V}} \frac{\rho(\mathbf{r}')}{|\mathbf{r}' - \mathbf{r}|} d\tau',} \tag{1.12}$$

where the integration is over the volume $\mathcal{V}$ of the distribution and the position vectors $\mathbf{r}$ and $\mathbf{r}'$ are measured from an origin O chosen arbitrarily.

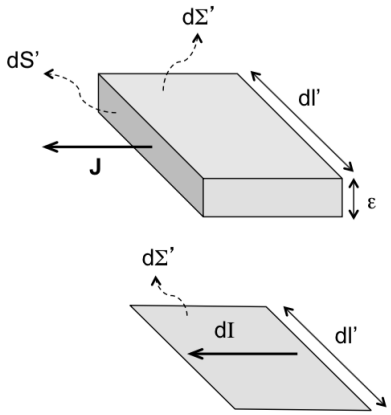
In writing equation (1.12), we have assumed that the reference point for the potential is the *same* for all the charge elements in the distribution, and is at *infinity*. This choice of the reference point can be made only if the charge distribution does not extend to infinity. Therefore, equation (1.12) only applies when the charge distribution is **localized**, that is, when the volume $\mathcal{V}$ is finite.

The scalar potential V given by equation (1.12) has to be a solution of Poisson's equation (1.7). As the cartesian components of the vector potential satisfy the same equation but with the source term being $-\mu_0 J_x$, $-\mu_0 J_y$ and $-\mu_0 J_z$ instead of $-\rho/\epsilon_0$, the solution of equation (1.9) has to be the same as that given by equation (1.12) after doing the appropriate replacement. Therefore, we obtain:

$$\boxed{\mathbf{A}(\mathbf{r}) = \frac{\mu_0}{4\pi} \iiint_{\mathcal{V}} \frac{\mathbf{J}(\mathbf{r}')}{|\mathbf{r}' - \mathbf{r}|} d\tau',} \tag{1.13}$$

where the integration is over the volume $\mathcal{V}$ of the **localized** current distribution. This expression is valid only when **cartesian** coordinates are used.

The vector potential above is a solution of Poisson's equation (1.9), but we should check that it also satisfies $\boldsymbol{\nabla} \cdot \mathbf{A} = 0$, as this is required to get Poisson's equation in the first place. It could be shown that this is indeed statisfied provided that the current distribution is steady (that is, when $\boldsymbol{\nabla} \cdot \mathbf{J} = 0$).



If the current flows through a wire which has a very small thickness ϵ, then the integral above can be written as:

$$\mathbf{A}(\mathbf{r}) = \frac{\mu_0}{4\pi} \iint_\Sigma \frac{\mathbf{J}(\mathbf{r}')\epsilon}{|\mathbf{r}' - \mathbf{r}|} d\Sigma',$$

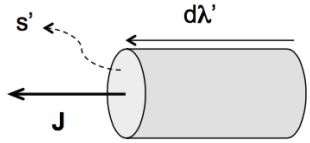
where we have assumed that $\mathbf{J}$ does not vary significantly over the thickness ϵ. We define the **surface current density** (or current per unit width) as:

$$\mathbf{K} = \mathbf{J}\epsilon = \frac{d\mathbf{I}}{dl'}, \tag{1.14}$$

where $dI = \mathbf{J} \cdot d\mathbf{S}'$ and the vector $d\mathbf{I}$ is orientated in the direction of the current.

Therefore, the vector pontential can be written as:

$$\mathbf{A}(\mathbf{r}) = \frac{\mu_0}{4\pi} \iint_\Sigma \frac{\mathbf{K}(\mathbf{r}')}{|\mathbf{r}' - \mathbf{r}|} d\Sigma'. \tag{1.15}$$



If the current flows through a wire which has a very small cross sectional area s', then the vector potential can be written as:

$$\mathbf{A}(\mathbf{r}) = \frac{\mu_0}{4\pi} \int_\Gamma \frac{I(\mathbf{r}')}{|\mathbf{r}' - \mathbf{r}|} d\boldsymbol{\lambda}', \tag{1.16}$$

where $I = \mathbf{J} \cdot \mathbf{s}'$ and we have assumed that $\mathbf{J}$ does not vary significantly over the area s'. Here Γ is the path along the wire.

The scalar and vector potentials given above are the solutions of Poisson's equations (1.7) and (1.9). However, the integrals are not in general easy to compute. Calculating these integrals is equivalent to solving directly Poisson's equation with the appropriate boundary conditions. As solving the equation directly is in general more tractable, this is the method which is often used to calculate the potentials, and it will be presented in section 1.1.3.

Uniqueness of solutions:

Consider a volume $\mathcal{V}$ bounded by a surface Σ. We specify the values of the potential V for all points on the surface. Then the solution to Poisson's (or Laplace's) equation subject to these so–called *Dirichlet* boundary conditions always exists and is unique.

Although it is not straightforward to prove that the solution exists, we can demonstrate easily that if it exists then it is unique. Let us assume that there are two solutions, V_1 and V_2, which satisfies Poisson's equation and the specified boundary conditions on the surface. Then $W = V_1 - V_2$ itself satisfies Laplace's equation and the condition that $W = 0$ on the surface. We consider the volume integral:

$$\mathcal{I} = \iiint_{\mathcal{V}} (\boldsymbol{\nabla} W)^2 \, d\tau.$$

If W is not a constant in the volume, then $\mathcal{I} > 0$, otherwise $\mathcal{I} = 0$. We write the identity $\boldsymbol{\nabla} \cdot (\alpha \mathbf{U}) = \mathbf{U} \cdot \boldsymbol{\nabla}\alpha + \alpha \boldsymbol{\nabla} \cdot \mathbf{U}$, where $\mathbf{U}$ is a vector and α is a scalar, with $\mathbf{U} = \boldsymbol{\nabla} W$ and $\alpha = W$. Using $\nabla^2 W = 0$, we obtain $\boldsymbol{\nabla} \cdot (W \boldsymbol{\nabla} W) = (\boldsymbol{\nabla} W)^2$, so that $\mathcal{I}$ can be written as:

$$\mathcal{I} = \iiint_{\mathcal{V}} \boldsymbol{\nabla} \cdot (W \boldsymbol{\nabla} W) d\tau = \iint_{\Sigma} W \boldsymbol{\nabla} W \cdot d\boldsymbol{\Sigma},$$

where we have used the divergence theorem to obtain the surface integral. Since $W = 0$ on the surface, the surface integral, and therefore $\mathcal{I}$, is zero. This implies that W is a constant in the volume $\mathcal{V}$. As $W = 0$ on the surface and W is continuous, $W = 0$ everywhere in the volume, and the solutions V_1 and V_2 are identical.

Similar result is obtained if we have *Neumann* boundary conditions instead of Dirichlet boundary conditions, that is to say we specify the values of $\partial V(\mathbf{r})/\partial n \equiv \hat{\mathbf{n}} \cdot \boldsymbol{\nabla} V$ for all points $\mathbf{r}$ on the surface, where $\hat{\mathbf{n}}$ is the unit vector perpendicular to the surface and pointing outward from the surface at $\mathbf{r}$.

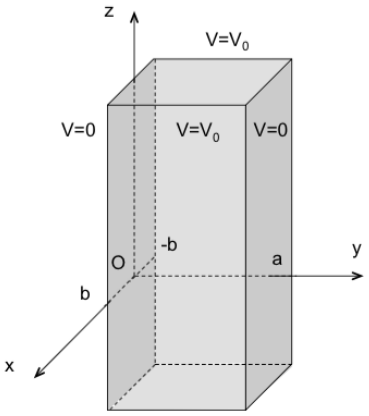
1.1.3 Solutions to Laplace's equation: separation of variables

Note that in general we are interested in calculating the potentials away from the distribution of charges or currents, that is to say where $\rho = 0$ and $\mathbf{J} = \mathbf{0}$. We are therefore going to focus on Laplace's, rather than Poisson's, equations.

Laplace's equation is a second–order partial differential equation, and does not have explicit solutions in three dimensions. However, solutions can be calculated when the variables are separable, as we are going to see in this section. The idea is to try to find a solution which is a product of independent functions, each of which depending on only one of the variables.

Cartesian coordinates:

We illustrate the method of variable separation in cartesian coordinates by studying a particular example. We consider a rectangular pipe which is infinite in the z–direction (meaning the length of the pipe in this direction is very large compared to the dimensions of its cross section).



The pipe is bounded by metal plates. The plates at $y = 0$ and $y = a$ are grounded, so that their potential is $V = 0$. The potential of the plates at $x = -b$ and $x = b$ is maintained at $V = V_0$. We want to calculate the potential V everywhere in the pipe. Cartesian coordinates are here a natural choice.

First we note that V does not depend on z, so Laplace's equation reduces to the two–dimensional equation:

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} = 0. \tag{1.17}$$

We look for solutions in the form $V(x, y) = F(x)G(y)$. Substituting into equation (1.17) and dividing by FG, we obtain:

$$\frac{1}{F}\frac{d^2F}{dx^2} + \frac{1}{G}\frac{d^2G}{dy^2} = 0.$$

As the first term on the left–hand–side can depend only on x, and the second term only on y, their sum is zero only if they are both constant. We therefore have:

$$\frac{1}{G}\frac{d^2G}{dy^2} = C, \quad \frac{1}{F}\frac{d^2F}{dx^2} = -C.$$

Let us assume that the constant C is positive, so that it can be written as $C = k^2$, where $k = +\sqrt{C}$. Then the solutions of the above equations are $F(x) = A_1 \cos(kx) + A_2 \sin(kx)$ and $G(y) = B_1 \mathrm{e}^{ky} + B_2 \mathrm{e}^{-ky}$. The boundary conditions at $y = 0$ and $y = a$, which are $G(0) = G(a) = 0$, imply $B_1 + B_2 = 0$ and $B_1 \mathrm{e}^{ka} + B_2 \mathrm{e}^{-ka} = 0$, so that $2B_1 \sinh(ka) = 0$. This can be satisfied only if $B_1 = 0$, and therefore $B_2 = 0$. As $V = 0$ everywhere in the pipe is not a solution, it means that C cannot be positive.

We then consider $C = -k^2$, so that $F(x) = A_1 \mathrm{e}^{kx} + A_2 \mathrm{e}^{-kx}$ and $G(y) = B_1 \cos(ky) + B_2 \sin(ky)$, with $k > 0$. The potential is symmetric with respect to x, that is to say $F(-x) = F(x)$, which implies $A_1 = A_2$. The boundary conditions at $y = 0$ and $y = a$

imply $B_1 = 0$ and $B_2 \sin(ka) = 0$, which requires $ka = n\pi$, with n a positive integer. The solutions to equation (1.17) which satisfy the boundary conditions at $y = 0$ and $y = a$ are then:

$$V(x, y) = \sum_{n=1}^{\infty} C_n \cosh(n\pi x/a) \sin(n\pi y/a), \tag{1.18}$$

where the C_n are constant. The solution we are looking for has to satisfy also the boundary condition at $x = b$ (the condition at $x = -b$ will then be satisfied automatically since the solution is symmetric with respect to x):

$$V(b, y) = \sum_{n=1}^{\infty} C_n \cosh(n\pi b/a) \sin(n\pi y/a) = V_0. \tag{1.19}$$

To calculate the C_n, we multiply the above equation by $\sin(m\pi y/a)$, with m a positive integer, and integrate over y from 0 to a:

$$\sum_{n=1}^{\infty} C_n \cosh(n\pi b/a) \int_0^a \sin(n\pi y/a) \sin(m\pi y/a) dy = V_0 \int_0^a \sin(m\pi y/a) dy.$$

Since:

$$\sin \frac{n\pi y}{a} \sin \frac{m\pi y}{a} = \frac{1}{2} \left[\cos \frac{(n-m)\pi y}{a} - \cos \frac{(n+m)\pi y}{a} \right],$$

the integral over y on the left–hand–side of the equation above is $a/2$ if $n = m$ and 0 otherwise. Therefore, this equation gives: $C_m \cosh(m\pi b/a) = 0$ if m is even and $4V_0/(m\pi)$ if m is odd.[2]

Finally, the potential is:

$$V(x, y) = \frac{4V_0}{\pi} \sum_{p=0}^{\infty} \frac{1}{2p+1} \frac{\cosh[(2p+1)\pi x/a]}{\cosh[(2p+1)\pi b/a]} \sin[(2p+1)\pi y/a].$$

Cylindrical coordinates:

An example of the method of variable separation in cylindrical coordinates will be studied in tutorial (Problem set 1).

[2]The method described here is rooted in the fact that the sine functions $\sin(n\pi y/a)$ are complete on the interval $0 \le y \le a$ and are orthogonal. *Completeness* means that any function $f(y)$ can be written as a linear combination of the sine functions in the interval $[0, a]$:

$$f(y) = \sum_{n=1}^{\infty} \alpha_n \sin(n\pi y/a),$$

where the α_n are constant. It is this property that ensures that the coefficients C_n satisfying equation (1.19) exist. *Orthogonality* means that the product of any two sine functions is zero:

$$\int_0^a \sin(n\pi y/a) \sin(m\pi y/a) dy = 0 \quad \text{for} \quad m \neq n.$$

It is this property that enables the C_n to be calculated.

Spherical coordinates:

Laplace's equation in spherical coordinates (r, θ, φ) takes the form:

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial V}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial V}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2 V}{\partial\varphi^2} = 0. \tag{1.20}$$

We look for solutions that are separable in the variables: $V(r,\theta,\varphi) = F(r)G(\theta)H(\varphi)$. Note that G and H have to be periodic with a period 2π. Substituting in equation (1.20) and multiplying by $r^2/(FGH)$, we obtain:

$$\frac{1}{F}\frac{d}{dr}\left(r^2\frac{dF}{dr}\right) + \frac{1}{G\sin\theta}\frac{d}{d\theta}\left(\sin\theta\frac{dG}{d\theta}\right) + \frac{1}{H\sin^2\theta}\frac{d^2H}{d\varphi^2} = 0. \tag{1.21}$$

If we multiply this equation by $\sin^2\theta$, the last term on the left–hand–side, $(d^2H/d\varphi^2)/H$, depends only on φ, whereas the first two terms depend on r and/or θ. Therefore, this last term has to be a constant:

$$\frac{1}{H}\frac{d^2H}{d\varphi^2} = K.$$

If $K > 0$, H is a linear superposition of $e^{\pm\sqrt{K}\varphi}$ terms, and cannot be 2π–periodic. Therefore, the constant has to be negative and we write $K = -m^2$, with m either positive or negative. This implies:

$$H(\varphi) = Ce^{im\varphi}, \tag{1.22}$$

where C is a constant and we have used complex notations. For H to be 2π–periodic, m has to be an integer. Substituting this solution into equation (1.21), we obtain:

$$\frac{1}{F}\frac{d}{dr}\left(r^2\frac{dF}{dr}\right) + \frac{1}{G\sin\theta}\frac{d}{d\theta}\left(\sin\theta\frac{dG}{d\theta}\right) - \frac{m^2}{\sin^2\theta} = 0. \tag{1.23}$$

The first term on the left–hand–side depends only on r, whereas the second and third terms depend only on θ. Therefore, the first term and the sum of the two other terms must each be a constant:

$$\frac{1}{F}\frac{d}{dr}\left(r^2\frac{dF}{dr}\right) = l(l+1), \tag{1.24}$$

$$\frac{1}{G\sin\theta}\frac{d}{d\theta}\left(\sin\theta\frac{dG}{d\theta}\right) - \frac{m^2}{\sin^2\theta} = -l(l+1). \tag{1.25}$$

(We write the constant in this form for reasons that will become clear later). Using the variable $x = \cos\theta$ instead of θ, equation (1.25) can be written in the form:

$$\frac{d}{dx}\left[(1-x^2)\frac{dG(x)}{dx}\right] + \left[l(l+1) - \frac{m^2}{1-x^2}\right]G(x) = 0. \tag{1.26}$$

This a well–known differential equation called the associated Legendre equation. The solutions are the **associated Legendre polynomials**:

$$G(x) = P_l^m(x), \quad \text{with } x = \cos\theta. \tag{1.27}$$

The constant l has to be a positive integer, otherwise the solution of equation (1.25) diverges for $\theta = 0$ or π. This is why we wrote the constant above as $l(l+1)$. Also m

can only take the integer values between $-l$ and l. The product of G and H is called a **spherical harmonic** and noted Y_l^m:

$$\boxed{Y_l^m(\theta,\varphi) = CP_l^m(\cos\theta)\mathrm{e}^{im\varphi},} \tag{1.28}$$

where the constant C depends on how the spherical harmonics are being normalised.

Finally, the radial equation (1.24) has the general solution:

$$F(r) = \alpha r^l + \frac{\beta}{r^{l+1}}, \tag{1.29}$$

where α and β are two constants.

The general solution of Laplace's equation (1.20) is obtained by summing over all the possible values of l and m:

$$\boxed{V(r,\theta,\varphi) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l}\left(\alpha_{lm}r^l + \frac{\beta_{lm}}{r^{l+1}}\right)Y_l^m(\theta,\varphi).} \tag{1.30}$$

In this course, we will in general focus on systems which have an **azimuthal symmetry**, which implies that V does not depend on φ. This corresponds to $\mathbf{m = 0}$ in the expression of H given by equation (1.22). The solutions of equation (1.26) with $m = 0$ are called **Legendre polynomials** and are noted $P_l(x)$, with $x = \cos\theta$. They are lth–order polynomials in x which contain only even or odd powers depending on wether l is even or odd, respectively. The first four Legendre polynomials are:

$$\boxed{P_0(x) = 1, \quad P_1(x) = x, \quad P_2(x) = \frac{1}{2}(3x^2-1), \quad P_3(x) = \frac{1}{2}(5x^3-3x).} \tag{1.31}$$

The general solution of Laplace's equation (1.20) in the axisymmetric case is therefore:

$$\boxed{V(r,\theta) = \sum_{l=0}^{\infty}\left(\alpha_l r^l + \frac{\beta_l}{r^{l+1}}\right)P_l(\cos\theta).} \tag{1.32}$$

The constants α_l and β_l are determined by the boundary conditions. We are guaranteed that these coefficients can always be calculated because, like the sine functions in the case of cartesian coordinates, the Legendre polynomials constitute a **complete** set of functions on the interval $-1 \leq x \leq 1$ (or, equivalently, $0 \leq \theta \leq \pi$).

An important property of the Legendre polynomials is that they are **orthogonal**:

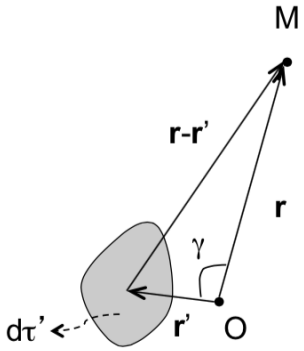
$$\int_{-1}^{+1} P_n(x)P_m(x)dx = \int_0^{\pi} P_n(\cos\theta)P_m(\cos\theta)\sin\theta d\theta = \begin{cases} 0, & \text{if } n \neq m, \\ \frac{2}{2n+1}, & \text{if } n = m. \end{cases} \tag{1.33}$$

This property is useful for calculating the constants α_l and β_l in equation (1.32) knowing the boundary conditions, in the same way the orthogonality of the sine functions was used to calculate the constants C_n in equation (1.19).

Note that *each* of the term in the sum (1.32) is itself a solution of Laplace's equation. Of course, a similar expression can be obtained for the three *cartesian* components of $\mathbf{A}$, as they satisfy Laplace's equation.

1.1.4 Multipole expansion and dipoles

Legendre polynomials were first introduced as the coefficients in the expansion of the $1/|\mathbf{r}' - \mathbf{r}|$ term which occurs in the expression of the gravitational potential, which has the same form as the electrostatic potential given by equation (1.12).



For $r' < r$:

$$\boxed{\frac{1}{|\mathbf{r}' - \mathbf{r}|} = \frac{1}{r}\sum_{l=0}^{\infty}\left(\frac{r'}{r}\right)^l P_l(\cos\gamma),} \tag{1.34}$$

where γ is the angle between the vectors $\mathbf{r}$ and $\mathbf{r}'$. This so–called *multipole expansion* is equivalent to a Taylor–Maclaurin series of $1/|\mathbf{r}' - \mathbf{r}|$ about $r'/r = 0$.

Electric dipole:

Using the expansion (1.34), the electrostatic potential due to a localised charge distribution and given by equation (1.12) can be written as:

$$V(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\sum_{l=0}^{\infty}\frac{1}{r^{l+1}}\iiint_{\mathcal{V}}(r')^l P_l(\cos\gamma)\rho(\mathbf{r}')d\tau'. \tag{1.35}$$

If we are calculating the potential far away from the charge distribution, then we can choose the origin O close to or in the distribution so that $r' \ll r$. In that case, the term $l = 0$ in the expansion is dominant. It is called the **monopole term** and it is given by:

$$V_0(\mathbf{r}) = \frac{1}{4\pi\epsilon_0 r}\iiint_{\mathcal{V}}\rho(\mathbf{r}')d\tau' = \frac{Q}{4\pi\epsilon_0 r},$$

where Q is the total charge of the distribution. This is consistent with the fact that, from far away, the distribution is seen as a point charge.
The next term, which corresponds to $l = 1$, is called the **dipole term** and is the dominant term if $Q = 0$. It is given by:

$$V_1(\mathbf{r}) = \frac{1}{4\pi\epsilon_0 r^2} \iiint_{\mathcal{V}} r' \cos\gamma \rho(\mathbf{r}') d\tau' = \frac{1}{4\pi\epsilon_0 r^2} \hat{\mathbf{r}} \cdot \iiint_{\mathcal{V}} \mathbf{r}' \rho(\mathbf{r}') d\tau',$$

where $\hat{\mathbf{r}} = \mathbf{r}/r$ and we have used $r' \cos\gamma = \hat{\mathbf{r}} \cdot \mathbf{r}'$. By definition, the **electric dipole moment** of the charge distribution is:

$$\boxed{\mathbf{p} = \iiint_{\mathcal{V}} \mathbf{r}' \rho(\mathbf{r}') d\tau',} \tag{1.36}$$

so that the dipole term in the expansion of the potential is:

$$\boxed{V_1(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \frac{\mathbf{p} \cdot \hat{\mathbf{r}}}{r^2}.} \tag{1.37}$$

If the distribution is made of point charges, the dipole moment is written as:

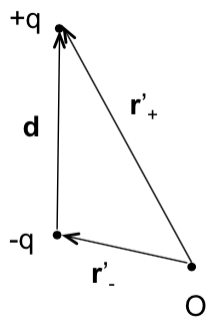
$$\mathbf{p} = \sum_i q_i \mathbf{r}'_i, \tag{1.38}$$

where the sum is over all the charges.

In general, the dipole moment as defined by equation (1.36) depends on the choice of the origine O from which the position vector $\mathbf{r}'$ is measured. If we chose an other origin O' from which the position vector is $\mathbf{r}''$, then the dipole moment becomes $\mathbf{p}'$ such that:

$$\mathbf{p}' = \iiint_{\mathcal{V}} \mathbf{r}'' \rho(\mathbf{r}'') d\tau'' = \iiint_{\mathcal{V}} \left(-\mathbf{OO}' + \mathbf{r}'\right) \rho(\mathbf{r}') d\tau' = -\mathbf{OO}' Q + \mathbf{p}.$$

Therefore, the dipole moment is independent of the choice of the origin of the coordinate system if $Q = 0$ (like in neutral molecules).



Let us consider a charge distribution which is a **physical dipole**, namely two equal and opposite charges $+q$ and $-q$ very close to each other (which means we are interested in the field at a distance from the charges large compared to their separation). The dipole moment is then $\mathbf{p} = q\mathbf{r}'_+ - q\mathbf{r}'_- = q(\mathbf{r}'_+ - \mathbf{r}'_-)$, that is to say:

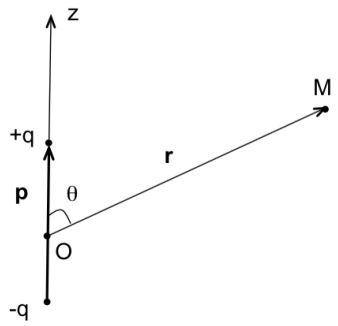
$$\boxed{\mathbf{p} = q\mathbf{d}.} \tag{1.39}$$

where $\mathbf{d}$ is the vector from the $-q$ charge to the $+q$ charge.

For this distribution of charges, $V_0 = 0$ as $Q = 0$. The dominant term is therefore the dipole term V_1.

By taking the gradient of V_1, we obtain the **electric field due to the dipole**:

$$\boxed{\mathbf{E}_1 = \frac{1}{4\pi\epsilon_0 r^3} \left[3 \left(\mathbf{p} \cdot \hat{\mathbf{r}}\right) \hat{\mathbf{r}} - \mathbf{p}\right].} \tag{1.40}$$



If we define a coordinate system with $\mathbf{p}$ at the origin O and along the z–axis, then the components of $\mathbf{E}$ in spherical coordinates (r, θ, φ) are:

$$\begin{aligned} E_r &= \frac{2p\cos\theta}{4\pi\epsilon_0 r^3}, \\ E_\theta &= \frac{p\sin\theta}{4\pi\epsilon_0 r^3}, \\ E_\varphi &= 0. \end{aligned}$$

Magnetic dipole:

Here we focus on the vector potential due to a current loop, which is given by the integral (1.16), with Γ being the contour delimiting the loop. Using the expansion in Legendre polynomials given by equation (1.34), we can write this integral as:

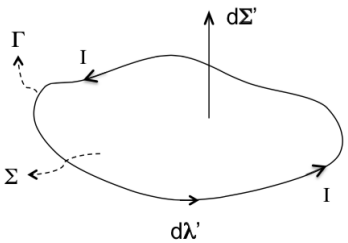
$$\mathbf{A}(\mathbf{r}) = \frac{\mu_0 I}{4\pi}\sum_{l=0}^{\infty}\frac{1}{r^{l+1}}\oint_\Gamma (r')^l P_l(\cos\gamma)\, d\boldsymbol{\lambda}', \tag{1.41}$$

where I is the current through the loop. As for the electric scalar potential, we define the **monopole term** corresponding to $l = 0$:

$$\mathbf{A}_0(\mathbf{r}) = \frac{\mu_0 I}{4\pi r}\oint_\Gamma d\boldsymbol{\lambda}'.$$

As the loop is closed, the integral is zero, so that **the monopole term is always zero**.

The dominant term in the expansion is therefore the $l = 1$ dipole term, and it will give an approximate expression of the vector potential at a point located far away from the loop (which means at a distance r from the loop large compared to the dimensions of the loop).



The **dipole term** is given by:

$$\begin{aligned} \mathbf{A}_1(\mathbf{r}) &= \frac{\mu_0 I}{4\pi r^2}\oint_\Gamma r'\cos\gamma\, d\boldsymbol{\lambda}', \\ &= \frac{\mu_0 I}{4\pi r^2}\oint_\Gamma (\hat{\mathbf{r}}\cdot\mathbf{r}')\, d\boldsymbol{\lambda}'. \end{aligned}$$

We use the equality[3]:

$$\oint_\Gamma (\hat{\mathbf{r}}\cdot\mathbf{r}')\, d\boldsymbol{\lambda}' = -\hat{\mathbf{r}}\times\boldsymbol{\Sigma},$$

[3] We write Stokes's theorem for the vector $(\hat{\mathbf{r}}\cdot\mathbf{r}')\mathbf{V}$, where $\mathbf{V}$ is a constant vector:

$$\oint_\Gamma (\hat{\mathbf{r}}\cdot\mathbf{r}')\mathbf{V}\cdot d\boldsymbol{\lambda}' = \iint_\Sigma \boldsymbol{\nabla}'\times\left[(\hat{\mathbf{r}}\cdot\mathbf{r}')\mathbf{V}\right]\cdot d\boldsymbol{\Sigma}'.$$

where $\boldsymbol{\Sigma}$ is the vector which modulus is the surface delimited by the loop and which orientation is given by the right–hand rule.

We define the **magnetic dipole moment** as:

$$\boxed{\mathbf{m} = I\boldsymbol{\Sigma},} \tag{1.42}$$

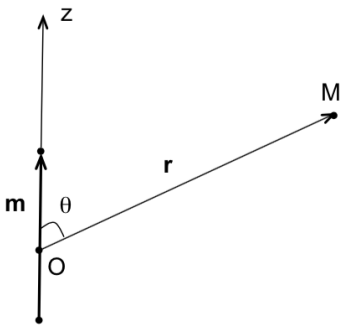
so that the dipole term in the expansion of the potential is:

$$\boxed{\mathbf{A}_1(\mathbf{r}) = \frac{\mu_0}{4\pi}\frac{\mathbf{m}\times\hat{\mathbf{r}}}{r^2}.} \tag{1.43}$$

By taking the curl of $\mathbf{A}_1$, we obtain the **magnetic field due to the dipole**:

$$\boxed{\mathbf{B}_1 = \frac{\mu_0}{4\pi r^3}\left[3\left(\mathbf{m}\cdot\hat{\mathbf{r}}\right)\hat{\mathbf{r}} - \mathbf{m}\right].} \tag{1.44}$$

This is very similar to the electric field due to an electric dipole and given by equation (1.40).



If we define a coordinate system with $\mathbf{m}$ at the origin O and along the z–axis, then the components of $\mathbf{B}$ in spherical coordinates (r, θ, φ) are:

$$\begin{aligned} B_r &= \frac{2\mu_0 m\cos\theta}{4\pi r^3}, \\ B_\theta &= \frac{\mu_0 m\sin\theta}{4\pi r^3}, \\ B_\varphi &= 0. \end{aligned}$$

The derivatives in $\boldsymbol{\nabla}'$ are taken with respect to the coordinates of $\mathbf{r}'$. Since $\mathbf{V}$ is constant, $\boldsymbol{\nabla}'\times[(\hat{\mathbf{r}}\cdot\mathbf{r}')\mathbf{V}] = -\mathbf{V}\times\boldsymbol{\nabla}'(\hat{\mathbf{r}}\cdot\mathbf{r}')$. If we write $\hat{\mathbf{r}}\cdot\mathbf{r}' = (xx' + yy' + zz')/r$, we see easily that $\boldsymbol{\nabla}'(\hat{\mathbf{r}}\cdot\mathbf{r}') = \hat{\mathbf{r}}$. Therefore, Stokes's theorem can be written as:

$$\mathbf{V}\cdot\oint_\Gamma(\hat{\mathbf{r}}\cdot\mathbf{r}')d\boldsymbol{\lambda}' = -\iint_\Sigma(\mathbf{V}\times\hat{\mathbf{r}})\cdot d\boldsymbol{\Sigma}' = -\mathbf{V}\cdot\iint_\Sigma\hat{\mathbf{r}}\times d\boldsymbol{\Sigma}'.$$

As this equality is satisfied for all constant vectors $\mathbf{V}$, it implies:

$$\oint_\Gamma(\hat{\mathbf{r}}\cdot\mathbf{r}')d\boldsymbol{\lambda}' = -\iint_\Sigma\hat{\mathbf{r}}\times d\boldsymbol{\Sigma}' = -\hat{\mathbf{r}}\times\iint_\Sigma d\boldsymbol{\Sigma}' = -\hat{\mathbf{r}}\times\boldsymbol{\Sigma}.$$

1.2 Time–dependent scalar and vector potentials

For time–dependent fields in the presence of electric charges of density ρ and electric currents of density $\mathbf{J}$, Maxwell's equations are:

$$\boldsymbol{\nabla}\cdot\mathbf{E} = \frac{\rho}{\epsilon_0}, \tag{1.45}$$

$$\boldsymbol{\nabla}\cdot\mathbf{B} = 0, \tag{1.46}$$

$$\boldsymbol{\nabla}\times\mathbf{E} = -\frac{\partial\mathbf{B}}{\partial t}, \tag{1.47}$$

$$\boldsymbol{\nabla}\times\mathbf{B} = \mu_0\mathbf{J} + \mu_0\epsilon_0\frac{\partial\mathbf{E}}{\partial t}. \tag{1.48}$$

In electrostatics, we have shown that, because $\boldsymbol{\nabla}\times\mathbf{E} = \mathbf{0}$, we can define a scalar potential V such that $\mathbf{E} = -\boldsymbol{\nabla}V$ (eq. [1.5]). In electrodynamics, $\boldsymbol{\nabla}\times\mathbf{E}$ is non zero and therefore the scalar potential cannot be defined that way.

In magnetostatics we have established that, since $\boldsymbol{\nabla}\cdot\mathbf{B} = \mathbf{0}$, we can define a vector potential $\mathbf{A}$ such that $\mathbf{B} = \boldsymbol{\nabla}\times\mathbf{A}$ (eq. [1.6]). This is still valid in electrodynamics, as $\mathbf{B}$ is always divergence free.
Substituting equation (1.6) into Faraday's law (1.47), we obtain:

$$\boldsymbol{\nabla}\times\mathbf{E} = -\frac{\partial}{\partial t}\left(\boldsymbol{\nabla}\times\mathbf{A}\right),$$

which is equivalent to:

$$\boldsymbol{\nabla}\times\left(\mathbf{E} + \frac{\partial\mathbf{A}}{\partial t}\right) = \mathbf{0}.$$

Since the quantity in parenthesis is curl free, it can be written as the gradient of a scalar potential V:

$$\mathbf{E} + \frac{\partial\mathbf{A}}{\partial t} = -\boldsymbol{\nabla}V.$$

Therefore, in electrodynamics, the scalar potential V and vector potential $\mathbf{A}$ are defined through:

$$\boxed{\begin{aligned}\mathbf{B} &= \boldsymbol{\nabla}\times\mathbf{A}, \qquad (1.49)\\ \mathbf{E} &= -\frac{\partial\mathbf{A}}{\partial t} - \boldsymbol{\nabla}V. \qquad (1.50)\end{aligned}}$$

When the fields are static, we recover $\mathbf{E} = -\boldsymbol{\nabla}V$.

1.2.1 Gauge transformations

As we have shown in section 1.1.1, the magnetic field is unchanged if we replace $\mathbf{A}$ by $\mathbf{A}' = \mathbf{A} + \boldsymbol{\nabla}\phi$, where ϕ is a scalar. But then, for $\mathbf{E}$ given by equation (1.50) to be unchanged, we must replace V by V' such that:

$$\frac{\partial\mathbf{A}'}{\partial t} + \boldsymbol{\nabla}V' = \frac{\partial\mathbf{A}}{\partial t} + \boldsymbol{\nabla}V,$$

which can also be written as:

$$\boldsymbol{\nabla} V' = \boldsymbol{\nabla} V - \frac{\partial}{\partial t} \left(\boldsymbol{\nabla} \phi \right).$$

This is satisfied if $V' = V - \partial\phi/\partial t$.
Therefore, $\mathbf{B}$ and $\mathbf{E}$ are unchanged if $\mathbf{A}$ and V are subject to the so–called **gauge transformation**[4]:

$$\boxed{\mathbf{A}' = \mathbf{A} + \boldsymbol{\nabla}\phi, \quad V' = V - \frac{\partial \phi}{\partial t}.} \tag{1.51}$$

As pointed out in section 1.1.2, the divergence of the vector potential can be chosen freely. Specifying the divergence of the vector potential is called **choosing a gauge**. In magnetostatics, we chose the **Coulomb gauge** $\boldsymbol{\nabla} \cdot \mathbf{A} = 0$. In electrodynamics, the divergence of the vector potential is chosen such as to simplify Maxwell's equations, as shown in the next section.

1.2.2 Lorenz gauge and Maxwell's equations

With $\mathbf{B}$ and $\mathbf{E}$ given by equations (1.49) and (1.50), Maxwell's equations (1.46) and (1.47) are satisfied. We now insert equation (1.50) into Gauss's law (1.45):

$$\nabla^2 V + \frac{\partial}{\partial t} \left(\boldsymbol{\nabla} \cdot \mathbf{A} \right) = -\frac{\rho}{\epsilon_0}. \tag{1.52}$$

Finally, we insert the expressions of $\mathbf{E}$ and $\mathbf{B}$ into equation (1.48):

$$\boldsymbol{\nabla} \times \left(\boldsymbol{\nabla} \times \mathbf{A} \right) = \mu_0 \mathbf{J} - \mu_0 \epsilon_0 \frac{\partial}{\partial t} \left(\frac{\partial \mathbf{A}}{\partial t} + \boldsymbol{\nabla} V \right).$$

Using the identity $\boldsymbol{\nabla} \times \left(\boldsymbol{\nabla} \times \mathbf{A} \right) = \boldsymbol{\nabla} \left(\boldsymbol{\nabla} \cdot \mathbf{A} \right) - \nabla^2 \mathbf{A}$, the above equation can be written under the form::

$$\nabla^2 \mathbf{A} - \mu_0 \epsilon_0 \frac{\partial^2 \mathbf{A}}{\partial t^2} - \boldsymbol{\nabla} \left(\boldsymbol{\nabla} \cdot \mathbf{A} + \mu_0 \epsilon_0 \frac{\partial V}{\partial t} \right) = -\mu_0 \mathbf{J}. \tag{1.53}$$

Using the Coulomb gauge $\boldsymbol{\nabla} \cdot \mathbf{A} = 0$ would enable equation (1.52) to be written as Poisson's equation: $\nabla^2 V = -\rho/\epsilon_0$. However, equation (1.53) could not be simplified with this gauge. Therefore, we instead use the **Lorenz gauge** given by:

$$\boxed{\boldsymbol{\nabla} \cdot \mathbf{A} = -\mu_0 \epsilon_0 \frac{\partial V}{\partial t}.} \tag{1.54}$$

Substituting into equations (1.52) and (1.53), we obtain Maxwell's equations under the form:

$$\boxed{\nabla^2 V - \frac{1}{c^2} \frac{\partial^2 V}{\partial t^2} = -\frac{\rho}{\epsilon_0},} \tag{1.55}$$

$$\boxed{\nabla^2 \mathbf{A} - \frac{1}{c^2} \frac{\partial^2 \mathbf{A}}{\partial t^2} = -\mu_0 \mathbf{J},} \tag{1.56}$$

[4]The word "gauge" was originally used to mean "size" or "scale" , as the "track gauge" used in rail transport, and which is the distance between the inner faces of the rails.

where we have used $1/c^2 = \mu_0\epsilon_0$. The operator:

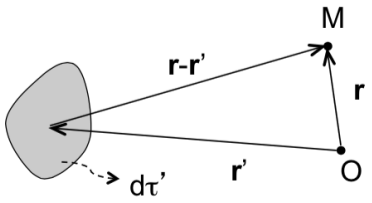
$$\Box \equiv \nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2},$$

which appears in the two equations above, is called the **d'Alembertian**. These equations are **inhomogeneous wave equations**. If the source terms on the right–hand–side are zero, we obtain homogeneous wave equations of the type that will be studied in chapter 4.

Instead of solving Maxwell's equations (1.45)–(1.48) for $\mathbf{E}$ and $\mathbf{B}$, we can therefore solve equations (1.55) and (1.56) for V and $\mathbf{A}$, and then calculate $\mathbf{B}$ and $\mathbf{E}$ using equations (1.49) and (1.50).

1.2.3 Retarded potentials

When the fields are static, the Lorenz gauge reduces to the Coulomb gauge, and equations (1.55) and (1.56) reduce to Poisson's equations $\nabla^2 V = -\rho/\epsilon_0$ and $\nabla^2\mathbf{A} = -\mu_0\mathbf{J}$. The solutions of these equations in integral form are given by equations (1.12) and (1.13).



If the fields are time–dependent, changes in the charges or currents contained in the volume element $d\tau'$ that occur at a time t are "felt" at the location of the point M only after a time $\Delta t = |\mathbf{r}' - \mathbf{r}|/c$, where c is the speed of light, since the electromagnetic field travels at the speed c.

Therefore, there is a delay between changes in the sources and the adjustment of the field at a given location. It can be shown that the solutions of equations (1.55) and (1.56) in the time–dependent case are the so–called **retarded potentials**:

$$V(\mathbf{r},t) = \frac{1}{4\pi\epsilon_0}\iiint_{\mathcal{V}} \frac{\rho(\mathbf{r}', t-\Delta t)}{|\mathbf{r}'-\mathbf{r}|}d\tau', \tag{1.57}$$

$$\mathbf{A}(\mathbf{r},t) = \frac{\mu_0}{4\pi}\iiint_{\mathcal{V}} \frac{\mathbf{J}(\mathbf{r}', t-\Delta t)}{|\mathbf{r}'-\mathbf{r}|}d\tau'. \tag{1.58}$$

Chapter 2

Electric fields in matter

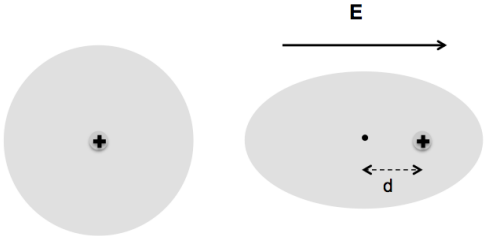
In first year, we studied electrostatic potentials and fields in the presence of charges and conductors. In conducting materials, there are charges which are free to move in the whole volume of the material. When an external electric field is applied, the free charges move under the action of the electric force. They re–arrange themselves in such a way that they create an internal electric field which cancels the applied field: the total electric field in the conductor is therefore zero.

In this chapter, we are going to study electric fields in **insulators**, also called **dielectrics**, in which electrons are bound to atoms or molecules. When an external field is applied, it can displace the electrons only to some extent. Therefore, the induced charge creates an internal field that opposes the external field, but does not cancel it entirely. The term *di–electric* comes from the greek $\delta\iota\alpha$ which means *through*, as electromagnetic fields can penetrate inside the material.

2.1 Polarization in dielectrics

2.1.1 Atomic dipoles

We first consider a dielectric made of neutral atoms. When an external electric field is applied in such a dielectric, the nucleus and the electron cloud that constitute an atom are moved in opposite directions. An equilibrium is reached when the electric force due to the external field balances the force of interaction between the nucleus and the electron cloud.



This is illustrated in the figure on the left, which gives a schematic view of an atom. The + sign indicates the centre of mass of the nucleus and the shaded cloud represents the electron distribution. The dot in the centre of the ellipse is the centre of mass of the electron distribution.

The nucleus reaches an equilibrium at a distance d from the centre of mass of the electron

distribution such that $\mathbf{E} = -\mathbf{E}'$, where $\mathbf{E}$ is the external field and $\mathbf{E}'$ is the field due to the electrons at the position of the nucleus. Note that the electron cloud becomes distorted as it moves under the action of the electric field. We neglect this distorsion, and will justify this approximation below. Assuming the electron cloud to be a uniformly charged sphere of radius a with total charge $-q$, Gauss's law gives:

$$E' = \frac{qd}{4\pi\epsilon_0 a^3}.$$

Therefore, at equilibrium:

$$E = \frac{qd}{4\pi\epsilon_0 a^3}.$$

As the centre of mass of the electron distribution and the nucleus do not coincide, the atom has a dipole moment $p = qd$ pointing toward the nucleus. Therefore:

$$\boxed{\mathbf{p} = \alpha\mathbf{E},} \tag{2.1}$$

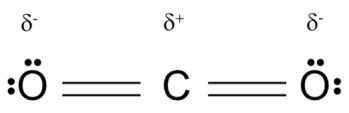
with $\alpha = 4\pi\epsilon_0 a^3$ being the **atomic polarizability**.
As noted above, the electron distribution is distorted by the external field. The flattening of the distribution (or ratio of the axes of the ellipse) is on the order of the ratio of the external field E to the field exerted by the nucleus on the electrons, which is on the order of $q/(4\pi\epsilon_0 a^2)$. Using the above expression for E, we obtain that the flattening is $\sim d/a$. For $a \sim 10^{-10}$ m (typical atomic dimension), the internal field $|q|/(4\pi\epsilon_0 a^2)$ is on the order of 10^{11} V m^{-1}. As this is much larger than any external large–scale steady field that could be imposed, the flattening d/a is very small and can therefore be neglected.

The model presented above is of course a very crude approximation. However, it gives a rather good prediction for the polarizability of atoms. The approximate size of a ground state hydrogen atom is given by the Bohr radius, which is $\simeq 0.5$ Å. The corresponding polarizability is therefore $\alpha/(4\pi\epsilon_0) \sim 10^{-31}$ m^3, close to the experimental value of 0.67×10^{-30} m^3. For atoms in their ground state, the measured polarizability ranges from about 0.2×10^{-30} m^3 for helium to about 60×10^{-30} m^3 for cesium. Noble gases, with filled valence shells, have low polarizability. By contrast, alkali metals, with only one electron in the valence shell, have high polarizability.

2.1.2 Molecular dipoles

Nonpolar molecules:

Some molecules do not have a permanent electric dipole moment. This is the case of the carbon dioxide molecule CO_2 for example, as represented in the figure below.



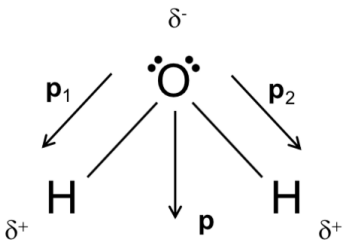
Lewis structure of the CO_2 molecule: the dots represent electrons, and the lines joining the atoms represent chemical bonds (each made of an electron pair). Carbon and oxygen have 4 and 6 valence electrons, respectively.

Because oxygen is more electronegative than carbon (which is indicated by the symbols δ^+ and δ^- on the figure above), there is an excess of positive charge on the carbon atom and an excess of negative charges on the oxygen atoms. But because of the symmetry of the molecule, there is no net electric dipole.

However, like neutral atoms, the molecule can become polarized when an electric field is applied, because the centres of mass of the negative and positive charges are moved in opposite directions. The polarizability depends on the direction of the applied field. For the CO_2 molecule, $\alpha/(4\pi\epsilon_0)$ is measured to be 4×10^{-30} or 1.8×10^{-30} m^3 depending on whether the electric field is applied parallel or perpendicular to the axis of the molecule.

Polar molecules:

Some molecules have a permanent electric dipole moment, like the water molecule H_2O represented on the figure below.



Lewis structure of the H_2O molecule. Hydrogen and oxygen have 1 and 6 valence electrons, respectively. The excesses of negative and positive charges on the oxygen and hydrogen atoms, respectively, induce the dipole moments $\mathbf{p}_1$ and $\mathbf{p}_1$. There is therefore a net dipole moment $\mathbf{p} = \mathbf{p}_1 + \mathbf{p}_2$.

The dipole moment of water, $p = 6.2 \times 10^{-30}$ C m, is rather large. If we had $p = ed$, then d would be about 0.4 Å. This makes water a very good solvent, as ionic compounds become dissociated in their positive and negative ions (the cations are attracted by the oxygen atoms of the water molecules whereas the anions are attracted by the hydrogen atoms).

In a dielectric made of polar molecules, the electric dipole moments are oriented randomly when no external electric field is applied. Therefore, no large scale polarization is present. When an external field is applied, it exerts a torque on the molecular dipoles. As a result, the molecules rotate until their dipoles align with the field. So, even though the applied field does not induce the molecular dipoles, it induces a large scale dipole.

2.1.3 Polarization

When an external electric field is present in a dielectric material, it either induces atomic or molecular dipole moments parallel to its own direction, or aligns already existing molecular dipoles along its own direction. In all cases, a large scale electric dipole moment is induced. The **dipole moment per unit volume** is called the **polarization vector** (or just *polarization*) and is noted $\mathbf{P}$. If there are N dipole moments $\mathbf{p}$ per unit volume, then $\mathbf{P} = N\mathbf{p}$.

2.2 Polarization charges and current

When an external field is applied, charges move slightly. As we have seen, this results in an induced polarization in the dielectric material. It also causes a net charge to appear in the material, while a charge with the opposite sign accumulates at the surface. They are called **polarization** or **bound** charges, as they cannot leave the material.

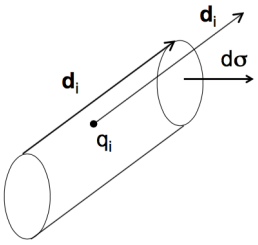
Moreover, if the polarization varies with time, charges keep moving and create a so–called **polarization current**.

2.2.1 Polarization charges

We consider an arbitrary domain of the dielectric material with a volume v and delimited by a closed surface σ. As we have seen above, the electric dipole moment $\mathbf{p}_i$ of an atom or a molecule can be modelled as the displacement $\mathbf{d}_i$ of a charge q_i. If there are n_i atoms or molecules of charge q_i per unit volume, then the polarization if given by:

$$\mathbf{P} = \sum_i n_i q_i \mathbf{d}_i. \tag{2.2}$$

We are now going to calculate the flux of charges through the surface σ due to the displacements $\mathbf{d}_i$.



The charges q_i which cross the surface element $d\sigma$ are contained in a cylinder of length d_i parallel to the vector $\mathbf{d}_i$ and cross–sectional area $d\sigma$. The volume of this cylinder is $\mathbf{d}_i \cdot d\boldsymbol{\sigma}$. Therefore, the total charge which crosses the surface element $d\sigma$ is $\sum_i n_i q_i \mathbf{d}_i \cdot d\boldsymbol{\sigma}$. Using the expression of $\mathbf{P}$ given above, this can be written as $\mathbf{P} \cdot d\boldsymbol{\sigma}$. The total charge which leaves the volume v is therefore $\iint_\sigma \mathbf{P} \cdot d\boldsymbol{\sigma}$. As the volume was initially neutral, it is left with a net charge $Q_p = -\iint_\sigma \mathbf{P} \cdot d\boldsymbol{\sigma}$.

Using the divergence theorem, we can also write:

$$Q_p = \iiint_v -\boldsymbol{\nabla} \cdot \mathbf{P} d\tau.$$

As this is valid for any volume v of the dielectric, it follows that the dielectric contains a density of **polarization charges per unit volume** given by:

$$\boxed{\rho_p = -\boldsymbol{\nabla} \cdot \mathbf{P}.} \tag{2.3}$$

If we now consider the whole volume of the dielectric, we get that the total charge which accumulates at the surface Σ of the dielectric is $\iint_\Sigma \mathbf{P} \cdot d\boldsymbol{\Sigma}$ (the displacements $\mathbf{d}_i$ are very small, so charges move only locally and dot not leave the material). This corresponds to a density of **polarization charges per unit surface**:

$$\boxed{\sigma_p = \mathbf{P} \cdot \hat{\mathbf{n}},} \tag{2.4}$$

where $\hat{\mathbf{n}}$ is the unit vector perpendicular to the surface area and directed outward.

2.2.2 Polarization current

The calculation we have done above is valid whether the polarization is static or not. When $\mathbf{P}$ varies with time, the displacement $\mathbf{d}_i$ of a charge q_i also varies with time. The charge therefore has a velocity $\mathbf{v}_i = d\mathbf{d}_i/dt$. This corresponds to a so–called polarization current which density per unit volume is:

$$\mathbf{J}_p = \sum_i n_i q_i \mathbf{v}_i = \sum_i n_i q_i \frac{d\mathbf{d}_i}{dt}.$$

Using the expression of the polarization given by equation (2.2), we obtain:

$$\boxed{\mathbf{J}_p = \frac{\partial \mathbf{P}}{\partial t}.} \tag{2.5}$$

This current is macroscopic and is an average of all the microscopic currents produced by the small displacements of the polarization (bound) charges. The polarization current is also called *bound current*, because it can flow only in the dielectric. It cannot be made to flow in wires outside the material, contrary to conduction curents, as polarization charges are bound. Like a current produced by the charges in a conductor, the polarization current induces magnetic effects.

To define the macroscopic quantities $\mathbf{P}$ and $\mathbf{J}_p$, which are continuous, we have implicitly assumed that the quantities could be averaged over a scale D small compared to the size of the dielectric but large compared to the size of atoms or molecules. Let us consider the case where the polarization is produced by an electromagnetic wave of period T and wavelength λ. The average cannot be done if λ is smaller than D or, equivalently, if T is smaller than the time it takes for electromagnetic waves to travel through the distance D. With D on the order of a few nm (10^{-9} m), it means that the dielectric cannot be described in terms of a macroscopic polarization to study the propagation of waves in the X–ray domain of the spectrum. For these waves, the discontinuous nature of the material has to be taken into account.

Charge conservation:

By taking the divergence of equation (2.5) and using equation (2.3), we obtain:

$$\boldsymbol{\nabla} \cdot \mathbf{J}_p = -\frac{\partial \rho_p}{\partial t}, \tag{2.6}$$

which is a continuity equation expressing the conservation of the polarization charge. This can be seen by writing the integral form of this equation:

$$\oiint_{\Sigma} \mathbf{J}_p \cdot d\boldsymbol{\Sigma} = -\iiint_{\mathcal{V}} \frac{\partial \rho_p}{\partial t} d\tau,$$

where we have used the divergence theorem, and the second integral is over the volume $\mathcal{V}$ delimited by the surface Σ. As the surface delimiting the volume is fixed, the term on the right–hand–side can also be written $-dQ_p/dt$, where Q_p is the total polarization charge in the volume. Therefore, the rate of change of Q_p is equal to the negative of the flux of current density outward through the surface which delimits the volume.

2.3 The electric field caused by polarized matter

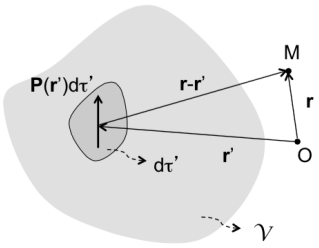
As a dielectric contains (polarization) charges, it produces an electric field, which will be superposed to the field that has produced the polarization in the first place. We are now going to calculate, in the **static** case, the electric potential from which the field produced by the polarization charges can be derived.

2.3.1 Field outside the dielectric

As we have seen in electrostatics, the electric potential created by a dipole moment $\mathbf{p}$ at a point M is:

$$V = \frac{1}{4\pi\epsilon_0}\frac{\mathbf{p}\cdot\hat{\mathbf{r}}}{r^2},$$

where $\mathbf{r}$ is the vector position of M measured from the dipole and $\hat{\mathbf{r}} = \mathbf{r}/r$ (see eq. [1.37]).



In a dielectric, the dipole moment is $\mathbf{P}$ per unit volume, so that the total potential at point M is:

$$V(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\iiint_{\mathcal{V}}\frac{\mathbf{P}(\mathbf{r}')d\tau'\cdot(\mathbf{r}-\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|^3}, \qquad (2.7)$$

where the position vectors $\mathbf{r}$ and $\mathbf{r}'$ are measured from a fixed origin O and the integration is over the volume $\mathcal{V}$ of the dielectric.

We note that:

$$\frac{\mathbf{r}-\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|^3} = \boldsymbol{\nabla}'\left(\frac{1}{|\mathbf{r}-\mathbf{r}'|}\right), \qquad (2.8)$$

where the derivatives are taken with respect to the coordinates $\mathbf{r}'$ defined from O (it can be shown easily by using cartesian coordinates). In addition, if α is a scalar and $\mathbf{A}$ a vector, we have the identity $\boldsymbol{\nabla}\cdot(\alpha\mathbf{A}) = \alpha(\boldsymbol{\nabla}\cdot\mathbf{A}) + \mathbf{A}\cdot(\boldsymbol{\nabla}\alpha)$. The potential at M can therefore be written as:

$$V(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\left\{\iiint_{\mathcal{V}}\boldsymbol{\nabla}'\cdot\left[\frac{1}{|\mathbf{r}-\mathbf{r}'|}\mathbf{P}(\mathbf{r}')\right]d\tau' - \iiint_{\mathcal{V}}\frac{1}{|\mathbf{r}-\mathbf{r}'|}\left[\boldsymbol{\nabla}'\cdot\mathbf{P}(\mathbf{r}')\right]d\tau'\right\}.$$

Using the divergence theorem, we obtain:

$$V(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\left\{\oiint_{\Sigma}\frac{1}{|\mathbf{r}-\mathbf{r}'|}\mathbf{P}(\mathbf{r}')\cdot d\boldsymbol{\Sigma}' - \iiint_{\mathcal{V}}\frac{1}{|\mathbf{r}-\mathbf{r}'|}\left[\boldsymbol{\nabla}'\cdot\mathbf{P}(\mathbf{r}')\right]d\tau'\right\},$$

where Σ is the closed surface that delimits the volume $\mathcal{V}$ of the dielectric. With the expressions of the polarization charges given by equations (2.3) and (2.4), the potential becomes:

$$V(\mathbf{r}) = \oiint_{\Sigma}\frac{\sigma_p(\mathbf{r}')d\Sigma'}{4\pi\epsilon_0|\mathbf{r}-\mathbf{r}'|} + \iiint_{\mathcal{V}}\frac{\rho_p(\mathbf{r}')d\tau'}{4\pi\epsilon_0|\mathbf{r}-\mathbf{r}'|}. \qquad (2.9)$$

The two terms on the right–hand side give the contributions from the surface charge density and volume charge density, respectively.

The corresponding electric field can be calculated from: $\mathbf{E}(\mathbf{r}) = -\boldsymbol{\nabla}V$, where the derivatives are taken with respect to the coordinates $\mathbf{r}$ defined from O.

This calculation confirms that the dielectric can be described either as a material having an induced polarization $\mathbf{P}$ or as a material with a distribution of charges given by equations (2.3) and (2.4). The two representations are **equivalent**.

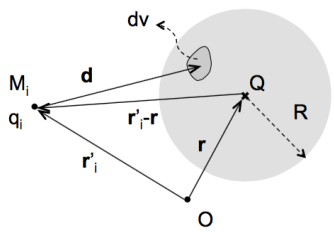
2.3.2 Field inside the dielectric

This section may be skipped on first reading.

In the previous section, we have adopted an idealised representation of a dielectric as a continous distribution of dipoles or charges. This is a good approximation when calculating the potential outside the dielectric, as the distance to the molecules or atoms is then large compared to their separation, and inhomogeneities within the dielectric average out. However, the situation is *a priori* more complicated when we study the electric field inside the dielectric, as it varies hugely over very short distances. For example, the field at a distance of 1 Å from an electron is enormous, on the order of 10^{11} V m^{-1}. A little further away from the electron, the field may become very small if contributions from different charges cancel. For most purposes, this so–called **microscopic** field is not of interest, as instruments detect only a value of the field averaged over many atomic distances, namely over a volume large compared to the size of an atom but small compared to the scale of the material. This averaged field is called **macroscopic**. Following Griffiths (section 4.2.3) and Feynman (volume II, chapter 11, section 4), we now show that the macroscopic electric field inside the material can be calculated in the same way as the electric field outside.

To calculate the macroscopic electric field $\mathbf{E}$ at a point Q in the material, we average the microscopic fields over a sphere of appropriate radius R around Q. Using the principle of superposition, we can write $\mathbf{E}$ as the sum of the average fields over the sphere due to the charges outside and inside the sphere: $\mathbf{E} = \mathbf{E}_{\text{out}} + \mathbf{E}_{\text{in}}$. The calculation we present below is a bit lengthy but does not require complicated integrals to be computed. See Jackson, Section 4.1, for a more elegant but more technical calculation of the average field over a sphere.

Contribution from the charges outside the sphere:



The field produced in a small volume element dv in the sphere by a charge q_i located at a point M_i outside the sphere is $q_i\mathbf{d}/(4\pi\epsilon_0 d^3)$, where $\mathbf{d}$ is the vector position of dv measured from M_i. The average field over the volume of the sphere due to the charge is therefore:

$$\langle \mathbf{E}_i \rangle = \frac{1}{\mathcal{V}} \iiint_{\mathcal{V}} \frac{q_i \mathbf{d}}{4\pi\epsilon_0 d^3}\, dv,$$

where $\mathcal{V}$ is the volume of the sphere.

But this is also the field that would be produced at M_i by a uniform charge density $-q_i/\mathcal{V}$ inside the sphere (the minus sign in the charge density comes from the orientation of $\mathbf{d}$). Using Gauss's theorem, we can therefore write $4\pi|\mathbf{r}'_i - \mathbf{r}|^2\langle\mathbf{E}_i\rangle = (-q_i/\epsilon_0)(\mathbf{r}'_i - \mathbf{r})/|\mathbf{r}'_i - \mathbf{r}|$, where the position vectors $\mathbf{r}$ and $\mathbf{r}'_i$ are measured from an origin O. The total field $\mathbf{E}_{\text{out}}$ is obtained by summing over the charges outside the sphere, so that:

$$\mathbf{E}_{\text{out}} = \sum_{i,\text{outside}} \langle\mathbf{E}_i\rangle = \sum_{i,\text{outside}} \frac{q_i(\mathbf{r} - \mathbf{r}'_i)}{4\pi\epsilon_0|\mathbf{r} - \mathbf{r}'_i|^3}.$$

We approximate the distribution of charges by a continuous distribution, so that $q_i = \rho(\mathbf{r}'_i)d\tau'$, with $\rho(\mathbf{r}'_i)$ being the charge density in a small volume element $d\tau'$ around M_i. We can then write the field as:

$$\mathbf{E}_{\text{out}} = \iiint_{\text{outside}} \frac{\rho(\mathbf{r}')(\mathbf{r} - \mathbf{r}')d\tau'}{4\pi\epsilon_o|\mathbf{r} - \mathbf{r}'|^3},$$

where we have dropped the subscript i.

We can see from the above expression that **the average field over the sphere due to the charges outside is equal to the field produced at the centre of the sphere**.

We can equivalently represent the distribution of charges by a distribution of dipoles. Therefore, we can write $\mathbf{E}_{\text{out}} = -\boldsymbol{\nabla}V_{\text{out}}$ (evaluated at Q) with:

$$V_{\text{out}}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \iiint_{\text{outside}} \frac{\mathbf{P}(\mathbf{r}')d\tau' \cdot (\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^3}, \tag{2.10}$$

(see eq. [2.7]) where $\mathbf{P}(\mathbf{r}') \equiv \mathbf{r}'\rho(\mathbf{r}')$ is the electric dipole moment per unit volume (eq.[1.36]).

Contribution from the charges inside the sphere:

Following the same argument as above, we note that the average field $\langle\mathbf{E}_i\rangle$ over the sphere due to a charge q_i located at a point M_i inside the sphere is equal to the field that would be produced at M_i by a uniform charge density $-q_i/\mathcal{V}$ inside the sphere. Gauss's theorem then gives:

$$4\pi|\mathbf{r}'_i - \mathbf{r}|^2\langle\mathbf{E}_i\rangle = \frac{1}{\epsilon_0}\frac{4}{3}\pi|\mathbf{r}'_i - \mathbf{r}|^2(\mathbf{r}'_i - \mathbf{r})\left(\frac{-q_i}{\mathcal{V}}\right).$$

With $\mathcal{V} = 4\pi R^3/3$, we obtain:

$$\langle\mathbf{E}_i\rangle = \frac{-q_i(\mathbf{r}'_i - \mathbf{r})}{4\pi\epsilon_0 R^3},$$

and the field due to all the charges inside the sphere is therefore:

$$\mathbf{E}_{\text{in}} = \sum_{i,\text{inside}} \langle\mathbf{E}_i\rangle = \sum_{i,\text{inside}} \frac{-q_i(\mathbf{r}'_i - \mathbf{r})}{4\pi\epsilon_0 R^3}.$$

By definition, the dipole moment of the sphere is:

$$\boldsymbol{\mathcal{P}} = \sum_{i,\text{inside}} q_i(\mathbf{r}'_i - \mathbf{r})$$

(see eq. [1.38] with the position of the charges being measured from the center of the sphere). We can then write the field under the form:

$$\mathbf{E}_{\text{in}} = -\frac{\boldsymbol{\mathcal{P}}}{4\pi\epsilon_0 R^3} = -\frac{\mathbf{P}}{3\epsilon_0}, \tag{2.11}$$

where $\mathbf{P} = 3\boldsymbol{\mathcal{P}}/(4\pi R^3)$ is the dipole moment per unit volume that we assume here to be uniform (as the sphere is very small). The field given by equation (2.11) is the field produced by a uniformly polarized sphere inside the sphere[1].

Therefore, **the average field over the sphere due to the charges inside is the same as the (uniform) field that would be produced in the sphere by assuming it to be uniformly polarized**.

The potential produced by the uniformly polarized sphere at a point with position vector $\mathbf{r}$ inside the sphere is:

$$V_{\text{in}}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \iiint_{\text{inside}} \frac{\mathbf{P}d\tau' \cdot (\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^3}, \tag{2.12}$$

which is the same as equation (2.10) with the integral being over the sphere and $\mathbf{P}$ assumed to be uniform. The electric field can be calculated from $\mathbf{E}_{\text{in}} = -\boldsymbol{\nabla} V_{\text{in}}$.

Total average field:

From the calculations above, we see that the macroscopic electric field at Q, which we define as being the average field over the sphere of radius R centered on Q, derives from a potential $V_{\text{out}} + V_{\text{in}}$ which can be written in exactly the same form as that given by equation (2.7).

Therefore, **the electric field can always be calculated by summing over the contribution from the dipoles (or equivalently the charges) which are in the dielectric**. This is true whether the field is calculated outside or inside the dielectric.

2.4 The electric displacement vector D

As we have seen above, an external field present in a dielectric induces a polarization of the material, which in turn produces an electric field. The total electric field, which is the superposition of these two fields, then modifies the polarization, which in turn affects the field, etc. In general, we are only interested in the total field, which is produced by all the charges present.

2.4.1 Gauss's law

The total charge density is $\rho = \rho_p + \rho_f$, where ρ_p are the **polarization, or bound**, charges, and ρ_f are all the other, **free**, charges. Free charges may be electrons in a

[1] This result will be shown in Problem Set 1.

conductor (outside the dielectric), or ions contained within the dielectric and not bound to atoms or molecules. The total field is given by Gauss's law:

$$\boldsymbol{\nabla} \cdot \mathbf{E} = \frac{\rho}{\epsilon_0} = \frac{\rho_p + \rho_f}{\epsilon_0}. \tag{2.13}$$

Using equation (2.3), this is equivalent to: $\boldsymbol{\nabla} \cdot (\epsilon_0 \mathbf{E} + \mathbf{P}) = \rho_f$.

We define the **electric displacement** vector $\mathbf{D}$ (where '$\mathbf{D}$' stands for 'displacement') by:

$$\boxed{\mathbf{D} \equiv \epsilon_0 \mathbf{E} + \mathbf{P},} \tag{2.14}$$

so that Gauss's law can be written in terms of the free charges only:

$$\boxed{\boldsymbol{\nabla} \cdot \mathbf{D} = \rho_f.} \tag{2.15}$$

The integral form of this equation is:

$$\oiint_{\Sigma} \mathbf{D} \cdot d\boldsymbol{\Sigma} = Q_f, \tag{2.16}$$

where Q_f is the total free charge enclosed by the surface Σ over which the integral is calculated.

It may seem from equation (2.15) that $\mathbf{D}$ depends only on the distribution of free charges. It is incorrect, as $\boldsymbol{\nabla} \times \mathbf{D}$ depends on $\mathbf{P}$. In the case where the fields are static, we have $\boldsymbol{\nabla} \times \mathbf{D} = \boldsymbol{\nabla} \times \mathbf{P}$, so that even if $\mathbf{P}$ is uniform in the volume of the dielectric, a discontinuity of $\mathbf{P}$ at the surface of the dielectric may induce a discontinuity of $\mathbf{D}$ there (see below).

2.4.2 Ampère's law

As we have mentioned in section 2.2.2, the polarization current due to a time–dependent electric field induces a magnetic field which has to be included in Ampère's law. The total current is $\mathbf{J} = \mathbf{J}_p + \mathbf{J}_f$, where $\mathbf{J}_p$ is the polarization, or bound, current, and $\mathbf{J}_f$ is the free current, that is to say the current produced by free (non bound) charges. Ampère's law therefore becomes:

$$\boldsymbol{\nabla} \times \mathbf{B} = \mu_0 \left(\mathbf{J}_p + \mathbf{J}_f \right) + \mu_0 \epsilon_0 \frac{\partial \mathbf{E}}{\partial t}. \tag{2.17}$$

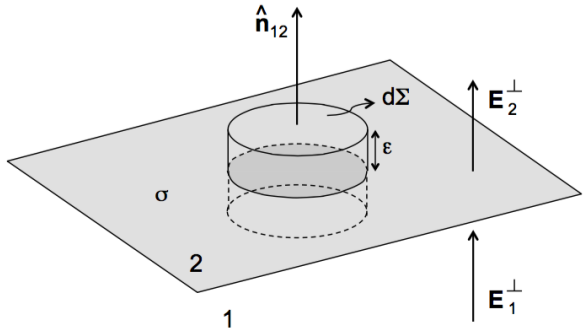
Using equations (2.5) and (2.14), we obtain:

$$\boxed{\boldsymbol{\nabla} \times \mathbf{B} = \mu_0 \mathbf{J}_f + \mu_0 \frac{\partial \mathbf{D}}{\partial t}.} \tag{2.18}$$

As above, using $\mathbf{D}$ instead of $\mathbf{E}$ enables the equation to be written in terms of the free current only. The name electric *displacement* field comes from the fact that $\mathbf{D}$ generates a *displacement current* $\partial \mathbf{D}/\partial t$, as can be seen from the above equation.

2.4.3 Boundary conditions on E and D

We consider a surface with a charge density $\sigma = \sigma_p + \sigma_f$ separating two media. Locally, the surface is flat.



We apply the integral form of Gauss's law:

$$\oiint \mathbf{E} \cdot d\mathbf{\Sigma} = \iiint \frac{\rho}{\epsilon_0} d\tau,$$

with the surface being that of a cylinder with length 2ϵ and cross sectional area $d\Sigma$ and ρ is the charge density in the cylinder.

The indices 1 and 2 refer to the regions of space below and above the surface, respectively. We note $\hat{\mathbf{n}}_{12}$ the unit vector perpendicular to the surface and pointing from medium (1) to medium (2). From the above equality we obtain: $\mathbf{E}_2 \cdot \hat{\mathbf{n}}_{12} d\Sigma - \mathbf{E}_1 \cdot \hat{\mathbf{n}}_{12} d\Sigma = (\sigma/\epsilon_0) d\Sigma$, where the minus sign before $\mathbf{E}_1$ comes from the orientation of the unit vector $\hat{\mathbf{n}}_{12}$ and $d\Sigma$ is small enough that σ is uniform over this surface element. We have taken the limit $\epsilon \to 0$ so that the flux of $\mathbf{E}$ through the lateral surface of the cylinder is negligible. If we note $\mathbf{E}^\perp = E^\perp \hat{\mathbf{n}}_{12}$, with $E^\perp = \mathbf{E} \cdot \hat{\mathbf{n}}_{12}$, the component of $\mathbf{E}$ perpendicular to the surface and very close to it, the above equality can be written in the following way:

$$\mathbf{E}_2^\perp - \mathbf{E}_1^\perp = \frac{\sigma}{\epsilon_0} \hat{\mathbf{n}}_{12}. \tag{2.19}$$

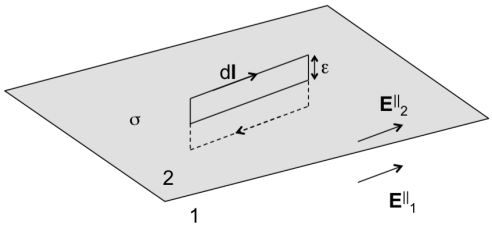
Gauss's law applied to $\mathbf{D}$ similarly gives:

$$\boxed{\mathbf{D}_2^\perp - \mathbf{D}_1^\perp = \sigma_f \hat{\mathbf{n}}_{12}.} \tag{2.20}$$

If there are **no free charges** on the surface, $\mathbf{D}^\perp$ is **continuous**.

If both regions 1 and 2 are dielectrics, we have $\mathbf{D}_2^\perp = \epsilon_0 \mathbf{E}_2^\perp + \mathbf{P}_2^\perp$ and $\mathbf{D}_1^\perp = \epsilon_0 \mathbf{E}_1^\perp + \mathbf{P}_1^\perp$, where $\mathbf{P}_1$ and $\mathbf{P}_2$ are the polarization vectors. With $\epsilon \to 0$, $P_1^\perp = \mathbf{P}_1 \cdot \hat{\mathbf{n}}_{12} = \sigma_{1,p}$ and $P_2^\perp = \mathbf{P}_2 \cdot \hat{\mathbf{n}}_{12} = -\sigma_{2,p}$, with $\sigma_{1,p}$ and $\sigma_{2,p}$ being the densities of polarization charges at the surface of the dielectrics 1 and 2, respectively. Equation (2.20) then becomes $\epsilon_0 \mathbf{E}_2^\perp - \epsilon_0 \mathbf{E}_1^\perp = \sigma_p \hat{\mathbf{n}}_{12}$ (we assume there is no free charges), with $\sigma_p = \sigma_{1,p} + \sigma_{2,p}$. This is the same as equation (2.19) with $\sigma_f = 0$.

We suppose that the fields are **static** so that $\mathbf{\nabla} \times \mathbf{E} = \mathbf{0}$.



Therefore $\oint \mathbf{E}\cdot d\mathbf{l} = 0$ along the contour represented on the figure, where the horizontal sides of the loop are parallel to the tangential component of the field $\mathbf{E}^{\parallel}$. Then, taking $\epsilon \to 0$, we obtain

$$\boxed{\mathbf{E}_2^{\parallel} - \mathbf{E}_1^{\parallel} = \mathbf{0},} \tag{2.21}$$

which means that $\mathbf{E}^{\parallel}$ is **continuous**.
Using equation (2.14), this gives:

$$\mathbf{D}_2^{\parallel} - \mathbf{D}_1^{\parallel} = \mathbf{P}_2^{\parallel} - \mathbf{P}_1^{\parallel}. \tag{2.22}$$

2.5 Linear dielectrics

Some dielectrics, called *electrets*, have a permanent polarization. However, such materials are rare. Most of the time, as we have seen in section 2.1, the polarization results from atomic or molecular dipoles lining up with an applied electric field. The polarization itself produces a field which in turn modifies the alignment of the microscopic dipoles. In many materials, the final state is a polarization that is proportional to the total electric field:

$$\boxed{\mathbf{P} = \epsilon_0 \chi_e \mathbf{E},} \tag{2.23}$$

where χ_e is a (positive) dimensionless constant called the **electric susceptibility**. This relation holds only when the electric field is not too large. For (very) large fields, at least on the order of a few 10^6 V m^{-1}, *breakdown* occurs, that is to say the dielectric material becomes conducting and the relation above is no longer valid.
The materials in which this relation is satisfied are called **linear**, **homogeneous** and **isotropic**, as χ_e does not depend on E , is the same everywhere in the material and in all directions, respectively.
When the electric field varies with time, the above relation holds only if the polarization adjusts instantaneously to the variations of the field. This is in general not the case when the field varies very rapidly, as it takes a finite time for the atomic or molecular dipoles to be modified. **When there is a delay between the variation of the field and the adjustment of the polarization, the electric susceptibility depends on frequency** (we will study this case later in the course, in section 5.6). In atoms and molecules, electrons can be moved on very short timescales, comparable to the period of visible light waves. Therefore, in a dielectric made of atoms or non polar molecules, the polarization can keep up with a changing field up to frequencies close to those of visible light, and the electric susceptibility is independent of frequency. However, in a dielectric made of polar molecules, dipoles have to re-oriente themselves when the electric field varies. The timescales involved are rather long, so that the electric susceptibility may be dependent on frequencies even when those are rather low. In liquid water for example,

the electric susceptibility is constant below 10^{10} Hz (microwaves) and becomes negligible at higher frequencies, when the dipoles cannot re-oriente themselves fast enough to keep up with the changes in the field. In ice just below 0° C, the transition is at around 10^3 Hz.

Using the expression of $\mathbf{D}$ given by equation (2.14), we obtain:

$$\mathbf{D} = \epsilon_0 \left(1 + \chi_e\right) \mathbf{E},$$

which is also written:

$$\boxed{\mathbf{D} = \epsilon \mathbf{E},} \tag{2.24}$$

with:

$$\boxed{\epsilon \equiv \epsilon_0 \left(1 + \chi_e\right).} \tag{2.25}$$

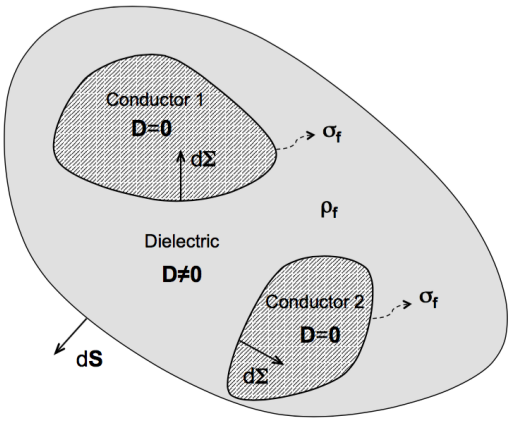
The constant ϵ is the **permittivity** of the material. In vacuum, $\mathbf{P}$ is zero, which implies $\chi_e = 0$, so that $\epsilon = \epsilon_0$. Therefore ϵ_0 is called the **permittivity of free space**. We further define the dimensionless **relative permittivity**, also called **dielectric constant**, by:

$$\boxed{\epsilon_r \equiv \frac{\epsilon}{\epsilon_0} = 1 + \chi_e.} \tag{2.26}$$

Note that the relation between $\mathbf{D}$ and $\mathbf{E}$ is a *constitutive relation* and cannot be derived from first principles.

2.6 Energy in the presence of dielectrics

Let us consider a dielectric which acquires some polarization $\mathbf{P}$ when placed in a **static** external field, itself generated by free charges. In the first year course, we have defined the energy of a system of charges as the work that has to be done to bring all the charges into the system, one by one, this energy being stored in the electric field. However, in the presence of a dielectric, if we were to define the energy as the work needed to bring both polarization and free charges into the system, we would not take into account the work needed to establish the polarization, that is to say the work needed to distort the atoms and molecules to create dipoles. Therefore, we instead define the energy as the work needed to build up the polarization (or, equivalently, the displacement vector) from zero to its final value, by bringing the *free* charges into the system, one by one.



The potential in which the free charges are moved in is due to both free and polarization charges. These free charges are distributed at the surface Σ of conductors and in some volume $\mathcal{V}$ outside the conductors (which is also occupied at least partially by the dielectric) . We note ρ_f and σ_f their volume and surface densities.

The work done to increase the charge densities by $\delta\rho_f$ and $\delta\sigma_f$ is:

$$\delta W = \iiint_{\mathcal{V}} V(\mathbf{r})\delta\rho_f(\mathbf{r})d\tau + \oiint_{\Sigma} V(\mathbf{r})\delta\sigma_f(\mathbf{r})d\Sigma,$$

where $V(\mathbf{r})$ is the potential (due to both free and polarization charges) at the position $\mathbf{r}$ of the charges.
From equation (2.20), with medium (2) being the conductor, we obtain: $\sigma_f d\Sigma = -\mathbf{D}\cdot d\mathbf{\Sigma}$, where $\mathbf{D}$ is the displacement vector outside the conductor (inside a conductor, $\mathbf{D} = \mathbf{0}$). Using also equation (2.15), we can write δW as:

$$\delta W = \iiint_{\mathcal{V}} V\boldsymbol{\nabla}\cdot(\delta\mathbf{D})d\tau - \oiint_{\Sigma} V\delta\mathbf{D}\cdot d\mathbf{\Sigma},$$

where $\delta\mathbf{D}$ is the variation of $\mathbf{D}$ induced by the changes in the charges (which induce a change in the polarization).
If we call S the surface that delimits the outer parts of the volume $\mathcal{V}$, then the divergence theorem gives:

$$\oiint_{\Sigma} V\delta\mathbf{D}\cdot d\mathbf{\Sigma} + \oiint_{S} V\delta\mathbf{D}\cdot d\mathbf{S} = \iiint_{\mathcal{V}} \boldsymbol{\nabla}\cdot(V\delta\mathbf{D})d\tau.$$

We put the surface S at infinity. At large distances, V and D vary like r^{-1} and r^{-2}, respectively, while dS increases as r^2. Therefore, the integral over S vanishes, and we obtain:

$$\delta W = \iiint_{\mathcal{V}} V\boldsymbol{\nabla}\cdot(\delta\mathbf{D})d\tau - \iiint_{\mathcal{V}} \boldsymbol{\nabla}\cdot(V\delta\mathbf{D})d\tau.$$

Substituting the identity $\boldsymbol{\nabla}\cdot(V\delta\mathbf{D}) = V\boldsymbol{\nabla}\cdot(\delta\mathbf{D}) + \delta\mathbf{D}\cdot\boldsymbol{\nabla}V$ and $\mathbf{E} = -\boldsymbol{\nabla}V$, this gives:

$$\boxed{\delta W = \iiint_{\mathcal{V}} \mathbf{E}\cdot\delta\mathbf{D}d\tau.} \tag{2.27}$$

The work that has to be done to establish the displacement vector from $\mathbf{0}$ to $\mathbf{D}$ is:

$$W = \int_0^D \delta W = \iiint_{\mathcal{V}} d\tau \int_0^D \mathbf{E}\cdot\delta\mathbf{D}.$$

When the field is decreased back to zero, if the curve $E(D)$ is the same as when D was increased, then:

$$\iiint_{\mathcal{V}} d\tau \int_0^D \mathbf{E}\cdot\delta\mathbf{D} = -\iiint_{\mathcal{V}} d\tau \int_D^0 \mathbf{E}\cdot\delta\mathbf{D}.$$

Therefore, the total work done to increase D from zero to its final value and decrease it back to zero is zero. The work done to establish the field can be viewed as being stored in the field reversibly: the energy is given back when the field is returned to zero.

If the dielectric is *linear*, that is to say $\mathbf{D}$ is proportional to $\mathbf{E}$:

$$\mathbf{E}\cdot\delta\mathbf{D} = \frac{1}{2}\delta\left(\mathbf{E}\cdot\mathbf{D}\right),$$

so that:

$$W = \frac{1}{2}\iiint_{\mathcal{V}} \mathbf{E}\cdot\mathbf{D}d\tau.$$

Since $\mathbf{E} = \mathbf{0}$ in conductors, this energy is stored in the field which is in the dielectric and vacuum (if any) that surround the conductors. From the expression of W, we obtain the **energy density for a linear dielectric**:

$$\boxed{U = \frac{1}{2}\mathbf{E} \cdot \mathbf{D}.} \tag{2.28}$$

In vacuum, $\mathbf{D} = \epsilon_0 \mathbf{E}$, so that we recover the expression of the energy density $U = \epsilon_0 E^2/2$ seen in the first year course.

Chapter 3

Magnetic fields in matter

When placed in a strong non–uniform field like that produced by a solenoid, materials are repelled or attracted by the field, even if only very weakly. These materials have become magnetic by interacting with a magnetic field.

In this chapter, we study how materials respond when placed in a magnetic field, and how the field is changed by the presence of the material.

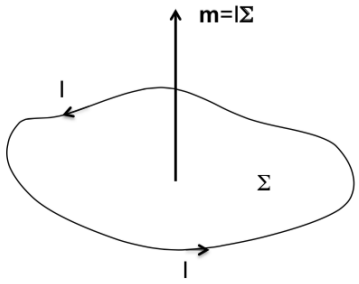
3.1 Magnetic materials

They are three types of magnetic materials: **diamagnetic** materials, which are weakly repelled by a solenoid, **paramagnetic** materials, which are more strongly attracted, and **ferromagnetic** materials, which are strongly pulled into the solenoid. These materials are able to respond to a magnetic field because they contain electric charges in motion, which are the electrons orbiting around nuclei and with a spin. These orbital motions and spins can be viewed as producing microscopic **magnetic dipole moments** which align with each other when an external magnetic field is applied, so that the material becomes magnetically polarized, or **magnetized**.

In paramagnetic materials, the magnetization induced by the external field is parallel to and in the same direction as the field, whereas it is opposite to the field in diamagnetic materials. Ferromagnetic materials stay magnetized even after the external field has been removed, so that the magnetization is not simply related to the external field. In this section, we discuss diamagnetic and paramagnetic materials. Ferromagnetism will be discussed later in the chapter.

3.1.1 Review: Torque and force on a magnetic dipole

In this paragraph, we review results which have been derived in the first year course.



We call magnetic dipole a loop in which flows a steady current I and with dimensions small compared to the distance at which we calculate its effects. The vector potential and magnetic field due to a dipole have been derived in section 1.1.4.

The **magnetic dipole moment** is:

$$\boxed{\mathbf{m} = I\boldsymbol{\Sigma},} \tag{3.1}$$

(see eq.[1.42]) where $\boldsymbol{\Sigma}$ is the vector perpendicular to the plane of the loop and in a direction related to the current by a right–hand–screw rule. The modulus of $\boldsymbol{\Sigma}$ is equal to the area delimited by the loop.

If the loop is placed in a magnetic field $\mathbf{B}$, it experiences a torque given by:

$$\boxed{\mathbf{N} = \mathbf{m}\times\mathbf{B}.} \tag{3.2}$$

To derive this expression, we have assumed that $\mathbf{B}$ was uniform, that is to say we have neglected the variations of $\mathbf{B}$ across the (small) loop.

The mechanical energy of the magnetic dipole in the field $\mathbf{B}$ is:

$$\boxed{U = -\mathbf{m}\cdot\mathbf{B}.} \tag{3.3}$$

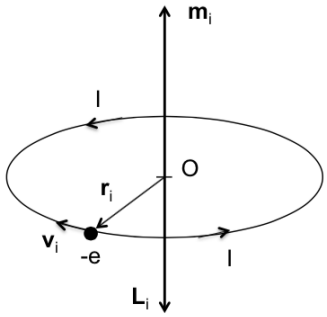
This is the mechanical work that has to be done to bring the steady current dipole into the field. It is the total energy of the dipole in the field only when $\mathbf{B}$ is uniform. When $\mathbf{B}$ is nonuniform, work is done on the conduction electrons in the loop when it is brought into the field. This is due to the emf induced in the current loop by the changing magnetic field through the surface of the dipole. Therefore, that electrical energy has to be provided or absorbed to maintain the current steady (see Feynman, chapter 15).

Even though U is not the total energy of the dipole in a nonuniform field, it can be used to derive the force $\mathbf{F}$ that acts on the dipole when the current I is steady: $\mathbf{F} = -\boldsymbol{\nabla}U$, that is to say:

$$\boxed{\mathbf{F} = \boldsymbol{\nabla}\left(\mathbf{m}\cdot\mathbf{B}\right).} \tag{3.4}$$

In the absence of any other effect, when a magnetic dipole is placed in an external magnetic field, it tends to align with $\mathbf{B}$. The situation where $\mathbf{m}$ and $\mathbf{B}$ are parallel corresponds indeed to zero torque and the lowest mechanical energy.

3.1.2 Diamagnetism



We can understand diamagnetism by adopting a simple classical model of an electron in orbit around a nucleus. We consider an electron i of charge $-e$ which moves with speed v_i on a circular orbit of radius r_i.

The electron completes $v_i/(2\pi r_i)$ orbits per second, so that the current around the circle (that is to say, the charge that passes through a fixed point on the circle per second) is:

$$I = \frac{ev_i}{2\pi r_i}.$$

The current flows in the direction opposite to the electron's motion. The magnetic field produced far away from the atom by this electron is that of a magnetic dipole with moment:

$$m_i = \pi r_i^2 I = \frac{ev_i r_i}{2}.$$

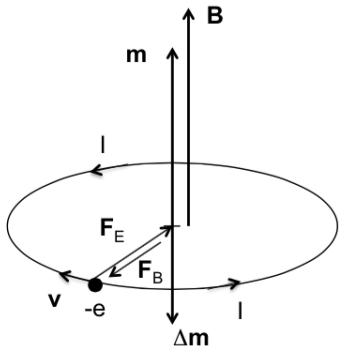
The orbital angular momentum of the electron is $\mathbf{L}_i = m_e \mathbf{r}_i \times \mathbf{v}_i$, where $\mathbf{r}_i$ is the position vector and m_e is the mass of the electron. As the orbit is circular, $L_i = m_e r_i v_i$. The dipole moment can therefore be written as:

$$\mathbf{m}_i = -\frac{e}{2m_e}\mathbf{L}_i. \tag{3.5}$$

The dipole moment of the atom is obtained by summing over all the electrons in the atom:

$$\mathbf{m} = -\frac{e}{2m_e}\sum_i \mathbf{L}_i.$$

In general, the orbital angular momenta of all the electrons in the atom are randomly oriented, so that $\sum_i \mathbf{L}_i = \mathbf{0}$. If, in addition, the resultant angular momentum due to the spin of the electrons is zero (see next section), the atom has **no permanent dipole moment**. Such an atom is called **diamagnetic**.



We now apply an external field **B**. We suppose that **B** is perpendicular to the plane of the orbit of the electron, and we also assume that the radius r of the orbit does not change[1](we drop the subscript 'i' for simplicity).

In the absence of an external magnetic field, the velocity v of the electron can be calculated by writing that the centripetal acceleration is due to the electrostatic force $\mathbf{F}_E$ exerted by the nucleus:

$$\frac{Ze^2}{4\pi\epsilon_0 r^2} = \frac{m_e v^2}{r}, \tag{3.6}$$

where Z is the number of charges of the nucleus. When a magnetic field is present, there is an extra force $\mathbf{F}_B = -e\mathbf{v}\times\mathbf{B}$ in the radial direction. In the case illustrated above, $\mathbf{F}_B$ is opposite to $\mathbf{F}_E$, so that the velocity of the particle is decreased by an amount Δv such that:

$$\frac{Ze^2}{4\pi\epsilon_0 r^2} - e(v-\Delta v)B = \frac{m_e(v-\Delta v)^2}{r}. \tag{3.7}$$

Substituting equation (3.6), we obtain:

$$e(v-\Delta v)B = \frac{m_e}{r}\left[-(v-\Delta v)^2 + v^2\right]. \tag{3.8}$$

Using $\Delta v \ll v$, which is always satisfied on Earth[2], this equation gives:

$$\Delta v = \frac{eBr}{2m_e}. \tag{3.9}$$

This decrease of the velocity leads to a change in angular momentum:

$$\Delta\mathbf{L} = m_e\mathbf{r}\times\Delta\mathbf{v} = \frac{er^2}{2}\mathbf{B},$$

which is aligned with **B**. According to equation (3.5), the corresponding change in the dipole moment is:

$$\Delta\mathbf{m} = -\frac{e}{2m_e}\Delta\mathbf{L} = -\frac{e^2r^2}{4m_e}\mathbf{B}, \tag{3.10}$$

which is **opposite** to the field. It is straightforward to show that the same result is obtained when the velocity of the electron is reversed. In that case indeed, the velocity is increased rather than decreased, and $\Delta\mathbf{m}$ is still opposite to **B**. This result is consistent

[1]See Purcell, Chapter 11, Section 5, for a proof of the fact that the radius of the orbit is not changed when a magnetic field is applied if the atom is not moving.

[2]The change in velocity is small compared to v if $F_B \ll F_E$. Writing $F_B^2 \ll F_E^2$ and using equation (3.6), this is equivalent to $B^2 \ll Zm_e/(4\pi\epsilon_0 r^3)$. With $r \sim 1$ Å, the condition is that B has to be small compared to about 10^5 T, which is met by far on Earth, where the strongest field ever produced is close to 10^3 T .

with Lenz's law, which states that when a field is introduced the current in the loop is modified in such a way as to oppose the effect of the field.

We have assumed above that the field was perpendicular to the orbit of the electron. Of course, in general, this is not the case, and the electric and magnetic forces make the orbit of the electron *precess* around the direction of **B**. In other words, in this situation where the magnetic moment is produced by an electron in orbit around a nucleus, the torque exerted by the magnetic field does not align the magnetic moment with **B** but makes it precess around **B** (this is called *Larmor precession*). However, it is possible to show that equation (3.10) is still valid in this case. As the calculation in the general case is lengthy, we will not give it in these notes.

The conclusion of this section is that, **when a material is placed in a magnetic field, all the atoms or molecules acquire a small dipole moment antiparallel to the field, so that a net dipole moment appears**. This is the phenomenon called **diamagnetism**, and it is due to the orbital angular momentum of the electrons. It is **universal**, meaning it always happens when a magnetic field is present. However, in materials in which atoms have a permanent dipole moment (due to the spin of the electrons), diamagnetism is a small effect as the moment induced by the field is negligible compared to the intrinsic moment of the atoms (see next section).

To obtain the total dipole moment due to diamagnetism, Δm given by equation (3.10) has to be multiplied by the total number of electrons. The number of nucleons per unit mass is the same in all materials (the mass of a nucleon being 1.6×10^{-27} kg), and in all atoms and molecules there is about 1 electron per 2 nucleons. In addition, if we take for r^2 in equation (3.10) the average value in an atom, then it is essentially the same in all atoms. Therefore, for a given external field, the total dipole moment *per unit mass* due to diamagnetism is roughly the same in all materials.

Although it is possible to gain some understanding of how diamagnetism is created using a classical model, as done in this section, it is important to point out that diamagnetism is inherently a *quantum* phenomenon.

3.1.3 Paramagnetism

In addition to their orbital angular momentum, electrons possess a **spin**, which is associated with an angular momentum $\hbar/2$, where $\hbar \equiv h/(2\pi)$ and h is the Planck's constant. Although this is a purely quantum property, its effects, as far as we are concerned in this section, are similar to those that would be produced if the electron were spinning around its own axis. In particular, the electron spin produces a magnetic field which, far enough away from the electron, is that of a magnetic dipole moment. The ratio of magnetic moment to angular momentum for the spin is $-e/m_e$, instead of $-e/(2m_e)$ as we have found for the orbital motion (eq. [3.5]). This cannot be explained with a classical model. When the electron is placed in an external magnetic field **B**, it experiences a torque given by equation (3.2), like in the classical case. This torque tends to align the dipole moment with **B**.

If electrons are associated in pairs, as is the case in most atoms and molecules, as spins within a pair are equal and opposite, there is no net dipole moment associated with the spin. In that case, a dipole moment can be created only through diamagnetism. But if unpaired electrons are present, either because the molecules have an odd number of electrons (which is rare) or because the electronic structure is such that some electrons remain unpaired, their spin dipole moments can reorient themselves when a magnetic field is applied, and **a net moment parallel to B and in the same direction** is obtained. This phenomenon is called **paramagnetism**. Although diamagnetism is also present in such a material, it produces a moment which is smaller than that produced by paramagnetism. The magnetic moment associated with the spin is indeed $eh/(4\pi m_e) = 0.93 \times 10^{-23}$ J T^{-1}, whereas Δm given by equation (3.10) is $2 \times 10^{-29} B$ J T^{-1}. However, note that, at high temperature, the dipole moment associated with the spin is prevented from aligning completely with **B** by thermal motions. Therefore, the total dipole moment is not just the number of unpaired electrons times the spin dipole moment (see Purcell, Chapter 11, Section 6, for a more thorough discussion).

3.1.4 Magnetization

When a material is placed in an external magnetic field, it becomes magnetic. We have seen above that this is due to a net dipole moment induced in the material by the external field.

The magnetic dipole moment per unit volume is called the **magnetization vector** (or just *magnetization*) and is noted **M**.

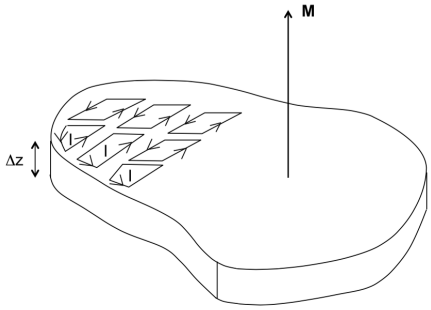
We have discussed diamagnetic and paramagnetic materials, in which **M** is parallel and opposite to or in the same direction as **B**, respectively. Later in this chapter, we will also discuss ferromagnets, which sustain a magnetization even in the absence of an external magnetic field.

3.2 Magnetization currents

In chapter 2, we have found that polarization bound charges appeared in a material placed in an external electric field. Similarly, we are now going to see how currents are produced in a material placed in an external magnetic field.

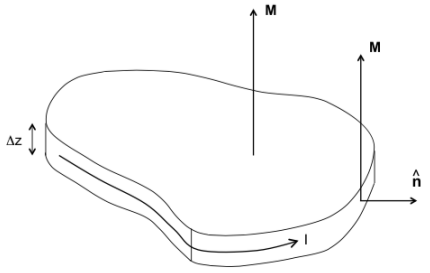
3.2.1 Surface currents

We first consider a magnetic material that has a uniform magnetization **M**, and take a slice of this material perpendicular to **M** and of thickness Δz. An element of surface area a on this slice has a volume $a\Delta z$ and therefore contains a dipole moment $Ma\Delta z$. The magnetic field it produces is the same as that produced by a loop of same surface area and thickness and with a current $I = M\Delta z$.



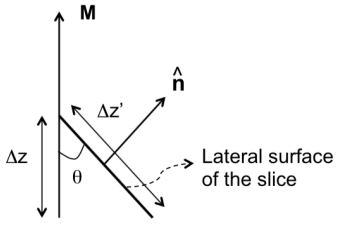
The total magnetization is the same as that we would get if the slice of material were paved with such loops, as illustrated on the figure. We can see that there is no current in the volume of the material, as the current from one loop is cancelled by that of an adjacent loop. However, there is a current I at the surface of the slice.

It is not due to a single charge moving all around the slice, but to lots of charges each moving a little bit.



I is the current that flows along the surface. Therefore, the surface current density is $K_m = I/\Delta z'$, where $\Delta z'$ is the height of the surface measured perpendicularly to the direction in which the current flows. The subscript 'm' indicates that the current is due to magnetization. It is bound, which means that it cannot be used outside of the surface.

On the figure above, the lateral surface of the slice is parallel to $\mathbf{M}$, so that $\Delta z' = \Delta z$. However, this is not the most general case.



In general, the lateral surface is inclined, and $\Delta z' = \Delta z/\cos\theta$. Therefore $K_m = (I/\Delta z)\cos\theta = M\cos\theta$.

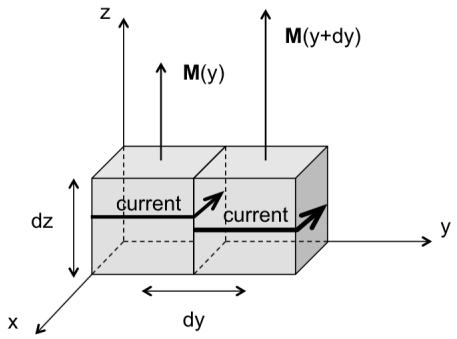
In vector form, the **surface current density** due to magnetization can be written as:

$$\boxed{\mathbf{K}_m = \mathbf{M}\times\hat{\mathbf{n}},} \tag{3.11}$$

where $\hat{\mathbf{n}}$ is the unit vector perpendicular to the surface and directed outward ($K_m = 0$ on the top and bottom surfaces of the slice, where $\mathbf{M}$ and $\hat{\mathbf{n}}$ are parallel). Like the bound polarization current, the current associated with the magnetization of the material induces magnetic effects.

3.2.2 Volume currents

When the magnetization is non uniform, the different loops which pave the slice of material no longer have the same current. We take the loops to be small enough so that the magnetization is uniform over the volume of each of them. The surface current derived above is unchanged, except that now $\mathbf{M}$ depends on the space coordinates, so that K_m may vary along the surface of the slice. But now there is also a current in the volume of the material, as the current from one loop is no longer cancelled by the current from an adjacent loop.



Let us assume that the magnetization is in the z direction and varies along the y direction. On the surface where two loops join, there is a current in the x direction which is given by:

$$I_x = [M_z(y+dy) - M_z(y)]\, dz = \frac{\partial M_z}{\partial y} dydz.$$

A component of the magnetization in the y direction and varying along the z direction would also produce a current in the x direction given by:

$$I_x = -\frac{\partial M_y}{\partial z} dydz.$$

The volume current density in the x direction, J_x, is the current per unit area flowing in the x direction. Therefore, J_x is equal to I_x times the number of loops per unit area, which is $1/(dydz)$. This gives:

$$J_x = \frac{\partial M_z}{\partial y} - \frac{\partial M_y}{\partial z},$$

which is the x component of $\boldsymbol{\nabla}\times\mathbf{M}$.
By calculating the current along the y and z direction in the same way, we obtain the **volume current density** due to magnetization:

$$\boxed{\mathbf{J}_m = \boldsymbol{\nabla}\times\mathbf{M}.} \tag{3.12}$$

Note that, contrary to conduction currents which can release energy through Joule heating for example, bound currents cannot result in energy dissipation as they are produced by the orbital motion or by the spin of electrons moving in vacuum.

3.3 The magnetic field caused by magnetized matter

As magnetic materials contain (bound) currents, they produce a magnetic field which will be superposed to the field that has produced the magnetization in the first place.

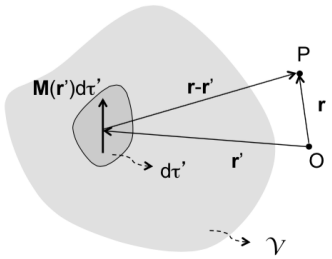
According to the classical model presented above, the magnetic field due to magnetic materials can be regarded as being produced by electric charges which move around nuclei. There is nothing special about these current loops, even though there are inside a material, and therefore the field they produce can be calculated in the same way as the field produced by current loops in vacuum. As there is no magnetic charge, $\boldsymbol{\nabla}\cdot\mathbf{B}=0$, whether the field is microscopic or macroscopic. Consequently, we can always define a vector potential $\mathbf{A}$ such that $\mathbf{B}=\boldsymbol{\nabla}\times\mathbf{A}$.

3.3.1 Field outside the material

As we have seen in chapter 1, the vector potential created by a dipole moment $\mathbf{m}$ at a point P is:

$$\mathbf{A}=\frac{\mu_0}{4\pi}\frac{\mathbf{m}\times\hat{\mathbf{r}}}{r^2},$$

(see eq.[1.43]) where $\mathbf{r}$ is the vector position of P measured from the dipole and $\hat{\mathbf{r}}=\mathbf{r}/r$.



In a magnetic material, the dipole moment is $\mathbf{M}$ per unit volume, so that the total vector potential at point P is:

$$\mathbf{A}(\mathbf{r})=\frac{\mu_0}{4\pi}\iiint_{\mathcal{V}}\frac{\mathbf{M}(\mathbf{r}')d\tau'\times(\mathbf{r}-\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|^3}, \tag{3.13}$$

where the position vectors $\mathbf{r}$ and $\mathbf{r}'$ are measured from a fixed origin O and the integration is over the volume $\mathcal{V}$ of the magnetic material.

Using:

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

(eq. [2.8]), where the derivatives are taken with respect to the coordinates $\mathbf{r}'$ defined from O, and the identity $\boldsymbol{\nabla}\times(\alpha\mathbf{V})=\alpha(\boldsymbol{\nabla}\times\mathbf{V})-\mathbf{V}\times(\boldsymbol{\nabla}\alpha)$, where α is a scalar and $\mathbf{V}$ is a vector, we obtain:

$$\mathbf{A}(\mathbf{r})=\frac{\mu_0}{4\pi}\left\{-\iiint_{\mathcal{V}}\boldsymbol{\nabla}'\times\left[\frac{1}{|\mathbf{r}-\mathbf{r}'|}\mathbf{M}(\mathbf{r}')\right]d\tau'+\iiint_{\mathcal{V}}\frac{1}{|\mathbf{r}-\mathbf{r}'|}\left[\boldsymbol{\nabla}'\times\mathbf{M}(\mathbf{r}')\right]d\tau'\right\}.$$

The first integral on the right–hand side can be written as a surface integral using the relation[3]:

$$\iiint_{\mathcal{V}} (\boldsymbol{\nabla}\times\mathbf{V})\, d\tau = -\oiint_{\Sigma} \mathbf{V}\times d\boldsymbol{\Sigma}, \tag{3.14}$$

where Σ is the closed surface that delimits the volume $\mathcal{V}$.
We then obtain:

$$\mathbf{A}(\mathbf{r}) = \frac{\mu_0}{4\pi}\left\{\oiint_{\Sigma} \frac{1}{|\mathbf{r}-\mathbf{r}'|}\mathbf{M}(\mathbf{r}')\times d\boldsymbol{\Sigma}' + \iiint_{\mathcal{V}} \frac{1}{|\mathbf{r}-\mathbf{r}'|}\left[\boldsymbol{\nabla}'\times\mathbf{M}(\mathbf{r}')\right] d\tau'\right\}.$$

With the expressions of the magnetization currents given by equations (3.11) and (3.12), the vector potential becomes:

$$\mathbf{A}(\mathbf{r}) = \frac{\mu_0}{4\pi}\oiint_{\Sigma} \frac{\mathbf{K}_m(\mathbf{r}')d\Sigma'}{|\mathbf{r}-\mathbf{r}'|} + \frac{\mu_0}{4\pi}\iiint_{\mathcal{V}} \frac{\mathbf{J}_m(\mathbf{r}')d\tau'}{|\mathbf{r}-\mathbf{r}'|}. \tag{3.15}$$

The two terms on the right–hand side give the contributions from the surface current density and volume current density, respectively.

The corresponding magnetic field can be calculated from: $\mathbf{B}(\mathbf{r}) = \boldsymbol{\nabla}\times\mathbf{A}$, where the derivatives are taken with respect to the coordinates $\mathbf{r}$ defined from O.

This calculation confirms that magnetic materials can be described either as materials having an induced magnetization $\mathbf{M}$ or as materials with a distribution of currents given by equations (3.11) and (3.12). The two representations are **equivalent**.

3.3.2 Field inside the material

This section may be skipped on first reading.

In the previous section, we have adopted an idealised representation of a magnetic material as a continous distribution of dipoles or currents. This is a good approximation when calculating the vector potential outside the material, as the distance to the molecules or atoms is then large compared to their separation, and inhomogeneities within the material average out.

[3] If $\mathbf{V}$ and $\mathbf{w}$ are two vectors, the divergence theorem for the vector $\mathbf{V}\times\mathbf{w}$ can be written:

$$\iiint_{\mathcal{V}} \boldsymbol{\nabla}\cdot(\mathbf{V}\times\mathbf{w})\, d\tau = \oiint_{\Sigma} (\mathbf{V}\times\mathbf{w})\cdot d\boldsymbol{\Sigma}.$$

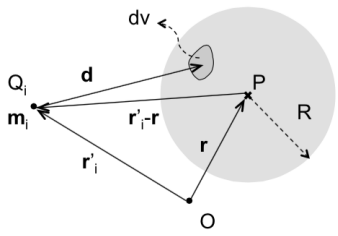
We now use $\boldsymbol{\nabla}\cdot(\mathbf{V}\times\mathbf{w}) = \mathbf{w}\cdot(\boldsymbol{\nabla}\times\mathbf{V}) - \mathbf{V}\cdot(\boldsymbol{\nabla}\times\mathbf{w})$. We chose $\mathbf{w}$ to be uniform, so that the second term on the right-hand side is zero. Furthermore, in the surface integral, $(\mathbf{V}\times\mathbf{w})\cdot d\boldsymbol{\Sigma} = \mathbf{w}\cdot(d\boldsymbol{\Sigma}\times\mathbf{V})$. The divergence theorem can therefore be written as:

$$\mathbf{w}\cdot\iiint_{\mathcal{V}} (\boldsymbol{\nabla}\times\mathbf{V})\, d\tau = -\mathbf{w}\cdot\oiint_{\Sigma} \mathbf{V}\times d\boldsymbol{\Sigma}.$$

As this is satisfied for any uniform vector $\mathbf{w}$, the integrals are equal and opposite.

However, the situation is *a priori* more complicated when studying the magnetic field inside the material, as it varies hugely over very short distances. In the same way as we defined a *macroscopic* electric field in a dielectric, we define here a macroscopic magnetic field (or, equivalently, vector potential) at a point P in the magnetic material by averaging the microscopic fields (or vector potentials) over a sphere of appropriate radius R around P. Using the principle of superposition, we can write the vector potential $\mathbf{A}$ as the sum of the average vector potentials over the sphere due to the dipole moments outside and inside the sphere: $\mathbf{A} = \mathbf{A}_{\text{out}} + \mathbf{A}_{\text{in}}$. In this section, we calculate $\mathbf{A}_{\text{out}}$ and $\mathbf{A}_{\text{in}}$ in a way very similar to the way we calculated V_{out} and V_{in} in a dielectric (section 2.3.2). Here again, the calculation is a bit lengthy but does not require complicated integrals to be computed. See Jackson, Section 5.6, for a more elegant but more technical calculation of the average field over a sphere.

Contribution from the dipole moments outside the sphere:



The vector potential produced in a small volume element dv in the sphere by a dipole moment $\mathbf{m}_i$ located at a point Q_i outside the sphere is $\mu_o \mathbf{m}_i \times \mathbf{d}/(4\pi d^3)$, where $\mathbf{d}$ is the vector position of dv measured from Q_i.

The average vector potential over the volume of the sphere due to the dipole moment $\mathbf{m}_i$ is therefore:

$$\langle \mathbf{A}_i \rangle = \frac{1}{\mathcal{V}} \iiint_{\mathcal{V}} \frac{\mu_o \mathbf{m}_i \times \mathbf{d}}{4\pi d^3} \, dv,$$

where $\mathcal{V}$ is the volume of the sphere. But this is also the vector potential that would be produced at Q_i by a uniform magnetization $-\mathbf{m}_i/\mathcal{V}$ inside the sphere (the minus sign in the magnetization comes from the orientation of $\mathbf{d}$). A sphere with a uniform magnetization $\mathbf{M}_s$ produces outside itself a vector potential which is the same as that produced by a dipole moment $\mathcal{V}\mathbf{M}_s$ located at the centre of the sphere[4]. Therefore:

$$\langle \mathbf{A}_i \rangle = \frac{\mu_o (-\mathbf{m}_i) \times (\mathbf{r}'_i - \mathbf{r})}{4\pi |\mathbf{r}'_i - \mathbf{r}|^3},$$

where $\mathbf{r}$ and $\mathbf{r}'_i$ are the position vectors measured from an origin O. The total vector potential $\mathbf{A}_{\text{out}}$ is obtained by summing over the dipole moments outside the sphere, so that:

$$\mathbf{A}_{\text{out}} = \sum_{i,\text{outside}} \langle \mathbf{A}_i \rangle = \frac{\mu_0}{4\pi} \sum_{i,\text{outside}} \frac{\mathbf{m}_i \times (\mathbf{r} - \mathbf{r}'_i)}{|\mathbf{r} - \mathbf{r}'_i|^3}.$$

[4]This result will be shown in Problem Set 1.

We approximate the distribution of dipole moments by a continuous distribution, so that $\mathbf{m}_i = \mathbf{M}(\mathbf{r}'_i)d\tau'$, with $\mathbf{M}(\mathbf{r}'_i)$ being the magnetization in a small volume element $d\tau'$ around Q_i. We can then write the vector potential as:

$$\mathbf{A}_{\text{out}} = \frac{\mu_0}{4\pi} \iiint_{\text{outside}} \frac{\mathbf{M}(\mathbf{r}')d\tau' \times (\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^3}, \tag{3.16}$$

where we have dropped the subscript i.

We can see from the above expression that **the average vector potential over the sphere due to the dipole moments outside is equal to the vector potential produced at the centre of the sphere**.

Contribution from the dipole moments inside the sphere:

We now assume that the point Q_i in the figure above is *inside* the sphere. The magnetic field produced in the small volume element dv in the sphere by the dipole moment $\mathbf{m}_i$ located at Q_i is:

$$\mathbf{B}_i = \frac{\mu_0}{4\pi d^3}\left[3\left(\mathbf{m}_i \cdot \hat{\mathbf{d}}\right) \cdot \hat{\mathbf{d}} - \mathbf{m}_i\right],$$

(see eq.[1.44]) where $\hat{\mathbf{d}} = \mathbf{d}/d$.
The average magnetic field over the volume of the sphere due to the dipole moment $\mathbf{m}_i$ is therefore:

$$\langle \mathbf{B}_i \rangle = \frac{1}{\mathcal{V}} \iiint_{\mathcal{V}} \frac{\mu_0}{4\pi d^3}\left[3\left(\mathbf{m}_i \cdot \hat{\mathbf{d}}\right) \cdot \hat{\mathbf{d}} - \mathbf{m}_i\right] dv,$$

where $\mathcal{V}$ is the volume of the sphere. But this is also the magnetic field that would be produced at Q_i by a uniform magnetization $\mathbf{m}_i/\mathcal{V}$ inside the sphere (note that the sign of the magnetization here is not reversed, as the expression of $\mathbf{B}_i$ does not depend on the orientation of $\hat{\mathbf{d}}$). A sphere with a uniform magnetization $\mathbf{M}_s$ produces inside itself a magnetic field[5] $2\mu_0\mathbf{M}_s/3$. Therefore, the average magnetic field $\langle \mathbf{B}_i \rangle$ over the sphere due to the dipole $\mathbf{m}_i$ at Q_i is:

$$\langle \mathbf{B}_i \rangle = \frac{2}{3}\mu_0 \frac{\mathbf{m}_i}{\mathcal{V}}.$$

The total magnetic field $\mathbf{B}_{\text{in}}$ is obtained by summing over the dipole moments inside the sphere, so that:

$$\mathbf{B}_{\text{in}} = \sum_{i,\text{inside}} \langle \mathbf{B}_i \rangle = \sum_{i,\text{inside}} \frac{2}{3}\mu_0 \frac{\mathbf{m}_i}{\mathcal{V}}.$$

Here again we approximate the distribution of dipole moments by a continuous distribution, so that $\mathbf{m}_i = \mathbf{M}(\mathbf{r}'_i)d\tau'$. The average field can then be written as:

$$\mathbf{B}_{\text{in}} = \frac{2}{3}\mu_0 \iiint_{\text{inside}} \frac{\mathbf{M}(\mathbf{r}')d\tau'}{\mathcal{V}},$$

where we have dropped the subscript i. We assume that the magnetization in the sphere is uniform, as the sphere is very small, so that the integral is equal to $\mathbf{M}$.

[5] This result will be shown in Problem Set 1.

Therefore, **the average field over the sphere due to the dipole moments inside is the same as the (uniform) field that would be produced in the sphere by assuming it to be uniformly magnetized**.

The vector potential produced by the uniformly magnetized sphere at a point with vector position $\mathbf{r}$ inside the sphere is:

$$\mathbf{A}_{\text{in}}(\mathbf{r}) = \frac{\mu_0}{4\pi} \iiint_{\text{inside}} \frac{\mathbf{M} d\tau' \times (\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^3},$$

which is the same as equation (3.16) with the integral being over the sphere and $\mathbf{M}$ assumed to be uniform.

Total average field:

From the calculations above, we see that the macroscopic magnetic field at P, which we define as being the average field over the sphere of radius R centered on P, derives from a vector potential $\mathbf{A}_{\text{out}} + \mathbf{A}_{\text{in}}$ which can be written in exactly the same form as that given by equation (3.13).

Therefore, **the magnetic field can always be calculated by summing over the contribution from the dipole moments (or equivalently the currents) which are in the magnetic material**. This is true whether the field is calculated outside or inside the material.

3.4 The auxiliary field H

As we have seen above, an external magnetic field present in a material induces a magnetization of the material, which in turn produces a magnetic field. The total magnetic field, which is the superposition of these two fields, then modifies the magnetization, which in turn affects the field, etc. In general, we are only interested in the total field, which is produced by all the currents present.

3.4.1 Ampère's law

The total current density is $\mathbf{J} = \mathbf{J}_m + \mathbf{J}_f$, where $\mathbf{J}_m$ are the **bound** currents, due to the magnetization of the material, and $\mathbf{J}_f$ are the **free** currents (due to the motion of charges in conductors). We assume here that the fields are **static** so that there is no polarization or displacement currents. Therefore, both bound and free currents generate a magnetic field which satisfies Ampère's law in the form:

$$\boldsymbol{\nabla} \times \mathbf{B} = \mu_0 \mathbf{J} = \mu_0 \left(\mathbf{J}_m + \mathbf{J}_f \right). \tag{3.17}$$

Using equation (3.12), this is equivalent to:

$$\boldsymbol{\nabla} \times \left(\frac{\mathbf{B}}{\mu_0} - \mathbf{M} \right) = \mathbf{J}_f.$$

We define the vector $\mathbf{H}$ by:

$$\boxed{\mathbf{H} = \frac{\mathbf{B}}{\mu_0} - \mathbf{M},} \tag{3.18}$$

so that Ampère's law can be written in terms of the free current only:

$$\boxed{\boldsymbol{\nabla} \times \mathbf{H} = \mathbf{J}_f.} \tag{3.19}$$

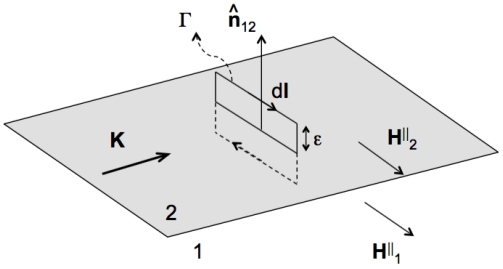
The integral form of this equation is:

$$\oint_\Gamma \mathbf{H} \cdot d\mathbf{l} = \iint_\Sigma \mathbf{J}_f \cdot d\boldsymbol{\Sigma} = I_f, \tag{3.20}$$

where I_f is the current flowing through the surface Σ of the loop delimited by the contour Γ.

It may seem from equation (3.19) that $\mathbf{H}$ depends only on the distribution of free currents. It is incorrect, as $\boldsymbol{\nabla} \cdot \mathbf{H} = -\boldsymbol{\nabla} \cdot \mathbf{M}$. In particular, the absence of free current ($\mathbf{J}_f = \mathbf{0}$) does not necessarily imply that $\mathbf{H} = \mathbf{0}$.

3.4.2 Boundary conditions on B and H



We consider a (locally flat) surface between two media with a current density $\mathbf{K} = \mathbf{K}_m + \mathbf{K}_f$. We apply the integral form of Ampère's law given by equation (3.20), where Γ is the contour represented on the figure.

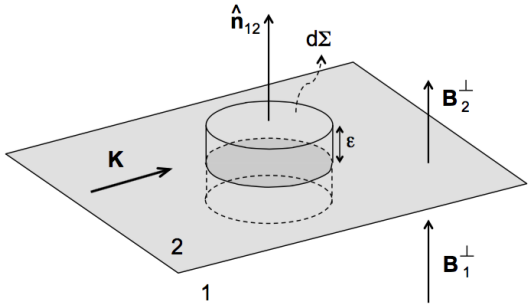
The horizontal sides of the contour are parallel to the tangential component of the field $\mathbf{H}^{\parallel}$, and therefore the surface delimited by the contour is perpendicular to $\mathbf{K}$. We orientate the contour in such a way that the current through the surface is positive. Then, taking $\epsilon \to 0$, we obtain:

$$H_2^{\parallel} dl - H_1^{\parallel} dl = K_f dl.$$

The above relation can also be written in the following way:

$$\boxed{\mathbf{H}_2^{\parallel} - \mathbf{H}_1^{\parallel} = \mathbf{K}_f \times \hat{\mathbf{n}}_{12},} \tag{3.21}$$

where $\hat{\mathbf{n}}_{12}$ is the unit vector perpendicular to the surface and pointing from medium (1) to medium (2). If there is **no free current** on the surface, $\mathbf{H}^{\parallel}$ is **continuous**.



We now apply $\oiint_\Sigma \mathbf{B} \cdot d\mathbf{\Sigma} = 0$ (which is implied by $\boldsymbol{\nabla} \cdot \mathbf{B} = 0$), where Σ is the surface of the cylinder represented on the figure. We note $\mathbf{B}^\perp = B^\perp \hat{\mathbf{n}}_{12}$, with $B^\perp = \mathbf{B} \cdot \hat{\mathbf{n}}_{12}$, the component of $\mathbf{B}$ perpendicular to the surface and very close to it.

We take the limit $\epsilon \to 0$ so that the flux of $\mathbf{B}$ through the lateral surface of the cylinder is negligible. Therefore, we obtain:

$$\boxed{\mathbf{B}_2^\perp - \mathbf{B}_1^\perp = \mathbf{0},} \tag{3.22}$$

which means that $\mathbf{B}^\perp$ is **continuous**.

3.5 Linear magnetic materials

We have seen that in diamagnetic and paramagnetic materials, an external magnetic field induces a magnetization which is (anti–)parallel to $\mathbf{B}$. Under most conditions, provided the field is not too strong, the magnetization is proportional to the total magnetic field $\mathbf{B}$, and we define the **magnetic susceptibility** χ_m through the relation:

$$\boxed{\mathbf{M} = \chi_m \mathbf{H}.} \tag{3.23}$$

χ_m is a dimensionless constant which is positive for paramagnetic materials and negative for diamagnetic materials. Note that this definition of χ_m is not analogous to the definition of the electric susceptibility. Indeed, to parallel the definition of $\chi_e = P/(\epsilon_0 E)$, the magnetic susceptibility should have been defined as $M/(B/\mu_0)$.
The magnetic materials in which this relation is satisfied are called **linear**, **homogeneous** and **isotropic**.

We have pointed out in section 3.1.2 that the total dipole moment per unit mass in a diamagnetic material is roughly the same in all materials, for a given applied field. The magnetic susceptibility, however, is defined as the ratio of the total dipole moment *per unit volume* to H. Therefore, it is not exactly the same for all materials, although it does not vary significantly. For purely diamagnetic materials, solid or liquid, and under normal conditions of pressure and temperature, the magnetic susceptibility is typically on the order of -10^{-5}.

In section 3.1.3, we have pointed out that the total dipole moment due to the spin of the electrons and induced by the external field decreases with increasing temperature, as thermal motions oppose the alignement of the spin dipole moment with the field. Therefore, the contribution to the magnetic susceptibility of paramagnetism decreases with temperature. The magnetic susceptibility of paramagnetic materials, which includes a positive contribution from paramagnetism and a negative contribution from diamagnetism, is on

the order of a few times 10^{-5} under normal conditions, but may be considerably larger at lower temperatures.

The magnetic susceptibility is smaller for gases than for solids or liquids as they have a smaller density of electrons.

Note that since χ_m is usually small compared to unity, $M \ll H$, which means that the magnetization does not contribute significantly to the total magnetic field. In other words, **the magnetic field produced by the magnetization of diamagnetic or paramagnetic materials is usually small compared to the external field** that produces the magnetization.

Substituting equation (3.18) into equation (3.23), we obtain:

$$\mathbf{B} = \mu_0 (1 + \chi_m) \mathbf{H},$$

which is also written:

$$\boxed{\mathbf{B} = \mu \mathbf{H},} \tag{3.24}$$

with:

$$\boxed{\mu \equiv \mu_0 (1 + \chi_m).} \tag{3.25}$$

The constant μ is the **permeability** of the material. In vacuum, $\mathbf{M}$ is zero, which implies $\chi_m = 0$, so that $\mu = \mu_0$. Therefore μ_0 is called the **permeability of free space**. We further define the dimensionless **relative permeability** by:

$$\boxed{\mu_r \equiv \frac{\mu}{\mu_0} = 1 + \chi_m.} \tag{3.26}$$

Note that the relation between $\mathbf{H}$ and $\mathbf{B}$ is a *constitutive relation* and cannot be derived from first principles.

We finally point out that, in a homogeneous linear material, the volume density of magnetization current can be expressed in terms of the volume density of free current in a simple way:

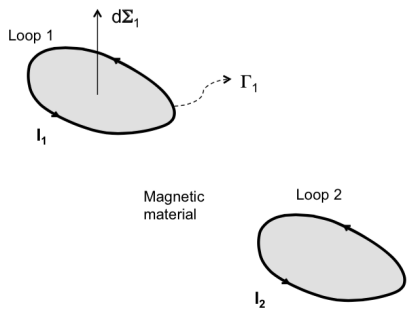
$$\mathbf{J}_m = \boldsymbol{\nabla} \times \mathbf{M} = \boldsymbol{\nabla} \times (\chi_m \mathbf{H}) = \chi_m \boldsymbol{\nabla} \times \mathbf{H} = \chi_m \mathbf{J}_f. \tag{3.27}$$

3.6 Energy in the presence of magnetic materials

In section 2.6, we have defined the energy in the presence of a dielectric as the work needed to build up the polarization from zero to its final value, by bringing the *free* charges into the system, one by one. Similarly, we define the energy in the presence of magnetic materials as the work needed to build up the magnetization (or, equivalently, the magnetic field) from zero to its final value, by establishing the *free* currents into the system. The magnetic field in which each current is established is due to the other free or bound currents. We consider here the case of **static** fields.

It is possible to model any distribution of free currents by a superposition of current loops (see Jackson, section 5.16). Therefore, we suppose that we have N loops (of which

we have represented only two on the figure), and we establish the current from 0 to its final value in each of them.



Let us suppose we increase the current slightly in one of the loops. This produces a change $\delta\mathbf{B}$ in the total magnetic field (due to both free and bound currents), which in turn produces a variation $\delta\Phi_k$ of the magnetic flux through the surface delimited by the loop k. The work that has to be done by the batteries supplying the current to counter the emf induced by $\delta\Phi_k$ is $\delta W_k = I_k \delta\Phi_k$, where I_k is the (free) current through loop k.

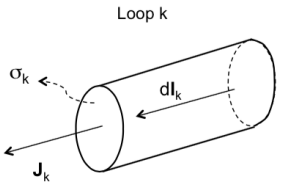
The variation of the flux is given by:

$$\delta\Phi_k = \oiint_{\Sigma_k} \delta\mathbf{B} \cdot d\mathbf{\Sigma}_k,$$

where Σ_k is the surface delimited by the loop k. Using $\mathbf{B} = \boldsymbol{\nabla}\times\mathbf{A}$, where $\mathbf{A}$ is the potential vector, we obtain:

$$\delta W_k = I_k \oiint_{\Sigma_k} (\boldsymbol{\nabla}\times\delta\mathbf{A}) \cdot d\mathbf{\Sigma}_k = I_k \oint_{\Gamma_k} \delta\mathbf{A} \cdot d\mathbf{l}_k,$$

where Γ_k is the contour of the loop k and we have used Stokes's theorem to get the last integral on the right–hand–side.



If we note σ_k the cross–sectional area of the wire along loop k, then $I_k = J_k\sigma_k$, where J_k is the density of (free) current in loop k. Therefore, $I_k d\mathbf{l}_k$ can be written as $\mathbf{J}_k d\tau$, where $d\tau$ is a volume element in the wire.

If we note V_k the volume of the wire along loop k , we can then write:

$$\delta W_k = \iiint_{V_k} \delta\mathbf{A} \cdot \mathbf{J_k} d\tau.$$

The work that has to be done to counter the emf induced in all the loops by the change δB is obtained by summing the above expression over k:

$$\delta W = \sum_k \delta W_k = \iiint_V \delta\mathbf{A} \cdot \mathbf{J}_f d\tau,$$

where the integral is over the volume V occupied by all the free currents $\mathbf{J}_f$. We can extend this integral over the volume $\mathcal{V}$ of the whole space, as only the regions where J_f is non zero will contribute anyway. We now use equation (3.19) to obtain:

$$\delta W = \iiint_{\mathcal{V}} \delta\mathbf{A} \cdot (\boldsymbol{\nabla}\times\mathbf{H})\, d\tau.$$

As $\boldsymbol{\nabla} \cdot (\mathbf{H} \times \delta\mathbf{A}) = \delta\mathbf{A} \cdot (\boldsymbol{\nabla} \times \mathbf{H}) - \mathbf{H} \cdot (\boldsymbol{\nabla} \times \delta\mathbf{A})$, the integral above can be transformed into:

$$\delta W = \iiint_{\mathcal{V}} \boldsymbol{\nabla} \cdot (\mathbf{H} \times \delta\mathbf{A})\, d\tau + \iiint_{\mathcal{V}} \mathbf{H} \cdot (\boldsymbol{\nabla} \times \delta\mathbf{A})\, d\tau.$$

By using the divergence theorem, the first integral on the right–hand–side can be written as the flux over the surface delimiting $\mathcal{V}$ of $\mathbf{H} \times \delta\mathbf{A}$. We can put this surface at infinity. At large distance, the loops are seen as dipoles, so that $H \sim r^{-3}$ and $A \sim r^{-2}$. This integral therefore vanishes. Since in addition $\boldsymbol{\nabla} \times \delta\mathbf{A} = \delta\mathbf{B}$, δW becomes:

$$\boxed{\delta W = \iiint_{\mathcal{V}} \mathbf{H} \cdot \delta\mathbf{B} d\tau.} \tag{3.28}$$

This equation is the equivalent of equation (2.27), which gives the energy in the presence of dielectrics, and is valid for all magnetic materials, linear or not.
The work that has to be done to establish the magnetic field from $\mathbf{0}$ to $\mathbf{B}$ (by establishing all the free currents in all the loops) is:

$$W = \int_0^B \delta W = \iiint_{\mathcal{V}} d\tau \int_0^B \mathbf{H} \cdot \delta\mathbf{B}.$$

When the field is decreased back to zero, if the curve $B(H)$ is the same as when B was increased, then:

$$\iiint_{\mathcal{V}} d\tau \int_0^B \mathbf{H} \cdot \delta\mathbf{B} = - \iiint_{\mathcal{V}} d\tau \int_B^0 \mathbf{H} \cdot \delta\mathbf{B}.$$

Therefore, the total work done to increase B from zero to its final value and decrease it back to zero is zero. The work done to establish the field can be viewed as being stored in the field reversibly: the energy is given back when the field is returned to zero.

In paramagnetic or diamagnetic materials, the relation between $\mathbf{H}$ and $\mathbf{B}$ is linear, so that:

$$\mathbf{H} \cdot \delta\mathbf{B} = \frac{1}{2} \delta \left(\mathbf{H} \cdot \mathbf{B} \right),$$

and:

$$W = \frac{1}{2} \iiint_{\mathcal{V}} \mathbf{H} \cdot \mathbf{B} d\tau.$$

This energy is stored in the magnetic field. From the expression of W, we obtain the **energy density for a linear magnetic material**:

$$\boxed{U = \frac{1}{2} \mathbf{H} \cdot \mathbf{B}.} \tag{3.29}$$

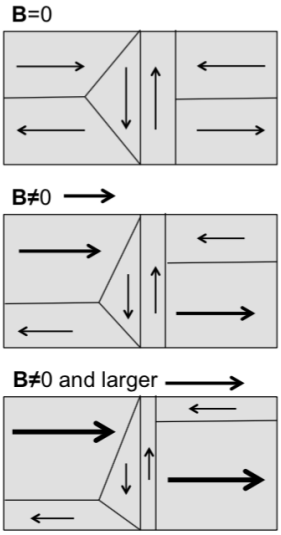
In vacuum, $\mathbf{H} = \mathbf{B}/\mu_0$, so that we recover the expression of the energy density $U = B^2/(2\mu_0)$ seen in the first year course.

3.7 Ferromagnetism

3.7.1 Microscopic description

In diamagnetic and paramagnetic materials, a magnetization is present only when an external magnetic field is present. When the field is removed, the magnetization vanishes. Ferromagnets are materials which can sustain a magnetization even after the external field has been removed. These materials are **non–linear**, since the magnetization is not simply proportional to **H**.

The magnetization in ferromagnets is due to paramagnetism, that is to say to the alignement of the spin dipole moments of unpaired electrons. But due to the structure of ferromagnetic materials (which are all solids), the energy is decreased when neighbouring atomic dipoles align with each other. The force responsible for this alignement is purely quantum mechanical (and is not the magnetic interaction between dipoles). In general, the dipoles have random orientations, and if one dipole wants to align with a neighbour, it will be pulled in another direction by another neighbour. However, if, for any reason, two neighbouring dipoles happen to be aligned with each other, they define a special direction in which other close dipoles are going to align. In a **domain** around these two dipoles, all the dipoles will then point in the same direction and there will be a net magnetization. Ferromagnetic materials are made of such domains, which size is typically between 10^{-3} and 1 mm^3.



If the material is not placed in a magnetic field, there is no net magnetization as different domains have a magnetization pointing in different directions. However, when a magnetic field is applied, the domains in which the magnetization is parallel or close to parallel with the field grow at the expense of the others. This is because at the boundary between two domains, misaligned dipoles compete with each other to impose a direction. When a field is applied, it exerts a torque which tends to align the dipoles with its own direction, and therefore enables the dipoles which are already in this direction to 'win' the competition. The boundaries between domains are then moved.

If the applied field is strong enough, only one single domain remains, and the material becomes **saturated** (the magnetization cannot be made larger for a given temperature).

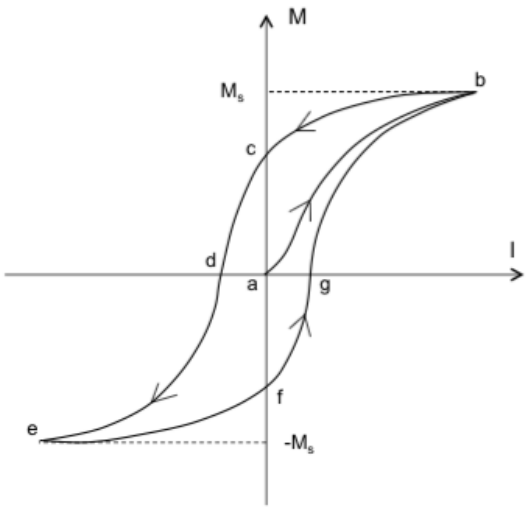
When the external field is removed, as it is energetically more favourable for the dipoles to be aligned with each other, a net magnetization in the direction of the field that was applied remains: the material has become a **permanent magnet**.

However, random thermal motions compete with the tendency of the dipoles to align with each other. Therefore, at high enough temperatures, the alignement is destroyed

and there is no remaining magnetization after the field is removed. The material then becomes paramagnetic. The transition between ferromagnetism and paramagnetism does not happen progressively as the temperature is increased. Instead, it happens at a precise temperature called the **Curie temperature**. This is an example of a **phase transition**.

3.7.2 Hysteresis loop

To illustrate how a ferromagnetic material responds to an external field, we vary the field and measure the magnetization. In practice, the material is inserted inside a solenoid in which we run a current I. By varying the current, we vary the magnetic field inside the solenoid.



We start with a ferromagnetic material that has initially no magnetization (point a on the graph). When the current is increased, the magnetization increases, slowly at first and then more rapidly as the boundaries of the domains move. At point b, the magnetization reaches the saturation value M_s. Increasing the current further does not change the magnetization. If the current is reduced from point b, the magnetization decreases slightly. When the current vanishes, there is a residual magnetization left in the material which has then become a permanent magnet (point c).

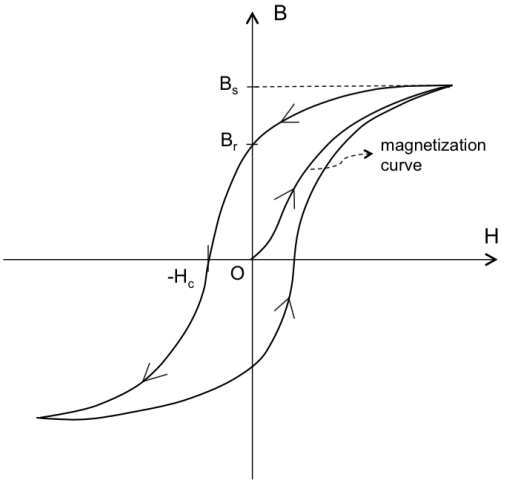
If the current is reversed, so that the external field reverses direction, the magnetization decreases until it vanishes (point d). Increasing the amplitude of the current further leads to saturation again (point e), but with a magnetization opposite to that at point b. If the amplitude of the current is then decreased to zero, a permanent magnet is obtained (point f). Once again, demagnetization is obtained by increasing the current in the other direction point g).

We see that, when the current is decreased, the magnetization does not come back to the same values as it took when the current was increased. The magnetization, therefore, does not depend only on I, but also on the magnetic history of the material. This phenomenon is called **hysteresis**, from a Greek word meaning *lagging behind*, and is a source of thermodynamic irreversibility.

The curve shown on the graph is called a **hysteresis loop**. As pointed out above, the value of M_s decreases when the temperature increases.

In general, the hysteresis loop is plotted in the H–B plane, rather than in the I–M plane shown above. If the external field is produced by a long solenoid, then $H = nI$, where n is the number of turns per unit length, so that H is entirely controlled by the experimentator and is not influenced by the ferromagnet. The total magnetic field is given

by $\mathbf{B} = \mu_0 (\mathbf{H} + \mathbf{M})$. The magnetization produced by ferromagnets is very large compared to H or, in other words, **the magnetic field produced by the magnetization in the ferromagnet is very large compared to the external magnetic field** produced by the solenoid. Therefore, $\mathbf{B} \simeq \mu_0 \mathbf{M}$, and the hysteresis loop in the H–B plane has the same shape as that shown above. Typically, ferromagnets at a temperature of 20° C have a saturated magnetization of $M_s \sim 10^6$ A m^{-1} which is reached for $H \sim 10^3$ A m^{-1}. These values correspond to $\mu_0 H \sim 10^{-3}$ T and $B \simeq \mu_0 M \sim 1$ T.



The magnetic field B_r that remains in the material when the external field is removed is called the **remanent** field (similarly, we define the remanent magnetization). For the total magnetic field to vanish, an external field $-H_c$ has to be applied, with H_c being called the **coercive force**.

3.7.3 Hysteresis loss

In section 3.6, we have seen that the energy necessary to establish a field B is regained when the field is decreased back to zero if the curve $B(H)$ is the same when the field is increased and when it is decreased. However, for ferromagnets, as the values taken by B when H is increased are not the same as those taken when H is decreased, the energy is not completely recovered: a fraction of the **energy is dissipated as heat in the ferromagnet**. The losses are partially due to the friction which occurs at the boundaries of the domains when the dipole moments reorient themselves under the action of the varying field.

From equation (3.28), the work done per unit volume to move from one point to another on the hysteresis curve is:

$$W = \int H dB,$$

where the integral is along the curve and we have used the fact that, in a ferromagnet, **H** and **B** are parallel (the magnetization is not proportional to the external field, but is nonetheless aligned with it). Therefore, the work done (by the batteries which generate the current) to take a unit volume of the ferromagnet around the hysteresis curve is:

$$W = \oint H dB,$$

where the integral is along the closed curve. The value of this integral is equal to the

surface area enclosed by the hysteresis curve. If the ferromagnet is taken around the hysteresis loop repeatedly by an alternating current of frequency f, the hysteresis loop is traversed f times per second, so that the hysteresis losses are linearly proportional to the frequency.

To minimize the losses, we have to use a **soft ferromagnet** which has a narrow hysteresis curve, that is to say a small coercive force and a small remanent magnetization. Ferromagnets with large hysteresis curves are called **hard**.

Chapter 4

Electromagnetic waves in vacuum

4.1 Maxwell's equations and boundary conditions

4.1.1 Local form of Maxwell's equations

As seen in the fist year course, for fields in the presence of electric charge of density ρ and electric current of density $\mathbf{J}$, Maxwell's equations are:

$$\boxed{\begin{aligned}
\boldsymbol{\nabla}\cdot\mathbf{E} &= \frac{\rho}{\epsilon_0}, && (4.1)\\
\boldsymbol{\nabla}\cdot\mathbf{B} &= 0, && (4.2)\\
\boldsymbol{\nabla}\times\mathbf{E} &= -\frac{\partial\mathbf{B}}{\partial t}, && (4.3)\\
\boldsymbol{\nabla}\times\mathbf{B} &= \mu_0\mathbf{J} + \mu_0\epsilon_0\frac{\partial\mathbf{E}}{\partial t}. && (4.4)
\end{aligned}}$$

Equations (4.1), (4.3) and (4.4) are, respectively, Gauss's law, Faraday's law, and Ampère's law with the inclusion of the displacement current (also known as Ampère–Maxwell's law).

4.1.2 Integral form of Maxwell's equations

Using the divergence and Stokes's theorems, the integral form of these equations can be written as:

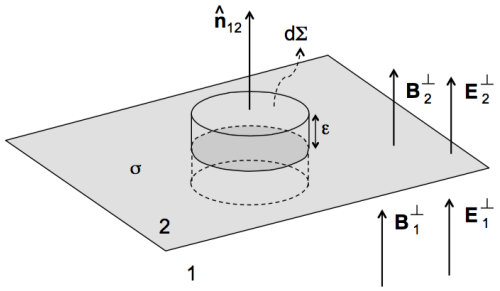
$$\begin{aligned}
\oiint_{\Sigma}\mathbf{E}\cdot d\boldsymbol{\Sigma} &= \iiint_{\mathcal{V}}\frac{\rho}{\epsilon_0}d\tau, && (4.5)\\
\oiint_{\Sigma}\mathbf{B}\cdot d\boldsymbol{\Sigma} &= 0, && (4.6)\\
\oint_{\Gamma}\mathbf{E}\cdot d\mathbf{l} &= -\frac{d}{dt}\iint_{\Sigma}\mathbf{B}\cdot d\boldsymbol{\Sigma}, && (4.7)\\
\oint_{\Gamma}\mathbf{B}\cdot d\mathbf{l} &= \mu_0\iint_{\Sigma}\mathbf{J}\cdot d\boldsymbol{\Sigma} + \mu_0\epsilon_0\frac{d}{dt}\iint_{\Sigma}\mathbf{E}\cdot d\boldsymbol{\Sigma}. && (4.8)
\end{aligned}$$

In equation (4.5), Σ is the (closed) surface that delimits the volume $\mathcal{V}$ whereas, in equations (4.7) and (4.8), Σ is the (open) surface delimited by the contour Γ.

4.1.3 Boundary conditions

In general, the fields $\mathbf{E}$ and $\mathbf{B}$ are discontinuous at a surface that carries a charge density and/or a current density. We have previously established the boundary conditions in dielectrics and in magnetic materials when the fields are static (sections 2.4.3 and 3.4.2). Here, we establish the boundary conditions on $\mathbf{E}$ and $\mathbf{B}$ at a surface with *free* charges and/or *free* currents and in the *time–dependent* case.

We consider a surface with a free charge density σ and a free current density $\mathbf{K}$ separating two empty regions. Locally, the surface is flat.



We apply the integral form of Gauss's law given by equation (4.5) with the surface Σ being that of the cylinder represented on the figure and ρ the charge density in the cylinder. Taking $\epsilon \to 0$, this implies:

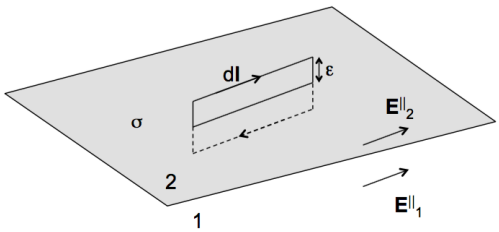
$$\boxed{\mathbf{E}_2^{\perp} - \mathbf{E}_1^{\perp} = \frac{\sigma}{\epsilon_0}\,\hat{\mathbf{n}}_{12},} \tag{4.9}$$

as shown in section 2.4.3 (see eq. [2.19]).

Similarly, equation (4.6) implies:

$$\boxed{\mathbf{B}_2^{\perp} - \mathbf{B}_1^{\perp} = \mathbf{0},} \tag{4.10}$$

as shown in section 3.4.2 (see eq. [3.22]).



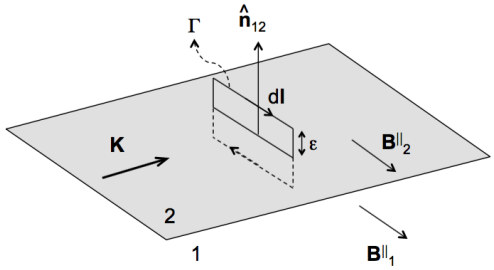
We now apply the integral form of Faraday's law given by equation (4.7) with the contour Γ being that of the loop represented on the figure (the horizontal sides of the loop are parallel to the tangential component of the fiel $\mathbf{E}^{\parallel}$). Then:

$$E_2^{\parallel} dl - E_1^{\parallel} dl = -\frac{d}{dt} \iint_{\Sigma} \mathbf{B} \cdot d\mathbf{\Sigma}.$$

As $\epsilon \to 0$, the vertical sides of the loop do not contribute to the integral over Γ. For the same reason, the flux of $\mathbf{B}$ through the surface vanishes. Therefore, we obtain:

$$\boxed{\mathbf{E}_2^{\parallel} - \mathbf{E}_1^{\parallel} = \mathbf{0},} \tag{4.11}$$

as in the static case (see eq. [2.21] in section 2.4.3).



Finally, we apply the integral form of Ampère–Maxwell's law given by equation (4.8) with the contour Γ being that of the loop represented on the figure. The horizontal sides of the contour are parallel to the tangential component of the field $\mathbf{B}^{\parallel}$, and therefore the surface delimited by the contour is perpendicular to $\mathbf{K}$.

We then obtain:

$$B_2^{\parallel} dl - B_1^{\parallel} dl = \mu_0 K dl + \mu_0 \epsilon_0 \frac{d}{dt} \iint_{\Sigma} \mathbf{E} \cdot d\mathbf{\Sigma}.$$

As above, the vertical sides of the loop do not contribute to the integral over Γ because $\epsilon \to 0$, and for the same reason the flux of $\mathbf{E}$ through the surface vanishes. The above equation can then be written as:

$$\boxed{\mathbf{B}_2^{\parallel} - \mathbf{B}_1^{\parallel} = \mu_0 \mathbf{K} \times \hat{\mathbf{n}}_{12},} \tag{4.12}$$

where $\hat{\mathbf{n}}_{12}$ is the unit vector perpendicular to the surface and oriented from region (1) to region (2). This is similar to the equation obtained for $\mathbf{H}$ in the static case (see eq. [3.21] in section 3.4.2).

Equations (4.9), (4.10), (4.11) and (4.12) can be recast in the following compact form:

$$\boxed{\begin{aligned} \mathbf{E}_2 - \mathbf{E}_1 &= \frac{\sigma}{\epsilon_0} \hat{\mathbf{n}}_{12} && (4.13) \\ \mathbf{B}_2 - \mathbf{B}_1 &= \mu_0 \mathbf{K} \times \hat{\mathbf{n}}_{12}. && (4.14) \end{aligned}}$$

Note that these equations would still be valid if region (1) were a conductor. In that case, $\mathbf{E}_1 = \mathbf{0}$ and it follows from equation (4.13) that **the electric field just outside a conductor is perpendicular to the surface and equal to** σ/ϵ_0.

4.2 Electromagnetic waves in vacuum

In vacuum (that is to say, away from the charges and currents that create the field) , $\rho = 0$ and $\mathbf{J} = \mathbf{0}$, so that Maxwell's equations become:

$$\boxed{\begin{aligned} \boldsymbol{\nabla} \cdot \mathbf{E} &= 0, && (4.15) \\ \boldsymbol{\nabla} \cdot \mathbf{B} &= 0, && (4.16) \\ \boldsymbol{\nabla} \times \mathbf{E} &= -\frac{\partial \mathbf{B}}{\partial t}, && (4.17) \\ \boldsymbol{\nabla} \times \mathbf{B} &= \mu_0 \epsilon_0 \frac{\partial \mathbf{E}}{\partial t}. && (4.18) \end{aligned}}$$

Taking the curl of (4.17) and (4.18) and using the identity:

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

which also applies to $\mathbf{B}$, we obtain the following **wave equations** for the fields $\mathbf{E}$ and $\mathbf{B}$:

$$\boxed{\nabla^2\mathbf{E} = \mu_0\epsilon_0\frac{\partial^2\mathbf{E}}{\partial t^2},} \tag{4.19}$$

$$\boxed{\nabla^2\mathbf{B} = \mu_0\epsilon_0\frac{\partial^2\mathbf{B}}{\partial t^2}.} \tag{4.20}$$

The above equations indicate that the three components of $\mathbf{E}$ and $\mathbf{B}$ may propagate in vacuum with the speed:

$$\boxed{c = \frac{1}{\sqrt{\mu_0\epsilon_0}},} \tag{4.21}$$

which is the speed of light.[1]

Note that each *cartesian* component of $\mathbf{E}$ and $\mathbf{B}$ satisfies the wave equation:

$$\nabla^2 F = \mu_0\epsilon_0\frac{\partial^2 F}{\partial t^2}. \tag{4.22}$$

However, this is not true for the cylindrical and spherical components[2] of $\mathbf{E}$ and $\mathbf{B}$.

The wave equation is a second–order linear partial differential equation. To solve it, it is necessary to specify initial and boundary conditions. For instance, to describe the vibrations of a string, which also satisfy a wave equation, we usually assume that the ends remain fixed (boundary conditions) and we specify the initial position and velocity along the string (initial conditions). The wave equation has a large number of solutions, for instance traveling plane waves, traveling spherical waves, standing waves, etc.

4.2.1 Monochromatic plane waves

The simplest and most fundamental electromagnetic waves are **plane waves**, for which F is uniform over every plane perpendicular to the direction of propagation. We assume that the wave propagates in the x–direction, so that F depends only on x and t. Equation (4.22) can therefore be written as:

$$\frac{\partial^2 F}{\partial x^2} - \frac{1}{c^2}\frac{\partial^2 F}{\partial t^2} = 0. \tag{4.23}$$

[1] Before Maxwell's theory of electromagnetism, ϵ_0 and μ_0 were two independant constants that were used to express $\mathbf{E}$ and $\mathbf{B}$ separately. When Maxwell discovered electromagnetic waves and showed that they propagate with speed $1/\sqrt{\mu_0\epsilon_0} = 3\times10^8$ m s^{-1}, it became apparent that light is an electromagnetic wave, and therefore $c = 1/\sqrt{\mu_0\epsilon_0}$. See Purcell, chapter 9.

[2] In cartesian coordinates, $\nabla^2\mathbf{E} = \nabla^2(E_x\hat{\mathbf{x}} + E_y\hat{\mathbf{y}} + E_z\hat{\mathbf{z}}) = \nabla^2(E_x)\hat{\mathbf{x}} + \nabla^2(E_y)\hat{\mathbf{y}} + \nabla^2(E_z)\hat{\mathbf{z}}$, where $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$ and $\hat{\mathbf{z}}$ are unit vectors. However, in cylindrical coordinates for example, $\nabla^2\mathbf{E} = \nabla^2(E_r\hat{\mathbf{r}} + E_\theta\hat{\boldsymbol{\theta}} + E_z\hat{\mathbf{z}})$, and as the unit vectors $\hat{\mathbf{r}}$ and $\hat{\boldsymbol{\theta}}$ depend on the coordinates, this is **not** equal to $\nabla^2(E_r)\hat{\mathbf{r}} + \nabla^2(E_\theta)\hat{\boldsymbol{\theta}} + \nabla^2(E_z)\hat{\mathbf{z}}$.

The most general solution to this equation is:

$$F(x,t) = f(x-ct) + g(x+ct), \tag{4.24}$$

where f and g are arbitrary functions describing plane waves traveling to the right (f) and to the left (g) with velocity c. Note that, because the wave equation is **linear**, the sum of two solutions is itself a solution.

Sinusoidal waves are a particular case of plane waves. They have the form:

$$F_k(x,t) = A\cos\left[k(x-ct)+\varphi\right] + B\cos\left[k(x+ct)+\psi\right], \tag{4.25}$$

where A and B are the amplitudes of the waves traveling to the right and to the left, respectively, φ and ψ are phase constants, and

$$k = \frac{2\pi}{\lambda}, \tag{4.26}$$

is the **wave number**, with λ being the **wavelength**. The phase of the wave traveling to the right (that is to say, the argument of the first cosine) is $k(x_0 - ct_0) + \varphi$ at position x_0 and time t_0. At time $t_0 + \Delta t$, the phase has the same value at position $x_0 + \Delta x$ such that $\Delta x - c\Delta t = 0$. Therefore, the phase moves to the right with velocity $\Delta x/\Delta t = c$. Similarly, the phase of the wave traveling to the left moves in that direction with the same speed. Therefore, c is called the **phase velocity** of the wave.
The **period** T of the wave is such that the phase of the wave moves forward or backward by 2π during T. Therefore:

$$T = \frac{2\pi}{kc}. \tag{4.27}$$

The **frequency** ν is the number of oscillations per unit time, and is given by:

$$\nu = \frac{1}{T} = \frac{kc}{2\pi} = \frac{c}{\lambda}. \tag{4.28}$$

We also define the **angular frequency** ω of the wave (often referred to as just 'frequency'), which is the number of radians the phase of the wave advances per unit time:

$$\omega = 2\pi\nu = kc. \tag{4.29}$$

Because the sinusoidal waves in equation (4.25) correspond to a single frequency, they are called **monochromatic**.

We note that F_k can be written as the real part of:

$$\tilde{F}_k = \tilde{A}e^{i(kx-\omega t)} + \tilde{B}e^{i(kx+\omega t)}, \tag{4.30}$$

with $\tilde{A} = Ae^{i\varphi}$ and $\tilde{B} = Be^{i\psi}$. Both the real and imaginary parts of $\tilde{F}_k$ are solutions of the wave equation (4.23), and therefore $\tilde{F}_k$ itself is a solution. As it is much easier to handle complex wave functions, which involve exponentials, rather than real wave functions,

which involve trigonometric functions, we will from now on use complex notations to write *linear* quantities. To avoid confusion, we will denote complex quantities with a *tilde*. **The physical quantity is the real part of the complex function** and is denoted by the same letter but without the *tilde*. When dealing with *non–linear* quantities, like energy density or Poynting vector, real numbers should be used (see below).

Using the Fourier integral theorem[3], we can write any complex wave function $\tilde{F}$ solution of equation (4.23) as a linear superposition of sinusoidal waves $\tilde{F}_k$:

$$\tilde{F}(x,t) = \int_0^{+\infty} \left[\tilde{A}\mathrm{e}^{\mathrm{i}(kx-\omega t)} + \tilde{B}\mathrm{e}^{\mathrm{i}(kx+\omega t)}\right] dk. \tag{4.31}$$

In the above expression, k has been chosen as the independent variable, but as there is a relationship between ω and k, we could have chosen ω instead.

In general, waves of different frequencies travel with different phase speed. Therefore, a **wave packet**, which is obtained by summing up over a range of frequencies, will change shape as it propagates, since its different components change phase with respect to one another. The shape of the wave packet, called *envelope*, moves at the so–called **group velocity** given by:

$$v_g = \frac{d\omega}{dk}. \tag{4.32}$$

Electromagnetic waves in vacuum all have the same phase velocity $c = \omega/k$, independent of ω (vacuum is non dispersive). Therefore, the group velocity of electromagnetic waves in vacuum is also c.

4.2.2 Electromagnetic waves

We consider plane waves. As they can be written as a superposition of monochromatic sinusoidal waves, we focus on fields of the form:

$$\tilde{\mathbf{E}}(\mathbf{r},t) = \tilde{\mathbf{E}}_0\mathrm{e}^{\mathrm{i}(\mathbf{k}\cdot\mathbf{r}-\omega t)}\ , \tag{4.33}$$

$$\tilde{\mathbf{B}}(\mathbf{r},t) = \tilde{\mathbf{B}}_0\mathrm{e}^{\mathrm{i}(\mathbf{k}\cdot\mathbf{r}-\omega t)}\ , \tag{4.34}$$

where $\tilde{\mathbf{E}}_0$ and $\tilde{\mathbf{B}}_0$ are the (constant) amplitudes of the fields, $\mathbf{r}$ is the position vector and $\mathbf{k} \equiv k\hat{\mathbf{n}}$ is the **wave vector**, with $\hat{\mathbf{n}}$ being the unit vector along the direction of

[3]This theorem states that any function $f(x)$ which is absolutely integrable can be written as:

$$f(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} \hat{f}(k)\mathrm{e}^{\mathrm{i}kx}dk,$$

with

$$\hat{f}(k) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} f(x)\mathrm{e}^{-\mathrm{i}kx}dx.$$

The function $\hat{f}$ is the Fourier transform of f, and f is the inverse Fourier transform to $\hat{f}$.

propagation (if the wave propagates along the x–direction, $\mathbf{k}\cdot\mathbf{r} = kx$, as in the previous subsection). In general, $\tilde{\mathbf{E}}_0$ and $\tilde{\mathbf{B}}_0$ are complex.

These fields have to satisfy Maxwell's equations. Using the relation $\boldsymbol{\nabla}\cdot\tilde{\mathbf{E}} = \mathrm{i}\mathbf{k}\cdot\tilde{\mathbf{E}}$ (and similarly for $\tilde{\mathbf{B}}$), equations (4.15) and (4.16) lead to:

$$\boxed{\hat{\mathbf{n}}\cdot\tilde{\mathbf{E}} = 0,} \tag{4.35}$$

$$\boxed{\hat{\mathbf{n}}\cdot\tilde{\mathbf{B}} = 0,} \tag{4.36}$$

which means that both $\tilde{\mathbf{E}}$ and $\tilde{\mathbf{B}}$ (and therefore also $\mathbf{E}$ and $\mathbf{B}$) are perpendicular to the direction of propagation. Electromagnetic waves in vacuum are therefore **transverse**.

Using the relation $\boldsymbol{\nabla}\times\tilde{\mathbf{E}} = \mathrm{i}\mathbf{k}\times\tilde{\mathbf{E}}$ (and similarly for $\tilde{\mathbf{B}}$), we can further show that equations (4.17) and (4.18) are both equivalent to:

$$\boxed{\tilde{\mathbf{B}} = \frac{\hat{\mathbf{n}}\times\tilde{\mathbf{E}}}{c},} \tag{4.37}$$

which implies that $\tilde{\mathbf{E}}$ and $\tilde{\mathbf{B}}$ are **in phase** and **perpendicular to each other.** The same result is obtained for $\mathbf{E}$ and $\mathbf{B}$. Note that this does not hold only for electromagnetic *monochromatic* plane waves, but for electromagnetic plane waves in general.

4.2.3 Polarization

The polarization of an electromagnetic wave refers to the direction of the electric vector. At a fixed point in space, $\mathbf{E}$ may either keep a constant direction (**linear polarization**), or sweeps around a circle or an ellipse at the frequency of the wave (**circular** or **elliptical polarization**).

Here we adopt real, rather than complex, notations. We consider a monochromatic plane wave propagating along the x–direction. The most general expression is:

$$\begin{aligned} E_x &= 0, \\ E_y &= E_{0y}\cos\left(kx - \omega t + \varphi_y\right), \\ E_z &= E_{0z}\cos\left(kx - \omega t + \varphi_z\right), \end{aligned} \tag{4.38}$$

where E_{0y}, E_{0z}, φ_y and φ_z are positive constants. To describe the polarization of the wave, we can study either the evolution of $\mathbf{E}$ at a fixed value of x when t varies, or the dependance of $\mathbf{E}$ on x at a fixed time t. Here we use the first method, that is to say we consider a fixed value of x. Without loss of generality we can take $x = 0$, so that:

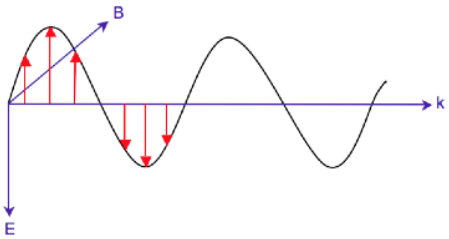
$$E_y = E_{0y}\cos\left(\omega t - \varphi_y\right), \quad \text{and} \quad E_z = E_{0z}\cos\left(\omega t - \varphi_z\right).$$

Linear polarization:

If $\varphi_z - \varphi_y = 0$ or π, then:

$$\frac{E_z}{E_y} = \pm\frac{E_{0z}}{E_{0y}}.$$

The electric vector vibrates while keeping a fixed direction. The electromagnetic wave is said to be linearly polarized, and the direction of polarization is that of the vector **E**.



Example of a linearly polarized wave.

Elliptical polarization:

Let us now consider the case where $\varphi_z - \varphi_y$ is not proportional to π. By shifting the origin of time, the electric field can be rewritten as:

$$E_y = E_{0y}\cos(\omega t)\,, \quad \text{and} \quad E_z = E_{0z}\cos(\omega t - \varphi)\,,$$

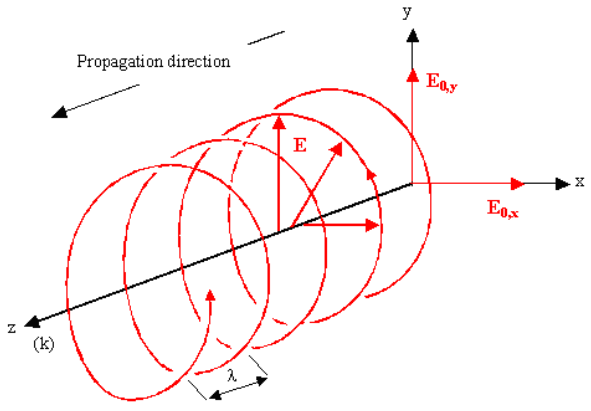
with $\varphi = \varphi_z - \varphi_y$. By eliminating the time from these two equations, we get:

$$\left(\frac{E_y}{E_{0y}}\right)^2 + \left(\frac{E_z}{E_{0z}}\right)^2 - 2\frac{E_y}{E_{0y}}\frac{E_z}{E_{0z}}\cos\varphi = \sin^2\varphi, \tag{4.39}$$

The end of the vector **E** therefore moves along an ellipse: the electromagnetic wave is elliptically polarized. Depending on whether the vector **E** rotates clockwise or counter–clockwise when the observer is facing into the oncoming wave, the polarization is said to be **right** or **left** elliptically polarized.

If $E_{0y} = E_{0z}$ and $\varphi = \pi/2$ or $3\pi/2$, equation (4.39) is that of a circle. In that case, the wave is **circularly** polarized.

An elliptically polarized wave can always be written as the sum of two out of phase waves which are linearly polarized in two directions perpendicular with each other.



Example of a circularly polarized wave.

Polarized and unpolarized light:

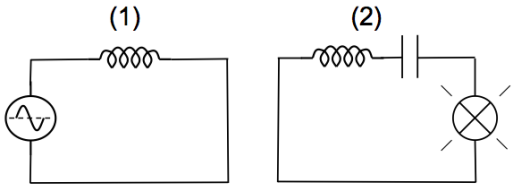
If a source is made of a collection of atoms or molecules which emit a strictly monochromatic light, the resultant electric field is the sum of linearly polarized waves which are *phase–locked*, that is to say which phases differ from one another by a constant. Such a source of light is said to be **coherent**, and the resultant electric field is itself polarized. Lasers and microwave ovens are examples of coherent sources.

By contrast, *natural light* (from a heated filament or the sun, for example) is not strictly monochromatic, as it is emitted by the de–excitation of atoms or molecules which act independently of each other. Each of them emits a short *wave train* with a specific polarization. As the phase difference between two wave trains with different frequencies is not constant (that is to say, the radiation is **incoherent**), the superposition of all the waves results in a randomly polarized, or unpolarized, electric field.

The superposition of unpolarized and polarized lights results in partially polarized light.

4.3 Energy and momentum transport by electromagnetic waves

Experimental facts suggest that electromagnetic waves transport energy. The most obvious fact is that the Earth receives energy from the Sun. This energy has travelled through (almost empty) space from the Sun to the Earth in the form of an electromagnetic wave.



Another example is illustrated in the figure on the left: a current through circuit (1) induces a current through circuit (2) which may be strong enough for the light bulb to glow, even though the two circuits are separated.

As there is no source of energy in circuit (2), this indicates that energy has been transferred through the air from circuit (1) to circuit (2).

Where is the energy as it travels? In this section, we are going to relate it to the electromagnetic field that carries it. Real notations are used throughout this section.

4.3.1 The Poynting vector

In electrostatics, we have calculated that a unit volume in which there is an electric field E has an amount of energy $\epsilon_0 E^2/2$ (eq. [2.28]). Similarly, in magnetostatics, we have established that a unit volume in which there is a magnetic field B has an amount of energy $B^2/(2\mu_0)$ (eq. [3.29]). Let us therefore assume that the energy density of an electromagnetic field is:

$$\boxed{u = \frac{\epsilon_0 E^2}{2} + \frac{B^2}{2\mu_0},} \tag{4.40}$$

even when the fields are time–dependent. Using $E^2 = \mathbf{E}\cdot\mathbf{E}$ and $B^2 = \mathbf{B}\cdot\mathbf{B}$, we obtain for the rate of change of u:

$$\frac{\partial u}{\partial t} = \epsilon_0 \frac{\partial \mathbf{E}}{\partial t}\cdot\mathbf{E} + \frac{1}{\mu_0}\frac{\partial \mathbf{B}}{\partial t}\cdot\mathbf{B}.$$

Equations (4.17) and (4.18) give $\partial\mathbf{E}/\partial t$ and $\partial\mathbf{B}/\partial t$ in terms of $\boldsymbol{\nabla}\times\mathbf{B}$ and $\boldsymbol{\nabla}\times\mathbf{E}$, so that:

$$\frac{\partial u}{\partial t} = \frac{1}{\mu_0}(\boldsymbol{\nabla}\times\mathbf{B})\cdot\mathbf{E} - \frac{1}{\mu_0}(\boldsymbol{\nabla}\times\mathbf{E})\cdot\mathbf{B}.$$

By using the vector identity $\boldsymbol{\nabla}\cdot(\mathbf{C}\times\mathbf{D}) = (\boldsymbol{\nabla}\times\mathbf{C})\cdot\mathbf{D} - (\boldsymbol{\nabla}\times\mathbf{D})\cdot\mathbf{C}$, we can finally write:

$$\frac{\partial u}{\partial t} = -\frac{1}{\mu_0}\boldsymbol{\nabla}\cdot(\mathbf{E}\times\mathbf{B}).$$

We define the **Poynting vector** by:

$$\boxed{\mathbf{S} = \frac{\mathbf{E}\times\mathbf{B}}{\mu_0},} \tag{4.41}$$

so that the rate of change of u can be written in the form:

$$\boxed{-\frac{\partial u}{\partial t} = \boldsymbol{\nabla}\cdot\mathbf{S}.} \tag{4.42}$$

This equation expresses the conservation of energy. It can be proved by integrating it over a given volume $\mathcal{V}$ delimited by a closed surface Σ:

$$-\iiint_{\mathcal{V}} \frac{\partial u}{\partial t}\, d\tau = \iiint_{\mathcal{V}} \boldsymbol{\nabla}\cdot\mathbf{S}\, d\tau. \tag{4.43}$$

As the surface delimiting the volume is fixed,

$$\iiint_{\mathcal{V}} \frac{\partial u}{\partial t}\, d\tau = \frac{d}{dt}\iiint_{\mathcal{V}} u\, d\tau.$$

Furthermore, using the divergence theorem, we can write:

$$\iiint_{\mathcal{V}} \boldsymbol{\nabla}\cdot\mathbf{S}\, d\tau = \oiint_{\Sigma} \mathbf{S}\cdot d\boldsymbol{\Sigma},$$

where $d\boldsymbol{\Sigma}$ is normal to the surface and pointing *outward*. Equation (4.43) then becomes:

$$-\frac{d}{dt}\iiint_{\mathcal{V}} u\, d\tau = \oiint_{\Sigma} \mathbf{S}\cdot d\boldsymbol{\Sigma}. \tag{4.44}$$

This shows that the negative of the rate at which the energy contained in the volume $\mathcal{V}$ varies is equal to the flux of the vector $\mathbf{S}$ outward through the surface Σ which delimits the volume. If energy flows out of the volume into another region, $\mathbf{S}\cdot d\boldsymbol{\Sigma} > 0$, and the energy contained in the volume decreases. Conversely, if energy flows into the volume from another region, $\mathbf{S}\cdot d\boldsymbol{\Sigma} < 0$, and the energy contained in the volume increases.

Since equation (4.44) is valid for any closed volume $\mathcal{V}$, **the Poynting vector S can be interpreted as the flow of energy per unit time per unit area through any surface**, closed or not. It is called the **power density**. The direction in which the energy flows is that of the vector $\mathbf{S}$.

4.3.2 Energy conservation for a system of charges and electromagnetic fields

In the previous section, we have established the conservation of energy of electromagnetic fields in vacuum. We now write the energy conservation of a combined system of particles and electromagnetic fields, and show that it can also be expressed using the Poynting vector.

Suppose we have a charge q moving with velocity $\mathbf{v}$ in a region of space where there is an electric field $\mathbf{E}$ and a magnetic field $\mathbf{B}$. The charge is subject to the Lorentz force, given by:

$$\mathbf{F} = q\left(\mathbf{E}+\mathbf{v}\times\mathbf{B}\right). \tag{4.45}$$

During a time interval dt, the charge moves over a distance $\mathbf{dl} = \mathbf{v}dt$. The work done on the charge by the electromagnetic field during dt is then:

$$dW = \mathbf{F}\cdot\mathbf{dl} = q\mathbf{E}\cdot\mathbf{v}dt,$$

and the power given to the charge by the electromagnetic field is:

$$\frac{dW}{dt} = q\mathbf{E}\cdot\mathbf{v}.$$

We now assume that the charge belongs to a distribution of charges which number density (number of charges per unit volume) is n. The power given by the electromagnetic field to the distribution of charges is:

$$\frac{dW}{dt} = \iiint_{\mathcal{V}} q\mathbf{E}\cdot\mathbf{v}n d\tau,$$

where the integral is over the volume of the distribution. By using the current density $\mathbf{J} = nq\mathbf{v}$, we obtain:

$$\boxed{\frac{dW}{dt} = \iiint_{\mathcal{V}} \mathbf{J}\cdot\mathbf{E}d\tau.} \tag{4.46}$$

This is the rate at which work is done on the charges in the volume $\mathcal{V}$.
The current density can be written as a function of $\mathbf{E}$ and $\mathbf{B}$ using Ampère–Maxwell's law (4.4), so that:

$$\mathbf{J}\cdot\mathbf{E} = \frac{1}{\mu_0}\left(\boldsymbol{\nabla}\times\mathbf{B}\right)\cdot\mathbf{E} - \epsilon_0\frac{\partial\mathbf{E}}{\partial t}\cdot\mathbf{E}.$$

Using the identity: $\boldsymbol{\nabla}\cdot(\mathbf{E}\times\mathbf{B}) = \mathbf{B}\cdot(\boldsymbol{\nabla}\times\mathbf{E}) - \mathbf{E}\cdot(\boldsymbol{\nabla}\times\mathbf{B})$ and Faraday's law (4.3), we obtain:

$$\mathbf{J}\cdot\mathbf{E} = -\frac{1}{\mu_0}\boldsymbol{\nabla}\cdot(\mathbf{E}\times\mathbf{B}) - \frac{1}{\mu_0}\frac{\partial\mathbf{B}}{\partial t}\cdot\mathbf{B} - \epsilon_0\frac{\partial\mathbf{E}}{\partial t}\cdot\mathbf{E} = -\frac{1}{\mu_0}\boldsymbol{\nabla}\cdot(\mathbf{E}\times\mathbf{B}) - \frac{1}{2}\frac{\partial}{\partial t}\left(\epsilon_0 E^2 + \frac{B^2}{\mu_0}\right).$$

Substituting into equation (4.46), we have:

$$\frac{dW}{dt} = -\iiint_{\mathcal{V}}(\boldsymbol{\nabla}\cdot\mathbf{S})d\tau - \iiint_{\mathcal{V}}\frac{\partial u}{\partial t}d\tau,$$

where we have used expressions (4.40) and (4.41) for the energy density u and the Poynting vector $\mathbf{S}$. As the surface delimiting the volume $\mathcal{V}$ of the charge distribution is fixed, the

time derivative can be taken out of the integral on the right–hand–side. We also use the divergence theorem to finally write the so–called **Poynting's theorem**:

$$\boxed{-\frac{d}{dt}\iiint_{\mathcal{V}} u\, d\tau = \oiint_{\Sigma} \mathbf{S}\cdot d\mathbf{\Sigma} + \frac{dW}{dt},} \tag{4.47}$$

where Σ is the surface delimiting the volume $\mathcal{V}$. This equality expresses conservation of energy. It is simliar to equation (4.44) but with the addition of the dW/dt term: the negative of the rate at which the energy contained in the volume $\mathcal{V}$ varies is equal to the rate at which energy flows out of the volume through the surface plus the rate at which work is done on the charges contained in the volume.

4.3.3 Application to monochromatic plane waves

In the case of electromagnetic plane waves, equation (4.37) gives $B^2 = E^2/c^2$. As $c = 1/\sqrt{\epsilon_0\mu_0}$, we obtain:

$$\frac{\epsilon_0 E^2}{2} = \frac{B^2}{2\mu_0}, \tag{4.48}$$

which indicates that there is an **equal amount of energy in the electric and magnetic fields**. The energy density (4.40) is therefore:

$$u = \epsilon_0 E^2 = \frac{B^2}{\mu_0}. \tag{4.49}$$

Using equation (4.37), we can write the Poynting vector (4.41) in the form:

$$\mathbf{S} = u\, c\, \hat{\mathbf{n}}. \tag{4.50}$$

As expected, the flow of energy is the energy density u times the group velocity c in the direction of propagation $\hat{\mathbf{n}}$ of the wave. Indeed, the energy carried by the wave through a surface area Σ perpendicular to the direction of propagation during a time interval Δt comes from a distance at most equal to $c\Delta t$. The amount of energy that crosses the surface during Δt is therefore $u\Sigma c\Delta t$, which corresponds to an energy per unit time per unit area uc.

As the power carried by the wave through a surface area Σ perpendicular to the direction of propagation is $S\Sigma$, and the plane wave is in theory infinite in the transverse direction, it carries an infinite amount of energy. This is of course physically unrealistic. Plane waves do not exist in reality. However, they are interesting to study because they may be a good *local* approximation to a more complicated wave. Also, as we have seen above, more realistic waves can be written as the superposition of plane waves.

Non–linearity of u and $\mathbf{S}$:

The expressions for u and $\mathbf{S}$ are non–linear: if E and B are multiplied by a factor α, u and $\mathbf{S}$ are multiplied by α^2. One consequence is that the principle of superposition does not apply to these quantities: **the Poynting vector corresponding to a superposition**

of electromagnetic waves is not the sum of the Poynting vectors corresponding to each wave. Also, because of the non–linearity, it is important to use real notations rather than complex notations to calculate u and $\mathbf{S}$.

Average values of u and S:

In the case of light, the frequency of the waves is so large that measurements of the energy usually encompass many cycles. The meaningful quantities are therefore averaged over time.

We consider a monochromatic plane wave propagating in the x–direction. The most general expression for $\mathbf{E}$ is given by equations (4.38). Remembering that $E^2 = \mathbf{E} \cdot \mathbf{E}$, the energy density (4.49) is therefore:

$$u = \epsilon_0 \left[E_{0y}^2 \cos^2 (kx - \omega t + \varphi_y) + E_{0z}^2 \cos^2 (kx - \omega t + \varphi_z) \right].$$

The time–average of the $\cos^2$ terms over a period is $1/2$, so that:

$$\langle u \rangle = \frac{\epsilon_0 E_0^2}{2}, \tag{4.51}$$

where $E_0^2 = E_{0y}^2 + E_{0z}^2$ and the brackets in $\langle u \rangle$ denote time–average. This result is independent of the polarization state of the wave. Similarly, we could show that the time–average energy density can be written as:

$$\langle u \rangle = \frac{B_0^2}{2\mu_0}. \tag{4.52}$$

Note that if we had calculated $\langle u \rangle$ using complex notations, we would have found $\langle u \rangle = 0$.[4]

The average power per unit area transported by the electromagnetic wave is called the **intensity**:

$$I \equiv \langle S \rangle = \frac{\epsilon_0 E_0^2}{2} c = \frac{B_0^2}{2\mu_0} c. \tag{4.53}$$

4.3.4 Momentum transport and radiation pressure

In the (classical) description of the electromagnetic field that we have given above, the energy of the field is distributed continuously, with an energy density u, and flows continuously at a rate per unit time and unit area corresponding to the Poynting vector $\mathbf{S}$. According to the quantum description, the electromagnetic field consists of discrete energy quanta, the photons. Each photon moves with the speed of light c and has an energy:

$$E = h\nu, \tag{4.54}$$

where ν is the frequency of the radiation and h is the Planck's constant. The corresponding momentum is:

$$p = \frac{E}{c} = \frac{h\nu}{c}. \tag{4.55}$$

[4] In complex notations, the components of $\mathbf{E}$ are $E_{0y} e^{i(kx - \omega t + \varphi_y)}$ and $E_{0z} e^{i(kx - \omega t + \varphi_z)}$ along the y– and z–directions, respectively. Therefore, $E^2 = E_{0y}^2 e^{2i(kx - \omega t + \varphi_y)} + E_{0z}^2 e^{2i(kx - \omega t + \varphi_z)}$. As $\langle \cos 2 (kx - \omega t + \varphi_z) \rangle = 0$, and similarly for the sine, E^2 in complex notations averages to 0 over a period of oscillations.

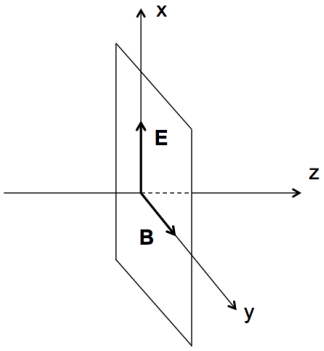
When a plane wave hits a material, it communicates its momentum to the surface. Let us consider a wave which hits a surface element $d\Sigma$ at normal incidence and is absorbed. The energy communicated by the wave to the surface during an interval of time dt is $Sd\Sigma dt$, where S is the modulus of the Poynting vector. According to equation (4.55), the momentum communicated to the surface element is therefore $Sd\Sigma dt/c$. The force (momentum per unit time) exerted by the wave onto the surface is thus $Sd\Sigma/c$. As the force is perpendicular to the surface (normal incidence), we can define a pressure (force per unit area) S/c. The **radiation pressure** P_{rad} is this pressure averaged over time:

$$P_{\text{rad}} = \frac{\langle S \rangle}{c} \equiv \frac{I}{c}, \tag{4.56}$$

where I is the intensity.
Using equation (4.53), we obtain: $P_{\text{rad}} = \epsilon_0 E_0^2/2$. If instead of being absorbed the wave were reflected, its momentum would reverse direction. Therefore, the momentum communicated to the surface would be twice as large as that calculated above, and the radiation pressure would accordingly be:

$$P_{\text{rad}} = \frac{2I}{c}. \tag{4.57}$$

This pressure force is due to the momentum given by the photons to the material. It can also be understood qualitatively in term of the electromagnetic field in the following way:



We assume that the wave that hits the surface travels from the left in the z–direction, with $\mathbf{E}$ and $\mathbf{B}$ being in the x– and y–directions, respectively. Electrons on the surface are given a velocity $\mathbf{v}$ along $-x$ by the electric field. The magnetic field then exerts a force $-e\mathbf{v}\times\mathbf{B}$ in the z–direction. The Lorentz force therefore has a component along $+z$, which corresponds to the radiation pressure.

Chapter 5

Electromagnetic waves in matter

5.1 Maxwell's equations in matter and boundary conditions

5.1.1 Maxwell's equations

Equations (4.1)–(4.4) are the most general local form of Maxwell's equations. They are valid in all possible environments, provided the charge density ρ and the current density $\mathbf{J}$ include all possible contributions. In the most general case, ρ has contribution from free[1] and polarization charges: $\rho = \rho_f + \rho_p$. Similarly, $\mathbf{J}$ has contribution from free currents and from currents produced by polarization and magnetization: $\mathbf{J} = \mathbf{J}_f + \mathbf{J}_p + \mathbf{J}_m$.
Therefore, equations (4.1) and (4.4) can be written under the form:

$$\boldsymbol{\nabla} \cdot \mathbf{E} = \frac{1}{\epsilon_0} \left(\rho_f + \rho_p \right), \tag{5.1}$$

and:

$$\boldsymbol{\nabla} \times \mathbf{B} = \mu_0 \left(\mathbf{J}_f + \mathbf{J}_p + \mathbf{J}_m \right) + \mu_0 \epsilon_0 \frac{\partial \mathbf{E}}{\partial t}. \tag{5.2}$$

Using the displacement vector:

$$\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P},$$

(eq. [2.14]) and the relation $\rho_p = -\boldsymbol{\nabla} \cdot \mathbf{P}$ (eq. [2.3]), equation (5.1) becomes $\boldsymbol{\nabla} \cdot \mathbf{D} = \rho_f$, as already shown in section 2.4 (see eq. [2.15]).
Similarly, using the auxiliary field:

$$\mathbf{H} = \frac{\mathbf{B}}{\mu_0} - \mathbf{M},$$

[1] What we call *free* charges are really the *excess* charges that would move in a conductor if an electric field were applied, until the charge density were zero. If the conductor is neutral, only surface charges are redistributed to cancel the applied field. More precisely, ρ_f can be written as:

$$\rho_f = \sum_{j,\text{free}} \rho_j + \sum_{n,\text{fixed molecules}} \rho_n,$$

where ρ_j is the charge density of the electrons and ρ_n is the charge density of the fixed molecules or atoms. See Jackson, section 6.6, for more details.

(eq. [3.18]) and the relations $\mathbf{J}_m = \boldsymbol{\nabla}\times\mathbf{M}$ (eq. [3.12]) and $\mathbf{J}_p = \partial\mathbf{P}/\partial t$ (eq. [2.5]), equation (5.2) becomes $\boldsymbol{\nabla}\times\mathbf{H} = \mathbf{J}_f + \partial\mathbf{D}/\partial t$.

Therefore, Maxwell's equations in matter can be summarised under the form:

$$\boxed{\begin{aligned}\boldsymbol{\nabla}\cdot\mathbf{D} &= \rho_f, && (5.3)\\ \boldsymbol{\nabla}\cdot\mathbf{B} &= 0, && (5.4)\\ \boldsymbol{\nabla}\times\mathbf{E} &= -\frac{\partial\mathbf{B}}{\partial t}, && (5.5)\\ \boldsymbol{\nabla}\times\mathbf{H} &= \mathbf{J}_f + \frac{\partial\mathbf{D}}{\partial t}. && (5.6)\end{aligned}}$$

5.1.2 Boundary conditions

In section 4.1.3, we derived the boundary conditions on $\mathbf{E}$ and $\mathbf{B}$ at a surface with free charges and/or free currents. The exact same calculations using the above form of Maxwell's equations lead to the boundary conditions on $\mathbf{E}$, $\mathbf{B}$, $\mathbf{D}$ and $\mathbf{H}$ at a surface with charges and/or currents having contribution from free sources, polarization and magnetization:

$$\boxed{\begin{aligned}\mathbf{D}_2^{\perp} - \mathbf{D}_1^{\perp} &= \sigma_f\,\hat{\mathbf{n}}_{12}, && (5.7)\\ \mathbf{E}_2^{\parallel} - \mathbf{E}_1^{\parallel} &= \mathbf{0}, && (5.8)\\ \mathbf{B}_2^{\perp} - \mathbf{B}_1^{\perp} &= \mathbf{0}, && (5.9)\\ \mathbf{H}_2^{\parallel} - \mathbf{H}_1^{\parallel} &= \mathbf{K}_f\times\hat{\mathbf{n}}_{12}, && (5.10)\end{aligned}}$$

where $\hat{\mathbf{n}}_{12}$ is the unit vector perpendicular to the surface and pointing from medium (1) to medium (2), $\mathbf{D}^{\perp} = D^{\perp}\hat{\mathbf{n}}_{12}$, with $D^{\perp} = \mathbf{D}\cdot\hat{\mathbf{n}}_{12}$, and similar definition for $\mathbf{B}^{\perp}$.

5.2 Electromagnetic waves in non–conducting linear media

In this section, we focus on media with no free charges or currents, which means $\rho_f = 0$ and $\mathbf{J}_f = \mathbf{0}$, and which are **linear**, so that $\mathbf{D} = \epsilon\mathbf{E}$ (eq. [2.24]) and $\mathbf{H} = \mathbf{B}/\mu$ (eq. [3.24]). We assume in addition that the media are **homogeneous** and **isotropic**, that is to say ϵ and μ are the same everywhere and in all directions. Then Maxwell's equations (5.3)–(5.6) become:

$$\begin{aligned}\boldsymbol{\nabla}\cdot\mathbf{E} &= 0, && (5.11)\\ \boldsymbol{\nabla}\cdot\mathbf{B} &= 0, && (5.12)\\ \boldsymbol{\nabla}\times\mathbf{E} &= -\frac{\partial\mathbf{B}}{\partial t}, && (5.13)\\ \boldsymbol{\nabla}\times\mathbf{B} &= \mu\epsilon\frac{\partial\mathbf{E}}{\partial t}, && (5.14)\end{aligned}$$

which are the same as equations (4.15)–(4.18), which describe electromagnetic waves in vacuum, but with μ_0 and ϵ_0 replaced by μ and ϵ, respectively. Therefore, electromagnetic waves that propagate through linear, homogeneous and isotropic (LHI) media have a structure similar to that of waves that propagate through vacuum. The **phase speed** of waves propagating through LHI media is:

$$\boxed{v_\varphi = \frac{1}{\sqrt{\mu\epsilon}} = \frac{c}{n},} \tag{5.15}$$

where:

$$\boxed{n \equiv \sqrt{\frac{\mu\epsilon}{\mu_0\epsilon_0}},} \tag{5.16}$$

is the **index of refraction**, or **refractive index**, of the medium. In most materials, $\mu \simeq \mu_0$, so that $n \simeq \sqrt{\epsilon_r}$.
In general, $\epsilon_r > 1$, so that $n > 1$ and light travels more slowly through matter than in vacuum.

Note that we have assumed when writing v_φ that $\mu\epsilon$ was real. This is not always the case, as we will see when studying waves propagating in media with frequency dependent permittivities (section 5.6 below). In this subsection however, we only consider media with permittivity ϵ and permeability μ that do not depend on ω and are real.
In that case, v_φ itself is independent of ω: waves with different frequencies travel with the same phase speed. In other words, there is **no dispersion**.
A medium through which a wave can propagate without dispersion is said to be **transparent** for that wave.

5.2.1 Ewald–Oseen extinction theorem

Above, we have shown that a wave which would propagate with the speed of light c in vacuum travels with a slower speed v_φ when penetrating in a linear material.

When an electromagnetic wave penetrates in a material, it moves the bound electrons around in such a way that electric and magnetic dipoles are induced. Each of these dipoles in turn produces an electromagnetic wave which has the same frequency as the applied field if the material is **linear**, and which also travels at speed c. According to the principle of superposition, the total field in the material is the sum of all these fields, including the original applied field. However, there is no wave travelling at speed c anymore after the original wave has entered the material. This is explained by the *Ewald–Oseen extinction theorem*, which states that the waves emitted by the dipoles can be written as the sum of a wave which exactly cancels out the original wave and phase–shifted waves which add up to a wave with a smaller phase speed.

5.2.2 Monochromatic plane waves

Like in the case of waves propagating in vacuum, we look for solutions of equations (5.11)–(5.14) under the form of monochromatic plane waves. Using complex notations, the most

general form of these solutions is:

$$\tilde{\mathbf{E}}(\mathbf{r},t) = \tilde{\mathbf{E}}_0 \mathrm{e}^{\mathrm{i}(\mathbf{k}\cdot\mathbf{r}-\omega t)} , \tag{5.17}$$

$$\tilde{\mathbf{B}}(\mathbf{r},t) = \tilde{\mathbf{B}}_0 \mathrm{e}^{\mathrm{i}(\mathbf{k}\cdot\mathbf{r}-\omega t)} , \tag{5.18}$$

where $\mathbf{k}$ is the wave vector. In general, $\tilde{\mathbf{E}}_0$ and $\tilde{\mathbf{B}}_0$ are complex. As for waves propagating in vacuum, equations (5.11) and (5.12) imply:

$$\boxed{\mathbf{k}\cdot\tilde{\mathbf{E}} = 0,} \tag{5.19}$$

$$\boxed{\mathbf{k}\cdot\tilde{\mathbf{B}} = 0,} \tag{5.20}$$

which means that electromagnetic waves in LHI media are **transverse**.

The equivalent of equation (4.19) for waves in vacuum is here:

$$\nabla^2\tilde{\mathbf{E}} = \mu\epsilon\frac{\partial^2\tilde{\mathbf{E}}}{\partial t^2}.$$

With the form of $\mathbf{E}$ given above, this equation yields the so–called **dispersion relation**:

$$\boxed{k^2 = \mu\epsilon\omega^2,} \tag{5.21}$$

which can also be written:

$$\boxed{\omega = kv_\varphi.} \tag{5.22}$$

The wavelength is defined by $\lambda = 2\pi/k$. With $k = \omega/v_\varphi = \omega n/c$, we obtain:

$$\lambda = \frac{2\pi c}{\omega}\frac{1}{n} = \frac{\lambda_0}{n}, \tag{5.23}$$

where λ_0 is the wavelength of the wave with same frequency that would propagate in vacuum. We note that $\lambda < \lambda_0$.

The **group velocity** of the waves is $v_g = d\omega/dk$ (eq. [4.32]), which is simply $v_g = v_\varphi$ when $\mu\epsilon$ does not depend on ω. In a non dispersive lossless medium, the group velocity is equal to the phase velocity, and this is also the velocity at which energy is transported.

Using the relation $\boldsymbol{\nabla}\times\tilde{\mathbf{E}} = \mathrm{i}\mathbf{k}\times\tilde{\mathbf{E}}$, we can further show that equation (5.13) leads to:

$$\boxed{\tilde{\mathbf{B}} = \frac{\mathbf{k}}{\omega}\times\tilde{\mathbf{E}},} \tag{5.24}$$

which implies that, like in vacuum, $\tilde{\mathbf{E}}$ and $\tilde{\mathbf{B}}$ are **perpendicular to each other.** Furthermore, **if $\mathbf{k}$ is real**, which requires $\mu\epsilon$ being real, **$\tilde{\mathbf{E}}$ and $\tilde{\mathbf{B}}$ are in phase**. Using equation (5.22) and the fact that $\mathbf{k}$ and $\mathbf{E}$ are perpendicular, we obtain from equation (5.24):

$$B = \frac{E}{v_\varphi}, \tag{5.25}$$

in real notations.

For a plane wave, we define the **intrinsic impedance** of the medium as:

$$\boxed{Z \equiv \frac{\tilde{E}_{\mathrm{ph}}}{\tilde{H}_{\mathrm{ph}}},} \tag{5.26}$$

where $\tilde{E}_{\text{ph}}$ and $\tilde{H}_{\text{ph}}$ are the phasor representations of the fields, that is to say the complex amplitude after the time–dependence has been factored out:

$$\tilde{E}(\mathbf{r},t) = \tilde{E}_{\text{ph}}(\mathbf{r})\mathrm{e}^{-\mathrm{i}\omega t}, \quad \tilde{H}(\mathbf{r},t) = \tilde{H}_{\text{ph}}(\mathbf{r})\mathrm{e}^{-\mathrm{i}\omega t}. \tag{5.27}$$

If $\tilde{E}$ and $\tilde{H}$ are not in phase, Z is complex. From equation (5.6), it can be seen that the units of H are amperes per metre. The units of E are volts per metre. Therefore, the units of Z are ohms, hence the name 'impedance'.
In the case we are studying here of a non–conducting medium, using equation (5.24) and $H = B/\mu$, we obtain:

$$\boxed{Z = \mu v_\varphi = \sqrt{\frac{\mu}{\epsilon}}.} \tag{5.28}$$

The impedance in vacuum is $Z_0 = \mu_0 c \simeq 377\ \Omega$.

5.2.3 Energy transport

In electrostatics, we have calculated that the unit volume of a material in which there is an electric field E and an associated polarization has an amount of energy $\mathbf{E}\cdot\mathbf{D}/2$ (eq. [2.28]). Similarly, in magnetostatics, we have established that the unit volume of a material in which there is a magnetic field B and an associated magnetization has an amount of energy $\mathbf{H}\cdot\mathbf{B}/2$ (eq. [3.29]). Let us therefore assume that the energy density of an electromagnetic field in a material is:

$$u = \frac{1}{2}\mathbf{E}\cdot\mathbf{D} + \frac{1}{2}\mathbf{H}\cdot\mathbf{B}, \tag{5.29}$$

even when the fields are time–dependent. In a linear medium, this expression can also be written as:

$$\boxed{u = \frac{\epsilon}{2}E^2 + \frac{1}{2\mu}B^2,} \tag{5.30}$$

which, as expected, is identical to equation (4.40) after replacing μ_0 and ϵ_0 by μ and ϵ, respectively.

In the same way as we have defined the Poynting vector (4.41) for waves propagating in vacuum, we define the Poynting vector for waves propagating in linear materials:

$$\boxed{\mathbf{S} = \frac{\mathbf{E}\times\mathbf{B}}{\mu}.} \tag{5.31}$$

In a material in which $\mu \simeq \mu_0$, the Poynting vector is the same as in vacuum. (Remember to take the real parts of $\mathbf{E}$ and $\mathbf{B}$ when calculating $\mathbf{S}$ or u). With $\mathbf{B}/\mu = \mathbf{H}$, we note that equation (5.31) can be written as:

$$\mathbf{S} = \mathbf{E}\times\mathbf{H}.$$

This is actually a more general expression of the Poynting vector, valid even in non–linear media. This could be shown by writing an energy conservation equation as in section 4.3.1.

For monochromatic plane waves given by equations (5.17) and (5.18), and using equation (5.24), we obtain the intensity:

$$I \equiv \langle S \rangle = \frac{1}{2\mu} \frac{k}{\omega} E_0^2, \tag{5.32}$$

where the brackets denote time–average. With $\omega/k = v_\varphi$ and $v_\varphi = 1/\sqrt{\mu\epsilon}$, this can also be written as:

$$I = \frac{\epsilon E_0^2}{2} \, v_g = \frac{B_0^2}{2\mu} \, v_g, \tag{5.33}$$

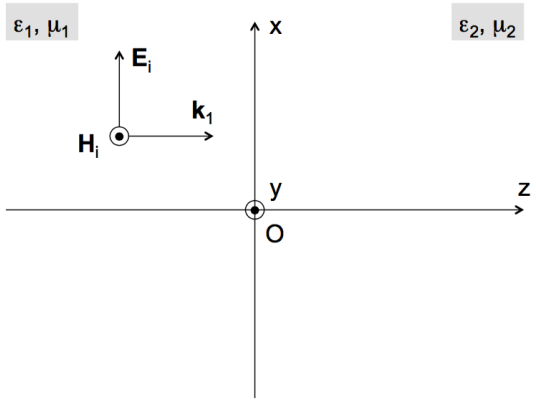
where we have used the fact that the group velocity v_g is equal to the phase velocity v_φ in a non dispersive medium. This is equivalent to equation (4.53) after replacing μ_0 and ϵ_0 by μ and ϵ, respectively. Here again, the energy is transported at the group velocity.

5.3 Reflection and transmission at the boundary between two linear media

We consider two linear media with permittivities ϵ_1 and ϵ_2, permeabilities μ_1 and μ_2, and separated by a plane surface.

5.3.1 Normal incidence

We first focus on a monochromatic plane wave hitting the interface at normal incidence.



We use cartesian coordinates with the interface being the $z = 0$ plane. An incident wave of frequency ω and polarized in the x–direction travels from the left to the right in the z–direction. We note $\mathbf{k}_1 = k_1 \hat{\mathbf{z}}$ the wave vector. When this wave hits the interface, it gives rise to both reflected and transmitted waves.

The incident wave can be written under the form:

$$\tilde{\mathbf{E}}_i(z,t) = E_{0i} \, \mathrm{e}^{\mathrm{i}(k_1 z - \omega t)} \, \hat{\mathbf{x}}, \tag{5.34}$$

$$\tilde{\mathbf{H}}_i(z,t) = H_{0i} \, \mathrm{e}^{\mathrm{i}(k_1 z - \omega t)} \, \hat{\mathbf{y}}. \tag{5.35}$$

We note ω_r, ω_t, $\mathbf{k}_r$ and $\mathbf{k}_t$ the frequencies and wave vectors of the reflected and transmitted waves. The electric field associated with those waves is therefore:

$$\tilde{\mathbf{E}}_r = \mathbf{E}_{0r} \, \mathrm{e}^{\mathrm{i}(\mathbf{k}_r \cdot \mathbf{r} - \omega_r t)}, \quad \tilde{\mathbf{E}}_t = \mathbf{E}_{0t} \, \mathrm{e}^{\mathrm{i}(\mathbf{k}_t \cdot \mathbf{r} - \omega_t t)}.$$

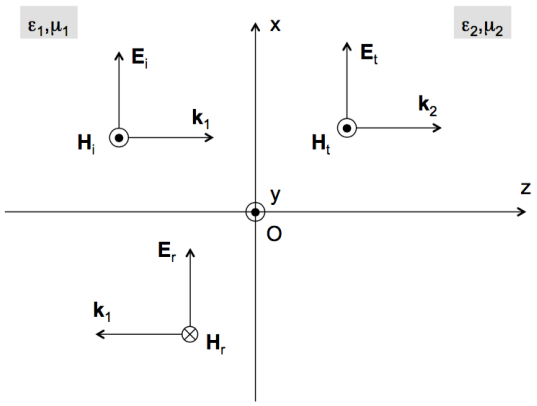
According to equation (5.8), the tangential component of the electric field is continuous at the interface, that is to say in the plane $z = 0$. Continuity of the x–component implies:

$$E_{0i}\ \mathrm{e}^{-\mathrm{i}\omega t} + E_{0r,x}\ \mathrm{e}^{\mathrm{i}(k_{r,x}x + k_{r,y}y - \omega_r t)} = E_{0t,x}\ \mathrm{e}^{\mathrm{i}(k_{t,x}x + k_{t,y}y - \omega_t t)},$$

where the subscripts x and y denote the x– and y–components. This condition has to be satisfied for all x, y and t, and therefore the exponential factors have to be equal, that is to say $-\omega t = k_{r,x}x + k_{r,y}y - \omega_r t = k_{t,x}x + k_{t,y}y - \omega_t t$. When $x = y = 0$, this implies $\omega = \omega_r = \omega_t$, which means that the reflected and transmitted waves have the same frequency as the incident wave. When $t = y = 0$, we obtain $0 = k_{r,x}x = k_{t,x}x$, which yields $k_{r,x} = k_{t,x} = 0$. Similarly, when $t = x = 0$, we obtain $k_{r,y} = k_{t,y} = 0$. Therefore, both $\mathbf{k}_r$ and $\mathbf{k}_t$ are in the z–direction, like the wave vector of the incident wave. Furthermore, using the dispersion relation (5.21) and the fact that all the waves have the same frequency, we see that $k_r = k_1$ whereas k_t has a different value, which we note k_2. As the reflected and transmitted waves travel toward $-z$ and $+z$, respectively, we can write $\mathbf{k_r} = -k_1\hat{\mathbf{z}}$ and $\mathbf{k_t} = k_2\hat{\mathbf{z}}$.

The fact that the reflected and transmitted waves have the same frequency as the incident wave can also be understood in the following way. When the incident wave hits the interface, it moves the bound electrons which are located there in such a way that electric and magnetic dipoles are induced. If the materials are linear, the frequency with which these dipoles oscillate is the same as the frequency of the fields that create them. In turn, the dipoles produce a transmitted and a reflected waves which have the same frequency.

The y–component of $\mathbf{E}$ is continuous at the interface, which means that $E_{0r,y} = E_{0t,y}$. Using equation (5.24), this yields $B_{0r,x}/k_1 = B_{0t,x}/k_2$, and therefore $\mu_1 H_{0r,x}/k_1 = \mu_2 H_{0t,x}/k_2$. As there are no free currents at the interface, equation (5.10) implies that the tangential component of $\mathbf{H}$ is continuous at the interface. This can be satisfied at the same time as the relation above only if $H_{0r,x} = H_{0t,x} = 0$, which also implies $E_{0r,y} = E_{0t,y} = 0$. Therefore, the reflected and transmitted waves have the same polarization as the incident wave.



Finally, the reflected and transmitted waves can be written under the form:

$$\tilde{\mathbf{E}}_r(z,t) = E_{0r}\ \mathrm{e}^{\mathrm{i}(-k_1 z - \omega t)}\ \hat{\mathbf{x}}, \qquad (5.36)$$

$$\tilde{\mathbf{H}}_r(z,t) = -H_{0r}\ \mathrm{e}^{\mathrm{i}(-k_1 z - \omega t)}\ \hat{\mathbf{y}}, \qquad (5.37)$$

$$\tilde{\mathbf{E}}_t(z,t) = E_{0t}\ \mathrm{e}^{\mathrm{i}(k_2 z - \omega t)}\ \hat{\mathbf{x}}, \qquad (5.38)$$

$$\tilde{\mathbf{H}}_t(z,t) = H_{0t}\ \mathrm{e}^{\mathrm{i}(k_2 z - \omega t)}\ \hat{\mathbf{y}}. \qquad (5.39)$$

The minus sign in $\tilde{\mathbf{H}}_r$ comes from equation (5.24) and is due to the fact that the reflected wave travels in opposite direction as the incident wave.

We have assumed on the figure above that both $\mathbf{E}_r$ and $\mathbf{E}_t$ have the same orientation as $\mathbf{E}_i$. It does not have to be the case, but if the orientation of $\mathbf{E}$ is changed then that of $\mathbf{H}$ has to be reversed accordingly. As we are going to see now, the sign of E_{0r} and E_{0t} is determined by the boundary conditions at the interface.
The continuity of the x– and y–components of $\mathbf{E}$ and $\mathbf{H}$, respectively, at the interface yields:

$$E_{0i} + E_{0r} = E_{0t}, \tag{5.40}$$

and:

$$H_{0i} - H_{0r} = H_{0t}. \tag{5.41}$$

Using the impedances $Z_1 = E_{0i}/H_{0i} = E_{0r}/H_{0r} = \mu_1 v_1$ and $Z_2 = E_{0t}/H_{0t} = \mu_2 v_2$ (see eq.[5.28]), where v_1 and v_2 are the phase velocities of the waves in media 1 and 2, respectively, equation (5.41) can be written as:

$$\frac{E_{0i}}{Z_1} - \frac{E_{0r}}{Z_1} = \frac{E_{0t}}{Z_2}. \tag{5.42}$$

Combining equations (5.40) and (5.42), we obtain:

$$E_{0r} = \frac{Z_2 - Z_1}{Z_1 + Z_2}\, E_{0i}, \quad \text{and} \quad E_{0t} = \frac{2Z_2}{Z_1 + Z_2}\, E_{0i}. \tag{5.43}$$

According to equation (5.33), the ratio of reflected intensity to incident intensity, which is called the **reflection coefficient** is:

$$R \equiv \frac{I_r}{I_i} = \left(\frac{E_{0r}}{E_{0i}}\right)^2 = \left(\frac{Z_2 - Z_1}{Z_1 + Z_2}\right)^2. \tag{5.44}$$

Similarly, we define the **transmission coefficient** as:

$$T \equiv \frac{I_t}{I_i} = \frac{\epsilon_2 v_2}{\epsilon_1 v_1}\left(\frac{E_{0t}}{E_{0i}}\right)^2 = \frac{\epsilon_2 v_2}{\epsilon_1 v_1}\left(\frac{2Z_2}{Z_1 + Z_2}\right)^2 = \frac{4Z_1 Z_2}{(Z_1 + Z_2)^2}, \tag{5.45}$$

where we have used $\epsilon_2 v_2/(\epsilon_1 v_1) = \mu_1 v_1/(\mu_2 v_2) = Z_1/Z_2$ (see eq. [5.15]). As expected from conservation of energy, we have:

$$R + T = 1. \tag{5.46}$$

In most materials, $\mu \simeq \mu_0$, so that $Z = \mu v_\varphi \simeq \mu_0 v_\varphi = \mu_0 c/n$, and equations (5.43), (5.44) and (5.45) can be written as:

$$E_{0r} = \frac{n_1 - n_2}{n_1 + n_2}\, E_{0i}, \quad E_{0t} = \frac{2n_1}{n_1 + n_2}\, E_{0i}, \tag{5.47}$$

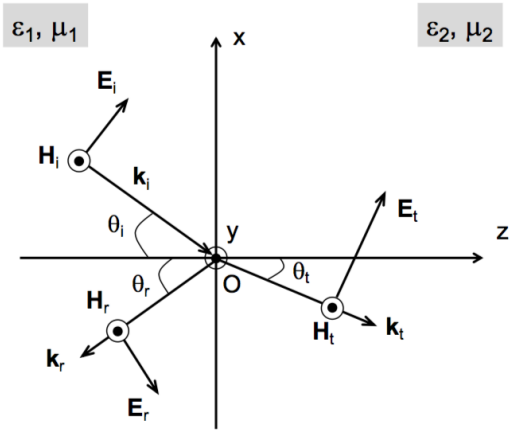
and:

$$R = \left(\frac{n_1 - n_2}{n_1 + n_2}\right)^2, \quad T = \frac{4n_1 n_2}{(n_1 + n_2)^2}. \tag{5.48}$$

The reflected wave is in phase with the incident wave if $n_1 > n_2$ and out of phase by π otherwise. The transmitted wave is always in phase with the incident wave.
If the wave is travelling from air ($n_1 = 1$) into glass ($n_2 = 1.5$), then $R = 0.04$ and $T = 0.96$. As expected, most of the energy is transmitted.

5.3.2 Oblique incidence and Brewster's angle

We now consider the case when the monochromatic plane wave is incident at an angle θ_i to the normal, with a polarization parallel to the plane of incidence.



The incident wave can be written under the form:

$$\tilde{\mathbf{E}}_i = \mathbf{E}_{0i}\, \mathrm{e}^{\mathrm{i}(\mathbf{k}_i\cdot\mathbf{r}-\omega t)}, \quad (5.49)$$

$$\tilde{\mathbf{H}}_i = \frac{1}{Z_1}\hat{\mathbf{k}}_i \times \tilde{\mathbf{E}}_i, \quad (5.50)$$

where $\hat{\mathbf{k}}_i = \mathbf{k}_i/k_i$.

This problem will be studied in detail in Problem Set 3. Here we only summarize the main results, starting with the three fundamental laws of geometrical optics:

- the three wavevectors $\mathbf{k}_i$, $\mathbf{k}_r$ and $\mathbf{k}_t$ are all in the same plane that also contains the normal to the surface z,
- $\theta_i = \theta_r$ (law of reflection),
- $n_2 \sin\theta_t = n_1 \sin\theta_i$ (law of refraction, or Snell's law).

The boundary conditions also yield the reflected and transmitted (complex) amplitudes, known as **Fresnel's equations**:

$$\tilde{E}_{0r} = \frac{Z_1\cos\theta_i - Z_2\cos\theta_t}{Z_1\cos\theta_i + Z_2\cos\theta_t}E_{0i}, \quad \text{and} \quad \tilde{E}_{0t} = \frac{2Z_2\cos\theta_i}{Z_1\cos\theta_i + Z_2\cos\theta_t}E_{0i}. \quad (5.51)$$

As can be seen from the equation giving $\tilde{E}_{0r}$, there is **no reflection** when $Z_1\cos\theta_i = Z_2\cos\theta_t$. The value of θ_i for which this happens is called the *Brewster's angle* and is noted θ_B. Using Snell's law, we obtain:

$$\cos\theta_B = \frac{Z_2}{Z_1}\sqrt{1-\left(\frac{n_1}{n_2}\right)^2\sin^2\theta_B}, \quad (5.52)$$

which yields:

$$\sin^2\theta_B = \frac{1-(Z_1/Z_2)^2}{(n_1/n_2)^2-(Z_1/Z_2)^2}. \quad (5.53)$$

For most materials, $\mu \simeq \mu_0$, so that $Z_1/Z_2 \simeq n_2/n_1$ and the equation above becomes:

$$\boxed{\tan\theta_B \simeq \frac{n_2}{n_1}}. \quad (5.54)$$

5.4 Electromagnetic waves in conductors

So far, in this section, we have focussed on the propagation of waves in media with no free charges or currents. We are now going to study wave propagation in media in which conduction currents are non zero, and which may also have some polarization and magnetization.

5.4.1 Skin depth

We assume that Ohm's law applies[2], that is to say the free current density is related to the electric field by:

$$\mathbf{J}_f = \sigma \mathbf{E}, \tag{5.55}$$

where σ is the conductivity of the material.
If we assume the material to be *linear*, Maxwell's equations (5.3)–(5.6) can then be written under the form:

$$\boldsymbol{\nabla} \cdot \mathbf{E} = \frac{\rho_f}{\epsilon}, \tag{5.56}$$

$$\boldsymbol{\nabla} \cdot \mathbf{B} = 0, \tag{5.57}$$

$$\boldsymbol{\nabla} \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \tag{5.58}$$

$$\boldsymbol{\nabla} \times \mathbf{B} = \mu\sigma\mathbf{E} + \mu\epsilon\frac{\partial \mathbf{E}}{\partial t}. \tag{5.59}$$

Here, μ and ϵ are real. The density of free charges[3] ρ_f satisfies the equation of charge conservation:

$$\boldsymbol{\nabla} \cdot \mathbf{J}_f = -\frac{\partial \rho_f}{\partial t}. \tag{5.60}$$

Combining this equation with equations (5.55) and (5.56) yields:

$$\frac{\partial \rho_f}{\partial t} = -\frac{\sigma}{\epsilon}\rho_f, \tag{5.61}$$

which solution is:

$$\rho_f = \rho_0 \mathrm{e}^{-t/\tau}, \tag{5.62}$$

where ρ_0 is the density of free charges at $t = 0$ and we have defined the *relaxation time* $\tau = \epsilon/\sigma$. In a perfect conductor, $\sigma = \infty$ and $\tau = 0$, which means that the charges instantaneously flow out of the material (to the surfaces). If an external field varying with a frequency ω is applied, the material is said to be a good conductor for $\tau \ll 1/\omega$ and a bad conductor for $\tau \gg 1/\omega$. Note that, in equation (5.59), $\mu\epsilon|\partial\mathbf{E}/\partial t| = \mu\epsilon\omega|\mathbf{E}|$, and

[2] Remember that Ohm's law is not a fundamental law of nature, but an empirical relationship that does not always apply. At high frequencies for example, larger than about 10^{11} Hz, the inertia of the conduction electrons is such that they cannot keep up with the very rapid variations of the field. A conductor in which Ohm's law applies is called an *ohmic conductor*.

[3] See the footnote in section 5.1.1 for the meaning of *free* charges.

this term is small compared to $\mu\sigma|\mathbf{E}|$ in a good conductor. In other words, in a good conductor, displacement currents are negligible compared to conduction currents.
After a time large compared to τ, all the free charges have flown out of the material, and we can set $\rho_f = 0$ in equation (5.56). We then take the curl of equations (5.58) and (5.59), and using the identity:

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

which also applies to $\mathbf{B}$, we obtain the following partial differential equations for $\mathbf{E}$ and $\mathbf{B}$:

$$\nabla^2\mathbf{E} = \mu\epsilon\frac{\partial^2\mathbf{E}}{\partial t^2} + \mu\sigma\frac{\partial\mathbf{E}}{\partial t}, \tag{5.63}$$

$$\nabla^2\mathbf{B} = \mu\epsilon\frac{\partial^2\mathbf{B}}{\partial t^2} + \mu\sigma\frac{\partial\mathbf{B}}{\partial t}. \tag{5.64}$$

As before, we look for solutions under the form of monochromatic plane waves with frequency ω, and choose the z–axis along the direction of propagation. Adopting complex notations, the most general form of these solutions for waves propagating in the $+z$–direction is:

$$\tilde{\mathbf{E}}(z,t) = \tilde{\mathbf{E}}_0 \mathrm{e}^{\mathrm{i}\left(\tilde{k}z-\omega t\right)}, \tag{5.65}$$

$$\tilde{\mathbf{B}}(z,t) = \tilde{\mathbf{B}}_0 \mathrm{e}^{\mathrm{i}\left(\tilde{k}z-\omega t\right)}, \tag{5.66}$$

where $\tilde{\mathbf{E}}_0$, $\tilde{\mathbf{B}}_0$ and $\tilde{k}$ are complex. Substituting these expressions into equations (5.63) and (5.64) yields the *dispersion relation*:

$$\tilde{k}^2 = \mu\epsilon\omega^2 + \mathrm{i}\mu\sigma\omega. \tag{5.67}$$

We write:

$$\tilde{k} = k' + \mathrm{i}k'', \tag{5.68}$$

where k' and k'' are real. Equation (5.67) then yields $k'^2 - k''^2 = \mu\epsilon\omega^2$ and $2k'k'' = \mu\sigma\omega$, so that:

$$k' = \omega\left[\frac{\mu\epsilon}{2}\left(1+\sqrt{1+\left(\frac{\sigma}{\epsilon\omega}\right)^2}\right)\right]^{1/2}, \tag{5.69}$$

$$k'' = \omega\left[\frac{\mu\epsilon}{2}\left(-1+\sqrt{1+\left(\frac{\sigma}{\epsilon\omega}\right)^2}\right)\right]^{1/2}, \tag{5.70}$$

where we have retained the solution $k' > 0$ as we are looking for solutions propagating in the $+z$–direction (note that $k'k''$ has to be positive). Equations (5.65) and (5.66) then become:

$$\tilde{\mathbf{E}}(z,t) = \tilde{\mathbf{E}}_0 \mathrm{e}^{-k''z}\mathrm{e}^{\mathrm{i}(k'z-\omega t)}, \tag{5.71}$$

$$\tilde{\mathbf{B}}(z,t) = \tilde{\mathbf{B}}_0 \mathrm{e}^{-k''z}\mathrm{e}^{\mathrm{i}(k'z-\omega t)}, \tag{5.72}$$

These equations show that the waves propagate with the phase speed:

$$v_\varphi = \frac{\omega}{k'}, \tag{5.73}$$

which in general depends on ω. Therefore, **a conductor is a dispersive medium**. In a bad conductor, $\sigma/(\epsilon\omega) \ll 1$ and equation (5.69) gives $k' \simeq \omega\sqrt{\mu\epsilon}$. This yields $v_\varphi \simeq 1/\sqrt{\mu\epsilon}$, as in a dielectric (see eq. [5.15]). Therefore, a bad conductor acts like a dielectric. This will be discussed again in section 5.6.5.

Equations (5.71) and (5.72) also show that the waves are **attenuated** as they propagate. The e–folding distance (distance over which the amplitude of the wave decreases by a factor e) is called the **skin depth** and is given by:

$$\delta = \frac{1}{k''}. \tag{5.74}$$

The skin depth is the distance through which the electromagnetic wave (and therefore free currents, since $\mathbf{J}_f = \sigma\mathbf{E}$) penetrates into the conductor.

- In a perfect conductor, $\sigma \to \infty$ and $k'' \to \infty$, so that $\delta \to 0$, which means that the wave does not penetrate into the material, as expected.
- In a good conductor, $\sigma/(\epsilon\omega) \gg 1$ and $\delta \simeq \sqrt{2/(\mu\sigma\omega)}$ depends on the frequency of the wave. If the frequency is very high, currents only flow in a very thin layer near the surface of the material[4]. Therefore, to carry a current, it is enough to coat any conductor with a very good conductor. Even at low, power–line frequencies (50 Hz in Europe), bundled conductors are a better choice than solid conductors to carry currents.
- In a bad conductor, $\sigma/(\epsilon\omega) \ll 1$ and $\delta \simeq (2/\sigma)\sqrt{\epsilon/\mu}$ is independent of ω. The skin depth in a bad conductor is much larger than in a good conductor.

Note that, when $\omega \to 0$, and if the conductivity is finite, $k'' \to 0$, so that $\delta \to \infty$. But we also have $k' \to 0$, so that $\tilde{\mathbf{E}}$ and $\tilde{\mathbf{B}}$ given by equations (5.71) and (5.72) are constant

[4]When $\sigma \gg \epsilon\omega$, the current density $\mathbf{J}_f$ is confined to such a small thickness just below the surface of the conductor that it is equivalent to an effective surface current $\mathbf{K}_f$. If we assume that the conductor extends from $z = 0$ to $z = \infty$, then:

$$\tilde{\mathbf{K}}_f = \int_0^\infty \tilde{\mathbf{J}}_f dz = \int_0^\infty \sigma\tilde{\mathbf{E}} dz = \int_0^\infty \sigma\tilde{\mathbf{E}}_0 \mathrm{e}^{\mathrm{i}(\tilde{k}z - \omega t)} dz = -\frac{\sigma}{\mathrm{i}\tilde{k}}\tilde{\mathbf{E}}_0 \mathrm{e}^{-\mathrm{i}\omega t}.$$

We still have $\boldsymbol{\nabla}\times\tilde{\mathbf{E}} = \mathrm{i}\tilde{\mathbf{k}}\times\tilde{\mathbf{E}}$ when $\tilde{k}$ is complex, so that equation (5.58) yields $\tilde{\mathbf{E}}_0 = \omega\tilde{\mathbf{B}}_0\times\hat{\mathbf{z}}/\tilde{k}$, where $\hat{\mathbf{z}}$ is the unit vector in the direction of propagation. Therefore, $\mathbf{K}_f = (i\omega\sigma/\tilde{k}^2)\tilde{\mathbf{B}}_0 \mathrm{e}^{-\mathrm{i}\omega t}\times\hat{\mathbf{z}}$. When $\sigma \gg \epsilon\omega$, equations (5.69) and (5.70) yield $k' = k'' \simeq \sqrt{\mu\sigma\omega/2}$. Therefore we can write $\mathbf{K}_f = (1/\mu)\tilde{\mathbf{B}}_0 \mathrm{e}^{-\mathrm{i}\omega t}\times\hat{\mathbf{z}}$. This is the boundary condition for the parallel component of the magnetic field near the surface of a perfect conductor, as can be seen from equation (5.10) with $\mathbf{H}_2^{\parallel} = \mathbf{0}$ (field inside the conductor), $\mathbf{H}_1^{\parallel} = \tilde{\mathbf{B}}_0 \mathrm{e}^{-\mathrm{i}\omega t}/\mu$ (field just outside the conductor, at $z = 0^-$) and $\hat{\mathbf{n}}_{12} = \hat{\mathbf{z}}$. Therefore, *a good conductor behaves like a perfect conductor*, with the idealized surface current replaced by an equivalent volume current distributed throughout a thickness which is very small but *nonetheless non zero*. [Note that we would have obtained the same result if we had assumed that the conductor were extending from whatever value of $z = z_{\mathrm{in}}$ to a value of $z = z_{\mathrm{out}}$ such that $z_{\mathrm{out}} - z_{\mathrm{in}} \gg \delta$.]

and uniform throughout the conductor. Equation (5.59) then implies that $\mathbf{E}$ is zero, which means that the electromagnetic wave does not penetrate into the material. This is consistent with the fact that any conductor is a good conductor when $\omega \to 0$.
The skin effect is due to electromagnetic induction. A time–varying magnetic field produces an emf which in turn creates currents. According to Lenz's law, they oppose the currents which produced them in the first place, so that the total current is reduced. The larger the conductivity, the larger the induced currents.
Another way to explain the skin effect is by considering the motion of the electrons in the conductor. When an electric field is applied, the electrons move toward the surfaces under the effect of the electric force and they themselves produce an electric field which opposes the applied field. If the electrons can follow the variations of the applied field, which is the case in a good conductor where $\tau \ll 1/\omega$, then at every instant the field produced by the electrons cancels the applied field, which is therefore zero in most of the volume of the conductor. In a bad conductor, where $\tau \gg 1/\omega$, the inertia of the electrons is such that they cannot keep up with the variations of the field. Before the electrons have time to reach the surfaces of the conductor, the applied field changes direction and the electrons turn around. Therefore, the field they produce cancels the applied field only in a restricted region of the conductor.

5.4.2 Impedance

We still have $\boldsymbol{\nabla}\times\tilde{\mathbf{E}} = \mathrm{i}\tilde{\mathbf{k}}\times\tilde{\mathbf{E}}$ when $\tilde{k}$ is complex, so that the relation given by equation (5.24) still holds. With $\tilde{\mathbf{k}} = \tilde{k}\hat{\mathbf{z}}$ and writing $\tilde{k} = |\tilde{k}|\mathrm{e}^{\mathrm{i}\phi}$, we can then obtain $\tilde{\mathbf{B}}$ from the expression (5.71) of $\tilde{\mathbf{E}}$:

$$\tilde{\mathbf{B}} = \frac{|\tilde{k}|}{\omega}\left(\hat{\mathbf{z}}\times\tilde{\mathbf{E}}_0\right)\mathrm{e}^{-k''z}\mathrm{e}^{\mathrm{i}(k'z-\omega t+\phi)}, \tag{5.75}$$

where $\phi = \tan^{-1}(k''/k')$. We see that **B lags behind E** by the angle ϕ.
The impedance Z of the conductor is then (see eq. [5.26]):

$$Z = \mu\frac{\tilde{E}_{\mathrm{ph}}}{\tilde{B}_{\mathrm{ph}}} = \mu\frac{\omega}{\tilde{k}} = \mu\frac{\omega}{|\tilde{k}|^2}\left(k' - \mathrm{i}k''\right). \tag{5.76}$$

In a good conductor, $\sigma/(\epsilon\omega) \gg 1$ so that $k' \simeq k'' \simeq \sqrt{\mu\sigma\omega/2}$, and:

$$Z \simeq \sqrt{\frac{\mu\omega}{2\sigma}}(1-\mathrm{i}) \simeq \frac{1-\mathrm{i}}{\delta\sigma}. \tag{5.77}$$

Note that:

$$|Z| = \sqrt{\frac{\mu\omega}{\sigma}} = \sqrt{\frac{\epsilon\omega}{\sigma}}\sqrt{\frac{\mu}{\epsilon}},$$

so that, in a good conductor where $\epsilon\omega/\sigma \ll 1$, the impedance is small compared to the impedance of a non–conducting medium with same permittivity and permeability. In particular, if $\epsilon = \epsilon_0$ and $\mu = \mu_0$, the impedance of the conductor is small compared to that of vacuum.

5.4.3 Energy transport

The Poynting vector is given by:

$$\mathbf{S} = \frac{\mathbf{E}\times\mathbf{B}}{\mu}, \tag{5.78}$$

where *real* notations have to be used here.
For $\mathbf{E}$, we take the real part of the field given by equation (5.71), so that:

$$\mathbf{E} = \mathbf{E}_0 \mathrm{e}^{-k''z} \cos\left(k'z - \omega t + \psi\right). \tag{5.79}$$

where we have written $\tilde{\mathbf{E}}_0 = \mathbf{E}_0 \mathrm{e}^{\mathrm{i}\psi}$.
For $\tilde{\mathbf{B}}$, we use equation (5.75) under the form:

$$\tilde{\mathbf{B}} = \frac{\tilde{k}}{\omega}\left(\hat{\mathbf{z}}\times\mathbf{E}_0\right)\mathrm{e}^{-k''z}\mathrm{e}^{\mathrm{i}(k'z-\omega t+\psi)}, \tag{5.80}$$

and therefore:

$$\mathbf{B} = \frac{1}{\omega}\left(\hat{\mathbf{z}}\times\mathbf{E}_0\right)\mathrm{e}^{-k''z}\left[k'\cos\left(k'z-\omega t+\psi\right) - k''\sin\left(k'z-\omega t+\psi\right)\right]. \tag{5.81}$$

Equation (5.78) then yields:

$$\mathbf{S} = \hat{\mathbf{z}}\frac{1}{\mu\omega}E_0^2\mathrm{e}^{-2k''z}\left[k'\cos^2\left(k'z-\omega t+\psi\right) - k''\sin\left(k'z-\omega t+\psi\right)\cos\left(k'z-\omega t+\psi\right)\right]. \tag{5.82}$$

We now perform a time–average over a period. The $\cos^2$ term gives $1/2$ whereas the $\cos\sin$ term gives zero. We then obtain:

$$\langle\mathbf{S}\rangle = \hat{\mathbf{z}}\frac{k'E_0^2}{2\mu\omega}\mathrm{e}^{-2k''z}, \tag{5.83}$$

which shows that the intensity $I \equiv \langle S\rangle$ of the wave decreases exponentially as it propagates with the absorption coefficient $2k''$.

5.5 Reflection and transmission at a conducting surface

We consider an electromagnetic monochromatic plane wave which travels in a non–conducting linear medium (which we label 1) and hits at normal incidence the plane surface of an ohmic conductor (medium 2). This is similar to the situation studied in section 5.3.1, except that here the intrinsic impedance of the conducting medium is a complex number. The boundary conditions are given by equations (5.7)–(5.10), and in principle there may be surface charges and currents. If we consider a thin layer of width d close to the surface of the conductor, then the surface current is given by the limit when $d \to 0$ of $\mathbf{J}_f d$. However, in an ohmic conductor, $\mathbf{J}_f = \sigma\mathbf{E}$. Therefore, if $\mathbf{E}$ stays finite as we approach the surface, $\mathbf{J}_f d$ vanishes as $d \to 0$. This means that there is *no surface current in an ohmic*

conductor[5]. Therefore, like for the reflection and transmission at the boundary between two dielectrics, the tangential components of $\mathbf{E}$ and $\mathbf{H}$ are continuous at the interface, which implies that the incident, reflected and transmitted waves have the same frequency, travel in the same direction, and are polarized in the same direction. Note that, as the electric field has no component perpendicular to the interface, equation (5.7) implies that there are no charges at the surface of the conductor.

As in section 5.3.1, we use cartesian coordinates with the interface being the $z = 0$ plane, and write the incident, reflected and transmitted waves as:

$$\tilde{\mathbf{E}}_i(z,t) = \tilde{E}_{0i}\, \mathrm{e}^{\mathrm{i}(k_1 z - \omega t)}\, \hat{\mathbf{x}}, \tag{5.84}$$

$$\tilde{\mathbf{H}}_i(z,t) = \tilde{H}_{0i}\, \mathrm{e}^{\mathrm{i}(k_1 z - \omega t)}\, \hat{\mathbf{y}}, \tag{5.85}$$

$$\tilde{\mathbf{E}}_r(z,t) = \tilde{E}_{0r}\, \mathrm{e}^{\mathrm{i}(-k_1 z - \omega t)}\, \hat{\mathbf{x}}, \tag{5.86}$$

$$\tilde{\mathbf{H}}_r(z,t) = -\tilde{H}_{0r}\, \mathrm{e}^{\mathrm{i}(-k_1 z - \omega t)}\, \hat{\mathbf{y}}, \tag{5.87}$$

$$\tilde{\mathbf{E}}_t(z,t) = \tilde{E}_{0t}\, \mathrm{e}^{\mathrm{i}(\tilde{k}_2 z - \omega t)}\, \hat{\mathbf{x}}, \tag{5.88}$$

$$\tilde{\mathbf{H}}_t(z,t) = \tilde{H}_{0t}\, \mathrm{e}^{\mathrm{i}(\tilde{k}_2 z - \omega t)}\, \hat{\mathbf{y}}, \tag{5.89}$$

where we have used complex notations. Note that the wave vector in the conductor is complex. The tangential components of $\mathbf{E}$ and $\mathbf{H}$ are continuous at the interface, which yields:

$$\tilde{E}_{0i} + \tilde{E}_{0r} = \tilde{E}_{0t}, \tag{5.90}$$

$$\tilde{H}_{0i} - \tilde{H}_{0r} = \tilde{H}_{0t}. \tag{5.91}$$

In the non–conducting medium, we have $\tilde{H}_{0i} = \tilde{E}_{0i}/Z_1$ and $\tilde{H}_{0r} = \tilde{E}_{0r}/Z_1$ with $Z_1 = \mu_1 v_1$, where v_1 is the phase velocity of the wave in the medium (see eq. [5.28]). In the conductor, $\tilde{H}_{0t} = \tilde{E}_{0t}/Z_2$ with $Z_2 = \mu_2\omega/\tilde{k}_2$ (see eq. [5.76]).
Combining equations (5.90) and (5.91), we therefore obtain:

$$\tilde{E}_{0r} = \frac{Z_2 - Z_1}{Z_1 + Z_2}\, \tilde{E}_{0i}, \quad \text{and} \quad \tilde{E}_{0t} = \frac{2Z_2}{Z_1 + Z_2}\, \tilde{E}_{0i}. \tag{5.92}$$

This is similar to the results obtained in section 5.3.1, except that here Z_2 is complex.

For a perfect conductor, $\sigma \to \infty$ and $|\tilde{k}_2| \to \infty$ (see eq. [5.69] and [5.70]), so that $Z_2 \to 0$. This yields $\tilde{E}_{0r} = -\tilde{E}_{0i}$ and $\tilde{E}_{0t} = 0$. As expected, the wave does not penetrate at all into the conductor and is completely reflected.

According to equation (5.83), the intensity which is transmitted at the interface, where $z = 0$, is $I_t = k'_2 E_{0t}^2/(2\mu_2\omega)$, with $E_{0t} = |\tilde{E}_{0t}|$ and $k'_2 = \mathrm{Re}(\tilde{k}_2)$. According to equation (5.32), the intensities which are incident and reflected at the interface are

[5] As shown in the footnote in the discussion below equation (5.74), a good conductor is equivalent to a perfect conductor, but with the idealized surface current replaced by a volume current distributed throughout some finite thickness below the surface. Therefore, in a perfect conductor we use Maxwell's equations with no volume current ($\mathbf{J}_f = \mathbf{0}$) and a surface current $\mathbf{K}_f$ which enters the boundary condition on $\mathbf{H}$, whereas in a good conductor, as studied here, we use Maxwell's equations with a non zero volume current $\mathbf{J}_f$ and no surface current ($\mathbf{K}_f = \mathbf{0}$) entering the boundary condition on $\mathbf{H}$. The two descriptions are equivalent and yield the same results

$I_i = k_1 E_{0i}^2/(2\mu_1\omega)$ and $I_r = k_1 E_{0r}^2/(2\mu_1\omega)$, respectively, with $E_{0i} = |\tilde{E}_{0i}|$ and $E_{0r} = |\tilde{E}_{0r}|$. Therefore, the reflection coefficient is:

$$R \equiv \frac{I_r}{I_i} = \frac{E_{0r}^2}{E_{0i}^2} = \left|\frac{\tilde{E}_{0r}}{\tilde{E}_{0i}}\right|^2 = \left|\frac{Z_2 - Z_1}{Z_1 + Z_2}\right|^2, \tag{5.93}$$

and the transmission coefficient is:

$$T \equiv \frac{I_t}{I_i} = \frac{\mu_1 k'_2}{\mu_2 k_1}\frac{E_{0t}^2}{E_{0i}^2} = \frac{\mu_1 k'_2}{\mu_2 k_1}\left|\frac{\tilde{E}_{0t}}{\tilde{E}_{0i}}\right|^2 = \frac{\mu_1 k'_2}{\mu_2 k_1}\left|\frac{2Z_2}{Z_1 + Z_2}\right|^2. \tag{5.94}$$

In a good conductor, $Z_2 \simeq \eta(1-\mathrm{i})$ with $\eta = \sqrt{\mu_2\omega/(2\sigma)}$ (see eq. [5.77]). As $Z_1 = \sqrt{\mu_1/\epsilon_1}$, we have:

$$\frac{\eta}{Z_1} = \sqrt{\frac{\epsilon_2\omega}{2\sigma}}\sqrt{\frac{\epsilon_1\mu_2}{\epsilon_2\mu_1}}.$$

The first factor on the right–hand–side is small compared to unity in a good conductor. Therefore, if the permittivities and permeabilities of the two media are not too different, we have $\eta/Z_1 \ll 1$. The reflection coefficient can then be written as:

$$R = \left|\frac{(\eta - Z_1) - \mathrm{i}\eta}{(\eta + Z_1) - \mathrm{i}\eta}\right|^2 = \frac{(\eta - Z_1)^2 + \eta^2}{(\eta + Z_1)^2 + \eta^2} \simeq \frac{1 - 2\eta/Z_1}{1 + 2\eta/Z_1} \simeq 1 - 4\frac{\eta}{Z_1},$$

to first order in η/Z_1. As expected, most of the energy is reflected. In a good conductor, $k'_2 \simeq \sqrt{\mu_2\sigma\omega/2}$ (see eq. [5.69]), so that the transmission coefficient can be written as:

$$T \simeq \frac{1}{2}\mu_1\frac{\omega}{k_1}\frac{1}{\eta}\left|\frac{2Z_2}{Z_1 + Z_2}\right|^2.$$

Using $\omega/k_1 = v_1 = 1/\sqrt{\mu_1\epsilon_1}$, we obtain:

$$T \simeq \frac{1}{2}\sqrt{\frac{\mu_1}{\epsilon_1}}\frac{1}{\eta}\frac{8\eta^2}{(\eta + Z_1)^2 + \eta^2} \simeq 4\frac{\eta}{Z_1},$$

to first order in η/Z_1. We verify that $R + T = 1$.

5.6 Electromagnetic waves in media with frequency dependent permittivity

In section 2.5, we have pointed out that the polarization does not always respond instantaneously to the variations of the field. At high frequency, there may be a delay between the adjustment of the atomic or molecular dipoles and the variation of the field. In that case, the electric susceptibility depends on frequency. This implies that waves with different frequencies travel with different speeds, a phenomenon called **dispersion**.

5.6.1 Dispersion of glass

In section 5.3.1, we have considered a wave hitting the interface between two linear media with *normal* incidence, and found that the reflected and transmitted waves were also travelling in the direction perpendicular to the interface. However, if the incident wave travels through medium 1 in a direction that makes an angle θ_1 with the normal to the interface, the transmitted wave may travel at a different angle θ_2 through medium 2, a process known as *refraction*. In that case, the transmitted wave is usually called *refracted* wave. According to Snell's law of refraction:

$$n_1 \sin\theta_1 = n_2 \sin\theta_2, \tag{5.95}$$

hence the name "index of refraction" for n.

As is well known from experience, a beam of incident "white" light is refracted when passing through a prism in such a way that the different colours are dispersed, that is to say they emerge from the prism at different angles. It means that the index of refraction n of the prism depends on the frequency of the waves. If the prism is made of glass, blue is bent more than red: the index of refraction increases (decreases) with frequency (wavelength). This is called *normal* dispersion.

5.6.2 A simple model to explain dispersion

To describe the physics of dispersive media, we adopt the so–called *Lorentz oscillator model*. It is essentially the same model as in chapter 2, where the polarization was explained in term of the displacement of bound charges under the effect of an external field, but taking into account the fact that charges (that is to say, electrons) do not respond instantaneously to an applied time–varying field. Although this classical model can only be an approximation to quantum models which alone can describe atomic–scale phenomena, it can successfully describe many experimental results.

We consider an electron of charge $q = -e$ which is bound to an atom or a molecule and is acted on by an electric field $\mathbf{E}(\mathbf{r}, t)$. Since the mass of the atom or molecule is much larger than that of the electron, we assume it stays fixed. As the displacement $\mathbf{x}$ of the electron from its equilibrium position caused by the electric field is small, the binding

force F_{bind} can be approximated by a spring–like harmonic force[6], that is to say it has the form $\mathbf{F}_{\text{bind}} = -m\omega_0^2\mathbf{x}$, where ω_0 is the frequency of the free oscillations. To account for the fact that the electron cannot adjust instantaneously to the variations of the applied field, we introduce a resistive force that damps the motion of the electron. This damping term is due to internal collisions in the material and radiation emitted by the electron (as any accelerating charge emits radiation). Resistive forces oppose the motion and do not act when the particle is at rest. The simplest form for the damping force is therefore $\mathbf{F}_{\text{damp}} = -m\gamma\dot{\mathbf{x}}$, where γ is a damping coefficient which has the dimensions of a frequency and the dot denotes derivative with respect to time. The equation of motion for the electron is then:

$$m\ddot{\mathbf{x}} = q\mathbf{E}(\mathbf{x}, t) - m\gamma\dot{\mathbf{x}} - m\omega_0^2\mathbf{x}, \tag{5.96}$$

where m is the mass of the electron. We have assumed here that the magnetic force was negligible compared to the electric force, which is usually the case[7]. We have also assumed that the electric field acting upon the electron was the applied field, which means that we have neglected the field produced by neighbouring molecules or atoms. This is true only in a dilute (low density) material.

As the displacement of the electron is small, the variations of $\mathbf{E}$ with position can be neglected, so that $\mathbf{E}$ can be evaluated at the average position of the electron. We consider an electromagnetic wave which has a frequency ω and we adopt complex notations. Here again, we denote complex quantities with a tilde. Remember that *physical* quantities are obtained by taking the real part. The electric field is then $\tilde{\mathbf{E}} = \mathbf{E}_0 \mathrm{e}^{-\mathrm{i}\omega t}$. In steady state, the displacement oscillates with the same frequency, that is to say $\tilde{\mathbf{x}}(t) \propto \mathrm{e}^{-\mathrm{i}\omega t}$, which implies $\dot{\tilde{\mathbf{x}}} = -\mathrm{i}\omega\tilde{\mathbf{x}}$ and $\ddot{\tilde{\mathbf{x}}} = -\omega^2\tilde{\mathbf{x}}$. Therefore, equation (5.96) yields:

$$\left(\omega_0^2 - \omega^2 - \mathrm{i}\omega\gamma\right)\tilde{\mathbf{x}} = \frac{q}{m}\mathbf{E}_0 \mathrm{e}^{-\mathrm{i}\omega t}. \tag{5.97}$$

The dipole moment contributed to by the electron is then the real part of:

$$\tilde{\mathbf{p}} = q\tilde{\mathbf{x}} = \frac{e^2/m}{\omega_0^2 - \omega^2 - \mathrm{i}\omega\gamma}\,\mathbf{E}_0 \mathrm{e}^{-\mathrm{i}\omega t}. \tag{5.98}$$

[6] We note U the potential energy which is associated with the binding force, and we assume that the motion is along the x–direction. Then $F_{\text{bind}} = -dU/dx$. If the equilibrium position is x_0, then for small displacements x around x_0 we can expand U in Taylor series:

$$U(x) = (x - x_0)\frac{dU}{dx}(x_0) + \frac{1}{2}(x - x_0)^2\frac{d^2U}{dx^2}(x_0) + \ldots,$$

where we have chosen the reference $U(x_0) = 0$. The first derivative is zero since the potential energy is extremum at an equilibrium point. Therefore, for small displacements:

$$F_{\text{bind}}(x) = -(x - x_0)\frac{d^2U}{dx^2}(x_0) + \ldots$$

If the equilibrium is stable, the second derivative is positive and then the force is of the form $F_{\text{bind}}(x) = -m\omega_0^2(x - x_0)$, where ω_0 is the frequency with which the particle freely oscillates around the equilibrium position.

[7] The electric force acting on the charge is $F_e = qE$, whereas the magnetic force is $F_m = |q\mathbf{v}\times\mathbf{B}|$, where $\mathbf{v}$ is the velocity of the charge. Therefore, $F_m/F_e \sim vB/E = v/v_\varphi$, where v_φ is the phase speed of the electromagnetic wave in the medium (see eq. [5.25]). Very often, $v \ll v_\varphi$ and the magnetic force can be neglected compared to the electric force.

The complex denominator indicates that **the dipole moment is out of phase with the electric field**. This can be seen by writing $\mathbf{p}$ under the form:

$$\tilde{\mathbf{p}} = \frac{e^2/m}{\sqrt{\left(\omega_0^2 - \omega^2\right)^2 + \omega^2\gamma^2}} \mathbf{E}_0 \mathrm{e}^{\mathrm{i}(-\omega t + \varphi)}, \tag{5.99}$$

where $\varphi = \tan^{-1}\left[\omega\gamma/\left(\omega_0^2 - \omega^2\right)\right]$ is the angle by which **p lags behind E**. This expression of the dipole moment is similar to the amplitude of a driven damped harmonic oscillator. Like in the case of an oscillator, we expect to find resonances when the frequency ω of the electromagnetic wave approaches ω_0. As we will see below, these resonances lead to large changes in the index of refraction of the medium and strong absorption of the wave.

We suppose that there are N molecules per unit volume with Z electrons per molecule. In general, different electrons have different natural frequencies ω_0 and different damping coefficients γ, which can be calculated using quantum mechanics. We assume that there are f_j electrons per molecule with natural frequency ω_j and damping coefficient γ_j. Then the polarization (dipole moment per unit volume) is:

$$\tilde{\mathbf{P}} = \frac{Ne^2}{m}\tilde{\mathbf{E}} \sum_j \frac{f_j}{\omega_j^2 - \omega^2 - \mathrm{i}\omega\gamma_j}. \tag{5.100}$$

The f_j are called *oscillator strengths* and they satisfy $\sum_j f_j = Z$.

The real part of the polarization given by equation (5.100) is not proportional to the real part of the electric field, which means that the material is not linear. However, we define a **complex susceptibility** $\tilde{\chi}_e$ such that:

$$\tilde{\mathbf{P}} = \epsilon_0 \tilde{\chi}_e \tilde{\mathbf{E}}, \tag{5.101}$$

that is to say:

$$\tilde{\chi}_e = \frac{Ne^2}{\epsilon_0 m} \sum_j \frac{f_j}{\omega_j^2 - \omega^2 - \mathrm{i}\omega\gamma_j}. \tag{5.102}$$

When $\omega \to 0$, $\tilde{\chi}_e$ is real and positive, and the polarization is equal to the static polarization. In that case, the polarization and the electric field are in phase. On the other hand, when ω is very large, the susceptibility becomes negative, which means that $\mathbf{P}$ is out–of–phase with $\mathbf{E}$ by π. In the limit $\omega \to \infty$, the susceptibility vanishes: the electrons cannot respond to the variations of the electric field.

Note that there is no real susceptibility obtained by taking the real part of $\tilde{\chi}_e$. If we write $\tilde{\chi}_e = \chi'_e + \mathrm{i}\chi''_e$, where χ'_e and χ''_e are real, then equation (5.101) implies:

$$\mathbf{P} = \mathrm{Re}(\tilde{\mathbf{P}}) = \epsilon_0 \left[\chi'_e \cos(\omega t) + \chi''_e \sin(\omega t)\right] \mathbf{E}_0. \tag{5.103}$$

The phase lag between $\mathbf{P}$ and $\mathbf{E}$ can be made explicit by writing the above expression under the form:

$$\mathbf{P} = \epsilon_0 \sqrt{(\chi'_e)^2 + (\chi''_e)^2} \cos(\omega t - \varphi)\mathbf{E}_0, \tag{5.104}$$

with $\varphi = \tan^{-1}(\chi''_e/\chi'_e)$.

The displacement vector in complex notation is:

$$\tilde{\mathbf{D}} = \epsilon_0 \tilde{\mathbf{E}} + \tilde{\mathbf{P}} = \epsilon_0 \left(1 + \tilde{\chi}_e\right) \tilde{\mathbf{E}}. \tag{5.105}$$

We define the **complex permittivity**:

$$\tilde{\epsilon} = \epsilon_0 \left(1 + \tilde{\chi}_e\right), \tag{5.106}$$

and the **complex relative permittivity**, also called **complex dielectric contant**:

$$\tilde{\epsilon}_r = \frac{\tilde{\epsilon}}{\epsilon_0} = 1 + \tilde{\chi}_e, \tag{5.107}$$

in the same way as we have defined the permittivity and dielectric constant in linear media for steady fields (see section 2.5). But again, here, there are no real permittivity or dielectric constant obtained by taking the real part of the complex quantities. As can be seen from equations (5.102) and (5.107), the complex relative permittivity depends on ω and is given by:

$$\boxed{\tilde{\epsilon}_r(\omega) = 1 + \frac{Ne^2}{\epsilon_0 m} \sum_j \frac{f_j}{\omega_j^2 - \omega^2 - \mathrm{i}\omega\gamma_j}.} \tag{5.108}$$

We note ϵ'_r and ϵ''_r the real and imaginary parts of $\tilde{\epsilon}_r$, respectively. We then have:

$$\epsilon'_r = 1 + \frac{Ne^2}{\epsilon_0 m} \sum_j \frac{f_j \left(\omega_j^2 - \omega^2\right)}{\left(\omega_j^2 - \omega^2\right)^2 + \omega^2\gamma_j^2}, \tag{5.109}$$

$$\epsilon''_r = \frac{Ne^2\omega}{\epsilon_0 m} \sum_j \frac{f_j \gamma_j}{\left(\omega_j^2 - \omega^2\right)^2 + \omega^2\gamma_j^2}. \tag{5.110}$$

5.6.3 Absorption and anomalous dispersion

In general, the damping coefficients γ_j are small compared to the **resonant frequencies** ω_j, so that when ω is not close to one of the ω_j the imaginary part of $\tilde{\epsilon}_r$ can be neglected. However, near a resonant frequency ω_j, the damping term becomes very important.

Wave equation:

We assume here that the permeability of the medium is $\mu = \mu_0$. In complex form, and in the absence of free charges and currents, Maxwell's equations (5.3)–(5.6) can be written as follows:

$$\boldsymbol{\nabla} \cdot \tilde{\mathbf{D}} = 0, \tag{5.111}$$

$$\boldsymbol{\nabla} \cdot \tilde{\mathbf{B}} = 0, \tag{5.112}$$

$$\boldsymbol{\nabla} \times \tilde{\mathbf{E}} = -\frac{\partial \tilde{\mathbf{B}}}{\partial t}, \tag{5.113}$$

$$\boldsymbol{\nabla} \times \tilde{\mathbf{B}} = \mu_0 \frac{\partial \tilde{\mathbf{D}}}{\partial t}. \tag{5.114}$$

With $\tilde{\mathbf{D}} = \tilde{\epsilon}\tilde{\mathbf{E}}$, equations (5.113) and (5.114) can be combined in the usual way to yields the wave equation:

$$\nabla^2\tilde{\mathbf{E}} = \mu_0\tilde{\epsilon}\frac{\partial^2\tilde{\mathbf{E}}}{\partial t^2}. \tag{5.115}$$

As before, we look for solutions under the form of monochromatic plane waves, and choose the z–axis along the direction of propagation. The most general form of these solutions for waves propagating in the $+z$–direction is:

$$\tilde{\mathbf{E}}(z,t) = \tilde{\mathbf{E}}_0 \mathrm{e}^{\mathrm{i}(\tilde{k}z-\omega t)}, \tag{5.116}$$

where $\tilde{E}_0$ and $\tilde{k}$ are complex. Substituting (5.116) into the wave equation (5.115) yields the *dispersion relation*:

$$\tilde{k}^2 = \mu_0\tilde{\epsilon}\omega^2. \tag{5.117}$$

If we write:

$$\tilde{k} = k' + \mathrm{i}k'', \tag{5.118}$$

where k' and k'' are real, then equation (5.117) implies $k'^2 - k''^2 = \mu_0\omega^2\epsilon'$ and $2k'k'' = \mu_0\omega^2\epsilon''$, where ϵ' and ϵ'' are the real and imaginary parts of $\tilde{\epsilon}$, respectively. As we are looking for waves propagating in the $+z$–direction, we select $k' > 0$. Then $k'' > 0$ if $\epsilon'' > 0$, which is the case when the damping coefficients γ_j are positive. A negative ϵ'' would lead to an amplification of the wave, as in a maser or laser. We will not consider this case here. Equation (5.116) can be written as:

$$\tilde{\mathbf{E}}(z,t) = \tilde{\mathbf{E}}_0 \mathrm{e}^{-k''z}\mathrm{e}^{\mathrm{i}(k'z-\omega t)}. \tag{5.119}$$

Therefore, the wave propagates in the z–direction with the phase velocity:

$$v_\varphi = \frac{\omega}{k'}\,, \tag{5.120}$$

and is **attenuated**.

Energy transport:

We calculate the Poynting vector in exactly the same way as in section 5.4.3. Taking $\mu = \mu_0$ in equation (5.83) we then obtain:

$$\langle\mathbf{S}\rangle = \hat{\mathbf{z}}\frac{k'E_0^2}{2\mu_0\omega}\mathrm{e}^{-2k''z}, \tag{5.121}$$

which shows that the intensity of the wave decreases exponentially as it propagates[8] with the **absorption coefficient** $2k''$.

[8] Energy dissipation in dielectrics with complex dielectric constant will be studied in Problem Set 3.

Complex index of refraction:

We define the **complex index of refraction** $\tilde{n}$ such that:

$$\tilde{n}^2 = \tilde{\epsilon}_r. \tag{5.122}$$

Then equation (5.117) yields:

$$\tilde{k} = \tilde{n}\frac{\omega}{c}. \tag{5.123}$$

If we write:

$$\tilde{n} = n' + \mathrm{i}n'', \tag{5.124}$$

where n' and n'' are real, then equation (5.120) becomes:

$$v_\varphi = \frac{c}{n'}\,. \tag{5.125}$$

Therefore, n' is called the **index of refraction**.

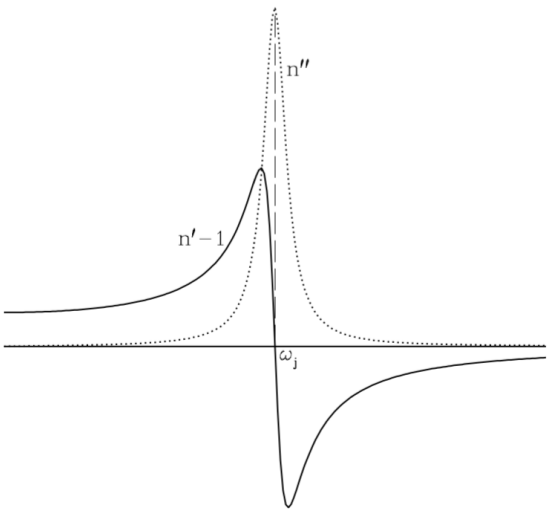
For gases, the second term on the right hand side in the expression (5.108) of $\tilde{\epsilon}_r$ is small compared to unity, so that we can approximate $\tilde{n}$ by:

$$\tilde{n} = \sqrt{\tilde{\epsilon}_r} \simeq 1 + \frac{Ne^2}{2\epsilon_0 m}\sum_j \frac{f_j}{\omega_j^2 - \omega^2 - \mathrm{i}\omega\gamma_j}. \tag{5.126}$$

The real and imaginary parts of $\tilde{n}$ are therefore given by:

$$n' \simeq 1 + \frac{Ne^2}{2\epsilon_0 m}\sum_j \frac{f_j\left(\omega_j^2 - \omega^2\right)}{\left(\omega_j^2 - \omega^2\right)^2 + \omega^2\gamma_j^2}, \tag{5.127}$$

$$n'' \simeq \frac{Ne^2\omega}{2\epsilon_0 m}\sum_j \frac{f_j\gamma_j}{\left(\omega_j^2 - \omega^2\right)^2 + \omega^2\gamma_j^2}. \tag{5.128}$$



The figure on the left shows $n' - 1$ (solid line) and n'' (dotted line) as a function of ω in the vicinity of a resonant frequency ω_j. The straight horizontal line is the zero axis. As pointed out above, dispersion of visible light in most transparent materials (air, glass, etc.) is such that the index of refraction increases with ω. This is called normal dispersion. We see in the figure that this is the case for all values of ω, except in the neighbourhood of the resonant frequency, where n' drops sharply. This is called **anomalous dispersion**.

Anomalous dispersion occurs in the region of maximum absorption (the absorption coefficient is $2k'' = 2\omega n''/c$), also called **resonant absorption**. The width of the absorption peak is on the order of γ_j, which is generally small compared to ω_j. We see that, for $\omega > \omega_j$, $n' < 1$, and therefore $v_\varphi > c$ from equation (5.125). This is not a concern as only the phase of the waves, not the energy, propagates at the phase velocity. In addition, we note that taking into account all the resonant frequencies in the expression of n' could bring it above 1.

In the model developed in this section, we have assumed that each electron was attached to an atom or molecule by a spring. It seems like a very crude representation of a far more complex reality. However, a detailed quantum mechanical analysis gives results which are remarkably close to those of this simple model. Here, we find that absorption occurs when the frequency of the wave is close to the natural frequency ω_j of the spring. In reality, absorption occurs when the frequency of the wave is close to that of a transition between two energy states in the material, like two energy levels of an electron around a nucleus. This is because an electron can make a transition from an energy level E_1 to a higher energy level E_2 only if it absorbs a photon which energy is equal to $E_2 - E_1$. When a wave enters a material, if its frequency matches that of a transition (that is to say if there is a resonance), it is absorbed by the electrons which then make the transition between the energy levels. On the other hand, if the frequency of the wave is not close to any of the transition frequencies in the material, there is no absorption. It is remarkable that, for the purpose of studying the interaction between wave and matter, the electrons can be described as being attached to an atom or molecule by a spring with a frequency that corresponds to a transition between energy levels.
In calculating $\tilde{\epsilon}_r$, we have neglected the contribution from the ions. Far from a resonance, their contribution would be small compared to that of the electrons as their mass is much larger. However, ions produce anomalous dispersion and resonant absorption at specific resonant frequencies, for example those associated with the stretching, bending and rotation of molecules. These frequencies are typically in the infrared or microwave region of the electromagnetic spectrum.

In a transparent medium like air, glass or water, there are no important resonances of the molecules in the visible range of frequencies. Therefore, when a beam of visible light travels through such a medium, there is no absorption and all the waves with the different frequencies are transmitted. This is why the medium is colorless. In glass, the resonant frequencies of the atoms are within the UV range, which is why the UV part of the spectrum is absorbed and we cannot get sunburnt through a window. Waves with frequencies in the infrared range are also absorbed as these frequencies are close to those with which atoms in some molecules vibrate. Coloured glasses can be produced by introducing atoms with mobile electrons which absorb a particular frequency in the visible spectrum. When a beam of "white" light enters a coloured glass, this frequency is absorbed and is not present in the beam which is transmitted to an observer, who therefore sees color. Note that the photons which are absorbed may be re-emitted at the same frequency, but as they will be radiated in all directions (scattering) they will still be

missing from the transmitted beam.

5.6.4 Group velocity

In general, a source of electromagnetic waves contains a spread of frequencies. Each sinusoidal component travels at the phase velocity, and if this velocity depends on the frequency then there is dispersion. However, if the medium has a real dielectric constant (and therefore is lossless), it can be shown that the packet as a whole travels with the *group velocity* $v_g = d\omega/dk$. If an energy density is associated with the magnitude of the wave, it is transported at the group velocity.
In a dissipative medium, the group velocity is defined as $v_g = d\omega/dk'$, and in general this is the velocity at which the signal and the energy propagate. However, this is not the case in a region of anomalous dispersion. To see why, we can write:

$$v_g = \left(\frac{dk'}{d\omega}\right)^{-1} = c\left[\frac{d}{d\omega}\left(n'\omega\right)\right]^{-1} = v_\varphi \frac{1}{\left(1 + \frac{\omega}{n'}\frac{dn'}{d\omega}\right)}.$$

For anomalous dispersion, $dn'/d\omega < 0$ and v_g may become larger than c and /or negative. The concept of group velocity becomes meaningless in that case.

5.6.5 Plasma frequency

A plasma is a ionised gas in which positive charges and electrons are unbound. The overall charge of a plasma is zero. In general, the electromagnetic field generated by the motion of particles in a plasma affect the other particles. Here, we consider the case of a very dilute plasma in which there is an external field, so that the external field dominates the local field.

The equation of motion for an electron in a plasma is:

$$m\ddot{\mathbf{x}} = q\mathbf{E}(\mathbf{x}, t) - m\gamma\dot{\mathbf{x}}, \tag{5.129}$$

which is similar to equation (5.96) but with $\omega_0 = 0$, as there is no binding force acting on the electron. In a plasma, the damping term comes from electron–ion collisions.

Collisionless plasma:

When $\omega \gg \gamma$, that is to say the wave frequency is much larger than the electron–ion collision frequency, we can describe the plasma using the model developed above with $\omega_j = 0$ and $\gamma_j = 0$. In principle, there is a free current due to the motion of the electrons. This has not been taken into account in Maxwell's equation (5.114). However, this equation contains a contribution from the polarization current. Therefore, here we attribute all the properties of the plasma to the dielectric constant ϵ.

With $\omega_j = 0$ and $\gamma_j = 0$, the dielectric constant given by equation (5.108) is real and can be written as:

$$\epsilon_r = 1 - \frac{Ne^2Z}{\epsilon_0 m\omega^2}, \tag{5.130}$$

where we have used $\sum_j f_j = Z$. As the contribution of a charge to the dielectric constant is $\propto 1/m$, we neglect the contribution from the positive charges. We define the **plasma frequency** ω_p as:

$$\boxed{\omega_p^2 = \frac{ne^2}{\epsilon_0 m},} \tag{5.131}$$

where $n \equiv NZ$ is the number of electrons per unit volume. Then the dielectric constant takes the form:

$$\boxed{\epsilon_r = 1 - \frac{\omega_p^2}{\omega^2}.} \tag{5.132}$$

Note that the same dielectric constant is obtained in a dielectric at frequencies much higher than the highest resonant frequency in the material. This can be seen by taking $\omega \gg \max(\omega_j)$ in equation (5.108).

With this expression of the dielectric constant, the dispersion relation (5.117) becomes:

$$k^2 = \mu_0 \epsilon_0 \omega^2 \left(1 - \frac{\omega_p^2}{\omega^2}\right), \tag{5.133}$$

which can also be written as:

$$k = \frac{\omega}{c}\sqrt{1 - \frac{\omega_p^2}{\omega^2}}. \tag{5.134}$$

Therefore, k is purely imaginary, which means that **waves cannot propagate, if** $\omega < \omega_p$. In that case, equation (5.119) gives:

$$\tilde{\mathbf{E}}(z,t) = \tilde{\mathbf{E}}_0 \mathrm{e}^{-k''z}\mathrm{e}^{-\mathrm{i}\omega t}, \tag{5.135}$$

with $k'' = |k|$, which shows that the waves are decaying standing waves. Since $k' = 0$, the time–averaged Poynting vector (5.121) is zero. This shows that a wave with frequency $\omega < \omega_p$ incident on a plasma is completely **reflected** (rather than absorbed).
The reason why the wave does not propagate in the medium for $\omega < \omega_p$ is the same as the reason for which there is no electric field in a perfect conductor. As long as the frequency is small enough (meaning smaller than ω_p), the free electrons in the plasma move under the electric force due to the external field fast enough that they can produce a field which cancels the external field. When the external field varies too fast ($\omega > \omega_p$), the electrons cannot keep up with its variations and the field they produce does not cancel the external field. Therefore the field in the plasma is non zero.

The *group velocity* of the wave, that is to say the velocity at which energy propagates, is given by $v_g = d\omega/dk$. Using equation (5.133), we obtain $v_g = kc^2/\omega$. Then equation (5.134) yields:

$$\boxed{v_g = c\sqrt{1 - \frac{\omega_p^2}{\omega^2}}.} \tag{5.136}$$

Of course, this expression is meaningful only for frequencies larger than ω_p for which waves can propagate. Note that, as expected, $v_g < c$. The phase speed is given by:

$$v_\varphi = \frac{c}{n} = \frac{c}{\sqrt{\epsilon_r}} = \frac{c}{\sqrt{1 - \frac{\omega_p^2}{\omega^2}}}. \tag{5.137}$$

We note that $v_g v_\varphi = c^2$.

Here we have described the properties of the plasma with a dielectric constant ϵ, so that we have used equation (5.114):

$$\boldsymbol{\nabla} \times \tilde{\mathbf{B}} = \mu_0 \epsilon \frac{\partial \tilde{\mathbf{E}}}{\partial t}. \tag{5.138}$$

If instead we describe the plasma as a conductor with a complex conductivity $\tilde{\sigma}$, then we have to use the following equation:

$$\boldsymbol{\nabla} \times \tilde{\mathbf{B}} = \mu_0 \left(\tilde{\mathbf{J}}_f + \epsilon_0 \frac{\partial \tilde{\mathbf{E}}}{\partial t} \right), \tag{5.139}$$

with $\tilde{\mathbf{J}}_f = \tilde{\sigma} \tilde{\mathbf{E}}$. As equations (5.138) and (5.139) have to be identical, this yields:

$$\tilde{\sigma} = \mathrm{i}\omega\epsilon_0 \left(1 - \epsilon_r\right) = \mathrm{i}\epsilon_0 \frac{\omega_p^2}{\omega} = \frac{\mathrm{i}ne^2}{\omega m}. \tag{5.140}$$

Note that this expression of $\tilde{\sigma}$ could also be obtained by writing $\tilde{\mathbf{J}}_f = -en\dot{\tilde{\mathbf{x}}}$. From equation (5.97), we see that $\dot{\tilde{\mathbf{x}}} = -\mathrm{i}\omega\tilde{\mathbf{x}} = -[\mathrm{i}e/(m\omega)]\tilde{\mathbf{E}}$, so that $\tilde{\mathbf{J}}_f = [\mathrm{i}ne^2/(m\omega)]\tilde{\mathbf{E}}$.
Since $\tilde{\sigma}$ is purely imaginary, the current is out of phase by 90° with the electric field. Therefore the power per unit volume dissipated by Ohmic heating and averaged over a period, $\langle P \rangle = \langle \mathbf{J}_f \cdot \mathbf{E} \rangle$, is zero: there is no transfer of energy between the wave and the plasma over a period.

Therefore, at high frequencies, there is no loss of wave energy due to ohmic heating, and the medium acts like a dielectric.

Plasma dominated by collisions:

When $\omega \ll \gamma$, we have to retain the damping term. We assume that the damping coefficient is the same for all the electrons. With $\omega_j = 0$, the dielectric constant given by equation (5.108) can then be approximated by:

$$\tilde{\epsilon}_r = 1 + \frac{\mathrm{i}ne^2}{\epsilon_0 m \gamma \omega}. \tag{5.141}$$

The same calculation as in the case of a collisionless plasma then gives:

$$\tilde{\sigma} = \mathrm{i}\omega\epsilon_0 \left(1 - \tilde{\epsilon}_r\right) = \frac{ne^2}{m\gamma}. \tag{5.142}$$

Here the conductivity is real, and $\langle P \rangle = \langle \mathbf{J}_f \cdot \mathbf{E} \rangle$ is non zero.

Therefore, at low frequencies, there is loss of wave energy due to ohmic heating, and the medium acts like an ohmic conductor.
The dispersion relation (5.117) yields:

$$\tilde{k}^2 = \mu_0 \epsilon_0 \omega^2 \left(1 + \frac{\mathrm{i}\sigma}{\epsilon_0 \omega}\right) = \mu_0 \epsilon_0 \omega^2 + \mathrm{i}\mu_0 \sigma \omega, \tag{5.143}$$

which is identical to the dispersion relation (5.67) derived in conductors with $\mu = \mu_0$ and $\epsilon = \epsilon_0$.

Chapter 6

Electromagnetism and special relativity

The material contained in this chapter is non–examinable.

This chapter is a very brief introduction to relativistic electrodynamics. The change of electric and magnetic fields under Lorentz transformations in simple cases are derived. Electrodynamics in tensor notation and the motion of relativistic particles in electromagnetic fields will be studied in third year.

6.1 Einstein's postulates

We have seen in previous chapters that electric and magnetic fields propagating in vacuum satisfy the wave equation: $\boldsymbol{\nabla}^2\mathbf{E}-(\partial^2\mathbf{E}/\partial t^2)/c^2=\mathbf{0}$. Let us assume this equation is satisfied in an inertial frame (F) with coordinates (x,y,z,t). If (F') is another inertial frame with coordinates (x',y',z',t') and moving with (constant) velocity v along the x–direction with respect to (F), then the Galilean transformation gives $x'=x-vt$, $y'=y$, $z'=z$ and $t'=t$. The wave equation above is then transformed into:

$$\boldsymbol{\nabla}'^2\mathbf{E}-\left(\frac{\partial^2}{c^2\partial t'^2}-\frac{2v}{c}\frac{\partial^2}{c\partial t'\partial x'}+\frac{v^2}{c^2}\frac{\partial^2}{\partial x'^2}\right)\mathbf{E}=\mathbf{0}.$$

This is a wave equation only to zeroth order in v/c. The wave equation describing the propagation of electromagnetic fields is therefore not form–invariant under a Galilean transformation. Likewise, applying a Galilean transformation to Maxwell's equations results in a set of equations which are not self–consistent.
It is to solve this problem that the Lorentz transformation was first introduced, in some preliminary form by Voigt in 1887 and in its final form by Lorentz in 1899. In 1905, Einstein published the theory of special relativity in a paper entitled *On the electrodynamics of moving bodies*, in which he formulated two postulates:

- The laws of physics are the same in all inertial frames of reference,

- The speed of light in vacuum has the same value c in all inertial frames. In other words, the speed of light is independent of the motion of its source.

Because electromagnetic waves have the same speed c in both frames (F) and (F'), the position of a point on the wavefront should satisfy $c^2t^2 - x^2 - y^2 - z^2 = 0$ in (F) and $c^2t'^2 - x'^2 - y'^2 - z'^2 = 0$ in (F'). Lorentz transformation is obtained when seeking a transformation between the sets of coordinates which leaves $c^2t^2 - x^2 - y^2 - z^2$ form–invariant.

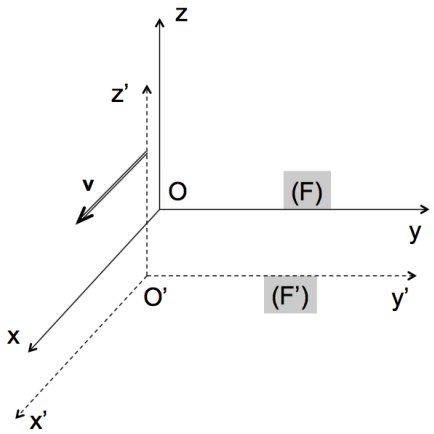
The special theory of relativity also explains why the emf obtained when moving a wire loop in a magnetic field is the same as when changing the magnetic field while keeping the loop at rest. When the loop moves relative to the magnetic field, a magnetic force is exerted on the charges in the loop and this results in an emf. For an observer moving with the loop though, it is the magnetic field which is changing. There is no magnetic force in that case, and the emf results from the electric field induced by the changing magnetic field. Therefore, what is viewed as a magnetic effect by an observer is an electric effect for another observer. As we will see below, magnetic and electric fields are transformed into each other when going from one inertial frame to another.

Although the theory of special relativity has its origin in electromagnetism, it applies to all physical phenomena and all forms of interactions. The one exception are large–scale gravitational interactions, which require the theory of general relativity and invariance under *any* mathematical coordinates transformation.

6.2 Review: Lorentz transformations

6.2.1 Transformation of coordinates

We define an *event* E as something that happens at a particular location (x, y, z) at a particular time t in an inertial frame (F). We consider another inertial frame (F') which has a (constant) velocity v along the x–axis with respect to (F). Lorentz transformations give the coordinates (t', x', y', z') of the event E in (F') to all orders in v/c:



with:

$$\boxed{\begin{aligned} x' &= \gamma\,(x - vt)\,, \\ y' &= y, \\ z' &= z, \\ t' &= \gamma\left(t - \frac{v}{c^2}x\right), \end{aligned}} \qquad \begin{aligned} &(6.1)\\ &(6.2)\\ &(6.3)\\ &(6.4) \end{aligned}$$

$$\gamma \equiv \frac{1}{\sqrt{1 - v^2/c^2}} \geq 1.$$

6.2.2 Time dilatation

We consider two events E_1 and E_2 that occur at the same point (x, y, z) in space as measured in the frame (F). The time interval between these events, as measured by an observer at rest in this same frame, is Δt_0. This time interval between two events occurring at the same point in space is called *proper time*. Using equation (6.4), we can calculate that the time interval between the two events as measured by an observer at rest in the frame (F') is:

$$\boxed{\Delta t' = \gamma \Delta t_0}. \tag{6.5}$$

This shows that the time interval is longer in (F') than in the frame where the two events occur at the same place. This can be expressed by saying that **moving clock run slow**. This is called **time dilatation**.

6.2.3 Lorentz contraction

We consider a stick which is at rest in the frame (F') and lies along the x'–axis. Its rest length, which is also called *proper length*, is $\Delta x_0 = |x'_2 - x'_1|$, where x'_1 and x'_2 are the coordinates of the ends of the stick. An observer at rest in (F) will measure the length Δx of the stick by recording the positions x_1 and x_2 of the ends of the stick *at the same time* in the frame (F). Equation (6.1) then yields:

$$\boxed{\Delta x = \frac{\Delta x_0}{\gamma}}. \tag{6.6}$$

If the stick lies along the y' or z' directions, its length is the same in (F) and (F'), as can be seen from equations (6.2) and (6.3). This shows that the length of an object *along the direction of motion* is shorter in (F) than in the frame where it is at rest. This can be expressed by saying that **moving objects are shortened**. This is called **Lorentz contraction**.

6.2.4 Velocity addition

We consider a particle that moves a distance dx in a time interval dt in the frame (F). The velocity of the particle as measured in (F) is therefore $u_x = dx/dt$. From equations (6.1) and (6.4), this corresponds to the particle having moved a distance $dx' = \gamma(dx - vdt)$ during a time interval $dt' = \gamma(dt - vdx/c^2)$ as measured in (F'), with $\gamma = 1/\sqrt{1 - v^2/c^2}$. The velocity in (F') is then:

$$u'_x = \frac{dx'}{dt'} = \frac{dx - vdt}{dt - vdx/c^2} = \frac{u_x - v}{1 - u_x v/c^2}. \tag{6.7}$$

If the particle moves a distance dy in a time interval dt in (F), the velocity of the particle in (F) is $u_y = dy/dt$. From equations (6.2) and (6.4), the velocity in (F') is then:

$$u'_y = \frac{dy'}{dt'} = \frac{dy}{\gamma\,(dt - vdx/c^2)} = \frac{u_y}{\gamma\,(1 - u_x v/c^2)}. \tag{6.8}$$

6.3 Relativistic electrodynamics

This section borrows from "Electricity and Magnetism", by Edward M. Purcell and David J. Morin.

To calculate how the electromagnetic field transforms when measured in different inertial frames moving with respect to each other, we need to consider distributions of charges that are either at rest or moving. We then start by discussing the fact that electric charge does not depend on its motion.

6.3.1 Invariance of electric charge

Because of the difficulty in measuring a charge in motion through the force it exerts on another charge, we define the electric charge Q in some volume $\mathcal{V}$ as being the flux of the electric field $\mathbf{E}$ through a surface Σ enclosing this volume multiplied by ϵ_0:

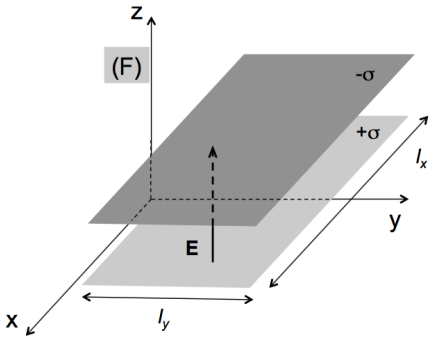
$$Q = \epsilon_0 \oiint_{\Sigma} \mathbf{E} \cdot d\mathbf{\Sigma}. \tag{6.9}$$

The surface Σ is fixed in a frame (F), and $\mathbf{E}$ is measured at time t and position (x, y, z) in (F) by the force on a test charge which is at rest in (F). It has been found experimentally that the integral above does not depend on the choice of the surface Σ delimiting the volume $\mathcal{V}$, whether Q is in motion or not. That is to say, Gauss's law is still valid when the charge carriers are in motion.

It has also been shown experimentally that the integral in equation (6.9) is the same in any inertial frame. In other words, Σ can be at rest in (F) or in any other inertial frame (F'), as long as it encloses the volume in which the charges are located and the force is measured in the frame in which the surface is at rest, Q defined by equation (6.9) is the same. This can also be expressed by saying that the charge Q does not depend on the motion of the charge carriers. This is called the **invariance of charge**.

6.3.2 Transformation of the electric field

We note $\mathbf{E}$ the electric field produced by some charge distribution and measured in an inertial frame (F) and $\mathbf{E}'$ the field produced by the same charge distribution but measured in another inertial frame (F').



We start with a simple case where the electric field is produced by two sheets of charge of uniform density $+\sigma$ and $-\sigma$ which are at rest in (F). If the separation of the sheets is small compared with their dimensions, then the electric field between them can be considered to be uniform and is equal to $\mathbf{E} = \sigma\hat{\mathbf{z}}/\epsilon_0$, as measured in (F).

We now consider an inertial frame (F') that moves toward the left with a (constant) velocity $\mathbf{v} = -v\hat{\mathbf{y}}$ relative to (F). To an observer in (F'), the dimensions of the sheets are $l'_x = l_x$ and $l'_y = l_y/\gamma$, with $\gamma = 1/\sqrt{1 - v^2/c^2}$. Since the total charge $Q = \sigma l_x l_y$ is invariant, the charge density measured from (F') is $\sigma' = Q/(l'_x l'_y)$, that is to say:

$$\boxed{\sigma' = \gamma\sigma}. \tag{6.10}$$

Gauss's law still applies in (F'), and because of the symmetries the field is still perpendicular to the sheets in this frame. Therefore, we obtain $\mathbf{E}' = \sigma'\hat{\mathbf{z}}/\epsilon_0$, which yields $\mathbf{E}' = \gamma\mathbf{E}$. Note that here the field is *perpendicular* to the direction of motion.
Let us now consider the case where the sheets are perpendicular to the y–axis, that is to say in the (x, z)–plane, so that the electric field is in the y–direction, *parallel* to the direction of motion. In that case, both l_x and l_z are unchanged when going from the frame (F) to the frame (F'). Only the separation of the sheets is contracted. Therefore, the surface charge density measured in (F') is the same as that measured in (F), $\sigma' = \sigma$, which yields $\mathbf{E}' = \mathbf{E}$.

The relations above have been obtained for a very simple charge distribution. However, the field at a given point contains all the information which is relevant to the theory of electromagnetism. Once the field is given in some inertial frame, we ought to be able to predict how it transforms in other frames without having to know anything about the details of the source of the field. We can therefore extrapolate from the above calculation that, if we note $E_\parallel$ and $E_\perp$ the components of the electric field parallel and perpendicular, respectively, to the velocity $\mathbf{v}$, produced by a charge distribution *stationary* in (F), and measured in (F), then the field measured in (F') has components $E'_\parallel$ and $E'_\perp$ parallel and perpendicular to $\mathbf{v}$, respectively, such that:

$$\boxed{E'_\parallel = E_\parallel \ ,} \tag{6.11}$$

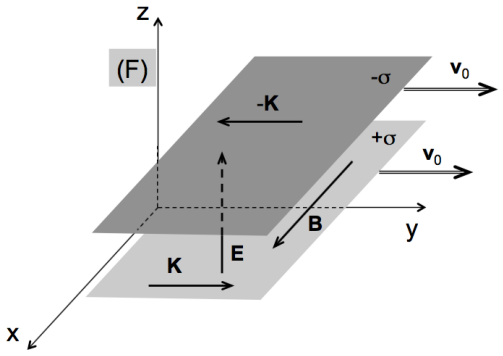
$$\boxed{E'_\perp = \gamma E_\perp \ .} \tag{6.12}$$

This result is valid only if the *charges are at rest in* (F).

6.3.3 Transformation of the electric and magnetic fields

In the section above, we assumed that the charges were at rest in (F), and therefore there was no magnetic field. We are now going to consider a system in which the charges move and produce a magnetic field.

Transformation of E_z and B_x:



We assume that the sheets of surface charge are moving with a constant velocity $\mathbf{v}_0 = v_0\hat{\mathbf{y}}$ relative to an inertial frame (F). The rest frame of the charges is therefore the inertial frame which moves with velocity $\mathbf{v}_0$ relative to (F).

If we draw an imaginary line segment of length L parallel to the x–axis on the sheet with the surface charge density σ, we see that the number of charges that cross this line segment during a time interval dt is $\sigma L v_0 dt$. The current through this line segment is therefore $dI = \sigma L v_0$. As the surface current density is $K = dI/L$, we have a surface current density $\mathbf{K} = \sigma v_0 \hat{\mathbf{y}}$ on the lower sheet and $-\mathbf{K}$ on the upper sheet.
The electric field between the plates measured by an observer in (F) is $\mathbf{E} = E_z\hat{\mathbf{z}}$ with:

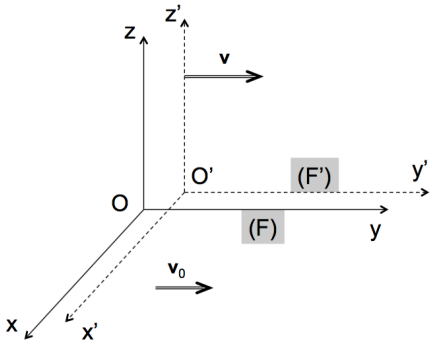
$$E_z = \sigma/\epsilon_0. \tag{6.13}$$

Here σ is the surface charge density measured by an observer at rest in (F) and, according to equation (6.10), we have:

$$\sigma = \gamma_0 \sigma_0, \tag{6.14}$$

where σ_0 is the surface charge density measured in the rest frame of the charges and $\gamma_0 = 1/\sqrt{1 - v_0^2/c^2}$. Using Ampère's law, it can be shown that the magnetic field between the two currents sheets as measured in (F) is $\mathbf{B} = B_x\hat{\mathbf{x}}$ with:

$$B_x = \mu_0 K = \mu_0 \sigma v_0. \tag{6.15}$$



We now consider an inertial frame (F') that moves with a (constant) velocity $\mathbf{v} = v\hat{\mathbf{y}}$ with respect to (F), and we want to calculate the electric field $\mathbf{E}'$ and the magnetic field $\mathbf{B}'$ as measured by an observer at rest in (F').

Using equation (6.7), the velocity $\mathbf{v'}_0$ of the sheets measured in the frame (F') is:

$$\mathbf{v'}_0 = \frac{v_0 - v}{1 - v_0 v/c^2}\,\hat{\mathbf{y}}. \tag{6.16}$$

According to equation (6.10), the surface charge density measured in (F') is $\sigma' = \sigma_0 \gamma'_0$, with $\gamma'_0 = 1/\sqrt{1 - v'^2_0/c^2}$. Using equation (6.14), this yields $\sigma' = \sigma\gamma'_0/\gamma_0$. Using equation (6.16), it can be shown that $\gamma'_0/\gamma_0 = \gamma(1 - v_0 v/c^2)$, with $\gamma = 1/\sqrt{1 - v^2/c^2}$. Therefore:

$$\sigma' = \sigma\gamma(1 - v_0 v/c^2). \tag{6.17}$$

The surface current density in the frame (F') is $\mathbf{K'} = \sigma' v'_0 \hat{\mathbf{y}}$ on the lower sheet and $-\mathbf{K'}$ on the upper sheet (note that v'_0 is either positive or negative). Using equations (6.16) and (6.17), we obtain:

$$\mathbf{K'} = \sigma\gamma\,(v_0 - v)\,\hat{\mathbf{y}}. \tag{6.18}$$

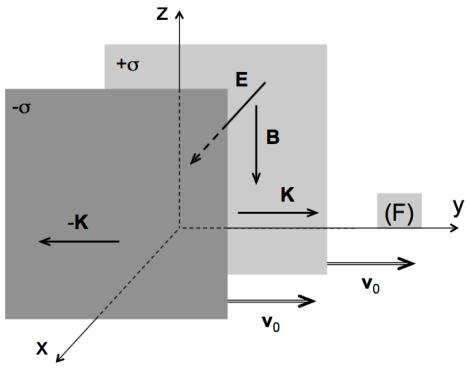
In (F'), we have $\mathbf{E'} = E'_z\hat{\mathbf{z}}$ with $E'_z = \sigma'/\epsilon_0$ and $\mathbf{B'} = B'_x\hat{\mathbf{x}}$ with $B'_x = \mu_0 K'$. This yields:

$$E'_z = \gamma\left(\frac{\sigma}{\epsilon_0} - \frac{v}{\epsilon_0\mu_0 c^2}\mu_0\sigma v_0\right) = \gamma\,(E_z - vB_x)\,, \tag{6.19}$$

$$B'_x = \gamma\left(\mu_0\sigma v_0 - \mu_0\epsilon_0 v\frac{\sigma}{\epsilon_0}\right) = \gamma\left(B_x - \frac{v}{c^2}E_z\right). \tag{6.20}$$

Note that these equations are the same as the Lorentz transformations (6.1) and (6.4) with x and t being replaced by E_z and B_x, respectively.

Transformation of E_x and B_z:



We have obtained transformations for E_z and B_x. To know how E_x and B_z transform, we now consider sheets which are perpendicular to the x–axis, with still $\mathbf{v}_0 = v_0\hat{\mathbf{y}}$ and $\mathbf{v} = v\hat{\mathbf{y}}$.

As the velocity of the sheets is still parallel to the plates, we have the same results as before but with E_z replaced by E_x and B_x replaced by $-B_z$. Therefore, using equations (6.19) and (6.20), we obtain:

$$E'_x = \gamma\,(E_x + vB_z)\,, \tag{6.21}$$

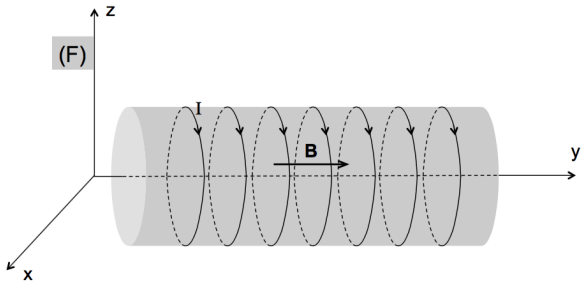
$$B'_z = \gamma\left(B_z + \frac{v}{c^2}E_x\right). \tag{6.22}$$

Transformation of E_y:

If now we put the sheets perpendicular to the y–axis, we have $\mathbf{E} = E_y\hat{\mathbf{y}}$ but there is no magnetic field. From the results of section 6.3.2, we know that the component of the electric field parallel to the motion is unchanged, so that:

$$E'_y = E_y. \tag{6.23}$$

Transformation of B_y:



Finally, to know how B_y transforms, we consider a long solenoid around the y-axis which is at rest in (F). The magnetic field produced by this solenoid and measured in (F) is $\mathbf{B} = B_y\hat{\mathbf{y}}$ with $B_y = \mu_0 nI$, where I is the current in the wire and n is the number of turns of wire per unit length.

If we note L the length of the solenoid in the frame (F) and N the total number of turns of wire, then $n = N/L$. In the frame (F') , the length of the solenoid is contracted and equal to $L' = L/\gamma$ (see eq. [6.6]). Therefore, the number of turns of wire per unit length as measured in (F') is $n' = N/L' = \gamma n$.
If dq is the amount of charge that passes a point at rest on the solenoid during the time interval dt, then the current I measured in (F) is dq/dt. This time interval as measured in (F') is $dt' = \gamma dt$ (see eq. [6.5]). Therefore, the current measured in (F') is $I' = dq/dt' = I/\gamma$. The magnetic field produced by the solenoid and measured in (F') is $\mathbf{B}' = B'_y\hat{\mathbf{y}}$ with $B'_y = \mu_0 n'I' = \mu_0 nI$, so that:

$$B'_y = B_y. \tag{6.24}$$

Transformation rules for E and B:

Here again, although the relations above have been obtained assuming simple charge and current distributions, they should hold whatever the source of the fields are. We can therefore summarise the transformations rules:

$$\boxed{\begin{array}{lll} E'_x = \gamma\left(E_x + vB_z\right), & E'_y = E_y, & E'_z = \gamma\left(E_z - vB_x\right) \\ B'_x = \gamma\left(B_x - \dfrac{v}{c^2}E_z\right), & B'_y = B_y, & B'_z = \gamma\left(B_z + \dfrac{v}{c^2}E_x\right), \end{array}} \tag{6.25, 6.26}$$

where the unprimed quantities are measured in an inertial frame (F) and the primed quantities are measured in another inertial frame (F'), which moves with velocity $\mathbf{v} = v\hat{\mathbf{y}}$ with respect to (F).

These relations can be written in a more compact form by using the following notations:

$$\mathbf{E} = \mathbf{E}_{\parallel} + \mathbf{E}_{\perp},$$
$$\mathbf{B} = \mathbf{B}_{\parallel} + \mathbf{B}_{\perp},$$
$$\mathbf{E}' = \mathbf{E}'_{\parallel} + \mathbf{E}'_{\perp},$$
$$\mathbf{B}' = \mathbf{B}'_{\parallel} + \mathbf{B}'_{\perp},$$

where the subscripts '$\parallel$' and '$\perp$' denote the components of the fields parallel and perpendicular to $\mathbf{v}$, respectively, both in (F) and (F'), with $\mathbf{v}$ being the velocity of (F') with respect to (F). Then equations (6.25) and (6.26) become:

$$\boxed{\begin{aligned} &\mathbf{E}'_{\parallel} = \mathbf{E}_{\parallel}, \qquad \mathbf{E}'_{\perp} = \gamma\left(\mathbf{E}_{\perp} + \mathbf{v}\times\mathbf{B}_{\perp}\right) && (6.27)\\ &\mathbf{B}'_{\parallel} = \mathbf{B}_{\parallel}, \qquad \mathbf{B}'_{\perp} = \gamma\left(\mathbf{B}_{\perp} - \frac{1}{c^2}\mathbf{v}\times\mathbf{E}_{\perp}\right), && (6.28)\end{aligned}}$$

We consider two interesting special cases:

- If $\mathbf{B} = \mathbf{0}$ (at a particular point) in (F), then $\mathbf{B}' = \gamma(v/c^2)\left(-E_z\hat{\mathbf{x}} + E_x\hat{\mathbf{z}}\right) = (v/c^2)\left(-E'_z\hat{\mathbf{x}} + E'_x\hat{\mathbf{z}}\right)$, which can also be written under the form:

$$\boxed{\mathbf{B}' = -\frac{1}{c^2}\mathbf{v}\times\mathbf{E}'}. \tag{6.29}$$

- If $\mathbf{E} = \mathbf{0}$ (at a particular point) in (F), then $\mathbf{E}' = \gamma v\left(B_z\hat{\mathbf{x}} - B_x\hat{\mathbf{z}}\right) = v\left(B'_z\hat{\mathbf{x}} - B'_x\hat{\mathbf{z}}\right)$, which can also be written under the form:

$$\boxed{\mathbf{E}' = \mathbf{v}\times\mathbf{B}'}. \tag{6.30}$$

Chapter 7

Transmission lines and waveguides

In A.C. circuits, current and voltage are assumed not to vary along the leads connecting different components. Any capacitance, inductance, resistance, gain, etc., in the circuit are modelled as idealised *lumped* (discrete) electrical components, instead of being continuously distributed along the circuit. It is assumed that voltage and current are transmitted instantaneously between these components along perfectly conducting wires.

However, in reality, the wires themselves have parasitic, or *stray* impedances. For example, there is a non zero capacitance between any two wires conducting the current in a circuit. There is also a self–inductance in the wires when a varying magnetic field propagates through them, as it induces and emf. These stray impedances, which are distributed along the whole length of the wires, cannot always be modelled as lumped components.

At high frequencies, when the wavelength of the signal propagating along a wire becomes comparable to or smaller than the length of the wire, the wave nature of the signal cannot be ignored, as it results in different phases at different locations in the wire. Said another way, the finite speed with which the signal propagates has to be taken into account at high frequencies, when voltage and current vary on a timescale comparable to or shorter than the time it takes for the signal to propagate along the wire. When this is the case, stray impedances along the wires cannot be ignored as they determine the speed of the signals along the wires, the phase shift between voltage and current and reflections at the end of the wires. Therefore, idealised A. C. circuits cannot be used to transport high frequency signals over large distances. Instead, we use *transmission lines* or *waveguides*.

A transmission line is made of two parallel conductors which have a fixed cross section anywhere along the length of the line and are separated by a dielectric. It is a two–port network that connects an input circuit which generates the signal to an output load. Although electromagnetic fields cannot penetrate far into good conductors, they can be guided across long distances by them.

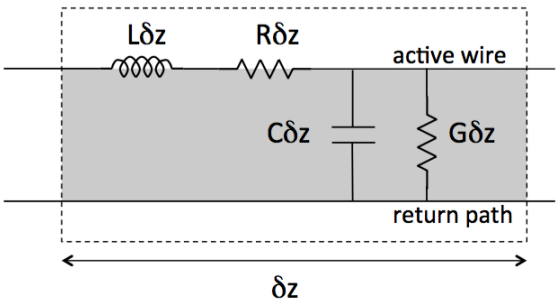
However, at very high frequencies (typically when the wavelength of the signal becomes comparable to the transverse size of the line), there is significant energy loss in a transmission line, due to the conductors, the dielectric and radiation. Therefore, a waveguide, which consists of a hollow metal tube, is preferred.

7.1 Transmission lines

7.1.1 Equivalent circuit of a transmission line

A transmission line consists of a conductor called an *active wire* that carries the (time–varying) voltage / current from the input circuit to the load, and another conductor that acts as a *return path* and is usually Earthed.

- The time–varying current in the conductors produces a magnetic field which flux through the area delimited by the wires (when the circuit is closed) varies with time. This induces an emf in the circuit and a related (self–)inductance. We note L the inductance per unit length.
- As a potential difference is maintained between the two conductors, a charge is induced on them and therefore there is a capacitance between them. We note C the capacitance per unit length.
- The conductors have a finite conductivity, and therefore energy is dissipated. We note R the resistance per unit length associated with this finite conductivity. If the conductors are perfect, their conductivity is infinite and $R = 0$.
- Finally, as the dielectric filling the space between the conductors is not perfectly insulating, there is a current flowing from one conductor to the other. We note G the conductance per unit length of the dielectric. If the dielectric is a perfect insulator, its resistance is infinite and $G = 0$.



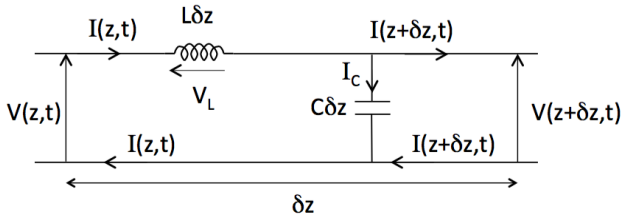
The equivalent circuit of a length δz of a transmission line is represented on the figure. The full transmission line can be modelled as an infinite series of such elementary circuits.

In many situations of interest, energy losses in the conductors and the dielectric are small and can be neglected. Therefore, hereafter, we will only consider *lossless* transmission lines, that is to say $R = 0$ and $G = 0$.

7.1.2 The telegrapher's equations and characteristic impedance

Single–wire transmission lines made of one conductor with the Earth as the return path were used at the end of the nineteenth century for telegraph transmission. The theory of such transmission lines was developed in the 1880s by Oliver Heaviside, who established the so–called *telegrapher's equations* which govern the propagation of voltage and current through the conductors.

The telegrapher's equations:



If $\delta z \ll \lambda$, where λ is the wavelength of the wave propagating through the transmission line, then Kirchoff's laws can be applied for the elementary circuit.

Note that, in a lossless line, the current in the return path is exactly the same as in the active wire.
We have:

$$V(z+\delta z,t) = V(z,t) - V_L = V(z,t) - L\delta z \frac{\partial I(z,t)}{\partial t}, \tag{7.1}$$

$$I(z+\delta z,t) = I(z,t) - I_C = I(z,t) - C\delta z \frac{\partial V(z+\delta z,t)}{\partial t}, \tag{7.2}$$

Using $V(z+\delta z,t) - V(z,t) = [\partial V(z,t)/\partial z]\delta z$, and similarly for I, yields:

$$\frac{\partial V(z,t)}{\partial z} = -L\frac{\partial I(z,t)}{\partial t}, \tag{7.3}$$

$$\frac{\partial I(z,t)}{\partial z} = -C\frac{\partial V(z+\delta z,t)}{\partial t}. \tag{7.4}$$

The term on the right–hand side in equation (7.4) can be written as:

$$\frac{\partial V(z+\delta z,t)}{\partial t} = \frac{\partial}{\partial t}\left(V(z,t) + \frac{\partial V(z,t)}{\partial z}\delta z\right).$$

When $\delta z \to 0$, the second term in the parenthesis tends towards 0, so that equations (7.3) and (7.4) become:

$$\boxed{\frac{\partial V}{\partial z} = -L\frac{\partial I}{\partial t},} \tag{7.5}$$

$$\boxed{\frac{\partial I}{\partial z} = -C\frac{\partial V}{\partial t},} \tag{7.6}$$

where all the quantities are evaluated at z and t. These are the *telegrapher's equations* for a lossless transmission line.

Wave propagation:

Taking the time derivative of equation (7.5) yields:

$$\frac{\partial}{\partial z}\left(\frac{\partial V}{\partial t}\right) = -L\frac{\partial^2 I}{\partial t^2}.$$

By replacing $\partial V/\partial t$ using equation (7.6), we then obtain:

$$\frac{1}{C}\frac{\partial^2 I}{\partial z^2} = L\frac{\partial^2 I}{\partial t^2},$$

and similarly for V. These are *wave equations* for the voltage and current:

$$\boxed{\frac{\partial^2 V}{\partial t^2} = \frac{1}{LC}\frac{\partial^2 V}{\partial z^2},} \tag{7.7}$$

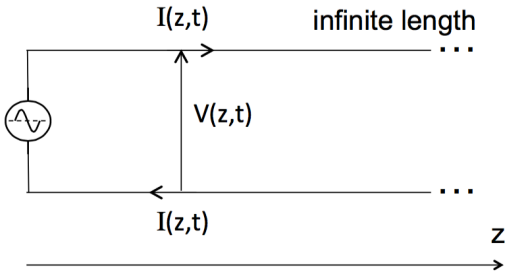
$$\boxed{\frac{\partial^2 I}{\partial t^2} = \frac{1}{LC}\frac{\partial^2 I}{\partial z^2},} \tag{7.8}$$

which show that V and I propagate through the transmission line with the speed:

$$\boxed{v = \frac{1}{\sqrt{LC}}}. \tag{7.9}$$

The fact that the speed at which signals propagate is determined by L and C is not surprising, as the propagation timescale is regulated by the time taken to accumulate charges on the capacitance and to build up currents across the inductance.

Characteristic impedance:



We consider a (imaginary) transmission line with infinite length. A generator delivers a voltage which varies with frequency ω. As the line is infinite, there is only a wave travelling in the plus z–direction, with no reflection.

We note V_+ and I_+ the voltage and the current on the line:

$$\begin{aligned} V_+(z,t) &= V_0 \cos(\omega t - kz), \\ I_+(z,t) &= I_0 \cos(\omega t - kz), \end{aligned}$$

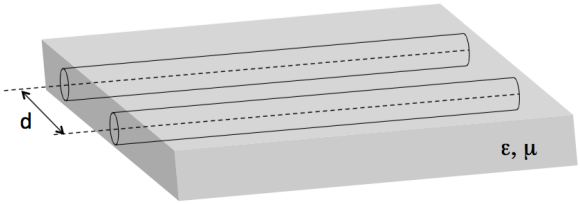
where k is the wavenumber. From equations (7.7) and (7.8), we obtain the dispersion relation $k = \omega\sqrt{LC} = \omega/v$. Substituting the above expressions for V_+ and I_+ in equation (7.5) then yields $kV_0 = L\omega I_0$, that is to say $V_0/I_0 = \sqrt{L/C}$. We define the **characteristic impedance** Z_0 of the line as the ratio V_+/I_+ for a wave travelling along the plus z–direction. Therefore we have:

$$\boxed{Z_0 = \sqrt{\frac{L}{C}}}. \tag{7.10}$$

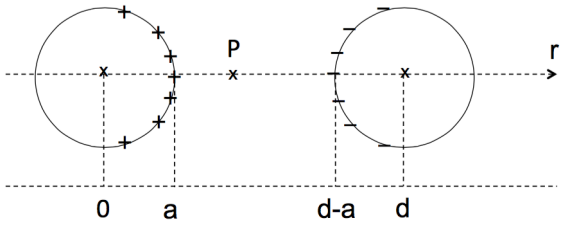
Note that V/I for a wave travelling along the minus z–direction is $-Z_0$. The characteristic impedance is resistive rather than reactive, even when the conductors and dielectric are perfect. From the point of view of the generator, an infinite line transmitting the energy it supplies is equivalent to a resistor $R = Z_0$ dissipating the same energy.

7.1.3 Parallel wire transmission line

Common type of transmission lines include coaxial line, parallel–plate (or strip) line[1] and two–wire line. The type of transmission line to be used depends on the cost, the frequency of the signal, the distance over which is has to be transported, how twisted the line has to be along the way, etc...



Here we study a two–wire transmission line, made of two identical cylindrical conductors of radius a separated by a distance $d \gg a$. The wires are embedded in a dielectric of permittivity ϵ and permeability μ.



The figure on the left shows a cross section of the transmission line. The potential difference which is maintained between the two conductors induces a surface charge density $\pm\sigma$ on them.

The distribution of charges at the surface of the cylinders is not axially symmetric, but this can be neglected as $a \ll d$. Therefore, the electric field produced by each of the cylinder is radial (with respect to cylindrical coordinates) and can be calculated using Gauss's law. The electric field and potential produced by a cylinder of radius a and surface charge density σ at a distance $\rho > a$ from its axis are $E(\rho) = a\sigma/(\epsilon\rho)$ and $V(\rho) = -(a\sigma/\epsilon)\ln\rho + V_0$, where V_0 is a constant.

The potential due to the two conductors at the point P shown on the figure above is then given by:

$$V(r) = -\frac{a\sigma}{\epsilon}\left[\ln r - \ln(d-r)\right] + V'_0 = \frac{a\sigma}{\epsilon}\ln\frac{d-r}{r} + V'_0,$$

for $a < r < d-a$, where V'_0 is a constant. Therefore, the potential difference between the surfaces of the two conductors is:

$$\Delta V \equiv V(a) - V(d-a) = \frac{2a\sigma}{\epsilon}\ln\frac{d-a}{a}.$$

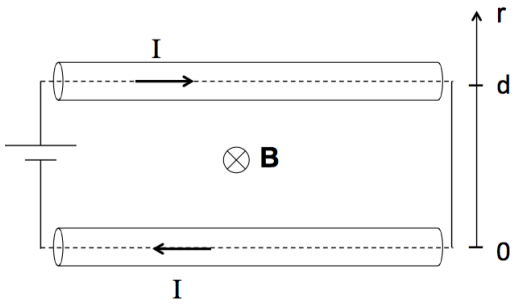
The charge on a length l of the conductors is $\pm Q = \pm 2\pi a l\sigma$, and the corresponding capacitance is $C_l = Q/\Delta V$. Therefore, the capacitance per unit length ($l = 1$) is:

$$\boxed{C = \frac{\pi\epsilon}{\ln(d/a)}}, \tag{7.11}$$

where we have used $d \gg a$.

[1] See Problem Set 4

To calculate the (self–)inductance, we make a closed circuit by joining the two conductors. Using Ampère's law, we can calculate that the magnetic field produced by a current I flowing along a cylinder of radius a is $B = \mu I/(2\pi\rho)$, where $\rho > a$ is the distance from the axis of the cylinder. This field is azimuthal.



The magnetic field due to the two cylinders at a distance r between a and $d-a$ from the axis of one of the cylinders and in the plane containing the two axes is:

$$B = \frac{\mu I}{2\pi}\left(\frac{1}{r} + \frac{1}{d-r}\right).$$

The flux of B through the surface delimited by the closed circuit is:

$$\phi = l\int_a^{d-a} B dr = \frac{\mu I l}{\pi}\ln\frac{d-a}{a},$$

where l is the width of the circuit. The (self–)inductance of the circuit is $L_l = \phi/I$. Therefore, the inductance per unit length ($l = 1$) is:

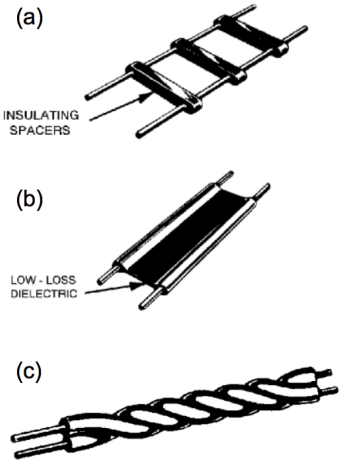
$$\boxed{L = \frac{\mu}{\pi}\ln(d/a)}, \tag{7.12}$$

where we have used $d \gg a$.

The speed of signals propagating through a two–wire transmission line is given by equation (7.9) and is therefore equal to:

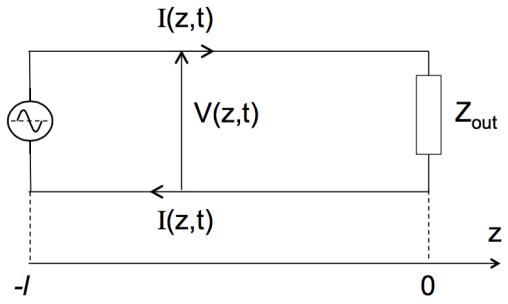
$$v = \frac{1}{\sqrt{\mu\epsilon}}, \tag{7.13}$$

which is the speed of light in the dielectric in which the transmission line is embedded. This is actually a very general result: *for any lossless transmission line, whatever the shape and the size of the conductors, the speed at which signals propagate is always equal to the speed of light in the material in which the conductors are embedded.*



Two–wire transmission lines can have different geometry depending on what they are used for. Parallel conductors separated by insulating spacers (a) are often used as power lines or rural telephone lines, whereas parallel conductors separated by a ribbon made of a low–loss dielectric (b) are used for connecting televisions to their antennas. However, these types of transmission lines have radiation losses due to the radiating fields produced by time–varying currents in the conductors. Losses can be reduced by using a twisted pair which consists of two conducting wires twisted together (c). Twisted lines are used for telephone lines of Local Area Networks (LAN).

7.1.4 Reflection by the load



We consider a transmission line with length l and characteristic impedance Z_0 which is terminated by a load with impedance Z_{out}. A generator delivers a voltage which varies with frequency ω.

In general, at least part of the signal delivered by the generator is reflected by the load, so that the general expressions for the voltage and current are, using complex notations (where complex quantities are denoted with a tilde):

$$\tilde{V}(z,t) = V_+ e^{j(\omega t - kz)} + \tilde{V}_- e^{j(\omega t + kz)}, \tag{7.14}$$

$$\tilde{I}(z,t) = I_+ e^{j(\omega t - kz)} + \tilde{I}_- e^{j(\omega t + kz)}, \tag{7.15}$$

where V_+ and I_+ are the amplitude of the incident waves whereas $\tilde{V}_-$ and $\tilde{I}_-$ are the amplitude of the reflected waves. Here V_+ and I_+ are real but the amplitude of the reflected waves may be complex if reflection induces a phase shift. We have $k = \omega/v$. As has been shown at the end of section 7.1.2, $I_+ = V_+/Z_0$ and $\tilde{I}_- = -\tilde{V}_-/Z_0$. Therefore, we obtain:

$$\tilde{V}(z,t) = V_0 e^{j(\omega t - kz)} + \tilde{\Gamma} V_0 e^{j(\omega t + kz)}, \tag{7.16}$$

$$\tilde{I}(z,t) = \frac{V_0}{Z_0} e^{j(\omega t - kz)} - \frac{\tilde{\Gamma} V_0}{Z_0} e^{j(\omega t + kz)}, \tag{7.17}$$

where we have defined $V_0 \equiv V_+$ and the reflection coefficient $\tilde{\Gamma} \equiv \tilde{V}_-/V_+$. In general, $\tilde{\Gamma}$ is complex.

Reflection coefficient:

At the end of the line, that is to say at $z = 0$, we must have $\tilde{V}/\tilde{I} = Z_{\text{out}}$, which yields:

$$Z_0 \frac{1+\tilde{\Gamma}}{1-\tilde{\Gamma}} = Z_{\text{out}}.$$

The reflection coefficient is therefore given by:

$$\boxed{\tilde{\Gamma} = \frac{Z_{\text{out}} - Z_0}{Z_{\text{out}} + Z_0}}. \tag{7.18}$$

The characteristic impedance Z_0 is real but Z_{out} is in general complex.

Input impedance:

The *input impedance* Z_{in} of the line is the impedance seen at the beginning of the line, that is to say at $z = -l$:

$$Z_{\text{in}} = \left.\frac{\tilde{V}}{\tilde{I}}\right|_{z=-l}. \tag{7.19}$$

Equations (7.16) and 7.17 yield:

$$Z_{\text{in}} = Z_0 \frac{\mathrm{e}^{\mathrm{j}kl} + \tilde{\Gamma}\mathrm{e}^{-\mathrm{j}kl}}{\mathrm{e}^{\mathrm{j}kl} - \tilde{\Gamma}\mathrm{e}^{-\mathrm{j}kl}}.$$

Using equation (7.18), we then obtain:

$$\boxed{Z_{\text{in}} = Z_0 \frac{Z_{\text{out}}\cos(kl) + \mathrm{j}Z_0 \sin(kl)}{Z_0 \cos(kl) + \mathrm{j}Z_{\text{out}} \sin(kl)}}, \tag{7.20}$$

which in general is a complex number. If $kl = n\pi$, where n is an integer, which is equivalent to $l = n\lambda/2$, we have $Z_{\text{in}} = Z_{\text{out}}$. In that case the transmission line acts as a simple conducting wire.

7.1.5 Impedance matching and quarter–wave transformer

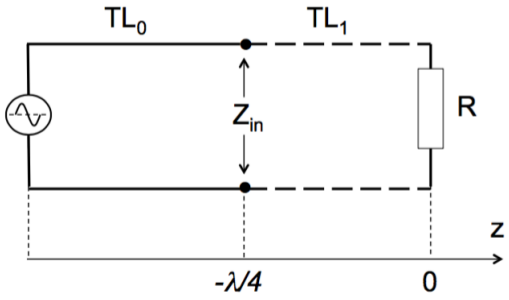
There is no reflection ($\tilde{\Gamma} = 0$) when $Z_{\text{out}} = Z_0$. In that case, all the power delivered by the generator is carried to the load by the wave travelling in the plus z–direction and is absorbed by the load. This is called *impedance matching*.

When the line is matched, $Z_{\text{in}} = Z_0 = Z_{\text{out}}$.

By contrast, the wave is completely reflected and no power is transmitted to the load when $|V_-| = |V_+|$, that it to say $|\Gamma| = 1$. Writing $Z_{\text{out}} = \alpha + \mathrm{j}\beta$, equation (7.18) then implies $(\alpha - Z_0)^2 = (\alpha + Z_0)^2$, and therefore $\alpha = 0$. For the wave to be completely reflected, Z_{out} then needs to be purely imaginary, or in other words the line has to be terminated by a purely reactive impedance.

In many situations, it is desirable to achieve impedance matching, and this can be done by using a *quarter–wave transformer*.

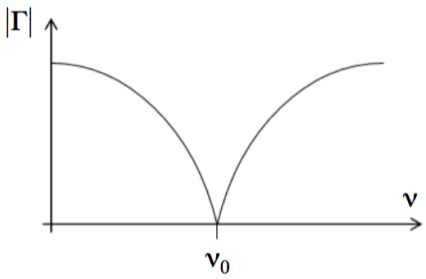
Let us consider, for example, the situation where a broadcast station uses a lossless transmission line TL_0 with characteristic impedance Z_0 between the transmitter and a tower mounted antenna. The antenna impedance is R (purely resistive). In general, some power is reflected back to the transmitter.



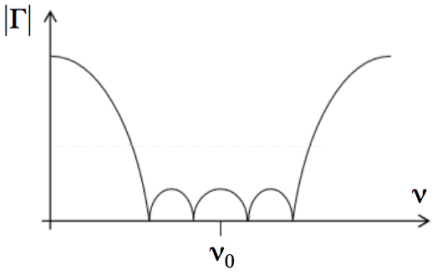
To match the antenna to TL_0, that is to say cancel any reflection at the end of TL_0, we add between them another lossless transmission line TL_1 with characteristic impedance Z_1 and length $\lambda/4$, where λ is the wavelength of the signal in TL_1.

We now calculate the value of Z_1 that gives impedance matching. At the beginning of TL_1, the input impedance is given by equation (7.20), in which we replace Z_0 by Z_1 and Z_{out} by R. Using $kl = (2\pi/\lambda)\lambda/4 = \pi/2$, we obtain $Z_{\text{in}} = Z_1^2/R$.

The load as seen by TL_0 is now Z_{in}. Therefore, to match the antenna to TL_0, we need $Z_0 = Z_{\text{in}}$, which yields $\boxed{Z_1 = \sqrt{Z_0 R}}$.



Since TL_1 is only a quarter–wavelength at a single frequency ν_0, it gives an exact matching only at this frequency, and an approximate matching only in a narrow–band of frequencies around this value. This is illustrated on the figure in the left, which shows the reflection coefficient $|\Gamma|$ at the end of TL_0 as a function of frequency ν.



To obtain a broad–band matching, we may use a *cascade* of $\lambda/4$ line sections with a characteristic impedance that varies gradually from one section to the next. Note however that the reflection coefficient is not exactly zero for all frequencies in the band, as illustrated on the figure (which corresponds to 4 quarter–wave line sections).

Another way to match the circuit in a narrow–band of frequencies is to insert a short–circuited stub of line in parallel at some point on the line[2].

[2]See Problem set 4

7.2 Waveguides and resonant cavities

The material contained in this section is non–examinable.

A waveguide is a hollow metallic cylinder with no end surfaces. When end surfaces are present, the cylinder is called a cavity. We consider a cylinder with a perfectly conducting surface and filled with a non–dissipative medium. In this situation, there is no energy loss as the waves propagate along the axis of the cylinder.

7.2.1 Generalities about waveguides

Adopting complex notations, the electric and magnetic fields inside the cylinder satisfy Maxwell's equations:

$$\boldsymbol{\nabla}\cdot\tilde{\mathbf{E}} = 0, \tag{7.21}$$

$$\boldsymbol{\nabla}\cdot\tilde{\mathbf{B}} = 0, \tag{7.22}$$

$$\boldsymbol{\nabla}\times\tilde{\mathbf{E}} = -\frac{\partial\tilde{\mathbf{B}}}{\partial t}, \tag{7.23}$$

$$\boldsymbol{\nabla}\times\tilde{\mathbf{B}} = \mu\epsilon\frac{\partial\tilde{\mathbf{E}}}{\partial t}, \tag{7.24}$$

where ϵ and μ are the permittivity and permeability, respectively, of the medium inside the cylinder.

As we have seen before, these equations lead to the wave equation satisfied by both $\tilde{\mathbf{E}}$ and $\tilde{\mathbf{B}}$:

$$\nabla^2\tilde{\mathbf{E}} - \mu\epsilon\frac{\partial^2\tilde{\mathbf{E}}}{\partial t^2} = 0, \tag{7.25}$$

$$\nabla^2\tilde{\mathbf{B}} - \mu\epsilon\frac{\partial^2\tilde{\mathbf{B}}}{\partial t^2} = 0. \tag{7.26}$$

We use a Cartesian coordinate system with the z–axis being the axis of the cylinder. The unit vectors in the x–, y– and z–directions are denoted $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$ and $\hat{\mathbf{z}}$, respectively. We look for waves with frequency[3] ω that propagate along the z–axis:

$$\tilde{\mathbf{E}}(x,y,z,t) = \tilde{\mathbf{E}}(x,y)\mathrm{e}^{\mathrm{i}(kz-\omega t)}, \quad \tilde{\mathbf{B}}(x,y,z,t) = \tilde{\mathbf{B}}(x,y)\mathrm{e}^{\mathrm{i}(kz-\omega t)}. \tag{7.27}$$

We will limit ourselves to situations where the wavenumber k is real.
We separate the fields into longitudinal components $\tilde{E}_z$, $\tilde{B}_z$ and transverse components $\tilde{\mathbf{E}}_t = \tilde{E}_x\hat{\mathbf{x}} + \tilde{E}_y\hat{\mathbf{y}}$, $\tilde{\mathbf{B}}_t = \tilde{B}_x\hat{\mathbf{x}} + \tilde{B}_y\hat{\mathbf{y}}$.

[3] Waves with a more general time–dependence can always be writen as a sum over monochromatic waves. Let us consider a function $F(\mathbf{r},t)$, with $\mathbf{r}$ the position vector, that satifies the wave equation $\nabla^2F - \mu\epsilon\partial^2F/\partial t^2$. $F(\mathbf{r},t)$ has the Fourier integral representation $F(\mathbf{r},t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}F(\mathbf{r},\omega)\mathrm{e}^{-\mathrm{i}\omega t}d\omega$ with the inverse transformation $F(\mathbf{r},\omega) = \int_{-\infty}^{\infty}F(\mathbf{r},t)\mathrm{e}^{\mathrm{i}\omega t}dt$. Then the Fourier transform $F(\mathbf{r},\omega)$ satisfies the wave equation $(\nabla^2+k^2)F(\mathbf{r},\omega) = 0$ with $k = \sqrt{\mu\epsilon}\omega$. We can therefore solve for $F(\mathbf{r},\omega)$ and obtain $F(\mathbf{r},t)$ using the Fourier integral representation.

Equation (7.23) yields:

$$\frac{\partial \tilde{E}_z}{\partial y} - \mathrm{i}k\tilde{E}_y = \mathrm{i}\omega\tilde{B}_x, \quad \mathrm{i}k\tilde{E}_x - \frac{\partial \tilde{E}_z}{\partial x} = \mathrm{i}\omega\tilde{B}_y, \quad \frac{\partial \tilde{E}_y}{\partial x} - \frac{\partial \tilde{E}_x}{\partial y} = \mathrm{i}\omega\tilde{B}_z.$$

The first two equations can be recast under the form:

$$\mathrm{i}k\tilde{\mathbf{E}}_t + \mathrm{i}\omega\hat{\mathbf{z}}\times\tilde{\mathbf{B}}_t = \boldsymbol{\nabla}_t\tilde{E}_z, \tag{7.28}$$

where $\boldsymbol{\nabla}_t$ is the transverse component of the $\boldsymbol{\nabla}$ operator.
Similarly, equation (7.24) yields:

$$\mathrm{i}k\tilde{\mathbf{B}}_t - \mathrm{i}\mu\epsilon\omega\hat{\mathbf{z}}\times\tilde{\mathbf{E}}_t = \boldsymbol{\nabla}_t\tilde{B}_z. \tag{7.29}$$

From this equation we can write::

$$\mathrm{i}\omega k\hat{\mathbf{z}}\times\tilde{\mathbf{B}}_t = -\mathrm{i}\mu\epsilon\omega^2\tilde{\mathbf{E}}_t + \omega\hat{\mathbf{z}}\times\boldsymbol{\nabla}_t\tilde{B}_z,$$

which we substitute into equation (7.28) to obtain:

$$\boxed{\tilde{\mathbf{E}}_t = \frac{\mathrm{i}}{\mu\epsilon\omega^2 - k^2}\left(k\boldsymbol{\nabla}_t\tilde{E}_z - \omega\hat{\mathbf{z}}\times\boldsymbol{\nabla}_t\tilde{B}_z\right).} \tag{7.30}$$

Similarly, substituting $\hat{\mathbf{z}}\times\tilde{\mathbf{E}}_t$ obtained from equation (7.28) into equation (7.29) yields:

$$\boxed{\tilde{\mathbf{B}}_t = \frac{\mathrm{i}}{\mu\epsilon\omega^2 - k^2}\left(k\boldsymbol{\nabla}_t\tilde{B}_z + \mu\epsilon\omega\hat{\mathbf{z}}\times\boldsymbol{\nabla}_t\tilde{E}_z\right).} \tag{7.31}$$

Equations (7.30) and (7.31) show that $\tilde{\mathbf{E}}$ *and* $\tilde{\mathbf{B}}$ *are entirely determined once the longitudinal components* $\tilde{E}_z$ *and* $\tilde{B}_z$ *are known.*

To calculate $\tilde{E}_z$ and $\tilde{B}_z$ themselves, we solve the z–component of the wave equations (7.25) and (7.26) that can be written under the form:

$$\boxed{\left[\boldsymbol{\nabla}_t^2 + \mu\epsilon\omega^2 - k^2\right]\tilde{E}_z = 0,} \tag{7.32}$$

$$\boxed{\left[\boldsymbol{\nabla}_t^2 + \mu\epsilon\omega^2 - k^2\right]\tilde{B}_z = 0,} \tag{7.33}$$

where $\nabla_t^2 = \partial^2/\partial x^2 + \partial^2/\partial y^2$ is the transverse part of he Laplacian operator. In addition, $\tilde{\mathbf{E}}$ and $\tilde{\mathbf{B}}$ satisfy the boundary conditions at the surface of the waveguide. Since the surface is a conductor, there is no electric field inside. If the magnetic field inside the surface were zero to start with, then it remains so as implied by $\boldsymbol{\nabla}\times\tilde{\mathbf{E}} = -\partial\tilde{B}/\partial t$. The boundary conditions (5.8) and (5.9) then yield, just outside the surface:

$$\boxed{\tilde{\mathbf{E}}^{\parallel} = \mathbf{0},} \tag{7.34}$$

$$\boxed{\tilde{\mathbf{B}}^{\perp} = \mathbf{0}.} \tag{7.35}$$

Equation (7.32) with the boundary condition (7.34) or, equivalently, equation (7.33) with the boundary condition (7.35), constitutes an *eigenvalue problem.* The values of

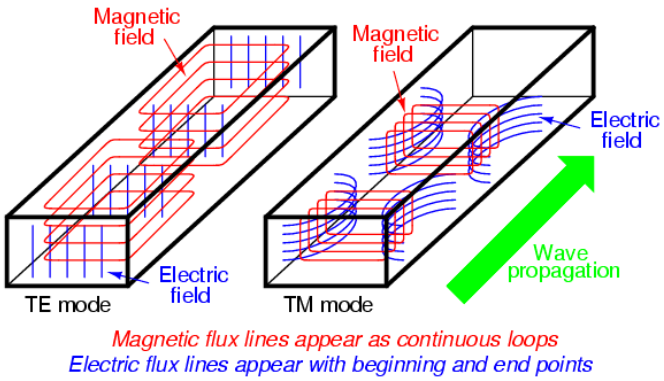
$\mu\epsilon\omega^2 - k^2$ that are permitted form a spectrum of eigenvalues, and the associated solutions $\tilde{E}_z$ (or, equivalently, $\tilde{B}_z$) form an orthogonal set of waveguide *modes*.

It is convenient to consider the following modes:

- Transverse Electric (TE) wave, which has $\tilde{E}_z = 0$ and $\tilde{B}_z \neq 0$,
- Transverse Magnetic (TM) wave, which has $\tilde{B}_z = 0$ and $\tilde{E}_z \neq 0$,
- Transverse Electric and Magnetic (TEM) wave, which has both $\tilde{E}_z = 0$ and $\tilde{B}_z = 0$.

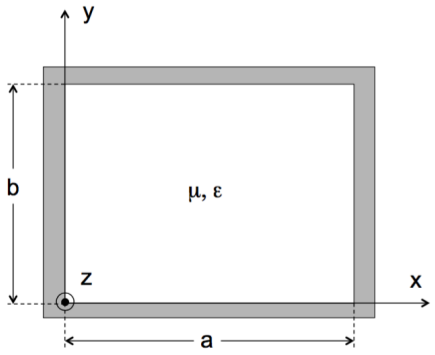
In general, a guided wave is a superposition of these different modes.

In a hollow wave guide of the type we consider here, TEM waves cannot exist. This can be shown by supposing such a wave exists. Then substituting $\tilde{B}_z = 0$ in equation (7.23) yields $\boldsymbol{\nabla}_t \times \tilde{\mathbf{E}}_t = 0$. This implies it exists a scalar potential V such that $\tilde{\mathbf{E}}_t = -\boldsymbol{\nabla}_t V$. In addition, substituting $\tilde{E}_z = 0$ in equation (7.21) yields $\boldsymbol{\nabla}_t \cdot \tilde{\mathbf{E}}_t = 0$. Therefore, V satisfies Laplace's equation $\nabla_t^2 V = 0$. As the surface of the waveguide is a conductor, it is an equipotential, that is to say we have the boundary condition $V = V_0$ at the surface. The function V which is equal to V_0 everywhere inside the waveguide satisfies Laplace's equation and the boundary condition. As the solution to Laplace's equation is unique given a boundary condition at the surface, it is $V = V_0$. This implies that $\tilde{\mathbf{E}} = \mathbf{0}$ everywhere in the waveguide, that is to say there is no wave.



Sketch of the field lines for TE and TM modes. (*from http://www.allaboutcircuits.com/textbook/alternating-current/chpt-14/waveguides/*)

7.2.2 TE waves in a rectangular wave guide



An important application of the results obtained in the previous section is the propagation of TE waves in a rectangular wave guide, which cross section is shown on the figure.

To calculate $\tilde{B}_z(x,y)$, we have to solve the wave equation (7.33). We use separation of variables: $\tilde{B}_z(x,y) = \tilde{F}(x)\tilde{G}(y)$. Substituting into equation (7.33) and dividing by $\tilde{F}\tilde{G}$, we obtain:

$$\frac{1}{\tilde{F}}\frac{d^2\tilde{F}}{dx^2} + \frac{1}{\tilde{G}}\frac{d^2\tilde{G}}{dy^2} + \left(\mu\epsilon\omega^2 - k^2\right) = 0.$$

As the first term on the left–hand–side can only depend on x, and the second term only on y, their sum is a constant only if they are both constant. We therefore have:

$$\frac{1}{\tilde{F}}\frac{d^2\tilde{F}}{dx^2} = C_x, \quad \frac{1}{\tilde{G}}\frac{d^2\tilde{G}}{dy^2} = C_y, \tag{7.36}$$

where C_x and C_y are two constants that may be complex and that satisfy:

$$C_x + C_y = -\mu\epsilon\omega^2 + k^2. \tag{7.37}$$

We write $C_x = -k_x^2$, where k_x may be complex. Then solving for $\tilde{F}$ in (7.36) we obtain $\tilde{F}(x) = A_1 e^{ik_x x} + A_2 e^{-ik_x x}$, where A_1 and A_2 are two (complex) constants. The boundary conditions (7.34) and (7.35) yield $\tilde{E}_z = 0$ and $\tilde{B}_x = 0$ at $x = 0$ and $x = a$. Substituting into equation (7.31), this implies that $\partial\tilde{B}_z/\partial x = 0$ at $x = 0$ and $x = a$. Therefore, $A_1 - A_2 = 0$ and $A_1 e^{ik_x a} - A_2 e^{-ik_x a} = 0$. We cannot have $A_1 = A_2 = 0$, as that would imply $B_z = 0$. As we are looking for solutions which are TE waves, that is to say with $E_z = 0$, and since TEM waves are not supported in a hollow wave guide (see previous section), we cannot have $B_z = 0$. This means that $A_1 = A_2 \neq 0$ and $e^{ik_x a} - e^{-ik_x a} = 0$. Writing $k_x = \alpha + i\beta$, where α and β are real, we obtain $\left(e^{-\beta a} - e^{\beta a}\right)\cos(\alpha a) = 0$ and $\left(e^{-\beta a} + e^{\beta a}\right)\sin(\alpha a) = 0$. This can only be achieved by having $\beta = 0$, so that $k_x = \alpha$ is real, and $\sin(\alpha a) = 0$. Therefore, $\alpha a = k_x a = m\pi$, where m is a positive integer, and $\tilde{F}(x) = 2A_1\cos(k_x x)$.

Similarly, the boundary conditions (7.34) and (7.35) yield $\tilde{E}_z = 0$ and $\tilde{B}_y = 0$ at $y = 0$ and $y = b$. Substituting into equation (7.31), this implies that $\partial\tilde{B}_z/\partial y = 0$ at $y = 0$ and $y = b$. As above, we then obtain $\tilde{G}(y) = 2A_3\cos(k_y y)$ with $k_y b = n\pi$, n being a positive

integer, and where A_3 is a (complex) constant.

The solutions to equation (7.33) which satisfy the boundary conditions at the surface of the waveguide are then:

$$\tilde{B}_z(x,y,z,t) = \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} \mathrm{TE}_{mn}(x,y)\mathrm{e}^{\mathrm{i}(kz-\omega t)}, \tag{7.38}$$

where the sum is over the so–called *TE_{mn} modes*:

$$\boxed{\mathrm{TE}_{mn} = C_{mn}\cos\left(\frac{m\pi x}{a}\right)\cos\left(\frac{n\pi y}{b}\right),} \tag{7.39}$$

with C_{mn} being a (complex) constant. Reflections at the conducting surface produce waves that interfere and add to give waves with standing patterns in the x– and y–directions while propagating in the z–direction. The other components of $\tilde{\mathbf{E}}$ and $\tilde{\mathbf{B}}$ can be obtained from equations (7.30) and (7.31).
Substituting C_x and C_y into equation (7.37) yields the wavenumber k that occurs for a given frequency ω:

$$k^2 = \mu\epsilon\omega^2 - \pi^2\left(\frac{m^2}{a^2}+\frac{n^2}{b^2}\right). \tag{7.40}$$

As can be seen from this equation, for a given mode TE_{mn}, there is a *cutoff frequency* ω_{mn} such that only waves with $\omega > \omega_{mn}$ can propagate. We have:

$$\boxed{\omega_{mn} = \frac{\pi}{\sqrt{\mu\epsilon}}\sqrt{\frac{m^2}{a^2}+\frac{n^2}{b^2}}.} \tag{7.41}$$

If $\omega < \omega_{mn}$, k is imaginary and the wave decays exponentially. Such modes are called *evanescent*.
We can write the wavenumber k, phase velocity $v_\varphi = \omega/k$ and group velocity $v_g = d\omega/dk$ (at which energy travels) in terms of the cutoff frequency:

$$k = \sqrt{\mu\epsilon}\sqrt{\omega^2-\omega_{mn}^2}\,, \quad v_\varphi = \frac{1}{\sqrt{\mu\epsilon}}\frac{1}{\sqrt{1-(\omega_{mn}/\omega)^2}}\,, \quad v_g = \frac{1}{\sqrt{\mu\epsilon}}\sqrt{1-(\omega_{mn}/\omega)^2}\,.$$

Since the wavenumber k is always smaller than the free–space value $\sqrt{\mu\epsilon}\omega$, the wavelength in the guide is always larger than the free-space value. Also, the phase velocity v_φ is always larger than the free–space value $1/\sqrt{\mu\epsilon}$.

A pointed out above, the TE_{mn} modes, being solutions of an eigenvalue problem, form an orthogonal set. Therefore, any signal with frequency ω can be written as a sum over m and n of TE_{mn} modes. If a signal with frequency ω is fed into the waveguide, only the terms in the sum over m and n which have $\omega_{mn} < \omega$ will propagate. The other terms will decay exponentially. For a non–monochromatic signal, there will be a superposition of waves with different frequencies (obtained by doing a Fourier transform of the signal), each of these waves being themselves a sum of TE_{mn} modes.

The TE_{10} mode:

If $a > b$, we see that the lowest cutoff frequency is obtained for $m = 1$ and $n = 0$, that is to say the lowest TE mode is TE_{10}. It can be shown that the only modes that satisfy Maxwell's equations and the boundary conditions in a rectangular waveguide are TE and TM modes. Therefore, any wave propagating along the axis of a rectangular waveguide is a combination of the TE and TM modes that can propagate. Besides, the TM mode with the lowest cutoff frequency is TM_{11}, and the associated cutoff frequency is larger than that of TE_{10}. The lowest cutoff frequency of all possible modes is therefore that of TE_{10} and is $\omega_{10} = \pi/(a\sqrt{\mu\epsilon})$. This means that a rectangular waveguide can only propagate signals with frequencies $\omega > \pi/(a\sqrt{\mu\epsilon})$, or with wavelengths $\lambda = 2\pi/(\sqrt{\mu\epsilon}\omega) < 2a$.

In general, when energy is fed into a waveguide in order to be carried across some distance, it is preferable to have only the lowest order mode propagate into the waveguide. In that case indeed, higher modes (with higher frequencies, or equivalently lower wavelengths) produced by distorsion of the field cannot propagate. In other words, noise decays exponentially along the guide.

This is usually achieved by using a rectangular waveguide with $a = 2b$. In such a waveguide, for the TE modes, $\omega_{20} = \omega_{01} = 2\omega_{10}$ and all the other cutoff frequencies are larger than ω_{20}. For TM_{11}, the cutoff frequency is 2.2 times that of the TE_{10} mode. Therefore, there is a range of frequencies from ω_{10} to $2\omega_{10}$ for which the TE_{10} mode is the only propagating mode. As an example, let us consider a radar wave with a wavelength $\lambda = 3$ cm and $\nu = c/\lambda = 10^{10}$ Hz. If a rectangular waveguide filled with air (so that $\mu\epsilon = 1/c^2$) has $a = 2$ cm and $b = 1$ cm, then $\nu_{10} = \omega_{10}/(2\pi) = c/(2a) = 7.5 \times 10^9$ Hz and the second lowest cutoff frequency is 1.5×10^{10} Hz. As only ν_{10} is smaller than ν, only the TE_{10} mode can be propagated. From equation (7.40), we calculate that the wavelength of the TE_{10} mode in the waveguide is $\lambda = 2\pi/k \simeq 4.5$ cm and the velocity at which energy flows down the waveguide is $v_g \simeq 2 \times 10^8$ m s^{-1}.

Using equations (7.38) and (7.39) , we obtain the z–component of the magnetic field corresponding to the TE_{10} mode:

$$\tilde{B}_z(x, y, z, t) = C \cos\left(\frac{\pi x}{a}\right) \mathrm{e}^{\mathrm{i}(kz-\omega t)}, \tag{7.42}$$

where C is a (complex) constant. Equation (7.30) with $\tilde{E}_z = 0$ yields the electric field:

$$\tilde{\mathbf{E}} = \frac{\mathrm{i}a\omega C}{\pi} \sin\left(\frac{\pi x}{a}\right) \mathrm{e}^{\mathrm{i}(kz-\omega t)} \hat{\mathbf{y}}, \tag{7.43}$$

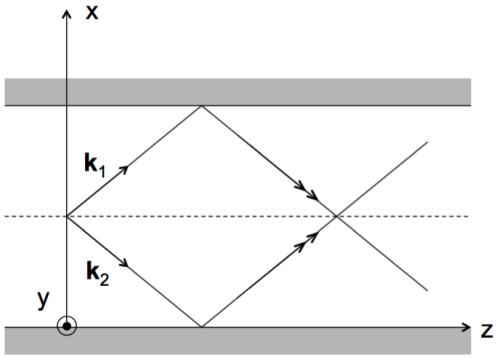
where we have used $\mu\epsilon\omega^2 - k^2 = (\pi/a)^2$ (eq. [7.40]). We can write:

$$\mathrm{i} \sin\left(\frac{\pi x}{a}\right) = \frac{1}{2}\left(\mathrm{e}^{\mathrm{i}\pi x/a} - \mathrm{e}^{-\mathrm{i}\pi x/a}\right),$$

so that equation (7.43) becomes:

$$\tilde{\mathbf{E}} = \frac{a\omega C}{2\pi}\left(\mathrm{e}^{\mathrm{i}(kz+\pi x/a-\omega t)} - \mathrm{e}^{\mathrm{i}(kz-\pi x/a-\omega t)}\right)\hat{\mathbf{y}}, \tag{7.44}$$

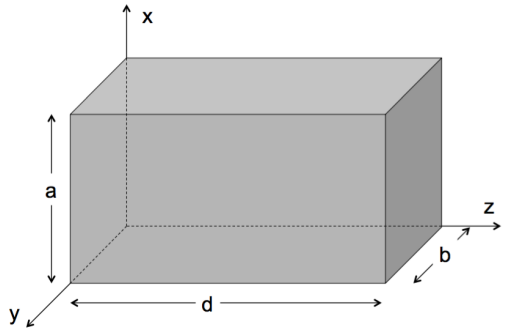
which shows that the TE_{10} mode is the superposition of two plane waves with wavevectors $\mathbf{k}_1 = (\pi/a)\hat{\mathbf{x}} + k\hat{\mathbf{z}}$ and $\mathbf{k}_2 = -(\pi/a)\hat{\mathbf{x}} + k\hat{\mathbf{z}}$.



These two waves reflect off the conducting surface and interfere to give a wave with a standing pattern in the x–direction while propagating in the z–direction.

7.2.3 Resonant cavities

A resonant cavity, or cavity resonator, is obtained by adding end surfaces in a waveguide. Reflection of the waves on these surfaces produce a standing pattern along the axis of the cylinder, so that the wave is "trapped". The cavity supports *resonant modes* which oscillate at *resonant frequencies* with a larger amplitude than non–resonant modes. Resonant cavities are similar to LC circuits but are used at higher frequencies (GHz) to select specific frequencies from a signal or produce waves with a specific frequency.



The simplest cavity for which we can calculate the resonant frequencies is the rectangular cavity obtained by closing the rectangular waveguide described above.

The waves in the cavity no longer propagate in the z–direction but are standing waves. As the boundary conditions in the x– and y–directions are the same as in a waveguide, we may use equations (7.38) and (7.39) for $\tilde{B}_z$ and write:

$$\tilde{B}_z(x,y,z,t) = \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} C_{mn}\cos\left(\frac{m\pi x}{a}\right)\cos\left(\frac{n\pi y}{b}\right)\left[K_1\sin(kz) + K_2\cos(kz)\right]\mathrm{e}^{-\mathrm{i}\omega t}, \tag{7.45}$$

where the z–dependence is that appropriate to standing waves, with K_1 and K_2 being constants. The boundary condition (7.35) leads to $\tilde{B}_z = 0$ at $z = 0$ and $z = d$, which implies $K_2 = 0$ and $kd = l\pi$, where l is an integer. Equation (7.45) then becomes:

$$\tilde{B}_z(x, y, z, t) = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \sum_{l=0}^{\infty} \Psi_{nml}(x, y, z) \mathrm{e}^{-\mathrm{i}\omega t}, \tag{7.46}$$

with:

$$\boxed{\Psi_{nml}(x, y, z) = C_{mnl} \cos\left(\frac{m\pi x}{a}\right) \cos\left(\frac{n\pi y}{b}\right) \sin\left(\frac{l\pi x}{d}\right),} \tag{7.47}$$

where C_{nml} is a constant. For TE fields, $\tilde{E}_z = 0$, and Maxwell's equations (7.21)–(7.24) enable $\tilde{\mathbf{E}}_t$ and $\tilde{\mathbf{B}}_t$ to be calculated. Given the expression (7.47) for $\tilde{B}_z$, the z–component of the wave equation (7.26) implies that ω can only take some particular values called *resonant frequencies* and defined by:

$$\boxed{\omega_{mnl} = \frac{\pi}{\sqrt{\mu\epsilon}} \sqrt{\frac{m^2}{a^2} + \frac{n^2}{b^2} + \frac{l^2}{d^2}}.} \tag{7.48}$$

There is an infinite number of resonant frequencies that form a discrete set. As the waveguide, the cavity is usually used so that only the low mode is excited, that is to say the mode with the lowest frequency, called the *fundamental* frequency. Equation (7.48) shows that the fundamental frequency is such that the associated wavelength is twice the largest dimension of the cavity. Therefore, resonant cavities are used for microwaves, which have wavelengths in the range 1 mm to 1 m.

Let us consider an electromagnetic signal like a pulse. It is composed of a range of frequencies which can be obtained by doing a Fourier transform of the time domain representation of the signal. If some of these frequencies are close enough to some resonant frequencies in the cavity, then the corresponding modes are excited when the signal is fed into the cavity (through a small aperture for example). In practice, resonant cavities are designed in such a way that the energy in the resonant modes is dissipated only after a very large number of oscillations. Therefore, resonant cavities can be used to store energy at a particular frequency, and tuning of the frequency can be made by changing the length of the cavity (using a piston).

Appendix A

Radiation

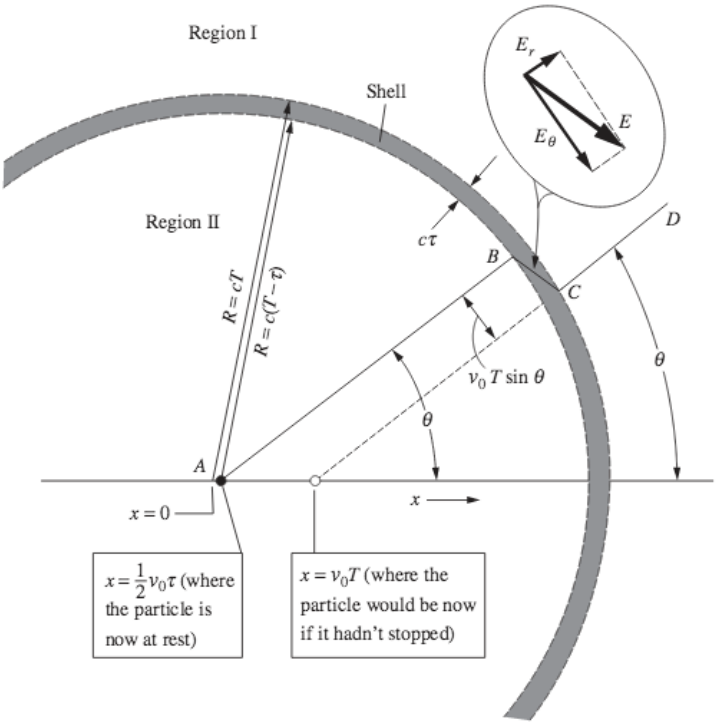
This chapter is off syllabus.

In general, electromagnetic waves are associated with a non zero Poynting vector which describes the transport of energy by the wave. The power passing through the surface of a sphere at infinity is equal to the flux of the Poynting vector through the sphere. When this is zero, the energy is not carried away to infinity. We define as *radiation* the energy flux that propagates to infinity.
If the electric and magnetic fields fall off as the inverse square of the distance, the Poynting vector falls off as the inverse distance to the fourth power, and the power through a sphere at infinity is zero. There is no radiation in that case.
Systems radiating energy include oscillating dipoles, as we will see in this chapter. They are very important as they enable communication through big distances. Electromagnetic radiation is what enables us to see the stars, communicate using satellites, receive information from the early universe through the microwave background, etc...

A.1 Power radiated by an accelerated point charge

It can be shown that the Poynting vector associated with the field of a point charge moving with constant velocity v is propotional to vE^2. As E varies inversely with the square of the distance, there is no radiation emitted by a charge moving at constant speed. The energy associated with the field is carried along by the charge when it moves, but does not escape to infinity. As we are going to show here, *only accelerated charges radiate energy.*

To calculate the energy radiated by an accelerated point charge, we follow the derivation given by Purcell (in an approach first proposed by J. J. Thomson). We consider a particle of charge q which has been moving with constant velocity $v_0 \ll c$ along the x–axis between the times $t = -\infty$ and $t = 0$. At $t = 0$, the particle is at $x = 0$ and starts to decelerate. It comes to a stop at time $t = \tau$. If a is the (constant) deceleration, then the velocity of the particle is $v = -at + v_0$, so that $\tau = v_0/a$. The position of the particle between $t = 0$ and $t = \tau$ is $x = -at^2/2 + v_0 t$, so that when the particle stops it is located at $x = v_0\tau/2$. We now consider the electric field due to this point charge at time $T \gg \tau$.

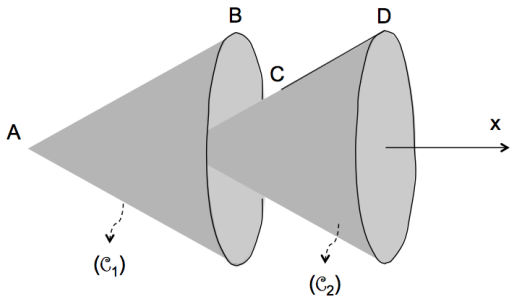


(From E. Purcell & D. Morin, *Electricity and Magnetism*)

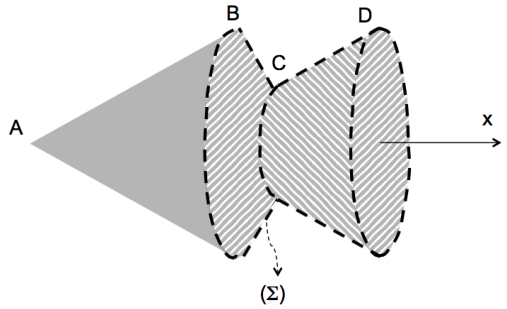
The information that the particle starts decelerating propagates from $x = 0$ at speed c. Therefore, at time $t = T$, this information has reached the distance cT from the point $x = 0$. Only observers within the sphere of radius cT centered on $x = 0$ know that the particle has decelerated. This means that for the observers beyond this sphere (region I on the figure), the electric field is that of a charge which is still moving at speed v_0. If the charge had not decelerated, it would be at $x = v_0 T$ at time $t = T$. Observers in region I therefore see the electric field produced by a charge at $x = v_0 T$ and moving with velocity v_0. (This is because the particle had been moving with constant velocity ever since $t = -\infty$, so the whole space has been filled with field lines following the charge moving at this constant velocity). The electric field produced by a charge moving at constant velocity points away from the *instantaneous* position of the charge Therefore, the electric field at point C on the figure, for example, is along the line segment CD.

The information that the particle has stopped propagates from $x = v_0\tau/2$ at speed c. Therefore, at time $t = T$, this information has reached the distance $c(T - \tau)$ from the point $x = v_0\tau/2$. This means that, for the observers within the sphere of radius $c(T - \tau)$ centered on $x = v_0\tau/2$ (region II on the figure), the electric field is that of a charge at rest at $x = v_0\tau/2$. Therefore, at point B on the figure, the electric field is along the line segment AB and is equal to $E_B = q/[4\pi\epsilon_0 c^2(T - \tau)^2]$. The point B has been chosen so that the angle θ between AB and the x–axis is the same as that between CD and the x–axis. Note that, since $v_0 \ll c$ and $\tau \ll T$, $x = v_0\tau/2 \ll cT$, that is to say the separation between the points located at $x = 0$ and at $x = v_0\tau/2$ is very small compared to the other distances in the problem, and can be neglected. Therefore, $E_B \simeq q/(4\pi\epsilon_0 R^2)$ with $R = cT$.

We are now going to calculate the electric field in the transition region, which is the spherical shell of thickness $c\tau$.



The field line CD is produced by a moving particle, and therefore the density of field lines in the cone $(\mathcal{C}_2)$ around the x–axis decreases as we approach the axis (see Problem 1 of Problem set 4). However, if $v_0 \ll c$, as we assume here, this can be neglected. Therefore, as the two cones $(\mathcal{C}_1)$ and $(\mathcal{C}_2)$ have the same opening angle, they carry the same amount of flux from the charge q.



This implies that the flux of $\mathbf{E}$ through the surface (Σ) is zero, where (Σ) is delimited by the bases of $(\mathcal{C}_1)$ and $(\mathcal{C}_2)$. This can be satisfied only if the line segments AB and CD are part of the same field line, connected by the segment BC.

As can be derived from the geometry of the problem, we then have $E_\theta/E_r = v_0 T \sin\theta/(c\tau)$, where E_r and E_θ are the radial and transverse components of the electric field in the shell, respectively. As the perpendicular component of $\mathbf{E}$ is continuous, we have $E_r = E_B$, which yields:

$$E_\theta = \frac{q v_0 \sin\theta}{4\pi\epsilon_0 c^2 \tau R},$$

were we have used $T = R/c$. As $v_0/\tau = a$, we obtain:

$$\boxed{E_\theta = \frac{qa\sin\theta}{4\pi\epsilon_0 c^2 R}}. \tag{A.1}$$

We obtain the very important result that E_θ is propotional to $1/R$, and not to $1/R^2$ as the field produced by a static charge. Since $E_r \propto 1/R^2$, we see that, when $R \to \infty$, or equivalently $T \to \infty$, $E \to E_\theta$ and $\mathbf{E}$ is transverse. There is an associated magnetic field $B_\varphi = E_\theta/c$ which is perpendicular to both $\hat{\mathbf{R}}$, the unit vector in the radial direction, and $\mathbf{E}$. Since $E_\theta B_\varphi \propto 1/R^2$, the Poynting vector through the sphere of radius R is finite and energy is radiated away. For this reason, we call E_θ and B_φ the *radiation fields* and denote them $E_{\rm rad}$ and $B_{\rm rad}$. The parts of the fields that fall off faster than $1/R$ with distance dominate near the charge and are called *near–zone* fields. The Poynting vector when $R \to \infty$ is:

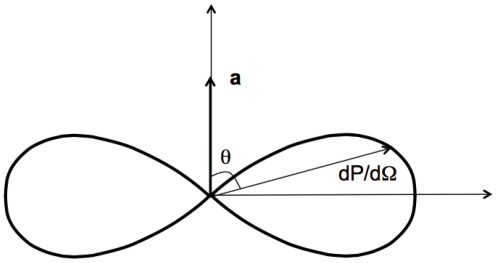
$$\mathbf{S} = \frac{1}{\mu_0}\mathbf{E}_{\rm rad}\times\mathbf{B}_{\rm rad} = \frac{E_{\rm rad}^2}{\mu_0 c}\hat{\mathbf{R}} = \epsilon_0 c E_{\rm rad}^2 \hat{\mathbf{R}} = \frac{q^2 a^2 \sin^2\theta}{16\pi^2\epsilon_0 c^3 R^2}\hat{\mathbf{R}}. \tag{A.2}$$

The energy per unit time (power) dP crossing an infinitesimal area $d\Sigma$ located at R and perpendicular to $\hat{\mathbf{R}}$ is $dP = Sd\Sigma$, that is to say:

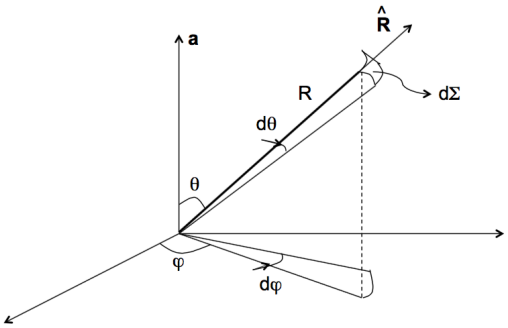
$$dP = \frac{q^2 a^2 \sin^2\theta}{4\pi\epsilon_0 c^3} \frac{d\Sigma}{4\pi R^2}. \tag{A.3}$$

The differential solid angle $d\Omega$ subtended by $d\Sigma$ at the origin is given by $d\Omega = d\Sigma/R^2$. Therefore, the power radiated into an infinitesimal solid angle is:

$$\boxed{dP = \frac{q^2 a^2 \sin^2\theta}{4\pi\epsilon_0 c^3} \frac{d\Omega}{4\pi}}. \tag{A.4}$$



The figure shows the radiation pattern, that is to say the curve drawn using polar coordinates (r, θ) with $r = dP/d\Omega \propto \sin^2\theta$. Note that *there is no energy radiated along the direction of motion.*



To obtain the total power P_{rad} radiated through the sphere of radius R at time T, we integrate dP over $d\Sigma = R^2 \sin\theta d\theta d\varphi$, or equivalently over $d\Omega = \sin\theta d\theta d\varphi$. This yields:

$$P_{\text{rad}} = \frac{q^2 a^2}{8\pi\epsilon_0 c^3} \int_0^\pi \sin^3\theta d\theta.$$

We have:

$$\int_0^\pi \sin^3\theta d\theta = \int_0^\pi \left(\sin\theta - \sin\theta\cos^2\theta\right) d\theta = \left[-\cos\theta + \frac{1}{3}\cos^3\theta\right]_0^\pi = \frac{4}{3},$$

so that:

$$\boxed{P_{\text{rad}} = \frac{q^2 a^2}{6\pi\epsilon_0 c^3}}. \tag{A.5}$$

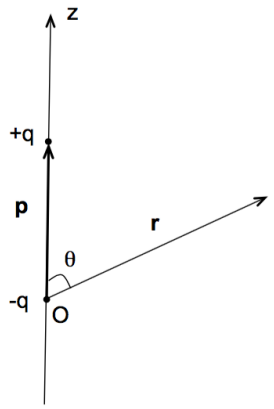
This is called **Larmor's formula**. As can be seen from this equation, P_{rad} does not depend on R or T, that is to say the power crossing the surface of a sphere with radius R at time T is the same as the power crossing the surface of a sphere with different radius at some other time. This is a statement of energy conservation. Although we have obtained Larmor's formula by considering a particle with a constant acceleration, it correctly gives the instantaneous rate of energy radiated by a charge with any acceleration, constant or variable. Therefore, is states that *any accelerated charged particle radiates a power proportional to the square of its acceleration.*

This formula is only valid for non–relativistic motions, that is to say for particles moving with velocity $v \ll c$. However, Larmor's formula is Lorentz invariant. This is because, when going from an inertial frame to another, energy and time transform in the same way. If dU' is the energy carried away by the radiation emitted between times $t' = 0$ and $t' = dt'$ measured in an intertial frame (F') in which the particle is at rest, then transforming to an inertial frame (F) in which the velocity of the particle is v, we have $dU = \gamma dU'$ and $dt = \gamma dt'$. Therefore, the power radiated by the charge is $dU/dt = dU'/dt'$. If the particle moves relativistically in (F), we can then transform to an intertial frame (F') where the particle is at rest *instantaneously*, and apply Larmor's formula in that frame.

A.2 Electric dipole radiation

A.2.1 Oscillating dipole

We consider a physical dipole in which the positive charge q has a harmonic motion along the axis of the dipole.



The position of the charge q is given by $z(t) = d \sin \omega t$, and the dipole moment is $\mathbf{p}(t) = qz(t)\,\hat{\mathbf{z}}$. As we have seen in chapter 1, the electric field produced by a dipole with moment $\mathbf{p}$ is proportional to $\mathbf{p}$. Therefore, this oscillating dipole produces an electric field which varies with frequency ω, and therefore also a magnetic field which varies with the same frequency. As one of the charges oscillates, this field is a radiation field far away from the dipole. A detailed calculation would show that the radiation zone is at distances $r \gg c/\omega$, that is to say $\boxed{r \gg \lambda}$, where $\lambda = 2\pi c/\omega$ is the wavelength of the electromagnetic field produced by the dipole.

We are going to calculate the power radiated by this dipole using the results of the previous section. These apply only when the velocity of the charge, which is $\dot{z} = \omega d \cos \omega t$, is small compared to c. This means $\omega d \ll c$, or $\boxed{d \ll \lambda}$. Note that the conditions $r \gg \lambda$ and $\lambda \gg d$ imply that $\boxed{r \gg d}$, which is the condition required for the system to be a perfect dipole. Such an oscillating dipole is called a *Hertzian dipole*.
The acceleration of the charge is $\ddot{z}(t) = -\omega^2 z(t)$. Using equation (A.4), we obtain that the power radiated per unit solid angle in the direction θ is:

$$\frac{dP}{d\Omega} = \frac{q^2\omega^4 z^2(t)\sin^2\theta}{16\pi^2\epsilon_0 c^3} = \frac{\omega^4 p^2(t)\sin^2\theta}{16\pi^2\epsilon_0 c^3}.$$

Averaged over one period of oscillations, this yields:

$$\boxed{\left\langle \frac{dP}{d\Omega} \right\rangle = \frac{\omega^4 p_0^2 \sin^2\theta}{32\pi^2\epsilon_0 c^3}}, \tag{A.6}$$

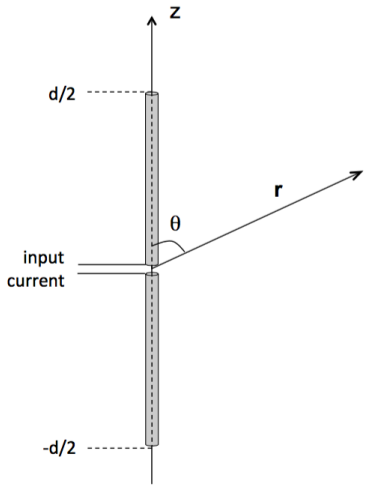
where $p_0 = qd$, The total power radiated by the dipole is given by Larmor's formula (A.5):

$$P_{\text{rad}} = \frac{\omega^4 p^2(t)}{6\pi\epsilon_0 c^3},$$

and averaged over a period of oscillations:

$$\boxed{\langle P_{\rm rad}\rangle = \frac{\omega^4 p_0^2}{12\pi\epsilon_0 c^3}}. \tag{A.7}$$

A.2.2 Antennas



An antenna is used to convert electric power into electromagnetic radiation, or vice versa. We consider a linear antenna which extends from $-d/2$ to $+d/2$ along the z–axis, with a small gap at $z = 0$ where a transmitter delivers current. The current in the antenna is given by: $I(z,t) = I(z)\cos\omega t$ with:

$$I(z) = I_0\left(1 - \frac{2|z|}{d}\right).$$

We note $\rho(z,t)$ the volume charge density in the antenna. The continuity (charge conservation) equation gives $-\partial\rho/\partial t = \boldsymbol{\nabla}\cdot\mathbf{J}$, where $\mathbf{J} = (I/s)\hat{\mathbf{z}}$ is the current density, with s being the cross section of the wire. This yields $-\partial\lambda/\partial t = \partial I/\partial z$, where $\lambda(z,t) = \rho(z,t)s$ is the charge per unit length in the antenna. Using the above expression for I, we obtain:

$$\lambda(z,t) = \pm\frac{2I_0}{\omega d}\sin\omega t,$$

where the upper (lower) sign applies for $z > 0$ ($z < 0$).
The dipole moment of this charge distribution is:

$$\mathbf{p}(t) = \hat{\mathbf{z}}\int_{-d/2}^{d/2} z\,dq,$$

where $dq = \lambda(z,t)dz$ is the charge contained in an infinitesimal length dz at position z along the wire (see eq. [1.36]). Therefore:

$$\mathbf{p}(t) = \frac{2I_0}{\omega d}\sin\omega t\left(-\int_{-d/2}^{0} z dz + \int_0^{d/2} z dz\right)\hat{\mathbf{z}} = \frac{I_0 d}{2\omega}\sin\omega t\,\hat{\mathbf{z}}.$$

The antenna is then equivalent to the oscillating dipole studied above, assuming $\boxed{d \ll \lambda}$, where λ is the wavelength of the electromagnetic field produced by the antenna. Therefore, the time average over a period of the power radiated per unit solid angle in the direction θ by the antenna is given by equation (A.6), after replacing p_0 by $I_0 d/(2\omega)$, that is to say:

$$\left\langle\frac{dP}{d\Omega}\right\rangle = \frac{I_0^2}{128\pi^2\epsilon_0 c}(kd)^2\sin^2\theta, \tag{A.8}$$

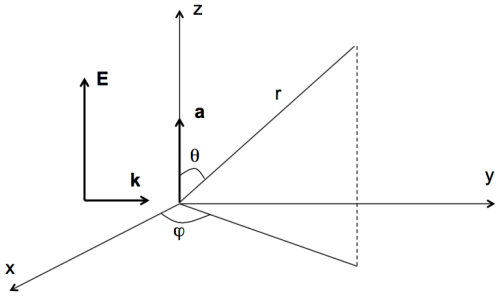
where $k = \omega/c = 2\pi/\lambda$ is the wavenumber. The time average over a period of the total power radiated by the antenna is given by equation (A.7):

$$\langle P_{\text{rad}} \rangle = \frac{I_0^2}{48\pi\epsilon_0 c}(kd)^2. \tag{A.9}$$

Note that antennas used to emit radio waves have a length which is not small compared to the wavelength of the electromagnetic field they produce. Therefore, the power they emit cannot be calculated using the formulas above.

A.3 Thomson scattering

When an electromagnetic wave is incident on a charged particle, it exerts a force on the particle. As the electric and magnetic fields vary periodically with time, the force induces oscillations of the particle which then behaves like an oscillating dipole. This dipole radiates energy in directions other than the direction of incidence: there is *scattering of the electromagnetic wave* by the particle. If the motion of the particle is non relativistic, the frequency of the radiation emitted by the particle is the same as that of the incident wave. This is called *Thomson scattering*.



We consider a particle of mass m and charge q located at the origin O of the coordinate system. We assume that the incident wave is polarized in the z–direction and propagates in the y–direction, so that the incident electric field can be written as $\mathbf{E}(y,t) = E_0 \cos(ky - \omega t)\,\hat{\mathbf{z}}$, where k is the wavenumber.

The force acting on the particle is $\mathbf{F} = q(\mathbf{E} + \mathbf{v}\times\mathbf{B})$, where $\mathbf{v}$ is the velocity of the particle. As $B = E/c$, we have $vB = (v/c)E \ll E$ for non relativistic motion, and therefore the magnetic force can be neglected. Then $\mathbf{F} \simeq q\mathbf{E}$, which results in an acceleration of the particle $\mathbf{a} = q\mathbf{E}(0,t)/m = a_0 \cos(\omega t)\,\hat{\mathbf{z}}$, where we have defined $a_0 = qE_0/m$. The particle therefore oscillates around O along the z–axis with the frequency ω.

Using equation (A.4), we obtain the time average over a period of the power radiated per unit solid angle in the direction θ by the particle:

$$\frac{dP}{d\Omega} = \left(\frac{q^2}{4\pi\epsilon_0 mc^2}\right)^2 \frac{\epsilon_0 E_0^2 c}{2} \sin^2\theta. \tag{A.10}$$

A.3.1 Thomson scattering cross section

We define the scattering cross section σ as:

$$\text{Scattered energy per unit time} = \sigma \times (\text{incident energy flux}).$$

As the energy flux is an energy per unit time and per unit area, it follows that σ is a *surface*. It is the equivalent area of the incident wavefront through which the energy per unit time is equal to the total power re–radiated by the particle. In other words, it is the effective area of the incident beam which is scattered.

We also define the cross section $d\sigma/d\Omega$ per unit solid angle as:

Scattered energy per unit time per unit solid angle = $d\sigma/d\Omega \times$ (incident energy flux).

It is the effective area of the incident beam which is scattered per unit solid angle.
Here we use quantities which are time averaged over a period of oscillations. The incident energy flux is then the time–averaged Poynting vector for the incident wave, and this is equal to $\epsilon_0 E_0^2 c/2$ (see eq. [4.53]). This yields:

$$\frac{d\sigma}{d\Omega} = \frac{dP/d\Omega}{\epsilon_0 E_0^2 c/2} = \left(\frac{q^2}{4\pi\epsilon_0 mc^2}\right)^2 \sin^2\theta. \tag{A.11}$$

If the particle is an electron with charge $q = -e$, then the total cross section is called *Thomson* scattering cross section and is noted $\sigma_{\rm T}$:

$$\sigma_{\rm T} = \int_0^\pi \frac{d\sigma}{d\Omega} 2\pi \sin\theta d\theta = \frac{8\pi}{3}\left(\frac{e^2}{4\pi\epsilon_0 mc^2}\right)^2. \tag{A.12}$$

We define the *classical electron radius* r_0 as:

$$r_0 = \frac{e^2}{4\pi\epsilon_0 mc^2}. \tag{A.13}$$

If we distribute the charge e at the surface of a sphere, then r_0 is the radius of the sphere whose electrostatic potential energy $e^2/(4\pi\epsilon_0 r_0)$ equals the rest mass mc^2 of the electron. We can write the Thomson cross section under the form:

$$\boxed{\sigma_{\rm T} = \frac{8\pi}{3} r_0^2}. \tag{A.14}$$

When spheres of negligible radii are scattered by a sphere of radius r_0, the cross section is πr_0^2. Therefore, an electron which scatters radiation behaves like a hard sphere of radius $\sim r_0$.

A.3.2 Thomson, Compton, resonant and Rayleigh scattering

Thomson scattering, as described in this section, is the *elastic* scattering of electromagnetic radiation by a *free* charged particle. The frequency of the radiation emitted by the particle is the same as that of the incident wave.

Compton scattering is an *inelastic* scattering of electromagnetic radiation by a *free* charged particle. The frequency of the radiation emitted by the particle is *not* the same as that of the incident wave.

In a quantum mechanical description of scattering, electromagnetic waves are made of photons. When a photon is scattered by a particle, it exchanges momentum and energy with the particle. The energy transferred to the particle makes it recoil. This effect is negligible as long as the energy of the incident photon ($h\nu$) is small compared to the rest mass energy of the particle (mc^2), in which case we are in the Thomson scattering regime. However, when $h\nu$ becomes comparable to mc^2, recoil of the free particle is no longer negligible and we are in the Compton scattering regime.

Resonant and **Rayleigh** scattering are the *elastic* scattering of an electromagnetic wave by an electron bound to a nucleus in an atom . The electron can be viewed as being bound to the nucleus by a spring with spring constant $m\omega_0^2$, where m is the electron's mass and ω_0 is the natural frequency at which the electron oscillates if perturbed. When an incident wave with frequency ω hits the atom, it forces oscillations of the electron at the frequency ω. If $\omega \gg \omega_0$, the electron behaves as if it were free, and we are in the Thomson scattering regime. The case $\omega \simeq \omega_0$ corresponds to resonant scattering, and is associated with a very large cross section. If $\omega \ll \omega_0$, the electron becomes a small radiating dipole and we are in the regime of Rayleigh scattering. In this case, the energy radiated by the electron varies like $\omega^4 \propto 1/\lambda^4$ (see problem 6 in Problem Set 4).

Scattering of a wave by an electron bound to an atom is described quantum mechanically as the absorption of a photon by the atom which excites the electron to a higher energy state. Scattering occurs when the atom drops down to its original energy state and re–emits the photon. If the atom collides with another atom while falling down to its original energy state and transfers the energy of the photon as kinetic energy to the colliding atom, there is **absorption** of the incident wave instead of scattering.