A2: Electromagnetism and Optics

Toby Adkins

June 20, 2016

Contents

1 Electromagnetism I **3**
- 1.1 Electric Fields in Matter 4
  - 1.1.1 Dipoles 4
  - 1.1.2 Polarisation 5
  - 1.1.3 The Electric Displacement Vector 6
- 1.2 Magnetic Fields in Matter 10
  - 1.2.1 Magnetic Dipoles 10
  - 1.2.2 Magnetisation 11
  - 1.2.3 The Auxiliary Magnetic Field 12
  - 1.2.4 Ferromagnetism 13
- 1.3 Boundary Conditions 15
  - 1.3.1 Perpendicular Fields 15
  - 1.3.2 Parallel Fields 15
- 1.4 Potentials 17
  - 1.4.1 Electric Potentials 17
  - 1.4.2 Magnetic Potentials 20
- 1.5 Electromagnetic Waves in a Vacuum 22
  - 1.5.1 Waves in a Vacuum 22
  - 1.5.2 The Poynting Vector 22
  - 1.5.3 Energy Conservation 23
  - 1.5.4 Energy and Momentum Transport 24
- 1.6 Electromagnetic Waves in Linear Media 25
  - 1.6.1 Normal Incidence 25
  - 1.6.2 Parallel Polarisation 27
  - 1.6.3 Perpendicular Polarisation 29
- 1.7 Electromagnetic Waves in Conductors 30
  - 1.7.1 Dispersion Relation 30
  - 1.7.2 Energy Transport 31
  - 1.7.3 Wave Impedance 31
- 1.8 Dispersion and Plasmas 32
  - 1.8.1 Plasmas 33

2 Electromagnetism II **36**
- 2.1 Radiation 37
  - 2.1.1 Larmor's Formula 37
  - 2.1.2 Scattering Types 38
  - 2.1.3 An Illustrative Example 39
- 2.2 Electromagnetism and Special Relativity 41
  - 2.2.1 Field Transformations 41
  - 2.2.2 Charge Invariance 42
- 2.3 Transmission Lines 43

2.3.1 Reflection and Transmission 45
2.3.2 Stubs 47

3 Optics 48
3.1 Fraunhofer Diffraction 49
3.1.1 Fourier Optics 50
3.1.2 Abbe's Theory of Imaging 53
3.1.3 Diffraction Limited Imaging 55
3.2 Spectroscopic Instruments 60
3.2.1 Fringe Formation 60
3.2.2 Some Important Definitions 60
3.2.3 The Michelson Interferometer 61
3.2.4 The Fabry-Perot Interferometer 66
3.3 Polarisation 70
3.3.1 Types of Polarisation 70
3.3.2 Birefringence 71

1. *Electromagnetism I*

This chapter aims to build on the concepts introduced in the first year CP2 Electromagnetism course, covering:

- Electric Fields in Matter
- Magnetic Fields in Matter
- Potentials
- Electromagnetic Waves in a Vacuum
- Electromagnetic Waves in Linear Media
- Electromagnetic Waves in Conductors
- Dispersion and Plasmas

After working through this chapter, students will be equipped to tackle a large variety of Electromagnetically related problems, and in a much more intuitive and succinct way in comparison to the CP2 course. It will be assumed that students are competent with the use of vector calculus, and in the manipulation of partial differential equations, as well as being able to recall Maxwell's equations as covered in the CP2 course. This is a significantly longer chapter than others previously included in these sets of notes, so it is recommended that readers look over it in sections, rather than attempting the entire thing in one go.

1.1 Electric Fields in Matter

Thus far in Electromagnetism, we have just considered systems of fixed charges, and calculated the potential assuming that this charge distribution was time-invariant. We will now look at what happens when materials are effected by external electric fields, in the particular case of *dielectrics*; materials where the electrons are bound to their atoms.

Consider the following simple model of a neutral atom; the nucleus can be modelled as a single point charge q, while the electron cloud is a uniformly charged sphere of radius a with total charge $-q$. The application of an external electric field will cause a separation of these two components, as shown below.

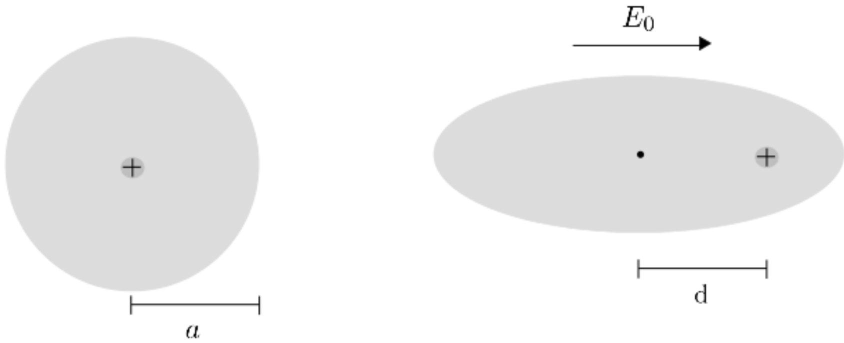

Figure 1.1: The original atom (left) and one with an external field applied (right)

The nucleus of the atom reaches equilibrium a distance d from it's original position when $\underline{E} = \underline{E}_0$, where $\underline{E}$ is the field due to the electrons at the position of the nucleus. Using Gauss' law,

$$E = \frac{qd}{4\pi\epsilon_0 a^3} = E_0 \quad \text{in equilibrium}$$

This means that we can write the *dipole moment* $p = qd$ as being proportional to the applied field

$$\underline{p} = \alpha \underline{E}_0$$

where $\alpha = 4\pi\epsilon_0 a^3$ is known as the *atomic polarisability*. Evidently, this means that the stronger the external field applied, the stronger the induced dipole field that is created as a result.

1.1.1 Dipoles

Before continuing with the results of this section, it might be worth recapping some of the important results that we know about electric dipoles. We know that for some dipole moment $\underline{p}$ that the potential is given by

$$V = \frac{1}{4\pi\epsilon_0} \frac{\underline{p} \cdot \underline{r}}{r^3}$$

With careful manipulation (most easily done using index notation, and remembering that $\partial_i r_j = \delta_{ij}$), it can be shown that the coordinate-free form of the electric field is given by

$$\boxed{\underline{E} = \frac{1}{4\pi\epsilon_0} \frac{1}{r^3} \left(3(\underline{p} \cdot \underline{\widehat{r}})\underline{\widehat{r}} - \underline{p}\right)} \tag{1.1}$$

Suppose that we have a dipole consisting of the positive charge at $\underline{r}_1$, and the negative charge at $\underline{r}_2$ such that $d = |\underline{r}_1 - \underline{r}_2|$ is small. Let the external electric field have value $\underline{E} + d\underline{E}$ at r_1 and $\underline{E}$ at $\underline{r}_2$. The force on the system is given by

$$\begin{aligned}\underline{F} &= \underline{F}_1 + \underline{F}_2 \\ &= q(\underline{E} + d\underline{E}) - q\underline{E} \\ &= qd\underline{E} \\ &= q\nabla(\underline{d} \cdot \underline{E})\end{aligned}$$

Noting that $\underline{p} = q\underline{d}$, we arrive at the result that

$$\boxed{\underline{F} = \nabla(\underline{p} \cdot \underline{E})} \tag{1.2}$$

This gives us the energy of the ensemble for 'free', as we know that $\underline{F} = -\nabla U$. Considering the torque on the system:

$$\begin{aligned}\underline{\Gamma} &= \underline{r}_1 \times \underline{F}_1 + \underline{r}_2 \times \underline{F}_2 \\ &= q((\underline{r}_1 - \underline{r}_2) \times \underline{E} + \underline{r}_1 \times d\underline{E}) \\ &= \underbrace{\underline{p} \times \underline{E}}_{\text{torque around dipole axis}} + \underbrace{\underline{r} \times \underline{F}}_{\text{torque around centre of coordinates}}\end{aligned}$$

Two perfect dipoles $\underline{p}_1$ and $\underline{p}_2$ are placed with their centres a distance d apart along the x-axis. $\underline{p}_1$ is orientated with it's dipole moment along $\widehat{\underline{x}}$, while $\underline{p}_2$ is orientated with it's dipole moment along $\widehat{\underline{y}}$. Determine the field each dipole experiences, and the torque on each dipole evaluated around its own centre. Why are these torques not equal and opposite?

The easiest way to answer this question is to use the form of the electric field given by (1.1). We find that in each case, the electric fields are given by

$$\begin{aligned}\underline{E}_1 &= \frac{1}{4\pi\epsilon_0}\frac{2p_1}{r^3}\,\widehat{\underline{x}} \\ \underline{E}_2 &= -\frac{1}{4\pi\epsilon_0}\frac{p_2}{r^3}\,\widehat{\underline{y}}\end{aligned}$$

Evaluating the torques using these fields:

$$\begin{aligned}\underline{\Gamma}_1 &= \underline{p}_1 \times \underline{E}_2 = \frac{1}{4\pi\epsilon_0}\frac{p_1 p_2}{d^3}\,\widehat{\underline{z}} \\ \underline{\Gamma}_2 &= \underline{p}_2 \times \underline{E}_1 = -\frac{1}{2\pi\epsilon_0}\frac{p_1 p_2}{d^3}\,\widehat{\underline{z}}\end{aligned}$$

These are evidently not equal and opposite as these are the torques of the dipoles around their own centre, and not their total torque. If we took into account the rotation around the centre of the coordinate system, then we would find that the torques do indeed balance.

1.1.2 Polarisation

Returning to the scenario above, suppose that we now have n small dipole moments per unit volume. If we are considering a sufficiently small volume of material, we can consider the electric field to be roughly uniform within it. This means that we can define the *polarisation* as the *number of dipole moments per unit volume*:

$$\boxed{\underline{P} = n\underline{p} = \sum_i n_i q_i \underline{d}_i} \tag{1.3}$$

This leads to three important changes within the material in question.

- Polarisation Volume Charge - Consider a uniform cylinder of oriented surface area element $d\underline{\Sigma}$. If the cylinder is of length $\underline{d}_i$, then the total charge that crosses $d\underline{\Sigma}$ is given by $\sum_i n_i q_i \underline{d}_i \cdot d\underline{\Sigma} = \underline{P} \cdot d\underline{\Sigma}$. The total charge that leaves the volume is given by the integral of this over the surface, which means that the divergence theorem can be used to show that
$$\rho_p = -\nabla \cdot \underline{P} \tag{1.4}$$
- Polarisation Surface Charge - If we now consider the whole surface bounding the entire dielectric, the charge that attempts to pass through the boundary can be considered as bound surface charge, given by
$$\sigma_p = \underline{P} \cdot \underline{\widehat{n}} \tag{1.5}$$
- Polarisation Current - Each of the charges moves a distance $\underline{d}_i$ as a result of the external electric field at some characteristic velocity $\underline{v}_i$. This means that we can write $\underline{J} = \sum_i n_i q_i \underline{v}_i$ by the definition of current density. This means that the polarisation current is clearly
$$\underline{J}_p = \frac{\partial \underline{P}}{\partial t} \tag{1.6}$$

Note that in all of these cases, the electrons are bound to their atoms; they cannot move more than a certain distance away from their equilibrium positions. This means that the polarisation current is a macroscopic effect created from the average of all the microscopic changes in the positions of the polarisation charges. The current is thus localised to the dielectric, and cannot be conducted into other substances. This is why many of these quantities are often referred to as *bound* quantities.

1.1.3 The Electric Displacement Vector

Let us now consider Gauss' Law in vector form. We now have another charge type to take account of; the polarisation charges ρ_p, as well as the free charges ρ_f.

$$\nabla \cdot \underline{E} = \frac{\rho}{\epsilon_0} = \frac{\rho_p + \rho_f}{\epsilon_0}$$

Considering (1.4), we can write that

$$\nabla \cdot (\epsilon_0 \underline{E} + \underline{P}) = \rho_f \longrightarrow \nabla \cdot \underline{D} = \rho_f$$

We call the quantity $\underline{D}$ *the electric displacement vector* defined as

$$\boxed{\underline{D} = \epsilon_0 \underline{E} + \underline{P}} \tag{1.7}$$

When dealing with fields associated with dielectrics, one should generally compute $\underline{D}$ first, as this allows us to get the other quantities from it with appropriate manipulation.

The electric field inside a dielectric is $\underline{E}_0$ and the polarisation is $\underline{P}$, such that $\underline{D}_0 = \epsilon_0 \underline{E}_0 + \underline{P}$. A cavity is hollowed out of the material, the size of which is not large enough to change the polarisation. Calculate the displacement vector $\underline{D}$ for the cases where the cavity is needle shaped, and where it is a thin circular wafer perpendicular to the polarisation.

In order to answer either of these cases, we need to superimpose a cavity that has polarisation $-\underline{P}$ in order for there to be no net polarisation within the cavity. We can model the first case as two small spheres of uniform charge separated by a distance d.

$$\sigma_p = \underline{P} \cdot \widehat{\underline{n}} = 0$$

for the length of the needle. For the ends

$$\begin{aligned} \sigma_p &= \underline{P} \\ q &= \sigma_p A \\ &= PA \end{aligned}$$

The field at the centre of the needle is that due to the point charges at either end.

$$\begin{aligned} \underline{E} &= \frac{1}{4\pi\epsilon_0}\frac{(PA)}{(\frac{1}{2}d)^2}\,\widehat{\underline{r}} - \frac{1}{4\pi\epsilon_0}\frac{(-PA)}{(\frac{1}{2}d)^2}\,\widehat{\underline{r}} \\ &= \frac{2}{\pi\epsilon_0}\frac{PA}{d}\,\widehat{\underline{r}} \end{aligned}$$

Letting A become vanishingly small, we find that $\underline{E} \to 0$. This means that we can assume that the electric field inside the cavity is not modified, and so we find that $\underline{D} = \underline{D}_0 - \underline{P}$.

The electric field inside the wafer can be modelled via that of a parallel plate capacitor with a charge density $\pm\sigma_p$. We know that $\sigma_p = \underline{P} \cdot \widehat{\underline{n}} = P$.

$$\begin{aligned} \underline{E} &= \left[\frac{\sigma_p}{2\epsilon_0} - \left(-\frac{\sigma_p}{2\epsilon_0}\right)\right]\widehat{\underline{P}} \\ &= \frac{\sigma_p}{\epsilon_0}\,\widehat{\underline{P}} \end{aligned}$$

This means that the total electric field is $\underline{E} = \underline{E}_0 + \underline{E}_c = \underline{E}_0 + 1/\epsilon_0\,\underline{P}$, and so $\underline{D} = \underline{D}_0$.

Linear Materials

Dielectric materials are called *linear*, *homogeneous*, and *isotropic* assuming that their polarisation varies linearly with the electric field; that is,

$$\underline{P} = \epsilon_0 \chi_e \underline{E}$$

The constant of proportionality χ_e is known as the *electric susceptibility* of the material, and is a measure of how the polarisation responds to the applied field. Substitute this result into the definition of $\underline{D}$:

$$\begin{aligned} \underline{D} &= \epsilon_0\underline{E} + \epsilon_0\chi_e\underline{E} \\ &= \underbrace{\epsilon_0(1+\chi_e)}_{\text{permittivity, }\epsilon}\,\underline{E} \end{aligned}$$

We define the coefficient of $\underline{E}$ above as the *permittivity* of the substance, and the *relative permittivity* as simply the ratio $\epsilon_r = \epsilon/\epsilon_0 = 1 + \chi_e$. In a vacuum, $\epsilon_r = 1$, as there is no polarisation. $\underline{D}$ can hence be written as

$$\boxed{\underline{D} = \epsilon\underline{E}} \tag{1.8}$$

This means that in many cases, computation with $\underline{D}$ becomes identical to those readers will be used to with $\underline{E}$, up to a constant.

For linear materials, it can be shown that the internal energy density is given by

$$\boxed{u_E = \frac{1}{2}\underline{E} \cdot \underline{D}} \tag{1.9}$$

Always remember when calculating the energy of systems to take account of the energy required to build up the system. For example, if calculating the energy of a charged sphere, one needs to calculate the energy for both $r \leq R$ *and* $r > R$.

Two long concentric, conducting cylinders of radii a and b, where $a < b$, are separated by a dielectric of permittivity ϵ_1. Find the capacitance per unit length of this system. The dielectric is now removed and the cylinders are placed vertically with one end in the surface of a non-conducting oil of permittivity ϵ_1 and mass density ρ. Taking into account the work done in maintaining a constant voltage on the capacitor, find an expression for z, the height above the surface, that the oil rises in the space between the cylinders.

Throughout this question, we will use a dash to prefer to per-unit-length quantities. Using Gauss's law for dielectrics with a cylindrical Gaussian surface:

$$\begin{aligned}
\int_{\partial V} \underline{D} \cdot d\underline{\Sigma} &= Q_{enclosed} \\
D(2\pi r\ell) &= Q'\,\ell \\
\underline{E} &= \frac{Q'}{2\pi r\epsilon_1}\,\widehat{\underline{r}}
\end{aligned}$$

We have assumed that the dielectric is linear. Finding the potential:

$$\begin{aligned}
V &= -\int \underline{E} \cdot d\underline{r} \\
&= \frac{Q'}{2\pi\epsilon_1}\log\left(\frac{b}{a}\right) \\
\frac{C}{\ell} &= \frac{Q'}{V} = \frac{2\pi\epsilon_1}{\log\left(\frac{b}{a}\right)}
\end{aligned}$$

This is the capacitance per-unit-length of the system. Suppose that the entire capacitor has length ℓ when it is immersed in the oil. Then, the total capacitance is given by

$$\begin{aligned}
C &= \underbrace{\frac{2\pi}{\log\left(\frac{b}{a}\right)}\epsilon_0(\ell - z)}_{\text{fraction out of oil}} + \underbrace{\frac{2\pi}{\log\left(\frac{b}{a}\right)}\epsilon_1 z}_{\text{fraction in oil}} \\
&= \frac{2\pi\epsilon_0}{\log\left(\frac{b}{a}\right)}\left(\ell + z(\epsilon_r - 1)\right)
\end{aligned}$$

Now we have to turn to the result that we saw last year when dealing with moving capacitor plates around, namely that

$$\boxed{dW = dW_{\text{Capacitor}} - dW_{\text{Battery}}} \tag{1.10}$$

For a battery, we know that $W_B = VQ$.

$$\begin{aligned}
dW &= d\left(\frac{1}{2}CV_0^2\right) - d(CV_0^2) \\
&= -\frac{1}{2}V_0^2 dC
\end{aligned}$$

as the battery imposes a constant voltage on the system by definition. This means that the reaction force due to the dielectric being "pulled out" of the system by gravity is given by

$$F_{\text{oil}} = -\frac{dW}{dz} = \frac{1}{2}V_0^2\frac{dC}{dz}$$

The system is in equilibrium when $F_{\text{oil}} = F_{\text{gravity}}$.

$$\begin{aligned}
\frac{1}{2}V_0^2\frac{dC}{dz} &= mg \\
\frac{1}{2}V_0^2\frac{d}{dz}\left(\frac{2\pi\epsilon_0}{\log\left(\frac{b}{a}\right)}\left(\ell + z(\epsilon_r - 1)\right)\right) &= \pi(b^2 - a^2)\rho z g \\
z &= \frac{V_0^2(\epsilon_r - 1)}{(b^2 - a^2)\rho g \log\left(\frac{b}{a}\right)}
\end{aligned}$$

This result behaves sensibly for some easy limits, such as $\epsilon_r = 1$ and $b \to a^+$.

1.2 Magnetic Fields in Matter

We are now going to look at what happens when materials are affected by external magnetic fields. We will adopt the simple Bohr model of the atom, where electrons orbit the nucleus at some fixed radius r_i, and at some fixed speed v_i. Each electron will complete $v_i/2\pi r_i$ orbits per second, and so the current is given by

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

Then, the magnetic moment for each electron is given by

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

The orbital angular momentum of the electron is given by $\underline{L}_i = m_e \underline{r}_i \times \underline{v}_i$, which simplifies to $L_i = m_e r_i v_i$ in the case of a circular orbit. This means that we can write the magnetic moment for each electron as

$$\underline{m}_i = -\frac{e}{2m_e}\underline{L}_i$$

This means that the magnetic moment of the atom is given by

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

In general, the orbital angular momenta of all the electrons in an atom will be randomly orientated, such that $\sum_i \underline{L}_i = 0$. However, when an external magnetic field is applied, these angular momenta align in a particular direction, that creates a net magnetic moment. In this case, there are two types of materials:

- Diamagnetic - These acquire a magnetic moment anti-parallel to the field. This means that the *magnetic susceptibility* χ_m is negative. We will introduce this more formally in coming sections.

- Paramagnetic - These acquire a magnetic moment parallel to the field. This means that χ_m is positive

Note that neither of these have a permanent magnetic moment; it is merely induced as a result of the external field. There are some permanently magnetic materials, which we will cover in Section (1.2.4).

1.2.1 Magnetic Dipoles

In a similar vein to with electric dipoles, let us recap some of the important results of magnetic dipoles. Suppose that we have some closed loop bounding an orientated surface $\underline{\Sigma}$ that carries a current $\underline{I}$. Then, we define the *magnetic dipole moment* as

$$\underline{m} = I\underline{\Sigma}$$

Then, using the results of Section (1.4.2), it can be shown that the magnetic vector potential of a magnetic dipole is

$$\boxed{\underline{A} = \frac{\mu_0}{4\pi}\frac{\underline{m}\times\underline{r}}{r^3}} \tag{1.11}$$

which is quite similar in form to that of the electric monopole; as is usual with magnetic fields, we replace the scalar product by a vector product.

Let us now assume we place the dipole in an external, uniform magnetic field $\underline{B}$. Then, the force on the dipole is given by

$$\underline{F} = \nabla(\underline{m} \cdot \underline{B})$$

Furthermore, the torque is given by

$$\underline{F} = \underline{m} \times \underline{B}$$

The torque will act in a sense such that it reduces its own magnitude, meaning that it attempts to align the normal of $\underline{m}$ with the magnetic field.

1.2.2 Magnetisation

Now, suppose that we have n atoms that have a magnetic moment m per unit volume. Then, the *magnetisation* of the system is given by

$$\boxed{\underline{M} = n \int d\underline{m}} \tag{1.12}$$

The integral form of this expression has been included to remind the reader that calculating the magnetic dipole moment may involve integration.

Calculate the magnetisation of a spinning spherical shell of radius R that has charge density σ.

In this case, we need to integrate over all of the horizontal surface area loops of the sphere that have radius $R\sin\theta$. Thus,

$$\begin{aligned} dI &= \omega\sigma R\sin\theta(Rd\theta) \\ A &= \pi(R\sin\theta)^2 \\ d\underline{m} &= \pi\omega\sigma R^4 \sin^3\theta \, d\theta \, \underline{\hat{z}} \\ \underline{m} &= \frac{4}{3}\pi\sigma\omega R^4 \, \underline{\hat{z}} \end{aligned}$$

This means that the magnetisation of the spherical shell is given by

$$\underline{M} = \sigma\omega R \, \underline{\hat{z}}$$

In a similar way to polarisation, magnetisation can create currents within the material.

- Surface Currents - Consider a slice of the material of width δz though which $\underline{M}$ is constant. The resultant magnetisation of this section is the same as if the entire surface where paved with small current loops. When considering the direction of the currents, the interior currents cancel each other out, so we are left with a single current around the outer edge of the material. These are the surface currents due to magnetisation, and are given by

$$\underline{K}_m = \underline{M} \times \underline{\hat{n}} \tag{1.13}$$

- Volume Currents - If we now consider $\underline{M}$ to be a variable in the z direction, then magnitude of the currents within each loop will vary with z, and so there will not be perfect cancellation. Considering this, it can be shown that the magnetisation gives rise to volume currents given by

$$\underline{J}_m = \nabla \times \underline{M} \tag{1.14}$$

Note that in a similar way to polarisation, these currents are bound. Furthermore, these are not dissipative, as they merely result from the orbital motion and/or spin of the electrons.

1.2.3 The Auxiliary Magnetic Field

Consider Ampere's law in vector form. We now have another type of current to take account of; the magnetisation current $\underline{J}_m$ as well as the free current $\underline{J}_f$.

$$\nabla \times \underline{B} = \mu_0 \underline{J} = \mu_0(\underline{J}_m + \underline{J}_f)$$

Considering (1.14), we can write that

$$\nabla \times \left(\frac{\underline{B}}{\mu_0} - \underline{M}\right) = \underline{J}_f \rightarrow \nabla \times \underline{H} = \underline{J}_f$$

We call the quantity $\underline{H}$ the *auxiliary magnetic field* defined as

$$\boxed{\underline{H} = \frac{\underline{B}}{\mu_0} - \underline{M}} \tag{1.15}$$

In a similar way to the electric displacement vector $\underline{D}$, it is often useful to find $\underline{H}$, and then use it to deduce other quantities.

Linear Materials

Magnetic materials are called *linear*, *homogeneous*, and *isotropic* assuming that their magnetisation varies linearly with the applied magnetic; that is,

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

The constant of proportionality χ_m is known as the *magnetic susceptibility* of the material, and is a measure of how the magnetisation responds to the applied field. Substitute this result into the definition of $\underline{H}$:

$$\begin{aligned} \underline{H} &= \frac{\underline{B}}{\mu_0} - \chi_m \underline{H} \\ \underline{B} &= \underbrace{\mu_0(1 + \chi_m)}_{\text{permeability, } \mu} \underline{H} \end{aligned}$$

We define the coefficient of $\underline{H}$ above as the *permeability* of the substance, and the *relative permeability* as simply the ratio $\mu_r = \mu/\mu_0 = 1 + \chi_m$. In a vacuum, $\mu_r = 1$, as there is no magnetisation. $\underline{B}$ can hence be written as

$$\boxed{\underline{B} = \mu \underline{H}} \tag{1.16}$$

This means that in many cases, computation with $\underline{H}$ becomes identical to those readers will be used to with $\underline{B}$, up to a constant.

Consider a toroidal solenoid of N turns of wire through which a current I passes, of mean radius R. There is a small air gap of width w in the solenoid, such that $w \ll R$. Calculate the magnetic flux density B_g in the gap. Compare the magnitude of the magnetic energy stored in the core, and in the air gap.

The best starting point in answering this question is Ampere's law for magnetic materials in integral form:

$$\begin{aligned} \oint \underline{H} \cdot d\underline{\ell} &= I_f = IN \\ H_{\text{core}}(2\pi R - w) + H_{\text{gap}}(w) &= IN \end{aligned}$$

We know that the magnetic field (but not necessarily the auxiliary field) must be continuous in the core. This means that $B_c = B_g$. Using (1.15),

$$\begin{aligned}\frac{B_c}{\mu}(2\pi R - w) + \frac{B_g}{\mu_0}w &= IN \\ B_g(2\pi R + (\mu_r - 1)w) &= IN\mu \\ B_g &= \frac{IN\mu}{2\pi R + w(\mu_r - 1)}\end{aligned}$$

Evidently, this is a sensible expression; we re-obtain the result for the non-magnetisable solenoid for $\mu_r = 1$. For this, we have to use the fact that the magnetic energy density is given by

$$\boxed{u_B = \frac{1}{2}\underline{B} \cdot \underline{H}} \tag{1.17}$$

Suppose that the cross-section of the core is A. Then we obtain

$$\begin{aligned}W_g &= \frac{1}{2\mu_0}B^2 wA \\ W_c &= \frac{1}{2\mu}B^2(2\pi R - w)A \\ \frac{W_g}{W_c} &= \mu_r \frac{w}{2\pi R - w}\end{aligned}$$

Interestingly, we find that for sensible values of $mu_r \backsim 1500$, R and w, we find that a large proportion of the magnetic energy is actually stored in the air-gap, rather than in the core of the solenoid. This is one of the reasons why solenoids sometimes have an air-gap in them.

1.2.4 Ferromagnetism

Ferromagnetic materials are non-linear magnetic materials that have a permanent magnetisation. This is due to *domains* in the material (regions where the local magnetic moments are all in the same direction) that become aligned under the application of external magnetic field. This alignment takes energy, and leaves the system in an energetically favourable state where many of the adjacent domains are aligned, resulting in a residual magnetisation. That is, *ferromagnetic materials have non-zero magnetisation in the absence of an external magnetic field.* The material will become magnetically *saturated* when all the domains within the material are aligned. This leads to the concept of a *hysteresis curve*. This plots the magnetic field $\underline{B}$ that results from the application of an external field $\underline{H}$ to the ferromagnetic material. This has a few key features:

- Magnetisation Curve (1) - This shows the response of the material under the applied field when it is initially un-magnetised
- Magnetisation Energy - The shaded area gives the energy required to fully magnetise the material
- Saturation field B_{sat} - The strength of the magnetic field in the material at saturation; i.e when all the domains are aligned
- Remnant Field B_r - The (permanent) field that remains when the external field has been reduced to zero
- Coercive Force (2) - This refers to the strength of the external field that is required to demagnetise the material

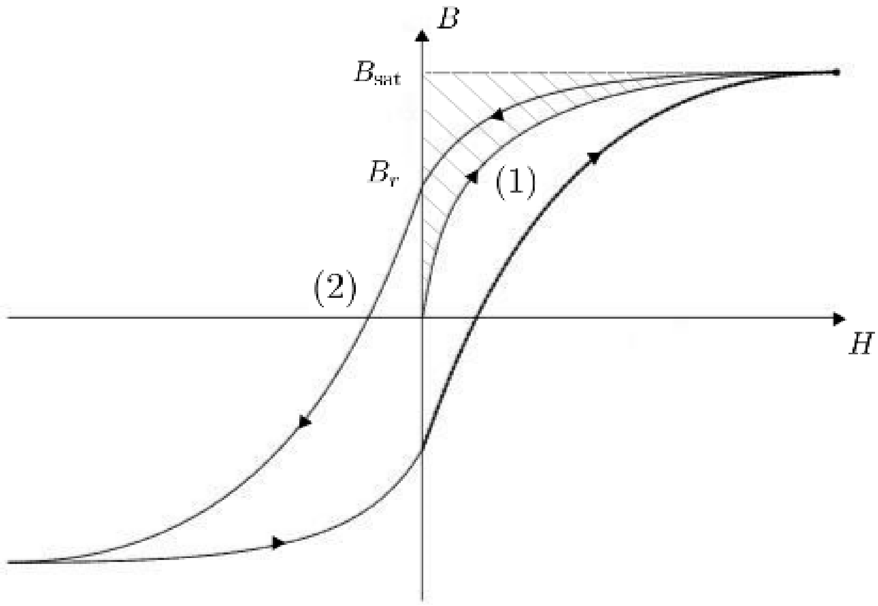

Figure 1.2: A typical hysteresis curve

It can be shown that the work per unit volume moving along the hysteresis curve is given by

$$\boxed{W = \oint \underline{H} \cdot d\underline{B}} \tag{1.18}$$

That is, the area enclosed by the curve. Evidently, this can often not be computed exactly, but can easily be approximated through rough estimation of the area in terms of a rectangle, for example.

Generally, materials that have a narrow hysteresis curve are called *soft*, while those with a wide curve are called *hard*; it takes less energy to go around a soft curve, than a hard one. This is why soft iron is advantageous when it is used as a core in transformers, for example. Iron has a very narrow hysteresis curve, meaning that the energy loss is minimised on application of an external magnetic flux. However, when this is used with alternating currents, there is a frequency limit to how quickly the domains can re-align with one another, limiting the effectiveness of the core.

1.3 Boundary Conditions

Now that we have outlined the way in which electric and magnetic fields are modified in different types of materials, we now turn to looking at the boundary conditions on these fields between two different materials. These results are derived in full generality, taking into account time varying fields. Note that these are summarised at the start of Section (1.4).

1.3.1 Perpendicular Fields

Consider the surface below bounding two media characterised by ϵ_1, μ_1 and ϵ_2, μ_2, and the Gaussian surface of side length ε. Let $\underline{\widehat{n}}_{12}$ be the normal pointing from medium 1 into medium 2.

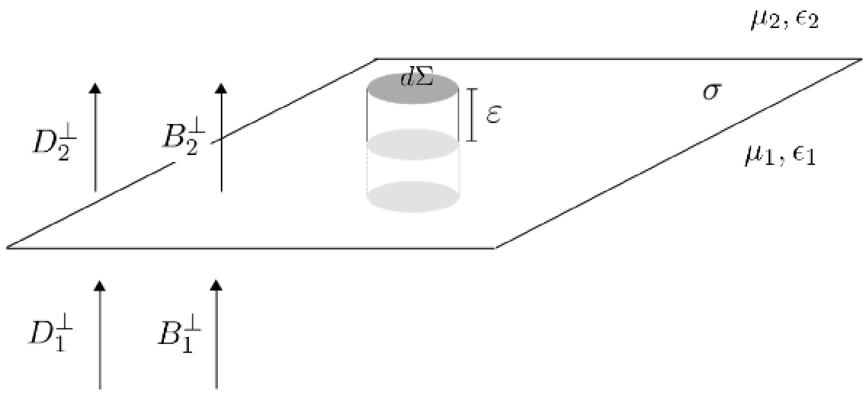

Figure 1.3: Deriving perpendicular field conditions

Consider $\nabla \cdot \underline{D} = \rho_f$. For $\varepsilon \to 0$

$$\left(\underline{D}_2^\perp - \underline{D}_1^\perp\right) d\Sigma = \rho_f \, \underline{\widehat{n}}_{12}$$
$$\to \underline{D}_2^\perp - \underline{D}_1^\perp = \sigma_f \, \underline{\widehat{n}}_{12}$$

Thus, the perpendicular $\underline{D}$ field is discontinuous at the boundary according to the surface charge density on that boundary. This makes sense, as this would chance the polarisation of the field.

Similarly, consider $\nabla \cdot \underline{B} = 0$. Again letting $\varepsilon \to 0$

$$\left(\underline{B}_2^\perp - \underline{B}_1^\perp\right) d\Sigma = 0$$
$$\to \underline{B}_2^\perp - \underline{B}_1^\perp = 0$$

This means that the perpendicular component of the magnetic field is *always* continuous at media boundaries.

1.3.2 Parallel Fields

Consider the surface below bounding two media characterised by ϵ_1, μ_1 and ϵ_2, μ_2, and the Amperian Loop of side length ε. Let $\underline{\widehat{n}}_{12}$ be the normal pointing from medium 1 into medium 2.

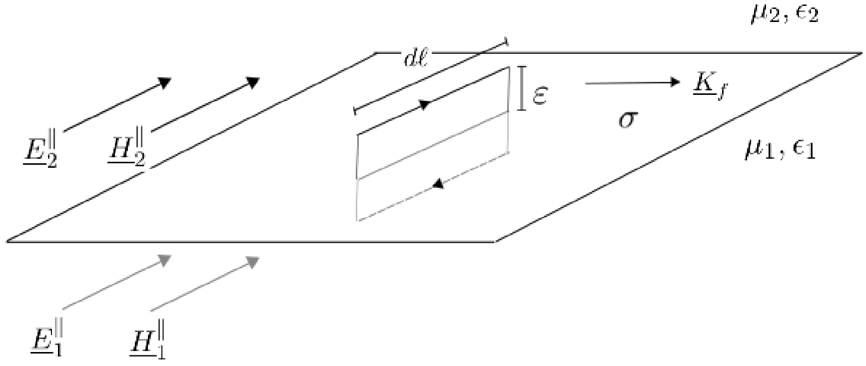

Figure 1.4: Deriving parallel field conditions

Consider $\nabla \times \underline{E} = -\frac{\partial B}{\partial t}$. We then have

$$\left(E_2^{||} - E_1^{||}\right) d\ell = -\frac{\partial}{\partial t} \int_S \underline{B} \cdot d\underline{\Sigma}$$

where $d\underline{\Sigma}$ is the surface bounded by the Amperian loop. For $\varepsilon \to 0$, this becomes negligible. We are thus left with

$$\to \underline{E}_2^{||} - \underline{E}_1^{||} = 0$$

This means that the parallel component of the electric field is always continuous at the boundary between media.

Similarly, consider $\nabla \times \underline{H} = \underline{J}_f + \frac{\partial D}{\partial t}$. Then

$$\left(H_2^{||} - H_1^{||}\right) d\ell = K_f d\ell + \frac{\partial}{\partial t} \int_S \underline{D} \cdot d\underline{\Sigma}$$

Through the same logic as before, the time-varying part of this expression is negligible, and so we arrive at

$$\to \underline{H}_2^{||} - \underline{H}_1^{||} = \underline{K}_f \times \underline{\widehat{n}}_{12}$$

This means that the parallel component of the auxiliary field is discontinuous according to the free surface currents. This makes sense, as this would change the magnetisation.

1.4 Potentials

In previous sections, we introduced the concepts of the $\underline{D}$ and $\underline{H}$ fields for describing electric and magnetic fields in matter respectively. In this light, we can write Maxwell's equations in a new form:

$$\nabla \cdot \underline{D} = \rho_f \tag{1.19}$$

$$\nabla \cdot \underline{B} = 0 \tag{1.20}$$

$$\nabla \times \underline{E} = -\frac{\partial \underline{B}}{\partial t} \tag{1.21}$$

$$\nabla \times \underline{H} = \underline{J}_f + \frac{\partial \underline{D}}{\partial t} \tag{1.22}$$

By imposing certain conditions on these equations, we can write our $\underline{E}$, $\underline{D}$, $\underline{B}$ and $\underline{H}$ fields as scalar or vector potentials, that will allow us to easily solve boundary value problems, as we shall see in the following sections.

It might also be useful to collate the results for the boundary conditions on these fields from the previous section here, particularly seeing as these are used extensively in the solution of potential problems. They are as follows:

$$\underline{D}_2^{\perp} - \underline{D}_1^{\perp} = \sigma_f \, \underline{\widehat{n}}_{12} \tag{1.23}$$

$$\underline{B}_2^{\perp} - \underline{B}_1^{\perp} = 0 \tag{1.24}$$

$$\underline{E}_2^{||} - \underline{E}_1^{||} = 0 \tag{1.25}$$

$$\underline{H}_2^{||} - \underline{H}_1^{||} = \underline{K}_f \times \underline{\widehat{n}}_{12} \tag{1.26}$$

These are relatively easy to commit to memory, but can can always be recalled from their derivations, as well as from Maxwell's Equations. Note that they are valid for all time-dependent scenarios.

1.4.1 Electric Potentials

The Electric field ($\underline{E}$), as we saw in the First Year CP2 course, can be described by a scalar potential. However, it must also be described using a vector potential under conditions. These conditions will be outlined in the following sections.

Scalar Potential

As many readers will already know, we can write the Electric field $\underline{E}$ as the gradient of a scalar potential V as $\underline{E} = -\nabla V$. This is assuming that there are no time dependant magnetic fields ($\dot{\underline{B}} = 0$) by (1.21). Assuming further that $\rho_f = 0$, meaning that by definition the polarisation $\underline{P}$ is uniform, $\underline{E}$ will satisfy Laplace's equation

$$\boxed{\nabla^2 V = 0} \tag{1.27}$$

The solution to this equation in Cartesian, Spherical and Cylindrical coordinates is included in the Mathematical Methods notes that accompany these, and so no further results will be derived here. In general, the result that proves the most useful is the solution to this equation in spherical coordinates, assuming that there is azimuthal symmetry (i.e that V is independent of ϕ), which is usually the case.

$$\boxed{V(r,\theta) = \sum_{\ell=0}^{\infty} \left(A_\ell r^\ell + \frac{B_\ell}{r^{\ell+1}} \right) P_\ell(\cos\theta)} \tag{1.28}$$

Consider a cylindrical hole of radius a and infinite length cut into a dielectric medium with relative electric permittivity ϵ_r (the interior can be treated as a vacuum). Inside the hole there are two line charges of infinite length with line charge densities $+\lambda$ and $-\lambda$, respectively. These line charges are arranged parallel to the z-axis, but displaced from it by an amount $\pm d/2$ along the x-axis (where $d \ll a$). Find the potential everywhere.

As is usual for these kind of questions, let us first consider a diagram of the situation.

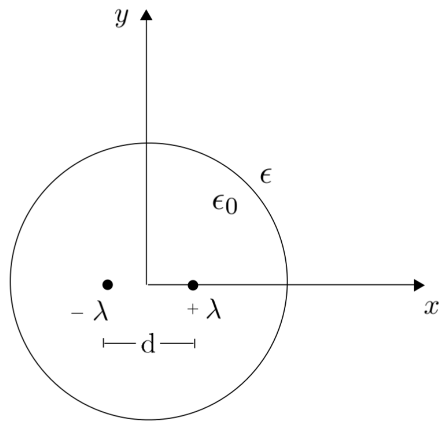

Figure 1.5: Double Boundary Problem

Evidently, at large distances from the origin, the two line charges will appear as an electric dipole. However, what does the field look like for $r \backsim d$? To answer this, we need to consider the potential of a line charge

$$V = -\frac{\lambda}{2\pi\epsilon_0}\log r + V_0$$

for some reference potential V_0. This means that the potential of the entire ensemble is given by

$$\begin{aligned} V_T &= V_+ + V_- \\ &= -\frac{\lambda}{2\pi\epsilon_0}\left[\log\left(\left|\underline{r} - \frac{d}{2}\underline{\widehat{x}}\right|\right) - \log\left(\left|\underline{r} + \frac{d}{2}\underline{\widehat{x}}\right|\right)\right] \end{aligned}$$

Taylor expanding for $d \ll r$:

$$\begin{aligned} \log\left(\left|\underline{r} \mp \frac{d}{2}\underline{\widehat{x}}\right|\right) &= \frac{1}{2}\log\left(r^2 \mp dr\cos\phi + \frac{d^2}{4}\right) \\ &\approx \log r + \frac{1}{2}\log\left(1 \mp \frac{d\cos\phi}{r}\right) \\ &\approx \log r \mp \frac{1}{2}\frac{d\cos\phi}{r} \end{aligned}$$

Putting these results together, we do find that

$$V_T = \frac{\lambda}{2\pi\epsilon_0}\frac{d\cos\phi}{r}$$

That is, the potential is of the form of a dipole for $d \ll r \ll a$. Evidently, in order to match the boundary conditions at $r \to 0$ and $r \to \infty$, we require only the $\ell = 1$ polynomial terms in (1.28). This means that the solution must be of the form

$$V(r,\phi) = \begin{cases} \left(A_1 r + \frac{B_1}{r}\right)\cos\phi & \text{for } r \leq a \\ \frac{B_2}{r}\cos\phi & \text{for } r > a \end{cases}$$

We have automatically done away with the linear term in r for $r > a$ as we require that the solution is properly bounded for large values, as well as the constant and logarithmic terms for similar reasons. Imposing boundary conditions:

- The potential needs to reduce to V_T for $r \to 0$

$$V \to \frac{\lambda}{2\pi\epsilon_0}\frac{d\cos\phi}{r}$$
$$B_1 = \frac{\lambda d}{2\pi\epsilon_0}$$

- Continuity of $\underline{D}^{\perp}$ across the boundary

$$\frac{\partial V}{\partial r}\Big|_{r=a^+} = \epsilon_r \frac{\partial V}{\partial r}\Big|_{r=a^-}$$
$$B_1 - a^2 A_1 = \epsilon_r B_2$$

- Continuity of $\underline{E}^{\parallel}$ across the boundary

$$\frac{\partial V}{\partial \theta}\Big|_{r=a^+} = \frac{\partial V}{\partial \theta}\Big|_{r=a^-}$$
$$B_1 + a^2 A_1 = B_2$$

Solving these equations, one quickly arrives at the solution of

$$V(r,\phi) = \begin{cases} \frac{\lambda}{2\pi\epsilon_0}\frac{d\cos\phi}{r}\left[1 + \left(\frac{1-\epsilon_r}{1+\epsilon_r}\right)\left(\frac{r}{a}\right)^2\right] & \text{for } r < a \\ \frac{\lambda}{2\pi\epsilon_0}\frac{d\cos\phi}{r}\frac{2}{1+\epsilon_r} & \text{for } r \geq a \end{cases}$$

Evidently, this holds for the boundary conditions, and we obtain sensible results if we let $\epsilon_r = 1$; this is the case for there being no boundary. A note to the wise; always do some 'sense-checks' on results that you obtain when doing these sorts of questions; they are very easy to do, and can often result in you finding an error, if it exists.

Vector Potential

In the cases where $\dot{\underline{B}} \neq 0$, we then have to include the vector magnetic potential $\underline{A}$ (see the following section) in order to satisfy Maxwell's equations. From (1.21)

$$\nabla \times \underline{E} = -\frac{\partial}{\partial t}(\nabla \times \underline{A})$$
$$\nabla \times \left(\underline{E} + \frac{\partial \underline{A}}{\partial t}\right) = 0$$

This means that we write the electric field as in a combined potential as

$$\underline{E} = -\nabla V - \frac{\partial \underline{A}}{\partial t}$$

This of course reduces to the familiar case in the absence of time-variant magnetic fields.

1.4.2 Magnetic Potentials

The Magnetic field ($\underline{H}$ or $\underline{B}$), is usually described by a vector potential. However, it can also be described using a vector potential under conditions. These conditions will be outlined in the following sections.

Scalar Potential

Considering (1.22), it is clear that in order to be able to define a scalar potential Ψ for the magnetic field $\underline{H}$, we require that $\underline{J}_f = \dot{\underline{D}} = 0$. Then

$$\boxed{\nabla^2 \Psi = 0} \tag{1.29}$$

assuming that $\nabla \cdot \underline{M} = 0$, which is always the case when there is a uniform polarisation. We thus have three conditions that need to be satisfied in order to be able to write $\underline{H}$ in the above form.

Consider a sphere of radius R made of a material with a permanent magnetisation $\underline{M} = M_0 \widehat{\underline{z}}$. The sphere is embedded in a vacuum, and is placed in a uniform magnetic field $\underline{B}_0 = B_0 \widehat{\underline{z}}$. There are no free currents. By first considering the continuity of the components of $\underline{H}$ and the scalar potential Ψ at the boundary, find the scalar potential everywhere. What is the magnetic field inside the sphere?

One very useful result worth remembering is expressing $\widehat{\underline{z}}$ in polar coordinates as

$$\boxed{\widehat{\underline{z}} = \cos\theta\, \widehat{\underline{r}} - \sin\theta\, \widehat{\underline{\theta}}} \tag{1.30}$$

This is because you will often be working with uniform fields in spherical coordinates, and so it is imperative to be able to convert back and forth between them quickly.

Turning our attention to the problem, we need to apply the boundary conditions (1.24) and (1.26). Clearly, Ψ is continuous as due to the second condition as there are no free currents. The first condition implies that

$$\begin{aligned} \underline{B}_2^\perp - \underline{B}_1^\perp &= 0 \\ \underline{H}_2^\perp - \underline{H}_1^\perp &= \underline{M} \\ H_2^\perp - H_1^\perp &= M_0 \cos\theta \end{aligned}$$

using the fact that $\underline{B} = \mu_0 \underline{H} + \underline{M}$. This means that there is a discontinuity in $\underline{H}$ at the boundary. Evidently, the solutions to Laplace's equation of Ψ are going to be of the form of (1.28), and so

$$\Psi(r,\theta) = \begin{cases} \sum_{\ell=0}^{\infty} A_1 r P_\ell(\cos\theta) & \text{for } r \le R \\ \sum_{\ell=0}^{\infty} \left(A_2 r + \frac{B_2}{r}\right) P_\ell(\cos\theta) & \text{for } r > R \end{cases}$$

We have already done away with the B_1 terms in the expansion as we require the solution to be finite for $r \to 0$.

$$\lim_{r\to\infty} \Psi = -\int dz\, H_z = -\frac{B_0}{\mu_0} z = -\frac{B_0}{\mu_0} r\cos\theta \tag{1.31}$$

This means that we require the $\ell = 1$ terms in the expansion in order to match the above condition. Imposing the remaining boundary conditions, we obtain the equations

$$\begin{aligned} A_1 &= A_2 + \frac{B_2}{R^3} \\ A_1 + A_2 + \frac{2B_2}{R^3} &= M_0 \end{aligned}$$

Solving these two equations, and finding A_2 by (1.31), we arrive at the find solution of

$$\Psi(r,\theta) = \begin{cases} \left(\frac{M_0}{3} - \frac{B_0}{\mu_0}\right) r\cos\theta & \text{for } r \leq R \\ \left(\frac{M_0 R^3}{3r^2} - \frac{B_0}{\mu_0} r\right)\cos\theta & \text{for } r > R \end{cases}$$

This means that we can calculate the magnetic field inside the sphere.

$$\begin{aligned} \underline{H}_{\text{in}} &= -\nabla\Psi_{\text{in}} = \frac{\underline{B_0}}{\mu_0} - \frac{\underline{M}_0}{3} \\ \underline{B}_{\text{in}} &= \underbrace{\underline{B}_0}_{\text{external field}} + \underbrace{\frac{2}{3}\mu_0 \underline{M}}_{\text{magnetisation field}} \end{aligned}$$

This would have been a horrible problem to solve if we had not written $\underline{H}$ as a scalar potential.

Vector Potential

Evidently, as $\nabla \cdot \underline{B} = 0$ in all cases, it is always well defined to write a vector potential such that

$$\underline{B} = \nabla \times \underline{A}$$

We can calculate this vector potential from

$$\boxed{\underline{A} = \frac{\mu_0}{4\pi}\int d\underline{\ell}\,\frac{I}{r'}} \tag{1.32}$$

Calculate the magnetic vector potential $\underline{A}$ at a point P a perpendicular distance r away from the centre of a wire of length $2L$ that carries a steady current I. Verify that this result agrees with that expected from Ampere's law.

In this case, $d\underline{\ell}$ is simply the line integral along the wire. Thus

$$\begin{aligned} \underline{A} &= \frac{\mu_0 I}{4\pi}\int_{-L}^{L} \frac{\widehat{\underline{z}}\,dz}{r'} \\ &= \frac{\mu_0 I}{4\pi}\int_{-L}^{L} \frac{\widehat{\underline{z}}\,dz}{\sqrt{z^2+r^2}} \\ &= \frac{\mu_0 I}{4\pi}\log\left(\frac{\sqrt{L^2+r^2}+L}{\sqrt{L^2+r^2}-L}\right)\widehat{\underline{z}} \end{aligned}$$

Then, the magnetic field is given by

$$\begin{aligned} \underline{B} &= -\frac{\partial A}{\partial z}\,\widehat{\underline{\phi}} \\ &= \frac{\mu_0 I}{2\pi r}\,\frac{L}{\sqrt{r^2+L^2}}\,\widehat{\underline{\phi}} \end{aligned}$$

Taking the limit as $L \to \infty$, we see that this clearly agrees with the Ampere's Law result that readers will be familiar with.

1.5 Electromagnetic Waves in a Vacuum

Before delving into new and exciting territory, it is customary to return to the mundane, and review some material that should already be known for the sake of familiarity. In such a vein, let us now examine electromagnetic waves in a vacuum (**gasp* has Christmas come early?*).

1.5.1 Waves in a Vacuum

Using the standard derivations set out in the CP2 notes (have a look at them if you are still unsure), we obtain the wave equations

$$\nabla^2 \underline{E} = \mu_0 \epsilon_0 \frac{\partial^2 \underline{E}}{\partial t^2} \quad \text{and} \quad \nabla^2 \underline{B} = \mu_0 \epsilon_0 \frac{\partial^2 \underline{B}}{\partial t^2}$$

which have solutions of the form

$$\underline{E} = \underline{E}_0 e^{i(\omega t - \underline{k}\cdot\underline{r})} \quad \text{and} \quad \underline{B} = \underline{B} e^{i(\omega t - \underline{k}\cdot\underline{r})}$$

respectively. Note that sometimes a tilde is used to denote complex quantities; this author finds that this is generally a waste of time; instead, assume that everything has the potential to be a complex quantity. Substituting these solutions into Maxwell's equations, we obtain the familiar relations of

$$\begin{aligned} \underline{k} \cdot \underline{E} &= 0 \\ \underline{k} \cdot \underline{B} &= 0 \\ \underline{B} &= \frac{i\underline{k} \times \underline{E}}{\omega} \end{aligned}$$

The electric and magnetic fields of this wave are orthogonal to one another, as well as being orthogonal to the direction of propagation, and so the wave is a *plane monochromatic wave*. Substitution of the solutions into the wave equations yields the *dispersion relation* of:

$$v_p = \frac{\omega}{k} = \frac{1}{\sqrt{\mu_0 \epsilon_0}}$$

Electromagnetic waves thus travel at the speed of light in a vacuum; nothing particularly surprising thus far.

1.5.2 The Poynting Vector

Some readers may already be familiar with the Poynting Vector, as it was introduced in the CP2 course, but a full treatment shall be detailed here. The *Poynting Vector* is defined as

$$\boxed{\underline{S} = \underline{E} \times \underline{H}} \tag{1.33}$$

This is the flux per-unit-area, per-unit-time of the electromagnetic radiation. It is thus very common to have to evaluate a flux-integral over some bounding surface. Up until this point, $d\underline{S}$ has been used to refer to the surface area element; evidently, this will just become confusing, so this will be changed to $d\underline{\Sigma}$

For the case where there is a steady increase in the current of $\dot{I}$, calculate the magnitude and direction of the Poynting vector $\underline{S}$ at the outer radius of an infinite solenoid of radius a. By performing appropriate integration, show that the energy stored in the electromagnetic field and the total work done are equal.

We shall use a dash to refer to per-unit-length quantities in this question. Assuming that $\epsilon_r = \mu_r = 1$, we can write that

$$\begin{aligned}
\int_{\partial S} \underline{E} \cdot d\underline{\ell} &= -\frac{\partial}{\partial t} \int \underline{B} \cdot d\underline{\Sigma} \\
E'(2\pi a) &= -\frac{\partial}{\partial t} \left(\mu_0 I n \pi a^2\right) \\
\underline{E'} &= -\frac{1}{2}\mu_0 n a \dot{I} \underline{\phi}
\end{aligned}$$

We will quote the result for the magnetic field around an infinite solenoid as

$$\underline{B} = \mu_0 I n \underline{\widehat{z}}$$

meaning that the Poynting vector is given by

$$\begin{aligned}
\underline{S} &= \frac{1}{\mu_0} \underline{E} \times \underline{B} \\
&= -\frac{1}{2}\mu_0 n^2 a I \frac{dI}{dt} \underline{\widehat{r}}
\end{aligned}$$

Then the work done is given by

$$\begin{aligned}
U' &= \iint \underline{S} \cdot d\underline{\Sigma}\, dt \\
&= -\frac{1}{2}\mu_0 n^2 a \int dS \int dt\, I \frac{dI}{dt} \\
&= -\frac{1}{2} \underbrace{(\mu_0 n^2 a^2 \pi)}_{\text{inductance per unit length, } L'} I^2 \\
&= -\frac{1}{2} L I^2
\end{aligned}$$

We note that this is opposite and equal to the energy required to raise the current in a magnetically inductive system from 0 to some value I, as required.

1.5.3 Energy Conservation

Consider the work done by an electric field on a single charge moving at velocity $\underline{v}$.

$$dW = \underline{F} \cdot d\underline{\ell} = q\underline{E} \cdot d\underline{\ell} = q\underline{E} \cdot \underline{v}\, dt$$

Assuming that this single charge is part of a larger ensemble of n charges per unit volume that are only weakly interacting, then we can write that

$$\begin{aligned}
\frac{dW}{dt} &= \int_V dV\, nq\underline{v} \cdot \underline{E} \\
&= \int_V dV\, \underline{J}_f \cdot \underline{E}
\end{aligned}$$

Thus, we find that the work done on charges in some volume is equal to the Ohmic Heating within that volume, as expected. Considering (1.22):

$$\begin{aligned}
-\underline{J}_f \cdot \underline{E} &= -\left(\nabla \times \underline{H} - \frac{\partial \underline{D}}{\partial t}\right) \cdot \underline{E} \\
&= -\left(\nabla \times \underline{H} - \frac{\partial \underline{D}}{\partial t}\right) \cdot \underline{E} + \underbrace{\left(\nabla \times \underline{E} + \frac{\partial \underline{B}}{\partial t}\right) \cdot \underline{H}}_{=0 \text{ from Maxwell's equations}}
\end{aligned}$$

$$
\begin{aligned}
&= \underline{E} \cdot \frac{\partial \underline{D}}{\partial t} + \underline{H} \cdot \frac{\partial \underline{B}}{\partial t} + \underline{H} \cdot \nabla \times \underline{E} - \underline{E} \cdot \nabla \times \underline{H} \\
&= \frac{\partial}{\partial t} \underbrace{\left(\frac{1}{2} \underline{E} \cdot \underline{D} + \frac{1}{2} \underline{B} \cdot \underline{H} \right)}_{\text{internal energy density, } u} + \nabla \cdot \underbrace{(\underline{E} \times \underline{H})}_{\text{Poynting Vector, } \underline{S}}
\end{aligned}
$$

Making the labelled substitutions, we arrive at the energy conservation result of

$$
\boxed{\frac{\partial u}{\partial t} + \nabla \cdot \underline{S} = -\underline{J}_f \cdot \underline{E}} \tag{1.34}
$$

Thus, it is clear that the rate of change of energy within a volume is a opposite and equal to the flux of energy through the surface bounding that volume plus the amount of work dissipated on the charges within that volume.

1.5.4 Energy and Momentum Transport

For plane electromagnetic waves, we know that $B^2 = E^2/c^2$. This means that we can write the internal energy density as

$$
u = \epsilon_0 E^2 \quad \text{or} \quad u = \frac{1}{\mu_0} B^2
$$

Then, the Poynting Vector can be written simply as $\underline{S} = uc\,\widehat{\underline{k}}$, where $\widehat{\underline{k}}$ is the direction of propagation of the wave. The *intensity* of a wave is the average power transported per unit area, and is given by

$$
\boxed{I = \langle \underline{S} \rangle} \tag{1.35}
$$

Always be very careful when calculating time-averages; always convert any complex quantities into their real parts before doing so, or else everything averages to zero (or has an undefined average!).

Let us consider the *radiation pressure* of the waves. The energy communicated to some surface area element in some time interval dt is $Sd\Sigma dt$, meaning that the momentum communicated is $Sd\Sigma dt/c$. The force (momentum per unit time) is then $Sd\Sigma/c$, meaning that the pressure (force per unit area) communicated to the surface is

$$
p_{rad} = \frac{I}{c}
$$

This corresponds to the component of the Lorentz force along the direction normal to the surface at incidence. If the wave is reflected, then the radiation pressure is doubled (in a similar way to calculating the pressure in Kinetic Theory). Note that the lower case 'p' has been used so as not to confuse it with power; the power of the wave is simply the intensity multiplied by the area.

Note that for 'macroscopic' quantities of Electromagnetic radiation, such as those associated with energy and momentum transport, it is usually fine to refer to quantities in terms of their time average. This is because, in general, the frequency of the electromagnetic radiation will be on the order of $\backsim 10^6$, and so the instantaneous value would vary so quickly as to be almost meaningless to talk about!

1.6 Electromagnetic Waves in Linear Media

We know that in perfectly linear media that $\rho_f = J_f = 0$. This means that Maxwell's equations reduce to the same as those in a vacuum, but with $\epsilon_r \neq 1$ and $\mu_r \neq 1$, and so we obtain the same wave equations, except with the replacements $\epsilon_0 \mapsto \epsilon$ and $\mu_0 \mapsto \mu$. This means that the dispersion relation is clearly

$$k^2 = \mu\epsilon\omega^2 \longrightarrow v_p = \frac{1}{\sqrt{\mu\epsilon}}$$

With this knowledge in mind, we can define two important quantities.

- Refractive Index - Most readers will already be familiar with this concept, but not necessarily aware of how it is defined. Suppose that we wanted to write the phase velocity of a wave as some proportion of the speed of light. As we know that it has to be less than c, it would make sense to write $v_p = c/n$ for $n > 1$. Using the definition of c and the dispersion relation for v_p above, we find that

$$\boxed{n = \sqrt{\mu_r\epsilon_r}} \tag{1.36}$$

  This is evidently a well-defined quantity as both μ_r and ϵ_r are bounded > 1. Note that this relation assumes that n is independent of the frequency of the wave being considered; we will consider waves where this is not the case later.

- Impedance - Again, most readers may have seen this term in the context of circuit theory, and not necessarily when applied to waves. For electromagnetic radiation, it is defined as

$$Z = \frac{E_{ph}}{H_{ph}} \tag{1.37}$$

  The subscript '$_{ph}$' stands for 'phasor', and simply refers to the complex amplitudes of the components after the time-dependant components have been factored out. For most waves, this reduces to the ratio of the magnitudes of the components. We find that

$$\boxed{Z = \sqrt{\frac{\mu}{\epsilon}} \backsim \mu_0\frac{c}{n}} \tag{1.38}$$

  The last of the above expressions follows from that fact that in most materials, $\mu \backsim \mu_0$. This is quite a useful expression, and well worth remembering for the coming sections.

Generally, the treatment of electromagnetic waves in different sorts of media is very similar. We first derive the wave equation, then assume plane-wave solutions, and substitute them in. This will then give us some dispersion relation that will allow us to simply the form of the wave.

1.6.1 Normal Incidence

Let us now turn to examining the reflection and transmission of electromagnetic waves at the boundaries between media. Readers will find that this is very similar to the reflection and transmission of one-dimensional waves at a boundary; this is simply an extension of those kind of problems to three dimensions, and using slightly different boundary conditions.

The first case we will consider is that where the light is normally incident on the infinite planes bounding the various regions. We shall use the convention that the bounding planes are planes of constant z, and hence that the variation in $\underline{E}$ and $\underline{H}$ occurs in planes of constant z in the case of normal incidence.

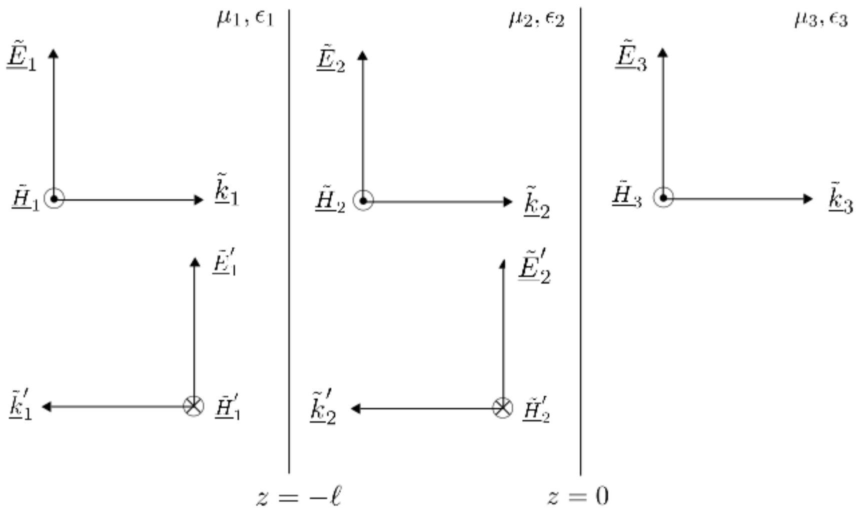

Figure 1.6: Double Boundary Problem

The above diagram shows the waves present in three consecutive regions, assuming that that the waves were incident from $z = -\infty$ on the left-hand side. We will treat the single boundary case, and then work up to this slightly more involved example.

Suppose that $\mu_2 = \mu_3$, $\epsilon_2 = \epsilon_3$, meaning that $\underline{E}_2 = \underline{E}_3$ and $\underline{E}'_2 = 0$ (there is no boundary, and so no reflected wave). We have thus reduced the problem to the single boundary at $z = -\ell$, which we can then shift to $z = 0$ without loss of generality. At normal incidence, all components of the $\underline{E}$ and $\underline{H}$ fields lie parallel to the bounding surfaces, and so we make use of (1.25) and (1.26) to write that

$$E_1 + E'_1 = E_2 \quad \text{and} \quad H_1 - H'_1 = H_2$$

The negative sign in front of the reflected $\underline{H}$ component is due to the fact that it changes direction in order to maintain the orthogonality of the components of the wave. Then, using the definition of Z

$$E_1 + E'_1 = E_2 \quad \text{and} \quad \frac{E_1}{Z_1} - \frac{E'_1}{Z_1} = \frac{E_2}{Z_2}$$

These are two equations in to unknowns that we can solve for the reflection and transmission coefficients of

$$\boxed{\frac{E_2}{E_1} = \frac{Z_2 - Z_1}{Z_2 + Z_1}} \quad \text{and} \quad \boxed{\frac{E'_1}{E_1} = \frac{2Z_2}{Z_2 + Z_1}} \tag{1.39}$$

These can easily be written in terms of the refractive indicies of the media by making use of (1.38) as required. Check the limiting cases of $Z_2 \to 0$ and $Z_2 \to \infty$; do they make sense?

We shall now turn our attention to the original problem of Figure (1.6). *What is the condition for no reflection at $z = -\ell$ for some value of ℓ?* Applying the boundary conditions at $z = 0$ we obtain

$$E_2 + E'_2 = E_3 \tag{1.40}$$

$$\frac{E_2}{Z_2} - \frac{E'_2}{Z_2} = \frac{E_3}{Z_3} \tag{1.41}$$

Similarly, imposing the boundary conditions at $z = -\ell$ yields

$$E_1 e^{ik_1\ell} + E_1' e^{-ik_1\ell} = E_2 e^{ik_2\ell} + E_2 e^{-ik_2\ell}$$
$$\frac{E_1}{Z_1} e^{ik_1\ell} - \frac{E_1'}{Z_1} e^{-ik_1\ell} = \frac{E_2}{Z_2} e^{ik_2\ell} - \frac{E_2'}{Z_2} e^{-ik_2\ell}$$

The usual value encountered in this sort of problem is $\ell = \lambda_2/4$ (quarter-wave plate), where λ_2 is the wavelength of the wave in the second medium. If this is the case, then the second set of equations become

$$E_1 e^{ik_1\ell} + E_1' e^{-ik_1\ell} = -i(E_2 - E_2') \tag{1.42}$$
$$\frac{E_1}{Z_1} e^{ik_1\ell} - \frac{E_1'}{Z_1} e^{-ik_1\ell} = -\frac{i}{Z_2}(E_2 + E_2') \tag{1.43}$$

We can use (1.40) and (1.41) to eliminate E_2 and E_2' from (1.42) and (1.43) to obtain

$$E_1 e^{ik_1\ell} + E_1' e^{-ik_1\ell} = -iE_3 \frac{Z_2}{Z_3}$$
$$E_1 e^{ik_1\ell} - E_1' e^{-ik_1\ell} = -iE_3 \frac{Z_1}{Z_2}$$

By inspection, we can only have $E_1' = 0$ in the above two equations if

$$\boxed{Z_2 = \sqrt{Z_1 Z_3}} \tag{1.44}$$

Thus, the impedance of this layer has to be equal to the geometric mean of the media on either side of it. This also happens to be the 'zero reflection condition' for similar problems in different systems, such as transmission lines.

1.6.2 Parallel Polarisation

Thus far, we have just considered the simple case in which there were no components of the $\underline{E}$ and $\underline{H}$ fields perpendicular to the bounding planes. Let us now delve into that whimsical world. We shall first consider the case where the light is *polarised* parallel to the *plane of incidence*, as shown in the figure below. The concept of polarisation is covered fully in Section (3.3), but for now it simply refers to the direction of the electric field component of the wave.

For the boundary to be properly defined, we require that the boundary conditions hold at every point within the $z = 0$ plane, and for all times. This means that the exponential factors must be equal. As the boundary conditions are themselves essentially a set vector equations, they must also hold component-wise. We can thus write that

$$\boxed{k_i \sin\theta_i = k_r \sin\theta_r = k_t \sin\theta_t} \tag{1.45}$$

From this, we can extract two familiar results.

- Law of Reflection - As $k_i = k_r$ as the waves are travelling in the same medium, we find that
$$\theta_i = \theta_r \tag{1.46}$$
- Snell's Law - Using the fact that $k \propto n$, we find that
$$n_i \sin\theta_i = n_t \sin\theta_t \tag{1.47}$$

It is quite satisfying that such seemingly 'fundamental' results simply drop out from a consideration of component-wise boundary conditions. In general, it is only worth remembering how to show (1.45), as these two results follow trivially from it.

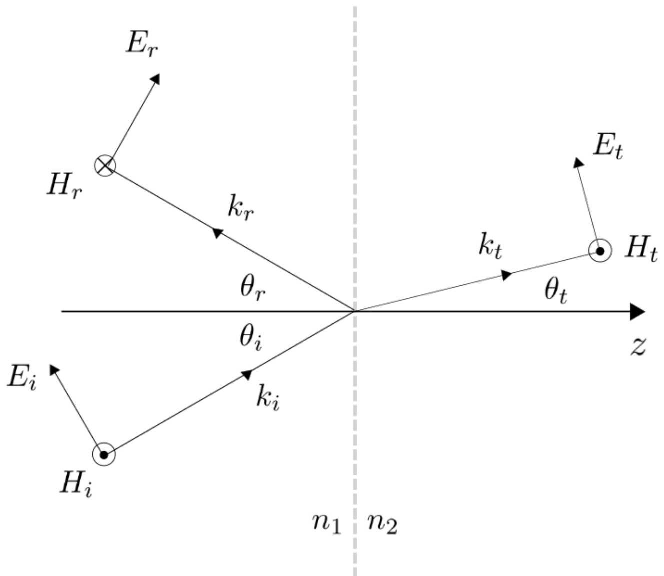

Figure 1.7: Parallel polarised light at non-normal incidence

As before, we can impose the conditions (1.25) and (1.26) on the field components, but taking about of the directions of said components. We then obtain

$$E_i \cos\theta_i + E_r \cos\theta_r = E_t \cos\theta_t$$
$$\frac{E_i}{Z_i} - \frac{E_r}{Z_i} = \frac{E_t}{Z_t}$$

As usual, we can simultaneously solve these equations to obtain the reflection coefficient as

$$\boxed{\left.\frac{E_r}{E_i}\right|^{\parallel} = \frac{Z_t \cos\theta_t - Z_i \cos\theta_i}{Z_t \cos\theta_t + Z_i \cos\theta_i}} \tag{1.48}$$

We will not explicitly derive the transmission coefficient here, as it is a relatively trivial calculation and is left as an exercise for the reader. Instead, let us ask ourselves the question of under what conditions we obtain no reflection?

$$\begin{aligned} 0 &= Z_t \cos\theta_t - Z_i \cos\theta_i \\ &= \sin 2\theta_t - \sin 2\theta_i \\ &= \cos(\theta_i + \theta_t)\sin(\theta_i - \theta_t) \end{aligned}$$

This means that the condition is clearly

$$\theta_i + \theta_t = \frac{\pi}{2}$$

Substituting this result into Snell's Law, we obtain an expression for *Brewster's Angle*

$$\boxed{\theta_B = \tan^{-1}\left(\frac{n_2}{n_1}\right)} \tag{1.49}$$

Do not get this confused with θ_c, the critical angle for *total internal reflection*; θ_B is always defined, where as θ_c is only defined for $n_1 < n_2$.

1.6.3 Perpendicular Polarisation

This is very similar to the case above, except we reverse the orientations of the $\underline{E}$ and $\underline{H}$ fields for each of the waves. This means, when we impose the boundary conditions, that we obtain

$$E_i + E_r = E_t$$
$$\frac{E_i}{Z_i}\cos\theta_i - \frac{E_r}{Z_i}\cos\theta_i = \frac{E_t}{Z_t}\cos\theta_t$$

Similar solving of these two equations as above yields the reflection coefficient as

$$\boxed{\frac{E_r}{E_i}^{\perp} = \frac{Z_t\cos\theta_i - Z_i\cos\theta_t}{Z_t\cos\theta_i + Z_i\cos\theta_t}} \tag{1.50}$$

Again, let us ask ourselves the same question about the conditions under which we obtain no reflection.

$$\begin{aligned} 0 &= Z_t\cos\theta_i - Z_i\cos\theta_t \\ &= \sin(\theta_i - \theta_t) \end{aligned}$$

This means that the condition is clearly

$$\theta_i = \theta_t$$

However, this is a meaningless condition, as for the angle to be unchanged, we would require that $n_1 = n_2$, meaning that there is in fact no boundary, and thus would be no reflection anyway.

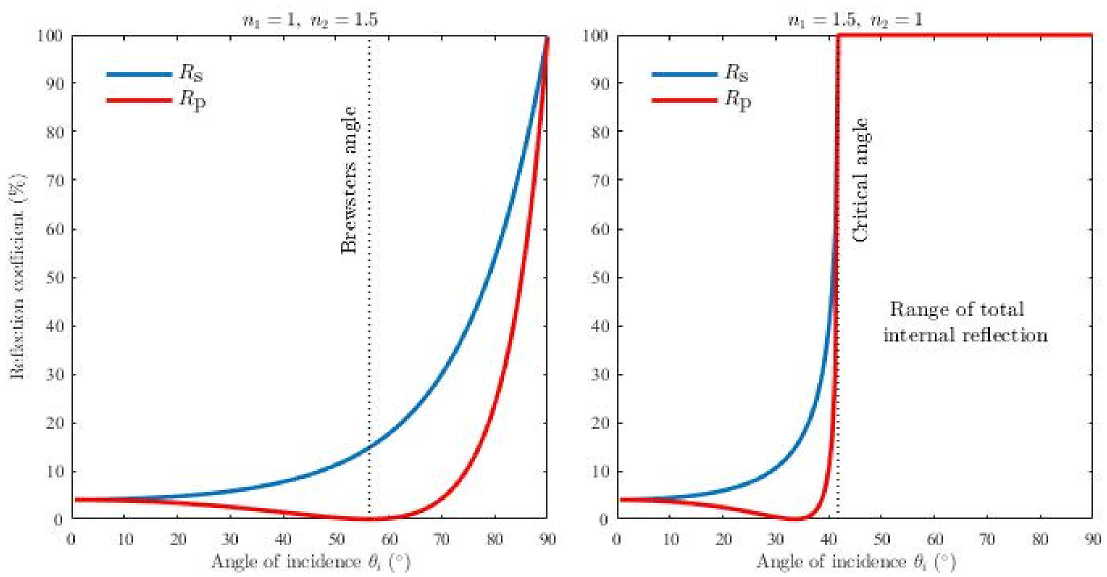

Figure 1.8: Reflection Coefficients. R_p corresponds to polarisation parallel to the plane of incidence, while R_s corresponds to perpendicular.

1.7 Electromagnetic Waves in Conductors

In a conductor, we cannot assume that $\rho_f = 0$ and $J_f = 0$. If we assume that the density of free charges satisfies charge conservation, and the conductor obeys Ohmic heating ($\underline{J}_f = \sigma\underline{E}$), we can write that:

$$\begin{aligned} \nabla \cdot \underline{J}_f &= -\frac{\partial \rho_f}{\partial t} \\ \frac{\partial \rho_f}{\partial t} &= -\frac{\sigma}{\epsilon}\rho_f \\ \rho_f &= \rho_0 e^{-t/\tau} \end{aligned}$$

The quantity τ is known as the *relaxation time* of the conductor. We have two limiting cases:

- Perfect Conductor - In such a material, we have that $\sigma \to \infty$, and so $\tau \to 0$. This means that the relaxation time is negligible, and so the free charges within the material can move instantaneously to cancel out the external fields.
- Perfect Insulator - In such a material, we have that $\sigma \to 0$, and so $\tau \to \infty$. This means that the relaxation time is very large, and so the free charges within the material cannot move at all the cancel the external fields.

For the remainder of this section, we will be working within the domain of the former case, though interested readers are welcome to work through the following derivations in the opposite limit to see if they can re-obtain the results of Section (1.6).

1.7.1 Dispersion Relation

In conductors, we thus have Maxwell's Equations in their full form as given in Section (1.4) (though assumed linear). Taking the curl of (1.21), and using (1.22), it follows quickly that

$$\nabla^2 \underline{E} = \underbrace{\mu\epsilon\frac{\partial^2 \underline{E}}{\partial t^2}}_{\text{Displacement current}} + \underbrace{\mu\sigma\frac{\partial \underline{E}}{\partial t}}_{\text{Conduction current}}$$

We obtain a similar wave equation for $\underline{B}$. Assuming plane-wave solutions as previously, we arrive at the dispersion relation of

$$\boxed{k^2 = i\mu\sigma\omega + \mu\epsilon\omega^2} \tag{1.51}$$

It can be useful at this stage to write $k = k' + ik''$, such that we can make some simplifications of this dispersion relation under certain conditions.

$$\begin{aligned} k'^2 - k''^2 &= \mu\epsilon\omega^2 \\ 2k'k'' &= \mu\sigma\omega \end{aligned}$$

Note that we can solve these equations to find explicit expressions for k' and k'', but this is usually not required. This is because, for a good conductor, the magnitude of the displacement current can be considered negligible in comparison to the Conduction current, such that $\sigma \gg \epsilon\omega$. The crossover frequency between the two regimes thus occurs for $\omega \backsim \sigma/\epsilon$. Under the good conductor condition, we have that

$$\begin{aligned} k'^2 - k''^2 \backsim 0 \to k' &= k'' \\ k^2 &= i\mu\sigma\omega \\ k &= \sqrt{\frac{\mu\sigma\omega}{2}}(1+i) \end{aligned}$$

where we have chosen the solution that is appropriately bounded for $\underline{r} \to \infty$. The quantity

$$\boxed{\delta = \sqrt{\frac{2}{\mu\sigma\omega}}} \tag{1.52}$$

is known as the *skin depth* of the conductor, and is a measure of the distance over it which it takes the amplitude of the signal to fall to $1/e$ of it's original value. Evidently, in the limit of high conductivity, $\delta \to 0$, and so for a perfect conductor, no fields actually penetrate into the material.

1.7.2 Energy Transport

Given that the amplitude of the wave is decaying, it would be interesting to look at how the energy is transported with the wave. Through the careful treatment of the real parts of the quantities involved, it can be shown that

$$\langle \underline{S} \rangle = \hat{\underline{z}} \frac{E_0^2}{2\mu_0} e^{-2k''z} \frac{k'}{\omega}$$

Similarly, it can be shown that

$$\langle u \rangle = \frac{E_0^2}{4} e^{-2k''z} \left(\frac{\sigma}{\omega} + \epsilon_0\right) \backsim \frac{E_0^2}{2\mu_0} e^{-2k''z} \left(\frac{k'}{\omega}\right)^2$$

The terms in the brackets are the magnetic and electric contributions to the energy respectively. Imposing $\sigma \gg \epsilon\omega_0$, we find the interesting result that the magnetic contribution always dominates; *the energy is mostly carried by the magnetic field in good conductors.* The velocity at which the energy is transported is given by

$$v_\varepsilon = \frac{\langle \underline{S} \rangle}{\langle u \rangle}$$

Comparing the two expressions above, we find that $v_\varepsilon = \delta\omega$, instead of the group velocity of the wave as is the usual result. This is because if $\lambda = 2\pi/k' = 2\pi\delta \backsim \delta$, then the envelope through which we define group velocity becomes undefined; it becomes meaningless to talk about group velocity!

1.7.3 Wave Impedance

If we substitute the solutions for $\underline{E}$ and $\underline{B}$ into (1.21), then we find that

$$\left(-ik - \frac{1}{\delta}\right) \tilde{E}_{ph} = \mu \tilde{H}_{ph}(-i\omega)$$

Re-arranging, and using the definition of Z, we find that

$$\boxed{Z = \frac{1}{\delta\sigma}(1 - i)} \tag{1.53}$$

In the limit of a perfect conductor, $\delta \to 0$, meaning that $Z \to \infty$. This is in accordance with the finding that the wave does not actually penetrate into the conductor.

1.8 Dispersion and Plasmas

Our aim in this section is to come up with an approximate model for *dispersion*; that is, the phenomenon where-by different frequencies travel at different rates within the same material. One such way of doing this is by the Lorentz oscillator model, where we model the electrons as being classically coupled to their atoms. This model evidently does not take account of the effects of Quantum Mechanics, but it does make some good predictive results.

Initially, let us assume that the oscillations of the electrons are small enough around their equilibrium point such that they can be approximated by a linear spring force $m\omega_0^2\underline{x}$, with some damping term $m\gamma\dot{\underline{x}}$. By Newton's Second Law, we can write the equation of motion of the electron as

$$m\ddot{\underline{x}} = q\underline{E} - m\gamma\dot{\underline{x}} - m\omega_0^2\underline{x}$$

Assuming that the external electric field $\underline{E}$ is sinusoidally varying, we can let $\underline{x} = e^{i\omega t}$

$$-\omega^2 m\underline{x} = q\underline{E} - im\omega\gamma\underline{x} - m\omega_0^2\underline{x}$$
$$\underline{x} = \frac{q\underline{E}}{m\left((\omega_0^2 - \omega^2) - i\gamma\omega\right)}$$

Let the average number density of electrons within the material be n. Recalling that $\underline{p} = q\underline{x}$:

$$\underline{P} = n\underline{p}$$
$$= \frac{nq^2}{m}\frac{\underline{E}}{\left((\omega_0^2 - \omega^2) - i\gamma\omega\right)}$$

Assuming that the material is linear, $\underline{P} = \epsilon_0\chi_e\underline{E}$. Then, as $\epsilon_r = 1 + \chi_e$, we find that

$$\epsilon_r = 1 + \frac{nq^2}{\epsilon_0 m}\frac{1}{(\omega_0^2 - \omega^2) - i\gamma\omega}$$

In reality, each coupled system can have an infinite number of energy levels and associated frequencies, as well as there being multiple oscillators (ie. more than one electron) per atom. Let us suppose that there are f_j electrons per molecule, that each have some natural frequency ω_j and some damping constant γ_j. This allows us to arrive at the slightly more comprehensive result of

$$\boxed{\epsilon_r(\omega) = 1 + \frac{nq^2}{\epsilon_0 m}\sum_j \frac{f_j}{(\omega_j^2 - \omega^2) - i\gamma_j\omega}} \tag{1.54}$$

There are two main types of dispersion that can arise:

- Normal dispersion - Lower frequencies travel slower than higher frequencies. This is what we observe in most materials
- Anomalous dispersion - Higher frequencies travel slower than lower frequencies. Dispersion is always said to be anomalous if $v_g > v_p$

It is important to bear in mind that in some cases of anomalous dispersion that the velocity at which the energy is transported is often not the group velocity, as this is often not well-defined. We saw this in the case of the conductor in the previous section.

1.8.1 Plasmas

Plasmas are gases of ionised particles that can be modelled as being roughly neutral. There are many interesting aspects of plasmas, such as Debye Shielding and Magnetic Drifts, that the author would love to be able to go into here, but for the sake of both brevity and sensibility, we shall concentrate on applying what we have learnt about dispersion to plasmas. In all plasmas, we have completely free electrons, and so there is no linear spring term binding them to nuclei.

Collisionless Plasma

Diffuse, high-temperature plasmas tend to be collisionless. We can have obtain an intuitive understanding of why this might be through Kinetic Theory; the faster the electrons are moving, the smaller their 'effective collision radius' (in the hard-sphere approximation), and so they are less likely to encounter another particle on it's trajectory. Of course, this is ignoring all electromagnetic interactions between the particles.

For such a plasma, we would expect the damping to be very small, and so we impose the condition that $\omega \gg \gamma$. This means that the equation of motion is simply

$$\underline{x} = -\frac{q}{m\omega^2}\underline{E}$$

At this stage, we actually have the decision to treat the plasma as either a dielectric, or a conductor with complex conductivity (this is fine; there is nothing that says that conductivity must be real).

- As a dielectric - In this case, we go through the same process as before, deriving an expression for $\underline{P} = n\underline{p}$, and subsequently finding ϵ_r. This turns out to be

$$\boxed{\epsilon_r(\omega) = 1 - \frac{\omega_p^2}{\omega^2}} \tag{1.55}$$

  where $\omega_p = ne^2/\epsilon_0 m$ is known as the *plasma frequency*. From the results of Section (1.6), we obtain a dispersion relation

$$\begin{aligned} k^2 &= \mu\epsilon\omega^2 \\ &= \mu_0\epsilon_0(\omega^2 - \omega_p^2) \end{aligned}$$

- As a conductor - In this case, we need to find the conductivity σ, as a conductor does not have a well-defined polarisation as in a dielectric.

$$\begin{aligned} \underline{J}_f &= nq\dot{\underline{x}} \\ &= i\frac{ne^2}{m\omega}\underline{E} \end{aligned}$$

  Assuming that the conductor obeys Ohms law

$$\begin{aligned} \underline{J}_f &= \sigma\underline{E} \\ \sigma &= i\frac{ne^2}{m\omega} \end{aligned}$$

  From the results of Section (1.7), we obtain the dispersion relation of

$$\begin{aligned} k^2 &= \mu\epsilon\omega^2 + i\omega\mu\omega \\ &= \mu\epsilon\omega^2 - \mu\frac{ne^2}{m} \\ &= \mu_0\epsilon_0(\omega^2 - \omega_p^2) \end{aligned}$$

We thus obtain the same dispersion relation (setting $\epsilon_r = 1$ and $\mu_r = 1$ in the conductor case); these two descriptions are thus equivalent in the case of a collisionless plasma. This means that we can write quite generally that

$$\boxed{k = \frac{1}{c}\sqrt{\omega^2 - \omega_p^2}} \tag{1.56}$$

Note how k becomes complex for $\omega < \omega_p$. This means that electromagnetic waves can only travel in a plasma for $\omega > \omega_p$, or else they are evanescent waves with a very short skin depth. But why is this the case physically? The argument is very similar to that for conductors. For frequencies below the plasma frequency, the electrons within the plasma are able to move quickly enough through it to compensate for the external fields, but this begins to break down as the frequency of the incident wave reaches and surpasses the plasma frequency.

A system is designed to measure the density of free electrons in the ionosphere. The density of free electrons is negligible below a height h_p and above this height it increases with height h as

$$n = a(h - h_p)$$

A transmitter on the ground sends short pulses of radiation of frequency ν vertically upwards. These are reflected at the point where the refractive index $n = 0$ and are then received on the ground at a time $\Delta\tau$ after emission. If for $\nu_1 = 5MHz$, $\Delta\tau_1 = 1.01ms$, and for $\nu_2 = 6MHz$, $\Delta\tau_2 = 1.16ms$, find values for a and h_p.

This is a hard, but also interesting question, which is why the author has decided to include it here. First of all, let us look at the group velocity of waves in a plasma.

$$\begin{aligned} v_p &= \frac{\omega}{k} = \frac{c}{n} \\ \omega^2 &= \omega_p^2 + c^2k^2 \\ 2\omega\frac{d\omega}{dk} &= 2c^2k \\ v_g &= \frac{c^2k}{\omega} \end{aligned}$$

This means that we obtain the useful plasma relation that

$$\boxed{v_p v_g = c^2} \tag{1.57}$$

Now to the question. We know that for $h < h_p$, $v_g = c$, and that for above this height, we have the relation that

$$c^2k^2 = \omega^2 - \frac{a(h - h_p)e^2}{\epsilon_0 m}$$

The travel time through the air is simply $t_a = 2h_p/c$. Through the plasma, we have that

$$t_p = 2\int \frac{dh}{v_g(k)} = 2\int \frac{1}{v_g}\frac{dh}{dk}dk$$

From above, we have that

$$\frac{dh}{dk} = -\frac{2c^2k\epsilon_0 m}{ae^2}$$

Then

$$t_p = -\frac{4\omega\epsilon_0 m}{ae^2}\int_{k_i}^{k_f} dk = \frac{4\omega^2\epsilon_0 m}{ae^2 c}$$

By definition, k_f is the wave-number at the point of reflection, which is zero (cannot propagate for n too large), and so $k_f = 0$. Similarly, $k_i = \omega/c$. From this, we are able to complete our calculation.

$$\begin{aligned}
\Delta\tau_2 - \Delta\tau_1 &= \frac{4\epsilon_0 m}{ae^2 c}4\pi^2(\nu_2^2 - \nu_1^2) \\
a &= \frac{16\epsilon_0 m\pi^2(\nu_2^2 - \nu_1^2)}{(\Delta\tau_2 - \Delta\tau_1)e^2 c} \\
&\approx 1.2 \times 10^7 \ \text{m}^{-3} \\
h_p &= \frac{c}{2}\left(\Delta\tau_1 - \frac{16\epsilon_0 m\pi^2\nu_1^2}{ae^2 c}\right) \\
&\approx 1 \times 10^5 \ \text{m}
\end{aligned}$$

Collision-dominated Plasma

Dense, low temperature plasmas tend to be dominated by collisions. In this case, we cannot neglect the damping term. Once again, we can treat the system in two ways:

- As a dielectric - In this case, we obtain a complex dielectric constant of

$$\epsilon_r = 1 + \frac{ine^2}{\epsilon_0 m\gamma\omega}$$

- As a conductor - Through an identical calculation as before, we obtain

$$\sigma = \frac{ne^2}{m\gamma}$$

Once again, we obtain the same dispersion relation via substitution into the appropriate expressions in both cases.

What should one take away from this? Generally, the methods for deriving the frequency dependant permittivity should be practised, as they allow one to describe free electrons in other scenarios quite effectively as well. Other than that, it is an interesting introduction into the world of Plasma Physics.

2. *Electromagnetism II*

This second chapter on Electromagnetism covers the following:

- Radiation
- Electromagnetism and Special Relativity
- Transmission Lines

The division of the material into two chapters is completely arbitrary. This second one covers quite separate and un-related topics that are stipulated in the syllabus, and so they can be omitted on first reading through the material. In fact, this author does not understand the decision to include them as part of the course in the first place, and hopes to change the syllabus to reflect this. Nevertheless, it is important that students make themselves familiar with this material.

2.1 Radiation

Radiation is the energy flux from a system that propagates through a sphere at $r = \infty$ (with an appropriate choice of coordinate system). If both the magnetic and electric fields fall of as the inverse square of the radius, the flux through such a sphere is zero. This means that radiation must occur in systems that have a different power-law dependence; this turns out to occur as a result of the oscillation/acceleration of dipoles. Note that this does not mean that all such systems produce radiation; certain symmetries (such as spherical) prevent radiation from actually being emitted, as the radiation fields may interfere destructively.

2.1.1 Larmor's Formula

Arguably, the most important result associated with electromagnetic radiation is *Larmor's Formula*, which states that

$$\boxed{P_{\text{rad}} = \frac{q^2 a^2}{6\pi\epsilon_0 c^3}} \tag{2.1}$$

That is, it is the statement that *any accelerated charged particle radiates energy proportional to the square of it's acceleration.* Note that this is the instantaneous acceleration, though usually one talks about the time averaged power. From this equation, it is clear that P_{rad} does not depend on the radius r, nor the time t. This is essentially a statement of energy conservation; the power crossing a given sphere at a given time must be the same as that crossing another sphere of a different radius at another time. For some dipole moment p and some solid angle $d\Omega$, the formula can be re-written as

$$\frac{dP_{\text{rad}}}{d\Omega} = \frac{1}{16\pi^2\epsilon_0 c^3}\left(\frac{d^2p}{dt^2}\right)^2 \sin^2\theta$$

An electron is released at rest and falls under the influence of gravity. In the first metre, what fraction of the potential energy lost is radiated away. Give an order of magnitude estimate. It can be assumed that the radiation lost is so small that it does not effect the motion of the electron.

Assume that the only force effecting the acceleration is gravity, and that the force remains roughly constant. This means that simply $a = g$. Let the height fallen be h. This means that potential energy lost is mgh, where m is the electron mass. The time taken to fall a height h is simply given by $t_h = \sqrt{2h/g}$. Then, the fraction lost as radiation is given by

$$\text{Fraction Lost} = \frac{P_{\text{rad}}\, t_h}{mgh} = \frac{q^2\sqrt{2gh}}{6\pi\epsilon_0 m h c^3} \backsim 10^{-21}$$

Thus, the amount of energy lost due to radiation is negligible in comparison to the potential energy lost, confirming our *a priori* assumption.

Proving Larmor's Formula

Now that we know the result, the question is, how did we arrive here? Let us go about proving Larmor's Formula, and in doing so consider some interesting effects of acceleration on charges.

First, let us consider the question: *what does the field of an accelerated charge look like?* The electric field of a charge moving at a constant velocity (or that is stationary) is purely radial, as shown in the left of Figure (2.1). However, when a charge is being accelerated, it takes time for the information about the instantaneous position of the particle to travel out to some radius r. This finite travel time actually means that components of the field in the $\hat{\underline{\theta}}$ direction are created, as shown on the right of Figure (2.1).

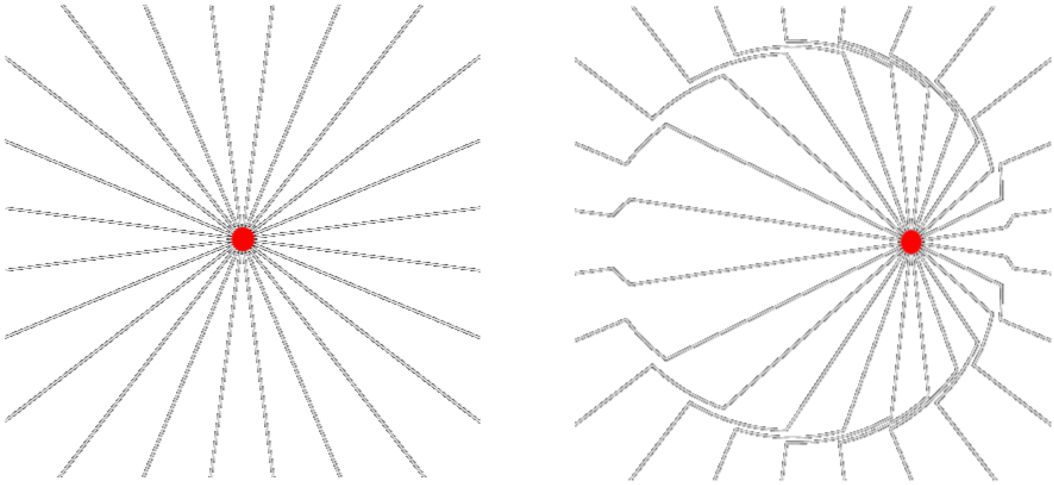

Figure 2.1: The electric field lines of a stationary charge (left) and an accelerated charge (right)

This component of the electric field, as well as the associated component of the magnetic field in the $\hat{\underline{\phi}}$ direction, are known as *radiation fields*; this is because their product is proportional to $1/r^2$, meaning that a finite power is radiated away to infinity. It is possible to shown this all mathematically, but the knowledge that the fields scale with $1/r$ is the most important information for deriving Larmor's Formula (up to numerical factors). This information has just been included in here for the sake of understanding.

Consider an accelerating charge that has an effective dipole moment $\underline{p}$. We now compute the Poynting vector for the system as

$$\underline{S}_{\text{rad}} = \frac{1}{\mu_0}\underline{E}\times\underline{B} = \frac{c}{\mu_0}\left|\underline{B}\right|^2\hat{\underline{r}} = \frac{1}{16\pi^2\epsilon_0 c^3}\left|\hat{\underline{r}}\times\ddot{\underline{p}}\right|^2\frac{\hat{\underline{r}}}{r^2}$$

The fact that the power is radiated radially follows from the directions of the $\underline{E}$ and $\underline{B}$ fields outlined above. Suppose now that the dipole oscillates in the $\hat{\underline{z}}$ direction, and define θ as normal in spherical coordinates. Then

$$\underline{S}_{\text{rad}} = \frac{1}{16\pi^2\epsilon_0 c^3}\frac{|\ddot{\underline{p}}|^2}{r^2}\sin^2\theta\,\hat{\underline{z}}$$

The total power radiated is computed by integrating over the solid angle $d\Omega$:

$$P_{\text{rad}} = \int \underline{S}_{\text{rad}}\cdot d\underline{\Sigma} = \frac{1}{16\pi^2\epsilon_0 c^3}\int_0^{2\pi}d\phi\int_0^{\pi}d\theta\sin^3\theta = \frac{|\ddot{\underline{p}}|^2}{6\pi\epsilon_0 c^3}$$

If we now suppose that the dipole is of the form $\underline{p} = \underline{p}_0 + \underline{p}(t)$, it follows quickly that $\ddot{\underline{p}} = qa$, where a is the acceleration of the electron. Substituting this in, we arrive at Larmor's Formula.

2.1.2 Scattering Types

When electromagnetic radiation is incident on atoms, it creates an electric dipole with the electron, meaning that it will radiate some of the incident energy away. This is known as electromagnetic *scattering*. We define the *scattering cross section* as

$$\boxed{\sigma(\omega) = \frac{\langle P_{\text{rad}} \rangle}{\langle S \rangle}} \tag{2.2}$$

This is essentially a measure of the amount of the incident power that is radiated. As such, we have the following scattering types:

- Thompson Scattering - This is the elastic scattering of EM radiation from a free electron. This means that we let $\omega_0 = 0$ in the equation of motion of the electron, meaning that the scattering cross section is a constant; or, put in another way, it does not depend on frequency
- Compton Scattering - This is the inelastic scattering of EM radiation, where the frequency of the emitted wave is not the same as the incident wave. This occurs when the energy of the photon is comparable to the rest mass energy of the electron $\hbar\omega \backsim m_e c^2$
- Rayleigh/Resonant Scattering - This is the elastic scattering of an electromagnetic wave by an electron bound to a nucleus in an atom. We can model the electron as being classically bound to it's atom by a spring force $m\omega_0^2$. The case where $\omega \backsim \omega_0$, we are in the regime of resonant scattering, and $\sigma \gg 1$. Rayleigh scattering occurs under the condition that $\omega \ll \omega_0$

Note that we usually only consider the effect of the electric field of the incident wave on the electron, and not that of the magnetic field. This is because the magnitude of the magnetic field force is much less than that of the electric field.

$$\left|\frac{F_B}{F_E}\right| = \frac{qvB}{qE} \backsim \frac{Z_0 \omega B_0}{E_0} \backsim \frac{\omega Z_0}{c}$$

This quantity is evidently much less than one in most materials, and for most frequencies.

2.1.3 An Illustrative Example

Consider coherent light incident from the sun on an atom in the atmosphere. Assuming we are in the Rayleigh Scattering regime, we can write the equation of motion of the electron as

$$\begin{aligned} m\underline{\ddot{x}} &= q\underline{E} - m\omega_0^2 \underline{x} \\ \underline{x} &= \frac{q\underline{E}_0}{m(\omega_0^2 - \omega^2)} \end{aligned}$$

where we have assumed that the electric field is of the form $\underline{E} = \underline{E}_0 \cos\omega t$. Then, the acceleration is given by

$$\ddot{x} = -\frac{qE_0\omega^2 \cos\omega t}{m(\omega_0^2 - \omega^2)}$$

Time averaging:

$$\begin{aligned} \langle \ddot{x}^2 \rangle &= \frac{E_0^2 \omega^4}{\left[m(\omega_0^2 - \omega^2)\right]^2} \cdot \frac{\omega_0}{2\pi} \int_0^{\frac{2\pi}{\omega_0}} dt \ \cos^2 \omega_0 t \\ &= \frac{1}{2} \frac{E_0^2 \omega^4}{\left[m(\omega_0^2 - \omega^2)\right]^2} \end{aligned}$$

This means that the radiated power is given by

$$P_{\text{rad}} = \underbrace{\left(\frac{1}{2}\epsilon_0 c E_0^2\right)}_{\text{average of }\underline{\text{S}}} \underbrace{\left(\frac{8\pi r_0^2}{3}\right) \frac{\omega^4}{(\omega_0^2 - \omega^2)^2}}_{\text{scattering cross section}}$$

where $r_0 = (\alpha^2\hbar^2)/(c^2)$. Writing it in this form is useful, as it allows us to see that the radiated power is proportional to the incident power, and the scattering cross section. In the Rayleigh scattering limit, $\omega \ll \omega_0$, and so we find that $P_{\text{rad}} \propto \omega^4$.

This limit is interesting to us as it allows us to explain a commonly mis-understood question; *why is the sky blue?* The light from the sun that is incident on atoms in the upper atmosphere causes said atoms to radiate energy, as we have demonstrated. However, as $P_{\text{rad}} \propto \omega^4$, the higher frequency end of the spectrum is favoured, and as such blue light is radiated with $\backsim 4.5$ times the power of red light, making the sky appear blue. Note that this is not the case if you are looking at areas of the sky near the sun, as the electrons are only able to radiate energy perpendicular to their oscillation, and this at angles to the incoming rays from the sun.

Conversely, for a sunset, the light from the sun has to travel through a much greater amount of atmosphere to reach the observer, as the path of the light rays are almost tangential to the surface of the Earth. This means that the majority of the blue light is already radiated/scattered before it reaches the observer, and so the sky appears red.

2.2 Electromagnetism and Special Relativity

This section shall serve as a brief introduction to relativistic electrodynamics, though this shall be covered in much more depth in the Third Year course. This author shall assume that the reader is familiar with Einstein's Postulates of Special Relativity, as well as the Lorentz Transformations, Time Dilation, Length Contraction and Relativistic Velocity addition. If the reader is unfamiliar with any of these concepts, it is recommended that they refresh their memory using the CP1 Special Relativity notes before reading on.

2.2.1 Field Transformations

We will now examine how both $\underline{E}$ and $\underline{B}$ fields transform under Special Relativity. These can be derived by considering the change in the surface charge density and surface current density (using the invariance of charge, see the following section) on parallel plates that have orientation parallel and/or perpendicular to the direction of relative motion between the two frames. The derivations will not be covered here, as this author believes knowledge of them is not required by the syllabus. The transformations between a frame F and frame F' moving at a velocity $\underline{v}$ with respect to F are thus

$$\underline{E}'_{\parallel} = \underline{E}_{\parallel} \tag{2.3}$$

$$\underline{B}'_{\parallel} = \underline{B}_{\parallel} \tag{2.4}$$

$$\underline{E}'_{\perp} = \gamma\left(\underline{E}_{\perp} + \underline{v} \times \underline{B}_{\perp}\right) \tag{2.5}$$

$$\underline{B}'_{\perp} = \gamma\left(\underline{B}_{\perp} - \frac{1}{c^2}\,\underline{v} \times \underline{E}_{\perp}\right) \tag{2.6}$$

where γ is defined as usual in Special Relativity. We can observe two interesting cases of these equations:

- If $\underline{B} = 0$ in frame F, then there is actually a magnetic field in F' given by
$$\underline{B}' = -\frac{1}{c^2}\,\underline{v} \times \underline{E}'$$
- If $\underline{E} = 0$ in frame F, then there is actually an electric field in F' given by
$$\underline{E}' = \underline{v} \times \underline{B}'$$

Both of these cases highlight the importance of the effects of Special Relativity on electrodynamics, as 'new' fields can appear if we previously had not been taking into account its effects. Furthermore, due to the way these transformations act, one could almost think of magnetism as simply a relativistic correction to electrostatics.

Astute readers may be asking the question as to whether there are any frame invariant products in relativistic electrodynamics, in the same way as (for example) $E^2 - p^2c^2$ was invariant for massive objects. In this case, it turns out that $\underline{E}_{\perp} \cdot \underline{B}_{\perp}$ is invariant, as follows.

$$\begin{aligned}
\underline{E}'_{\perp} \cdot \underline{B}'_{\perp} &= \gamma^2\left(\underline{E}_{\perp} \cdot \underline{B}_{\perp} - \frac{1}{c^2}\,\underline{E}_{\perp} \cdot (\underline{v} \times \underline{E}_{\perp}) + (\underline{v} \times \underline{B}_{\perp}) \cdot \underline{B}_{\perp} - \frac{1}{c^2}\,(\underline{v} \times \underline{B}_{\perp}) \cdot (\underline{v} \times \underline{E}_{\perp})\right) \\
&= \gamma^2\left(\underline{E}_{\perp} \cdot \underline{B}_{\perp} - \frac{1}{c^2}\,((\underline{v} \cdot \underline{v})\,((\underline{E}_{\perp} \cdot \underline{B}_{\perp}) - (\underline{B}_{\perp} \cdot \underline{v})\,(\underline{E}_{\perp} \cdot \underline{v}))\right) \\
&= \gamma^2\left(\underline{E}_{\perp} \cdot \underline{B}_{\perp} - \frac{v^2}{c^2}\underline{E}_{\perp} \cdot \underline{B}_{\perp}\right) \\
&= \underline{E}_{\perp} \cdot \underline{B}_{\perp}
\end{aligned}$$

Similarly, it can also be shown that $E^2 - c^2B^2$ is also frame invariant.

2.2.2 Charge Invariance

As it is quite difficult to measure the value of a charge in motion, we define the charge as

$$\boxed{q = \epsilon_0 \int_{\partial V} \underline{E} \cdot d\underline{\Sigma}} \tag{2.7}$$

Now consider a charge q that is at rest at the origin of frame F. Consider the electric field at some point P at (r, θ) in the z-x plane in frame F. Our aim is then to find the field at some point P' at (r', θ') in a frame F' moving at a velocity $\underline{v} = (v, 0, 0)$ relative to $\underline{F}$. Using the Lorentz Transforms, it is clear that

$$\begin{aligned} x &= \gamma\left(x' + vt\right) = \gamma r'_x \\ y &= y' = r'_y = 0 \\ z &= z' = r'_z \end{aligned}$$

From (2.7), we can write the electric field in frame F' as

$$\underline{E}' = \frac{\gamma q}{4\pi\epsilon_0} \frac{\underline{r}'}{r'^3}$$

Recalling that $\underline{r}' = \left(x'^2 + y'^2 + z'^2\right)^{1/2}$, we can use the relations above to show that

$$\underline{E}' = \frac{q}{4\pi\epsilon_0} \frac{1 - \beta^2}{\left(1 - \beta^2 \sin^2\theta'\right)^{3/2}} \frac{\widehat{\underline{r}'}}{r'^2}$$

where $\beta = v/c$ as normal. It is easy to show that this reduces to the expression for $\beta \ll 1$ that we would obtain by applying the Galilean transformations to the usual expression for $\underline{E}$. Let us compute the flux of $\underline{E}'$ through a sphere of radius $\underline{r}'$ in frame F':

$$\begin{aligned} \int \underline{E}' \cdot d\underline{\Sigma}' &= \frac{q}{4\pi\epsilon_0} \frac{1}{\gamma^2} \int d\theta' d\phi' \, \frac{\sin\theta'}{\left(1 - \beta^2 \sin^2\theta'\right)^{3/2}} \\ &= \frac{q}{2\epsilon_0\gamma^2} \int_0^\pi d\theta' \, \frac{\sin\theta'}{\left(1 - \beta^2 \sin^2\theta'\right)^{3/2}} \\ &= \frac{q}{2\epsilon_0\gamma^2} \int_{-1}^{1} du \, \frac{1}{(1 - \beta^2 u^2)^{3/2}} \end{aligned}$$

where the substitution $u = \cos\theta'$ has been made. Evaluating this integral, it is clear that the right-hand-side is equal to q/ϵ_0. This means that *charge is invariant* across frames, regardless of whether they are relativistic or not. This is quite a powerful result, as it means that Gauss' Law can be applied independent of the consideration of frames.

2.3 Transmission Lines

Thus far in our consideration of AC circuits, we have considered voltage and current to be constant along the wires connecting the active components; any capacitance, inductance, resistance, voltage gain etc. is all 'lumped' together in the components. This assumption, while not immediately obvious, does seem un-physical, as there will evidently be changes in the voltage and the current along the wires as they are not perfect conductors.

A transmission line consists 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. Suppose that a semi-infinite transmission line has capacitance per unit length C' and inductance per unit length L'. Now consider the voltage and current at z and $z + dz$ as below. Note that the arrows show the direction of current flow.

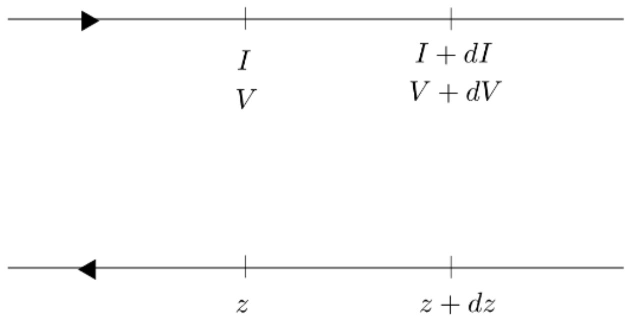

Figure 2.2: A schematic transmission line

Applying the normal formulae for inductors and capacitors, the changes in voltage and current over the interval $[z, z + dz]$ are given by

$$\underbrace{dV = -L' dz \frac{\partial I}{\partial t}}_{\text{due to inductance}} \quad \text{and} \quad \underbrace{dI = -C' dz \frac{\partial V}{\partial t}}_{\text{due to capacitance}}$$

Note that we are assuming that there are no parasitic losses due to resistance within the line. We can re-arrange the above two expressions to obtain

$$\boxed{\begin{aligned} \frac{\partial V}{\partial z} &= -L' \frac{\partial I}{\partial t} \qquad (2.8) \\ \frac{\partial I}{\partial z} &= -C' \frac{\partial V}{\partial t} \qquad (2.9) \end{aligned}}$$

These equations are known as the *telegraph equations* that define the behaviour of the voltage and current. Taking the spatial derivative of the first equation, and the time derivative of the second:

$$\frac{\partial^2 V}{\partial z^2} = -L' \frac{\partial^2 I}{\partial t \partial z} \quad \text{and} \quad \frac{\partial^2 I}{\partial z \partial t} = -C' \frac{\partial^2 V}{\partial t^2}$$

Using the fact that derivatives commute, we can put these two equations together to obtain

$$\frac{\partial^2 V}{\partial z^2} = L' C' \frac{\partial^2 V}{\partial t^2}$$

We can find a similar equation for the current through applying the same derivatives but in the reverse order. This means that both current and voltage behave as waves in the circuit, propagating at

$$\boxed{v_p = \frac{1}{\sqrt{L'C'}}} \tag{2.10}$$

Now, suppose that the voltage takes the functional form $V(z,t) = f(z+vt) + g(z-vt)$. Substitute this into (2.9), we find that $I(z,t)$ must be of the form

$$I(z,t) = \sqrt{\frac{C'}{L'}} f(z+vt) - \sqrt{\frac{C'}{L'}} g(z-vt)$$

We thus define

$$\boxed{Z_0 = \frac{V}{I} = \sqrt{\frac{L'}{C'}}} \tag{2.11}$$

as the *characteristic impedance* of the transmission line. Note that the impedance is defined at any particular point in the line as the ratio of V and I.

A transmission line is formed by a wire of radius a placed a distance d in a vacuum above an infinite conducting plane, where $d \gg a$. Find the characteristic impedance of the system. Assume that the wire has charge per unit length λ, and carries a current I

We can treat this system using the method image charges. Replace the conducting plane by another wire a distance $2d$ below the original wire, with charge per unit length $-\lambda$ and current $-I$. Then, the electric field at some point satisfying $z > 0$ along the line joining the two wires is given by

$$E(z) = \frac{\lambda}{2\pi\epsilon_0(d-z)} - \frac{\lambda}{2\pi\epsilon_0(d+z)}$$

Integrating to find the potential:

$$\begin{aligned}
\Delta V &= -\int_0^{d-a} dz\, E(z) \\
&= -\frac{\lambda}{2\pi\epsilon_0} \int_0^{d-a} dz \left(\frac{1}{d-z} - \frac{1}{d+z}\right) \\
&= \frac{\lambda}{2\pi\epsilon_0} \log\left(\frac{2d-a}{a}\right) \\
&\backsim \frac{\lambda}{2\pi\epsilon_0} \log\left(\frac{2d}{a}\right)
\end{aligned}$$

We have made use of the fact that $d \gg a$ in the last step. This means that the capacitance per unit length is given by

$$C' = \frac{Q}{\ell\Delta V} = \frac{\lambda}{\Delta V} = \frac{2\pi\epsilon_0}{\log\left(\frac{2d}{a}\right)}$$

Now for the inductance per unit length. By Ampere's law, it is easy to show that

$$B(z) = \frac{\mu_0 I}{2\pi}\left(\frac{1}{d-z} - \frac{1}{d+z}\right)$$

Calculating the magnetic flux:

$$\begin{aligned}\phi &= \int \underline{B} \cdot d\underline{\Sigma} \\ &= \ell \int_0^{d-a} dz \, \frac{\mu_0 I}{2\pi} \left(\frac{1}{d-z} - \frac{1}{d+z} \right) \\ &\backsim \frac{\mu_0 I \ell}{2\pi} \log\left(\frac{2d}{a}\right)\end{aligned}$$

where we have again made use of $d \gg a$. This means that the inductance per unit length is given by

$$L' = \frac{1}{\ell}\frac{\partial \phi}{\partial I} = \frac{\mu_0}{2\pi} \log\left(\frac{2d}{a}\right)$$

This means that the characteristic impedance of the line is given by

$$Z_0 = \frac{1}{2\pi}\sqrt{\frac{\mu_0}{\epsilon_0}} \log\left(\frac{2d}{a}\right)$$

Typically, $Z_0 \backsim 10^2\ \Omega$. This means that, to a rough estimate, we require that $d \backsim 3a$, meaning our assumption that $d >> a$ is only valid for systems of much higher impedance.

2.3.1 Reflection and Transmission

Consider a line of length ℓ, with characteristic impedance Z, that is terminated by some impedance Z_T.

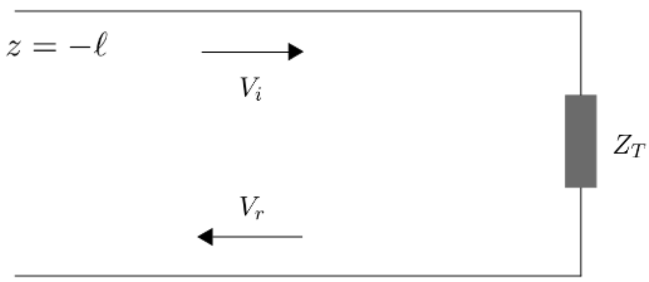

Figure 2.3: A transmission line terminated by Z_T

The solutions to V and I will be the sum of the reflected and transmitted waves:

$$\begin{aligned}V(z,t) &= V_i e^{i(\omega t - kz)} + V_r e^{i(\omega t + kz)} \\ I(z,t) &= \frac{V_i}{Z_0} e^{i(\omega t - kz)} - \frac{V_r}{Z_0} e^{i(\omega t + kz)}\end{aligned}$$

Note the negative sign in-front of the second term in the expression for current; this comes from the functional form used to derive (2.11). The boundary condition in this case is that the ratio of V to I is equal to the terminating impedance at $z = 0$, by definition. From this, it is easy to show that

$$\boxed{r = \frac{V_r}{V_i} = \frac{Z_T - Z_0}{Z_T + Z_0}} \tag{2.12}$$

Let us consider some limiting cases:

– $Z_T = 0$ - All the energy is reflected with a phase shift of π.

$$V_{reflected} \to -V_{incident}$$

– $Z_T = Z_0$ - No energy is reflected. This is called matched impedance, where all the power is transmitted to the terminating load.

$$V_{reflected} \to 0$$

– $Z_T \to \infty$ - All the energy is reflected but with no phase shift.

$$V_{reflected} \to V_{incident}$$

We now define a quantity

$$\boxed{Z_i = Z_0 \frac{e^{ik\ell} + re^{-ik\ell}}{e^{ik\ell} - re^{-ik\ell}}} \tag{2.13}$$

as being the *input impedance* of the system at $z = -\ell$. This is quite simply the impedance seen by the system at said point. Note that we can re-obtain the condition for impedance matching by letting $Z_i = Z_0$ in this expression.

A leak develops at one point in an infinite transmission line. The resistance of the leak is equal to the characteristic impedance of the line. Show that $1/9$ *of the power in the incident wave is reflected and* $4/9$ *is dissipated in the leak.*

We can model this problem by considering the leak as being some impedance Z_ℓ, as shown in the following Figure.

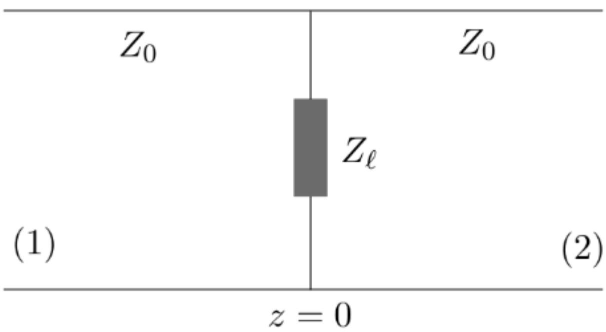

Figure 2.4: A leaky transmission line

We need to consider the waves in each of the regions (1) and (2).

$$\left.\begin{aligned} V_1 &= V_i e^{i(\omega t - kz)} + V_r e^{i(\omega t + kz)} \\ I_1 &= \tfrac{V_i}{Z_0} e^{i(\omega t + kz)} - \tfrac{V_r}{Z_0} e^{i(\omega t - kz)} \end{aligned}\right\} \text{Region (1)}$$

$$\left.\begin{aligned} V_2 &= V_t e^{i(\omega t - kz)} \\ I_2 &= \tfrac{V_t}{Z_0} e^{i(\omega t - kz)} \end{aligned}\right\} \text{Region (2)}$$

Now, the boundary conditions of this problem are that the voltage is continuous, while the current is discontinuous over the leak, namely:

$$V_1 = V_2 \quad \text{and} \quad I_1 = I_2 + \frac{V}{Z_\ell}$$

Defining $L = Z_0/Z_\ell$, it is quite easy to show algebraically that

$$\frac{V_r}{V_i} = -\frac{1}{1+2/L} = -\frac{1}{3}$$
$$\frac{V_t}{V_i} = \frac{2}{2+L} = \frac{2}{3}$$

However, we are told that $Z_\ell = Z_0$, and so these expressions reduce to the expressions in the lighter type above. Recalling that $P_f = \left|\frac{V_f}{V_i}\right|^2$, we have shown the result.

2.3.2 Stubs

Stubs are simply transmission lines that are connected to the system by a single end, and terminated at the other end by either a short-circuit, or an open circuit. These are typically used to replace capacitors and inductors in high frequency circuits, as under these conditions these components perform poorly due to parasitic reactance. Note that stubs can only behave in this way when their length ℓ satisfies certain conditions, which can be found by equating the stub impedance (see below) with the required impedance. However, at the higher frequencies, the associated wavelength is sufficiently small such that the stub does not have to be very long.

- Open Circuit Stub - Let us assume the end of the stub that is not connected to the system is terminated by an open circuit. In this case, we can either use the condition that $I = 0$ at the end ($z = -\ell$), or look at limiting cases of (2.13).

$$Z_T \to \infty$$
$$r \to 1$$
$$Z_{\text{open}} \to -iZ_0 \cot k\ell$$

- Short Circuit Stub - Let us assume the end of the stub that is not connected to the system is terminated by an short circuit. In this case, we can either use the condition that $V = 0$ at the end, or again look at limiting cases of (2.13).

$$Z_T \to 0$$
$$r \to -1$$
$$Z_{\text{short}} \to iZ_0 \tan k\ell$$

It immediately follows that

$$\boxed{Z_{\text{open}} Z_{\text{short}} = Z_0^2} \tag{2.14}$$

This actually has an interesting practical application; one can find the characteristic impedance of a transmission line by measuring Z_{open} and Z_{short} by attaching a stub to it. This is used frequently in electrical maintenance.

Note that with stubs, it is generally useful to match *admittances*, defined as $Y = 1/Z$. For example, for no reflection back to the source, we need the $Y_{\text{stub}} + Y_{\text{impedance}} = Y_0$ when the stub is placed in parallel.

3. *Optics*

This chapter aims to introduce some further concepts in the field of Optics, including:

- Fraunhofer Diffraction
- Spectroscopic Instruments
- Polarisation

It is assumed that readers are familiar with the basic assumptions and conventions associated with both geometric and wave optics, as these were covered in the Optics Section of the CP2 paper in First Year. Generally, it is not hard to gain an understanding of the fundamentals of Optics, but students often find it difficult to apply this learned knowledge to problems. The answer on this front, as it is in most cases, is practise.

3.1 Fraunhofer Diffraction

As we have outlined in Section (1.5), light behaves as a superposition of orthogonal electric and magnetic fields, and as such, must be a solution to the wave equation. This has either plane wave solutions, or spherical wave solutions ($\propto 1/r$). As we have seen previously, we take take the amplitudes of the waves from point sources, and add them together to give the total resultant amplitude $U(\theta)$. Then, using the fact that the intensity is given by

$$\boxed{I(\theta) \propto |U(\theta)|^2} \tag{3.1}$$

any non-negligible phase difference between the two point sources will lead to an interference pattern when properly imaged. Now let us recall Huygens' Principle:

> *Every point on a wave-front may be considered a source of secondary spherical wavelets which spread out in the forward direction at the speed of light. The new wave-front is the tangential surface to all of these secondary wavelets.*

This means that our amplitude treatment of the interference of point-like sources can be extended to finite apertures; we can sum up the contributions from each 'point' in the aperture (that each act as a source for Huygens' spherical waves) through integration to find the total amplitude. This idea forms the basis of most wave optics calculations.

Fraunhofer diffraction is a special case of this, and occurs where *the phase difference between optical paths from the source to the image plane via some point in the aperture is a linear function of position in said aperture.* The resultant intensity pattern is observed at infinity, or more conveniently in the focal plane of a lens that collects the diffracted rays. A question worth asking at this stage is as to how far do we need to be away from the diffracting aperture to be able to work within the Faunhofer regime. Consider the figure below.

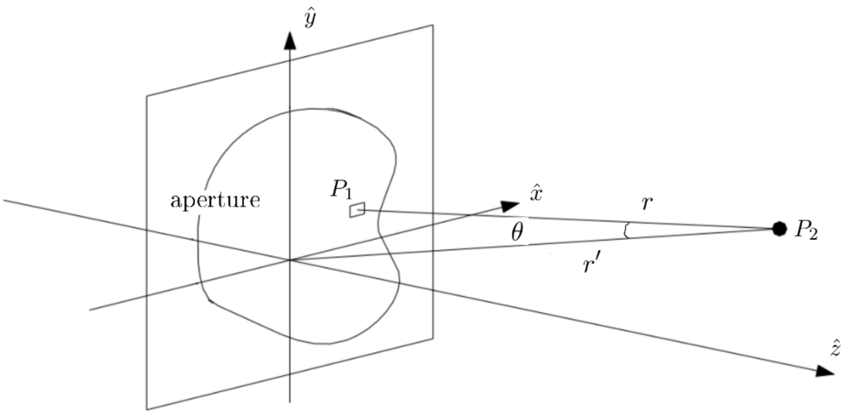

Figure 3.1: Diffraction by some general aperture

Let the coordinates $(x, y, 0)$ represent a point P_1 within the aperture, and the coordinates (x', y', z') represent some point P_2 at a distance r' from the centre of the aperture. We can let $y = y'$ without loss of generality. Then

$$r = \sqrt{(x'-x)^2 + z'^2} \backsim r' - x\sin\theta + \frac{x^2}{2r'}$$

where we have expanded under the assumption that $x \ll r'$, and observed that under the small angle approximation, $\sin\theta \backsim x/r'$. Under the Fraunhofer condition, we require that the phase difference introduced by the last term is negligible:

$$\frac{kx^2}{2r'} \ll 2\pi$$

This means that for some general aperture of 'size' a, wavelength λ and a distance R from the screen, the Fraunhofer condition can be written as

$$\boxed{\frac{a^2}{2\lambda} \ll R} \tag{3.2}$$

Note that this is not how we *define* Fraunhofer Diffraction; it is a consequence of the Fraunhofer condition, not the other way around.

3.1.1 Fourier Optics

With these principles in mind, we will now turn to explicit consideration of diffraction calculations. It can be shown that the Fresnel-Kirchoff diffraction integral is of the form

$$U(\theta) = \frac{i}{\lambda} \underbrace{\int dS}_{\text{integral over aperture}} \overbrace{U_0}^{\text{incoming amplitude}} \underbrace{\frac{e^{ikr}}{r}}_{\text{Huygens' Wave}} \overbrace{\eta(\underline{n}, \underline{r})}^{\text{obliquity factor}}$$

The normalisation constant in-front of the integral aside, it is pretty clear that this expression is simply summing up individual Huygens waves across the entire surface of the aperture. The *obliquity factor* accounts for the fact that the direction of propagation of the wave is not included in Huygen's Principle; the spherical waves spread out in all directions from a given point. Its explicit form is not an important consideration here; just know that it has value zero for waves travelling in the negative direction.

Under the Fraunhofer regime, we can make a series of assumptions to allow us to simplify this integral, as follows:

1. Ignore the obliquity factor - For small angles around the optical axis, $\eta(\underline{n}, \underline{r}) \backsim 1$.

2. Far Field Approximation - If we assume that we are viewing the diffraction under the condition given by Equation (3.2), then the variation in the distance to different points in the image plane will be negligible. This means that we can ignore the factor of $1/r$.

3. Fraunhofer Condition - If we impose the condition that the phase difference has to be a linear function of position, we can separate out this from the exponential factor

   $$e^{ikr} = e^{ikr_0} e^{ikx\sin\theta} \propto e^{ikx\sin\theta}$$

   We can absorb the remaining factor into our normalisation assuming that r_0 is roughly constant, which is the case in the Far Field Approximation.

4. Region of Integration - As we are no longer integrating over all r, we can then restrict the integral to one dimension: $dS \longrightarrow dx$.

Having made all of these simplifications, we arrive at the important expression of

$$\boxed{U(\theta) = U_0 \int dx\, T(x)\, e^{ikx\sin\theta}} \tag{3.3}$$

The function $T(x)$ is the aperture's *transmission function*. It is then clear that the resultant amplitude pattern is simply the Fourier transform of the transmission function; as soon as we know the transmission function, it is almost trivial to find the resultant diffraction pattern. Some common transmission functions include:

- Single Slit of width a: $T(x) = \begin{cases} 1 & \text{for } |x| < a/2 \\ 0 & \text{otherwise} \end{cases}$
- Double slits separated by d: $T(x) = \delta(x - d/2) + \delta(x + d/2)$
- Diffraction grating of spacing d: $T(x) = \sum_{m=0}^{N-1} \delta(x - md)$

Once we start dealing with diffraction as a Fourier transform, results from the mathematics of Fourier transforms become useful. For example, recall the convolution theorem:

> *The Fourier transform of a convolution*
>
> $$h(x) = f(x) * g(x) = \int_{-\infty}^{\infty} dx'\, g(x - x')f(x')$$
>
> *is given by the product of the Fourier transforms of the individual functions that make up the convolution.*

This means that we can find the diffraction pattern resulting from a convolution of two known transmission functions by the product of the amplitude patterns resulting from each transmission function. For example, a triangular function of height $1/a$ and width $2a$ is the convolution of two 'top-hat' single split functions of height $1/a$ and width a. If we then want two triangular functions as our finite amplitude profiles for double slits, we can then simply multiply this by the double split pattern. In this way, we can build up quite complicated diffraction transmission functions through the convolution of known transmission functions.

The factor $\beta = k \sin\theta$ is known as the *spatial frequency*; the number of cycles per visual degree of angle. High spatial frequencies correspond to fine frequencies and sharp edges for images, while low spatial frequencies correspond to the broader features; so an image where the high spatial frequencies are filtered out will be blurry and low-resolution. We can intentionally use this by blocking off part of the Fourier plane if we want to limit our imaging to high or low spatial frequencies. The transmission function can be represented as a sum of sinusoidal components, which become a series of delta functions upon applying a Fourier transform. For example, consider $T(x) = 1 + \cos(\omega_s x)$ for $\omega_s = 2\pi/d_s$ i.e. a periodic structure of size d_s.

$$U(\theta) \propto \int_{-\infty}^{\infty} dx\, e^{ikx\sin\theta} = \delta(0) + \delta\left(\sin\theta - \frac{\lambda}{d}\right) + \delta\left(\sin\theta + \frac{\lambda}{d}\right)$$

This means that there is only zeroth and first order diffraction from this single spatial frequency. Evidently, Blocking some of these will change the nature of the image. This can be achieved because each spatial frequency (corresponding to a different value of d_s) is diffracted at a different angle and so ends up at a different point on the Fourier plane; regions of high spatial frequency occur at large angles, while regions of low spatial frequency occur at low angles.

An Illustrative Example

Find the intensity pattern for a grating of N slits of width a, each separated by a distance d. Comment on the intensity pattern in the cases where $d = \ell a$ and $d \longrightarrow a$. Find the *resolving power of the grating*

Using the apparatus that we have set-up above, we can write that

$$I(\theta) = |U_1(\theta)\, U_2(\theta)|^2$$

where $U_1(\theta)$ corresponds to the amplitude resulting from the single slit, and $U_2(\theta)$ corresponds to the amplitude resulting from the delta-function diffraction grating.

Calculating the resultant amplitudes:

$$U_1(\theta) \propto \int_{-a/2}^{a/2} dx\, e^{ikx\sin\theta} \propto \frac{e^{ik\frac{a}{2}\sin\theta} - e^{-ik\frac{a}{2}\sin\theta}}{ik\sin\theta} \propto \text{sinc}\left(\frac{1}{2}ka\sin\theta\right)$$

This is the typical single slit pattern that we encountered in the Optics course during the First Year course.

$$U_2(\theta) \propto \int_{-\infty}^{\infty} dx\, e^{ikx\sin\theta} \sum_{m=0}^{N-1} \delta(x - md) \propto \underbrace{1 + e^{ikd\sin\theta} + e^{2ikd\sin\theta} + e^{3ikd\sin\theta} + \dots}_{\text{sum as a series}}$$
$$\propto \frac{1 - e^{iNkd\sin\theta}}{1 - e^{ikd\sin\theta}} \propto \frac{\sin\left(\frac{1}{2}Nkd\sin\theta\right)}{\sin\left(\frac{1}{2}kd\sin\theta\right)}$$

Putting these together, we arrive at the final result of

$$\boxed{I(\theta) = \frac{I(0)}{N^2}\, \underbrace{\text{sinc}^2\left(\frac{1}{2}ka\sin\theta\right)}_{\text{envelope}}\, \underbrace{\frac{\sin^2\left(\frac{1}{2}Nkd\sin\theta\right)}{\sin^2\left(\frac{1}{2}kd\sin\theta\right)}}_{\text{diffraction pattern}}} \tag{3.4}$$

The normalisation factor comes from the fact that we want the intensity pattern to have value $I(0)$ at $\theta = 0$; expanding the sine functions for small values of their argument allows us to see that this is the correct normalisation factor.

The maxima of the intensity pattern occur at the values for which the sine function on the denominator is zero, namely:

$$kd\sin\theta = 2\pi p$$

where p is an integer corresponding to the order of the maxima. On the other hand, the minima occur both at the zeros of the sinc and sine functions:

$$Nkd\sin\theta = 2\pi n \quad \text{and} \quad ka\sin\theta = 2\pi m$$

for integers n and m. This means that there are $N-1$ subsidiary maxima from the diffraction pattern between each principal maxima. What happens if we now let $d = \ell a$? It is quickly obvious that this means that every ℓ^{th} maxima of the diffraction pattern coincides with a minima of the envelope, and so is missing from our diffraction pattern. In the limiting case where $d \longrightarrow a$:

$$\lim_{d\to a} I(\theta) \propto \frac{\cancel{\sin\left(\frac{1}{2}ka\sin\theta\right)}}{\frac{1}{2}ka\sin\theta}\, \frac{\sin^2\left(\frac{1}{2}Nka\sin\theta\right)}{\cancel{\sin^2\left(\frac{1}{2}ka\sin\theta\right)}} \propto \text{sinc}^2\left(\frac{1}{2}k(Na)\sin\theta\right)$$

Thus, it becomes equivalent to a single slit of width Na. This is because the spatial frequency of the grating is on the order of that of a single slit, and so one cannot distinguish between the different slits.

The Reflection Grating

Another type of optical grating is that of the reflection grating, consisting of a series of reflective mirrors in the place of the slits. This system can add a phase difference depending on the incident angle. Consider the figure below. Evidently, for each individual mirror surface, $\phi = \theta$. In the same perpendicular distance ℓ from the surface, the incoming light ray has horizontal optical path length $k \sin\phi$, while the outgoing ray has $k \sin\theta$. This introduces a phase difference of

$$\boxed{\delta = kd(\sin\theta - \sin\phi)} \tag{3.5}$$

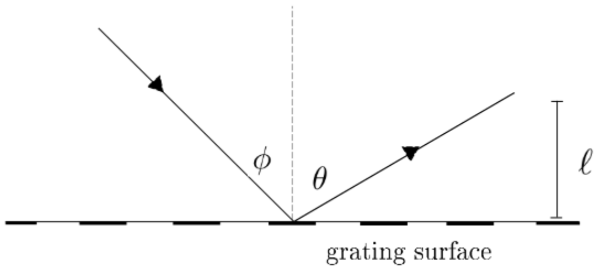

Figure 3.2: Phase difference for a reflection grating

This is known as *the grating equation*. Note that if $\theta = -\phi$, the light will return through the source slit.

We obtain the same intensity pattern as the diffraction grating, except with this new phase term. In zeroth order, $\theta = \phi$ (corresponding to reflection off individual mirrors), meaning that we cannot use it as a spectrometer in the lowest order. Furthermore, the intensity is significantly reduced as a result of the fact that only half of the grating surface is reflective. This issue can be fixed by *blazing*; the mirrors are tiled by an angle γ to the surface, allowing us to shift the central maxima away from the zeroth order to areas of higher dispersion.

3.1.2 Abbe's Theory of Imaging

Abbe's Theory of Imaging characterises the action of an optical system as a Fourier transform of the initial object, part of which is sampled and inverse Fourier transformed by the imaging system, such as the lens. This helps to explain why diffraction patterns are limited by the the resolution of the imaging system, as we will see in the next section.

Let us now go about demonstrating this. Consider an 'object' $f(x)$ (such as a diffraction aperture) that is illuminated by plane, monochromatic light of wavelength λ. The diffracted rays ass through a lens of focal length f placed at a distance u from the object, as shown in Figure (3.3).

Let r_1 be the optical path from the object plane (x) to the image plane (x'). This means that we can write $r_1 = \underline{xx'} = \underline{ox'} - x\sin\theta$. We will now make use of the Fresnel-Kirchoff diffraction integral as given by Equation (3.3), except including the In the focal plane of the lens:

$$\widetilde{f}(x') = \frac{i}{\lambda}\int dx\, e^{ikr_1}\, f(x) = \frac{ie^{ik\underline{ox'}}}{\lambda} \underbrace{\int dx\, e^{-ikx\sin\theta}\, f(x)}_{\text{Fourier transform } F(k\sin\theta)}$$

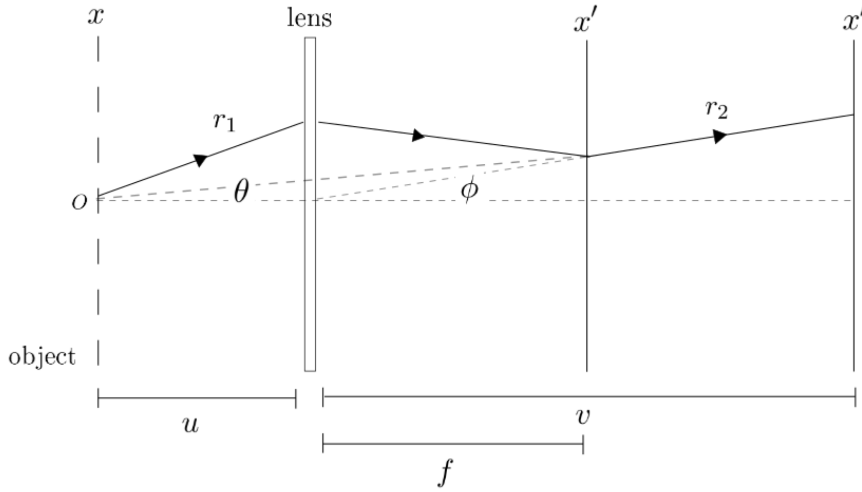

Figure 3.3: Demonstrating Abbe's Theory of Imaging

Then, considering the resultant amplitude in the image plane:

$$f_I(x') = \frac{i}{\lambda} \int dx' \, e^{ikr_2} \, \widetilde{f}(x') = \frac{-1}{\lambda^2} \int dx' \, e^{ik(r_2 + \underline{ox'})} \, F(k \sin\theta)$$

This means that the image plane at v is the conjugate plane of the source, and is the inverse Fourier transform of the amplitude sampled by the lens, as stated above.

Consider two slits of width a who's centres are separated by a distance d. Find an expression for the intensity pattern as a function of the angle θ between the direction of observation and the normal to the slits. The diffraction pattern is imaged by a lens of focal length f at a distance u from the lens. In the case where $d = 3a$, make a sketch of the intensity distribution: (i) in the plane of the slits, (ii) in the focal plane of the lens, and (iii) in the image plane of the lens at a distance v behind the lens. Note that $v > a^2/\lambda$.

Clearly, the intensity pattern will be as result of the convolution of the pattern due to a single slit of width a, and two point sources separated by a distance d. By the convolution theorem:

$$U(\theta) = \underbrace{U_1(\theta)}_{\text{finite single slit}} \quad \underbrace{U_2(\theta)}_{\text{double slits}}$$

Let us calculate these amplitudes, making use of the transmission functions given in Section (3.1.1).

$$U_1(\theta) = U_0 \int_{-a/2}^{a/2} dx \, e^{-ikx\sin\theta} \propto \text{sinc}\left(\frac{1}{2} ka \sin\theta\right)$$

and

$$U_2(\theta) = U_0 \int_{-\infty}^{\infty} dx \, \delta(x + d/2) + \delta(x - d/2) \propto \cos\left(\frac{1}{2} kd \sin\theta\right)$$

The intensity pattern is thus given by

$$I(\theta) = I(0) \, \text{sinc}^2\left(\frac{1}{2} ka \sin\theta\right) \cos^2\left(\frac{1}{2} kd \sin\theta\right)$$

The intensity pattern in the slit is just the mod-square of the result of the convolution of the two transmission functions; that is, two 'top-hat' functions, as shown above. In the plane of the lens, we observe the normal Fraunhoffer diffraction pattern, with every third interference maxima missing due to a zero of the envelope (recall that $d = 3a$).

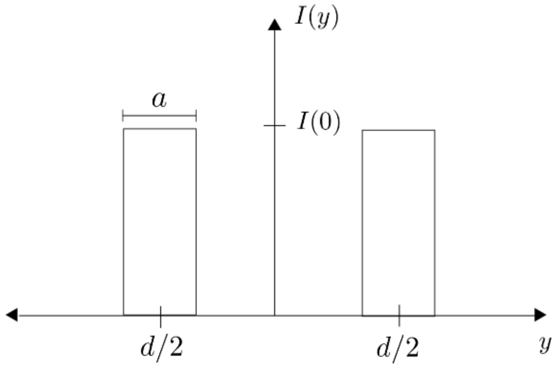

Figure 3.4: The intensity distribution in the plane of the slits

At a distance v from the slits, we will now observe the same pattern as in the plane of the slits, except that it has been magnified by a factor of v/u due to the lens. There will also be a reduction in clarity or detail due to the finite diameter of the lens, leading to diffraction limited imaging (see the following section).

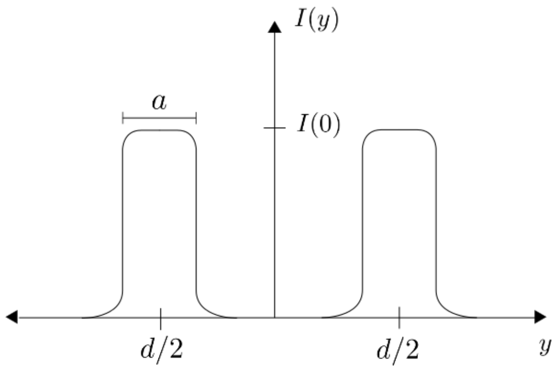

Figure 3.5: The intensity distribution at a distance v from the slits

Now we can ask ourselves the question as to what occurs when the lens is removed. As we are told that $v \gg a^2/\lambda$, we will still get a Fraunhoffer diffraction pattern without the lens, except that it will be magnified by a factor of $(v + u)$ due to propagation. This is because the Fraunhoffer diffraction pattern is formed by angles of equal inclination. In the context of Abbe's theory of imaging, the inverse Fourier transform is due to the propagation of the wave-front. This only occurs at v because there is phase information encoded in the waves at the plane of the lens that takes into account where the object is (i.e. of the distance u).

3.1.3 Diffraction Limited Imaging

Following on from Abbe's Theory of Imaging, the minimum dimension of the structure in the object corresponds to the maximum angle of incidence for which the objective lens can collect light (and so depends on the size of the objective lens) according to:

$$\boxed{\sin\theta_{max} = \frac{\lambda}{d_{\min}}} \tag{3.6}$$

Note that the right-hand side of this equation should be multiplied by a factor of 1.22 if the aperture is circular. This is also the minimum angular separation for which two point sources resolved under the Rayleigh Criterion:

> *Two objects are considered distinguishable if and only if the limit of the maxima of one falls on the minima of the other.*

These two factors together imply that the resolution of a system is limited by diffraction. In the subsequent sections, we will investigate some optical devices that are limited in this way.

Plano-convex Lens

Consider a plano-convex lens, as shown in the figure below. The smallest spatial scales of the object are only sampled if, under the small angle approximation:

$$\theta = \frac{\lambda}{d} < \frac{D}{2u} \longrightarrow d_{min} = \frac{2\lambda u}{D}$$

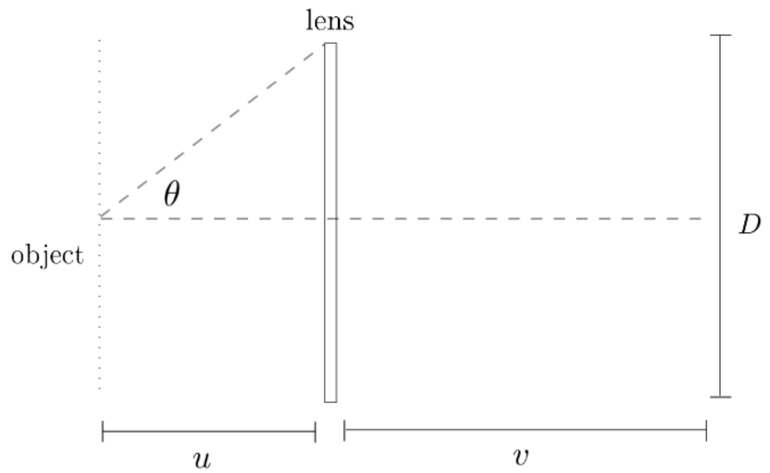

Figure 3.6: A simple plano-convex lens

If the lens is now immersed in some medium with refractive index n_0, this becomes

$$\boxed{n_0 \sin\theta_{max} = \frac{\lambda}{d_{min}} = NA} \tag{3.7}$$

NA stands for *numerical aperture*, which is in fact defined by this equation; it is a dimensionless number that characterises the range of angles over which a system can accept or emit light.

Compound Microscope

The compound microscope consists of two lenses, an objective lens that forms a real, inverted image, and the eyepiece lens that allows the real image to be viewed at infinity.

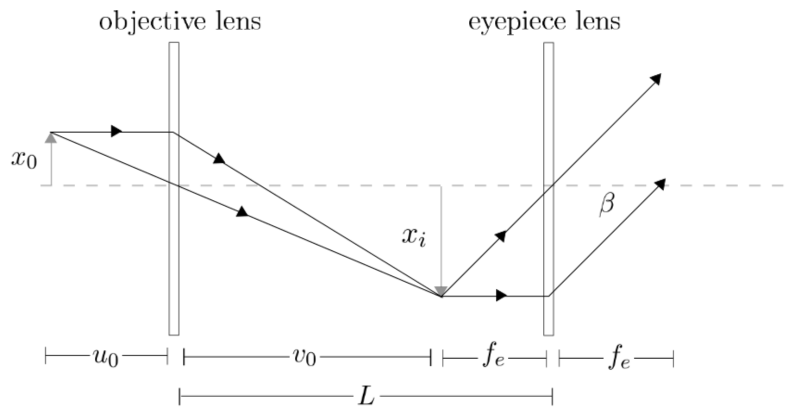

Figure 3.7: A schematic diagram of a compound microscope

The angular size seen in the microscope is β. The angular size subtended at the *near point* $D \backsim 25$ cm of the eye is $\alpha = x_o/D$. It is clear from the figure that $\beta = x_i/f_e$. Then, the total magnification is

$$M = \frac{\beta}{\alpha} = \frac{D}{f_e}\frac{x_i}{x_o} = \frac{D}{f_e}\frac{v_0}{u_0}$$

Thus, the total magnification of the system is the product of the magnifications that result from the individual lenses. Under normal conditions, typically $u_0 \backsim f_0$ and $v_0 \backsim L$, where L is known as the *tube length*. This means that we can write the magnification in the form

$$\boxed{M = \frac{D}{f_e}\frac{L}{f_0}} \tag{3.8}$$

Telescope

Telescopes are essentially used to increase the angular separation between two points in the observation plane, resulting in magnification. This can either be done by reflection or refraction. Schematically, telescopes appear very similar to Figure (3.7), except that the rays are incident parallel to one another at some angle α, and that $v_0 = f_0$. This means that the magnification of the telescope can simply be shown to be the ratio of the objective focal length to the eyepiece focal length; namely:

$$\boxed{M = \frac{f_0}{f_e}} \tag{3.9}$$

What about it's resolution? The easiest way to argue this is in reverse by asking the question 'what would a point source look like if projected from a telescope. Well, it must simply be the point spread function of the telescope. If the diameter of the telescope is D, then the minimum resolvable angle is

$$\boxed{\theta_{\text{resolution}} \backsim 1.22\frac{\lambda}{D}} \tag{3.10}$$

Optical Fibre

These utilise total internal reflection to keep the light travelling inside the fibre. Consider a cylinder of refractive index n_f, clad in an insulator of refractive index n_i ($n_i < n_f$).

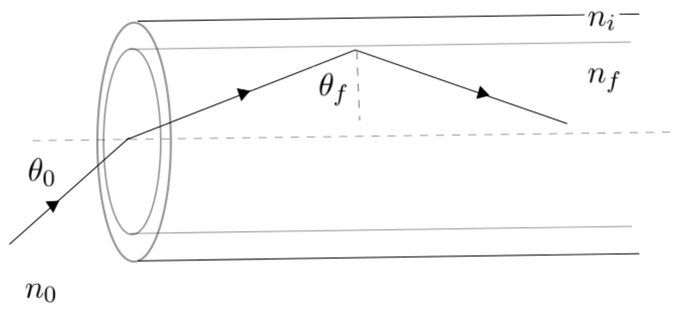

Figure 3.8: The end of an optical fibre

Recall Snell's law at the interface between the cylinder and the cladding.

$$n_f \sin\theta_f = n_i \sin\theta_i$$

When $\sin\theta_f > n_i/n_f$, that is, greater than the critical angle, then we have total internal reflection. In the case of an optical fibre, the numerical aperture is the maximum angle at which a ray can enter the fibre and still undergo total internal reflection. Apply Snell's law at the end of the fibre:

$$n_0 \sin\theta_0 = n_f \sin\left(\frac{\pi}{2} - \theta_f\right) = \cos\theta_f = \sqrt{1 - \sin^2\theta_f}$$

But the condition for total internal reflection is that $\sin\theta_f > n_i/n_f$, such that

$$n_f \cos\theta_f = n_0 \sin\theta_0 < \sqrt{n_f^2 - n_i^2}$$

We thus define

$$\boxed{NA = \sqrt{n_f^2 - n_i^2} = n_0 \sin\theta_0^{\max}} \tag{3.11}$$

Note that this is slightly different to the definition that we have used for lenses and their diffraction limit. What the two definitions have in common, however, is that they both set a maximum angle at which the component can accept information.

Diffraction Limited Brightness

Evidently, as systems are limited by the maximum angle to which they can collect and subsequently emit light, the energy flux into the system is limited. The $F\#$ (said 'F-number'), measures the amount of light that a lens can collect. We know that the energy collected is proportional to the area of the entrance pupil. The flux density is then related both to the image area ($\propto f^2$) and the energy collected ($\propto D^2$). Thus

$$\text{Flux Density } \propto \frac{\text{Energy collected}}{\text{Image area}} \propto \left(\frac{D}{f}\right)^2$$

We thus defined the $F\#$ as

$$\boxed{F\# = \frac{f}{D}} \tag{3.12}$$

This means that the flux density is proportional to one over the square of the $F\#$. If the system is diffraction limited, then the image area is approximately $(\lambda/D)^2$. This means that the flux density is proportional to D^4, where D is the diameter of the lends. This actually means that a small change in the size of the aperture will give a bigger gain in image brightness than if one is not limited by diffraction.

Finite Sources

We are now going to look at the case where the source pertaining to an optical system is finite, and thus limited by diffraction. Suppose that we have a finite source of width s, that is placed in the back focal plane of a collimating lens of focal length f. Then by Equation (3.6), we know that the wavelength spread that results from diffraction is

$$\Delta\lambda \backsim d_{\text{min}}\theta_{\text{max}} \backsim \frac{sD}{f}$$

We define the *bandpass* of an optical system as *the range of wavelengths that the system collects*, given by

$$\boxed{\Delta\lambda_{\text{ bandpass}} = \frac{sD}{f}} \tag{3.13}$$

for a lens of diameter D. This means that wavelengths outside this range will not be collected by the optical system.

It is also worth considering the angular shift in the diffraction pattern that results from the finite nature of the source slit. Suppose again that the source slit is in the back focal plane of a collimating lens of focal length f_1. The diffraction pattern is observed in the focal plane of a focal length f_2. The half angle subtended at the lens by the rays leaving the slit is given by

$$\alpha \backsim \frac{s}{2f_1}$$

This means that the principle maxima will experience a linear shift of $x_0 = \alpha f_2$ in the plane of observation. For this effect to be neglected, we require that $x_0 \ll x_p$, where x_p is the linear position of the p^{th} principle maximum.

3.2 Spectroscopic Instruments

We are now going to delve into the realm of spectroscopy (loosely, the measurement of the intensity as a function of wavelength), and take a look at some spectroscopic instruments. As such, we will be using notation that is more familiar in this domain; we will usually refer to wave-number, instead of wavelength, defined as

$$\bar{\nu} = \frac{1}{\lambda}$$

3.2.1 Fringe Formation

Thus far, we have just been assuming that the light we have been considering is coherent, and monochromatic. As such, we are able to add wave amplitudes, as the single wavelength can interfere with itself to produce the fringe pattern. However, for incoherent waves, or waves of different wavelengths, we add the intensity patterns, as each wavelength does not effect the others. In all cases, fringes will occur where the sum (over all sources and wavelengths) of intensity is constructive.

Fringes are said to be *localised* if they can only be seen at some subset of places where the beams cross. They are then *non-localised* if they can be seen everywhere the beams cross. Generally, *point sources* tend to produce non-localised fringes, as the rays leaving a point source interfere constructively each time that they cross. *Extended sources* can be modelled as a series of incoherent point sources. These produce non-localised fringes for every point within the source, but there is only some finite region in which the sum of the intensity patterns is constructive, and will give good fringes, instead of washing out to a constant. This means that extended sources, in general, produce localised fringes.

Fringes of equal inclination are formed by combining parallel rays that arrive at infinity. As a result, these will be localised at infinity, and will be circular rings. This is because the fringe locations only depend on angle from the central axis, meaning that we can make arbitrary rotations associated with cylindrical symmetry.

3.2.2 Some Important Definitions

There are some important definitions that we need to cover before investigating two common interferometers in the coming sections. These are as follows

- Free Spectral Range ($\Delta\bar{\nu}_{\text{FSR}}$) - This is the largest wave-number difference at which adjacent diffraction orders do not overlap. To calculate the free spectral range, find the change in $\bar{\nu}$ that increases the phase difference δ by 2π.

- Instrumental Width ($\Delta\bar{\nu}_{\text{INST}}$) - This measures the width of the wave-number peaks as the phase difference is changed. To calculate the instrumental width, equate the full-width-half-maximum (FWHM) for the peaks to the change in phase difference $\Delta\delta$.

- Resolving Power (RP) - This measures the smallest wave-number or wavelength difference that the instrument can resolve. For order p, it is defined by

$$\boxed{RP = \frac{\lambda}{\Delta\lambda_{\text{INST}}} = \frac{\bar{\nu}}{\Delta\bar{\nu}_{\text{INST}}}} \tag{3.14}$$

We shall put more of these definitions into practise as we have a look at both the Michelson and Fabry-Perot interferometers.

3.2.3 The Michelson Interferometer

The Michelson interferometer is common configuration for optical interferometry and was invented by Albert Abraham Michelson. A schematic diagram of the apparatus can be seen in the figure below.

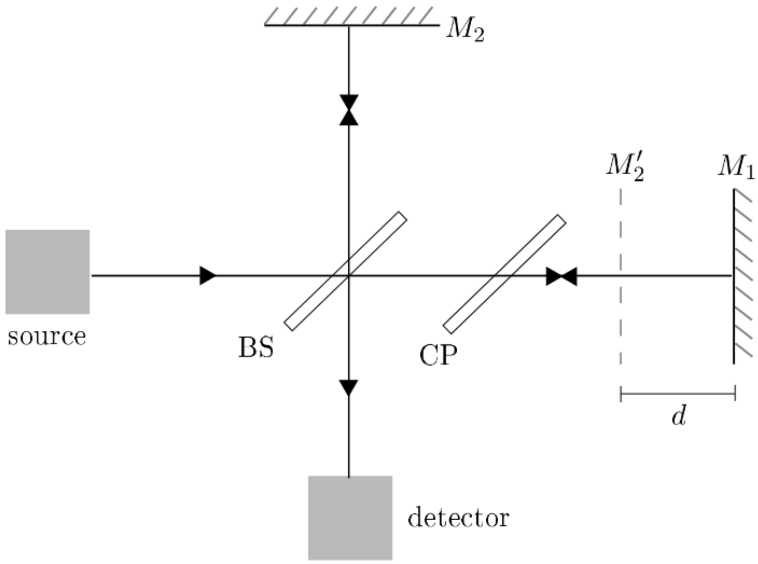

Figure 3.9: The Michelson Interferometer

The *beam splitter* (BS) created two orthogonal beams that travel down the arms of the apparatus, and reflect off the two mirrors M_1 and M_2. M_2' represents the image of M_2 in-front of M_1. The *compensating plate* is designed to account for the fact that the vertical beam has a longer optical path length as a result of spending more time in the optically denser beam splitter. It thus ensures that the phase difference between the two beams is due solely to the mirror separation d, and so will be zero at zeroth order.

Circular fringes are created when the mirrors are orientated at right angles to one another, and are both perpendicular to the optical axis. This creates the symmetry required for circular fringes. They are thus fringes of equal inclination, localised at infinity, as they are created by the phase difference between outgoing, parallel rays. Straight, equally spaced fringes are observed when there is a small angle between M_1 and M_2', and when there is only a small separation between the mirrors. These are localised in the plane of the mirrors.

Interference Pattern

To find the interference pattern, we first need to derive the phase difference that results from some mirror separation d. Consider the equivalent set-up for the Michelson apparatus shown in the figure below. Beam (1) travels an extra $AB+BC$ between the mirror surfaces. This means that the path difference is given by:

$$\Delta x = AB + BC = AB + AB\cos 2\theta = \frac{d}{\cos\theta}(1+\cos 2\theta) = 2d\cos\theta$$

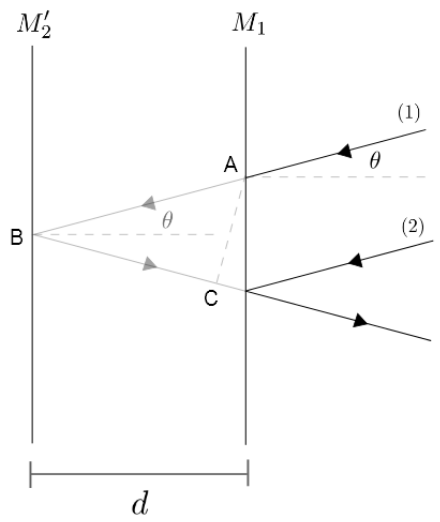

Figure 3.10: Equivalent Set-up for Michelson Interferometer

Assuming now that the medium separating the mirrors has a refractive index n, then the total phase difference between the two beams is given by

$$\boxed{\delta = 4\pi n\bar{\nu} d\cos\theta} \tag{3.15}$$

The resultant amplitude is then given simply by the sum of the amplitudes of the individual beams, namely

$$U(\delta) = U_0 + U_0 e^{i\delta} = U_0\left(1 + e^{i\delta}\right)$$

This means that the intensity profile for a Michelson that is illuminated by a single wavelength is given by

$$\boxed{I(\delta) = \frac{I_0}{2}\left[1 + \cos\delta\right]} \tag{3.16}$$

The normalisation has come from the fact that we have to re-obtain the incident intensity at zero phase difference.

Fringes

There is a central dark fringe at $\theta = 0$. Let this have order p_0. Then, clearly

$$2d = p_0\lambda \longrightarrow p_0 = \frac{2d}{\lambda}$$

If one substitutes sensible values for λ and d, they will discover that this is in fact not only an integer, but a very large number! This is because it turns out that the highest order occurs for $\theta = 0$, with the order decreasing for higher angles. This is because the linear displacement from the centre is very large for large angles, and so the phase difference between the beams becomes vanishingly small in comparison to their propagation.

Let us now consider the order for angles $\theta \neq 0$. We shall define our order relative to p_0.

$$2d\cos\theta = (p_0 - p)\lambda$$

$$2d\left(1 - \frac{\theta_p^2}{2} + \dots\right) = p_0\lambda - p\lambda$$

$$\theta_p^2\, d \backsim p\lambda$$

where we have ignored terms of $\mathcal{O}(\theta^3)$ and higher. This means that each order is roughly located at an angular displacement of

$$\boxed{\theta_p = \sqrt{\frac{p\lambda}{d}}} \tag{3.17}$$

Let us now investigate how the order changes when we vary some of the parameters of the system. In both cases below, we will be working in the regime where $\cos\theta \backsim 1$.

- Variation with d - From the definition of δ given by Equation (3.15), is is clear that

$$\lambda \Delta p = 2\Delta d$$

- Variation with λ - Again using the definition of δ:

$$\frac{\Delta p}{\Delta \lambda} = -\frac{2d}{\lambda^2} \xrightarrow{\Delta p \backsim 1} \Delta\lambda_{\text{ FSR}} \backsim \frac{\lambda^2}{2d}$$

The second expression is the change in wavelength for fringes of different orders to overlap. If this is less than the spread of wavelengths, then we will not see fringes.

With Multiple Wavelengths

Suppose that the beams now contain two wavelengths with associated wavenumbers $\bar{\nu}_1$ and $\bar{\nu}_2$, both with the same intensity. As stated in Section (3.2.1), we add the intensities of different wave-numbers. Using double-angle formulae, and for $\theta = 0$, it is easy to show that the interference pattern is given by

$$\boxed{I(d) = \frac{I(0)}{2}\left[1 + \underbrace{\cos\left(4\pi d\,\frac{\bar{\nu}_1 + \bar{\nu}_2}{2}\right)}_{\text{interference}} \overbrace{\cos\left(4\pi d\,\frac{\bar{\nu}_1 - \bar{\nu}_2}{2}\right)}^{\text{envelope}}\right]} \tag{3.18}$$

This means that we obtain the characteristic 'beat' pattern, as shown in the figure overleaf. The beat pattern will disappear at the zeros of the envelope, namely:

$$4\pi d\,\frac{\bar{\nu}_1 - \bar{\nu}_2}{2} = \frac{\pi}{2} + p\pi \longrightarrow d = \frac{2p+1}{4|\bar{\nu}_1 - \bar{\nu}_2|}$$

If the intensities of the two wave-numbers are not the same, there is incomplete cancellation, and the beat pattern does not go completely to zero, as in Figure (3.11), but the relationship above still approximately holds.

For these wavelengths to be resolvable, we require that we can scan the interferometer far enough through d such that this whole cycle of cancellation is visible - then we can tell that two wavelengths are present. This condition amounts to

$$4\pi d\,\frac{\Delta\bar{\nu}_{\text{ INST}}}{2} = \pi$$

meaning that the instrumental width of the Michelson is given by

$$\boxed{\Delta\bar{\nu}_{\text{ INST}} = \frac{1}{2d}} \tag{3.19}$$

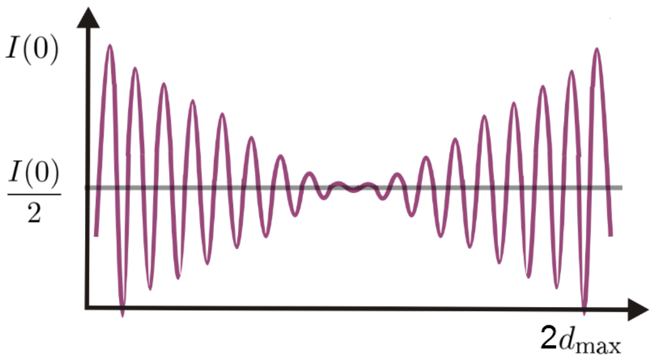

Figure 3.11: Two wave-number beat pattern

This means that the resolving power is straightforwardly given by

$$\boxed{RP = \frac{\bar{\nu}}{2d}} \tag{3.20}$$

Suppose that we have three wave-numbers $\bar{\nu}$, $\bar{\nu}+\delta\bar{\nu}$, and $\bar{\nu}-\delta\bar{\nu}$. The intensity of the latter two is half that of the first. Find an expression for the intensity observed at the detector.

As usual, the resultant intensity is just the sum of the intensities from each of the components.

$$\begin{aligned} I(d) &= 2a[1+\cos(4\pi\bar{\nu}d)] + a[1+\cos(4\pi d(\bar{\nu}-\delta\bar{\nu}))] + a[1+\cos(4\pi d(\bar{\nu}+\delta\bar{\nu}))] \\ &= a[4 + 2\cos(4\pi\bar{\nu}d) + 2\cos(4\pi\bar{\nu}d)\cos(4\pi d\delta\bar{\nu})] \\ &= a\left[4 + 2\cos(4\pi\bar{\nu}d)\left[1+\cos(4\pi d\delta\bar{\nu})\right]\right] \end{aligned}$$

Thus, the final pattern can be written as

$$I(d) = \frac{I(0)}{2}\left[1 + \cos^2(2\pi\delta\bar{\nu}d)\cos(4\pi\bar{\nu}d)\right]$$

Doppler Broadening

Doppler broadening is the broadening of spectral lines due to the Doppler effect caused by a distribution of velocities of atoms or molecules. Different velocities of the emitting particles result in different Doppler shifts, the cumulative effect of which is the line broadening. This essentially means that we can no longer consider our wave-number profile to be consisting of a series of perfect delta-functions; that is, each has a finite spectrum. This creates a finite *coherence length* after which no interference pattern can be observed.

Suppose that we place the axis of observation along the z-axis. Let $\bar{\nu}_0$ be the original wave-number of the source, and that v_z is the component of the velocity of an atom in the source along z. Then the shifted wave-number is given by

$$\bar{\nu} - \bar{\nu}_0 = \frac{v_z}{c}\bar{\nu}_0$$

We know that from Kinetic theory that the distribution of velocities is given by:

$$p(v_z) \propto e^{-\frac{mv_z^2}{2k_BT}}$$

This means that we can write the power spectrum of the source at $\bar{\nu}_0$ as

$$\boxed{p(\bar{\nu}) = \frac{\bar{\nu}_0}{\sqrt{\pi}}\left(\frac{c}{v_{th}}\right) e^{-\frac{c^2}{v_{th}^2}\left(\frac{\bar{\nu}-\bar{\nu}_0}{\bar{\nu}_0}\right)^2}} \tag{3.21}$$

Fourier Transform Spectroscopy

The Michelson Interferometer can be used for Fourier transform spectroscopy to find the power spectrum $p(\bar{\nu})$ of the source. The intensity pattern observed at the detector is simply the cosine Fourier transform of $p(\bar{\nu})$, which can then be found by computing a subsequent inverse Fourier transform.

$$I(x) = \text{const.} + \int d\bar{\nu}\, p(\bar{\nu}) \cos(2kx) \tag{3.22}$$

$$p(\bar{\nu}) = \int dx\, (I(x) - I_0) \cos(2kx) \tag{3.23}$$

We define the ***visibility*** of the fringe pattern as

$$\boxed{V = \frac{I_{\text{max}} - I_{\text{min}}}{I_{\text{max}} + I_{\text{min}}}} \tag{3.24}$$

where $I_{\text{max}} - I_{\text{min}}$ is the difference in intensity between the intensities at the light and dark fringes; that is, the difference between the cosine term of the interference pattern taking values ± 1. Essentially, it is a measure of the width of the envelope of the intensity pattern. For a monochromatic source, this is simply unity.

Suppose that we have a monochromatic source that is Doppler broadened. What is the visibility of the pattern observed? Using Equation (3.22)

$$\begin{aligned} I(x) &= \text{const.} + \int d\bar{\nu}\, e^{-\frac{c^2}{v_{th}^2}\left(\frac{\bar{\nu}-\bar{\nu}_0}{\bar{\nu}_0}\right)^2} \cos(4\pi\bar{\nu}x) \\ &= \text{const.} + \int du\, e^{-\frac{c^2}{v_{th}^2}\left(\frac{u}{\bar{\nu}_0}\right)^2} \cos(4\pi(u+\bar{\nu}_0)x) \end{aligned}$$

Note that we have ignored the normalisation constant from $p(\bar{\nu})$ as this will cancel in our calculation of V. We make use of the result that

$$\int_{-\infty}^{\infty} dx\, e^{-a^2x^2} \cos(bx + c) = \frac{\sqrt{\pi}}{a} e^{-b^2/4a^2} \cos(c)$$

to find that the intensity pattern is given by

$$I(x) = \text{const.} + \frac{\sqrt{\pi}}{2a} e^{-(2\pi x)^2\left(\frac{v_{th}}{c}\bar{\nu}_0\right)^2} \cos(4\pi\bar{\nu}_0 x)$$

This is the normal cosine interference pattern that we associated with a single wave-number, except modulated by the exponential factor that defines the coherence length of the pattern. This means that the visibility is given by

$$V = e^{-(2\pi x)^2\left(\frac{v_{th}}{c}\bar{\nu}_0\right)^2} = e^{-\frac{8\pi^2 k_B T \bar{\nu}_0^2}{mc^2}x^2} = e^{-x^2/\alpha^2}$$

Evidently, $V \to 0$ for $x \to \infty$ as the spacing will exceed the characteristic coherence length α, and we expect that there is simply the mean average value left in the interference pattern. Suppose that the monochromatic source corresponds to a spectral emission line from a distant star. If given data for the intensity as a function of the separation x, this means that we can estimate α, and thus the temperature of the star.

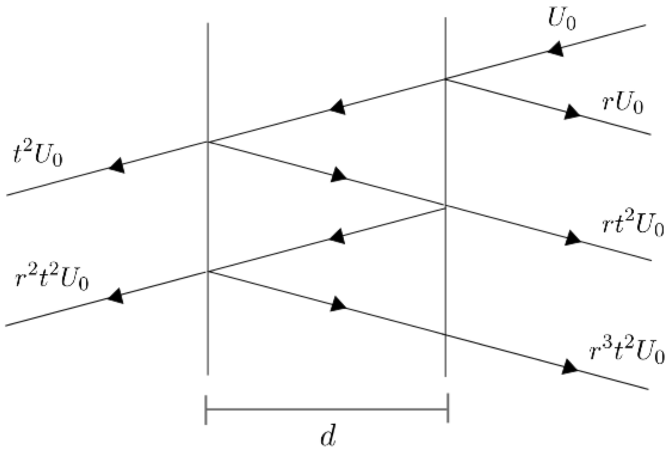

Figure 3.12: Path of light beam through the FPI

3.2.4 The Fabry-Perot Interferometer

Also known as the Fabry-Perot Etalon, the Fabry-Perot interferometer (FPI) consists of two surfaces of reflectivity $R = |r|^2$ and transmittance $T = |t|^2$.

The phase difference between adjacent rays is clearly given by (3.15). As a result of the similarity with the Michelson (in terms of configuration), the fringes created by the FPI will have the same localisation as in the Michelson, as well as the same relationship for θ_p given by Equation (3.17). Calculating the transmitted amplitude:

$$U(\delta) = t^2U_0 + r^2t^2U_0e^{i\delta} + r^4t^2U_0e^{i2\delta} + \dots = U_0t^2 \underbrace{\left(1 + r^2e^{i\delta} + r^4e^{i2\delta} + \dots\right)}_{\text{infinite series}} = U_0t^2\,\frac{1}{1 - r^2e^{i\delta}}$$

$$I(\delta) \propto \frac{1}{1 + r^4 - 2r^2\cos\delta} \propto \frac{1}{1 + \frac{4R}{(1-R)^2}\sin^2\left(\frac{\delta}{2}\right)}$$

We can thus write the transmitted intensity of the FPI as

$$\boxed{I(\delta) = \frac{I_0}{1 + \frac{4\mathcal{F}^2}{\pi^2}\sin^2\left(\frac{\delta}{2}\right)}} \tag{3.25}$$

where the quantity $\mathcal{F}$ is known as the *finesse*, given by

$$\boxed{\mathcal{F} = \frac{\pi\sqrt{R}}{1 - R}} \tag{3.26}$$

Clearly, as $0 < R < 1$, higher finesse corresponds to higher reflectivity. The finesse can be interpreted (roughly) as the number of 'effective' rays that contribute to the output intensity. To see why this, we need to consider the intensity of the outgoing rays. The first ray has intensity I_0, the second R^2I_0, the N-th $R^{2N}I_0$. Suppose that the decrease in intensity corresponding to the N-th ray is given by

$$R^{2N} = e^{-6} \;\longrightarrow\; 2N\log R = -6$$

Assuming that $R \backsim 1$ (which is valid in most cases), we can expand the logarithm in terms of a small parameter $\beta = 1 - R$:

$$2N\log(1-\beta) \backsim -2N\beta \backsim -2N(1-R)$$

It follows that

$$N = \frac{3}{1-R} \backsim \mathcal{F}$$

Evidently, this is a rough approximation; we have been quite arbitrary with our definition of the decay in intensity required for us to neglect a ray. However, it does give a bit more of an intuitive way of thinking about the finesse.

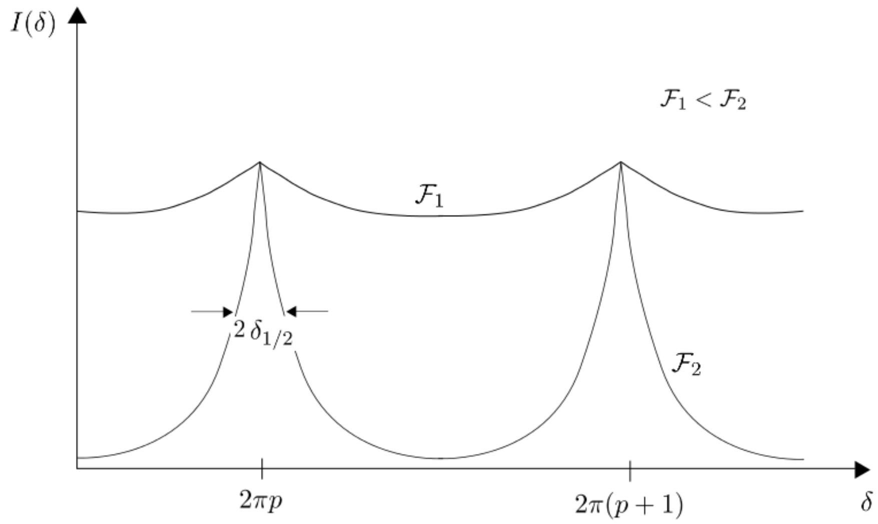

Figure 3.13: Intensity pattern for an FPI

Some Relevant Calculations

In a similar vein to with the Michelson, let us calculate some important quantities associated with the FPI.

- Free Spectral Range - As stated in Section (3.2.2), we need to find the change in $\bar{\nu}$ that causes a change in δ by 2π. Thus:

$$\Delta\delta = 4\pi n \Delta\bar{\nu} d\cos\theta$$
$$2\pi = 4\pi n \Delta\bar{\nu}_{\text{FSR}} d\cos\theta$$

  Thus, it is clear that

$$\boxed{\Delta\bar{\nu}_{\text{FSR}} = \frac{1}{2nd\cos\theta}} \tag{3.27}$$

- Half Width at Half Maximum - We need to let $I(\delta_{1/2}) = I_0/2$.

$$\frac{1}{2} = \frac{1}{1 + \frac{4\mathcal{F}^2}{\pi^2}\sin^2\left(\frac{1}{2}(\delta_{1/2} + 2\pi p)\right)}$$
$$\sin^2\left(\frac{\delta_{1/2}}{2}\right) = \frac{\pi^2}{4\mathcal{F}^2}$$

  Assuming that $\delta_{1/2}$ is small, we find that

$$\boxed{\delta_{1/2} \backsim \frac{\pi}{\mathcal{F}}} \tag{3.28}$$

- Instrumental Width - We can use the previous result to equate the FWHM to the change in δ:
$$\frac{2\pi}{\mathcal{F}} = 4\pi n \Delta\bar{\nu}_{\,\mathrm{INST}} d\cos\theta$$
resulting in
$$\boxed{\Delta\bar{\nu}_{\,\mathrm{INST}} = \frac{1}{\mathcal{F}}\Delta\bar{\nu}_{\,\mathrm{FSR}}} \tag{3.29}$$
- Resolving Power - Using the definition given by Equation (3.14), it is clear that
$$\boxed{RP = p\mathcal{F}} \tag{3.30}$$
for some order p. We have used the fact that fringes are located at $2d\cos\theta = p\lambda = \frac{p}{\bar{\nu}}$.

Analysing a Spectrum

In order to be able to interpret a spectrum consisting of two wave-numbers separated by $\Delta\bar{\nu}$, we must satisfy the condition that
$$\boxed{\Delta\bar{\nu}_{\,\mathrm{INST}} < \Delta\bar{\nu} < \Delta\bar{\nu}_{\,\mathrm{FSR}}} \tag{3.31}$$
If this is satisfied, then the q^{th} order peak from one wave-number will be between the p^{th} and the $(p+1)^{th}$ orders of the other. However, there is a certain ambiguity; the q^{th} order of $\bar{\nu} + \Delta\bar{\nu}$ is either near the p^{th} or $(p+1)^{th}$ order of $\bar{\nu}$, as shown in the following figure.

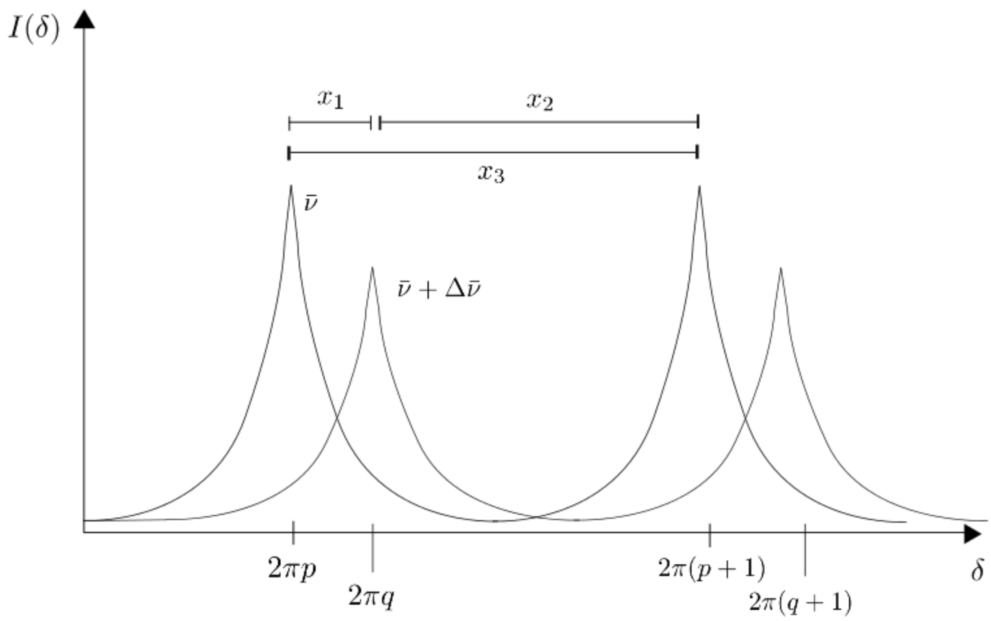

Figure 3.14: Resolving two wave-numbers

This means that we have either
$$\frac{\Delta\bar{\nu}}{\Delta\bar{\nu}_{\,\mathrm{FSR}}} = \frac{x_1}{x_1+x_2} \quad \text{or} \quad \frac{x_2}{x_1+x_2}$$
We can convert these to angular measurements by
$$\frac{\Delta\bar{\nu}}{\Delta\bar{\nu}_{\,\mathrm{FSR}}} = \frac{\cos\theta_q - \cos\theta_p}{\cos\theta_{(p+1)} - \cos\theta_p} \quad \text{or} \quad \frac{\cos\theta_{(p+1)} - \cos\theta_q}{\cos\theta_{(p+1)} - \cos\theta_p}$$

We can then convert to radial measurements by observing that $\theta^2 \propto r^2$. Of course, this does not solve out ambiguity, but merely gives us two possible values. The way to resolve the ambiguity would be to take another set of measurements at a different value for d, which should give another two values, only one of which will be the same as that in the original data.

There are also other considerations that have to be taken into account when analysing a spectrum, including:

- Etalon Design - The parameters that we have to tune are the optical thickness (nd) and the reflectivity (R) that will affect the Finesse, in order to satisfy the condition given by Equation (3.31). Ideally, one would choose the spacing such that the fringes corresponding to the second wave-number lie in the middle of an order. Otherwise, if the peaks lie close together, a large Finesse, and thus reflectivity, is required to distinguish the peaks.

- Illumination Type - We should also deal with the case where the light passing through the etalon is not a continuous beam, as we have thus far assumed, but a single laser pulse that satisfies $\tau\Delta\omega \backsim 1$, where $\Delta\omega$ is the frequency width of the pulse.

  $$\Delta\omega = \frac{2\pi c}{\lambda^2}\Delta\lambda_{\text{source}} \quad \longrightarrow \quad \Delta\lambda_{\text{source}} = \frac{\lambda^2}{2\pi c\tau}$$

  From the resolving power,

  $$\frac{\lambda}{\Delta\lambda_{\text{INST}}} = p\mathcal{F} = \frac{2d}{\lambda}\mathcal{F}$$

  Evidently, it is useless to have $\Delta\lambda_{\text{INST}} > \Delta\lambda_{\text{source}}$, because the etalon will be attempting to analyse spectral information that is not there. This places an upper limit on the useful finesse, and thus reflectivity, given by

  $$\mathcal{F} \leq \frac{\pi c\tau}{d}$$

- Parallelism - If the reflecting plates are not quite parallel, deviating at maximum by h, then this introduces on average an error of $2h$ every time this mirror is visited. We can estimate the effect of this by arguing that

  $$\Delta\bar{\nu}_{\text{ INST}} = \frac{1}{2nd\mathcal{F}} \backsim \frac{1}{\text{maximum optical path}}$$

  This means that the ray makes roughly $\mathcal{F}$ trips through the etalon. For coherence to hold, we require that $\lambda > 2h\mathcal{F}$. In other words, an upper bound on the practical Finesse is set by

  $$\boxed{\mathcal{F} < \frac{1}{2h\bar{\nu}}} \tag{3.32}$$

The optimum set-up for the FPI will thus evidently depend on the spectrum being analysed, and so the above points always need to be considered when configuring the apparatus.

3.3 Polarisation

The *polarisation* of a light wave is defined as the direction of the electric field as the wave propagates. For the analysis of this property, we often split the light up into orthogonal components, and use real wave representations. For light propagating along z:

$$E_x = E_{0x}\cos(kz - \omega t)$$
$$E_y = E_{0y}\cos(kz - \omega t + \phi(t))$$

Note that we have made the arbitrary choice to absorb the phase factor into the y component; this could also be put in the x component, or even both.

3.3.1 Types of Polarisation

There are four types of polarisation that are possible. These will depend on both the relative magnitude of E_{0x} and E_{0y}, as well as the behaviour of $\phi(t)$.

Un-polarised

Un-polarised light occurs where $\phi(t)$ is *stochastic* in nature; that is, it is random, or uncorrelated. This means that the direction of the electric field is not well behaved, and so we cannot give it a determinate value. Instead, we can decompose un-polarised light into two orthogonal polarisations.

Linear Polarisation

Linear polarisation occurs where both the direction and amplitude of the electric field remains constant. This means that we require $\phi(t) = 0$. The direction of polarisation is at some angle α to the y-axis, given by

$$\tan\alpha = \frac{E_{0x}}{E_{0y}}$$

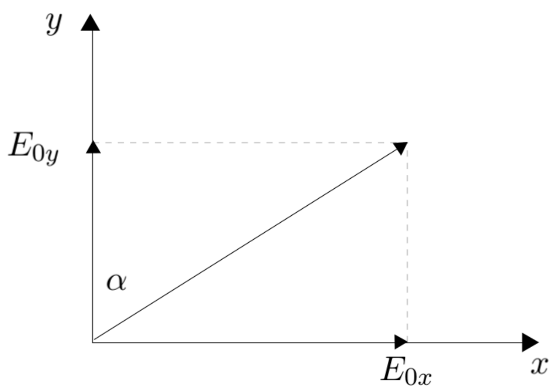

Figure 3.15: Orientation of linearly polarised light

Circular Polarisation

Circular polarisation occurs where the amplitude remains constant ($E_{0x} = E_{0y}$), and that the direction of orientation rotates at a constant angular frequency around the direction of propagation. For this, we require that the phase difference is given by

$$\boxed{\phi = \pm\frac{\pi}{2}} \tag{3.33}$$

Light with the positive sign above is known as *left-hand* polarised light, as it rotates anti-clockwise when looking back along the direction of propagation. Similarly, the negative sign is known as *right-hand* polarised. We can write these two polarisation types as

$$\begin{aligned}
\underline{E}_r &= E_0\cos(kz-\omega t)\,\widehat{\underline{x}} + E_0\sin(kz-\omega t)\,\widehat{\underline{y}} \\
\underline{E}_\ell &= E_0\cos(kz-\omega t)\,\widehat{\underline{x}} - E_0\sin(kz-\omega t)\,\widehat{\underline{y}}
\end{aligned}$$

meaning that the superposition of left and right hand polarised light gives linearly polarised light along the x-axis:

$$\underline{E} = \underline{E}_r + \underline{E}_\ell = 2E_0\cos(kz-\omega t)\,\widehat{\underline{x}}$$

For circularly polarised light, the direction of the $\underline{E}$ field is given by

$$\tan\alpha = \frac{E_y}{E_x} = \pm\tan(kz-\omega t)$$

Elliptical Polarisation

Elliptical polarisation occurs for unequal field amplitudes ($E_{0x} \neq E_{0y}$), and where ϕ is not necessarily equal to $\pm\pi/2$. This means that in general we can write

$$\begin{aligned}
E_x &= E_{0x}\cos(kz-\omega t) \\
E_y &= E_{0y}\cos(kz-\omega t+\phi) = E_{0y}(\cos(kz-\omega t)\cos\phi - \sin(kz-\omega t)\sin\phi)
\end{aligned}$$

Then:

$$\frac{E_y}{E_{0y}} - \frac{E_x}{E_{0x}}\cos\phi = \sin(kz-\omega t)\sin\phi = -\left(1-\frac{E_x^2}{E_{0x}^2}\right)^{1/2}\sin\phi$$

Squaring both sides,

$$\left(\frac{E_y}{E_{0y}} - \frac{E_x}{E_{0x}}\cos\phi\right)^2 = \left(1-\frac{E_x^2}{E_{0x}^2}\right)\sin^2\phi$$

Re-arranging, we arrive at

$$\boxed{\left(\frac{E_x}{E_{0x}}\right)^2 + \left(\frac{E_y}{E_{0y}}\right)^2 - 2\left(\frac{E_x}{E_{0x}}\right)\left(\frac{E_y}{E_{0y}}\right)\cos\phi = \sin^2\phi} \tag{3.34}$$

This is the general equation for an ellipse. We can diagonalise this as a quadratic for to find the eigenvectors, which will give the directions of the axes of the ellipse. In the case where $\phi = \frac{n\pi}{2}$, the ellipse is aligned with the x and y axes.

3.3.2 Birefringence

In some crystals, the relative permeability of the substance can depend on the orientation of the electric within it, rather than being isotropic. In particular, in birefringent or uniaxial materials, the refractive index that a light way experiences depends on whether the electric field vector is *perpendicular* (the ordinary ray, n_0), or *parallel* (the extraordinary ray, n_e) to the optical axis of the material. As the wave propagates through such a material, this will introduce a phase difference between the perpendicular and parallel components given by

$$\boxed{\Delta\phi = |n_e - n_0|k\ell} \tag{3.35}$$

where ℓ is the distance travelled by the light in the material. This fact is exploited by *wave-plates* in order to change the polarisation of light. In particular, there are two types:

- Quarter-wave plate - As one could guess, in a quarter-wave plate $\ell \propto \lambda/4$. This introduces a phase shift between the components of $\Delta\phi = \pm\pi/2$. This can be used to convert between linearly polarised, and circular/elliptically polarised light. To produce circularly polarised light from linearly polarised light, first place a linear polariser at 45° to the x-y components to ensure that equal intensities of each are transmitted. This will mean that $E_{0x} = E_{0y}$ in the circularly polarised light. Introducing a quarter-wave plate will then create the circularly polarised light. Similarly, one can convert in the opposite direction by using a polariser to determine the angle at which the transmitted light is of minimum or maximum intensity. Then, a quarter-wave plate aligned along one of these directions (corresponding to the minor and major axes of the ellipse respectively) will create linearly polarised light. The direction to the optical axis of the wave-plate will be given by

$$\tan\alpha = \frac{b}{a}$$

  where a and b are the major and minor axes of the ellipse respectively.

- Half-wave plate - As one could have guessed, $\ell \propto \lambda/2$. This can be used to rotate linearly polarised light. Suppose that the electric field is initially at an angle α to the optical axis, the component perpendicular will experience a phase shift of π. This amounts to an anticlockwise rotation of the polarisation by an angle 2α. This means that if we require a rotation by 2α, we need to place the wave-plate with the optical axis at an angle α to the polarisation, which can be determined using a polariser.

We can use a combination of these two plates to change the polarisation of light almost arbitrarily. Do not forget to use linear polarisers to determine polarisation directions.

Some Examples

Following are some examples of some questions involving birefringent materials, and wave-plates.

- *Estimate the thickness of a zero-order quarter wave calcite plate for light of wavelength* 589 *nm* ($n_0 = 1.658$ *and* $n_e = 1.486$)*. Find the thickness of an order-100 plate, and explain why this is more practical.*

  Evidently, we require $\Delta\phi = \pi/2$. Thus:

$$\frac{\pi}{2} = |n_0 - n_e|\frac{2\pi}{\lambda}\ell \longrightarrow \ell = \frac{\lambda}{4|n_0 - n_e|} \backsim 0.856\,\mu\text{m}$$

  For a order-p plate, we make the transformation that $\Delta\phi \mapsto \Delta\phi + 2\pi p$. Thus, for $p = 100$, the thickness is $\ell \backsim 0.343\,\text{mm}$. This is significantly more practical that the zeroth order plate as the latter is too thin to handle and manipulate properly.

- *A laser-light filter consists of a thin, plane-parallel quartz plate cut with its faces parallel to the optic axis. Laser light strikes the plate at an angle of* 55° *to the normal. The light is plane polarized in the plane of incidence and at* 45° *to the optic axis of the* *quartz. Slight maxima in intensity are found in the transmitted light for wavelengths around* 540 *nm and* 1080 *nm. No maxima are found at any longer wavelengths. Explain these observations. Given that the difference in principal refractive indices for quartz is* 9×10^{-3} *estimate the thickness of the plate.*

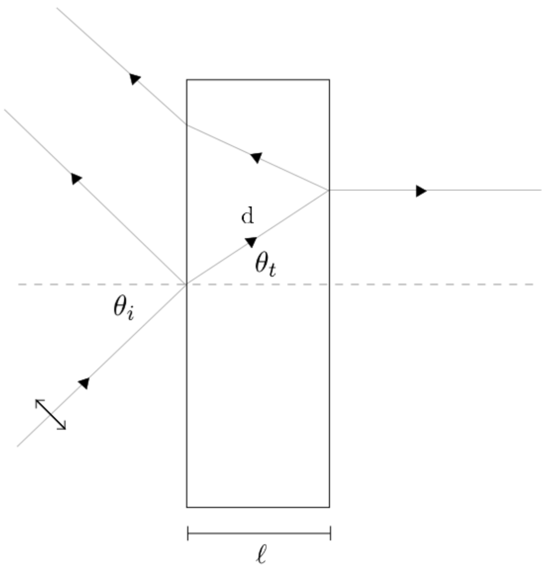

Figure 3.16: A diagram of the quartz plate

As $\theta_i = 55°$, this is close to Brewster's angle, meaning that very little light is reflected. At certain values of λ, the plate acts as a plate an integer multiple of λ thick, meaning that there is total transmission, which creates the slight maxima in intensity.

$$k\Delta nd = 2\pi \longrightarrow \lambda_{\text{max}} = d\Delta n \longrightarrow d \backsim 1.2 \times 10^{-4}\,\text{m}$$

To calculate the thickness of the plate, we want to find θ_t. Using Snell's law:

$$n_{\text{air}} \sin\theta_i = n_{\text{glass}} \sin\theta_t$$

Using $n_{\text{air}} \backsim 1$ and $n_{\text{glass}} \backsim 1.5$, we find that $\theta_t \backsim 33°$. With simple trigonometry, this gives $\ell \backsim 100\,\mu$m.

- *The incoming light is a combination of un-polarised and elliptically polarised light. When using a polarising filter, maximum intensity is observed when the axis of the filter is vertical, which is twice the minimum intensity (observed with axis horizontal). The beam is then passed through a quater-wave plate with it's 'fast-axis' vertical, followed by another polariser. It is now found that the maximum is at* $\alpha \backsim 33.21°$. *Find the ratio of the intensities of the elliptically polarised to un-polarised light.*

Assume that the un-polarised light is isotropically distributed, and has magnitude E_0. Let E_1 be the component of the elliptically polarised light oriented along the vertical (major axis), and E_2 be the component orientated along the horizontal. Then:

$$I_{\text{max}} = E_0^2 + E_1^2$$
$$I_{\text{min}} = E_0^2 + E_2^2$$

We are told that

$$\frac{E_0^2 + E_1^2}{E_0^2 + E_2^2} = 2 \longrightarrow E_0^2 = E_1^2 - 2E_2^2$$

After the quarter-wave plate, we have that

$$\tan\alpha = \frac{b}{a} = \frac{E_2}{E_1} \longrightarrow E_2^2 = \tan^2\alpha E_1^2$$

Putting these together, we find that $E_1^2 \backsim 7E_0^2$, and so that $I_{\max}/I_{\min} \backsim 5$.

Prisms

We can cut the uniaxial crystal into the form of a prism, which allows us to further manipulate light based on its polarisation.

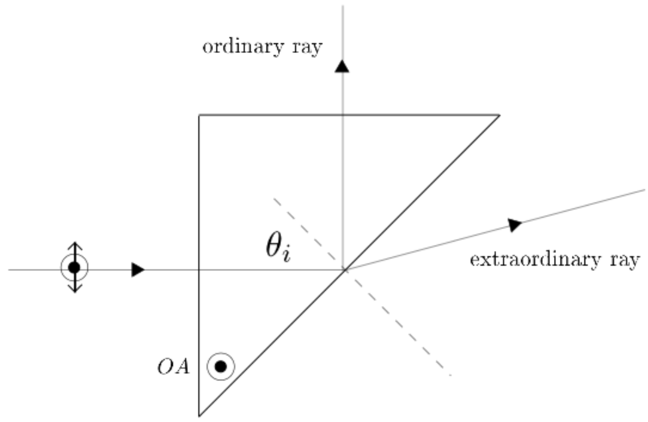

Figure 3.17: A birefringent prism

The extraordinary ray can pass if $\theta_i < \theta_c$ (critical angle), while the ordinary ray is totally internally reflected assuming that $n_0 > n_e$. Let us consider the particular case of the *Wollaston Prism*, as shown in the figure below. Suppose that the light incident on it is un-polarised. We can decompose this into components both parallel and perpendicular to the optical axis in the prism (1).

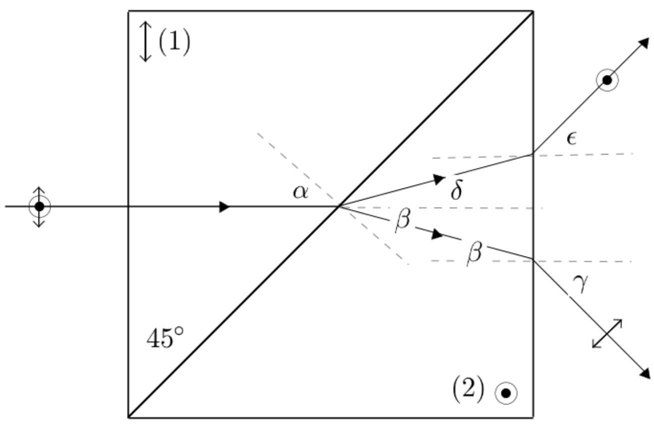

Figure 3.18: A Wollaston Prism

We will refer to our polarisation directions with respect to the optical axis of (1). Let us consider the two cases:

1. Parallel - In region (1), it experiences n_e, and in region (2), it experiences n_0. As $n_0 > n_e$, it will refract towards the normal. Applying Snell's Law at the two boundaries:

$$n_e \sin\alpha = n_0 \sin(\alpha - \beta)$$
$$n_0 \sin\beta = n_{\text{air}} \sin\gamma$$

As we know that $\sin\alpha = 45°$, this can then be solved for γ assuming we have values for the refractive indexes.

2. Perpendicular - In region (1), it experiences n_0, and in region (2), it experiences n_e. This means that it will refract away from the normal. Again applying Snell's Law at the two boundaries:

$$n_0 \sin\alpha = n_e \sin(\alpha + \delta)$$
$$n_e \sin\delta = n_{\text{air}} \sin\epsilon$$

Likewise, we can solve for ϵ.

The angle $\theta = \gamma + \epsilon$ is known as the *angle of divergence* of the two rays. If $\sin\alpha > n_e/n_0$, we would have total internal reflection of the perpendicularly polarised light, while the parallel component would behave as before.

If we change the optical axis in region (1) from $\updownarrow$ to $\leftrightarrow$, the polarisation direction that is initially $\updownarrow$ will remain undeviated, while the other will be refracted. This is an advantageous alternative as if there is a range of wavelengths, any refraction causes dispersion. There is no dispersion for the $\updownarrow$ polarised light, meaning that it would be good for separating a particular polarisation from another for spectral analysis.

Interference of Polarised Light

There is no interference between polarisations that are orthogonal, meaning that the total intensity is the sum of the intensity patterns of the components. If we are splitting an initially un-polarised beam, subsequent polarisations are uncorrelated, which causes fringes to becomes washed out. If a filter is placed immediately after the source, there is a common phase factor, but the polarisations will still be orthogonal. This can be fixed with the use of wave-plates.

For example, suppose that a source for a double-slit interference experiment emits incoherent light that is un-polarised. One of the slits is covered with a polarising material that transmits only horizontal polarisation, and the other vertical polarisation. To obtain interference, one would place a polariser at 45° before the slits to make the transmitted components coherent, and then a half-wave plate at 45° to one of the polarisations after the slits to rotate it to be parallel to the other.

Polarised Light and Brewster's Angle

Suppose that a beam of polarised light is incident on a boundary between two media at θ_B, as defined in Equation (1.49). We know that the reflected and transmitted (refracted) rays are created as a result of the action of electrons as dipoles within the material. These are only able to radiate energy perpendicular to the axis along which they oscillate. This means that the dipoles created by the light polarised in the plane of incidence are only able to radiate along the transmitted ray. On the other hand, the dipoles that are created by light polarised perpendicular to the plane of incidence can radiate both along the transmitted ray, but also the reflected ray (as the dipole axis is perpendicular to both). This means that the transmitted ray is partially polarised, while the reflected ray only has polarisation perpendicular to the plane of incidence. In this way, we can create polarised light from un-polarised light through reflection from a surface at θ_B.

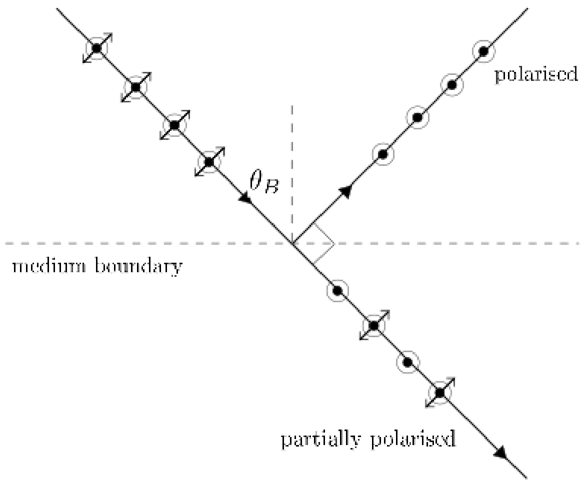

Figure 3.19: Polarisation upon reflection at θ_B