A2: Electromagnetism and Optics

Toby Adkins

June 20, 2016

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.

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

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

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

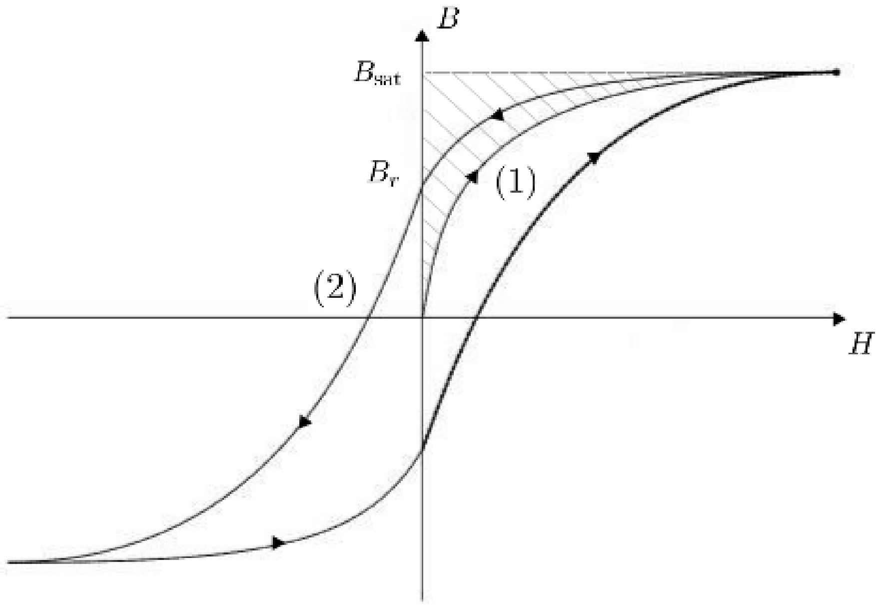

Figure 1.2: A typical hysteresis curve

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.

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

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

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.

This means that the radiated power is given by

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

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.

2.3 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.

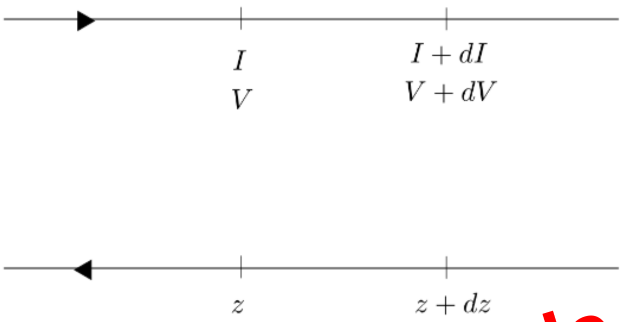

Figure 2.2: A schematic transmission line

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

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

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

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

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

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

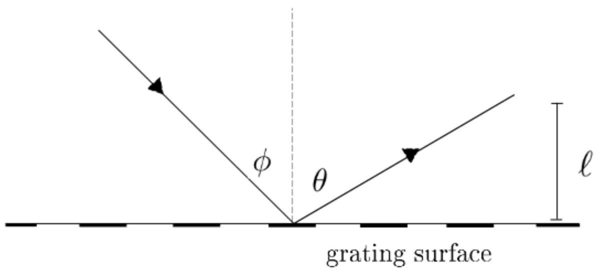

Figure 3.2: Phase difference for a reflection grating

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:

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

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.

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.

- Instrumental Width - We can use the previous result to equate the FWHM to the change in δ:

$$\frac{2\pi}{\mathcal{F}} = 4\pi n \Delta\bar{\nu}_{\text{INST}} d\cos\theta$$

resulting in

$$\boxed{\Delta\bar{\nu}_{\text{INST}} = \frac{1}{\mathcal{F}}\Delta\bar{\nu}_{\text{FSR}}} \tag{3.29}$$

- 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

$$\boxed{\Delta\bar{\nu}_{\text{INST}} < \Delta\bar{\nu} < \Delta\bar{\nu}_{\text{FSR}}} \tag{3.31}$$

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.

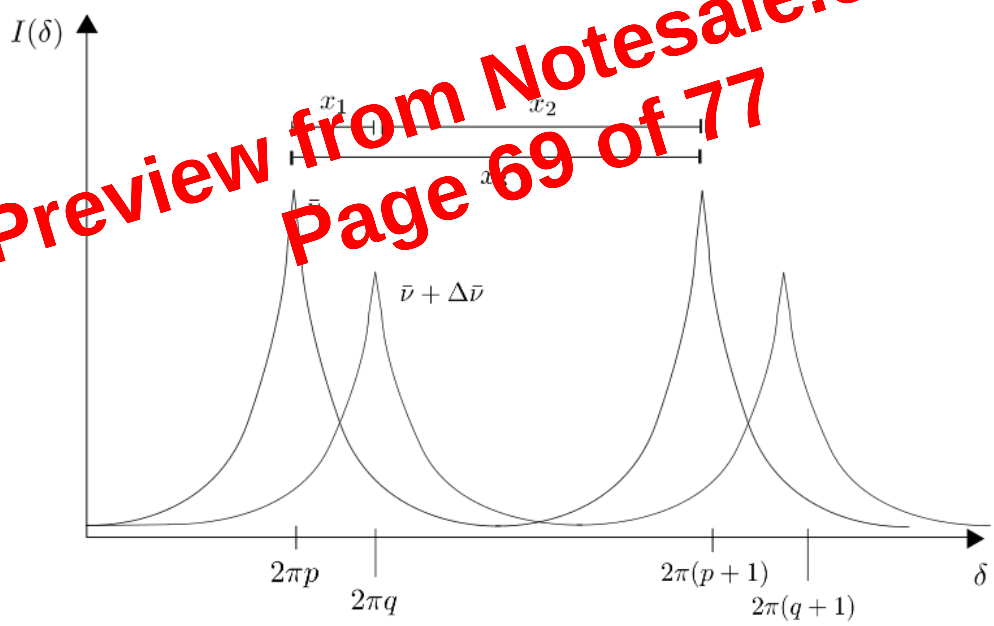

Figure 3.14: Resolving two wave-numbers

This means that we have either

$$\frac{\Delta\bar{\nu}}{\Delta\bar{\nu}_{\text{FSR}}} = \frac{x_1}{x_1+x_2} \quad \text{or} \quad \frac{x_2}{x_1+x_2}$$

We can convert these to angular measurements by

$$\frac{\Delta\bar{\nu}}{\Delta\bar{\nu}_{\text{FSR}}} = \frac{\cos\theta_q - \cos\theta_p}{\cos\theta_{(p+1)} - \cos\theta_p} \quad \text{or} \quad \frac{\cos\theta_{(p+1)} - \cos\theta_q}{\cos\theta_{(p+1)} - \cos\theta_p}$$

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.

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

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.