Optics

Alexander Byrne

Michaelmas 2020

1 Introduction

According to Huygens, each point on a wavefront generates a spherical "wavelet" which collectively carry the wave along. This can be used to derive Snell's Law, but also the diffraction pattern at a point on a screen a distance L away from an aperture. (Actually Huygens' Principle would lead to a backward-propagating wave as well, which is not observed, but we will be interested only in points pretty much directly in front of the source, so this point can be disregarded.) (Also the wavelets after the aperture have an amplitude of $-i/\lambda$ relative to the primary ones, but we will be interested only in the relative amplitudes of the diffraction pattern, so we ignore this too.) The wave amplitude decreases as $1/r$, so that the intensity decays as an inverse-square to conserve energy. Thus the wavelet generated by the element at (x, y) in the aperture plane thus has an amplitude proportional to:

$$d\psi_{(x,y)} \propto h(x, y)\frac{e^{ikr_1(x,y)}}{r_1(x, y)}\,\mathrm{d}x\,\mathrm{d}y$$

where $h(x, y)$ is the *aperture function* (typically just 1 for the aperture bit and 0 for the blocking screen) and $r_1(x, y)$ is the distance from the source to the element (x, y). Dropping all the functional dependences for aesthetics, the resultant amplitude at the point $P = (x_0, y_0)$ on the measuring screen is then proportional to:

$$\psi_P \propto \iint_\Sigma h(x, y)\frac{e^{ikr_1}}{r_1}\frac{e^{ikr_2}}{r_2}\,\mathrm{d}x\,\mathrm{d}y \tag{1}$$

where $r_2(x, y, x_0, y_0)$ is the distance from the element (x, y) to the point of interest $P = (x_0, y_0)$ (so $r_2 \geqslant L$), and Σ is the aperture plane, whose extent by the way we characterise by a size l for convenience, i.e. $x^2 + y^2 \leqslant l^2$.

2 Fraunhofer Diffraction

We look for solutions to the above in a certain regime:

- $r_1 \ggg l$. This means that $r_1(x,y)$ is essentially constant for all x, y, as is e^{ikr_1}.
- $L \gg l^2/\lambda$. This turns out to define Fraunhofer diffraction as the regime where *the phase varies linearly across the aperture*, as we will see.

With the first restrictions in mind, we can modify the integral (1) to:

$$\psi_P \propto \iint_\Sigma h(x,y)\frac{e^{ikr_2}}{r_2}\,\mathrm{d}x\,\mathrm{d}y \tag{2}$$

r_2 requires a closer look. It can be easily seen that

$$\begin{aligned}
r_2^2 &= L^2 + (x_0 - x)^2 + (y_0 - y)^2 \\
&= \underbrace{L^2 + x_0^2 + y_0^2}_{R^2} - 2(x_0 x + y_0 y) + x^2 + y^2 \\
&= R^2\left(1 - 2\frac{x_0 x + y_0 y}{R^2} + \frac{x^2+y^2}{R^2}\right) \\
\Rightarrow kr_2 = \frac{2\pi}{\lambda}r_2 &\approx \frac{2\pi}{\lambda}R\left(1 - \frac{x_0 x + y_0 y}{R^2} + \frac{x^2+y^2}{2R^2}\right) \\
&= \frac{2\pi R}{\lambda} - \frac{2\pi}{\lambda}\frac{x_0 x + y_0 y}{R} + \underbrace{\frac{\pi}{\lambda}\frac{x^2+y^2}{R}}_{\approx 0} \\
&\approx \frac{2\pi}{\lambda}\left(R - \frac{x_0 x + y_0 y}{R}\right) = kR - k\frac{x_0}{R}x - k\frac{y_0}{R}y
\end{aligned}$$

and as for the factor of $1/r_2$, we assume that to be constant $(1/R)$ and ignore the variation, which is likely to be small anyway. Large enough to detect the phase difference, which cycles around and is sometimes zero, but small enough to neglect on the denominator. Thus the only thing remaining to obtain the Fraunhofer diffraction equation is to define $(p,q) = (kx_0/R, ky_0/R)$, which themselves are often defined as $(k\sin\xi, k\sin\theta)$. We then have:

$$\psi_P \propto \iint_\Sigma h(x,y)e^{-ipx-iqy}\,\mathrm{d}x\,\mathrm{d}y \tag{3}$$

– the Fraunhofer Diffraction Equation, which is simply a 2-dimensional Fourier transform. If Σ, or equivalently $h(x,y)$, are constant in e.g. the x direction, then:

$$\psi_P \propto \int_\Sigma h(y)e^{-iqy}\,\mathrm{d}y \tag{4}$$

2.1 Diffraction Gratings

A diffraction grating of N slits has an aperture function:

$$h(y) = \sum_{n=0}^{N-1} \delta(y - nd)$$

where d is the *spacing* of the grating. Therefore the amplitude at a point q is given by:

$$\begin{aligned}
\psi_P &\propto \sum_{n=0}^{N-1} e^{-iqnd} \\
&= \frac{1 - e^{-iqNd}}{1 - e^{-iqd}} \\
&\propto \frac{e^{iqNd/2} - e^{-iqNd/2}}{e^{iqd/2} - e^{-iqd/2}} \\
&\propto \frac{\sin(qNd/2)}{\sin(qd/2)}
\end{aligned}$$

This is an important result; the square of this gives the intensity. We see that there are *primary maxima* wherever the denominator vanishes

$$\frac{qd}{2} = m\pi \Rightarrow q = \frac{2m\pi}{d} \qquad \text{(Primary maxima)}$$

and *subsidiary maxima* wherever the numerator is ± 1

$$\frac{qNd}{2} = \left(M + \frac{1}{2}\right)\pi \Rightarrow q = \frac{2(M + 1/2)}{Nd} \qquad \text{(Subsidiary maxima)}$$

for integers m and M — the *orders* of the peaks.

Diffraction gratings can be used to separate light of different wavelengths, and thus are useful in spectroscopy. The first minimum around the central maximum is found for $qNd/2 = \pi$, i.e. $q = 2\pi/Nd$. This is also the distance from any primary maximum to the nearest minimum. Now q depends on the wavelength of the light being used; consider trying to separate two wavelengths of light: λ and $\lambda + \delta\lambda$. The first minimum of λ around the primary maximum of order m occurs at $q = 2\pi\theta/\lambda = 2m\pi/d + 2\pi/Nd$, so the angle is:

$$\theta = \frac{m\lambda}{d} + \frac{\lambda}{Nd}$$

If this lies on top of the mth order *maximum* of $\lambda + \delta\lambda$, then according to the *Rayleigh criterion* the two wavelengths count as resolved. This means that $q' = 2\pi\theta/(\lambda + \delta\lambda) = 2m\pi/d$, so

$$\theta = \frac{m(\lambda + \delta\lambda)}{d}$$

Thus the "chromatic resolving power" of the grating (given by $\lambda/\delta\lambda$) is found by setting these two expressions equal:

$$\begin{aligned} \frac{m(\lambda + \delta\lambda)}{d} &= \frac{m\lambda}{d} + \frac{\lambda}{Nd} \\ \frac{m\delta\lambda}{d} &= \frac{\lambda}{Nd} \\ \frac{\lambda}{\delta\lambda} &= mN \end{aligned}$$

so more slits (and higher orders) enable greater chromatic resolution.

2.2 2D Apertures

Often the aperture function $h(x, y)$ is separable in x and y, in which case the Fraunhofer diffraction pattern in 2 dimensions only takes double the work in these cases, e.g. rectangular apertures.

For a circular aperture, this is not the case, though we can simplify things as all the dimensions are equivalent, i.e. the amplitude is symmetric in p and q, so we can express the amplitude ψ_P solely in terms of, say, θ. Unfortunately, though, turns out the solution for ψ_P is only expressible as a Bessel function of the 1st kind:

$$\psi_P \propto \frac{J_1\left(\frac{\pi D\theta}{\lambda}\right)}{\frac{\pi D\theta}{\lambda}}$$

where D is the diameter of the aperture. The region of this function inside the first zero ($J_1(3.8317) \approx 0$) is known as the *Airy disc*; the angular radius of the disc is therefore $3.8317\lambda/\pi D \approx 1.2197\lambda/D$. One might use a lens in order to simulate a circular aperture; if the screen is at the focal point of the lens, the radius of the disc is therefore $1.22\lambda f/D$. If one has an instrument designed to create such images (e.g. a telescope) using a circular aperture, this is the limit to the angular resolution of the device. There may be other factors like atmospheric disturbances making it even worse, but if diffraction is the only thing preventing further precision, the device is said to be "diffraction limited".

3 Fresnel Diffraction

Suppose we are no longer at a distance where $L \gg l^2/\lambda$. This is the domain of Fresnel diffraction, of order l^2/λ or smaller away from the aperture. We begin the analysis by supposing that the (point) source is a distance a behind the aperture, and considering the amplitude at the origin, so that $r_1 = \sqrt{a^2 + x^2 + y^2}$ and $r_2 = \sqrt{L^2 + x^2 + y^2}$. Thus the original formula 1 becomes:

$$\psi_P(0,0) \propto \iint_\Sigma h(x,y) \frac{e^{ik\sqrt{a^2+x^2+y^2}}}{\sqrt{a^2+x^2+y^2}} \frac{e^{ik\sqrt{L^2+x^2+y^2}}}{\sqrt{L^2+x^2+y^2}} \,\mathrm{d}x\,\mathrm{d}y \tag{5}$$

This requires simplification. The denominator can be taken as roughly constant (as before) (for now), but variations in x and y are not incomparable with λ, which k depends on. We therefore approximate $r_1 + r_2$ as:

$$\begin{aligned} r_1 + r_2 &= \sqrt{a^2 + x^2 + y^2} + \sqrt{L^2 + x^2 + y^2} \\ &\approx a + \frac{x^2+y^2}{2a} + L + \frac{x^2+y^2}{2L} + \mathcal{O}(x^4) \\ &= a + L + \frac{x^2+y^2}{2\mathsf{R}} \end{aligned}$$

where $\mathsf{R}^{-1} = a^{-1} + L^{-1}$ (we are very much running out of letters here!). Thus the amplitude can then be written as:

$$\psi_P(0,0) \propto \iint_\Sigma h(x,y) \exp\left(ik\frac{x^2+y^2}{2\mathsf{R}}\right) \mathrm{d}x\,\mathrm{d}y \tag{6}$$

which funnily enough was the term we threw away in Fraunhofer, but with R instead of R. Changing variables to simplify things:

$$u = x\sqrt{\frac{k}{\pi\mathsf{R}}} \qquad\qquad v = y\sqrt{\frac{k}{\pi\mathsf{R}}}$$

$$\Rightarrow \psi_P(0,0) \propto \iint_\Sigma h(u,v) \exp\left(i\frac{\pi u^2}{2}\right) \exp\left(i\frac{\pi v^2}{2}\right) \mathrm{d}x\,\mathrm{d}y \tag{7}$$

This is a *hard* integral. We will for simplicity only include problems where the double integral reduces to a single (either by being constant in one dimension, or by separation), and where $h = 1$ or 0. We then have:

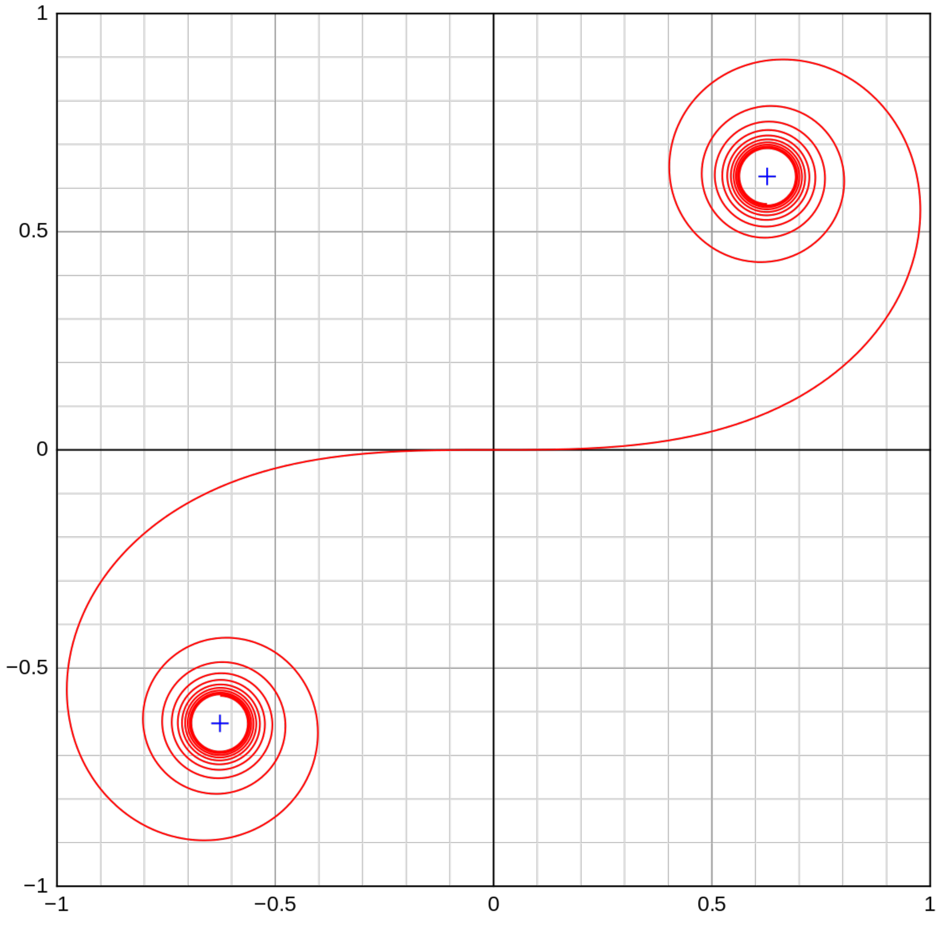
$$\Rightarrow \psi_P(0,0) \propto \int_{w_1}^{w_2} \exp\left(i\frac{\pi u^2}{2}\right) \mathrm{d}u \tag{8}$$

where $[w_1, w_2]$ is the open region of the aperture.

Defining the *Fresnel Integrals* as:

$$C(w) = \int_0^w \cos\left(\frac{\pi u^2}{2}\right) \mathrm{d}u \qquad S(w) = \int_0^w \sin\left(\frac{\pi u^2}{2}\right) \mathrm{d}u$$

We can then evaluate (8) if we can evaluate the Fresnel integrals (there'll have to be some subtractions and stuff but that's not too bad). Unfortunately these integrands have no antiderivative, so they must be evaluated numerically. A neat way of doing this is to make use of the *Cornu Spiral*, the locus of all points $\big(C(w), S(w)\big)$ in the complex plane:



The Cornu Spiral is interesting for many reasons, among which that on travelling along the curve from the origin for a length w, one arrives at exactly the point $\big(C(w), S(w)\big)$; the spiral is often labelled with values of w along it. Integrals like (8) can therefore be evaluated by converting it to one or more

Fresnel Integrals and then evaluating these using the Cornu Spiral.

The above enables one to evaluate the amplitudes *at* $(x_0, y_0) = (0, 0)$; what if we want to know the intensity at some other point? Well I'm not sure how I feel about this, but we *change the origin* and then apparently everything else is still alright. The axis is chosen so that the source and the point of interest are on it, and the *bounds of the aperture* are adjusted accordingly.

3.1 Circular Aperture

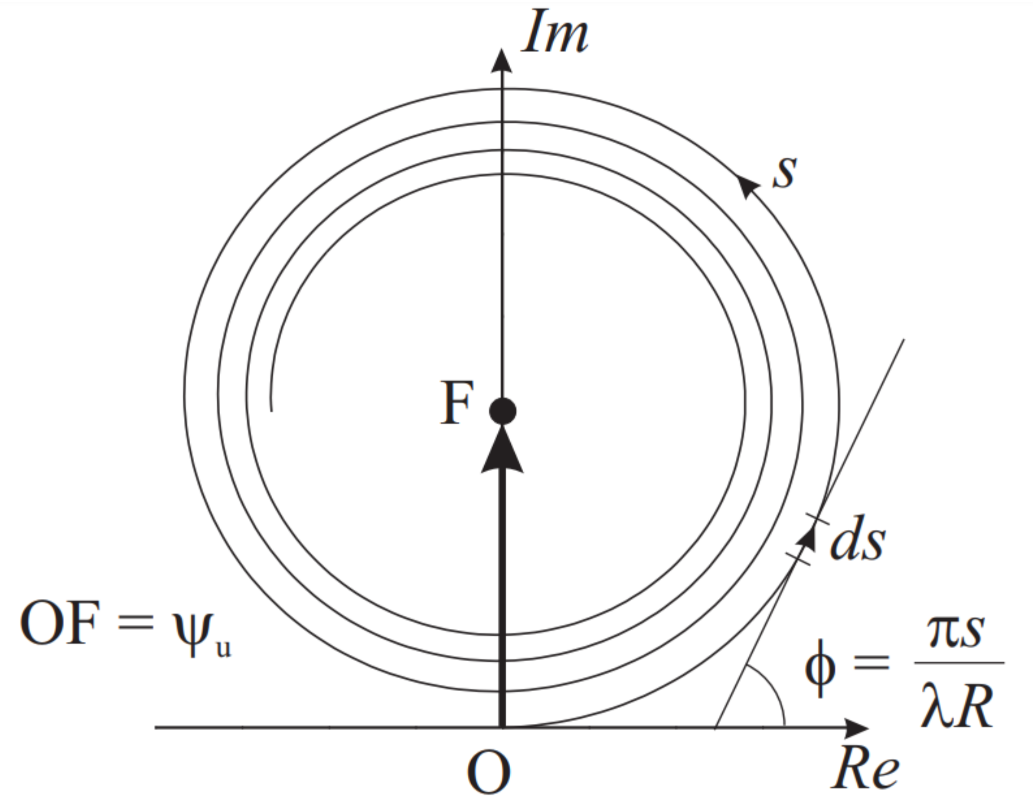
Suppose $h(x, y) = 1$ for $x^2 + y^2 \leqslant S^2$, i.e. a circular aperture of radius S. We can then convert equation 5 to:

$$\begin{aligned}\psi_P(0,0) &\propto \int_{\rho=0}^{\rho=S} \frac{e^{ik\sqrt{a^2+\rho^2}}}{\sqrt{a^2+\rho^2}} \frac{e^{ik\sqrt{L^2+\rho^2}}}{\sqrt{L^2+\rho^2}} 2\pi\rho \,\mathrm{d}\rho \\ &\propto \int_0^S \frac{1}{\sqrt{L^2+\rho^2}} \exp\left(ik\frac{\rho^2}{2\mathsf{R}}\right) 2\rho \,\mathrm{d}\rho\end{aligned}$$

Then, changing variables to $s = \rho^2$, we obtain:

$$\psi_P(0,0) \propto \int_{s=0}^{s=S^2} \frac{1}{\sqrt{L^2+s}} \exp\left(ik\frac{s}{2\mathsf{R}}\right) \mathrm{d}s$$

This integral is not evaluated, but conceptualised. The pre-exponential factor is a slowly-decreasing function of s, and if we were to get far enough from the origin that the obliquity factor K gets involved, that would also be a slowly-decreasing function of s. The exponential itself can be thought of in a phasor sense as an arrow, whose orientation linearly increases with s; just how much it changes by depends on the size of s relative to k and R. So what the integral does is add together a series of infinitesimal arrows, each in a more anticlockwise direction, and slightly shorter, than the last. This leads to the visualisation of the integral on the following page:



The distance from O (the origin of the Argand Diagram) to the point on the spiral corresponding to $s = S^2$ is $\psi_P(0,0)$, so as S is increased, $\psi_P(0,0)$ initially increases, reaches a maximum when the argument ϕ of the exponential reaches π, and then a minimum when $\phi = 2\pi$. The central circular region of the *aperture* for which $\phi < \pi$ (i.e. the first semicircle of the phasor diagram) is termed the "1st Fresnel half-period zone". The radius of the 1st half-period zone is then found:

$$\begin{aligned} k\frac{\rho_1^2}{2\mathsf{R}} &= \pi \\ \Rightarrow \rho_1 &= \sqrt{\frac{2\pi\mathsf{R}}{k}} \\ &= \sqrt{\lambda\mathsf{R}} \end{aligned}$$

The 2nd half-period zone is the *annulus* on the aperture for which ϕ is in between π and 2π. It has an inner radius of $\sqrt{\lambda\mathsf{R}}$ and an outer radius easily shown to be $\rho_2 = \sqrt{2\lambda\mathsf{R}}$. Similarly, the nth half-period zone has an inner radius of $\sqrt{(n-1)\lambda\mathsf{R}}$ and an outer radius of $\sqrt{n\lambda\mathsf{R}}$. Interestingly, the area of each half-period zone is $\pi(\rho_n^2 - \rho_{n-1}^2) = \pi\lambda\mathsf{R}$, the same for all n.

The line OF on the imaginary axis corresponds to the limit $s \to \infty$, i.e. $S \to \infty$, i.e. an aperture with an infinite radius, i.e. no obstacles at all. Interestingly, if one introduces a small circular obstruction on the axis, then the integral is instead goes from a certain non-zero point up to ∞. This corresponds on the phasor diagram to an amplitude arrow drawn from some point on the outer circumference, to F (wherein the spiral goes at infinity). Provided the object isn't too big (which would lead to starting the arrow on inner spirals), the magnitude of the resulting amplitudes is roughly the same as OF, i.e. the same intensity as if there were no spot present. This leads to a bright spot appearing at $(0,0)$, known as Poisson's (or Arago's, or Fresnel's) Spot; this is a special case of *Babinet's Theorem*, which isn't too difficult to show.

To consider the behaviour of a circular aperture off-axis (i.e. $\psi(x,y)$ for $x,y \neq 0$), as before we adjust the origin so that the axis is between the source and (x,y). We must then effectively move the aperture over the Fresnel zones, adding areas of the odd half-zones and subtracting those of the even half-zones.

A Fresnel zone plate is designed to block out every other half-zone (often blocking the 1st, 3rd... and admitting the 2nd, 4th...), so the resultant amplitude is found from adding semicircles end-to-end in the Argand diagram. As the size of the Fresnel zones depend on R, and thus L, a zone plate only works at a certain distance (L) away from it – not unlike a lens with a certain focal length $f = L$. Because the incoming wave is assumed to be a plane wave, a is assumed large, and so $L \approx \mathsf{R}$, so $f \approx \mathsf{R} = \rho_1^2/\lambda$. This focal length is therefore highly wavelength-dependent, i.e. much chromatic aberration will be present.

On moving the observation point closer to the axis (i.e. reducing L, and thus R), the size of the Fresnel half-zones changes, and when L (and ess. R) are reduced by a factor of 2, each open area of the half-plate admits *two* half-zones, which cancel each other out and the overall amplitude and intensity goes to 0. Similar things occur for $L' = L/4, L/6$ etc. However, when $L' = L/3$ etc., each open area of the half-plate admits an odd number of half-zones, and there are subsidiary maxima. These maxima gradually decrease in size, however, due to the reductions caused by K.

4 Interferometry

Consider Ψ, a superposition of two general waves $\Re\psi_1 e^{-i\omega_1 t} + \Re\psi_2 e^{-i\omega_2 t} = \Re(\psi_1 e^{-i\omega_1 t} + \psi_2 e^{-i\omega_2 t}) = \Re\hat{\Psi}$. The intensity of Ψ is given by its square; since $\Re(A)\Re(B) = \frac{1}{2}\Re(AB + AB^*)$, we have:

$$\begin{aligned}
I = \Psi^2 =& \frac{1}{2}\Re(\hat{\Psi}^2 + \hat{\Psi}\hat{\Psi}^*) \\
=& \frac{1}{2}\Re(\hat{\Psi}^2) + \frac{1}{2}|\hat{\Psi}|^2 \\
=& \frac{1}{2}\Re(\psi_1^2 e^{-2i\omega_1 t} + 2\psi_1\psi_2 e^{-i(\omega_1+\omega_2)t} + \psi_2^2 e^{-2i\omega_2 t}) \\
&+ \frac{1}{2}(|\psi_1|^2 + 2\Re\psi_1^*\psi_2 e^{-i(\omega_2-\omega_1)t} + |\psi_2|^2)
\end{aligned}$$

Now most detectors have a response time of such a length (usu. above 1ns) that variation of frequency ω_1 or ω_2 will simply average to 0, whereas that of frequency $\omega_2 - \omega_1$ might not – we might be dealing with 700nm light interfering with 697nm light, for instance. The intensity detected by the detector, $\langle I \rangle$, is then:

$$\begin{aligned}
\langle I \rangle &= \frac{1}{2}|\psi_1|^2 + \Re\psi_1^*\psi_2\langle e^{-i(\omega_2-\omega_1)t}\rangle + \frac{1}{2}|\psi_2|^2 \\
&= \frac{1}{2}|\psi_1|^2 + \frac{1}{2}|\psi_2|^2 + \underbrace{|\psi_1||\psi_2|\Re\langle e^{i(\phi_2-\phi_1-(\omega_2-\omega_1)t)}\rangle}_{\text{interference term}}
\end{aligned}$$

where $\phi_2 - \phi_1$ is the relative phase of the two waves, and the time-averaging is done over the integration time of the detector (over all time, the interference term will also go to 0). The first two terms here are simply the intensities of the two individual waves; the interference term is what will be important in this section.

There are two parts to this term: the relative phase, and the relative frequency; the former is most important. The phases of independent sources (e.g. emissions from two different atoms) typically vary in time randomly, so interference is almost only seen when the two waves originate from the same source – often by splitting the light from a single source, sending them through different paths, and recombining them later. An advantage of this is that the frequencies will also be the same, so we will not have to worry about the relative frequency term. Even with the subtraction, the frequencies have to be *very* close together to be able to detect any time variation.

4.1 Michelson Interferometer and Fourier Transform Spectroscopy

The Michelson Interferometer consists of a beam splitter and two mirrors. The point light source is split in two by the beam splitter, and the two waves propagate at 90°. They reflect off of their relative mirrors and return to the beam splitter, where half of each simply returns back to the source, and the other half interfere; the resulting intensity is recorded by a detector. Depending on the difference between the two path lengths, $\phi_2 - \phi_1$ can create constructive or destructive interference; one mirror is kept fixed (to an accuracy of $\ll \lambda$ by necessity) and the other moved relative to it.

We have $|\psi_i|^2 = I_0$, $\omega_1 = \omega_2$ and $\phi_2 - \phi_1 = kx$ where x is the path difference – note that this is double the difference in the distances between the mirrors and the splitter.:

$$\begin{aligned}\langle I\rangle(x) &= \frac{1}{2}I_0 + \frac{1}{2}I_0 + \sqrt{I_0}\sqrt{I_0}\Re e^{ikx} \\ &= I_0(1 + \Re e^{ikx})\end{aligned}$$

so if x is varied linearly in time, the output of the interferometer's detector will be $I_0(1 + \cos kvt)$ – fringes.

If two wavelengths are present ($k_0 \pm \Delta k$), there is still no interference observed between the two wavelengths (their frequencies are likely still too different). The output is therefore simply the sum of the two fringe patterns, and a beating pattern is seen as a function of x, varying as $1 + \cos(k_0 x)\cos(\Delta k x)$ (as can be seen by adding these two waves together, assuming they have the same amplitude). One then defines the *fringe contrast* ν as:

$$\nu = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}}$$

At a point where the modulating sinusoid is 0, the fringes (due to the high-frequency sinusoid) are very dim – the fringe contrast is low; where the modulating sinusoid is 1, the fringes are bright and easily distinguished.

If a broadband light is used, with an intensity $2S(k)\,\mathrm{d}k$ between the wavenumbers k and $k + \mathrm{d}k$ and with not much overlap, we then have:

$$\langle I\rangle(x) = \int_0^\infty 2S(k)(1 + \Re e^{ikx})\,\mathrm{d}k$$

$$
\begin{aligned}
&= 2\Re \int_0^\infty S(k)(1+e^{ikx})\,\mathrm{d}k \\
&= \int_0^\infty S(k)(1+e^{ikx})\,\mathrm{d}k + \left(\int_0^\infty S(k)(1+e^{ikx})\,\mathrm{d}k\right)^* \\
&= \int_0^\infty S(k)(1+e^{ikx})\,\mathrm{d}k + \int_0^\infty S(k)^*(1+e^{-ikx})\,\mathrm{d}k
\end{aligned}
$$

We then define $S(k)$ for negative k for convenience: $S(-k) = S(k)^*$. Then,

$$
\begin{aligned}
\langle I\rangle(x) &= \int_0^\infty S(k)(1+e^{ikx})\,\mathrm{d}k + \int_0^\infty S(-k)(1+e^{-ikx})\,\mathrm{d}k \\
&= \int_0^\infty S(k)(1+e^{ikx})\,\mathrm{d}k + \int_{-\infty}^0 S(k)(1+e^{ikx})\,\mathrm{d}k \\
&= \int_{-\infty}^\infty S(k)(1+e^{ikx}) \\
&= I_{\text{tot}} + \mathscr{F}^{-1}\left[S(k)\right]
\end{aligned}
$$

$$
\Rightarrow S(k) = \mathscr{F}\left[\langle I\rangle(x) - I_{\text{tot}}\right]
$$

where I_{tot} is the total intensity of the light (integrated across all wavelengths). However, because these "Fourier Transform Spectrometers" have only a finite length and the Fourier Transform has infinite bounds, the measured spectrum $\mathsf{S}(k)$ cannot be equal to $S(k)$. This can be thought of as being due to the infinitely-defined $\langle I\rangle(x)$ being what one *would* measure if one had an infinitely long spectrometer, and the measured intensity function $\langle \mathsf{I}\rangle(x)$ being equal to a $\langle I\rangle$ multiplied by a Π function. For a spectrometer whose mirror can be moved over a total range $X/2$, the range in x is in fact X; supposing that the centre of the mirror's range (i.e. $X/4$ from either end) corresponds to a path difference of 0, x is therefore measured between $-X/2$ and $X/2$, so the Π function is $\Pi(x;X)$. We then have:

$$
\begin{aligned}
\mathsf{S}(k) &= \mathscr{F}\left[\left(\langle I\rangle(x) - I_{\text{tot}}\right)\Pi(x;X)\right] \\
&\propto \mathscr{F}\left[\langle I\rangle(x) - I_{\text{tot}}\right] * \mathscr{F}\left[\Pi(x;X)\right] \\
&= S(k) * \operatorname{sinc}\left(\frac{kX}{2}\right)
\end{aligned}
$$

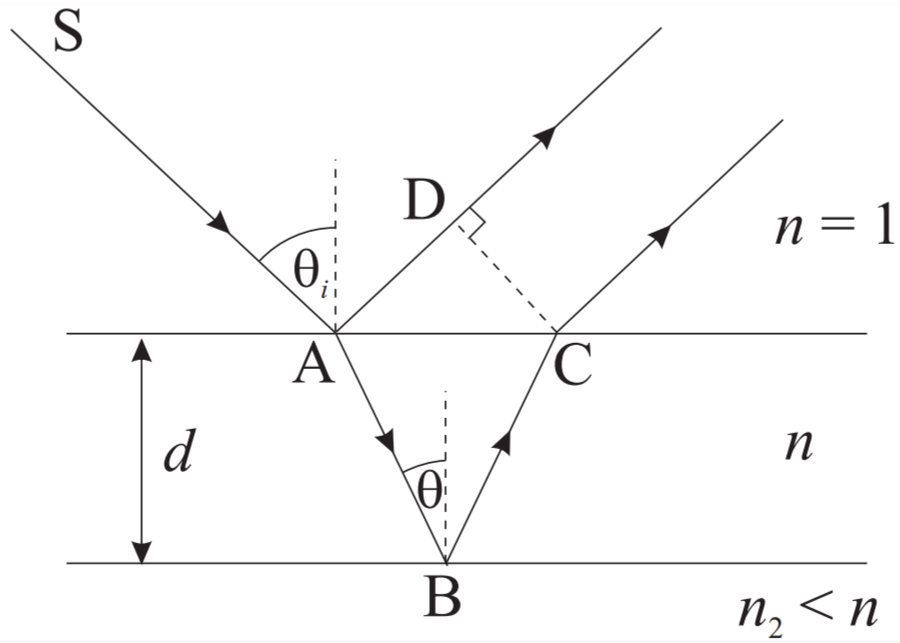
i.e. the true frequency spectrum is smeared out. Returning to the case where two wavelengths were present at $k_0 \pm \Delta k$, we see that for the two delta functions in $S(k)$ to still be resolvable after the smearing, the sinc at

$k_0+\Delta k$ must lie on the minimum of the sinc at $k_0-\Delta k$. Thus $2\Delta k = 2\pi/X$. To derive the spectral resolution $\Delta\lambda/\lambda_0$, we note that $k\lambda = 2\pi$, which is constant, and so $k\,\mathrm{d}\lambda + \lambda\,\mathrm{d}k = 0$, so $|\Delta\lambda|/\lambda_0 = \mathrm{d}k/k_0$. The difference in wavenumber between $k \pm \Delta k$ is $\mathrm{d}k = 2\Delta k$, so we have:

$$\begin{aligned}\frac{|\Delta\lambda|}{\lambda_0} &= \frac{2\Delta k}{k_0} = \frac{2\pi}{X}\frac{\lambda_0}{2\pi} \\ &= \frac{\lambda_0}{X}\end{aligned}$$

which can be compared with the resolution of a diffraction grating, $1/mN$. FTS therefore usually has higher (spectral) resolution than diffraction gratings, but requires many measurements rather than just a single image.

4.2 Thin Film Interference



The path difference between the interfering waves must be calculated taking into account the refractive index n of the thin film; n reduces c, so reduces λ, so increases k. Additionally the reflection at the lower surface introduces a π phase shift (as the impedance of the air is smaller than the impedance of the film). The total phase difference is therefore:

$$\begin{aligned}\phi_2 - \phi_1 &= k(n\ AB + n\ BC - AD) + \pi \\ &= k\left(\frac{2nd}{\cos\theta} - 2d\tan\theta\sin\theta_i\right) + \pi\end{aligned}$$

Using Snell's Law,

$$\begin{aligned}
&= k\left(\frac{2nd}{\cos\theta} - 2d\tan\theta n\sin\theta\right) + \pi \\
&= k\frac{2nd}{\cos\theta}(1-\sin^2\theta) + \pi \\
&= 2nkd\cos\theta + \pi
\end{aligned}$$

The intensity is therefore obtained as:

$$\langle I\rangle = I_0\left(1 - \Re e^{2inkd\cos\theta}\right)$$

or equivalently expressed in terms of λ. This assumes that the amplitudes of the two beams are equal, but this can be shown to be the case only when the impedances (and thus ns) of the two media are the same! Otherwise the fringes occur at the same frequency, but with an offset and smaller amplitude, giving a lower fringe contrast.

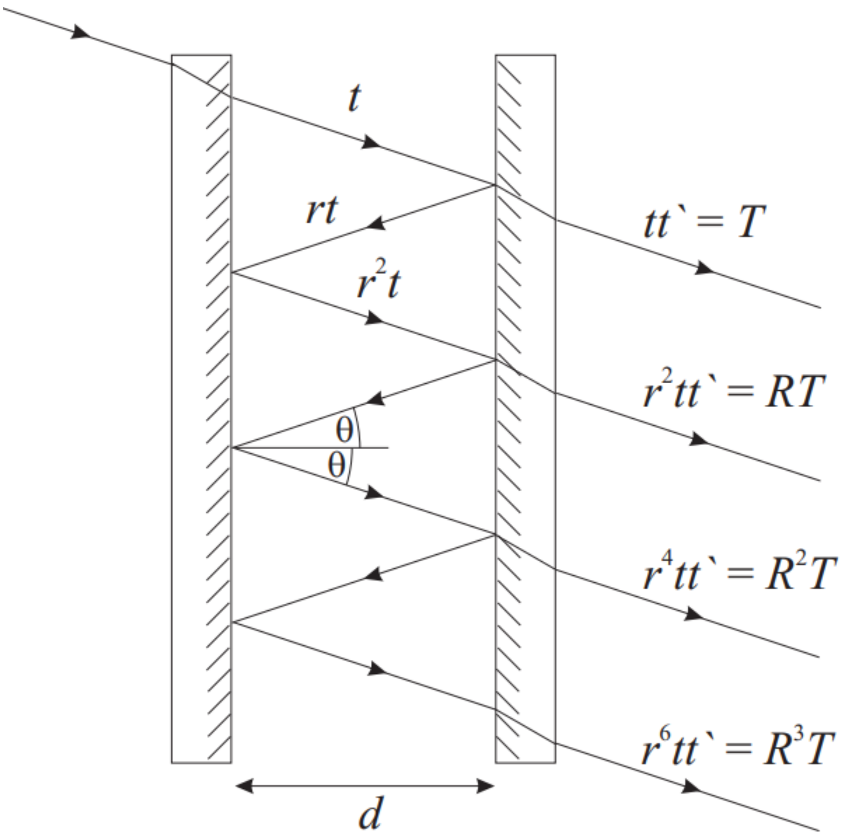
Maxima occur for $2nkd\cos\theta = (2m+1)\pi$, i.e.:

$$nd\cos\theta = \frac{2m+1}{4}\lambda$$

This can be viewed as either a condition on the θs for which the maxima occur (*fringes of equal inclination*), or a condition of the d (fringes of equal *thickness*) for a film which varies in thickness.

4.3 Fabry-Pérot Etalon

The above analysis ignored the possibility of repeated reflections in the film, which would cause a third wave to join, and a 4th etc. This is because for most films (e.g. bubbles) the reflection coefficient is as low as 0.04, but the Fabry-Pérot etalon makes use of higher reflection coefficients of around 0.97, consisting of essentially some air sandwiched between two mirrors. Assuming the reflection and transmission coefficients r and t are real, it can be seen on the diagram below that each interfering wave has an amplitude of r^2 and a phase of $2kd\cos\theta$ both relative to the previous wave (this is essentially the TFI phase with $n = 1$, and with an extra π phase shift because of the additional reflection). We note here that the FPE is usually used at essentially normal incidence, so the successive wave is r^2e^{2ikd} relative to the previous.



Taking into account the transmission coefficients, the total output amplitude is then:

$$\begin{aligned} A &= \sum_{n=1}^{\infty} r^{2n} e^{2inkd} tt' \\ &= \sum_{n=1}^{\infty} R^n e^{in\delta} T \\ &= \frac{T}{1 - Re^{i\delta}} \end{aligned}$$

where $R = r^2$, $T = tt'$, and $\delta = 2kd$. The intensity is then given by:

$$\begin{aligned} I = |A|^2 &= \frac{T^2}{(1 - R\cos\delta)^2 + R^2 \sin^2\delta} \\ &= \frac{T^2}{1 - 2R\cos\delta + R^2} \end{aligned}$$

which also gives a fringe pattern as a function of δ, that is, of d. The half-width of the peaks $\delta_{1/2}$ is the (typically small) change in δ required to reduce the intensity to half its maximum value (the max value is achieved for $\delta = 2\pi m$), which is $T^2/(1-R)^2$. We therefore have:

$$\begin{aligned}
1 - 2R\cos\delta_{1/2} + R^2 &= 2(1-R)^2) \\
1 - 2R + R\delta_{1/2}^2 + R^2 &= 2(1-R)^2 \\
\delta_{1/2}^2 &= \frac{(1-R)^2}{R} \\
\delta_{1/2} &= \frac{1-R}{\sqrt{R}}
\end{aligned}$$

The *finesse* $\mathcal{F}$ of an FPE is the width between successive peaks in δ (equal to 2π) divided by the FWHM of the peaks in δ (equal to $2\delta_{1/2}$). Thus we have:

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

$\mathcal{F}$ becomes very high as $R \to 1$; for $r = 0.97$, $\mathcal{F} = 52$.

If two wavelengths are present, there will be two sets of fringes. The chromatic resolving power of the etalon, $\lambda/\Delta\lambda$ is derived as follows:

$$\begin{aligned}
\delta = 2kd \Rightarrow \mathrm{d}\delta = 2d\,\mathrm{d}k &= \delta\frac{\mathrm{d}k}{k} \\
2\delta_{1/2} &= \delta\frac{\Delta\lambda}{\lambda} = 2m\pi\frac{\Delta\lambda}{\lambda} \\
\frac{\lambda}{\Delta\lambda} &= m\frac{2\pi}{2\delta_{1/2}} \\
&= m\mathcal{F}
\end{aligned}$$

which can be very high, as the FPE is often used at high m

A final quantity of interest is the *free spectral range*: the range of wavelengths for which the mth order peak of the lowest wavelength overlaps with the m+1th of the highest wavelength. We have $2d = m\lambda$, so $0 = m\,\mathrm{d}\lambda + \lambda\,\mathrm{d}m$, where m is large so taking a differential is justified. $\mathrm{d}\lambda$ is the free spectral range and $\mathrm{d}m$ is 1; we therefore have:

$$\mathrm{FSR} = \frac{\lambda}{m}$$

– not very big. The FPE is therefore best at distinguishing very accurately between two very similar wavelengths.