12

Interference and Diffraction

All waves display the phenomena of interference and diffraction which arise from the superposition of more than one wave. At each point of observation within the interference or diffraction pattern the phase difference between any two component waves of the same frequency will depend on the different paths they have followed and the resulting amplitude may be greater or less than that of any single component. Although we speak of separate waves the waves contributing to the interference and diffraction pattern must ultimately derive from the same single source. This avoids random phase effects from separate sources and guarantees coherence. However, even a single source has a finite size and spatial coherence of the light from different parts of the source imposes certain restrictions if interference effects are to be observed. This is discussed in the section on spatial coherence on p. 360. The superposition of waves involves the addition of two or more harmonic components with different phases and the basis of our approach is that laid down in the vector addition of Figure 1.11. More formally in the case of diffraction we have shown the equivalence of the Fourier transform method on p. 287 of Chapter 10.

Interference

Interference effects may be classified in two ways:

1. Division of amplitude
2. Division of wavefront

1. *Division of amplitude.* Here a beam of light or ray is reflected and transmitted at a boundary between media of different refractive indices. The incident, reflected and transmitted components form separate waves and follow different optical paths. They interfere when they are recombined.

2. *Division of wavefront.* Here the wavefront from a single source passes simultaneously through two or more apertures each of which contributes a wave at the point of superposition. Diffraction also occurs at each aperture.

The Physics of Vibrations and Waves, 6th Edition H. J. Pain

The difference between interference and diffraction is merely one of scale: in optical diffraction from a narrow slit (or source) the aperture is of the order of the wavelength of the diffracted light. According to Huygens Principle every point on the wavefront in the plane of the slit may be considered as a source of secondary wavelets and the further development of the diffracted wave system may be obtained by superposing these wavelets.

In the interference pattern arising from two or more such narrow slits each slit may be seen as the source of a single wave so the number of superposed components in the final interference pattern equals the number of slits (or sources). This suggests that the complete pattern for more than one slit will display both interference and diffraction effects and we shall see that this is indeed the case.

Division of Amplitude

First of all we consider interference effects produced by division of amplitude. In Figure 12.1 a ray of monochromatic light of wavelength λ in air is incident at an angle i on a plane parallel slab of material thickness t and refractive index $n > 1$. It suffers partial reflection and transmission at the upper surface, some of the transmitted light is reflected at the lower surface and emerges parallel to the first reflection with a phase difference determined by the extra optical path it has travelled in the material. These parallel beams meet and interfere at infinity but they may be brought to focus by a lens. Their optical path difference is seen to be

$$\begin{aligned} n(\mathrm{AB} + \mathrm{BD}) - \mathrm{AC} &= 2n\mathrm{AB} - \mathrm{AC} \\ &= 2nt/\cos\theta - 2t\tan\theta\sin i \\ &= \frac{2nt}{\cos\theta}(1 - \sin^2\theta) = 2nt\cos\theta \end{aligned}$$

(because $\sin i = n\sin\theta$).

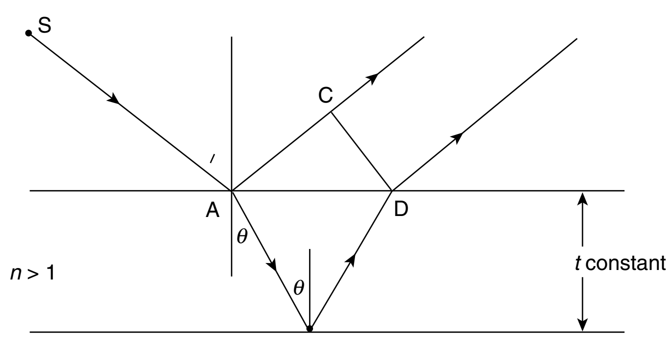


Figure 12.1 Fringes of constant inclination. Interference fringes formed at infinity by division of amplitude when the material thickness t is constant. The mth order bright fringe is a circle centred at S and occurs for the constant θ value in $2nt\cos\theta = (m + \frac{1}{2})\lambda$

This path difference introduces a phase difference

$$\delta = \frac{2\pi}{\lambda} 2nt\cos\theta$$

but an additional phase change of π rad occurs at the upper surface.

The phase difference δ between the two interfering beams is achieved by writing the beam amplitudes as

$$y_1 = a(\sin\omega t + \delta/2) \quad \text{and} \quad y_2 = a\sin(\omega t - \delta/2)$$

with a resultant amplitude

$$\begin{aligned} R &= a[\sin(\omega t + \delta/2) + \sin(\omega t - \delta/2) \\ &= 2a\sin\omega t\cos\delta/2 \end{aligned}$$

and an intensity

$$I = R^2 = 4a^2\sin^2\omega t\cos^2\delta/2$$

Figure 12.2 shows the familiar $\cos^2\delta/2$ intensity fringe pattern for the spatial part of I.

Thus, if $2nt\cos\theta = m\lambda$ (m an integer) the two beams are anti-phase and cancel to give zero intensity, a minimum of interference. If $2nt\cos\theta = (m + \frac{1}{2})\lambda$ the amplitudes will reinforce to give an interference maximum.

Since t is constant the locus of each interference fringe is determined by a constant value of θ which depends on a constant angle i. This gives a circular fringe centred on S. An extended source produces a range of constant θ values at one viewing position so the complete pattern is obviously a set of concentric circular fringes centred on S and formed at infinity. They are fringes of *equal inclination* and are called Haidinger fringes. They are observed to high orders of interference, that is values of m, so that t may be relatively large.

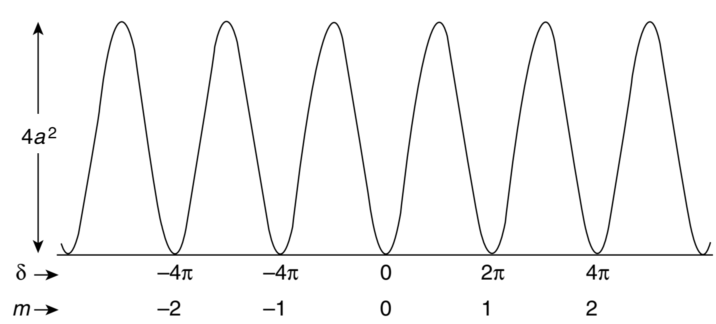


Figure 12.2 Interference fringes of $\cos^2$ intensity produced by the division of amplitude in Figure 12.1. The phase difference $\delta = 2\pi nt\cos\theta/\lambda$ and m is the order of interference

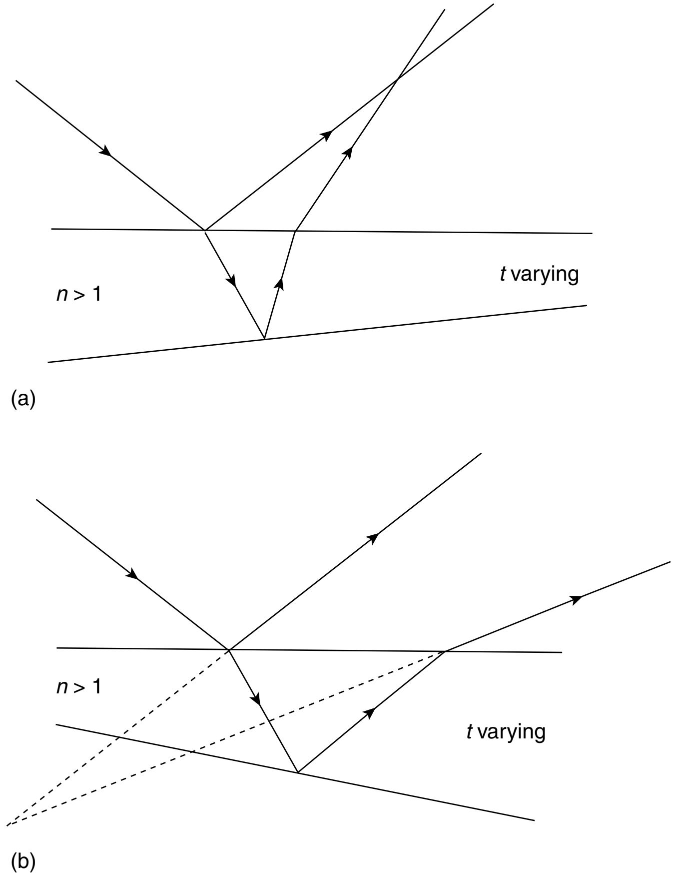


Figure 12.3 Fringes of constant thickness. When the thickness t of the material is not constant the fringes are localized where the interfering beams meet (a) in a real position and (b) in a virtual position. These fringes are almost parallel to the line where $t = 0$ and each fringe defines a locus of constant t

When the thickness t is not constant and the faces of the slab form a wedge, Figure 12.3a and b the interfering rays are not parallel but meet at points (real or virtual) near the wedge. The resulting interference fringes are localized near the wedge and are almost parallel to the thin end of the wedge. When observations are made at or near the normal to the wedge $\cos\theta \sim 1$ and changes slowly in this region so that $2nt\cos\theta \approx 2nt$. The condition for bright fringes then becomes

$$2nt = (m + \tfrac{1}{2})\lambda$$

and *each fringe locates a particular value of the thickness t of the wedge* and this defines the patterns as *fringes of equal thickness*. As the value of m increases to $m + 1$ the thickness of the wedge increases by $\lambda/2n$ so the fringes allow measurements to be made to within a fraction of a wavelength and are of great practical importance.

The spectral colours of a thin film of oil floating on water are fringes of constant thickness. Each frequency component of white light produces an interference fringe at that film thickness appropriate to its own particular wavelength.

In the laboratory the most familiar fringes of constant thickness are Newton's Rings.

Newton's Rings

Here the wedge of varying thickness is the air gap between two spherical surfaces of different curvature. A constant value of t yields a circular fringe and the pattern is one of concentric fringes alternately dark and bright. The simplest example, Figure 12.4, is a plano convex lens resting on a plane reflecting surface where the system is illuminated from above using a partially reflecting glass plate tilted at 45°. Each downward ray is partially reflected at each surface of the lens and at the plane surface. Interference takes

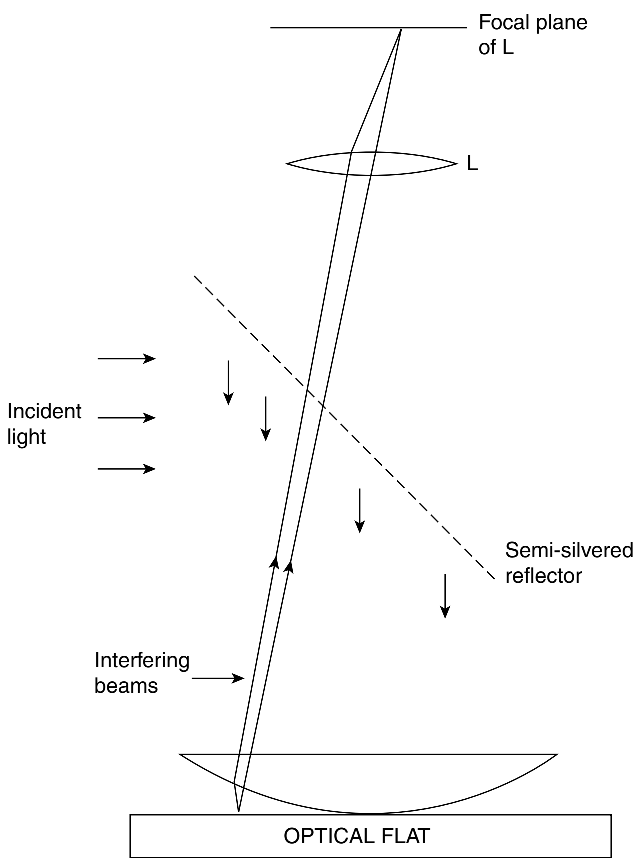


Figure 12.4 Newton's rings of interference formed by an air film of varying thickness between the lens and the optical flat. The fringes are circular, each fringe defining a constant value of the air film thickness

place between the light beams reflected at each surface of the air gap. At the lower (air to glass) surface of the gap there is a π rad phase change upon reflection and the centre of the interference fringe pattern, at the point of contact, is dark. Moving out from the centre, successive rings are light and dark as the air gap thickness increases in units of $\lambda/2$. If R is the radius of curvature of the spherical face of the lens, the thickness t of the air gap at a radius r from the centre is given approximately by $t \approx r^2/2R$. In the mth order of interference a bright ring requires

$$2t = (m + \tfrac{1}{2})\lambda = r^2/R$$

and because $t \propto r^2$ the fringes become more crowded with increasing r. Rings may be observed with very simple equipment and good quality apparatus can produce fringes for $m > 100$.

(Problem 12.1)

Michelson's Spectral Interferometer

This instrument can produce both types of interference fringes, that is, *circular fringes of equal inclination at infinity* and *localized fringes of equal thickness.* At the end of the nineteenth century it was one of the most important instruments for measuring the structure of spectral lines.

As shown in Figure 12.5 it consists of two identical plane parallel glass plates G_1 and G_2 and two highly reflecting plane mirrors M_1 and M_2. G_1 has a partially silvered back face, G_2 does not. In the figure G_1 and G_2 are parallel and M_1 and M_2 are perpendicular. Slow, accurately monitored motion of M_1 is allowed in the direction of the arrows but the mounting of M_2 is fixed although the angle of the mirror plane may be tilted so that M_1 and M_2 are no longer perpendicular.

The incident beam from an extended source divides at the back face of G_1. A part of it is reflected back through G_1 to M_1 where it is returned through G_1 into the eye or detector. The remainder of the incident beam reaches M_2 via G_2 and returns through G_2 to be reflected at the back face of G_1 into the eye or detector where it interferes with the beam from the M_1 arm of the interferometer. The presence of G_2 assures that each of the two interfering beams has the same optical path in glass. This condition is not essential for fringes with monochromatic light but it is required with a white light source where dispersion in glass becomes important.

An observer at the detector looking into G_1 will see M_1, a reflected image of M_2 (M_2', say) and the images S_1 and S_2' of the source provided by M_1 and M_2. This may be represented by the linear configuration of Figure 12.6 which shows how interference takes place and what type of firnges are produced.

When the optical paths in the interferometer arms are equal and M_1 and M_2 are perpendicular the planes of M_1 and the image M_2' are coincident. However a small optical path difference t between the arms becomes a difference of $2t$ between the mirrored images of the source as shown in Figure 12.6. The divided ray from a single point P on the extended source is reflected at M_1 and M_2 (shown as M_2') but these reflections appear to

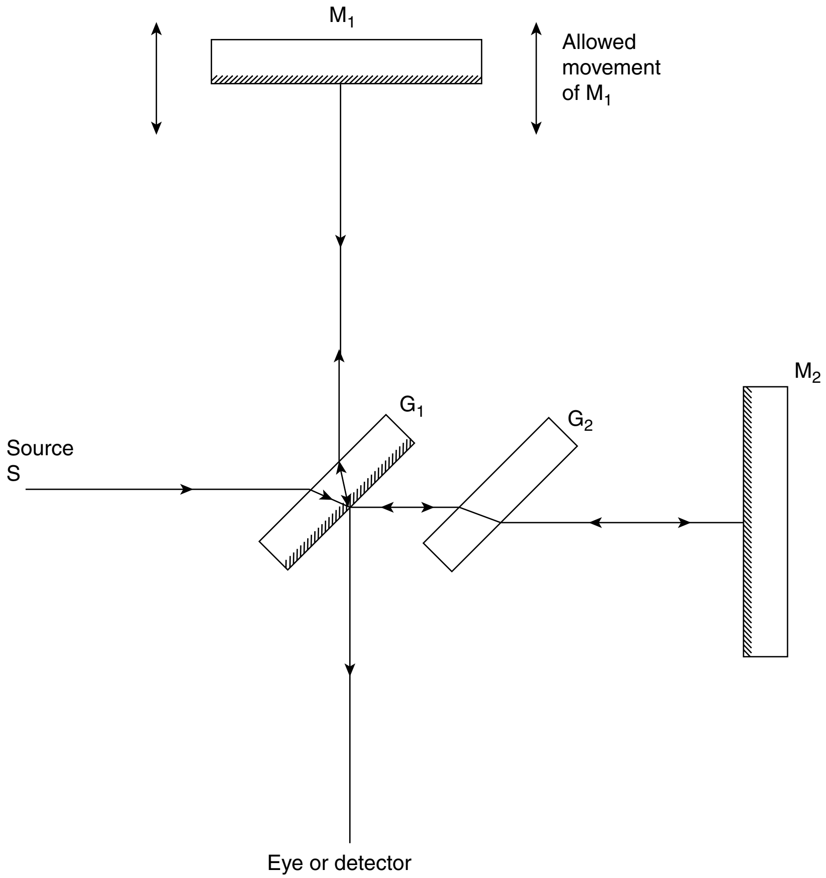


Figure 12.5 Michelson's Spectral Interferometer. The beam from source S splits at the back face of G_1, and the two parts are reflected at mirrors M_1 and M_2 to recombine and interfere at the eye or detector. G_2 is not necessary with monochromatic light but is required to produce fringes when S is a white light source

come from P_1 and P_2' in the image planes of the mirrors. The path difference between the rays from P_1 and P_2' is evidently $2t\cos\theta$. When $2t\cos\theta = m\lambda$ a maximum of interference occurs and for constant θ the interference fringe is a circle. The extended source produces a range of constant θ values and a pattern of concentric circular fringes of constant inclination.

If the path difference t is very small and the plane of M_2 is now tilted, a wedge is formed and straight localized fringes may be observed at the narrowest part of the wedge. As the wedge thickens the fringes begin to curve because the path difference becomes more strongly dependent upon the angle of observation. These curved fringes are always convex towards the thin end of the wedge.

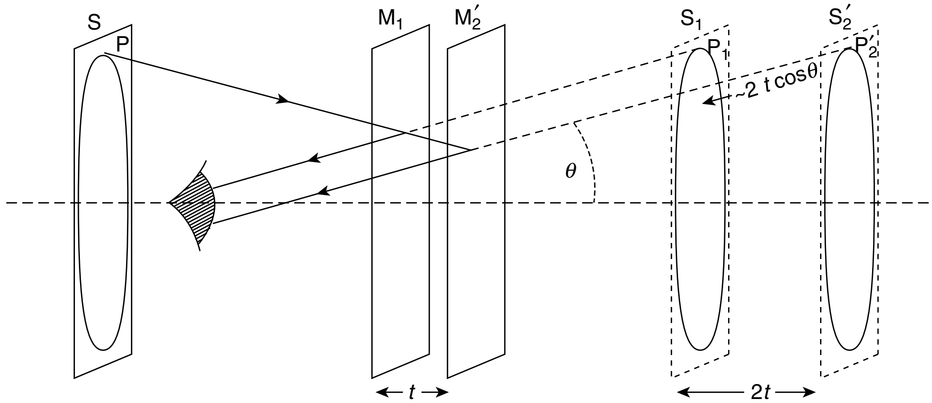


Figure 12.6 Linear configuration to show fringe formation by a Michelson interferometer. A ray from point P on the extended source S reflects at M_1, and appears to come from P_1 in the reflected plane S_1. The ray is reflected from M_2 (shown here as M_2') and appears to come from P_2' in the reflected plane S_2'. The path difference at the detector between the interfering beams is effectively $2t\cos\theta$ where t is the difference between the path lengths from the source S to the separate mirrors M_1 and M_2

The Structure of Spectral Lines

The discussion on spatial coherence, p. 362, will show that two close identical sources emitting the same wavelength λ produce interference fringe systems slightly displaced from each other (Figure 12.17).

The same effect is produced by a *single* source, such as sodium, emitting two wavelengths, λ and $\lambda - \Delta\lambda$ so that the maxima and minima of the $\cos^2$ fringes for λ are slightly displaced from those for $\lambda - \Delta\lambda$. This displacement increases with the order of interference m until a value of m is reached when the maximum for λ coincides with a minimum for $\lambda - \Delta\lambda$ and the fringes disappear as their visibility is reduced to zero.

In 1862, Fizeau, using a sodium source to produce Newton's Rings, found that the fringes disappeared at the order $m = 490$ but returned to maximum visibility at $m = 980$. He correctly identified the presence of two components in the spectral line.

The visibility

$$(I_{\max} - I_{\min})/(I_{\max} + I_{\min})$$

equals zero when

$$m\lambda = (m + \tfrac{1}{2})(\lambda - \Delta\lambda)$$

and for $\lambda = 0.5893\,\mu\text{m}$ and $m = 490$ we have $\Delta\lambda = 0.0006\,\mu\text{m}$ (6 Å), which are the accepted values for the D lines of the sodium doublet.

Using his spectral interferometer, Michelson extended this work between the years 1890 and 1900, plotting the visibility of various fringe systems and building a mechanical harmonic analyser into which he fed different component frequencies in an attempt to reproduce his visibility curves. The sodium doublet with angular frequency components ω and $\omega + \Delta\omega$ produced a visibility curve similar to that of Figures 1.7 and 4.4 and was easy to interpret. More complicated visibility patterns were not easy to reproduce and the modern method of Fourier transform spectroscopy reverses the procedure by extracting the frequency components from the observed pattern.

Michelson did however confirm that the cadmium red line, $\lambda = 0.6438\,\mu\text{m}$ was highly monochromatic. The visibility had still to reach a minimum when the path difference in his interferometer arms was 0.2 m.

Fabry – Perot Interferometer

The interference fringes produced by division of amplitude which we have discussed so far have been observed as reflected light and have been produced by only two interfering beams. We now consider fringes which are observed in transmission and which require multiple reflections. They are fringes of constant inclination formed in a pattern of concentric circles by the Fabry–Perot interferometer. The fringes are particularly narrow and sharply defined so that a beam consisting of two wavelengths λ and $\lambda - \Delta\lambda$ forms two patterns of rings which are easily separated for small $\Delta\lambda$. This instrument therefore has an extremely high resolving power. The main component of the interferometer is an etalon Figure 12.7 which consists of two plane parallel glass plates with identical highly reflecting inner surfaces S_1 and S_2 which are separated by a distance d.

Suppose a monochromatic beam of unit amplitude, angular frequency ω and wavelength (in air) of λ strikes the surface S_1 as shown. A fraction t of this beam is transmitted in passing from glass to air. At S_2 a further fraction t' is transmitted in passing from air to glass to give an emerging beam of amplitude $tt' = T$. The reflection coefficient at the air–S_1 and air–S_2 surfaces is r so each subsequent emerging beam is parallel but has an amplitude factor $r^2 = R$ with respect to its predecessor. Other reflection and transmission losses are common to all beams and do not affect the analysis. Each emerging beam has a phase lag $\delta = 4\pi d \cos\theta/\lambda$ with respect to its predecessor and these parallel beams interfere when they are brought to focus via a lens.

The vector sum of the transmitted interfering amplitudes together with their appropriate phases may be written

$$\begin{aligned} A &= T\,\mathrm{e}^{\mathrm{i}\omega t} + TR\,\mathrm{e}^{\mathrm{i}(\omega t-\delta)} + TR^2\,\mathrm{e}^{\mathrm{i}(\omega t-2\delta)} \ldots \\ &= T\,\mathrm{e}^{\mathrm{i}\omega t}[1 + R\,\mathrm{e}^{-\mathrm{i}\delta} + R^2\,\mathrm{e}^{-\mathrm{i}2\delta} \ldots \end{aligned}$$

which is an infinite geometric progression with the sum

$$A = T\,\mathrm{e}^{\mathrm{i}\omega t}/(1 - R\,\mathrm{e}^{-\mathrm{i}\delta})$$

This has a complex conjugate

$$A^* = T\,\mathrm{e}^{-\mathrm{i}\omega t}/(1 - R\,\mathrm{e}^{\mathrm{i}\delta})$$

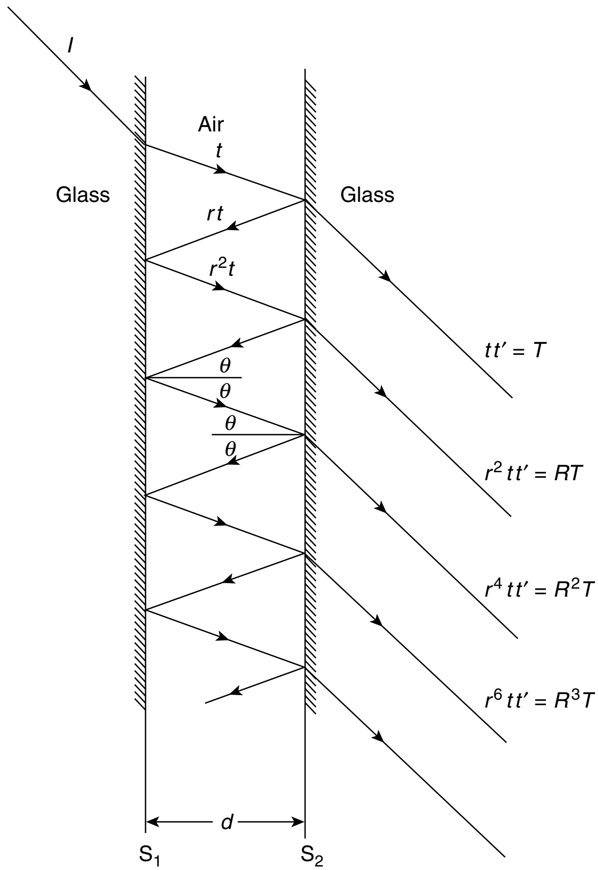


Figure 12.7 S_1 and S_2 are the highly reflecting inner surfaces of a Fabry--Perot etalon with a constant air gap thickness *d*. Multiple reflections produce parallel interfering beams with amplitudes *T*, *RT*, R^2T, etc. each beam having a phase difference

$$\delta = 4\pi d\cos\theta/\lambda$$

with respect to its neighbour

If the incident unit intensity is I_0 the fraction of this intensity in the transmitted beam may be written

$$\frac{I_t}{I_0} = \frac{AA^*}{I_0} = \frac{T^2}{(1 - R\,e^{-i\delta})(1 - R\,e^{i\delta})} = \frac{T^2}{(1 + R^2 - 2R\cos\delta)}$$

or, with

$$\cos\delta = 1 - 2\sin^2\delta/2$$

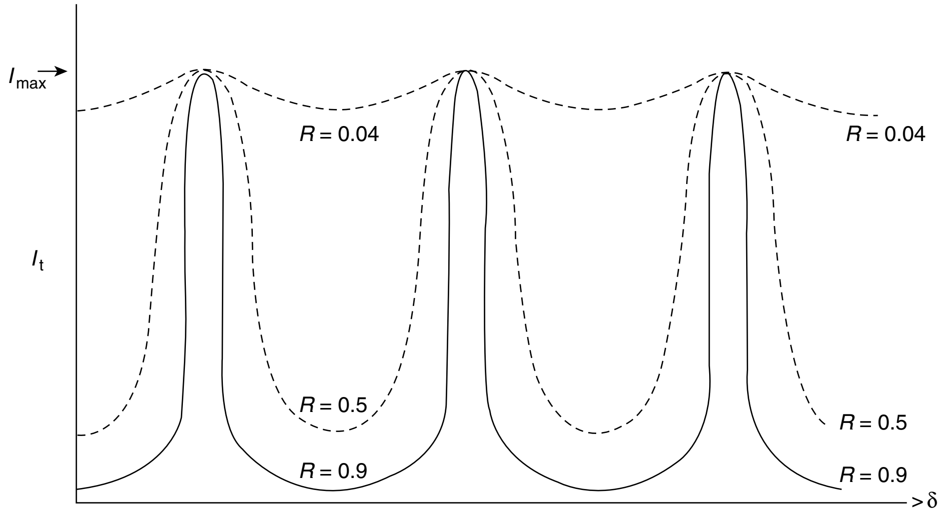


Figure 12.8 Observed intensity of fringes produced by a Fabry–Perot interferometer. Transmitted intensity I_t versus δ. $R = r^2$ where r is the reflection coefficient of the inner surfaces of the etalon. As R increases the fringes become narrower and more sharply defined

as

$$\frac{I_t}{I_0} = \frac{T^2}{(1-R)^2 + 4R\sin^2\delta/2} = \frac{T^2}{(1-R)^2}\frac{1}{1 + [4R\sin^2\delta/2/(1-R)^2]}$$

But the factor $T^2/(1-R)^2$ is a constant, written C so

$$\frac{I_t}{I_0} = C \cdot \frac{1}{1 + [4R\sin^2\delta/2/(1-R)^2]}$$

Writing $CI_0 = I_{max}$, the graph of I_t versus δ in Figure 12.8 shows that as the reflection coefficient of the inner surfaces is increased, the interference fringes become narrow and more sharply defined. Values of $R > 0.9$ may be reached using the special techniques of multilayer dielectric coating. In one of these techniques a glass plate is coated with alternate layers of high and low refractive index materials so that each boundary presents a large change of refractive index and hence a large reflection. If the *optical* thickness of each layer is $\lambda/4$ the emerging beams are all in phase and the reflected intensity is high.

Resolving Power of the Fabry – Perot Interferometer

Figure 12.8 shows that a value of $R = 0.9$ produces such narrow and sharply defined fringes that if the incident beam has two components λ and $\lambda - \Delta\lambda$ the two sets of fringes should be easily separated. The criterion for separation depends on the shape of the fringes:

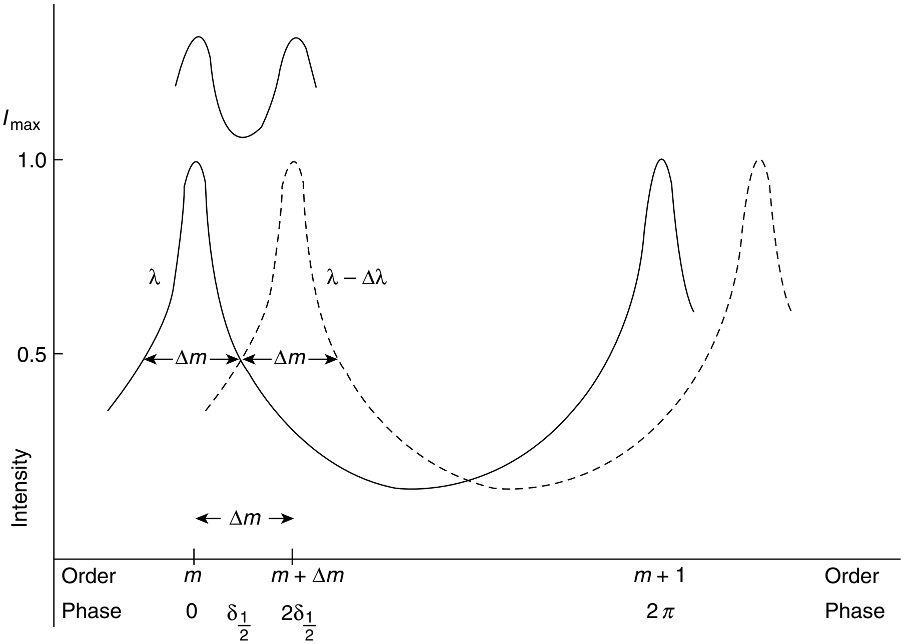


Figure 12.9 Fabry–Perot interference fringes for two wavelength λ and $\lambda - \Delta\lambda$ are resolved at order m when they cross at half their maximum intensity. Moving from order m to $m + 1$ changes the phase δ by 2π rad and the full 'half-value' width of each maximum is given by $\Delta m = 2\delta_{1/2}$ which is also the separation between the maxima of λ and $\lambda - \Delta\lambda$ when the fringes are just resolved

the diffraction grating of p. 373 uses the Rayleigh criterion, but the fringes here are so sharp that they are resolved at a much smaller separation than that required by Rayleigh.

Here the fringes of the two wavelengths may be resolved when they cross at half their maximum intensities; that is, at $I_t = I_{max}/2$ in Figure 12.9.

Using the expression

$$I_t = I_{max} \cdot \frac{1}{1 + \dfrac{4R \sin^2 \delta/2}{(1 - R)^2}}$$

we see that $I_t = I_{max}$ when $\delta = 0$ and $I_t = I_{max}/2$ when the factor

$$4R \sin^2 \delta/2/(1 - R)^2 = 1$$

The fringes are so narrow that they are visible only for very small values of δ so we may replace $\sin \delta/2$ by $\delta/2$ in the expression

$$4R \sin^2 \delta/2/(1 - R)^2 = 1$$

to give the value

$$\delta_{1/2} = \frac{(1-R)}{R^{1/2}}$$

as the phase departure from the maximum, $\delta = 0$, which produces the intensity $I_t = I_{max}/2$ for wavelength λ. Our criterion for resolution means, therefore, that the maximum intensity for $\lambda - \Delta\lambda$ is removed an extra amount $\delta_{1/2}$ along the phase axis of Figure 12.9. This axis also shows the order of interference *m* at which the wavelengths are resolved, together with the order $m + 1$ which represents a phase shift of $\delta = 2\pi$ along the phase axis.

In the *m*th order of interference we have

$$2d\cos\theta = m\lambda$$

and for fringes of equal inclination (θ constant), logarithmic differentiation gives

$$\lambda/\Delta\lambda = -m/\Delta m$$

Now $\Delta m = 1$ represents a phase change of $\delta = 2\pi$ and the phase difference of $2.\delta_{1/2}$ which just resolves the two wavelengths corresponds to a change of order

$$\Delta m = 2.\delta_{1/2}/2\pi$$

Thus, the resolving power, defined as

$$\frac{\lambda}{\Delta\lambda} = \left|\frac{m}{\Delta m}\right| = \frac{m\pi}{\delta_{1/2}} = \frac{m\pi R^{1/2}}{(1-R)}$$

The equivalent expression for the resolving power in the *m*th order for a diffracting grating of *N* lines (interfering beams) is shown on p. 376 to be

$$\frac{\lambda}{\Delta\lambda} = mN$$

so we may express

$$N' = \pi R^{1/2}/(1-R)$$

as the effective number of interfering beams in the Fabry–Perot interferometer.

This quantity N' is called the *finesse* of the etalon and is a measure of its quality. We see that

$$N' = \frac{2\pi}{2\delta_{1/2}} = \frac{1}{\Delta m} = \frac{\text{separation between orders } m \text{ and } m+1}{\text{'half value' width of } m\text{th order}}$$

Thus, using one wavelength only, the ratio of the separation between successive fringes to the narrowness of each fringe measures the quality of the etalon. A typical value of $N' \sim 30$.

Free Spectral Range

There is a limit to the wavelength difference $\Delta\lambda$ which can be resolved with the Fabry–Perot interferometer. This limit is reached when $\Delta\lambda$ is such that the circular fringe for λ in the mth order coincides with that for $\lambda - \Delta\lambda$ in the $m + 1$th order. The pattern then loses its unique definition and this value of $\Delta\lambda$ is called the *free spectral range*.

From the preceding section we have the expression

$$\frac{\lambda}{\Delta\lambda} = -\frac{m}{\Delta m}$$

and in the limit when $\Delta\lambda$ represents the free spectral range then

$$\Delta m = 1$$

and

$$\Delta\lambda = -\lambda/m$$

But $m\lambda = 2d$ when $\theta \simeq 0$ so the free spectral range

$$\Delta\lambda = -\lambda^2/2d$$

Typically $d \sim 10^{-2}$ m and for λ (cadmium red) $= 0.6438$ microns we have, from $2d = m\lambda$, a value of

$$m \approx 3 \times 10^4$$

Now the resolving power

$$\frac{\lambda}{\Delta\lambda} = mN'$$

so, for

$$N' \approx 30$$

the resolving power can be as high as 1 part in 10^6.

Central Spot Scanning

Early interferometers recorded flux densities on photographic plates but the non-linear response of such a technique made accurate resolution between two wavelengths tedious and more difficult. This method has now been superseded by the use of photoelectronic detectors which have the advantage of a superior and more reliable linearity. Moreover, the response of such a device with controlled vibration of one mirror of the etalon allows the variation of the intensity across the free spectral range to be monitored continuously.

The vibration of the mirror, originally electro-mechanical, is now most often produced by using a piezoelectric material on which to mount one of the etalon mirrors. When a voltage is applied to such a material it changes its length and the distance d between the etalon mirrors can be varied. The voltage across the piezoelectric mount is tailored to produce the desired motion.

Changing d by $\lambda/2$ is equivalent to changing Δm by 1, which corresponds to a scan of the free spectral range, $\Delta\lambda$, when $\lambda/\Delta\lambda = |m/\Delta m|$ (Figure 12.9).

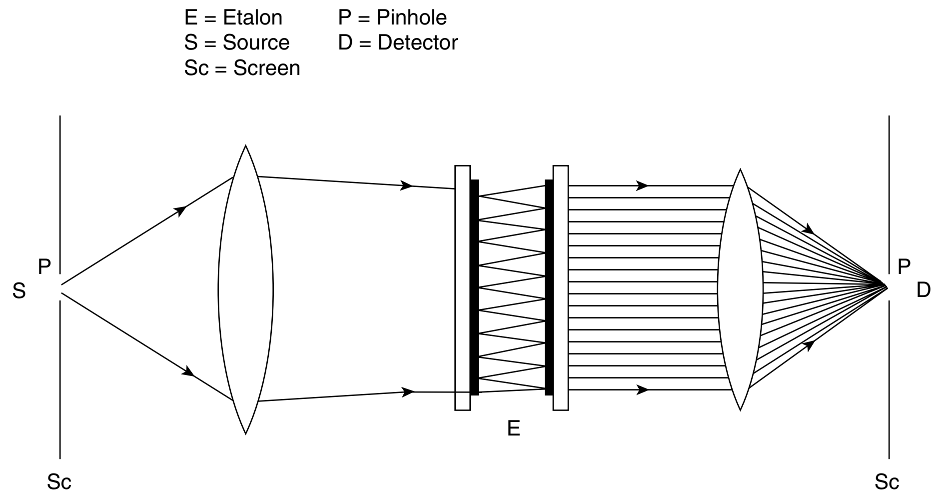


Figure 12.10 Fabry–Perot etalon central spot scanning. The distance between the etalon mirrors changes when one mirror vibrates on its piezoelectric mount. The free spectral range is scanned over many vibration cycles at a central spot and a stationary trace is obtained on the oscilloscope screen

One of the most common experimental arrangements is that of central spot scanning (Figure 12.10). Where the earlier photographic technique recorded the flux density over a wide region for a short period, central spot scanning focuses on a single point in space for a long period over many cycles of the etalon vibration. Matching the time base of the oscilloscope to the vibration period of the etalon produces a stationary trace on the screen which can be measured directly in addition to being filmed for record purposes.

The Laser Cavity

The laser cavity is in effect an extended Fabry–Perot etalon. Mirrors coated with multi-dielectric films described in the next section can produce reflection coefficients $R \approx 0.99$ and the amplified stimulated emission in the laser produces a beam which is continuously reflected between the mirror ends of the cavity. The high value of R allows the amplitudes of the beam in opposing directions to be taken as equal, so a standing wave system is generated (Figure 12.11) to form a longitudinal mode in the cavity.

The superposed amplitudes after a return journey from one mirror to the other and back are written for a wave number k and a frequency $\omega = 2\pi\nu$ as

$$\begin{aligned} E &= A_1(\mathrm{e}^{\mathrm{i}(\omega t - kx)} - \mathrm{e}^{\mathrm{i}(\omega t + kx)}) \\ &= A_1(\mathrm{e}^{-\mathrm{i}kx} - \mathrm{e}^{\mathrm{i}kx})\,\mathrm{e}^{\mathrm{i}\omega t} = -2\mathrm{i}A_1 \sin kx\, \mathrm{e}^{\mathrm{i}\omega t} \end{aligned}$$

of which the real part is $E = 2A_1 \sin kx \sin \omega t$.

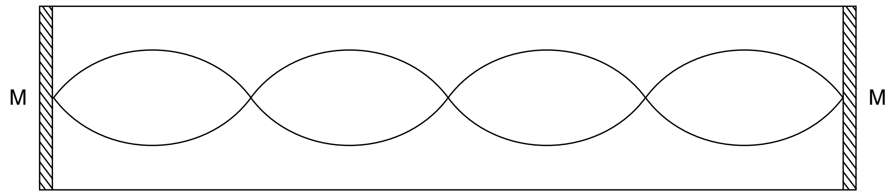


Figure 12.11 A longitudinal mode in a laser cavity which behaves as an extended Fabry–Perot etalon with highly reflecting mirrors at each end. The standing wave system acquires an extra $\lambda/2$ for unit change in the mode number m. A typical output is shown in Figure 12.12

If the cavity length is L, one round trip between the mirrors creates a phase change of

$$\phi = -2Lk + 2\alpha = -\frac{4\pi L}{c}\nu + 2\alpha$$

where α is the phase change on reflection at each mirror.

For this standing wave mode to be maintained, the phase change must be a multiple of 2π, so for m an integer

$$\phi = m2\pi = \frac{4\pi L}{c}\,\nu - 2\alpha$$

or

$$\nu = \frac{mc}{2L} + \frac{\alpha c}{2\pi L}$$

When m changes to $m + 1$, the phase change of 2π corresponds to an extra wavelength λ for the return journey; that is, an extra $\lambda/2$ in the standing wave mode. A series of longitudinal modes can therefore exist with frequency intervals $\Delta\nu = c/2L$ determined by a unit change in m.

The intensity profile for each mode and the separation $\Delta\nu$ is best seen by reference to Figure 12.9, where $\phi \equiv \delta$ is given by the unit change in the order of interference from m to $m + 1$.

The intensity profile for each cavity mode is that of Figure 12.9, where the full width at half maximum intensity is given by the phase change

$$2\delta_{1/2} = \frac{2(1 - R)}{R^{1/2}}$$

where R is the reflection coefficient. This corresponds to a full width intensity change over a frequency $d\nu$ generated by the phase change

$$d\phi = \frac{4\pi L}{c}\,d\nu \quad \text{in the expression for } \phi \text{ above}$$

The width at half maximum intensity for each longitudinal mode is therefore given by

$$\frac{4\pi L}{c}\,d\nu = \frac{2(1-R)}{R^{1/2}}$$

or

$$d\nu = \frac{(1-R)c}{R^{1/2}2\pi L}$$

For a laser 1 m long with $R = 0.99$, the longitudinal modes are separated by frequency intervals

$$\Delta\nu = \frac{c}{2L} = 1.5 \times 10^{8}\,\text{Hz}$$

Each mode intensity profile has a full width at half maximum of

$$d\nu = 10^{-2}\frac{c}{2\pi} \approx 4.5 \times 10^{5}\,\text{Hz}$$

For a He–Ne laser the mean frequency of the output at 632.8 nm is 4.74×10^{14} Hz. The pattern for $\Delta\nu$ and $d\nu$ is shown in Figure 12.12, where the intensities are reduced under the dotted envelope as the frequency difference from the mean is increased.

The finesse of the laser cavity is given by

$$\frac{\Delta\nu}{d\nu} = \frac{1.5 \times 10^{8}}{4.5 \times 10^{5}} \approx 300$$

for the example quoted.

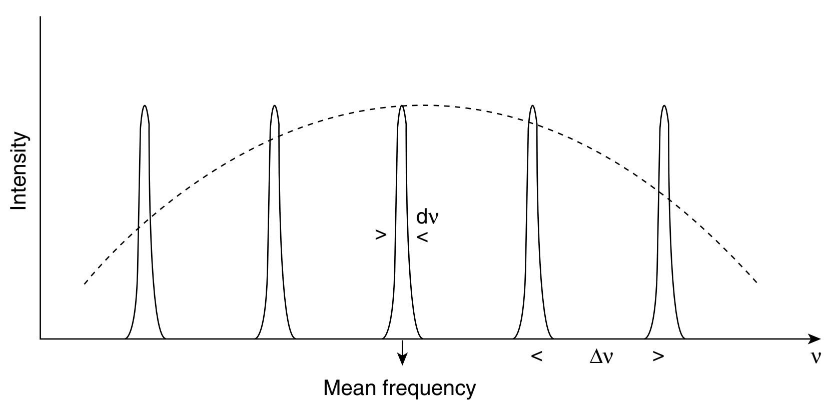


Figure 12.12 Output of a laser cavity. A series of longitudinal modes separated by frequency intervals $\Delta\nu = c/2L$, where c is the velocity of light and L is the cavity length. The modes are centred about the mean output frequency and are modulated under the dotted envelope. For a He–Ne laser 1 m long the separation $\Delta\nu$ between the modes $\approx$ 300 full widths of a mode intensity profile at half its maximum value

The intensity of each longitudinal mode is of course, amplified by each passage of the stimulated emission. Radiation allowed from out of one end represents the laser output but the amplification process is dominant and the laser produces a continuous beam.

Multilayer Dielectric Films

We have just seen that in the *m*th order of interference the resolving power of a Fabry–Perot interferometer is given by

$$\lambda/\Delta\lambda \equiv mN'$$

where the finesse or number of interfering beams

$$N' = \pi R^{1/2}/(1-R) = \pi r/(1-r^2)$$

and r is the reflection coefficient of the inner surfaces of the etalon.

It is evident that as $r \to 1$ the values of N' and the resolving power become much larger. The value of r can be increased to more than 99% by using a metallic coating on the inner surfaces of the etalon or by depositing on them a multilayer of dielectric films with alternating high and low refractive indices. For a given monochromatic electromagnetic wave each layer or film has an optical thickness of $\lambda/4$.

The reflection coefficient r for such a wave incident on the surface of a higher refractive index film is increased because the externally and internally reflected waves are in phase; a phase change of π occurs only on the outer surface and is reinforced by the π phase change of the wave reflected at the inner surface which travels an extra $\lambda/2$ optical distance.

High values of r result from films of alternating high and low values of the refractive index because reflections from successive boundaries are in phase on return to the front surface of the first film. Those retarded an odd multiple of π by the extra optical path length per film also have a π phase change on reflection to make a total of 2π rad.

We consider the simplest case of a monochromatic electromagnetic wave in a medium of refractive index n_1, *normally incident* on a single film of refractive index n'_1, and thickness d'_1. This film is deposited on the surface of a material of refractive index n'_2, which is called a substrate (Figure 12.13). The phase lag for a single journey across the film is written δ.

The boundary conditions are that the components of the E and H fields parallel to a surface are continuous across that surface. We write these field amplitudes as E_f and $H_f = nE_f$ for the forward-going wave to the right in Figure 12.13 and E_r and $H_r = nE_r$ for the reflected wave going to the left.

We see that at surface 1 the boundary conditions for the electric field E are

$$E_{f1} + E_{r1} = E'_{f1} + E'_{r1} \tag{12.1a}$$

and for the magnetic field

$$n_1E_{f1} - n_1E_{r1} = n'_1E'_{f1} - n'_1E'_{r1} \tag{12.1b}$$

where the negative sign for the reflected amplitude arises when the $\mathbf{E} \times \mathbf{H}$ direction of the wave is reversed (see Figure 8.7).

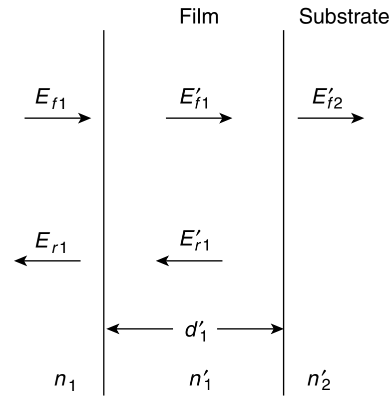


Figure 12.13 A thin dielectric film is deposited on a substrate base. At each surface an electromagnetic wave is normally incident, as E_{fi}, in a medium of refractive index n_i and is reflected as E_{ri}. A multilayer stack of such films, each of optical thickness $\lambda/4$ and of alternating high and low refractive indices can produce reflection coefficients >99%

At surface 2 in Figure 12.13, E'_{f1} arrives with a phase lag of δ with respect to E'_{f1} at surface 1 but the E'_{r1} wave at surface 2 has a phase δ in advance of E'_{r1} at surface 1, so we have the boundary conditions

$$E'_{f1}\,\mathrm{e}^{-i\delta} + E'_{r1}\,\mathrm{e}^{i\delta} = E'_{f2} \tag{12.1c}$$

and

$$n'_1 E'_{f1}\,\mathrm{e}^{-i\delta} - n'_1 E'_{r1}\,\mathrm{e}^{i\delta} = n'_2 E'_{f2} \tag{12.1d}$$

We can eliminate E'_{f1} and E'_{r1} from equations (12.1a)–(12.1d) to give

$$1 + \frac{E_{r1}}{E_{f1}} = \left(\cos\delta + i\frac{n'_2}{n'_1}\sin\delta\right)\frac{E'_{f2}}{E_{f1}} \tag{12.2}$$

and

$$n_1 - n_1\frac{E_{r1}}{E_{f1}} = (in'_1\sin\delta + n'_2\cos\delta)\frac{E'_{f2}}{E_{f1}} \tag{12.3}$$

which we can express in matrix form

$$\begin{bmatrix} 1 \\ n_1 \end{bmatrix} + \begin{bmatrix} 1 \\ -n_1 \end{bmatrix}\frac{E_{r1}}{E_{f1}} = \begin{bmatrix} \cos\delta & i\sin\delta/n'_1 \\ in'_1\sin\delta & \cos\delta \end{bmatrix}\begin{bmatrix} 1 \\ n'_2 \end{bmatrix}\frac{E'_{f2}}{E_{f1}}$$

We write this as

$$\begin{bmatrix} 1 \\ n_1 \end{bmatrix} + \begin{bmatrix} 1 \\ -n_1 \end{bmatrix} r = M_1 \begin{bmatrix} 1 \\ n_2' \end{bmatrix} t$$

where $r = E_{r1}/E_{f1}$ is the reflection coefficient at the first surface and $t = E'_{f2}/E_{f1}$ is the transmitted coefficient into medium n_2' (a quantity we shall not evaluate).

The 2×2 matrix

$$M_1 = \begin{bmatrix} \cos\delta & i\sin\delta/n_1' \\ in_1'\sin\delta & \cos\delta \end{bmatrix}$$

relates r and t across the first film and is repeated with appropriate values of n_i' for each successive film. The product of these 2×2 matrices is itself a 2×2 matrix as with the repetitive process we found in the optical case of p. 325.

Thus, for N films we have

$$\begin{bmatrix} 1 \\ n_1 \end{bmatrix} + \begin{bmatrix} 1 \\ -n_1 \end{bmatrix} R = M_1 M_2 M_3 \cdots M_N \begin{bmatrix} 1 \\ n'_{N+1} \end{bmatrix} T, \tag{12.4}$$

where $R = E_{r1}/E_{f1}$ as before and $T = E'_{f(N+1)}/E_{f1}$ the transmitted coefficient after the final film. *Note, however, that E_{r1} in R is now the result of reflection from all the film surfaces and that these are in phase.*

The typical matrix M_3 relates r to t across the third film and the product of the matrices

$$M_1 M_2 M_3 \cdots M_N = M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}$$

is a 2×2 matrix.

Eliminating T from the two simultaneous equations (12.4) we have, in terms of the coefficients of M

$$R = \frac{A - B}{A + B} \tag{12.5}$$

where

$$A = n_1(M_{11} + M_{12} n'_{N+1})$$

and

$$B = (M_{21} + M_{22} n'_{N+1})$$

If we now consider a system of two films, the first of higher refractive index n_H and the second of lower refractive index n_L, where each has an optical thickness $d = \lambda/4$, then the phase $\delta = \pi/2$ for each film and

$$M_1 M_2 = \begin{bmatrix} 0 & i/n_H \\ in_H & 0 \end{bmatrix} \begin{bmatrix} 0 & i/n_L \\ in_L & 0 \end{bmatrix} = \begin{bmatrix} -n_L/n_H & 0 \\ 0 & -n_H/n_L \end{bmatrix}$$

A stack of N such pairs, $2N$ films in all with alternating n_H and n_L, produces

$$M_1M_2\cdots M_{2N} = [M_1M_2]^N = \begin{bmatrix} \left(\dfrac{-n_L}{n_H}\right)^N & 0 \\ 0 & \left(\dfrac{-n_H}{n_L}\right)^N \end{bmatrix}$$

giving R the total reflection coefficient from equation (12.5) equal to

$$R = \frac{\left(\dfrac{-n_L}{n_H}\right)^N - \left(\dfrac{-n_H}{n_L}\right)^N}{\left(\dfrac{-n_L}{n_H}\right)^N + \left(\dfrac{-n_H}{n_L}\right)^N}$$

We see that as long as $n_H \neq n_L$, then as $N \to \infty$, $R \to 1$ and this value may be used in our derivation of the expressions for the finesse and resolving power of the Fabry–Perot interferometer.

Multilayer stacks using zinc sulphate $(n_H = 2.3)$ and cryolite $(n_L = 1.35)$ have achieved R values > 99.5%.

Note that all the 2×2 matrices and their products have determinants equal to unity which states that the column vectors represent a quantity which remains invariant throughout the matrix transformations.

(Problem 12.2)

The Thin Film Optical Wave Guide

Figure 12.14 shows a thin film of width d and refractive index n along which light of frequency ν and wave number k is guided by multiple internal reflections. The extent of the

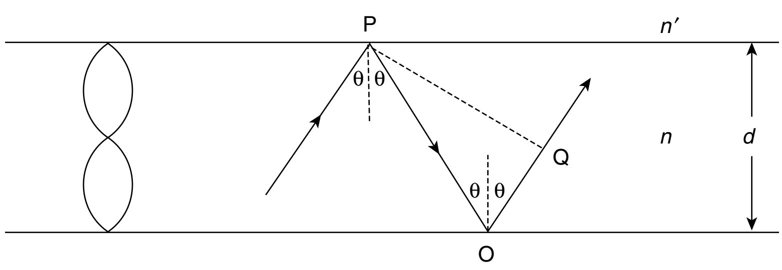


Figure 12.14 A thin dielectric film or fibre acts as an optical wave guide. The reflection angle θ must satisfy the relation $n\sin\theta \geq n'$, where n' is the refractive index of the coating over the film of refractive index n. Propagating modes have standing wave systems across the film as shown and constructive interference occurs on the standing wave axis where the amplitude is a maximum. Destructive interference occurs at the nodes

wave guide is infinite in the direction normal to the page. The internal reflection angle θ must satisfy

$$n \sin\theta \geq n'$$

where n' is the refractive index of the medium bounding the thin film surfaces. Each reflected ray is normal to a number of wave fronts of constant phase separated by λ, where $k = 2\pi/\lambda$ and constructive interference is necessary for any mode to propagate. Reflections may take place at any pair of points P and O along the film and we examine the condition for constructive interference by considering the phase difference along the path POQ, taking into account a phase difference α introduced by reflection at each of P and Q.

Now

$$\mathrm{PO} = \cos\theta/d$$

and

$$\mathrm{OQ} = \mathrm{PO}\cos 2\theta$$

so with

$$\cos 2\theta = 2\cos^2\theta - 1$$

we have

$$\mathrm{PO} + \mathrm{OQ} = 2d\cos\theta$$

giving a phase difference

$$\Delta\phi = \phi_Q - \phi_P = -\frac{2\pi\nu}{c}(n\,2d\cos\theta) + 2\alpha$$

Constructive interference requires

$$\Delta\phi = m\,2\pi$$

where m is an integer, so we write

$$m\,2\pi = \frac{2\pi\nu}{c}\,n\,2d\cos\theta - 2\pi\Delta m$$

where

$$\Delta m = 2\alpha/2\pi$$

represents the phase change on reflection.

Radiation will therefore propagate only when

$$\cos\theta = \frac{c(m + \Delta m)}{\nu\,2nd}$$

for $m = 0, 1, 2, 3$.

The condition $n \sin\theta \geq n'$ restricts the values of the frequency ν which can propagate. If $\theta = \theta_m$ for mode m and

$$\cos\theta_m = (1 - \sin^2\theta_m)^{1/2}$$

then

$$n \sin\theta_m \geq n'$$

becomes

$$\cos\theta_m \leq \left[1 - \left(\frac{n'}{n}\right)^2\right]^{1/2}$$

and ν must satisfy

$$\nu \geq \frac{c(m + \Delta m)}{2d(n^2 - n'^2)^{1/2}}$$

The mode $m = 0$ is the mode below which ν will not propagate, while Δm is a constant for a given wave guide. Each mode, Figure 12.14, is represented by a standing wave system across the wave guide normal to the direction of propagation. Constructive interference occurs on the axis of this wave system where the amplitude is a maximum and destructive interference is indicated by the nodes.

Division of Wavefront

Interference Between Waves from Two Slits or Sources

In Figure 12.15 let S_1 and S_2 be two equal sources separated by a distance *f*, each generating a wave of angular frequency ω and amplitude *a*. At a point P sufficiently distant from S_1 and S_2 only plane wavefronts arrive with displacements

$$y_1 = a \sin(\omega t - kx_1) \quad \text{from } S_1$$

and

$$y_2 = a \sin(\omega t - kx_2) \quad \text{from } S_2$$

so that the phase difference between the two signals at P is given by

$$\delta = k(x_2 - x_1) = \frac{2\pi}{\lambda}(x_2 - x_1)$$

This phase difference δ, which arises from the path difference $x_2 - x_1$, depends only on x_1, x_2 and the wavelength λ and not on any variation in the source behaviour. This requires that there shall be no sudden changes of phase in the signal generated at either source – such sources are called *coherent*.

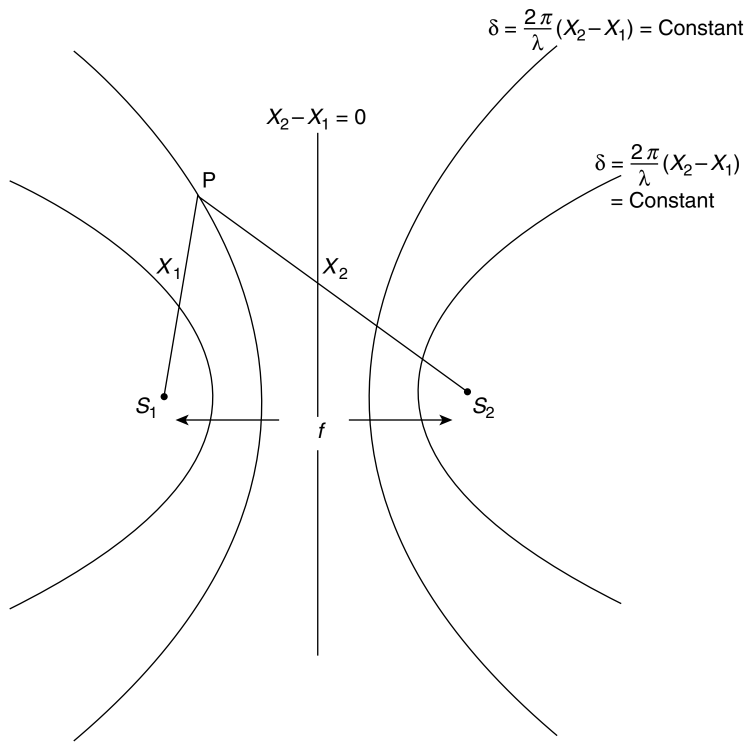


Figure 12.15 Interference at P between waves from equal sources S_1 and S_2, separation f, depends only on the path difference $x_2 - x_1$. Loci of points with constant phase difference $\delta = (2\pi/\lambda)(x_2 - x_1)$ are the family of hyperbolas with S_1 and S_2 as foci

The superposition of displacements at P gives a resultant

$$R = y_1 + y_2 = a[\sin(\omega t - kx_1) + \sin(\omega t - kx_2)]$$

Writing $X \equiv (x_1 + x_2)/2$ as the average distance from the two sources to point P we obtain

$$kx_1 = kX - \delta/2 \quad \text{and} \quad kx_2 = kX + \delta/2$$

to give

$$\begin{aligned} R &= a[\sin(\omega t - kX + \delta/2) + \sin(\omega t - kX - \delta/2)] \\ &= 2a\sin(\omega t - kX)\cos\delta/2 \end{aligned}$$

and an intensity

$$I = R^2 = 4a^2\sin^2(\omega t - kX)\cos^2\delta/2$$

When

$$\cos\frac{\delta}{2} = \pm 1$$

the spatial intensity is a maximum,

$$I = 4a^2$$

and the component displacements reinforce each other to give *constructive interference*. This occurs when

$$\frac{\delta}{2} = \frac{\pi}{\lambda}(x_2 - x_1) = n\pi$$

that is, when the path difference

$$x_2 - x_1 = n\lambda$$

When

$$\cos\frac{\delta}{2} = 0$$

the intensity is zero and the components cancel to give *destructive interference*. This requires that

$$\frac{\delta}{2} = (2n+1)\frac{\pi}{2} = \frac{\pi}{\lambda}(x_2 - x_1)$$

or, the path difference

$$x_2 - x_1 = (n + \tfrac{1}{2})\lambda$$

The loci or sets of points for which $x_2 - x_1$ (or δ) is constant are shown in Figure 12.15 to form hyperbolas about the foci S_1 and S_2 (in three dimensions the loci would be the hyperbolic surfaces of revolution).

Interference from Two Equal Sources of Separation *f*

Separation $f \gg \lambda$. Young's Slit Experiment

One of the best known methods for producing optical interference effects is the Young's slit experiment. Here the two coherent sources, Figure 12.16, are two identical slits S_1 and S_2 illuminated by a monochromatic wave system from a single source equidistant from S_1 and S_2. The observation point P lies on a screen which is set at a distance l from the plane of the slits.

The intensity at P is given by

$$I = R^2 = 4a^2\cos^2\frac{\delta}{2}$$

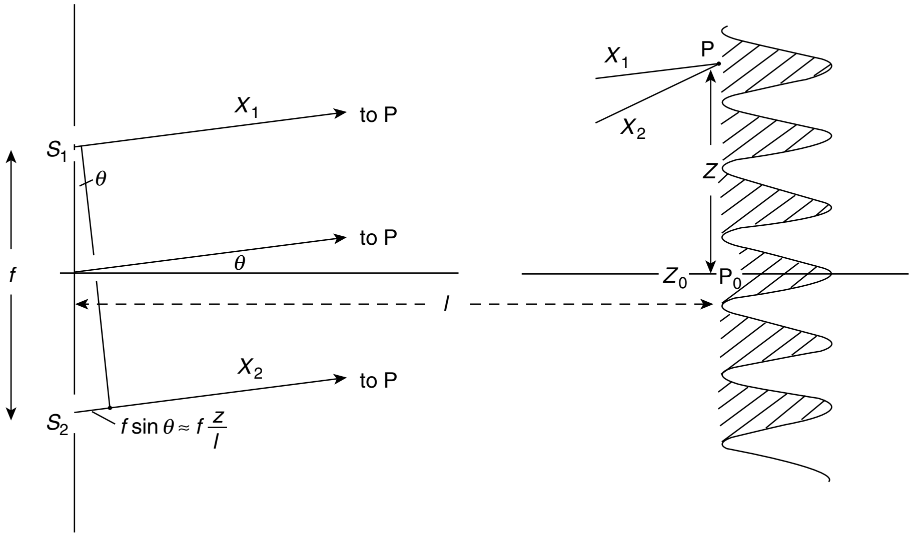


Figure 12.16 Waves from equal sources S_1 and S_2 interfere at P with phase difference $\delta = (2\pi/\lambda)(x_2 - x_1) = (2\pi/\lambda)f\sin\theta \approx (2\pi/\lambda)f(z/l)$. The distance $l \gg z$ and f so S_1P and S_2P are effectively parallel. Interference fringes of intensity $I = I_0\cos^2\delta/2$ are formed in the plane PP_0

and the distances $PP_0 = z$ and slit separation f are both very much less than l (experimentally $\approx 10^{-3}\,l$). This is indicated by the break in the lines x_1 and x_2 in Figure 12.16 where S_1P and S_2P may be considered as sufficiently parallel for the path difference to be written as

$$x_2 - x_1 = f\sin\theta = f\frac{z}{l}$$

to a very close approximation.

Thus

$$\delta = \frac{2\pi}{\lambda}(x_2 - x_1) = \frac{2\pi}{\lambda}f\sin\theta = \frac{2\pi}{\lambda}f\frac{z}{l}$$

If

$$I = 4a^2\cos^2\frac{\delta}{2}$$

then

$$I = I_0 = 4a^2 \quad \text{when} \quad \cos\frac{\delta}{2} = 1$$

that is, when the path difference

$$f\frac{z}{l} = 0, \quad \pm\lambda, \quad \pm 2\lambda, \quad \ldots \pm n\lambda$$

and

$$I = 0 \quad \text{when} \quad \cos\frac{\delta}{2} = 0$$

that is, when

$$f\frac{z}{l} = \pm\frac{\lambda}{2}, \quad \pm\frac{3\lambda}{2}, \quad \pm(n + \tfrac{1}{2})\lambda$$

Taking the point P_0 as $z = 0$, the variation of intensity with z on the screen P_0P will be that of Figure 12.16, a series of alternating straight bright and dark fringes parallel to the slit directions, the bright fringes having $I = 4a^2$ whenever $z = n\lambda l/f$ and the dark fringes $I = 0$, occurring when $z = (n + \frac{1}{2})\lambda l/f$, n being called the *order of interference* of the fringes. The zero order $n = 0$ at the point P_0 is the central bright fringe. The distance on the screen between two bright fringes of orders n and $n + 1$ is given by

$$z_{n+1} - z_n = [(n+1) - n]\frac{\lambda l}{f} = \frac{\lambda l}{f}$$

which is also the physical separation between two consecutive dark fringes. The spacing between the fringes is therefore constant and independent of n, and a measurement of the spacing, l and f determines λ.

The intensity distribution curve (Figure 12.17) shows that when the two wave trains arrive at P exactly out of phase they interfere destructively and the resulting intensity or energy flux is zero. Energy conservation requires that the energy must be redistributed, and that lost at zero intensity is found in the intensity peaks. The average value of $\cos^2\delta/2$ is $\frac{1}{2}$, and the dotted line at $I = 2a^2$ is the average intensity value over the interference system which is equal to the sum of the separate intensities from each slit.

There are two important points to remember about the intensity interference fringes when discussing diffraction phenomena; these are

- The intensity varies with $\cos^2\delta/2$.
- The maxima occur for path differences of zero or integral numbers of the wavelength, whilst the minima represent path differences of odd numbers of the half-wavelength.

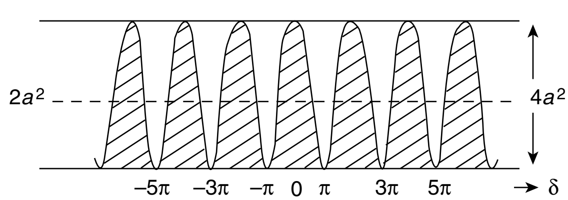


Figure 12.17 Intensity of interference fringes is proportional to $\cos^2\delta/2$, where δ is the phase difference between the interfering waves. The energy which is lost in destructive interference (minima) is redistributed into regions of constructive interference (maxima)

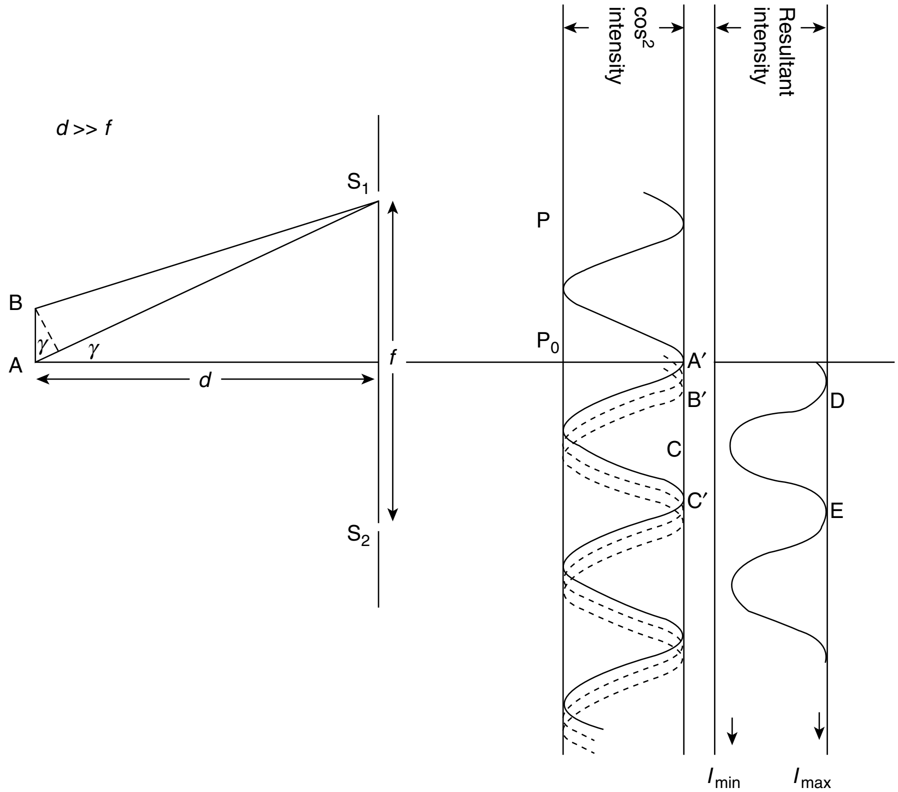


Figure 12.18 The point source A produces the $\cos^2$ interference fringes represented by the solid curve A′C′. Other points on the line source AB produce $\cos^2$ fringes (the displaced broken curves B′) and the observed total intensity is the curve DE. When the points on AB extend A′B′ to C the fringes disappear and the field is uniformly illuminated

Spatial Coherence In the preceding section nothing has been said about the size of the source producing the plane wave which falls on S_1 and S_2. If this source is an ideal *point* source A equidistant from S_1 and S_2, Figure 12.18, then a single set of $\cos^2$ fringes is produced. But every source has a finite size, given by AB in Figure 12.18, and each point on the line source AB produces its own set of interference fringes in the plane PP_0; the eye observing the sum of their intensities.

If the solid curve A′C′ is the intensity distribution for the point A of the source and the broken curves up to B′ represent the corresponding fringes for points along AB the resulting intensity curve is DE. Unless A′B′ extends to C the variations of DE will be seen as faint interference bands. These intensity variations were quantified by Michelson, who defined the

$$\text{Visibility} = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}}$$

The $\cos^2$ fringes from a point source obviously have a visibility of unity because the minimum intensity $I_{\min} = 0$.

When $A'B'$ of Figure 12.18 $= A'C$, the point source fringe separation (or a multiple of it) the field is uniformly illuminated, fringe visibility $= 0$ and the fringes disappear.

This occurs when the path difference

$$\mathrm{AS}_2 - \mathrm{BS}_1 \approx \mathrm{AB}\sin\gamma = \lambda/2 \quad \text{where } \mathrm{AS}_2 = \mathrm{AS}_1.$$

Thus, the requirement for fringes of good visibility imposes a limit on the finite size of the source. Light from points on the source must be *spatially coherent* in the sense that $\mathrm{AB}\sin\gamma \ll \lambda/2$ in Figure 12.18.

But for $f \ll d$,

$$\sin\gamma \approx f/2d$$

so the coherence condition becomes

$$\sin\gamma = f/2d \ll \lambda/2\mathrm{AB}$$

or

$$\mathrm{AB}/d \ll \lambda/f$$

where AB/d measures the angle subtended by the source at the plane $\mathrm{S}_1\mathrm{S}_2$.

Spatial coherence therefore requires that the angle subtended by the source

$$\ll \lambda/f$$

where f is the linear size of the diffracting system. (Note also that λ/f measures $\theta(\sim z/l)$ the angular separation of the fringes in Figure 12.16.)

As an example of spatial coherence we may consider the production of Young's interference fringes using the sun as a source.

The sun subtends an angle of 0.018 rad at the earth and if we accept the approximation

$$\frac{\mathrm{AB}}{d} \ll \frac{\lambda}{f} \approx \frac{\lambda}{4f}$$

with $\lambda = 0.5\,\mu\mathrm{m}$,
we have

$$f \sim \frac{0.5}{4(0.018)} \sim 14\,\mu\mathrm{m}$$

This small value of slit separation is required to meet the spatial coherence condition.

Separation $f \ll \lambda(kf \ll 1$ where $k = 2\pi/\lambda)$

If there is a zero phase difference between the signals leaving the sources S_1 and S_2 of Figure 12.16 then the intensity at some distant point P may be written

$$I = 4a^2\cos^2\frac{\delta}{2} = 4I_\mathrm{s}\cos^2\frac{kf\sin\theta}{2} \approx 4I_\mathrm{s},$$

where the path difference $S_2P - S_1P = f \sin\theta$ and $I_s = a^2$ is the intensity from each source.

We note that, since $f \ll \lambda (kf \ll 1)$, the intensity has a very small θ dependence and the two sources may be effectively replaced by a single source of amplitude $2a$.

Dipole Radiation ($f \ll \lambda$)

Suppose, however, that the signals leaving the sources S_1 and S_2 are anti-phase so that their total phase difference at some distant point P is

$$\delta = (\delta_0 + kf \sin\theta)$$

where $\delta_0 = \pi$ is the phase difference introduced at source.

The intensity at P is given by

$$\begin{aligned} I &= 4 I_s \cos^2 \frac{\delta}{2} = 4 I_s \cos^2\left(\frac{\pi}{2} + \frac{kf \sin\theta}{2}\right) \\ &= 4 I_s \sin^2\left(\frac{kf \sin\theta}{2}\right) \\ &\approx I_s (kf \sin\theta)^2 \end{aligned}$$

because

$$kf \ll 1$$

Two anti-phase sources of this kind constitute a *dipole* whose radiation intensity $I \ll I_s$ the radiation from a single source, when $kf \ll 1$. The efficiency of radiation is seen to depend on the product kf and, for a fixed separation f the dipole is a less efficient radiator at low frequencies (small k) than at higher frequencies. Figure 12.19 shows the radiation intensity I plotted against the polar angle θ and we see that for the dipole axis along the direction $\theta = \pi/2$, completely destructive interference occurs only on the perpendicular axis $\theta = 0$ and $\theta = \pi$. There is no direction (value of θ) giving completely constructive interference. The highest value of the radiated intensity occurs along the axis $\theta = \pi/2$ and $\theta = 3\pi/2$ but even this is only

$$I = (kf)^2 I_s,$$

where

$$kf \ll 1$$

The directional properties of a radiating dipole are incorporated in the design of transmitting aerials. In acoustics a loudspeaker may be considered as a multi dipole source, the face of the loudspeaker generating compression waves whilst its rear propagates rarefactions. Acoustic reflections from surrounding walls give rise to undesirable interference effects which are avoided by enclosing the speaker in a cabinet. Bass reflex or phase inverter cabinets incorporate a vent on the same side as the speaker face at an acoustic distance of half a wavelength from the rear of the speaker. The vent thus acts as a second source *in phase* with the speaker face and radiation is improved.

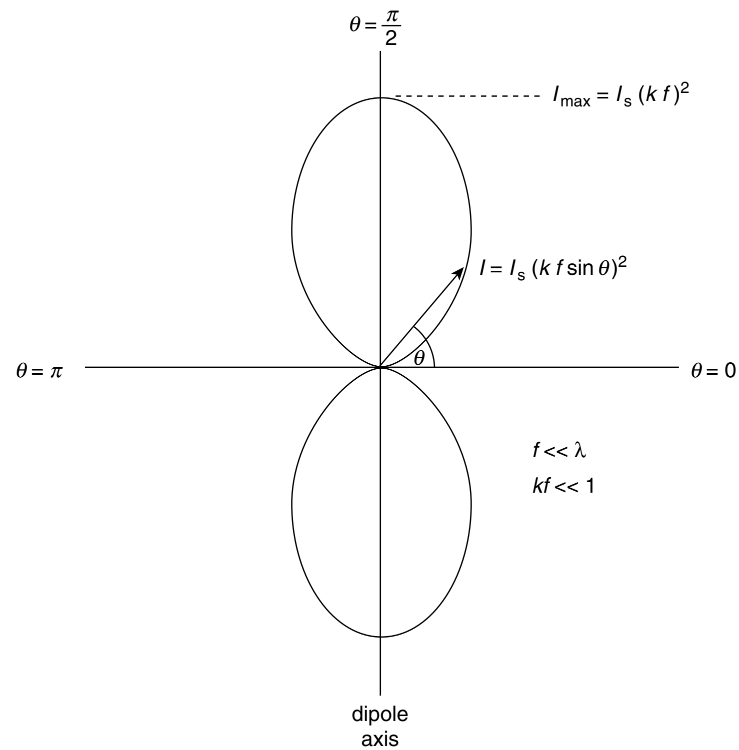


Figure 12.19 Intensity I versus direction θ for interference pattern between waves from two equal sources, π rad out of phase (dipole) with separation $f \ll \lambda$. The dipole axis is along the direction $\theta = \pm\pi/2$

(Problems 12.3, 12.4, 12.5)

Interference from Linear Array of *N* Equal Sources

Figure 12.20 shows a linear array of N equal sources with constant separation f generating signals which are all in phase ($\delta_0 = 0$). At a distant point P in a direction θ from the sources the phase difference between the signals from two successive sources is given by

$$\delta = \frac{2\pi}{\lambda} f \sin\theta$$

and the resultant at P is found by superposing the equal contribution from each source with the constant phase difference δ between successive contributions.

But we found from Figure 1.11 that the resultant of such a superposition was given by

$$R = a\frac{\sin(N\delta/2)}{\sin(\delta/2)}$$

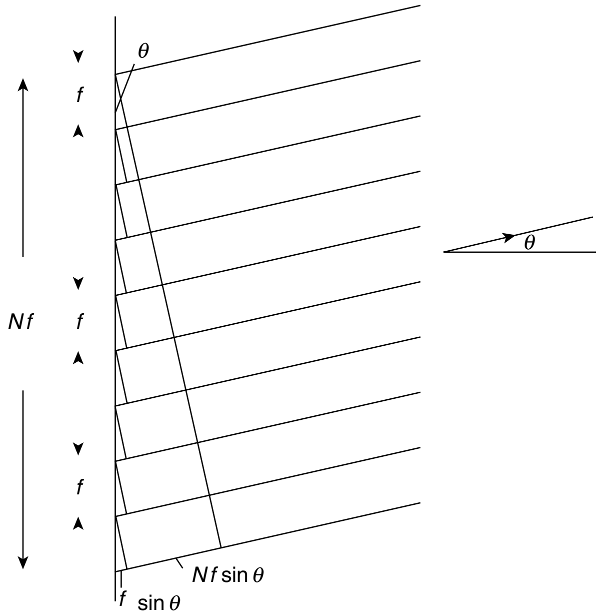


Figure 12.20 Linear array of N equal sources separation f radiating in a direction θ to a distant point P. The resulting amplitude at P (see Figure 1.11) is given by

$$R = a[\sin N(\delta/2)/\sin(\delta/2)]$$

where a is the amplitude from each source and

$$\delta = (2\pi/\lambda) f \sin\theta$$

is the common phase difference between successive sources

where a is the signal amplitude at each source, so the intensity may be written

$$I = R^2 = a^2 \frac{\sin^2(N\delta/2)}{\sin^2(\delta/2)} = I_s \frac{\sin^2(N\pi f \sin\theta/\lambda)}{\sin^2(\pi f \sin\theta/\lambda)}$$
$$= I_s \frac{\sin^2 N\beta}{\sin^2\beta}$$

where I_s is the intensity from each source and $\beta = \pi f \sin\theta/\lambda$.

If we take the case of $N = 2$, then

$$I = I_s \frac{\sin^2 2\beta}{\sin^2\beta} = 4I_s \cos^2\beta = 4I_s \cos^2\frac{\delta}{2}$$

which gives us the Young's Slit Interference pattern.

We can follow the intensity pattern for N sources by considering the behaviour of the term $\sin^2 N\beta/\sin^2\beta$.

We see that when

$$\beta = \frac{\pi}{\lambda}\sin\theta = 0 \pm \pi \pm 2\pi, \text{etc.}$$

i.e. when

$$f\sin\theta = 0,\ \pm\lambda,\ \pm 2\lambda \ldots \pm n\lambda$$

constructive interference of the order n takes place, and

$$\frac{\sin^2 N\beta}{\sin^2\beta} \to \frac{N^2\beta^2}{\beta^2} \to N^2$$

giving

$$I = N^2 I_{\mathrm{s}}$$

that is, a very strong intensity at the *Principal Maximum* condition of

$$f\sin\theta = n\lambda$$

We can display the behaviour of the $\sin^2 N\beta/\sin^2\beta$ term as follows

$$\begin{array}{rl} \text{Numerator } \sin^2 N\beta \text{ is zero for } N\beta \to & 0\pi \ldots N\pi \ldots 2N\pi \\ & \downarrow \quad\;\; \downarrow \quad\;\; \downarrow \\ \text{Denominator } \sin^2\beta \text{ is zero for } \beta \to & 0 \;\ldots\; \pi \;\ldots\; 2\pi \end{array}$$

The coincidence of zeros for both numerator and denominator determine the Principal Maxima with the factor N^2 in the intensity, i.e. whenever $f\sin\theta = n\lambda$.

Between these principal maxima are $N-1$ points of zero intensity which occur whenever the numerator $\sin^2 N\beta = 0$ but where $\sin^2\beta$ remains finite.

These occur when

$$f\sin\theta = \frac{\lambda}{N},\quad \frac{2\lambda}{N}\ldots(n-1)\frac{\lambda}{N}$$

The $N-2$ subsidiary maxima which occur between the principal maxima have much lower intensities because none of them contains the factor N^2. Figure 12.21 shows the intensity curves for $N = 2, 4, 8$ and $N \to \infty$.

Two scales are given on the horizontal axis. One shows how the maxima occur at the order of interference $n = f\sin\theta/\lambda$. The other, using units of $\sin\theta$ as the ordinate displays two features. It shows that the separation between the principal maxima in units of $\sin\theta$ is λ/f and that the width of half the base of the principal maxima in these units is λ/Nf (the same value as the width of the base of subsidiary maxima). As N increases not only does the principal intensity increase as N^2 but the width of the principal maximum becomes very small.

As N becomes very large, the interference pattern becomes highly directional, very sharply defined peaks of high intensity occurring whenever $\sin\theta$ changes by λ/f.

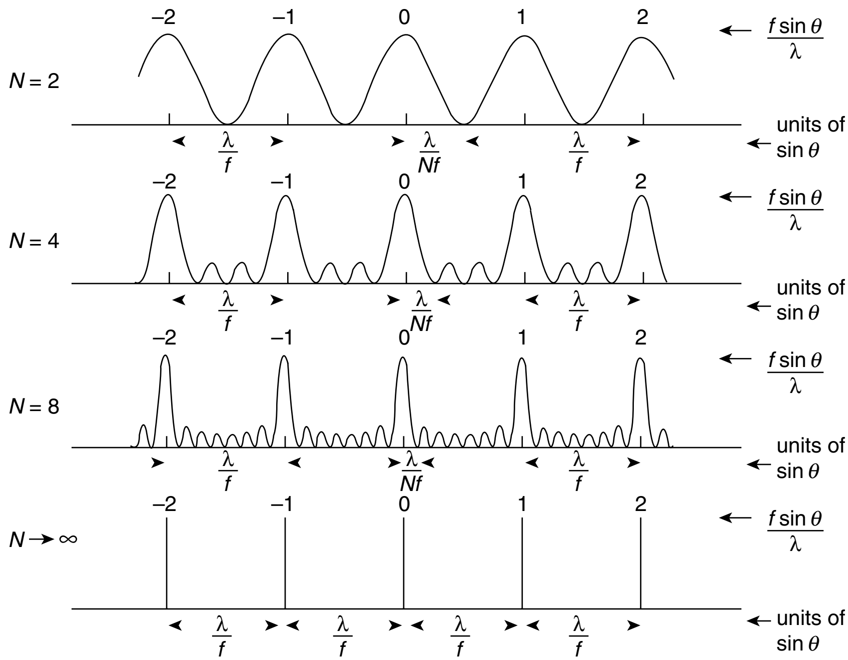


Figure 12.21 Intensity of interference patterns from linear arrays of N equal sources of separation f. The horizontal axis in units of $f \sin\theta/\lambda$ gives the spectral order n of interference. The axis in units of $\sin\theta$ shows that the separation between principal maxima is given by $\sin\theta = \lambda/f$ and the half-width of the principal maximum is given by $\sin\theta = \lambda/Nf$

The directional properties of such a linear array are widely used in both transmitting and receiving aerials and the polar plot for $N = 4$ (Figure 12.22) displays these features. For N large, such an array, used as a receiver, forms the basis of a radio telescope where the receivers (sources) are set at a constant (but adjustable) separation f and tuned to receive a fixed wavelength. Each receiver takes the form of a parabolic reflector, the axes of which are kept parallel as the reflectors are oriented in different directions. The angular separation between the directions of incidence for which the received signal is a maximum is given by $\sin\theta = \lambda/f$.

(Problems 12.6, 12.7)

Diffraction

Diffraction is classified as Fraunhofer or Fresnel. In Fraunhofer diffraction the pattern is formed at such a distance from the diffracting system that the waves generating the pattern may be considered as plane. A Fresnel diffraction pattern is formed so close to the diffracting system that the waves generating the pattern still retain their curved characteristics.

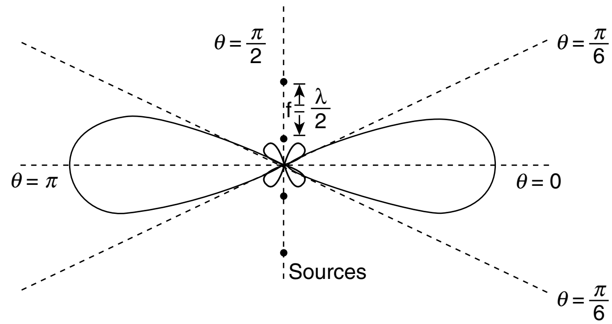


Figure 12.22 Polar plot of the intensity of the interference pattern from a linear array of four sources with common separation $f = \lambda/2$. Note that the half-width of the principal maximum is $\theta = \pi/6$ satisfying the relation $\sin\theta = \lambda/Nf$ and that the separation between principal maxima satisfies the relation that the change in $\sin\theta = \lambda/f$

Fraunhofer Diffraction

The single narrow slit. Earlier in this chapter it was stated that the difference between interference and diffraction is merely one of scale and not of physical behaviour.

Suppose we contract the scale of the N equal sources separation f of Figure 12.20 until the separation between the first and the last source, originally Nf, becomes equal to a distance d where d is now assumed to be the width of a narrow slit on which falls a monochromatic wavefront of wavelength λ where $d \sim \lambda$. Each of the large number N equal sources may now be considered as the origin of secondary wavelets generated across the plane of the slit on the basis of Huygens' Principle to form a system of waves diffracted in all directions.

When these diffracted waves are focused on a screen as shown in Figure 12.23 the intensity distribution of the diffracted waves may be found in terms of the aperture of the slit, the wavelength λ and the angle of diffraction θ. In Figure 12.23 a plane light wave falls normally on the slit aperture of width d and the waves diffracted at an angle θ are brought to focus at a point P on the screen PP_0. The point P is sufficiently distant from the slit for all wavefronts reaching it to be plane and we limit our discussion to *Fraunhofer Diffraction*.

Finding the amplitude of the light at P is the simple problem of superposing all the small contributions from the N equals sources in the plane of the slit, taking into account the phase differences which arise from the variation in path length from P to these different sources. We have already solved this problem several times. In Chapter 10 we took it as an example of the Fourier transform method but here we reapply the result already used in this chapter on p. 364, namely that the intensity at P is given by

$$I = I_s \frac{\sin^2 N\beta}{\sin^2 \beta} \quad \text{where} \quad N\beta = \frac{\pi}{\lambda} Nf \sin\theta$$

is half the phase difference between the contributions from the first and last sources. But now $Nf = d$ the slit width, and if we replace β by α where $\alpha = (\pi/\lambda)\, d \sin\theta$ is now half

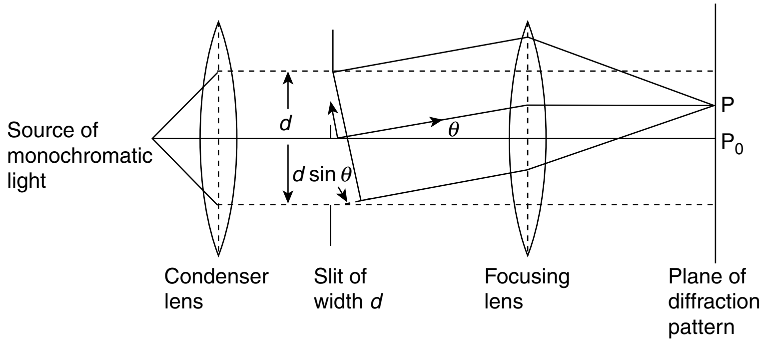


Figure 12.23 A monochromatic wave normally incident on a narrow slit of width d is diffracted through an angle θ and the light in this direction is focused at a point P. The amplitude at P is the superposition of all the secondary waves in the plane of the slit with their appropriate phases. The extreme phase difference from contributing waves at opposite edges of the slit is $\phi = 2\pi d \sin\theta/\lambda = 2\alpha$

the phase difference between the contributions from the opposite edges of the slit, the intensity of the diffracted light at P is given by

$$I = I_s = \frac{\sin^2(\pi/\lambda)d\sin\theta}{\sin^2(\pi/\lambda N)d\sin\theta} = I_s\frac{\sin^2\alpha}{\sin^2(\alpha/N)}$$

For N large

$$\sin^2\frac{\alpha}{N} \to \left(\frac{\alpha}{N}\right)^2$$

and we have

$$I = N^2 I_s\frac{\sin^2\alpha}{\alpha^2} = I_0\frac{\sin^2\alpha}{\alpha^2}$$

(recall that in the Fourier Transform derivation on p. 289,

$$I_0 = \frac{d^2h^2}{4\pi^2}$$

where h was the amplitude from each source).

Plotting $I = I_0(\sin^2\alpha/\alpha^2)$ with $\alpha = (\pi/\lambda)d\sin\theta$ in Figure 12.24 we see that its pattern is symmetrical about the value

$$\alpha = \theta = 0$$

where $I = I_0$ because $\sin\alpha/\alpha \to 1$ as $\alpha \to 0$. The intensity $I = 0$ whenever $\sin\alpha = 0$ that is, whenever α is a multiple of π or

$$\alpha = \frac{\pi}{\lambda}d\sin\theta = \pm\pi\ \pm 2\pi\ \pm 3\pi, \text{etc.}$$

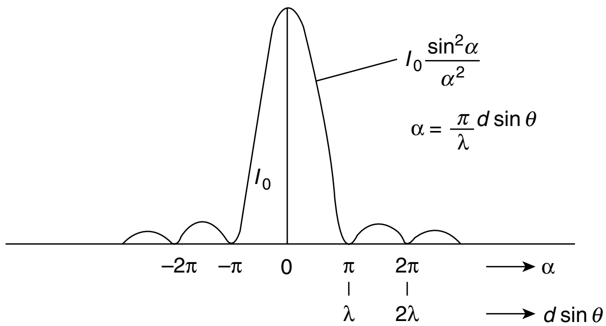


Figure 12.24 Diffraction pattern from a single narrow slit of width d has an intensity $I = I_0 \sin^2 \alpha/\alpha^2$ where $\alpha = \pi \mathrm{d} \sin\theta/\lambda$

giving

$$d \sin\theta = \pm\lambda \pm 2\lambda \pm 3\lambda, \text{etc.}$$

This condition for *diffraction minima* is the same as that for *interference maxima* between two slits of separation d, and this is important when we consider the problem of light transmission through more than one slit.

The intensity distribution maxima occur whenever the factor $\sin^2 \alpha/\alpha^2$ has a maximum; that is, when

$$\frac{\mathrm{d}}{\mathrm{d}\alpha}\left(\frac{\sin\alpha}{\alpha}\right)^2 = \frac{\mathrm{d}}{\mathrm{d}\alpha}\left(\frac{\sin\alpha}{\alpha}\right) = 0$$

or

$$\frac{\cos\alpha}{\alpha} - \frac{\sin\alpha}{\alpha^2} = 0$$

This occurs whenever $\alpha = \tan\alpha$, and Figure 12.25 shows that the roots of this equation are closely approximated by $\alpha = \pm 3\pi/2, \pm 5\pi/2$, etc. (see problem at end of chapter on exact values).

Table 12.1 shows the relative intensities of the subsidiary maxima with respect to the principal maximum I_0.

The rapid decrease in intensity as we move from the centre of the pattern explains why only the first two or three subsidiary maxima are normally visible.

Scale of the Intensity Distribution

The width of the principal maximum is governed by the condition $d \sin\theta = \pm\lambda$. A constant wavelength λ means that a decrease in the slit width d will increase the value of $\sin\theta$ and will widen the principal maximum and the separation between subsidiary maxima. The narrower the slit the wider the diffraction pattern; that is, in terms of a Fourier transform the narrower the pulse in x-space the greater the region in k- or wave number space required to represent it.

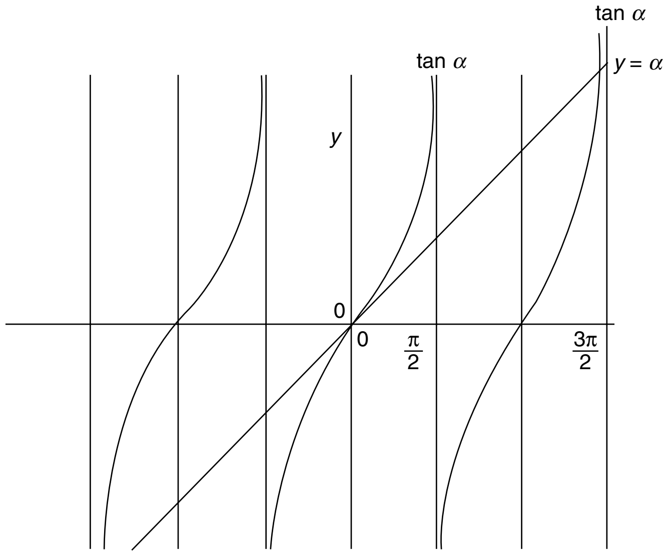


Figure 12.25 Position of principal and subsidiary maxima of single slit diffraction pattern is given by the intersections of $y = \alpha$ and $y = \tan\alpha$

Table 12.1

α	$\frac{\sin^2\alpha}{\alpha^2}$	$\frac{I_0 \sin^2\alpha}{\alpha^2}$
0	1	I_0
$\frac{3\pi}{2}$	$\frac{4}{9\pi^2}$	$\frac{I_0}{22.2}$
$\frac{5\pi}{2}$	$\frac{4}{25\pi^2}$	$\frac{I_0}{61.7}$
$\frac{7\pi}{2}$	$\frac{4}{49\pi^2}$	$\frac{I_0}{121}$

(Problems 12.8, 12.9)

Intensity Distribution for Interference with Diffraction from *N* Identical Slits

The extension of the analysis from the example of one slit to that of *N* equal slits of width *d* and common spacing *f*, Figure 12.26, is very simple.

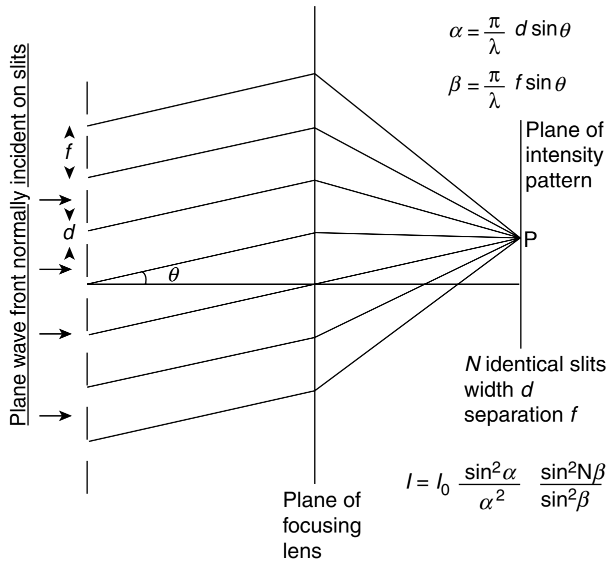


Figure 12.26 Intensity distribution for diffraction by N equal slits is

$$I = I_0 \frac{\sin^2 \alpha}{\alpha^2} \frac{\sin^2 N\beta}{\sin^2 \beta}$$

the product of the diffraction intensity for one slit, $I_0 \sin^2 \alpha/\alpha^2$ and the interference intensity between N sources $\sin^2 N\beta/\sin^2 \beta$, where $\alpha = (\pi/\lambda) d \sin\theta$ and $\beta = (\pi/\lambda) f \sin\theta$

To obtain the expression for the intensity at a point P of diffracted light from a single slit we considered the contributions from the multiple equal sources across the plane of the slit.

We obtained the result

$$I = I_0 \frac{\sin^2 \alpha}{\alpha^2}$$

by contracting the original linear array of N sources of spacing f on p. 364. If we expand the system again to recover the linear array, where each source is *now* a slit giving us the diffraction contribution

$$I_s = I_0 \frac{\sin^2 \alpha}{\alpha^2}$$

we need only insert this value at I_s in the original expression for the interference intensity,

$$I = I_s \frac{\sin^2 N\beta}{\sin^2 \beta}$$

on p. 364 where

$$\beta = \frac{\pi}{\lambda} f \sin\theta$$

to obtain, for the intensity at P in Figure 12.26, the value

$$I = I_0 \frac{\sin^2\alpha}{\alpha^2} \frac{\sin^2 N\beta}{\sin^2\beta},$$

where

$$\alpha = \frac{\pi}{\lambda} d \sin\theta$$

Note that this expression combines the diffraction term $\sin^2\alpha/\alpha^2$ for each slit (source) and the interference term $\sin^2 N\beta/\sin^2\beta$ from N sources (which confirms what we expected from the opening paragraphs on interference). The diffraction pattern for any number of slits will always have an envelope

$$\frac{\sin^2\alpha}{\alpha^2} \quad \text{(single slit diffraction)}$$

modifying the intensity of the multiple slit (source) interference pattern

$$\frac{\sin^2 N\beta}{\sin^2\beta}$$

Fraunhofer Diffraction for Two Equal Slits ($N = 2$)

When $N = 2$ the factor

$$\frac{\sin^2 N\beta}{\sin^2\beta} = 4\cos^2\beta$$

so that the intensity

$$I = 4I_0 \frac{\sin^2\alpha}{\alpha^2} \cos^2\beta$$

the factor 4 arising from N^2 whilst the $\cos^2\beta$ term is familiar from the double source interference discussion. The intensity distribution for $N = 2, f = 2d$, is shown in Figure 12.27. The intensity is zero at the diffraction minima when $d\sin\theta = n\lambda$. It is also zero at the interference minima when $f\sin\theta = (n + \frac{1}{2})\lambda$.

At some value of θ an *interference maximum* occurs for $f\sin\theta = n\lambda$ at the same position as a *diffraction minimum* occurs for $d\sin\theta = m\lambda$.

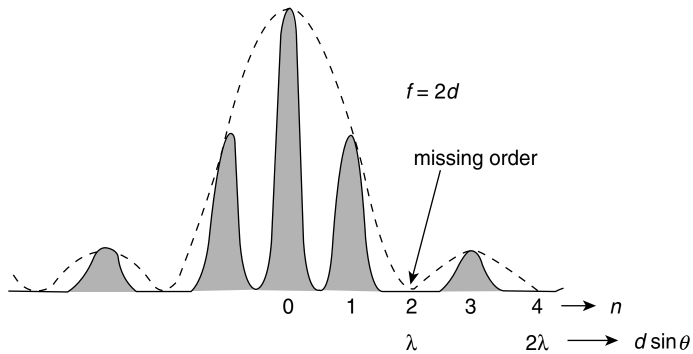


Figure 12.27 Diffraction pattern for two equal slits, showing interference fringes modified by the envelope of a single slit diffraction pattern. Whenever diffraction minima coincide with interference maxima a fringe is suppressed to give a 'missing order' of interference

In this case the diffraction minimum suppresses the interference maximum and the order n of interference is called a *missing order.*

The value of n depends upon the ratio of the slit spacing to the slit width for

$$\frac{n\lambda}{m\lambda} = \frac{f\sin\theta}{d\sin\theta}$$

i.e.

$$\frac{n}{m} = \frac{f}{d} = \frac{\beta}{\alpha}$$

Thus, if

$$\frac{f}{d} = 2$$

the missing orders will be $n = 2, 4, 6, 8$, etc. for $m = 1, 2, 3, 4$, etc.
The ratio

$$\frac{f}{d} = \frac{\beta}{\alpha}$$

governs the scale of the diffraction pattern since this determines the number of interference fringes between diffraction minima and the scale of the diffraction envelope is governed by α.

(Problem 12.10)

Transmission Diffraction Grating (*N* Large)

A large number N of equivalent slits forms a transmission diffraction grating where the common separation f between successive slits is called the *grating space*.

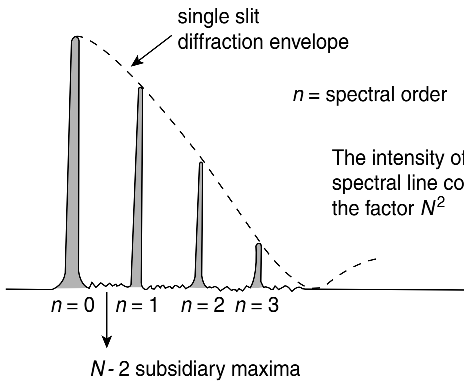


Figure 12.28 Spectral line of a given wavelength produced by a diffraction grating loses intensity with increasing order n as it is modified by the single slit diffraction envelope. At the principal maxima each spectral line has an intensity factor N^2 where N is the number of lines in the grating

Again, in the expression for the intensity

$$I = I_0 \frac{\sin^2 \alpha}{\alpha^2} \frac{\sin^2 N\beta}{\sin^2 \beta}$$

the pattern lies under the single slit diffraction term (Figure 12.28).

$$\frac{\sin^2 \alpha}{\alpha^2}$$

The principal interference maxima occur at

$$f \sin \theta = n\lambda$$

having the factor N^2 in their intensity and these are observed as *spectral lines* of order n. We see, however, that the intensities of the spectral lines of a given wavelength decrease with increasing spectral order because of the modifying $\sin^2 \alpha/\alpha^2$ envelope.

Resolving Power of Diffraction Grating

The importance of the diffraction grating as an optical instrument lies in its ability to resolve the spectral lines of two wavelengths which are too close to be separated by the naked eye. If these two wavelengths are λ and $\lambda + d\lambda$ where $d\lambda/\lambda$ is very small the *Resolving Power* for any optical instrument is given by the ratio $\lambda/d\lambda$.

Two such lines are just resolved, according to Rayleigh's Criterion, when the maximum of one falls upon the first minimum of the other. If the lines are closer than this their separate intensities cannot be distinguished.

If we recall that the spectral lines are the principal maxima of the interference pattern from many slits we may display Rayleigh's Criterion in Figure 12.29 where the *n*th order spectral lines of the two wavelengths are plotted on an axis measured in units of $\sin\theta$. We have already seen in Figure 12.21 that the half width of the spectral lines (principal maxima) measured in such units is given by λ/Nf where N is now the number of grating lines (slits) and f is the grating space. In Figure 12.29 the *n*th order of wavelength λ occurs when

$$f\sin\theta = n\lambda$$

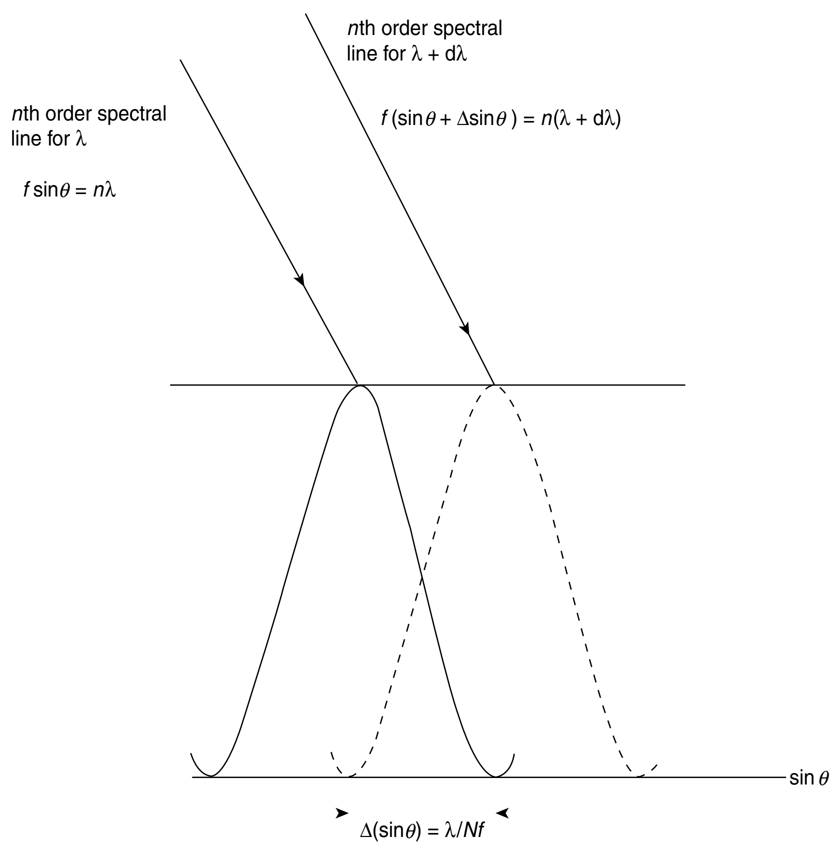


Figure 12.29 Rayleigh's criterion states that the two wavelengths λ and $\lambda + d\lambda$ are just resolved in the *n*th spectral order when the maximum of one line falls upon the first minimum of the other as shown. This separation, in units of $\sin\theta$, is given by λ/Nf where N is the number of diffraction lines in the grating and f is the grating space. This leads to the result that the resolving power of the grating $\lambda/d\lambda = nN$

whilst the nth order for $\lambda + d\lambda$ satisfies the condition

$$f[\sin\theta + \Delta(\sin\theta)] = n(\lambda + \mathrm{d}\lambda)$$

so that

$$f\Delta(\sin\theta) = n\,\mathrm{d}\lambda$$

Rayleigh's Criterion requires that the fractional change

$$\Delta(\sin\theta) = \frac{\lambda}{Nf}$$

so that

$$f\Delta(\sin\theta) = n\,\mathrm{d}\lambda = \frac{\lambda}{N}$$

Hence the Resolving Power of the diffraction grating in the nth order is given by

$$\frac{\lambda}{\mathrm{d}\lambda} = Nn$$

Note that the Resolving Power increases with the number of grating lines N and the spectral order n. A limitation is placed on the useful range of n by the decrease of intensity with increasing n due to the modifying diffraction envelope

$$\frac{\sin^2\alpha}{\alpha^2} \quad \text{(Fig. 12.28)}$$

Resolving Power in Terms of the Bandwidth Theorem

A spectral line in the nth order is formed when $f\sin\theta = n\lambda$ where $f\sin\theta$ is the path difference between light coming from two successive slits in the grating. The extreme path difference between light coming from opposite ends of the grating of N lines is therefore given by

$$Nf\sin\theta = Nn\lambda$$

and the time difference between signals travelling these extreme paths is

$$\Delta t = \frac{Nn\lambda}{c}$$

where c is the velocity of light.

The light frequency $\nu = c/\lambda$ has a resolvable differential change

$$|\Delta\nu| = c\frac{|\Delta\lambda|}{\lambda^2} = \frac{c}{Nn\lambda}$$

because $\Delta\lambda/\lambda = 1/Nn$ (from the inverse of the Resolving Power).

Hence

$$\Delta\nu = \frac{c}{Nn\lambda} = \frac{1}{\Delta t}$$

or $\Delta\nu\,\Delta t = 1$ (the Bandwidth Theorem).

Thus, the frequency difference which can be resolved is the inverse of the time difference between signals following the extreme paths

$$(\Delta\nu\,\Delta t = 1 \quad \text{is equivalent of course to } \Delta\omega\,\Delta t = 2\pi)$$

If we now write the extreme path difference as

$$Nn\lambda = \Delta x$$

we have, from the inverse of the Resolving Power, that

$$\frac{\Delta\lambda}{\lambda} = \frac{1}{Nn}$$

so

$$\frac{|\Delta\lambda|}{\lambda^2} = \Delta\left(\frac{1}{\lambda}\right) = \frac{\Delta k}{2\pi} = \frac{1}{Nn\lambda} = \frac{1}{\Delta x}$$

where the wave number $k = 2\pi/\lambda$.

Hence we also have

$$\Delta x\,\Delta k = 2\pi$$

where Δk is a measure of the resolvable wavelength difference expressed in terms of the difference Δx between the extreme paths.

On pp. 70 and 71 we discussed the quality factor Q of an oscillatory system. Note that the resolving power may be considered as the Q of an instrument such as the diffraction grating or a Fabry–Perot cavity for

$$\frac{\lambda}{\Delta\lambda} = \left|\frac{\nu}{\Delta\nu}\right| = \frac{\omega}{\Delta\omega} = Q$$

(Problems 12.11, 12.12, 12.13, 12.14)

Fraunhofer Diffraction from a Rectangular Aperture

The value of the Fourier transform method of Chapter 10 becomes apparent when we consider plane wave diffraction from an aperture which is finite in two dimensions.

Although Chapter 10 carried through the transform analysis for the case of only one variable it is equally applicable to functions of more than one variable.

In two dimensions, the function $f(x)$ becomes the function $f(x, y)$, giving a transform $F(k_x, k_y)$ where the subscripts give the directions with which the wave numbers are associated.

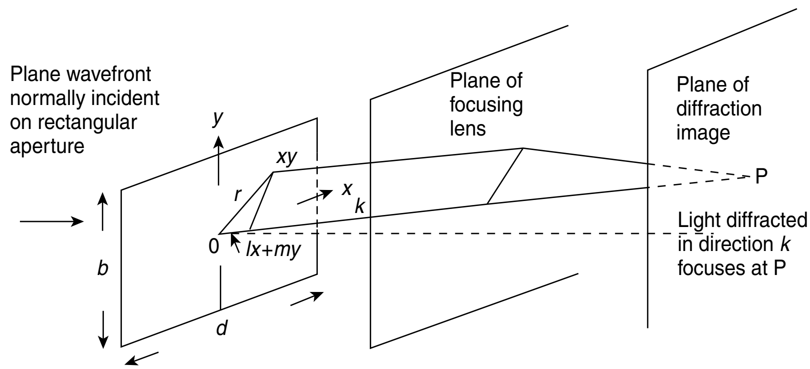


Figure 12.30 Plane waves of monochromatic light incident normally on a rectangular aperture are diffracted in a direction **k.** All light in this direction is brought to focus at P in the image plane. The amplitude at P is the superposition of contributions from all the typical points, x, y in the aperture plane with their appropriate phase relationships

In Figure 12.30 a plane wavefront is diffracted as it passes through the rectangular aperture of dimensions d in the x-direction and b in the y-direction. The vector **k**, which is normal to the diffracted wavefront, has direction cosines l and m with respect to the x- and y-axes respectively. This wavefront is brought to a focus at point P, and the amplitude at P is the superposition of the contributions from all points (x, y) in the aperture with their appropriate phases.

A typical point (x, y) in the aperture may be denoted by the vector **r**; the difference in phase between the contribution from this point and the central point O of the aperture is, of course, $(2\pi/\lambda)$ (path difference). But the path difference is merely the projection of the vector **r** upon the vector **k**, and the phase difference is $\mathbf{k} \cdot \mathbf{r} = (2\pi/\lambda)(lx + my)$, where $lx + my$ is the projection of **r** on **k**.

If we write

$$\frac{2\pi l}{\lambda} = k_x \quad \text{and} \quad \frac{2\pi m}{\lambda} = k_y$$

we have the Fourier transform in two dimensions

$$F(k_x, k_y) = \frac{1}{(2\pi)^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, \mathrm{e}^{-\mathrm{i}(k_x x + k_y y)}\, \mathrm{d}x\, \mathrm{d}y$$

where $f(x, y)$ is the amplitude of the small contributions from the points in the aperture.

Taking $f(x, y)$ equal to a constant a, we have $F(k_x, k_y)$ the amplitude in k-space at P

$$= \frac{a}{(2\pi)^2} \int_{-d/2}^{+d/2} \int_{-b/2}^{+b/2} \mathrm{e}^{-\mathrm{i}k_x x}\, \mathrm{e}^{-\mathrm{i}k_y y}\, \mathrm{d}x\, \mathrm{d}y$$

$$= \frac{a}{4\pi^2} bd \frac{\sin\alpha}{\alpha} \frac{\sin\beta}{\beta}$$

where

$$\alpha = \frac{\pi l d}{\lambda} = \frac{k_x d}{2}$$

and

$$\beta = \frac{\pi m b}{\lambda} = \frac{k_y b}{2}$$

Physically the integration with respect to y evaluates the contribution of a strip of the aperture along the y direction, and integrating with respect to x then adds the contributions of all these strips with their appropriate phase relationships.

The intensity distribution of the rectangular aperture is given by

$$I = I_0 \frac{\sin^2 \alpha}{\alpha^2} \frac{\sin^2 \beta}{\beta^2}$$

and relative intensities of the subsidiary maxima depend upon the product of the two diffraction terms $\sin^2 \alpha/\alpha^2$ and $\sin^2 \beta/\beta^2$.

These relative values will therefore be numerically equal to the product of any two terms of the series

$$\frac{4}{9\pi^2}, \quad \frac{4}{25\pi^2}, \quad \frac{4}{49\pi^2}, \quad \text{etc.}$$

The diffraction pattern from such an aperture together with a plan showing the relative intensities is given in Figure 12.31.

Fraunhofer Diffraction from a Circular Aperture

Diffraction through a circular aperture presents another two-dimensional problem to which the Fourier transform technique may be applied.

As in the case of the rectangular aperture, the diffracted plane wave propagates in a direction **k** with direction cosines l and m with respect to the x- and y-axes (Figure 12.32a). The circular aperture has a radius r_0 and any point in it is specified by polar coordinates (r, θ) where $x = r\cos\theta$ and $y = r\sin\theta$. This plane wavefront in direction **k** is focused at a point P in the plane of the diffraction pattern and the amplitude at P is the superposition of the contributions from all points (r, θ) in the aperture with their appropriate phase relationships. The phase difference between the contribution from a point defined (x, y) and that from the central point of the aperture is

$$\frac{2\pi}{\lambda}\,(\text{path difference}) = \frac{2\pi}{\lambda}(lx + my) = k_x x + k_y y \tag{12.6}$$

as with the rectangular aperture, so that the Fourier transform becomes

$$F(k_x k_y) = \frac{1}{(2\pi)^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, \mathrm{e}^{-\mathrm{i}(k_y x + k_y y)}\, \mathrm{d}x\, \mathrm{d}y \tag{12.7}$$

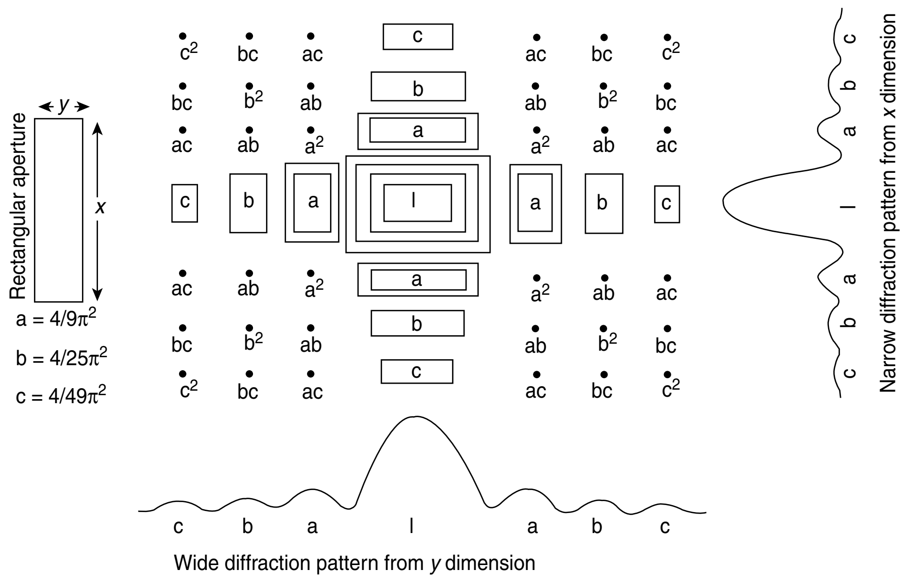


Figure 12.31 The distribution of intensity in the diffraction pattern from a rectangular aperture is seen as the product of two single-slit diffraction patterns, a wide diffraction pattern from the narrow dimension of the slit and a narrow diffraction pattern from the wide dimension of the slit. This 'rotates' the diffraction pattern through 90° with respect to the aperture

If we use polar coordinates, $f(x, y)$ becomes $f(r, \theta)$ and dx dy becomes $r\,\mathrm{d}r\,\mathrm{d}\theta$, where the limits of θ are from 0 to 2π. Moreover, because of the circular symmetry we may simplify the problem. The amplitude or intensity distribution along any radius of the diffraction pattern is sufficient to define the whole of the pattern, and we may choose this single radial direction conveniently by restricting **k** to lie wholly in the xz plane (Figure 12.32b) so that $m = k_y = 0$ and the phase difference is simply

$$\frac{2\pi}{\lambda} lx = k_x x = k_x r \cos\theta$$

Assuming that $f(r, \theta)$ is a constant amplitude a at all points in the circular aperture, the transform becomes

$$F(k_x) = \frac{a}{2\pi} \int_0^{2\pi} \mathrm{d}\theta \int_0^{r_0} \mathrm{e}^{-\mathrm{i}k_x r \cos\theta} r\,\mathrm{d}r \qquad (12.8)$$

This can be integrated by parts with respect to r and then term by term in a power series for $\cos\theta$, but the result is well known and conveniently expressed in terms of a *Bessel function* as

$$F(k_x) = \frac{a r_0}{k_x} J_1(k_x r_0)$$

where $J_1(k_x r_0)$ is called a Bessel function of the first order.

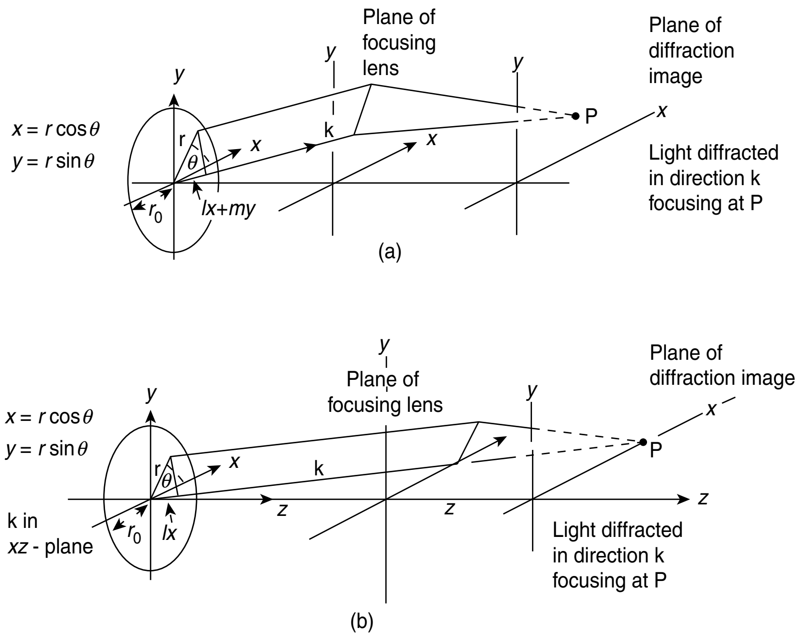


Figure 12.32 (a) A plane monochromatic wave diffracted in a direction **k** from a circular aperture is focused at a point P in the image plane. Contributions from all points x, y in the aperture superpose at P with appropriate phase relationships. (b) The direction **k** of (a) is chosen to lie wholly in the xz-plane to simplify the analysis. No generality is lost because of circular symmetry. The variation of the amplitude of diffracted light along any one radius determines the complete pattern

Bessel functions are series expansions which are analogous to sine and cosine functions. Where sines and cosines are those functions which satisfy rectangular boundary conditions defined in Cartesian coordinates, Bessel functions satisfy circular or cylindrical boundary conditions requiring polar coordinates.

Standing waves on a circular membrane, e.g. a drum, would require definition in terms of Bessel functions.

The Bessel function of order n is written

$$J_n(x) = \frac{x^n}{2^n n!}\left(1 - \frac{x^2}{2 \cdot 2n+2} + \frac{x^4}{2 \cdot 4 \cdot 2n+2 \cdot 2n+4} \cdots\right)$$

so that

$$J_1(x) = \frac{x}{2} - \frac{x^3}{2^2 4} + \frac{x^5}{2^2 4^2 6} - \frac{x^7}{2^2 4^2 6^2 8}$$

The expression $a^2 r_0^2 [J_1(k_x r_0)/k_x r_0]^2$, which measures the intensity along any radius of the diffraction pattern due to a circular aperture is normalized and plotted in Figure 12.33.

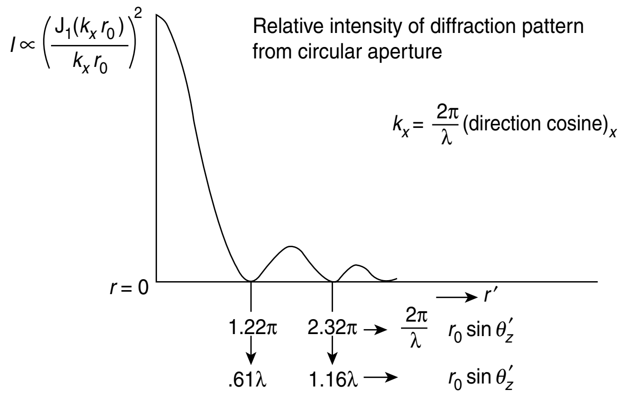


Figure 12.33 Intensity of the diffraction pattern from a circular aperture of radius r_0 versus r', the radius of the pattern. The intensity is proportional to $[J_1(k_x r_0)/k_x r_0]^2$, where J_1 is Bessel's function of order 1. The pattern consists of a central circular principal maximum surrounded by a series of concentric rings of minima and subsidiary maxima of rapidly diminishing intensity

$J_1(k_x r_0)$ has an infinite number of zeros, and the diffraction pattern is formed by an infinite number of light and dark concentric rings. The first dark band will occur at the first zero of $J_1(k_x r_0)$ which is given by $k_x r_0 = 1.219\pi$.

However,

$$k_x r_0 = \frac{2\pi}{\lambda} l r_0 = \frac{2\pi}{\lambda} r_0 \sin\theta_z'$$

where θ_z' is the angle between the vector $\mathbf{k}$ and the z-axis and defines the angle of diffraction. The first minimum therefore occurs at $r_0 \sin\theta_z' = 0.61\lambda$ and the next minimum at $r_0 \sin\theta_z' = 1.16\lambda$.

If the aperture were square with a side length $2r_0$ (the diameter of the circle) the first dark fringe would be at $r_0 \sin\theta_z' = 0.5\lambda$ and the second at $r_0 \sin\theta_z' = \lambda$.

As the radius of the circular aperture is reduced the value of θ_z' for the first minimum is increased and the whole pattern expands. This reminds us that a reduction of the pulse in x-space requires an increase in wave number or k-space to represent it.

We may write equation (12.8) as

$$F(k_x) = \frac{a}{2\pi} \int_0^{r_o} \int_0^{2\pi} \mathrm{e}^{-\mathrm{i}k_x \cdot r\cos\theta} r\,\mathrm{d}r d\theta$$

where $\int_0^{2\pi} \mathrm{e}^{-\mathrm{i}k_x \cdot r\cos\theta} d\theta = 2\pi J_0(k_x r)$ and J_0 is the Bessel function of order zero.

Then

$$F(k_x) = a \int_0^{r_0} J_0(k_x r) r dr$$

Now $J_1(k_x r)$ and $J_0(k_x r) r dr$ are related by

$$\int_0^{k_x r_0} J_0(k_x r) k_x r d(k_x r) = k_x r_0 J_1(k_x r_0)$$

giving

$$F(k_x) = a\pi r_0^2 \left[\frac{2J_1(k_x r_0)}{k_x r_0}\right]$$

where r_0 is the radius of the aperture.

The Intensity

$$I = I_0 \left[\frac{J_1(k_x r_0)}{k_x r_0}\right]^2$$

with the curve shown in Figure 12.33.

Fraunhofer Far Field Diffraction

If we remove the focusing lens in Figure 12.32 and leave the aperture open or place the lens within it we have the conditions for far field diffraction, Figure 12.34, where R'_0 the distance from $\tilde{O}$ to P' is $\gg$ distances in the aperture and image planes from the optic axis. The aperture is uniformly illuminated by a distant monochromatic source and a small area $d\tilde{s} = d\tilde{x}d\tilde{y}$ in the aperture is $\ll \lambda^2$, where λ is the wavelength.

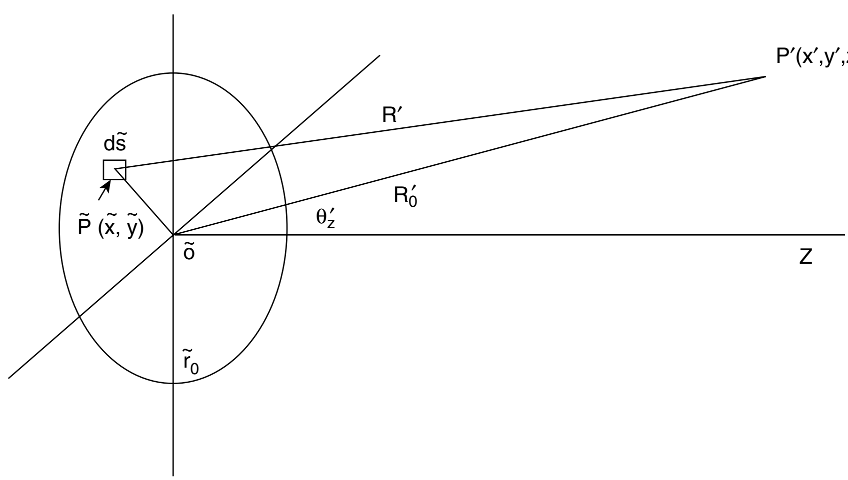


Figure 12.34 In Fraunhofer far field diffraction the distance from the aperture to the image point P' is $\gg$ distances in the aperture and image planes from the optic axis. The electric field at P' is the integral of the spherical waves from small areas $d\tilde{s}$ in the aperture plane and the resulting intensity pattern is that of Figure 12.33. It is known as the Airy disc

The electric field at P' due to the spherical wave from $d\tilde{s}$ is

$$dE_{P'} = \frac{\tilde{E}}{R'} e^{i\omega t - kR'} d\tilde{s}$$

Where $\tilde{E}e^{i\omega t}$ is the field at $d\tilde{s}$

Now

$$R'^2 = z'^2 + (x' - \tilde{x})^2 + (y' - \tilde{y})^2$$

and

$$R_0'^2 = z'^2 + x'^2 + y'^2$$

which combine to give

$$R' = R_0'[1 + (\tilde{x}^2 + \tilde{y}^2)/R_0'^2 - 2(x'\tilde{x} + y'\tilde{y})/R_0'^2]^{1/2}$$

and $R_0'^2 \gg (\tilde{x}^2 + \tilde{y}^2)$

so we write

$$R' = R_0'[1 - 2(x'\tilde{x} + y'\tilde{y})/R_0'^2]^{1/2}$$

and if we neglect higher terms

$$\begin{aligned} R' &= R_0'[1 - (x'\tilde{x} + y'\tilde{y})/R_0'^2] \\ &= R_0' - \frac{x'\tilde{x}}{R_0'} - \frac{y'\tilde{y}}{R_0'} \end{aligned}$$

We use this value for R' in the expression for $dE_{p'}$ to give the total field at P' as

$$E_{P'} = \frac{\tilde{E}e^{i\omega t - kR_0'}}{R_0'} \int\int_{\text{aperture}} e^{ik\frac{(x'\tilde{x} + y'\tilde{y})}{R_0'}} d\tilde{s}$$

Comparison with equation (12.6) shows that $k\tilde{x}/R_0' = kl$ and $k\tilde{y}/R_0' = km$ of that equation and proceeding via polar co-ordinates we obtain the same value for the intensity of the diffraction pattern,

i.e.

$$I = I_0 \left(\frac{J_1(kr_0 \sin\theta_2')}{kr_0 \sin\theta_2'} \right)^2 \qquad \text{in Figure 12.33}$$

This far field diffraction pattern is known as the Airy disc, Figure 12.35, and its size places a limit on the resolving power of a telescope. When the two components of a double star with an angular separation $\Delta\phi$ are viewed through a telescope with an objective lens of focal length l and diameter d their images will appear as two Airy discs separated by the angle $\Delta\phi$. The two diffraction patterns will be resolved if $\Delta\phi$ is much wider than the

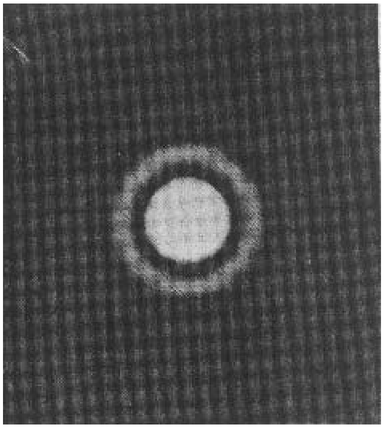


Figure 12.35 Photograph of an Airy disc showing the central bright disc, the first dark ring and the first subsidiary maximum. Compare this with Figure 12.33

angluar width of a disc but not if it is much less. Lord Rayleigh's criterion (Figure 12.29) gives the critical angle $\Delta\phi$ for resolution as that when the maximum of one disc falls on the first minimum of the other λ, Figure 12.36. Figure 12.33 then gives

$$\Delta\phi = \frac{0.61\lambda}{r_0} = \frac{1.22\lambda}{d}$$
$$(\Delta\phi = \Theta'_z \quad \text{in Figure 12.33})$$

where λ is the rediated wavelength.

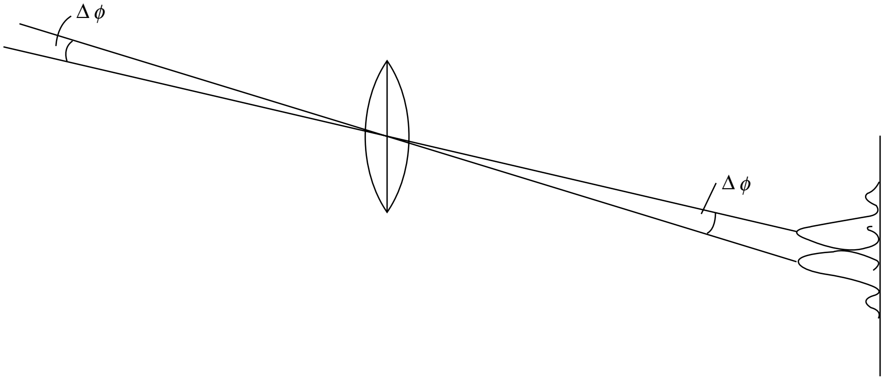


Figure 12.36 Two stars with angular separation $\Delta\phi$ form separate Airy disc images when viewed through a telescope. Rayleigh's criterion (Figure 12.29) states that the these images are resolved when the central maximum of one falls upon the first minimum of the other

This condition is known as diffraction-limited resolution. A poor quality lens will introduce aberrations and will not meet this criterion.

The Michelson Stellar Interferometer

In the discussion on Spatial Coherence (p. 360) we saw that the relative displacement of the interference fringes from separate sources 1 and 2 led to a partial loss of the visibility of the fringes defined as

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

and eventually when the displacement was equal to half a fringe width $V = 0$ and there was a complete loss of contrast.

Michelson's Stellar Interferomenter (1920) used this to measure the angular separation between the two components of a double star or, alternatively, the angular width of a single star.

Initially, we take the simplest case to illustrate the principle and then discuss the practical problems which arise. We assume in the first instance that light from the stars is monochromatic with a wavelenght λ_0. Michelson used four mirros M_1 M_2 M_3 M_4 mounted on a girder with two slits S_1 and S_2 in front of the lens of an astronomical telescope, Figure 12.37. The slits were perpendicular to the line joining the two stars. The separation h of the outer pair of mirrors ($\sim$meters) was increased until the fringes observed in the focal plane of the objective just disappeared. Assuming zero path difference between M_1M_2 P_0 and M_4 M_3 P_0 the light from star A will form its zero order fringe maximum at P_0 and its first order fringe maximum at P_1, due to a path difference $S_2N = d\sin\theta = \lambda_0$ so the fringe spacing is determined by d, the separation between the inner mirrors M_2 and M_3.

The condition for fringe disappearance is that rays from star B will form a first order maximum fringe midway between P_0 and P_1, that is, when

$$C\mathrm{M}_1\mathrm{M}_2\mathrm{S}_1P_0 - \mathrm{M}_4\mathrm{M}_3\mathrm{S}_2P_0 = C\mathrm{M}_1 = h\sin\phi = \lambda_0/2$$

The condition for fringe disappearnce is therefore determined by h while the angular size of the fringes depends on d so there is an effective magnification of h/d over a fringe system produced by the slits alone.

The angles θ and ϕ are small and the minimum value of h is found which produces $V = 0$ so that the fringes disappear at

$$h\phi = \lambda_0/2 \qquad \text{or} \qquad h = \frac{\lambda}{2\phi}$$

Measurement of h thus determines the double-star angular separation.

Several assumptions have been made in this simple case presentation. First, that the intensities of the light radiated by the stars are equal and that they are coherent soruces. In

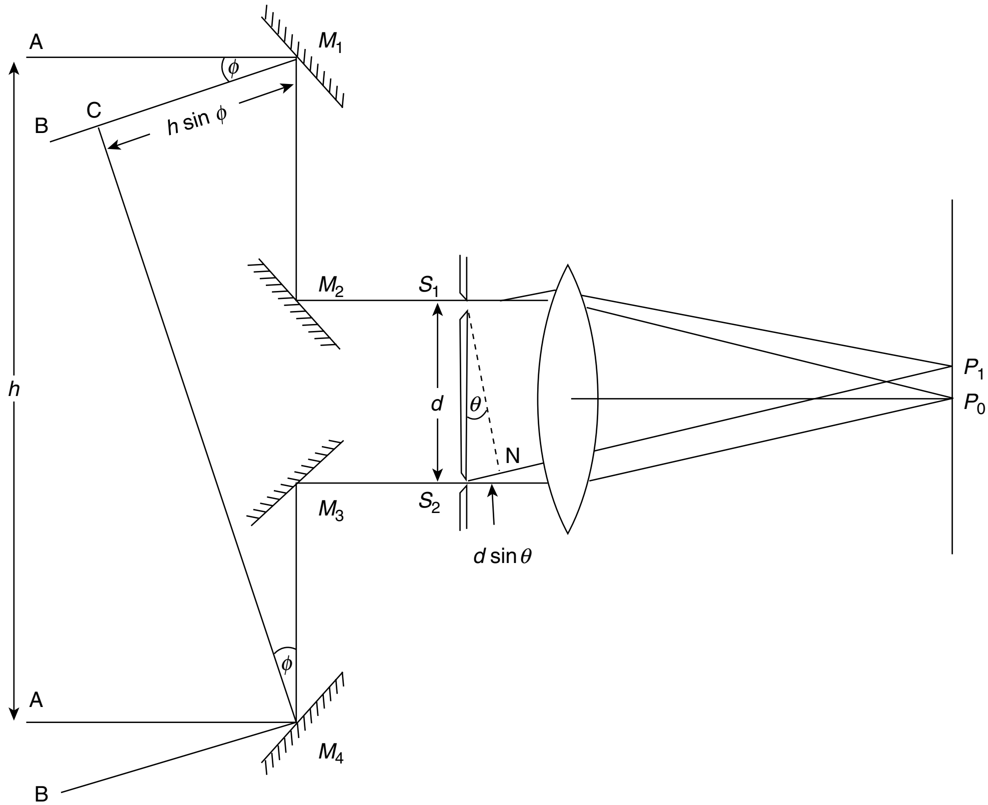


Figure 12.37 In the Michelson stellar interferometer light from stars A and B strike the movable outer mirrors M_1 and M_4 to be reflected via fixed mirrors M_2 and M_3 through two slits S_1 and S_2 and a lens to form interference fringes. Light from Star A forms its zero order fringe at P_0 and its first order fringe at P_1 when $S_2N = d\sin\theta = \lambda_0$. The minimum separation h of M_1M_4 is found for light from B to reduce the fringe visibility to zero, that is, when the path difference $h = \sin\phi = \lambda_0/2$. The angles are so small that θ and ϕ replace their sines. Note that the fringe separation depends on d, but the fringe visibility is governed by h

fact, even if the sources are incoherent their radiation is essentially coherent at the interferometer. Second, the radiation is not monochromatic and only a few fringes around the zero order were visible so λ_0 must be taken as a mean wavelength. Finally, the introduction of a lens into the system inevitably creates Airy discs and the visibility must be expressed in terms of the Airy disc intensity distribution. This results in

$$V = 2\left(\frac{J_1(u)}{u}\right)$$

where

$$u = \pi h\phi/\lambda_0$$

If this visibility is plotted against $h\phi/\lambda_0$ its first zero occurs at 1.22 so the fringes disappear when $h = 1.22\,\lambda_0/\phi$.

In fact, Michelson first used his interferometer in 1920 to measure the angular diameter of the star Betelgeuse the colour of which is orange. His astronomical telescope was the 2.54 m (100 in.) telescope of the Mt. Wilson Observatory. A mean wavelength $\lambda_0 = 570 \times 10^{-9}$m was used and the fringes vanished when $h = 3.07$ m to give an angular diameter $\phi = 22.6 \times 10^{-8}$ radians or 0.047 arc seconds. The distance of Betelgeuse from the Earth was known and its diameter was calculated to be about 384×10^6km, roughly 280 times that of the Sun. This magnitude is greater than that of the orbital diameter of Mars around the Sun.

The Convolution Array Theorem

This is a very useful application of the Convolution Theorem p. 297 5th edn, when one of the members is the sum of a series of δ functions.

e.g.

$$\begin{aligned} g(x) &= f_1(x) \otimes \sum_m \delta(x - x_{\mathrm{m}}) \\ &= \int_{-\infty}^{\infty} f_1(x') \sum_m \delta(x - x' - x_m)\mathrm{d}x' \\ &= \sum_m f_1(x - x_m) \end{aligned}$$

This is a linear addition of functions each of the form $f_1(x)$ but shifted to new origins at $x_m (m = 1, 2, 3 \ldots)$, Figure 12.38.

The convolution theorem gives the Fourier Transform of $g(x)$ as

$$F[g(x)] = F[f_1(x)]F\left[\sum_m \delta(x - x_m)\right]$$

i.e.

$$F(k_x) = F_1[f_1(x)] \sum_m \mathrm{e}^{-ik_x x_m}$$

so the transform of the spatially shifted local function is just the product of the transform of the local function and a phase factor.

This is the Array Theorem which we now apply in a more rigorous approach to the effect of diffraction on the interference fringes in Young's slit experiment (p. 358) where the illuminating source is equidistant from both slits.

The Array Theorem may be applied to any combination of identical apertures but Young's experiment involves only the two rectangular (slits) pulses in Figure 12.39a. Here, $f_1(x)$ is a rectangular pulse of width d and the x_m values above are $x_m = \pm\, a/2$.

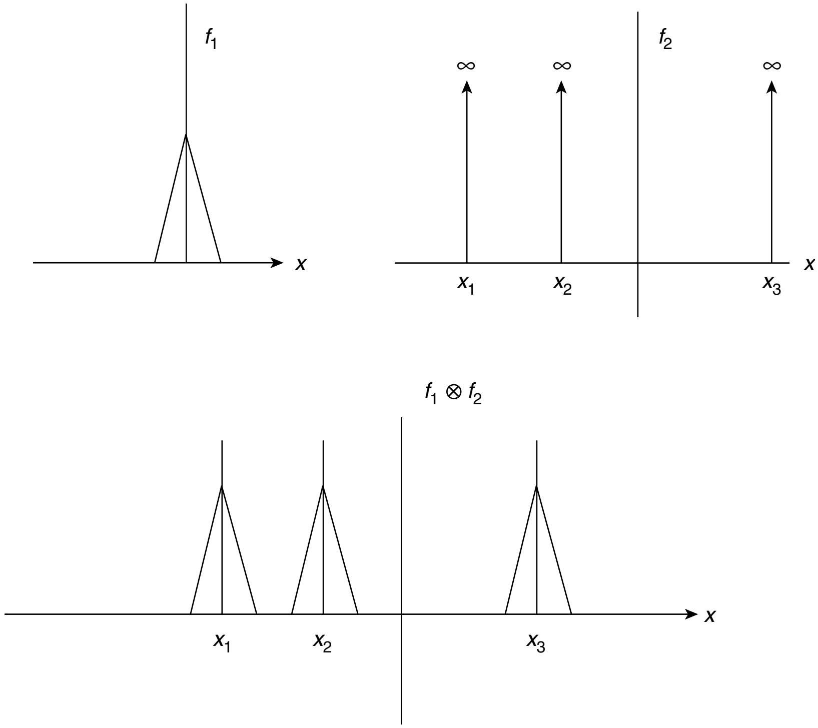


Figure 12.38 In the convolution array theorem a function $f_1(x)$ is convolved with a series of Dirac functions which shift it to new origins

Thus, we have the transform amplitude

$$F(k_x) = F_1(k_x) \sum_m e^{-ik_x x_m}$$

where $k_x = k \cdot x = kx \sin\theta$ and k in Figure 13.39b is the vector direction from $x = -a/2$ to a point P on the diffraction-interference pattern. p. 288 gives

$$F_1(k_x) \propto \frac{\sin\alpha}{\alpha}$$

where

$$\alpha = \frac{\pi}{\lambda} d \sin\theta$$

The second term, a phase factor, is

$$\sum_m e^{-ik_x x_m} = [e^{ik_x a/2} + e^{-ik_x a/2}] = 2\cos k_x a/2$$

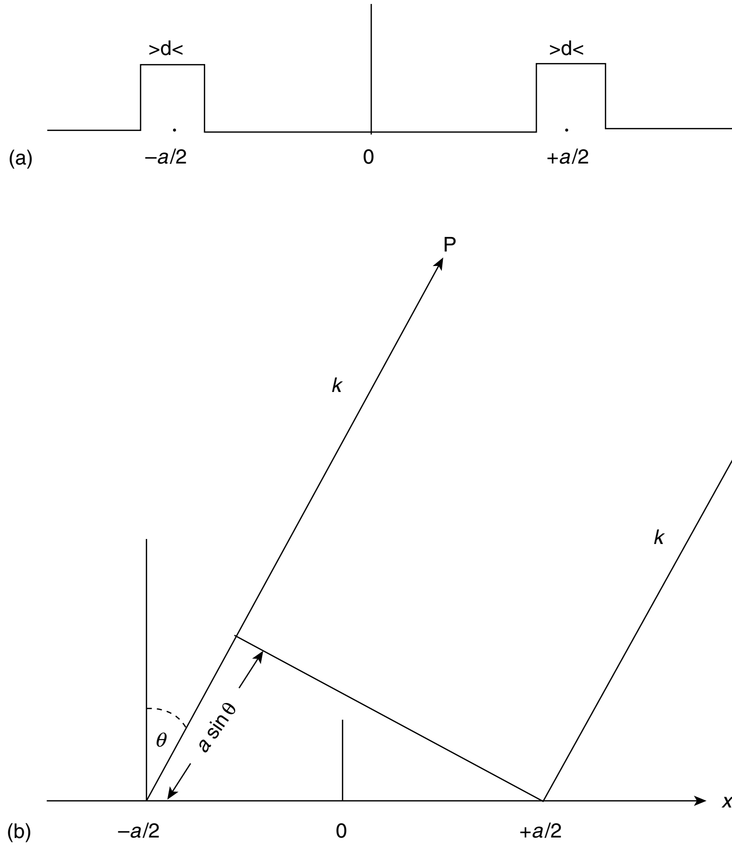


Figure 12.39 Young's double slit experiment represented in convolution array theorem (a) by two reactangular pulses and (b) with a path difference in the direction **k** of $d \sin \phi$ where a is the separation between the pulse centres

We may equate $k_x a/2$ with $\delta/2$ on p. 358 where $\delta = \frac{2\pi}{\lambda}(x_2 - x_1)$ is the phase difference at point P due to the path difference from the two sources. Here, $k_x a/2 = ka \sin\theta/2 = \pi a \sin\theta/\lambda$ (Figure 13.39b). When $\cos k_x a/2 = 1$ for maximum constructive interference

$$ka \sin\theta/2 = \frac{\pi}{\lambda} a \sin\theta = n\pi$$

i.e.

$$a \sin\theta = n\lambda$$

The amplitude squared or intensity is, therefore

$$I \propto \frac{\sin^2\alpha}{\alpha^2} 4\cos^2(\delta/2)$$

a cos^2 interference system modulated by a diffraction envelope as shown in Figure 12.27

This method can be extended to produce the pattern for a diffraction grating of N identical slits.

The Optical Transfer Function

The modern method of testing an optical system, e.g. a lens, is to consider the object as a series of Fourier frequency components and to find the response of the system to these frequencies. A test chart with a sinusoidal distribution of intensity would make a suitable object for this purpose. The function of the lens or optical system is considered to be that of a linear operator which transforms a sinusoidal input into an undistorted sinusoidal output.

The linear operator is defined in terms of the Optical Transfer Function (OTF) which may be real or complex. The real part, the Modulation Transfer Function (MTF), measures the effect of the lens on the amplitude of the sinusoidal input; the complex element is the Phase Transfer Function (PTF), a shift in phase when aberrations are present. If there are no aberrations and the effect on the image is limited to diffraction the PTF is zero.

Changing the amplitude of the object frequency components affects the contrast between different parts of the image compared with the corresponding parts of the object. We shall evaluate this effect at the end of the analysis.

We shall assume that the object is space invariant and incoherent. Space invariance means that the only effect of moving a point source over the object is to change the location of the image. When an object is incoherent its intensity or irradiance varies from point to point and all contributions to the final image are added under the integral sign.

Over a small area $\mathrm{d}x\,\mathrm{d}y$ of the object the radiated flux will be $I_0(x, y)\mathrm{d}x\,\mathrm{d}y$ and this makes its contribution to the image intensity. In addition, every point source on the object creates a circular diffraction pattern (Airy disc) around the corresponding image point so the resulting intensity of the image at (x', y') will be

$$\mathrm{d}\,I'(x', y') = I_0(x, y)O(x, y, x'y')\mathrm{d}x\,\mathrm{d}y$$

where $O(x, y, x'y')$ is the radially symmetric intensity distribution of the diffraction pattern (Airy disc). In this context it is called the Point Spread Function (PSF).

Adding all contributions gives the image intensity

$$I'(x', y') = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I_{\mathrm{o}}(x, y)O(x, y, x'y')\mathrm{d}x\,\mathrm{d}y$$

If, as we shall assume for simplicity, the magnification is unity, there is a one-to-one correspondence between the point (x, y) on the object and the centre of its diffraction pattern in the image plane. Using (x, y) as the coordinate of this centre the value of $O(x, y, x', y')$ at any other point (x', y') in the diffraction pattern is given by

$$O(x' - x,\ y' - y)$$

Thus, the intensity or irradiance at any image point may be written

$$I'(x', y') = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I_0(x, y) O(x' - x, y' - y) \mathrm{d}x \, \mathrm{d}y$$

This is merely the two-dimensional form of the convolution we met on p. 293 and we reduce it to one dimension by writing

$$I'(x') = \int_{-\infty}^{\infty} I_0(x) \, O(x' - x) \mathrm{d}x = \int_{-\infty}^{\infty} I_0(x' - x) \, O(x) \mathrm{d}x$$

because the convolution theorem of p. 297 allows us to exchange the variables of the functions under the convolution integral.

This is evidently of the form

$$I' = I_0 \otimes O$$

with Fourier Transforms

$$F(I') = F(I_0) \cdot F(O)$$

The choice of one dimension which adds clarity to the following analysis tranforms the PSF to a Line Spread Function (LSF) by cutting a narrow slice from the three-dimensional PSF. This is achieved by using a line source represented by a Dirac δ function, the sifting property of which isolates an infinitesimally narrow section of the PSF.

The shape of the three-dimensional PSF may be imagined by rotating Figure 12.33 about its vertical axis for a complete revolution. The profile of a slice along the diameter through the centre of the PSF is then the intensity of Figure 12.33 together with its reflection about the vertical axis. Any other slice, not through the centre, will have a similar profile but will differ in some details, e.g. its minimum values will not be zero, Figure 12.40.

Thus, in one dimension, replacing $O(x)$ by $L(x)$ the LSF, we have

$$I'(x') = \int_{-\infty}^{\infty} I_0(x' - x) \, L(x) \mathrm{d}x$$

or

$$I' = I_0 \otimes L = L \otimes I_0$$

with

$$F(I') = F(I_0) \cdot F(L) = F(L) \cdot F(I_0)$$

Let us write the intensity distribution of an object frequency component in one dimension as $a + b\cos k_x x$, where b modulates the cosine and a is a positive d.c. bias greater than b so

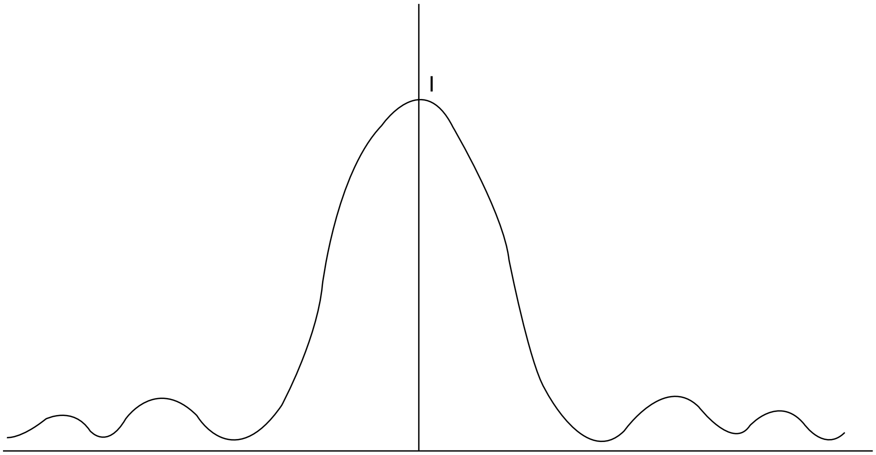


Figure 12.40 The profile of the Line Spread Function $L(x)$ is formed by cutting an off-centre slice from the three-dimensional Point Spread Function: $L(x)$ is the area under the curve. Note that the minimum values of $L(x)$ are non-zero, unlike the curve of Figure 12.33

that the intensity is always positive. Then, in the convolution above

$$I_0 = a + b\cos k_x(x' - x)$$

and the image intensity at x' is

$$\begin{aligned} I'(x') &= \int_{-\infty}^{\infty} [a + b\cos k_x(x' - x)]L(x)\,\mathrm{d}x \\ &= \int_{-\infty}^{\infty} L(x)[a + b\cos k_x(x' - x)\,\mathrm{d}x \end{aligned}$$

We remove the x' terms from the integral by expanding the cosine term to give

$$I'(x') = a\int_{-\infty}^{\infty} L(x)\mathrm{d}x + b\cos k_x x' \int_{-\infty}^{\infty} L(x)\cos k_x x\mathrm{d}x + b\sin k_x x' \int_{-\infty}^{\infty} L(x)\sin k_x x\,\mathrm{d}x \tag{12.9}$$

The integrals in the second and third terms on right-hand side of this equation are, repectively, the cosine and sine Fourier transforms from pp. 285, 286.

If we write

$$C(k_x) = \int_{-\infty}^{\infty} L(x)\cos k_x x\mathrm{d}x$$

and

$$S(k_x) = \int_{-\infty}^{\infty} L(x)\sin k_x x\mathrm{d}x$$

we have

$$C(k_x) - i\,S(k_x) = \int_{-\infty}^{\infty} L(x)e^{-ik_x x}dx = F(L_x) = M(k_x)e^{-i\phi(k_x)}$$

where

$$M(k_x) = [C(k_x)^2 + S(k_x)^2]^{1/2}$$

is the MTF and $e^{-i\phi(k_x)}$ is the PTF with

$$\tan\phi = S(k_x)/C(k_x)$$

The OTF is, therefore, the Fourier transform of the LSF.

If the LSF is symmetrical, as in the case of the diffraction pattern, the odd terms in $S(k_x)$ are zero, so the phase change $\phi = 0$ and the OTF is real.

For a given frequency component n we can normalize $L(x)$ to give

$$L_n(x) = \frac{L(x)}{\int_{-\infty}^{\infty} L_n(x)\mathrm{d}x} = 1$$

so that equation (12.9) becomes

$$\begin{aligned} I'(x') &= a + M(k_x)b(\cos k_x x'\cos\phi - \sin k_x x'\sin\phi) \\ &= a + M(k_x)b(\cos k_x x' + \phi) \end{aligned}$$

In the absence of aberrations, that is, in the symmetric diffraction limited case, $\phi = 0$. I_0 is shown in Figure 12.41(a) and $I'(x')$ in Figure 12.41(b) where $\phi \neq 0$ due to aberrations.

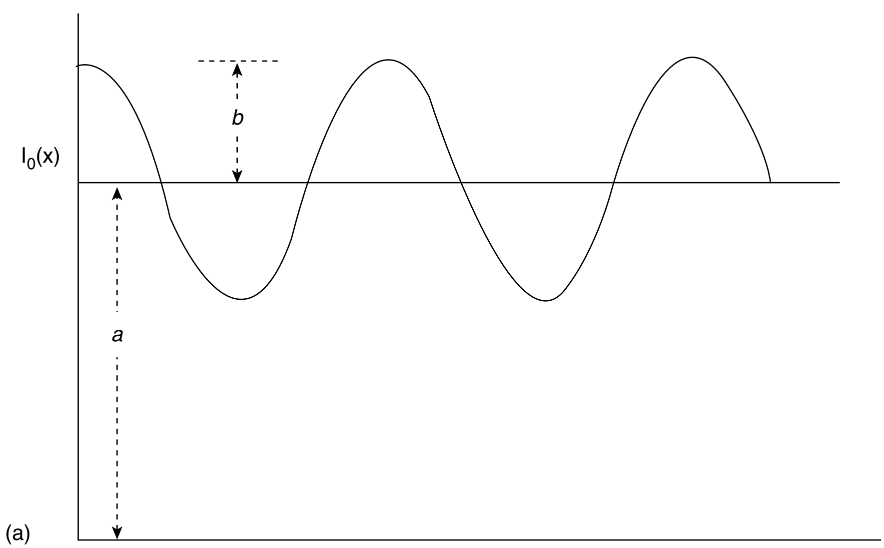


Figure 12.41 (a) The object frequency component $a + b\cos k_x x$ is modified by the Optical Transfer Function

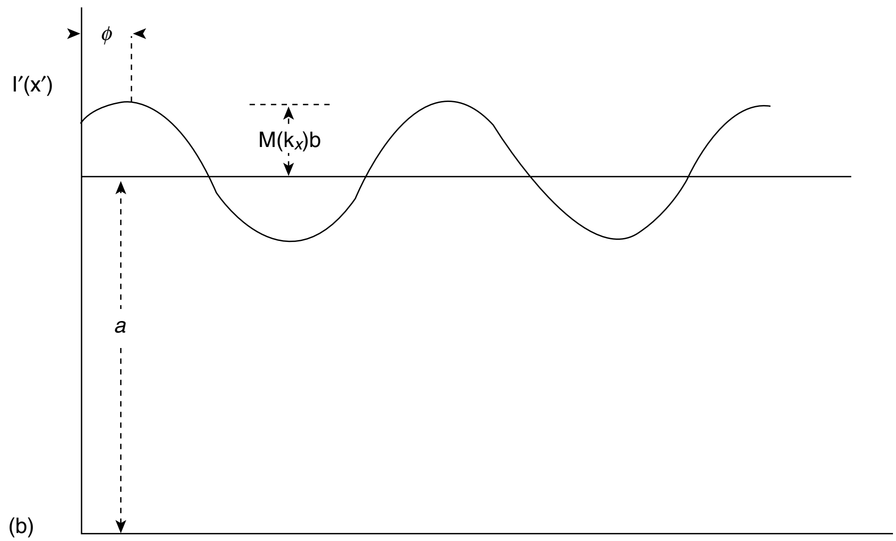


Figure 12.41 (b) In the image component $a + M(k)b\cos\,(k_x x' + \phi)$, $M(k)$ is the Modulation Transfer Function, which is < 1 and the phase change ϕ results from aberrations. The contrast in the image is less than that in the object. Note that in (b) ϕ is negative in the expression $\cos(k_x x' + \phi)$

The effect of the MTF on the amplitude of the frequency components is to reduce the contrast between parts of the image compared with corresponding parts of the object.

We have already met an expression for the contrast which we called Visibility on p. 360. Thus, we can write

$$\text{Contrast} = \frac{I_{\text{max}} - I_{\text{min}}}{I_{\text{mix}} + I_{\text{min}}} = \frac{(a+b)-(a-b)}{(a+b)+(a-b)} = \frac{b}{a} \qquad \text{for the object}$$

The image contrast $M(k_x)b/a < b/a$ so the image contrast is less than that of the object.

Fresnel Diffraction

The Straight Edge and Slit

Our discussion of Fraunhofer diffraction considered a plane wave normally incident upon a slit in a plane screen so that waves at each point in the plane of the slit were in phase. Each point in the plane became the source of a new wavefront and the superposition of these wavefronts generated a diffraction pattern. At a sufficient distance from the slit the superposed wavefronts were plane and this defined the condition for Fraunhofer diffraction. Its pattern followed from summing the contributions from these waves together with their relative phases and on p. 21 we saw that these formed an arc of constant length. When the contributions were all in phase the arc was a straight line but as the relative phases increased the arc curved to form *closed* circles of decreasing radii. The length of the chord joining the ends of the arc measured the resulting amplitude of the superposition and the square of that length measured the light intensity within the pattern.

Nearer the slit where the superposed wavefronts are not yet plane but retain their curved character the diffraction pattern is that of Fresnel. There is no sharp division between Fresnel and Fraunhofer diffraction, the pattern changes continuously from Fresnel to Fraunhofer as the distance from the slit increases.

The Fresnel pattern is determined by a procedure exactly similar to that in Fraunhofer diffraction, an arc of constant length is obtained but now it convolutes around the arms of a pair of joined spirals, Figure 12.42, and not around closed circles.

An understanding of Fresnel diffraction is most easily gained by first considering, not the slit, but a straight edge formed by covering the lower half of the incident plane wavefront with an infinite plane screen. The undisturbed upper half of the wavefront will contribute one half of the total spiral pattern, that part in the first quadrant.

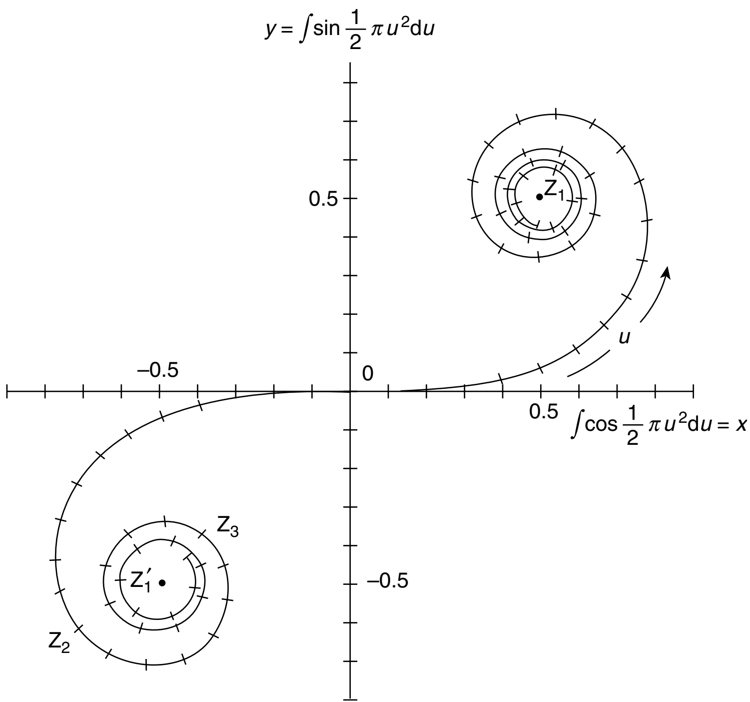


Figure 12.42 Cornu spiral associated with Fresnel diffraction. The spiral in the first quadrant represents the contribution from the upper half of an infinite plane wavefront above an infinite straight edge. The third quadrant spiral results from the downward withdrawal of the straight edge. The width of the wavefront contributing to the diffraction pattern is correlated with the length u along the spiral. The upper half of the wavefront above the straight edge contributes an intensity $(OZ_1)^2$ which is the square of the length of the chord from the origin to the spiral eye. This intensity is 0.25 of the intensity $(Z_1Z_1')^2$ due to the whole wavefront

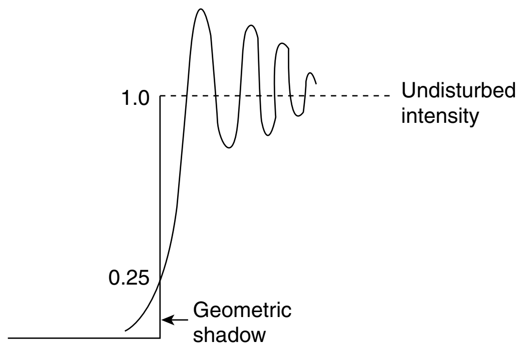


Figure 12.43 Fresnel diffraction pattern from a straight edge. Light is found within the geometric shadow and fringes of varying intensity form the observed pattern. The intensity at the geometric shadow is 0.25 of that due to the undisturbed wavefront

The Fresnel diffraction pattern from a straight edge, Figure 12.43, has several significant features. In the first place light is found beyond the geometric shadow; this confirms its wave nature and requires a Huygens wavelet to contribute to points not directly ahead of it (see the discussion on p. 305). Also, near the edge there are fringes of intensity greater and less than that of the normal undisturbed intensity (taken here as unity). On this scale the intensity at the geometric shadow is exactly 0.25.

To explain the origin of this pattern we consider the point O at the straight edge of Figure 12.44 and the point P directly ahead of O. The line OP defines the geometric shadow. Below O the screen cuts off the wavefront. The phase difference between the contributions to the disturbance at P from O and from a point H, height h above O is given by

$$\Delta(h) = \frac{2\pi}{\lambda}(\mathrm{HP} - \mathrm{OP}) \simeq \frac{2\pi}{\lambda}\frac{1}{2}\frac{h^2}{l}$$

where $\mathrm{OP} = l$ and higher powers of h^2/l^2 are neglected.

We now divide the wavefront above O into strips which are parallel to the infinite straight edge and we call these strips ‘half period zones’. This name derives from the fact that the width of each strip is chosen so that the contributions to the disturbance at P from the lower and upper edges of a given strip differ in phase by π radians.

Since the phase $\Delta(h) \propto h^2$ we shall not expect these strips or half period zones to be of equal width and Figure 12.45 shows how the width of each strip decreases as h increases. The total contribution from a strip will depend upon its area; that is, upon its width. The amplitude and phase of the contribution at P from a narrow strip of width dh at a height h above O may be written as $(\mathrm{d}h)\,\mathrm{e}^{\mathrm{i}\Delta}$ where $\Delta = \pi h^2/\lambda l$.

This contribution may be resolved into two perpendicular components

$$\mathrm{d}x = \mathrm{d}h \cos\Delta$$

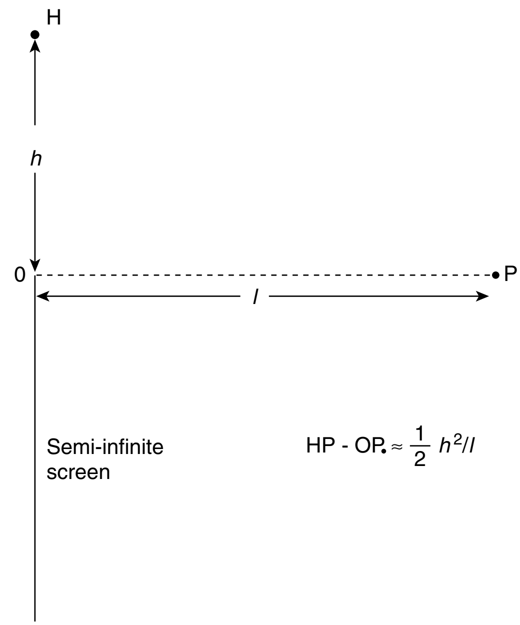


Figure 12.44 Line OP normal to the straight edge defines the geometric shadow. The wavefront at height h above O makes a contribution to the disturbance at P which has a phase lag of $\pi h^2/\lambda l$ with respect to that from O. The total disturbance at P is the vector sum (amplitude and phase) of all contributions from the wavefront section above O

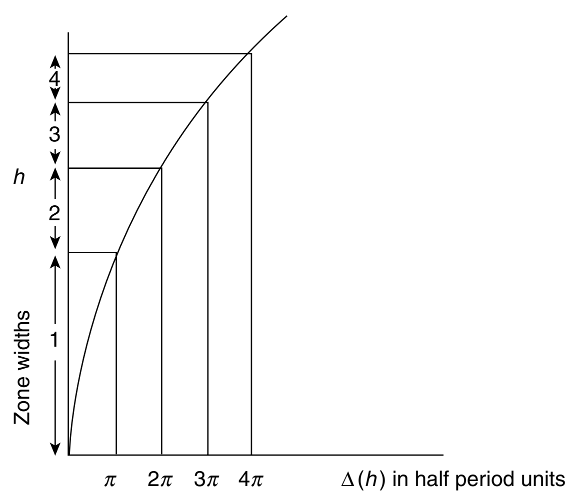


Figure 12.45 Variation of the width of each half period zone with height h above the straight edge

and

$$\mathrm{d}y = \mathrm{d}h \sin \Delta$$

If we now plot the vector sum of these contributions the total disturbance at P from that section of the wavefront measured from O to a height h will have the component values $x = \int \mathrm{d}x$ and $y = \int \mathrm{d}y$. These integrals are usually expressed in terms of the dimensionless variable $u = h(2/\lambda l)^{1/2}$, the physical significance of which we shall see shortly.

We then have $\Delta = \pi u^2/2$ and $\mathrm{d}h = (\lambda l/2)^{1/2}\,\mathrm{d}u$ and the integrals become

$$x = \int \mathrm{d}x = \int_0^u \cos(\pi u^2/2)\,\mathrm{d}u$$

and

$$y = \int \mathrm{d}y = \int_0^u \sin(\pi u^2/2)\,\mathrm{d}u$$

These integrals are called Fresnel's Integrals and the locus of the coordinates x and y with variation of u (that is, of h) is the spiral in the first quadrant of Figure 12.42. The complete figure is known as Cornu's spiral.

As h, the width of the contributing wavefront above the straight edge, increases, we measure the increasing length u from 0 along the curve of the spiral in the first quadrant unit, as h and $u \to \infty$ we reach Z_1 the centre of the spiral eye with coordinate $x = \frac{1}{2}, y = \frac{1}{2}$.

The tangent to the spiral curve is

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \tan\frac{\pi u^2}{2}$$

and this is zero when the phase

$$\Delta(h) = \pi h^2/\lambda l = \pi u^2/2 = m\pi$$

where m is an integer so that $u = \sqrt{(2m)}$ relates u, the distance measured along the spiral to m the number of half period zones contributing to the disturbance at P. The total intensity at P due to all the half period zones above the straight edge is given by the square of the length of the 'chord' OZ_1. This is the intensity at the geometric shadow.

Suppose now that we keep P fixed as we slowly withdraw the screen vertically downwards from O. This begins to reveal contributions to P from the lower part of the wavefront; that is, the part which contributes to the Cornu spiral in the third quadrant. The length u now includes not only the whole of the upper spiral arm but an increasing part of the lower spiral until, when u has extended to Z_2 the 'chord' $\mathrm{Z}_1\mathrm{Z}_2$ has its maximum value and this corresponds to the fringe of maximum intensity nearest the straight edge. Further withdrawal of the screen lengthens u to the position Z_3 which corresponds to the first minimum of the fringe pattern and the convolutions of an increasing length u around the spiral eye will produce further intensity oscillations of decreasing magnitude until, with the final removal of the screen, u is now the total length of the spiral and the square of the 'chord' length Z_1Z_1' gives the undisturbed intensity of unit value. But $Z_1Z_1' = 2Z_1O$ so that the undisturbed intensity $(Z_1Z_1')^2$ is a factor of four greater than $(Z_1O)^2$ the intensity at the geometric shadow.

The Fresnel diffraction pattern from a slit may now be seen as that due to a fixed height h of the wavefront equal to that of the slit width. This defines a fixed length u of the spiral between the end points of which the 'chord' is drawn and its length measured and squared to give the intensity. At a given distance from the slit the intensity at a point P in the diffraction pattern will correlate with the precise location of the fixed length u along the spiral. At the centre of the pattern P is symmetric with respect to the upper and lower edges of the slit and the fixed length u is centred about O (Figure 12.46). As P moves across the pattern towards the geometric shadow the length u moves around the convolutions of the spiral. In the geometric shadow this length is located entirely within the first or third quadrant of the spiral and the magnitude of the 'chord' between its ends is less than OZ_1. When the slit is wide enough to produce the central minimum of the diffraction pattern in Figure 12.47 the length u is centred at O with its ends at Z_3 and Z_4 in Figure 12.46.

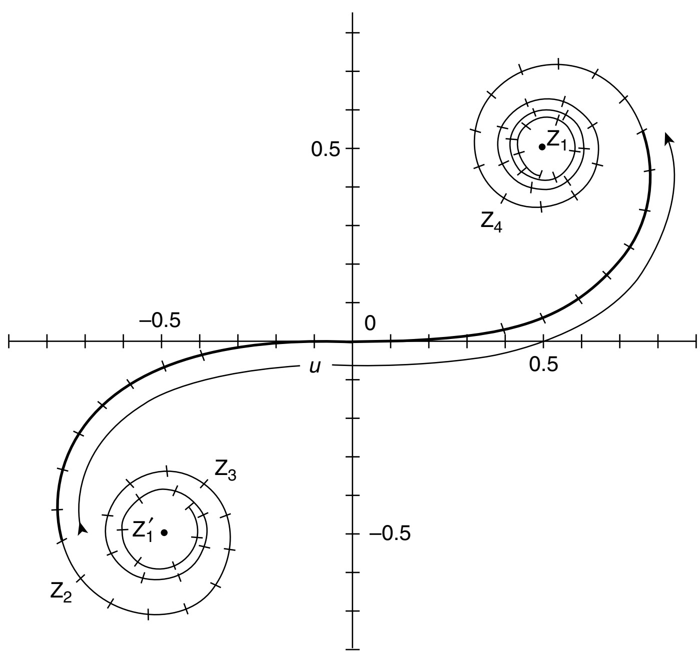


Figure 12.46 The slit width h defines a fixed length u of the spiral. The intensity at a point P in the diffraction pattern is correlated with the precise location of u on the spiral. When P is at the centre of the pattern u is centred on O and moves along the spiral as P moves towards the geometric shadow. Within the geometric shadow the chord joining the ends of u is less than OZ_1

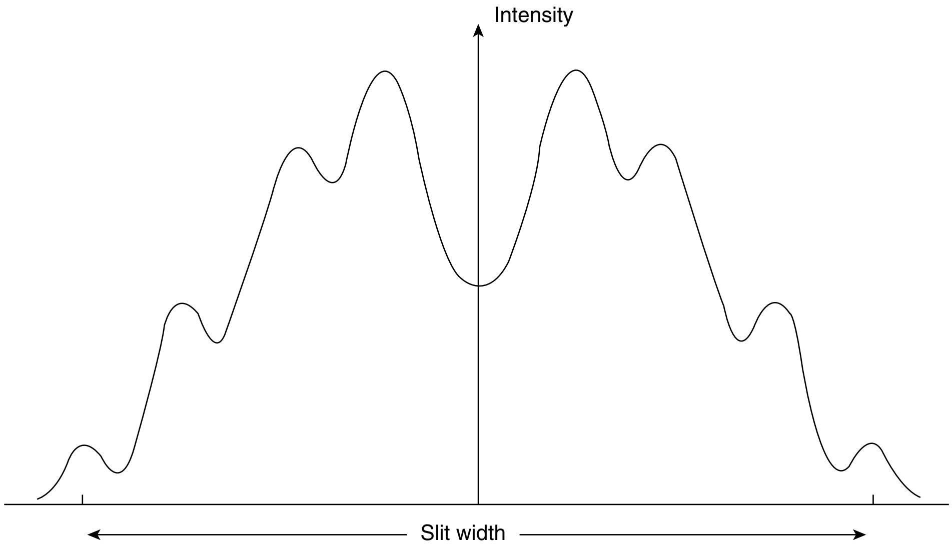


Figure 12.47 Fresnel diffraction pattern from a slit which is wide enough for the spiral length u to be centred at 0 and to end on points Z_3 and Z_4 of Figure 12.46. This produces the intensity minimum at the centre of the pattern

Circular Aperture (Fresnel Diffraction)

In this case the half period zones become annuli of decreasing width. If r_n is the mean radius of the half period zone whose phase lag is $n\pi$ with respect to the contribution from the central ring the path difference in Figure 12.48 is given by

$$\mathrm{NP} - \mathrm{OP} = \Delta = n\lambda/2 = \tfrac{1}{2}r_n^2/l$$

Unlike the rectangular example of the straight edge where the area of the half period zone was proportional to its width dh each zone here has the same area equal to $\pi\lambda l$.

Each zone thus contributes equally to the disturbance at P except for a factor arising from the rigorous Kirchhoff theory which, until now, we have been able to ignore. This is the so-called obliquity factor $\cos\chi$ where χ is shown in the figure. This factor is negligible for small values of n but its effect is to reduce a zone contribution as n increases. A large circular aperture with many zones produces, in the limit, an undisturbed normal intensity on the axis and from Figure 12.49 where we show the magnitude and phase from successive half zones we see that the sum of these vectors which represents the amplitude produced by an undisturbed wave is only half of that from the innermost zone.

It is evident that if alternate zones transmit no light then the contributions from the remaining zones would all be in phase and combine to produce a high intensity at P similar

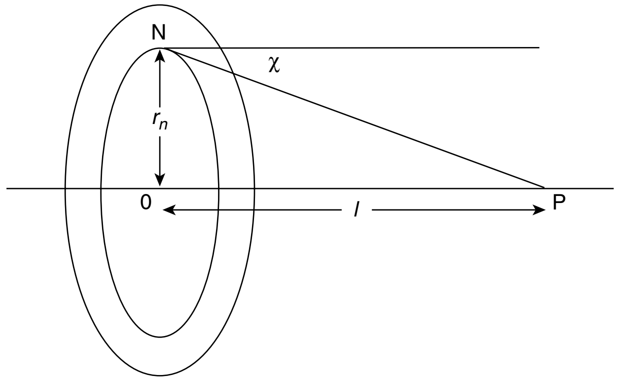


Figure 12.48 Fresnel diffraction from a circular aperture. The mean radius r_n defines the half period zone with a phase lag of $n\pi$ at P with respect to the contribution from the central zone. The obliquity angle χ which reduces the zone contribution at P increases with n

to the focusing effect of a lens. Such circular 'zone plates' are made by blacking out the appropriate areas of a glass slide, Figure 12.50. A further refinement increases the intensity still more. If the alternate zone areas are not blacked out but become areas where the *optical* thickness of the glass is reduced, via etching, by $\lambda/2$ the light transmitted through these zones is advanced in phase by π rad so that the contributions from all the zones are now in phase.

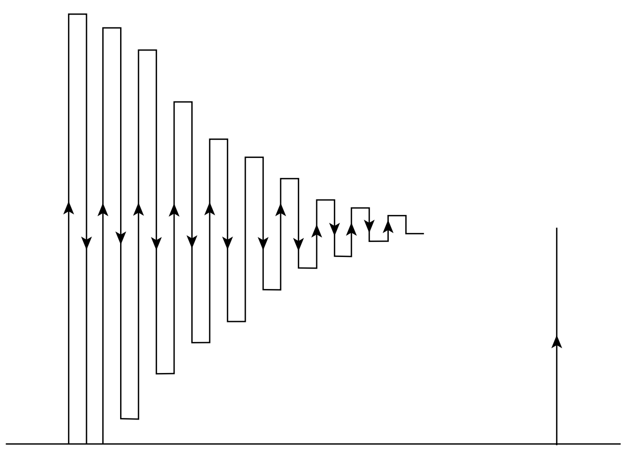


Figure 12.49 The vector contributions from successive zones in the circular aperture. The amplitude produced by an undisturbed wave is seen to be only half of that from the central zone. Removing the contributions from alternate zones leaves the remainder in phase and produces a very high intensity. This is the principle of the zone plate of Figure 12.50

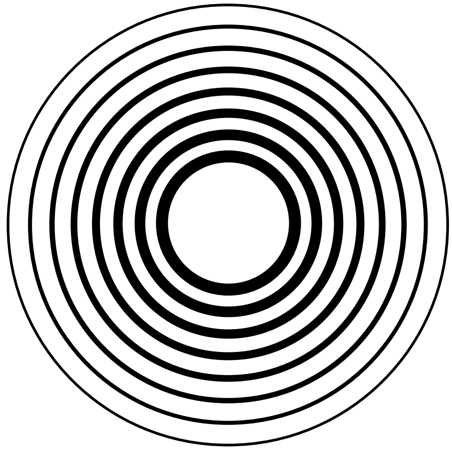


Figure 12.50 Zone plate produced by removing alternate half zones from a circular aperture to leave the remaining contributions in phase

Holography

Why is it that when we observe an object we see it in three dimensions but when we photograph it we obtain only a flat two dimensional distribution of light intensity? The answer is that the photograph has lost the information contained in the *phase* of the incident light. Holographic processes retain this information and a hologram reconstructs a three-dimensional image.

The principle of holography was proposed by Gabor in 1948 but its full development needed the intense beams of laser light. A hologram requires two coherent beams and the holographic plate records their interference pattern. In practice both beams derive from the same source, one serves as a direct reference beam the other is the wavefront scattered from the object.

Figure 12.51 shows one possible arrangement where a partly silvered beam splitter passes the direct reference beam and reflects light on to the object which scatters it on to the photographic plate. Mirrors or deviating prisms are also used to split the incident beam.

In Figure 12.51 let the reference beam amplitude be $A_0\,\mathrm{e}^{\mathrm{i}\omega t}$. If the holographic plate lies in the yz plane both the amplitude and phase of scattered light which strikes a given point (y, z) on the plate will depend on these co-ordinates. We simplify the analysis by considering only the y co-ordinate shown in the plane of the paper and we represent the scattered light in amplitude and phase as a function of y, namely

$$A(y)\,\mathrm{e}^{\mathrm{i}(\omega t+\phi(y))}$$

It is this information we shall wish to recover.

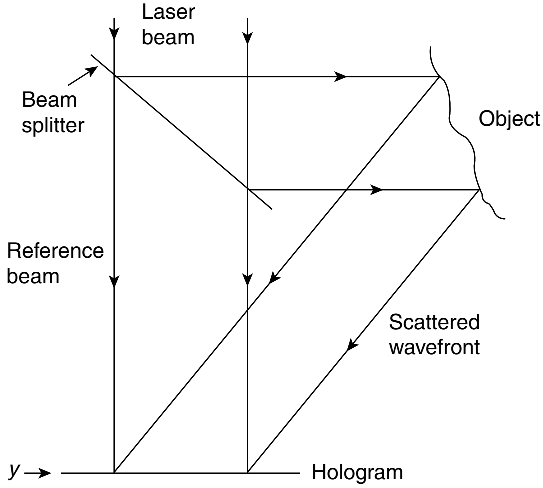


Figure 12.51 The hologram records the interference between two parts of the same laser beam. The original beam is divided by the partially silvered beam splitter to form a direct reference beam and a wavefront scattered from the object. The amplitude and phase information contained in the scattered wavefront must be preserved and recovered

We may now write the resulting amplitude at y (after removing the common $e^{i\omega t}$ factor) as

$$A = A_0 + A(y)\,e^{i\phi(y)}$$

The intensity is therefore

$$\begin{aligned} I = AA^* &= [A_0 + A(y)\,e^{i\phi(y)}][A_0 + A(y)\,e^{-i\phi(y)}] \\ &= A_0^2 + A(y)^2 + A_0 A(y)[e^{i\phi(y)} + e^{-i\phi(y)}] \end{aligned}$$

The holographic plate records this intensity as shown in Figure 12.52 where the reference intensity A_0^2 is modulated by the terms which contain $A(y)$ and $\phi(y)$, the original scattered amplitude and phase information. This modulation shows of course as contrasting interference fringes whose local intensity is governed by the amplitude $A(y)$ and whose distribution along the y axis is determined by the phase $\phi(y)$. The wavefront scattered by the object is now reconstructed to form the holographic image. This is done by shining the reference beam through the processed hologram which acts as a diffraction grating. The greater the recorded intensity the lower the transmitted amplitude. If the developed photographic emulsion possessed idealized characteristics the relation between the transmitted amplitude of the reference beam and the *exposure* would be linear.

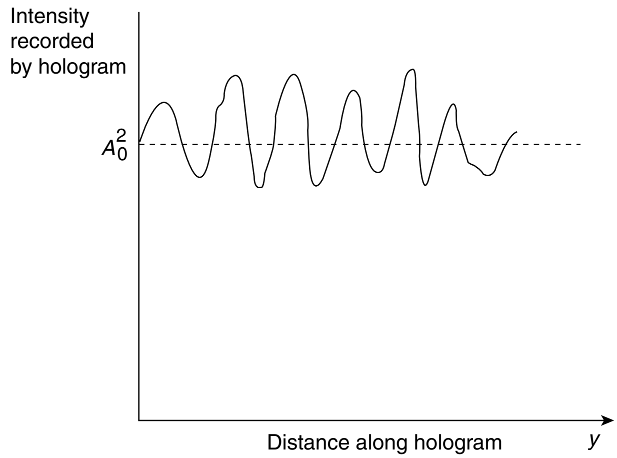


Figure 12.52 Total intensity recorded as a function of y by the holographic plate. The direct reference beam intensity A_0^2 is modulated by information from the scattered wavefront. This shows as variations in the intensity of an interference fringe pattern

Exposure defines the product of incident intensity and exposure time. The curve relating the characteristics for a real holographic emulsion is shown in Figure 12.53 and this is linear only over a limited range near the centre indicated by the dotted lines. This imposes several conditions on practical holography.

In the first place the exposure must be correctly chosen at the value E_C. Secondly, the value of the reference beam intensity A_0^2 must be chosen to produce the correct transmitted amplitude T_0 on the vertical axis of Figure 12.53. This value of T_0 is at the centre of the linear range. Finally, the modulation of A_0^2 by the scattered intensity $A(y)^2$ in Figure 12.53 must be small enough for the transmission of the modulated signal to remain within the linear range of the characteristic curve. Excursions outside this range introduce non-linear distortions by generating extra Fourier frequency components (the situation is similar to that for characteristic curves in electronic amplifiers).

Experimentally this final restriction requires $A(y) \ll A_0$.

Shining the reference beam through the processed hologram produces a transmitted *amplitude*

$$\begin{aligned} A_0 T &= A_0^3 + A_0^2 A(y)\,e^{i\phi(y)} + A_0^2 A(y)\,e^{-i\phi(y)} \\ &= A_0^2[A_0 + A(y)\,e^{i\phi(y)} + A(y)\,e^{-i\phi(y)}] \end{aligned}$$

where we have neglected the $A(y)^2$ term as $\ll A_0^2$ and have written the negative and positive exponential terms separately. This has a profound physical significance for we see that apart from the common constant factor A_0^2, the observed transmitted beam has three components $A_0, A(y)\,e^{i\phi(y)}$ and $A(y)\,e^{-i\phi(y)}$.

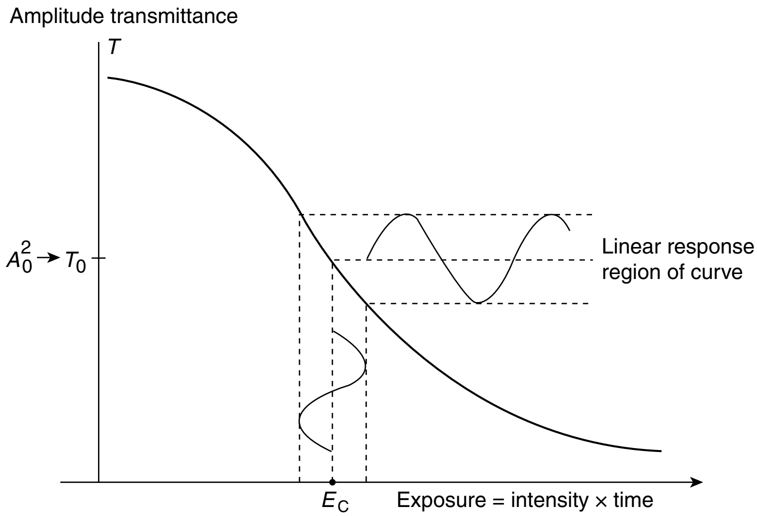


Figure 12.53 Characteristic curve of a real holographic emulsion (transmittance versus exposure). Only the central linear section of the curve is used. The transmittance T_0 (governed by the reference beam intensity A_0^2) is chosen with the critical exposure E_C to produce the central point on the linear part of the curve. Information from the scattered wavefront must keep the modulations within the linear range for faithful reproduction free from distortion

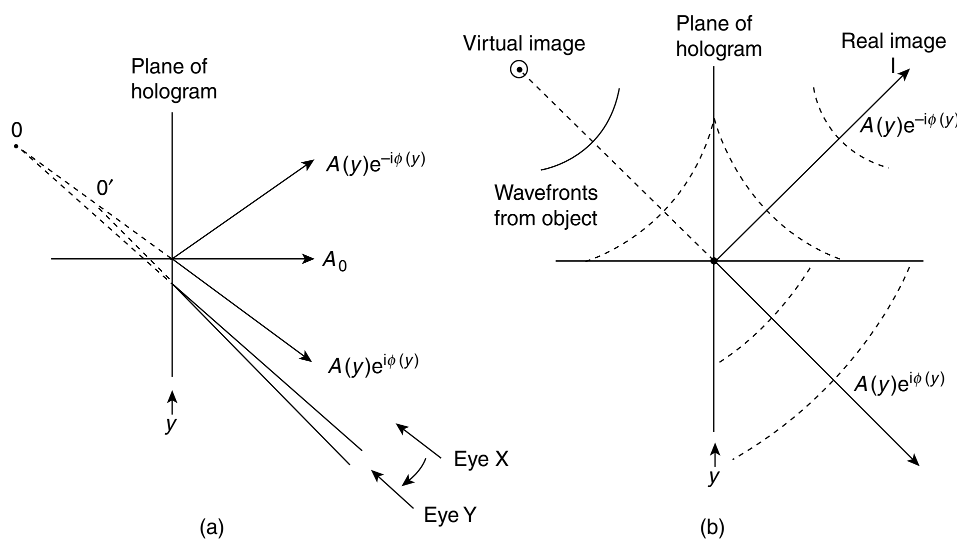


Figure 12.54 (a) Shining the reference beam through the processed hologram produces three components $A_0, A(y)\,e^{i\phi(y)}$ and $A(y)\,e^{-i\phi(y)}$ in the directions shown. Movement of the eye from X to Y about the component $A(y)\,e^{i\phi(y)}$ resolves the separate points 0 and 0′ on the image of the object to reveal its three dimensional nature. (b) This image at 0 is seen to be virtual while the image associated with the component $A(y)\,e^{-i\phi(y)}$ is real. This real image is 'phase reversed' and the object appears 'inside out'

The first term, A_0, shows that the incident reference beam has continued beyond the hologram to form the central beam of Figure 12.54a. The second component $A(y)\,e^{i\phi(y)}$ has the same form in amplitude and phase as the original wavefront scattered from the object. As shown in Figure 12.54b it is seen to be a wavefront diverging from a virtual image of the object having the same size and three dimensional distribution as the object itself. Moving the eye across this beam in 12.54a exposes a different section OO′ of the virtual image to produce a three dimensional effect.

The third component of the transmitted beam is identical with the second except for the phase reversal; it has a negative exponential index. It forms another image, in this case a real image often referred to as ‘pseudoscopic’. It is an image of the original object turned inside out. All contours are reversed, bumps become dents and the closest point on the original object when viewed directly by the observer now becomes the most distant.

Problem 12.1

Suppose that Newton’s Rings are formed by the system of Figure 12.4 except that the plano convex lens now rests centrally in a concave surface of radius of curvature R_1 and not on an optical flat. Show that the radius r_n of the nth dark ring is given by

$$r_n^2 = R_1 R_2 n\lambda/(R_1 - R_2)$$

where R_2 is the radius of curvature of the lens and $R_1 > R_2$ (note that R_1 and R_2 have the same sign).

Problem 12.2

Light of wavelength λ in a medium of refractive index n_1 is normally incident on a thin film of refractive index n_2 and optical thickness $\lambda/4$ which coats a plane substrate of refractive index n_3. Show that the film is a perfect anti-reflector $(r = 0)$ if $n_2^2 = n_1 n_3$.

Problem 12.3

Two identical radio masts transmit at a frequency of 1500 kc s^{-1} and are 400 m apart. Show that the intensity of the interference pattern between these radiators is given by $I = 2I_0[1 + \cos(4\pi \sin\theta)]$, where I_0 is the radiated intensity of each. Plot this intensity distribution on a polar diagram in which the masts lie on the 90°–270° axis to show that there are two major cones of radiation in opposite directions along this axis and 6 minor cones at 0°, 30°, 150°, 180°, 210° and 330°.

Problem 12.4

(a) Two equal sources radiate a wavelength λ and are separated a distance $\lambda/2$. There is a phase difference $\delta_0 = \pi$ between the signals at source. If the intensity of each source is I_s, show that the intensity of the radiation pattern is given by

$$I = 4I_s \sin^2\left(\frac{\pi}{2}\sin\theta\right)$$

where the sources lie on the axis $\pm\pi/2$.

Plot I versus θ.

(b) If the sources in (a) are now $\lambda/4$ apart and $\delta_0 = \pi/2$ show that

$$I = 4I_s\left[\cos^2\frac{\pi}{4}(1 + \sin\theta)\right]$$

Plot I versus θ.

Problem 12.5

(a) A large number of identical radiators is arranged in rows and columns to form a lattice of which the unit cell is a square of side d. Show that all the radiation from the lattice in the direction θ will be in phase at a large distance if $\tan\theta = m/n$, where m and n are integers.

(b) If the lattice of section (a) consists of atoms in a crystal where the rows are parallel to the crystal face, show that radiation of wavelength λ incident on the crystal face at a grazing angle of θ is scattered to give interference maxima when $2d\sin\theta = n\lambda$ (Bragg reflection).

Problem 12.6

Show that the separation of equal sources in a linear array producing a principal maximum along the line of the sources ($\theta = \pm\pi/2$) is equal to the wavelength being radiated. Such a pattern is called 'end fire'. Determine the positions (values of θ) of the secondary maxima for $N = 4$ and plot the angular distribution of the intensity.

Problem 12.7

The first multiple radio astronomical interferometer was equivalent to a linear array of $N = 32$ sources (receivers) with a separation $f = 7\,\mathrm{m}$ working at a wavelength $\lambda = 0.21\,\mathrm{m}$. Show that the angular width of the central maximum is 6 min of arc and that the angular separation between successive principal maxima is $1°42'$.

Problem 12.8

Monochromatic light is normally incident on a single slit, and the intensity of the diffracted light at an angle θ is represented in magnitude and direction by a vector $\mathbf{I}$, the tip of which traces a polar diagram. Sketch several polar diagrams to show that as the ratio of slit width to the wavelength gradually increases the polar diagram becomes concentrated along the direction $\theta = 0$.

Problem 12.9

The condition for the maxima of the intensity of light of wavelength λ diffracted by a single slit of width d is given by $\alpha = \tan\alpha$, where $\alpha = \pi d\sin\theta/\lambda$. The approximate values of α which satisfy this equation are $\alpha = 0, +3\pi/2, +5\pi/2$, etc. Writing $\alpha = 3\pi/2 - \delta, 5\pi/2 - \delta$, etc. where δ is small, show that the real solutions for α are $\alpha = 0,\ \pm1.43\pi, \pm2.459\pi, \pm3.471\pi$, etc.

Problem 12.10

Prove that the intensity of the secondary maximum for a grating of three slits is $\frac{1}{9}$ of that of the principal maximum if only interference effects are considered.

Problem 12.11

A diffraction grating has N slits and a grating space f. If $\beta = \pi f\sin\theta/\lambda$, where θ is the angle of diffraction, calculate the phase change $\mathrm{d}\beta$ required to move the diffracted light from the principal maximum to the first minimum to show that the half width of the spectral line produced by the grating is given by $\mathrm{d}\,\theta = (nN\cot\theta)^{-1}$, where n is the spectral order. (For $N = 14,000, n = 1$ and $\theta = 19°$, $\mathrm{d}\theta \approx 5\,\mathrm{s}$ of arc.)

Problem 12.12

(a) Dispersion is the separation of spectral lines of different wavelengths by a diffraction grating and increases with the spectral order n. Show that the dispersion of the lines when projected by a lens of focal length F on a screen is given by

$$\frac{\mathrm{d}l}{\mathrm{d}\lambda} = F\frac{\mathrm{d}\theta}{\mathrm{d}\lambda} = \frac{nF}{f}$$

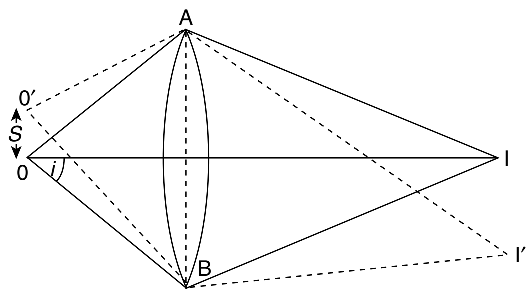
for a diffraction angle θ and the nth order, where l is the linear spacing on the screen and f is the grating space.
(b) Show that the change in linear separation per unit increase in spectral order for two wavelengths $\lambda = 5 \times 10^{-7}$ m and $\lambda_2 = 5.2 \times 10^{-7}$ m in a system where $F = 2$ m and $f = 2 \times 10^{-6}$ m is 2×10^{-2} m.

Problem 12.13

(a) A sodium doublet consists of two wavelength $\lambda_1 = 5.890 \times 10^{-7}$ m and $\lambda_2 = 5.896 \times 10^{-7}$ m. Show that the minimum number of lines a grating must have to resolve this doublet in the third spectral order is ≈ 328.
(b) A red spectral line of wavelength $\lambda = 6.5 \times 10^{-7}$ m is observed to be a close doublet. If the two lines are just resolved in the third spectral order by a grating of 9×10^4 lines show that the doublet separation is 2.4×10^{-2} m.

Problem 12.14

Optical instruments have circular apertures, so that the Rayleigh criterion for resolution is given by $\sin\theta = 1.22\lambda/a$, where a is the diameter of the aperture.



s exaggerated. Consider OB || O′B
OA || O′A

Two points O and O′ of a specimen in the object plane of a microscope are separated by a distance s. The angle subtended by each at the objective aperture is $2i$ and their images I and I′ are just resolved. By considering the path difference between O′A and O′B show that the separation $s = 1.22\lambda/2\sin i$.

Summary of Important Results

Interference: Division of Wavefront (Two Equal Sources)

Intensity

$$I = 4I_s \cos^2 \delta/2$$

where

$$I_s = \text{source intensity}$$

and

$$\delta = \left[\frac{2\pi}{\lambda}(\text{path difference})\right] \text{is phase difference}$$

Interference (N Equal Sources – Separation f)

$$I = I_s \frac{\sin^2 N\beta}{\sin^2 \beta} \quad \text{where} \quad \beta = \frac{\pi}{\lambda} f \sin\theta$$

Principal Maxima

$$I = N^2 I_s \quad \text{at}\, f \sin\theta = n\lambda$$

Fraunhofer Diffraction (Single Slit – Width d)

Intensity

$$I = I_0 \frac{\sin^2 \alpha}{\alpha^2} \quad \text{where} \quad \alpha = \frac{\pi}{\lambda} d \sin\theta$$

Intensity Distribution from N Slits (Width d – Separation f)

$$I = I_0 \frac{\sin^2 \alpha}{\alpha^2} \frac{\sin^2 N\beta}{\sin^2 \beta}$$

(interference pattern modified by single slit diffraction envelope).

Resolving Power of Transmission Grating

$$\frac{\lambda}{d\lambda} = nN$$

where n is spectral order and N is number of grating lines:
Expressible in terms of Bandwidth Theorem as

$$\Delta\nu \Delta t = 1$$

where $\Delta\nu$ is resolvable frequency difference and Δt is the time difference between extreme optical paths.

Resolving power

$$\frac{\lambda}{\Delta\lambda} = \left|\frac{\nu}{\Delta\nu}\right| = \frac{\omega}{\Delta\omega} = Q$$

where Q is the quality factor of the system.