Week 7: Interference

Superposition: Till now we have mostly discusssed single waves. While discussing group velocity we did talk briefly about superposing more than one wave. We will now focus on superposition in more detail now. Superposition of waves and the resulting consequences are crucial for understanding a large number of phenomena involving interference and diffraction. At the outset, let us recall that **the principle of linear superposition** holds for the fields associated with the waves under consideration (eg. electric and magnetic fields for an electromagnetic wave). This is because the wave equation is **linear** and a linear sum of two solutions is also a solution. We must keep this in mind as we go along.

Let us first discuss the simple case of adding the electric fields of two plane electromagnetic waves of the same frequency, in vacuum, but moving along different directions $\vec{\mathbf{k}}_\mathbf{1}$ and $\vec{\mathbf{k}}_\mathbf{2}$. We write the electric fields as

$$\vec{\mathbf{E}}_\mathbf{1} = \vec{\mathbf{E}}_\mathbf{01} \cos\left(\vec{\mathbf{k}}_\mathbf{1} \cdot \vec{\mathbf{r}} - \omega t + \epsilon_1\right) \tag{1}$$

$$\vec{\mathbf{E}}_\mathbf{2} = \vec{\mathbf{E}}_\mathbf{02} \cos\left(\vec{\mathbf{k}}_\mathbf{2} \cdot \vec{\mathbf{r}} - \omega t + \epsilon_2\right) \tag{2}$$

$\epsilon_{1,2}$ are arbitrary constant phases. The resultant electric field is just $\vec{\mathbf{E}} = \vec{\mathbf{E}}_\mathbf{1} + \vec{\mathbf{E}}_\mathbf{2}$. What is the **intensity** or **irradiance** due to the superposed field? We know that

$$I = \epsilon_0 c \langle E^2 \rangle_T \tag{3}$$

with T much greater than τ (the time period of the field oscillations). We shall forget the factor $\epsilon_0 c$ henceforth and only look at $I = \langle E^2 \rangle_T$. This is because we are not very interested in the absolute values of the intensities. Let us calculate E^2 first. This turns out to be,

$$\begin{aligned} E^2 &= \left(\vec{\mathbf{E}}_\mathbf{1} + \vec{\mathbf{E}}_\mathbf{2}\right) \cdot \left(\vec{\mathbf{E}}_\mathbf{1} + \vec{\mathbf{E}}_\mathbf{2}\right) \\ &= |\vec{\mathbf{E}}_\mathbf{01}|^2 \cos^2\left(\vec{\mathbf{k}}_\mathbf{1} \cdot \vec{\mathbf{r}} - \omega t + \epsilon_1\right) + |\vec{\mathbf{E}}_\mathbf{01}|^2 \cos^2\left(\vec{\mathbf{k}}_\mathbf{2} \cdot \vec{\mathbf{r}} - \omega t + \epsilon_2\right) \\ &\quad + 2\vec{\mathbf{E}}_\mathbf{01} \cdot \vec{\mathbf{E}}_\mathbf{02} \cos\left(\vec{\mathbf{k}}_\mathbf{1} \cdot \vec{\mathbf{r}} - \omega t + \epsilon_1\right) \cos\left(\vec{\mathbf{k}}_\mathbf{2} \cdot \vec{\mathbf{r}} - \omega t + \epsilon_2\right) \end{aligned} \tag{4}$$

In the last term, we isolate the time-dependent parts by breaking the $\cos\left(\vec{\mathbf{k}}_\mathbf{1} \cdot \vec{\mathbf{r}} - \omega t + \epsilon_1\right)$ and $\cos\left(\vec{\mathbf{k}}_\mathbf{2} \cdot \vec{\mathbf{r}} - \omega t + \epsilon_2\right)$. Thereafter, taking a time-average, we obtain

$$I = \frac{1}{2}|\vec{\mathbf{E}}_\mathbf{01}|^2 + \frac{1}{2}|\vec{\mathbf{E}}_\mathbf{02}|^2 + \vec{\mathbf{E}}_\mathbf{01} \cdot \vec{\mathbf{E}}_\mathbf{02} \cos\left(\left(\vec{\mathbf{k}}_\mathbf{1} - \vec{\mathbf{k}}_\mathbf{2}\right) \cdot \vec{\mathbf{r}} + \epsilon_1 - \epsilon_2\right) \tag{5}$$

With redefinitions and assuming $\vec{\mathbf{E}}_{01}$ and $\vec{\mathbf{E}}_{02}$ are parallel vectors, we can write the above as

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2} \cos \delta \tag{6}$$

where δ is defined as

$$\delta = \left(\vec{\mathbf{k}}_{\mathbf{1}} - \vec{\mathbf{k}}_{\mathbf{2}} \right) \cdot \vec{\mathbf{r}} + \epsilon_1 - \epsilon_2 \tag{7}$$

and $I_1 = \frac{1}{2}|\vec{\mathbf{E}}_{\mathbf{01}}|^2$, $I_2 = \frac{1}{2}|\vec{\mathbf{E}}_{\mathbf{02}}|^2$. Note that if $\vec{\mathbf{E}}_{01}$ and $\vec{\mathbf{E}}_{02}$ are perpendicular to each other $I = I_1 + I_2$. The term associated with the phase δ is known as the **interference term**. The interference term is the cause behind the variations in the intensity. When $\delta = 2m\pi$ we see that $I = I_1 + I_2 + 2\sqrt{I_1 I_2}$ whereas when $\delta = (2m+1)\pi$ we have $I = I_1 + I_2 - 2\sqrt{I_1 I_2}$. If we have $I_1 = I_2$ then the intensities become $4I_1$ or zero. For visible wavelengths, we will therefore see regions where there is no light (dark) and regions where there is a lot of light (bright). The redistribution of light will be regular and is quantified. Such a regular redistribution of the light intensity is known as a **fringe** in optics. We shall deal with the occurence and the properties of such fringes which can arise in various diverse situations. Precise measurements of the properties of such fringes can lead to precise measurements of wavelengths, the thickness of films and various other quantities in science and engineering. In the above, we have talked about plane waves. If we have spherical waves how do things change? The electric fields can now be written as

$$\vec{\mathbf{E}}_{\mathbf{1}} = \vec{\mathbf{E}}_{\mathbf{01}}(r_1) e^{i(kr_1 - \omega t + \epsilon_1)} \quad ; \quad \vec{\mathbf{E}}_{\mathbf{2}} = \vec{\mathbf{E}}_{\mathbf{02}}(r_2) e^{i(kr_2 - \omega t + \epsilon_2)} \tag{8}$$

where r_1 and r_2 are the radii of the overlapping wavefronts at the point of observation P (see Figure 1). The δ in this case is

$$\delta = k\left(r_1 - r_2\right) + \epsilon_1 - \epsilon_2 \tag{9}$$

We now make a crucial assumption – **the separation between the sources is small** and **the region of interference is also small**. With these assumptions, the $\vec{\mathbf{E}}_{\mathbf{01}}$ and $\vec{\mathbf{E}}_{\mathbf{02}}$ can be assumed to be independent of r_1 and r_2. Therefore, assuming $\vec{\mathbf{E}}_{\mathbf{01}} = \vec{\mathbf{E}}_{\mathbf{02}}$, we have

$$I = 4I_0 \cos^2 \frac{1}{2}\left[k(r_1 - r_2) + \epsilon_1 - \epsilon_2\right] \tag{10}$$

as the expression for the intensity. We have maxima and minima for (assuming $\epsilon_1 = \epsilon_2$),

$$r_1 - r_2 = 2\pi \frac{m}{k} = m\lambda \ \ (maxima) \quad ; \quad r_1 - r_2 = \pi \frac{m'}{k} = \frac{m'}{2}\lambda \ \ (minima) \tag{11}$$

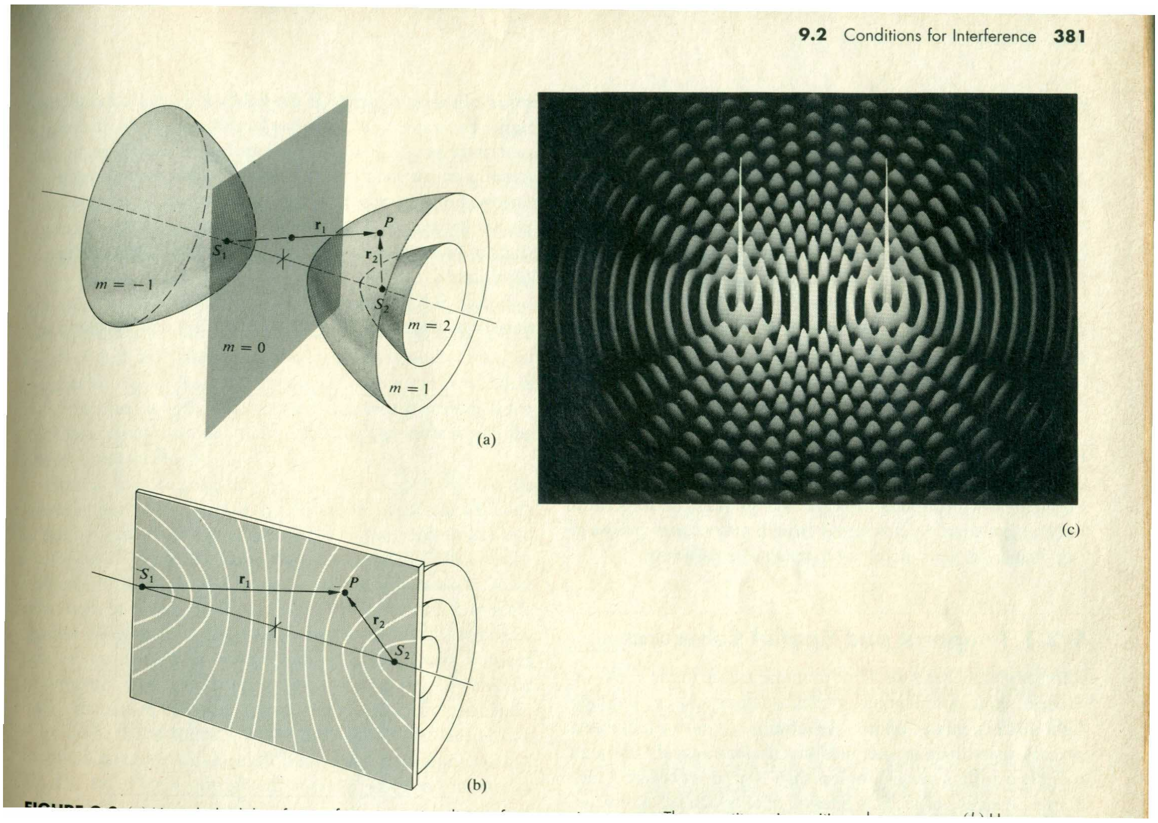


FIG. 1: Superposition of two spherical waves (figure from E. Hecht (Optics))

with $m = 0, \pm 1, \pm 2....$ and $m' = \pm 1, \pm 3.....$

What are these surfaces? They are hyperboloids of revolution (see Figure 1). If we assume a screen perpendicular to the $m = 0$ plane we get somewhat straight fringes as shown in Figure 1. On the other hand, if we place a screen parallel to the $m = 0$ surface, we will see approximately circular fringes.

We now turn to the topic of **coherence** which is intimately related to the question: when do we really see the fringes?

Coherence (qualitative): Coherence is understood w.r.t. the spatial and temporal prop-

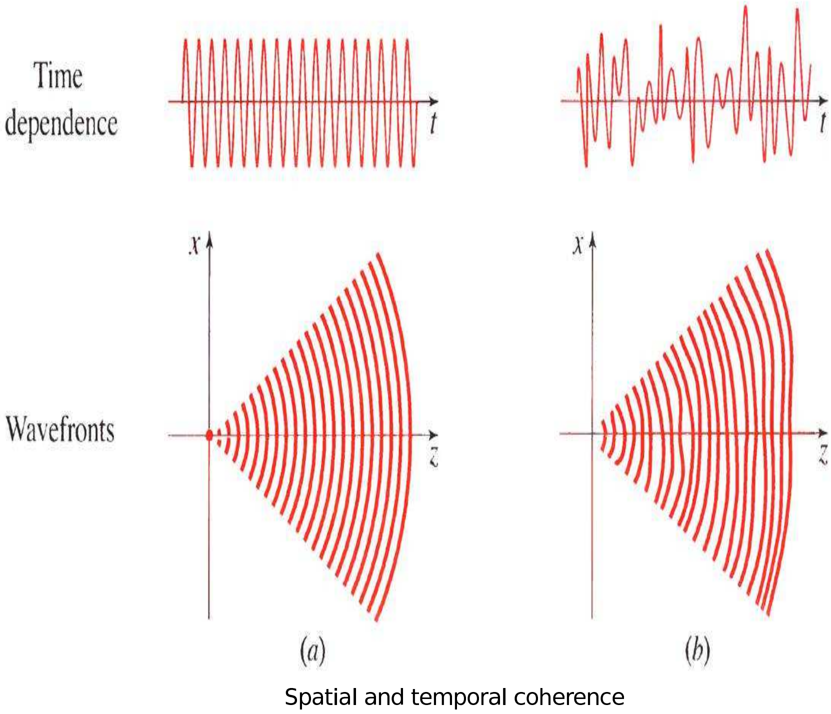


FIG. 2: Spatial and temporal coherence

erties of emissions from sources and therefore we shall talk about **spatial and temporal coherence** separately. Loosely speaking, temporal coherence is associated with the **finite frequency band-width** of a source, i.e. how close a source is to being **monochromatic**. Spatial coherence is related to the **finite spatial dimensions** of a source and hence to the **shape of the wavefront** as it propagates.

Let us now try to understand these features pictorially. Figure 2 shows perfect coherence on the L. H. S. (top-temporal, bottom-spatial). On the R. H. S (top) we have a superposition of different frequencies which result in the formation of wave groups or wave-trains–they are approximately sinusoidal over a limited period of time–this time is the **coherence time** τ_c and the corresponding temporal coherence length $l_c = c\tau_c$. Both these quantities are shown explicitly in Figure 3 once again. In Figures 2 and 3 (bottom figures) we illustrate spatial coherence (L. H. S) and spatial in-coherence (R. H. S). Spatial in-coherence, as mentioned earlier arises due to the finite dimensions of any source (for example, finite width for a slit source). Thus, a spatially incoherent wavefront is distorted in shape (as shown if Figs. 2 and 3). In Figure 4, we illustrate the three possible cases: (i) spatially and temporally coherent, (ii) temporal incoherence but spatial coherence (iii) spatial incoherence but temporal coherence. The figures are self-explanatory. Figures 2-3 are taken from Saleh and Teich, Fundamentals of Photonics (2007), Figure 4 is from a web-source. We will discuss spatial and temporal coherence in a more quantitative way later.

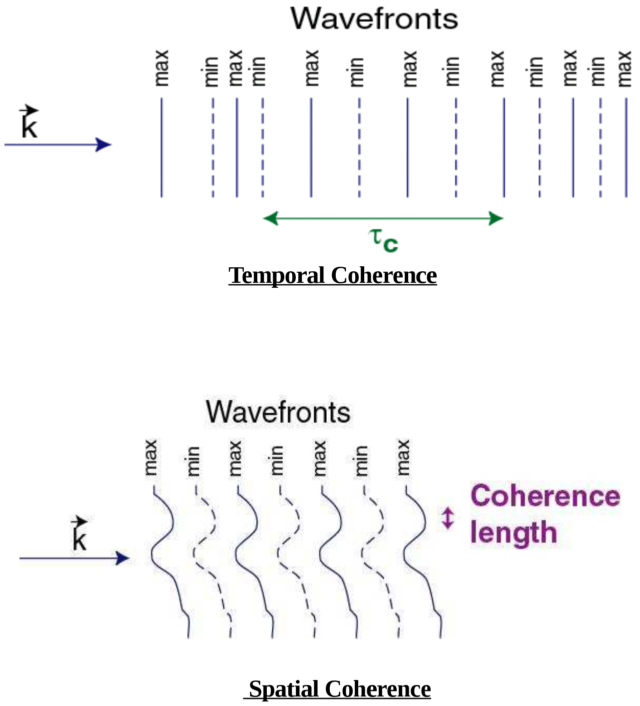


FIG. 3: Coherence time,length, spatial coherence

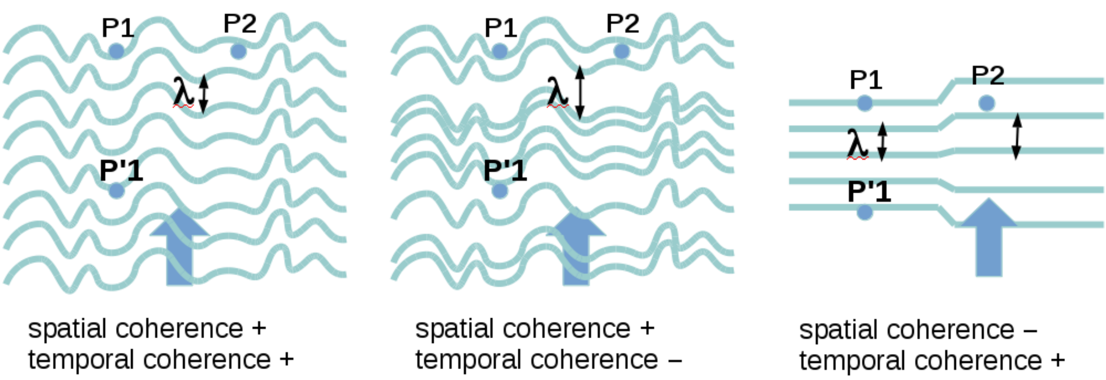


FIG. 4: Different situations

Huygens' principle: Long before we had a proper understanding of wave optics, Huygens, based on pure intuition formulated a principle which, remarkably, holds till today. Here is what Huygens' said.

Every point on a propagating wavefront serves as a source of spherical secondary wavelets, such that the wavefront at a later time is the envelope of these wavelets. If the propagating wave has a frequency ν and is transmitted through the medium at a speed v_t, then the secondary wavelets have the same frequency and speed.

Later, Fresnel added the notion of **interference** to Huygens' principle and Kirchoff put this principle on firm mathematical ground using Maxwell's theory of electrodynamics.

For a spherical waves, while using Huygens' principle, we just draw **hemispheres** , ignoring the **backward** wave. This issue was indeed addressed and solved by Kirchoff. A discussion on this is however beyond the scope of this course.

Young's double slit: Let us now turn to Young's double slit experiment which is the first example of an **interferometer** or more precisely a **wavefront splitting interferometer**. Generically, **interferometers** are devices which produce a **regular redistribution of light** known as a **fringe**. Thomas Young made this device–his use of two slits produced two coherent beams which could produce observable interference.

Figure 5 (top) provides a three dimensional view of what is happening in the Young's double slit experiment. Note the incident cylindrical wavefront which is split at the two slits and which then interferes to produce the new intensity distribution. The lower figure in Figure 5 sets up the details in a transverse section.

We may work out the details as follows. Assume the coordinates of the various points in the lower figure in Fig. 5 as follows (choosing the plane of the figure as the xy plane). The origin is at the centre between the two slits. The top slit has coordinates $(0, \frac{d}{2})$ and the bottom one has $(0, -\frac{d}{2})$. The coordinates of the point of observation in the screen are (D, y) and the distance from the origin to the point P is r, the line joining the origin to P makes an angle θ with the x axis. The unit vectors representing the propagation vectors of the two waves from slit 1 and slit 2 are

$$\hat{\mathbf{k}}_{\mathbf{1}} = \frac{D\hat{\mathbf{i}} + (y - \frac{d}{2})\hat{\mathbf{j}}}{\sqrt{D^2 + (y - \frac{d}{2})^2}} \tag{12}$$

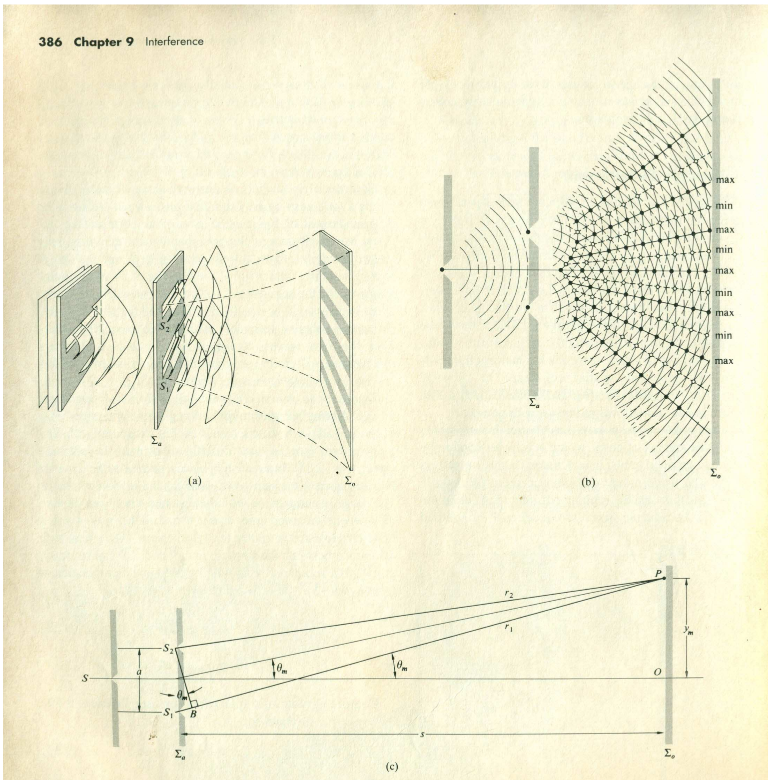


FIG. 5: Young's double slit, figure from Hecht, Optics

$$\hat{\mathbf{k}}_{\mathbf{2}} = \frac{D\hat{\mathbf{i}} + (y + \frac{d}{2})\hat{\mathbf{j}}}{\sqrt{D^2 + (y + \frac{d}{2})^2}} \tag{13}$$

The vector $\vec{\mathbf{r}} = D\hat{\mathbf{i}} + y\hat{\mathbf{j}}$. Therefore one can evaluate $\left(\vec{\mathbf{k}}_{\mathbf{2}} - \vec{\mathbf{k}}_{\mathbf{1}}\right) \cdot \vec{\mathbf{r}} = k\left(\hat{\mathbf{k}}_{\mathbf{2}} - \hat{\mathbf{k}}_{\mathbf{1}}\right) \cdot \vec{\mathbf{r}}$. Using the approximation that D is large, i.e. the screen is very far away, we can assume $\sqrt{D^2 + (y - \frac{d}{2})^2} \approx r \approx \sqrt{D^2 + (y + \frac{d}{2})^2}$. This finally gives

$$\left(\vec{\mathbf{k}}_{\mathbf{2}} - \vec{\mathbf{k}}_{\mathbf{1}}\right) \cdot \vec{\mathbf{r}} \approx \frac{dyk}{r} \tag{14}$$

Since $y = r\sin\theta$ and $D = r\cos\theta$ and further, assuming θ is small so that $\sin\theta \approx \theta$, we get the phase δ as

$$\delta = \frac{2\pi d\theta}{\lambda} \tag{15}$$

Therefore, we finally obtain the intensity as

$$I = 2I_0\left(1 + \cos\frac{2\pi d\theta}{\lambda}\right) \tag{16}$$

The bright fringes appear for

$$y_m = \frac{mr\lambda}{d} \quad ; \quad \theta_m = \frac{m\lambda}{d} \tag{17}$$

and the fringe-width is

$$\Delta y = \frac{r\lambda}{d} \tag{18}$$

Figure 6 shows further details about Young's double slit including a photograph of the actual fringes that occur. There are several other devices which are similar to the Young's double slit set-up–these include **Fresnel's mirrors**, **Fresnel biprism** and **Lloyd's mirror**. The analysis of interference in each of them appear as problems in **Tutorial 6**.

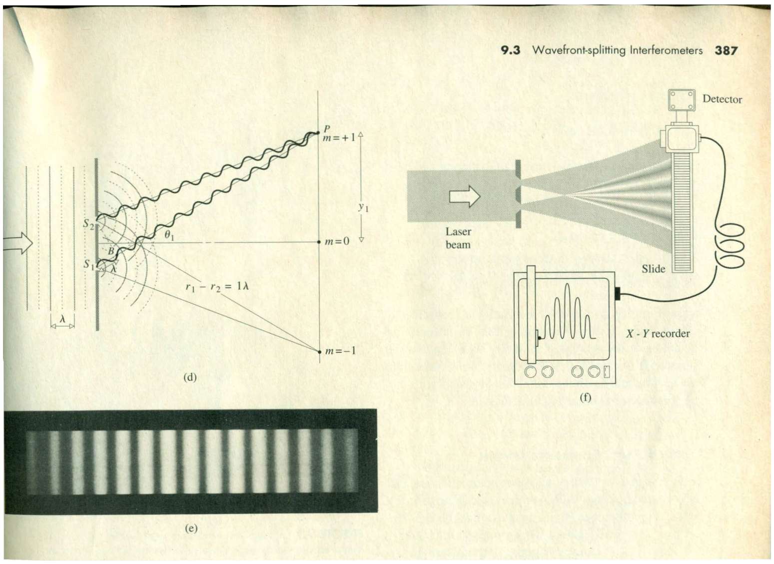


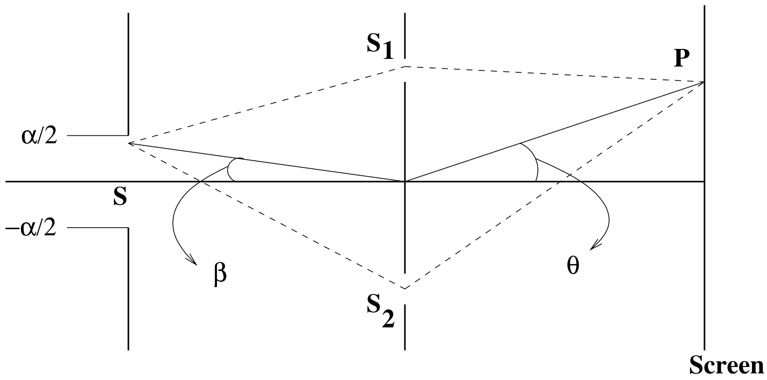
FIG. 6: Young's double slit, figure from Hecht, Optics

The Young's double slit experiment can be used to measure the spatial coherence of a source. Let us assume that the source from which light is incident on the double slit system is wide. Its angular width is α, i.e. w.r.t the origin it spreads over the angles $\frac{\alpha}{2}$ to $\frac{\alpha}{2}$. Let us consider a point at an angular position β (where $\beta \neq 0$). If we consider the intensity distribution due to this point source we will find that the intensity changes to

$$I(\theta, \beta) = 2I_0 \left(1 + \cos \frac{2\pi d\,(\theta + \beta)}{\lambda}\right) \tag{19}$$

Exercise: Show that the above expression is indeed correct.

If we now superpose the intensities due to the point sources across the width α (note we are not considering interference between these sources, all the point sources are mutually incoherent), we will get



Young's double slit: wide slit, spatial coherence.

FIG. 7: Young's double slit, wide slit, spatial coherence

$$I(\theta) = \frac{1}{\alpha}\int_{-\frac{\alpha}{2}}^{\frac{\alpha}{2}} I(\theta,\beta)d\beta = 2I_0 + \frac{2I_0}{\alpha}\left(\sin\frac{2\pi d\left(\theta - \frac{\alpha}{2}\right)}{\lambda} - \sin\frac{2\pi d\left(\theta + \frac{\alpha}{2}\right)}{\lambda}\right)\frac{\lambda}{2\pi d} \tag{20}$$

$$= 2I_0\left[1 + \frac{\sin\frac{\pi d\alpha}{\lambda}}{\frac{\pi d\alpha}{\lambda}}\cos\frac{2\pi d\theta}{\lambda}\right] \tag{21}$$

Therefore, we have

$$I = 2I_0\left(1 + (sinc\ u)\ \cos\frac{2\pi d\theta}{\lambda}\right) \tag{22}$$

where $u = \frac{\pi d\alpha}{\lambda}$. Note that we once again get a *sinc* function in the final expression. What does this *sinc* function do. If $u \to 0$ then the *sinc* function becomes 1 and we get back our earlier expression for the intensity. $u \to 0$ means that $\alpha \to 0$ and the slit is very close to a point. But if $u = m\pi$ $(m \neq 0)$ then we find that the so-called interference contribution vanishes (i.e. there are no fringes at all) and $I = 2I_0$! Thus, the *sinc* u piece seems to decide whether we will see clear fringes or not–in other words, it tells us about the **visibility** of fringes.

If we rewrite the intensity as

$$I = 2I_0\left(1 + C_{12}\ \cos\frac{\pi d\theta}{\lambda}\right) \tag{23}$$

then, we can show quite easily that

$$C_{12} = \frac{I_{max} - I_{min}}{I_{max} + I_{min}} = V \tag{24}$$

In general, the **visibility** is defined as

$$V = |C_{12}| = \frac{I_{max} - I_{min}}{I_{max} + I_{min}} \tag{25}$$

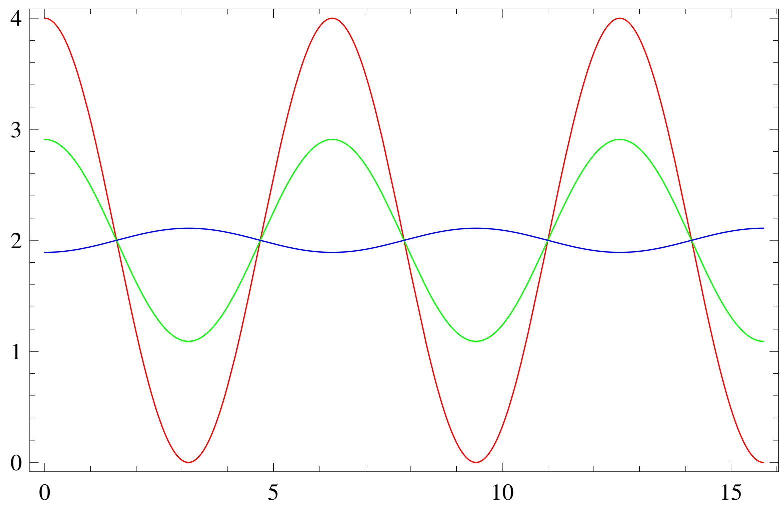


FIG. 8: Young's double slit, spatial coherence, $u = 1$ (red), $u = 2$ (green), $u = 10$ (blue)

If $V = 1$ then the fringes are sharply visible. If $V = 0$ they are not visible. Why do the fringes loose visibility? The cause seems to be the finite width of the source which results in various different wavefronts which illuminate the double slit–thereby destroying spatial coherence. The visibility is thus a measure of spatial coherence. While doing the Young's experiment, one can measure the maximum and minimum intensities and calculate V–it will surely never be exactly equal to one and its value will be able to tell us about the extent of the primary source.