Light and Interference

What is light?

Einstein's brilliant notion about "photons" still does not fully illuminate the nature of light.

The photon is a particle-like model, but clearly light does not actually consist of classical particles.

Particles do not diffract.

They do not interfere.

But light does.

Yet, this model demolishes the idea that light is a classical wave.

Indeed, the nature of electromagnetic radiation after Einstein seemed more ephemeral than ever; depending on the physical conditions, light seemed to behave either like a classical wave or like a classical particle.

The more one thinks about this so-called wave-particle duality, the more con- fusing it seems.

Understanding Diffraction the Classical Way - First Pass

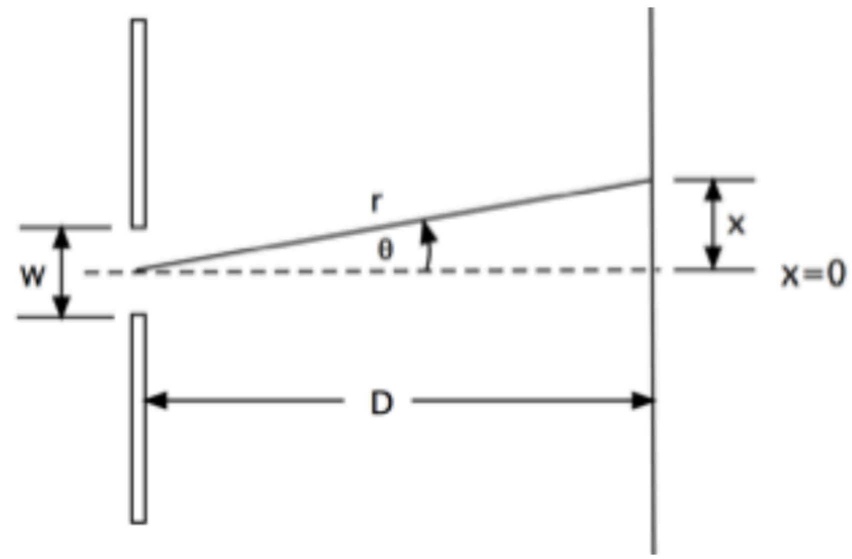
Diffraction, which was first observed by Leonardo daVinci, is often considered to be the signature of a wave.

Diffraction occurs when ripples in a pond encounter a pair of logs that are close together, when light passes through a narrow slit in a window shade, or when x-rays scatter from a crystal.

In each case, we can explain the distinctive pattern that forms using classical wave theory.

Let us quickly review the most important characteristics of wave phenomena.

A schematic of a single-slit diffraction experiment with light is shown in the figure below.



Single slit experimental setup

The above diagram represents a highly simplified single-slit diffraction apparatus.

At the detector, the distance x is measured from a point on a perpendicular(dotted line) from the midpoint of the slit.

The analysis of the experiment is performed in terms of the radial distance r from the midpoint of the slit to the point at x and the corresponding angle θ.

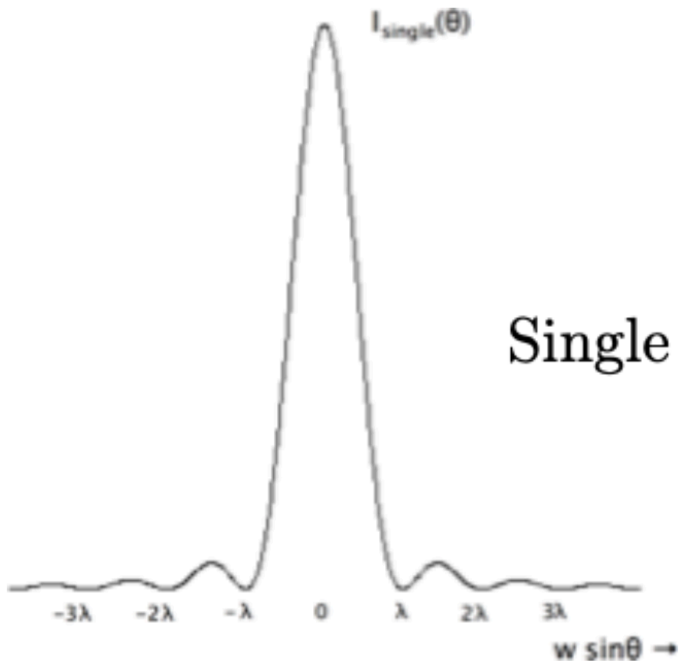
In the actual experiment D >> w, i.e., the width of the slit is very much smaller than the distance to the screen(the detector).

Monochromatic (single frequency or wavelength) light of frequency f from a distant source is incident on a barrier in which there is a single slit of width w.

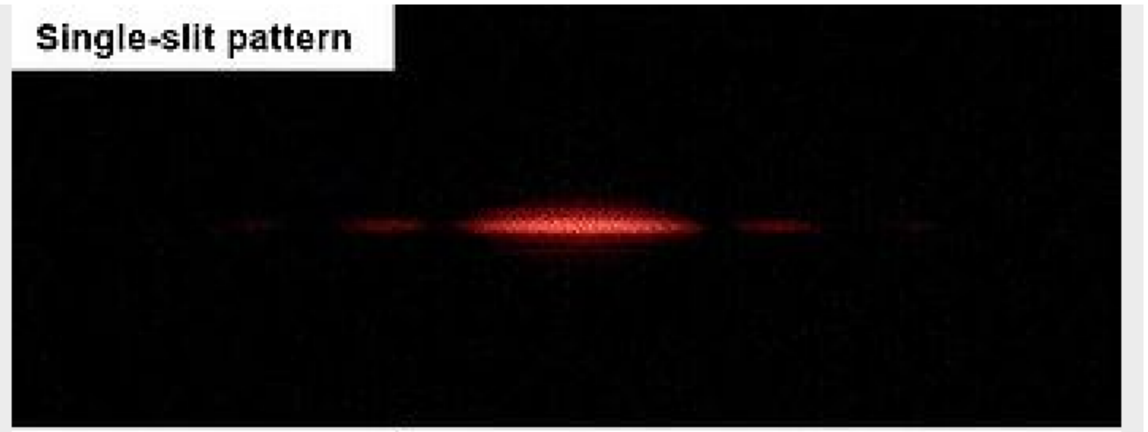
The width of the slit must be comparable to wavelength $\lambda = c/f$ of the radiation(typically 5×10^{-5} cm) if the slit is to appreciably diffract the light; for example, to diffract visible light enough to observe this phenomenon, we require a slit of width $w \approx 10^{-4}$ cm.

Light diffracted by the slit falls on a detector such as a photographic plate or a photocell or a screen, located at a distance D far to the right of the slit.

The detector measures the energy delivered by the diffracted wave as a function of the distance x as shown in the figure below.



Single slit diffraction pattern.



Light diffracted by the single-slit apparatus forms a beautiful pattern at the detector as shown above.

This pattern is characterized by a very bright central band located directly opposite the center of the slit, surrounded by a series of alternating bright and dark regions secondary bands.

If we experiment with the frequency control of the light source and study the resulting diffraction patterns, then we find that the separation between adjacent bright bands is proportional to the wavelength λ of the incident radiation.

The phenomenon is called **Frauenhofer** diffraction.

To understand the origin of this pattern, we must digress and spend some time studying wave motion.

First, we ask what is a wave?

The essential feature of wave motion is that a disturbance of some kind is transmitted from one place to another.

A local effect can be linked to a distant cause and there is a time lag between cause and effect that depends on the properties of the medium and finds its expression in the velocity of the wave.

The most important wave for our purposes is the so-called traveling wave (assume propagation in the x−direction) described mathematically by the expression

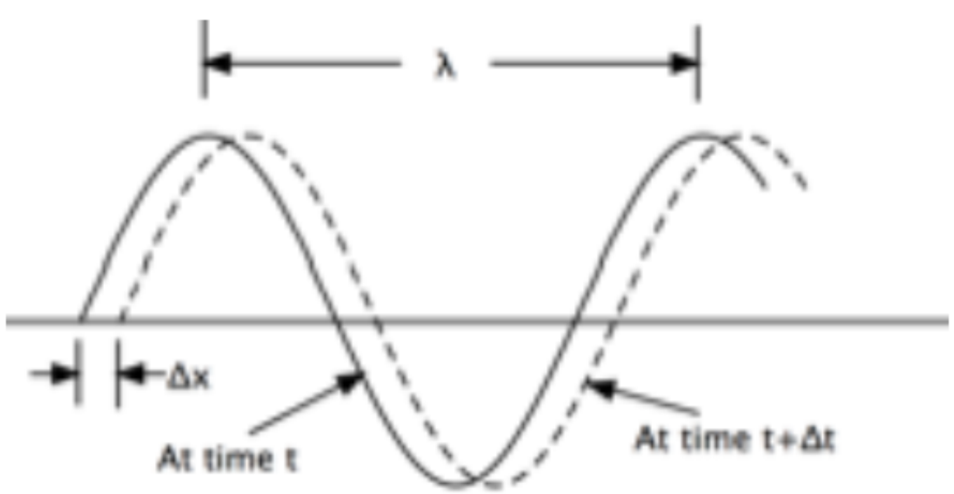
$$y(x,t) = A\sin\left(\frac{2\pi}{\lambda}(x - vt)\right)$$

where y(x,t) describes the **shape** of the wave in space and time, A is the wave amplitude, λ is the **wavelength**, v is **velocity of propagation** of the wave, and f = v/λ is the **frequency** of the wave.

If we imagine that time is frozen at some particular instant (i.e., we take a photograph of the wave), say at t = 0, the shape of the wave in space(as shown by the solid or dashed line in the figure below) is a sine wave

$$y(x,0) = A\sin\left(\frac{2\pi}{\lambda}x\right) = A\sin(kx)$$

with a distance λ between any pair of identical(having the same value and slope) points on the wave (the wavelength).



Traveling wave at a fixed instant of time.

Now let us fix attention on any one value of y, corresponding to certain values of x and t, and ask where we find that same value of y at a slightly later time t + Δt.

If the appropriate location is designated as x + Δx, then we must have

$$y(x,t) = y(x + \Delta x, t + \Delta t)$$

or

$$\begin{aligned}\sin\left(\frac{2\pi}{\lambda}(x - vt)\right) &= \sin\left(\frac{2\pi}{\lambda}(x + \Delta x - v(t + \Delta t))\right)\\ &= \sin\left(\frac{2\pi}{\lambda}(x - vt) + \frac{2\pi}{\lambda}(\Delta x - v\Delta t)\right)\end{aligned}$$

This says that the values of Δx and Δt are related by the equation

$$\Delta x - v\Delta t = 0 \Rightarrow v = \frac{\Delta x}{\Delta t}$$

This implies that the wave(the propagating disturbance) as a whole is moving in the positive x–direction with speed v as shown in the figure above, where we see the incremental displacement of wave traveling in the positive x–direction.

If we look at the wave at only one location (say x = 0) we have the expression

$$y(0,t) = -A\sin\left(\frac{2\pi}{\lambda}vt\right) = -A\sin(2\pi ft) = -A\sin(\omega t)$$

or the shape(any point on the wave) of the wave oscillates with frequency f.

Note that $\omega = 2\pi f$ is called the angular frequency.

The wave is usually written in the useful form

$$y(x,t) = A\sin(kx - \omega t) \quad k = \frac{2\pi}{\lambda} \;, \; \omega = 2\pi f$$

It turns out that waves can be combined or added together.

The process is called **superposition**.

Suppose that we have two waves traveling over the same path (x–direction).

Then the resultant shape of the propagating disturbance at any instant of time is given by the *simple sum* of the two shapes, that is, if we have two propagating waves

$$y_1(x,t) = A\sin\left(\frac{2\pi}{\lambda_1}(x - vt)\right)$$
$$y_2(x,t) = A\sin\left(\frac{2\pi}{\lambda_2}(x - vt)\right)$$

then the total disturbance at any point in space and time is given by

$$y(x,t) = y_1(x,t) + y_2(x,t) = A\sin\left(\frac{2\pi}{\lambda_1}(x - vt)\right) + A\sin\left(\frac{2\pi}{\lambda_2}(x - vt)\right)$$

Since we have assumed that the two waves are propagating with the same speed, it turns out that the combined disturbance will move with unchanging shape.

The shape of the combination or superposition is most easily determined if we put t = 0 (since we get the same shape for any t value).

We then have

$$y(x,0) = A\sin\left(\frac{2\pi}{\lambda_1}x\right) + A\sin\left(\frac{2\pi}{\lambda_2}x\right)$$

If the two wavelengths are not very different from one another, then we get the superposition of two traveling waves of slightly different wavelength as shown the figure below, that is, we get a high frequency sinusoidal wave modulated by a low frequency sinusoidal wave.

Superposition of two traveling waves of slightly different wavelengths.

Clearly the superposed wave shape is a combination of a very high frequency (small wavelength) wiggle modulated by a low frequency(large wavelength) outer envelope.

In discussing superposed waves (and interference and diffraction later) it is convenient to introduce the reciprocal of the wavelength.

The quantity $k = 1/\lambda$ is called the wave number; it is the number of complete wavelengths per unit distance (not necessarily an integer).

In terms of wave numbers, the equation for the superposed wave form can be written as follows:

$$y(x,0) = A\left[\sin\left(2\pi\bar{k}_1 x\right) + \sin\left(2\pi\bar{k}_2 x\right)\right] \qquad \text{where } \bar{k}_1 = 1/\lambda_1 \text{ and } \bar{k}_2 = 1/\lambda_2.$$

Using the trigonometric identity

$$\sin A + \sin B = 2\cos\frac{1}{2}(A-B)\sin\frac{1}{2}(A+B)$$

that you learned in high school, we can now write this as

$$y(x,0) = 2A\underbrace{\cos\left[\pi(\bar{k}_1 - \bar{k}_2)x\right]}_{\text{low frequency}}\underbrace{\sin\left[\pi(\bar{k}_1 + \bar{k}_2)x\right]}_{\text{high frequency}}$$

With this basic knowledge of waves we can now return to the discussion of the single slit diffraction pattern.

The Huygens-Fresnel Principle

Suppose that a disturbance occurs at some point in a tank of water - a small object is dropped into the water, for example, or the surface is touched with a pencil point.

Then an expanding circular wave pulse is created.

The observed pulse expands at some well-defined speed v.

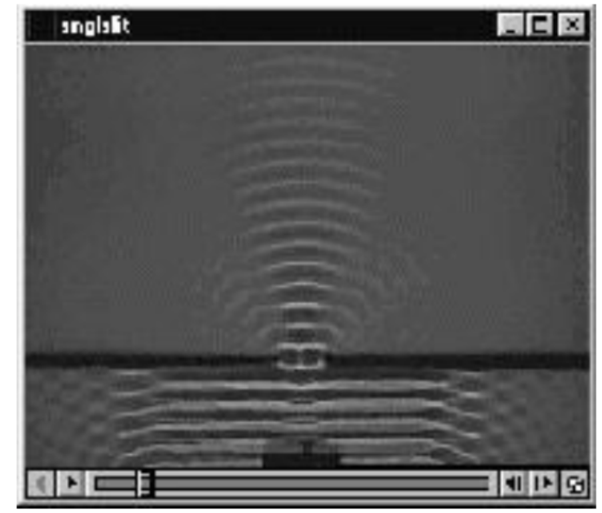
If the pulse is created at the origin at $t = 0$, then particles of the medium (water) at a distance r from the origin are set in motion at $t = r/v$.

It was Huygens' view that the effects occurring at $r + \Delta r$ at time $t + \Delta r/v$ were directly caused by the agitation of the medium at r at time t(cause and effect), thus treating the disturbance very explicitly as something handed on from point to adjacent point through the medium.

This picture of things was probably suggested by the observed behavior of ripples on water.

In particular, if wave traveling outward from a source encounter a barrier with only a tiny aperture in it (tiny meaning a width small compared to the wavelength), then this aperture appears to act just like a new point source, from which circular wave spread out as shown in the two figures below.

Theoretical simulation



Experimental water wave

Double-Slit Interference using Huygens Principle

We now consider what happens when an advancing wavefront is obstructed by barriers.

From the standpoint of Huygens' principle, each unobstructed point on the original wavefront acts as a new source and the disturbance beyond the barrier is the superposition(sum) of all the waves spreading out from these secondary sources.

For a traveling circular(spherical) wave (not one-dimensional as we have been discussing) starting out from r = 0 at t = 0, we have the form

$$y(r,t) = A\cos\left(\frac{2\pi}{\lambda}(r - vt) + \beta\right) = A\cos\left(\omega(t - r/v) - \beta\right)$$

The quantity

$$\frac{2\pi}{\lambda}(r - vt) + \beta$$

phase or the value of the phase at r = 0 and t = 0.

Because all of the secondary sources are derived from the same original wave, there is a well-defined relation between their initial phases.

This says that all the sources, primary and secondary, are **in phase** or are **coherent**.

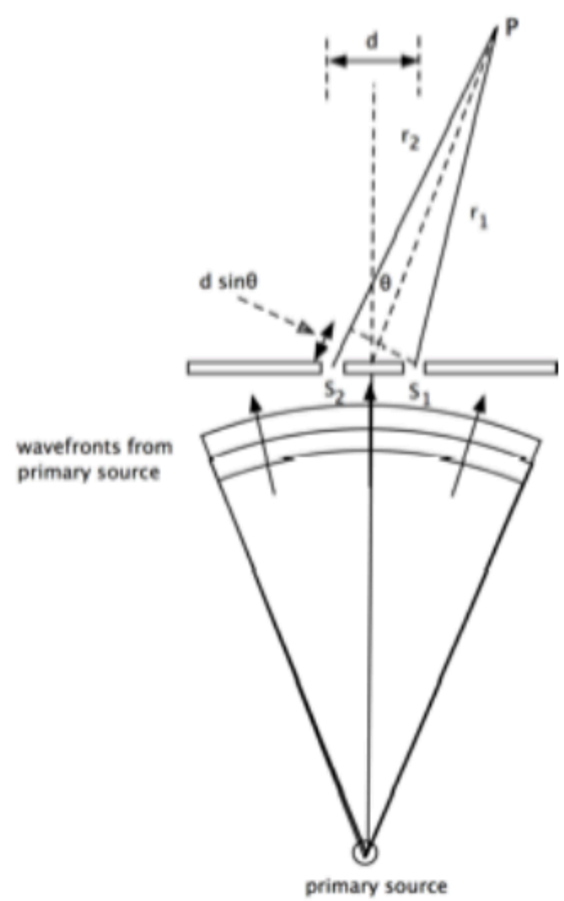
If this were not true, they would be **incoherent**.

This condition of coherence implies, in turn, a definite phase relation among the secondary disturbances as they arrive at some distance point beyond the barrier.

As a result there exists a characteristic interference pattern in the region on the far side of the barrier that does not fluctuate in appearance.

The simplest situation, and one that is basic to the analysis of all others, is to have the original wave completely obstructed except at two arbitrarily narrow apertures.

In this two-dimensional system, the two slits act as new point sources, according to Huygen's principle, as shown in the figure below.



Spherical wave interacting with two slits and generating a pattern

In the figure above we indicate a wavefront approaching two slits S_1 and S_2, which are assumed to be very narrow (but identical).

For simplicity we assume that the slits are the same distance from some point source which acts as the primary source of the wave. Thus, the secondary sources are completely in phase with each other, that is, the phase $\omega(t - r/v)$ is the same for the primary wavefront when it arrives at both slits and thus for for each new source (we choose $\beta = 0$ for simplicity).

If the original source generates sinusoidal waves, then the secondary sources also generate sinusoidal waves.

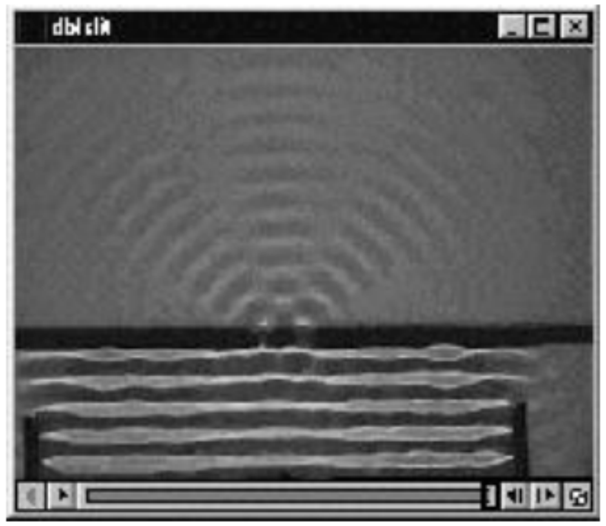
At an arbitrary point P(could be on a screen), the disturbance is obtained by adding together the contributions arriving at a given instant from S_1 and S_2.

In general we need to consider two characteristic effects:

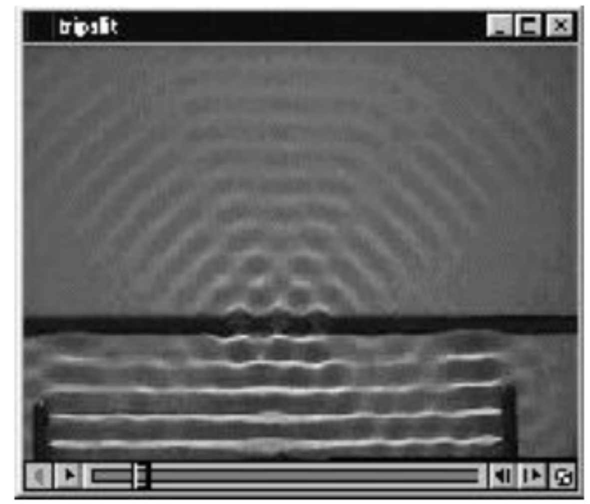
1. The disturbances arriving at P from S_1 and S_2, are different in amplitude(value of A) for two reasons. First, the distances r_1 and r_2 are different, and the amplitude generated by an expanding disturbance falls off with increasing distance from the source (like 1/r). Second, the angles θ_1 and θ_2 are different and this also affects the amplitudes. We will concentrate on situations for which the distances r_1 and are very large compared to the separation between the slits d. In these situations, the differences between the amplitudes from the two slits at P is negligible and we ignore the difference.

2. There is a phase difference between the disturbances at P corresponding to the different amounts of time it takes each wave to arrive at P. The time difference is $(r_2 - r_1)/v$ where v is the wave speed.

It is this phase difference which dominates the general appearance of the resultant interference pattern. Ripple tank examples for two and three slits are shown in the two figures below.



Two slit pattern - water waves



Three slit pattern - water waves

Clearly there exist nodal lines along which the resultant disturbance is almost zero (called **destructive interference**) at all times.

It is easy to calculate their positions.

Between the nodal lines there is another set of lines of maximum displacement, where the resultant disturbance reaches its greatest value (called **constructive interference**).

The important parameter that governs the general appearance of the interference pattern is the dimensionless ratio of the slit separation d to the wavelength λ.

The condition for interference maxima is

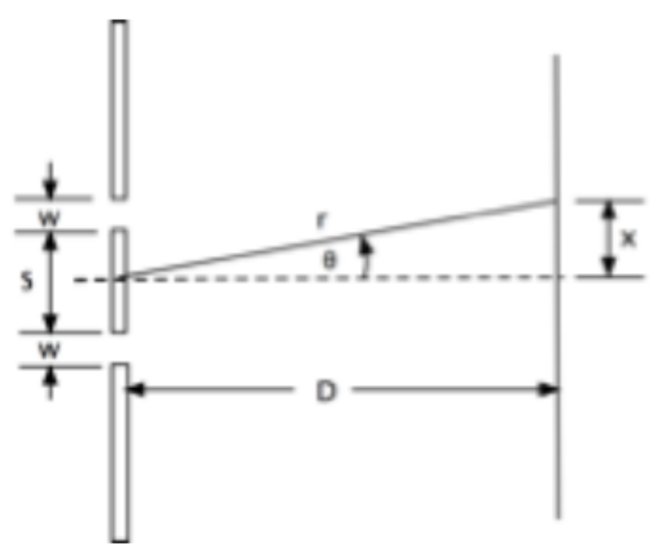
$$d \sin\theta_n = n\lambda \rightarrow \sin\theta_n = \frac{n\lambda}{d}$$

and the amplitude at some arbitrary direction is given by

$$A(\theta) = 2A_0 \cos\left(\frac{\pi d \sin\theta}{\lambda}\right)$$

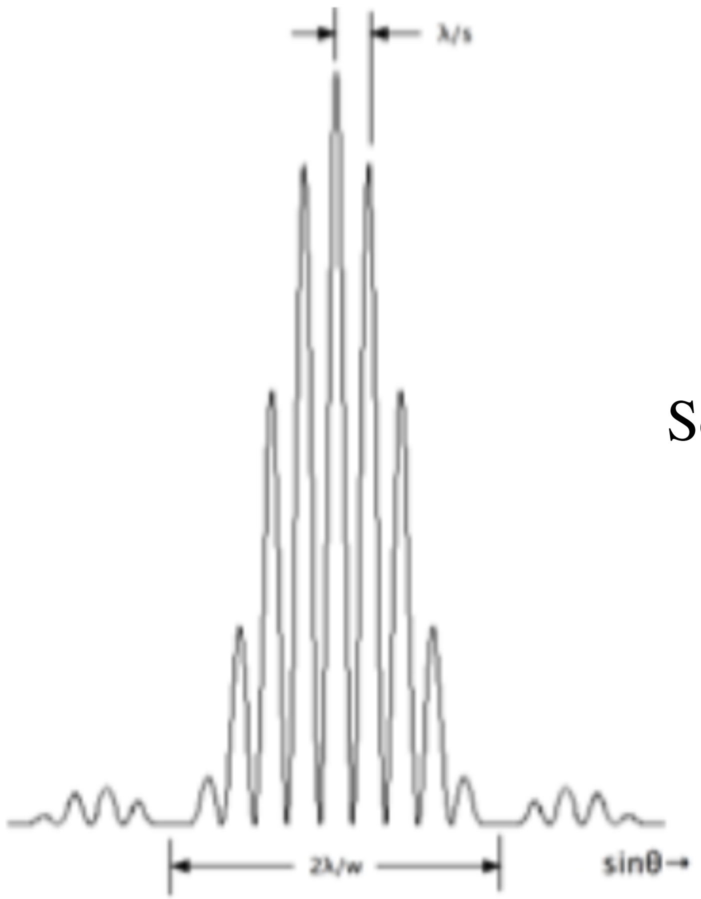
We then see that the interference at a large distance from the slits is essentially a directional effect, that is, if the positions of the nodes and the interference maxima are observed along a line parallel to the line joining the two slits (we put a screen out there), the linear separation of adjacent maxima (or zeroes) increase in proportion to the distance D from the slits ($\sin\theta \approx x/D$).

If we now turn to the schematic for the double slit setup shown in the figure below

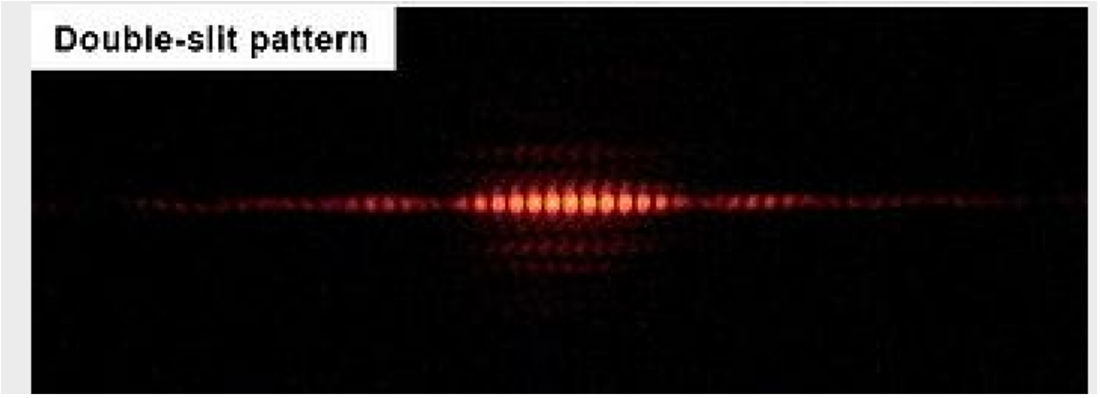


Two slit schematic

then, in this case, the detector or screen shows an interference pattern like the one in the figure below ($s \gg \lambda$)



Screen interference pattern



The separation between maxima is

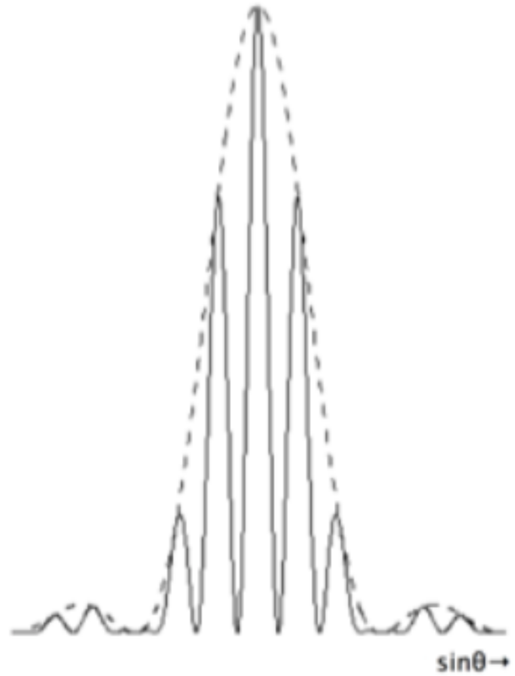
$$\sin\theta = \frac{\lambda}{s} \rightarrow x = D\sin\theta = \frac{D\lambda}{s}\left(\text{or } \frac{D\lambda}{d}\right)$$

The pattern above does not assume the slit width w is zero. The equation for this pattern is

$$I_{double}(\theta) = 4I_0\left[\frac{\sin\left(\pi\frac{w\sin\theta}{\lambda}\right)}{\pi\frac{w\sin\theta}{\lambda}}\right]^2\cos^2\left(\pi\frac{s\sin\theta}{\lambda}\right)$$

where $I_0 = A_0^2$

The intensity wiggles are shown in the figure below where they are modulated by the single-slit pattern



Theoretical prediction

You might think that the same effect could be achieved by firing two lasers (why bother with the slits?).

Unfortunately, this doesn't work.

As the two lasers are independent light sources, there is no guarantee that the phases of the waves from each laser start out in a fixed pattern.

The path difference still produces a phase difference at the screen, but whether this is constructive or destructive depends on how the waves started out, which could vary from one moment to the next.

All you would get is a rapidly and randomly shifting pattern, which just looks like featureless illumination on the screen.

A single laser beam covering both slits will split into two parts with the same phase as the light passes through the slits.

It is worth emphasizing this aspect of the experiment: the effect can only be made to work when waves from a single source pass through both slits at the same time.