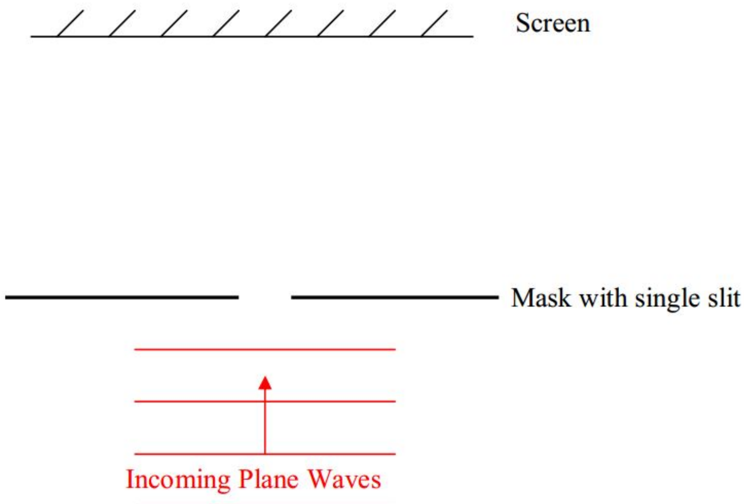
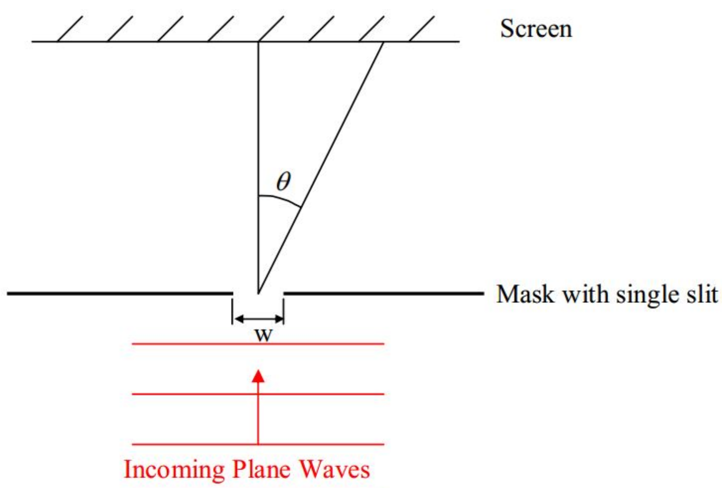
B23: Single-Slit Diffraction

Single-slit diffraction is another interference phenomenon. If, instead of creating a mask with two slits, we create a mask with one slit, and then illuminate it, we find, under certain conditions, that we again get a pattern of light and dark bands. It is not the same pattern that you get for two-slit interference, but, it's quite different from the single bright line in the straight-ahead direction that you might expect. Here's how it comes about. Firstly, here's the setup:

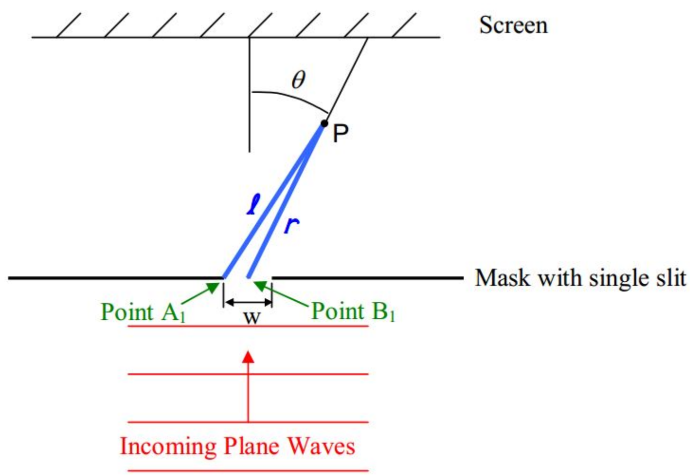


Again, we get a bright fringe in the straight-ahead position on the screen. From there, working out to either side, we get bands that alternate between dark and bright. The first maximum to the right or left of the central maximum is not nearly as bright as the central maximum. And each maximum after that is less bright than the maximum preceding it. As far as the analysis goes, I want to start with the minima. Consider an imaginary line extending out from the midpoint of the slit all the way to the screen.



So now the question is, "Under what conditions will there be completely destructive interference along a line such as the one depicted to be at angle θ above?" To get at the answer, we first divide the slit in half. I'm going to enlarge the mask so that you can see what I mean.

Now I imagine dividing side A up into an infinite number of pieces and side B up the same way. When the slit is illuminated by the light, each piece becomes a point source. Consider the first point source (counting from the left) on side A and the first point source (again counting from the left) on side B. These two point sources are a distance $w/2$ apart, where w is the width of the slit. If the light from these two point sources (which are in phase with each other because they are really both part of the same incoming plane wave), interferes completely destructively, at some angle θ with respect to the straight-ahead direction, then the light from the second point source on side A and the second point source on side B will also interfere with each other completely destructively because these two point sources are also $w/2$ apart. The same goes for the third-from-the-left point sources on both sides, the fourth, the fifth, and so on, ad infinitum. So, all we need is to establish the condition that makes the light from the leftmost point source on side A (overall, the leftmost point of the slit) interfere completely destructively with the leftmost point source on side B (overall, essentially the midpoint of the slit). So, consider any point P on a proposed line of minima.

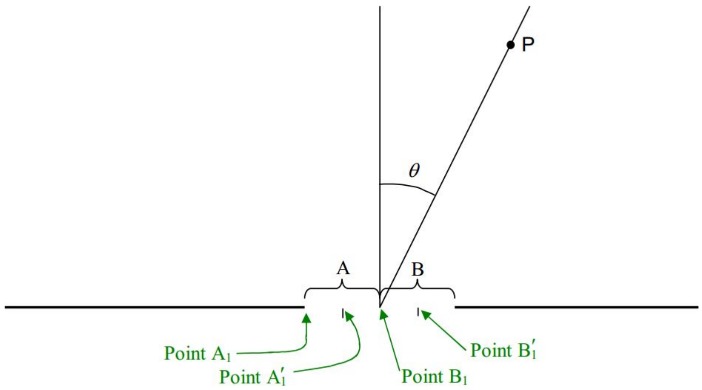


The distance between the two point sources is $w/2$. From the analysis done for the case of two-slit interference, we know that this results in a path difference $\ell - r = \frac{w}{2}\sin\theta$. And, as you know, the condition for completely destructive interference is that the path difference is half a wavelength, or, any integer number of wavelengths plus a half a wavelength. So, we have a minimum along any angle θ (less than 90°) such that:

$$(m+\frac{1}{2})\lambda = \frac{w}{2}\sin\theta \quad (m=0,1,2,\dots)$$

Now we turn our attention to the question of diffraction maxima. I should warn you that this analysis takes an unexpected turn. We do the exact same thing that we did to locate the minima, except that we set the path difference $\ell - r$ equal to an integer number of wavelengths (instead of a half a wavelength plus an integer number of wavelengths). This means that the path to point P from the leftmost point on side A (position A_1) of the slit, differs by an integer number of wavelengths, from the path to point P from the leftmost point on side B (position B_1). This will also hold true for the path from A_2 vs. the path from B_2. It will hold true for the path from A_3 vs. the path from B_3 as well. Indeed, it will hold true for any pair of corresponding points, one from side A and one point from side B. So, at point P, we have maximally constructive interference for every pair of corresponding points along the width of the slit. There is, however, a problem. While, for any pair of points, the oscillations at P will be maximal; that just means that P is at an angle that will make the amplitude of the oscillations of the electric field due to the pair of points maximal. But the electric field due to the pair of points will still be oscillating, e.g. from max up, to 0, to max down, to 0, and back to max up. And, these oscillations will not be in synchronization with the maximal oscillations due to other pairs of points. So, the grand total will not necessarily correspond to an intensity maximum. The big difference between this case and the minima case is that, in contrast to the time varying maximal oscillations just discussed, when a pair of contributions results in an electric field amplitude of zero, the electric field due to the pair is always zero. It is constant at zero. And, when every pair in an infinite sum of pairs contributes zero to the sum, at every instant in time, the result is zero. In fact, in our attempt to locate the angles at which maxima will occur, we have actually found some more minima. We can see this if we sum the contributions in a different order.

Consider the following diagram in which each half of the slit has itself been divided up into two parts:



If the path difference between “A_1 to P” and “B_1 to P”, is one wavelength, then the path difference between “ A_1 to P” and “ A_1 'to P” must be half a wavelength. This yields completely destructive interference. Likewise for the path difference between “ B_1 to P” and “ B_1 'to P”. So, for each half of the slit (with each half itself being divided in half) we can do the same kind of pair-wise sum that we did for the whole slit before. And, we get the same result—an infinite number of zero contributions to the electric field at P. All we have really done is to treat each half of the slit the way we treated the original slit. For the entire slit we found

$$(m+\frac{1}{2})\lambda=\frac{w}{2}\sin\theta \quad (m=0,1,2,\dots)$$

Here we get the same result but with w itself replaced by $w/2$ (since we are dealing with half the slit at a time.) So now we have:

$$(m+\frac{1}{2})\lambda=\frac{w}{4}\sin\theta \quad (m=0,1,2,\dots)$$

Let’s abandon our search for maxima, at least for now, and see where we are in terms of our search for minima. From our consideration of the entire slit divided into two parts, we have $(m+\frac{1}{2})\lambda=\frac{w}{2}\sin\theta$ which can be written as $(2m+1)\lambda=w\sin\theta$ meaning that we have a minimum when:

$$w\sin\theta=1\lambda,3\lambda,5\lambda,\dots$$

From our consideration of each half divided into two parts (for a total of four parts) we have $(m+\frac{1}{2})\lambda=\frac{w}{4}\sin\theta$ which can be written $(4m+2)\lambda=w\sin\theta$ meaning that we have a minimum when:

$$w\sin\theta=2\lambda,6\lambda,10\lambda,14\lambda,\dots$$

If we cut each of the four parts of the slit in half so we have four pairs of two parts, each $\frac{w}{8}$ in width, we find minima at $(m+\frac{1}{2})\lambda=\frac{w}{8}\sin\theta$ which can be written $(8m+4)\lambda=w\sin\theta$ meaning that we have a minimum when:

$$w\sin\theta=4\lambda,12\lambda,20\lambda,28\lambda,\dots$$

If we continue this process of splitting each part of the slit in two and finding the minima for each adjacent pair, ad infinitum, we eventually find that we get a minimum when $w\sin\theta$ is equal to any integer number of wavelengths.

$$w\sin\theta=1\lambda,2\lambda,3\lambda,4\lambda,\dots$$

a result which write as

$$m\lambda=w\sin\theta \quad (m=1,2,3,\dots) \tag{B23.1}$$

We still haven’t found any maxima. The only analytical way to determine the angles at which maxima occur is to do a full-fledged derivation of the intensity of the light as a function of position, and then mathematically solve for the maxima. While, this is not really as hard as it sounds, let’s save that for an optics course and suffice it to say that, experimentally, we find maxima approximately midway between the minima. This includes the straight-ahead (0°) direction except that the straight-ahead maximum, a.k.a. the central maximum, is exactly midway between its neighboring minima.

Conditions Under Which Single-Slit Diffraction and Two-Slit Interference Occurs

To see the kinds of interference patterns that we have been talking about in this and the preceding chapter, certain conditions need to be met. For instance, in order to see one set of bright fringes in the two-slit interference experiment we need monochromatic light. Translated literally, from the Latin, monochromatic means one-color. Monochromatic light is single-frequency light. Strictly monochromatic light is an idealization. In practice, light that is classified as being essentially monochromatic, actually consists of an infinite set of frequencies that are all very close to the nominal frequency of the light. We refer to the set of frequencies as a band of frequencies. If all the frequencies in the set are indeed very close to the nominal frequency of the light, we refer to the light as narrow-band radiation.

If you illuminate a single or double slit with light consisting of several discrete (individual) wavelengths of light, you get a mix of several interference/diffraction patterns. If you illuminate a single or double slit with a continuum of different frequencies, you find that minima from light of one wavelength are "filled in" by maxima and/or intermediate-amplitude oscillations of light of other wavelengths. Depending on the slit width and (in the case of two-slit interference) slit separation, and the wavelengths of the light, you may see a spectrum of colors on the screen.

In order for the kind of interference that we have been talking about to occur, the light must be coherent. The light must be temporally coherent (coherent with respect to time). While it applies to any part of a wave, I am going to talk about it in terms of crests. In temporally coherent light, one wave crest is part of the same wave that the preceding wave crest is a part of. In light having a great deal of temporal coherence this holds true for thousands of crests in a row. In light with very little temporal coherence this may hold true for only one or two crests in a row. Another way of stating it is to say that light that consists of a bunch of little wave pulses is temporally incoherent and light that consists of long continuous waves is temporally coherent. The long continuous wave can be called a "wave train". In terms of wave trains, light that is temporally incoherent consists of lots of short wave trains, whereas light that is temporally coherent consists of a relatively small number of long wave trains. The kinds of interference we have been talking about involve one part of a wave passing through a slit or slits in a mask, interfering with another part of the same wave passing through the same mask at a later time. In order for the latter to indeed be part of the same wave, the light must consist of long wave trains, in other words, it must be temporally coherent. Now, if the wave crest following a given wave crest is not part of the same wave, the distance from one wave crest to the next will be different for different crests. This means the wavelengths are different and hence the frequencies are different. Thus the light is not monochromatic. Under the opposite circumstances, the light is monochromatic. So, monochromatic light is temporally coherent.

The other condition is that the light must be spatially coherent. In the context of light that is normally incident on plane masks, this means that the wavefronts must be plane and they must have extent transverse to the direction in which the light is traveling. In the case of two-slit interference for instance, spatial coherence means that the light at one slit really is in phase with the light at the other slit. In the case of single-slit diffraction, spatial coherence means that light passing through the right half of the slit is in phase with light passing through the left half of the slit.