Chapter 4: Two-Beam Interference

Two-beam interference:

Because light waves are repetitive, with electric fields that swing alternately positive and negative (that is, they reverse direction sinusoidally), interesting things can happen when two (or more) of them arrive at the same place, but are delayed by differing amounts of time. When they arrive "in phase," so that crests meet crests, their effects "add up" or reinforce each other, and we get "constructive interference." However, if one wave arrives half a cycle behind the other (or 180° "out of phase"), so that crest meets trough, their effects cancel so that there is no net vibration, and we get "destructive interference." For intermediate phase shifts, intermediate results occur, so that the total vibration intensity can be negligible or enormous, depending on small time shifts between the waves.

We will begin by looking at a few examples of interference effects in everyday life. They are not extremely common, because the small time shifts involved are hard to control, and the effects average out if the phase shift varies during an observation time. Longer time delays can produce interference effects only for highly coherent or single frequency waves, which are also fairly rare (except in the holography lab!).

Interference examples

Soap films

Whenever kids blow soap bubbles, they enjoy the swirling play of colors on the soap film, which becomes increasingly intense until just before the film darkens and finally bursts. Similarly for oil films on water—the color of the reflected light depends on the thickness of the film. These effects are caused by the interference of waves reflected from the front and back of the films. The same time delay can cause the interference to be additive or destructive, depending on the frequency or wavelength of the light, so that red light might be reinforced and blue light extinguished in one area, and vice versa in another with just a tiny difference in film thickness.

Polarized glasses, colors observed with

Polarized sunglasses are wonderful for blocking reflected glare light, which tends to be horizontally polarized, but they cause strange color patterns to sometimes appear in car rear windows, stretched plastic sheets, and so forth. These color interference effects arise because glass and plastic sheets under mechanical stress will delay different polarizations of light by different amounts, and the sunglasses cause the two waves to combine and possibly interfere to form bands of color.

Radio fading

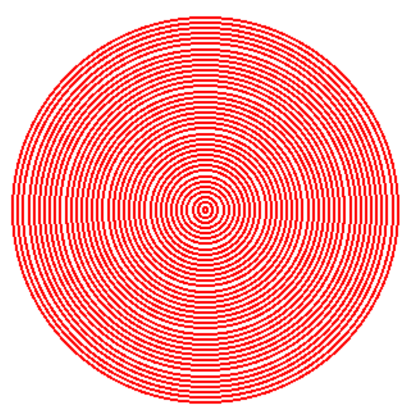
While driving around in hilly countryside, listening to the car radio, it is not uncommon for the signal to fade in and out almost randomly (this is more common with FM radios). This is caused not because the car is moving in and out of "radio shadows" caused by the hills, but because waves are reflected by the hills and combine to reinforce or cancel, depending on location (called "multi-path reception"). Similarly, a propeller plane flying over a TV antenna can cause a fluttering of the image due to the reception of multiple weak signals.

Audio beats

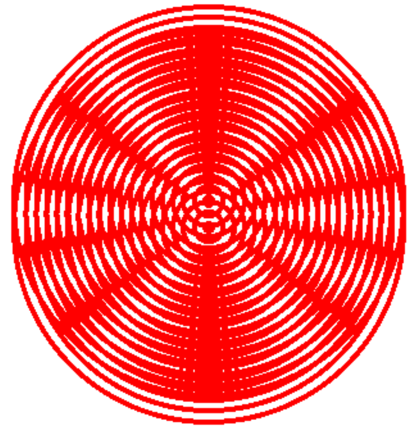
Familiar to musicians is the phenomenon of "acoustical beats," often used for tuning up stringed instruments such as guitars and pianos. Two strings are plucked, and one tuned until the "vibrato" or "beating" effect becomes slower and slower, and eventually freezes. When there is beating, the two strings are at nearly the same frequency, but slowly going in and out of phase. Their emissions thus add up and cancel alternately. When they are exactly the same frequency, and are "phase-synchronized" or "coherent," the sound can be weak or strong depending on whether they are "in" or "out" of phase, or somewhere in between. Usually, they don't stay tuned for long, though! Likewise, the amount of sound from a tuning fork varies markedly as it turns—the tines are out of phase at some angles.

Moiré fringes

You may not have thought of it this way, but the *moiré* patterns ("more-ay" is an English mispronunciation of a French mis-transcription, "mwah-ray," of a Persian word (unpronounceable?) for a watermarked taffeta fabric) that are formed between two repetitive optical patterns is also a case of wave interference. You may have seen these between two pieces of window screening, chain link fences at a distance, muslin curtains, and so forth (printers worry about them in color half-tone printing, too!). The mathematics of multiplying two repetitive or wave-like patterns is the same as the mathematics of adding and squaring them (that is, interference!), a fact that we will exploit on several occasions.



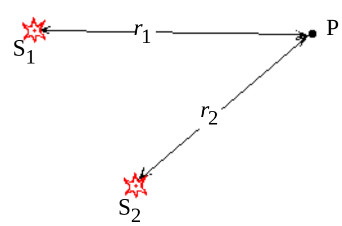
Wave interference is taught in many high school physics classes with the aid of ripple tanks. There, two bobbing corks launch shallow waves across a "pond" of constant depth, and the overall depth of the combined waves is a measure of the total "intensity." That is a wonderful way to learn this stuff, and I recommend that you find a ripple tank to play with if at all possible. In the meantime, we will have to depend on a simpler optical demonstration of the same effects. Luckily, we can see most of the relevant phenomena almost as clearly with moiré fringe patterns. We can think of a pattern of concentric equally-spaced circles as a “snapshot” of a slice through a spherical wave as it propagates outward from a central point source. A slow-motion movie would show the circles slowly expanding, and a new wave emerging from the center, until the pattern looked just the same as seen one oscillation period earlier.



When two spherical wave patterns are laid on top of each other, a distinctive pattern of dark and light bands, or “moiré fringes” appears (I am not aware of a logical reason to call the banded components of these patterns “fringes”—if you think of one, let me know!). The dark fringes occur where the dark rings of one pattern overlay the light rings of the other, and the lighter fringes occur where the dark rings overlay dark rings so that some of the light ring area is visible (that is, the rings are “in phase”). If the slow-motion movie of the waves were played on from this point, although the rings/waves would move outward at exactly the same speed, the regions of dark and light moiré fringes would remain in the same places. In the example shown here, one set of ripples comes from a source that is 2.5 wavelengths above the source of the other. Everywhere straight up and down, the waves arrive 2.5 cycles “out of phase” and so produce a dark region (as though the waves were “canceling each other out”). But in areas exactly to the right and left, the waves arrive “in phase” and so those areas are brighter. As we move around the edge of the overlapped circles, we move from areas where the waves are “in phase,” half a cycle out of phase, a whole cycle out of phase (and thus back in phase), and so on, until we reach the maximum phase difference of 2.5 cycles.

Quantitative discussion of interference contrast:

We can fairly easily describe the interference effects between <u>two</u> mutually coherent light sources in more quantitative and mathematical terms (more sources is much harder!). By “mutually coherent,” we mean that each source is not only a point source of well-defined frequency, but that the oscillation of each source is locked into phase with that of the other. Typically, this means that the two beams of light came from the same laser, via a system of beamsplitters and mirrors, but we can think of them as two separate sources, S_1 and S_2, that are somehow synchronized (by atomic clocks, etc.). We can imagine that the phase of the emission from one can be adjusted at will, and that the observation point, *P*, can be freely moved about in space (well, 2-D space in this diagram) to change the time delays between it and either or both of the two sources.



Each source, S_i, emits a spherical wave, which arrives at the observation point, *P*, after a time delay of $\tau = r_i/c$. This causes a phase delay of $\phi = (2p/\lambda)\, r_i$. The absolute phase of each wave is unobservable, because optical frequencies are so high, but the phase *difference* between the two waves will determine whether there is no vibration intensity observed at *P*, a little, or a lot. If the amplitudes of the two waves are <u>equal</u> at *P*, and they arrive “in phase,” they will add together and the total intensity (defined as the average of the square of their sum) will be four times as great as the intensity of either of the waves separately. If they arrive “out of phase,” one will be positive when the other is negative, and they will cancel out exactly, and the total intensity will be zero! This itself is odd enough. Consider that there could be one laser death ray headed straight at you, and another (coherent) beam coming in from the side, there are supposed to be places where there is NO total intensity, where it might be safe to stand (but if one of the beams is suddenly blocked, you get fried!). The in-between cases, where the waves are not equal in amplitude and the phase difference is somewhere between 0 and p radians (0° and 180°) need more mathematics to be defined precisely.

<u>Mathematical discussion:</u>

In this section we will grind through the derivation of the “interference equation” at a simple “shop math” level, so it will take about a page to finish.

Depending on your own level of math background, you might be able to show the same results in only three lines using complex algebra and phasor notation; that approach will immediately follow this section so that you can see how we link the two together for possible future reference.

The expression for the wave amplitude (electric field) measured at point P from source S_1 is given by:

$$E_1(P,t) = \frac{U_1}{r_1}\sin\left(2\pi\nu t - \frac{2\pi}{\lambda} r_1\right) . \quad (1)$$

Likewise, the amplitude for the second wave, from source S_2, is given by

$$E_2(P,t) = \frac{U_2}{r_2}\sin\left(2\pi\nu t - \frac{2\pi}{\lambda} r_2\right) . \quad (2)$$

That is, the two waves have the same frequency, ν, and thus the same wavelength, λ. To make life a little more generalized, we will refer to these waves in the more general terms of their amplitudes and phases measured at P.

$$E_1(P,t) = a_1(P)\sin\left(2\pi\nu t - \phi_1(P)\right) , \quad (3)$$

and similarly for the wave from S_2,

$$E_2(P,t) = a_2(P)\sin\left(2\pi\nu t - \phi_2(P)\right) . \quad (4)$$

Intensity: The irradiance, or "intensity" as we will more commonly call it, is proportional to the average over time (a brief time, perhaps a few microseconds) of the square of the magnitude of the electric field vector of the total light field. The proportionality factor depends on the units of the discussion; we will use MKS for the time being, so that the factor becomes $\varepsilon_0 c$. It is the squaring and averaging that produces all of the interesting results, not the units.

The total electric wavefield is, summing the two waves:

$$E_{\text{total}}(P,t) = a_1(P)\sin\left(2\pi\nu t - \phi_1(P)\right) + a_2(P)\sin\left(2\pi\nu t - \phi_2(P)\right) . \quad (5)$$

We are discussing the wave's electric field as a scalar quantity here, so the squared-magnitude is simply the arithmetic square (with P omitted on the right side to save space):

$$\begin{aligned} \left|E_{\text{total}}(P,t)\right|^2 &= \left(a_1 \sin\left(2\pi\nu t - \phi_1\right) + a_2 \sin\left(2\pi\nu t - \phi_2\right)\right)\cdot\left(a_1 \sin\left(2\pi\nu t - \phi_1\right) + a_2 \sin\left(2\pi\nu t - \phi_2\right)\right) \\ &= \left(a_1^2 \sin^2\left(2\pi\nu t - \phi_1\right)\right) + \left(a_2^2 \sin^2\left(2\pi\nu t - \phi_2\right)\right) + 2a_1a_2 \sin\left(2\pi\nu t - \phi_1\right)\sin\left(2\pi\nu t - \phi_2\right) \quad (6) \\ &= \left(a_1^2 \sin^2\left(2\pi\nu t - \phi_1\right)\right) + \left(a_2^2 \sin^2\left(2\pi\nu t - \phi_2\right)\right) + a_1a_2 \cos\left(\phi_1 - \phi_2\right) - a_1a_2 \cos\left(4\pi\nu t - \phi_1 - \phi_2\right). \end{aligned}$$

Note that the last step invokes some familiar trig identities. Recalling that the time average of sin$t$ (and cos$t$) is 0.0, and the time average of $\sin^2 t$ is 0.5, we find that

$$\left\langle \left|E_{\text{total}}(P,t)\right|^2 \right\rangle_{\text{time avg.}} = \frac{a_1^2(P)}{2} + \frac{a_2^2(P)}{2} + a_1(P)\,a_2(P)\cos\left(\phi_1(P) - \phi_2(P)\right) . \quad (7)$$

The total intensity is then (in MKS units),

$$\begin{aligned} I_{\text{total}}(P) &\equiv \varepsilon_0 c \left\langle \left|E_{\text{total}}(P,t)\right|^2 \right\rangle_{\substack{\text{time}\\\text{average}}} \quad (\text{W/m}^2) \\ &= \varepsilon_0 c \frac{a_1^2(P)}{2} + \varepsilon_0 c \frac{a_2^2(P)}{2} + \varepsilon_0 c\, a_1(P)\,a_2(P)\cos\left(\phi_1(P) - \phi_2(P)\right) \quad (8) \\ &= I_1(P) + I_2(P) + 2\sqrt{I_1(P)\cdot I_2(P)}\cos\left(\phi_1(P) - \phi_2(P)\right) . \end{aligned}$$

$$I_{\text{total}} = I_1 + I_2 + 2\sqrt{I_1 \cdot I_2}\cos\left(\phi_1 - \phi_2\right)$$

It is the last form that is the most familiar in optics, in which the proportionality constants even out and the result is expressed in terms of the intensities of the waves by themselves and the cosine of the phase difference between them.

complex amplitude proof

The same proof can be compressed if we consider instead the complex amplitude, $u_i(P)$, of each of the waves. The complex amplitude of each wave, and its complex conjugate (denoted by an asterisk), are defined as (using Gaskill's notation here[1]),

$$u_i(P) = a_i(P)\, e^{j\phi_i(P)}, \qquad u_i^*(P) = a_i(P)\, e^{-j\phi_i(P)}. \tag{9}$$

The real measurable field may be recovered as

$$E_i(P,t) = \mathrm{Im}\left\{u_i^*(P)\, e^{j2\pi\nu t}\right\}.$$

Consistent with Eq. 8, we define the intensity of a single wave in terms of its complex amplitude as

$$I_i(P) \equiv \frac{\varepsilon_0 c}{2} u_i(P)\, u_i^*(P) = \frac{\varepsilon_0 c}{2} a_i^2\,. \tag{10}$$

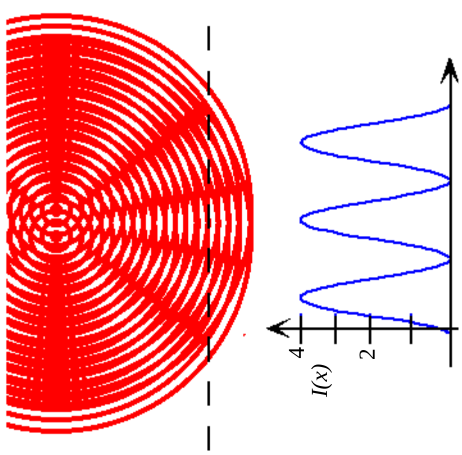
Similarly, for a summation of many waves, the total intensity in terms of the total complex amplitude, $u_{\text{total}}(P) = \sum_i u_i(P)$, is

$$I_{\text{total}}(P) = \frac{\varepsilon_0 c}{2} \left|u_{\text{total}}(P)\right|^2 = \frac{\varepsilon_0 c}{2} u_{\text{total}}(P)\, u_{\text{total}}^*(P). \tag{11}$$

With these preliminaries in place, we deal entirely in terms of the complex amplitudes, and can readily show that, letting i=1 and 2 in turn:

$$\begin{aligned} \left|u_{\text{total}}(P)\right|^2 &= u_{\text{total}}(P) \cdot u_{\text{total}}^*(P) \\ &= \left(a_1 e^{j\phi_1} + a_2 e^{j\phi_2}\right)\left(a_1 e^{-j\phi_1} + a_2 e^{-j\phi_2}\right) \\ &= a_1^2 + a_2^2 + 2a_1 a_2 \cos(\phi_1 - \phi_2)\,. \end{aligned} \tag{12}$$

This is the desired result when plugged into the definition of the intensity.



Equal beam case, conservation of energy

In the case seen above, we can imagine sketching the intensity observed along the right-hand edge of the interference pattern. Assuming that the intensities of the two beams are unity when measured separately, we see that when turned on together, we do not get a uniform reading of two, but rather that the energy "bunches up" to give four in some places, and zero in others. Simple interference patterns pose some of the mostly deeply reaching questions of modern physics. Here we see that the principle of *conservation of energy* does not always apply in the micro-scale, but only as an average over several cycles of the interference pattern.

Unequal beams; heterodyne gain

An interesting effect in interference patterns is that the variations of intensity are usually much greater than the intensity of the weaker of the two beams. That is, if a weak beam overlaps a strong beam the contrast of the fringe pattern, or its "visibility," is usually much greater than the visibility of the weak beam by itself; interference provides a kind of amplification, analogous to the "heterodyne gain" of radio electronics. Let the ratio of the beam intensities be given by $K = I_{\text{strong}}/I_{\text{weak}}$. The variation of intensity is then given by

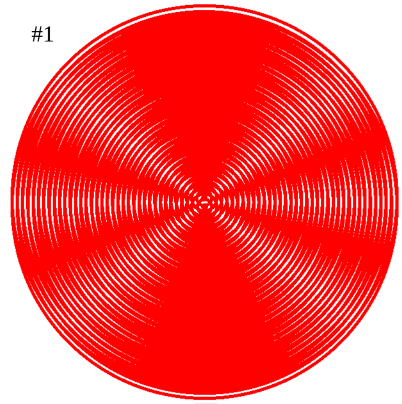
$$I_{\max} - I_{\min} = 2\sqrt{I_{\text{weak}} I_{\text{strong}}} = 2\sqrt{K}\, I_{\text{weak}}\,. \tag{13}$$

The “visibility” of a fringe pattern is defined as the ratio of the variation of the total intensity to its average intensity, times two:

$$V = \frac{I_{max} - I_{min}}{I_{max} + I_{min}}. \qquad (14)$$

A V of 0.01 is usually near the threshold of visibility of the human eye, depending on the fringe spacing, which means that a beam that is only one forty-thousandth (1/40,000) the intensity of the stronger beam (completely invisible as an incoherent addition) could produce an easily visible interference pattern! This causes lots of problems when we try to make holograms in the laboratory!

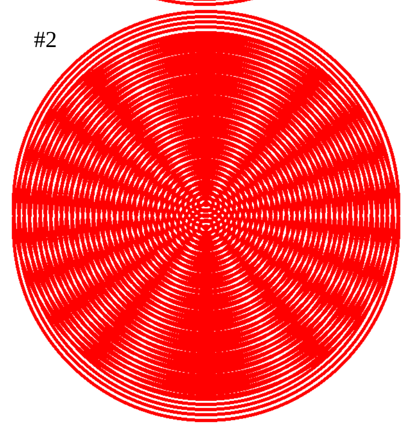
#1



The geometry of interference fringes:

We have learned about the magnitudes of interference effects, and their extreme sensitivity to weak beams, but we will generally be more interested in the geometry of these fringe effects. In particular, we will want to know *where* these moiré-like fringes are formed, and what their *spacings* and *orientations* are. These will eventually determine where light goes when it is diffracted by a hologram (as opposed to how much light goes there).

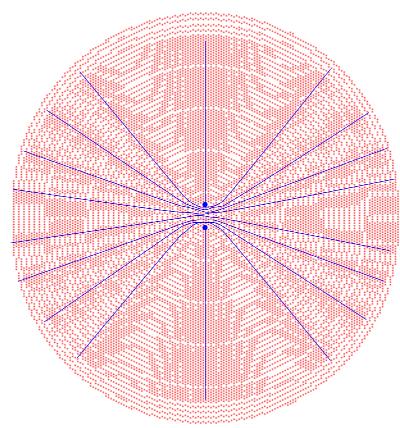
#2



We have already been introduced to the moiré fringe analog of the ripple tank, and the two-point interference patterns that it produces. Now we will look at the same phenomena with a finer grating scale, in order to reduce the visibility of the circular rings and emphasize the fringe patterns that they create.

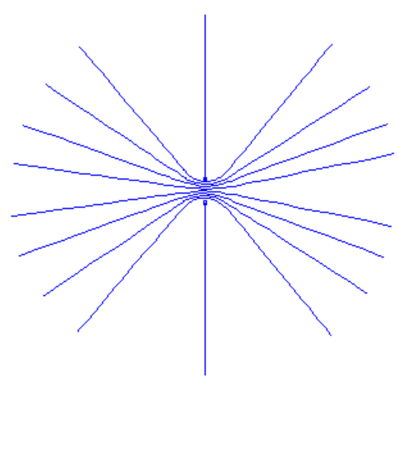
As the vertical separation of the two sources increases, the number of fringes around the perimeter of the circle increases. In the first sketch, the sources are 1.5 wavelengths apart, so that there are two dark fringes between the 12 o’clock and 3 o’clock positions (centered at 12:00 and 2:20). In the second sketch, the sources are 4.5 fringes apart, so there are five dark fringes in the same angular region (at 12, 1:15, 1:45, 2:15, and 2:45). As the sources separate further, more fringes emerge, and the angular spacing between them decreases. This kind of experimenting is best done with samples of such patterns right in your hands!

The fringe patterns are a little indistinct, especially as pixellated here, but we can draw center lines through them with the aid of a little mathematical insight.



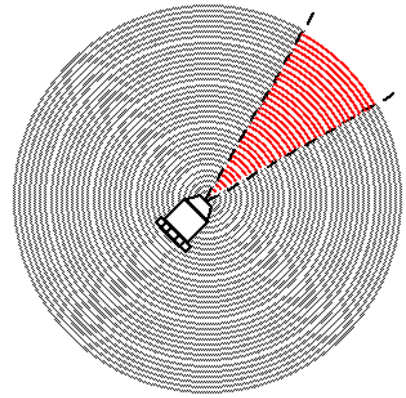
Near the edges of the circles, the fringes seem to be straight lines, aimed between the two sources. In fact, they are mathematical hyperbolas, and arc around between the sources to emerge on the other side. The fringes are the loci of points of equal path difference between the two source points (the foci of the hyperbolas). If the sources really were point sources in 3-dimensional space, these fringes would be hyperbolas of revolution nested one within the other. Between the two sources, the fringes are equally spaced at half-wavelength intervals. As they move outward, they approach the straight-line asymptotes typical of hyperbolas.

hy•per•bo•la (hí pûr'bé lé) n. pl. -las: 1. the set of points in a plane whose distances to two fixed points in the plane have a constant difference; a curve consisting of two branches, formed by the intersection of a plane with a right circular cone when the plane makes a greater angle with the base than does the generator of the cone. Equation: $x^2/a^2 - y^2/b^2 = 1.0$.

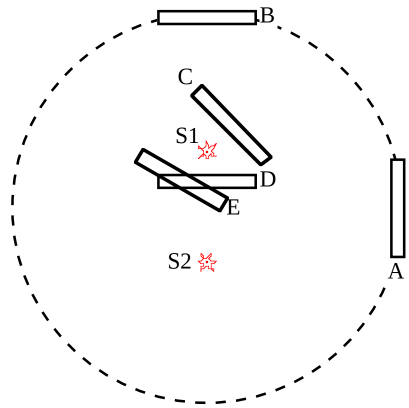


Spherical waves

As a rule, we will be dealing with sources that radiate light in only a fairly limited angle, perhaps 30° for light spread by a microscope lens. Thus we are interested at any one time in only a small region of the patterns we have been describing so far. Even so, we can use the overall pattern as a kind of "road map" of the various domains of holography, in which we will consider just one area at a time, as shown on the next sketch.



Here, the various types of holography are mapped out as domains with respect to the locations of the two sources used. "A" signifies "diffraction gratings," which we will study first. Then comes "B," the "holographic lenses," or "in-line" or "Gabor" holograms (where S1 becomes the prototype for the "object" and S2 for the "reference source"). Combined, their mathematics allows us to discuss "off-axis transmission" or "Leith-Upatnieks holograms" at location "C," which will extend to include "image plane" and "rainbow" holograms. Then we will move to reflection holograms, first the "single beam" or "Denisyuk" type, at location "D," and then the "off-axis" reflection hologram at location "E."



side-by-side: linear fringes

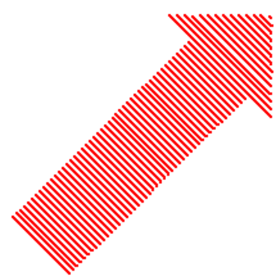
When we are at location "A," the interference fringes are straight lines radiating from a point midway between the sources, and they intersect the recording plane at equally-spaced points, which become lines if we consider them in three-dimensional space (contentions that we will prove later on).

in-line: Fresnel zone plate

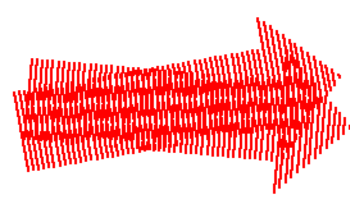
At location "B," the interference fringes are also straight lines radiating from a point midway between the sources, but they intersect the recording plane in circles, not lines, and the circles are not equally spaced—they become closer and closer as we move away from the line that passes through both sources.

Plane waves (in *x*-*z* plane)

If we are far from a point source of radiation, and considering the waves only over a limited region, the waves can be approximated as flat or "plane" wavefronts. In this region, we say that the light is "collimated" or that the "rays" are all parallel. This is a common case for star light, for example, but in the laboratory it is often quite difficult to produce exact plane waves. We usually mean that waves are "plane" if their departure from exact planarity is small compared to a wavelength of light over the aperture we are interested in (a quarter of a wavelength tolerance is typical). We often sketch a portion of such a wave as a large arrow, pointed perpendicular to the wavefronts (which is the direction of propagation of the plane wave in most media), with the wavefronts more or less visible within it, and loosely refer to this as a light "ray" (a "ray bundle" might be more accurate).

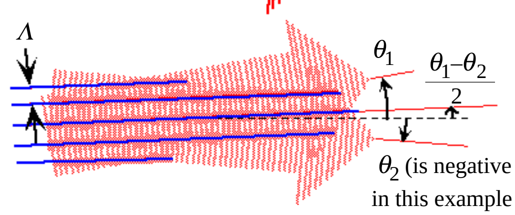


When two plane waves cross, the interference pattern between them takes on a fairly simple characteristic shape. The fringes are now strictly straight lines (the graphics here may make them wobble a bit) that are parallel and equally spaced. Their spacing decreases as the angle between the rays increases, and the line of the fringes bisects the angle between the two rays. These effects are really best explored by working with moiré patterns between pieces cut from parallel-line patterns on acetate (it helps the contrast if the ratio of dark/clear areas is around 1:1).



To get a little more quantitative about it, this is probably the time to state that the angle of the fringes is the average of the two ray angles,

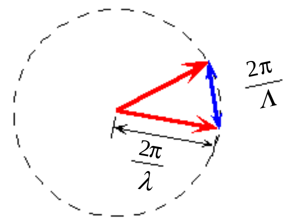
$$\theta_{\text{fringe}} = \frac{\theta_1 + \theta_2}{2}, \qquad (15)$$



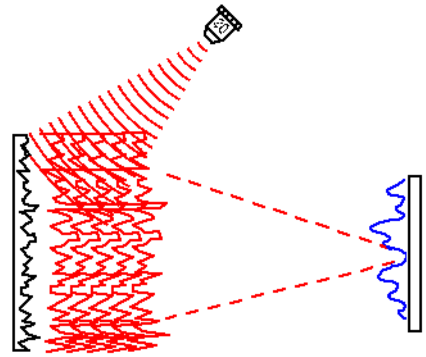
and the spacing between the fringes, which we will call Λ, is determined by the angle between the rays and the wavelength of the light:

$$\frac{2\pi}{\Lambda} = \frac{2\pi}{\lambda} 2\sin\left(\frac{|\theta_1 - \theta_2|}{2}\right). \qquad (16)$$

(The last page of the chapter shows how this distance, Λ, is related to the grating spacing, d, that we will see later)



It is sometimes easier to remember these in geometrical terms, with a vector representing the fringe pattern that is the difference between the vectors representing the two rays. These vectors all have lengths that are proportional to the reciprocal of the scale of the pattern they represent (here the wavelengths, λ and Λ), and a direction that is perpendicular to their wavefronts or fringes, and are generally known as **K**-vectors when the 2π is included. They are our introduction to *reciprocal space*!



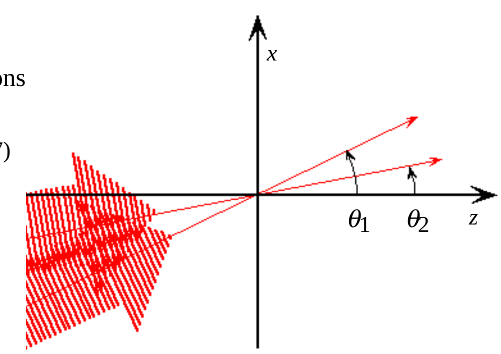
Laser speckle

We have talked about many kinds of interference patterns here, but not about the one kind that we see most often with laser light, and which is perhaps the most difficult to explain; the one we call "laser speckle." It is the gritty or sandy appearance of laser beams when played upon a diffusing or matte surface (like paper or paint). The microscopic roughness of the surface, which is what causes it to scatter light in all directions, creates many, many overlapping waves with randomized phases. When these waves cross again, such as when focused by the lens of your eye, they produce a randomized intensity pattern with high contrast. Try looking at a speckle pattern through a pinhole (made by pinching your fingertips together) and seeing how their size changes; watch how they move as you move your head from side to side (repeat without your glasses, if you usually wear them). A rigorous discussion of laser speckle requires the mathematics of random process theory, but your TA can probably convince you of the reasonableness of some simple rules. I can only warn you that Prof. Gabor once referred to laser speckle as "holographic enemy number ONE!" So, you had better figure on understanding it one of these days, if only for your own protection. Fortunately, experienced holographers eventually no longer see speckle as intensely as novice holographers seem to.

Simple interference patterns:

With this background, we can now consider a few interference patterns produced by simple optical setups, using the expression of Eq. 8 in slightly different form to emphasize the usefulness of the "phase footprints" found in Ch. 3, so that the phases, and resulting total intensity, are expressed as functions of x and y in the observation plane, usually at z=0.

$$I_{\text{total}}(x,y) = I_1(x,y) + I_2(x,y) + 2\sqrt{I_1(x,y)\cdot I_2(x,y)}\,\cos\big(\phi_1(x,y) - \phi_2(x,y)\big)\,. \qquad (17)$$



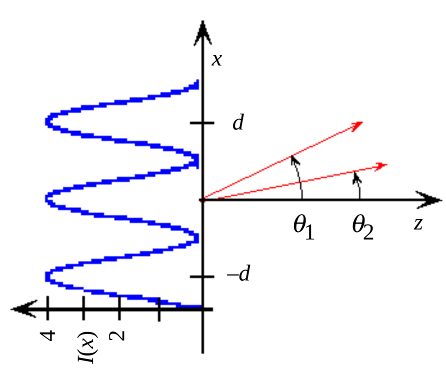
Overlapping plane waves

Consider two plane waves incident at angles θ_1 and θ_2, as shown in the sketch. Each has unit intensity, so I_1=I_2=1.0, and their phase footprints are:

$$\begin{aligned} \phi_1(x,y) &= \frac{2\pi}{\lambda} x \sin\theta_1\,, \\ \phi_2(x,y) &= \frac{2\pi}{\lambda} x \sin\theta_2\,. \end{aligned} \qquad (18)$$

Simply plugging this information into the expression above (again assuming that the intensity of each source at the hologram plane is unity) yields

$$\begin{aligned} I_{\text{total}}(x,y) &= 1 + 1 + 2\sqrt{1\cdot 1}\,\cos\left(\frac{2\pi}{\lambda} x\sin\theta_1 - \frac{2\pi}{\lambda} x\sin\theta_2\right) \\ &= 2 + 2\cos\left(\frac{2\pi}{\lambda} x\big(\sin\theta_1 - \sin\theta_2\big)\right), \end{aligned} \qquad (19)$$

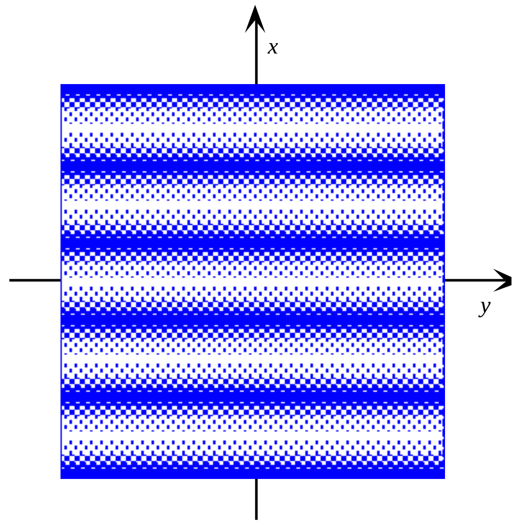


which is a sinusoidal variation of intensity as x increases, reaching a new peak at multiples of the distance d, given by

$$d = \frac{\lambda}{\sin\theta_1 - \sin\theta_2}\ , \tag{20}$$

so that the spatial frequency of the pattern, f, is given by

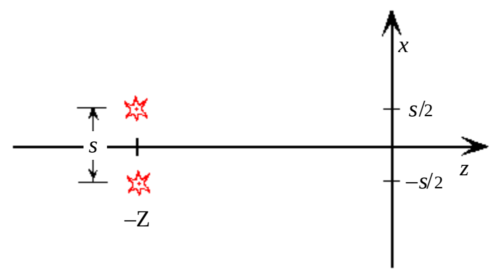
$$f = \frac{\sin\theta_1 - \sin\theta_2}{\lambda}\ . \tag{21}$$



A comment about *spatial frequency*:
Researchers in coherent optics often refer to patterns in terms of their "spatial frequency" (usually measured in cycles per millimeter), reflecting the grounding of the field in communication theory. As a two-dimensional (and occasionally three-dimensional) extension of temporal frequency concepts (cycles per second, referred to as Hertz or Hz), spatial frequency thinking makes the extension of signal analysis concepts fairly straightforward. Depending how we assign numbers to the beams, the results for d and f could well come out negative. By convention, we will always consider the spacing and spatial frequency to be **positive** numbers (negative frequencies are more common in linear systems theory), so there really should be "magnitude bars" around the right sides of Eqns. 20 and 21.

Note that there is no variation of either wave's phase in the y-direction, and thus no variation of I_{total} with y. The intensity pattern in the x,y plane will be a series of parallel bands of graded intensity.

<u>Side-by-side point sources</u>
Consider now the case where two coherent point sources of light, S_1 and S_2, are at the same distance from the hologram plane, at z=–Z, and at equal distances from the z-axis, at x_1=+s/2 and x_2=–s/2. The intensity of each source at the hologram plane is unity, and their phase footprints are



$$\begin{aligned} \phi_1(x,y) &= \frac{2\pi}{\lambda}\mathrm{Z} + \frac{\pi}{\lambda \mathrm{Z}}\left(\left(x-\frac{s}{2}\right)^2 + y^2\right), \\ \phi_2(x,y) &= \frac{2\pi}{\lambda}\mathrm{Z} + \frac{\pi}{\lambda \mathrm{Z}}\left(\left(x+\frac{s}{2}\right)^2 + y^2\right). \end{aligned} \tag{22}$$

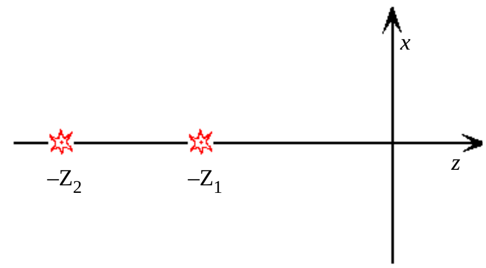
Plugging these into the master interference equation then gives

$$\begin{aligned} I_{\text{total}}(x,y) &= 1+1+2\sqrt{1\cdot 1}\cos\left(\frac{\pi}{\lambda \mathrm{Z}}\left(\left(x-\frac{s}{2}\right)^2 + y^2\right) - \frac{\pi}{\lambda \mathrm{Z}}\left(\left(x+\frac{s}{2}\right)^2 + y^2\right)\right) \\ &= 2 + 2\cos\left(\frac{2\pi}{\lambda}x\frac{\mathrm{s}}{\mathrm{Z}}\right). \end{aligned} \tag{23}$$

This pattern has the same form as that shown above, and if we can arrange it so that (s/Z) is equal to the difference of the sines of the angles, the spatial frequency of the pattern will even be the same. Which is to say that the phase contributions due to the sphericity of the waves "cancel out" if the interfering sources have the <u>same</u> sphericity; e.g., they are at the same distance. We caution that this is true only for small s/Z and for fringes near (x,y)=(0,0), which is often the case. The general principle that waves need not be exactly planar to make the plane wave approximation useful still stands, though.

In-line point sources

Here, the point sources are arranged one in front of the other, at $-Z_1$, and the second at $-Z_2$. The phase footprints are now



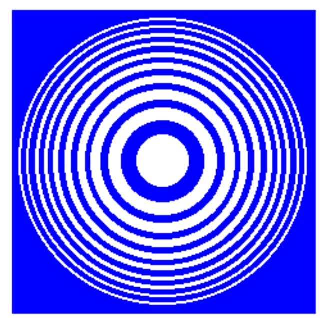
$$\phi_1(x,y) = \frac{2\pi}{\lambda} Z_1 + \frac{\pi}{\lambda Z_1}\left(x^2+y^2\right) \ ,$$
$$\phi_2(x,y) = \frac{2\pi}{\lambda} Z_2 + \frac{\pi}{\lambda Z_2}\left(x^2+y^2\right) \ . \qquad (24)$$

The leading terms in both are constant phases, and we will assume for the moment that they are both exact multiples of 2π, equivalent to zero, and can safely be ignored. Plugging the rest of the terms into the master interference equation (again assuming that the intensity of both waves at the hologram plane is unity) then gives a characteristic intensity pattern:

$$I_{\text{total}}(x,y) = 1+1+2\sqrt{1\cdot 1}\cos\left(\frac{\pi}{\lambda Z_1}\left(x^2+y^2\right) - \frac{\pi}{\lambda Z_2}\left(x^2+y^2\right)\right)$$
$$= 2+2\cos\left(\frac{\pi}{\lambda}\left(\frac{1}{Z_1}-\frac{1}{Z_2}\right)\left(x^2+y^2\right)\right) \ . \qquad (25)$$

Now we are dealing with something quite different! This pattern is a function of both *x* and *y*, and in a combination that makes it a function only of the distance, *r*, from the (*x*,*y*)=(0,0) point. That is, the pattern has rotational symmetry about the (0,0) point, and thus consists of some kind of pattern of concentric circles, however spaced. In fact, the spacing is also an important matter, so we will examine it in some detail.

We will consider first a general function of radius, *r*, described by



$$I(r) = 1+\cos\left(2\pi \frac{r^2}{a^2}\right) \ . \qquad (26)$$

This has a maximum at the origin, r=0, and another maximum (or bright ring) at $r=a$. The third maximum is at $r=\sqrt{2}a$, and in general, the *n*-th maximum is at a radius $r=\sqrt{n}a$. Which is to say that the bright rings are not equally spaced, but the spacing slowly shrinks as we move outward; in fact the area between successive bright rings is a constant! A pattern of this sort was first devised by the French mathematician Augustin Jean Fresnel (1788-1827), and generally bears the name “Fresnel zone plate” in his honor. Actually, Fresnel’s zone plate is a binarized “on-or-off” version of this pattern, and we holographers tend to call this continuous-scale version a “Gabor zone plate.” In the case of our interferometric exposure, the scale factor becomes

$$a = \sqrt{2\lambda \frac{z_1 z_2}{Z_2 - Z_1}} \ . \qquad (27)$$

When this pattern exposes a piece of film, the resulting transmittance pattern is found to have some interesting focusing patterns that we will soon explore in some detail!

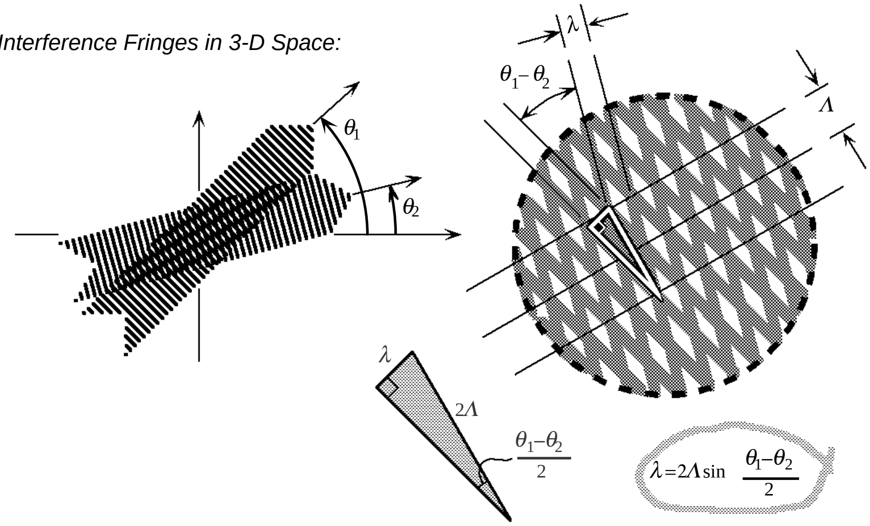
Conclusions:

The notion of “interference” defies some of our intuitive notions of “conservation of energy” on a small scale, but once it becomes a natural way of “seeing” things, it explains many interesting wave-optical phenomena. There are many, many categories of interference phenomena, as any book on physical optics will reveal. Here, we will limit our attention to the interference of waves from two spatially-separated coherent sources, as this is the simplest model for understanding holography. Later we will generalize from point-like sources to large-area diffuse sources, but the underlying concepts will stay the same.

With the help of the “phase footprints” of some common wavefronts, we can become quite quantitative about the intensities of some interference patterns of interest. But it is the geometry of the patterns—the directions, spacings, and shapes of the resulting fringes—that is of most interest to us for most of this course. That information follows directly from simply subtracting the "phase footprints," something we can do mathematically, or by looking at moiré fringes!

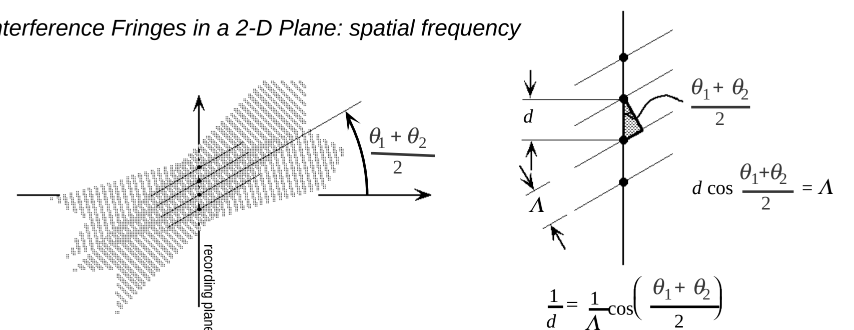
References:
1. J.D. Gaskill, *Linear Systems, Fourier Transforms, and Optics,* John Wiley & Sons, New York, 1978 (ISBN 0-471-29288-5), "Ch. 10: The Propagation and Diffraction of Optical Wave Fields."
Contrast this with the classic J.W. Goodman text, *Introduction to Fourier Optics,* McGraw-Hill Book Co., New York, 2nd Edn. 1996, "Ch. 3: Foundations of Scalar Diffraction Theory," which uses the opposite sign convention for spatial phase. It is sometimes said that the main differences between electrical engineers and physicists can be explained by " j = –i ," their respective symbols for the square root of minus one having opposite signs!

Interference Fringes in 3-D Space:



$$\lambda = 2\Lambda \sin \frac{\theta_1 - \theta_2}{2}$$

Interference Fringes in a 2-D Plane: spatial frequency



$$d \cos \frac{\theta_1 + \theta_2}{2} = \Lambda$$

$$\frac{1}{d} = \frac{1}{\Lambda} \cos\left(\frac{\theta_1 + \theta_2}{2}\right)$$

$$= \frac{2}{\lambda} \sin\left(\frac{\theta_1 - \theta_2}{2}\right) \cdot \cos\left(\frac{\theta_1 + \theta_2}{2}\right)$$

$$= \frac{1}{\lambda} \left(\sin\theta_1 - \sin\theta_2\right)$$

"spatial frequency" = $|f|$ $\qquad f = \dfrac{\sin\theta_1 - \sin\theta_2}{\lambda}$