Chapter 9: Holographic Lenses and In-Line "Gabor" Holograms

Introduction:

In this chapter we will re-examine the interference pattern formed by two in-line point sources from a different point of view, and re-examine diffraction by that pattern in wavefront terms, confirming the behavior indicated by the ray-tracing results. The pattern is found to have many of the properties of conventional glass refracting lenses, so we will take a minute to review those properties for a moment too. Then, we will use these elements as a way of describing the operation of Gabor's original in-line type of transmission hologram, along with all its shortcomings.

Transition to wavefront curvature:

Up to now, we have usually been referring to the phase-footprint of a source as being a function of the "location" of the source, the distance and angle of the source from the observation plane. Beginning now, we will emphasize a subtle change and usually refer instead to the same phase footprint as being a function of the <u>curvature</u> and <u>inclination</u> of the wavefront as seen at the observation plane. Instead of referring to the source location, which gave a sense of physical concreteness to the discussions, we will refer to the properties of the wavefront itself, which is all we can measure at the observation plane, after all. This "disembodiment" of the waves, as it were, will help us to avoid becoming dependent on a particular coordinate system choice, and will also make it easier to discuss wavefronts that might not correspond to single point sources or images. However, in most cases the correspondence is close enough that we will be able to oscillate between the phase footprint, the wavefront curvature and inclination, and the source/image location without any effort. Indeed, we have been doing so all along without commenting on it!

<u>definition of "inclination" and "radius of curvature"</u>:

If we take a snapshot of a wavefront just as it comes to the observation surface, we can (in principle) make the physical measurements needed to characterize it with a ruler and a protractor. First, we construct a plane tangent to the wavefront at the (0,0) location, the center of our local system of coordinates (typically the plane of the hologram). The perpendicular to that tangent plane defines the inclination, and presumably lies in the x-z plane, so only one angle is sufficient to define it, the angle θ that we have referred to. We might need to resort to direction cosines if the perpendicular sticks out of the plane, so that $\sin \theta$ becomes ℓ, and so forth.

Unless it is a plane wave, the wavefront surface will separate from the tangent plane as we move in either the x or y direction, and the rate of separation increases as we move further away. Which is to say that the wavefront arrives at x later than it would if it were a plane wave, because of its curvature toward the left. In general, we describe the separation by a distance $\delta(x,y)$ where

$$\delta(x,y) = \frac{1}{2}\frac{1}{R_x}\left(x\cos\theta_0\right)^2 + \frac{1}{2}\frac{1}{R_y}y^2 \ . \qquad (1)$$

Note that $(x\cos\theta_0)$ is the distance along the tangent plane for an inclined wave.

In some cases, the wavefront may have different curvatures in the x- and y-directions (that is, have a cylindrical or astigmatic component), depending on how it was generated. If the curvatures are equal, which means the wave came from a point source or is focused from/to a well-defined point virtual/real image, we will not use an x- or y-subscript, and save subscripts to denote the identity of the wave (object, reference, illumination, m-th order output, etc.).

<u>positive & negative = diverging & converging</u>

Our definition of δ includes a sign; it is positive when the wave is bulging outward, or is *convex*, as it travels forward. That is, when the wave is *diverging*. When the wave is curved inwards, or is *concave*, it is said to be *converging*. This happens when a wave is being focused toward a point real image in space a distance R from the hologram. In this case, the wave at x arrives sooner than for a plane wave, and the δ is said to be negative. We describe the wave mathematically as having a <u>negative</u> radius of curvature, and again perhaps with different radii in the x- and y-directions.

This picture is easy to understand if the wave is traveling from left to right, as light usually has been doing in our optical diagrams. However, we will soon have to deal with situations where we must let the light travel from right to left instead. A snapshot of a wave converging from right to left looks just like a wave diverging from left to right (we have included dashed lines here as a hint of the past, but these will not always be available), but the δ of the two cases
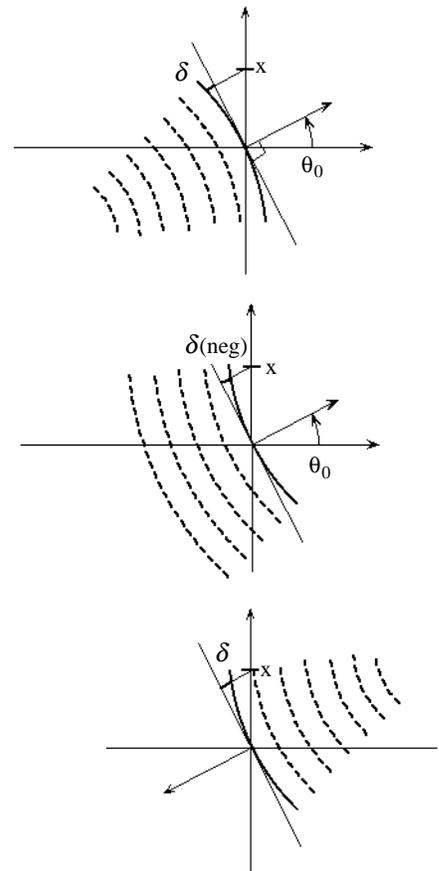
have opposite signs, regardless. What we are really interested in is the phase footprint, and the phase is proportional to the δ with its correct sign. That is, the margin of a diverging spherical wave always arrives later than the equally-inclined plane wave, no matter what direction it is headed in! Which is to say that <u>diverging</u> waves will always have <u>positive</u> radii of curvature, regardless of what direction they are traveling in. Likewise, <u>converging</u> waves will always have <u>negative</u> radii of curvature.

Phase footprints, again:

From this geometrical discussion, we hope to re-derive the "phase footprints" of our rogue's gallery of simple wavefronts. The equation above is really the most general case we will need, so it suffices to work backward from that! The phase delay follows directly from $\phi(x,y)=(2\pi/\lambda)\delta(x,y)$.

<u>inclined plane wave</u>
The increase in distance from an inclined plane is a linear function of distance up the x-axis, and we get the familiar phase footprint

$$\phi_{\substack{\text{plane}\\\text{wave}}}(x,y)=\phi_0+\frac{2\pi}{\lambda}\sin\theta_0\,x\,.\qquad(2)$$

<u>general case</u>
To find the phase footprint of an inclined spherical wave, it is only necessary to add the extra phase term due to the extra distance, the δ found above, giving

$$\phi_{\substack{\text{off-axis}\\\text{spherical}\\\text{wave}}}(x,y)=\phi_0+\frac{2\pi}{\lambda}\sin\theta_0\,x+\frac{\pi}{\lambda}\left(\frac{\cos^2\theta_0}{R_x}x^2+\frac{1}{R_y}y^2\right).\qquad(3)$$

further terms
A more complete expansion would include higher-order terms such as those below, which we will ignore except when they are needed for discussions of aberrations (especially spherical aberration and coma) as special topics:

$$\begin{aligned}\phi_{\text{higher}}(x,y)=&-\frac{\pi}{\lambda}\frac{\cos^2\theta_0\,\sin\theta_0}{R_x^2}x^3\\&-\frac{\pi}{\lambda}\frac{\sin\theta_0}{R_xR_y}x\,y^2\\&+\frac{\pi}{8\lambda}\frac{\cos^2\theta_0\left(3-5\cos 2\theta_0\right)}{R_x^3}x^4\\&+\frac{\pi}{4\lambda}\frac{1-3\cos^2\theta_0}{R_x^2R_y}x^2\,y^2\qquad(4)\\&-\frac{\pi}{4\lambda}\frac{1}{R_y^3}y^4\\&+O[x^5,y^6]\,.\end{aligned}$$

<u>on-axis spherical wave</u>
Reducing θ_0 to zero brings us back to where we started, to the phase footprint of an on-axis spherical wave from a source a distance R from the hologram plane, except that now we are also prepared to deal with non-spherical or *astigmatic* waves that have different radii of curvature in perpendicular directions.

$$\phi_{\substack{\text{on-axis}\\\text{wave}\\\text{spherical}}}(x,y)=\phi_0+\frac{\pi}{\lambda}\left(\frac{1}{R_x}x^2+\frac{1}{R_y}y^2\right).\qquad(5)$$

In-line interference, again:

Revisiting our familiar in-line interference case again in the new terminology, we must begin with the same old interference equation, where now we assume that both of the spherical (stigmatic) waves have an intensity of 0.25 at the hologram plane, but different radii of curvature (due to the difference in distance of their sources, which we are not supposed to know directly, but of course $R_i = -z_i$):

$$I_{\text{total}} = I_A + I_B + 2\sqrt{I_A \cdot I_B}\,\cos\left(\phi_A - \phi_B\right) \ . \tag{6}$$

The phase footprints are, where we now introduce λ_1 to represent the recording wavelength, given by

$$\phi_A(x,y) = \phi_A + \frac{\pi}{\lambda_1 R_A}\left(x^2 + y^2\right),$$

$$\phi_B(x,y) = \phi_B + \frac{\pi}{\lambda_1 R_B}\left(x^2 + y^2\right), \tag{7}$$

so that the exposing intensity pattern becomes the familiar Gabor zone plate:

$$I_{\text{total}}(x,y) = 0.5 + 0.5\cos\left(\frac{\pi}{\lambda_1}\left(\frac{1}{R_A} - \frac{1}{R_B}\right)\left(x^2 + y^2\right)\right). \tag{8}$$

Now we expose and process a holographic material to this pattern to produce a transmittance pattern that is its exact replica. That is, the transmittance pattern has the form

$$t_{\text{amp}}(x,y) = 0.5 + 0.5\cos\left(\frac{\pi}{\lambda_1}\left(\frac{1}{R_A} - \frac{1}{R_B}\right)\left(x^2 + y^2\right)\right). \tag{9}$$
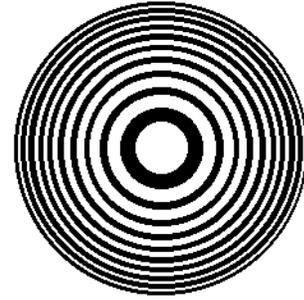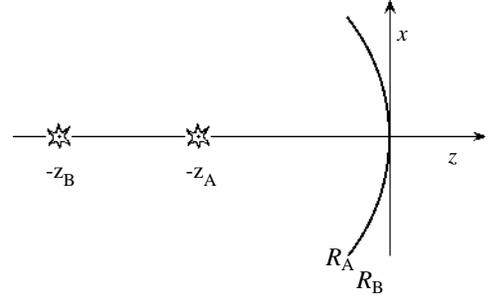
Transmittance proof of the focus equation:

Next, we illuminate this transmittance pattern with yet a third spherical wavefront, the illumination wave, with a curvature R_c. described by

$$E_{\text{ill}}(x,y,t) = \sin\left(2\pi\nu_2 t - \frac{\pi}{\lambda_2 R_c}\left(x^2 + y^2\right)\right). \tag{10}$$

The output wave is then given by

$$E_{\text{out}}(x,y,t) = t_{\text{ampl}}(x,y) \cdot E_{\text{ill}}(x,y,t)$$

$$= \frac{1}{2}\left(1 + \cos\left(\frac{\pi}{\lambda_1}\left(\frac{1}{R_A} - \frac{1}{R_B}\right)\left(x^2 + y^2\right)\right)\right) \cdot \sin\left(2\pi\nu_2 t - \frac{\pi}{\lambda_2 R_c}\left(x^2 + y^2\right)\right)$$

$$= \frac{1}{2}\sin\left(2\pi\nu_2 t - \frac{\pi}{\lambda_2 R_c}\left(x^2 + y^2\right)\right)$$

$$+ \frac{1}{2}\sin\left(2\pi\nu_2 t - \frac{\pi}{\lambda_2}\left(\frac{\lambda_2}{\lambda_1}\left(\frac{1}{R_A} - \frac{1}{R_B}\right) + \frac{1}{R_c}\right)\left(x^2 + y^2\right)\right)$$

$$+ \frac{1}{2}\sin\left(2\pi\nu_2 t - \frac{\pi}{\lambda_2}\left(-\frac{\lambda_2}{\lambda_1}\left(\frac{1}{R_A} - \frac{1}{R_B}\right) + \frac{1}{R_c}\right)\left(x^2 + y^2\right)\right). \tag{11}$$

These terms represent three spherical waves with three different curvatures, which we will designate as the $m=0$, $+1$, and -1 diffraction orders. The locations of the corresponding foci follow from $R_i = -z_i$. Things get a little more complicated if any of the waves are off-axis, but the principles are the same.

The three output wave curvatures can be represented by a single formula,

$$\frac{1}{R_m} = m\frac{\lambda_2}{\lambda_1}\left(\frac{1}{R_A} - \frac{1}{R_B}\right) + \frac{1}{R_C} \ , \tag{12}$$

which is the same as the first term of the paraxial expansion of Chapter 7, and also follows the *focus law* for refractive lenses. Thus we see (yet again) that a simple holographic lens may be represented as paired positive and negative lenses, with equal and opposite focal lengths, and perhaps higher-order pairs with one-half and one-third the focal length, etc. We will usually call it the "one-over-R" equation in our discussions. Thus our expectations from local ray-tracing analyses are confirmed by a global transmittance or "linear systems theoretic" analysis.

$$\frac{1}{R_m} = m \frac{\lambda_2}{\lambda_1}\left(\frac{1}{R_A} - \frac{1}{R_B}\right) + \frac{1}{R_C}$$

off-axis illumination

If the illumination of Eq. 10 includes an off-axis term,

$$E_{\text{ill}}(x,y,t) = \sin\left(2\pi\nu_2 t - \frac{2\pi}{\lambda_2}\sin\theta_c\ x - \frac{\pi}{\lambda_2 R_c}\left(\cos^2\theta_c\ x^2 + y^2\right)\right),\quad (13)$$

then the transmitted wavefronts will have the same linear phase term imposed upon them, and the radius of curvature of the wavefront (at least in the y-direction, as we shall see later) stays the same while it gains an overall tip in the x-direction given by

$$\sin\theta_m = m\frac{\lambda_2}{\lambda_1}\left(\sin\theta_A - \sin\theta_B\right) + \sin\theta_C,\qquad (14)$$

where we have also allowed for a tip in the object and/or reference beams. We will usually call this the "sine-theta" equation in discussions to follow. Now the focus locations follow from

$$\sin\theta_m = m\frac{\lambda_2}{\lambda_1}\left(\sin\theta_A - \sin\theta_B\right) + \sin\theta_C$$

$$\begin{aligned}x_m &= -R_m \sin\theta_m, \\ z_m &= -R_m \cos\theta_m.\end{aligned}\qquad (15)$$

With these generalized properties of the Gabor zone plate in hand, we can go ahead to apply them to help us understand the imaging properties of an early and simple type of hologram, the "in-line" or "Gabor" hologram.

In-line (Gabor) holograms:

Now the object consists of a multiplicity of point sources arrayed near, but not entirely on, the z-axis so as to represent a three-dimensional object that reflects light that is coherent with the distant and on-axis reference point source. There are, as you know by now, some practical problems in getting all the beams to the hologram without shadowing, but let's accept this simple picture for now. Each of the image points produces its own interference pattern that adds in intensity with those from the other points, plus some cross terms, to produce a superposition of Gabor zone plates in the hologram. These, in turn, produce a multiplicity of spherical waves, each in several orders, that combine to produce arrays of point images that replicate the three-dimensionality of the object.
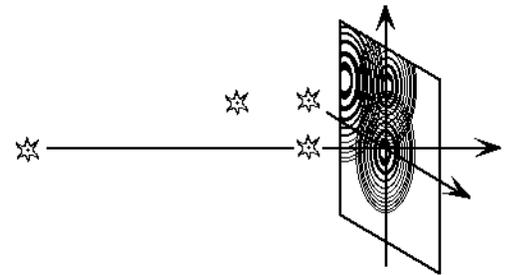
multiple image points

To defend this simple superposition principle, we need to look at the mathematics of interference again briefly. Let the reference and object waves be represented by completely general phase footprints (which means that this concept will apply to off-axis holograms too), where the reference beam intensity is K times that of the combined object beams. Just to set the stage, we will assume that there are N uniform object beams, all of equal (say, one over N, so they total to unity) intensity, and a uniform reference beam of intensity K (which is the beam ratio). Then we can think of the total wave as given by

$$E_{\text{total}}(x,y,0,t) = \left(\sqrt{\frac{2}{\varepsilon_0 c}}\right)\sqrt{K}\cos\left(2\pi\nu t - \phi_{\text{ref}}(x,y)\right) + \left(\sqrt{\frac{2}{\varepsilon_0 c}}\right)\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\cos\left(2\pi\nu t - \phi_i(x,y)\right)$$

$$= \left(\sqrt{\frac{2}{\varepsilon_0 c}}\right)\sqrt{K}\cos\left(2\pi\nu t - \phi_{\text{ref}}(x,y)\right) + \left(\sqrt{\frac{2}{\varepsilon_0 c}}\right)\frac{1}{\sqrt{N}}\cos\left(2\pi\nu t - \phi_1(x,y)\right) + \left(\sqrt{\frac{2}{\varepsilon_0 c}}\right)\frac{1}{\sqrt{N}}\cos\left(2\pi\nu t - \phi_2(x,y)\right) + \cdots$$

(16)

Now, the total intensity is given by the time-average of the square of this expression, which we can look at term by term as

$$I_{\text{total}}(x,y) = \left(\sqrt{K}\cos\left(2\pi vt - \phi_{\text{ref}}(x,y)\right) + \frac{1}{\sqrt{N}}\cos\left(2\pi vt - \phi_1(x,y)\right) + \frac{1}{\sqrt{N}}\cos\left(2\pi vt - \phi_2(x,y)\right) + \cdots \right)$$

$$\times \left(\sqrt{K}\cos\left(2\pi vt - \phi_{\text{ref}}(x,y)\right) + \frac{1}{\sqrt{N}}\cos\left(2\pi vt - \phi_1(x,y)\right) + \frac{1}{\sqrt{N}}\cos\left(2\pi vt - \phi_2(x,y)\right) + \cdots \right)$$

$$= K + \frac{1}{N} + \frac{1}{N} + \frac{1}{N} + \cdots$$

$$+ 2\sqrt{\frac{K}{N}}\cos\left(\phi_1(x,y) - \phi_{\text{ref}}(x,y)\right) + 2\sqrt{\frac{K}{N}}\cos\left(\phi_2(x,y) - \phi_{\text{ref}}(x,y)\right) + \cdots$$

$$+ 2\frac{1}{N}\cos\left(\phi_1(x,y) - \phi_2(x,y)\right) + 2\frac{1}{N}\cos\left(\phi_1(x,y) - \phi_3(x,y)\right) + \cdots$$

(17)

= intensity of the beams by themselves ($N+1$ terms)

+ holographic cross terms (object $\times$ reference, N terms)

+ object - object cross terms (object $\times$ object, $N(N-1)/2$ terms) .

The new class of interference terms is the third one, the "object-object" cross terms, which will have a total diffraction efficiency of roughly ($1/K$) times the sum of the second or "holographic" cross terms. The second class of terms is the superposition set (or image set), where the contribution of each from each of the points of the object appears in a simple sum with the others. This constitutes the demonstration of "superposition" of holographic waves, which allows us to decompose a dimensional object into a 3-D array of points, separated by the resolving power of the optical system, and to trace the optical fate of each point separately, as if it alone existed in the exposure and reconstruction steps. In fact, we will trace the fates of only a few key points to obtain a representation of the spatial imaging parameters of the system, such as the 3-D location and magnification of the images.

object self-interference terms:
Note that the third class, or object-object self-interference terms, depend only on the differences between the phases of the two object points concerned. If these points are roughly at the same distance from the hologram, then the resulting pattern will consist of parallel fringes of roughly constant spatial frequency – an example of interference from side-by-side points. The resulting transmittance pattern will have $N(N-1)/2$ of these terms, with those of highest spatial frequency corresponding to interference between points at the extreme opposite boundaries of the object. The resulting transmittance pattern and image reconstruction is often termed "object self-interference noise," or "object shape-dependent noise," or "intermodulation noise." We shall see that it introduces a noisome fourth component into the array of output wave components of the hologram.

multiple point images
We will begin by discussing a very simple image consisting of three points: A, B, and C. The central object point, A (located a $(x,z)=(0,z_{\text{obj}})$), will serve to establish the central locations of the resulting images, and the longitudinally displaced point, B (located at $z_B=z_{\text{obj}}+\Delta z$) will serve to establish the longitudinal magnification of the images, M_{long}. The laterally displaced point, C (located at $(x,z)=(\Delta x,z_{\text{obj}})$), will then serve to establish their lateral magnifications, M_{lateral} .

location of the virtual image
To find the $m=+1$ or "virtual" image of the central point, A, which in more general terms we should call the "true" image, we resort to the focusing or "$1/R$-equation," Eq.12, with the variables adapted for this occasion:

$$\frac{1}{R_m} = m\frac{\lambda_2}{\lambda_1}\left(\frac{1}{R_{\text{obj}}} - \frac{1}{R_{\text{ref}}} \right) + \frac{1}{R_{\text{ill}}} \; , \qquad (18)$$
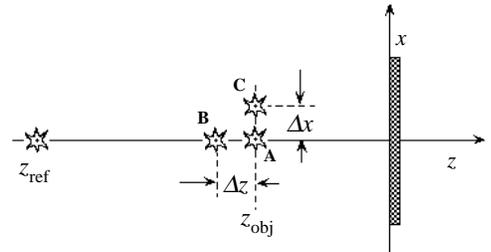
where the wavefront curvatures, the R_i, are related to the locations, the z_i, by $R_i = -z_i$. For the $m=+1$ case we then have:

$$\frac{1}{\underset{\text{virtual}}{R_{+1}}} = \frac{\lambda_2}{\lambda_1}\left(\frac{1}{R_{\text{obj}}} - \frac{1}{R_{\text{ref}}}\right) + \frac{1}{R_{\text{ill}}} \qquad (19)$$

Consider the specific case of an object point 300 mm from the plate, and the reference point 1000 mm from the plate. Recording is at 633 nm, and reconstruction is at 543 nm with a point source 1200 mm from the hologram. Cranking through the $1/R$ equation gives the $m=+1$ image location as $(x,z)=(0,-353)$ (a virtual image).

location of the <u>conjugate</u> image

The $m=-1$ image is usually referred to as the "conjugate" image (because it and the "true" image are paired), or the "real" image (because it is often focused "downstream" of the hologram plane). Again, it is only necessary to plug the familiar terms into the $1/R$ equation, remembering to get the sign of m correct and to interpret a negative radius of curvature properly (both are constant bugaboos in this class!). Under the same conditions as above, the conjugate image will be found at $(x,z)=(0,856)$ (a real image). Note that if the object is far enough away, relative to the reference source, the conjugate image will become a virtual image instead. If both the reference and illumination beams were collimated ($R_{\text{ref}} = R_{\text{ill}} = \infty$), then the real and conjugate images would be at equal distances from the hologram, but on opposite sides!

higher-order image locations:

No mysteries here: just plug in $m=2,-2,3,-3,4,-4$ and so forth into the $1/R$ equation, and figure out the corresponding image locations. The positive-m images will all lie between the true image and the hologram, and the negative-m images will be between the conjugate image and the hologram (if the conjugate image is a real image, that is). Under the same conditions as above, those higher-order images will be found at $x = 0$, $z = -206,315,-146,193,-113,139$, respectively.

If the reference and illumination sources are far enough away (that is, much farther than the object), we can approximate the $1/R$ equation (12, 18) as

$$R_m \approx \frac{R_{\text{obj}}}{m} + \frac{R_{\text{obj}}^2}{R_{\text{ref}}} - \frac{R_{\text{obj}}^2}{m\,R_{\text{ill}}} \qquad (20)$$
$$\approx \frac{R_{\text{obj}}}{m}.$$

That is, the higher-order images are at distances that are at integer fractions of the object distance. As the reference or illumination sources move toward the plate (compared to the object) that simple approximation starts to fall apart.

object self-interference noise image:

We said before that there would be components of the hologram transmittance pattern that would be diffraction gratings with a maximum spatial frequency determined by interference between points at the extremes of the object boundaries. For smaller spatial frequencies, there will be many possible pairs contributing grating components, and the density of those gratings becomes greater as the resulting spatial frequency approaches zero. Because these gratings are very nearly parallel-fringed with constant spatial frequency, the images they produce will appear at roughly the same plane as the illumination source. Because there are so many gratings, the images appear as a diffuse "halo" around the illumination source. For small object and illumination, we can say that the angle between the illumination source and the edge of the halo, as seen from the hologram, is the same as the angle between the extremes of the object in the same azimuth. The brightness of the halo is not very great, and tapers from a maximum near the illumination source to zero near the edge, often with a nearly-linear falloff. A careful mathematical analysis of the halo pattern
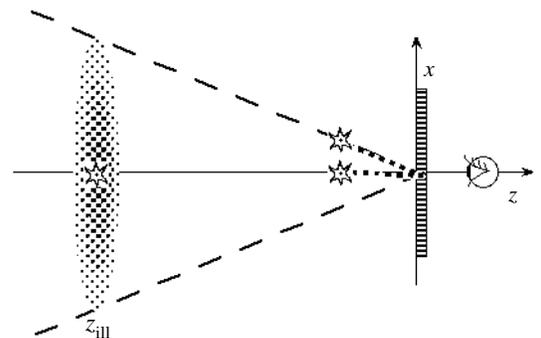
reveals that its intensity is proportional to the "autocorrelation function" of the object intensity pattern, and it can have some shape of its own if the object shape is suitably complex.

longitudinal magnification:
If the object point moves to $z+\Delta z$, away from the plate let's say, then the image point will move out to $z_{+1}+\Delta z'$, where $\Delta z'$ is given by

$$\Delta z' = \frac{dR_{out}}{dR_{obj}}\Delta z_{obj} \; , \tag{21}$$

and the longitudinal magnification follows from differentiation of the $1/R$ equation as

$$MAG_{long} = \frac{\Delta z'}{\Delta z_{obj}} = m\frac{\lambda_2}{\lambda_1}\left(\frac{R_m}{R_{obj}}\right)^2 . \tag{22}$$

Note that the longitudinal magnification will be negative for negative m. Which is to say that parts of the objects that are closer to the plate will be imaged closer to the hologram, no matter which side of the hologram they are focused to.

(the rest of this page is intentionally left blank)

pseudoscopic images

An implication of a negative longitudinal magnification is that the depth of the image will be reversed, as seen from the observer's downstream location. Such an image is said to be "pseudoscopic," and corresponds to what you see at the 3-D movies when you have the polarized glasses on upside-down (so that the right eye sees the left eye image, etc.; just putting the glasses on backwards doesn't work, by the way). Later, we will see that shaded surfaces give conflicting cues to the depth, but in simple Gabor holograms the inversion of depth is often easy to perceive, and to demonstrate with real images focused on cardboard or ground glass screens.
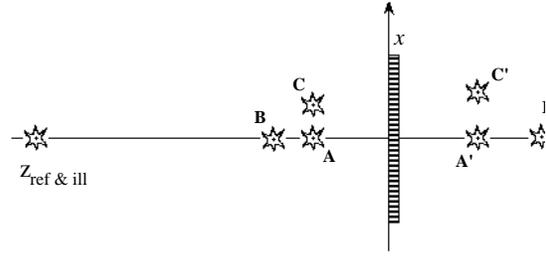
lateral magnification

The side-to-side or lateral magnification of a hologram image can be determined from fairly simple geometrical considerations for the in-line Gabor hologram, and from a slightly more generalized view that extends to off-axis holograms too. We will pursue both in this section.

"zero-frequency-point" geometrical analysis

Consider that point C is at the tip of an arrow, standing erect on the z-axis. The hologram of the tip of the arrow will be a Gabor zone plate centered on the ZFP or "zero-frequency point," which is defined by the intersection with the recording plate of a line drawn from the reference source to the tip, C, and extended to the plane of the plate.
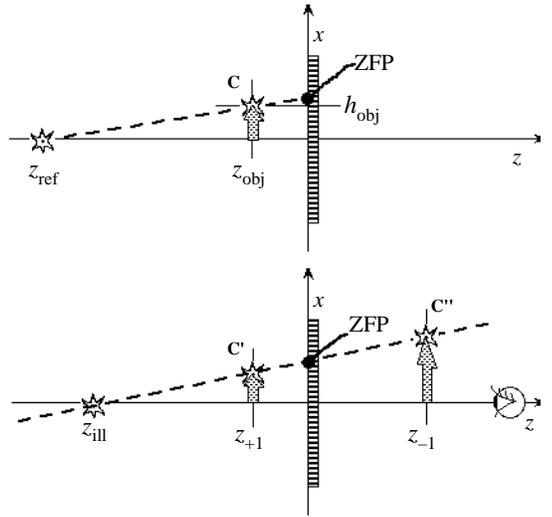
Upon illumination of that Gabor zone plate with a point source, all images of the arrow tip will be formed somewhere along the "central ray," an infinitely-long line that passes through both the illumination source point and the hologram ZFP point. For all values of m and for all wavelengths, λ_2, the tips of the images must lie on that central ray. If we know the z_m, we can easily find the height of the image (the height of the ray at that point), and hence the lateral magnification of the image. While this approach lends itself well to a graphical solution, it also yields useful analytical results.

First we find the height of the zero-frequency point, h_{ZFP}, from similar triangles:

$$\frac{h_{ZFP}}{-z_{ref}} = \frac{h_{obj}}{-z_{ref} + z_{obj}} . \qquad (23)$$

The height of the image, h_{image}, is similarly determined by

$$\frac{h_{image}}{z_{image} - z_{ill}} = \frac{h_{ZFP}}{-z_{ill}} . \qquad (24)$$

Combining terms gives the image height as

$$h_{image} = \frac{z_{image} - z_{ill}}{-z_{ill}} \frac{-z_{ref}}{-z_{ref} + z_{obj}} h_{obj} . \qquad (25)$$

The lateral magnification is defined as the ratio of the image height to the object height,

$$MAG_{lateral} \equiv \frac{h_{image}}{h_{obj}} = \frac{\dfrac{z_{image}}{z_{ill}} - 1}{\dfrac{z_{obj}}{z_{ref}} - 1} . \qquad (26)$$

And with a few cranks, invoking Eq. 18, we find that

$$MAG_{lateral} = m \frac{\lambda_2}{\lambda_1} \frac{R_{image}}{R_{obj}} . \qquad (27)$$

-p.8-

The lateral magnification will be positive, and the image will be erect, except when an odd-*m* image is virtual. Two special cases (which can be combined) are of interest: 1) if the illumination is collimated, the image size will be independent of the reconstructing wavelength. 2) if the reference beam is also collimated, the magnification will be unity, regardless of object distance.

Comparing the longitudinal and lateral magnifications, Eqs. 22 & 27, we find that

$$MAG_{long} = \frac{\lambda_1}{m\,\lambda_2}\,M_{lateral}^2 \;. \qquad (28)$$

In conventional optics, the longitudinal magnification is always the square of the lateral magnification; in holography, we have the opportunity to change the wavelength after recording, and to observe a higher-order image, if that is useful. There is also a possibility of scaling the hologram pattern up or down, which we will not analyze here. Of course, the images will usually have higher-order aberrations that will obscure their details (except for the *m*=+1 image when the illumination is an exact replica of the reference beam), but these simple rules will give the locations and sizes of the images with good accuracy.

angular subtend method:
A different way to think about lateral magnification is based on considering the angle subtended by the object as seen from the center of the hologram. Call this angle Ω_1 (capital "omega", shown here as negative), given by

$$\Omega_1 = \frac{h_{obj}}{-z_{obj}} \;. \qquad (29)$$

Interference between the top and bottom points of the object produces a grating with a spatial frequency, *f*, that is given by

$$f = \sin\Omega_1 / \lambda_1 \approx \Omega_1 / \lambda_1 \;. \qquad (30)$$

Upon illumination by an on-axis source, the output light will be diffracted by the spatial frequency, *f*, through an angle Ω_2, which might be different from Ω_1 if the order is different than +1, or if the wavelength changes:

$$\Omega_2 = \sin^{-1}\left(\lambda_2\,m\,f\right) \approx \lambda_2\,m\,f \approx m\frac{\lambda_2}{\lambda_1}\Omega_1 \;. \qquad (31)$$
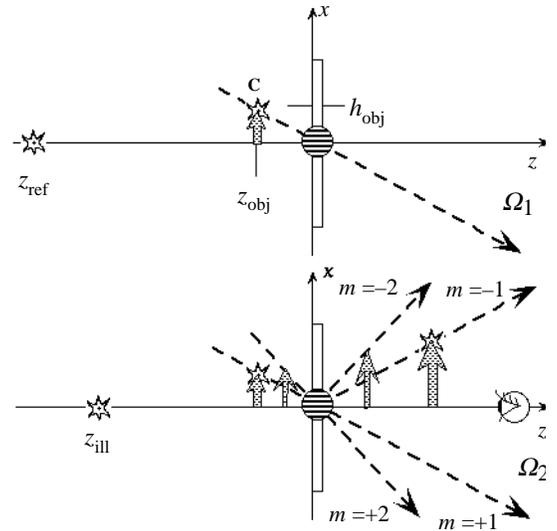
Wherever the output images are, their end points must lie on the rays defined by this equation. By plugging in the *z*-axis locations found before (or the radii of curvature of the output wavefronts), we can find the height of any of the output images. The lateral magnification is then given by

$$MAG_{lateral} \equiv \frac{h_{image}}{h_{obj}} = \frac{z_{image}\cdot\Omega_2}{z_{obj}\cdot\Omega_1}$$

$$= m\frac{\lambda_2}{\lambda_1}\frac{R_{image}}{R_{obj}} \;, \qquad (32)$$

which is what we wanted to show, as the two points of view must produce equivalent results.

off-axis holograms
So far, we have assumed that the reference and illumination beams are directly on axis. For off-axis holograms, we would look instead at the difference in spatial frequencies of the gratings corresponding to the object points separated by Ω_1. That difference of spatial frequencies would give rise to a difference of diffracted angles, which would again define rays along which the top and bottom of all images must lie. Invoking simple differential calculus (as infrequently as possible), and letting the Ω angles become vanishingly small, we describe the small difference of spatial frequency as δf, where

$$\delta f = \frac{1}{\lambda_1}\left[\left(\sin\left(\theta_{obj} + \Omega_1\right) - \sin\theta_{ref}\right) - \left(\sin\theta_{obj} - \sin\theta_{ref}\right)\right]$$
$$= \frac{\Omega_1}{\lambda_1}\cos\theta_{obj}.$$

(33)

The two gratings are superimposed during the exposure, and produce separate transmittance terms in the final hologram. Illumination by an off-axis illumination beam produces output beams that must satisfy a similar relationship,

$$\delta f = \frac{1}{\lambda_2}\left[\left(\sin\left(\theta_{out} + \Omega_2\right) - \sin\theta_{ill}\right) - \left(\sin\theta_{out} - \sin\theta_{ill}\right)\right]$$
$$= \frac{\Omega_2}{\lambda_2}\cos\theta_{out}.$$

(34)

The lateral magnification is then given by

$$MAG_{lateral} \equiv \frac{h_{image}}{h_{obj}} = \frac{z_{image}\cdot\Omega_2}{z_{obj}\cdot\Omega_1}$$
$$= m\frac{\lambda_2}{\lambda_1}\frac{R_{image}}{R_{obj}}\frac{\cos\theta_{obj}}{\cos\theta_{image}}.$$

(35)

Whenever the central object and output beam angles are equal (which is usually the case for $m=+1$) we find that the lateral magnification is again given by the in-line hologram result, shown in Eqs. 27 and 32.

As for the longitudinal magnification, there was nothing in the derivation of Eq. 22 that depended on the angles of any of the rays, so that result applies to off-axis as well as on-axis holograms. However, because the 1/R equation will later be seen to apply only to horizontally focused light, Eq. 22 predicts the longitudinal magnification only for vertical line features (about which we will see much more later on).

The two ways of thinking about image formation, via the "ZFP & central ray method" and via the "central angles, Ω_1 and Ω_2, method" produce identical results. It is a little early to know which you will find easier to remember and to use as a problem-solving tool—you will probably just plug in the formulas, after all! We will soon move into the domain of off-axis holograms, where the ZFP approach becomes only approximate but the central angle approach extends accurately. So, make your peace with both before we have to move along!

what's wrong with "in-liners"?
The in-line hologram broke new intellectual ground in 1948. Nobody believed that it was possible to reproduce the phases as well as the amplitude of a wavefront, but Gabor showed photographs that proved it was true. The Nobel Prize in Physics for 1971 finally made the significance of these ideas clear to everyone, once the laser had made some important improvements in holographic technique possible. We have had to move beyond Gabor's configuration because of visual problems with the images it produced. Here is a brief catalog, which you can probably add to:

1. glare of illumination
The fact that the desired image is directly in line with the illumination source can make for some very uncomfortable viewing, unless the image has somehow been made very bright. There ought to be a way to block the zero order light using polarizers, but this would require a special needle-like photographic grain to be invented.

2. visibility of the conjugate image
This is a more profound problem: the oppositely-diffracted or "conjugate" image is also directly in line with the desired image! Although it is far enough from the desired image to be substantially out of focus, it still provides a noisy coherent background, which especially degrades the resolution at the edges of the object.

3. back-lit objects
In Gabor's configuration, the only light available for the object comes from scattered reference beam light. This means that a) there isn't much light, b) only forward-scattering and translucent objects can be used, and c) the reference beam will have holes in it, the shadows of the objects. We can get around all of these problems by using a

beam-splitter to bring reflected object light into line with a reference beam, but this defeats some of the other advantages of Gabor holography.

    4. "halo" noise

The halo noise due to object self-interference, or intermodulation, is also centered around the illumination beam, so even those parts of the image not in line with the source, and not obscured by the conjugate image, have a contrast-reducing flare light behind them!

what's OK about "in-liners?"

In spite of the shortcomings of in-line transmission holograms mentioned above, a few are still made each year to take advantage of a few of the characteristics of this hologram type.

    1. low stability required

Because the path length difference between the reference and scattered object beam is relatively insensitive to object motion, the exposure system need not be as mechanically stable as for more advanced hologram types. Also, because the fringe spatial frequencies are typically quite low, the stability of the film holder need not be as high as usual.

    2. holograms can be very big

Very few optical parts are needed, and the low stability allows large pieces of film to be spread out and exposed with a minimum of precautions.

    3. coarse fringes = high-speed film

Because highest-resolution emulsions are not needed, coarser-grained higher speed emulsions may be used (e.g., A-G 10E75 instead of 8E75, a 20X speed gain). This allows much shorter exposure times, again reducing the stability requirements, or the use of a smaller laser or larger film areas. On the other hand, it is only the finest-grained emulsions that produce bright and clear bleached holograms.

    4. only low-coherence laser light needed

The path length variations between the reference beam and the scattered object light are much smaller than for off-axis holograms. Thus the etalon may usually be taken out of an ion laser, and more power obtained, without degrading the image. Very small holograms can even be made without a laser, using a mercury arc source (as Gabor did).

    5. wide-band illumination OK

As we will see, the spectral smear caused by white light illumination "points" in the direction of the light source. The closer we are to looking directly into the light source, the more nearly "end on" we are seeing the smear, and the less visible it becomes. Instead, the various colors stack on top of each other, and an achromatic image appears. As we move to the side of the hologram, color fringing does become apparent, however.

Overall, the disadvantages have vastly outweighed the advantages, and holography has moved on to off-axis techniques to separate the various image components, and allow holography to broaden the scope of its imaging capabilities.

Conclusions:

In-line transmission holography quite literally involves a combination of diffraction gratings to change the direction of light, and holographic lenses to focus the light in three-dimensional space, and form images. The first of these elements, diffraction gratings, has been covered in the chapter on diffraction. Holographic focusing is more subtle, and we have looked at it in two ways, by ray-tracing, and by wavefront curvatures. The second of these gives us the tools to record and reconstruct spherical and astigmatic wavefronts without regard to actual or virtual source locations or image foci. Finally, we showed how the simple imaging properties of several holographic lenses, in the form of Gabor zone plates, can be combined to predict the three-dimensional imaging properties of in-line holograms of extended objects.

This is nearly the end of the road as far as learning new optical concepts goes; only two big new ideas lie ahead. From here, we take these building-block ideas and make lots of new kinds of holograms, exploring the properties of each as we go. That is a little like saying that Ohm's law and the Kirchhoff equations can explain all of electronics, but it is more than a little bit true! The "*sine-theta*" and "*one-over-R*" equations will serve similar roles in explaining what we see as we move ahead.