Big Idea: Consider viewing a scene through a glass window. As you move your head, your view of the scene changes. When an ideal hologram of a scene is illuminated, the electric field that is transmitted through the hologram is exactly the same as the electric field emitted by the original scene. If you replace the glass window with a hologram of the scene, as you move your head, your view of the scene changes because the electric field reaching your eye from the ideal hologram is exactly the same as the electric field reaching your eye from the original scene. Thus, the hologram reproduces both the amplitude and the phase of the electric field emitted by the source. In contrast, if you put a photograph behind the window you see the same thing no matter how you move your head. The electric field emitted by the photograph is completely different from the electric field emitted by the original scene, though the intensity of the light from the photograph is the same as the intensity of the light from the scene when it reaches the window.
Outline:
1. Overview of holograms
2. Review of making a hologram that reproduces a point source, which you did at the end of lab 3.
3. General theory of holograms

1. Overview of holograms:
   a. **Goal:** Make the phase and amplitude of the electric field from the hologram the same as the phase and amplitude from the source. Make this happen by creating an array of point sources at the window that emit an electric field with the same phase an amplitude that the source.

2. Make a hologram of a point source.
   a. **Strategy 1: Huygen’s Wavelet Calculation**
      i. The hologram is the same as the optical mask you created using the LCD at the end of lab 3. That mask focused the plane wave laser by allowing light to pass through the mask if the plane wave had the same phase as the spherical wave, and blocked light where the plane wave have the opposite phase to the spherical wave. This is the same as making a hologram of a spherical wave.
      ii. Review of Lab 3 solutions
          1. Find positions where the plane wave and spherical wave have the same phase, which correspond to positions where the LCD should transmit all of the incoming plane wave.
             a. These are infinitely narrow rings, so no light would get through. To get light through, loosen the condition so that you allow light in places that nearly have the same phase. You can’t let all the light through, or you will get the plane wave back.
          2. Find positions where the plane wave and spherical wave are out of phase, which correspond to positions where the LCD should block all of the incoming plane wave.
             a. These are also infinitely narrow rings, but rings with nearly the same radius emit light that is almost exactly out of phase with the spherical wave.
          3. Arrive at functions that black all of the light that is exactly out of phase, transmit all of the light that is exactly in phase, and connect the two rings using some function that determines the transmission of the LCD at all other positions (see Appendix A for results)
Wavefront derivation of the positions of the transparent radii viewed along the x axis

The optical mask is located in the (0,0,-f) plane. The planar and spherical wave fronts are shown by the black lines and black spheres, respectively. The spherical wave converges onto the origin at (x,y,z)=(0,0,0). An identical interference pattern would be produced by a diverging spherical wave centered at the origin since the wave fronts for the two waves are the same. The only difference is that the kr vector for the diverging wave would point away from the origin. Define

\[ \rho = \sqrt{x^2 + y^2}, \]

as the distance from the z axis in the plane defined by z=-f. Let \( r(x,y) \) be the distance from the origin to a position \((x,y,z=-f)\) in the plane of the LCD. So \( r \) the radius vector in spherical coordinates is given by

\[ r = \sqrt{x^2 + y^2 + f^2} = \sqrt{\rho^2 + f^2}. \]

The positions where the planar and spherical waves are in phase are indicated by positions where wavefronts cross. We have chosen the special case where the phase of the plane wave matches the phase of the spherical wave at the origin. Given that choice, the spherical wave and the plane wave when the path difference varies by an integer multiple of a wavelength, corresponding to \( r \) values such that

\[ r-f = n \lambda. \]

**Substituting for \( r \) and rearranging the equation above gives in phase radii**

\[ \rho = \sqrt{x^2 + y^2} = \sqrt{n \lambda (2f + \lambda n)}. \]

**Out of phase radii are given by**

\[ \rho = \sqrt{x^2 + y^2} = \sqrt{(n+1/2) \lambda (2f + \lambda (n+1)/2))}. \]

Transparent radii in the plane of the LCD (z=-f) viewed along the z axis

\[ \rho = \sqrt{x^2 + y^2}, \]

In phase radii indicated by the red circles. These are the positions where the LCD will be transparent

\[ \rho_{\text{transparent}} = \sqrt{x^2 + y^2} = \sqrt{n \lambda (2f + \lambda n)}. \]

Similarly, out of phase radii correspond to positions in plane of the LCD (z=-f) where the plane wave and the spherical wave have a phase difference of \( \pi \). These are positions where the LCD mask should not pass any light.

\[ \rho_{\text{opaque}} = \sqrt{x^2 + y^2} = \sqrt{(n+1/2) \lambda (2f + \lambda (n+1)/2))}. \]

The mask also represents a converging or diverging spherical wave centered at (-2f,0,0). Transmission of light through the mask does not focus light after the LCD, so it is hard to observe this hologram in your apparatus.
b. **Strategy 2: Make the hologram by interfering a plane wave and spherical wave on the surface of film**
   i. Experimental setup using a concave mirror

   ![Diagram of hologram creation](image)

   **Explanation:** Instead of calculating the positions where a spherical and plane wave would interfere constructively, you can simply allow a plane wave and a spherical wave to interfere with each other on the surface of a photographic plate. The places where the two waves interfere constructively (2nπ phase difference) will produce bright light, whereas the places where they interfere destructively ((2n+1)π phase difference) will be dark. Intermediate phase differences will produce an intermediate intensity.

   Note a spherical wave centered at z=+R, which is shown in the diagram below, would create exactly the same interference pattern on the film as the spherical wave centered at z=−R that was shown in the diagram above.

   ![Diagram of spherical wave interference](image)

   To allow the film to become a hologram, you must choose photographic film that becomes transparent where it is exposed to light and opaque where it is not exposed to light. After the film is developed, the film will transmit at points where the plane wave and the spherical wave were in phase, but will block light where they were out of phase. These are exactly the
conditions that we required for the gray scale patterns derived from the Huygen’s wavelet treatment in the previous section.

After the film has developed, you can remove the mirror and look at the hologram.

**Note:** If you shine a flashlight through this hologram while you look through the hologram, you will see the focused spot which looks like a bright star in empty space. For a discussion of the connection between a plane wave hologram and experiments you did in labs 2 and 3, please see appendix B.

**Calculation of the hologram based on the interference pattern**

A photographic film, indicated by the red line, positioned at $z = f$. The film is then exposed to a total electric field that is the sum of the electric field due to a plane wave and the electric field due to a spherical wave converging on the origin. At points where the waves match, the interference will be constructive creating bright regions. Destructive interference will create dark regions. The developed film will convert a plane wave to a converging spherical wave if the film becomes transparent where the light is bright and opaque where there is no light.

**Making the hologram by exposing the film to the interfering light waves**

After the film is developed, one can display the resulting hologram that converts the plane wave to a spherical wave.
The total field at the photographic plate will be the sum of the E-field due to the plane wave, $E_p$, and the e-field due to the spherical wave, $E_s$.

$$E_{\text{total}} = E_p + E_s.$$ 

If one uses complex notation to describe the E-fields, the total time averaged intensity is given by

$$<\text{Intensity}> = \frac{1}{2} E_{\text{total}} E_{\text{total}}^*.$$ 

For simplicity assume that the plane wave and spherical waves have unit amplitude in the plane of the LCD. Of course the spherical wave amplitude rolls off as $1/r$, but if the focal length $f$ is large and the size of the LCD is small than $r \sim f$ independent of the x and y position in the LCD screen. Thus,

$$<\text{Intensity}> = \frac{1}{2} (E_p + E_s) \frac{1}{2} (E_p + E_s)^* =$$

$$= \frac{1}{2} [\exp(-i(kz - wt + \phi)) + \exp(-i(-kr - wt))] [\exp(i(kz - wt)) + \exp(i(-kr - wt))]$$

$$= \frac{1}{2} (1 + \exp(-i(kz + \phi + kr - wt)) + 1 + \exp(i(kz + \phi + kr - wt))]$$

$$= 1 + \cos[k( r + z) + \phi]$$

$$= \frac{1}{2} \cos^2[(k( r + z) + \phi)/2]$$

Since we are interested in the intensity at the film where $z = -f$, this expression becomes

$$= \frac{1}{2} \cos^2[(k( r - f) + \phi)/2]$$

where $\phi$ corresponds to the phase difference between the plane wave and the spherical wave at $(x,y,z) = (0,0,-f)$ in the plane of the film. This is the phase that we chose in the Huygen’s wavelet derivation. The expression above then says that the hologram will have bright rings at positions $r$ such that

$$k( r - f)/2 = n \pi$$

This is exactly our Huygen’s wavelet result. Similarly, there will be an opaque region where

$$k( r - f)/2 = (2n + 1) \pi$$

Thus, our hologram result agrees with our Huygen’s result, but provides us with a nice grayscale expression for intensity transition at every point in the LCD plane. In addition, it allows us to consider the general case where the phase difference between the two waves is not zero at $(x,y,z) = (0,0,-f)$ The constructive interference condition BETWEEN successive rings remains the same. The propagation difference must still be $n\ell$; however, the center is no longer bright. Thus, the rings are shifted, as illustrated below. For $\phi = 2n\pi$, the center is bright because the plane wave
and the spherical wave are in phase at \((x,y,z)=(0,0,-f)\). For \(\phi=(2n+1)\pi\) the center is dark because the spherical wave and the plane wave are exactly out of phase at at \((x,y,z)=(0,0,-f)\).

\[
\phi=0 \quad \phi=\pi/2 \quad \phi=\pi \quad \phi=3\pi/2 \quad \phi=2\pi
\]

Deep Physics connecting the Fresnel Zone plate with mirages and Feynman’s path integral treatment of quantum mechanics.

The derivation above showed that an LCD mask with appropriately chosen transparent regions could convert a plane wave into a collection of Huygen’s wavelets in the LCD plane such that the wavelets that have the nearly same phase as a chosen converging spherical wave. As a result, most the light from the wavelet sources interferes constructively at the origin, producing a bright focused spot. The ray optics picture of focusing requires not only a bright spot at the origin, but no light anywhere else. How is it that all of those Huygen’s wavelets produce no intensity except at the origin?

The ideal answer would be that they interfere destructively at every point \textbf{EXCEPT} the origin. In reality, the focused spot at the origin is not a point, the light is spread out slightly around the origin. In addition, there is a bit of light outside of the focused spot. The width of the central spot and the intensity in the surrounding rings depends on details of the ring transparency, as well as the number of rings in the zone plate. An ideal lens would have an infinite number of rings, but any real mask has only a finite number of rings. Below we will briefly consider the effect of having only a finite number of rings.

\textbf{Focused spots due to a finite number of rings:}

Two dimensional plots of results for 1 and 2 rings are shown below. For the one ring case there are several rings visible outside the central spot, whereas for 2 rings only one additional ring is clearly visible and the central spot has become narrower.
In order to more clearly illustrate the effects of additional rings, it is useful to consider plots of the intensity along the x axis. Results are shown below for 1 ring (red), 2 rings (orange), 4 rings (green), 8 rings (blue), and 16 rings (purple).

You can see that for just one ring, rings outside the central focused spot still have substantial intensity. As more source rings are added, the intensity becomes more and more confined to the origin. The narrowing of the central region is illustrated by the last figure which includes all of the plots, but confines the x axis to the region from -2 lambda to +2 lambda.

More broadly, if one is considering the intensity associated with waves, the total intensity will be small except at positions where the waves interfere constructively. This is the origin of mirages, one of which is shown below from a Wikipedia posting by Brocken Inaglory, and a video can be seen at https://www.youtube.com/watch?feature=player_embedded&v=S9CztTYuGqg
Usually, people say that “light travels in straight lines,” but the light in mirages travels in a curved line as illustrated in the Encyclopedia Britannica illustration shown on the right below. The common explanation is that light is following the curved line because the “optical path is shorter” because hot air has a lower index of refraction, as you observed in lab 2. Thus, traveling a longer distance through a medium with a lower index of refraction can produce a shorter optical path that traveling along a straight line through a higher index medium. This leads to the question: how does light know that the curved path will be shorter given that it is not a straight line? Feynman’s answer is that light takes EVERY path, but we only see it along paths that interfere constructively just as in the Fresnel zone plate experiment we see light only at the focused spot where all of the fields interfere constructively even though the electric fields due to individual source rings in the zone plate have significant amplitude outside of the central spot.

Particles follow a similar rule, which is the basis for Feynman’s path integral version of quantum mechanics. Particles travel every possible path, but we only observe them on paths where nearby trajectories interfere constructively because they have nearly the same phase, which means that the path is either minimal (shortest) or maximal (longest).

**Generalization to an arbitrary source:**

1. **Physics view 1:** The E-field due to any arbitrary source can be expressed as the sum of the E-fields due to point sources. Thus, you can add up the holograms due to all of the point sources in your object in order to arrive at the hologram for the entire source.

2. **Physics view 2:** Position your film in a plane of constant z. Let that plane of constant z receive light from a plane wave source and from the plane wave source that is reflected by your object. The resulting interference pattern on your film will create an appropriate transmission mask that will transform your plane wave into the electric field from your object. Not by definition the electric field of your plane was Ep will be independent of the x and y coordinates in your film.
Traditionally, the setups were more complicated. Below is the setup that I used as an undergrad illustrated using the Mach-Zender diagram from Lab 2. The Wikipedia version is shown on the right, along with the Wikipedia illustration of how the final hologram can be viewed. Remember that there are four possible wavefronts that could produce each particular hologram.

Holograms using Mach-Zender

3. Math view justification for view 2: The total field at the photographic plate will be the sum of the E-field due to the plane wave, $E_p$, and the e-field due to the image, $E_{image}$.

$$E_{total} = E_p + E_{image}.$$ 

If one uses complex notation to describe the E-fields, the total time averaged intensity is given by

$$<\text{Intensity}> = \frac{1}{2} E_{total} E_{total}^*$$

For simplicity assume that the plane wave and spherical waves have unit amplitude in the plane of the LCD. Of course the spherical wave amplitude rolls off as $1/r$, but if the focal length $f$ is large and the size of the LCD is small than $r~f$ independent of the $x$ and $y$ position in the LCD screen. Thus,

$$<\text{Intensity} (x,y,z=\text{film position}=z_{\text{film}})> = \frac{1}{2} (E_p + E_{image}) (E_p + E_{image})^* =$$
\[
= \frac{1}{2} \left( |E_p|^2 + |E_{\text{image}}|^2 + (E_p \cdot E_{\text{image}}^* + E_{\text{image}}^* \cdot E_p) \right)
\]
\[
= \frac{1}{2} |E_p|^2 + \frac{1}{2} |E_{\text{image}}|^2 + |E_p| |E_{\text{image}}| \cos(\Delta \Phi)
\]

Where \(\Delta \Phi\) is the phase difference between the phase of the plane wave and the phase of the electric field due to the image at position \((x,y, z_{\text{film}})\).

Illuminate with \(E_p\), result is
\[
E_{\text{hologram}} = (\frac{1}{2} |E_p|^2 + \frac{1}{2} |E_{\text{image}}|^2 + |E_p| |E_{\text{image}}| \cos(\Delta \Phi)) E_p
\]

Thus, the light has the phase of the plane wave. At positions where \(\Delta \Phi=0\) this is also the phase of the source wave.

Special cases: points where \(\Delta \Phi=0\)

At those positions, the amplitude of the transmitted E-field is
\[
E_{\text{hologram}} = (\frac{1}{2} |E_p|^2 + \frac{1}{2} |E_{\text{image}}|^2 + |E_p| |E_{\text{image}}| \cos(\Delta \Phi)) E_p
\]

If \(|E_p| >> |E_{\text{image}}|\), the \(|E_{\text{image}}|^2\) term can be ignored. By definition, for a plane wave \(E_p\) is independent of \(x\) and \(y\). The transmitted E-field is then proportional to
\[
E_{\text{hologram}} \text{ is proportional to } |E_p|^2 \left(1 + \frac{|E_{\text{image}}|}{|E_p|} \right) E_p
\]

The constant term just provides a background, so at positions where \(\Delta \Phi=0\) the transmitted electric field has the same phase as the source \(E_{\text{field}}\) and an amplitude proportional to the source field. Thus, those points will reproduce the source field. These are the idea points in your hologram. If those points are sparse, not enough light gets through. To get more light into your hologram, you have to let through some plane wave light whose phase does not exactly match the phase of the source

\[
E_{\text{hologram}} \text{ is proportional to } |E_p|^2 \left(1 + \frac{|E_{\text{image}}|}{|E_p|} \cos(\Delta \Phi) \right)
\]

The since \(|E_p| >> |E_{\text{image}}|\) and \(E_p\) is independent of \(x\) and \(y\), the spatially varying term in
\[
E_{\text{hologram}} \text{ is proportional to } |E_{\text{image}}| | \cos^2(\Delta \Phi/2)
\]

Thus, it has the same amplitude as the image at that point and the intensity has an additional term that is weighted in favor of positions where the phase of the plane wave matches the phase of the source. Some other light gets through, making the reproduction imperfect; however, those points provide enough intensity to make the hologram visible. If the film has a non-linear response to light, other functions are possible.
Appendix A: Mask transmission functions that produce focused spots

**Black and White Cases:** For each \( n \), one can find the radius corresponding to the center of the white zone. \( R[n] = \sqrt{n \lambda (2f + \lambda n)} \), where \( n \) is an integer. Similarly, the center of the black zones are at \( R[n+1/2] = \sqrt{(n+1/2)\lambda (2f + \lambda (n+1/2))} \). The half width of the dark fringe on the side nearest the origin can be chosen as \( \alpha (R[n+1/2] - r[n]) \), and the width on the other side can be chosen as \( \alpha (R[n+1] - R[n+1/2]) \). \( \alpha \) allows one to tune the fraction of the pattern that is open. The result for \( \alpha = 1/4 \) is on the left below. The one on the center has \( \alpha = 1/3 \), and the one on the right has \( \alpha = 1/2 \).

**Gray Scale Cases:** Gray scale allows mask where the fraction of the plane wave transmitted can be varied continuously from 0 to 1. A hologram naturally creates such a gray scale mask. When making a hologram, the interference between the plane reference wave and the wave scattered from the object creates an interference pattern on the film. That interference pattern is proportional to the time averaged intensity created by the interfering light beams. If \( E_1 \) is the electric field from the reference plane wave and \( E_2 \) is the electric field from the scattering object, then the time averaged intensity at any position on the film is \( \frac{1}{2} (E_1 + E_2) (E_1^* + E_2^*) = |E_1|^2 + |E_2|^2 + 2 |E_1||E_2| \cos[\Delta \Phi] \). If \( |E_1| = |E_2| = 1 \), then the time average intensity is proportional to \( \cos^2[\Delta \Phi/2] \).

Let \( E_1 \) be the plane wave with e-field \( A_p \exp[i(kz-wt)] \), and \( E_2 \) the spherical wave with e-field \( A_s \exp[i(kr-wt)/r] \). \( r = \sqrt{x^2 + y^2 + f^2} \). We have chosen a plane where \( z = f \), so this becomes \( r = \sqrt{x^2 + y^2 + f^2} \). Since \( |x|, |y| << z \), the \( \rho \) in the numerator can be replaced by \( f \), so \( E_2 = A_s \exp[i(k(x^2 + y^2 + f^2) - wt)/f] \).

Thus \( \Delta \Phi = k(\sqrt{x^2 + y^2 + f^2} - f) \).

The graph on the left below is \( \cos^2[k(r-f)/2] = \cos^2[k(\sqrt{f^2 + \rho^2} - f)] \).

For the graph the distance units are chosen to be equal to the wavelength \( \lambda \). In those units, \( R = 10 \) and the x and y axis range from -20 to 20.

It resembles the result obtained in the black and white case when \( \alpha = 1/3 \) was chosen. The actual overlap between the alpha=1/3 function and the \( \cos^2[k(\sqrt{f^2 + \rho^2} - f)/2] \) pattern is shown in the upper right image where the black and white grating is shown in blue and white and the gray scale appears underneath it. One can get black and white from grayscale by rounding the grayscale function. Rounding converts any value < 0.5.
to 0 and any value >0.5 to 1. The result for Round[Cos²[k (Sqrt[f²+ρ²]-f)/2]] is shown in the lower left image. The lower right image is a graph of Cos²[k (Sqrt[f²+ρ²]-f)/2]. with the x and y axis limits are changed to -40 to 40, providing more rings in the image.

The Cos²[k (Sqrt[f²+ρ²]-f)/2]. is the function that would be obtained using a hologram, but it is not the only possible grayscale function.

Cos⁴[k (Sqrt[f²+ρ²]-f)/2]. is another possible function. The resulting pattern is shown on the left below. Abs[Cos²[k (Sqrt[f²+ρ²]-f)/2]], is shown in the center and Sqrt[Abs[Cos²[k (Sqrt[f²+ρ²]-f)/2]].] is shown in the right.
Appendix B: Plane Wave Holograms

In optics Lab 1, you made a Mach-Zender interferometer and found that you could create horizontal or vertical fringes on a screen at the output by adjusting the relative angle between the two beams in the interferometer. A description from the lab solution is reproduced below. The green and black bars are crude cartoons of the intensity pattern. A more accurate black and white version is shown to the right of the schematic cartoon.

Slightly misaligned plane waves

One can make a plane wave hologram by replacing the viewing screen in the Mach-Zender experiment with holographic film, as illustrated below where the red line corresponds to the position of the holographic film.
Lab 3 had a section called **Fourier transform optics beyond slits** that included a subsection called periodic gratings. The smooth grating was exactly the grating that would appear on the holographic film in the illustration above. In Lab 3, you sent a plane wave through the grating and observed two spots in the Fourier transform plane, which correspond to the two plane waves that propagate to the right of the holographic film in the illustration above. The interference of EITHER beam with the incoming plane wave would produce the intensity pattern in the hologram, which is why the hologram produces BOTH beams even though it was created using only 1.

This is true in general. The output of the hologram includes two different E-fields. One E-field corresponds to the forward propagating field that would produce the intensity pattern in the hologram if it interfered with the incoming reference plane wave. The other E-field corresponds to the backward propagating field that would produce the intensity pattern if it interfered with the incoming reference plane wave.

For the case of the case where the holographic film is in the $z=-f$ plane, if the hologram produces a converging spherical wave that focuses at the origin, then the second E-field is a diverging spherical wave with origin at $(x,y,z)=(0,0,-2f)$. Since it is diverging, the resulting light is very dim in the forward propagating direction. Thus, it just formed part of the background that you observed around the focused spot produced by the Fresnel lens. The two image beams are illustrated below.
Two Holographic Image Beams in the Spherical Wave Hologram

Converging Spherical Wave

\[ k_z \]
\[ k_r \]

Film at \( z = -f \)

Diverging Spherical Wave Would Produce the Identical Interference Pattern in the Film

\[ \rightarrow k_z \]
\[ \rightarrow k_r \]
where $\phi$ corresponds to the phase difference between the plane wave and the spherical wave at $(x,y,z)=(0,0,-f)$ in the plane of the film. This is the phase that we chose in the Huygen’s wavelet derivation. The expression above then says that the hologram will have bright rings at positions $r$ such that

$$k( r-f)/2 = n \pi$$

This is exactly our Huygen’s wavelet result. Similarly, there will be an opaque region where

$$k( r-f)/2 = (2n+1) \pi$$