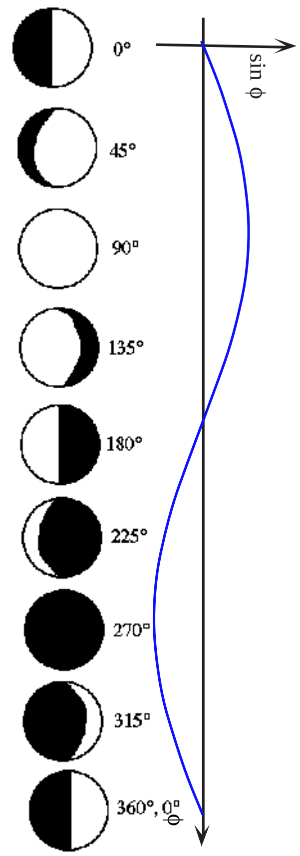
# Chapter 3: Waves and Phases

Our discussion so far has described light waves as three-dimensional phenomena, especially their electric fields as functions of $(x,y,z,t)$. But for what follows we want to concentrate on the behavior of waves as they cross specific two-dimensional planes, with their time dependence suppressed (because they will all come from the same laser, and have the same frequency). We need a way to describe the shape of the wave in mathematical terms. To do this, we will introduce the notion of the phase of the wavefront, as determined by its relative delay in arriving at various points on the measurement plane. The observation of the pulsing from a point source in space is delayed, relative to the source, by a time proportional to the straight-line path length between the source and the observation point. If the source is repetitive or cyclic, from a sine-wave source for example, then we can also express the delay as a fraction of the repetition time or period, $T$. This in turn is represented as a fraction of 360 degrees or $2\pi$ radians, the angle that a wheel turns in generating a full cycle of a sinusoid. As a rule, we are not interested in the number of whole cycles of delay, but in the fraction beyond the nearest whole cycle.




**Wave phase**

Our most common notions of phase probably come from the "phases of the moon," and that is not a bad place to start! The moon goes through its cycles very reliably and repetitively, with a period of 28 days. We think of the phases as full, half (coming and going, or waxing and waning), and new-moon (with gibbous in between somewhere, and ignoring eclipses). We can think of wave phases in much the same way, except that the fraction of a full cycle is measured in degrees, with 360° representing a full cycle, or the entire repetition time (or "period"). Describing the moon's cycle by a sinusoidal variation that is roughly the illuminated area (more nearly, $0.5 + 0.5\sin(2\pi t/T)$), a waxing half moon would be the zero degree mark, a full moon at the 90° mark, a waning half moon the 180°, a new moon at 270°, and the waxing new moon at 360° or 0° (they look the same in this modulo-360° math).

Formally, we would describe a wave as having an amplitude and a phase, with the frequency, $\nu$ (the Greek letter "new"), usually being left implicit. Thus we would say for a spherical wave in 3-D space:

$$E(r,t) = \frac{E_0}{r}\sin\left(2\pi\frac{\left(t-\frac{r}{c}\right)}{T}\right) = \frac{E_0}{r}\sin\left(2\pi\left(\nu t - \frac{r}{\lambda}\right)\right) \qquad (1)$$

$$= E(r)\sin\bigl(2\pi\nu t - \phi(r)\bigr),$$

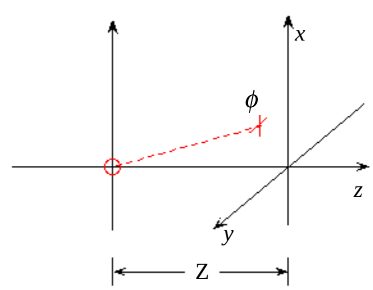
where the electric-field amplitude (as a function of distance from the source, $r$), $E(r)$, is

$$E(r) = \frac{E_0}{r} = \frac{E_0}{\sqrt{x^2+y^2+z^2}}, \qquad (2)$$

and the phase, denoted by $\phi$ (the Greek letter "fie" or "fee") as a function of $r$, is

$$\phi(r) = \frac{2\pi}{\lambda}r = \frac{2\pi}{\lambda}\sqrt{x^2+y^2+z^2}. \qquad (3)$$

The fun begins when we look at specific analytical expressions for $\phi(r)$ and try to guess what kind of wave produced them, and where it is going! To do this, we will limit our attention to a single $x$-$y$ plane in space, and try to identify the characteristic phase patterns, or "phase footprints" of some typical waves.



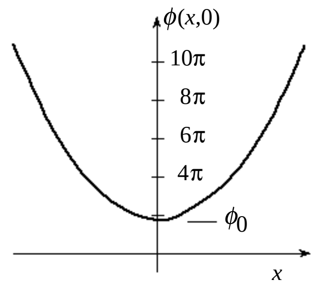

An on-axis spherical wave:

Imagine that a point source is located at (0,0,0), and our observation plane is located at $z = +Z$, as sketched on the right. What then is the form of the phase function, $\phi(x,y)$, in this plane? We have seen that the phase increases linearly with distance from the source, and so increases as we move away from the

$(x,y)=(0,0)$ point, on the $z$-axis. Further, the phase stays the same as we move in a circle around the $z$-axis, because the distance from the source is a constant as the line from the source to the observation point sweeps out a cone. Plugging in the expression from Eq. 2, which we will use over and over again, we find

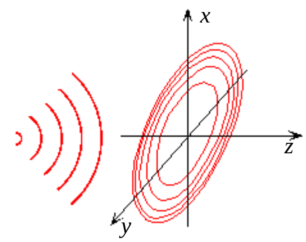
$$\phi(r) = \frac{2\pi}{\lambda}\sqrt{x^2+y^2+z^2}$$
$$= \frac{2\pi}{\lambda}\sqrt{x^2+y^2+\mathrm{Z}^2} \ . \qquad (4)$$

To simplify the equations we must apply some approximations: namely, that the angles of the lines from the source to the observation plane are small, so that the $x,y$ distances are much smaller than Z. In this case, we can use the binomial theorem to simplify the equations to:




$$\phi(x,y,\mathrm{Z}) = \frac{2\pi}{\lambda}\sqrt{x^2+y^2+\mathrm{Z}^2}$$
$$= \frac{2\pi}{\lambda}\mathrm{Z}\sqrt{1+\left(\frac{x}{\mathrm{Z}}\right)^2+\left(\frac{y}{\mathrm{Z}}\right)^2} \qquad (5)$$
$$\approx \frac{2\pi}{\lambda}\mathrm{Z} + \frac{\pi}{\lambda \mathrm{Z}}\left(x^2+y^2\right) \ .$$

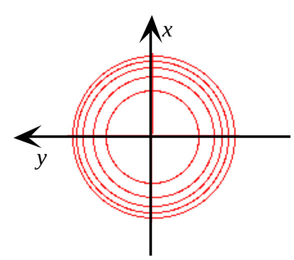
Now, we can identify Z as equal to the radius of curvature, $R$, of the spherical wave when it first reaches the observation plane. And, because this observation plane may be anywhere along the $z$ axis, we will consider the phase to be a function only of the coordinates within the plane, $x$ and $y$. The phase pattern of a diverging spherical wave, what we will call its "phase footprint," is then:



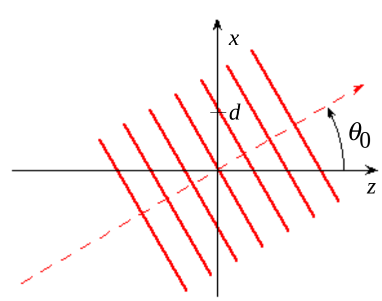

$$\phi(x,y) = \frac{2\pi}{\lambda}\mathrm{R} + \frac{\pi}{\lambda \mathrm{R}}\left(x^2+y^2\right) . \qquad (6)$$

The first term represents a constant (over the plane) phase delay due to the time that it took the wave to get to the plane at all; we can call it $\phi_0$. Because the phase of the source is itself unknown, we will usually ignore this constant phase, and emphasize the term that has variation with $x$ and $y$, calling it the "footprint" pattern that will reveal the shape of the wave that caused it. In this case it is a "parabolic" term, with only second order terms, but in both $x$ and $y$, and with the same coefficient for both directions.

On-axis spherical wave phase footprint:
This phase pattern is worth looking at in more graphical detail, as seen from the $+z$ direction and sketched here. The concentric circles represent the loci of points of equal phase, which we assign to be zero degrees (successive multiples of 360°, that is) for simplicity. The lines are circles because the phase is constant wherever $x^2+y^2$ is a constant. But the circles have radii that increases more and more slowly as the phase increases; if the $n$-th circle represents a phase of $n{\cdot}360°$ greater than the center, then the radius of the $n$-th circle is proportional to the square root of $n$. That means that the area between successive circles is a constant, by the way!

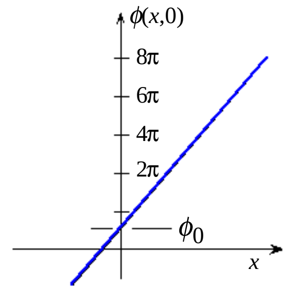



Off-axis plane wave:
Let's now consider the case of a point source so far away that the wavefronts hitting the observation plane (at $z$=0) are effectively flat. The source is at $-\mathrm{X}_0$, $-\mathrm{Z}_0$, which are both very large compared to the size of the observation plane, but in a ratio that determines the inclination of the wave vector—which is perpendicular to the wavefront—to be $\theta_0$ degrees to the $z$-axis, where $\theta_0 = \tan^{-1}(\mathrm{X}_0/\mathrm{Z}_0)$. The waves crossing the $x$-$y$ plane as seen here:



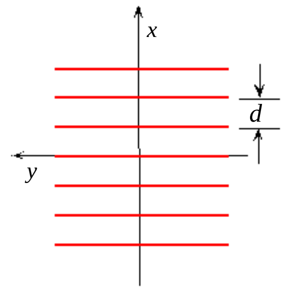

The higher up the $x$-axis we go, the farther away from the source we find ourselves, so that the wavefront phase increases with increasing $x$. If $\phi_0$.is subtracted from the phase at $x$=0 (as usual), the phase increases linearly with distance, $x$. Further, from the geometry (or by substitution into Eq. 3), we can see that

$$\phi(x,y) = \phi_0 + \frac{2\pi}{\lambda}\sin\theta_0 \cdot x \ . \tag{7}$$




Off-axis plane wave phase footprint:
The phase is constant along any $x$ = *constant* line, independent of $y$, because the plane wave is inclined purely vertically. As seen from the +$z$ direction, we can sketch the lines where the phase is equal to zero degrees (again), which are now straight horizontal lines. We find that the spacing of the lines, d, is inversely proportional to the (sine of the) angle of inclination, $\theta_0$.

$$d = \frac{\lambda}{\sin\theta_0} \ . \tag{8}$$

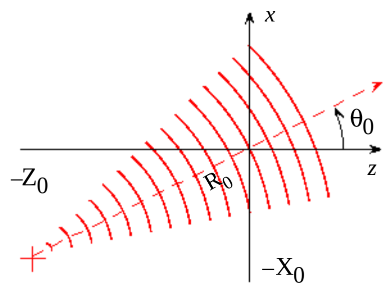



Note: If instead the source were off to the side, somewhere along the $y$-axis ($X_0$=0), the constant-$\phi$ lines would become vertical. The spacing of the constant-phase lines depends only on the angle of the wave vector to the $z$-axis; as the source revolves around the $z$-axis, the lines rotate around the $x$,$y$=0 origin, staying perpendicular to the plane defined by the $z$-axis and the wave vector. We will only rarely consider waves not in the $x$-$z$ plane, though.

Off-axis spherical waves:
The most general case of a spherical wave is the off-axis diverging wave, which forces us to grapple with a few strange new ideas. Now, we consider the source to be at -$X_0$,-$Z_0$, where these are not very large numbers (that is, we have to take wavefront curvature into account at last).

Converting to polar coordinates, we can express the radius of curvature and inclination of the wavefront at the origin as:

$$R_0 = \sqrt{X_0^2 + Z_0^2}$$
$$\theta_0 = \tan^{-1}\left(\frac{X_0}{Z_0}\right) . \tag{9}$$




Now the problem is to find $r$ (the distance from the source) as a function of $x$ and $y$ (the location in the $z$=0 plane). Because we are interested in only a small area around the origin, we can express $r$ as a power series expansion with acceptable accuracy:

$$r(x,y) = \left((x+X_0)^2 + y^2 + (z+Z_0)^2\right)^{\frac{1}{2}}$$
$$\approx R_0 + Ax + By + Cx^2 + Dy^2 + Exy + \ldots \ . \tag{10}$$

The derivation of the coefficients is left as an exercise for the reader, but the answers are shown without too much difficulty to be:

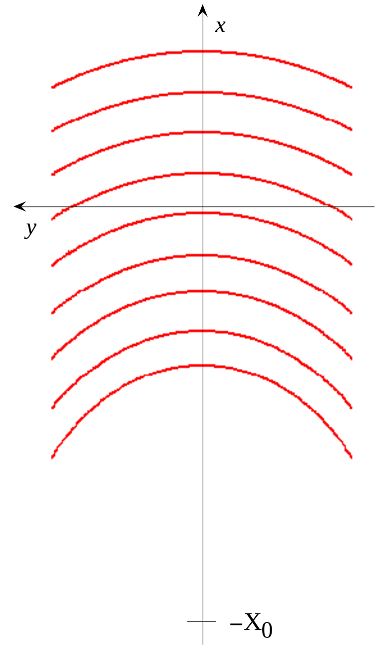
$$A = \sin\theta_0$$
$$B = 0$$
$$C = \frac{\cos^2\theta_0}{2R_0}$$
$$D = \frac{1}{2R_0}$$
$$E = 0 \ . \tag{11}$$

Plugging these into the expression for the phase then gives

$$\phi(x,y) = \phi_0 + \frac{2\pi}{\lambda}\sin\theta_0 \cdot x + \frac{\pi}{\lambda R_0}\left(\cos^2\theta_0 \cdot x^2 + y^2\right) , \tag{12}$$

which is what we will use in all our future work. Note that it seems like a logical combination of the on-axis spherical waves and off-axis plane waves we have seen so far, except for the peculiar $\cos^2\theta_0$ term; you should satisfy yourself that this term is correct, if not logical, before going any further! The wavefront itself has, by the definition of a spherical wave, equal physical curvatures in the *x*- and *y*-directions, namely $R_0$. However, the "mathematical curvatures," or coefficients of the second-order phase terms, are different, which is likely to be confusing.

Off-axis spherical wave phase footprint:
Our sketch exaggerates the difference, but connotes how the "phase footprint" of this generalized spherical wave might look. You should satisfy yourself that it reduces to both the on-axis spherical wave case and the off-axis plane wave case under the proper conditions (namely, $\theta_0 = 0$ or $R_0 = \infty$).

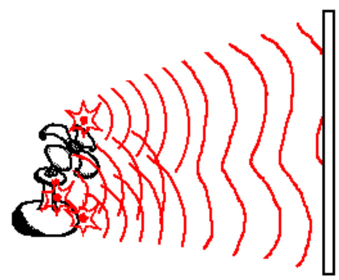



Given a sketch of a wavefront, or better yet its analytical expression, we are now prepared to work backwards and determine its radius of curvature and inclination in a general spherical-wave case. It may even be that the wavefront has two radii of curvature, different ones in the *x*- and *y*- axis directions, but that is a story that has yet to come, and when it does it will be called *astigmatism*!

**The local inclination and divergence of a complex wave:**

In analyzing holograms, we will often be dealing with complex waves that are reflected by complex three-dimensional objects. We will model such waves in two different ways:

1) Most often, we will treat complex wavefronts as the composite of many spherical wavefronts emitted by point-like areas on the surface. When the recording and playback processes can be shown to be roughly linear, what is true for each of the spherical waves individually will be true for their combination, even for millions of them at a time. This approach is akin to the linear systems theory style of electrical engineering analysis, where complex waveforms are built up from a superposition of sinusoidal (Fourier) elemental components.



2) In other cases, we will consider even a complex wave to have a slowly varying amplitude and modest wavefront curvature at every point on the recording medium. Which is to say that the amplitude and curvatures are constant over areas that are several wavelengths on a side. This, in turn, requires that the complex objects subtend an angle that is much less than 180° in any direction (usually well under 30°). Then, we can model the wave at every small area as a spherical wavelet (or perhaps an astigmatic wavelet). We can then calculate the wavefront for that area at the output side of the hologram, stitch the small areas together, and predict what the entire output wave will be like (or at least what its relationship to the input wave will be). This is the "patchwise-spherical" model, and will be especially useful if we have to deal with strong non-linearities in the recording/playback response.

**Conclusions:**

There are many ways to describe waves, each of which highlights an aspect that is important to a certain kind of problems. Laser light is highly coherent, which means that it seems to come from a well-defined point source, and that it is of a single frequency, so we don't have to worry about dealing carefully with the frequency or wavelength of the light, and can concentrate instead on its spatial variations. Our concerns with interference and diffraction will make a careful account of the phases of wavefronts especially important. Phase will tell us "where the light goes." We will be a little less interested in the wave amplitudes, as we are a little less interested in "how much light gets there," for the time being anyway.