# Notes on Huygens' Principle



In the 17th Century, Christiaan Huygens (1629–1695) proposed what we now know as Huygens' Principle. We often invoke Huygens' Principle as one of the fundamental concepts of waves and wave optics, but textbooks seldom justify it or even explain it in much detail. A typical statement of the principle is "every point on a wavefront acts as a source of a new wavefront, propagating radially outward." To understand what Huygens meant by this, we first need to consider the state of science in Huygens' time.

René Descartes (1564–1642) had recently founded a philosophy based on the notion that all truth could be deduced from a few first principles. Part of Descartes' success was his lucid theory of the universe and how it works. In fact, for the first time since the Copernican world view disrupted the Aristotelean world view, the universe again seemed wholly fathomable. Descartes proposed that all physical process could be described in terms of matter and motion. As matter moved, it collided, until the motion formed a vortex. The universe consisted of an infinite number of adjacent vortices. Matter, through this collision process, had been broken down into chunks of different sizes. The smallest chunks were called "first matter." First matter settled to the center of the vortex, got very hot and formed the sun. Other bodies of first matter coalesced to form the core of the planets. Some large pieces of matter were cold and adhered to the surface of the sun as sun spots. To the surfaces of the smaller bodies of first matter, the planets, larger amounts of this cold matter adhered until the planets became entirely coated in it. This coarse, cold matter was termed "third matter." Matter gave extent, so that there was no matter without space and no space without matter. In between the sun and the earth, another sort of matter must then exist. This was called "second matter," and was thought of as the chunks of matter intermediate in size between first and third matter. Descartes described light as a "pressure" in the second matter. Just as water instantaneously comes out of a faucet when we open a valve, light instantaneously traveled from the sun to the earth when the sun applied pressure to the second matter.

Huygens generally subscribed to Descates' world view, but he worked to quantify what Descartes had only described in words. (This work of Huygens, which produced the Law of Conservation of Momentum and the $1 / R^2$ dependence of the centrifugal force, was instrumental in establishing modern physics.) Huygens took exception with Descartes, however, on his theory of light. Huygens reasoned that light could not be a pressure. For a person to see an object, the object must exert a pressure on the second matter in a direction from the object to the observer's eye. If two observers were looking at each other, there would be one pressure from the first observer to the second and another in the opposite direction from the second observer to the first. The two pressures would cancel, and neither person could see the other person's eyes. Huygens then proposed that light was a disturbance in the second matter. A light source would cause a disturbance in the second matter adjacent to it, and that matter would cause the matter adjacent to it to be disturbed, and so on.

In terms of this model, Huygens' Principle is just that 1) disturbances are passed on from one piece of matter to another, and 2) each piece of matter becomes a point source with light propagating radially outward in all directions from it. Huygens is often considered to be one of the early advocates of a wave theory of light, but he didn't propose a wave train with crests and troughs, wavelength and frequency. He thought of light as just a disturbance passing from one piece of matter to another.

One logical conclusion of this principle related to the propagation of wavefronts. Huygens' idea of a wavefront was the maximum extent a disturbance reached at a given instant. If we know the wavefront of a light pulse at one time, Huygens' Principle then allows us to predict the wavefront at later times. Let's see how this works for one specific case. We take a small source that has just emitted a pulse. The wavefront is  represented by the larger circle in the Figure 1.

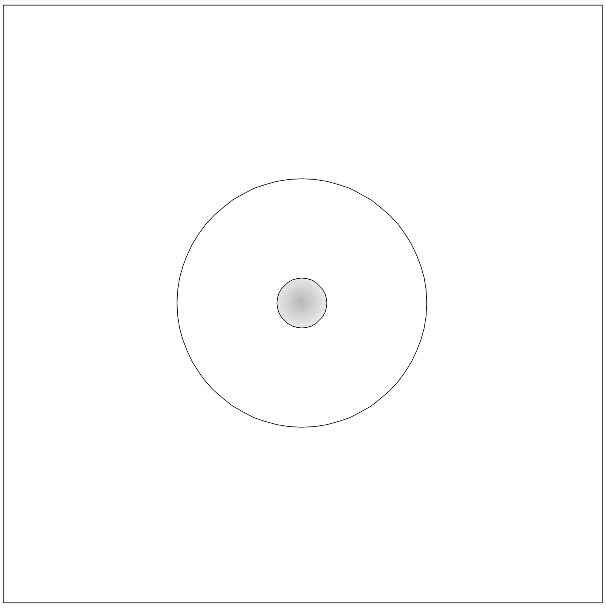

**Figure 1** A source with a wavefront emanating from it.

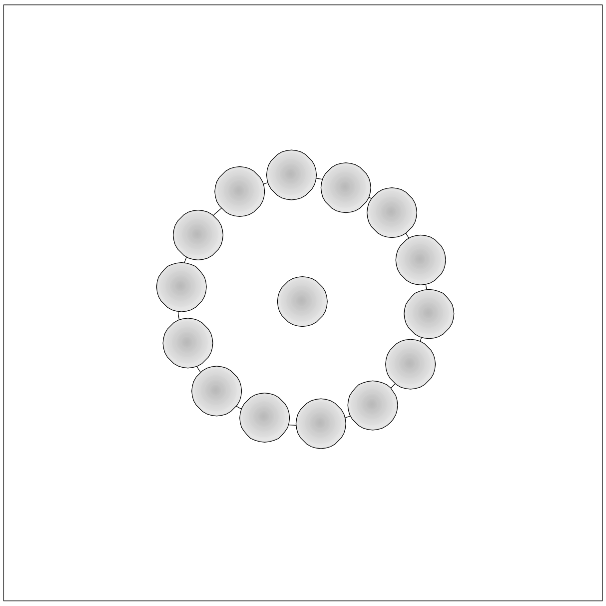

**Figure 2** Secondary sources are placed on the wavefront.

Each point on this wavefront can be considered a new point source. We really should draw an infinite number of sources, but we'll be content with just a few.

When each of these sources in turn emit a new pulse, they form a new wavefront. The new wavefront (heavy solid line) must be the outside envelope of all the wavelets emitted by the point sources located on the old wavefront (heavy dashed line).

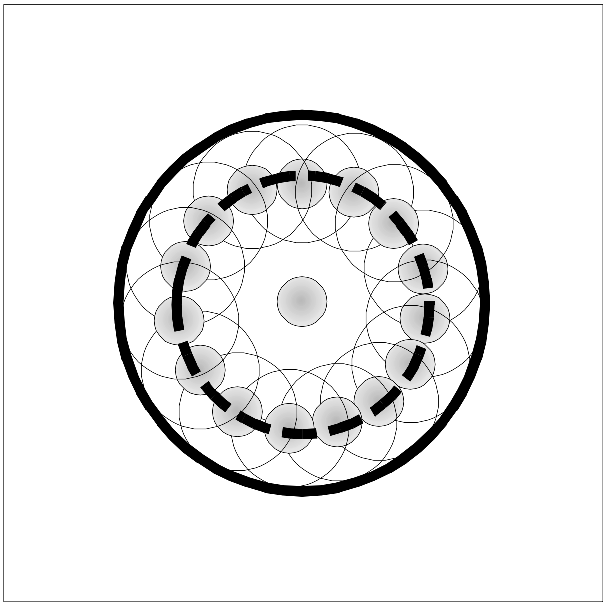

**Figure 3** The secondary sources emit wavelets which determine the position of the new wavefront.

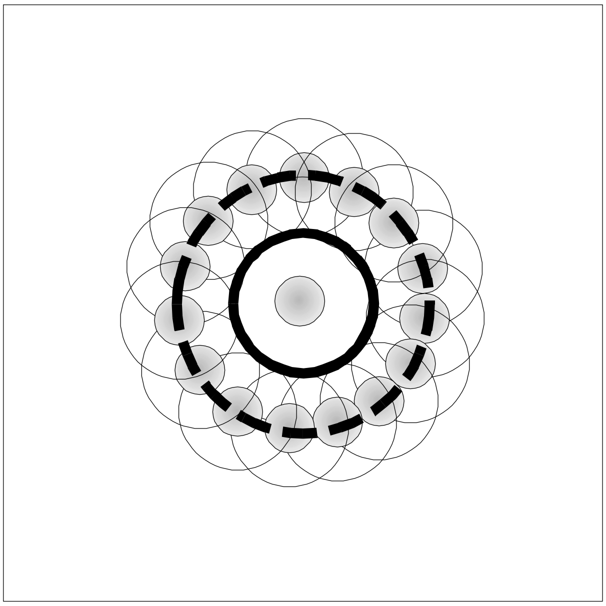

**Figure 4** If the wave were traveling inward, the new wavefront would be at smaller radius.

On the other hand, if the pulse were traveling inward, we would have to choose the inside envelope of all these wavelets to be the new wavefront.

Modern textbooks apply Huygens' Principle to argue that a wavefront passing through a narrow opening forms a circular wavefront on the opposite side of the opening. In the case to the right, plane wavefronts (dashed straight lines) travels upward and passes through a slit, making a circular wavefront (dashed semicircle) on the far side of the slit. It is ironic that Huygens believed that light could not bend around corners and had to argue that only part of the circular wavefront would survive when light went through a slit – the part that did not bend around the corners of the slit!

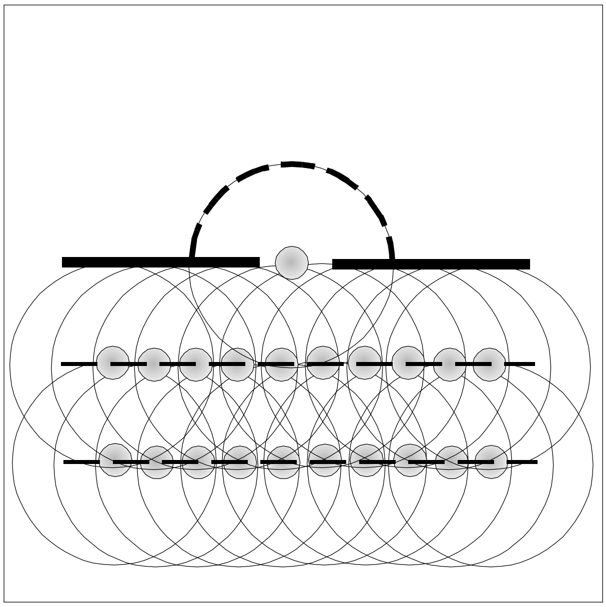

**Figure 5** A plane wave passing through an opening produces a spherical wavefront.

Whereas we apply Huygens' Principle to wave optics, Huygens' applied it to the question of refraction. If we take a plane wavefront of light striking an interface between two different materials at an angle, we can demonstrate that the wavefront will bend. A geometric analysis of the situation leads to Snell's Law. In fact, in parts of the world Snell's Law is known as Huygens' Law.

In class, our entire discussion of wave optics will be based on Huygens' Principle. We will consider what happens when light passes through two very narrow slits and the waves emerging from the slits interfere. Then we will generalize this to many slits, and to a wide slit which is treated as an infinite number of narrow slits placed close together. Finally we will study the interference of light through multiple wide slits.

But if Huygens' Principle is only a qualitative principle, how well can we trust it to give believable results when we apply the principle to wave optics? The answer is that we really don't apply Huygens' Principle to wave optics. We apply a different principle roughly similar to Huygens' Principle, but one which works.

To understand this, let's look at the problem from the point of view of electromagnetic fields. Since light is an electromagnetic wave, this is a reasonable approach. First, we begin with a small, positive point charge, $+q$. Now we need the charge to interact with something. In general, the interaction of electromagnetic fields with matter is complicated and tied into quantum mechanics, but we can understand the behavior of a perfect conductor. When a perfect conductor is in an electric field, electrons in the conductor feel a force and move until there remains no electric field at all in the conductor. We then place our positive charge in the vicinity of an infinite, grounded conducting plane. We know that the positive charge causes a total accumulation of charge $-q$ on the conductor and that the charge is distributed in such a way that the electric field is zero inside the conductor and on the opposite side of the conductor from the charge. In Figure 6, we draw electric field lines for the charge (dashed) and the plane (dotted) separately. Note that the field below the plane cancels out. The net field above the plane is shown in Figure 7.

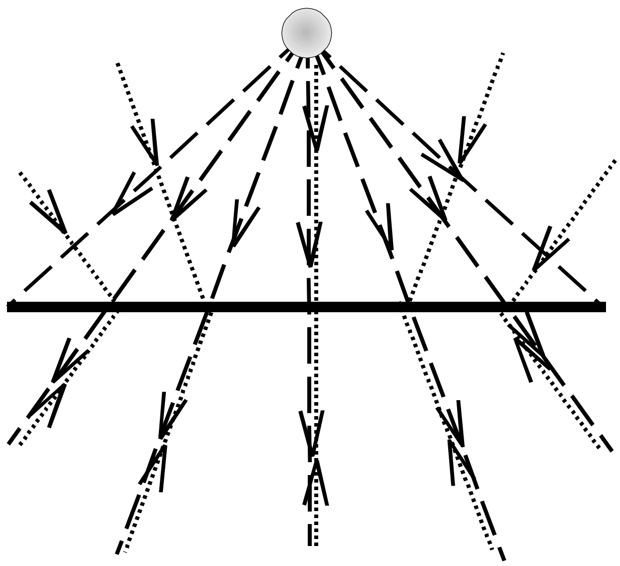

**Figure 6** The electric field of a positive charge placed near a conducting plane.

Now, let us "freeze" the negative charge on the plane so that it cannot move and then drill

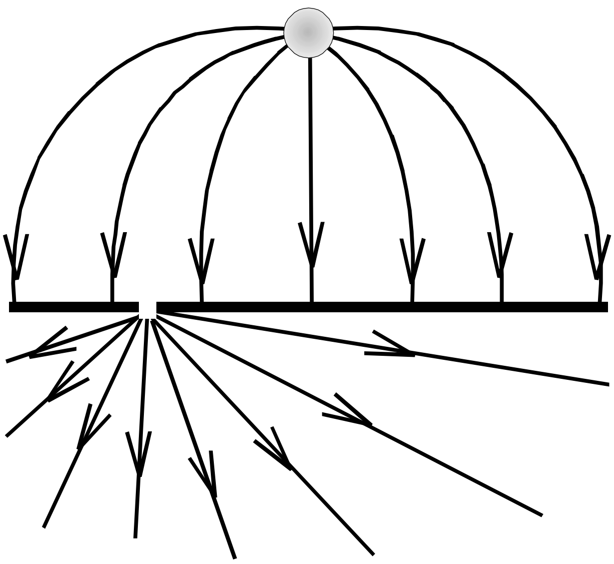

**Figure 7** The electric field with a hole in the plane.

a small hole in the plane. This is equivalent to adding a bit of positive charge on the plane to counteract the negative charge. The only field below the plane is that of this small positive charge on the plane. On the far side of the plane, the entire system of positive charge, plane, and hole is equivalent to a point positive charge on the plane.

In reality, we can't freeze charge on the conductor. A ring of negative charge will surround the hole and tend to cancel the field of the hole. But we can apply a similar argument to the case of light passing through an opening.

**Figure 8** An electric dipole.

Consider an oscillating electric dipole. A dipole can be thought of as a rod with a positive charge on one end and an equal and opposite negative charge on the other end. The net charge on the dipole is therefore zero. If the dipole is oscillating, the charge on each end varies sinusoidally, so sometimes the left end is positive with the right end negative, and sometimes the right end is positive with the left negative. At a distance from the dipole, this oscillating electric field, joined by the magnetic field required by Maxwell's Equations, constitutes electromagnetic radiation.

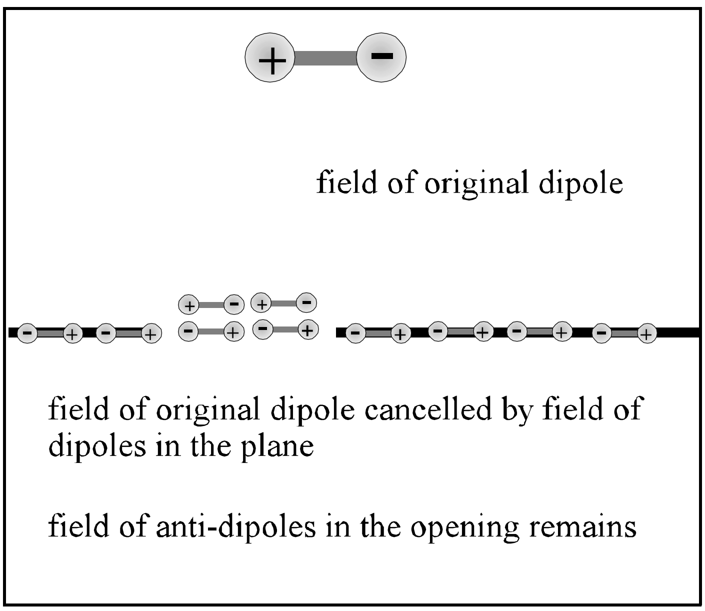


**Figure 9** An oscillating dipole near a conducting plane.

Now we place the oscillating dipole above a perfectly conducting plane. The dipole will cause the electrons in the plane to oscillate. The positive ions in the plane will, however, remain fixed. A fixed positive ion with an electron oscillating around it is effectively the same as an oscillating dipole in the plane. We thus can describe the effects of the electrons oscillating in the plane as a collection of oscillating dipole sources. As in the case of a static electric field, the electric and magnetic fields inside the plane and on the opposite side of the plane are zero.

We then drill a hole in the plane. This is the same as adding to the plane a large number of “anti-dipoles” oscillating opposite to the plane’s dipoles in the region of the hole. Since the net charge on the dipoles is zero, and the dipoles are rapidly oscillating, the addition of these anti-dipoles has little effect on the conducting plane. The field below the plane was zero before the anti-dipoles were added. After they are added, the field becomes the field of the anti-dipoles alone. If the oscillations are in the visible range of the spectrum, we can say that the net result in this region is the same as if light were emitted by sources located at the hole.

Note that we could have chosen a perfectly conducting surface of any shape surrounding the original dipole (or going off to infinity as in the case of the plane) and arrived at a similar conclusion. Since the electric field outside the conductor is zero, the dipoles in the conductor must produce a field that exactly cancels the field of the original dipole within the conductor and outside the conductor (on the side opposite the source). Now let us try a thought experiment: we remove the original dipole but somehow keep the dipoles in the conductor oscillating exactly the same way as with the original dipole present. On the outside of the conductor is a field identical to that of the original dipole, only opposite in sign. If we want to go a step further, we can change the signs of the charges on each dipole in the conductor to make a field that is identical to original field. In terms of light, a thin layer of conducting material can cause the cancellation of the electric field and stop light from passing through it. This leads to another conclusion related to Huygens’ Principle:

> Surround a light source by an arbitrary surface. The light outside the surface can be described by a suitably chosen collection of sources on the surface.

Let’s look at one consequence of this conclusion. If the conductor actually produces a field that cancels the original source, it must really be emitting light out of phase with source, but otherwise identical to it. This implies that the conductor must also emit light on the same side as the original source. In fact, if the conductor were a flat plane, the light emitted on the source side of the conductor would look exactly as if the source were on the opposite side of the conductor. We, of course, know this process as reflection.

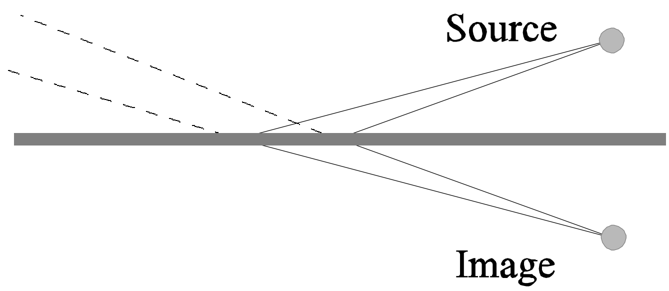

**Figure 10** The reflection of a source from a conducting plane.

We used a conducting plane to derive this result. We did this because we needed the field of the original dipole to be canceled on the far side of the plane by the field of the oscillating electrons in the plane. If the electrons in the plane were not free to oscillate, then this result would not be valid. However, for electromagnetic radiation in the visible range, we know that the electromagnetic field goes to zero on the far side of any opaque plane, whether or not it is a conductor. We have to consider quantum mechanical effects in explaining the details of light absorption, but if the plane is opaque, we can argue that the situation is equivalent to light passing through a conducting plane. This leads to the important conclusion that is usually called Huygens' Principle:

> Light passing through an opening in a plane is equivalent to the light emanating from a collection of point sources located at the opening.

In order to apply this principle to wave optics, we need to consider one last thing: the finite propagation time of electromagnetic fields. If a charged sphere is stationary, it produces an electric field throughout all of space. If we wiggle the sphere, then the field will also "wiggle." However, a charged object at great distance from the sphere is not instantaneously affected by the wiggle; if it were, we could send signals at an infinite speed. It turns out that the change in the electric field propagates outward from the sphere at the speed of light. This is equivalent to saying that electromagnetic radiation travels at the speed of light. Thus, a pulse produced by a source at point S and time $t = 0$, will arrive at a point P over a range of time rather than a single instant. Conversely, the signal arriving at point P at a time $t$ was produced at the source over a range of times rather than at one instant. Applying this same logic to a dipole oscillation, it is evident that light arriving at point P at a time $t$ will have been produced by the source at different phases of oscillation depending on the path of the light through the aperture. This means that one part the light emitted from the source will interfere with other parts of the light. The intensity of the light will be strongly dependent on our choice of P.

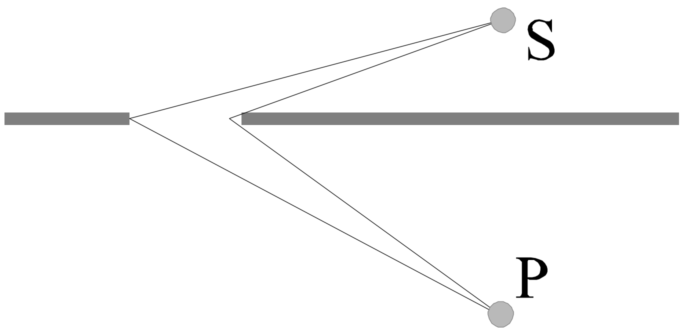


**Figure 12** Two paths of different length from a source, S, to a receiver, P.

Fortunately, these effects can be readily calculated. Let us set up a problem as shown in Figure 12. Light from a source, S, is normally incident on an aperture of area $\Delta A$. The intensity of the light is measured at a point P which is coplanar with S and the center of the aperture. For simplicity, we can consider the dipole to be oscillating in and out of the page at point S. In this case the electric field is in the usual form for

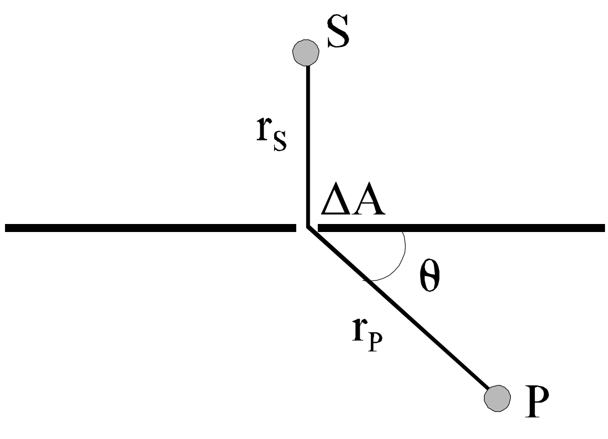


**Figure 11** Geometry for the diffraction integral.

a wave propagating from a point source in three dimensions:

$$E_A = E_0 \frac{\sin(kr_S - \omega t)}{r_S}.$$

The field at point P can be considered to arise from "anti-dipoles" oscillating at the aperture. The strength of the reradiation from the aperture is proportional to the amplitude of $E_A$, which is just $E_0 / r_S$. We then can consider the wave emitted from these anti-dipoles arriving at P to be of the form:

$$E_P = f \frac{E_0}{r_S} \frac{\sin(kr_S + kr_P - \omega t)}{r_P} \Delta A.$$

where $f$ is a constant. The area of the aperture, $\Delta A$, is included because the amplitude of the reradiated wave is the sum of the amplitudes from each dipole, and the total number of dipoles in the aperture is proportional to the area of the aperture.

Finally, if we have a larger hole, we can add up the sum of the electric fields from each small element of aperture, $dA$. This gives for time $t=0$ :

$$E_P(\theta) = fE_0 \int \frac{\sin(kr_S + kr_P)}{r_P r_S} dA.$$

We must remember that $r_S$ and $r_P$ are functions of $\theta$.

This integral is called the Huygens Diffraction Integral. In this course, we will take cases where $r_S$ and $r_P$ are approximately the same for all values of $\theta$, so that all we have to do is the integral over the sine function, and this integral we can evaluate geometrically using phasors. But it is important to conceptually understand Huygens' Principle and recognize that there is a reasonable physical basis for its application. When you take a course in optics, you will learn to deal with diffraction problems using fewer approximations than we have done here, but the basic conceptual idea is still valid and may well be helpful to you in the future.