Chapter 3
Scatter Calculations and Diffraction Theory

"In theory there is no difference between theory and practice. In practice there is."
– Yogi Berra

This chapter outlines the important elements of diffraction theory and gives several key results that pertain to the interpretation of measured scatter data. These results are employed in Chapters 4 and 8 to relate measured scatter from reflective surfaces to the corresponding surface roughness and to consider various methods of scatter prediction. In Chapter 9, the diffraction theory results presented here are combined with the polarization concepts found in Chapter 5 and used to outline a technique for separating surface scatter from that due to subsurface defects and contamination. A complete development of diffraction theory is well beyond the scope of this book; however, excellent texts on the subject are available, and these will be referenced in the basic review presented in the next three sections. Some relatively new diffraction results are presented in Sections 3.4–3.6. The following discussions assume that the reader has some familiarity with EM field theory and the required complex math notation. Appendix A is a brief review of the elements of field theory, and Appendix B gives details of some diffraction calculations.

3.1 Overview

When light from a point source passes through an aperture or past an edge, it expands slightly into the shadowed region. The result is that the shadow borders appear fuzzy instead of well defined. The effect is different from the one obtained by illuminating an object with an extended light source (such as the shadow of your head on this book) where the width of the reading lamp also contributes to an indistinct shadow. Well-collimated light sources (sunlight, for example) also produce fuzzy shadow edges. This bending effect, which illustrates the failure of light to travel in exactly straight lines, is called diffraction and is analyzed through the wave description of light.

As explained in Appendix A, the propagation of light is described in terms of the transverse electric field $E(t, r)$, where r denotes position, and t is time. The value k is $2\pi/\lambda$, and v is the light frequency. The expression in Eq. (3.1) is for a wave

traveling in the direction of increasing r:

$$E(t, r) = \text{Re}[e(r)e^{j(kr-2\pi\nu t)}].\tag{3.1}$$

$$E(r) = e(r)e^{jkr}.\tag{3.2}$$

Phasor notation is used (the "real part" is understood), and the dependence on time, which will appear in all terms, is dropped for convenience as indicated in Eq. (3.2). The term $e(r)$ gives spatial dependence. Quantities shown in **bold** are vectors, indicating that they denote the polarization direction. Three common cases given below are for a plane wave traveling from $r = 0$, a spherical wave diverging from $r = 0$, and a spherical wave converging to $r = 0$. The value $\mathbf{E}_0$ is a constant in space and time. The power of the converging and diverging waves, which is proportional to $1/r^2$, follows the expected inverse square law:

$$E(r) = E_0 e^{jkr} \quad \text{Plane wave.}\tag{3.3}$$

$$E(r) = \frac{E_0}{r}e^{jkr} \quad \text{Diverging.}\tag{3.4}$$

$$E(r) = \frac{E_0}{r}e^{-jkr} \quad \text{Converging.}\tag{3.5}$$

An infinitely wide plane wave can be thought of as being made up of an infinite number of spherical waves. Imagine the spherical waves originating at each point along a constant phase plane of the plane wave. Superimposing the spherical waves at some distance results in equal forward (Z direction) contributions and equal but opposite contributions in the XY directions. Thus, at each point, the spherical contributions sum to a forward-propagating wave with equal amplitude and phase—in other words, a plane wave. If the original wave has amplitude or phase variations, then the result will not be a plane wave, but summing the spherical components will give the new wavefront. This is the essence of the calculations in the next few paragraphs; variations are the result of approximations made to simplify the mathematics.

Figure 3.1 shows the diffraction geometry for light transmitted through an aperture in the x, y plane. The aperture, centered at $r = 0$, is typically illuminated by a point source (diverging), a collimated beam (plane wave), or a converging beam (virtual point source). In general, the aperture modulates the transmitted light in both amplitude and phase. The modulated light leaving the aperture is given by $\mathbf{E}(x, y)$, and the object of the diffraction calculation is to find the resulting electric field $\mathbf{E}(x_s, y_s)$ in the observation plane, located a distance R from the aperture. The source could also be located on the $z > 0$ side of a reflective aperture (or sample).

Amplitude modulations, caused by changes in aperture reflectance or transmittance, are expressed by variations in $\mathbf{e}(x, y)$. For example, a slit aperture changes from zero transmittance to unity and back again with no phase modulation. Phase modulations are caused by index of refraction changes in transmitting
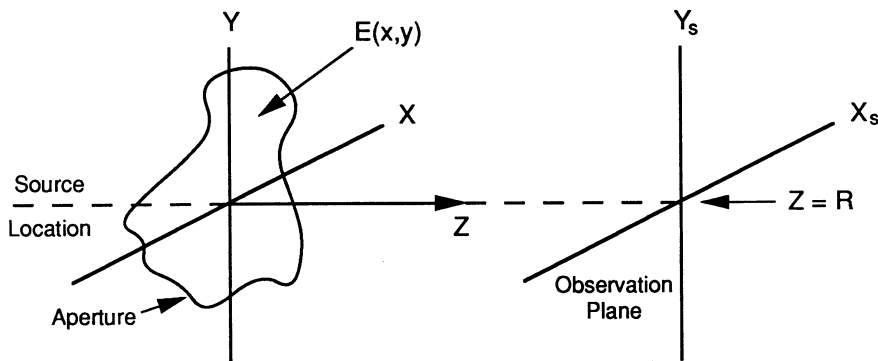
Figure 3.1 Geometry for diffraction from an aperture in the x, y plane to the x_s, x_y plane.

samples and by surface roughness on reflecting samples and are expressed by changes in the exponential component of $\mathbf{E}(r)$.

A useful exercise is the calculation of diffracted light from a slit aperture without the benefit of using a diffraction theory result. The general nature of the solution, the approximations required, and the limitations of such an approach become immediately obvious. Consider a slit aperture of width L to be centered on the x, y plane along the y axis, as shown in Fig. 3.2. A plane wave traveling along the z axis is incident upon the aperture, and diffraction is to be observed at the x_s, y_s plane located at $z = R$. The assumption is made that $R \gg L$. Use is made of the Huygens principle, which is an intuitive statement that wavefronts can be constructed by allowing each point in a field to radiate as a spherical source. The new wavefront, downstream, is then found from the envelope of the spherical fronts. Early diffraction results depended on variations of this reasoning, even though there are some obvious problems. For example, what do we do about the backward-traveling wave? Polarization issues are ignored. We will also ignore any field–aperture interaction, and assume that the field exists as presented by the source right up to the aperture edge, where it drops to zero in a sudden discontinuity. This is a true "back-of-the-envelope" calculation, whose purpose is to develop insight for the more complicated issues to follow.

Two rays leaving from $x = 0$ and $x = L/2$ and eventually interfering at coordinate C_s on the x_s axis are shown in the diagram. The path difference of the two waves is the small distance h shown in the figure. Making use of the small-angle assumptions gives

$$h = Lx_s/2R. \tag{3.6}$$

At some value of x_s, h will reach the value $\lambda/2$, and the two waves will cancel in the observation plane. Within the limitations of the small-angle assumption, the same reasoning holds for all the other pairs of rays separated by $L/2$ at the aperture and reaching point C_s. Thus, the condition

$$x_s = \pm n\lambda R/L \tag{3.7}$$
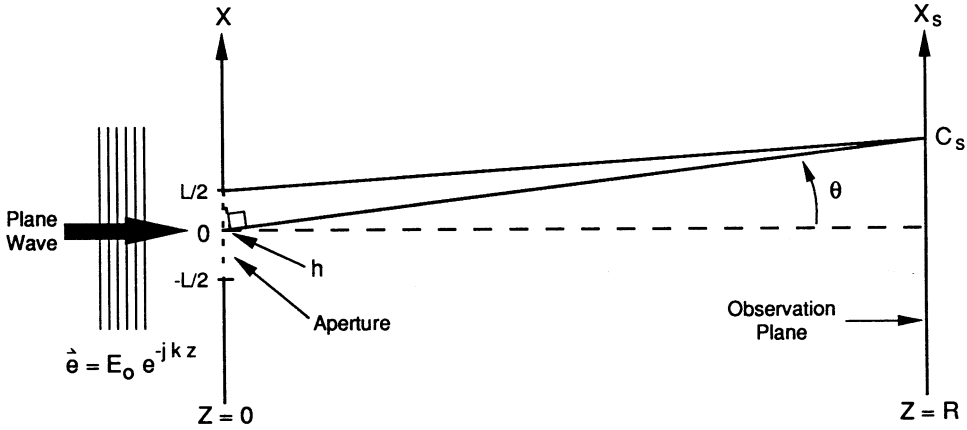
Figure 3.2 Plane wave diffraction from a slit.

(where n is an integer) will result in a zero-intensity value on the otherwise illuminated x_s axis. The relative intensity pattern can also be found. Ignoring polarization issues, the spherically expanding wave from a differential source dE over dx, located at x in the aperture, will have an amplitude proportional to dx and inversely proportional to the distance from x. The resulting differential scalar amplitude from the differential source may be evaluated in the observation plane, where K has been used as a proportionality constant:

$$dE_s = \frac{Kdx}{\sqrt{R^2 + (x_s - x)^2}} e^{jk\sqrt{R^2+(x_s-x)^2}}.\tag{3.8}$$

The approach is to integrate over x from $-L/2$ to $L/2$, and thus obtain the total field strength at C_s. In order to perform the integral easily, some assumptions are made to simplify the expression for the distance r_s between x and C_s. In the amplitude component, the distance is approximated as R; however, in the phase component, this is inappropriate, as distance errors of only half a wavelength change the sign with which a particular component is summed. For the phase term, the radical can be expanded as

$$
\begin{aligned}
\sqrt{R^2 + (x_s - x)^2} &= R\left[1 + \frac{(x_s - x)^2}{R^2}\right]^{1/2} \\
&= R\left[1 + \frac{(x_s - x)^2}{2R^2} - \frac{(x_s - x)^4}{8R^4} + \cdots\right] \\
&= R + \frac{x_s^2}{2R} - \frac{x_s x}{R} + \frac{x^2}{2R} - \frac{x_s^4}{8R^3} \cdots
\end{aligned}\tag{3.9}
$$

$$\underbrace{\qquad\qquad}_{\text{Fraunhofer}}$$

$$\underbrace{\qquad\qquad\qquad}_{\text{Fresnel}}$$

Each additional term makes the integral more accurate and more difficult to evaluate. The two common approximations have been named after the men who made them, as indicated. The approximations are better for large R. This has led to the terminology *getting to the far field*, which usually implies that the Fraunhofer approximation is accurate enough to predict experimental results. If a source is used that converges at the observation plane, then a term is introduced that cancels the $x^2/2R$ term, making the Fraunhofer and Fresnel approximations identical. For the example at hand, we will proceed with the Fraunhofer approximation and evaluate the integral as follows:

$$E_s = \frac{KL}{R} e^{jk(R+x_s^2/2R)} \int_{-L/2}^{L/2} e^{-j(2\pi x_s x/\lambda R)} dx, \qquad (3.10)$$

$$E_s = \frac{KL}{jR} e^{jk(R+x_s^2/2R)} \operatorname{sinc}\left(\frac{x_s L}{\lambda R}\right), \qquad (3.11)$$

where

$$\operatorname{sinc}(\alpha) = \frac{\sin(\pi\alpha)}{\pi\alpha} \quad \text{for any real argument } \alpha.$$

Squaring the absolute value of the electric field and dividing by twice the impedance of free space η_0 gives the time-average power density I_s (watts per unit area) as a function of x_s:

$$I_s = \frac{1}{2\eta_0}\left[\frac{KL}{R}\right]^2 \operatorname{sinc}^2\left(\frac{x_s L}{\lambda R}\right). \qquad (3.12)$$

This relationship is plotted in Fig. 3.3. Notice that the zero intensity values appear at the locations predicted earlier by Eq. (3.7). This means that the Fraunhofer approximation is equivalent to the same small-angle approximation. Patterns very much like the one in Fig. 3.3 can be observed by placing a small slit in a HeNe laser beam. The inverse aperture, a small block, is easier to do. A piece of hair works just fine. Using the above relationships and measurements of the diffraction pattern made with a ruler allows the hair diameter to be calculated (and makes a great classroom demonstration). The proportionality constant K has not been evaluated, but this can be accomplished by integrating over the observation plane and applying the conservation of energy.

Another observation is worth making: the sinc function is the Fourier transform of the slit aperture [sometimes expressed as rect(x/L)]. In fact, Eq. (3.10) shows this explicitly. In this context, the quantity $(x_s/\lambda R)$ may be viewed as a spatial frequency propagating in the x direction in the aperture plane. Notice that it has units of inverse length, as required in Chapter 1. In fact, this is the same expression for spatial frequency that is obtained from the grating equation at normal incidence and small angles. Retracing our steps back through the development, it is easy to see that if an aperture function other than unity had been applied, the Fraunhofer