3.0 PROBLEMS IN OPTICS

Much of the work done by the Thermal Radiation Group involves the development of radiative models of radiometric instruments. The Monte-Carlo ray-trace (MCRT) method is generally used, but the design objectives vary with the measurements being sought. For instance, the design of a solar aureolemeter used to measure the concentration and size distribution of atmospheric aerosols is a continuing topic of investigation in the Thermal Radiation Group. This is an imaging instrument that measures the circumsolar radiation pattern produced by atmospheric aerosols. These measurements are important because the scattering produced by aerosols in the region around the sun is much greater than that due to other atmospheric constituents. The image quality is of principle concern in the design of this instrument [Deepak and Adams 1983; Nakajima, *et al.*, 1983]. On the other hand, the CERES instrument is not an imaging instrument, but a radiometer. Energy throughput, rather than the image quality, is the principle concern in the design of this instrument. As described by Walkup [1996], the area in which the two disciplines of imaging and radiometry merge is imaging radiometry, illustrated in Figure 3.1. This best describes the type of investigations conducted within the Thermal Radiation Group.

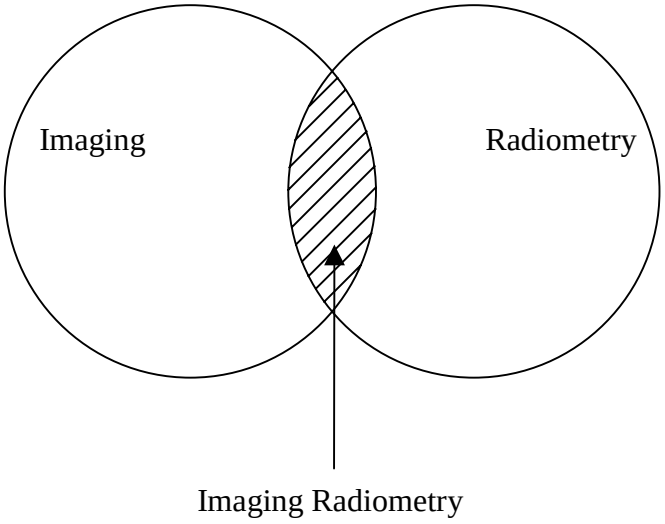


Figure 3.1 Combination of imaging and radiometry disciplines, (borrowed from Walkup [1996]).

3.1 Ideal versus real behavior in optical systems

Much of the work presented in this thesis in some way pertains to the deviation of the true performance of instruments from simplified models of their ideal behavior. The ideal behavior of optical systems is often called Gaussian optics. This first-order theory is used for preliminary design specifications, and can be used for mirror placement and for locating the ideal image point in an optical system. Gaussian optics describes a point-to-point, object-to-image relationship. That is, when rays leave a point within the object plane, they will arrive at a single point in the image plane. True optical performance of instruments deviates from this idealized behavior. The degree of this deviation can be described by five different aberrations: spherical aberration, coma, astigmatism, field curvature, and distortion. For the purposes of the current effort, it is important to describe spherical aberration in an optical system. For a detailed description of the other aberrations, refer to Walkup [1996].

3.1.1 Spherical aberration

Spherical aberration is the only aberration that occurs on-axis. Because of this aberration, rays reflected from conic reflectors do not converge to the ideal infinitely

small spot size as they cross the optical axis at the Gaussian image plane. Instead, a blur circle of finite size is formed at the focal plane. Rays that traverse the telescope near the optical axis behave ideally, passing through the Gaussian image point, while those reflecting far from the optical axis do not. The point at which this blur circle is a minimum is called the "circle of least confusion," as illustrated in Figure 3.2. It is important in the design and assembly of an imaging telescope that this circle of least confusion fall in the detector plane. For a non-imaging telescope such as CERES, the telescope is assembled such that the circle of least confusion is located at the plane containing the precision aperture that serves to shape the image cast on the detector beneath.

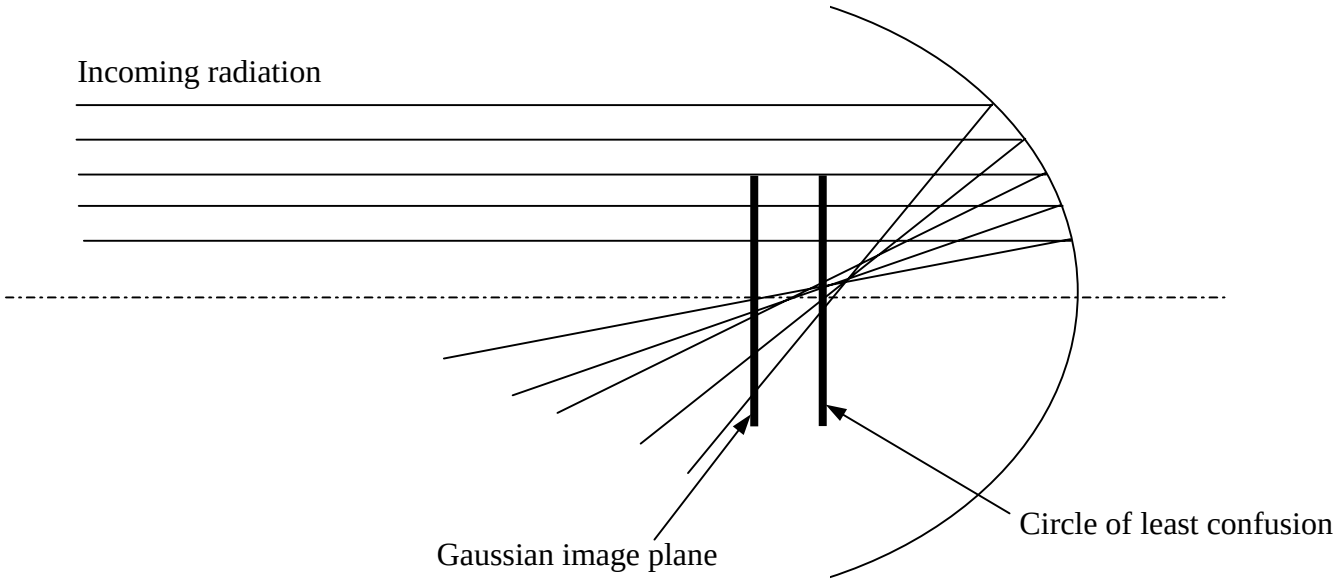


Figure 3.2 Illustration of spherical aberration (borrowed from Walkup [1996]).

3.2 Ideal versus real propagation of electromagnetic waves

The idealized behavior of the propagation of electromagnetic energy is sometimes referred to as the ideal ray approximation. This model ignores diffraction, and rays are modeled as straight lines that are perpendicular to the propagating wavefronts. As electromagnetic waves enter apertures or pass by obstacles, the ideal ray approximation does not account for diffraction effects that are known to occur. The correction of this idealization in instrument modeling is a major topic of this thesis. The following discussion of diffraction serves to provide the necessary background for understanding the work presented in Chapter 4.0.

3.2.1 Diffraction

3.2.1.1 Basic ideas

Diffraction and interference are actually one and the same: the constructive and destructive interference of in-phase or out-of-phase waves arriving at the same point. The convention has developed to call this additive wave property *interference* when it deals with only a few waves, and to call it *diffraction* when it involves many waves [Hecht and Zajac, 1974]. Diffraction occurs when electromagnetic waves pass through apertures or past obstacles, causing the waves to deviate from their original lines of travel. The degree to which this divergence occurs depends upon the wavelength of the entering radiation relative to the aperture dimensions. For the case where the wavelength of the approaching wave is much less than the aperture dimensions, diffraction effects are slight. In this case, the ideal ray approximation is valid. The resulting intensity on an observation screen placed behind the aperture is in the shape of a top hat, with no light arriving outside the area which is formed by the projection of the aperture area on the screen. This case is illustrated in Figure 3.3 (a). When the wavelength of the light is on the order of the aperture size, diffraction effects are more pronounced as shown in Figure 3.3 (b). If the entering energy is monochromatic, a pattern of alternating minima and maxima will form on an observation screen placed behind the aperture. Energy will spread out to areas that would be predicted to be in shadow by the ideal ray approximation. The shape and span of the resulting diffraction pattern depends on the aperture size, the wavelength of the entering energy, and the distance between the

observation screen and the aperture. Finally, when the wavelength of the approaching energy is much greater than the aperture dimensions, the aperture behaves as a point source that emits spherical waves, (i.e. a slit behaves as a line of point sources) as illustrated in Figure 3.3 (c). In this case, the energy will spread out in the region where a shadow would be predicted, but no pattern of maxima and minima will be observed.

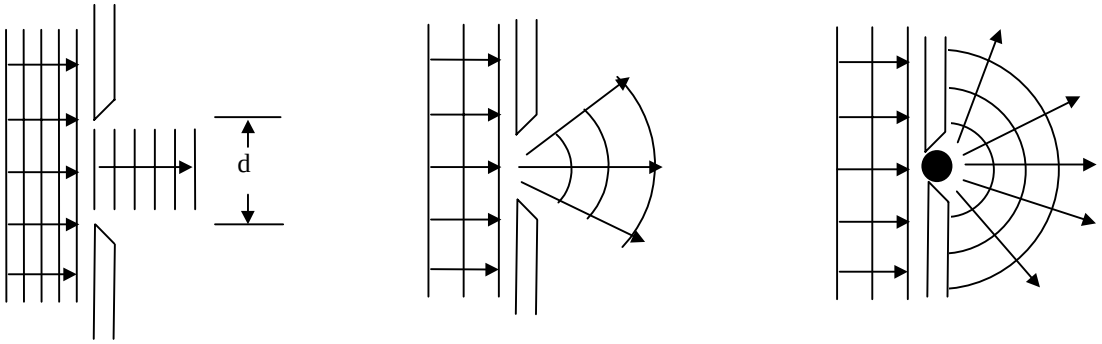


Figure 3.3 The behavior of radiation as it passes through an infinite slit: (a) the wavelength of entering radiation is much less than the slit width, (b) the wavelength is approximately equal to the slit width, and (c) the wavelength is much greater than the slit width (borrowed from Serway [1994]).

The diffraction models that are presented in Chapter 4.0 are used to simulate the behavior of monochromatic, coherent radiant energy as it passes through an aperture, and the results are compared to closed-form analytical solutions of the expected interference pattern for validation. Upon validation, these models will then be used to simulate diffraction for cases in which the approaching energy is neither coherent nor monochromatic. Such is the case for applications such as instruments used to measure the Earth's radiation energy budget.

3.2.1.2 The two diffraction regimes

Diffraction can be classified into two regimes: Fresnel, or near-field, and Fraunhofer, or far-field, diffraction. Fraunhofer diffraction occurs when the light source and the

observation screen upon which the diffraction pattern is observed are both effectively at infinite distances from the aperture causing the diffraction. Fresnel diffraction occurs when either (or both) the source or the screen are at finite distances from the aperture. In the development of a radiative model of an instrument, one must determine how crucial the modeling of diffraction is. The first step in this assessment is to determine which of the two diffraction regimes governs the diffraction. For the case in which energy is entering through a rectangular slit, and the source is at infinity, the parameter

$$\Delta\xi = a\left(\frac{2}{\lambda z}\right)^{\frac{1}{2}}, \tag{3.1}$$

where a is the slit width, z is the distance between the aperture and the observation screen, and λ is the wavelength of the entering energy, is used for the purpose of categorizing diffraction as Fraunhofer or Fresnel [Wyatt, 1987; Haskell, 1970]. $\Delta\xi$ =1 is taken to be the transitional configuration between the two regimes. For cases in which $\Delta\xi$>1, Fresnel diffraction prevails, while for cases in which $\Delta\xi$<1, Fraunhofer diffraction prevails. Figure 3.4 illustrates this transition, brought on by moving the observation screen away from the aperture. In this case, the point source is at infinity with respect to the slit. As the observation screen is moved away from the aperture, the rays become approximately parallel upon their arrival at the observation screen, meeting the criterion for Fraunhofer diffraction. Note that a similar transition would occur if the wavelength of the entering radiation were increased while holding the parameters z and a constant. Such is the case in the practical application presented in Chapter 4.0.

The diffraction patterns shown in Figure 3.4 are based on the closed-form analytical solution reported by Haskell [1970]. The intensity along the screen is given in terms of the Fresnel integrals

$$C(u) = \int_0^u \cos\left[\frac{1}{2}\pi u^2\right] du, \tag{3.2 a}$$

and

$$S(u) = \int_0^u \sin\left[\frac{1}{2}\pi u^2\right] du, \tag{3.2 b}$$

and the parameters

$$\xi_1 = \frac{1}{2}\Delta\xi\left[\left(\xi_o / \Delta\xi\right) - 1\right], \tag{3.3}$$

$$\xi_2 = \frac{1}{2}\Delta\xi\left[\left(\xi_o / \Delta\xi\right) + 1\right], \tag{3.4}$$

where $\Delta\xi$ is as defined in equation 3.1 and

$$\xi_o = -2x\left(\frac{2}{\lambda z}\right)^{\frac{1}{2}}. \tag{3.5}$$

Thus the intensity along an observation screen, as illustrated in Figure 3.4 is given by

$$I(x) = \frac{1}{2}\left[C(\xi_2) - C(\xi_1)\right]^2 + \frac{1}{2}\left[S(\xi_2) - S(\xi_1)\right]^2. \tag{3.6}$$

Equation 3.6 can be used to determine diffraction patterns in either the Fresnel or Fraunhofer regime. However, Fraunhofer diffraction can be described by simplified equations found in Laufer [1996].

Monochromatic, coherent radiation of wavelength, $\lambda = 0.001$ μm approaching infinite slit of width, a = 0.1 μm from a source at infinity.

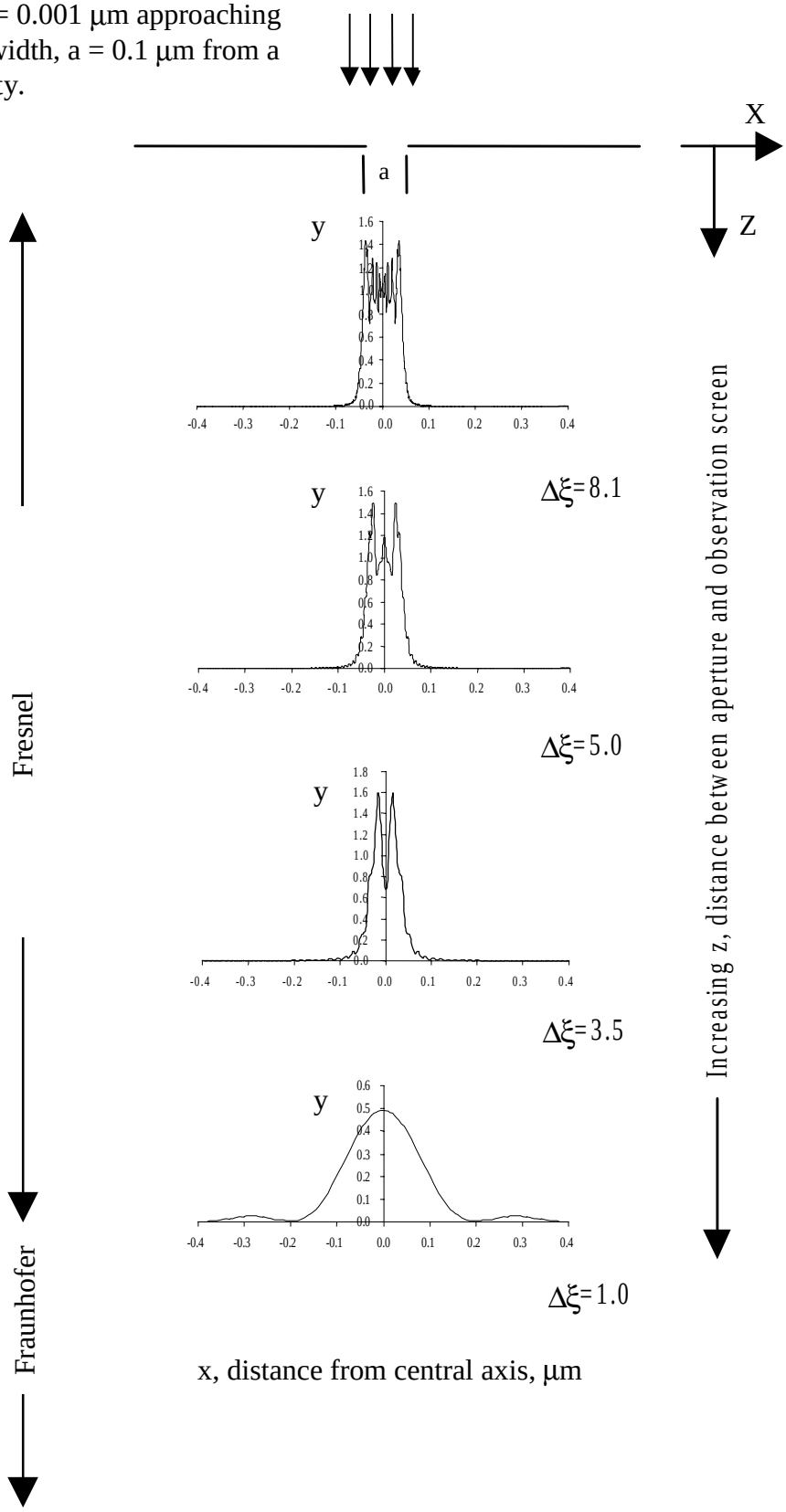
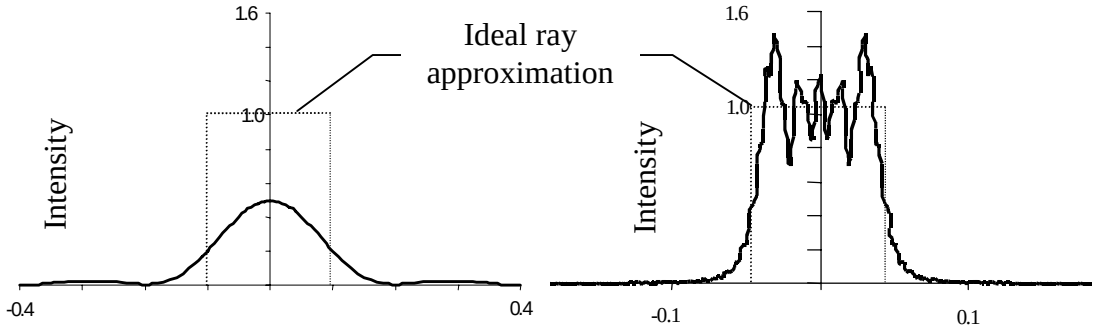


Figure 3.4 Illustration of the transition from Fresnel to Fraunhofer diffraction and the resulting change in the intensity distribution, I(x) where y = I(x)

Table 3.1 summarizes the characteristics of the two diffraction regimes. The modeling capabilities mentioned refer to the models that are presented in Chapter 4.0.

Table 3.1 Comparing characteristics of Fraunhofer and Fresnel diffraction.

Fraunhofer	**Fresnel**
♦ Far-field	♦ Near-field
♦ Simpler analytical solutions.	♦ Complex analytical solutions.
♦ Instrument throughput not good.	♦ Instrument throughput good.
♦ Possibility of modeling phase using new ray-trace approach (Model 2) and MCRT method.	♦ Cannot model phase using MCRT method.
♦ Good modeling is more important, ideal ray trace deviates greatly from actual behavior (Figure 3.5 (a)).	♦ Good modeling is not always as important, ideal ray trace approximates the true behavior well in very near field (Figure 3.5 (b)).



(a) (b)

Figure 3.5 Illustration of the deviation of the ideal ray approximation from the true intensity pattern resulting from the passage of energy through a slit 0.1 μm wide (a) for a Fraunhofer diffraction situation, and (b) for a Fresnel diffraction situation.

Figures 3.5 (a) and (b) illustrate the importance of modeling diffraction well in the Fraunhofer regime, and the reduced importance of modeling diffraction well in the

Fresnel regime. Because energy is distributed far into the geometrical shadow in Fraunhofer cases, much of the on-axis energy is dispersed out of the optical path. For this reason, Fraunhofer diffraction is generally avoided in imaging systems. Ideally, an effort should be made to ensure that diffraction occurs in the Fresnel region in the early design stages of an instrument [Wyatt, 1987].

3.2.1.3 A brief history of diffraction

The diffraction of electromagnetic waves has puzzled and challenged scientists, engineers and physicists for more than five centuries, and some of these challenges remain even today. The authors of most optics and physics books state that diffraction is one of the most complicated problems to be dealt with in optics. Diffraction is of academic importance, as it serves as evidence of the wave nature of light. It is of practical significance to engineers and scientists as it limits the resolving power of optical devices, and often results in less than ideal instrument throughput.

Challenges posed by the phenomenon of diffraction are interwoven throughout the history of optical science. Every generation of scientists since the 1400s has made a contribution to further our understanding of diffraction. The following briefly summarizes the stages of diffraction knowledge, borrowing from the historical summary by Hecht and Zajac [1974], and other authors. Throughout this historical summary, many important principles are described which serve as background for the diffraction models presented in Chapter 4.0.

The first reference to the phenomena of diffraction was made in the work of Leonardo Da Vinci (1452-1519). It was not until 1665 that diffraction was accurately described by Professor Francesco Maria Grimaldi in his observations of bands of light within the shadow of a rod illuminated by a small source [Born, 1970]. Soon after, Robert Hooke (1635-1703) performed experiments in which he also observed diffraction patterns. These observations of diffraction prompted the beginning of the wave theory of light, and the great debate as to whether light is corpuscular (stream of particles) or wave-like. Christian Huygens (1629-1695), a proponent of the wave theory, continued to extend the

wave theory of light. One of his significant contributions, Huygens' principle which states that "every point on a primary wavefront serves as the source of spherical secondary wavelets such that the primary wavefront at some later time is the envelope of these wavelets" is illustrated in Figure 3.6 (a). This model, which treats the secondary wavelets as radiating equally in all directions, and the envelope of the wavelets as the intensity at an observation screen behind the aperture, is unable to predict the expected interference pattern due to the diffraction of energy. It also predicts a backward propagating wave, which is not observed in reality. Thomas Young (1773-1829) was the next important contributor to the wave theory of light despite ridicule from many of his colleagues who supported the corpuscular theory. Young provided the first clear demonstration of the wave nature of light in 1801. His double slit experiment showed that light coming from a single source, arriving at a point by two separate paths, combines both constructively and destructively, resulting in an interference pattern on an observation screen placed behind the two slits, as illustrated in Figure 3.6 (b).

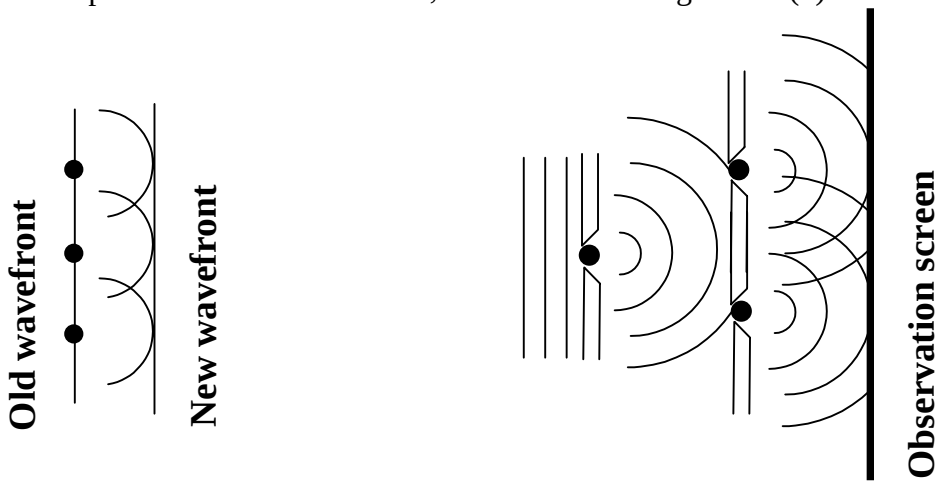


Figure 3.6 (a) Illustration of Huygens' principle for plane waves propagating to the right (excluding the backwards-propagating wave), and (b) illustration of Young's double-slit experiment.

Augustin Jean Fresnel (1788-1827) added to Huygens' ideas in order to provide a model which could predict the intensity variations expected by diffracted energy. He added an undefined angular dependence factor, K, called the obliquity factor, which serves to model the directionality of the emission of secondary wavelets coming from a point

source. The Huygens-Fresnel principle states that "every unobstructed point of a wavefront, at a given instant in time, serves as a source of secondary wavelets of the same frequency as the primary wave. The amplitude of the optical field at any point beyond is the superposition of all of these wavelets (considering their amplitudes and relative phases)". Fresnel began to describe diffraction on a mathematical basis and was able to predict diffraction patterns of light when it passed by obstacles. Kirchoff was the first to put the conceptual Huygens-Fresnel principle on a more mathematical basis, proposing a specific form of the obliquity factor, K which is used in Chapter 4.0. Instead of conceptual arguments to describe diffraction phenomena, mathematical descriptions began to be sought and used more extensively. The determination of an exact solution for a particular diffracting configuration is amongst the most difficult to be dealt with in optics [Hecht and Zajac, 1974]. The first such solution was published in 1896 by Arnold Sommerfeld (1868-1951). This solution utilized the electromagnetic theory of light, and involved an infinitely thin, opaque, perfectly conducting plane screen. Note that rigorous solutions such as the one published by Sommerfeld exist for very few configurations of practical interest today. Because of this, numerical methods based on conceptual diffraction models and other approximations are often used to determine the effect of diffraction on configurations of interest.

3.2.1.4 Modern methods of dealing with diffraction

A number of methods for modeling and/or characterizing diffraction in an optical instrument have been developed in this century, taking advantage of advancements in computing power. Some of these methods are described briefly in the following paragraphs, and references are cited which provide more extensive details.

Airy disk construction

Circular baffles, apertures, or obscurations are quite common in optical systems. For such configurations, the importance of diffraction in their optical performance can be approximated by a simple calculation. The size of the diffraction blur or Airy disk, named in honor of British mathematician, Lord George Biddell Airy (1801-1892), is given by

$$D_1 = 2.44 \lambda f , \tag{3.7}$$

where D_1 is the diameter of the diffraction blur, λ is the wavelength of the entering energy, and f is the focal length of the optical system divided by the aperture diameter [Riedl, 1997]. For the case of energy entering through a circular aperture, D_1 defines the area on the observation screen containing 84 percent of the incident energy, assuming no other aberrations are present. The diameter, D_2, which defines the area that will contain 91 percent of the incident energy is given by

$$D_2 = 4.48 \lambda f . \tag{3.8}$$

It is important to note in instrument design that there is no point in decreasing the diffraction blur size below the size of the blur circle due to the geometrical aberrations present in the optical system. Likewise, an optical system can never perform better than the limits that are predicted by diffraction, thus optical designers may reach a point where further minimization of geometrical aberrations would be in vain. Another important situation, a circular aperture with a central obscuration, is commonly found in telescopes such as CERES. Details of the treatment of this situation are provided by Riedl [1997] and by Walker [1998]. Note that the addition of a circular obscuration in the center of a circular aperture causes the incident energy to become more dispersed, thus the Airy disk diameters are larger. The degree of this dispersion of energy out of the desired optical path depends on the diameter of the obscuration relative to the diameter of the aperture, characterized by the percent obscuration. The larger the percent obscuration, the more the incident energy is dispersed. For this reason, systems with obscurations greater than 50 percent are rarely used [Walker, 1998]. For further discussions regarding the influence of diffraction on instrument imaging performance, the interested reader is referred to Braun [1970], Tschunko and Sheehan [1971], and Boivin [1977].

Method of moments

The method of moments can be used to approximate the diffraction pattern of energy entering an aperture. This method, based on the method of weighted residuals, is a common tool within the electrical engineering culture, and can be applied to treat the propagation of electromagnetic waves of any wavelength. For details of this method see Hubing [1991]. Graves *et al.* [1976] address the modeling of the electromagnetic field

penetration through a circular aperture in both the near and far field using this method. Results are shown to compare well with theory.

Geometrical theory of diffraction

The geometrical theory of diffraction provides a means of supplementing standard geometric optics with the ability to model diffraction. Keller [1957, 1962] and Balanis and Peters [1968] published descriptions and sample applications of this method, which treats diffraction as an edge effect, as originally proposed by Thomas Young [Keller, 1962]. In addition to the usual rays, the geometric theory of diffraction introduces new diffracted rays which originate as a ray strikes an edge or vertex. As an example of the application of this method, the treatment of radiation as it enters an aperture is described by Keller [1957, 1962]. Rays which strike an edge become diffracted rays. When the incident ray is normal to the edge, the diffracted rays produced are also normal to the edge (i.e. all lie in the same plane), and leave in all directions within that plane, as shown in Figure 3.7 (a). When waves approach the edge obliquely, the diffracted wave fronts are defined by cones emanating from the edge, as illustrated in Figure 3.7 (b). These obliquely diffracted rays follow the *law of edge diffraction*, which states that a diffracted ray and the corresponding incident ray make equal angles with the edge at the point of diffraction. The incident and diffracted rays lie on opposite sides of the plane normal to the edge at the point of diffraction [Keller, 1962]. All rays are assigned an initial phase upon their arrival to the aperture opening, and their final phase upon arrival to an observation screen is proportional to the optical length of a ray from this point. Keller [1957, 1962] and Balanis and Peters [1968] demonstrate good agreement when comparing results from the application of the geometrical theory of diffraction with the closed-form analytical solution of the expected diffraction pattern for several cases.

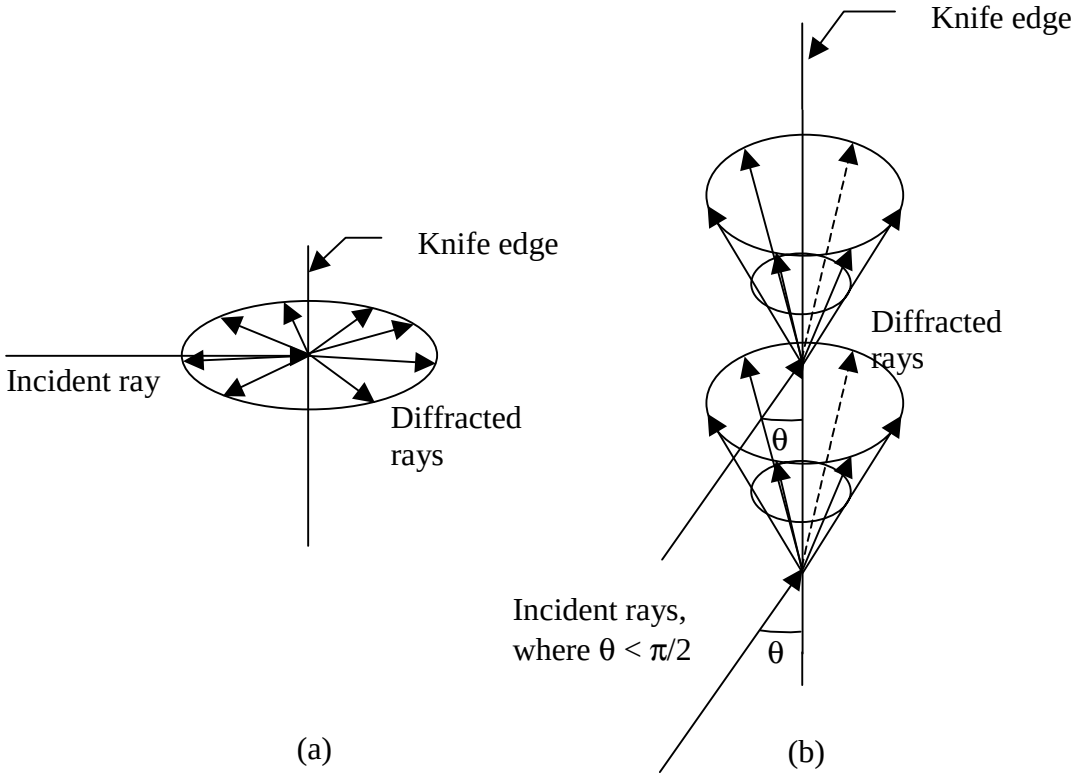


Figure 3.7 Illustration of the formation of diffracted rays, (a) when an incident ray strikes perpendicular to the edge, and (b) when incident rays strike the edge obliquely (borrowed from Keller [1962]).

The geometrical theory of diffraction was implemented in the GUERAP II code, a computer program for the analysis of the stray radiation rejection capabilities of optical systems, as briefly described by Boyce [1977]. A subsequent version of this code, GUERAP III, replaced the use of this methodology with a newer technique, referred to in this thesis as the statistical method for dealing with diffraction [Morbey and Hutchings, 1993].

Statistical method for dealing with diffraction

This method is based upon Heisenberg's uncertainty principle and the particle theory of light, and is convenient for use in the Monte-Carlo ray-trace environment [Chou, 1974]. This approach does not keep track of phase, so the resulting pattern approximates the minima and maxima which occur alongside the central maxima. Heinsch and Chou [1971] and Likeness [1977] briefly describe the concepts of this approach, omitting most

of the details necessary for its implementation. The details needed for the application have been rediscovered by the author of this thesis and are discussed at length in Chapter 4.0.

Use of Huygens' principle to model diffraction

The use of Huygens' principle to model an aperture as many point sources emitting isotropically in all directions has been implied by a technique described by Sinnott [1987]. The diffraction pattern resulting from the passage of starlight through a telescope is desired. The Monte-Carlo technique is applied, sending random parallel rays through the telescope. Ray path lengths are determined upon their arrival at an observation screen, and the net phase at a given point is given by the sum of the phase of all rays arriving at that point. Results are given as a two-dimensional scattergram, and are shown to compare qualitatively with analytical results (i.e. the spatial distribution of energy compares with the expected diffraction distribution). Results are not, however, quantitatively compared with theory (i.e. comparison of the relative magnitude of secondary maxima to central maxima is not made). Further discussion and modification of this technique are proposed in Chapter 4.0.