

Optical Engineering

OpticalEngineering.SPIEDigitalLibrary.org

Understanding diffraction grating behavior: including conical diffraction and Rayleigh anomalies from transmission gratings

James E. Harvey
Richard N. Pfisterer

SPIE.

James E. Harvey, Richard N. Pfisterer, "Understanding diffraction grating behavior: including conical diffraction and Rayleigh anomalies from transmission gratings," *Opt. Eng.* **58**(8), 087105 (2019), doi: 10.1117/1.OE.58.8.087105.

Understanding diffraction grating behavior: including conical diffraction and Rayleigh anomalies from transmission gratings

James E. Harvey* and Richard N. Pfisterer
Photon Engineering, LLC, Tucson, Arizona, United States

Abstract. With the wide-spread availability of rigorous electromagnetic (vector) analysis codes for describing the diffraction of electromagnetic waves by specific periodic grating structures, the insight and understanding of nonparaxial parametric diffraction grating behavior afforded by approximate methods (i.e., scalar diffraction theory) is being ignored in the education of most optical engineers today. Elementary diffraction grating behavior is reviewed, the importance of maintaining consistency in the sign convention for the planar diffraction grating equation is emphasized, and the advantages of discussing “conical” diffraction grating behavior in terms of the direction cosines of the incident and diffracted angles are demonstrated. Paraxial grating behavior for coarse gratings ($d \gg \lambda$) is then derived and displayed graphically for five elementary grating types: sinusoidal amplitude gratings, square-wave amplitude gratings, sinusoidal phase gratings, square-wave phase gratings, and classical blazed gratings. Paraxial diffraction efficiencies are calculated, tabulated, and compared for these five elementary grating types. Since much of the grating community erroneously believes that scalar diffraction theory is only valid in the paraxial regime, the recently developed linear systems formulation of nonparaxial scalar diffraction theory is briefly reviewed, then used to predict the nonparaxial behavior (for transverse electric polarization) of both the sinusoidal and the square-wave amplitude gratings when the $+1$ diffracted order is maintained in the Littrow condition. This nonparaxial behavior includes the well-known Rayleigh (Wood’s) anomaly effects that are usually thought to only be predicted by rigorous (vector) electromagnetic theory. © The Authors. Published by SPIE under a Creative Commons Attribution 4.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.OE.58.8.087105](https://doi.org/10.1117/1.OE.58.8.087105)]

Keywords: paraxial and nonparaxial diffraction grating behavior; “conical” diffraction in direction cosine space; Rayleigh (Wood’s) anomalies from sinusoidal and square-wave amplitude gratings.

Paper 190651 received May 14, 2019; accepted for publication Jul. 29, 2019; published online Aug. 28, 2019.

1 Introduction

The fundamental diffraction problem consists of two parts: (i) determining the effects of introducing the diffracting aperture (or grating) upon the field immediately behind the screen and (ii) determining how it affects the field downstream from the diffracting screen (i.e., what is the field immediately behind the grating and how does it propagate).

A “diffraction grating” is an optical element that imposes a “periodic” variation in the amplitude and/or phase of an incident electromagnetic wave.¹ It thus produces, through constructive interference, a number of discrete diffracted orders (or waves) which exhibit dispersion upon propagation. Diffraction gratings are thus widely used as dispersive elements in spectrographic instruments,^{2–5} although they can also be used as beam splitters or beam combiners in various laser devices or interferometers. Other applications include acousto-optic modulators or scanners.⁶

One example of a diffraction grating would be a periodic array of a large number of very narrow slits. This would be a binary amplitude grating (completely opaque or completely transparent). Consider the cylindrical Huygens’ wavelet produced at each narrow slit when the grating is illuminated by a normally incident plane wave as shown in Fig. 1. It is clear to see that there will be constructive interference only in those

discrete directions where the optical path difference from adjacent slits is an integral number of wavelengths (i.e., phase differences in multiples of 2π). Every point P in the focal plane of the lens that satisfies this condition will exhibit a primary maximum. The angular width of this interference maximum depends upon the number of slits making up the grating. Figure 2 illustrates the one-dimensional profile of the Fraunhofer diffraction pattern of an array of slits as we progress from two slits (Young’s interference pattern) to three slits, to five slits, and to eleven slits.

The trend is evident. In the limit of a large number of very narrow slits, the primary interference maxima (diffraction orders) become narrower and narrower, with more and more ($n - 2$) small secondary maxima in between them.

The first reported observation of diffraction grating effects was made in 1785 when Francis Hopkinson (one of the signers of the declaration of independence and George Washington’s first Secretary of the Navy) observed a distant street lamp through a fine silk handkerchief. He noticed that this produced multiple images, which to his astonishment did not change location with motion of the handkerchief. He mentioned his discovery to the astronomer David Rittenhouse. Rittenhouse recognized the observed phenomenon as a diffraction effect and promptly made a diffraction grating by wrapping fine wire around the threads of a pair of fine pitch screws. Knowing the pitch of his screws in terms of the Paris inch, he determined the approximate wavelength of light.⁷

*Address all correspondence to James E. Harvey, E-mail: jimh@photoneng.com

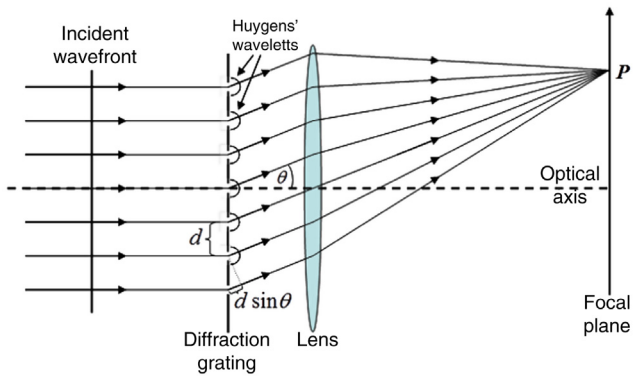


Fig. 1 Illustration of paths of superposed optical disturbances that interfere constructively to produce discrete diffracted orders.

Since the diffraction angle for a given order varies with wavelength, a diffraction grating produces angular dispersion. This angular dispersion is illustrated in Fig. 3 for a grating with a period $d = 10 \mu\text{m}$. Diffracted orders for wavelengths 450, 550, and 650 nm are plotted versus angle.

Spectral resolution and diffraction efficiency are quantities of practical interest in many diffraction grating applications. The diffraction efficiency is defined as the fraction of the incident optical power that appears in a given diffracted order of the grating. Note from Fig. 3 that the zero order exhibits no dispersion, and there is twice as much dispersion in the second order as there is in the first order.

Diffraction gratings can be categorized according to several different criteria: their geometry, material, their efficiency behavior, their method of manufacture, or their intended application. We thus talk about:

- amplitude and phase gratings;
- reflection and transmission gratings;
- binary gratings;
- symmetrical and blazed gratings;
- plane and concave gratings;
- ruled, holographic, and lithographic gratings;
- Bragg type and Raman-Nath type gratings;
- waveguide gratings; and
- fiber gratings.

We acknowledge that this list of grating types is non-exhaustive and nonexclusive but none-the-less is useful for comparing and contrasting grating performance for different gratings types, characteristics, and manufacturing techniques.

Joseph von Fraunhofer began his detailed study of diffraction gratings about 1821. He built the first ruling engine for fabricating reflection gratings on metallic substrates. His insight into the diffraction process led him to predict that diffraction efficiency behavior would “strain even the cleverest of physicists,” which it did for the next 150 years. Many of Fraunhofer’s findings were written up in great detail, so we are entirely justified in calling him the father of diffraction grating technology.^{8,9}

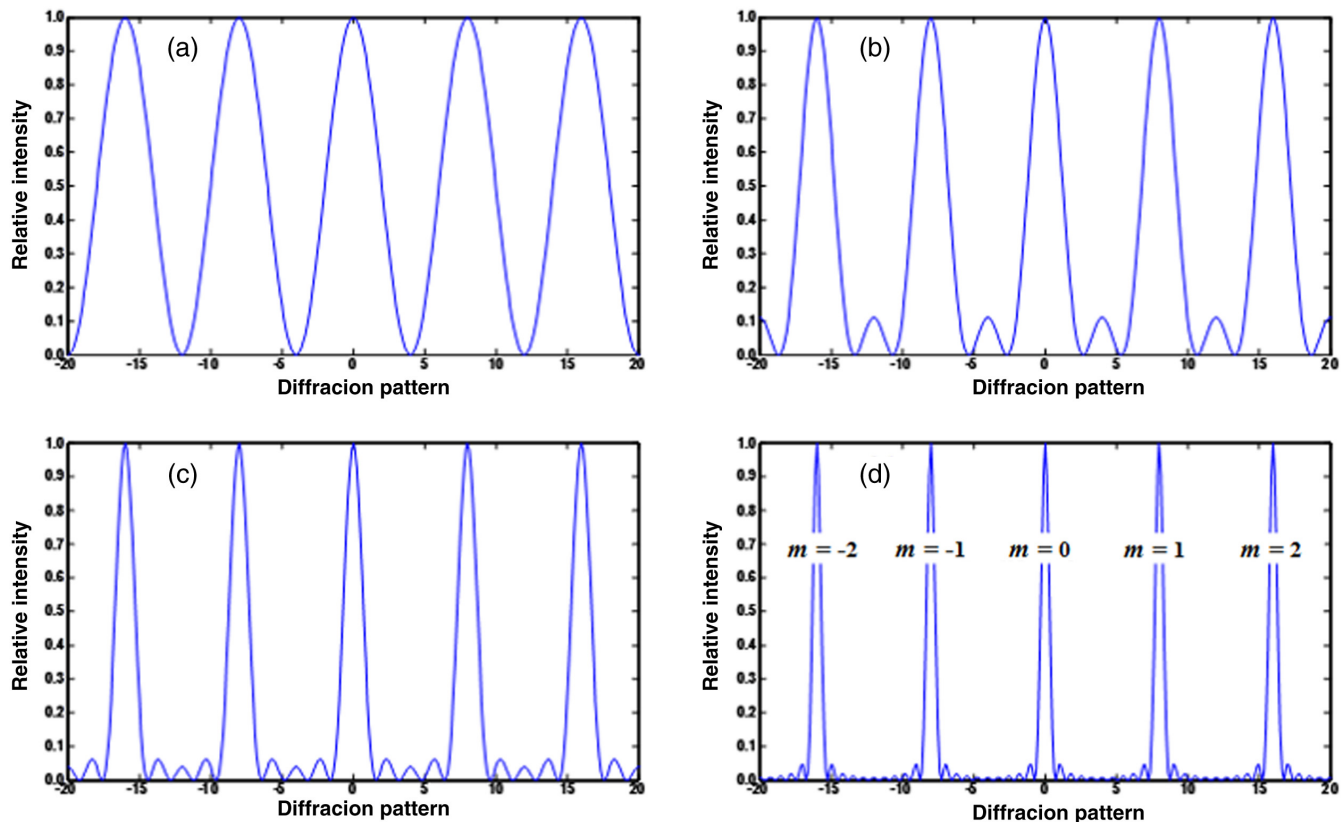


Fig. 2 The Fraunhofer diffraction pattern of an array of equally spaced narrow slits is illustrated as the number of slits is increased: (a) two slits, (b) three slits, (c) five slits, and (d) eleven slits.

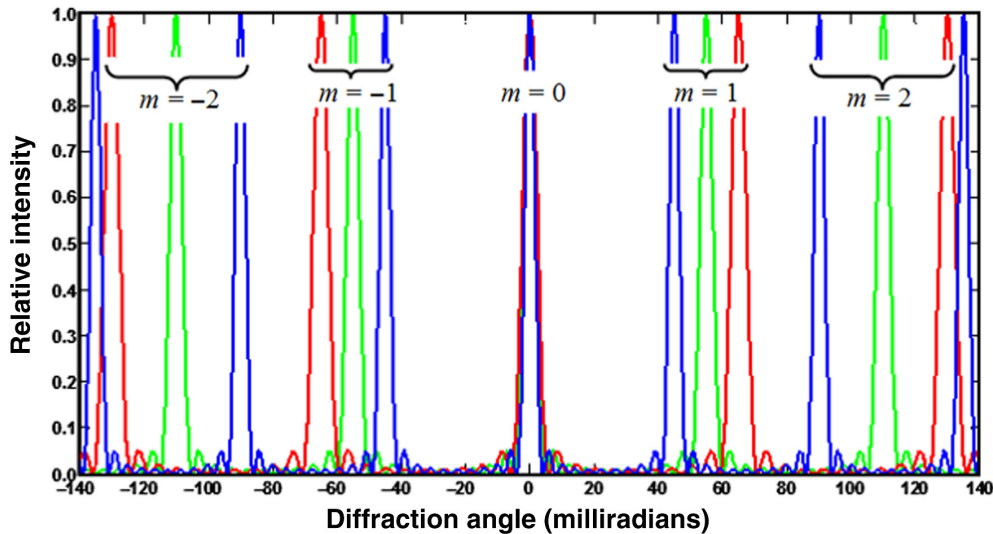


Fig. 3 Illustration of angular dispersion produced by a diffraction grating.

A whole new era of spectral analysis opened up with Rowland’s famous paper in 1882. He constructed sophisticated ruling engines and invented the “concave grating,” a device of spectacular value to modern spectroscopists.¹⁰

John Strong, quoting G. R. Harrison, stated in a JOSA article in 1960—It is difficult to point to another single device that has brought more important experimental information to every field of science than the diffraction grating. The physicist, the astronomer, the chemist, the biologist, the metallurgist, all use it as a routine tool of unsurpassed accuracy and precision, as a detector of atomic species to determine the characteristics of heavenly bodies and the presence of atmospheres in the planets, to study the structures of molecules and atoms, and to obtain a thousand and one items of information without which modern science would be greatly handicapped.”¹¹

A troublesome aspect of the multiple order behavior of diffraction gratings is that adjacent higher order spectra frequently overlap. In fact, in Fig. 3, one can see the third-order principle maximum for blue light almost overlapping the second-order red principle maximum. One can readily show that the second order for wavelengths 100, 200, and 300 nm is diffracted into the same directions as the first order for wavelengths 200, 400, and 600 nm.

Two generalizations to the behavior of gratings must now be discussed. First, if the individual slits making up the grating have significant width (in order to transmit more light), the Fraunhofer diffraction pattern of an individual slit will form an envelope function modulating the strength of the discrete diffracted orders.^{12–15} For the case illustrated in Fig. 4, we have chosen the width of the slits to be one-third of the slit separation. You will note that every third diffracted

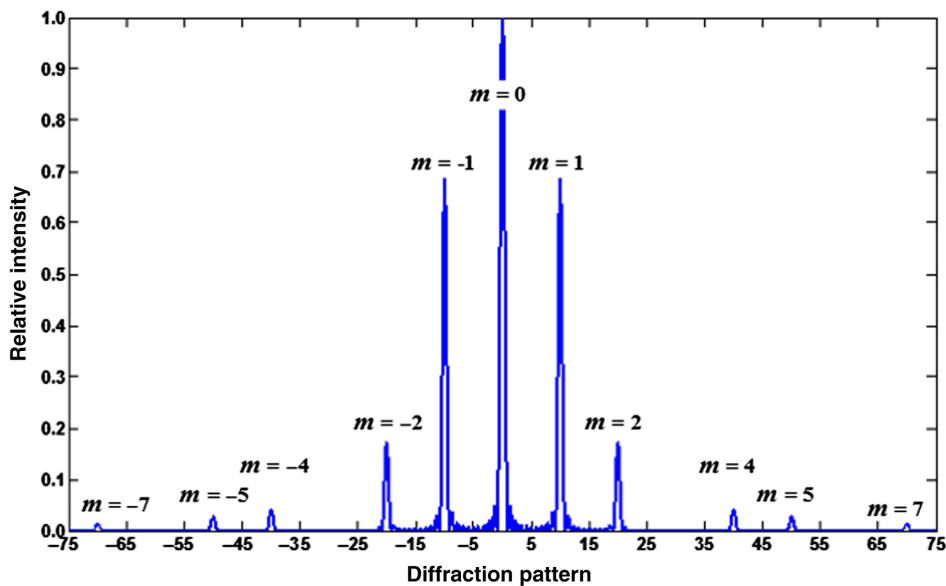


Fig. 4 The Fraunhofer diffraction pattern of an array of eleven equally spaced slits whose width is one-third of their spacing.

order is absent. This is caused by the envelope function going to zero at those locations.

The second generalization includes the situation where the light is incident upon the grating at an arbitrary angle θ_i rather than normal incidence. This situation will be taken care of by including the incident angle in the grating equation discussed in Sec. 2, where we will review the planar grating equation and the sign convention for numbering the various diffracted orders.

The more general phenomenon of “conical” diffraction that occurs with large obliquely incident angles will be discussed in Sec. 3 and the parametric behavior will be shown to be particularly simple and intuitive when formulated and displayed in terms of the direction cosines of the incident and diffracted angles. In Sec. 4, we will use the remarkably intuitive direction cosine diagram to portray the conical grating behavior exhibited in the presence of large obliquely incident beams and arbitrary orientation of the grating. Section 5 examines the paraxial diffraction efficiency behavior of several elementary grating types. Section 6 will review the underlying concepts of nonparaxial scalar diffraction theory and apply them to the sinusoidal and square-wave amplitude gratings when the +1 diffracted order is maintained in the “Littrow condition.” This nonparaxial behavior includes the well-known Rayleigh (Wood’s) anomaly effects that are usually thought to only be predicted by rigorous (vector) electromagnetic theory.¹⁶

A summary, statement of conclusions, and an extensive set of references will then complete this paper.

2 Planar Grating Equation and Sign Convention

Monochromatic light of wavelength λ incident upon a refractive transmission grating (interface between two dielectric media exhibiting a periodic surface relief pattern) of spatial period d at an angle of incidence θ_i will be diffracted into the discrete angles θ_m according to the following (planar) grating equation:^{3,4,16–18}

$$n' \sin \theta_m - n \sin \theta_i = -m\lambda/d, \quad m = 0, \pm 1, \pm 2, \pm 3, \quad (1)$$

where n is the refractive index of the media on the incident side of the diffracting surface, n' is the refractive index of the media containing transmitted diffracted light, and m is an integer called the order of diffraction. The sign of m is arbitrary and determines the sign convention for labeling diffracted orders.

The equation for a reflection grating can be obtained by setting $n' = -n$, just as we do when tracing rays from a reflecting surface:⁴

$$\sin \theta_m + \sin \theta_i = m\lambda/nd, \quad m = 0, \pm 1, \pm 2, \pm 3. \quad (2)$$

Note that setting $m = 0$ in Eq. (1) results in θ_0 having the same sign as θ_i . Likewise, setting $m = 0$ in Eq. (2) results in θ_0 having the opposite sign as θ_i . We have thus adopted a sign convention that conforms to that used in geometrical optics whereby all angles are directional quantities measured from optical axes or surface normals to refracted or reflected rays. These directional angles are “positive if counterclockwise,” and “negative if clockwise.” An “angle” here is the smaller of the two angles that a ray forms with the axis or surface normal.

For a thin diffraction grating in air, we thus have $n = n' = 1$, and the two grating equations can be combined to yield

$$\sin \theta_m \mp \sin \theta_i = \mp m\lambda/d, \quad m = 0, \pm 1, \pm 2, \pm 3. \quad (3)$$

Here the minus signs describe a transmission grating and the plus signs describe a reflection grating as illustrated in Fig. 5. Note from this figure that the zero order corresponds to the directly transmitted or specularly reflected beam.

The arrangement of the diffracted orders is the same for the two gratings except they are reflected about the plane of the reflection grating. Note also that the algebraic signs of two directional angles are different if they are measured on different sides of the grating normal. A final useful observation is that for both the transmission and the reflection grating, the positive diffracted orders lie on the same side of the grating normal as the incident beam; whereas the negative

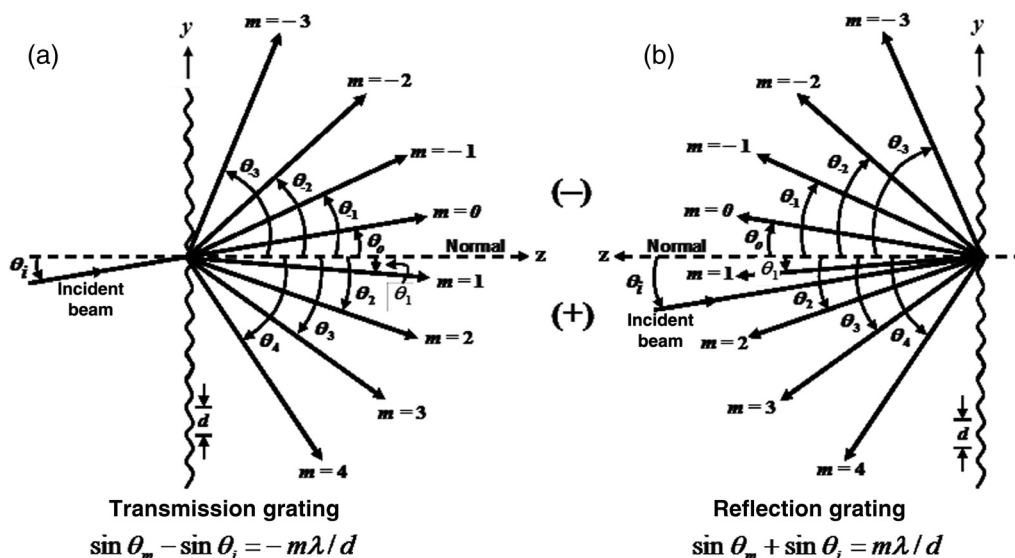


Fig. 5 Illustration of a thin (a) transmission grating and (b) reflection grating.

diffracted orders lie on the opposite side of the grating normal from the incident beam. A “plus” sign has thus been placed on the lower side of the grating normal in Fig. 5 and a “minus” sign has been placed on the upper side of the grating normal as an indicator of our sign convention. Some authors absorb the minus sign on the right side of Eq. (3) into the m , thus achieving a seemingly simpler equation. However, this results in a different sign convention for labeling the diffracted orders.

We have specifically chosen the form of Eq. (3) not only to maintain the sign convention for directional angles used almost exclusively in geometrical optics and optical design ray trace codes (positive if counterclockwise and negative if clockwise), but also to be consistent with the sign convention for labeling diffraction grating order numbers used by the popular *Diffraction Grating Handbook* published and distributed free by the Newport Corporation (formerly Richardson Grating Laboratory).¹⁹

The above grating equations are restricted to the special case where the grating grooves/lines are oriented perpendicular to the plane of incidence, i.e., the plane containing the incident beam and the normal to the grating surface. For this situation, all of the diffracted orders lie in the plane of incidence.

The more general phenomenon of conical diffraction that occurs with large obliquely incident angles is rarely discussed in elementary optics or physics text books. However, the formulation of a nonparaxial scalar diffraction theory²⁰⁻²³ provides a simple and intuitive means of gaining additional insight into this nonparaxial diffraction grating behavior.

3 Conical Diffraction in Direction Cosine Space

Consider diffraction from a conventional linear reflection grating. However, suppose the incident light strikes the grating at a large oblique angle (represented by direction cosines α_i and β_i) as illustrated in Fig. 6. The resulting diffraction behavior is described by the following grating equation written in terms of the direction cosines of the propagation

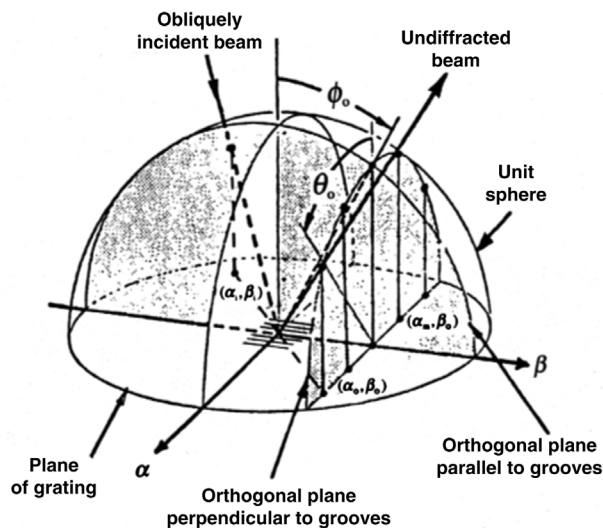


Fig. 6 Illustration of the position of the diffracted orders in real space and direction cosine space for an arbitrary (skew) obliquely incident beam.

vectors of the incident beam and the diffracted orders (the grooves are assumed to be parallel to the y axis):²⁴

$$\alpha_m + \alpha_i = m\lambda/d, \quad \beta_m + \beta_i = 0, \quad (4)$$

where

$$\alpha_m = \sin \theta_m \cos \phi_0, \quad \alpha_i = -\sin \theta_0 \cos \phi_0, \quad \beta_i = -\sin \phi_0. \quad (5)$$

The diffracted orders now propagate along the surface of a cone and will strike the observation hemisphere in a cross section that is not a great circle, but instead a latitude slice as illustrated for a reflection grating in Fig. 6. Note that the direction cosines are obtained by merely projecting the respective points on the hemisphere down onto the plane of the aperture and normalizing to a unit radius. Even for large angles of incidence and large diffracted angles, the various diffracted orders are equally spaced and lie on a straight line only in the direction cosine space.

This behavior is even more evident in Fig. 7, in which the location of the incident beam and the diffracted orders are displayed in direction cosine space for a reflection grating whose grooves are parallel to the y or β axis. The diffracted orders are always exactly equally spaced in direction cosine space and lie in a straight line perpendicular to the orientation of the grating grooves. From Eq. (4), this equidistant spacing of diffracted orders is readily shown to be equal to the nondimensional quantity λ/d . The diffracted orders that lie inside the unit circle are real and propagate, and the diffracted orders that lie outside the unit circle are evanescent (and thus do not propagate).

For a reflection grating, the undiffracted zero order always lies diametrically opposite the origin of the $\alpha - \beta$ coordinate system from the incident beam. As the incident angle is varied, the diffraction pattern (size, shape, separation, and orientation of diffracted orders) remains unchanged but merely shifts its position maintaining the above relationship between the zero order and the incident beam. Note also that when the plane of incidence is perpendicular to the

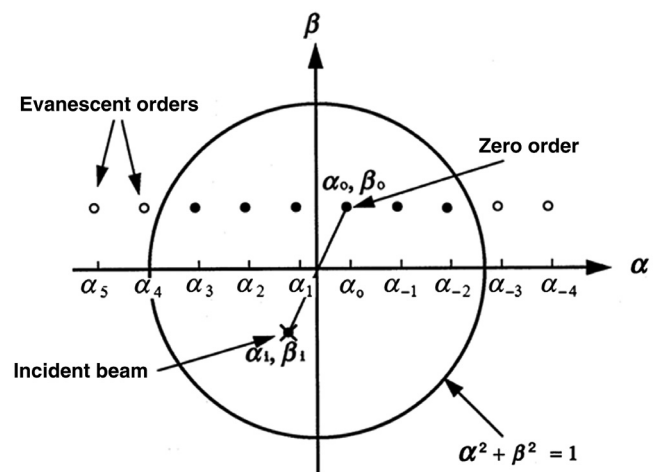


Fig. 7 Relative position of diffracted orders and incident beam in direction cosine space for a reflection grating. Diffracted orders outside the unit circle are evanescent.

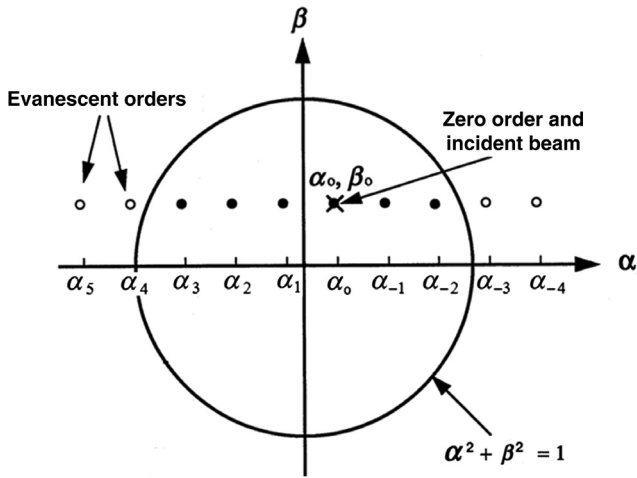


Fig. 8 Relative position of diffracted orders and incident beam in direction cosine space for a transmission grating. The zero order and incident beam are superposed.

grating grooves ($\phi_0 = 0$), Eq. (4) reduces to the familiar grating equation presented in Eq. (3).

For a transmission grating, with our sign convention, the diffraction angle for the zero order is equal to the incident angle ($\theta_0 = \theta_i$). Thus the coordinates of the location in the direction cosine diagram representing the zero order and the incident beam are superposed as illustrated as shown in Fig. 8.

As with the case of the reflection grating, the diffracted orders remain equally spaced and in a straight line as the incident angle is changed, i.e., the size, separation, and orientation of diffracted orders again remains unchanged, merely shifting its position such that the zero order remains superposed upon the incident beam.

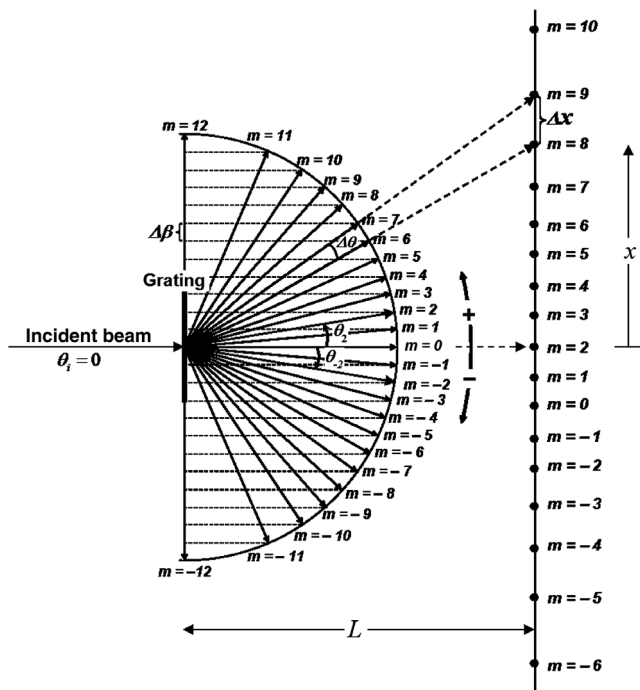


Fig. 9 Graphical illustration of the relationship between $\Delta\theta$, Δx , and $\Delta\beta$.

Figure 9 illustrates the propagating diffracted orders that would exist if a beam were normally incident upon a transmission diffraction grating with $\lambda/d = 0.08333$. There would be precisely 25 propagating diffracted orders including the two at ± 90 deg. The uniform diffracted order spacing in direction cosine space $\Delta\beta$ is contrasted with the increasing angular spacing $\Delta\theta$, and the even more rapidly increasing linear spacing Δx , when the diffracted orders are projected upon a plane observation screen.

For normal incidence, the diffraction grating equation yields

$$m = \frac{d}{\lambda} \sin \theta_m, \quad \text{thus} \quad \frac{dm}{d\theta_m} = \frac{d}{\lambda} \cos \theta_m. \quad (6)$$

Taking the reciprocal of this derivative and writing it as a ratio of differences, we have

$$\frac{\Delta\theta_m}{\Delta m} = \frac{\lambda}{d \cos \theta_m}. \quad (7)$$

Setting Δm equal to unity, we obtain the following expression for the angular spacing of “adjacent” diffracted orders as a function of diffracted angle:

$$\Delta\theta_m = \frac{\lambda}{d \cos \theta_m}. \quad (8)$$

Similarly, from Fig. 9, we can see that

$$x_m = L \tan \theta_m, \quad (9)$$

where L is the distance between the grating and the observation screen.

Taking the derivative of x_m with respect to m , we obtain

$$\frac{\Delta x_m}{\Delta m} = \frac{dx_m}{dm} = \frac{dx_m}{d\theta_m} \frac{d\theta_m}{dm} = \frac{\lambda}{d} L \left(\frac{1}{\cos \theta_m} + \frac{\sin^2 \theta_m}{\cos^3 \theta_m} \right). \quad (10)$$

Again setting Δm equal to unity yields an expression for the linear spacing of adjacent diffracted orders projected upon a plane observation screen as a function of diffracted angle

$$\Delta x_m = \frac{\lambda}{d} L \left(\frac{1}{\cos \theta_m} + \frac{\sin^2 \theta_m}{\cos^3 \theta_m} \right). \quad (11)$$

Plotting the expressions provided by Eqs. (8) and (11) provides a graphical comparison of the relative spacing between adjacent diffracted orders Δx , $\Delta\theta$, and $\Delta\beta$.

Figure 10 indicates that both Δx and $\Delta\theta$ asymptotically approaches infinity for diffracted angles of 90 deg, whereas $\Delta\beta$ remains constant for all diffracted angles. When projected upon a plane screen, the spacing of adjacent diffracted orders increases by a factor of two (100% increase) at a diffraction angle of merely 38 deg. The angular spacing of adjacent diffracted orders increases by a factor of two at a diffraction angle of 60 deg. If only a 5% increase in Δx were allowed, the diffraction angle would have to be held below 10 deg. For $\Delta\theta$, a 5% increase is observed at a diffraction angle of 18 deg.

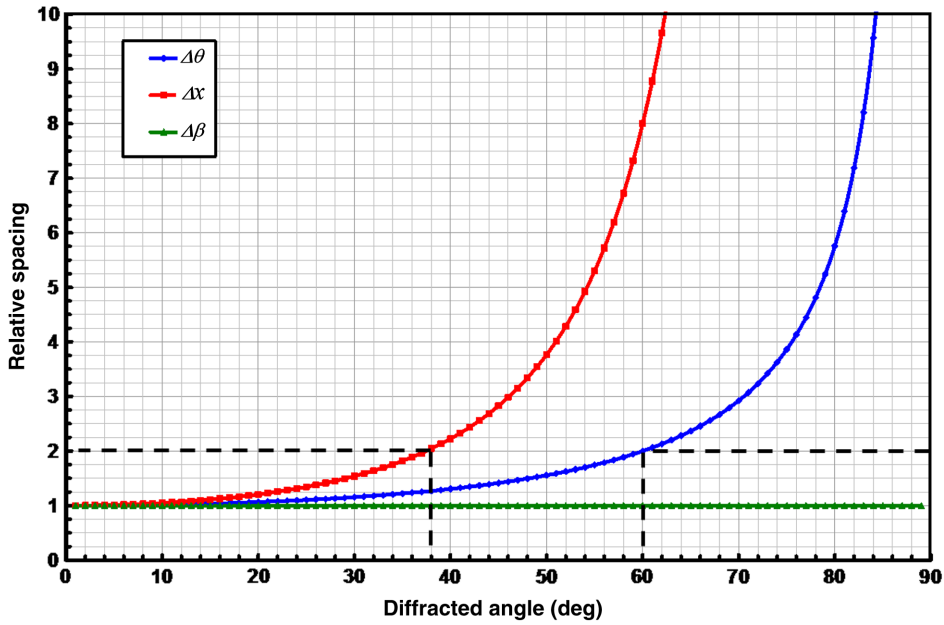


Fig. 10 Graphical illustration of the relative spacing between adjacent diffracted orders Δx , $\Delta\theta$, and $\Delta\beta$.

4 General Grating Equation and the Direction Cosine Diagram

For obliquely incident beams and arbitrarily oriented gratings, a complicated three-dimensional diagram is required to depict the diffraction behavior in real space.²⁵ However, the direction cosine diagram provides a simple and intuitive means of determining the diffraction grating behavior even for these general cases. The general grating equation for

a reflection grating with arbitrarily oriented lines (grooves) is given by²⁴

$$\alpha_m + \alpha_i = \left(\frac{m\lambda}{d}\right) \sin \psi \quad \beta_m + \beta_i = \left(\frac{m\lambda}{d}\right) \cos \psi, \quad (12)$$

where ψ is the angle between the direction of the grating grooves and the α axis. Note that Eq. (12) still reduces to

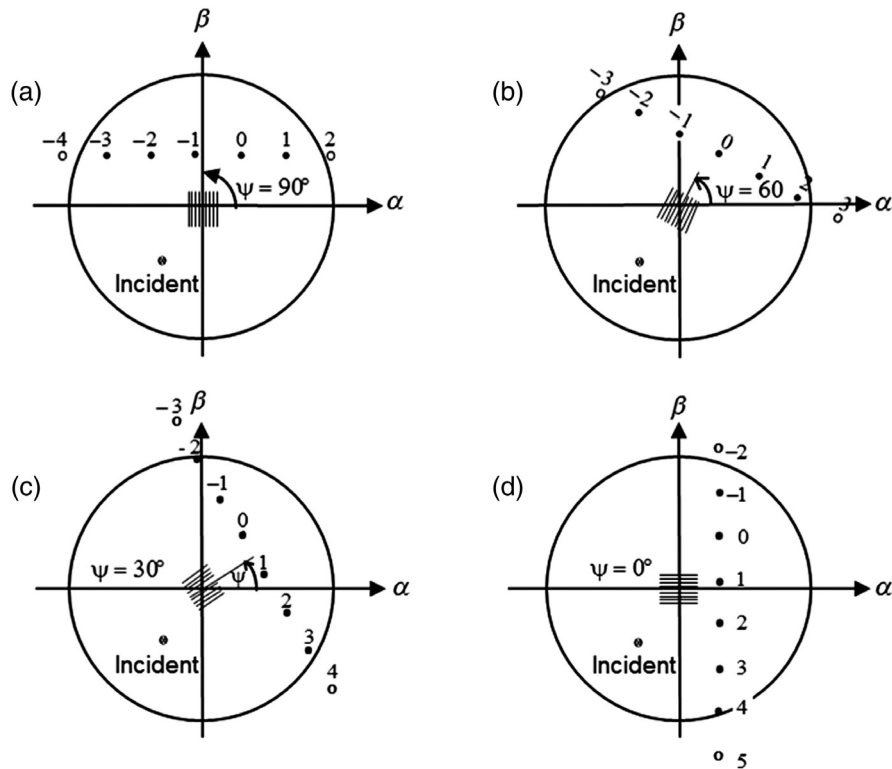


Fig. 11 Direction cosine diagrams for four orientations of a grating with period $d = 3\lambda$ illuminated with an obliquely incident beam ($\alpha_i = -0.3$ and $\beta_i = -0.4$): (a) $\psi = 90$ deg, (b) $\psi = 60$ deg, (c) $\psi = 30$ deg, (d) $\psi = 0$ deg.

Eq. (3) when $\psi = 90$ deg. Figure 11 illustrates the direction cosine diagram for a beam obliquely incident ($\alpha_i = -0.3$ and $\beta_i = -0.4$) upon the same reflection grating discussed above ($d = 3\lambda$) for different orientations of the grating.

Note that in all cases, the zero order is diametrically opposite to the origin from the incident beam and the diffracted orders remain equally spaced in a straight line. However, this line is rotated about the zero order such that it is always perpendicular to the grating grooves. This simple behavior of conical diffraction from linear gratings when expressed in direction cosine space provides understanding and insight not provided by most textbook treatments. It is interesting to note that Rowland expressed the grating equation in terms of direction cosines in a paper published over 125 years ago.²⁶

We have demonstrated that when the grating equation is expressed in terms of the direction cosines of the propagation vectors of the incident beam and the diffracted orders, even wide-angle diffraction phenomena (including conical diffraction from arbitrarily oriented gratings) is shift invariant with respect to variations in the incident angle. New insight and an intuitive understanding of diffraction behavior for arbitrary grating orientation were then shown to result from the use of a simple direction cosine diagram.

5 Paraxial Grating Behavior (Coarse Gratings)

In this section, we discuss the paraxial predictions of diffraction efficiency for five basic types of diffraction gratings: sinusoidal amplitude gratings, square-wave amplitude gratings, sinusoidal phase gratings, square-wave phase gratings, and the classical blazed grating (sawtooth groove profile). The paraxial diffraction efficiencies of various diffracted orders will then be tabulated and compared for these five elementary grating types. For all cases, transverse electric (TE) polarization for the incident beam has been assumed.

If the grating is placed immediately behind an aberration-free positive lens of focal length f that is uniformly illuminated by a normally incident plane wave as illustrated in Fig. 12, the Fraunhofer diffraction pattern produced in the back focal plane of the lens is given by^{27,28}

$$E_2(x_2, y_2) = \frac{E_0}{\lambda^2 f^2} \left| \mathcal{F}\{t_A(x_1, y_1)\} \Big|_{\substack{\xi=x_2/\lambda f \\ \eta=y_2/\lambda f}} \right|^2, \quad (13)$$

where E_0 is the irradiance of the incident beam and $\mathcal{F}\{\}$ denotes the Fourier transform operation:

$$\mathcal{F}\{t(x_1, y_1)\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} t_A(x_1, y_1) \exp[-i2\pi(x_1\xi + y_1\eta)] dx_1 dy_1. \quad (14)$$

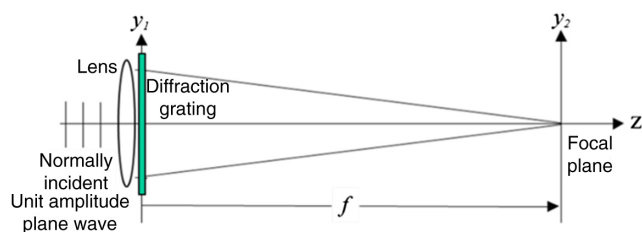


Fig. 12 Geometry for producing a Fraunhofer diffraction pattern of an aperture (or transmission grating) in the back focal plane of a lens.

Both Goodman²⁷ and Gaskill²⁸ discussed in some detail both the Fraunhofer and the Fresnel approximations and the geometrical criteria for each. Goodman, in particular, showed that the cosine obliquity factor in the more general Huygens–Fresnel principle must be approximately unity for both the Fraunhofer and the Fresnel approximations to be valid. It is this requirement that limits our diffraction angles to be paraxial angles.

The classical definition of a paraxial ray is that the ray must lie close to, and make a small angle with, the optical axis, i.e.,^{29,30}

$$\sin \theta \sim \theta, \quad \tan \theta \sim \theta, \quad \text{and} \quad \cos \theta \sim 1. \quad (15)$$

This paraxial requirement obviously places strong limitations on the applicability of the results of this section concerning the grating period-to-wavelength ratio d/λ . The paraxial expressions in Eq. (15) are accurate to within 5% if the angle does not exceed about 18 deg. Although scalar diffraction theory is known to predict diffraction grating performance for TE-polarized light, not transverse magnetic (TM) or unpolarized light,²² at these paraxial angles there will be very little difference between the diffraction efficiency for the two orthogonal polarizations.

5.1 Sinusoidal Amplitude Grating

The complex amplitude transmittance of a thin sinusoidal amplitude grating can be written as

$$t_A(x_1, y_1) = \left[\frac{1}{2} + \frac{a}{2} \cos(2\pi x_1/d) \right] \text{rect}\left(\frac{x_1}{w}, \frac{y_1}{w}\right). \quad (16)$$

We have assumed that the grating is bounded by a square aperture of width w . The parameter a represents the peak-to-peak variation in amplitude transmittance and d is the spatial period of the grating. Figure 13(a) shows a two-dimensional image of the grating, and Fig. 13(b) illustrates a profile of the amplitude transmittance in the x direction.

If this grating is placed immediately behind an aberration-free positive lens of focal length f that is uniformly illuminated by a normally incident plane wave as illustrated in Fig. 12, the Fraunhofer diffraction pattern produced in the back focal plane of the lens is given by Eq. (13).

Applying the scaling theorem and the convolution theorem of Fourier transform theory,²⁸ we can write the Fourier transform of Eq. (16) as

$$\mathcal{F}\{t_A(x_1, y_1)\} = \left[\frac{1}{2} \delta(\xi, \eta) + \frac{a}{4} \delta\left(\xi + \frac{1}{d}, \eta\right) + \frac{a}{4} \delta\left(\xi - \frac{1}{d}, \eta\right) \right] ** w^2 \text{sinc}(w\xi, w\eta), \quad (17)$$

where $**$ is the symbolic notation for the two-dimensional convolution operation.²⁸

Due to the replication property of convolution with a delta function, and since the two-dimensional function is separable into the product of two one-dimensional functions:²⁸

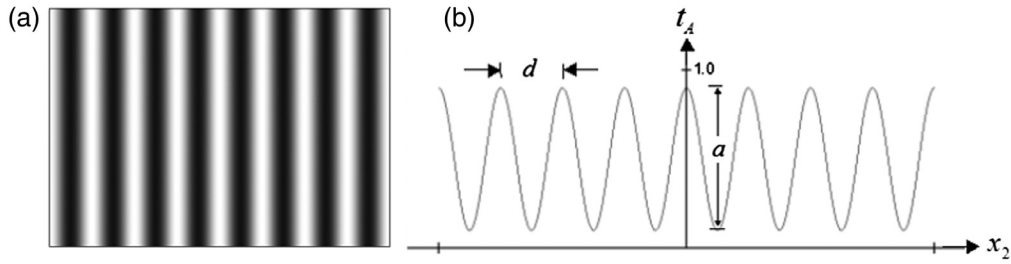


Fig. 13 (a) Two-dimensional image of sinusoidal amplitude grating and (b) profile of amplitude transmittance in the x direction.

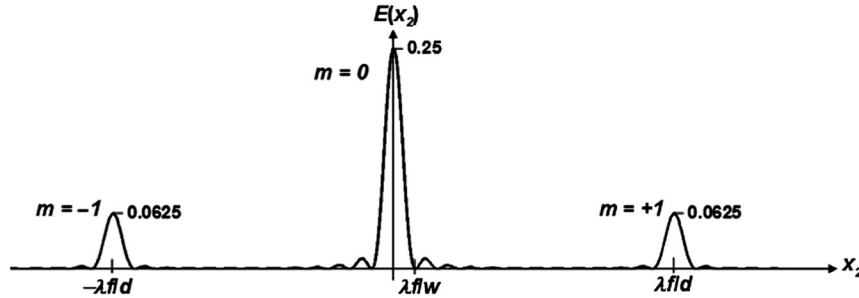


Fig. 14 Irradiance profile of the Fraunhofer diffraction pattern produced by a thin sinusoidal amplitude grating.

$$\mathcal{F}\{t_A(x_1, y_1)\} \Big|_{\substack{\xi=x_2/\lambda f \\ \eta=y_2/\lambda f}} = w^2 \operatorname{sinc}\left(\frac{y_2}{\lambda f/w}\right) \left[\frac{1}{2} \operatorname{sinc}\left(\frac{x_2}{\lambda f/w}\right) + \frac{a}{4} \operatorname{sinc}\left(\frac{x_2 + \lambda f/d}{\lambda f/w}\right) + \frac{a}{4} \operatorname{sinc}\left(\frac{x_2 - \lambda f/d}{\lambda f/w}\right) \right]. \quad (18)$$

If there are many grating periods within the aperture, then $w \gg d$, and there will be negligible overlap between the three sinc functions; hence, there will be no cross terms in the squared modulus of this sum. Substituting this into Eq. (13) thus yields the diffracted irradiance distribution in the focal plane of the lens:

$$E(x_2, y_2) = \frac{E_0 w^4}{\lambda^2 f^2} \operatorname{sinc}^2\left(\frac{y_2}{\lambda f/w}\right) \left[\underbrace{\frac{1}{4} \operatorname{sinc}^2\left(\frac{x_2}{\lambda f/w}\right)}_{m=0} + \underbrace{\frac{a^2}{16} \operatorname{sinc}^2\left(\frac{x_2 + \lambda f/d}{\lambda f/w}\right)}_{m=+1} + \underbrace{\frac{a^2}{16} \operatorname{sinc}^2\left(\frac{x_2 - \lambda f/d}{\lambda f/w}\right)}_{m=-1} \right]. \quad (19)$$

We thus have three discrete diffracted waves or “orders,” each of which is a scaled replica of the Fraunhofer diffraction pattern of the square aperture bounding the grating. The central diffraction lobe is called the “zero order,” and the two side lobes are called the plus and minus “first orders.” The spatial separation of the first orders from the zero order is $\lambda f/d$, whereas the width of the main lobe of all orders is $2\lambda f/w$ as shown in Fig. 14.

The diffraction efficiency is defined as the fraction of the incident optical power that appears in a given diffracted order (usually the +1 order) of the grating. Integrating the irradiance distribution representing a given diffracted order and dividing by the incident optical power $P_o = E_0 w^2$ gives the diffracted efficiency for that order. Since, for any b and x_o

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{b^2} \operatorname{sinc}^2\left(\frac{x-x_o}{b}, \frac{y}{b}\right) = 1, \quad (20)$$

it is clear that the efficiencies are just the coefficients of the three sinc^2 terms in the curly brackets of Eq. (18). These efficiencies are tabulated in Table 1.

The +1 diffracted order thus contains at most (if the quantity a is equal to unity) 6.25% of the optical power incident upon a sinusoidal amplitude grating. This very low diffraction efficiency is not adequate for many applications. As seen in Table 1, the sum of the efficiencies of all three orders is only equal to $1/4 + a^2/8$. The rest of the incident optical power is lost through absorption by the grating.

We will find later in Sec. 6 that a nonparaxial analysis indicates somewhat better performance for certain combinations of grating period and incident angle.

Table 1 Diffraction efficiencies for Fig. 14.

Order #	Efficiency
0	0.25
+1	$a^2/16$
-1	$a^2/16$

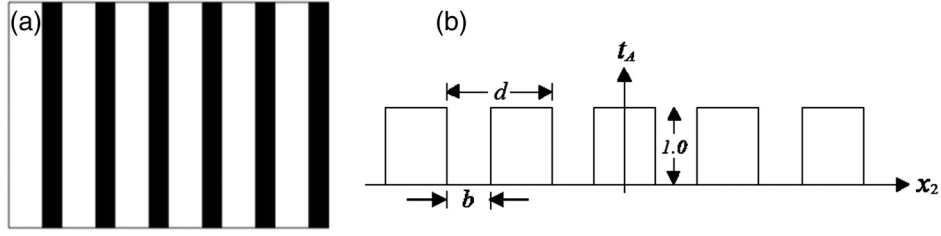


Fig. 15 (a) Two-dimensional image of square-wave amplitude grating (b) and profile of amplitude transmittance in the x direction.

5.2 Square-Wave Amplitude Grating

The complex amplitude transmittance of a thin square-wave amplitude grating can be written as

$$t_A(x_1, y_1) = \left[\text{rect}\left(\frac{x_1}{b}\right) (1) \right. \\ \left. * * \frac{1}{d} \text{comb}\left(\frac{x_1}{d}\right) \delta(y_1) \right] \text{rect}\left(\frac{x_1}{w}, \frac{y_1}{w}\right), \quad (21)$$

where d is the period of the grating, and $b < d$ is the width of the transparent slits in this binary amplitude grating. We have again assumed that the grating is bounded by a square aperture of width w . When $2b = d$, we have a 50% duty cycle square-wave grating, i.e., the transparent slits are separated by opaque lines of width equal to the width of the slits. Such a square-wave amplitude grating is known as a Ronchi ruling. Figure 15(a) shows a two-dimensional image of this grating, and Fig. 15(b) illustrates a profile of the amplitude transmittance in the x direction.

Again, applying the scaling theorem and the convolution theorem of Fourier transform theory,²⁸ we can write

$$\mathcal{F}\{t_A(x_1, y_1)\} = [b \text{sinc}(b\xi) \delta(\eta)] [\text{comb}(d\xi) (1)] \\ * * w^2 \text{sinc}(w\xi, w\eta). \quad (22)$$

However, since the sinc function is separable and the two-dimensional convolution of two separable functions can be written as the product of two one-dimensional convolutions, the above equation can be written as

$$\mathcal{F}\{t_A(x_1, y_1)\} = w^2 \frac{b}{d} \{ [\text{sinc}(b\xi)] [d \text{comb}(d\xi)] \\ * \text{sinc}(w\xi) \} \text{sinc}(w\eta). \quad (23)$$

Also the product of a sinc function with a comb function can be written as an infinite sum of shifted and scaled delta functions,²⁸ hence,

$$\mathcal{F}\{t_A(x_1, y_1)\} = w^2 \frac{b}{d} \left\{ \left[\sum_{m=-\infty}^{\infty} \text{sinc}\left(\frac{mb}{d}\right) \delta(\xi - m/d) \right] \right. \\ \left. * \text{sinc}(w\xi) \right\} \text{sinc}(w\eta). \quad (24)$$

Due to the replication property of convolution with delta functions, we can now write the quantity in the curly bracket as an infinite series of shifted and scaled sinc functions, thus eliminating the convolution operation from the above equation:

$$\mathcal{F}\{t_A(x_1, y_1)\} = w^2 \frac{b}{d} \left[\sum_{m=-\infty}^{\infty} \text{sinc}\left(\frac{mb}{d}\right) \right. \\ \left. \times \text{sinc}\left(\frac{\xi - m/d}{1/w}\right) \right] \text{sinc}(w\eta). \quad (25)$$

Evaluating this function at spatial frequencies $\xi = x_2/\lambda f$ and $\eta = y_2/\lambda f$, and again writing as a two-dimensional function, we obtain

$$\mathcal{F}\{t_A(x_1, y_1)\} \Big|_{\substack{\xi=x_2/\lambda f \\ \eta=y_2/\lambda f}} \\ = w^2 \frac{b}{d} \left[\sum_{m=-\infty}^{\infty} \text{sinc}\left(\frac{mb}{d}\right) \text{sinc}\left(\frac{x_2 - m\lambda f/d}{\lambda f/w}, \frac{y_2}{\lambda f/w}\right) \right]. \quad (26)$$

Since $w \gg d$, there is again negligible overlap between the discrete diffracted orders, and there will be no cross terms in the squared modulus of this sum. The Fraunhofer diffraction pattern predicted by Eq. (13) for a square-wave amplitude grating is thus given by

$$E(x_2, y_2) = \frac{E_0 w^4 b^2}{\lambda^2 f^2 d^2} \left[\sum_{m=-\infty}^{\infty} \text{sinc}^2\left(\frac{mb}{d}\right) \right. \\ \left. \times \text{sinc}^2\left(\frac{x_2 - m\lambda f/d}{\lambda f/w}, \frac{y_2}{\lambda f/w}\right) \right]. \quad (27)$$

There is thus a myriad of diffracted orders produced by the square-wave amplitude grating as shown in Fig. 16. However, they are rapidly attenuated by the sinc^2 envelope function. The irradiance distribution representing the m 'th diffracted order is thus given by

$$E_m(x_2, y_2) = E_0 w^2 \frac{b^2}{d^2} \text{sinc}^2\left(\frac{mb}{d}\right) \left[\frac{1}{(\lambda f/w)^2} \right. \\ \left. \times \text{sinc}^2\left(\frac{x_2 - m\lambda f/d}{\lambda f/w}, \frac{y_2}{\lambda f/w}\right) \right]. \quad (28)$$

The optical power contained in the m 'th diffraction order is obtained by integrating the above irradiance distribution over all space in the $x_2 - y_2$ plane; however, due to Eq. (20), the integral of the quantity in curly brackets is just unity, and we simply obtain

$$P_m(x_2, y_2) = E_0 w^2 \frac{b^2}{d^2} \text{sinc}^2\left(\frac{mb}{d}\right). \quad (29)$$

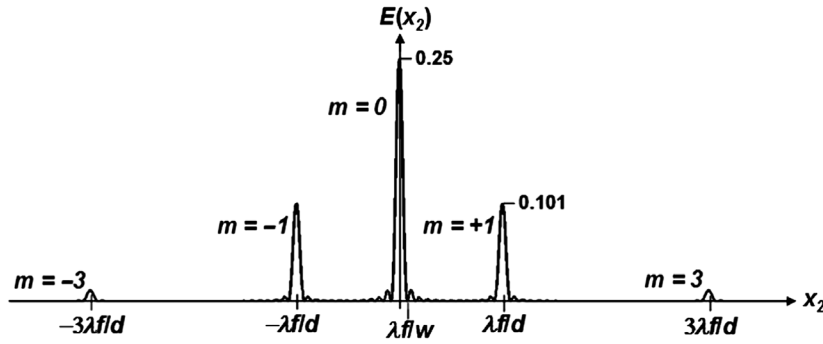


Fig. 16 Irradiance profile of the Fraunhofer diffraction pattern produced by a square-wave amplitude grating.

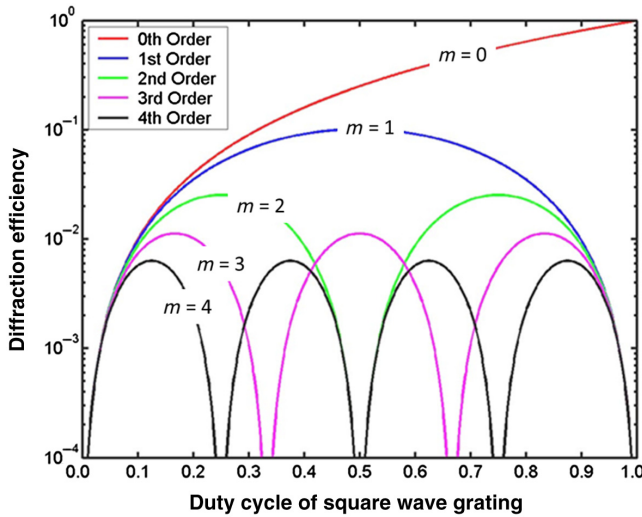


Fig. 17 Diffraction efficiency of the first several orders produced by a square-wave amplitude grating as a function of the width of the transparent slits relative to the grating period.

The diffraction efficiency of the m 'th diffracted order is just the above optical power divided by the optical power in the incident beam, $P_o = E_0 w^2$, or

$$\text{efficiency} \equiv \frac{P_m(x_2, y_2)}{P_o} = \frac{b^2}{d^2} \text{sinc}^2\left(\frac{mb}{d}\right). \quad (30)$$

One can readily calculate that a square-wave amplitude grating with transparent and opaque strips of equal width ($b = d/2$) results in only 10% of the incident optical power being diffracted into the +1 order. This is a little better than we achieved with the sinusoidal amplitude grating, but still not adequate for many applications. Figure 17 illustrates the diffraction efficiency of the first several orders as a function of the parameter b/d .

In spite of the fact that increasing the parameter b/d reduces the absorption of the grating, we see that for $b/d > 0.5$, all of the additional transmitted power, plus some, goes into the zero order, with the efficiency of the +1 order actually diminishing with increasing b/d .

Table 2 lists the efficiency for the first several orders for $b/d = 0.5$. Note that the efficiency of all even orders is identically zero because the zeros of the envelope function in Eq. (28) fall exactly upon the even diffracted orders. We can also see from Fig. 17 that the maximum efficiency

Table 2 Diffraction efficiencies for $b/d = 0.5$.

Order #	Efficiency
0	0.250
± 1	0.101
± 2	0.000
± 3	0.011
± 4	0.000

that can be achieved for the second order is 0.025 for $b/d = 0.25$ or 0.75.

5.3 Sinusoidal Phase Grating

One of the disadvantages of amplitude gratings is that much of the incident optical power is lost through absorption, whereas phase gratings can be made with virtually no absorption losses. Transmission phase gratings can consist of periodic index of refraction variations, or of a periodic surface relief structure, in a thin transparent optical material. Reflection phase gratings are merely a surface relief grating covered with some highly reflective material.

Following Goodman,²⁷ a thin sinusoidal phase grating can be defined by the amplitude transmittance function:

$$t_A(x_1, y_1) = \exp\left[i\frac{a}{2} \sin(2\pi x_1/d)\right] \text{rect}\left(\frac{x_1}{w}, \frac{y_1}{w}\right), \quad (31)$$

where we have ignored a factor representing the average phase delay through the grating. The parameter a represents the peak-to-peak excursion of the sinusoidal phase variation. The grating, bounded by a square aperture of width w , is again placed immediately behind an aberration-free lens that is illuminated with a normally incident plane wave of uniform irradiance E_0 as shown in Fig. 13.

Making use of the Bessel function identity²⁷

$$\exp\left[i\frac{a}{2} \sin(2\pi x_1/d)\right] = \sum_{m=-\infty}^{\infty} J_m\left(\frac{a}{2}\right) \exp(i2\pi m x_1/d), \quad (32)$$

where J_m is a Bessel function of the first kind, order m , and the fact that the exponential Fourier transforms into a shifted

delta function,²⁸ it is readily shown that, within the paraxial limitation, the irradiance distribution in the back focal plane of the lens is given by

$$E(x_2, y_2) = E_0 w^2 \left\{ \sum_{m=-\infty}^{\infty} J_m^2 \left(\frac{a}{2} \right) \left[\frac{1}{(\lambda f/w)^2} \times \text{sinc}^2 \left(\frac{x_2 - m\lambda f/d}{\lambda f/w}, \frac{y_2}{\lambda f/w} \right) \right] \right\}, \quad (33)$$

and the diffraction efficiency of the m 'th diffracted order of a perfectly conducting sinusoidal phase grating is given by the following well-known expression:^{1,22,27,31,32}

$$\text{efficiency} \equiv \frac{P_m(x_2, y_2)}{P_o} = J_m^2 \left(\frac{a}{2} \right), \quad (34)$$

where $a = 4\pi h/\lambda$ and h is the peak to peak groove depth of the sinusoidal reflection grating.

The conservation of energy is easily shown for this perfectly conducting paraxial ($d \gg \lambda$) reflection grating at normal incidence because the sum over m from $-\infty$ to ∞ of the squared Bessel function in Eq. (33) is equal to unity.

Since the Fraunhofer diffraction integral implicitly contains the paraxial approximation, the diffraction angle is proportional to displacement on the focal plane containing the Fraunhofer diffraction patterns

$$\theta_x = \tan^{-1}(x_2/f) \approx x_2/f, \quad \theta_y = \tan^{-1}(y_2/f) \approx y_2/f. \quad (35)$$

Recalling our definitions of radiometric quantities, it is clear that the diffracted intensity distribution (radiant power per unit solid angle) emanating from the grating is thus proportional to the diffracted irradiance distribution (radiant power per unit area) incident upon the focal plane as given by Eq. (33):

$$I(\theta_x, \theta_y) = I_0 \sum_{m=-\infty}^{\infty} J_m^2 \left(\frac{a}{2} \right) \left[\frac{1}{(\lambda f/w)^2} \times \text{sinc}^2 \left(\frac{x_2 - m\lambda f/d}{\lambda f/w}, \frac{y_2}{\lambda f/w} \right) \right]. \quad (36)$$

Figure 18 illustrates the diffracted intensity distribution as a function of diffraction angle θ_x and groove depth h , for a sinusoidal "reflection" grating with period $d = 20\lambda$ operating at normal incidence.

The maximum value of $J_1^2(a/2)$ is 0.3386 and occurs for $a = 3.68$, corresponding to a groove depth of $h = 0.293\lambda$. The diffraction efficiency of the first few orders for this value of a is tabulated in Table 3. Note that the energy falls off rapidly, with 99.88% of the diffracted radiant power contained in diffracted orders $|m| \leq 3$. This paraxial model is accurate only for very coarse gratings ($d \gg \lambda$).

The paraxial behavior described by Eq. (36) above leads to the common misconception that it is impossible to get more than 33.86% of the incident energy into the first diffracted order with a sinusoidal phase grating. "Nothing could be further from the truth!" In fact, if you decrease the grating period, the diffracted angles increase and the higher orders eventually go evanescent. When only the zero and ± 1 diffracted orders remain, changing the incident angle will cause

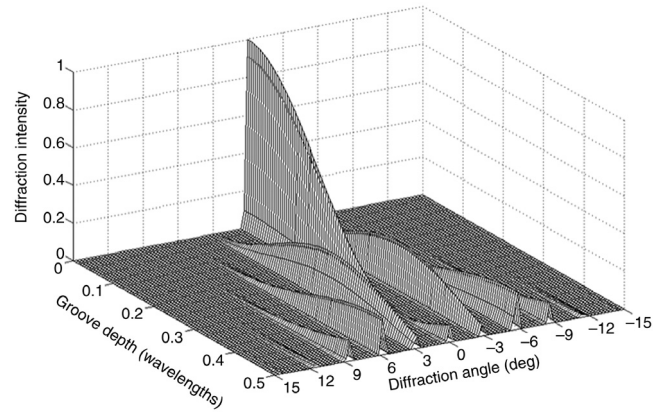


Fig. 18 Diffracted intensity distribution as predicted by the above paraxial model for a sinusoidal reflection grating of period $d = 20\lambda$ operating at normal incidence.

Table 3 Diffraction efficiencies for $a = 3.68$, corresponding to $h = 0.293\lambda$.

Order #	Efficiency
0	1.003×10^{-1}
± 1	3.386×10^{-1}
± 2	9.970×10^{-2}
± 3	1.093×10^{-2}
± 4	6.320×10^{-4}

the -1 order to go evanescent. Then one can vary the groove depth to squelch the energy in the zero order. For a perfectly conducting sinusoidal reflectance grating, we can thus get 100% of the incident energy in the $+1$ diffracted order!³³

In addition to being a paraxial ($d \gg \lambda$) grating, if the sinusoidal reflection grating is also shallow (i.e., the groove depth is much less than a wavelength of the incident light), then the diffraction efficiency of the first orders of the sinusoidal reflection grating can be approximated by

$$\text{efficiency} \equiv J_1^2(a/2) \approx a^2/16. \quad (37)$$

Figure 19 compares the predicted diffraction efficiency of this approximation with the results of Eq. (34) for a perfectly

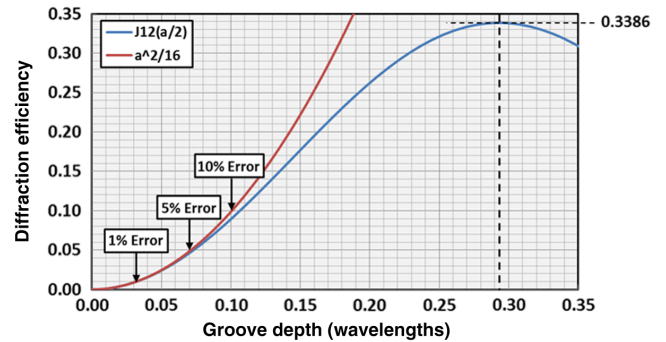


Fig. 19 Comparison of diffracted efficiency of a sinusoidal phase grating as predicted by Eq. (34) and the common approximation for shallow (smooth) gratings expressed in Eq. (37).

conducting surface ($R = 1$) and illustrates how shallow the grating must be to satisfy various error tolerances. Note that the above approximation exhibits only a 1% error in the prediction of diffraction efficiency of the +1 diffracted order at $h = 0.0318\lambda$, a 5% error at $h = 0.0702\lambda$, and a 10% error at $h = 0.098\lambda$.

5.4 Square-Wave Phase Grating

Let us first look at a special case of a rectangular phase grating where the peak-to-peak phase step is equal to π (this should result in zero efficiency for the zero diffracted order) and a duty cycle of $b/d = 0.5$ as illustrated in Fig. 20. From Euler's equation

$$\exp(i\phi) = \cos(\phi) + i \sin(\phi), \quad (38)$$

we readily see that $\exp(i\phi)$ is equal to -1 when $\phi = \pi$ and $+1$ when $\phi = 0$ as illustrated in Fig. 21. The complex amplitude transmittance of this rectangular phase grating bounded by a square aperture of width w thus can be written as

$$t_A(x_1, y_1) = \left\{ \left[2 \operatorname{rect}\left(\frac{x_1}{d/2}\right) (1) ** \frac{1}{d} \operatorname{comb}\left(\frac{x_1}{d}\right) \delta(y_1) \right] - 1 \right\} \times \operatorname{rect}\left(\frac{x_1}{w}, \frac{y_1}{w}\right). \quad (39)$$

Following the discussion of the square-wave amplitude grating, we obtain a Fraunhofer diffraction pattern given by

$$E(x_2, y_2) = \frac{E_0 w^4}{\lambda^2 f^2} \left[\sum_{m=-\infty}^{\infty} \operatorname{sinc}^2\left(\frac{m}{2}\right) \times \operatorname{sinc}^2\left(\frac{x_2 - m\lambda f/d}{\lambda f/w}, \frac{y_2}{\lambda f/w}\right) \right], \quad (40)$$

except that the zero diffracted order is absent. Continuing, we obtain

$$\text{efficiency} \equiv \frac{P_m(x_2, y_2)}{P_o} = \operatorname{sinc}^2\left(\frac{m}{2}\right) \quad \text{for } m \neq 0. \quad (41)$$

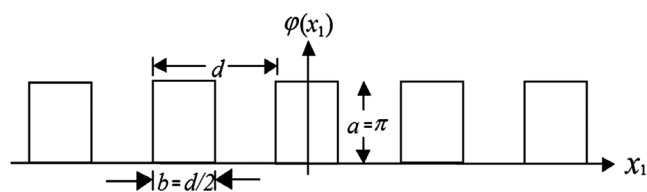


Fig. 20 Phase variation for a special case of a rectangular phase grating with a peak-to-peak phase step of π and a 50% duty cycle.

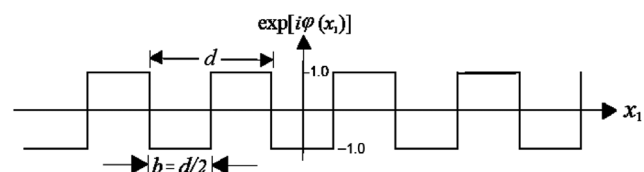


Fig. 21 Complex amplitude for a rectangular phase grating with a peak-to-peak phase step of π and a 50% duty cycle.

Table 4 Diffraction efficiencies for a square-wave phase grating with a π phase step.

Order #	Efficiency
0	0.000
± 1	4.053×10^{-1}
± 2	0.000
± 3	4.503×10^{-2}
± 4	0.000
± 5	1.621×10^{-2}

Table 4 thus lists the efficiency for the first several orders for this special case of a rectangular phase grating. Note that the π phase step has eliminated the zero order, and the efficiency of all other even orders is identically zero because the zeros in the envelope function in Eq. (40) fall exactly upon the even diffracted orders. This thus maximizes the efficiency of the remaining orders.

We have thus seen that the maximum efficiency of the +1 diffracted order (in the paraxial limit) increases from 0.0625 for a sinusoidal amplitude grating, to 0.1013 for a rectangular amplitude grating, to 0.3386 for a sinusoidal phase grating, and to 0.4053 for a rectangular phase grating.

Before we proceed to discuss the classical blazed grating, we want to derive the general solution for the diffraction behavior of an "arbitrary rectangular phase grating." This derivation will lay the groundwork for studying the behavior of diffraction gratings with "arbitrary groove shapes."

For a rectangular phase grating with an arbitrary phase step, the complex amplitude transmittance can be written as

$$t_A(x_1, y_1) = \exp[i\phi(x_1)], \quad (42)$$

where the phase variation is given by

$$\phi(x_1) = a \operatorname{rect}\left(\frac{x_1}{b}\right) (1) ** \frac{1}{d} \operatorname{comb}\left(\frac{x_1}{d}\right) \delta(y_1). \quad (43)$$

This phase variation is illustrated graphically in Fig. 22.

Since this is an even function, it can be decomposed into a discrete cosine Fourier series. The Fourier series coefficients for the above periodic function can be shown to be given by

$$c_n = 2a \frac{b}{d} \operatorname{sinc}\left(\frac{nb}{d}\right), \quad (44)$$

thus

$$\phi(x_1) = \frac{a}{2} + \sum_{n=1}^{\infty} c_n \cos(2\pi n x_1 / d). \quad (45)$$

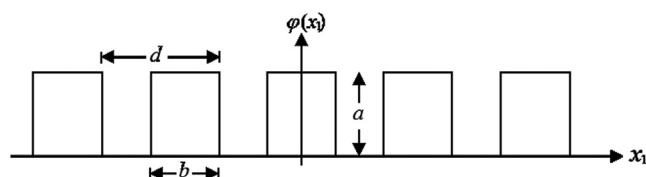


Fig. 22 Phase variation for a rectangular phase grating.

However, we can ignore the constant term resulting from the fact that $\phi(x_1)$ as illustrated above does not have a zero mean. The rectangular phase variation is thus represented as a superposition of cosinusoidal phase variations:

$$\phi(x_1) = \sum_{n=1}^{\infty} c_n \cos(2\pi n x_1 / d). \quad (46)$$

A thin rectangular phase grating can thus be defined by the amplitude transmittance function:

$$t_A(x_1, y_1) = \exp \left[i \sum_{n=1}^{\infty} c_n \cos(2\pi n x_1 / d) \right]. \quad (47)$$

But this can be written as the infinite product:

$$t_A(x_1, y_1) = \prod_{n=1}^{\infty} \{ \exp[i c_n \cos(2\pi n x_1 / d)] \}. \quad (48)$$

Making use of the Bessel function identity³⁴

$$\exp[i z \cos(\theta)] = \sum_{m=-\infty}^{\infty} i^m J_m(z) \exp(im\theta), \quad (49)$$

we have an infinite product of infinite sums, which upon Fourier transforming results in an infinite array of convolutions of infinite sums of delta functions:

$$\begin{aligned} \mathcal{F}\{t_A(x_1, y_1)\} = & \left\{ \left[\sum_{m=-\infty}^{\infty} J_m(c_n) \delta(\xi - n m / d) \right]_{n=1} \right. \\ & * \left[\sum_{m=-\infty}^{\infty} J_m(c_n) \delta(\xi - n m / d) \right]_{n=2} \\ & * \left[\sum_{m=-\infty}^{\infty} J_m(c_n) \delta(\xi - n m / d) \right]_{n=3} \\ & * \cdots * \left. \left[\sum_{m=-\infty}^{\infty} J_m(c_n) \delta(\xi - n m / d) \right]_{n=\infty} \right\}. \end{aligned} \quad (50)$$

Although the above expression might at first appear to be rather unwieldy, it is rather easily solved numerically with the array operations provided with the MATLAB software package. In fact, the above operation results in an array of delta functions that represents the diffracted orders produced by the rectangular phase grating. The squared moduli of the coefficients of those terms are the efficiencies of the diffracted orders.

We can now readily calculate the diffraction efficiencies for a paraxial rectangular phase grating with an arbitrary phase step and duty cycle. Figure 23 graphically illustrates the efficiency of the first few diffracted orders produced by a rectangular phase grating with a phase step of π as a function of the duty cycle (b/d). Note that when b/d equals either zero or unity, that no phase variations exist, and all of the diffracted energy remains in the undiffracted beam (zero order). Also note that for $b/d = 0.5$, we obtain the same results as those tabulated in Table 4.

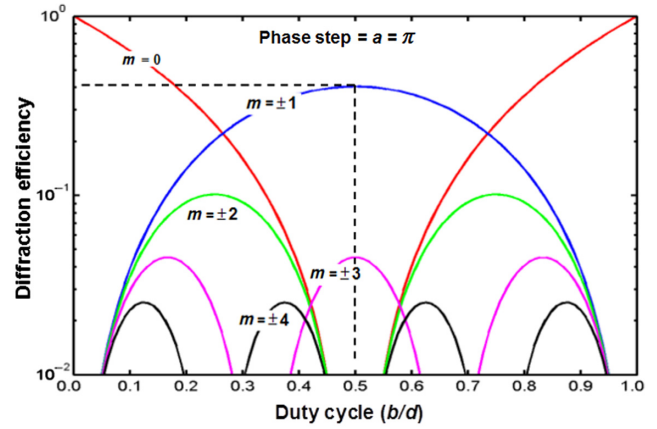


Fig. 23 Efficiency of the first few diffracted orders produced by a rectangular phase grating with a phase step of π as a function of the duty cycle (b/d).

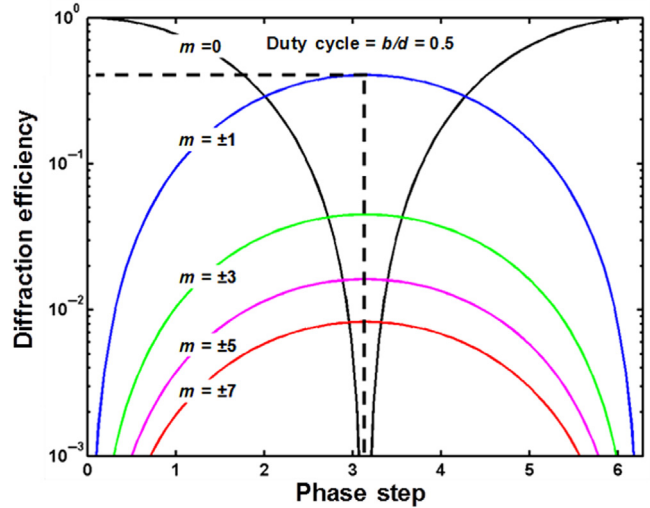


Fig. 24 Efficiency of the first few diffracted orders produced by a rectangular phase grating with a duty cycle of 0.5 as a function of the phase step a .

Similarly, Fig. 24 graphically illustrates the efficiency of the first few diffracted orders produced by a rectangular phase grating with a duty cycle of 0.5 as a function of the phase step a . Note that the even orders are absent. Equation (50) and Figs. 23 and 24 combined constitute a rather unique and comprehensive graphical display of the parametric paraxial performance of square-wave phase gratings.

The above technique can also be used to calculate the paraxial diffraction efficiencies of a reflection grating with arbitrary groove shape by merely supplying the appropriate Fourier coefficients in Eq. (44).

5.5 Classical Blazed Grating

The concept of a blazed grating is that each groove should be so formed that independently, by means of geometrical optics, it redirects the incident light in the direction of a chosen diffracted order, thus making it appear to “blaze” when viewed from that direction. Lord Rayleigh was first to describe the ideal groove shape in 1874.³⁵ He wrote:

“... the retardation should gradually alter by a wavelength in passing over each element of the grating and then fall back to its previous value, thus springing suddenly over a wavelength.” He was not very optimistic about achieving such geometry, but 36 years later, in 1910, Wood³⁶ produced the first grating that we would call “blazed” for use in the infrared. He did this with a tool of carborundum, ruled into copper.

A missing insight that we now take for granted was provided by John Anderson in 1916 while working at the Mt. Wilson Observatory. He demonstrated that superior gratings could be produced by “burnishing” (plastic deformation of the surface) rather than cutting the grooves into the substrate.³⁷ The material thus had to be soft enough to accept local deformation and at the same time be highly polished.

The classical blazed grating is thus a reflection grating with a sawtooth groove profile as shown in Fig. 25. Such gratings have been manufactured for over 150 years by scribing, or burnishing, a series of grooves upon a good optical surface. Originally, this surface was one of highly polished speculum metal.

A major advance in the development of diffraction gratings was the discovery by John Strong in 1936 that vacuum deposited aluminum on glass is a far superior medium into which to rule grating grooves than speculum metal, which had been almost universally used for nearly a century.³⁸ Therefore, in recent times, diffraction gratings have been ruled in thin layers of aluminum or gold deposited upon a glass substrate.

Blazed gratings can be designed for a particular wavelength, incident angle, and diffracted order. The blaze angle θ_B of the grating is given by

$$\theta_B = \tan^{-1}(h/d), \quad (51)$$

where h is the groove depth and d is the grating period.

For a paraxial grating designed to operate at normal incidence, the groove depth must be equal to

$$h = n_B \lambda_B / 2, \quad (52)$$

where n_B is the blaze (or design) order and λ_B is the blaze (or design) wavelength.

The specularly reflected plane wavefront segments will then be out of phase by precisely 2π , thus producing constructive interference for that wavelength and diffracted order. Stated another way, the reflected phase variation over one period of the above grating can be written as

$$\phi(x_1) = \frac{2\pi}{\lambda} \text{OPD}(x_1) = \frac{2\pi}{\lambda} 2h \frac{x_1}{d} = 2\pi n_B \lambda_B x_1 / (\lambda d). \quad (53)$$

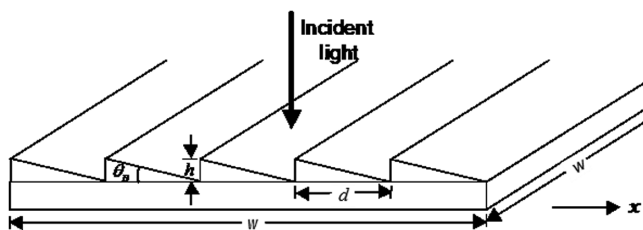


Fig. 25 Classical reflection blazed grating with normally incident light.

Making use of the replication properties of convolution with a comb function, the complex amplitude transmittance (or reflectance in this case) of a grating blazed for the n 'th order and operating at the blaze wavelength can thus be written as

$$t_A(x_1) = \text{rect}\left(\frac{x_1}{d}\right) \exp(-i2\pi n_B \lambda_B x_1 / \lambda d) * \frac{1}{d} \text{comb}\left(\frac{x_1}{d}\right). \quad (54)$$

Using the scaling theorem and the convolution theorem of Fourier transform theory, we can write

$$\mathcal{F}\{t_A(x_1)\} = \text{sinc}[d(\xi - n_B \lambda_B / \lambda d)] [d \text{comb}(d\xi)]. \quad (55)$$

The irradiance of the Fraunhofer diffraction pattern in the $x_2 - y_2$ observation plane a distance z from the grating is proportional to the squared modulus of the Fourier transform of the complex amplitude distribution emerging from the grating:

$$E_2(x_2) \propto \frac{1}{\lambda z} |\mathcal{F}\{t_A(x_1)\}|_{\xi=x_2/\lambda z}^2, \quad (56)$$

$$E_2(x_2) \propto \text{sinc}^2\left[\frac{x_2 - (n_B \lambda_B / \lambda) \lambda z / d}{\lambda z / d}\right] \frac{1}{\lambda z / d} \text{comb}\left(\frac{x_2}{\lambda z / d}\right). \quad (57)$$

When operating at the blaze wavelength $\lambda = \lambda_B$, the peak of the sinc^2 function is centered on the n_B 'th diffracted order and all of the other delta functions (diffracted orders) fall on the zeros of the sinc^2 function. All of the reflected energy is thus diffracted into the n_B 'th diffracted order. Figure 26 shows a plot of diffraction efficiency versus $x_2 \times \lambda z / d$ for a coarse grating blazed to operate in the second order at normal incidence for a wavelength of 550 nm. If $d \gg n_B \lambda_B$, we can be assured, from the planar grating equation, Eq. (3), that the n_B 'th order will be diffracted at a paraxial angle and this predicted behavior will be accurate.

If the incident angle is nonzero, there would be an additional linear phase variation over the entire grating (not each facet individually). Equation (54) describing the complex amplitude distribution emerging from the reflecting blazed grating would thus have to be modified as follows:

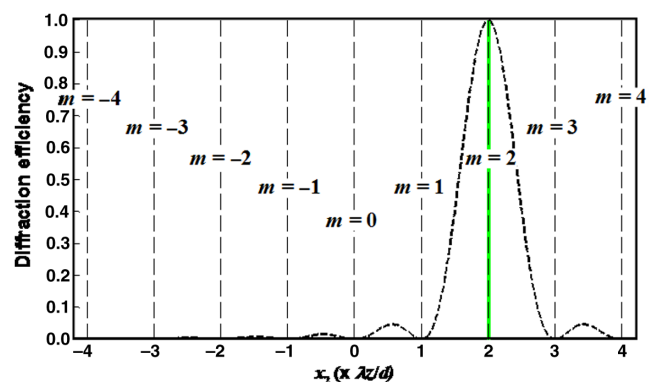


Fig. 26 Illustration of the 100% efficiency achieved by a perfectly reflecting blazed grating designed to operate at normal incidence in the second diffracted order.

$$t_A(x_1) = \left[\text{rect}\left(\frac{x_1}{d}\right) \exp(-i2\pi n_B \lambda_B x_1 / \lambda d) \right. \\ \left. * \frac{1}{d} \text{comb}\left(\frac{x_1}{d}\right) \right] \exp\left(-i2\pi \frac{\theta_0}{\lambda} x_1\right), \quad (58)$$

where the diffraction angle of the zero order (angle of reflection) is merely the negative of the incident angle, i.e., $\theta_0 = -\theta_i$. Again, using the scaling theorem and the convolution theorem of Fourier transform theory, we obtain

$$\mathcal{F}\{t_A(x_1)\} = \{\text{sinc}[d(\xi - n_B \lambda_B / \lambda d)] [d \text{comb}(d\xi)]\} \\ * \delta(\xi - \theta_0 / \lambda). \quad (59)$$

Evaluating at $\xi = x_2 / \lambda z$ and substituting into Eq. (56) yields the following expression for the diffraction pattern projected onto a screen at a distance z from the grating:

$$E_2(x_2) \propto \text{sinc}^2 \left[\frac{x_2 - \left(\frac{n_B \lambda_B}{\lambda} + \frac{\theta_0 d}{\lambda}\right) \frac{\lambda z}{d}}{\frac{\lambda z}{d}} \right] \frac{1}{\frac{\lambda z}{d}} \text{comb} \left[\frac{x_2 - \left(\frac{\theta_0 d}{\lambda}\right) \frac{\lambda z}{d}}{\frac{\lambda z}{d}} \right]. \quad (60)$$

Introducing an arbitrary incident angle will thus shift both the sinc^2 envelope function and the diffracted orders by precisely the same amount. Therefore, under “paraxial” conditions, the diffraction efficiency does not change with incident angle. For example, if we illuminate the above grating blazed for the second order with an incident angle equal to the blaze angle ($\theta_i = \theta_B$), the incident beam will strike the individual facets at normal incidence and the second order will be retroreflected as illustrated in Fig. 27. This situation ($\theta_i = \theta_2$) is referred to as the Littrow condition for the second order,¹⁹ and the efficiency will remain at 100% as shown in Fig. 28. The zero order will of course be specularly reflected from the plane of the grating, and the +1 order will be diffracted normal to the plane of the grating.

The product of a sinc^2 function with a comb function can be written as an infinite sum of shifted and scaled delta functions,²⁸ each of which represents a different diffracted order. Equation (60) can, therefore, be rewritten as

$$E_2(x_2) \propto \sum_{m=-\infty}^{\infty} \text{sinc}^2 \left[\frac{m - (n_B \lambda_B / \lambda + \theta_0 d / \lambda) \lambda z / d}{\lambda z / d} \right] \\ \times \delta[x_2 - (\theta_0 d / \lambda) \lambda z / d]. \quad (61)$$

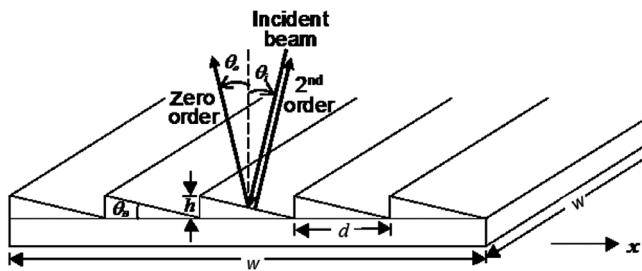


Fig. 27 Blazed grating satisfying the Littrow condition for the second order.

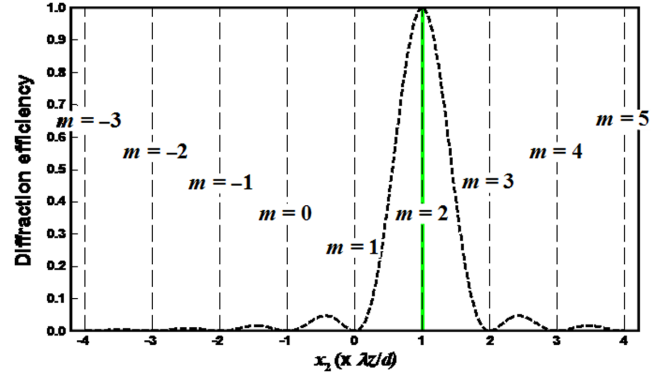


Fig. 28 Illustration of the 100% efficiency achieved by a perfectly reflecting blazed grating satisfying the Littrow condition for the second diffracted order.

For polychromatic light, we can represent the resulting diffracted orders with a summation over the discrete diffracted orders of an integral over some spectral band $\Delta\lambda = \lambda_2 - \lambda_1$:

$$E_2(x_2) \propto \sum_{m=-\infty}^{\infty} \int_{\lambda_1}^{\lambda_2} \text{sinc}^2 \left[\frac{m - (n_B \lambda_B / \lambda + \theta_0 d / \lambda) \lambda z / d}{\lambda z / d} \right] \\ \times \delta[x_2 - (\theta_0 d / \lambda) \lambda z / d]. \quad (62)$$

Figure 29 schematically illustrates the dispersive behavior over the visible spectrum of a grating blazed for the first order at a wavelength 500 nm. The seven classical discrete colors: red ($\lambda_1 = 650$ nm), orange ($\lambda_2 = 600$ nm), yellow ($\lambda_3 = 550$ nm), green ($\lambda_4 = 500$ nm), blue ($\lambda_5 = 450$ nm), indigo ($\lambda_6 = 400$ nm), and violet ($\lambda_7 = 350$ nm) are obtained by replacing the integral in the above equation by a discrete summation:

$$E_2(x_2) \propto \sum_{m=-\infty}^{\infty} \sum_{\lambda=\lambda_1}^{\lambda_7} \text{sinc}^2 \left[\frac{m - (n_B \lambda_B / \lambda + \theta_0 d / \lambda) \lambda z / d}{\lambda z / d} \right] \\ \times \delta[x_2 - (\theta_0 d / \lambda) \lambda z / d]. \quad (63)$$

Figure 30 illustrates that the dispersion is indeed doubled if the grating is blazed for the second diffracted order. Note

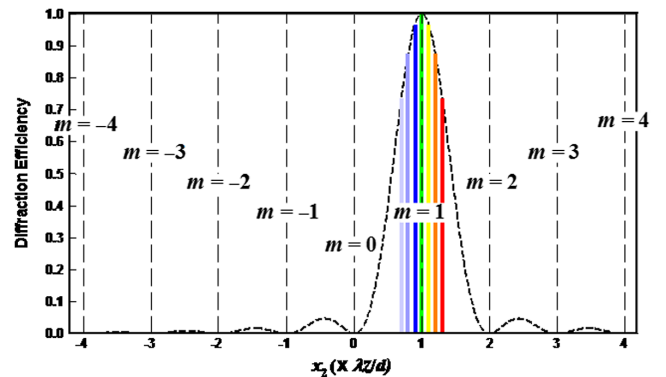


Fig. 29 Illustration of the dispersion produced over the visible spectrum by a grating blazed for a wavelength of 500 nm in the first diffractive order.

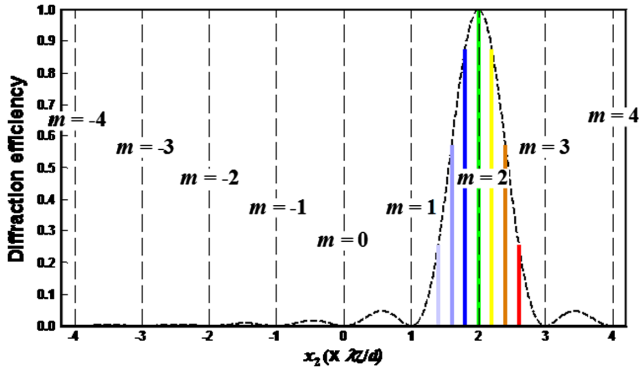


Fig. 30 Illustration that the dispersion is proportional to the diffracted order number.

also that the diffraction efficiency is substantially reduced for all wavelengths other than the blaze wavelength.

In this section, we have systematically described in detail the paraxial behavior of five different classical grating types: the sinusoidal amplitude grating, the square-wave amplitude grating, the sinusoidal phase grating, the square-wave phase grating, and the blazed reflection grating (sawtooth profile). The result of the paraxial diffraction efficiency analyses of these five grating types is summarized in Table 5.

6 Nonparaxial Scalar Diffraction Theory

As discussed briefly in Sec. 1–Sec. 5, it is well-known that the paraxial irradiance distribution on a plane in the far field (Fraunhofer region) of a diffracting aperture is given by the squared modulus of the Fourier transform of the complex amplitude distribution emerging from the diffracting aperture.^{27,28} A slight variation of Eq. (13), without the presence of the lens, can be written as

$$E(x_2, y_2) = \frac{E_0}{\lambda^2 z^2} \left| \mathcal{F}\{U_o^+(x_1, y_1)\} \Big|_{\xi=\frac{x_2}{\lambda z}, \eta=\frac{y_2}{\lambda z}} \right|^2. \quad (64)$$

Here $U_o^+(x_1, y_1) = U_o^-(x_1, y_1)t_1(x_1, y_1)$ is the complex amplitude distribution emerging from the diffracting aperture of complex amplitude transmittance $t_1(x_1, y_1)$, and $U_o^-(x_1, y_1)$ is the complex amplitude incident upon the lens.

$$\begin{aligned} L'(\alpha, \beta - \beta_0) &= K \frac{\lambda^2}{A_s} \left| \mathcal{F}\{U_o'(\hat{x}, \hat{y}; 0) \exp(i2\pi\beta_0\hat{y})\} \right|^2 & \text{for } \alpha^2 + \beta^2 \leq 1 \\ L'(\alpha, \beta - \beta_0) &= 0 & \text{for } \alpha^2 + \beta^2 > 1. \end{aligned} \quad (67)$$

For large incident and/or diffracted angles, the diffracted radiance distribution function will be truncated by the unit circle in direction cosine space. Evanescent waves are then produced and the equation for diffracted radiance must be renormalized. The renormalization factor in Eq. (67) is given by^{20–23}

$$K = \frac{\int_{\alpha=-\infty}^{\infty} \int_{\beta=-\infty}^{\infty} L(\alpha, \beta - \beta_0) d\alpha d\beta}{\int_{\alpha=-1}^1 \int_{\beta=-\sqrt{1-\alpha^2}}^{\sqrt{1-\alpha^2}} L(\alpha, \beta - \beta_0) d\alpha d\beta} \quad (68)$$

and only differs from unity if the diffracted radiance distribution function extends beyond the unit circle in direction cosine space (i.e., only if evanescent waves are produced).

Table 5 Paraxial efficiencies of various grating types (optimized for +1 order).

Grating type	Zero order	First order	Second order	Third order	Fourth order
Sinusoidal amplitude	0.250	0.0625	N/A	N/A	N/A
Square-wave amplitude	0.250	0.101	0.000	0.011	0.000
Sinusoidal phase	0.1003	0.3386	0.0997	0.0109	0.0006
Square-wave phase	0.0000	0.4053	0.0000	0.0450	0.0000
Classical blazed	0.0000	1.0000	0.0000	0.0000	0.0000

The spatial frequencies ξ and η are the reciprocal variables in Fourier transform space. Also the Fresnel diffraction integral is given by the Fourier transform of the product of the aperture function with a quadratic phase factor.^{27,28} Implicit in both the Fresnel and the Fraunhofer approximation is a “paraxial limitation” that restricts their use to small diffraction angles and small angles of incidence.^{27,28} This paraxial limitation severely restricts the conditions under which this conventional linear systems formulation of scalar diffraction theory adequately describes real diffraction phenomena.

A linear systems approach to modeling nonparaxial scalar diffraction phenomena has been developed by normalizing the spatial variables by the wavelength of light:^{20–23}

$$\hat{x} = x/\lambda, \quad \hat{y} = y/\lambda, \quad \hat{z} = z/\lambda, \text{ etc.} \quad (65)$$

The reciprocal variables in Fourier transform space become the “direction cosines” of the propagation vectors of the plane wave components in the angular spectrum of plane waves discussed by Ratcliff,³⁹ Goodman,²⁷ and Gaskill:²⁸

$$\alpha = \hat{x}/\hat{r}, \quad \beta = \hat{y}/\hat{r}, \quad \text{and} \quad \gamma = \hat{z}/\hat{r}. \quad (66)$$

By incorporating sound radiometric principles into scalar diffraction theory, it becomes evident that the squared modulus of the Fourier transform of the complex amplitude distribution emerging from the diffracting aperture yields “diffracted radiance (not irradiance or intensity)”^{20–23.}

In spite of the fact that it is almost universally believed that—“in no way can scalar theory deal with cut-off anomalies,”⁴⁰ the renormalization factor K in Eq. (67) and defined by Eq. (68) enables this linear systems formulation of nonparaxial scalar diffraction theory to predict and model the well-known Wood’s (Rayleigh) anomalies¹⁶ that occur in diffraction efficiency behavior for simple cases of amplitude transmission gratings discussed in the following two sections of this paper.

This renormalization process is also consistent with the law of conservation of energy. However, it is significant that this linear systems formulation of nonparaxial scalar diffraction theory has been derived by the application of Parseval’s theorem and not by merely heuristically imposing the law of conservation of energy.^{20–23}

6.1 Rayleigh Anomalies from Sinusoidal Amplitude Transmission Gratings

Since many individual measurements are required to completely characterize the efficiency behavior of a given grating, it has become commonplace to make diffraction efficiency measurements with a given diffracted order in the Littrow condition.¹⁹ For transmission gratings, a given diffracted order satisfies the Littrow condition if $\theta_m = -\theta_i$. For reflection gratings, the Littrow condition is satisfied if the given diffracted order is antiparallel to the incident beam, i.e., $\theta_m = \theta_i$. This allows the experimenter to leave the detector and the source in a fixed location and merely rotate the grating between measurements.

As previously shown in Table 1 of Sec. 5.1, for a narrow beam normally incident upon a paraxial sinusoidal amplitude grating with modulation of unity, five-eighths of the incident energy is absorbed and three-eighths of it is transmitted. Twenty-five percent of the total incident energy is contained in the zero order and six and one-quarter percent is contained in both the +1 and the -1 orders.

If the +1 diffracted order is in the Littrow condition ($\theta_1 = -\theta_i$) as shown in Fig. 31, the grating equation expressed in Eq. (3) results in the following expression for the incident angle

$$\theta_i = \sin^{-1}(0.5\lambda/d). \quad (69)$$

Substituting Eq. (69) into Eq. (3) yields

$$\sin \theta_m = -\left(m - \frac{1}{2}\right) \frac{\lambda}{d}. \quad (70)$$

Hence, the +1 and -1 diffracted orders produced by a sinusoidal amplitude grating propagate at angles:

$$\theta_1 = -\sin^{-1}\left(\frac{1\lambda}{2d}\right) \quad \text{and} \quad \theta_{-1} = \sin^{-1}\left(\frac{3\lambda}{2d}\right). \quad (71)$$

Note that the sign of these two angles are consistent with the sign convention previously illustrated in Fig. 5. Figure 31 illustrates this situation for $\lambda/d = 0.4$.

As the grating is rotated to increase λ/d , both the angle of incidence and the diffraction angles increase. If we use Eq. (71) to calculate at what value of λ/d the -1 diffracted order goes evanescent, $\theta_{-1} = \pi/2$, we obtain

$$\lambda/d = 2/3 = 0.667. \quad (72)$$

Clearly, the total amount of energy transmitted through this thin grating does not vary as the angle of incidence of

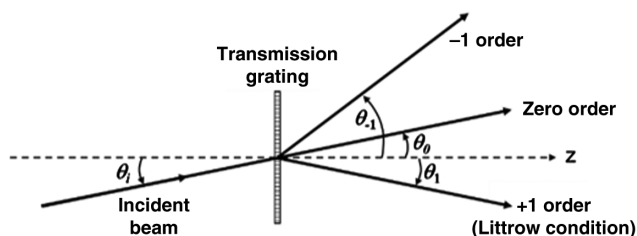


Fig. 31 Diffraction configuration for a sinusoidal amplitude transmission grating with the +1 diffracted order satisfying the Littrow condition when $\lambda/d = 0.4$.

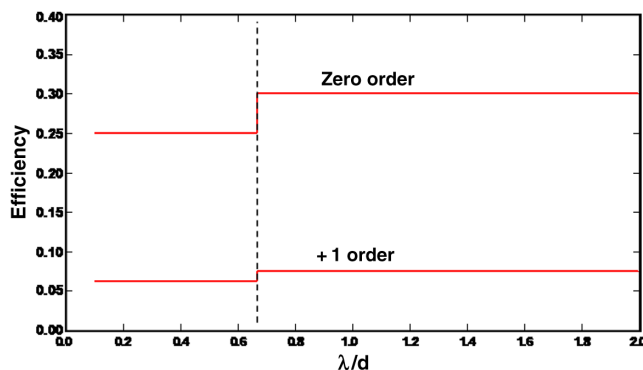


Fig. 32 Illustration of Rayleigh anomalies from a sinusoidal amplitude transmission grating with the +1 diffracted order satisfying the Littrow condition.

the narrow beam is increased. Thus when the -1 diffracted order goes evanescent, the energy that was contained in it (6.25% of the incident energy) is redistributed into the two remaining propagating orders (the Rayleigh anomaly phenomenon).

According to Eq. (68), the renormalization constant K is equal to

$$K = \frac{\eta_{-1} + \eta_0 + \eta_1}{\eta_0 + \eta_1} = \frac{0.0625 + 0.25 + 0.0625}{0.25 + 0.0625} = 1.2, \quad (73)$$

where η_m is the diffraction efficiency of the m 'th diffracted order. The diffraction efficiency of a sinusoidal amplitude diffraction grating is plotted versus λ/d in Fig. 32.

Note the 20% increase in diffraction efficiency of both the zero and the +1 diffracted order at $\lambda/d > 0.667$.⁴¹ It is thus possible to get a maximum diffraction efficiency of 0.075 for the +1 order with a sinusoidal amplitude grating. In spite of this increase over the paraxial prediction of Sec. 5.1, this low diffraction efficiency combined with the fact that precision sinusoidal amplitude gratings are difficult to fabricate explains why they are rarely used for practical applications.

6.2 Rayleigh Anomalies from Square-Wave Amplitude Gratings

The paraxial behavior of the square-wave amplitude grating was discussed in detail in Sec. 5.2. Equation (28) indicated that there is a myriad of diffracted orders produced; however, they are rapidly attenuated by a sinc^2 envelope function. For a 50% duty cycle square-wave amplitude grating ($d = 2b$), the zeros of the envelope function fall precisely on the even diffracted orders as illustrated in Fig. 33. We see from Eq. (28) and Fig. 33 that the diffraction efficiency of the m 'th diffracted order is given by

$$\eta_m = \frac{1}{4} \text{sinc}^2\left(\frac{m}{2}\right). \quad (74)$$

The paraxial diffraction efficiencies of the first 19 diffracted orders of a square-wave amplitude grating with a 50% duty cycle are listed in Table 6. Note that 25% of the incident energy is contained in the zero diffracted order,

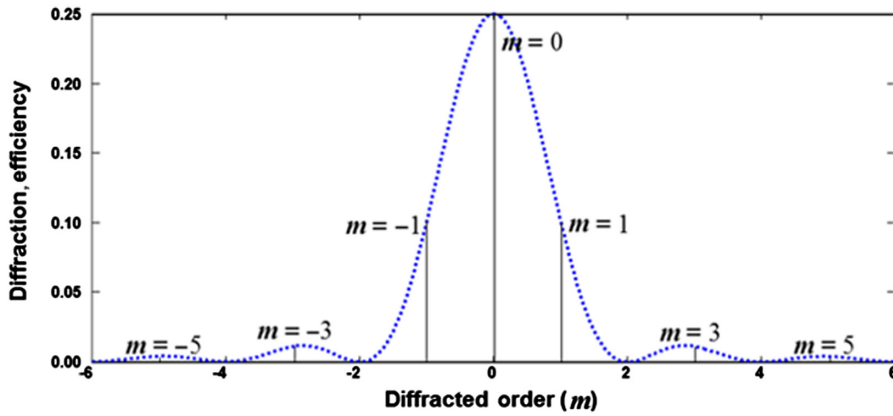


Fig. 33 Schematic illustration of diffraction orders for a 50% duty cycle square-wave amplitude grating. Note that all even orders are absent.

Table 6 Diffraction Efficiencies for the 1st 19 diffracted orders of a square-wave amplitude grating with $b/d = 0.5$.

Order #	Efficiency
0	0.2500
±1	0.1013
±2	0.0000
±3	0.0113
±4	0.0000
±5	0.0041
±6	0.0000
±7	0.0021
±8	0.0000
±9	0.0013

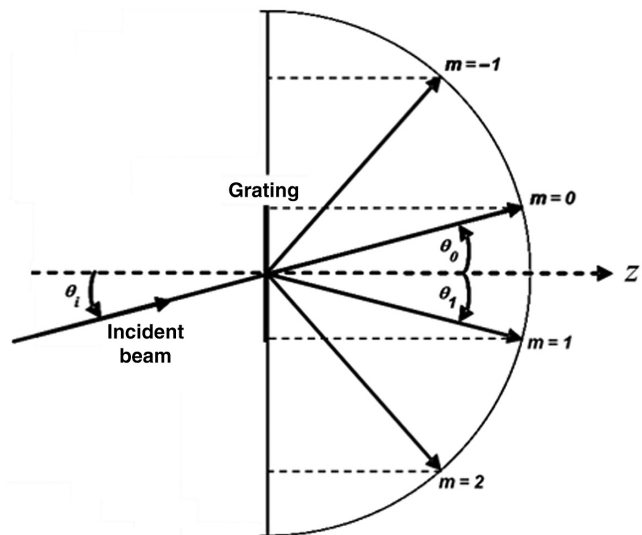


Fig. 34 Illustration of diffraction orders for a transmission grating with $\lambda/d = 0.25$ and the +1 diffracted order satisfying the Littrow condition.

all even orders are identically zero, and the remaining diffracted orders contain another 25%. The remaining 50% of the energy in the incident beam is absorbed by the opaque strips making up the square-wave amplitude grating.

When operating in the Littrow condition, the diffracted orders are distributed symmetrically about the grating normal as shown in Fig. 34. For small λ/d , there are many diffracted orders, but they all have small diffraction angles. As λ/d is increased, both the angle of incidence and the diffraction angles increase, and the higher diffracted orders start going evanescent.

Since the diffracted orders are distributed symmetrically about the grating normal, a positive and a negative order always go evanescent simultaneously. Figure 34 illustrates the situation for a transmission grating with $\lambda/d = 0.25$ and the +1 diffracted order satisfying the Littrow condition ($\theta_1 = -\theta_i$).

Using Eq. (69) to calculate at what value of λ/d the +2 diffracted order goes evanescent, we obtain

$$\sin(-\pi/2) = -1 = -\left(2 - \frac{1}{2}\right) \frac{\lambda}{d} \quad \text{or} \quad \lambda/d = 2/3. \quad (75)$$

Similarly, the -1 order goes evanescent when

$$\sin(\pi/2) = 1 = -\left(-1 - \frac{1}{2}\right) \frac{\lambda}{d} \quad \text{or} \quad \lambda/d = 2/3. \quad (76)$$

We likewise discover that the -2 and +3 diffracted orders go evanescent when $\lambda/d = 2/5$, and the -3 and +4 diffracted orders go evanescent when $\lambda/d = 2/7$, etc.

Hence, when plotting diffraction efficiency versus λ/d , there can be at most only two propagating orders (the zero order and the +1 that is being maintained in the Littrow condition) for $\lambda/d > 2/3$. All other orders are evanescent.

As with the sinusoidal amplitude grating, the total amount of energy transmitted through a square-wave amplitude grating does not vary as the angle of the incident beam is increased. Thus as each pair of diffracted orders goes evanescent, the energy that was contained by them is redistributed into the remaining propagating orders (again the Rayleigh grating anomaly phenomenon) according to the nonparaxial scalar diffraction theory summarized earlier in this section. The renormalization constant K is equal to

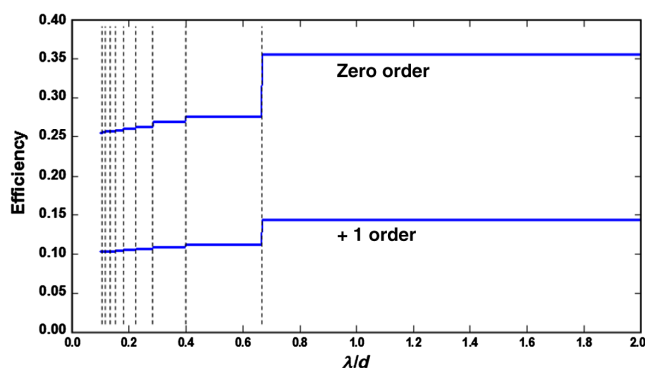


Fig. 35 Illustration of Rayleigh anomalies from a square-wave amplitude transmission grating with the +1 order satisfying the Littrow condition.

$$K = \frac{\sum_{m=-\infty}^{\infty} \eta_m}{\sum_{\text{prop. orders}} \eta_m} = \frac{0.5}{\sum_{\text{prop. orders}} \eta_m}, \quad (77)$$

where η_m is the diffraction efficiency of the m 'th diffracted order.

The diffraction efficiency of the zero order and the +1 order which is maintained in the Littrow condition for a square-wave amplitude diffraction grating is plotted versus λ/d in Fig. 35.

Note in Fig. 35, the incremental increase in diffraction efficiency of both the zero and the +1 diffracted order as successive pairs of diffracted orders go evanescent.⁴¹ A major increase is observed at $\lambda/d > 0.667$ when the -1 order goes evanescent, after which the renormalization factor has a value of

$$K = \frac{0.5}{\eta_0 + \eta_1} = \frac{0.5}{0.25 + 0.1013} = 1.4233. \quad (78)$$

It is thus possible to get a maximum diffraction efficiency of 0.1442 for the +1 order with a square-wave amplitude grating. This is a 42.3% increase over the paraxial value of 0.1013.

7 Summary and Conclusions

Elementary diffraction grating behavior (including diffraction efficiency and dispersion) was reviewed and early challenges in the development of diffraction grating fabrication technology were discussed. The importance of maintaining consistency in the sign convention for the planar diffraction grating equation was emphasized. The advantages of discussing conical diffraction grating behavior in terms of the direction cosines of the incident and diffracted angles were demonstrated, particularly for oblique incident angles and arbitrary grating orientation.

The paraxial grating behavior for coarse gratings ($d \gg \lambda$) was derived and displayed graphically for five elementary grating types: the sinusoidal amplitude grating, the square-wave amplitude grating, the sinusoidal phase grating, the square-wave phase grating, and the classical blazed grating (sawtooth groove profile). Paraxial diffraction efficiencies for various diffracted orders were calculated, tabulated, and compared for these five elementary grating types.

Since much of the grating community erroneously believes that scalar diffraction theory is only valid in the paraxial regime ($d \gg \lambda$), it was emphasized that this limitation is due to an "unnecessary" paraxial approximation in the traditional Fourier treatment of scalar diffraction theory, not a limitation of scalar theory itself. The development of a linear systems formulation of "nonparaxial scalar diffraction theory"^{20–23} was thus briefly reviewed, then used to predict the nonparaxial behavior of both the sinusoidal and the square-wave amplitude transmission gratings when the +1 diffracted order is maintained in the Littrow condition. This nonparaxial behavior included the well-known Rayleigh anomaly effects that are usually thought to require rigorous (vector) electromagnetic theory.

A companion paper, *Understanding Diffraction Grating Behavior, Part II* is currently in progress and will discuss in detail the limits of applicability of nonparaxial scalar diffraction theory to sinusoidal reflection (holographic) gratings as a function of the grating period to wavelength ratio.

References

1. M. Born and E. Wolf, *Principles of Optics*, 6th ed., p. 401, 598, Pergamon Press, Oxford (1980).
2. M. C. Hutley, *Diffraction Gratings*, p. 57, Academic Press, Inc., New York (1982).
3. D. J. Schroeder, *Astronomical Optics*, Academic Press, San Diego, California (1987).
4. D. Malacara, *Geometrical and Instrumental Optics*, p. 210, Academic Press, Inc., San Diego, California (1988).
5. W. L. Wolfe, *Introduction to Grating Spectrometers*, p. 55, SPIE Press, Bellingham, Washington (1997).
6. B. E. A. Saleh and M. C. Teich, *Fundamentals of Photonics*, 2nd ed., p. 804, Wiley Interscience, Hoboken, New Jersey (2012).
7. D. Rittenhouse, "An optical problem proposed by F. Hopkinson and solved," *Trans. Am. Phil. Soc.* **2**, 202–206 (1786).
8. J. Fraunhofer, "Kurtzer Bericht von den Resultaten neuerer Versuche uber die Gesetze des Lichtes, und die Theorie derselben," *Gilbert's Ann. Phys.* **74**, 337–378 (1823).
9. J. Fraunhofer, "Uber die Brechbarkeit des Electrishes Lichtes," *K. Acad. D. Wiss. Zu Munchen*, pp. 61–62 (1824).
10. H. Rowland, "Preliminary notice of results accomplished in the manufacture and theory of gratings for optical purposes," *Philos. Mag.* **13**, 469–474 (1882).
11. J. Strong, "The John Hopkins University and diffraction gratings," *J. Opt. Soc. Am.* **50**, 1148–1152 (1960).
12. R. S. Longhurst, *Geometrical and Physical Optics*, John Wiley and Sons, New York (1957).
13. F. A. Jenkins and H. E. White, *Fundamentals of Optics*, McGraw-Hill Book Co., New York (1957).
14. M. V. Klein, *Optics*, John Wiley and Sons, Inc., New York (1970).
15. E. Hecht, *Optics*, 2nd ed., Addison Wesley Publishing Co., Reading, Massachusetts (1987).
16. R. W. Wood, "On a remarkable case of uneven distribution of light in a diffraction grating spectrum," *Philos. Mag.* **4**, 396–402 (1902).
17. P. Mouraloux and J. McDonald, *Geometrical Optics and Optical Design*, Oxford University Press, New York (1997).
18. K. Moore, *ZEMAX Optical Design Program User's Guide*, Version 9.0, Focus Software, Inc. (2000).
19. C. Palmer and E. Loewen, *Diffraction Grating Handbook*, 7th ed., Newport Corp., Rochester, New York (2005).
20. J. E. Harvey et al., "Diffracted radiance: a fundamental quantity in a non-paraxial scalar diffraction theory," *Appl. Opt.* **38**, 6469–6481 (1999).
21. J. E. Harvey et al., "Diffracted radiance: a fundamental quantity in a non-paraxial scalar diffraction theory: errata," *Appl. Opt.* **39**, 6374–6375 (2000).
22. J. E. Harvey, A. Krywonos, and D. Bogunovic, "Non-paraxial scalar treatment of sinusoidal phase gratings," *J. Opt. Soc. Am. A* **23**, 858–865 (2006).
23. A. Krywonos, "Predicting surface scatter using a linear systems formulation of non-paraxial scalar diffraction," PhD Dissertation, University of Central Florida (2006).
24. J. E. Harvey and C. L. Vernold, "Description of diffraction grating behavior in direction cosine space," *Appl. Opt.* **37**, 8158–8160 (1998).
25. M. G. Moharam and T. K. Gaylord, "Three-dimensional vector coupled-wave analysis of planar-grating diffraction," *J. Opt. Soc. Am.* **73**, 1105–1112 (1983).

26. H. A. Rowland, "XXXIX. Gratings in theory and practice," *Philos. Mag.* **35**(216), 397–419 (1893).
27. J. W. Goodman, *Introduction to Fourier Optics*, McGraw-Hill, New York (1968).
28. J. D. Gaskill, *Linear Systems, Fourier Transforms, and Optics*, Wiley, New York (1978).
29. R. Kingslake, *Optical System Design*, p. 32, Academic Press, Inc., Orlando, Florida (1983).
30. W. J. Smith, *Modern Optical Engineering*, 2nd ed., p. 32, McGraw-Hill, Inc., New York (1990).
31. T. K. Gaylord and M. G. Moharam, "Analysis and applications of optical diffraction by gratings," *Proc. IEEE* **73**, 894–937 (1985).
32. C. V. Raman and N. S. N. Nath, "The diffraction of light by high frequency sound waves," *Proc. Indian Acad. Sci. A* **2**, 406–412 (1935).
33. J. E. Harvey and R. N. Pfisterer, "Parametric diffraction efficiency of non-paraxial sinusoidal reflection gratings," *Proc. SPIE* **10375**, 103750B (2017).
34. I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals Series and Products*, p. 973, Academic Press, New York (1965).
35. J. W. Strutt (Lord Rayleigh), "XII. On the manufacture and theory of diffraction gratings," *Philos. Mag.* **47**, 81–93 (1874).
36. R. Wood, "LXXXV. The Echelette Grating for the Infra-red," *Philos. Mag.* **20**, 770–778 (1910).
37. H. D. Babcock, "Bright diffraction gratings," *J. Opt. Soc. Am.* **34**, 1–5 (1944).
38. J. Strong, "The evaporation process and its application to the aluminizing of large telescope mirrors," *Astrophys. J.* **83**, 401–423 (1936).
39. J. A. Ratcliff, "Some aspects of diffraction theory and their application to the ionosphere," *Rep. Prog. Phys.* **19**, 188–267 (1956).
40. E. G. Loewen and E. Popov, *Diffraction Gratings and Applications*, p. 376, Marcel Dekker, New York (1997).
41. D. Bogunovic, "Non-paraxial scalar treatment of diffraction grating behavior," MS Thesis, University of Central Florida (2002).

James E. Harvey received his PhD in optical sciences from the University of Arizona. He is a retired associate professor from CREOL at the College of Optics and Photonics of the University of Central Florida, and currently a senior optical engineer with Photon Engineering, LLC, Tucson, Arizona, USA. He is credited with more than 220 publications and conference presentations in diverse areas of applied optics. He is a member of OSA and a fellow and past board member of SPIE.

Richard N. Pfisterer received his bachelor's and master's degrees in optical engineering from the Institute of Optics at the University of Rochester in 1979 and 1980, respectively. He is a co-founder and a president of Photon Engineering at LLC. Previously, he was the head of optical design at TRW (now Northrup-Grumman) and a senior optical engineer at Breault Research Organization. He is credited with more than 20 articles and conference presentations in the areas of optical design, stray light analysis, and phenomenology. He is a member of OSA and SPIE.