Nonparaxial scalar treatment of sinusoidal phase gratings

James E. Harvey, Andrey Krywonos, and Dijana Bogunovic

Center for Research and Education in Optics and Lasers, (CREOL), P.O. Box 162700, 4000 Central Florida Boulevard, University of Central Florida, Orlando, Florida 32816

Received May 2, 2005; revised August 19, 2005; accepted August 24, 2005; posted October 14, 2005 (Doc. ID 61867)

Scalar diffraction theory is frequently considered inadequate for predicting diffraction efficiencies for grating applications where $\lambda/d > 0.1$. It has also been stated that scalar theory imposes energy upon the evanescent diffracted orders. These notions, as well as several other common misconceptions, are driven more by an unnecessary paraxial approximation in the traditional Fourier treatment of scalar diffraction theory than by the scalar limitation. By scaling the spatial variables by the wavelength, we have previously shown that diffracted radiance is shift invariant in direction cosine space. Thus simple Fourier techniques can now be used to predict a variety of wide-angle (nonparaxial) diffraction grating effects. These include (1) the redistribution of energy from the evanescent orders to the propagating ones, (2) the angular broadening (and apparent shifting) of wide-angle diffracted orders, and (3) nonparaxial diffraction efficiencies predicted with an accuracy usually thought to require rigorous electromagnetic theory.

OCIS codes: 050.0050, 050.1950, 050.1940, 050.2770, 260.1960.

1. INTRODUCTION

Since the late 1960s, holographic gratings, fabricated by the exposure of photoresist by a stationary sinusoidal interference fringe field, have become commonplace. The photoresist substrate is chemically developed after exposure to produce a master holographic grating with sinusoidal groove profiles. These master holographic gratings are routinely coated and replicated, as are master ruled gratings.

Prior to the widespread use of holographic gratings, the diffraction characteristics of sinusoidal phase gratings were of interest primarily because other groove profiles (lamellar and blazed gratings) can be Fourier analyzed into a superposition of sinusoidal profiles.[1] Likewise, arbitrary scattering surfaces are routinely modeled as a superposition of sinusoidal surfaces of different amplitudes, periods, and orientations.[2–4]

Scalar diffraction theory is frequently considered inadequate for predicting diffraction efficiencies for grating applications where $\lambda/d > 0.1$.[5–7] It has also been stated that scalar theory imposes energy upon the evanescent diffracted orders.[5] These notions, as well as several other common misconceptions, are actually driven more by an *unnecessary paraxial approximation* in the traditional Fourier treatment of scalar diffraction theory than the scalar limitation. We have found only one recent paper in the literature that seems to support our claim that scalar diffraction theory may indeed be valid for feature sizes much smaller than popular opinion seems to dictate.[8]

To minimize any chance for confusion concerning our motivation or claims in this paper, we want to emphasize here that we are not trying to replace rigorous electromagnetic (vector) theory with scalar theory. And we are not claiming that our new nonparaxial scalar theory somehow provides us with the sometimes dramatic distinctions in the behavior of orthogonal polarizations of light that can be provided by electromagnetic theory. However, we have developed a more convenient Fourier treatment of scalar theory (not restricted to paraxial applications). Furthermore, the nonparaxial diffraction behavior predicted by this scalar theory agrees well with the behavior of TE-polarized light, not TM or unpolarized light.

2. HISTORICAL BACKGROUND

In a review article on diffraction gratings in 1984, Maystre[9] discussed a variety of rigorous vector theories including the Rayleigh method, the Waterman method, his own integral vector method, and other differential and modal methods. He presented a comparison (duplicated here as Fig. 1) of the diffraction efficiency (TE polarization) of the first diffracted order of a perfectly conducting sinusoidal grating ($h/d = 0.20$) in the Littrow condition as calculated by the classical Beckmann–Kirchhoff theory and his own rigorous integral vector theory.[9] For this special case ($\theta_m = -\theta_i$ in Maystre's sign convention), the Beckmann geometric factor[2] reduces to

$$F = \sec\theta_m \frac{1+\cos(\theta_i+\theta_m)}{\cos\theta_i+\cos\theta_m} = \frac{1}{\cos^2\theta_m}, \tag{1}$$

and the classical Beckmann–Kirchhoff theory predicts a diffraction efficiency of

$$\eta_1 = J_1^2\left(k\frac{h}{2}\cos\theta_m\right) \Big/ \cos^4\theta_m, \tag{2}$$

where h is the peak-to-peak amplitude of the sinusoidal grating surface profile.

We have added to Fig. 1 the paraxial scalar prediction of diffraction efficiency for sinusoidal phase gratings provided by Goodman[10] and others[11–13]:

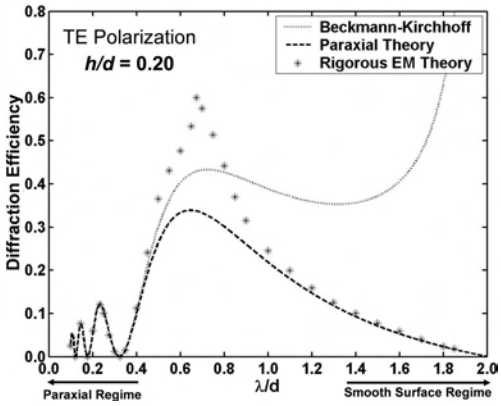

Fig. 1. Diffraction grating efficiency of the first order of a sinusoidal phase grating ($h/d=0.20$) in the Littrow condition as predicted by the nonrigorous Beckmann–Kirchhoff theory, the paraxial scalar theory, and the rigorous integral vector theory. EM, electromagnetic.

$$\eta_m = J_m^2(a/2), \tag{3}$$

where $a=2kh\cos\theta_i$.

The format of Fig. 1 is commonly used to display diffraction efficiency data because the Littrow condition ($\theta_m-\theta_i$, in our standard sign convention) allows one to leave the detector and the source fixed and to merely rotate the grating between measurements. An angular deviation of 2–8 deg between the fixed source and detector is often used for convenience (thus not strictly satisfying the Littrow condition).[14] Every data point in the above diffraction efficiency curve therefore requires a different incident angle. For small λ/d, there are many diffracted orders, but they all have small diffraction angles; hence the left edge of the curve is the paraxial regime. As the grating is rotated to increase λ/d, both the angle of incidence and the diffraction angles increase, and the higher diffracted orders start going evanescent. For the right two thirds of the curves in Fig. 1 there can be at most only two propagating orders, the zero order and the +1 order, the latter being maintained in the Littrow condition. All other orders are evanescent.

Note that all three curves agree well for $\lambda/d<0.4$, i.e., in the paraxial regime. The Beckmann–Kirchhoff theory is better than the paraxial theory for $\lambda/d<0.8$ but blows up for values of $\lambda/d>1.5$. The paraxial scalar theory agrees well not only in the paraxial regime but also in the long wavelength region, i.e., the smooth surface regime. Although many authors state categorically that scalar theory can be applied only in the paraxial regime ($\lambda/d<0.1$), they would probably concede that in the smooth surface regime the higher orders of a sinusoidal phase grating contain a negligible fraction of the energy and the predicted efficiency is not significantly affected when those orders go evanescent. It is in the midportion of the above diffraction efficiency curve that the traditional paraxial scalar diffraction theory is totally inadequate. However, we will now show that it is not the scalar limitation, but instead the unnecessary paraxial approximation implicit in the traditional Fourier treatment of scalar diffraction theory that has prevented its useful application to diffraction grating problems.

3. LINEAR SYSTEM'S FORMULATION OF NONPARAXIAL SCALAR DIFFRACTION THEORY

The fundamental diffraction problem consists of two parts: (i) determining the effects of introducing the diffracting screen (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). Harvey *et al.* have generalized Goodman's Fourier treatment of scalar diffraction theory to include new insight into the phenomenon of diffraction throughout the whole space in which it occurs.[15–17] By using a scaled coordinate system in which all of the spatial variables are normalized by the wavelength of the light,

$$\hat{x}=x/\lambda, \quad \hat{y}=y/\lambda, \quad \hat{z}=z/\lambda, \quad \text{etc.}, \tag{4}$$

the reciprocal variables in Fourier-transform space become the direction cosines of the propagation vectors of the plane-wave components

$$\alpha=\hat{x}/\hat{r}, \quad \beta=\hat{y}/\hat{r}, \quad \gamma=\hat{z}/\hat{r} \tag{5}$$

in the angular spectrum of plane waves discussed by Ratcliff,[18] Goodman,[10] and Gaskill.[19] Recall that this angular spectrum approach leads to a transfer function of free space whose Fourier transform (the impulse response of the diffraction process) is a precise mathematical description of a Huygens wavelet, complete with a cosine obliquity factor and a $\pi/2$ phase delay.[10,15] The convolution of that impulse response with the optical disturbance emerging from a diffracting aperture is precisely the familiar Rayleigh–Sommerfeld diffraction integral for near-field ($z\gg\lambda$) diffraction:

$$U(\hat{x}_2,\hat{y}_2;\hat{z}) = -i\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} U_0(\hat{x}_1,\hat{y}_1;0)\frac{\hat{z}}{\hat{\ell}}\frac{\exp(i2\pi\hat{\ell})}{\hat{\ell}}\mathrm{d}\hat{x}_1\mathrm{d}\hat{y}_1, \tag{6}$$

where ℓ is the distance between an arbitrary point in the diffracting aperture (grating) to an arbitrary point in the observation plane.

The Rayleigh–Sommerfeld diffraction integral is rather unwieldy to solve explicitly for most problems of practical interest. The Fresnel and Fraunhofer diffraction formulas are obtained by retaining only the first two terms in the binomial expansion for the quantity $\hat{\ell}$ in the exponent of the Rayleigh–Sommerfeld diffraction integral. These explicit approximations impose severe restrictions upon both the size of the aperture (relative to the observation distance) and the diffraction angle. In fact, Goodman states that the Fresnel approximation is equivalent to a paraxial approximation.[10] So that we do not impose these restrictions, all terms from the binomial expansion must be retained. This can be accomplished by rewriting Eq. (6) as a Fourier-transform integral of a generalized pupil function that includes phase variations that resemble conventional aberrations.[20,21] Any departures of the actual diffracted wave field from that predicted by the Fourier transform of the aperture function are shown to have the same functional form as the conventional wavefront

aberrations of imaging systems. These aberrations, which are inherent to the diffraction process, are precisely the effects ignored when making the usual Fresnel and Fraunhofer approximations. Significant insight is obtained by recognizing that near-field diffraction patterns are merely aberrated Fraunhofer diffraction patterns. It follows that Fresnel diffraction patterns are merely defocused Fraunhofer diffraction patterns.[20–22]

When a spherical wave is incident upon a diffracting aperture and the observation space is a hemisphere centered upon the aperture, the phase variations mentioned above are frequently negligible.[15] For normal incidence, the diffracted wave field on the hemisphere is then given directly by the Fourier transform of the aperture function

$$U(\alpha,\beta;\hat{r}) = \gamma[\exp(i2\pi\hat{r})/(i\hat{r})]\mathcal{F}\{U_0(\hat{x},\hat{y};0)\}, \tag{7}$$

where $\mathcal{F}$ is the Fourier transform operator, $E_0 = |U_0(\hat{x},\hat{y};0)|^2$ is the irradiance in the plane of the diffracting aperture, and $P_0 = E_0 A_s$ is the total radiant power passing through an aperture of area A_s. Furthermore, this Fourier-transform relationship is valid not merely over a small region about the optical axis, but over the entire hemisphere (with certain restrictions depending upon the residual phase variations).[20]

Now consider the situation in which the incident radiation strikes the diffracting aperture at an angle θ_i as illustrated in Fig. 2. This is equivalent to introducing a linear phase variation across the aperture and attenuating the irradiance in the plane of the aperture by the factor $\gamma_i = \cos(\theta_i)$. By applying the shift theorem of Fourier-transform theory to Eq. (7), we find that the complex amplitude distribution in direction cosine space is a function of $\beta - \beta_0$,

$$U(\alpha,\beta-\beta_0;\hat{r}) = \gamma[\exp(i2\pi\hat{r})/(i\hat{r})]\mathcal{F}\{U_0'(\hat{x},\hat{y};0)\exp(i2\pi\beta_0\hat{y})\} \tag{8}$$

where

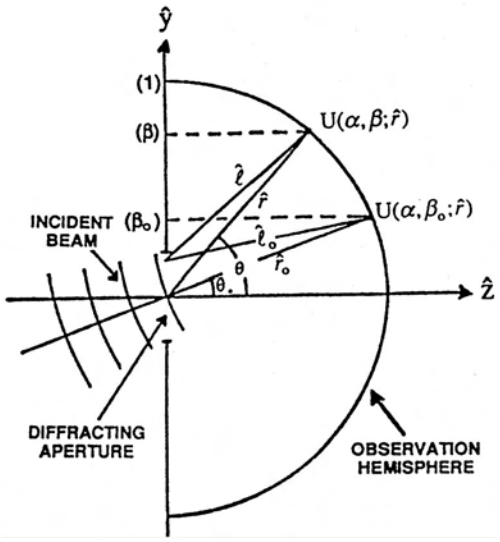

Fig. 2. Geometric configuration when the incident beam strikes the diffracting aperture at an arbitrary angle.

$$U_0'(\hat{x},\hat{y};0) = \sqrt{\gamma_i}U_0(\hat{x},\hat{y};0). \tag{9}$$

Here β is the direction cosine of the position vector of the observation point, and β_0 is the direction cosine of the position vector of the undiffracted beam. Note that the direction cosines are obtained by merely projecting the respective points on the hemisphere back onto the plane of the aperture and normalizing to a unit radius. The complex amplitude distribution at an arbitrary point on the hemisphere can now be said to be a function of the distance of the observation point from the undiffracted beam in direction cosine space. Furthermore, $\gamma = \cos\theta$ is just a cosine obliquity factor.

Hence nonparaxial diffraction phenomena have been shown to be linear, shift-invariant phenomena with respect to incident angle if we formulate the problem in terms of the direction cosines of the propagation vectors. Furthermore, for a uniformly illuminated diffracting aperture of area A_s, it is the diffracted radiance (not irradiance or intensity) that is shift invariant in direction cosine space[16]:

$$L(\alpha,\beta-\beta_0) = \frac{\lambda^2}{A_s}|\mathcal{F}\{U_0'(\hat{x},\hat{y};0)\exp(i2\pi\beta_0\hat{y})\}|^2. \tag{10}$$

Rayleigh's (Parseval's) theorem from Fourier-transform theory states that the integral over all space of the squared modulus of any function is equal to the integral over all space of the squared modulus of its Fourier transform.[10,16] Hence we can write

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|U_0'(\hat{x},\hat{y};0)\exp(i2\pi\beta_0\hat{y})|^2 d\hat{x}\, d\hat{y}$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|\mathcal{F}\{U_0'(\hat{x},\hat{y};0)\exp(i2\pi\beta_0\hat{y})\}|^2 d\alpha\, d\beta. \tag{11}$$

Substituting Eq. (10) into Eq. (11),

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|U_0'(\hat{x},\hat{y};0)\exp(i2\pi\beta_0\hat{y})|^2 d\hat{x}\, d\hat{y}$$

$$= \frac{A_s}{\lambda^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}L(\alpha,\beta-\beta_0)d\alpha\, d\beta. \tag{12}$$

Recall that only those plane-wave components that lie inside the unit circle in direction cosine space ($\alpha^2+\beta^2 \leq 1$) are real and propagate. Those that lie outside of the unit circle are imaginary and are referred to as evanescent waves (and thus do not propagate).[10,15,16] All (real) space is therefore represented by a unit circle in the two-dimensional direction cosine space. Hence all of the radiant power emanating from the diffracting aperture is contained in that portion of the diffracted radiance distribution function lying inside a unit circle in direction cosine space (the direction cosines of a vector must satisfy the equation $\alpha^2+\beta^2+\gamma^2=1$). Therefore Eq. (12) can also be written as

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|U_0'(\hat{x},\hat{y};0)\exp(i2\pi\beta_0\hat{y})|^2\mathrm{d}\hat{x}\ \mathrm{d}\hat{y} = \frac{A_s}{\lambda^2}\int_{-1}^{1}\int_{-\sqrt{1-\alpha^2}}^{\sqrt{1-\alpha^2}}L'(\alpha,\beta-\beta_0)\mathrm{d}\alpha\ \mathrm{d}\beta, \tag{13}$$

where we have used $L'(\alpha,\beta-\beta_0)$ to indicate the real diffracted radiance distribution that lies inside of the unit circle. Note that the left side of Eq. (13) is merely the integral of the radiant exitance over the (scaled) aperture. It is therefore proportional to the total radiant power transmitted through the diffracting aperture, which diminishes with the cosine of the incident angle.

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|U_0'(\hat{x},\hat{y};0)\exp(i2\pi\beta_0\hat{y})|^2\mathrm{d}\hat{x}\ \mathrm{d}\hat{y} = \frac{P_T(\theta_i)}{\lambda^2} = \frac{E_0A_s\gamma_i}{\lambda^2}. \tag{14}$$

Substituting Eqs. (14) into Eqs. (12) and (13), it is now evident that, for the case of a uniformly illuminated diffracting aperture, the total radiant power diffracted from the aperture is just the area of the aperture times the integral of the diffracted radiance:

$$P_T = A_s\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}L(\alpha,\beta-\beta_0)\mathrm{d}\alpha\ \mathrm{d}\beta = A_s\int_{-1}^{1}\int_{-\sqrt{1-\alpha^2}}^{\sqrt{1-\alpha^2}}L'(\alpha,\beta-\beta_0)\mathrm{d}\alpha\ \mathrm{d}\beta, \tag{15}$$

where $L(\alpha,\beta-\beta_0)$ is given by Eq. (10) and

$$L'(\alpha,\beta-\beta_0) = K\,L(\alpha,\beta-\beta_0). \tag{16}$$

The real radiance distribution function is thus merely a renormalized version of the original radiance distribution function $L(\alpha,\beta-\beta_0)$.[16] The renormalization constant K is given by the ratio of the integral of $L(\alpha,\beta-\beta_0)$ over infinite limits to the integral of $L(\alpha,\beta-\beta_0)$ over the unit circle in direction cosine space[16]:

$$K = \frac{\displaystyle\int_{\alpha=-\infty}^{\infty}\int_{\beta=-\infty}^{\infty}L(\alpha,\beta-\beta_0)\mathrm{d}\alpha\ \mathrm{d}\beta}{\displaystyle\int_{\alpha=-1}^{1}\int_{\beta=-\sqrt{1-\alpha^2}}^{\sqrt{1-\alpha^2}}L(\alpha,\beta-\beta_0)\mathrm{d}\alpha\ \mathrm{d}\beta}. \tag{17}$$

For those cases where the diffracted radiance distribution function given by Eq. (10) extends beyond the unit circle in direction cosine space, the real radiance distribution is given by

$$L'(\alpha,\beta-\beta_0) = K\frac{\lambda^2}{A_s}|\mathcal{F}\{U_0'(\hat{x},\hat{y};0)\exp(i2\pi\beta_0\hat{y})\}|^2 \quad \text{for } \alpha^2+\beta^2 \leq 1$$

$$L'(\alpha,\beta-\beta_0) = 0 \quad \text{for } \alpha^2+\beta^2 > 1. \tag{18}$$

The renormalization constant differs from unity only if the radiance distribution function extends beyond the unit circle in direction cosine space (i.e., only if evanescent waves are produced). The well-known Wood's anomalies that occur in diffraction grating efficiency measurements are entirely consistent with this predicted renormalization in the presence of evanescent waves.[22] These equations have also been applied to more general scattering surfaces and have successfully explained some otherwise nonintuitive surface scatter effects.[16] 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[10] and not by merely heuristically imposing the law of conservation of energy. The physics has not changed from the Rayleigh–Sommerfeld theory, but we have reformulated it into a Fourier treatment that can be easily solved for a wide variety of nonparaxial applications.

4. NONPARAXIAL (WIDE-ANGLE) BEHAVIOR OF SINUSOIDAL PHASE GRATINGS

There are substantial advantages of modeling diffraction grating behavior (especially for conical diffraction configurations with large obliquely incident beams) in terms of the direction consines of the propagation vectors of the diffracted orders and the incident beam.[23] In particular, use of direction cosine diagrams can be shown to simplify the analysis of the number and angular location of the various propagating orders for different grating periods and orientations. In addition, several qualitative aspects of the intensity distributions diffracted from gratings become apparent. These include the broadening and apparent shifting of diffracted orders at large diffraction angles.[16,24]

However, in this paper we will use the nonparaxial scalar diffraction theory summarized in Section 3 to make quantitative predictions of diffraction efficiency for perfectly conducting sinusoidal phase gratings. Hopefully, these results will dispel several common misconceptions that continue to persist due to the paraxial scalar prediction of Eq. (3). Two related misconceptions are (i) the notion that scalar theory imposes energy on the evanescent diffracted orders[5] and (ii) the prediction, from Eq. (3) and illustrated in Fig. 1, that it is impossible to obtain a diffraction efficiency greater than 0.3386 for the first diffracted order of a sinusoidal phase grating.

We could model either a uniformly illuminated grating of finite size or a small beam of some specific size and shape underfilling a large grating. In the first case the total radiant power being transmitted through the aperture is reduced by the $\cos\theta_i$, whereas in the second case the total transmitted radiant power remains the same but the beam footprint in the plane of the grating is increased by $\cos\theta_i$ and the exitance emerging from the plane of the grating is reduced by $\cos\theta_i$. Both cases should yield the same results for diffraction efficiency, which is defined as the ratio of the radiant power in a given diffracted order to the radiant power in the incident beam. We have chosen to analyze the first case to be consistent with Goodman's paraxial treatment.[10]

Using a symbolic notation for special functions popularized by Goodman[10] and Gaskill,[19] a square sinusoidal

phase grating (of finite size, $b\times b$) can be defined by the complex amplitude transmittance function

$$t(\hat{x},\hat{y}) = \mathrm{rect}\left(\frac{\hat{x}}{\hat{b}},\frac{\hat{y}}{\hat{b}}\right)\exp\left[i\frac{a}{2}\sin(2\pi\hat{y}/\hat{d})\right]. \tag{19}$$

Illuminating this grating with a uniform amplitude plane wave with an angle of incidence θ_i and using the following mathematical identity

$$\exp\left[i\frac{a}{2}\sin(2\pi\hat{y}/\hat{d})\right] = \sum_{m=-\infty}^{\infty} J_m\left(\frac{a}{2}\right)\exp(i2\pi m\hat{y}/\hat{d}), \tag{20}$$

we can now rewrite Eq. (10) as

$$L(\alpha,\beta-\beta_0) = \gamma_i E_0\frac{\lambda^2}{A_s}\left|\sum_{m=-\infty}^{\infty}\mathcal{F}\left\{\mathrm{rect}\left(\frac{\hat{x}}{\hat{b}},\frac{\hat{y}}{\hat{b}}\right)\times\exp(i2\pi\beta_i\hat{y})J_m\left(\frac{a}{2}\right)\exp(i2\pi m\hat{y}/\hat{d})\right\}\right|^2. \tag{21}$$

This is an expression for the diffracted radiance emanating from a sinusoidal phase grating with its grating lines (grooves) aligned parallel to the x axis when illuminated with a uniform amplitude plane wave whose propagation vector lies in the $y-z$ plane. We thus have planar diffraction (all of the diffracted orders lie in the $y-z$ plane) and the grating equation can be written as

$$\beta_m + \beta_i = m/\hat{d}, \tag{22}$$

where $\beta_m = \sin\theta_m$ and $\beta_i = \sin\theta_i$.

Applying the convolution theorem of Fourier-transform theory to Eq. (21), we obtain

$$L(\alpha,\beta-\beta_0) = \gamma_i E_0\frac{\lambda^2}{A_s}\left|\sum_{m=-\infty}^{\infty}\mathcal{F}\left\{\mathrm{rect}\left(\frac{\hat{x}}{\hat{b}},\frac{\hat{y}}{\hat{b}}\right)\exp(i2\pi\beta_i\hat{y})\right\} **\mathcal{F}\left\{J_m\left(\frac{a}{2}\right)\exp(i2\pi m\hat{y}/\hat{d})\right\}\right|^2, \tag{23}$$

where $**$ denotes a two-dimensional convolution operation.[19] Applying the shift theorem and the similarity theorem of Fourier-transform theory, and noting that the quantity $\exp(i2\pi m\hat{y}/\hat{d})$ merely Fourier transforms into a shifted delta function, $\delta(\alpha,\beta-m/\hat{d})$, we have

$$L(\alpha,\beta-\beta_0) = \gamma_i E_0\frac{\lambda^2}{A_s}\left|\sum_{m=-\infty}^{\infty}\frac{1}{1/\hat{b}^2}\mathrm{sinc}\left(\frac{\alpha}{1/\hat{b}},\frac{\beta+\beta_i}{1/\hat{b}}\right) **J_m\left(\frac{a}{2}\right)\delta(\alpha,\beta-m/\hat{d})\right|^2. \tag{24}$$

Since any function convolved with a delta function merely replicates that function at the location of the delta function,

$$L(\alpha,\beta-\beta_0) = \gamma_i E_0\frac{\lambda^2}{A_s}\left|\sum_{m=-\infty}^{\infty}J_m\left(\frac{a}{2}\right)\frac{1}{1/\hat{b}^2}\times\mathrm{sinc}\left(\frac{\alpha}{1/\hat{b}},\frac{\beta+\beta_i-m/\hat{d}}{1/\hat{b}}\right)\right|^2. \tag{25}$$

From the grating equation expressed in Eq. (22), the argument of the sinc function becomes

$$L(\alpha,\beta-\beta_0) = \gamma_i E_0\frac{\lambda^2}{A_s}\left|\sum_{m=-\infty}^{\infty}J_m\left(\frac{a}{2}\right)\frac{1}{1/\hat{b}^2}\mathrm{sinc}\left(\frac{\alpha}{1/\hat{b}},\frac{\beta-\beta_m}{1/\hat{b}}\right)\right|^2. \tag{26}$$

Since $\Delta\beta = 1/\hat{d} \gg 1/\hat{b}$, there is negligible overlap between the individual sinc functions; hence there are no cross terms in the squared modulus of the above summation. Factoring $\hat{b}^2$ outside of the summation sign and noting that $A_s = b^2$, we can write

$$L(\alpha,\beta-\beta_0) = \gamma_i E_0\sum_{m=-\infty}^{\infty}J_m^2\left(\frac{a}{2}\right)\left[\frac{1}{1/\hat{b}^2}\mathrm{sinc}^2\left(\frac{\alpha}{1/\hat{b}},\frac{\beta-\beta_m}{1/\hat{b}}\right)\right]. \tag{27}$$

Substituting Eq. (27) into Eq. (16) and then into Eq. (15) and interchanging the order of the summation and the integral operation, we have

$$P_T = \gamma_i E_0 K A_s\sum_{m=\min}^{\max}J_m^2\left(\frac{a}{2}\right)\int_{-1}^{1}\int_{-\sqrt{1-\alpha^2}}^{\sqrt{1-\alpha^2}}\times\left[\frac{1}{1/\hat{b}^2}\mathrm{sinc}^2\left(\frac{\alpha}{1/\hat{b}},\frac{\beta-\beta_m}{1/\hat{b}}\right)\right]\mathrm{d}\alpha\,\mathrm{d}\beta, \tag{28}$$

where the summation is taken only over the diffracted orders lying inside the unit circle in direction cosine space (i.e., the propagating diffracted orders). The quantity in square brackets is merely a unit volume sinc^2 function.[19] Since $\hat{b} \gg 1$, the sinc^2 function is very narrow and we will assume that, even for large diffracted angles, it lies completely inside the unit circle; hence the above integral is equal to unity and

$$P_T = \gamma_i E_0 K A_s\sum_{m=\min}^{\max}J_m^2\left(\frac{a}{2}\right). \tag{29}$$

The radiant power in a given diffracted order is thus clearly equal to

$$P_m = \gamma_i E_0 K A_s J_m^2\left(\frac{a}{2}\right), \tag{30}$$

and the diffraction efficiency of the mth diffracted order is given by

$$\eta_m = \frac{P_m}{P_T} = \frac{J_m^2\left(\frac{a}{2}\right)}{\sum_{m=\min}^{\max}J_m^2\left(\frac{a}{2}\right)}. \tag{31}$$

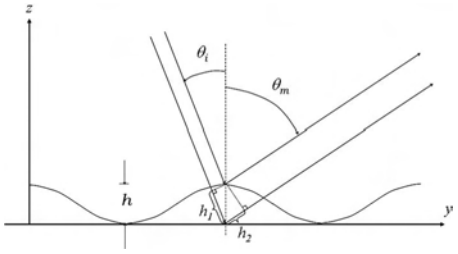

Fig. 3. Illustration of the peak-to-peak phase variation introduced into a given diffracted order by reflection from a sinusoidal surface.

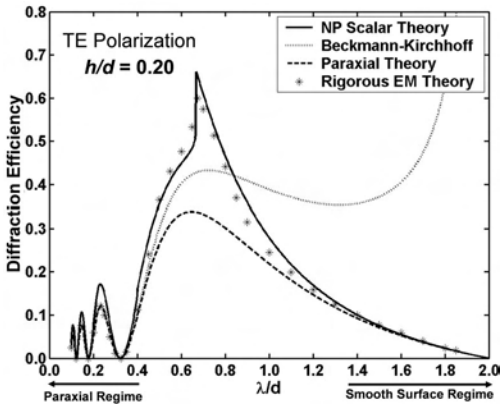

Fig. 4. Diffraction grating efficiency of the first order of a perfectly conducting sinusoidal phase grating ($h/d=0.20$) in the Littrow condition as predicted by the Beckmann–Kerchhoff theory, the paraxial scalar theory, the nonparaxial (NP) scalar diffraction theory presented in this paper, and a rigorous integral vector theory. EM, electromagnetic.

Note that the nonparaxial Eq. (31) for diffraction efficiency reduces to the paraxial Eq. (3) when the denominator is a summation from minus infinity to plus infinity. Recall that the quantity a in the argument of the above Bessel functions is the peak-to-peak phase variation introduced by the grating. Figure 3 illustrates that a varies not only with the angle of incidence, but with the diffracted order as well:

$$a = (2\pi/\lambda)(h_1 + h_2) = 2\pi\hat{h}(\cos\theta_i + \cos\theta_m). \qquad (32)$$

Equations (31) and (32) can now be used to calculate the diffraction efficiency of a perfectly conducting sinusoidal reflection grating. Figure 4 shows these predictions for the first order in the Littrow condition and compares them with the predictions shown in Fig. 1.

The above nonparaxial scalar diffraction theory provides remarkably good agreement with rigorous integral electromagnetic theory, not merely in the paraxial regime and the smooth surface (shallow grating) regime, but over the entire range of λ/d.

Similar predictions for sinusoidal reflection gratings at h/d values of 0.05, 0.15, and 0.30 are shown in Figs. 5–7, respectively. As expected, Fig. 5 shows that all of the theories agree well with rigorous calculations for low values of h/d. The rigorous data in Figs. 5–7 are taken from Ref. 1, p. 185.

Figure 6 not only shows good agreement between our nonparaxial scalar theory and rigorous calculations, it also demonstrates that our nonparaxial scalar diffraction

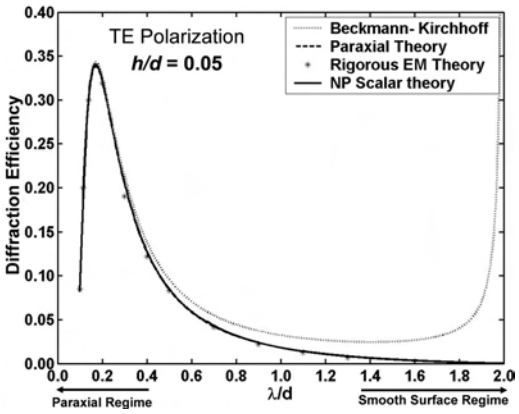

Fig. 5. Diffraction grating efficiency of the first order of a sinusoidal phase grating ($h/d=0.05$) in the Littrow condition as predicted by the Beckmann–Kerchhoff theory, the paraxial scalar theory, our nonparaxial (NP) scalar diffraction theory, and a rigorous integral vector theory. EM, electromagnetic.

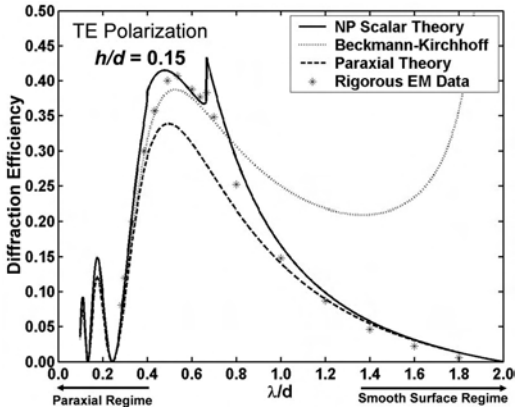

Fig. 6. Diffraction grating efficiency of the first order of a perfectly conducting sinusoidal phase grating ($h/d=0.15$) in the Littrow condition as predicted by the Beckmann–Kerchhoff theory, the paraxial scalar theory, our nonparaxial (NP) scalar diffraction theory, and a rigorous integral vector theory. EM, electromagnetic.

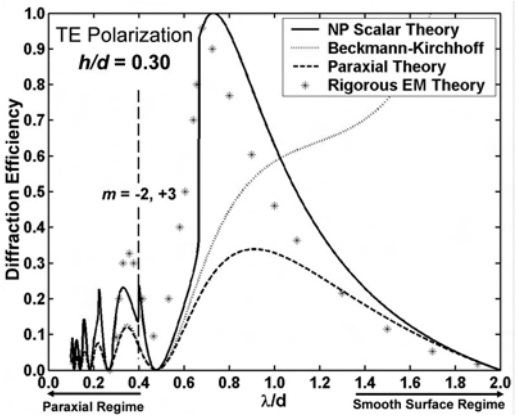

Fig. 7. Diffraction grating efficiency of the first order of a perfectly conducting sinusoidal phase grating ($h/d=0.30$) in the Littrow condition as predicted by the Beckmann–Kerchhoff theory, the paraxial scalar theory, our nonparaxial (NP) scalar diffraction theory, and a rigorous integral vector theory. EM, electromagnetic.

theory indeed predicts the Rayleigh anomalies[1,7,25] that occur when a propagating order goes evanescent. Note the abrupt increase in diffraction efficiency at $\lambda/d=0.67$. This is precisely the value of λ/d at which the −1 and +2 diffracted orders go evanescent.

Likewise, Fig. 7 shows that our nonparaxial scalar theory continues to predict the major features of the diffraction efficiency curve even for h/d values of 0.30. The diminishing agreement for $\lambda/d>0.6$ is probably due to the fact that our theory does not include shadowing and multiple-scattering effects. However, it does still predict a second Rayleigh anomaly at $\lambda/d=0.40$ where the −2 and +3 diffracted orders go evanescent.

Note that there is a substantial difference in the peak values of the oscillatory behavior in the paraxial regime as predicted by the paraxial and the nonparaxial scalar theories in Figs. 4, 6, and 7. The diffraction efficiency predicted for the nonparaxial scalar theory is given by Eq. (31). As λ/d becomes very small (paraxial regime), the number of diffracted orders becomes very large and the denominator approaches unity resulting in the well-known result given by $\eta_m=J_m^2(a/2)$. Clearly for a finite number of propagating orders the denominator is less than unity. This results in a higher value for the nonparaxial prediction. Our nonparaxial scalar theory assumes that the energy previously contained in the evanescent orders is distributed uniformly over the remaining propagating orders. This may or may not be an accurate assumption. Additional rigorous data are needed to quantitatively evaluate the deviation between the nonparaxial scalar theory and the rigorous theory in the paraxial regime.

5. SUMMARY AND CONCLUSIONS

Apparently much of the grating community believes that scalar diffraction theory is valid only for $\lambda/d<0.1$.[5–7] We believe 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. We have therefore summarized the development of a linear system's formulation of nonparaxial scalar diffraction theory that greatly expands the range of parameters over which accurate predictions can be made with simple Fourier techniques. This theory was then applied to the prediction of diffraction efficiency from perfectly conducting sinusoidal phase (holographic) gratings. These diffraction efficiency predictions were compared with those of the classical Beckmann–Kirchhoff theory, a simple paraxial theory, and a rigorous integral vector theory. The results for the diffraction efficiency of the +1 order in the Littrow condition were plotted in the usual manner as a function of λ/d for values from 0.1 to 2.0. Excellent agreement with rigorous electromagnetic theory was shown for h/d values up to 0.3. This is quite remarkable since this entire range ($0.1>\lambda/d>2.0$) lies outside of what is usually thought of as the scalar regime ($\lambda/d<0.1$). Furthermore, our nonparaxial scalar theory even predicts the Rayleigh anomalies[1,7,22] that are associated with the redistribution of energy from the evanescent orders to the remaining propagating orders. Hopefully, these results completely dispel the notion that scalar theory imposes energy upon the evanescent orders.[5] We are fully aware that there are two distinct types of efficiency anomalies: Rayleigh anomalies and resonance anomalies. The resonance anomalies are an electromagnetic effect that depends upon the optical constants of a specific material, and thus do require a rigorous vector theory (they should be nonexistent in the perfectly conducting gratings modeled in this paper). Also, we have only compared our scalar predictions with rigorous calculations for TE polarization. We will report at a later date on our attempt to quasi-vectorize our nonparaxial scalar theory to see if we can predict some of the anomalous TM behavior.

Corresponding author James E. Harvey's e-mail address is harvey@creol.ucf.edu.

REFERENCES

1. R. Petit, *Electromagnetic Theory of Gratings* (Springer-Verlag, 1980), p. 98.
2. P. Beckman and A. Spizzichino, *The Scattering of Electromagnetic Waves from Rough Surfaces* (Pergamon, 1963).
3. J. M. Bennett and L. Mattsson, *Introduction to Surface Roughness and Scattering*, 2nd ed. (Optical Society of America, 1999).
4. J. C. Stover, *Optical Scattering, Measurement and Analysis*, 2nd ed. (SPIE, 1995).
5. D. A. Gremaux and N. C. Gallager, "Limits of scalar diffraction theory for conducting gratings," Appl. Opt. **32**, 1048–1953 (1993).
6. D. A. Pommet, M. G. Moharam, and E. B. Grann, "Limits of scalar diffraction theory for diffractive phase elements," J. Opt. Soc. Am. A **11**, 1827–1834 (1994).
7. E. G. Loewen and E. Popov, *Diffraction Gratings and Applications* (Marcel Dekker, 1997).
8. S. D. Mellin and G. P. Nordin, "Limits of scalar diffraction theory and an iterative angular spectrum algorithm for finite aperture diffractive optical element design," Opt. Express **8**, 705–722 (2001).
9. D. Maystre, "Rigorous vector theories of diffraction gratings," in *Progress in Optics XXI*, E. Wolf, ed. (Elsevier Science, 1984).
10. J. W. Goodman, *Introduction to Fourier Optics* (McGraw-Hill, 1968).
11. M. Born and E. Wolf, *Principles of Optics* (Pergamon, 1980), p. 598.
12. T. K. Gaylord and M. G. Moharam, "Analysis and applications of optical diffraction by gratings," Proc. IEEE **73**, 894–937 (1985).
13. C. V. Raman and N. S. N. Nath, "The diffraction of light by high frequency sound waves," Proc. Ind. Acad. Sci. A **2**, 406–413 (1935).
14. C. Palmer, *Diffraction Grating Handbook*, 4th ed. (Richardson Grating Laboratory, 2000), p. 15.
15. J. E. Harvey, "Fourier treatment of near-field scalar diffraction theory," Am. J. Phys. **47**, 974–980 (1979).
16. J. E. Harvey, C. L. Vernold, A. Krywonos, and P. L. Thompson, "Diffracted radiance: a fundamental quantity in a nonparaxial scalar diffraction theory," Appl. Opt. **38**, 6469–6481 (1999).
17. J. E. Harvey, C. L. Vernold, A. Krywonos, and P. L. Thompson, "Diffracted radiance: a fundamental quantity in a nonparaxial scalar diffraction theory: errata," Appl. Opt. **39**, 6374–6375 (2000).
18. J. A. Ratcliff, "Some aspects of diffraction theory and their application to the ionosphere," in *Reports on Progress in Physics*, A. C. Strickland, ed. (Physical Society, 1956), Vol. XIX.
19. J. D. Gaskill, *Linear Systems, Fourier Transforms, and Optics* (Wiley, 1978).

20. J. E. Harvey and R. V. Shack, "Aberrations of diffracted wave fields," Appl. Opt. **17**, 3003–3009 (1978).
21. J. E. Harvey, A. Krywonos, and D. Bogunovic, "Tolerance on defocus precisely locates the far field (exactly where is that far field anyway?)" Appl. Opt. **41**, 2586–2588 (2002).
22. R. W. Wood, "On a remarkable case of uneven distribution of light in a diffraction grating spectrum," Philos. Mag. **4**, 396–410 (1902).
23. J. E. Harvey and C. L. Vernold, "Description of diffraction grating behavior in direction cosine space," Appl. Opt. **37**, 8158–8160 (1998).
24. J. E. Harvey and E. A. Nevis, "Angular grating anomalies: effects of finite beam size upon wide-angle diffraction phenomena," Appl. Opt. **31**, 6783–6788 (1992).
25. A. Hessel and A. A. Oliner, "A new theory of Wood's anomalies on optical gratings," Appl. Opt. **4**, 1275–1297 (1965).