Diffraction by volume gratings: approximate solution in terms of boundary diffraction coefficients

J. T. Sheridan* and L. Solymar

Department of Engineering Science, University of Oxford, Parks Road, Oxford OX1 3PJ, UK

Received July 17, 1991; revised manuscript received February 14, 1992; accepted March 13, 1992

An approximate analytic method is presented that permits the calculation of all the output waves, both backward and forward traveling, for slanted, transmission-type volume gratings replayed at the Bragg angle and perfectly index matched to their surroundings. The predictions of the resulting analytic expressions are compared with numerical results produced by using the rigorous coupled-wave method and found to be accurate over a large range of parameters to the first order of the permittivity variation of the volume grating, all second-order terms being assumed negligible.

1. INTRODUCTION

It is well known that coupled-wave differential equations are eminently suitable for solving diffraction problems in the theory of volume gratings. There is a basic distinction between approximate theories that need only the first derivatives of the field quantities in the coupled differential equations and rigorous theories that retain the second derivatives as well (for a recent review, see Ref. 1). In the former approach only the electric fields are matched at the boundaries. It is therefore unsuitable for finding the amplitudes of the waves reflected by the grating. If the aim is to find the amplitudes of such waves, one needs to resort to one of the rigorous theories that require extensive numerical calculations; they have the further disadvantage that the associated physical picture is rather complicated. The question arises whether one could devise an approximate theory that gives the amplitudes of such beams at least in the first order of the grating amplitude. The magnitudes of such small spurious beams are of great interest in the design of some holographic devices, e.g., head-up displays. An approximate first-order theory based on perturbing the Fresnel reflection coefficients was recently presented for unslanted transmission gratings. It was also shown that the results of the approximate theory agreed for a wide range of parameters with those based on the rigorous theory of Gaylord and Moharam. The method of perturbing the Fresnel coefficients implies that the boundary diffraction is caused by the periodic variation of the dielectric constant along the boundary. Hence it has no knowledge of what is happening inside the grating, in medium II, and cannot therefore account for grating slant. The aim of the present paper is to extend the ideas discussed in Ref. 3 so as to derive an approximate method for the calculation of diffraction by slanted gratings. For simplicity we restrict the investigation to those slanted gratings in which the interaction of the backward-traveling waves with the volume grating can be ignored. We believe, however, that the method described has much more general validity. Its essence is that it separates boundary and volume diffraction in a manner reminiscent of the geometrical theory of diffraction.

2. FORMULATION OF THE PROBLEM

The basic configuration investigated is shown in Fig. 1. It is treated as a two-dimensional problem, i.e., no variation is assumed in the y direction. The average dielectric constant of all three media \( \varepsilon_0 \) is equal to unity, but medium II contains a phase grating characterized by the grating vector \( \mathbf{K} \) and the dielectric constant variation:

\[
\varepsilon_{II} = 1 + \varepsilon_m \cos(K \cdot r) = 1 + \varepsilon_m \cos(K_x x + K_z z), \tag{1}
\]

where \( r = x \hat{x} + z \hat{z} \) and \( \hat{x} \) and \( \hat{z} \) are unit vectors in the x and the z directions, respectively. Note that \( \varepsilon_m \), the dielectric modulation, is assumed to be much smaller than unity, which is nearly always the practical situation. Wave 0 with an electric polarization in the y direction is incident at an angle \( \theta_0 \) from medium I upon medium II. It is assumed to satisfy the Bragg condition and to give rise to a diffracted wave in the direction \( \theta_1 \). The corresponding Ewald diagram is shown in Fig. 2, where

\[
\rho_0 = \rho_{0x} \hat{x} + \rho_{0z} \hat{z}, \quad \rho_1 = \rho_{1x} \hat{x} + \rho_{1z} \hat{z}, \tag{2}
\]

are the wave vectors of the two beams. According to the rigorous theory there will also be reflected waves in medium I corresponding to the reflections of waves 0 and 1.

The waves mentioned above may be described mathematically as follows:

Incident wave:

\[
A_{0,\text{in}} \exp[-j(\rho_{0x} x + \rho_{0z} z)]. \tag{3}
\]

Reflected waves:

\[
R_{l,\text{in}} \exp[-j(\rho_{0x} x - \rho_{0z} z)], \tag{4}
\]

\[
R_{l,\text{out}} \exp[-j(\rho_{1x} x - \rho_{1z} z)]. \tag{5}
\]

The waves transmitted into medium III are

\[
A_{0,\text{out}} \exp[-j(\rho_{0x} x + \rho_{0z} z)], \tag{6}
\]

\[
A_{1,\text{out}} \exp[-j(\rho_{1x} x + \rho_{1z} z)]. \tag{7}
\]
Next we need the tangential component of the magnetic field, which may be obtained from one of Maxwell's equations as

$$H_e = \frac{1}{j\omega \mu} \frac{dE_x}{dz},$$

where \(\omega\) is the angular frequency of the input wave and \(\mu\) is the permeability of the medium, which is assumed to be equal to the free-space value. We may then write \(H_e\) in medium I as

$$H_e^I = \frac{1}{j\omega \mu} (-\rho_0 \exp(-j\rho_0 \cdot r) + \rho_0 R_0^I \times \exp[-j(\rho_0 x - \rho_0 z)] + \rho_1 R_1^I \times \exp[-j(\rho_1 x - \rho_1 z)])$$

and in medium II as

$$H_e^II = \frac{1}{j\omega \mu} \left[ \exp(-j\rho_0 \cdot r) \left( \frac{dA_0}{dz} - j\rho_0 A_0 \right) \right.$$ \left. + \exp(-j\rho_1 \cdot r) \left( \frac{dA_1}{dz} - j\rho_1 A_1 \right) \right].$$

Matching \(H_e\) on the boundary, \(z = 0\), we obtain the equations

$$-j\rho_0 (1 - R_0^I) = \frac{dA_0}{dz} \bigg|_{z=0} - j\rho_0 A_0(0),$$

$$+j\rho_1 R_1^I = \frac{dA_1}{dz} \bigg|_{z=0} - j\rho_1 A_1(0).$$

There is now a little complication owing to the appearance of the derivatives of the field quantities. They can, however, be eliminated with the aid of the well-known first-order differential equations of approximate coupled-wave theory (valid in medium II when \(\varepsilon_m \ll 1\), in which our notation take the form for illumination at the Bragg angle\(^6\):

$$\cos \theta_0 \frac{dA_0}{dz} + j \frac{\varepsilon_m \beta}{4} A_1 = 0,$$

$$\cos \theta_1 \frac{dA_1}{dz} + j \frac{\varepsilon_m \beta}{4} A_0 = 0,$$

where \(\beta = |\rho_0| = |\rho_1|\).

Expressing the derivatives of \(A_0\) and \(A_1\) from the above equations and substituting them into Eqs. (13) and (14),

$$1 + R_0^I = A_0(0),$$

$$R_1^I = A_1(0).$$

Fig. 1. Schematic representation of the input and the output waves.

Fig. 2. Ewald diagram showing the wave vectors of the two forward-traveling beams.

3. CONCEPT OF BOUNDARY DIFFRACTION COEFFICIENTS AND THEIR DERIVATION

Since rigorous solutions of Maxwell's equations are usually complicated and often just unattainable, it is desirable to introduce some approximate method that can combine physical intuition with mathematical simplicity. One such method is known as the geometrical theory of diffraction,\(^5\) in which geometrical optics is supplemented by assigning diffraction coefficients to various types of discontinuity. In our case this means that the general problem of grating diffraction is divided into two parts, boundary diffraction and volume diffraction, and both are treated by approximate methods.

What we mean by boundary diffraction is shown in Fig. 3(a), where wave 0 is incident upon the I-II boundary and gives rise to two forward-diffracted waves and two backward-diffracted (or reflected) waves. The forward-diffracted waves are written in the usual form:

$$A_0(z) \exp(-j\rho_0 \cdot r), \quad A_1(z) \exp(-j\rho_1 \cdot r),$$

i.e., their amplitudes will be expected to vary in the \(z\) direction. For the moment we are interested only in their values at the boundaries, i.e., \(A_0(0)\) and \(A_1(0)\). Thus the immediate problem is to find the four diffracted quantities \(R_0^I, R_1^I, A_0(0),\) and \(A_1(0)\) in terms of the input quantity \(A_{0,in}\) (which is usually taken to be of unit amplitude). We can find the necessary relations by matching the tangential components of the electric and the magnetic fields at the boundary, \(z = 0\). It may be easily seen from expressions (3)–(5) that the matching of the electric field yields the equations

$$1 + R_0^I = A_0(0),$$

$$R_1^I = A_1(0).$$

Fig. 3. (a) Waves that can be generated by the \(z = 0\) boundary with the \(A_{0,in}\) wave incident. (b) Reflections from the I–II and the II–III boundaries.
respectively, we obtain
\[\rho_{01}(1 - R_i') = \frac{\varepsilon_m \beta}{4 \cos \theta_0} A_1(0) + \rho_{01} A_0(0), \quad (17)\]
\[\rho_{11} R_i' = -\frac{\varepsilon_m \beta}{4 \cos \theta_1} A_0(0) - \rho_{11} A_1(0). \quad (18)\]

We now have four equations and four unknowns yielding the solutions
\[A_0(0) = A_{0, in} = 1, \quad (19)\]
\[R_i' = 0, \quad (20)\]
\[R_i' = A_1(0) = D_{10} A_{1, in} = D_{10}, \quad (21)\]
where
\[D_{10} = -\frac{\varepsilon_m}{8 \cos^2 \theta_1}. \quad (22)\]

It is not particularly surprising that \(A_0(0) = A_{0, in}\); since the boundary represents only a first-order discontinuity it may be expected that the transmitted wave goes straight through. It may also be expected that \(R_i' = 0\) since the average dielectric constants are the same in both media. The interesting result is the amount of diffraction into the forward- and the backward-diffracted beams that depends linearly on the dielectric constant modulation (as it should) and otherwise depends only on \(\theta_1\), the diffracting angle of wave 1.

If instead of wave 0 it is wave 1 that is incident, we can go through the same calculation, which yields
\[A_1(0) = A_{1, in}, \quad (23)\]
\[R_i' = 0, \quad (24)\]
\[R_i' = A_0(0) = D_{01} A_{1, in}, \quad (25)\]
where \(A_{1, in}\) is the amplitude of the incident wave and
\[D_{01} = -\frac{\varepsilon_m}{8 \cos^2 \theta_0}. \quad (26)\]

Note that the diffraction coefficient from wave 1 into wave 0 has the same expression as before but depends now on \(\theta_0\), the diffracting angle of wave 0.

We may now ask what will happen on the II-III boundary. According to our physical picture, that would also cause boundary diffraction. There is no need, however, to derive again the boundary diffraction coefficients. They may be obtained by a simple physical argument illustrated by Fig. 3. At the I-II boundary the input wave \(A_{0, in}\) gives rise, for example, to the backward-diffracted wave \(R_i'\). We may similarly expect wave 0 at the II-III boundary to give rise to a backward-diffracted wave in the same direction with a certain diffraction coefficient. Let us now assume that \(d\), the thickness of the grating, goes to zero, i.e., the boundary disappears. Consequently there should not be any backward diffraction, which can happen only if the diffraction coefficients at the two boundaries are of opposite signs. That also enables us to find the diffraction coefficients at the I-II boundary when the wave is incident from medium II.

We are now in a position to describe completely the role of the boundary in causing diffraction by relating all output waves to all input waves with the aid of a boundary diffraction matrix.

For the I-II boundary we find
\[\begin{bmatrix} A_0(0) \\ A_1(0) \end{bmatrix} = \begin{bmatrix} 1 & D_{01} \\ 0 & D_{01} \end{bmatrix} \begin{bmatrix} A_{0, in} \\ A_{1, in} \end{bmatrix}, \quad (27)\]
where the backward-traveling waves 0 and 1 in medium II are written in the form
\[R_0''(z)\exp[-j(\rho_{01} x - \rho_{01} z)], \quad R_1''(z)\exp[-j(\rho_{12} x - \rho_{12} z)] \quad (28)\]
so that they are incident upon the I-II boundary with amplitudes \(R_0''(0)\) and \(R_1''(0)\), respectively.

We may similarly obtain for the boundary diffraction matrix of boundary II-III
\[\begin{bmatrix} A_{0, out} \\ A_{1, out} \end{bmatrix} = \begin{bmatrix} 1 & -D_{01} \\ 0 & -D_{01} \end{bmatrix} \begin{bmatrix} A_0(d) \\ A_1(d) \end{bmatrix}, \quad (29)\]
where we have disregarded any waves incident from medium III since they are of no direct interest in our problem.

4. CALCULATION OF THE OUTPUT BEAMS

Let us now assume that wave 0 is incident from medium I upon medium II with wave vector \(\rho_0\) and amplitude \(A_{0, in}\) and calculate the amplitudes of all the output beams by using our method. For determining \(A_0(d)\) and \(A_1(d)\) we can just use the differential equations [Eqs. (15) and (16)] valid in medium II. Since both \(A_0(d)\) and \(A_1(d)\) are, in general, zero-order quantities, we can neglect the two boundary diffractions and obtain the well-known results

\[A_0(d) = A_{0, in} \cos \nu, \quad A_1(d) = -j A_{0, in} \left(\frac{\cos \theta_0}{\cos \theta_1}\right)^{1/2} \sin \nu, \quad (30)\]

Fig. 4. Waves traveling inside the volume grating. The \(R_i'\) backward-traveling wave has only one component produced by the \(z = d\) boundary. The \(R_i''\) beam has two contributions from the opposite boundaries.
Z = O B z = d

Fig. 5. Path difference of wave 1 accumulated by traversing twice the volume grating.

where

\[ \nu = \frac{\beta \varepsilon_m d}{4(\cos \theta_1 \cos \theta_0)^{1/2}}. \]

Let us now see what happens according to our physical picture as wave 0 is incident from medium I upon medium II at an angle \( \theta_0 \). At boundary I-II the reflected wave \( R_0 \) will be zero because the average dielectric constants are equal. The backward-diffracted \( R_1' \) wave will, however, be present, as shown in Fig. 4. We know already that the transmitted wave and the forward-diffracted wave will appear with amplitudes \( A_0(d) \) and \( A_1(d) \) at the II-III boundary, where again there will be no direct reflections. There will, however, be backward diffraction resulting in the waves \( R_0' \) and \( R_1' \), as shown again in Fig. 4. Note that both waves are off Bragg. For a slightly slanted grating they will be only slightly off Bragg, so they will still interact with the volume grating. For simplicity we assume, however, that in the cases to be investigated the slant angle is sufficiently large so that both waves will be strongly off Bragg and the interaction with the volume grating can be ignored. They will diffract at the I-II boundary, but that can be disregarded since that leads to second-order quantities. It is clear then from Fig. 4 that there will be only one contribution to \( R_0' \) but two contributions to \( R_1' \). How could the two contributions be combined? What is the right phase? For that we need to find the path-length difference between the waves in the geometry shown in Fig. 5.

We consider an input ray incident upon point O, which is the end point of a grating fringe on the first boundary. The same fringe intersects the second boundary at Q, i.e., \( OQ \) is the line of a grating fringe. The wave front of the incident wave is denoted by the dashed line \( AB \). The \( R_1' \) wave is represented by a ray originating at point O, and its wave front is denoted by the dashed line \( CD \). The input wave then travels a distance \( a \) from the \( AB \) wave front to the \( PQ \) wave front. We now regard the wave diffracted backward by the second boundary as originating from Q, i.e., we represent it by a ray drawn from Q. This ray intersects the \( CD \) line (that is, the wave front of the spurious wave diffracted backward by the boundary) at point F. Denoting the distance \( QF \) by \( b \), we have now obtained the path difference between the rays diffracted by the first and the second boundaries. It is equal to \( a + b \). From the geometry of Fig. 5 we may obtain

\[ a + b = 2d \cos \theta_1. \]

We are now in a position to find the amplitude of wave 1 in medium I by combining the contributions from the two boundaries. The contributions from the first boundary is \( D_{00} A_{00,1} \). The amplitude of the transmitted beam at the second boundary is \( A_{01,1} \) from Eqs. (30). This is now diffracted backward with a diffraction coefficient \( -D_{01} \) in agreement with the discussion in Section 3. It will combine with the diffracted wave from the first boundary with a phase difference of \( 2\beta d \cos \theta_1 \), resulting in

\[ R_{1'}^I = -\frac{e_m}{8 \cos^2 \theta_1} [1 - \cos(\nu) \exp(-j2\beta d \cos \theta_1)], \]

and the formula for \( R_{0'}^I \) may be similarly obtained as

\[ R_{0'}^I = -\frac{e_m}{8 \cos^2 \theta_0} \sin(\nu) \exp(-j2\beta d \cos \theta_0). \]

Fig. 6. Intensities of the two backward-traveling spurious waves (a) \( |R_0'/A_{01,1}|^2 \) and (b) \( |R_1'/A_{01,1}|^2 \) plotted against \( d/\lambda \), the grating thickness in wavelengths. The grating recording waves are incident at \( \theta_0 = -20^\circ \) and \( \theta_1 = +80^\circ \). The grating is replayed on Bragg at \( \theta_0 = -20^\circ \). \( e_m = 0.02 \). Solid curves: rigorous theory; dashed curves: approximate theory.
5. RESULTS

In the results presented, all the gratings are assumed to be replayed on Bragg, and the variations of the two backward-traveling spurious wave intensities as a function of grating thickness are presented. In all cases both the rigorous coupled-wave result as obtained from the rigorous theory of Gaylord and Moharam (solid curves) and the corresponding approximate analytic result (dashed curves) are shown on the same graph.

The grating is assumed to be strongly slanted. In our first example we take the recording waves incident at $\theta_0 = -20^\circ$ and $\theta_1 = +80^\circ$. For a sufficiently small value of $\varepsilon_m$ it must be true that the first-order approximate theory gives the same result as the rigorous one. It may indeed be seen in Fig. 6, where $|R_0/A_{0,\text{in}}|^2$ and $|R_1/A_{0,\text{in}}|^2$ are plotted against $d/\lambda$ for an input beam at $-20^\circ$, that there is good agreement for $\varepsilon_m = 0.02$. As the modulation increases the agreement may be expected to deteriorate. This is clearly shown in Figs. 7 and 8, where the same quantities are plotted for the same set of parameters for $\varepsilon_m = 0.05$ and 0.1, respectively. A modulation of 10% is of course very large and not normally achievable in a practical situation. But even in that case both the order of magnitude and the rough shape of the curves are correctly predicted. If the slant is less, i.e., the recording angles are less asymmetrical, we may expect some improvement in the validity of the approximate theory. Choosing $\theta_0 = -20^\circ$ and $\theta_1 = 50^\circ$ for the recording angles and replaying again at $\theta_0 = -20^\circ$, we obtained the exact and the approximate curves shown in Figs. 9(a) and 9(b). Comparing them with the curves displayed in Figs. 8(a) and 8(b), respectively, it is obvious that we have improved agreement.

We would get similar results if the replay angle were chosen at the other Bragg angle, $\theta_1$. In fact we investigated a large number of cases, and up to a modulation of $\varepsilon_m = 0.1$ the agreement was always reasonable. The assumption that the reflected beams are far off Bragg is also not necessary for the validity of the approximate method. If that assumption is abandoned the approximate formulas are a little more complicated because the waves reflected at the second boundary interact with the volume grating. But this can be accounted for by using a first-order theory.

Our approximate formulas were derived in a heuristic manner. The only proof for their validity is that in the limit of small modulation the approximate results tend to agree with the exact ones. Independent proof for a slanted grating is not possible since the only methodology
diffraction is separated into a boundary diffraction and a volume diffraction problem relying on first-order approximations. The resulting physical picture is an attractive one, and the approximate formulas derived are extremely simple. We have shown the approximations to be valid for a large range of parameters by comparing them with numerical results derived from the rigorous theory of Moharam and Gaylord. It is our belief that the concepts introduced in Ref. 3 and in the present paper are of general validity. We believe that for any practical case, when \( \varepsilon_m \) is smaller than 0.1, the separation of the problem into boundary and volume diffraction should always be possible. Higher diffraction orders could be included by deriving boundary diffraction coefficients for the higher orders and then using coupled-wave theory for diffraction within the grating. We hope to produce further evidence in favor of this conjecture.

**ACKNOWLEDGMENTS**

The authors acknowledge the funding provided for this research by the Sciences and Engineering Research Council and Pilkingtons PLC. We thank A. Ramsbottom of Pilkingtons PLC for several stimulating discussions. J. T. Sheridan currently holds an Alexander von Humboldt Fellowship in the Department of Physics of the University of Erlangen–Nuremberg, Erlangen, Germany.

*Present address, Lehrstuhl Angewandte Optik, Physikalisches Institut der Universität Erlangen–Nürnberg, Staudstrasse 7/B2, D-8520 Erlangen, Germany.

**REFERENCES**