Notes regarding and rederivation of Knop's square grating results for the TE polarization

M. Sadgrove
Center for Photonic Innovations, The University of Electro-Communications, Chofu
(Dated: September 1, 2014)

In [*J. Opt. Soc. Am. 68, 1206 (1978)*] K. Knop applied rigourous diffraction theory to square gratings of arbitrary depth and period to derive the scattered field of a grating in both near and far-field regimes. The present note gives a detailed reworking of Knop's derivation which clarifies some amibiguities and also uses symbols and a setting in line with [*Optics Letters, 38, 2542 (2013)*]. The aim is to clarify Knop's formulation to allow the calculation of the near-field optical trapping potential.

INTRODUCTION

The term "rigorous diffraction theory" (RDT) implies the calculation of the diffracted field from some object in which no or very few assumptions are made. Typically it is difficult to find an analytical solution to Maxwell's equations in such cases, which is why solutions with limited ranges of validity, such as Kirchoff's diffraction law, are commonly used.

Here, our motivation is to assess the potential created by the *near field* diffraction pattern due to a dielectric square grating. The diffraction pattern is further modified by a nanofiber mounted on the grating (the presence of the nanofiber is not considered in the present note) and in principle, atoms may be trapped only hundreds of nanometers from the surface of the nanofiber and / or the grating. Thus, methods such us Kirchoff's which rely on the field being calculated at a distance L from the grating much larger than the incident light wavelength λ cannot be used.

Instead, the aforementioned rigorous diffraction theory approach must be used. The RDT solution in the case of a square grating was already calculated by Knop in Ref. [1] in 1978. Note that Knop's derivation itself is a relatively straightforward extension of Burckhardt's rigorous diffraction theory for a sinusoidal grating [2]. The motivations for reviewing Knop's derivation, confined to the simplest case of TE polarization, are the following:

- Knop's paper [1] is essentially a letter, and the important derivations take up only a page. Thus they do not provide a good introduction to rigorous diffraction theory.
- One of the most important equations in Knop's paper, Eq.(17), is rendered ambiguous by a typesetting error. Although it is possible to guess the correct form of Knop's equation, this is somewhat unsatisfactory as the basis of further research.
- Knop's result was formulated before the personal computer era and well before the advent of sophisticated, matrix-based numerical programming languages such as Matlab. Thus, although his derivation is naturally expressed in matrix form, he gives his final expressions, such as the aforementioned Eq.(17), in matrix-element form, no doubt to aid calculation using the 8kB memory HP calculator that was used to produce his results! In the rederivation, we keep the results in matrix form, which allows more rapid application of the results in Matlab or similar matrix enabled analysis software.

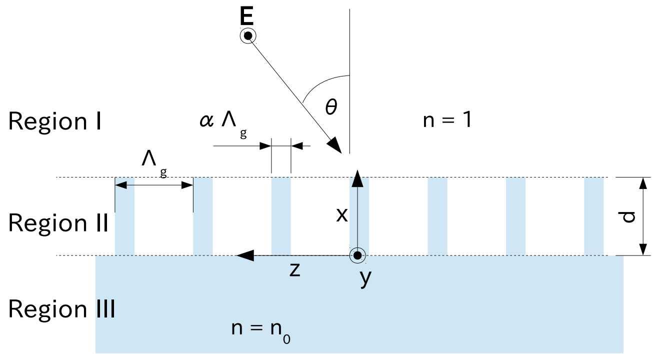

FIG. 1. Depiction of the difraction grating and incident plane wave.

REVIEW OF KNOP'S DERIVATION FOR THE TE POLARIZATION

The system considered in the derivation is shown in Fig. 1. Compared with [1], we have swapped the x and z axes so they correspond to the typical notation used at the Center for Photonic Innovations. Additionally, we use notation for the grating parameters taken from [3], where Λ_g is the grating period, α is the grating duty cycle, and d is the grating depth. We will analyse the problem for TE polarized light (following Knop, this indicates that the electric field is *parallel* to the grating slats) incident from above the grating (i.e. on the air side). In practice, we will wish to solve the problem for light incident from the grating dielectric side. However, it is trivial to switch to this incidence condition once the problem has been solved, so for simplicity, we use the same incidence condition as Knop.

Definition of grating refractive index

Assuming that there is a slat centered at the origin, a single period of the grating defined for this central slat is given by:

$$n(z) = \begin{cases} 1, & \alpha\Lambda_g < 2|z| \leq \Lambda_g \\ n_0, & 0 \leq 2|z| \leq \alpha\Lambda_g \end{cases} \quad (1)$$

The grating is assumed to be infinite in extent with $n(z) = n(z + \Lambda_g)$ defining its periodicity in the z direction.

General solution (regions I and III)

Since we are solving for the TE polarization, as defined above, only the E_x component of the electric field vector is non-zero. We can therefore identify three non-zero electromagnetic field components:

$$E_y \equiv E(x, z) \quad (2)$$

$$H_x = \frac{\mathrm{i}}{k_0}\frac{\partial E}{\partial z} \quad (3)$$

$$H_z = -\frac{\mathrm{i}}{k_0}\frac{\partial E}{\partial x}, \quad (4)$$

where a time dependence $\exp(i\omega t)$ is assumed for all fields, but left out of the notation, and $k_0 = 2\pi/\lambda$ is the wave number of the incident light. In addition, the wave equation must hold:

$$\frac{\partial^2 E}{\partial^2 x} + \frac{\partial^2 E}{\partial^2 z} + n(x, z)^2 k_0^2 E = 0. \quad (5)$$

Solutions to the wave equation 5 in region I (in the air above the grating) and region III (in the substrate material of the grating) are straightforward plane waves. In particular, we have the general solution in these regions:

$$E(x, z) = \exp[\mathrm{i}(pz + qx)]. \quad (6)$$

Substitution of (6) into (5) reveals that

$$p^2 + q^2 = n^2 k_0^2. \quad (7)$$

To proceed, we constrain the solution family given by Eqs. (6) and (7) by imposing a reasonable physical assumption. The assumption is the following:

Assumption 1: The solution should be periodic with a base (fundmental) period of Λ_g.

I do not attempt to justify this assumption, but note that it is reasonable given the assumption of a perfectly periodic grating of infinite extent, and that the assumption should be equivalent to imposing Bloch boundary conditions on the solution for the field.

The simplest way to guarantee Assumption 1 is to require that p, the wavenumber of the solution along the direction perpendicular to the grating slats, is an integer multiple of the grating spatial frequency $1/\Lambda_g$. In the general case of a non-zero incident angle θ, we also add a constant term to the wavenumber giving

$$p \to p_\ell = 2\pi\ell/\Lambda_g + k_0 \sin\theta. \quad (8)$$

Then, Eq.(7) gives

$$q \to q_\ell = \sqrt{n^2 k_0^2 - p_l^2}. \quad (9)$$

SOLUTIONS IN REGIONS I AND III

We now consider more specific solutions for the fields in regions I and III.

Firstly, the field in region I may be considered to be a combination of the incident field $\exp[\mathrm{i}(p_0 z - q_0 x)]$ and the reflected field from the grating. In particular, the field will have the form

$$E_I(x, z) = \exp[\mathrm{i}(p_0 z - r_0 x)] + \sum_\ell R_\ell \exp[\mathrm{i}(p_\ell z + r_\ell x)], \quad (10)$$

where the R_ℓ terms are the reflection coefficients for the ℓth partial wave in the solution, and the r_ℓ terms are the wavenumbers in the y direction in air. From Eq. (9), we find that

$$r_\ell = \begin{cases} \sqrt{k_0^2 - p_\ell^2} & k_0 \geq |p_\ell| \\ \mathrm{i}\sqrt{p_\ell^2 - k_0^2} & k_0 \leq |p_\ell| \end{cases} \quad (11)$$

Similarly, in region III we have

$$E_{III}(x, z) = \sum_\ell T_\ell \exp[\mathrm{i}(p_\ell z - t_\ell x)], \quad (12)$$

where the T_ℓ terms are the transmission coefficients for the ℓth partial wave in the solution, and the t_ℓ terms are the wavenumbers in the y direction in the substrate material.From Eq. (9), we find that

$$t_\ell = \begin{cases} \sqrt{n_0^2 k_0^2 - p_\ell^2} & n_0 k_0 \geq |p_\ell| \\ \mathrm{i}\sqrt{p_\ell^2 - n_0^2 k_0^2} & n_0 k_0 \leq |p_\ell| \end{cases} \quad (13)$$

Notes on the formulation of the problem

Although the above working is essentially the formulation of the problem rather than its full solution, we can already see some important facts about the solution from Eqs.(10) through (13).

Firstly, we note that we are interested in the *transmitted* field given by Eq. (12) for our purposes. In this case we note that for diffracted orders (i.e. partial waves in the solution with $\ell > 0$) to be transmitted, and thus for there to be a diffraction pattern created, we must have $n_0 k_0 \geq |p_\ell|$ by Eq (13). For normally incident light, this implies that

$$\lambda \leq n_0 \Lambda_g \quad (14)$$

must be satisfied for first order diffraction to occur. This is important for the research at at the Center for Photonic

Innovations because the grating patterns we typically work with have *sub-wavelength* periods in order to implement Bragg mirrors. In particular, it may be difficult to manufacture on-fiber gratings using the femto-second ablation technique [4] for periods of order $\Lambda_g \sim \lambda$. On the other hand, it should not be difficult to manufacture an external grating which satisfies this creterion.

SOLUTION INSIDE THE GRATING (REGION II)

We now turn to the solution in region II which is non-trivial. Here the general solution (6) is still valid, as is Assumption 1, but because the index of refraction is a function of z, the simple relation given by Eq. (7) cannot be used to find the wavenumber in the x direction.

We first introduce a preliminary solution as follows:

$$e_{II}(x,z) = \sum_{\ell} E_\ell \exp[\mathrm{i}(p_\ell z + gx)], \tag{15}$$

where E_ℓ is the coefficient of the ℓth partial wave, and g is the wavenumber in the y direction whose form is still to be determined.

To proceed, we Fourier expand the refractive index profile. Specifically, because the refractive index enters the wave equation raised to the second power, we expand n^2:

$$n^2(z) = \sum_s \alpha_s \exp(\mathrm{i}2\pi sz/\Lambda_g), \tag{16}$$

where α_s is the Fourier coefficient given by

$$\alpha_s = \frac{1}{\Lambda_g}\int_{-\Lambda_g/2}^{\Lambda_g/2} n^2(z)\exp(-\mathrm{i}2\pi sz/\Lambda_g)\mathrm{d}z. \tag{17}$$

Derivation of the eigenvalue equation

We will now substitute Eqs. (15) and (16) into the wave equation (5) and simplify to give the eigenvalue equation which is Eq. (13) in Knop [1].

The substitution described above gives

$$\sum_\ell -E_\ell p_l^2 \exp[\mathrm{i}(p_\ell z + gx)] + \sum_\ell -E_\ell g^2 \exp[\mathrm{i}(p_\ell z + gx)] + k_0^2 \sum_s \alpha_s \exp(\mathrm{i}2\pi sz/\Lambda_g) \sum_\ell -E_\ell \exp[\mathrm{i}(p_\ell z + gx)] = 0 \tag{18}$$

For enhanced clarity, we rewrite the term on the second line of Eq. (18) as follows:

$$\begin{aligned} &k_0^2 \sum_s \alpha_s \exp(\mathrm{i}2\pi sz/\Lambda_g) \sum_\ell -E_\ell \exp[\mathrm{i}(p_\ell z + gx)] \\ &= k_0^2 \sum_s \sum_\ell \alpha_s E_\ell \exp[\mathrm{i}(\{2\pi[s+\ell]/\Lambda_g) + k_0 \sin\theta\}z + gx)] \\ &= k_0^2 \sum_s \sum_\ell \alpha_s E_\ell \exp[\mathrm{i}(p_{s+\ell} z + gx)], \end{aligned} \tag{19}$$

where we have used the definition of p_ℓ given in Eq. (8). For additional clarity, we label the terms in Eq. (18) as follows:

$$\chi_1 = \sum_\ell -E_\ell p_l^2 \exp[\mathrm{i}(p_\ell z + gx)], \tag{20}$$

$$\chi_2 = \sum_\ell -E_\ell g^2 \exp[\mathrm{i}(p_\ell z + gx)], \tag{21}$$

$$\chi_3 = k_0^2 \sum_s \sum_\ell \alpha_s E_\ell \exp[\mathrm{i}(p_{s+\ell} z + gx)], \tag{22}$$

allowing us to write Eq. (18) as follows:

$$\chi_1 + \chi_2 + \chi_3 = 0. \tag{23}$$

To proceed, we note that Eq. (18) involves a sum of plane waves or partial waves in z, which are harmonics of the fundamental period solution with wavenumber $p_1 = 2\pi/\Lambda_g$ (note that a non-zero incident angle θ merely adds the constant $k\sin\theta$ to all the p_ℓ).

These plane waves are orthogonal, so if we take the inner product $\langle f, g\rangle = \int_{\Lambda_g/2}^{\Lambda_g/2} fg\mathrm{d}z$ between Eq. (18) and a plane wave state with $\ell = \eta$,$\exp[\mathrm{i}p_\eta z]$, for any integer η, we obtain the orthogonality condition

$$\langle \exp[\mathrm{i}(p_\ell z + gx)], \exp[\mathrm{i}(p_\eta z + gx)]\rangle = \delta_{\ell,\eta}\exp(\mathrm{i}2gx), \tag{24}$$

where δ is the Kroenecker delta symbol. Taking inner products with each of the $\chi_i, i = 1,2,3$ in Eq. (20) and $\exp[\mathrm{i}(p_\eta z + gy)$ we find

$$\langle \chi_1, \exp[\mathrm{i}(p_\eta z + gx)\rangle = -E_\eta p_\eta^2 \exp[\mathrm{i}2gx], \tag{25}$$

$$\langle \chi_2, \exp[\mathrm{i}(p_\eta z + gx)\rangle = -E_\eta g^2 \exp[\mathrm{i}2gx], \tag{26}$$

$$\langle \chi_3, \exp[\mathrm{i}(p_\eta z + gx)\rangle = k_0^2 \sum_\ell \alpha_{\eta-\ell} E_\ell \exp[\mathrm{i}2gx], \tag{27}$$

where the result in the case of χ_3 is obtained by noting that the orthogonality condition kills all terms except the term where $s + \ell = \eta \Rightarrow s = \eta - \ell$.

Finally, by cancelling the factor $\exp[\mathrm{i}2gx]$ which is common to all the inner products, and with reference to Eq. (23), we can rewrite the infinite sum of Eq. (18) as an infinite number of equations

$$-E_\eta p_\eta^2 - E_\eta g^2 + k_0^2 \sum_\ell \alpha_{\eta-\ell} E_\ell = 0. \tag{28}$$

If we further rearrange the terms as follows:

$$\sum_{\ell}[k_0^2\alpha_{\eta-\ell}E_\ell - \delta_{\ell,\eta}p_\ell^2 E_\ell] = E_\eta g^2, \tag{29}$$

we can recognize the result as an eigenvalue equation of the form

$$\hat{A}\mathbf{E} = g^2\mathbf{E}, \tag{30}$$

where $\hat{A}_{\eta,\ell} = k_0^2\alpha_{\eta-\ell} - \delta_{\ell,\eta}p_\ell^2$ and $E = [\ldots, E_{-N}, \ldots, E_l, \ldots, E_N, \ldots]^T$ is a vector composed of the E_ℓ values. Note that although there are an inifinite number of components in principle, for numerical calculations we will choose some finite N such that there are $2N+1$ entries in $\mathbf{E}$.

In general, linear algebra tells us that there will be as many eigenvalues g as there are entries in $\mathbf{E}$. Where therefore rewrite the eigenvalue equation as follows:

$$\hat{A}\mathbf{E} = g_n^2\mathbf{E}, \tag{31}$$

where the integer n labels the eigenvalues. From here on, we will assume that the eigenvalue equation can be solved and that the set of eigenvalues g_n is known. Furthermore, for each distinct eigenvalue g_n, there will be an associated eigenvector $\mathbf{E}_n$. We will label the components of these eigenvectors $E_{\ell,n}$. Finally we note that each g_n corresponds to a physically allowable wavenumber in the y-direction, with both incoming (negative) and outgoing (positive) waves possible for each wavenumber. With this in mind, we may write the field in Region II as follows:

$$E_{II}(x,z) = \sum_{\ell,n} E_{\ell,n}\exp(\mathrm{i}p_\ell z) \times [A_n\exp(\mathrm{i}g_n x) + B_n\exp(-\mathrm{i}g_n x)], \tag{32}$$

where A_n and B_n are respectively the amplitudes of the incoming and outgoing partial wave with wave number g_n.

APPLICATION OF BOUNDARY CONDITIONS

Having defined the field in all three areas defined in Fig. 1, we now apply continuity conditions at the boundaries of each region. We explicitly require the following conditions in order for the Electro-magnetic field to be continuous everywhere:

Boundary conditions

- $E(x,z)$ itself must obviously be continuous at the boundaries,
- $\partial E(x,z)/\partial x$ must be continuous and
- the above two conditions must hold for any and all values of z.

With these conditions required, the contiunity requirements may be written as the following four equations

$$E_I(x=d,z) = E_{II}(x=d,z) \tag{33a}$$

$$\frac{\partial E_I}{\partial x}(x=d,z) = \frac{\partial E_{II}}{\partial x}(x=d,z) \tag{33b}$$

$$E_{II}(x=0,z) = E_{III}(x=0,z) \tag{33c}$$

$$\frac{\partial E_{II}}{\partial x}(x=0,z) = \frac{\partial E_{III}}{\partial x}(x=0,z). \tag{33d}$$

Using Eqs. (12), (32) and (12), we can write Eqs. (33) as

$$\exp[\mathrm{i}(p_0 z - r_0 d)] + \sum_\ell R_\ell \exp[\mathrm{i}(p_\ell z + r_\ell d)] = \sum_{\ell,n} E_{\ell,n}\exp(\mathrm{i}p_\ell z)[A_n\exp(\mathrm{i}g_n x) + B_n\exp(-\mathrm{i}g_n x)] \tag{34a}$$

$$-\mathrm{i}r_0\exp[\mathrm{i}(p_0 z - r_0 d)] + \mathrm{i}\sum_\ell r_\ell R_\ell\exp[\mathrm{i}(p_\ell z + r_\ell d)] = \mathrm{i}\sum_{\ell,n} E_{\ell,n}\exp(\mathrm{i}p_\ell z)[A_n g_n\exp(\mathrm{i}g_n x) - B_n g_n\exp(-\mathrm{i}g_n x)] \tag{34b}$$

$$\sum_{\ell,n} E_{\ell,n}\exp(\mathrm{i}p_\ell z)[A_n + B_n] = \sum_\ell T_\ell\exp[\mathrm{i}p_\ell z] \tag{34c}$$

$$\mathrm{i}\sum_{\ell,n} E_{\ell,n}\exp(\mathrm{i}p_\ell z)[A_n g_n - B_n g_n] = -\mathrm{i}\sum_\ell T_\ell t_\ell\exp[\mathrm{i}p_\ell z]. \tag{34d}$$

Next, we apply the inner product with the function $\exp[\mathrm{i}p_\eta z]$ to each equation. Then Eqs. (34) become

$$\delta_{\eta,0}\exp[\mathrm{i}r_\eta d] + R_\eta \exp[\mathrm{i}r_\eta d] = \sum_n E_{\eta,n}[A_n \exp(\mathrm{i}g_n x) + B_n \exp(-\mathrm{i}g_n x)] \tag{35a}$$

$$-\delta_{\eta,0}r_\eta \exp[\mathrm{i}r_\eta d] + r_\eta R_\eta \exp[\mathrm{i}r_\eta d] = \sum_n E_{\eta,n}[A_n g_n \exp(\mathrm{i}g_n x) - B_n g_n \exp(-\mathrm{i}g_n x)] \tag{35b}$$

$$\sum_n E_{\eta,n}[A_n + B_n] = T_\eta \tag{35c}$$

$$\sum_n E_{\eta,n}[A_n g_n - B_n g_n] = -T_\eta t_\eta. \tag{35d}$$

We would now like to express Eqs. (35) in matrix form. To do so, we note the following: Expressions of the sort $\sum_n E_{\eta,n}[A_n \exp(\mathrm{i}g_n x) + B_n \exp(-\mathrm{i}g_n x)]$ may be interpreted as a component of a matrix multiplication where the ηth row of the matrix $\hat{E}_{\eta,n}$ is multiplying the column vector $(\mathbf{Y})_n = A_n \exp(\mathrm{i}g_n x) + B_n \exp(-\mathrm{i}g_n x)$. Furthermore, a vector of the form $(\mathbf{V})_n = A_n \exp(\mathrm{i}g_n d)$ may be written as a matrix multiplication $V = \hat{e}_g \mathbf{A}$, where $(\hat{e}_g)_{\eta,n} = \delta_{\eta,n}\exp(\mathrm{i}g_n d)$.

In order to write Eqs. (35) in matrix form, we therefore define the following matrices:

$$(\hat{M}_{\mathrm{eg}})_{\eta,n} = \delta_{\eta,n}\exp(\mathrm{i}g_n d) \tag{36a}$$
$$(\hat{M}_{\mathrm{er}})_{\eta,n} = \delta_{\eta,n}\exp(\mathrm{i}r_\eta d) \tag{36b}$$
$$(\hat{M}_{\mathrm{g}})_{\eta,n} = \delta_{\eta,n} g_\eta \tag{36c}$$
$$(\hat{M}_{\mathrm{t}})_{\eta,n} = \delta_{\eta,n} t_n \tag{36d}$$
$$(\hat{M}_{\mathrm{r}})_{\eta,n} = \delta_{\eta,n} r_n. \tag{36e}$$

These matrices are all diagonal and thus trivially invertible. We also note that the matrix $\hat{E}$, as defined above, is the matrix whose nth column is the eigenvector associated with g_n.

In addition, we define the vectors

$$(\mathbf{V}_{\mathrm{er}})_\eta = \delta_{\eta,0}\exp(\mathrm{i}r_\eta d) \tag{37a}$$
$$(\mathbf{V}_{\mathrm{rer}})_\eta = \delta_{\eta,0} r_\eta \exp(\mathrm{i}r_\eta d) \tag{37b}$$
$$(\mathbf{T})_\eta = T_\eta \tag{37c}$$
$$(\mathbf{R})_\eta = R_\eta \tag{37d}$$
$$(\mathbf{A})_n = A_n \tag{37e}$$
$$(\mathbf{B})_n = B_n. \tag{37f}$$

We now move all terms in Eqs. (35) which depend on the coefficient vectors $\mathbf{T}, \mathbf{R}, \mathbf{A}$ and $\mathbf{B}$ to the left hand side of the equation and express Eqs. (35) in matrix form as follows:

$$\hat{E}\hat{M}_{\mathrm{eg}}\mathbf{A} + \hat{E}\hat{M}^*_{\mathrm{eg}}\mathbf{B} - \hat{M}_{\mathrm{er}}\mathbf{R} = \mathbf{V}_{\mathrm{er}} \tag{38a}$$
$$\hat{E}\hat{M}_{\mathrm{eg}}\hat{M}_{\mathrm{g}}\mathbf{A} + \hat{E}\hat{M}^*_{\mathrm{eg}}\hat{M}_{\mathrm{g}}\mathbf{B} - \hat{M}_{\mathrm{r}}\hat{M}_{\mathrm{er}}\mathbf{R} = \mathbf{V}_{\mathrm{rer}} \tag{38b}$$
$$\hat{E}\mathbf{A} + \hat{E}\mathbf{B} - \mathbf{T} = 0 \tag{38c}$$
$$\hat{E}\hat{M}_{\mathrm{g}}\mathbf{A} - \hat{E}\hat{M}_{\mathrm{g}}\mathbf{B} + \hat{M}_{\mathrm{t}}\mathbf{T} = 0. \tag{38d}$$

Our goal is to find the coefficients $\mathbf{T}$ of the transmitted partial waves. Inspection of Eqs. (38a) reveals that we have four equations and four unknowns (i.e. $\mathbf{T}, \mathbf{R}, \mathbf{A}$ and $\mathbf{B}$). Assuming the equations are consistent, we can therefore solve for any one of the aforementioned four unknowns using elimination of variables. We will now eliminate variables in order to find $\mathbf{T}$.

SOLUTION OF THE CONTINUITY EQUATIONS FOR THE TRANSMISSION COEFFICIENTS

We deal with the simplest of Eqs. (38a). Rearranging Eqs. (38c) and (38d) by applying $\hat{E}^{-1}$ and $\hat{M}_{\mathrm{g}}^{-1}\hat{E}^{-1}$ respectively, we find

$$\mathbf{A} + \mathbf{B} = \hat{E}^{-1}\mathbf{T} \tag{39}$$
$$\mathbf{A} - \mathbf{B} = -\hat{M}_{\mathrm{g}}^{-1}\hat{E}^{-1}\hat{M}_{\mathrm{t}}\mathbf{T} \tag{40}$$

Next, we form the sum and difference of Eqs.(39) and (40) giving, respectively,

$$\mathbf{A} = \frac{1}{2}[\hat{E}^{-1} - \hat{M}_{\mathrm{g}}^{-1}\hat{E}^{-1}\hat{M}_{\mathrm{t}}]\mathbf{T} \tag{41}$$
$$\mathbf{B} = \frac{1}{2}[\hat{E}^{-1} + \hat{M}_{\mathrm{g}}^{-1}\hat{E}^{-1}\hat{M}_{\mathrm{t}}]\mathbf{T} \tag{42}$$

We proceed by multiplying Eq. (38a) by $\hat{M}_{\mathrm{r}}$ and subtracting Eq. (38b) to give

$$[\hat{M}_{\rm r}\hat{E}\hat{M}_{\rm eg} - \hat{E}\hat{M}_{\rm eg}\hat{M}_{\rm g}]\mathbf{A} + [\hat{M}_{\rm r}\hat{E}\hat{M}^*_{\rm eg} - \hat{E}\hat{M}^*_{\rm eg}\hat{M}_{\rm g}]\mathbf{B} = \hat{M}_{\rm r}\mathbf{V}_{\rm er} + \mathbf{V}_{\rm rer}. \quad (43)$$

Finally, we substitute Eqs. (41) and (42) into Eq. (43) to give the following equation involving T $\mathbf{T}$:

$$\frac{1}{2}\{[\hat{M}_{\rm r}\hat{E}\hat{M}_{\rm eg} - \hat{E}\hat{M}_{\rm eg}\hat{M}_{\rm g}][\hat{E}^{-1} - \hat{M}_{\rm g}^{-1}\hat{E}^{-1}\hat{M}_{\rm t}] + [\hat{M}_{\rm r}\hat{E}\hat{M}^*_{\rm eg} - \hat{E}\hat{M}^*_{\rm eg}\hat{M}_{\rm g}][\hat{E}^{-1} + \hat{M}_{\rm g}^{-1}\hat{E}^{-1}\hat{M}_{\rm t}]\}\mathbf{T} = \hat{M}_{\rm r}\mathbf{V}_{\rm er} + \mathbf{V}_{\rm rer}. \quad (44)$$

We can then express $\mathbf{T}$ as

$$\mathbf{T} = \hat{U}^{-1}[\hat{M}_{\rm r}\mathbf{V}_{\rm er} + \mathbf{V}_{\rm rer}], \quad (45)$$

where after expanding and rearranging the LHS of Eq. (44), we find that $\hat{U}$ is given by

$$\hat{U} = \hat{M}_{\rm r}\hat{E}\hat{C}_{\rm g}\hat{E}^{-1} - {\rm i}\hat{M}_{\rm r}\hat{E}\hat{S}_{\rm g}\hat{M}_{\rm g}^{-1}\hat{E}^{-1}\hat{M}_{\rm t} - {\rm i}\hat{E}\hat{S}_{\rm g}\hat{M}_{\rm g}\hat{E}^{-1} + \hat{E}\hat{C}_{\rm g}\hat{E}^{-1}\hat{M}_{\rm t}, \quad (46)$$

where $(\hat{C}_{\rm g})_{\eta,n} = \delta_{\eta,n}\cos(g_\eta d)$ and $(\hat{S}_{\rm g})_{\eta,n} = \delta_{\eta,n}\sin(g_\eta d)$. Although the expression for $\hat{U}$ looks complicated, Eq. 45 is arguably the most useful form of the solution for computation with modern matrix based numerical software such as Matlab. Nonetheless, we now go on to calculate the component-wise values of $\hat{U}$ and the RHS of Eq. 45 in order to compare our result with that given by Knop.

COMPARISON WITH KNOP'S RESULT

Components of RHS of Eq. (45)

We first perform the relatively simple task of comparing the RHS of Eq. (45) with Knop's result. We want to find the components of the vector $\hat{M}_{\rm r}\mathbf{V}_{\rm er} + \mathbf{V}_{\rm rer}$. Expanding into components based on the definitions in Eqs. (36a) and (37a), we find

$$\begin{aligned}(\hat{M}_{\rm r}\mathbf{V}_{\rm er} + \mathbf{V}_{\rm rer})_n &= \sum_\eta \delta_{\eta,n} r_n \delta_\eta, 0\exp(-{\rm i}r_\eta d) \\ &\quad + \delta_{\eta,0} r_\eta \exp(-{\rm i}r_\eta d) \\ &= \delta_{0,n} 2r_n \exp(-{\rm i}r_n d). \quad (47)\end{aligned}$$

That is, the RHS is a vector containing a single element $2r_0\exp(-{\rm i}r_0 d)$ at the zeroth position in the vector, in agreement with Knop's Eq. (16) [1].

Components of $\hat{U}$

We first divide $\hat{U}$ into a sum of four matrices:

$$\hat{U} = \hat{U}_1 - {\rm i}\hat{U}_2 - {\rm i}\hat{U}_3 + \hat{U}_4, \quad (48)$$

where

$$\hat{U}_1 = \hat{M}_{\rm r}\hat{E}\hat{C}_{\rm g}\hat{E}^{-1} \quad (49)$$

$$\hat{U}_2 = \hat{M}_{\rm r}\hat{E}\hat{S}_{\rm g}\hat{M}_{\rm g}^{-1}\hat{E}^{-1}\hat{M}_{\rm t} \quad (50)$$

$$\hat{U}_3 = \hat{E}\hat{S}_{\rm g}\hat{M}_{\rm g}\hat{E}^{-1} \quad (51)$$

$$\hat{U}_4 = \hat{E}\hat{C}_{\rm g}\hat{E}^{-1}\hat{M}_{\rm t}. \quad (52)$$

We will find the components of each matrix by noting the following facts about matrix multiplication:

1. Multiplication of two aribtrary matrices in terms of their components is given by
$$(\hat{A}\hat{B})_{l,m} = \sum_n (\hat{A})_{l,n}(\hat{B})_{n,m}. \quad (53)$$
2. When a diagonal matrix $\hat{D}$ premultiplies an arbitrary matrix $\hat{A}$, the components have the form
$$(\hat{D}\hat{A})_{l,m} = (\hat{D})_{l,l}(\hat{A})_{l,m}. \quad (54)$$
For diagonal matrix postmultplication, we have, of course,
$$(\hat{A}\hat{D})_{l,m} = (\hat{A})_{l,m}(\hat{D})_{m,m}. \quad (55)$$
Because all of our matrices other than $\hat{E}$ are diagonal, the evaluation of the components is straighforward.

Components of $\hat{U}_1$

We proceed from right to left: Apply Eq. (54) to show

$$(\hat{C}_{\rm g}\hat{E}^{-1})_{l,m} = (\hat{C}_{\rm g})_{l,l}(\hat{E}^{-1})_{l,m} = \cos(g_l d)(E^{-1})_{l,m}.$$

Next, apply Eq. (53) to show

$$(\hat{E}\hat{C}_{\rm g}\hat{E}^{-1})_{l,m} = \sum_n (\hat{E})_{l,n}\cos(g_n d)(E^{-1})_{n,m}.$$

Finally, we apply Eq. (54) again to give

$$\begin{aligned}(\hat{U}_1)_{l,m} &= (\hat{M}_{\rm r}\hat{E}\hat{C}_{\rm g}\hat{E}^{-1})_{l,m} \\ &= r_l \sum_n (\hat{E})_{l,n}\cos(g_n d)(E^{-1})_{n,m}. \quad (56)\end{aligned}$$

We will proceed in the same manner below for the other matrices $\hat{U}_s$ but with less working shown.

Components of $\hat{U}_2$

$$(\hat{E}^{-1}\hat{M}_{\rm t})_{l,m} = (\hat{E}^{-1})_{l,m}(\hat{M}_{\rm t})_{m,m} = (\hat{E}^{-1})_{l,m}t_m.$$

$$\hat{S}_{\rm g}\hat{M}_{\rm g}^{-1}\hat{E}^{-1}\hat{M}_{\rm t} = \sin(g_l)g_l(\hat{E}^{-1})_{l,m}t_m.$$

$$\begin{aligned}\hat{U}_2 &= \hat{M}_{\rm r}\hat{E}\hat{S}_{\rm g}\hat{M}_{\rm g}^{-1}\hat{E}^{-1}\hat{M}_{\rm t} \\ &= r_l\sum_n E_{l,n}\sin(g_n)(1/g_n)(\hat{E}^{-1})_{n,m}t_m.\end{aligned} \quad (57)$$

Components of $\hat{U}_3$

$$\hat{U}_3 = \hat{E}\hat{S}_{\rm g}\hat{M}_{\rm g}\hat{E}^{-1} = \sum_n(\hat{E})_{l,n}\sin(g_n d)g_n(E^{-1})_{n,m}. \quad (58)$$

Components of $\hat{U}_4$

$$\hat{U}_4 = \hat{E}\hat{C}_{\rm g}\hat{E}^{-1}\hat{M}_{\rm t} = \sum_n(\hat{E})_{l,n}\cos(g_n d)(E^{-1})_{n,m}t_m. \quad (59)$$

Finally, we can find the components of $\hat{U}$ itself:

$$\begin{aligned}\hat{U}_{l,m} &= (\hat{U}_1)_{l,m} - {\rm i}(\hat{U}_2)_{l,m} - {\rm i}(\hat{U}_3)_{l,m} + (\hat{U}_4)_{l,m} \\ &= (r_l + t_m)\sum_n(\hat{E})_{l,n}(\hat{E}^{-1}_{n,m}\cos(g_n d) \\ &\quad -{\rm i}\sum_n\left(g_n + \frac{r_l t_m}{g_n}\right)(\hat{E})_{l,n}(\hat{E}^{-1}_{n,m}\sin(g_n d)\end{aligned} \quad (60)$$

Knop's Eq. 17 in [1] contains a typo which makes the result ambiguous. However, the only sane interpretation of Knop's result is exactly the equation we have derived above.

KNOP'S SOLUTION IN THE CASE OF INCIDENCE FROM THE SUBSTRATE SIDE

The situation analysed by Knop and which was analysed above involves a plane wave incident from the air side of the grating. However, in the situation we are interested in analysing, a nanofiber lies on top of the grating and a laser is illuminated from "below" the grating (i.e. coming from the grating substrate side). It is essentially trivial to change the above formalism to accomodate this difference. The simplest way is just to swap the indices of refraction in regions I and III. (Region II is not affected). To do this we alter Eqs. (11) and (13) to read as follows:

$$r_\ell = \begin{cases} \sqrt{n_0^2k_0^2 - p_\ell^2} & n_0k_0 \geq |p_\ell| \\ {\rm i}\sqrt{p_\ell^2 - n_0^2k_0^2} & n_0k_0 \leq |p_\ell| \end{cases} \quad (61)$$

$$t_\ell = \begin{cases} \sqrt{k_0^2 - p_\ell^2} & k_0 \geq |p_\ell| \\ {\rm i}\sqrt{p_\ell^2 - k_0^2} & k_0 \leq |p_\ell| \end{cases} \quad (62)$$

With the above redefinitions, the derivation continues completely unchanged, and the results (45) and (46) are unchanged.

Matlab code for solving Knop's equations in the case of a photonic crystal nanofiber

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Solve the Maxwell equations for a grating using "rigorous
% diffraction theory" after Knop, J. Opt. Soc. Am 68, 1206 (1978)
% "Rigorous diffraction theory" just means solve Maxwell's equations
% without approximations no matter how tedious it gets!
%
%
% NB: I am *only* solving fo the "EP" polarization here.
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
clear all
close all

% Parameters of the experiment
% We take 1um as our length standard. All
% lengths should be taken to be in um
lambda = 0.85 %e-6; % Input wavelength
k0 = 2*pi/lambda;
theta0 = 0; % Input angle
ng = 1.45; % Silica grating's refractive index
d=0.9 % Grating period
a = 2 %e-6; % Grating depth - this is just the crater depth
b=0.08; % Grating duty cycle

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Proceed with Knop's method
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% Set up our l space. This defines the number of diffraction orders
N=2; % There will be N/2+1 diffraction orders included in the calculation
     %(including the 0th order)
     % At this point it's important to note the following: Knop's analysis
```

```
     % is watertight formally, but in terms of numerical calculations it
     % starts to fail for deep gratings when evanescent orders are
     % included. If you are only interested in the travelling wave
     % transmitted field, as we will be for trapping concerns, just check
     % for what value of l the transmitted wave-numbers t become
     % imaginary, and choose N so that only the travelling wave diffraction
     % orders are included. For example, for the kind of gratings we're
     % interested in where the grating period is only a few 10s of percent
     % bigger than the incident wavelength, the first order diffraction is
     % sufficient, and we can set N=2.

l=(-N/2):(N/2);
ll = (-N):N;
% Define wave numbers
p = 2*pi*l/d + k0*sin(theta0) % x wave number
r = sqrt(ng^2*k0^2-p.^2) % reflected y wave number (in region 1)
t = sqrt(k0^2-p.^2) % Transmitted y wave number (in region 3)



% Fourier coefficients
alphas = (ng^2-1)*sin(ll*pi*b)./(ll*pi);
alphas(find(ll==0)) = 1 - b + b*ng^2;

% Solve the eigenvalue equation.
% This simply means that we construct the matrix from Knop eq. 13
% and then let Matlab do the hard work of finding the eigenvalues and
% vectors
[eta,ell] = meshgrid(l,l);
alphaind = eta-ell+N+1 % This is the index into the alpha values for constructing the matrix

At = k0^2*(alphas(alphaind)); % If you're not familiar with
                                        % matlab, this looks a bit strange,
                                        % but basically if just reshapes
                                        %  the alphaNN matrix of Fourier
                                        % coefficients into a matrix

Chi = At - diag(p.^2); % This provides the matrix for Knop's Eq. 13.
[Es,g2] = eig(Chi)
g = sqrt(g2)

% After solving the eigenvalue problem, all that remains is to define some
% matrices which arise from the boundary conditions. The matrices are only
% needed in order to make the matrix version of Knop's Eq. (17)

opG = diag(exp(i*diag(g)*a)) % eig returns a diagonal matrix for the
                            %eigenvalues
opC = diag(exp(-imag(diag(g))*a).*cos(real(diag(g)*a))) % cos of eigvals
opS = diag(exp(-imag(diag(g))*a).*sin(real(diag(g)*a))) % sin of "
opT = diag(exp(i*t*a))
opr = diag(r)
opt = diag(t)
opg = g % eig returns a diagonal matrix for the eigen values, so fine as is

% We need the following definition for Eq. (16)
V =  zeros(length(l),1);
V(find(l==0)) = 2*r(find(l==0))*exp(-i*r(find(l==0))*a)

% Knop's U in matrix form (as derived by MS - could be wrong, but have
% compared for a few values and it looks correct)
U1 = opr*Es*opC*inv(Es)
U2 = opr*Es*opS*inv(opg)*inv(Es)*opt
U3 = Es*opS*opg*inv(Es)
U4 = Es*opC*inv(Es)*opt
U = U1 - i*U2 -i*U3 + U4;
check2 = inv(U)

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% To use the U matrix from Knop's Eq. 17, uncomment below.
% However, the matrix expression above can be shown to
% be equivalent to the component-wise definition below.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
 iEs = inv(Es);
 U = zeros(length(l),length(l));
 for jj = 1:length(l)
  for kk = 1:length(l)
    U(jj,kk) = (r(jj)+t(kk)) * Es(jj,:)*(diag(real(opG)).*iEs(:,kk)) -...
        i*(Es(jj,:)*(diag(g).*diag(imag(opG)).*iEs(:,kk)) +...
        r(jj)*t(kk)*Es(jj,:)*(diag(inv(g)).*diag(imag(opG)).*iEs(:,kk)));
  end
 end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% Finally, calculate the transmission coefficients
T = U\V % Knop Eq. (16)

whos

check = sum(abs(T(:)).^2.*t(:)/r(find(l==0)))


%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Plotting
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% Plot the transmission coefficients to
% check they look sensible
figure(1)
clf
plot(p,abs(T).^2,'bo-')
xlabel('l')
ylabel('T_l')

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Finally, plot what we really want to see:
% the diffraction pattern!
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

x=linspace(-6,6,500);
```

```
y = linspace(-3.0,-0.00005,2000);
[X,Y] = meshgrid(x,y);
E = zeros(size(X));
% Construct the E field (Eq (7))
for ss = 1:length(l)
  E = E + T(ss) * exp(i*(p(ss)*X - t(ss)*Y));
end


figure(2)
clf
hold on

% Plot intensity map over x,y
contourf(X,Y,abs(E).^2)

% Finally labels
xlabel('x (um)','FontSize',18)
ylabel('y (um)','FontSize',18)
title('|E|^2','FontSize',18)
colorbar
set(gca,'FontSize',14)
```


[1] K. Knop, J. Opt. Soc. Am. **68**, 1206 (1978).

[2] C. B. Burckhardt, J. Opt. Soc. Am. **56**, 1502 (1966).

[3] M. Sadgrove, R. Yalla, K. P. Nayak, and K. Hakuta, Optics Letters, **38**, 2542 (2013).

[4] K.P. Nayak and K. Hakuta, Opt. Express **21**, 2480-2490 (2013).