Explaining the success of Kogelnik's coupled-wave theory by means of perturbation analysis: discussion

Heinz-Jürgen Schmidt,* Mirco Imlau, and Kay-Michael Voit

Universität Osnabrück, Fachbereich Physik, Barbarastr. 7, 49069 Osnabrück, Germany
**Corresponding author: hschmidt@uos.de*

Received December 17, 2013; revised March 9, 2014; accepted April 1, 2014;
posted April 1, 2014 (Doc. ID 203083); published May 2, 2014

The problem of diffraction of an electromagnetic wave by a thick hologram grating can be solved by the famous Kogelnik's coupled-wave theory (CWT) to a very high degree of accuracy. We confirm this finding by comparing the CWT and the exact result for a typical example and propose an explanation in terms of perturbation theory. To this end we formulate the problem of diffraction as a matrix problem following similar well-known approaches, especially rigorous coupled-wave theory (RCWT). We allow for a complex permittivity modulation and a possible phase shift between refractive index and absorption grating and explicitly incorporate appropriate boundary conditions. The problem is solved numerically exact for the specific case of a planar unslanted grating and a set of realistic values of the material's parameters and experimental conditions. Analogously, the same problem is solved for a two-dimensional truncation of the underlying matrix that would correspond to a CWT approximation but without the usual further approximations. We verify a close coincidence of both results even in the off-Bragg region and explain this result by means of a perturbation analysis of the underlying matrix problem. Moreover, the CWT is found not only to coincide with the perturbational approximation in the in-Bragg and the extreme off-Bragg cases, but also to interpolate between these extremal regimes.

OCIS codes: (140.3490) Lasers, distributed-feedback; (060.2420) Fibers, polarization-maintaining; (060.3735) Fiber Bragg gratings; (060.2370) Fiber optics sensors.
http://dx.doi.org/10.1364/JOSAA.31.001158

1. INTRODUCTION

The recording of elementary phase and absorption Bragg gratings in thick dielectric media by means of the holographic principle [1,2] obeys its potential in a variety of visionary applications in the field of photonics, such as high-density holographic data storage [3] or real-time, real three-dimensional holographic displays [4]. Widely established are holographically recorded Bragg gratings for distributed feedback lasers, laser beam combiners, and wavelength division multiplexers in telecommunication networks [5,6]. But also blazed holographic surface gratings are state-of-the-art optical systems for chirped pulse amplification in ultrafast laser sources.

Correct analysis of the phase and amplitude of the reconstructed signal wave diffracted from a hologram is inevitable for such applications. But it is also the key point of material characterization by means of holographic spectroscopy [7]. Boundary conditions for the derivation of the analysis have to be chosen adequately to experimental conditions and material parameters. Several publications in the literature face this type of analysis in detail for isotropic media [7–17], with Kogelnik's *coupled-wave theory (CWT) for thick hologram gratings* [18] being the most recognized one. Anisotropic media were studied thoroughly by Montemezzani and Zgonik [19] and Sturman *et al.* [20] by vectorial theories and consideration of the tensorial properties of the permittivity. This enabled the analysis of such important classes of photosensitive materials as semiconductors and sillenites. For a recent comparison of experiments on holographic reconstruction and different approaches in the far-off-Bragg regime, see [21].

Since numerical solutions of holographic diffraction problems based on rigorous approaches like rigorous coupled-wave theory (RCWT) are possible but laborious, there has always been a demand for relatively simple analytical approximations, the most prominent being the above-mentioned CWT ansatz by Kogelnik [18]. The success of this approximation is impressive but largely unexplained except for physical plausibility arguments. In the work at hand, we present a justification of CWT in terms of first-order perturbation theory. In Section 2 the exact solution of the diffraction problem for arbitrarily thick crystals and periodic complex permittivity is formulated as an infinite sum over "reflection generations" (see Fig. 1) that are mutually connected by appropriate boundary conditions at the faces of the crystal. Each generation is given by a superposition of the eigenvectors of an infinite-dimensional matrix $\mathbb{B}$. Following [7], the CWT approximation can be viewed as the restriction of $\mathbb{B}$ to a two-dimensional subspace spanned by the zeroth and first diffraction orders. Further approximations that are commonly used in the CWT approach, such as the replacement of a second-order differential equation with a first-order one, as well as the neglect of boundary conditions, are not necessary in the present context and hence can be avoided. This has the advantage that we can fully concentrate on the effect of the restriction to a two-dimensional subspace and need not bother about other approximations. Thus we also obtain an extension of CWT to

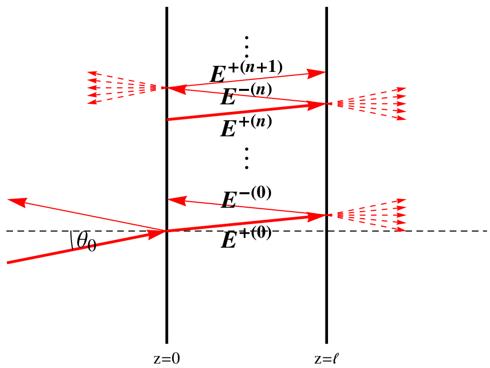

Fig. 1. Schematic sketch of the exact solution of the diffraction problem by the superposition of various generations. The thick red arrow at the left-hand side symbolizes an incoming plane wave. $\mathbf{E}^{+(n)}$ and $\mathbf{E}^{-(n)}$ denote the nth generations of the special solutions of the Helmholtz equation inside the grating. The sign $\pm$ refers to the two possible directions of the wave vectors. $\mathbf{E}^{+(n)}$ and $\mathbf{E}^{-(n)}$ are related by boundary conditions at $z = \ell$ and, similarly, $\mathbf{E}^{-(n)}$ and $\mathbf{E}^{+(n+1)}$ by boundary conditions at $z = 0$; see Eqs. (51) and (52). The superposition of the transmitted solutions at $z > \ell$, symbolized by small pencils of rays at the right-hand side, can be measured and compared with the theoretical predictions.

the case of a possible phase shift between refractive index and absorption grating not yet considered in the literature.

The numerical solutions of both theories, the exact one and the CWT, for realistic values of the material's parameters and experimental conditions are compared for the two relevant diffraction orders in Section 3 and found to be in excellent agreement. The explanation of this agreement in Section 4 is based on the observation that the values in the secondary diagonals of $\mathbb{B}$ fall off with powers of a perturbation parameter $\gamma \sim 10^{-3} \ldots 10^{-4}$. It turns out that for the in-Bragg case degenerate zeroth-order perturbation theory already provides a sensible approximation that is equivalent to the CWT approximation. In the far-off-Bragg region nondegenerate first-order perturbation theory leads to a restriction to the three-dimensional space of diffraction orders $m = -1, 0, 1$. Hence the use of CWT can be justified if the amplitude A_{-1} is relatively small, which is satisfied for the parameters used. Moreover, we argue that CWT can be viewed as an interpolation between the extreme in-Bragg and far-off-Bragg regimes. We close with a short summary in Section 5.

2. DIFFRACTION AS AN EIGENVALUE PROBLEM

In this section we want to present an exact solution of the diffraction problem outlined above in terms of the solution of an infinite-dimensional eigenvalue problem. We do not consider any approximations. The approach is thus similar to that of [9] but more explicit. As sketched in Fig. 1, the solution is represented as a superposition of a doubly infinite family of special solutions $\mathbf{E}^{+(n)}$ and $\mathbf{E}^{-(n)}$ of the Helmholtz equation, called *generations*. $\mathbf{E}^{-(n)}$ and $\mathbf{E}^{+(n)}$ are related by suitable boundary conditions at $z = \ell$ that guarantee the continuity of the tangential components of the electric and magnetic fields. Analogously for $\mathbf{E}^{-(n)}$ and $\mathbf{E}^{+(n+1)}$ at $z = 0$. The zeroth generation $\mathbf{E}^{+(0)}$ is defined as a solution satisfying the boundary conditions together with an incoming and a reflected plane wave at $z = 0$.

All generations are solutions of the Helmholtz equation with a complex periodic permittivity in the slab $0 \le z \le \ell$. By means of the ansatz of an x-periodic modulation of the well-known solution with constant permittivity, the problem can be reduced to the eigenvalue problem of an infinite-dimensional matrix $\mathbb{B}$. Hence our solution is "formal" in the sense that it depends on the solution of an eigenvalue problem, which can, however, be solved with arbitrary accuracy by standard numerical methods.

A. Homogeneous Case

We consider the slab $0 \le z \le \ell$ with infinite extension into the x and y directions and with constant permittivity

$$\epsilon = (n_0(1 + i\kappa_0))^2. \tag{1}$$

For $z < 0$ and $z > \ell$ we assume vacuum permittivity. For $z < 0$ there is an incoming monochromatic plane wave

$$\mathbf{E}_I = \begin{pmatrix} 0 \\ E_\mathrm{I} \\ 0 \end{pmatrix} e^{-i\omega t} \exp(ik_0(x \sin\theta_0 + z \cos\theta_0)), \tag{2}$$

plus a reflected wave

$$\mathbf{E}_\mathrm{III} = \begin{pmatrix} 0 \\ E_\mathrm{III} \\ 0 \end{pmatrix} e^{-i\omega t} \exp(ik_0(x \sin\theta_0 - z \cos\theta_0)). \tag{3}$$

As usual, we set $\omega = ck_0$, c being the vacuum velocity of light, and $\lambda_0 = (2\pi/k_0)$. For the electric field inside the slab we make the ansatz of a damped plane wave

$$\mathbf{E}_\mathrm{II} = \begin{pmatrix} 0 \\ E_\mathrm{II} \\ 0 \end{pmatrix} e^{-i\omega t} \exp(ik(x \sin\theta + \cos\theta) - qz \cos\theta). \tag{4}$$

The Helmholtz equation

$$\Delta \mathbf{E}_\mathrm{II} + k_0^2 \epsilon \mathbf{E}_\mathrm{II} = \mathbf{0} \tag{5}$$

and the boundary conditions at $z = 0$ lead to

$$k = \frac{1}{2} k_0 \sqrt{1 + 2n_0^2(1 - \kappa_0^2) - \cos(2\theta_0) + \xi}, \tag{6}$$

$$\xi = \frac{1}{\sqrt{2}} (3 + 8n_0^2(-1 + \kappa_0^2 + n_0^2(1 + \kappa_0^2)^2 \\ -4(1 + 2n_0^2(-1 + \kappa_0^2)) \cos 2\theta_0) + \cos 4\theta_0)^{1/2}, \tag{7}$$

$$q = \frac{k}{\kappa_0} \left(-1 + \left(\frac{k}{k_0 n_0} \right)^2 + \kappa_0^2 \right), \tag{8}$$

$$k \sin\theta = k_0 \sin\theta_0 (\text{Snell's law}), \tag{9}$$

$$\frac{E_\mathrm{III}}{E_I} = \frac{-(k + iq) \cos\theta + k_0 \cos\theta_0}{(k + iq) \cos\theta + k_0 \cos\theta_0}, \tag{10}$$

$$\frac{E_{\mathrm{II}}}{E_{\mathrm{I}}} = \frac{2k_0 \cos \theta_0}{(k+iq)\cos\theta + k_0 \cos\theta_0}, \tag{11}$$

where the last two equations are Fresnel's formulas.

B. Periodic Case

It turns out that the definitions for the exact solution and those for the CWT case are completely analogous. In order to allow for this we introduce an index set $\mathcal{Z}$ that is identical with the set $\mathbb{Z}$ of integers in the general case, and $\mathcal{Z} = \{0, 1\}$ in the CWT case. In the following the main text refers to the general case, whereas appropriate remarks about the CWT case are put into brackets.

We assume that the complex permittivity varies periodically in the x direction and hence can be represented by the Fourier series

$$\epsilon(x) = \sum_{\nu\in\mathbb{Z}} \tilde{\epsilon}_\nu \exp(iK\nu x). \tag{12}$$

We identify $\tilde{\epsilon}_0$ with $(n_0(1+i\kappa_0))^2$ in Eq. (1) and consider an x-periodic modulation of the solution (4) for the homogeneous case

$$\mathbf{E} = \begin{pmatrix} 0 \\ E_{\mathrm{II}} \\ 0 \end{pmatrix} e^{-i\omega t} \exp(ik(x\sin\theta + \cos\theta) - qz\cos\theta) \times \sum_{m\in\mathcal{Z}} \mathcal{A}_m(z)\exp(-imKx). \tag{13}$$

Inserting Eq. (13) into the Helmholtz equation yields

$$\begin{aligned} 0 = {} & \exp(ik(x\sin\theta + \cos\theta) - qz\cos\theta)\sum_{m\in\mathcal{Z}}\exp(-imKx) \\ & \{[(ik\sin\theta - imK)^2 + (ik-q)^2\cos^2\theta]\mathcal{A}_m(z) \\ & + 2(ik-q)\cos\theta\mathcal{A}'_m(z) \\ & + \mathcal{A}''_m(z) + k_0^2\sum_{\nu\in\mathbb{Z}}\tilde{\epsilon}_\nu\mathcal{A}_{m+\nu}(z)\}. \end{aligned} \tag{14}$$

Since the $m=0$ part of Eq. (13) satisfies the Helmholtz equation for the homogeneous case, the following holds:

$$0 = -k^2 + (q^2 - 2ikq)\cos^2\theta + k_0^2\tilde{\epsilon}_0. \tag{15}$$

This together with the redefinition,

$$\epsilon_\nu = \begin{cases} \tilde{\epsilon}_\nu & \text{if } \nu \neq 0, \\ 0 & \text{if } \nu = 0, \end{cases} \tag{16}$$

can be used to simplify Eq. (14). We obtain the following system of equations:

$$\begin{aligned} 0 = {} & mK(2k\sin\theta - mK)\mathcal{A}_m(z) + 2(ik-q)\cos\theta\mathcal{A}'_m(z) \\ & + \mathcal{A}''_m(z) + k_0^2\sum_{\nu\in\mathbb{Z}}\epsilon_\nu\mathcal{A}_{m+\nu}(z), \end{aligned} \tag{17}$$

for all $m \in \mathcal{Z}$. This can be formulated as a matrix equation of the form

$$0 = 2(ik-q)\cos\theta\mathcal{A}'(z) + \mathcal{A}''(z) + \mathbb{B}\mathcal{A}(z), \tag{18}$$

if we introduce the infinite-dimensional vector

$$\mathcal{A}(z) = \begin{pmatrix} \vdots \\ \mathcal{A}_{-1}(z) \\ \mathcal{A}_0(z) \\ \mathcal{A}_1(z) \\ \vdots \end{pmatrix} \tag{19}$$

and the complex matrix $\mathbb{B}$ with entries

$$\mathbb{B}_{m,m+\nu} = \begin{cases} k_0^2\epsilon_\nu & \text{if } \nu\neq 0, \\ mK(2k\sin\theta - mK) & \text{if } \nu = 0, \end{cases} \tag{20}$$

for $m, \nu \in \mathcal{Z}$. [In the CWT case $\mathbb{B}$ is a 2×2 matrix and Eq. (18) does not hold exactly but at most approximately.] We will assume that $\mathbb{B}$ has a complete (although not orthonormal) system of eigenvectors or "eigenmodes" $\phi^{(\mu)}$ with eigenvalues $b^{(\mu)}$ in the Hilbert space $\ell^2(\mathcal{Z})$:

$$\mathbb{B}\phi^{(\mu)} = b^{(\mu)}\phi^{(\mu)}, \qquad \mu\in\mathcal{Z}. \tag{21}$$

The vector $\mathcal{A}(z)$ can be expanded in terms of the eigenmodes in the form

$$\mathcal{A}(z) = \sum_{\mu\in\mathcal{Z}} a_\mu(z)\phi^{(\mu)}. \tag{22}$$

Then Eq. (18) is equivalent to an infinite system of decoupled ordinary second-order differential equations

$$0 = 2(ik-q)\cos\theta\, a'_\mu(z) + a''_\mu(z) + b^{(\mu)}a_\mu(z), \qquad \mu\in\mathcal{Z}, \tag{23}$$

with solutions

$$a_\mu(z) = c_\mu\exp(\beta_\mu z), \qquad c_\mu\in\mathbb{C}, \tag{24}$$

$$\beta_\mu = (q-ik)\cos\theta\left(1 \pm \sqrt{1 - \frac{b^{(\mu)}}{(q-ik)^2\cos^2\theta}}\right). \tag{25}$$

The sign $\pm$ in Eq. (25) has to be chosen in such a way that $\Re(\beta_\mu) < 0$ and hence the solution (13) will decay exponentially w. r. t. z. The corresponding coefficients will be denoted by c_μ^+. Then, after some simplifications and by defining

$$\alpha_\mu \equiv \pm\sqrt{(q-ik)^2\cos^2\theta - b^\mu}, \qquad \text{such that } \Re(\alpha_\mu) < 0, \tag{26}$$

Eq. (13) assumes the form

$$\mathbf{E}^+ = \begin{pmatrix} 0 \\ E \\ 0 \end{pmatrix} e^{-i\omega t} \exp(ikx \sin\theta) \times \sum_{m\in\mathcal{Z}} \exp(-imKx) \sum_{\mu\in\mathcal{Z}} c_\mu^+ \exp(\alpha_\mu z)\phi_m^\mu, \tag{27}$$

setting $E \equiv E_{\mathrm{II}}$ in what follows. The analogous solution that decays exponentially from right to left will be denoted by

$$\mathbf{E}^- = \begin{pmatrix} 0 \\ E \\ 0 \end{pmatrix} e^{-i\omega t} \exp(ikx \sin\theta) \times \sum_{m\in\mathcal{Z}} \exp(-imKx) \sum_{\mu\in\mathcal{Z}} c_\mu^- \exp(-\alpha_\mu z)\phi_m^\mu. \tag{28}$$

We will introduce some additional notations in order to simplify the boundary conditions for the final solution. First let us write

$$a_\mu^\pm(z) \equiv c_\mu^\pm \exp(\pm\alpha_\mu z) = a_\mu^\pm(0)\exp(\pm\alpha_\mu z), \tag{29}$$

and

$$A_m^\pm(z) \equiv \sum_{\mu\in\mathcal{Z}} c_\mu^\pm \exp(\pm\alpha_\mu z)\phi_m^{(\mu)} = \sum_{\mu\in\mathcal{Z}} a_\mu^\pm(z)\phi_m^{(\mu)}. \tag{30}$$

The last equation can be written in matrix form as

$$A^\pm(z) = \Phi a^\pm(z). \tag{31}$$

The solutions (27) and (28) then assume the form

$$\mathbf{E}^\pm = \begin{pmatrix} 0 \\ E \\ 0 \end{pmatrix} e^{-i\omega t} \exp(ikx \sin\theta) \sum_{m\in\mathcal{Z}} \exp(-imKx) A_m^\pm(z). \tag{32}$$

Moreover, introducing the diagonal matrix $\Delta(z)$ with entries

$$\Delta(z)_{\mu\nu} \equiv \delta_{\mu\nu} \exp(\alpha_\mu z), \tag{33}$$

and

$$T_z \equiv \Phi\Delta(z)\Phi^{-1}, \tag{34}$$

we can write

$$A^+(z) = \Phi a^+(z) = \Phi\Delta(z)a^+(0) \tag{35}$$

$$= (\Phi\Delta(z)\Phi^{-1})A^+(0) \tag{36}$$

$$= T_z A^+(0) = T_{z-\ell} A^+(\ell). \tag{37}$$

The last identity follows from

$$T_z T_u = T_{z+u}, \qquad \text{and} \qquad T_{-z} = T_z^{-1}. \tag{38}$$

The reverse z development can be written analogously as

$$A^-(z) = \Phi a^-(z) = \Phi\Delta(-z)a^-(0) \tag{39}$$

$$= (\Phi\Delta(-z)\Phi^{-1})A^-(0) \tag{40}$$

$$= T_{-z}A^-(0) = T_{\ell-z}A^-(\ell). \tag{41}$$

There is a certain analogy between T_z and the time evolution operator U_t in quantum mechanics. However, T_z is not unitary due to damping effects.

C. Boundary Conditions

In the last subsection we have found the general form of solutions $\mathbf{E}^+$ and $\mathbf{E}^-$ of the Helmholtz equation with periodic permittivity in the slab $0 \le z \le \ell$. These solutions have to be superposed in order to satisfy the boundary conditions at $z = 0$ and $z = \ell$ together with incoming and outgoing waves. As indicated above we will consider various "generations" of solutions that are accordingly pairwise matched. The zeroth generation $\mathbf{E}^{+(0)}$ is an electric field of the form (27). It satisfies the boundary conditions at $z = 0$ together with an incoming and a reflected plane wave if it coincides with the homogeneous solution (4) in an infinitesimal vicinity of $z = 0$. This can be achieved by choosing

$$A_m^{+(0)}(0) = \delta_{m0}E, \quad \text{for all } m \in \mathcal{Z}. \tag{42}$$

The next problem is to match $\mathbf{E}^{+(0)}$ and the backward solution $\mathbf{E}^{-(0)}$ at $z = \ell$. We will formulate this problem generally for the nth generation. The matching can only be realized for plane waves; hence we have to consider each order separately. For the m th order we have the boundary conditions of three waves, the forward solution

$$\mathbf{E}_m^{+(n)}(z) = \begin{pmatrix} 0 \\ E \\ 0 \end{pmatrix} e^{-i\omega t} \exp(ikx \sin\theta) \exp(-imKx) A_m^{+(n)}(z), \tag{43}$$

the corresponding reflected wave

$$\mathbf{E}_m^{-(n)}(z) = \begin{pmatrix} 0 \\ \rho E \\ 0 \end{pmatrix} e^{-i\omega t} \exp(ikx \sin\theta) \times \exp(-imKx) A_m^{+(n)}(2\ell - z), \tag{44}$$

and the transmitted plane wave

$$\mathbf{E}_m^{\tau(n)}(z) = \begin{pmatrix} 0 \\ \tau E \\ 0 \end{pmatrix} e^{-i\omega t} \exp(ik_0 x(\sin\theta_m x + \cos\theta_m(z - \ell)). \tag{45}$$

The condition of continuous tangential components of the resulting electric field at $z = \ell$ leads to the usual Snell's law

$$k \sin\theta - mK = k_0 \sin\theta_m, \tag{46}$$

where the case of total reflection with complex θ_m is included. In this case the transmitted wave decays exponentially for $z > \ell$. Additionally, we obtain the condition

$$\tau = (1+\rho)A_m^{+(n)}(\ell). \tag{47}$$

From the continuity of the tangential component of the magnetic field we obtain

$$\tau = (1-\rho)\frac{1}{ik_0 \cos\theta_m}\sum_{\mu\in\mathbb{Z}} c_\mu^+ \alpha_\mu \exp(\alpha_\mu \ell)\phi_m^{(\mu)} \equiv (1-\rho)\tilde{A}_m^{+(n)}(\ell). \tag{48}$$

Hence we have derived some analogue of the Fresnel formulas, namely

$$\rho = \frac{\tilde{A}_m^{+(n)}(\ell) - A_m^{+(n)}(\ell)}{\tilde{A}_m^{+(n)}(\ell) + A_m^{+(n)}(\ell)}, \tag{49}$$

and

$$\tau = \frac{2\tilde{A}_m^{+(n)}(\ell)A_m^{+(n)}(\ell)}{\tilde{A}_m^{+(n)}(\ell) + A_m^{+(n)}(\ell)}. \tag{50}$$

Summarizing, the boundary conditions at $z=\ell$ imply some nonlinear mapping from the vector $A^{+(n)}(\ell)$ onto the vector $A^{-(n)}(\ell)$ by means of

$$A_m^{-(n)}(\ell) = \rho A_m^{+(n)}(\ell) = \frac{\tilde{A}_m^{+(n)}(\ell) - A_m^{+(n)}(\ell)}{\tilde{A}_m^{+(n)}(\ell) + A_m^{+(n)}(\ell)}A_m^{+(n)}(\ell). \tag{51}$$

The analogous calculation at $z=0$ yields

$$A_m^{+(n+1)}(0) = \rho A_m^{-(n)}(0) = \frac{\tilde{A}_m^{-(n)}(0) - A_m^{-(n)}(0)}{\tilde{A}_m^{-(n)}(0) + A_m^{-(n)}(0)}A_m^{-(n)}(0). \tag{52}$$

These transformation formulas constitute a system of recursion relations, which, together with the initial condition (42) and the evolution equations (36) and (40), completely determine the vectors $A^{+(n)}(z)$ and $A^{-(n)}(z)$ for all $n\in\mathbb{N}$. The exact solution (resp. the CWT solution) then can be written as

$$\mathbf{E} = \begin{pmatrix} 0 \\ E \\ 0 \end{pmatrix} e^{-i\omega t}\exp(ikx\sin\theta) \times \sum_{m\in\mathbb{Z}}\exp(-imKx)\sum_{n\in\mathbb{N}}(A_m^{+(n)}(z) + A_m^{-(n)}(z)). \tag{53}$$

It is very plausible that the infinite sum over $n\in\mathbb{N}$ converges due to the exponential damping in the sequence of generations.

3. COMPARISON BETWEEN EXACT NUMERICAL SOLUTIONS AND KOGELNIK'S COUPLED-WAVE THEORY

It is the aim of this paper to explain the success of the CWT by arguments from perturbation theory. In view of the extensive literature on the CWT it is not necessary to provide an extended study of its validity. Rather it will be sufficient to consider just one realistic case. Our arguments to be presented in the next sections presuppose that the Fourier coefficients of the permittivity satisfy $\epsilon_\nu \sim \gamma^{|\nu|}$ for some small perturbational parameter γ. Hence we will consider a scenario in which this is the case. We will choose a permittivity of the form

$$\epsilon(x) = [(n_0 + n_1\cos(Kx+\Phi_n)) \times (1 + i(\kappa_0 + \kappa_1\cos(Kx+\Phi_\kappa)))]^2, \tag{54}$$

which can be regarded as a test hologram written by the superposition of two identical, mutually tilted coherent plane waves; see [7]. Consequently,

$$K = 2k_0 \sin\theta_B, \tag{55}$$

which defines the Bragg angle θ_B as the angle by which the two writing beams are tilted relative to the optical axis. By Snell's law (9) we can define another Bragg angle θ_b

$$K = 2k\sin\theta_b, \tag{56}$$

which refers to the values inside the crystal.

For the values shown in Table 1 it follows that the parameters n_1 and κ_1 that determine the amplitude of the modulation are almost four orders of magnitude smaller than the mean value n_0 (or, what amounts to the same, $|n_0(1+i\kappa_0)|$) of the permittivity. Hence the above condition for applying perturbation theory is satisfied.

The exact solution (53) of the diffraction problem involves two infinite summations, namely over all Fourier orders $m\in\mathbb{Z}$ and over all generations $n\in\mathbb{N}$. Moreover, it depends on the eigenvalues and eigenvectors of the the infinite-dimensional matrix $\mathbb{B}$. For a numerical evaluation of Eq. (53) we have to truncate this matrix. We choose a truncation to values $|m|\le 10$. For the CWT the matrix $\mathbb{B}$ is, by definition, a truncated 2×2 matrix. Next consider the summation over $n\in\mathbb{N}$. Here we recursively calculate the contribution of the generations for the exact solution or for the CWT according to Eqs. (51) and (52). This recursion loop will be exited if the relative correction of the squared amplitudes turns out to be less than 10^{-10}. As the crucial quantity to be calculated we choose exactly what would be obtained by a measurement of the transmitted intensity, namely the square amplitude $|A_1|^2$ of the superposition of all transmitted waves with diffraction order $m=1$. For a test we have also calculated $|A_0|^2$ and $|A_{-1}|^2$, but not displayed these results as figures. As the independent variables of the calculation we choose the thickness ℓ of the crystal sample and the angle θ_0 of the incident plane wave. The latter is varied between $0.5\theta_B$ and $1.5\theta_B$ in steps of

Table 1. Parameters Used for the Exact Solution and the CWT Solution

Description	Symbol	Value
Wavelength	λ_0	5.43×10^{-7} m
Permittivity parameters according to Eq. (54)	n_0	2.1776
	n_1	0.00026
	Φ_n	0
	K	3.12622×10^{6} m^{-1}
	κ_0	4.76237×10^{-4}
	κ_1	2.3812×10^{-4}
	Φ_κ	$\pi/2$
Bragg angle	θ_B	0.1355

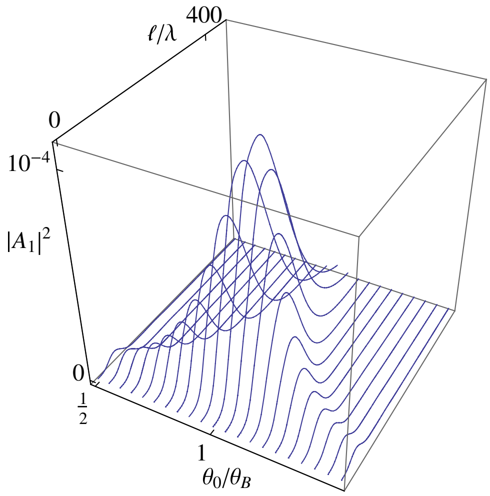

Fig. 2. Square amplitude $|A_1|^2$ of the first diffraction order of the transmitted wave according to the exact solution of the diffraction problem and the values shown in Table 1. The two independent variables are the thickness of the crystal ℓ in units of the wavelength λ_0 and the angle θ_0 of the incident plane wave in units of the Bragg angle θ_B.

$0.1\theta_B$. The thickness ℓ is varied from λ to $10^3\lambda$, where $\lambda = (2\pi/k)$ is the wavelength of the $m = 0$ component of the solution inside the crystal; see Subsection 2.A. Moreover, we have chosen the physical units such that the amplitude of the incoming plane wave assumes the value $E_\mathrm{I} = 1$.

The result of this calculation for the exact solution is shown in Fig. 2. We stress that the various curves do not show the z dependence of some solution inside the crystal, but rather the transmitted intensity $|A_1|^2$ of diffraction order $m = 1$ as a function of the thickness ℓ. As expected, this intensity is maximal for $\theta_0 = \theta_B$ and for values of ℓ of a few hundred wavelengths. The corresponding figure for the CWT is not shown since the difference δ to the exact solution is below visibility. Rather we have displayed the relative difference $\delta/\max_\ell |A_1|^2$ as a function of θ_0 and ℓ in Fig. 3. Obviously the relative difference increases with $|\theta_0 - \theta_B|$ but never exceeds the values of 10^{-3}. Moreover, the exact solution curves

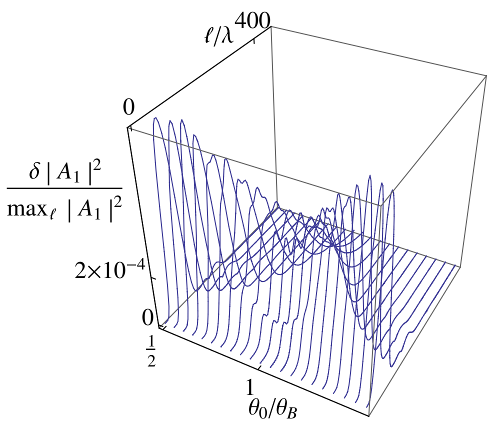

Fig. 3. Difference $\delta|A_1|^2$ between the square amplitudes of the exact solution according to Fig. 2 and those of the corresponding CWT divided by the maximum of $|\tilde{A}_1|^2$ w. r. t. ℓ and fixed θ_0. The independent variables are the same as in Fig. 2. Note that the relative difference increases with the deviation of θ_0 from the Bragg angle but never exceeds 10^{-3}.

for the Fourier order $m = -1$ will be discussed in the next section. We also mention that the quality of the CWT as an approximation for $|A_0|^2$ is even better than for $|A_1|^2$. We have not displayed the corresponding curves since the zeroth Fourier order is less relevant for experimental purposes.

These results show a surprisingly good overall quality of the CWT, at least for the case under consideration, that seems to contradict the results of [9]. The difference may, however, be due to the effect of damping not considered in [9].

4. CWT AND PERTURBATION ANALYSIS

We write the matrix $\mathbb{B}$ in a slightly more general form as

$$\begin{pmatrix} \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\ \ddots & b^-\gamma & B_{-2} & b^+\gamma & 0 & 0 & 0 & 0 & \ddots \\ \ddots & 0 & b^-\gamma & B_{-1} & b^+\gamma & 0 & 0 & 0 & \ddots \\ \ddots & 0 & 0 & b^-\gamma & 0 & b^+\gamma & 0 & 0 & \ddots \\ \ddots & 0 & 0 & 0 & b^-\gamma & B_1 & b^+\gamma & 0 & \ddots \\ \ddots & 0 & 0 & 0 & 0 & b^-\gamma & B_2 & b^+\gamma & \ddots \\ \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \end{pmatrix}. \tag{57}$$

Here we have introduced a formal perturbation parameter γ in order to distinguish between the different orders of magnitudes of the entries in $\mathbb{B}$. According to the parameters used in Section 3, $\gamma \sim 10^{-3}\ldots10^{-4}$, which means that perturbation theory should be the method of choice and would provide excellent approximations. All terms of second or higher order in γ are set to zero since we will content ourselves with first-order perturbation. Hence a tridiagonal matrix with constant secondary diagonals results. Moreover, it is convenient to divide the original matrix (20) by the factor k_0^2 in order to obtain diagonal elements of order 1.

In general it will be the case that the unperturbed diagonal elements B_n, $n \in \mathbb{Z}$, are pairwise different and we can invoke nondegenerate perturbation theory. But there is an important exception: comparison with Eq. (20) shows that $B_1 = K(2k \sin\theta - K)/k_0^2$ and hence $B_1 = 0 = B_0$ in the in-Bragg case where $2k \sin\theta = K = 2k \sin\theta_b$ or $\theta = \theta_b$. This means that, in the in-Bragg case, we have to use (twofold) degenerate perturbation theory, which is markedly different from the nondegenerate case. Still in the case $\theta \approx \theta_b$, such that B_1 is of the order γ, degenerate perturbation theory would work. On the other hand, nondegenerate perturbation theory requires that $B_1 = K(2k \sin\theta - K)/k_0^2$ should be of order 1, and hence only works in the extreme off-Bragg case. Thus a strict perturbational approach would be faced with the problem to find approximations in the intermediate regime between the almost-Bragg case and the extreme off-Bragg case.

Before discussing this problem we will investigate the two cases of perturbation analysis in more detail. Let us begin with the twofold degenerate case, i.e., $B_1 = 0 = B_0$, in Eq. (57). Here perturbation theory tells us to first diagonalize the submatrix U within the degenerated eigenspace of the unperturbed matrix, namely

$$U \equiv \begin{pmatrix} 0 & b^+\gamma \\ b^-\gamma & 0 \end{pmatrix}, \tag{58}$$

where we have neglected the terms quadratic in γ according to the rules of first-order perturbation. The resulting two eigenvalues $x_{0,1} = \pm\gamma\sqrt{b^+b^-}$ together with the remaining diagonal elements A_n are eigenvalues of $\mathbb{B}$ in the zeroth order of γ. The corresponding eigenvectors of U together with the remaining standard unit vectors $|m\rangle$, $m \in \mathbb{Z}$, $m \neq 0$, 1 are the zeroth-order eigenvectors of $\mathbb{B}$. The next corrections are linear in γ. As explained in Section 2 the eigenvalues and eigenvectors of $\mathbb{B}$ are used to obtain the exact solution. Hence the zeroth-order perturbation theory as well as the CWT already provide excellent approximations of the measurable squared amplitudes $|A_0|^2$ and $|A_1|^2$ in the in-Bragg case. In the off-Bragg case the zeroth order only yields the leading approximation of $|A_0|^2$, not of $|A_1|^2$, as we will see below.

As an interesting fact, albeit without consequences for the remainder of the paper, we mention that in the in-Bragg case the diagonal elements of $\mathbb{B}$ read $\mathbb{B}_{mm} = K^2 m(1-m)$ and hence exhibit a certain reflection symmetry. This can be most conveniently utilized by adopting a slightly asymmetric truncation of the Fourier orders from $-m$ to $m+1$. Let $\mathbb{B}^{(k)}$ denote the $k \times k$ principal submatrix of the truncated matrix $\mathbb{B}$ and p_k its characteristic polynomial. Then, due to the mentioned reflection symmetry, the characteristic polynomial p of the truncated matrix $\mathbb{B}$ factorizes according to

$$p = \left(p_{m+1} - \gamma\sqrt{b^+b^-}p_m\right)\left(p_{m+1} + \gamma\sqrt{b^+b^-}p_m\right). \tag{59}$$

Hence $\mathbb{B}$ can be exactly diagonalized when truncated according to the values $m = -3, \ldots, 4$.

We now turn to the extreme off-Bragg case $|\theta - \theta_B| \sim |\theta_B|$. Here the diagonal elements B_m of $\mathbb{B}$ are pairwise different, and we may apply nondegenerate perturbation theory. The eigenvalues $b^{(m)}$ of $\mathbb{B}$ coincide with B_m up to corrections quadratic in γ. The corresponding eigenvectors assume the form

$$\phi_\nu^{(m)} = \delta_{m\nu} + \gamma(\delta_{\nu,m+1}\Psi_{m,\nu} + \delta_{\nu,m-1}\Psi_{m,\nu}) + \mathcal{O}(\gamma^2). \tag{60}$$

This implies that the diffraction problem with dominating order $m = 0$ can be well approximated by restriction to the three-dimensional subspace with Fourier orders $m = -1$, 0, 1. The restriction of CWT to the smaller two-dimensional subspace with $m = 0$, 1, however, remains to be justified; see below. The linear approximation to the zeroth generation of the exact solution (35) can be written as follows (we suppress the generation index and the superscript$^+$):

$$\begin{aligned} A(z) &= \Phi\Delta(z)\Phi^{-1}A(0) \\ &= \Delta_0(z)A(0) + \gamma(-\Delta_0(z)\Psi A(0) + \Psi\Delta_0(z)A(0)) + \mathcal{O}(\gamma^2), \end{aligned} \tag{61}$$

where the matrix elements of Ψ are implicitly defined by Eq. (60) and $\Delta_0(z)$ is the zeroth perturbation order approximation of Eq. (33), i.e.,

$$(\Delta_0(z))_{mm} = \exp\left(-\sqrt{(q-ik)^2\cos^2\theta - B_m}\,z\right). \tag{62}$$

First-order perturbation theory of tridiagonal matrices yields the explicit expression

$$\Psi_{m,m\pm1} = \frac{b^\pm}{B_m - B_{m\pm1}}. \tag{63}$$

Together with Eq. (42) this yields

$$A_{\pm1}(z) = -\frac{b^\pm}{B_{m\pm1}}(e^{\alpha z} - e^{\mu^\pm z}) + \mathcal{O}(\gamma^2), \tag{64}$$

where

$$\alpha \equiv -(q-ik)\cos\theta, \tag{65}$$

and

$$\mu^\pm \equiv -\sqrt{(q-ik)^2\cos^2\theta - B_{\pm1}}. \tag{66}$$

It can be shown that the above approximation of $A_1(z)$ coincides with that of CWT up to corrections quadratic in γ. This explains the success of the CWT in the off-Bragg case, as far as the Fourier order $m = 1$ is concerned.

Before turning to the question of why the Fourier order $m = -1$ can be neglected, it will be appropriate to briefly discuss the approximate solution (64). Let us first ignore the prefactor and consider its absolute square

$$\begin{aligned} |e^{\alpha z} - e^{\mu^\pm z}|^2 &= |e^{(\alpha_1 + i\alpha_2)z} - e^{(\mu_1^\pm + i\mu_2^\pm)z}|^2 \\ &= e^{2\alpha_1 z} + e^{2\mu_1^\pm z} - 2e^{(\alpha_1+\mu_1^\pm)z} \times \cos((\alpha_2 - \mu_2^\pm)z). \end{aligned} \tag{67}$$

If this expression assumes a maximal value at $z_\pm$ it follows that

$$0 = \alpha_1 e^{(\alpha_1-\mu_1^\pm)z_\pm} + \mu_1^\pm e^{-(\alpha_1-\mu_1^\pm)z_\pm} + A^\pm \sin((\alpha_2 - \mu_2^\pm)z_\pm + \phi^\pm), \tag{68}$$

where

$$A^\pm \equiv \sqrt{(\alpha_1 + \mu_1^\pm)^2 + (\alpha_2 - \mu_2^\pm)^2}, \tag{69}$$

and

$$\phi^\pm \equiv \arctan\frac{\alpha_1 + \mu_1^\pm}{\alpha_2 - \mu_2^\pm}. \tag{70}$$

For an approximate solution of Eq. (68) we make some assumptions that are satisfied for the parameters of Table 1 and $\theta_0 \approx (1/2)\theta_B$. First we assume that the exponential terms in Eq. (68) can be linearized about the values $z = 0$ and, secondly, that the sin function in Eq. (68) can be linearized about the value $z = (\pi/(\alpha_2 - \mu_2^\pm))$. This gives an approximation of the z value for the first maximum of Eq. (67) of the form

$$z_\pm \approx \frac{\alpha_1 + \mu_1^\pm + A^\pm(\pi + \phi^\pm)}{-(\alpha_1 - \mu_1^\pm)^2 + A^\pm(\alpha_2 - \mu_2^\pm)}. \tag{71}$$

We have chosen $\theta_0 = (1/2)\theta_B$ in what follows since the maximum of $|A_{-1}(z)|^2$ is larger for small values of θ_0. For the actual parameters an even cruder approximation is possible:

$$z_\pm \approx \frac{\pi + \phi^\pm}{\alpha_2 - \mu_2^\pm}; \tag{72}$$

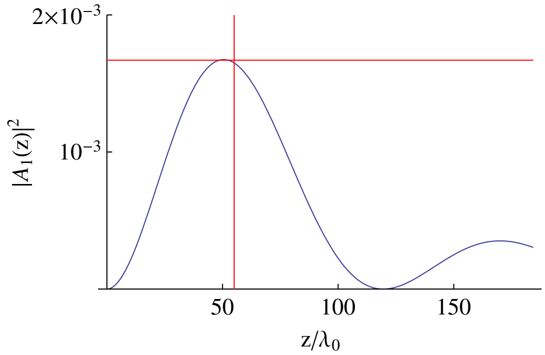

Fig. 4. Absolute square $|A_1(z)|^2$ of the $m = 1$ component of the exact solution (61) and the approximate coordinates (72), (73) of its first maximum. The parameters are chosen according to Table 1 and $\theta_0 = (1/2)\theta_B$.

see Figs. 4 and 5. As a suitable approximation for the maximum of the intensity we similarly obtain

$$|A_{\pm1}(z_\pm)|^2 \approx \left|\frac{\gamma b^\pm}{B_{\pm1}}\right|^2 |e^{\alpha_1 z_\pm} + e^{\mu_1^\pm z_\pm}|^2. \tag{73}$$

Now we can compare the two maximal values $|A_1(z_+)|^2$ and $|A_{-1}(z_-)|^2$. For the actual parameters the ratio of the prefactors is

$$\left|\frac{b^+ B_{-1}}{b^- B_1}\right|^2 \approx 82.1, \tag{74}$$

whereas

$$\left|\frac{A_1(z_+)}{A_{-1}(z_-)}\right|^2 \approx 51.5. \tag{75}$$

The difference between Eqs. (74) and (75) comes from the fact that for the actual parameters $z_- \approx 1.04426 \times 10^{-5}$ m $< z_+ \approx 2.98878 \times 10^{-5}$ m and hence the exponential terms in $|A_{-1}(z_-)|^2$ are larger than those in $|A_1(z_+)|^2$.

Summarizing, first-order perturbation theory explains well the possible restriction to the subspace with $m = -1$, 0, 1 and the good approximation of $A_1(z)$ by CWT. On the other hand, the dominance of the Fourier order of $m = 1$ over that with $m = -1$ by a factor of 50 for the intensity is a consequence of the chosen parameters and may be different for other cases.

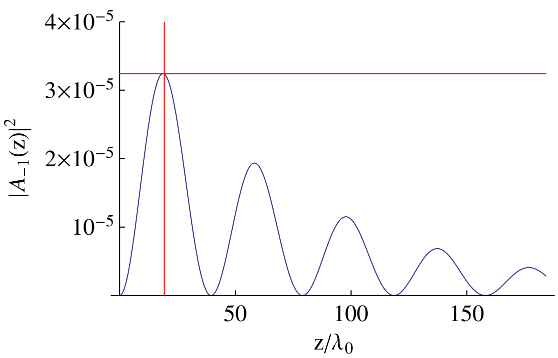

Fig. 5. Absolute square $|A_{-1}(z)|^2$ of the $m = -1$ component of the exact solution (61) and the approximate coordinates (72), (73) of its first maximum. The parameters are chosen according to Table 1 and $\theta_0 = (1/2)\theta_B$.

Finally, we come back to the problem how to treat the domain between the almost-Bragg case and the extreme off-Bragg case, where no perturbation theory seems to be available. We have shown that CWT coincides with the results of perturbational analysis in both extremal cases for the relevant perturbation order (zeroth-order in the in-Bragg case and first-order in the extreme off-Bragg case). Hence our problem has an obvious solution: the CWT can be viewed as the natural interpolation between these extreme cases. The quality of the numerical approximation of CWT also in the intermediate region (see Section 3) supports this view. We have thus obtained an additional explanation of the surprising success of the CWT that goes beyond its mere coincidence with perturbation theory in the extreme cases.

5. SUMMARY

If one has to solve holographic diffraction problems one can either seek numerical solutions based on rigorous approaches or use relatively simple analytical approximations as Kogelnik's CWT. In this paper we have tried to bridge the gap between these different strategies by explaining the success of Kogelnik's approximation in terms of perturbation theory applied to the underlying infinite matrix problem. Our explanation might be the starting point for investigating the limits of CWT, which is beyond the scope of the work at hand since we have restricted ourselves to a realistic but limited set of physical parameters.

REFERENCES

1. D. Gabor, "A new microscopic principle," Nature **161**, 777–778 (1948).
2. M. Imlau, M. Fally, H. Coufal, G. Burr, and G. Sincerbox, "Holography and optical storage," in *Springer Handbook of Lasers and Optics*, F. Träger, ed. (Springer, 2007), pp. 1205–1249.
3. M. Imlau, S. Odoulov, T. Woike, and S. Bieringer, "Holographic data storage," in *Nanoelectronics and Information Technology*, R. Waser, ed. (Wiley-VCH, 1999), pp. 727–750.
4. S. Tay, P. A. Blanche, R. Voorakaranam, A. V. Tunc, W. Lin, S. Rokutanda, T. Gu, D. Flores, P. Wang, G. Li, P. St. Hilaire, J. Thomas, R. A. Norwood, M. Yamamoto, and N. Peyghambarian, "An updatable holographic three-dimensional display," Nature **451**, 694–698 (2008).
5. D. Cotter, R. J. Manning, K. J. Blow, A. D. Ellis, A. E. Kelly, D. Nesset, D. I. D. Phillips, A. J. Poustie, and D. C. Rogers, "Nonlinear optics for high-speed digital information processing," Science **286**, 1523–1528 (1999).
6. M. J. O'Mahony, C. Politi, D. Klonidis, R. Nejabati, and D. Simeonidou, "Future optical networks," J. Lightwave Technol. **24**, 4684–4696 (2006).
7. K.-M. Voit and M. Imlau, "Holographic spectroscopy: wavelength-dependent analysis of photosensitive materials by means of holographic techniques," MDPI Materials **6**, 334–358 (2013).
8. E. Guibelalde, "Coupled wave analysis for out-of-phase mixed thick hologram gratings," Opt. Quantum Electron. **16**, 173–178 (1984).
9. M. G. Moharam and T. K. Gaylord, "Rigorous coupled-wave analysis of planar-grating diffraction," J. Opt. Soc. Am. **71**, 811–818 (1981).
10. N. Uchida, "Calculation of diffraction efficiency in hologram gratings attenuated along the direction perpendicular to the grating vector," J. Opt. Soc. Am. **63**, 280–287 (1973).
11. N. V. Kukhtarev, V. B. Markov, S. G. Odulov, M. S. Soskin, and V. L. Vinetskii, "Holographic storage in electrooptic crystals. I. Steady state," Ferroelectrics **22**, 949–960 (1978).
12. N. V. Kukhtarev, V. B. Markov, S. G. Odulov, M. S. Soskin, and V. L. Vinetskii, "Holographic storage in electrooptic crystals. II.

Beam coupling–light amplification," Ferroelectrics **22**, 961–964 (1978).

13. P.St.J. Russell, "Optical volume holography," Phys. Rep. **71**, 209–312 (1981).
14. J. Strait, J. D. Reed, and N. V. Kukhtarev, "Orientational dependence of photorefractive two-beam coupling in InP:Fe," Opt. Lett. **15**, 209–211 (1990).
15. B. I. Sturman, D. J. Webb, R. Kowarschik, E. Shamonina, and K. H. Ringhofer, "Exact solution of the Bragg-diffraction problem in sillenites," J. Opt. Soc. Am. B **11**, 1813–1819 (1994).
16. J. T. Sheridan, "Generalization of the boundary diffraction method for volume gratings," J. Opt. Soc. Am. A **11**, 649–656 (1994).
17. M. R. Gleeson, J. Guo, and J. T. Sheridan, "Optimisation of photopolymers for holographic applications using the non-local photo-polymerization driven diffusion model," Opt. Express **19**, 22423–22436 (2011).
18. H. Kogelnik, "Coupled wave theory for thick hologram gratings," Bell Syst. Tech. J. **48**, 2909–2947 (1969).
19. G. Montemezzani and M. Zgonik, "Light diffraction at mixed phase and absorption gratings in anisotropic media for arbitrary geometries," Phys. Rev. E **55**, 1035–1047 (1997).
20. B. I. Sturman, E. V. Podivilov, K. H. Ringhofer, E. Shamonina, V. P. Kamenov, E. Nippolainen, V. V. Prokofiev, and A. A. Kamshilin, "Theory of photorefractive vectorial wave coupling in cubic crystals," Phys. Rev. E **60**, 3332–3352 (1999).
21. M. Prijatelj, J. Klepp, Y. Tomita, and M. Fally, "Far-off-Bragg reconstruction of volume holographic gratings: a comparison of experiment and theories," Phys. Rev. A **87**, 063810 (2013).