Green's function method to study thin diffraction gratings

Daniel A. Travo,[1,*] Rodrigo A. Muniz,[1] M. Liscidini,[2] and J. E. Sipe[1]

[1]*Department of Physics, University of Toronto, Toronto, Ontario M5S 1A7, Canada*
[2]*Department of Physics, University of Pavia, Pavia I-27100, Italy*

The anomalous features in diffraction patterns first observed by Wood over a century ago have been the subject of many investigations, both experimental and theoretical. The sharp, narrow structures—and the large resonances with which they are sometimes associated—arise in numerous studies in optics and photonics. In this paper we present an analytical method to study diffracted fields of optically thin gratings that highlights the nonanalyticities associated with the anomalies. Using this approach we can immediately derive diffracted fields for any polarization in a compact notation. While our equations are approximate, they fully respect energy conservation in the electromagnetic field, and describe the large exchanges of energy between incident and diffracted fields that can arise even for thin gratings.

I. INTRODUCTION

Over a century ago, Wood observed anomalies in the angular dependence of light reflected from a metal sheet [1], and since then there have been many studies of these anomalies and their applications in optics and photonics, particularly as they arise in the reflection from surfaces on which a grating is intentionally deposited. Maystre [2] has recently presented a detailed review of the history of research in this field. Generally, two mechanisms have been identified as sources of the anomalies: The first is a transition between propagation and evanescence in one of the diffracted orders [3], and the second is the excitation of a leaky mode within the grating region [4]. These anomalies have been discussed for a number of years [5] and some ambiguity exists in the literature concerning the distinction between these mechanisms and the terms used to refer to them. In this paper, we follow the convention that anomalies related to a transition in a diffracted order are referred to as *Rayleigh anomalies*, and those associated with the resonant excitation of a leaky mode in the grating region are referred to as *Wood anomalies*. Maradudin [6] has recently shown that the resonant excitation of any surface wave in a substrate below the grating, by scattering from the grating, can lead to anomalies in the reflectance as well; we also refer to these as Wood anomalies.

Rayleigh anomalies are square-root-like, sharp, narrow peaks that arise in the irradiance of both the specularly reflected and diffracted beams. Wood anomalies are associated with extraordinary increases in the specular reflectance [7–9], and have seen wide use in applications where gratings serve as filters [10–12], modulators [13,14], and sensors [15,16]. Because even very thin gratings can lead to very large effects on the reflectivity in the region of these anomalies, the simplest perturbation theories are not sufficient to describe them; it is essential to consider the full interaction between diffracted and specularly reflected beams. Over the years a wide range of approaches have arisen to treat such systems. These include guided-mode techniques such as coupled-mode theory [17–20], transfer-matrix approaches [10], and a variety

of robust numerical techniques based on finite-element and RCWA (scattering matrix) methods [21–24].

In this paper we present a semianalytic method for the treatment of thin gratings, with advantages that are not all present in earlier work. We consider the grating structure shown in Fig. 1(a) in this first communication. Based on a Green's function formalism [25] that treats the scattered light in terms of its *s*- and *p*-polarized components, the method leads to an immediate identification of the features in the scattering equations that describe the anomalies, and allows for the easy inclusion of effects of surface waves of the substrate as well as leaky modes in the grating region. Light with any plane of incidence, any polarization, and at any incident angle is treated, and anisotropy in the response of the material in the grating region is included; the substrate can consist of an arbitrary set of layers with uniaxial optical properties. The description of the reflected and diffracted light is necessarily approximate, since we simplify our equations based on the grating being thin, but it is nonetheless completely robust with respect to energy conservation: In the absence of any absorption in the material media, at whatever number the inclusion of diffracted and evanescent fields in the calculation is truncated, the approximate equations respect conservation of energy in the electromagnetic field, despite large exchanges of energy between diffracted and reflected fields. For simple incidence configurations, and if only a few diffracted orders are important, the set of equations to be solved is small and the physics easily identified. This is an important advantage, since gratings are now being used to access resonances for enhanced sensing applications [26,27] and in novel 2D materials such as graphene [28–31]. Our simple but robust treatment of the optics of the grating should allow for such work to focus on the physics of the medium being probed.

The outline of the paper is as follows. In Sec. II we first treat the simpler, symmetric grating structure shown in Fig. 1(b). In the limit of a thin grating, we show how the scattering equations lead naturally to the assignment of a uniform dielectric tensor for a layer associated with the grating region; see Fig. 1(c). The scattering by the grating can be best understood as occurring with this as part of the background optical response, and it is the waveguide modes of this nominal layer that become the resonances associated with the Wood

---
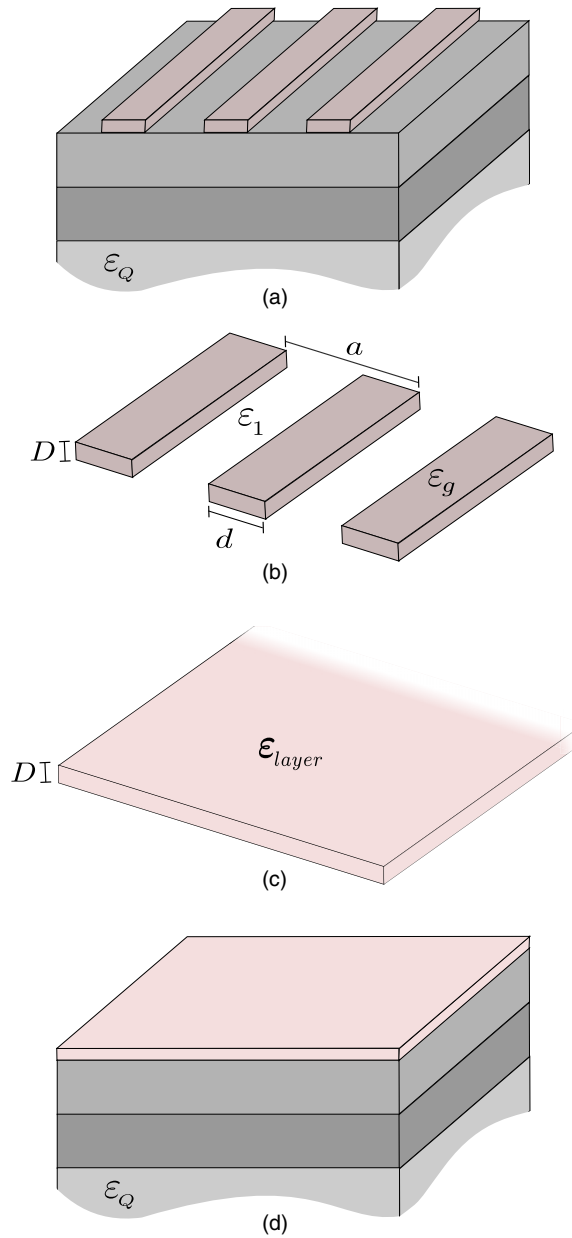
*dtravo@physics.utoronto.ca

205140-1

FIG. 1. The general structure studied and discussed in this paper (not to scale). (a) A thin grating placed on top of a multilayer structure with a substrate with relative dielectric constant ε_Q. The grating parameters are the same as in (b), and the relative dielectric constant of the cladding is ε_1. (b) An isolated grating with relative dielectric tensor ε_g suspended in a medium with relative dielectric constant ε_1. (c) An effective dielectric slab with relative dielectric tensor $\boldsymbol{\varepsilon}_{\text{layer}}$ that characterizes the average properties of the grating relation. (d) The corresponding effective planar structure consisting of the effective dielectric slab and the multilayer structure.

anomalies, discussed alongside Rayleigh anomalies in Sec. III and identified in the resulting scattering equations in Sec. IV. In Sec. IV A we build a scattering matrix for the problem. This can of course be done in many ways, but we adopt an approach that leads to a proof that the equations respect energy conservation, and allows for an easy generalization to include an arbitrary layered substrate [Fig. 1(a)]. These

equations are separated by polarization and simplified in Sec. IV B for a simple configuration chosen as an example. In Sec. IV C a two-wave-vector model is used to derive analytic expressions for the scattered fields alongside a discussion of their poles that signal the Wood anomalies. We discuss how the Wood anomalies associated with the waveguide modes of the grating region [Fig. 1(c)] are modified—or disappear— in the presence of the substrate in Sec. IV D and present, as a sample calculation, results for a simple silicon grating atop a glass substrate and confirm the validity of our approximate treatment by comparison with convergent, numerically exact calculations. We focus on a dielectric system to highlight the pronounced spectral features associated with the excitation of a Wood anomaly. As derived the method presented is applicable to a wider range of materials, such as metallic gratings. Our conclusions are presented in Sec. V. Some of the details of the derivations, a discussion of waveguide dispersion, and our proof of energy conservation are relegated to appendices.

II. THIN GRATINGS

We begin by considering a grating in the region $-D/2 < z < D/2$, with the rest of space taken to be filled with an isotropic dielectric; see Fig. 1(b). In the presence of a field $\mathbf{E}_{\text{in}}(\mathbf{r},t)$ incident on the grating region we write the total field as

$$\mathbf{E}(\mathbf{r},t) = \mathbf{E}_{\text{in}}(\mathbf{r},t) + \mathbf{E}_{sc}(\mathbf{r},t), \tag{1}$$

where $\mathbf{E}_{sc}(\mathbf{r},t)$ is the scattered field. We take all such time-dependent fields $\mathbf{F}(\mathbf{r},t)$ to be stationary,

$$\mathbf{F}(\mathbf{r},t) = \mathbf{F}(\mathbf{r})e^{-i\omega t} + \text{c.c.},$$

and we assume the refractive index of the surrounding dielectric, $n_1 = \sqrt{\varepsilon_1}$, is real at frequency ω. Denoting by $\mathbf{R}$ the projection in the xy plane of a position vector $\mathbf{r} = \mathbf{R} + z\hat{z}$, we take $\hat{e}$ to be the unit vector in the xy plane that identifies the direction in which the susceptibility varies, and write the (possibly complex) spatially dependent (tensor) susceptibility in the grating region as $\boldsymbol{\chi}(\zeta)$, where $\zeta = \hat{e} \cdot \mathbf{R}$.

It will be convenient to Fourier-transform our field amplitudes $\mathbf{F}(\mathbf{r})$ only in the xy plane,

$$\mathbf{F}(\mathbf{r}) = \mathbf{F}(\mathbf{R}; z) = \int \frac{d\boldsymbol{\kappa}}{(2\pi)^2} e^{i\boldsymbol{\kappa} \cdot \mathbf{R}} \mathbf{F}(\boldsymbol{\kappa}; z), \tag{2}$$

where $\boldsymbol{\kappa}$ has x and y components, so for example

$$\mathbf{E}(\mathbf{R}; z) = \mathbf{E}_{\text{in}}(\mathbf{R}; z) + \mathbf{E}_{sc}(\mathbf{R}; z),$$
$$\mathbf{E}(\boldsymbol{\kappa}; z) = \mathbf{E}_{\text{in}}(\boldsymbol{\kappa}; z) + \mathbf{E}_{sc}(\boldsymbol{\kappa}; z). \tag{3}$$

In the usual example of an incident plane wave, the incident field will be characterized by a single $\boldsymbol{\kappa}_{\text{in}}$, and there will be scattered fields characterized by $\boldsymbol{\kappa}_{\text{in}} + m\mathbf{K}$, where m ranges over all positive and negative integers, and

$$\mathbf{K} = \frac{2\pi}{a}\hat{e}, \tag{4}$$

where a is the fundamental period of the grating. For $|\boldsymbol{\kappa}_{\text{in}} + m\mathbf{K}| > \tilde{\omega}n_1$, where $\tilde{\omega} \equiv \omega/c$, the scattered fields are evanescent, confined to the neighborhood of the grating region; for $|\boldsymbol{\kappa}_{\text{in}} + m\mathbf{K}| < \tilde{\omega}n_1$ the scattering leads to diffracted fields that can carry energy away from the grating region.

We write

$$\boldsymbol{\chi}(\zeta) = \boldsymbol{\chi}_1 + \boldsymbol{\chi}_{\text{add}}(\zeta), \qquad (5)$$

where in dyadic form

$$\boldsymbol{\chi}_1 = (\varepsilon_1 - 1)(\hat{\boldsymbol{x}}\hat{\boldsymbol{x}} + \hat{\boldsymbol{y}}\hat{\boldsymbol{y}} + \hat{\boldsymbol{z}}\hat{\boldsymbol{z}})$$

is just the susceptibility that would be present were the background dielectric extended into the grating region, and $\boldsymbol{\chi}_{\text{add}}(\zeta)$ is an additional contribution that is responsible for the ζ dependence of the susceptibility in the grating region. We assume that one of the principal axes of $\boldsymbol{\chi}_{\text{add}}(\zeta)$ is the z axis, and we choose the x and y axes to coincide with the other principal axes,

$$\boldsymbol{\chi}_{\text{add}}(\zeta) = \hat{\boldsymbol{x}}\hat{\boldsymbol{x}}\chi^{xx}_{\text{add}}(\zeta) + \hat{\boldsymbol{y}}\hat{\boldsymbol{y}}\chi^{yy}_{\text{add}}(\zeta) + \hat{\boldsymbol{z}}\hat{\boldsymbol{z}}\chi^{zz}_{\text{add}}(\zeta). \qquad (6)$$

We let $\mathbf{P}(\mathbf{r})$ denote the polarization in the grating region, above and beyond that which would result if the grating region consisted solely of the background isotropic dielectric medium. Then

$$\mathbf{P}(\mathbf{R}; z) = \epsilon_0 \boldsymbol{\chi}_{\text{add}}(\zeta) \cdot \mathbf{E}(\mathbf{R}; z),$$
$$\text{for } -D/2 < z < D/2. \qquad (7)$$

The scattered field contribution to (3) is determined by

$$\mathbf{E}_{sc}(\boldsymbol{\kappa}; z) = \int_{-D/2}^{D/2} dz' \mathbf{G}(\boldsymbol{\kappa}; z - z') \cdot \mathbf{P}(\boldsymbol{\kappa}; z'), \qquad (8)$$

where the Green's function [25] is

$$\mathbf{G}(\boldsymbol{\kappa}; z - z') = \mathbf{g}(\boldsymbol{\kappa}; z - z') - \frac{1}{\epsilon_0 \varepsilon_1} \delta(z - z')\hat{\boldsymbol{z}}\hat{\boldsymbol{z}}, \qquad (9)$$

with

$$\mathbf{g}(\boldsymbol{\kappa}; z - z') = \frac{i\tilde{\omega}^2}{2\epsilon_0 w_1} \theta(z - z') e^{iw_1(z-z')}(\hat{\boldsymbol{s}}\hat{\boldsymbol{s}} + \hat{\boldsymbol{p}}_{1+}\hat{\boldsymbol{p}}_{1+})$$
$$+ \frac{i\tilde{\omega}^2}{2\epsilon_0 w_1} \theta(z' - z) e^{-iw_1(z-z')}(\hat{\boldsymbol{s}}\hat{\boldsymbol{s}} + \hat{\boldsymbol{p}}_{1-}\hat{\boldsymbol{p}}_{1-}),$$

and where

$$\hat{\boldsymbol{s}} = \hat{\boldsymbol{\kappa}} \times \hat{\boldsymbol{z}}, \quad \hat{\boldsymbol{p}}_{1\pm} = \frac{\kappa\hat{\boldsymbol{z}} \mp w_1\hat{\boldsymbol{\kappa}}}{\tilde{\omega}n_1} \qquad (10)$$

identify the s- and p-polarized field components of the radiated fields. Here $w_1 \equiv \sqrt{\tilde{\omega}^2\varepsilon_1 - \kappa^2}$, where $\kappa = |\boldsymbol{\kappa}|$; to ensure proper radiation conditions, the square root is made unique by taking $\text{Im}\sqrt{Z} \geqslant 0$, and taking $\text{Re}\sqrt{Z} \geqslant 0$ if $\text{Im}\sqrt{Z} = 0$.

Of the two terms on the right-hand side of (9), the second will typically lead to the larger contribution for our thin gratings of interest, and it can be dealt with explicitly. If we define a modified field,

$$\mathbf{E}_{\text{mod}}(\boldsymbol{\kappa}; z) = \mathbf{E}_{\text{in}}(\boldsymbol{\kappa}; z) + \int_{-D/2}^{D/2} \mathbf{g}(\boldsymbol{\kappa}; z - z') \cdot \mathbf{P}(\boldsymbol{\kappa}; z') dz', \qquad (11)$$

we have

$$\mathbf{E}(\mathbf{R}; z) = \mathbf{E}_{\text{mod}}(\mathbf{R}; z) - \frac{1}{\epsilon_0 \varepsilon_1}\hat{\boldsymbol{z}}\hat{\boldsymbol{z}} \cdot \mathbf{P}(\mathbf{R}; z),$$

and we can write the expression (7) for the polarization as

$$\mathbf{P}(\mathbf{R}; z) = \epsilon_0 \boldsymbol{\chi}_{\text{mod}}(\zeta) \cdot \mathbf{E}_{\text{mod}}(\mathbf{R}; z), \qquad (12)$$

where

$$\boldsymbol{\chi}_{\text{mod}}(\zeta) = \hat{\boldsymbol{x}}\hat{\boldsymbol{x}}\chi^{xx}_{\text{mod}}(\zeta) + \hat{\boldsymbol{y}}\hat{\boldsymbol{y}}\chi^{yy}_{\text{mod}}(\zeta) + \hat{\boldsymbol{z}}\hat{\boldsymbol{z}}\chi^{zz}_{\text{mod}}(\zeta), \qquad (13)$$

with

$$\chi^{xx}_{\text{mod}}(\zeta) \equiv \chi^{xx}_{\text{add}}(\zeta),$$
$$\chi^{yy}_{\text{mod}}(\zeta) \equiv \chi^{yy}_{\text{add}}(\zeta),$$
$$\chi^{zz}_{\text{mod}}(\zeta) \equiv \frac{\chi^{zz}_{\text{add}}(\zeta)}{1 + \chi^{zz}_{\text{add}}(\zeta)/\varepsilon_1}.$$

From (11) we see that as $D \to 0$ we typically have $\mathbf{E}_{\text{mod}}(\mathbf{R}; z) \to \mathbf{E}_{\text{in}}(\mathbf{R}; z)$, and so in that limit $\boldsymbol{\chi}_{\text{mod}}(\zeta)$ can be understood as an effective local susceptibility relating the (excess) polarization to the *incident field*, rather than to the field in the grating region itself.

So far we have made no approximations, and an exact description of the scattering could proceed by numerically solving (11) and (12) for any specified $\mathbf{E}_{\text{in}}(\boldsymbol{\kappa}; z)$. Instead, we develop an approximate description of the scattering based on the condition that the thickness D of the grating region is much less than the wavelength of light, $\tilde{\omega}n_1 D \ll 1$, and as well that the variation in z of the scattered fields over the grating region is negligible for $\boldsymbol{\kappa}$ of interest, $|w_1|D \ll 1$. This leads to the ansatz that a number of fields can be taken as independent of z within the grating region,

$$\mathbf{F}(\boldsymbol{\kappa}; z) \to \mathbf{F}(\boldsymbol{\kappa}) \quad \text{for } -D/2 < z < D/2, \qquad (14)$$

and for such fields we write

$$\mathbf{F}(\mathbf{R}) = \int \frac{d\boldsymbol{\kappa}}{(2\pi)^2} e^{i\boldsymbol{\kappa}\cdot\mathbf{R}} \mathbf{F}(\boldsymbol{\kappa}).$$

A full solution of the fields in the grating material would show variations over a distance on the order of $(\tilde{\omega}\sqrt{|\varepsilon_g|^{\text{max}}})^{-1}$, where $|\varepsilon_g|^{\text{max}}$ is the maximum of the absolute values of the eigenvalues of $\boldsymbol{\varepsilon}_g$, the dielectric tensor of the grating material. So we expect the condition $\tilde{\omega}\sqrt{|\varepsilon_g|^{\text{max}}}D \ll 1$ is also required for this approximation to be valid, as we confirm in our example calculations in Sec. IV. Naturally for the incident field we simply take $\mathbf{E}_{\text{in}}(\boldsymbol{\kappa}) = \mathbf{E}_{\text{in}}(\boldsymbol{\kappa}; 0)$, while to determine $\mathbf{P}(\boldsymbol{\kappa})$ self-consistently we approximate the field $\mathbf{E}_{\text{mod}}(\boldsymbol{\kappa}; z)$ as uniform over the grating region by taking $\mathbf{E}_{\text{mod}}(\boldsymbol{\kappa}; z) \to \mathbf{E}_{\text{mod}}(\boldsymbol{\kappa})$, where

$$\mathbf{E}_{\text{mod}}(\boldsymbol{\kappa}) = \mathbf{E}_{\text{in}}(\boldsymbol{\kappa}) + \frac{1}{D}\int_{-D/2}^{D/2} dz \int_{-D/2}^{D/2} dz' \mathbf{g}(\boldsymbol{\kappa}; z - z') \cdot \mathbf{P}(\boldsymbol{\kappa}).$$

In the limit $|w_1|D \ll 1$ this leads to

$$\mathbf{E}_{\text{mod}}(\boldsymbol{\kappa}) = \mathbf{E}_{\text{in}}(\boldsymbol{\kappa}) + \mathbf{g}(\boldsymbol{\kappa}) \cdot \mathbf{P}(\boldsymbol{\kappa}), \qquad (15)$$

where

$$\mathbf{g}(\boldsymbol{\kappa}) = \frac{i\tilde{\omega}^2 D}{2\epsilon_0 w_1}\left[\hat{\boldsymbol{s}}\hat{\boldsymbol{s}} + \frac{1}{2}(\hat{\boldsymbol{p}}_{1+}\hat{\boldsymbol{p}}_{1+} + \hat{\boldsymbol{p}}_{1-}\hat{\boldsymbol{p}}_{1-})\right]$$
$$= \frac{i\tilde{\omega}^2 D}{2\epsilon_0 w_1}\hat{\boldsymbol{s}}\hat{\boldsymbol{s}} + \frac{iw_1 D}{2\epsilon_0 \varepsilon_1}\hat{\boldsymbol{\kappa}}\hat{\boldsymbol{\kappa}} + \frac{i\kappa^2 D}{2\epsilon_0 \varepsilon_1 w_1}\hat{\boldsymbol{z}}\hat{\boldsymbol{z}},$$

and in this limit Eq. (12) reduces to

$$\mathbf{P}(\mathbf{R}) = \epsilon_0 \boldsymbol{\chi}_{\text{mod}}(\zeta) \cdot \mathbf{E}_{\text{mod}}(\mathbf{R}). \qquad (16)$$

Within these approximations the fields in the grating region are determined by the solution of (15) and (16).

At this point it is useful to separate out the spatial average of our various quantities. In particular for $\chi_{\mathrm{mod}}(\zeta)$ we have

$$\langle \chi_{\mathrm{mod}} \rangle = \frac{1}{a} \int_{-a/2}^{a/2} \chi_{\mathrm{mod}}(\zeta) d\zeta,$$

and we put

$$\chi_v(\zeta) \equiv \chi_{\mathrm{mod}}(\zeta) - \langle \chi_{\mathrm{mod}} \rangle. \tag{17}$$

Our equations (15) and (16) can then be written as

$$\mathbf{P}(\boldsymbol{\kappa}) = \epsilon_0 \langle \chi_{\mathrm{mod}} \rangle \cdot \mathbf{E}_{\mathrm{mod}}(\boldsymbol{\kappa}) + \mathbf{P}_v(\boldsymbol{\kappa}),$$

$$\mathbf{E}_{\mathrm{mod}}(\boldsymbol{\kappa}) = \mathbf{E}_{\mathrm{in}}(\boldsymbol{\kappa}) + \mathbf{g}(\boldsymbol{\kappa}) \cdot \mathbf{P}(\boldsymbol{\kappa}), \tag{18}$$

where

$$\mathbf{P}_v(\mathbf{R}) = \epsilon_0 \chi_v(\zeta) \cdot \mathbf{E}_{\mathrm{mod}}(\mathbf{R}) \tag{19}$$

is the only contribution from the variation of the effective susceptibility with ζ. If we define χ_{layer} according to

$$\chi_{\mathrm{layer}}^{xx} \equiv \langle \chi_{\mathrm{mod}}^{xx} \rangle = \langle \chi_{\mathrm{add}}^{xx} \rangle,$$

$$\chi_{\mathrm{layer}}^{yy} \equiv \langle \chi_{\mathrm{mod}}^{yy} \rangle = \langle \chi_{\mathrm{add}}^{yy} \rangle,$$

$$\frac{\chi_{\mathrm{layer}}^{zz}}{1 + \chi_{\mathrm{layer}}^{zz}/\varepsilon_1} \equiv \langle \chi_{\mathrm{mod}}^{zz} \rangle = \left\langle \frac{\chi_{\mathrm{add}}^{zz}}{1 + \chi_{\mathrm{add}}^{zz}/\varepsilon_1} \right\rangle, \tag{20}$$

where the last equation is to be solved for $\chi_{\mathrm{layer}}^{zz}$, we can identify χ_{layer} as the effective (excess) susceptibility of the thin layer that would lead to the optical response of the grating region were the variation $\chi_v(\zeta)$ in the effective susceptibility ignored; this is the scenario sketched in Fig. 1(c). For if we would return to (7) and (8), take $\chi_{\mathrm{add}}(\zeta) \to \chi_{\mathrm{layer}}$, and repeat the derivation and approximations leading to (18), we would recover precisely those equations with $\mathbf{P}_v(\boldsymbol{\kappa})$ absent, and with the components of $\langle \chi_{\mathrm{mod}} \rangle$ replaced by the components of χ_{layer} according to (20). Note that if we write the full relative dielectric tensor in the grating region as $\boldsymbol{\varepsilon}(\zeta) \equiv \boldsymbol{\varepsilon}_1 + \chi_{\mathrm{add}}(\zeta)$, and the full relative dielectric tensor associated with χ_{layer} as $\boldsymbol{\varepsilon}_{\mathrm{layer}} \equiv \boldsymbol{\varepsilon}_1 + \chi_{\mathrm{layer}}$, where $\boldsymbol{\varepsilon}_1 = \varepsilon_1(\hat{x}\hat{x} + \hat{y}\hat{y} + \hat{z}\hat{z})$, we have

$$\varepsilon_{\mathrm{layer}}^{xx} = \langle \varepsilon^{xx} \rangle,$$

$$\varepsilon_{\mathrm{layer}}^{yy} = \langle \varepsilon^{yy} \rangle,$$

$$\frac{1}{\varepsilon_{\mathrm{layer}}^{zz}} = \left\langle \frac{1}{\varepsilon^{zz}} \right\rangle. \tag{21}$$

We return to Eqs. (18), and can now understand them as describing the scattering due to a variation in the effective excess susceptibility, $\chi_v(\zeta)$, in the presence of a uniform background dielectric tensor $\boldsymbol{\varepsilon}_{\mathrm{layer}}$ in the grating region. Below we will construct an expression for $\mathbf{P}_v(\boldsymbol{\kappa})$, and then these equations can be solved consistently for $\mathbf{P}(\boldsymbol{\kappa})$. Once that is done we can construct the scattered fields above the grating region ($z > D/2$) and below the grating region ($z < -D/2$). We denote these by $\mathbf{E}_{sc}^+(\boldsymbol{\kappa}; z)$ and $\mathbf{E}_{sc}^-(\boldsymbol{\kappa}; z)$, respectively, and they follow immediately from the general expression (8) for the scattered field [25]; we have

$$\mathbf{E}_{sc}^\pm(\boldsymbol{\kappa}; z) = e^{\pm i w_1 z} \mathbf{E}_{sc}^\pm(\boldsymbol{\kappa}), \tag{22}$$

where

$$\mathbf{E}_{sc}^\pm(\boldsymbol{\kappa}) = \mathbf{G}^\pm(\boldsymbol{\kappa}) \cdot \mathbf{P}(\boldsymbol{\kappa}) \tag{23}$$

and

$$\mathbf{G}^\pm(\boldsymbol{\kappa}) = \frac{i \tilde{\omega}^2 D}{2\epsilon_0 w_1}(\hat{s}\hat{s} + \hat{p}_{1\pm}\hat{p}_{1\pm}), \tag{24}$$

and we have again assumed $|w_1|D \ll 1$. Now the incident field satisfies the Maxwell equations with a uniform relative dielectric constant ε_1, and so everywhere in space it is of the form

$$\mathbf{E}_{\mathrm{in}}(\boldsymbol{\kappa}; z) = \mathbf{E}_{\mathrm{in}}^+(\boldsymbol{\kappa}; z) + \mathbf{E}_{\mathrm{in}}^-(\boldsymbol{\kappa}; z), \tag{25}$$

where

$$\mathbf{E}_{\mathrm{in}}^\pm(\boldsymbol{\kappa}; z) = e^{\pm i w_1 z} \mathbf{E}_{\mathrm{in}}^\pm(\boldsymbol{\kappa}) \tag{26}$$

[cf. (22)]. Then given any $\boldsymbol{\kappa}$, for $z > D/2$ we label the full upward-propagating (or evanescent) fields as $\mathbf{E}_{\mathrm{out}}^+(\boldsymbol{\kappa}) \exp(i w_1 z)$, while for $z < -D/2$ we label the full downward-propagating (or evanescent) fields as $\mathbf{E}_{\mathrm{out}}^- \exp(-i w_1 z)$; we clearly have

$$\mathbf{E}_{\mathrm{out}}^\pm(\boldsymbol{\kappa}) = \mathbf{E}_{\mathrm{in}}^\pm(\boldsymbol{\kappa}) + \mathbf{E}_{sc}^\pm(\boldsymbol{\kappa}). \tag{27}$$

Before solving for these fields, we identify how the Rayleigh and Wood anomalies are captured in their calculation.

III. RAYLEIGH AND WOOD ANOMALIES

Returning to the expression (24) for $\mathbf{G}^\pm(\boldsymbol{\kappa})$, and writing $\hat{p}_{1\pm}$ in terms of $\hat{\kappa}$ and $\hat{z}$, we see that in this basis of real unit vectors there are terms in $\mathbf{G}^\pm(\boldsymbol{\kappa})$ proportional to w_1, and terms proportional to $1/w_1$ These are both nonanalytic in κ, since w_1 is purely real for $\kappa < \tilde{\omega} n_1$, purely imaginary for $\kappa > \tilde{\omega} n_1$, and vanishes at $\kappa = \tilde{\omega} n_1$; $1/w_1$ thus diverges at $\kappa = \tilde{\omega} n_1$. The transition from real to imaginary w_1 can arise as the angle of incidence is varied, and κ is associated with a diffracted order that becomes evanescent in the background dielectric as κ first approaches and then exceeds $\tilde{\omega} n_1$. Of course, although the $\mathbf{G}^\pm(\boldsymbol{\kappa})$ diverge as $w_1 \to 0$, the $\mathbf{E}_{sc}^\pm(\boldsymbol{\kappa})$ do not; the same nonanalyticity as $w_1 \to 0$ appears in $\mathbf{g}(\boldsymbol{\kappa})$, since

$$\mathbf{g}(\boldsymbol{\kappa}) = \tfrac{1}{2}\big[\mathbf{G}^+(\boldsymbol{\kappa}) + \mathbf{G}^-(\boldsymbol{\kappa})\big],$$

and once the expression for $\mathbf{P}_v(\boldsymbol{\kappa})$ is included the self-consistent solution of the set of equations (18) leads to finite fields everywhere at all κ, as we show in detail below. This is enforced by the coupling among the different diffracted and evanescent orders, and by the coupling between each of them to the specularly reflected and transmitted fields; the source of these couplings is of course the grating that is itself responsible for the existence of the diffracted and evanescent orders themselves. Another consequence of these couplings is that the nonanalyticity associated with the passing of a diffracted order into evanescence appears as well in the expressions for the amplitudes of the other diffracted orders, and in those of the specularly reflected and transmitted fields. These are the Rayleigh anomalies.

Another nonanalyticity implicit in these equations can be revealed by inserting the second of (18) into the first and formally solving for $\mathbf{P}(\boldsymbol{\kappa})$,

$$\mathbf{P}(\boldsymbol{\kappa}) = [\mathbf{I} - \epsilon_0 \langle \chi_{\mathrm{mod}} \rangle \cdot \mathbf{g}(\boldsymbol{\kappa})]^{-1}$$
$$\times [\epsilon_0 \langle \chi_{\mathrm{mod}} \rangle \cdot \mathbf{E}_{\mathrm{in}}(\boldsymbol{\kappa}) + \mathbf{P}_v(\boldsymbol{\kappa})], \tag{28}$$

where $\mathbf{I} = \hat{x}\hat{x} + \hat{y}\hat{y} + \hat{z}\hat{z}$ is the unit dyadic. The expression (28), in which the grating contributions above the averaged optical response are contained in $\mathbf{P}_v(\boldsymbol{\kappa})$, is valid as long as $[\mathbf{I} - \epsilon_0 \langle \boldsymbol{\chi}_{\text{mod}} \rangle \cdot \mathbf{g}(\boldsymbol{\kappa})]^{-1}$ has no divergent components, and this holds as long as the determinant of a matrix representing $[\mathbf{I} - \epsilon_0 \langle \boldsymbol{\chi}_{\text{mod}} \rangle \cdot \mathbf{g}(\boldsymbol{\kappa})]$ does not vanish. In the special case where $\varepsilon_{\text{layer}}^{xx} = \varepsilon_{\text{layer}}^{yy} \equiv \varepsilon_{\text{layer}}^{\parallel}$, which holds if the dielectric constants of the materials satisfy $\varepsilon^{xx} = \varepsilon^{yy}$ [recall (20) and (21)], the matrix $\boldsymbol{\varepsilon}_{\text{layer}}$ can be easily written out in the $(\hat{s}, \hat{\kappa}, \hat{z})$ basis, since $\hat{x}\hat{x} + \hat{y}\hat{y} = \hat{s}\hat{s} + \hat{\kappa}\hat{\kappa}$; we find

$$
\det[\mathbf{I} - \epsilon_0 \langle \boldsymbol{\chi}_{\text{mod}} \rangle \cdot \mathbf{g}(\boldsymbol{\kappa})]
$$

$$
= \left[1 - \frac{i\tilde{\omega}^2 D}{2w_1} \left(\varepsilon_{\text{layer}}^{\parallel} - \varepsilon_1 \right) \right] \left[1 - \frac{iw_1 D}{2\varepsilon_1} \left(\varepsilon_{\text{layer}}^{\parallel} - \varepsilon_1 \right) \right]
$$

$$
\times \left[1 - \frac{i\kappa^2 D}{2w_1} \frac{\varepsilon_{\text{layer}}^{\perp} - \varepsilon_1}{\varepsilon_{\text{layer}}^{\perp}} \right], \tag{29}
$$

where we have put $\varepsilon_{\text{layer}}^{\perp} \equiv \varepsilon_{\text{layer}}^{zz}$. For reasonable $\varepsilon_{\text{layer}}^{\parallel}$ the middle bracketed term in (29) cannot vanish, since by assumption $|w_1| D \ll 1$; thus the determinant in (29) can vanish only if one of the following conditions is met:

$$
1 - \frac{i\tilde{\omega}^2 D}{2w_1} \left(\varepsilon_{\text{layer}}^{\parallel} - \varepsilon_1 \right) = 0,
$$

$$
1 - \frac{i\kappa^2 D}{2w_1} \frac{\varepsilon_{\text{layer}}^{\perp} - \varepsilon_1}{\varepsilon_{\text{layer}}^{\perp}} = 0. \tag{30}
$$

In Appendix B we show that the first of (30) is the dispersion relation for the fundamental s-polarized mode, and the second for the fundamental p-polarized mode, of a thin enough planar uniaxial waveguide with relative dielectric tensor $(\hat{x}\hat{x} + \hat{y}\hat{y})\varepsilon_{\text{layer}}^{\parallel} + \hat{z}\hat{z}\varepsilon_{\text{layer}}^{\perp}$, bounded above and below by a uniform isotropic dielectric with dielectric constant ε_1; recall that in the limit of a thin enough planar waveguide at most one waveguide mode of each polarization exists. Around the values of κ where they vanish, the left-hand sides of (30) can be written as proportional to $(\kappa - \kappa_S)$ and $(\kappa - \kappa_P)$, respectively, where at frequency ω the s- and p-polarized waveguide modes have wave numbers κ_S and κ_P, respectively; if there is no absorption, κ_S and κ_P are real. Thus the nonanalyticities of $[\mathbf{I} - \epsilon_0 \langle \boldsymbol{\chi}_{\text{mod}} \rangle \cdot \mathbf{g}(\boldsymbol{\kappa})]^{-1}$ are poles, on the real κ axis if there is no absorption, associated with the waveguide modes of the "effective waveguide" established by the average optical response in the grating region.

Despite these divergences, the solution of (28) for $\mathbf{P}(\boldsymbol{\kappa})$ is again always finite. The waveguide modes exist for $\kappa > \tilde{\omega} n_1$, "beyond the light line," and no physical field incident from infinity can be described by nonzero $\mathbf{E}_{\text{in}}(\boldsymbol{\kappa})$ for κ in the range of the divergences. Of course, by coupling through the grating, $\mathbf{P}_v(\boldsymbol{\kappa})$ can acquire $\boldsymbol{\kappa}$ components for κ at wave numbers near or at the waveguide modes if the angle of incidence of the incident field is properly chosen, as we see in detail below. However, a grating that allows $\mathbf{P}_v(\boldsymbol{\kappa})$ to acquire those $\boldsymbol{\kappa}$ components from the incident field will also couple part of any field that $\mathbf{P}_v(\boldsymbol{\kappa})$ generates back to the wave vector of the incident field, thus modifying the effective incident field driving $\mathbf{P}_v(\boldsymbol{\kappa})$ and ameliorating the response; the effective waveguide pole is moved off the real κ axis, as we illustrate in an example later.

Another consequence is that the resonant structure associated with one of the evanescent orders being close to an effective waveguide mode will lead, through coupling by the grating, to resonant structures in other diffracted and evanescent orders, and in the specularly reflected and transmitted fields. These are the Wood anomalies.

Thus within the approximation of a thin-grating region even a schematic discussion as presented above can identify Rayleigh and Wood anomalies with nonanalyticities in the response of the grating structure to an incident field: Rayleigh anomalies are associated with square-root divergences as a diffracted order becomes evanescent, and Wood anomalies are associated with pole divergences as an evanescent order approaches an effective waveguide mode of the grating region. Full calculations within this approximation presented below will confirm this connection, and show that our equations, while approximate, exhibit exact energy conservation. As well, since for thin-grating regions the dispersion relations of the effective waveguide modes lie close to the light line, we can expect a complicated response because the resonances associated with the anomalies, considered independently, lie close to each other. This is considered in some examples presented in Sec. IV.

IV. COUPLED WAVE VECTOR EQUATIONS

We now turn to the solution for the fields in the presence of a grating $\chi(\zeta)$ of the form (5), where since $\chi_v(\zeta)$ is taken as periodic with period a, we can expand it in a Fourier series

$$
\boldsymbol{\chi}_v(\zeta) = \sum_m \boldsymbol{\chi}_{v[m]} e^{im\mathbf{K}\cdot\mathbf{R}},
$$

where m ranges over the integers and $\mathbf{K}$ is given by (4); here

$$
\boldsymbol{\chi}_{v[m]} = \frac{1}{a} \int_{-a/2}^{a/2} e^{-imK\zeta} \boldsymbol{\chi}_v(\zeta) d\zeta,
$$

with $K = |\mathbf{K}|$. Note that by virtue of the definition (17) of $\chi_v(\zeta)$ we have $\chi_{v[0]} = 0$. Since $\mathbf{P}_v(\mathbf{R})$ is the response (19) to $\mathbf{E}_{\text{mod}}(\mathbf{R})$ due to $\chi_v(\zeta)$, we seek a solution for our fields of the form

$$
\mathbf{F}(\boldsymbol{\kappa}) = (2\pi)^2 \sum_m \delta(\boldsymbol{\kappa} - \boldsymbol{\kappa}_{\text{in}} - m\mathbf{K}) \mathcal{F}(\boldsymbol{\kappa}_{\text{in}} + m\mathbf{K}),
$$

where

$$
\mathbf{F}(\mathbf{R}) = \sum_m \mathcal{F}(\boldsymbol{\kappa}_m) e^{i(\boldsymbol{\kappa}_{\text{in}} + m\mathbf{K})\cdot\mathbf{R}}, \tag{31}
$$

and here and henceforth we put

$$
\boldsymbol{\kappa}_m \equiv \boldsymbol{\kappa}_{\text{in}} + m\mathbf{K}.
$$

Here $\boldsymbol{\kappa}_{\text{in}}$ characterizes the incident field, but we actually allow the incident field $\mathbf{E}_{\text{in}}(\mathbf{R})$ to be of the general form (31), with $\mathcal{E}_{\text{in}}(\boldsymbol{\kappa}_m)$ nonzero for $m \neq 0$; in later sections we will consider a grating above a substrate, and terms with $m \neq 0$ will arise from reflection of scattered light off the substrate. Using the expansion (31) in (19) we have

$$
\mathcal{P}_v(\boldsymbol{\kappa}_m) = \epsilon_0 \sum_{m'} \boldsymbol{\chi}_{v[m-m']} \cdot \mathcal{E}_{\text{mod}}(\boldsymbol{\kappa}_{m'}), \tag{32}
$$

for example; equations for the Fourier components of other quantities will be given below. The set of these equations can be organized as matrix equations in many ways; below we present one approach that is both useful for calculations, and allows for an easy proof of energy conservation even when the number of Fourier components is truncated.

A. *S*-matrix equations

To complete a calculation we approximate sums over m by a restriction to $|m| \leqslant N$, where the threshold integer N includes at least all diffracted, propagating orders. For each field $\mathbf{F}(\mathbf{R})$ we then introduce $\bar{\mathcal{F}}$, a column of columns

$$\bar{\mathcal{F}} = \begin{bmatrix} \bar{\mathcal{F}}(\boldsymbol{\kappa}_N) \\ \bar{\mathcal{F}}(\boldsymbol{\kappa}_{(N-1)}) \\ \vdots \\ \bar{\mathcal{F}}(\boldsymbol{\kappa}_{(-N)}) \end{bmatrix}, \tag{33}$$

where each $\bar{\mathcal{F}}(\boldsymbol{\kappa}_m)$ is a column with the three Cartesian components of $\mathcal{F}(\boldsymbol{\kappa}_m)$,

$$\bar{\mathcal{F}}(\boldsymbol{\kappa}_m) = \begin{bmatrix} \hat{\boldsymbol{x}} \cdot \mathcal{F}(\boldsymbol{\kappa}_m) \\ \hat{\boldsymbol{y}} \cdot \mathcal{F}(\boldsymbol{\kappa}_m) \\ \hat{\boldsymbol{z}} \cdot \mathcal{F}(\boldsymbol{\kappa}_m) \end{bmatrix}, \tag{34}$$

and so the full column $\bar{\mathcal{F}}$ has $3(2N+1)$ elements. For the tensors we introduce $(2N+1) \times (2N+1)$ matrices with elements that are themselves 3×3 matrices; thus in each of these there are $3(2N+1) \times 3(2N+1)$ elements in all. We put

$$\bar{g} = \begin{bmatrix} \bar{g}_{NN} & \bar{0} & \cdots & \bar{0} \\ \bar{0} & \bar{g}_{(N-1)(N-1)} & \cdots & \bar{0} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{0} & \bar{0} & \cdots & \bar{g}_{(-N)(-N)} \end{bmatrix},$$

a block-diagonal matrix where $\bar{0}$ indicates a 3×3 matrix of zeros, and the 3×3 matrices $\bar{g}_{mm}$ are given by

$$\bar{g}_{mm} = \begin{bmatrix} [\hat{\boldsymbol{x}} \cdot \mathbf{g}(\boldsymbol{\kappa}_m) \cdot \hat{\boldsymbol{x}}] & [\hat{\boldsymbol{x}} \cdot \mathbf{g}(\boldsymbol{\kappa}_m) \cdot \hat{\boldsymbol{y}}] & [\hat{\boldsymbol{x}} \cdot \mathbf{g}(\boldsymbol{\kappa}_m) \cdot \hat{\boldsymbol{z}}] \\ [\hat{\boldsymbol{y}} \cdot \mathbf{g}(\boldsymbol{\kappa}_m) \cdot \hat{\boldsymbol{x}}] & [\hat{\boldsymbol{y}} \cdot \mathbf{g}(\boldsymbol{\kappa}_m) \cdot \hat{\boldsymbol{y}}] & [\hat{\boldsymbol{y}} \cdot \mathbf{g}(\boldsymbol{\kappa}_m) \cdot \hat{\boldsymbol{z}}] \\ [\hat{\boldsymbol{z}} \cdot \mathbf{g}(\boldsymbol{\kappa}_m) \cdot \hat{\boldsymbol{x}}] & [\hat{\boldsymbol{z}} \cdot \mathbf{g}(\boldsymbol{\kappa}_m) \cdot \hat{\boldsymbol{y}}] & [\hat{\boldsymbol{z}} \cdot \mathbf{g}(\boldsymbol{\kappa}_m) \cdot \hat{\boldsymbol{z}}] \end{bmatrix}.$$

For example, let the associated polarization vectors associated with $\boldsymbol{\kappa}_m$ be $\hat{\boldsymbol{s}}_m$ and $\hat{\boldsymbol{p}}_{m\pm}$, such that

$$\hat{\boldsymbol{s}}_m = \hat{\boldsymbol{\kappa}}_m \times \hat{\boldsymbol{z}}, \quad \hat{\boldsymbol{p}}_{1\pm,m} = \frac{\kappa_m \hat{\boldsymbol{z}} \mp w_1(\boldsymbol{\kappa}_m) \hat{\boldsymbol{\kappa}}_m}{\tilde{\omega} n_1}, \tag{35}$$

with $\hat{\boldsymbol{\kappa}}_m = \boldsymbol{\kappa}_m / |\boldsymbol{\kappa}_m|$ and $w_1(\boldsymbol{\kappa}_m) = \sqrt{\tilde{\omega}^2 \varepsilon_1 - \kappa_m^2}$ [compare (10)]. If we then let ϕ_m indicate the rotation in the xy plane between the sets of unit orthogonal vectors $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ and $(\hat{\boldsymbol{s}}_m, \hat{\boldsymbol{\kappa}}_m)$ (see Fig. 2), we have

$$\bar{g}_{mm}$$

$$= \frac{i \tilde{\omega}^2 D}{2 \epsilon_0 w_1(\boldsymbol{\kappa}_m)} \begin{bmatrix} 1 - \frac{\kappa_m}{\tilde{\omega}^2 \varepsilon_1} \sin^2 \phi_m & \frac{\kappa_m}{2\tilde{\omega}^2 \varepsilon_1} \sin 2\phi_m & 0 \\ \frac{\kappa_m}{2\tilde{\omega}^2 \varepsilon_1} \sin 2\phi_m & 1 - \frac{\kappa_m}{\tilde{\omega}^2 \varepsilon_1} \cos^2 \phi_m & 0 \\ 0 & 0 & \frac{\kappa_m}{\tilde{\omega}^2 \varepsilon_1} \end{bmatrix}.$$
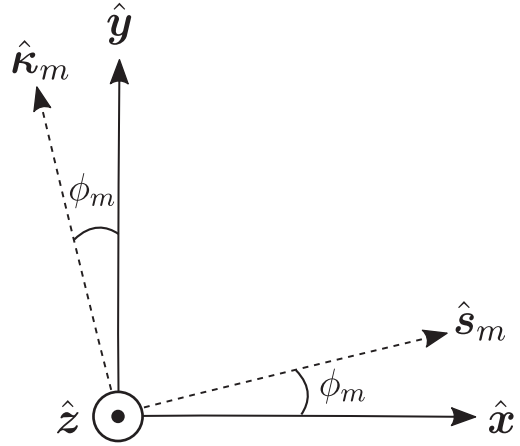
FIG. 2. Plane containing the set of axes $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ and basis vectors $(\hat{\boldsymbol{s}}_m, \hat{\boldsymbol{\kappa}}_m)$.

Block-diagonal matrices $\bar{G}^{\pm}$ and $\bar{\chi}_o$ are defined similarly, where the blocks $\bar{\chi}_{o;mm}$ of $\bar{\chi}_o$ are all identical, $\bar{\chi}_{o;mm} = \text{diag}(\langle \chi_{\text{mod}}^{xx} \rangle, \langle \chi_{\text{mod}}^{yy} \rangle, \langle \chi_{\text{mod}}^{zz} \rangle)$. The matrix of matrices $\bar{\chi}_v$ representing the grating is not block diagonal, but is given by

$$\bar{\chi}_v = \begin{bmatrix} \bar{0} & \bar{\chi}_{v;N(N-1)} & \cdots & \bar{\chi}_{v;N(-N)} \\ \bar{\chi}_{v;(N-1)(N)} & \bar{0} & \cdots & \bar{\chi}_{v;(N-1)(-N)} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{\chi}_{v;(-N)(N)} & \bar{\chi}_{v;(-N)(N-1)} & \cdots & \bar{0} \end{bmatrix}, \tag{36}$$

where

$$\bar{\chi}_{v;mm'}$$
$$= \begin{bmatrix} \hat{\boldsymbol{x}} \cdot \boldsymbol{\chi}_{v[m-m']} \cdot \hat{\boldsymbol{x}} & 0 & 0 \\ 0 & \hat{\boldsymbol{y}} \cdot \boldsymbol{\chi}_{v[m-m']} \cdot \hat{\boldsymbol{y}} & 0 \\ 0 & 0 & \hat{\boldsymbol{z}} \cdot \boldsymbol{\chi}_{v[m-m']} \cdot \hat{\boldsymbol{z}} \end{bmatrix},$$

and the diagonal elements of $\bar{\chi}_v$ vanish because $\boldsymbol{\chi}_{v[0]} = 0$. In this notation Eq. (32) for the $\mathcal{P}_v(\boldsymbol{\kappa}_m)$ can be written in full matrix form as

$$\bar{\mathcal{P}}_v = \epsilon_0 \bar{\chi}_v \bar{\mathcal{E}}_{\text{mod}},$$

and combining this with the matrix form of (18) we find

$$\bar{\mathcal{P}} = \epsilon_0 (\bar{\chi}_o + \bar{\chi}_v) \bar{\mathcal{E}}_{\text{mod}},$$

$$\bar{\mathcal{E}}_{\text{mod}} = \bar{\mathcal{E}}_{\text{in}} + \bar{g} \bar{\mathcal{P}},$$

with a formal solution

$$\bar{\mathcal{P}} = \epsilon_0 (\bar{\chi}_o + \bar{\chi}_v) [\bar{1}_3 - \epsilon_0 \bar{g} (\bar{\chi}_o + \bar{\chi}_v)]^{-1} \bar{\mathcal{E}}_{\text{in}}, \tag{37}$$

where $\bar{1}_j$ denotes the $j(2N+1) \times j(2N+1)$ unit matrix, with j an integer. Introducing columns $\bar{\mathcal{E}}_{sc}^{\pm}$ to describe the scattered fields, from (23) we then have

$$\bar{\mathcal{E}}_{sc}^{\pm} = \epsilon_0 \bar{G}^{\pm} (\bar{\chi}_o + \bar{\chi}_v) [\bar{1}_3 - \epsilon_0 \bar{g} (\bar{\chi}_o + \bar{\chi}_v)]^{-1} \bar{\mathcal{E}}_{\text{in}}.$$

Separating out the upward-propagating (or evanescent) contributions of the incident field from the corresponding downward propagations [see (26)], we have $\bar{\mathcal{E}}_{\text{in}} = \bar{\mathcal{E}}_{\text{in}}^{+} + \bar{\mathcal{E}}_{\text{in}}^{-}$, and introducing columns for the full outward-propagating (or

evanescent) fields for $z > D/2$ and $z < -D/2$ [see (27)], with new columns defined as indicated we can write

$$\bar{\mathcal{E}}_{\text{out}}^+ = \epsilon_0 \bar{G}^+ (\bar{\chi}_o + \bar{\chi}_v)[\bar{1}_3 - \epsilon_0 \bar{g}(\bar{\chi}_o + \bar{\chi}_v)]^{-1}$$
$$\times (\bar{\mathcal{E}}_{\text{in}}^+ + \bar{\mathcal{E}}_{\text{in}}^-) + \bar{\mathcal{E}}_{\text{in}}^+,$$
$$\bar{\mathcal{E}}_{\text{out}}^- = \epsilon_0 \bar{G}^- (\bar{\chi}_o + \bar{\chi}_v)[\bar{1}_3 - \epsilon_0 \bar{g}(\bar{\chi}_o + \bar{\chi}_v)]^{-1}$$
$$\times (\bar{\mathcal{E}}_{\text{in}}^+ + \bar{\mathcal{E}}_{\text{in}}^-) + \bar{\mathcal{E}}_{\text{in}}^-. \quad (38)$$

The subcolumns of $\bar{\mathcal{E}}_{\text{in}}^\pm$, $\bar{\mathcal{E}}_{\text{in}}^\pm(\boldsymbol{\kappa}_m)$, contain the three Cartesian components of $\mathcal{E}_{\text{in}}^\pm(\boldsymbol{\kappa}_m)$ [recall (34)]. However, these are not independent, since for any $\boldsymbol{\kappa}_m$ there are only s- and p-polarized components,

$$\mathcal{E}_{\text{in}}^\pm(\boldsymbol{\kappa}_m) = \hat{\boldsymbol{s}}_m \mathbb{E}_{\text{in};s}^\pm(\boldsymbol{\kappa}_m) + \hat{\boldsymbol{p}}_{1\pm,m} \mathbb{E}_{\text{in};p}^\pm(\boldsymbol{\kappa}_m),$$

with two independent amplitudes $\mathbb{E}_{\text{in};s}^\pm(\boldsymbol{\kappa}_m)$ and $\mathbb{E}_{\text{in};p}^\pm(\boldsymbol{\kappa}_m)$. As such we can write

$$\bar{\mathcal{E}}_{\text{in}}^\pm(\boldsymbol{\kappa}_m) = \begin{bmatrix} \hat{\boldsymbol{x}} \cdot \mathcal{E}_{\text{in}}^\pm(\boldsymbol{\kappa}_m) \\ \hat{\boldsymbol{y}} \cdot \mathcal{E}_{\text{in}}^\pm(\boldsymbol{\kappa}_m) \\ \hat{\boldsymbol{z}} \cdot \mathcal{E}_{\text{in}}^\pm(\boldsymbol{\kappa}_m) \end{bmatrix}$$
$$= \bar{\sigma}_{\text{in}}^\pm(\boldsymbol{\kappa}_m) \bar{\mathbb{E}}_{\text{in}}^\pm(\boldsymbol{\kappa}_m),$$

where

$$\bar{\sigma}_{\text{in}}^\pm(\boldsymbol{\kappa}_m) = \begin{bmatrix} (\hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{s}}_m) & (\hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{p}}_{1\pm,m}) \\ (\hat{\boldsymbol{y}} \cdot \hat{\boldsymbol{s}}_m) & (\hat{\boldsymbol{y}} \cdot \hat{\boldsymbol{p}}_{1\pm,m}) \\ (\hat{\boldsymbol{z}} \cdot \hat{\boldsymbol{s}}_m) & (\hat{\boldsymbol{z}} \cdot \hat{\boldsymbol{p}}_{1\pm,m}) \end{bmatrix} \quad (39)$$

is a 3×2 matrix for each of the $(+)$ and $(-)$ examples, and

$$\bar{\mathbb{E}}_{\text{in}}^\pm(\boldsymbol{\kappa}_m) = \begin{bmatrix} \mathbb{E}_{\text{in};s}^\pm(\boldsymbol{\kappa}_m) \\ \mathbb{E}_{\text{in};p}^\pm(\boldsymbol{\kappa}_m) \end{bmatrix} \quad (40)$$

is a column of two elements. Constructing the full column for all $\boldsymbol{\kappa}_m$ components of $\bar{\mathcal{E}}_{\text{in}}^\pm$ we have [recall (33)]

$$\bar{\mathcal{E}}_{\text{in}}^\pm = \bar{\sigma}_{\text{in}}^\pm \bar{\mathbb{E}}_{\text{in}}^\pm, \quad (41)$$

where

$$\bar{\mathbb{E}}_{\text{in}}^\pm = \begin{bmatrix} \bar{\mathbb{E}}_{\text{in}}^\pm(\boldsymbol{\kappa}_N) \\ \bar{\mathbb{E}}_{\text{in}}^\pm(\boldsymbol{\kappa}_{(N-1)}) \\ \vdots \\ \bar{\mathbb{E}}_{\text{in}}^\pm(\boldsymbol{\kappa}_{(-N)}) \end{bmatrix}, \quad (42)$$

which for each of the $(+)$ and $(-)$ examples is a column with $2(2N+1)$ elements, once all the $\bar{\mathbb{E}}_{\text{in}}^\pm(\boldsymbol{\kappa}_m)$ are written out. Further,

$$\bar{\sigma}_{\text{in}}^\pm = \begin{bmatrix} \bar{\sigma}_{\text{in}}^\pm(\boldsymbol{\kappa}_N) & \bar{0} & \cdots & \bar{0} \\ \bar{0} & \bar{\sigma}_{\text{in}}^\pm(\boldsymbol{\kappa}_{(N-1)}) & \cdots & \bar{0} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{0} & \bar{0} & \cdots & \bar{\sigma}_{\text{in}}^\pm(\boldsymbol{\kappa}_{(-N)}) \end{bmatrix}, \quad (43)$$

which for each of the examples is a $3(2N+1) \times 2(2N+1)$ matrix, once all the elements of the $\bar{\sigma}_{\text{in}}^\pm(\boldsymbol{\kappa}_m)$ are written out; here $\bar{0}$ are 3×2 matrices with all elements vanishing.

Similarly, for each $\bar{\mathcal{E}}_{\text{out}}^\pm(\boldsymbol{\kappa}_m)$ in $\bar{\mathcal{E}}_{\text{out}}^\pm$ there will be only s- and p-polarized components,

$$\mathcal{E}_{\text{out}}^\pm(\boldsymbol{\kappa}_m) = \hat{\boldsymbol{s}}_m \mathbb{E}_{\text{out};s}^\pm(\boldsymbol{\kappa}_m) + \hat{\boldsymbol{p}}_{1\pm,m} \mathbb{E}_{\text{out};p}^\pm(\boldsymbol{\kappa}_m),$$

which we can immediately see will be identified by the $\mathbf{G}^\pm(\boldsymbol{\kappa}_m)$ [see (24)] that appear in $\bar{G}^\pm$. Nonetheless, we can formally extract those amplitudes $\mathbb{E}_{\text{out};s,p}^\pm(\boldsymbol{\kappa}_m)$ by writing

$$\bar{\mathbb{E}}_{\text{out}}^\pm = \bar{\sigma}_{\text{out}}^\pm \bar{\mathcal{E}}_{\text{out}}^\pm, \quad (44)$$

where

$$\bar{\mathbb{E}}_{\text{out}}^\pm = \begin{bmatrix} \bar{\mathbb{E}}_{\text{out}}^\pm(\boldsymbol{\kappa}_N) \\ \bar{\mathbb{E}}_{\text{out}}^\pm(\boldsymbol{\kappa}_{(N-1)}) \\ \vdots \\ \bar{\mathbb{E}}_{\text{out}}^\pm(\boldsymbol{\kappa}_{(-N)}) \end{bmatrix}$$

is a $2(2N+1)$ column for each example $(\pm)$ once the elements of

$$\bar{\mathbb{E}}_{\text{out}}^\pm(\boldsymbol{\kappa}_m) = \begin{bmatrix} \mathbb{E}_{\text{out};s}^\pm(\boldsymbol{\kappa}_m) \\ \mathbb{E}_{\text{out};p}^\pm(\boldsymbol{\kappa}_m) \end{bmatrix}$$

are written out, and

$$\bar{\sigma}_{\text{out}}^\pm = \begin{bmatrix} \bar{\sigma}_{\text{out}}^\pm(\boldsymbol{\kappa}_N) & \bar{0} & \cdots & \bar{0} \\ \bar{0} & \bar{\sigma}_{\text{out}}^\pm(\boldsymbol{\kappa}_{(N-1)}) & \cdots & \bar{0} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{0} & \bar{0} & \cdots & \bar{\sigma}_{\text{out}}^\pm(\boldsymbol{\kappa}_{(-N)}) \end{bmatrix}, \quad (45)$$

where the $\bar{0}$ denote 2×3 matrices with all their elements vanishing, and

$$\bar{\sigma}_{\text{out}}^\pm(\boldsymbol{\kappa}_m) = \begin{bmatrix} (\hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{s}}_m) & (\hat{\boldsymbol{y}} \cdot \hat{\boldsymbol{s}}_m) & (\hat{\boldsymbol{z}} \cdot \hat{\boldsymbol{s}}_m) \\ (\hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{p}}_{1\pm,m}) & (\hat{\boldsymbol{y}} \cdot \hat{\boldsymbol{p}}_{1\pm,m}) & (\hat{\boldsymbol{z}} \cdot \hat{\boldsymbol{p}}_{1\pm,m}) \end{bmatrix}. \quad (46)$$

Using (41) and (44) in (38), we can write

$$\bar{\mathbb{E}}_{\text{out}}^+ = \bar{\mathbb{T}}_g \bar{\mathbb{E}}_{\text{in}}^+ + \bar{\mathbb{R}}_g \bar{\mathbb{E}}_{\text{in}}^-,$$
$$\bar{\mathbb{E}}_{\text{out}}^- = \bar{\mathbb{R}}_g \bar{\mathbb{E}}_{\text{in}}^+ + \bar{\mathbb{T}}_g \bar{\mathbb{E}}_{\text{in}}^-,$$

where $\bar{\mathbb{R}}_g$ and $\bar{\mathbb{T}}_g$ are $2(2N+1) \times 2(2N+1)$ matrices,

$$\bar{\mathbb{T}}_g = \bar{\sigma}_{\text{out}}^\pm \{\bar{1}_3 + \epsilon_0 \bar{G}^\pm(\bar{\chi}_o + \bar{\chi}_v)[\bar{1}_3 - \epsilon_0 \bar{g}(\bar{\chi}_o + \bar{\chi}_v)]^{-1}\} \bar{\sigma}_{\text{in}}^\pm,$$
$$\bar{\mathbb{R}}_g = \bar{\sigma}_{\text{out}}^\pm \epsilon_0 \bar{G}^\pm(\bar{\chi}_o + \bar{\chi}_v)[\bar{1}_3 - \epsilon_0 \bar{g}(\bar{\chi}_o + \bar{\chi}_v)]^{-1} \bar{\sigma}_{\text{in}}^\mp. \quad (47)$$

Since we consider the same dielectric above and below the grating, the transmission and reflection properties are the same whether light is incident from above or below; thus the expressions (47) are the same whether the $+$ or $-$ matrices on the right-hand side of the equations are used in their evaluation.

Finally, combining the two columns $\bar{\mathbb{E}}_{\text{out}}^+$ and $\bar{\mathbb{E}}_{\text{out}}^-$, each with $2(2N+1)$ elements, to form one column with $4(2N+1)$ elements, and likewise for $\bar{\mathbb{E}}_{\text{in}}^+$ and $\bar{\mathbb{E}}_{\text{in}}^-$, we can form a $4(2N+1) \times 4(2N+1)$ scattering matrix $\mathscr{S}$,

$$\mathscr{S} \equiv \begin{bmatrix} \bar{\mathbb{T}}_g & \bar{\mathbb{R}}_g \\ \bar{\mathbb{R}}_g & \bar{\mathbb{T}}_g \end{bmatrix}, \quad (48)$$
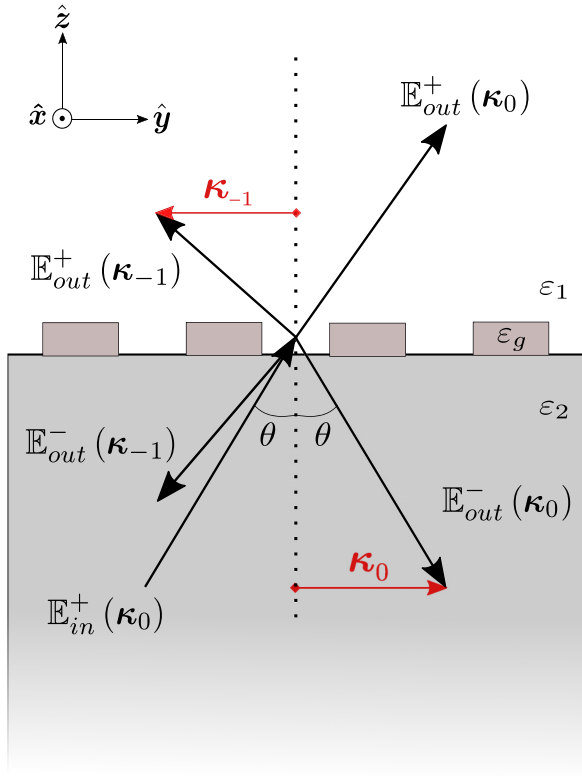
FIG. 3. A simple 1D grating configuration with the grating oriented such that $\hat{e} = \hat{y}$. Incident, reflected, transmitted, and diffracted rays are shown by black (thick) lines, and labeled by the notation used to indicate their field amplitudes; the projection of the wave vectors on the xy plane is shown by red (thin) lines.

so that

$$\begin{bmatrix} \bar{\mathbb{E}}^+_{\text{out}} \\ \bar{\mathbb{E}}^-_{\text{out}} \end{bmatrix} = \mathscr{S} \begin{bmatrix} \bar{\mathbb{E}}^+_{\text{in}} \\ \bar{\mathbb{E}}^-_{\text{in}} \end{bmatrix}. \tag{49}$$

With the equations in this form, a proof of energy conservation is possible, and is presented in Appendix C. That proof, and the equations from which it was derived, hold for any orientation of the grating direction $\hat{e}$ in the xy plane and any plane of incidence.

B. A simple configuration

In this subsection we simplify the equations above for a common scenario of interest: We take the grating susceptibility to be uniaxial, $\chi^{xx}_{\text{add}}(\zeta) = \chi^{yy}_{\text{add}}(\zeta)$ [recall (5) and (6)], choose $\hat{e} = \hat{y}$, and assume the plane of incidence contains $\hat{z}$ and $\hat{e} = \hat{y}$, as illustrated in Fig. 3; for the isolated grating treated above and in this section, we have $\varepsilon_2 = \varepsilon_1$. The wave vectors κ_m that are relevant here are then either in the $\hat{y}$ or $-\hat{y}$ direction, so $\hat{k}_m \cdot \hat{y} = \text{sgn}(\hat{k}_m \cdot \hat{y}) = \hat{s}_m \cdot \hat{x}$; the form of the expressions (39) for the $\bar{\sigma}^{\pm}_{\text{in}}(\kappa_m)$ simplifies to

$$\bar{\sigma}^{\pm}_{\text{in}}(\kappa_m) = \begin{bmatrix} (\hat{k}_m \cdot \hat{y}) & 0 \\ 0 & \mp \frac{w_1(\kappa_m)}{\tilde{\omega} n_1}(\hat{k}_m \cdot \hat{y}) \\ 0 & \frac{\kappa_m}{\tilde{\omega} n_1} \end{bmatrix},$$

and similarly for the form of the expressions (46) for $\bar{\sigma}^{\pm}_{\text{out}}(\kappa_m)$, which are the transpose of the $\bar{\sigma}^{\pm}_{\text{in}}(\kappa_m)$. When these are assembled into $\bar{\sigma}^{\pm}_{\text{in}}$ and $\bar{\sigma}^{\pm}_{\text{out}}$ in (43) and (45) and the results used in (47) for $\bar{\mathbb{R}}_g$ and $\bar{\mathbb{T}}_g$, we find that because of the high symmetry of the problem each of these $2(2N+1) \times 2(2N+1)$ matrices can be reorganized into two $(2N+1) \times (2N+1)$ matrices, one relevant for s-polarized light and one for p-polarized light. For each polarization the relevant matrices can then be combined into a $2(2N+1) \times 2(2N+1)$ scattering matrix, and in place of (49) we have two sets of equations,

$$\begin{bmatrix} \bar{\mathbb{E}}^+_{\text{out},\alpha} \\ \bar{\mathbb{E}}^-_{\text{out},\alpha} \end{bmatrix} = \mathscr{S}_\alpha \begin{bmatrix} \bar{\mathbb{E}}^+_{\text{in},\alpha} \\ \bar{\mathbb{E}}^-_{\text{in},\alpha} \end{bmatrix}, \tag{50}$$

where $\alpha = s, p$, each $\bar{\mathbb{E}}^+_{\text{in},\alpha}$ is a $(2N+1)$ element column,

$$\bar{\mathbb{E}}^-_{\text{in},\alpha} = \begin{bmatrix} \mathbb{E}^-_{\text{in},\alpha}(\kappa_N) \\ \mathbb{E}^-_{\text{in},\alpha}(\kappa_{N-1}) \\ \vdots \\ \mathbb{E}^-_{\text{in},\alpha}(\kappa_{-N}) \end{bmatrix}, \tag{51}$$

[compare (40), (42)], and likewise for $\bar{\mathbb{E}}^+_{\text{in},\alpha}$ and $\bar{\mathbb{E}}^{\pm}_{\text{out},\alpha}$, and where

$$\mathscr{S}_\alpha \equiv \begin{bmatrix} \mathbb{T}_{g,\alpha} & \mathbb{R}_{g,\alpha} \\ \mathbb{R}_{g,\alpha} & \mathbb{T}_{g,\alpha} \end{bmatrix},$$

with

$$\mathbb{T}_{g,s} = \bar{\beta} \left[\bar{1}_1 - \frac{i\tilde{\omega}^2 D}{2} \bar{w}_1^{-1} \bar{\chi}^{\parallel} \right]^{-1} \bar{\beta},$$

$$\mathbb{R}_{g,s} = \mathbb{T}_{g,s} - \bar{1}_1, \tag{52}$$

and

$$\mathbb{T}_{g,p} = \left[\bar{1}_1 - \frac{iD}{2\varepsilon_1} \bar{\kappa} \bar{\chi}^{\perp} \bar{\kappa} \right]^{-1} + \bar{\beta} \left[\bar{1}_1 - \frac{iD}{2\varepsilon_1} \bar{\chi}^{\parallel} \bar{w}_1 \right]^{-1} \bar{\beta} - \bar{1}_1,$$

$$\mathbb{R}_{g,p} = \left[\bar{1}_1 - \frac{iD}{2\varepsilon_1} \bar{\kappa} \bar{\chi}^{\perp} \bar{\kappa} \right]^{-1} - \bar{\beta} \left[\bar{1}_1 - \frac{iD}{2\varepsilon_1} \bar{\chi}^{\parallel} \bar{w}_1 \right]^{-1} \bar{\beta}. \tag{53}$$

Here each of $\mathbb{T}_{g,s}$, $\mathbb{R}_{g,s}$, $\mathbb{T}_{g,p}$, and $\mathbb{R}_{g,p}$ is a $(2N+1) \times (2N+1)$ matrix. The matrices $\bar{\beta}$, $\bar{\kappa}$, and $\bar{w}_1$ are diagonal matrices of the same dimension, $\bar{\beta} = \text{diag}(\hat{k}_N \cdot \hat{y}, \hat{k}_{N-1} \cdot \hat{y}, \ldots, \hat{k}_{-N} \cdot \hat{y})$, $\bar{\kappa} = \text{diag}(|\kappa_N|, |\kappa_{N-1}|, \ldots, |\kappa_{-N}|)$, and $\bar{w}_1 = \text{diag}(w_1(\kappa_N), w_1(\kappa_{N-1}), \ldots, w_1(\kappa_{-N}))$, with $w_1(\kappa) = \sqrt{\tilde{\omega}^2 n_1^2 - \kappa^2}$. Finally, $\bar{\chi}^{\parallel}$ and $\bar{\chi}^{\perp}$ are $(2N+1) \times (2N+1)$ matrices with (mm') elements $\bar{\chi}^{\parallel}_{mm'} = \delta_{mm'} \langle \chi^{xx}_{\text{mod}} \rangle + (\hat{x} \cdot \chi_{v[m-m']} \cdot \hat{x}) = \delta_{mm'} \langle \chi^{yy}_{\text{mod}} \rangle + (\hat{y} \cdot \chi_{v[m-m']} \cdot \hat{y})$ and $\bar{\chi}^{\perp}_{mm'} = \delta_{mm'} \langle \chi^{zz}_{\text{mod}} \rangle + (\hat{z} \cdot \chi_{v[m-m']} \cdot \hat{z})$. We note that the relation between $\mathbb{T}_s$ and $\mathbb{R}_s$ is simple because the reference vectors $\hat{s}_m$ for the fields are all the same or differ simply by a minus sign; while that between $\mathbb{T}_p$ and $\mathbb{R}_p$ is more complicated because, even for a particular κ_m, the z components of $\hat{p}_{1+,m}$ and $\hat{p}_{1-,m}$ are identical, but the y components differ by a sign [see (35)].

C. An example

The expressions (52) and (53) for the reflection and transmission matrices, and indeed the more general expressions (47), can be used to calculate specular reflection and

transmission, and diffraction, for the choice of any number $2N + 1$ of wave vectors κ_m in the calculation. However, in certain circumstances further approximations are possible. For example, if the grating period a [see Fig. 1(b)] is small enough, then for at least some angles of incidence there will be only one propagating diffracted order ($m = -1$) in addition to the specularly reflected and transmitted fields (see Fig. 3, again with $\varepsilon_1 = \varepsilon_2$). A choice of $2N + 1 = 3$ could be adopted, but since the field associated with κ_1 is evanescent we can neglect that field and still respect energy conservation in a lossless structure if we keep only the fields at κ_0 and κ_{-1}, simply neglecting the fields at κ_1. If we do this, and consider the simple excitation scenario presented above, each of the $\mathbb{T}_{g,\alpha}$ and $\mathbb{R}_{g,\alpha}$ is a 2×2 matrix, and the resulting equations for the specularly reflected and transmitted fields, and the diffracted fields, can be solved easily. We refer to this as the "two-wave-vector model." Considering an incident field from $z = -\infty$, for s polarization we find specularly transmitted and reflected fields

$$\frac{\mathbb{E}^+_{\text{out},s}(\kappa_0)}{\mathbb{E}^+_{\text{in},s}(\kappa_0)} = U_s^{-1}\left(1 - \frac{i\tilde{\omega}^2 D}{2w_1(\kappa_{-1})}\bar{\chi}^{\parallel}_{00}\right),$$

$$\frac{\mathbb{E}^-_{\text{out},s}(\kappa_0)}{\mathbb{E}^+_{\text{in},s}(\kappa_0)} = \frac{\mathbb{E}^+_{\text{out},s}(\kappa_0)}{\mathbb{E}^+_{\text{in},s}(\kappa_0)} - 1, \tag{54}$$

and upward and downward diffracted fields that are equal in amplitude,

$$\frac{\mathbb{E}^\pm_{\text{out},s}(\kappa_{-1})}{\mathbb{E}^+_{\text{in},s}(\kappa_0)} = U_s^{-1}\left(\frac{-i\tilde{\omega}^2 D}{2w_1(\kappa_{-1})}\bar{\chi}^{\parallel}_{(-1)0}\right), \tag{55}$$

where

$$U_s = \left(1 - \frac{i\tilde{\omega}^2 D}{2w_1(\kappa_0)}\bar{\chi}^{\parallel}_{00}\right)\left(1 - \frac{i\tilde{\omega}^2 D}{2w_1(\kappa_{-1})}\bar{\chi}^{\parallel}_{00}\right)$$
$$+ \frac{\tilde{\omega}^4 D^2}{4w_1(\kappa_0)w_1(\kappa_{-1})}\bar{\chi}^{\parallel}_{0(-1)}\bar{\chi}^{\parallel}_{(-1)0},$$

while for p polarization we find specularly transmitted and reflected fields

$$\frac{\mathbb{E}^+_{\text{out},p}(\kappa_0)}{\mathbb{E}^+_{\text{in},p}(\kappa_0)} = U_z^{-1}\left(1 - \frac{i\kappa_{-1}^2 D}{2w_1(\kappa_{-1})\varepsilon_1}\bar{\chi}^{\perp}_{00}\right)$$
$$+ U_\kappa^{-1}\left(1 - \frac{iw_1(\kappa_{-1})D}{2\varepsilon_1}\bar{\chi}^{\parallel}_{00}\right) - 1,$$

$$\frac{\mathbb{E}^-_{\text{out},p}(\kappa_0)}{\mathbb{E}^+_{\text{in},p}(\kappa_0)} = U_z^{-1}\left(1 - \frac{i\kappa_{-1}^2 D}{2w_1(\kappa_{-1})\varepsilon_1}\bar{\chi}^{\perp}_{00}\right)$$
$$- U_\kappa^{-1}\left(1 - \frac{iw_1(\kappa_{-1})D}{2\varepsilon_1}\bar{\chi}^{\parallel}_{00}\right), \tag{56}$$

and upward and downward diffracted fields

$$\frac{\mathbb{E}^\pm_{\text{out},p}(\kappa_{-1})}{\mathbb{E}^+_{\text{in},p}(\kappa_0)} = U_z^{-1}\left(\frac{i\kappa_0\kappa_{-1}D}{2w_1(\kappa_{-1})\varepsilon_1}\bar{\chi}^{\perp}_{(-1)0}\right)$$
$$\pm U_\kappa^{-1}\left(\frac{iw_1(\kappa_0)D}{2\varepsilon_1}\bar{\chi}^{\parallel}_{(-1)0}\right), \tag{57}$$

where

$$U_\kappa = \left(1 - \frac{iw_1(\kappa_0)D}{2\varepsilon_1}\bar{\chi}^{\parallel}_{00}\right)\left(1 - \frac{iw_1(\kappa_{-1})D}{2\varepsilon_1}\bar{\chi}^{\parallel}_{00}\right)$$
$$+ \frac{w_1(\kappa_0)w_1(\kappa_{-1})D^2}{4\varepsilon_1^2}\bar{\chi}^{\parallel}_{0(-1)}\bar{\chi}^{\parallel}_{(-1)0},$$

$$U_z = \left(1 - \frac{i\kappa_0^2 D}{2w_1(\kappa_0)\varepsilon_1}\bar{\chi}^{\perp}_{00}\right)\left(1 - \frac{i\kappa_{-1}^2 D}{2w_1(\kappa_{-1})\varepsilon_1}\bar{\chi}^{\perp}_{00}\right)$$
$$+ \frac{\kappa_0^2\kappa_{-1}^2 D^2}{4\varepsilon_1^2 w_1(\kappa_0)w_1(\kappa_{-1})}\bar{\chi}^{\perp}_{0(-1)}\bar{\chi}^{\perp}_{(-1)0}.$$

Of course, the diffracted fields only appear for $\kappa_{-1} < \tilde{\omega}n_1$, and the expressions above are to be used only in that range. The more complicated form of the results for p-polarized light arises because of the two components ($\hat{\kappa}$ and $\hat{z}$) of the light that arise, as opposed to the simpler results for s polarization where there is only one component ($\hat{s}$).

As an example, we consider a grating with thickness $D = 25$ nm consisting of an isotropic medium with refractive index 3.5 embedded in vacuum; we take the grating period to be $a = 1.25$ μm, with a fill fraction of one-half ($d/a = 0.5$), and consider incident s-polarized light at a wavelength of 1.55 μm. In Fig. 4 we plot the relative irradiance of the radiated electric fields in this system,

$$\mathcal{I}^+_s(\kappa_m) = \frac{|\mathbb{E}^+_{\text{out},s}(\kappa_m)|^2 n_1\cos\theta_1(\kappa_m)}{|\mathbb{E}^+_{\text{in},s}(\kappa_0)|^2 n_Q\cos\theta},$$

$$\mathcal{I}^-_s(\kappa_m) = \frac{|\mathbb{E}^-_{\text{out},s}(\kappa_m)|^2\cos\theta_Q(\kappa_m)}{|\mathbb{E}^+_{\text{in},s}(\kappa_0)|^2\cos\theta}, \tag{58}$$
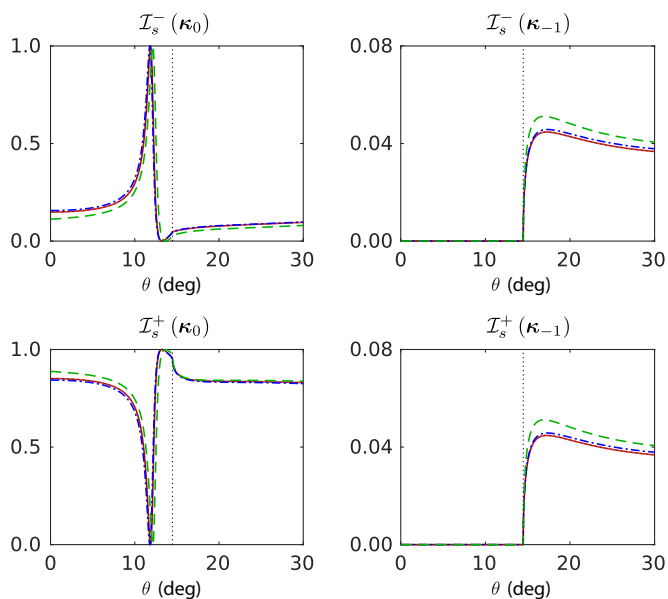
FIG. 4. Comparison between the numerically exact relative irradiances (red solid curves), the irradiances predicted by the full thin-grating model with $2N + 1 = 7$ (blue dash-dot curves), and the irradiances predicted from the two-wave-vector model (green dashed curves). The calculation is for a 25 nm thick grating with $a = 1.24$ μm, $d = 0.62$ μm, and a refractive index of $n_g = 3.5$ suspended in vacuum and subject to an s-polarized field incident at angle θ at a wavelength of 1.55 μm. The vertical, dotted, black line marks the Rayleigh anomaly.

a transfer matrix for the whole structure, in terms of which the optical properties of the structure can be calculated. To do this, consider first light characterized by a single κ in the presence of the multilayer structure of Fig. 1(a), but *without* the presence of the grating. The transfer matrix of the multilayer structure relating upward and downward propagating (or evanescent) amplitudes of light just above the multilayer in the medium with relative dielectric constant ε_1 [at $z = (-D/2)^+$] to upward and downward propagating (or evanescent) amplitudes of light at the largest z to which the substrate, with relative dielectric constant ε_Q, extends, takes the form

$$\mathbb{M}_{1Q} = \begin{bmatrix} \mathbb{T}_{Q1} - \mathbb{R}_{1Q}(\mathbb{T}_{1Q})^{-1}\mathbb{R}_{Q1} & \mathbb{R}_{1Q}(\mathbb{T}_{1Q})^{-1} \\ -(\mathbb{T}_{1Q})^{-1}\mathbb{R}_{Q1} & (\mathbb{T}_{1Q})^{-1} \end{bmatrix}. \quad (63)$$

This is a 4×4 matrix, but as long as the layered materials are isotropic or uniaxial it will be composed of 2×2 block matrices $\mathbb{T}_{ij} = \mathrm{diag}(\mathbb{T}_{ij}^s, \mathbb{T}_{ij}^p)$ where $\mathbb{T}_{ij}^{s,p}$ is the Fresnel coefficient for the transmitted s- or p-polarized fields from ε_i to ε_j and $\mathbb{R}_{ij}$ is similarly defined for their reflected counterparts [25]. We can immediately extend this to a transfer matrix $\bar{\mathbb{M}}_{1Q}$ of the layered structure involving all our $(2N + 1)$ κ_m of interest by writing

$$\bar{\mathbb{M}}_{1Q} = \begin{bmatrix} \bar{\mathbb{T}}_{Q1} - \bar{\mathbb{R}}_{1Q}(\bar{\mathbb{T}}_{1Q})^{-1}\bar{\mathbb{R}}_{Q1} & \bar{\mathbb{R}}_{1Q}(\bar{\mathbb{T}}_{1Q})^{-1} \\ -(\bar{\mathbb{T}}_{1Q})^{-1}\bar{\mathbb{R}}_{Q1} & (\bar{\mathbb{T}}_{1Q})^{-1} \end{bmatrix}, \quad (64)$$

a $4(2N + 1) \times 4(2N + 1)$ matrix where $[\bar{\mathbb{T}}_{1Q}]_{mm} = \mathrm{diag}(\mathbb{T}_{1Q}^s(\kappa_m), \mathbb{T}_{1Q}^p(\kappa_m))$ and other terms are similarly defined.

We can now construct a transfer matrix for the full structure shown in Fig. 1(a) by imagining an infinitesimal layer of material with relative dielectric constant ε_1 inserted between the bottom of the grating structure and the top of the highest layer in the multilayer structure below. Then the transfer matrix relating the upward and downward propagating (or evanescent) field amplitudes just above the grating to the upward and downward propagating (or evanescent) field amplitudes at the largest z in the substrate is given by

$$\bar{\mathbb{M}}'_{1Q} = \bar{\mathbb{M}}_g \bar{\mathbb{M}}_1(D/2)\bar{\mathbb{M}}_{1Q},$$

where

$$\bar{\mathbb{M}}_1(D/2) = \begin{bmatrix} \bar{\mathbb{L}}^+ & \bar{0} \\ \bar{0} & \bar{\mathbb{L}}^- \end{bmatrix}$$

is composed of $2(2N + 1) \times 2(2N + 1)$ block matrices with $\bar{\mathbb{L}}^{\pm}(\kappa_m) = e^{\pm i w_1(\kappa_m)D/2}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and propagates the fields from the center of the grating at $z = 0$ to the position of the substrate at $z = -D/2$. Through simple algebra we can write the elements of $\bar{\mathbb{M}}'_{1Q}$ as

$$\bar{\mathbb{M}}'_{1Q} = \begin{bmatrix} \bar{\mathbb{T}}'_{Q1} - \bar{\mathbb{R}}'_{1Q}(\bar{\mathbb{T}}'_{1Q})^{-1}\bar{\mathbb{R}}'_{Q1} & \bar{\mathbb{R}}'_{1Q}(\bar{\mathbb{T}}'_{1Q})^{-1} \\ -(\bar{\mathbb{T}}'_{1Q})^{-1}\bar{\mathbb{R}}'_{Q1} & (\bar{\mathbb{T}}'_{1Q})^{-1} \end{bmatrix}, \quad (65)$$

where

$$\bar{\mathbb{T}}'_{1Q} = \bar{\mathbb{T}}_{1Q}\bar{\mathbb{L}}^+[\bar{\mathbb{I}}_2 - \bar{\mathbb{R}}_g\bar{\mathbb{R}}_{1Q}(\bar{\mathbb{L}}^+)^2]^{-1}\bar{\mathbb{T}}_g,$$

$$\bar{\mathbb{T}}'_{Q1} = \bar{\mathbb{T}}_g\bar{\mathbb{R}}_{1Q}(\bar{\mathbb{L}}^+)^2[\bar{\mathbb{I}}_2 - \bar{\mathbb{R}}_g\bar{\mathbb{R}}_{1Q}(\bar{\mathbb{L}}^+)^2]^{-1}(\bar{\mathbb{R}}_{1Q})^{-1}\bar{\mathbb{T}}_{Q1}\bar{\mathbb{L}}^-,$$

$$\bar{\mathbb{R}}'_{1Q} = \bar{\mathbb{R}}_g + \bar{\mathbb{T}}_g\bar{\mathbb{R}}_{1Q}(\bar{\mathbb{L}}^+)^2[\bar{\mathbb{I}}_2 - \bar{\mathbb{R}}_g\bar{\mathbb{R}}_{1Q}(\bar{\mathbb{L}}^+)^2]^{-1}\bar{\mathbb{T}}_g,$$

$$\bar{\mathbb{R}}'_{Q1} = \bar{\mathbb{R}}_{Q1} + \bar{\mathbb{T}}_{1Q}\bar{\mathbb{L}}^+[\bar{\mathbb{I}}_2 - \bar{\mathbb{R}}_g\bar{\mathbb{R}}_{1Q}(\bar{\mathbb{L}}^+)^2]^{-1}\bar{\mathbb{R}}_g\bar{\mathbb{T}}_{Q1}\bar{\mathbb{L}}^+ \quad (66)$$

are easily identified as the transmission and reflection matrices of the entire structure [compare (62), (64)].

The analytic structure of the new Fresnel matrices (66) is inherited from that of the isolated grating (47) and from the $\bar{\mathbb{R}}_{ij}$ and $\bar{\mathbb{T}}_{ij}$ of the multilayer below it. Besides the poles of $\bar{\mathbb{T}}_g$ and $\bar{\mathbb{R}}_g$ signaling the waveguide modes in the isolated grating structure, we can in general expect poles in $\bar{\mathbb{R}}_{ij}$ and $\bar{\mathbb{T}}_{ij}$ signaling the presence of waveguide modes in the multilayer. The positions of the poles in the new Fresnel matrices (66) will exhibit the interaction between these excitations, and we will turn to a general analysis of the new excitations in a later publication. Here we focus on a first application our thin-grating model in the presence of a substrate, and on some of the qualitative features that arise from the interaction. Thus we consider the simplest multilayer structure possible, taking the substrate with relative dielectric constant ε_Q to extend up to $z = -D/2$. The grating, suspended in a medium of dielectric constant ε_1 which also serves as a cladding, then resides on a semi-infinite substrate of dielectric constant ε_Q, which we now relabel ε_2 (see Fig. 3). Taking ε_2 to be real and positive there are no modes associated with the substrate, and so the only effect of the substrate will be to modify the modes identified by the poles of the isolated grating structure [see Fig. 1(b)].

Insight into the nature of this modification can be gleaned from recalling the simplest picture of the grating region as an effective anisotropic slab [see Fig. 1(c)]. In a symmetric environment both s- and p-polarized waveguide modes exist, but for an environment with $\varepsilon_2 \neq \varepsilon_1$ (see Fig. 3) the modes will not survive if the asymmetry is large enough. In such a situation we expect the Wood anomalies will vanish, although of course the Rayleigh anomalies will remain. We demonstrate how our thin-grating model describes this situation by considering a grating with $D = 25$ nm, with a dielectric constant $\varepsilon_g = (3.5)^2$ appropriate for silicon, a period of $a = 1.8$ μm, and a fill fraction $d/a = 0.4$ in vacuum ($\varepsilon_1 = 1$), located above a substrate of fused silica [$\varepsilon_2 = (1.44)^2$] and subject to s-polarized light from below at a vacuum wavelength of $\lambda = 1.55$ μm. For the resulting $\varepsilon_{\mathrm{layer}}$ we would require an $\varepsilon_1 > (1.38)^2$ for a waveguide mode to be contained within the guiding layer, so the asymmetry here is too great to allow for Wood anomalies, and only Rayleigh anomalies should survive. Assuming $\varepsilon_2 > \varepsilon_1$, this can be confirmed by solving (B1) for ε_1 with $\kappa = \tilde{\omega}\sqrt{\varepsilon_2}$. In agreement with this simple argument, the reflected, transmitted, and diffracted light intensities exhibits only cusplike Rayleigh anomalies, as seen in Fig. 5. In blue (dashed-dot) we plot a calculation with $(2N + 1) = 7$ wave vectors using (49), while in red (solid) we plot the exact result found numerically from the approach of Whittaker and Culshaw [22,23]. There is excellent qualitative and good quantitative agreement between the results of the thin-grating model and the exact result, especially considering that a parameter $2\pi\sqrt{\varepsilon_g}D/\lambda$, which should obviously be small for our thin-grating approximations (14) to be valid, is here about 0.35. There is only a significant relative correction in the diffracted intensities at κ_{-1}, where the diffracted intensities themselves are very small.

In order for this grating to exhibit a Wood anomaly, we need to move to a more symmetric system. For dielectric systems, as the mismatch between ε_1 and ε_2 is decreased,

FIG. 5. Comparison between relative irradiance calculations that are numerically exact (red solid curves) and those predicted by the full thin-grating model with $2N + 1 = 7$ (blue dash-dot curves) around a Rayleigh anomaly. The calculation is for a 25 nm thick grating with $a = 1.8$ μm, $d = 0.72$ μm, and a refractive index of $n_g = 3.5$. The system has a substrate with index $n_Q = 1.44$, a vacuum cladding, and is subject to an s-polarized incident field from the substrate at angle θ and with a vacuum wavelength of 1.55 μm.

FIG. 6. Comparison of relative irradiance calculations that are numerically exact (red solid curves) and those predicted by the full thin-grating model (blue dash-dot curves). The calculation is for a 25 nm thick grating with $a = 1.8$ μm, $d = 0.72$ μm, and a refractive index of $n_g = 3.5$. The system has a substrate with index $n_Q = 1.44$, a cladding with index $n_1 = 1.42$, and is subject to an s-polarized field, incident from the substrate at angle θ with a vacuum wavelength 1.55 μm.
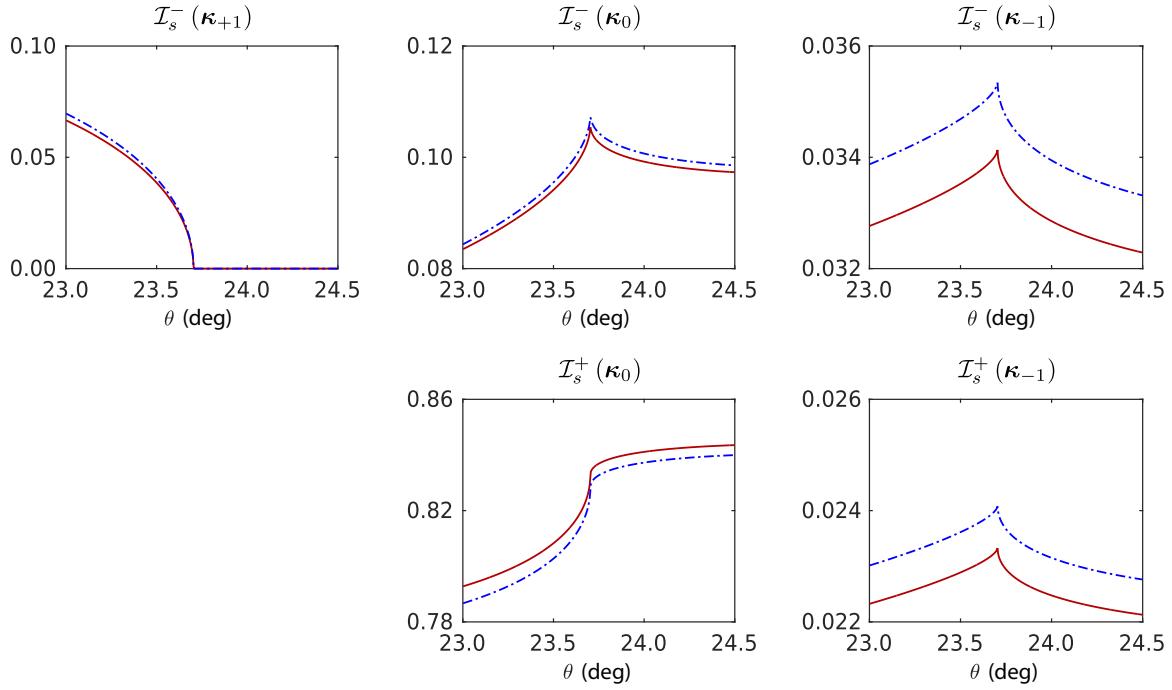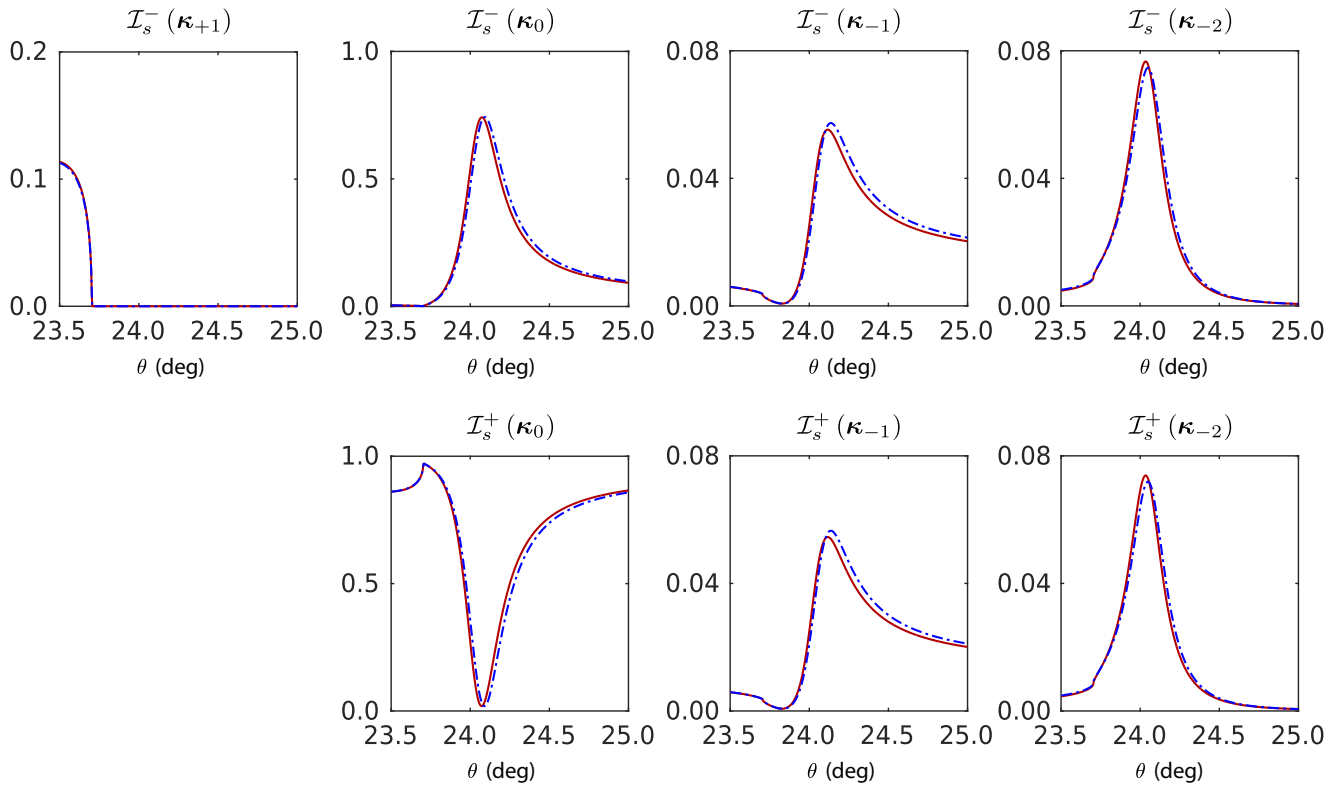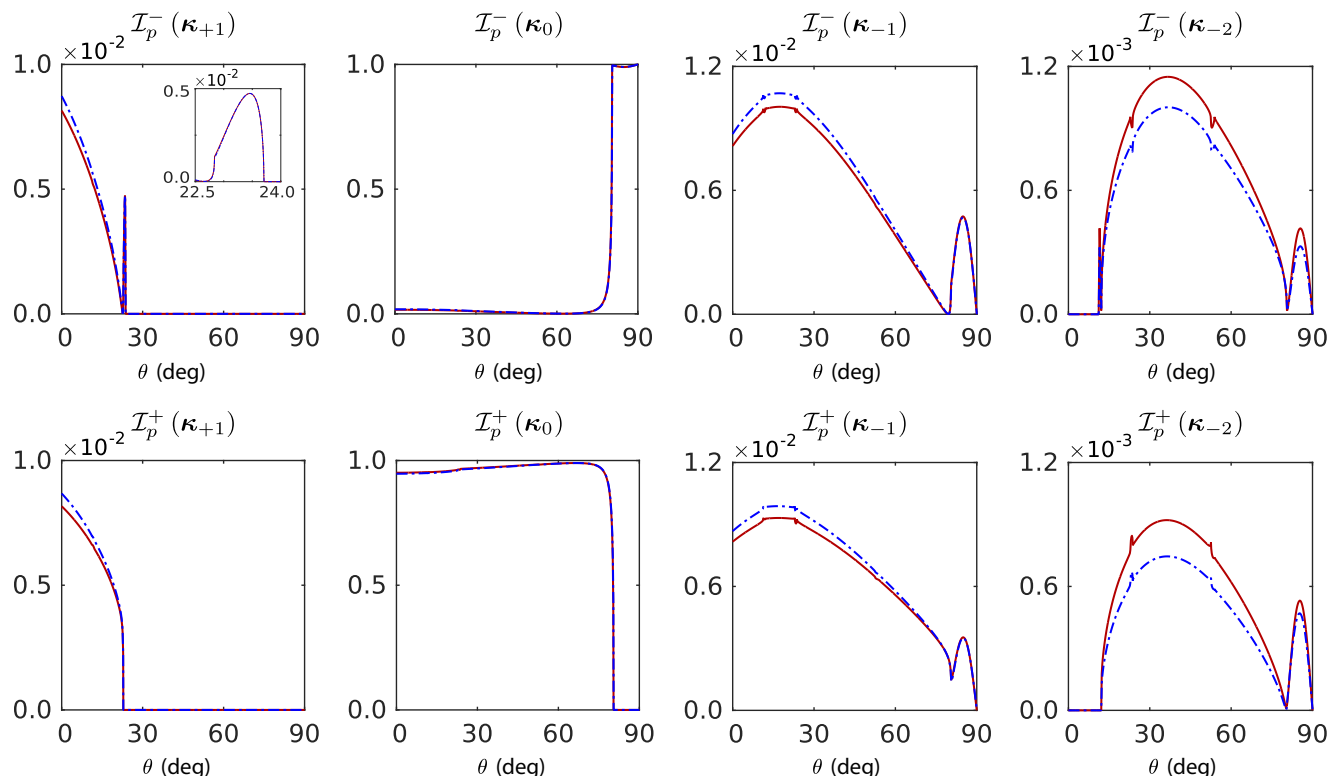
FIG. 7. Comparison of relative irradiance calculations that are numerically exact (red solid curves) and those predicted by the full thin-grating model (blue dash-dot curves). The calculation is for a 25 nm thick grating with $a = 1.8$ μm, $d = 0.72$ μm, and a refractive index of $n_g = 3.5$. The system has a substrate with index $n_Q = 1.44$, a cladding with index $n_1 = 1.42$, and is subject to a p-polarized field, incident from the substrate at angle θ with a vacuum wavelength 1.55 μm. The insert in the graph of $\mathcal{I}_p^-(\kappa_{+1})$ shows the detail around $\theta = 23.5°$.

first an s-polarized mode appears, and then a p-polarized mode. Thus one can expect situations where there will be a Wood anomaly associated with s-polarized light and not with p-polarized light, but not the reverse. This result follows naturally from our approach that associated with the grating region an effective layer response. For the current grating the cutoff for p-polarized light is $\varepsilon_1 = (1.4396)^2$, only very slightly lower than the substrate $\varepsilon_2 = (1.44)^2$; recall that the cutoff for s-polarized light is $\varepsilon_1 = (1.38)^2$. To move into this regime, we raise ε_1 to $(1.42)^2$, and keep all other parameters the same; for this value the simple argument used above predicts a waveguide mode for s-polarized light, but not for p-polarized light. In accord with this, the calculated reflected, transmitted, and diffracted light irradiances shown in Figs. 6 and 7 exhibit Wood and Rayleigh anomalies for s-polarized light, but only Rayleigh anomalies for p-polarized light. Again in blue (dashed-dot) we plot a calculation using (49) with $2N + 1 = 7$, while in red (solid) we give the results from a full numerical calculation using the approach of Whittaker and Culshaw [22,23]. For s-polarized light we focus on the region around the Wood anomaly; note that with the field incident from the substrate, which has a higher index than the cladding, the forward-diffracted fields become evanescent before the backward-diffracted fields. For p-polarized light we plot the response for all incident angles; the absence of a Wood anomaly leaves somewhat unremarkable results for specularly reflected and transmitted light, but yields several noteworthy features in the diffracted components, which

can propagate up to $m = -3$. Rayleigh anomalies when the $m = +1$, -2, and -3 diffracted orders transition between evanescence and propagation lead to Rayleigh anomalies that appear as nonanalyticities in the $m = -1$ and $m = -2$ beams. Additionally, for angles of incidence beyond that which would yield total internal reflection were the grating absent, the specular reflectance does not remain at unity. A small dip in the specular reflectance follows the Rayleigh anomaly associated with this transition, which is compensated by an increase in the irradiance of the remaining diffracted components. They display peaks over this range, which finally drop to zero as the incidence approaches grazing. We note excellent agreement between the thin-grating results (49) and the exact calculation throughout the plots in Figs. 6 and 7, with the largest relative disagreements appearing only when the intensities involved are very small.

Although not shown, we note that if the asymmetry between cladding and substrate is decreased further so that a p-polarized Wood anomaly appears, we observe a small shift between its location as predicted by (49) and the full numerical results, which does not occur for the s-polarized Wood anomaly shown in Fig. 6. For both s- and p-polarized Wood anomalies, the disagreements with the full numerical calculations increase as the dimensionless optical thickness parameters $\tilde{D}_s$ and $\tilde{D}_p$, given by (B4) and (B5) in Appendix B, approach unity. In that Appendix we show that this signals the breakdown of our approximate treatment of the waveguide modes in the effective anisotropic slab.

V. CONCLUSIONS

In this work we have presented a treatment for the optical response of thin gratings. Although approximate, it nonetheless respects energy conservation exactly, even if there are large exchanges of energy between specular and diffracted fields, and between specularly transmitted and reflected fields. These large exchanges are associated with Rayleigh and Wood anomalies. Our Green's function approach makes it easy to see how the anomalies arise from the structure of the equations that describe the specular and diffracted fields, with square-root singularities associated with Rayleigh anomalies and poles with Wood anomalies; the poles are linked to effective waveguide modes of the grating region that are easily identified in the thin-grating limit. This helps in understanding the optical response even if a set of coupled wave vector equations must be solved for the specular and diffracted fields. Yet, where only a few wave vectors are important, analytic expressions can be given directly for the specular and diffracted fields. Comparison with full numerical solutions of a 1D grating response confirms that our approximate solution is in excellent agreement with the exact response, even near the anomalies.

We expect that the development of approximate yet accurate treatments of thin gratings, such as the one presented here, will play an important role in enabling their use as probes of optical systems. The calculations can be made more easily than full numerical treatments, and the physics can be identified in the reasonably simple sets of equations that are used in calculations.

ACKNOWLEDGMENTS

APPENDIX A: FOURIER COMPONENTS FOR A RECTANGULAR GRATING

In this section we evaluate $\boldsymbol{\chi}_{v[m]}$ for a one-dimensional rectangular grating, such as the one presented in Fig. 1(b), composed of isotropic materials. Recalling (17), we have

$$\boldsymbol{\chi}_{v[m]} = \frac{1}{a} \int_{-a/2}^{a/2} e^{-imK\zeta} [\boldsymbol{\chi}_{\mathrm{mod}}(\zeta) - \langle \boldsymbol{\chi}_{\mathrm{mod}} \rangle] d\zeta, \qquad (A1)$$

where from (13) and our definition of $\boldsymbol{\varepsilon}(\zeta) \equiv \boldsymbol{\varepsilon}_1 + \boldsymbol{\chi}_{\mathrm{add}}(\zeta)$, we can write

$$\chi_{\mathrm{mod}}^{\parallel}(\zeta) = \varepsilon^{\parallel}(\zeta) - \varepsilon_1,$$

$$\chi_{\mathrm{mod}}^{\perp}(\zeta) = \varepsilon_1 \left(1 - \frac{\varepsilon_1}{\varepsilon^{\perp}(\zeta)} \right).$$

Note that for the system considered, we have

$$\varepsilon^{\perp}(\zeta) = \varepsilon^{\parallel}(\zeta) = \begin{cases} \varepsilon_g, & |\zeta| \leqslant d/2, \\ \varepsilon_1, & |\zeta| > d/2, \end{cases}$$

within a single period of our grating. Additionally, we note that $\langle \boldsymbol{\chi}_{\mathrm{mod}} \rangle$, found from (20), yields $\langle \chi_{\mathrm{mod}}^{\parallel} \rangle = \varepsilon_{\mathrm{layer}}^{\parallel} - \varepsilon_1$ and

$\langle \chi_{\mathrm{mod}}^{\perp} \rangle = \varepsilon_1 (1 - \varepsilon_1 / \varepsilon_{\mathrm{layer}}^{\perp})$ where $\varepsilon_{\mathrm{layer}}^{\parallel}$ and $\varepsilon_{\mathrm{layer}}^{\perp}$ are found from (21). At this point we can evaluate (A1) to find

$$\chi_{v[m]}^{\parallel} = (\varepsilon_{\mathrm{layer}}^{\parallel} - \varepsilon_1) \mathrm{sinc} \left(\frac{mKd}{2} \right),$$

$$\chi_{v[m]}^{\perp} = \varepsilon_1 \left(1 - \frac{\varepsilon_1}{\varepsilon_{\mathrm{layer}}^{\perp}} \right) \mathrm{sinc} \left(\frac{mKd}{2} \right).$$

APPENDIX B: WAVEGUIDE MODES

Here we compare the exact and approximate dispersion relations of the waveguide modes of a thin uniaxial slab. For $\hat{\boldsymbol{z}}$ perpendicular to the slab we take the relative dielectric tensor to be $\boldsymbol{\varepsilon}_{\mathrm{layer}} = \varepsilon_{\mathrm{layer}}^{\parallel}(\hat{\boldsymbol{x}}\hat{\boldsymbol{x}} + \hat{\boldsymbol{y}}\hat{\boldsymbol{y}}) + \varepsilon_{\mathrm{layer}}^{\perp}\hat{\boldsymbol{z}}\hat{\boldsymbol{z}}$, with the cladding and substrate of the slab taken to be isotropic media respectively characterized by relative dielectric constants ε_1 and ε_2. The exact solution for the waveguide modes [32] of this system is

(a) *s* polarization

(b) *p* polarization
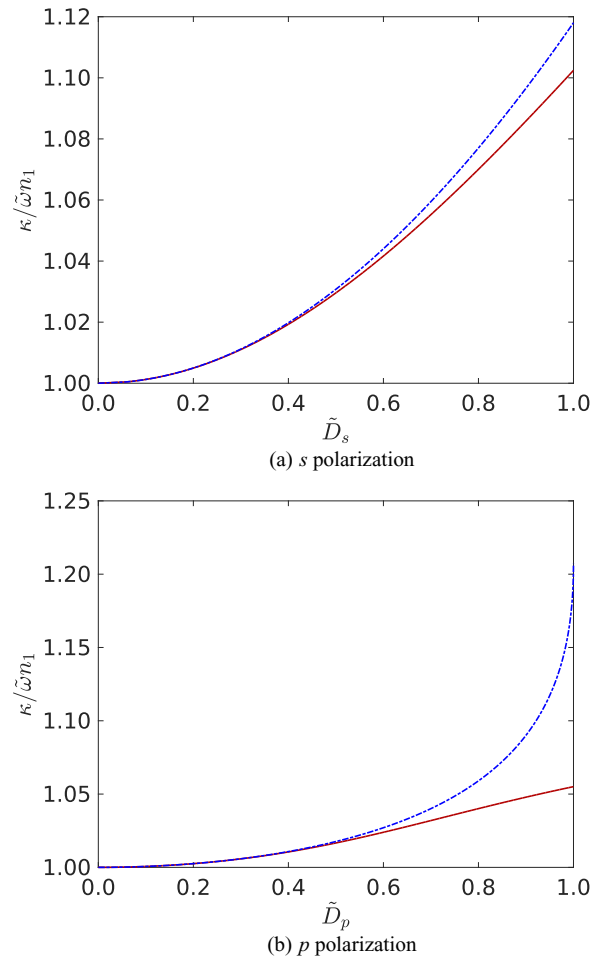
FIG. 8. Comparison of exact (solid red curves) and approximate (blue dash-dot curves) waveguide mode dispersion calculations over a range of slab thicknesses. The calculation is for a thin uniaxial slab characterized by $\varepsilon_{\mathrm{layer}}^{\parallel} = 6.11$ and $\varepsilon_{\mathrm{layer}}^{\perp} = 3.03$, suspended in a medium with index 1.42, subject to either an (a) *s*-polarized or (b) *p*-polarized incident field with a vacuum wavelength of 1.55 μm. The thickness parameters $\tilde{D}_s$ and $\tilde{D}_p$ are given by (B4) and (B5), respectively.

given by

$$\cot(hD) = \frac{h^2 - qp}{h(q+p)}, \tag{B1}$$

where

$$h = \sqrt{\left(\frac{\varepsilon_{\text{layer}}^{\parallel}}{\varepsilon_{\text{layer}}^{b}}\right)(\tilde{\omega}^2 \varepsilon_{\text{layer}}^{b} - \kappa^2)},$$

$$q = a_q \sqrt{\tilde{\omega}^2(\varepsilon_{\text{layer}}^{b} - \varepsilon_1) - \frac{\varepsilon_{\text{layer}}^{b}}{\varepsilon_{\text{layer}}^{\parallel}} h^2},$$

$$p = a_p \sqrt{\tilde{\omega}^2(\varepsilon_{\text{layer}}^{b} - \varepsilon_2) - \frac{\varepsilon_{\text{layer}}^{b}}{\varepsilon_{\text{layer}}^{\parallel}} h^2},$$

and where $\varepsilon_{\text{layer}}^{b} = \varepsilon_{\text{layer}}^{\parallel}$ and $a_q = a_p = 1$ for s polarization, and $\varepsilon_{\text{layer}}^{b} = \varepsilon_{\text{layer}}^{\perp}$, $a_q = \varepsilon_{\text{layer}}^{\parallel}/\varepsilon_1$, and $a_p = \varepsilon_{\text{layer}}^{\parallel}/\varepsilon_2$ for p polarization. Taking $\varepsilon_1 = \varepsilon_2$, approximate dispersion relations can be determined directly from (30); solving those equations yields

$$\frac{\kappa^2}{\tilde{\omega}^2 \varepsilon_1} = 1 + \frac{1}{4}\tilde{D}_s^2, \tag{B2}$$

for s polarization, and

$$\frac{\kappa^2}{\tilde{\omega}^2 \varepsilon_1} = \frac{2}{1 + \sqrt{1 - \tilde{D}_p}}, \tag{B3}$$

for p polarization [33], where

$$\tilde{D}_s = \tilde{\omega} n_1 \left(\frac{\varepsilon_{\text{layer}}^{\parallel}}{\varepsilon_1} - 1\right) D, \tag{B4}$$

$$\tilde{D}_p = \tilde{\omega} n_1 \left(1 - \frac{\varepsilon_1}{\varepsilon_{\text{layer}}^{\perp}}\right) D. \tag{B5}$$

The approximate dispersion relations can be shown to agree with the lowest order solution of the exact relations to first order in the grating thickness. The exact and approximate dispersion relations are shown in Fig. 8 over a range of waveguide thicknesses. For the purposes of the calculation we set $\varepsilon_{\text{layer}}^{\parallel} = 6.11$ and $\varepsilon_{\text{layer}}^{\perp} = 3.03$, with $\varepsilon_1 = (1.42)^2$, as used in Fig. 6 with the absence of any diffraction.

For both polarizations we begin to see significant deviations in Fig. 8 as $\tilde{D}_j \rightarrow 1$, where $j = s, p$. The s-polarized case has a relative deviation of 1.5% at $\tilde{D}_s = 1$, which corresponds to a thickness of 86 nm, while the p-polarized case has a deviation of 27% at $\tilde{D}_p = 1$. The significantly larger deviation in the p-polarized results can be attributed to two factors. The first is that for $\tilde{D}_p > 1$ the square root in the denominator of (B2) becomes imaginary, giving a firm cutoff for its valid comparison to the exact solution, and the second is due to the fact that $\varepsilon_{\text{layer}}^{\parallel} > \varepsilon_{\text{layer}}^{\perp}$ for the cases considered in this paper. While the chosen cutoff for the s-polarized calculation was found to be 86 nm, the breakdown of the p-polarized case occurs at 520 nm, a thickness well beyond our underlying assumption that $w_1 D \ll 1$. To provide a better comparison to the s-polarized case, we note that (B3) has a relative deviation of approximately 1.5% at $\tilde{D}_p = 0.7$ which corresponds to a thickness of 364 nm.

APPENDIX C: ENERGY CONSERVATION

Here we confirm that our approximate treatment of diffraction and scattering across the grating satisfies energy conservation exactly in the limit of no absorption. To do this, we start with the difference between the total irradiance of the outgoing and incident fields

$$\Delta I = 2c\epsilon_0 n_1 \sum_m (|\bar{\mathbb{E}}_{\text{out}}^+(\boldsymbol{\kappa}_m)|^2 + |\bar{\mathbb{E}}_{\text{out}}^-(\boldsymbol{\kappa}_m)|^2)\mathbb{W}(\boldsymbol{\kappa}_m)\cos\theta(\boldsymbol{\kappa}_m) - 2c\epsilon_0 n_1 \sum_m (|\bar{\mathbb{E}}_{\text{in}}^+(\boldsymbol{\kappa}_m)|^2 + |\bar{\mathbb{E}}_{\text{in}}^-(\boldsymbol{\kappa}_m)|^2)\mathbb{W}(\boldsymbol{\kappa}_m)\cos\theta(\boldsymbol{\kappa}_m), \tag{C1}$$

where $\mathbb{W}(\boldsymbol{\kappa}_m) \equiv 1$ for propagating fields, and $\mathbb{W}(\boldsymbol{\kappa}_m) \equiv 0$ for evanescent fields, such that (C1) considers the difference in the incoming and outgoing energy from the grating via propagating fields; also $\cos\theta(\boldsymbol{\kappa}_m) = w_1(\boldsymbol{\kappa}_m)/\tilde{\omega} n_1$. Denoting by $\bar{\theta}$ the diagonal matrix with elements $\bar{\theta}_{mm} = 2c\epsilon_0 n_1 \cos\theta(\boldsymbol{\kappa}_m)$ we can write the difference in irradiance as

$$\Delta I = [\bar{\mathbb{E}}_{\text{out}}^{+*} \quad \bar{\mathbb{E}}_{\text{out}}^{-*}]\begin{bmatrix} \bar{\mathbb{W}} & \bar{0} \\ \bar{0} & \bar{\mathbb{W}} \end{bmatrix}\begin{bmatrix} \bar{\theta} & \bar{0} \\ \bar{0} & \bar{\theta} \end{bmatrix}\begin{bmatrix} \bar{\mathbb{W}} & \bar{0} \\ \bar{0} & \bar{\mathbb{W}} \end{bmatrix}\begin{bmatrix} \bar{\mathbb{E}}_{\text{out}}^+ \\ \bar{\mathbb{E}}_{\text{out}}^- \end{bmatrix} - [\bar{\mathbb{E}}_{\text{in}}^{+*} \quad \bar{\mathbb{E}}_{\text{in}}^{-*}]\begin{bmatrix} \bar{\mathbb{W}} & \bar{0} \\ \bar{0} & \bar{\mathbb{W}} \end{bmatrix}\begin{bmatrix} \bar{\theta} & \bar{0} \\ \bar{0} & \bar{\theta} \end{bmatrix}\begin{bmatrix} \bar{\mathbb{W}} & \bar{0} \\ \bar{0} & \bar{\mathbb{W}} \end{bmatrix}\begin{bmatrix} \bar{\mathbb{E}}_{\text{in}}^+ \\ \bar{\mathbb{E}}_{\text{in}}^- \end{bmatrix}, \tag{C2}$$

where $\bar{0}$ is a $2(2N+1) \times 2(2N+1)$ matrix of zeros, and $\bar{\mathbb{W}}$ is a $2(2N+1) \times 2(2N+1)$ block-diagonal matrix with diagonal elements

$$\bar{\mathbb{W}}(\boldsymbol{\kappa}_m) = \mathbb{W}(\boldsymbol{\kappa}_m)\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

After recalling (48), (C2) becomes

$$\Delta \mathbb{I} = [\bar{\mathbb{E}}_{\text{in}}^{+*} \quad \bar{\mathbb{E}}_{\text{in}}^{-*}]\begin{bmatrix} \bar{\mathbb{W}} & \bar{0} \\ \bar{0} & \bar{\mathbb{W}} \end{bmatrix}\left(\mathscr{S}^{\dagger}\begin{bmatrix} \bar{\theta}' & \bar{0} \\ \bar{0} & \bar{\theta}' \end{bmatrix}\mathscr{S} - \begin{bmatrix} \bar{\theta}' & \bar{0} \\ \bar{0} & \bar{\theta}' \end{bmatrix}\right)\begin{bmatrix} \bar{\mathbb{W}} & \bar{0} \\ \bar{0} & \bar{\mathbb{W}} \end{bmatrix}\begin{bmatrix} \bar{\mathbb{E}}_{\text{in}}^+ \\ \bar{\mathbb{E}}_{\text{in}}^- \end{bmatrix}, \tag{C3}$$

where $\bar{\theta}' = \bar{\mathbb{W}}\bar{\theta}\bar{\mathbb{W}}$ and we have made use of the fact that $\bar{\mathbb{W}}^2 = \bar{\mathbb{W}}$.

For convenience we introduce the matrix

$$\bar{\mathcal{C}}^{\pm} = \epsilon_0 \bar{G}^{\pm} \bar{\chi}_{\text{tot}} (\bar{1}_3 - \epsilon_0 \bar{g} \bar{\chi}_{\text{tot}})^{-1}, \tag{C4}$$

where $\bar{\chi}_{\text{tot}} = \bar{\chi}_o + \bar{\chi}_v$. Using (C4), we can write our scattering matrix from (48) as

$$\mathscr{S} = \begin{bmatrix} \bar{\sigma}_{\text{out}}^+ & \bar{0} \\ \bar{0} & \bar{\sigma}_{\text{out}}^- \end{bmatrix} \left(\bar{1}_6 + \begin{bmatrix} \bar{\mathcal{C}}^+ & \bar{\mathcal{C}}^+ \\ \bar{\mathcal{C}}^- & \bar{\mathcal{C}}^- \end{bmatrix} \right) \begin{bmatrix} \bar{\sigma}_{\text{in}}^+ & \bar{0} \\ \bar{0} & \bar{\sigma}_{\text{in}}^- \end{bmatrix}, \tag{C5}$$

and can then reduce (C3) to

$$\Delta I = \begin{bmatrix} \bar{\mathbb{E}}_{\text{in}}^{+*} \bar{\mathbb{W}} \bar{\sigma}_{\text{out}}^+ & \bar{\mathbb{E}}_{\text{in}}^{-*} \bar{\mathbb{W}} \bar{\sigma}_{\text{out}}^- \end{bmatrix}$$
$$\times \left(\begin{bmatrix} \bar{\Theta}^+ & \bar{0} \\ \bar{0} & \bar{\Theta}^- \end{bmatrix} \begin{bmatrix} \bar{\mathcal{C}}^+ & \bar{\mathcal{C}}^+ \\ \bar{\mathcal{C}}^- & \bar{\mathcal{C}}^- \end{bmatrix} + \begin{bmatrix} \bar{\mathcal{C}}^{+\dagger} & \bar{\mathcal{C}}^{-\dagger} \\ \bar{\mathcal{C}}^{+\dagger} & \bar{\mathcal{C}}^{-\dagger} \end{bmatrix} \begin{bmatrix} \bar{\Theta}^+ & \bar{0} \\ \bar{0} & \bar{\Theta}^- \end{bmatrix} + \begin{bmatrix} \bar{\mathcal{C}}^{+\dagger} & \bar{\mathcal{C}}^{-\dagger} \\ \bar{\mathcal{C}}^{+\dagger} & \bar{\mathcal{C}}^{-\dagger} \end{bmatrix} \begin{bmatrix} \bar{\Theta}^+ & \bar{0} \\ \bar{0} & \bar{\Theta}^- \end{bmatrix} \begin{bmatrix} \bar{\mathcal{C}}^+ & \bar{\mathcal{C}}^+ \\ \bar{\mathcal{C}}^- & \bar{\mathcal{C}}^- \end{bmatrix} \right) \begin{bmatrix} \bar{\sigma}_{\text{in}}^+ \bar{\mathbb{W}} \bar{\mathbb{E}}_{\text{in}}^+ \\ \bar{\sigma}_{\text{in}}^- \bar{\mathbb{W}} \bar{\mathbb{E}}_{\text{in}}^- \end{bmatrix}, \tag{C6}$$

where $\bar{\Theta}^{\pm} = \bar{\sigma}_{\text{in}}^{\pm} \bar{\theta}' \bar{\sigma}_{\text{out}}^{\pm}$.

In simplifying (C6) we find

$$\Delta I = \begin{bmatrix} \bar{\mathbb{E}}_{\text{in}}^{+*} & \bar{\mathbb{E}}_{\text{in}}^{-*} \end{bmatrix} \begin{bmatrix} \bar{\bar{\mathbb{I}}}_{++} & \bar{\bar{\mathbb{I}}}_{+-} \\ \bar{\bar{\mathbb{I}}}_{-+} & \bar{\bar{\mathbb{I}}}_{--} \end{bmatrix} \begin{bmatrix} \bar{\mathbb{E}}_{\text{in}}^+ \\ \bar{\mathbb{E}}_{\text{in}}^- \end{bmatrix}, \tag{C7}$$

where

$$\bar{\bar{\mathbb{I}}}_{\pm\pm} = \bar{\mathbb{W}} \bar{\sigma}_{\text{out}}^{\pm} (\bar{\Theta}^{\pm} \bar{\mathcal{C}}^{\pm} + \bar{\mathcal{C}}^{\pm\dagger} \bar{\Theta}^{\pm} + \bar{\mathcal{C}}^{+\dagger} \bar{\Theta}^+ \bar{\mathcal{C}}^+ + \bar{\mathcal{C}}^{-\dagger} \bar{\Theta}^- \bar{\mathcal{C}}^-) \bar{\sigma}_{\text{in}}^{\pm} \bar{\mathbb{W}},$$
$$\bar{\bar{\mathbb{I}}}_{\pm\mp} = \bar{\mathbb{W}} \bar{\sigma}_{\text{out}}^{\pm} (\bar{\Theta}^{\pm} \bar{\mathcal{C}}^{\pm} + \bar{\mathcal{C}}^{\mp\dagger} \bar{\Theta}^{\mp} + \bar{\mathcal{C}}^{+\dagger} \bar{\Theta}^+ \bar{\mathcal{C}}^+ + \bar{\mathcal{C}}^{-\dagger} \bar{\Theta}^- \bar{\mathcal{C}}^-) \bar{\sigma}_{\text{in}}^{\mp} \bar{\mathbb{W}}. \tag{C8}$$

Before further proceeding we note that through the use of the identity $\bar{\sigma}_{\text{out}}^{\pm} \bar{\sigma}_{\text{in}}^{\pm} = \bar{1}_2$ we can write

$$\bar{\Theta}^{\pm} \bar{G}^{\pm} = i \tilde{\omega} c D \bar{\sigma}_{\text{in}}^{\pm} \bar{\mathbb{W}} \bar{\sigma}_{\text{out}}^{\pm}, \tag{C9}$$

which then allows us to write

$$\bar{\Theta}^{\pm} \bar{\mathcal{C}}^{\pm} = i \tilde{\omega} \epsilon_0 c D \bar{\sigma}_{\text{in}}^{\pm} \bar{\mathbb{W}} \bar{\sigma}_{\text{out}}^{\pm} \bar{\chi}_{\text{tot}} (\bar{1}_3 - \epsilon_0 \bar{g} \bar{\chi}_{\text{tot}})^{-1},$$
$$\bar{\mathcal{C}}^{\pm\dagger} \bar{\Theta}^{\pm} = -i \tilde{\omega} \epsilon_0 c D (\bar{1}_3 - \epsilon_0 \bar{\chi}_{\text{tot}}^{\dagger} \bar{g}^{\dagger})^{-1} \bar{\chi}_{\text{tot}}^{\dagger} \bar{\sigma}_{\text{in}}^{\pm} \bar{\mathbb{W}} \bar{\sigma}_{\text{out}}^{\pm}, \tag{C10}$$

where in the latter expression we have made use of the fact that $(\bar{\sigma}_{\text{in}}^{\pm} \bar{\mathbb{W}} \bar{\sigma}_{\text{out}}^{\pm})^{\dagger} = \bar{\sigma}_{\text{in}}^{\pm} \bar{\mathbb{W}} \bar{\sigma}_{\text{out}}^{\pm}$. Next, note that when we premultiply the $3(2N + 1) \times 3(2N + 1)$ block-diagonal matrix $\bar{G}^+$ by the $3(2N + 1) \times 3(2N + 1)$ block-diagonal matrix $\bar{\sigma}_{\text{in}}^+ \bar{\mathbb{W}} \bar{\sigma}_{\text{out}}^+$ we get a block-diagonal matrix $\bar{\sigma}_{\text{in}}^+ \bar{\mathbb{W}} \bar{\sigma}_{\text{out}}^+ \bar{G}^+$ in which each block associated with a propagating (diffracted or scattered) order equals the corresponding block of $\bar{G}^+$, but which vanishes if the block is associated with an evanescent order. In the same way the blocks of $\bar{\sigma}_{\text{in}}^- \bar{\mathbb{W}} \bar{\sigma}_{\text{out}}^- \bar{G}^-$ associated with propagating orders equal the corresponding blocks of $\bar{G}^-$, while those associated with evanescent orders vanish. Recalling that $\bar{g}(\kappa_m) = \frac{1}{2}[\bar{G}^+(\kappa_m) + \bar{G}^-(\kappa_m)]$, we have $\bar{\sigma}_{\text{in}}^+ \bar{\mathbb{W}} \bar{\sigma}_{\text{out}}^+ \bar{G}^+ + \bar{\sigma}_{\text{in}}^- \bar{\mathbb{W}} \bar{\sigma}_{\text{out}}^- \bar{G}^- = \bar{g} - \bar{g}^{\dagger}$, where the sum on the left-hand side will either equal $2\bar{g}$ or $\bar{0}$ as $\bar{g}(\kappa_m) = -\bar{g}^{\dagger}(\kappa_m)$ for propagating orders, while $\bar{g}(\kappa_m) = \bar{g}^{\dagger}(\kappa_m)$ for evanescent orders. With this in mind and from the use of (C10) we have

$$\bar{\mathcal{C}}^{+\dagger} \bar{\Theta}^+ \bar{\mathcal{C}}^+ + \bar{\mathcal{C}}^{-\dagger} \bar{\Theta}^- \bar{\mathcal{C}}^- = -i \tilde{\omega} \epsilon_0^2 c D (\bar{1}_3 - \epsilon_0 \bar{\chi}_{\text{tot}}^{\dagger} \bar{g}^{\dagger})^{-1} \bar{\chi}_{\text{tot}}^{\dagger} (\bar{g} - \bar{g}^{\dagger}) \bar{\chi}_{\text{tot}} (\bar{1}_3 - \epsilon_0 \bar{g} \bar{\chi}_{\text{tot}})^{-1}. \tag{C11}$$

Finally, through the substitution of the expressions for $\bar{\Theta}^{\pm} \bar{\mathcal{C}}^{\pm}$, $\bar{\mathcal{C}}^{\pm\dagger} \bar{\Theta}^{\pm}$, (C10) and (C11), into (C8), and some simple algebra, we find $\bar{\bar{\mathbb{I}}}_{\pm\pm} = \bar{\Gamma}^{\pm} \bar{\sigma}_{\text{in}}^{\pm} \bar{\mathbb{W}}$ and $\bar{\bar{\mathbb{I}}}_{\pm\mp} = \bar{\Gamma}^{\pm} \bar{\sigma}_{\text{in}}^{\mp} \bar{\mathbb{W}}$, where

$$\bar{\Gamma}^{\pm} = i \tilde{\omega} \epsilon_0 c D \bar{\mathbb{W}} \bar{\sigma}_{\text{out}}^{\pm} (\bar{1}_3 - \epsilon_0 \bar{\chi}_{\text{tot}}^{\dagger} \bar{g}^{\dagger})^{-1} [\bar{\chi}_{\text{tot}} - \bar{\chi}_{\text{tot}}^{\dagger}] (\bar{1}_3 - \epsilon_0 \bar{g} \bar{\chi}_{\text{tot}})^{-1}. \tag{C12}$$

For nonabsorbing gratings $\bar{\chi}_{\text{tot}}$ is Hermitian, $\bar{\bar{\mathbb{I}}}_{\pm\pm} = \bar{\bar{\mathbb{I}}}_{\pm\mp} = 0$, and $\Delta I = 0$ from (C7). If absorption is present (C12) and (C7) can be used to calculate its effect on the energy balance between incident and scattered fields.

---

[1] R. W. Wood, Philos. Mag., Ser. 6 **4**, 396 (1902).

[2] D. Maystre, in *Plasmonics: From Basics to Advanced Topics*, edited by S. Enoch and N. Bonod (Springer, Berlin, 2012), Chap. 2, pp. 39–83.

[3] L. Rayleigh, Proc. R. Soc. A **79**, 399 (1907).

[4] A. Hessel and A. A. Oliner, Appl. Opt. **4**, 1275 (1965).

[5] U. Fano, J. Opt. Soc. Am. **31**, 213 (1941).

[6] A. A. Maradudin, I. Simonsen, J. Polanco, and R. M. Fitzgerald, J. Opt. **18**, 024004 (2016).

[7] F. Falco, T. Tamir, and K. M. Leung, J. Opt. Soc. Am. A **21**, 1621 (2004).

[8] B. S. Zhao, G. Meijer, and W. Schöllkopf, Phys. Rev. Lett. **104**, 240404 (2010).

[9] C. Yi, Y. J. Yoo, Y. J. Kim, K. W. Kim, Y. Lee, and J. Y. Rhee, J. Phys. D **49**, 195103 (2016).

[10] S. S. Wang and R. Magnusson, Appl. Opt. **34**, 2414 (1995).

[11] G. Niederer, H. P. Herzig, J. Shamir, H. Thiele, M. Schnieper, and C. Zschokke, Appl. Opt. **43**, 1683 (2004).

[12] D. L. Brundrett, E. N. Glytsis, and T. K. Gaylord, Opt. Lett. **23**, 700 (1998).

[13] M. Grande, M. A. Vincenti, T. Stomeo, G. V. Bianco, D. de Ceglia, N. Aközbek, V. Petruzzelli, G. Bruno, M. De Vittorio, M. Scalora, and A. D'Orazio, Opt. Express **22**, 31511 (2014).

[14] M. Grande, M. A. Vincenti, T. Stomeo, G. V. Bianco, D. de Ceglia, N. Aközbek, V. Petruzzelli, G. Bruno, M. De Vittorio, M. Scalora, and A. D'Orazio, Opt. Express **23**, 21032 (2015).

[15] H. Mukundan, A. S. Anderson, W. K. Grace, K. M. Grace, N. Hartman, J. S. Martinez, and B. I. Swanson, Sensors **9**, 5783 (2009).

[16] M. Eitan, Z. Iluz, Y. Yifat, A. Boag, Y. Hanein, and J. Scheuer, ACS Photonics **2**, 615 (2015).

[17] J. E. Sipe and C. M. de Sterke, J. Lightwave Technol. **19**, 1892 (2001).

[18] A. Yariv, IEEE J. Quantum Electron. **9**, 919 (1973).

[19] S. M. Norton, T. Erdogan, and G. M. Morris, J. Opt. Soc. Am. A **14**, 629 (1997).

[20] D. A. Bykov and L. L. Doskolovich, Opt. Express **23**, 19234 (2015).

[21] L. Li, J. Opt. Soc. Am. A **13**, 1024 (1996).

[22] D. M. Whittaker and I. S. Culshaw, Phys. Rev. B **60**, 2610 (1999).

[23] M. Liscidini, D. Gerace, L. C. Andreani, and J. E. Sipe, Phys. Rev. B **77**, 035324 (2008).

[24] A. Ahmed, M. Liscidini, and R. Gordon, IEEE Photonics J. **2**, 884 (2010).

[25] J. E. Sipe, J. Opt. Soc. Am. B **4**, 481 (1987).

[26] M. Liscidini and J. E. Sipe, Appl. Phys. Lett. **91**, 253125 (2007).

[27] M. Liscidini and J. E. Sipe, J. Opt. Soc. Am. B **26**, 279 (2009).

[28] L. Ju, B. Geng, J. Horng, C. Girit, M. Martin, Z. Hao, H. A. Bechtel, X. Liang, A. Zettl, Y. R. Shen, and F. Wang, Nat. Nanotechnol. **6**, 630 (2011).

[29] W. Gao, J. Shu, C. Qiu, and Q. Xu, ACS Nano **6**, 7806 (2012).

[30] J. Fu, B. Lv, R. Li, R. Ma, W. Chen, and Z. Wang, Plasmonics **12**, 209 (2017).

[31] Y.-X. Peng, M.-D. He, Z.-J. Li, K.-J. Wang, S. Li, J.-B. Li, J.-Q. Liu, M. Long, W.-D. Hu, and X. Chen, Opt. Commun. **382**, 86 (2017).

[32] W. Liao, X. Chen, Y. Chen, Y. Xia, and Y. Chen, J. Opt. Soc. Am. A **21**, 2196 (2004).

[33] A second formal solution of the second of (30) is unphysical, in that it corresponds to values of $\kappa^2/\tilde{\omega}\varepsilon_1$ so large that the approximations leading to the derivation of (30) are invalid.