General model of optical frequency conversion in homogeneous media: Application to second-harmonic generation in an ϵ-near-zero waveguide

Jin Jer Huang,[1,*] Xin Lu Zhang,[1] Liu Yang Zhang,[2] and Jian Xin Zhang[1]
[1]*College of Science, Tianjin Polytechnic University, Tianjin 300387, China*
[2]*College of Applied Science, Harbin University of Science and Technology, West Campus, Harbin 150080, China*
(Received 23 November 2016; revised manuscript received 4 April 2017; published 19 July 2017)

Traditional optical frequency conversion model is well improved in this work. In terms of the dyadic Green's function method, a set of coupled-amplitude equations is reduced under a proposed transition layer assumption, accompanying the simultaneous integral equations. The model, as a generalization of the current frequency conversion theory, is aimed at any one-dimensional thin film or bulk nonlinear structure, allowing for arbitrary optical anisotropy and absorption without pumping and propagating limitations. The assumption reasonably simplifies the strict nonlinear boundary conditions and enables the equations to yield exact radiative field solutions. A field-enhanced phase-matching configuration is designed for second harmonic generation in a lossy ϵ-near-zero material. The high contrast of refractive indices between a substrate (silicon) and the material traps the harmonic wave inside and constructs a natural mirror reflection waveguide. A simulation in the lowest guided mode predicts an efficiency enhancement proportional to the relative wave impedance to the fifth power under a resonant condition.

DOI: 10.1103/PhysRevA.96.013836

I. INTRODUCTION

Nonlinear optics has rapidly developed since it started in the early 1960s after the laser emerged. Optical frequency conversion (FC), as a kind of representative nonlinear process, e.g., harmonic generation, frequency mixing, and in a broad sense stimulated optical scattering, was found and substantially applied to laser technology. FC, specifically harmonic generation, in a single film is mainly for nonlinearity measurement to characterize new materials and applications in surface optical microscopy and sensing [1–3]. Apart from applications in film coating, multilayer heterogenous films have richer functions originating from a combination of nonlinearity and Fabry-Pérot effect. With the advent of photonic crystals [4], nonlinear optics in photonic band-gap structures and microcavity structures were of interest due to light confinement or field enhancement realized by the increased mode density at band edges [4–8]. The nonlinearity boost opened a way to create miniature photonic devices. Moreover, the field enhancement can be greatly stimulated in metamaterials by localized surface plasmons and surface plasmon polaritons [9–11], and commonly is in accordance with the near-zero permittivity (ϵ), permeability (μ), or both, which interestingly produce a near-zero refractive index. The ϵ-near-zero (ENZ) artificial materials were especially found to demonstrate numerous extreme optical properties [12,13], such as optical tunneling [14–16], perfect absorption [17–19], enhanced emission [20–22], optical cloaking [23–25], enhanced optical nonlinearities [26–31], and so on. Generally, the Drude- or Lorentz-type model of permittivity demonstrates an $\epsilon = 0$ crossing point near the cutoff frequency (ENZ frequency) in metals and some semiconductors, like indium tin oxide, which show a plasma-like behavior [27,32–34]. Nanotechnology offers an abundant of routes to tailor material parameters, including ϵ. Up to now, silicon-based photonic crystals and/or dielectric metamaterials [35–37] and plasma-based metamaterials [38–40] are the most promising candidates for nonlinear ENZ nanostructures. In these ENZ materials, the field enhancement comes from the magnification of the longitudinal electric field of a TM-polarized wave at oblique incidence, based on the continuity of the longitudinal electric displacement across an interface [41]. Harmonic generation stimulated by field enhancement has created a large rise in efficiency on the order of 10^2 in experiments [42–44]. Importantly, a phase-mismatch-free scheme was realized in a zero-index multilayer metamaterial in 2013 [45]. It was an important advance in application of bulk metamaterials and found omnidirectional FC without phase matching (PM). Another ideal design is to locate the FC at Dirac points (zero Bloch vector at the Γ point) in nonlinear photonic crystals [46], expected to be realized in practice. Though the zero-index condition is not satisfied in the ENZ case, it still has the ability to raise FC efficiency under a quite loose PM condition. It is found in this paper that the conversion efficiency can be enhanced further in the ENZ waveguide under a PM condition during multipass second harmonic generation (SHG).

For these subwavelength nanostructures, the effective medium approximation can be well applied [47–51], in which the medium is homogenized as a normal bulk or layered material with retrieved effective parameters (linear and nonlinear susceptibilities). Therefore, a theoretical treatment on the FC is accessible without resorting to any complicated model associated with their internal constructions. The FC process was first modeled by a set of coupled-wave equations, governed by nonlinear polarization, which successfully uncovered nonlinear optical phenomena in various bulk materials [1,52]. In the infinite plane wave case of weak absorption, they turn into coupled-amplitude equations (CAE, which we call the standard model), as are widely applied now. In addition, a few other methods have been developed for the FC in nonlinear films. The first is to solve the nonlinear wave equations directly through an inversion method given known wave vectors. The solution is formulated as a superposition of particular solutions

*hjinjer@126.com

and homogeneous solutions [53–57]. The second is by the dyadic [58–61] or matrix [62] Green's function skill, based on Berreman equations [63], to describe the solution as an integral. In particular, Sipe got an exact result in the isotropic case [59], which was used in simple ENZ materials [43,44]. However, for the nanostructures, the foregoing methods are incapable of or too complicated for, in many cases, describing new extreme optical phenomena, in that the ENZ materials possibly take on large energy losses (like perfect absorption), high anisotropy [64–67], and abnormal interfaces [5,68–70]. A theoretical refinement is requisite in the face of these extreme situations. Because of the following reasons, we start with the standard model to further the FC theory by stages. First, it demonstrates the principle of FC in by far the clearest and easiest way. The foremost concept in nonlinear optics, PM, is explicitly derived from the simple solution of a CAE [52], which also holds in a quite small-sized material. The CAE is quite easy to use for any pumping strength since it presents a propagating FC process with simple mathematics: the first-order differential equation. The standard model is the best candidate for representing dynamic behaviors in a FC process and for visualizing wave interaction and transmission.

At the microscopic scale, some phase-mismatched processes become noticeable for the final FC as well. In a thin film, the FC is a bidirectional process and can be divided into a forward frequency conversion (FFC) and a backward frequency conversion (BFC) [1,71–73]. The traditional CAE does not include the BFC due to the well-known slowly varying amplitude approximation (SVAA), and so cannot be directly applied in wavelength-level materials. The BFC, obscured by the SVAA, plays an important role in nonlinear materials in the range of less than one wavelength. Hence, an exact application of the CAE should take in the BFC. Recently, it has been implemented for the third harmonic generation in layer structures by an intuitive extension [72,73]. Another limitation of the standard model is the normal incidence, as this is the very reason it has to be abandoned in oblique harmonic generation [74]. Oblique incidence is required for generality and practice; e.g., a field-enhanced mode in an ENZ metamaterial is near grazing incidence inside [41]. Finally, the transmission of frequency converted waves is subjected to strict nonlinear boundary conditions (BCs). The strict BCs take account of all fields induced by the nonlinearity on the boundaries. They were first considered by Bloembergen *et al.*, who investigated the transmission of harmonic waves [54], i.e., the reflection and refraction through an interface. But now in most cases, they are directly replaced by the simple BCs confined to the free electromagnetic waves. The legitimacy of the *a priori* application should be evaluated in current theories. The problems addressed above have to been considered in a different model, and so become the central issues in this paper.

We therefore in this paper scrutinize the FC in a nonlinear medium and present an application of SHG in a simplified ENZ waveguide. The transmission modes in bidirectional FC are analyzed in Sec. II, where the transition-layer assumption of ignoring the nonlinear source at the interfaces is given as well. A set of CAEs is derived in Sec. III by the dyadic Green's function method, exhibiting a form similar to the original CAE. Subsequent discussions are conducted on the supplementary BCs and the nonradiative (NR) field. Importantly, the CAEs are extended to include a general permittivity tensor in Sec. IV. The SHG in a nonlinear planar ENZ waveguide, formed by silicon slabs, is expounded under the improved CAEs in Sec. V, where the harmonic wave, as a lowest guided mode, is selected at the ENZ wavelength. It is found that the SHG possesses a noncollinear PM in a nearly counterpropagating geometry. A simulation in PM configuration shows that the conversion efficiency can be enhanced from the third to the fifth power of the relative wave impedance in the weak absorption case, where, however, only a square enhancement exists for a wavelength-thick nonlinear ENZ material without PM.

II. THEORY PREPARATION

A. Nonlinear wave equations

Nonlinear optics was established by introducing a nonlinear polarization vector $\mathbf{P}^{\mathrm{NL}}$ (radiating source), which can be expanded as a power series of electric field [1]. It is a driving source to activate various radiations and excitations. The FC is a radiation process controlled by the nonlinear polarization and yields electromagnetic waves at the same frequency. We now consider a linearly polarized nonlinear source inside a homogeneous film, ignoring the optical activity effect, whose strength oscillates in accordance with a wave vector $\mathbf{k}_{\mathrm{P}}$. The Maxwell's equations of electromagnetic fields with the angular frequency ω are

$$\nabla \times \mathbf{E} = i\omega \mathbf{B}, \tag{1a}$$

$$\nabla \times \mathbf{H} = -i\omega(\mathbf{D}^{\mathrm{L}} + \mathbf{P}^{\mathrm{NL}}), \tag{1b}$$

$$\nabla \cdot \mathbf{D}^{\mathrm{L}} = -\nabla \cdot \mathbf{P}^{\mathrm{NL}}, \tag{1c}$$

$$\nabla \cdot \mathbf{B} = 0, \tag{1d}$$

where $\mathbf{E}$, $\mathbf{D}$, $\mathbf{H}$, and $\mathbf{B}$ are the electric field, electric displacement, magnetic vector, and magnetic induction, respectively. In Eqs. (1), a definition $\mathbf{D} \equiv \mathbf{D}^{\mathrm{L}} + \mathbf{P}^{\mathrm{NL}}$ is used where the superscript L denotes the linear part of the electric displacement and NL denotes the nonlinear polarization. The term about current density is simply folded into the linear electric displacement. Compared to the typical Maxwell's equations in a conducting medium, the term $-i\omega\mathbf{P}^{\mathrm{NL}}$ can be regarded as a nonlinear polarization current density $\mathbf{J}$, and at the same time $-\nabla \cdot \mathbf{P}^{\mathrm{NL}}$ is a nonlinear polarization charge density ρ_{C}. The continuity condition $\nabla \cdot \mathbf{J} + \partial_t \rho_{\mathrm{C}} = 0$ is naturally satisfied. So the film is a special conducting material carrying prescribed current and charge distributions. Therefore, solving Eqs. (1), in some sense, is a radiation problem.

Through a direct derivation from Maxwell's Eqs. (1) [75], the electrical field $\mathbf{E}$ obeys the equations

$$(\nabla \times \nabla \times \mathbf{E}) - k_0^2 \boldsymbol{\epsilon} \cdot \mathbf{E} = \frac{k_0^2}{\varepsilon_0} \mathbf{P}^{\mathrm{NL}}, \tag{2a}$$

$$\varepsilon_0 \nabla \cdot (\boldsymbol{\epsilon} \cdot \mathbf{E}) = -\nabla \cdot \mathbf{P}^{\mathrm{NL}}, \tag{2b}$$

where ε_0 is the vacuum permittivity constant, $\boldsymbol{\epsilon}$ is the permittivity tensor, c is the light speed in vacuum, and a convenient definition $k_0^2 \equiv \mu\omega^2/c^2$ is used with the relative permeability μ, a constant. A complex-valued extension of $\boldsymbol{\epsilon}$, of course, predicts ohmic loss in a metallic medium, as will be considered in the next section. Equation (2b) just is

Eq. (1c) or the divergence of Eq. (2a), revealing the origin of polarization charges from the irrotational part of the nonlinear polarization. The electric field $\mathbf{E}$ is commonly expressed as a radiative field $\mathbf{E}_{\mathrm{R}}$ (electromagnetic wave) adding a NR field $\mathbf{E}_{\mathrm{NR}}$. Field description will abide by the complex analytic signal representation, where a real-valued field strength can be represented as $2\mathrm{Re}(\mathbf{E}) = \mathbf{E}$+c.c. Besides, the time factor $\exp(-i\omega t)$ is dropped as a convention which will be taken throughout the paper.

It is suggested that if a nonlinear magnetization is considered, a nonlinear term $i\mu_0\omega\nabla \times \mathbf{M}^{\mathrm{NL}}$ will appear on the right-hand side of Eq. (2a), where μ_0 is the vacuum permeability and $\mathbf{M}^{\mathrm{NL}}$ is the nonlinear magnetization vector. Essentially the extra term does not change the mathematics in Eqs. (2), so the ensuing treatment is extendable.

B. Eigenmodes of forward and backward transmission modes

Without losing generality, the medium can be taken as an anisotropic crystalline slab with a thickness l and two smooth parallel surfaces. In a situation of general oblique propagating, field symmetry strongly constrains the propagating directions of free waves [59]. It is a consequence of light interference and propagation. The following results are suitable for uniaxial and biaxial media without a directional limitation on optic axes. However, we still stipulate a simple situation for a better illustration. In this way, the film is supposed to be a uniaxial crystal-like medium with the optic axis being in the plane of incidence, shown in Fig. 1(a) where an acute angle θ_0 is made by $\mathbf{k}_{\mathrm{P}}$ with the z direction.

It is appropriate to accept that the film system (including the shape of the film, the nonlinear source, and the initial and boundary conditions) possesses a space translational symmetry in the x direction [1]. Therefore, the FC, as a part of the system, has the same symmetry and must own a wave vector whose x component is $k_x = k_{\mathrm{P}} \cos\theta_0$ for both the forward and backward branches, where the "forward" indicates a propagating direction close to $\mathbf{k}_{\mathrm{P}}$ and the "backward" represents one in a reflected direction. It is well known in crystal optics that there are two kinds of permitted eigenmodes: One mode is the extraordinary light having a refractive index related to its propagating direction and having a walk-off effect, while the other is the ordinary light having a fixed refractive index. Considering these transmission modes, there are two sets of wave vectors $(\mathbf{k}_{\mathrm{F1}}, \mathbf{k}_{\mathrm{F2}})$ and $(\mathbf{k}_{\mathrm{B1}}, \mathbf{k}_{\mathrm{B2}})$, where the subscript 1 is for the transverse magnetic (TM) or extraordinary light and 2 for the transverse electric (TE) or ordinary with polarizations perpendicular to each other indicated by $\hat{\mathbf{a}}_{\mathrm{F1}}$ ($\hat{\mathbf{a}}_{\mathrm{F2}}$) and $\hat{\mathbf{a}}_{\mathrm{B1}}$ ($\hat{\mathbf{a}}_{\mathrm{B2}}$), in which F and B denote FFC and BFC, respectively. In this case, the backward wave vector $\mathbf{k}_{\mathrm{B}}$, not like the case of normal incidence, is not the inversion of the forward one $\mathbf{k}_{\mathrm{F}}$, since they have the same wave-vector component k_x. The Huygens principle in optics helps us to understand the paired orientation. It regards every point in a slice located at an arbitrary position ($z = z'$) as a secondary radiating source [75], emitting a spherical wave, as sketched in the middle of Fig. 1(a). Hence, the wave fronts formed by the envelopes suggest the forward- and backward-propagating directions. The above results are for a source with a fixed wave vector. When there are several nonlinear sources with

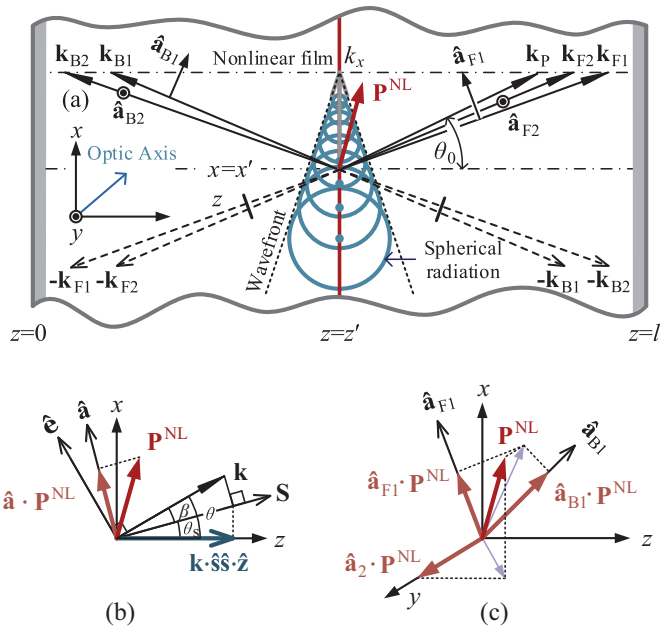

FIG. 1. Schematic of the eigenmodes of the electromagnetic waves generated in a FC process, driven by a planar sheet nonlinear source at z' with a fixed wave vector $\mathbf{k}_{\mathrm{P}}$, where the forward and backward wave vectors of the four possible modes are illustrated (a) with seriate spherical wavefronts radiated at the sampled secondary source points. The geometries of the transmission modes in the plane of incidence (b) and of the modes (c) in three-dimensional (3D) coordinates are shown, where $\hat{\mathbf{a}}_2$ for $\hat{\mathbf{a}}_{\mathrm{F2}}$ or $\hat{\mathbf{a}}_{\mathrm{B2}}$, $\mathbf{S}$ for a Poynting vector.

wave vectors having different k_x, more eigenmodes will be involved, as is possible for multiple reflection inside or for an artificially patterned structure [5,69], where the engineered surface permits numerous spatial frequencies of different orders.

The polarization is taken as the total electric moment per unit volume, e.g., in a dielectric material, which comes from massive molecule dipoles. So a molecule dipole, as the basic unit of polarization source, radiates electromagnetic waves in all directions. Numerous aligned molecules radiate as an antenna array does. The wave vectors thereby exist in all directions. In this sense, a superposition of various electromagnetic radiations from a phased array of dipoles enables an effective FC, which only takes place in some directions of interference enhancement, such as $\mathbf{k}_{\mathrm{F}}$, $\mathbf{k}_{\mathrm{B}}$ in our case.

C. Transition layer assumption

As the boundaries are concerned, there is a surface nonlinearity at the interface due to a breaking of spatial symmetry of nonlinearity in the normal direction and a discontinuity of field distribution [3,58]. The surface nonlinearity can be added into bulk terms and will not be considered separately.

The strict BCs are inconvenient to apply for differential and integral equations since they relate to all nonlinear fields. The boundary issue, nevertheless, can be circumvented by a transitional process without incurring an appreciable influence. As Fig. 2 shows, we assume there is an extremely

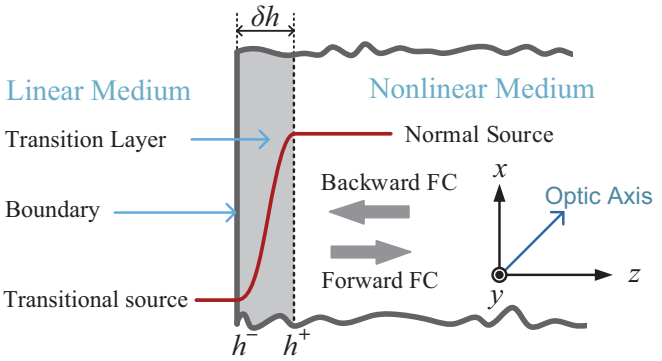

FIG. 2. Transition layer of an infinitely small thickness δh in which the nonlinear polarization grows quickly to the normal from a linear medium to a nonlinear medium. One similar transition layer is on the other side, where the nonlinear polarization goes quickly down to zero.

thin transition layer on each surface, where the nonlinear susceptibility rises (falls) quickly from zero (the normal) and reaches the normal (zero) when light goes past the layer. The layers will then not affect the reflection and refraction of free waves.

At the beginning of the transition layer $z = 0^-$ or l^+, there will not be any FC process. The layer thickness δh is far less than l so it will not bring a detectable disturbance in the entire FC process and we can formally let $\delta h \to 0$ to gradually remove the influence of the transition layers. Since the FC is an accumulated radiating process, it will be described as an integration over all sheet sources along the propagating direction. The possible errors, relying on the variation of radiant flux generated from the transition region, decrease with the layer size. So, the transition layer assumption is expected to predict an exact result when close to the limitation. In this way, an effective field point is confined in the interior sandwiched by the two transition layers, which removes the influence of interface effect and makes the BCs easy to apply. In summary, we briefly recap the transition layer assumption as three points: (i) the derivative of the amplitude $\mathcal{A}$ of a radiative field and induced NR fields are zero:

$$\frac{\partial \mathcal{A}}{\partial z} = 0, \quad \mathbf{E}_{\mathrm{NR}} = \mathbf{H}_{\mathrm{NR}} = 0, \tag{3}$$

at $z = 0$ and l, which are the supplementary BCs; (ii) a field point is within the slab interior $0^+ < z < l^-$; and (iii) the radiative fields can be taken as free waves on the boundaries, as a deduction of point (i).

Actually, the assumption seems more like an equivalent relation than an approximative condition, so it gives rise to exact solutions of a FC, based on the supplementary BCs. This assumption was directly used by Shen to explain BFC [1] and was implicitly applied in standard [52] and integral models [59] by accepting assumption point (iii). The validity of the assumption is associated with the resulting NR field and will be discussed in Sec. III C.

III. MODELING

A. Dyadic Green's function

We now consider the oblique propagation of infinite plane waves as that in Sec. II B and simply let the optic axis of the uniaxial film be in the plane of incidence, the x-z plane. Variations of fields will then be limited in the (x, z) domain. For convenience, all hatted letters below denote unit vectors associated with some related physical quantities in use.

The formulation below will adhere to the dyadic Green's function method from Eq. (2a), which will demonstrate an intuitive physics from its clear analytic representation. In terms of Eq. (2a), the involved Green's function $\boldsymbol{G}$ can be written as [76]

$$(\nabla \times \nabla \times \boldsymbol{G}) - k_0^2 \boldsymbol{\epsilon} \cdot \boldsymbol{G} = \boldsymbol{I} \exp[ik_x(x - x')]\delta(z - z'), \tag{4}$$

where $\boldsymbol{G} = \boldsymbol{G}(\mathbf{r}, \mathbf{r}')$ with $\mathbf{r} = x\hat{\mathbf{x}} + z\hat{\mathbf{z}}$ and $\mathbf{r}' = x'\hat{\mathbf{x}} + z'\hat{\mathbf{z}}$ represents a particular dyadic field at the field position $\mathbf{r}$, generated at the source position $\mathbf{r}'$ (see Fig. 1), $\delta(\cdot)$ is the Dirac function. $\boldsymbol{G}$ commonly includes all the wave components radiated by the highly localized planar source at z' multiplied by the unit dyadic (matrix) $\boldsymbol{I}$, whose element G_{ij} denotes the ith component of the radiative field produced by a unit nonlinear source polarized along the jth coordinate axis. Out of the exchange symmetry of Green's function between the two points $\mathbf{r}$ and $\mathbf{r}'$, it must contain the phasor $\exp[ik_x(x - x')]$, which is the phase distribution on the source plane. For the orientation of the optic axis in the given reference system, $\boldsymbol{\epsilon}$ has a simple matrix form

$$\begin{bmatrix} \epsilon_{11} & 0 & \epsilon_{13} \\ 0 & \epsilon_{22} & 0 \\ \epsilon_{31} & 0 & \epsilon_{33} \end{bmatrix}, \tag{5}$$

where the subscripts 1, 2, 3 correspond to x, y, z coordinates, respectively. Since the permittivity tensor must be symmetric, we have $\epsilon_{13} = \epsilon_{31}$. To convert Eq. (4) into a matrix equation, we first use the identity

$$\nabla \times \nabla \times \boldsymbol{G} = (\nabla\nabla - \nabla^2 \boldsymbol{I}) \cdot \boldsymbol{G} \tag{6}$$

to transfer the operations of cross product into ones of dot product. Second, the operator on the left-hand side of Eq. (4) should be organized as a symmetrical matrix

$$\begin{bmatrix} -\frac{\partial^2}{\partial z^2} - k_0^2\epsilon_{11} & 0 & ik_x\frac{\partial}{\partial z} - k_0^2\epsilon_{13} \\ 0 & k_x^2 - \frac{\partial^2}{\partial z^2} - k_0^2\epsilon_{22} & 0 \\ ik_x\frac{\partial}{\partial z} - k_0^2\epsilon_{31} & 0 & k_x^2 - k_0^2\epsilon_{33} \end{bmatrix} \tag{7}$$

by use of matrix (5) and identity (6). Here, the replacement $\partial/\partial x \to ik_x$ is performed due to the mentioned phasor. From now on, the phasor will be dropped out in G_{ij} for a concise manipulation except in its dyadic representation. Finally, a matrix equation results from Eq. (4) by insertion of matrix (5). It is found that the TE component in $\boldsymbol{G}$ with the polarization normal to the paper is independent of other components. That means the dyadic Green's function can be decomposed into a direct sum of G_{22} and a 2×2 submatrix $\boldsymbol{G}_{\mathrm{t}}$ belonging to the (x,z) subspace, i.e., $\boldsymbol{G} = G_{22} \oplus \boldsymbol{G}_{\mathrm{t}}$ ($G_{22}\hat{\mathbf{y}}\hat{\mathbf{y}} + \boldsymbol{G}_{\mathrm{t}}$ in a dyadic form), ensured by the similar decomposition $\boldsymbol{\epsilon} = \epsilon_{22} \oplus \boldsymbol{\epsilon}_{\mathrm{t}}$. From

matrix (7), G_{22} satisfies

$$\left(\frac{\partial^2}{\partial z^2} - k_x^2 + k_2^2\right) G_{22} = -\delta(z - z'). \tag{8}$$

In this case, the forward and backward waves have the same wave constant $k_{F2} = k_{B2} = k_2$. Just like that in an isotropic medium, a planar source at z' will radiate symmetrically and the electromagnetic waves radiate forward and backward with the same amplitude. Via a standard method, the solution of Eq. (8) can be obtained by a combination of homogeneous solutions, i.e.,

$$G_{22}(z,z') = \frac{i}{2k_{2z}}\{\Theta(z - z')\exp[ik_{2z}(z - z')] + \Theta(z' - z)\exp[-ik_{2z}(z-z')]\}, \tag{9}$$

where $k_{2z} = \mathbf{k}_2 \cdot \hat{\mathbf{z}}$, $k_0^2\epsilon_{22} = k_2^2$, and Θ is the Heaviside function. Note that the bidirectional radiative fields possess the same wave-vector component k_x. For $\boldsymbol{G}_t$, Eq. (4) gives rise to two TM modes propagating along $\mathbf{k}_{F1}$ and $\mathbf{k}_{B1}$ illustrated by the sketch in Fig. 1. Combining matrix (7) and Eq. (4) reduces into Eq. (8) plus a 2×2 matrix equation of $\boldsymbol{G}_t$ with a form

$$\begin{bmatrix} \frac{\partial^2}{\partial z^2} + k_0^2\epsilon_{11} & k_0^2\epsilon_{13} - ik_x\frac{\partial}{\partial z} \\ k_0^2\epsilon_{31} - ik_x\frac{\partial}{\partial z} & k_0^2\epsilon_{33} - k_x^2 \end{bmatrix} \begin{bmatrix} G_{11} & G_{13} \\ G_{31} & G_{33} \end{bmatrix} = -\delta(z - z') \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{10}$$

On equating each matrix element in the two sides of Eq. (10), we get four differential equations. They are two homogeneous equations from the off-diagonal elements

$$\left(k_0^2\epsilon_{31} - ik_x\frac{\partial}{\partial z}\right) G_{11} + \left(k_0^2\epsilon_{33} - k_x^2\right) G_{31} = 0, \tag{11a}$$

$$\left(\frac{\partial^2}{\partial z^2} + k_0^2\epsilon_{11}\right) G_{13} + \left(k_0^2\epsilon_{13} - ik_x\frac{\partial}{\partial z}\right) G_{33} = 0, \tag{11b}$$

and two nonhomogeneous ones from the diagonal elements

$$\left(ik_x\frac{\partial}{\partial z} - k_0^2\epsilon_{31}\right) G_{13} + \left(k_x^2 - k_0^2\epsilon_{33}\right) G_{33} = \delta(z - z'), \tag{12a}$$

$$\left(ik_x\frac{\partial}{\partial z} - k_0^2\epsilon_{13}\right) G_{31} - \left(\frac{\partial^2}{\partial z^2} + k_0^2\epsilon_{11}\right) G_{11} = \delta(z - z'). \tag{12b}$$

Since $G_{31} = G_{13}$ due to the symmetry of Eq. (10), only three of them are independent. It is easy to find that if G_{11} is known, then G_{13} and G_{33} can be determined step by step. Eliminating the off-diagonal components by grouping Eqs. (11a) and (12b) leads to

$$\left[\epsilon_{33}\frac{\partial^2}{\partial z^2} + 2ik_x\epsilon_{13}\frac{\partial}{\partial z} + k_0^2\det(\boldsymbol{\epsilon}) - \epsilon_{11}k_x^2\right] G_{11} = -\xi_0\delta(z - z') \tag{13}$$

with $\xi_0 = \epsilon_{33}k_0^2 - k_x^2$. The homogeneous solutions are of bidirectional electromagnetic waves with the wave vectors $\mathbf{k}_F$ and $\mathbf{k}_B$, where the subscript 1 denoting the first eigenmode is omitted for concision and the longitudinal components of the two wave vectors are $k_{Fz} = \mathbf{k}_F \cdot \hat{\mathbf{z}}$ and $k_{Bz} = -\mathbf{k}_B \cdot \hat{\mathbf{z}}$, respectively. By use of the relation $\partial/\partial z \to ik_{Fz}$ or $-ik_{Bz}$, the resulting wave constant is

$$k_{Fz/Bz} = \epsilon_{33}^{-1}[\sqrt{\xi_0\det(\boldsymbol{\epsilon}_t)} \mp k_x\epsilon_{13}] \tag{14}$$

with "−" for F and "+" for B, as can alternatively be achieved by the Fresnel equation in crystal optics. As such, a related polarization direction $\hat{\mathbf{a}}_{F/B}$ can be determined. The nonhomogeneous solution of Eq. (13) represents the x component of the electric field radiated by the same x component of the planar source at z'. For the component of source along $\hat{\mathbf{x}}$, the source plane can be taken as an interface with zero thickness. Thus, the continuity condition of the in-plane ($\hat{\mathbf{x}}$) field across the interface requires an identical radiative field component for bidirectional FC. Hence, we can introduce a tentative form similar to solution (9)

$$G_{11} = G_0[\mathrm{Ph}_F(z,z') + \mathrm{Ph}_B(z,z')], \tag{15}$$

where $\mathrm{Ph}_n(z,z') \equiv \Theta[\pm(z - z')]\exp[\pm ik_{nz}(z - z')]$ represents a unidirectional wave with $n = F$ in the "+" case or B in the "−" case, and the constant G_0 will be determined by a direct algebraic calculation when putting the ansatz into Eq. (13). At last, we have

$$G_0 = \frac{i\xi_0}{k_0^2\epsilon_{33}(k_{Fz} + k_{Bz})} = i\frac{\hat{\mathbf{a}}_F \cdot \hat{\mathbf{x}}}{2\mathbf{k}_F \cdot \hat{\mathbf{s}}_F} = i\frac{\hat{\mathbf{a}}_B \cdot \hat{\mathbf{x}}}{2\mathbf{k}_B \cdot \hat{\mathbf{s}}_B} \tag{16}$$

with the help of Eq. (14) and Eqs. (A6), (A8), and (A9) in Appendix A. Here, $\hat{\mathbf{s}}$ indicates the unit vector of the time-averaged complex Poynting vector $\mathbf{S} = \mathbf{E} \times \mathbf{H}^*$, perpendicular to the polarization direction $\hat{\mathbf{a}}_{F/B}$ according to the definition of energy flux density. Evidently, only the eigenmodes with a polarization direction having a nonzero x component can be generated. Finally, the whole solutions are as follows:

$$G_{11} = \frac{1}{2}\sum_{n=F,B}\frac{\pm i(\hat{\mathbf{a}}_n \cdot \hat{\mathbf{x}})^2}{(\mathbf{k} \cdot \hat{\mathbf{s}}\hat{\mathbf{s}})_n \cdot \hat{\mathbf{z}}}\mathrm{Ph}_n(z,z'), \tag{17a}$$

$$G_{13} = \frac{1}{2}\sum_{n=F,B}\frac{\pm i\hat{\mathbf{a}}_n \cdot \hat{\mathbf{x}}\hat{\mathbf{a}}_n \cdot \hat{\mathbf{z}}}{(\mathbf{k} \cdot \hat{\mathbf{s}}\hat{\mathbf{s}})_n \cdot \hat{\mathbf{z}}}\mathrm{Ph}_n(z,z'), \tag{17b}$$

$$G_{33} = \frac{1}{2}\sum_{n=F,B}\frac{\pm i(\hat{\mathbf{a}}_n \cdot \hat{\mathbf{z}})^2}{(\mathbf{k} \cdot \hat{\mathbf{s}}\hat{\mathbf{s}})_n \cdot \hat{\mathbf{z}}}\mathrm{Ph}_n(z,z') - \frac{\delta(z - z')}{k_0^2\epsilon_{33}}, \tag{17c}$$

where "+" for $n = F$ and "−" for $n = B$, and Eq. (A9) is used to form the Green's function fields. In the derivation process, the relations used in Eqs. (17) are regulated in Appendix A. Each component of the dyadic Green's function in Eqs. (17) demonstrates a clear bidirectional radiation process of the simultaneous FFC and BFC. The factor $(\mathbf{k} \cdot \hat{\mathbf{s}}\hat{\mathbf{s}})_n \cdot \hat{\mathbf{z}} = \mathbf{k}_n \cdot \hat{\mathbf{s}}_n\hat{\mathbf{s}}_n \cdot \hat{\mathbf{z}}$ in the denominators can be written as $\hat{\mathbf{a}}_n \times (\mathbf{k}_n \times \hat{\mathbf{a}}_n) \cdot \hat{\mathbf{z}}$, illustrating the energy flux density of an electromagnetic wave of a unit amplitude through the x-y plane. The singular distribution including $\delta(z - z')$ in G_{33} represents an induced local field, formed by an extremely thin dipole layer, the planar NR source. The nonlinear area polarization charges produce an electric field normal to the layer. The Green's function will

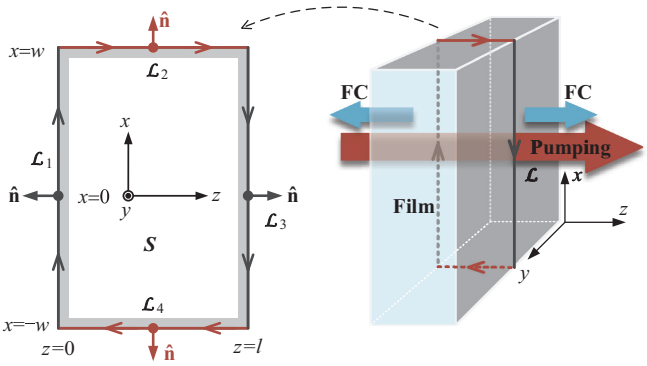

FIG. 3. Clockwise-directed curve $\mathcal{L}$ over the film section S, a rectangle, which is divided into four parts as $\mathcal{L}_1$ and $\mathcal{L}_3$ with a length $2w$ along the x direction, $\mathcal{L}_2$ and $\mathcal{L}_4$ with a length l along the z direction, respectively.

turn into the same expression as that of Sipe's formulation, $\overleftrightarrow{G}_{EP}$ in original Eq. (2.43) [59], when neglecting anisotropy.

B. Coupled-amplitude equations

To acquire a radiative field solution, the Green's integral formula should be carried out over an area through the film. Let us see a narrow rectangular area S as a section of the film, shown in Fig. 3, where S is bounded by a closed curve $\mathcal{L}$ with an outward unit vector $\hat{\mathbf{n}}$. Accordingly, the electric field $\mathbf{E}$ can be described as an area integral over the source space S and a close line integral around its boundary $\mathcal{L}$

$$2w\mathbf{E}(\mathbf{r}) = \frac{k_0^2}{\varepsilon_0}\int_S \boldsymbol{G}(\mathbf{r},\mathbf{r}')\cdot\mathbf{P}^{\mathrm{NL}}(\mathbf{r}')dS + \oint_{\mathcal{L}}(\boldsymbol{G}\times\nabla'\times\mathbf{E}' - \mathbf{E}'\times\nabla'\times\boldsymbol{G})\cdot\hat{\mathbf{n}}\,d\mathcal{L}, \tag{18}$$

where $\mathbf{E}' = \mathbf{E}(\mathbf{r}')$ and the primed derivative ∇' applies at the source point $\mathbf{r}'$. Equation (18) comes from the Green's vector identity

$$\int_S [\,\mathbf{E}'\cdot(\nabla'\times\nabla'\times\boldsymbol{G}) - \boldsymbol{G}\cdot(\nabla'\times\nabla'\times\mathbf{E}')\,]dS = \oint_{\mathcal{L}}(\boldsymbol{G}\times\nabla'\times\mathbf{E}' - \mathbf{E}'\times\nabla'\times\boldsymbol{G})\cdot\hat{\mathbf{n}}\,d\mathcal{L} \tag{19}$$

in terms of the exchange symmetry $\mathbf{E}'\cdot\boldsymbol{\epsilon}\cdot\boldsymbol{G} = \boldsymbol{G}\cdot\boldsymbol{\epsilon}\cdot\mathbf{E}'$. The closed line integral is evaluated along the curve $\mathcal{L}$, around a rectangle area S, in the outward direction denoted by the unit vector $\hat{\mathbf{n}}$, which is $-\hat{\mathbf{z}}$ at $z'=0$ and $\hat{\mathbf{z}}$ at $z'=l$. When pushing w to infinity, the integrals on the paths $\mathcal{L}_2$, $\mathcal{L}_4$ are small enough compared with those on the paths $\mathcal{L}_1$, $\mathcal{L}_3$, and so can be neglected in Eq. (19). In addition, the integrands are independent of the variable x', so the integrals on the two lines ($\mathcal{L}_1$ and $\mathcal{L}_3$) can be taken easily for an arbitrary vector function $\mathbf{F}$: $(2w)^{-1}\int_{-w}^{w}\mathbf{F}dx' = \mathbf{F}$, which is used in Eq. (18).

A similar case exists in the area integral over S. So, Eq. (18) can be further written as

$$\mathbf{E}(\mathbf{r}) = \frac{k_0^2}{\varepsilon_0}\int_0^l \boldsymbol{G}(\mathbf{r},\mathbf{r}')\cdot\mathbf{P}^{\mathrm{NL}}(\mathbf{r}')dz' + [(\nabla'\times\mathbf{E}')\cdot(\hat{\mathbf{n}}\times\boldsymbol{G}) - (\hat{\mathbf{n}}\times\mathbf{E}')\cdot(\nabla'\times\boldsymbol{G})]|_{z'=0}^{z'=l}. \tag{20}$$

Given all the transmission modes, independent of each other, the electric field and the nonlinear polarization can be factored as

$$\mathbf{E}(\mathbf{r}) = \sum_{n,m}\mathcal{A}_{nm}(z)\exp(i\mathbf{k}_{nm}\cdot\mathbf{r})\hat{\mathbf{a}}_{nm} + \mathbf{E}_{\mathrm{NR}}(\mathbf{r}), \tag{21a}$$

$$\mathbf{P}^{\mathrm{NL}}(\mathbf{r}) = \varepsilon_0\mathcal{P}^{\mathrm{NL}}(z)\exp(i\mathbf{k}_{\mathrm{P}}\cdot\mathbf{r})\hat{\mathbf{p}}, \tag{21b}$$

where $\mathbf{E}_{\mathrm{NR}}$ is the NR field, having the same phasor as the nonlinear polarization, n indicates F or B, $m = 1,2$ represent different polarization modes of the coupling waves. According to point (iii) in the transition layer assumption, we have $\nabla' \to \pm i\mathbf{k}_{nm}$. Hence, inserting Green's functions (17) into Eq. (20) yields

$$\mathcal{A}_{\mathrm{F}m}(z) = \mathcal{A}_{\mathrm{F}m}(0) + \frac{ik_0^2\hat{\mathbf{a}}_{\mathrm{F}m}\cdot\hat{\mathbf{p}}}{2(\mathbf{k}\cdot\hat{\mathbf{s}}\hat{\mathbf{s}})_{\mathrm{F}m}\cdot\hat{\mathbf{z}}}\times\int_0^z\mathcal{P}^{\mathrm{NL}}(z')\exp(i\Delta k_{\mathrm{F}m}z')dz', \tag{22a}$$

$$\mathcal{A}_{\mathrm{B}m}(z) = \mathcal{A}_{\mathrm{B}m}(l) - \frac{ik_0^2\hat{\mathbf{a}}_{\mathrm{B}m}\cdot\hat{\mathbf{p}}}{2(\mathbf{k}\cdot\hat{\mathbf{s}}\hat{\mathbf{s}})_{\mathrm{B}m}\cdot\hat{\mathbf{z}}}\times\int_z^l\mathcal{P}^{\mathrm{NL}}(z')\exp(i\Delta k_{\mathrm{B}m}z')dz' \tag{22b}$$

with the phase mismatch

$$\Delta k_{nm} \equiv (\mathbf{k}_{\mathrm{P}} - \mathbf{k}_{nm})\cdot\hat{\mathbf{z}}, \tag{23}$$

where supplementary BCs (3) and assumption point (ii) are imposed. The factor $(\mathbf{k}\cdot\hat{\mathbf{s}}\hat{\mathbf{s}})_{nm}\cdot\hat{\mathbf{z}}$ in the denominators, for ease of computation, can be written as $k_{nm}\cos\beta\cos\theta_{\mathrm{s}}$ [the angles β and θ_{s} are shown in Fig. 1(b)]. The initial amplitudes $\mathcal{A}_{\mathrm{F}m}(0)$ and $\mathcal{A}_{\mathrm{B}m}(l)$ will be determined by the BCs. On the ground of the physical implication of the Green's function, Eqs. (22) are, in essence, the superposition of dephased free electromagnetic waves. The integral equations, equivalently, correspond to the following differential equations:

$$\frac{d\mathcal{A}_{nm}}{dz} = \frac{ik_0^2\hat{\mathbf{a}}_n\cdot\hat{\mathbf{p}}}{2(\mathbf{k}\cdot\hat{\mathbf{s}}\hat{\mathbf{s}})_{nm}\cdot\hat{\mathbf{z}}}\mathcal{P}^{\mathrm{NL}}(z)\exp(i\Delta k_{nm}z). \tag{24}$$

The equations are able to describe a complete FC process instead, each of which formally has an expression similar to the original CAE. In practice, the nonlinear source may include a number of wave vectors, in the presence of diverse nonlinear coupling processes or multiple light pumping. Hence, the equations above can be immediately extended to a general form

$$\frac{d\mathcal{A}_{nm}}{dz} = \sum_{\mathbf{k}_{\mathrm{P}}}\frac{ik_0^2\hat{\mathbf{a}}_{nm}\cdot\hat{\mathbf{p}}(\mathbf{k}_{\mathrm{P}})}{2(\mathbf{k}\cdot\hat{\mathbf{s}}\hat{\mathbf{s}})_{nm}\cdot\hat{\mathbf{z}}}\mathcal{P}^{\mathrm{NL}}(z,\mathbf{k}_{\mathrm{P}})\times\exp[\,i\Delta k_{nm}(\mathbf{k}_{\mathrm{P}})z], \tag{25}$$

where all $\mathbf{k}_{\mathrm{P}}$s must have the mutual component k_x in the x direction. If there are different values of k_x, like noncollinear pumping and diffraction from structured interfaces [5,69], more CAEs belonging to different k_x should be built in an analogous way. It can be seen that there is an obvious obliquity effect, i.e., the PM takes place in the normal direction denoted by Eq. (23), and the obliquity factor $\hat{\mathbf{s}}\cdot\hat{\mathbf{z}}=\cos\theta_{\mathrm{s}}$ shows that an effective light interaction is along the direction of the energy flux, not that of the wave vector. For a specific FC process, the term in the numerator of the CAEs is commonly written as a well-known form in the electric dipole approximation

$$\begin{aligned}\hat{\mathbf{a}}\cdot\hat{\mathbf{p}}\mathcal{P}^{\mathrm{NL}}&=\hat{\mathbf{a}}\cdot\boldsymbol{\chi}^{(q)}(\omega=\textstyle\sum\omega_i)(\overset{q}{\cdot})\boldsymbol{\mathcal{B}}_1(\omega_1)\cdots\boldsymbol{\mathcal{B}}_q(\omega_q)D_q\\&=\hat{\mathbf{a}}\cdot\boldsymbol{\chi}^{(q)}(\overset{q}{\cdot})\hat{\mathbf{b}}_1\cdots\hat{\mathbf{b}}_q\mathcal{B}_1\cdots\mathcal{B}_qD_q\\&=\chi^{(q)}_{\mathrm{eff}}\mathcal{B}_1\cdots\mathcal{B}_qD_q\end{aligned}\tag{26}$$

for a qth-order nonlinear process, where $(\overset{q}{\cdot})$ denotes q times contractions or dot products, $\boldsymbol{\mathcal{B}}_i=\hat{\mathbf{b}}_i\mathcal{B}_i$ are the pumping fields in specified polarization directions, and D_q is the degeneracy factor. The calculation of the effective susceptibility should be taken in the crystallographic axis coordinate system, where the nonlinearity possesses a simplest form due to space symmetry. Thus a determination of the effective susceptibility $\chi^{(q)}_{\mathrm{eff}}$ is pivotal to a FC process. Equations (25) are rigorous, which will be discussed in the next subsection, and can be applied in any scale FC. The nonlinear magnetization can be considered in the equations without effort, and is omitted here. This is the central result in this paper, indicative of a generalization of the standard model to the oblique propagation without SVAA. In Sec. IV, it will be confirmed that Eq. (25) hold in a general case, e.g., a case in which the optic axis is not in the plane of incidence or the medium has an optical biaxial symmetry, where the triad $(\hat{\mathbf{k}},\ \hat{\mathbf{s}},\ \hat{\mathbf{z}})$ is possibly a noncoplanar orientation. Under the transition layer assumption, the transmission through boundaries with a pumping from outside can be readily handled through a known method, like the matrix optics.

The final equations are also suitable for lossy media by introducing a complex permittivity tensor, which gives rise to the complex wave vectors $\mathbf{k}_{\mathrm{P}}$ and $\mathbf{k}$. The phase distribution $\mathbf{k}\cdot\mathbf{r}$ includes an imaginary part proportional to z [75]. In this case, the constant amplitude surface, parallel to the interface, generally does not coincide with that of the constant phase front, so the complex wave vector $\mathbf{k}$ can be equivalently factored as a real part $\mathbf{k}_{\mathrm{R}}=\omega c^{-1}n_{\mathrm{R}}\hat{\mathbf{k}}_{\mathrm{R}}$ adding an imaginary component $\mathbf{k}_i=k_i\hat{\mathbf{z}}=\omega c^{-1}n_i\hat{\mathbf{z}}$, which, by $\mathbf{k}=\omega c^{-1}\tilde{\mathbf{n}}$, denotes a complex vector index of refraction

$$\begin{aligned}\tilde{\mathbf{n}}&\equiv n_x\hat{\mathbf{x}}+(n_z+in_i)\hat{\mathbf{z}}=n_x\hat{\mathbf{x}}+\tilde{n}_z\hat{\mathbf{z}}\\&=n_{\mathrm{R}}\hat{\mathbf{k}}_{\mathrm{R}}+in_i\hat{\mathbf{z}}.\end{aligned}\tag{27}$$

Hence, all development in this paper can straightforwardly be extended to lossy media with a complex index component $\tilde{n}_z$. For a light ray, k_i in fact is an effective coefficient of absorption (σ) in the z direction. The refractive indices in this case generally have a directional dependence on θ, as will be shown later.

There are remarkable features for the renewed model in the lossy media. The unit vectors $\hat{\mathbf{s}}$, $\hat{\mathbf{a}}$, defined in Appendix A, and $\hat{\mathbf{p}}$ become complex valued for the TM waves (similar in the TE case), which brings additional phase factors in fields. Therefore, a longitudinal component is present in the electromagnetic field, which no longer is an ordinary linear polarization. This peculiar polarization state is wholly represented by the complex unit vector $\hat{\mathbf{a}}$. Specifically, it is proportionate to $\tilde{n}_z\hat{\mathbf{x}}-n_x\hat{\mathbf{z}}$ due to $\tilde{\mathbf{n}}\cdot\hat{\mathbf{a}}=0$ in the isotropic approximation. So the electric field has a relation

$$\begin{aligned}\mathbf{E}\propto\ \hat{\mathbf{a}}&=a_x\hat{\mathbf{x}}+a_z\hat{\mathbf{z}}=(\tilde{n}_z\hat{\mathbf{x}}-n_x\hat{\mathbf{z}})(\tilde{\mathbf{n}}\cdot\tilde{\mathbf{n}})^{-1/2}\\&=[|\tilde{n}_z|e^{i\phi_x(\theta)}\hat{\mathbf{x}}+n_xe^{i\phi_z(\theta)}\hat{\mathbf{z}}]|\tilde{\mathbf{n}}\cdot\tilde{\mathbf{n}}|^{-1/2},\end{aligned}\tag{28}$$

which demonstrates that the fields in the x and z components carry different phases. For instance, the nonlinear source $\hat{\mathbf{a}}\cdot\mathbf{P}$ in a second-order nonlinear process leads to an effective susceptibility

$$\begin{aligned}\chi^{(2)}_{\mathrm{eff}}&=\hat{\mathbf{a}}\cdot\boldsymbol{\chi}^{(2)}(\omega=\omega_1+\omega_2):\hat{\mathbf{b}}_1\hat{\mathbf{b}}_2\\&=\sum_{i,j,k}a_i\chi^{(2)}_{ijk}b_{1j}b_{2k},\end{aligned}\tag{29}$$

where $\hat{\mathbf{b}}_{1/2}$ represents the polarization directions of two pumping fields. Expression (29) implies that the complex components from $\hat{\mathbf{a}}$ and $\hat{\mathbf{b}}$ will cause various interferences by superposition, which is more complicated than that of a loss-free medium. In addition, the factor $\mathbf{k}\cdot\hat{\mathbf{s}}\hat{\mathbf{s}}\cdot\hat{\mathbf{z}}$ reduces to $\omega c^{-1}\tilde{n}_z$ in the isotropic case. In an ENZ material, the complex refractive index $\tilde{n}_z$ will play an important role in the enhanced nonlinearity. Comparatively, the original model is only associated with a real refractive index n_{R} in a normal incidence case, the difference in which will be attributed to the weak absorption approximation. It should be noted that the amplitude in this case is not $\mathcal{A}$ but

$$A(z)\equiv\mathcal{A}(z)\exp(-k_iz),\tag{30}$$

according to a general definition. It can be proved that a real Poynting vector can be expressed as

$$\mathbf{S}_{\mathrm{R}}\equiv\mathbf{S}+\mathbf{S}^*=2n_{\mathrm{R}}(\mu_0\mu c)^{-1}|A|^2\hat{\mathbf{s}}_{\mathrm{R}},\tag{31}$$

where the real flux direction $\hat{\mathbf{s}}_{\mathrm{R}}\neq\hat{\mathbf{k}}_{\mathrm{R}}$, exhibiting the anisotropy in another way.

C. Nonradiative field and boundary conditions

The last term in solution (17c) gives rise to the rest part of the particular solution

$$\mathbf{E}_{\mathrm{NR}}(\mathbf{r})=-\frac{1}{\varepsilon_0\epsilon_{33}}\sum_{\mathbf{k}_{\mathrm{P}}}\mathbf{P}^{\mathrm{NL}}(z,\mathbf{k}_{\mathrm{P}})\cdot\hat{\mathbf{z}}\hat{\mathbf{z}}.\tag{32}$$

This NR field is along the normal direction to the film, independent of the direction of polarization. It can be taken as an inductive electric field in response to the nonlinear polarization current and/or charge density. From the point of view of energetics, the radiated energy is continuously transformed from the nonlinear polarization energy formed by a pumping process [1], whereas a fraction of the energy is stored in the medium to form a bound field. It is found from the Poynting vector that the stored polarization energy seems

to flow and to be guided in the film, carried by the rotational component of the NR field. In fact, it will be seen that the energy flux is canceled off by the growing radiative fields. One may not pay attention to the local field inside the film, but it has to be considered in reflection and transmittance through the boundaries. However, equivalently in physics, we can use the transition layer assumption to reasonably wipe out the NR field. Hence, the BCs are still suitable for the radiative fields, but with supplementary condition (3). This conclusion will be justified below.

Two widely used field BCs are

$$\hat{\mathbf{z}} \times \delta\mathbf{E} = 0, \quad \hat{\mathbf{z}} \times \delta\mathbf{H} = 0, \tag{33}$$

where $\delta\mathbf{E}$ and $\delta\mathbf{H}$ are the differences of the electric fields and magnetic vectors across a boundary, respectively, which means the continuity of their tangential components. The radiative field can be taken as a superposition of free electromagnetic waves locally, so they satisfy the first boundary condition in Eqs. (33) automatically. There is no NR field outside, so it can be well understood that the continuity of the tangential component of radiative fields on the boundary makes the NR field survival only in the normal direction inside, so it is absent in the first equation of Eqs. (33). Since the boundary terms in Eq. (20) are a collection of forward and backward free waves and can be absorbed into $\mathcal{A}_{\mathrm{F}m}(0)$ and $\mathcal{A}_{\mathrm{B}m}(l)$, the general solution forms as Eqs. (22) still hold without assumption (3) and will be used in a subsequent analysis, for it does not lose a generality.

We can decompose the total field $\mathbf{E}$ into three kinds of electric fields as $\mathbf{E}_{\mathrm{F}m}$, $\mathbf{E}_{\mathrm{B}m}$, and $\mathbf{E}_{\mathrm{NR}}$ with a clear indication by the subscripts as usual. It is known that the magnetic vector $\mathbf{H}$ is a curl of the electric field, i.e., $(i\omega\mu_0\mu)^{-1}\nabla\times\mathbf{E}$, so the magnetic vector within the nonlinear medium relates to

$$\begin{aligned}\nabla\times\mathbf{E} = {} & i\sum_{n,m}\mathbf{k}_{nm}\times\mathbf{E}_{nm} + i\mathbf{k}_{\mathrm{P}}\times\mathbf{E}_{\mathrm{NR}} \\ & + \hat{\mathbf{z}}\times\sum_{n,m}\frac{\partial\mathcal{A}_{nm}}{\partial z}\exp(i\mathbf{k}_{nm}\cdot\mathbf{r})\hat{\mathbf{a}}_{nm}\end{aligned} \tag{34}$$

by a constant, where the relation $\nabla\times\mathbf{E}_{\mathrm{NR}} = i\mathbf{k}_{\mathrm{P}}\times\mathbf{E}_{\mathrm{NR}}$ is applied, as directly follows from Eq. (32). On the right-hand side of Eq. (34), the first Σ term represents the sum of the propagable magnetic field components of all transmission modes, the second term is the local (or bound) magnetic field induced by the local electric field, and the last Σ indicates another kind of local magnetic field inspired by the radiative fields. Next, we will prove via Eq. (25) that all the local magnetic fields with the wave vector $\mathbf{k}_{\mathrm{P}}$ vanish, i.e.,

$$i\mathbf{k}_{\mathrm{P}}\times\mathbf{E}_{\mathrm{NR}} + \hat{\mathbf{z}}\times\sum_{n,m}\frac{\partial\mathcal{A}_{nm}}{\partial z}\exp(i\mathbf{k}_{nm}\cdot\mathbf{r})\hat{\mathbf{a}}_{nm} = 0, \tag{35}$$

so Eq. (34) will become

$$\nabla\times\mathbf{E} = i\sum_{n,m}\mathbf{k}_{nm}\times\mathbf{E}_{nm}. \tag{36}$$

In such a way, the local field does contribute but does not appear in the magnetic vector formula which renders a sound property that the strict nonlinear BCs are formally irrelevant to the NR field. For the sake of brevity, the second subscript m for different modes are simply omitted until the end of this subsection.

The x component of Eq. (35), pertinent to the TE light polarized along the $\hat{\mathbf{y}}$ direction, turns into

$$\hat{\mathbf{p}}\cdot\left(\frac{\hat{\mathbf{a}}_{\mathrm{F}}\hat{\mathbf{a}}_{\mathrm{F}}}{\mathbf{k}_{\mathrm{F}}\cdot\hat{\mathbf{s}}_{\mathrm{F}}\hat{\mathbf{s}}_{\mathrm{F}}\cdot\hat{\mathbf{z}}} + \frac{\hat{\mathbf{a}}_{\mathrm{B}}\hat{\mathbf{a}}_{\mathrm{B}}}{\mathbf{k}_{\mathrm{B}}\cdot\hat{\mathbf{s}}_{\mathrm{B}}\hat{\mathbf{s}}_{\mathrm{B}}\cdot\hat{\mathbf{z}}}\right)\times\hat{\mathbf{z}} = 0. \tag{37}$$

Owing to the identities $\hat{\mathbf{a}}_{\mathrm{F}} = \hat{\mathbf{a}}_{\mathrm{B}}$, $\hat{\mathbf{k}}_{\mathrm{F/B}} = \hat{\mathbf{s}}_{\mathrm{F/B}}$, and $\mathbf{k}_{\mathrm{F}}\cdot\hat{\mathbf{z}} = -\mathbf{k}_{\mathrm{B}}\cdot\hat{\mathbf{z}}$, the validity of Eq. (37) is straightforward. The y component of Eq. (35) is about the TM light polarized in the x-z plane and gives rise to

$$\hat{\mathbf{p}}\cdot\left[k_0^2\left(\frac{\hat{\mathbf{a}}_{\mathrm{F}}}{2\mathbf{k}_{\mathrm{F}}\cdot\hat{\mathbf{s}}_{\mathrm{F}}} - \frac{\hat{\mathbf{a}}_{\mathrm{B}}}{2\mathbf{k}_{\mathrm{B}}\cdot\hat{\mathbf{s}}_{\mathrm{B}}}\right) + \frac{k_x}{\epsilon_{33}}\hat{\mathbf{z}}\right] = 0, \tag{38}$$

where the relations $|\hat{\mathbf{z}}\times\hat{\mathbf{a}}_{\mathrm{F}}| = \hat{\mathbf{s}}_{\mathrm{F}}\cdot\hat{\mathbf{z}}$, $|\hat{\mathbf{z}}\times\hat{\mathbf{a}}_{\mathrm{B}}| = -\hat{\mathbf{s}}_{\mathrm{B}}\cdot\hat{\mathbf{z}}$ are used. Substituting Eqs. (A6) and (A8) into Eq. (38) makes it hold for an arbitrary $\hat{\mathbf{p}}$, so the proof ends.

Equations (35) or (36) can be developed into two exact nonhomogeneous BCs at $z = 0, l$, suitable for integral solution (20). However, they are still difficult to apply in CAEs. In most cases, the radiative field solutions are concerned only, so we can make a finer simplification on magnetic fields to readily figure out the boundary problem without breaking condition (35). The cheapest way is to negate each term in Eq. (35) by setting the nonlinear source zero at the interfaces. It follows that the radiative fields propagate like free electromagnetic waves with constant amplitudes near the boundaries, which is just what the transition layer assumption amounts to. Hence, the simple BCs, a combination of conditions (33) and supplementary conditions (3), thoroughly solve the boundary issue. Equation (36) also shows the vanishing of the total local magnetic field and therefore an absence of local energy flux inside the film.

In another way without the transition layer assumption, by applying condition (36) to the boundaries, Eq. (20) generates the same solutions as Eqs. (22) and (32), where a reduction of the curl of the dyadic Green's function, a crucial step to derive Eqs. (22) from condition (36), is elaborated in Appendix B. This formulation further validates the transition layer assumption by the uniqueness theorem of time-varying electromagnetic field.

Certainly, conditions (3) differ from Eq. (35) in that the latter still demands a growing field amplitude but the former leads to a locally invariant one near the boundaries. However, assumptive conditions (3) are more convenient to use if we can make a little sacrifice of the exactness on surfaces. It should be mentioned that in the traditional standard model, an approximation is hidden behind the nonlinear polarization, in which

$$\mathbf{P}_{\mathrm{R}}^{\mathrm{NL}}(\omega) = \varepsilon_0\boldsymbol{\chi}^{(2)}:\mathbf{E}_{\mathrm{R}}(\omega_1)\mathbf{E}_{\mathrm{R}}(\omega_2) + \cdots, \quad \omega = \omega_1 \pm \omega_2,$$

is addressed to replace the well-known expansion of the nonlinear polarization

$$\begin{aligned}\mathbf{P}^{\mathrm{NL}}(\omega) = {} & \varepsilon_0\boldsymbol{\chi}^{(2)}:\mathbf{E}(\omega_1)\mathbf{E}(\omega_2) + \cdots \\ = {} & \varepsilon_0\boldsymbol{\chi}^{(2)}:[\mathbf{E}_{\mathrm{R}}(\omega_1)\mathbf{E}_{\mathrm{R}}(\omega_2) + \mathbf{E}_{\mathrm{NR}}(\omega_1)\mathbf{E}_{\mathrm{NR}}(\omega_2) \\ & + \mathbf{E}_{\mathrm{R}}(\omega_1)\mathbf{E}_{\mathrm{NR}}(\omega_2) + \mathbf{E}_{\mathrm{NR}}(\omega_1)\mathbf{E}_{\mathrm{R}}(\omega_2)] + \cdots.\end{aligned}$$

To be more exact, they are different since the last expression permits some nonlinear interactions associated with the NR fields. For the second-order $\chi^{(2)}$ process in nonlinear optics, the ordinary frequency converted radiative field is only related to the leading term including $\mathbf{E}_{\mathrm{R}}(\omega_1)\mathbf{E}_{\mathrm{R}}(\omega_2)$. However, there are other polarization sources proportional to, such as $\mathbf{E}_{\mathrm{NR}}(\omega_1)\mathbf{E}_{\mathrm{NR}}(\omega_2)$, inducing classic dipole electromagnetic radiation, $\mathbf{E}_{\mathrm{R}}(\omega_1)\mathbf{E}_{\mathrm{NR}}(\omega_2)$ and $\mathbf{E}_{\mathrm{NR}}(\omega_1)\mathbf{E}_{\mathrm{R}}(\omega_2)$, leading to a coupling between the two kinds of fields. These radiation processes together, requiring some special nonzero components of $\chi^{(2)}$, produce a real total radiative field $\mathbf{E}_{\mathrm{R}}(\omega)$. They will play a role in intense field physics or in nonlinear ENZ materials. In particular, for ENZ materials with $\epsilon_{33} \to 0$, $\mathbf{E}_{\mathrm{NR}}$ may produce a stronger interaction and broaden a nonlinear process by like producing cascaded harmonic generation.

IV. GREEN'S FUNCTION FOR A GENERAL PERMITTIVITY TENSOR

For a general permittivity tensor $\boldsymbol{\epsilon}$ without vanishing elements, it becomes complicated to handle the equation of the dyadic Green's function. However, the FC mechanism, i.e., the nonlinear source's bidirectional radiation of four transmission modes, does not change in essence. Here, we only discuss a physical generalization of the Green's function and the CAEs.

An arbitrary dyadic Green's function can be expanded as

$$\boldsymbol{G}(\mathbf{r},\mathbf{r}') = \boldsymbol{G}_{\mathrm{NR}}\delta(z-z')\exp[ik_x(x-x')] + \sum_{n,m}\boldsymbol{G}_{nm}\mathrm{Ph}_{nm}(\mathbf{r},\mathbf{r}'), \tag{39}$$

where the definition $\mathrm{Ph}_{nm}(\mathbf{r},\mathbf{r}') \equiv \Theta[\pm(z-z')]\exp[i\mathbf{k}_{nm}\cdot(\mathbf{r}-\mathbf{r}')]$ is the like of $\mathrm{Ph}_n(z,z')$ in Eq. (15), $\boldsymbol{G}_{nm}$ represents a single dyad being associated with the radiative field polarized along $\hat{\mathbf{a}}_{nm}$, and $\boldsymbol{G}_{\mathrm{NR}}$ is a dyad element of the NR field. $\hat{\mathbf{a}}_{nm}$ still denotes an eigenmode of four kinds of electromagnetic waves, which are commonly classified as "slow" and "fast" modes according to their refractive indices. In terms of the meaning of the Green's function, they satisfy

$$\boldsymbol{G}_{nm}\cdot\hat{\mathbf{p}} \propto \hat{\mathbf{a}}_{nm}, \quad \boldsymbol{G}_{\mathrm{NR}}\cdot\hat{\mathbf{p}} \propto \hat{\mathbf{z}} \tag{40}$$

for an arbitrary unit vector $\hat{\mathbf{p}}$, denoting the direction of the nonlinear polarization. The last relation in Eq. (40) holds because the NR field must be normal to the film surface, along the z axis, as a nonpropagating field. It can be proved that there are two kinds of unit vectors: $\hat{\mathbf{a}}_{nm}^{\perp 1}$ and $\hat{\mathbf{a}}_{nm}^{\perp 2}$, and they should be perpendicular to $\hat{\mathbf{a}}_{nm}$ and so to $\boldsymbol{G}_{nm}$, i.e.,

$$\hat{\mathbf{a}}_{nm}^{\perp 1,2}\cdot\hat{\mathbf{a}}_{nm} = 0, \quad \hat{\mathbf{a}}_{nm}^{\perp 1,2}\cdot\boldsymbol{G}_{nm} = 0. \tag{41}$$

In an analogous way, we have

$$\hat{\mathbf{x}}\cdot\boldsymbol{G}_{\mathrm{NR}} = 0, \quad \hat{\mathbf{y}}\cdot\boldsymbol{G}_{\mathrm{NR}} = 0. \tag{42}$$

All these lead to

$$\boldsymbol{G}_{nm} \propto \hat{\mathbf{a}}_{nm}\hat{\mathbf{a}}'_{nm}, \quad \boldsymbol{G}_{\mathrm{NR}} \propto \hat{\mathbf{z}}\hat{\mathbf{N}} \tag{43}$$

since the rank of $\boldsymbol{G}_{nm}$ and $\boldsymbol{G}_{\mathrm{NR}}$ is 1, where $\hat{\mathbf{a}}'_{nm}$ and $\hat{\mathbf{N}}$ are two unknown unit vectors, but are fixed by the symmetry of $\boldsymbol{G}$. Based on a 3×3 matrix equation, similar to Eq. (10), it can be found that the Green's function still is a symmetrical dyadic function for the forward ($z > z'$) and backward ($z < z'$) radiation and for the NR case ($z = z'$), so there is only one suggestion consistent to Eq. (B1) in Appendix B, i.e., $\hat{\mathbf{a}}'_{nm} = \hat{\mathbf{a}}_{nm}$, $\hat{\mathbf{N}} = \hat{\mathbf{z}}$. Consequently, Eq. (39) can be rewritten as

$$\boldsymbol{G}(\mathbf{r},\mathbf{r}') = \sum_{n,m}\mathcal{C}_{nm}\hat{\mathbf{a}}_{nm}\hat{\mathbf{a}}_{nm}\mathrm{Ph}_{nm}(\mathbf{r},\mathbf{r}') + \mathcal{C}_{\mathrm{NR}}\delta(z-z')\hat{\mathbf{z}}\hat{\mathbf{z}}\exp[ik_x(x-x')], \tag{44}$$

where $\mathbf{k}_{nm}$ and $\hat{\mathbf{a}}_{nm}$ can be determined in the crystal optics method and the coefficients $\mathcal{C}_{nm}$ and $\mathcal{C}_{\mathrm{NR}}$ need to be figured out. Certainly, a formal solution (44) must obey Eqs. (4) and (20). By inserting Green's function (44) into Eq. (20), we obtain

$$\begin{aligned}\mathbf{E}(\mathbf{r}) = k_0^2\sum_{m=1,2}\Bigg[&\mathcal{C}_{\mathrm{F}m}\int_0^z\mathcal{P}^{\mathrm{NL}}(z')\exp(i\Delta k_{\mathrm{F}m}z')dz' \\ &\times\ \exp(i\mathbf{k}_{\mathrm{F}m}\cdot\mathbf{r})\hat{\mathbf{a}}_{\mathrm{F}m}\hat{\mathbf{a}}_{\mathrm{F}m}\cdot\hat{\mathbf{p}} \\ &+\ \exp(i\mathbf{k}_{\mathrm{B}m}\cdot\mathbf{r})\hat{\mathbf{a}}_{\mathrm{B}m}\hat{\mathbf{a}}_{\mathrm{B}m}\cdot\hat{\mathbf{p}} \\ &\times\ \mathcal{C}_{\mathrm{B}m}\int_z^l\mathcal{P}^{\mathrm{NL}}(z')\exp(i\Delta k_{\mathrm{B}m}z')dz'\Bigg] \\ &+\ \mathcal{C}_{\mathrm{NR}}\mathbf{P}^{\mathrm{NL}}(\mathbf{r})\cdot\hat{\mathbf{z}}\hat{\mathbf{z}} + \mathbf{E}_{\mathrm{EM}}(\mathbf{r}).\end{aligned} \tag{45}$$

The terms in the square bracket of Eq. (45) are the radiative fields with varying amplitudes, characterizing the FFC and BFC processes. $\mathbf{E}_{\mathrm{EM}}$ comes from the boundary terms in Eq. (20), i.e.,

$$\mathbf{E}_{\mathrm{EM}}(\mathbf{r}) = [(\boldsymbol{\nabla}'\times\mathbf{E}')\cdot(\hat{\mathbf{n}}\times\boldsymbol{G}) - (\hat{\mathbf{n}}\times\mathbf{E}')\cdot(\boldsymbol{\nabla}'\times\boldsymbol{G})]|_{z'=0}^{z'=l}, \tag{46}$$

which is the propagating form of free electromagnetic waves. It can therefore be written as another form:

$$\mathbf{E}_{\mathrm{EM}}(\mathbf{r}) \equiv \sum_{n,m}\mathcal{A}'_{nm}\exp(i\mathbf{k}_{nm}\cdot\mathbf{r})\hat{\mathbf{a}}_{nm}. \tag{47}$$

From the expression of $\mathbf{E}_{\mathrm{EM}}$ and Eq. (45), it is found that $\mathcal{A}'_{nm}$ must be the radiative field amplitude on the boundaries. So, we have

$$\mathcal{A}'_{\mathrm{F}m} = \mathcal{A}_{\mathrm{F}m}(0), \quad \mathcal{A}'_{\mathrm{B}m} = \mathcal{A}_{\mathrm{B}m}(l). \tag{48}$$

However, there are some cross terms of the four transmission modes in Eq. (46). These cross terms imply that the variation of the electromagnetic wave of one eigenmode influences the propagation of the others. This should not happen since all the transmission modes are independent. Then we have to accept the following relation,

$$[\hat{\mathbf{a}}_{nm}\times(\mathbf{k}\times\hat{\mathbf{a}})_{n'm'} + \hat{\mathbf{a}}_{n'm'}\times(\mathbf{k}\times\hat{\mathbf{a}})_{nm}]\cdot\hat{\mathbf{z}} = 0, \tag{49}$$

to remove the cross terms, where at least $n \neq n'$ or $m \neq m'$. The verification of relation (49) is shown in Appendix C. Denoted by Eq. (C7), this condition is equivalent to the fact that the Poynting vectors of different transmission modes present no interference to each other in the normal direction. According to the transition layer assumption and the relation

$$[\hat{\mathbf{a}}\times(\mathbf{k}\times\hat{\mathbf{a}}) = \mathbf{k}\cdot\hat{\mathbf{s}}\hat{\mathbf{s}}]_{nm}, \tag{50}$$

we obtain

$$\mathcal{C}_{nm} = \pm i[2(\mathbf{k}\cdot\hat{\mathbf{s}}\hat{\mathbf{s}})_{nm}\cdot\hat{\mathbf{z}}]^{-1} \tag{51}$$

again through Eqs. (46) and (48), where "+" denotes the FFC and "−" denotes the BFC. Then, Eqs. (22) and (25) hold in a general case.

The NR field can be found without the transition layer assumption. The radiative fields propagate and reflect through the film interfaces as electromagnetic waves do. Besides, any free wave with the same magnitude and eigenmode as the radiative field must satisfy conditions (33), so we conclude that the dyadic Green's function does not produce any local magnetic field; otherwise, it will break the magnetic BC in (33). This yields

$$\hat{\mathbf{z}} \times \sum_{m=1,2} (\mathcal{C}_{\mathrm{F}m} \hat{\mathbf{a}}_{\mathrm{F}m} \hat{\mathbf{a}}_{\mathrm{F}m} - \mathcal{C}_{\mathrm{B}m} \hat{\mathbf{a}}_{\mathrm{B}m} \hat{\mathbf{a}}_{\mathrm{B}m}) - i k_x \mathcal{C}_{\mathrm{NR}} \hat{\mathbf{y}} \hat{\mathbf{z}} = 0. \tag{52}$$

Relation (52), as a generalized form of Eqs. (37) and (38), does not restrict itself on the boundaries. In fact, relation (52) is consistent to Eqs. (48) and also is a necessary condition in Eq. (4) to get rid of the derivative of the Dirac δ function $\delta'(z-z')$, where the details are omitted here. At the same time, identity (36) can result again by this relation.

Now, to acquire the coefficient $\mathcal{C}_{\mathrm{NR}}$, let us see the z component of Eq. (2a)

$$\varepsilon_0 \hat{\mathbf{z}} \cdot (\boldsymbol{\nabla} \times \boldsymbol{\nabla} \times \mathbf{E}) - \varepsilon_0 k_0^2 \hat{\mathbf{z}} \cdot \boldsymbol{\epsilon} \cdot \mathbf{E} = k_0^2 \hat{\mathbf{z}} \cdot \mathbf{P}^{\mathrm{NL}}, \tag{53}$$

where the curl of the magnetic field can be expressed as

$$\boldsymbol{\nabla} \times \boldsymbol{\nabla} \times \mathbf{E} = \sum_{n,m} \left(i \hat{\mathbf{z}} \frac{\partial}{\partial z} - \mathbf{k}_{nm} \right) \times (\mathbf{k} \times \hat{\mathbf{a}} \mathcal{A})_{nm} \tag{54}$$

by insertion of identity (36). Substituting Eqs. (54) and (C1) into Eq. (53), we get

$$-\varepsilon_0 \hat{\mathbf{z}} \cdot \boldsymbol{\epsilon} \cdot \mathbf{E}_{\mathrm{NR}} = \hat{\mathbf{z}} \cdot \mathbf{P}^{\mathrm{NL}}. \tag{55}$$

It is just Eq. (32) for a fixed $\mathbf{k}_{\mathrm{P}}$ and directly leads to $\mathcal{C}_{\mathrm{NR}} = -(k_0^2 \epsilon_{33})^{-1}$.

Green's function (44) can be taken as projection operators onto different transmission modes due to its symmetrical dyads. It makes Eq. (25) universal to deal with an arbitrarily oriented nonlinear source.

V. APPLICATION TO SHG IN A NONLINEAR ENZ WAVEGUIDE

A. Phase matching and waveguide amplification

Nonlinear materials exhibiting ENZ properties are in a position to stimulate nonlinearity, such as the second and third harmonic generations [26,27]. The strong field enhancement due to the ENZ mode was reported to account for a highly efficient conversion process. To estimate the conversion efficiency inside an ENZ material, we consider the SHG in a simple isotropic slab of thickness l. For convenience, we regularize the field intensity [Eq. (31)] as $I = 2|E|^2/(Z_0 Z_{\mathrm{r}})$, where $Z_0 = \sqrt{\mu_0/\varepsilon_0} \approx 377\Omega$ is the wave impedance in vacuum and $Z_{\mathrm{r}} \equiv \sqrt{\mu/\epsilon}$ with $\boldsymbol{\epsilon} = \epsilon \boldsymbol{I}$ is the relative impedance in the ENZ material. Simply excluding resonance, absorption, phase mismatch, and BFC in the slab, the harmonic amplitude, according to Eqs. (25), has an estimation of

$$\mathcal{A} \approx \frac{k_0^2 \chi_{\mathrm{eff}}}{2 k_2 \cos\theta} \mathcal{B}^2 l \approx \frac{\omega \chi_{\mathrm{eff}}}{4c \cos\theta} Z_0 Z_{\mathrm{r}1} Z_{\mathrm{r}2} I_1 l, \tag{56}$$

where the subscript "1" indicates the fundamental wave whose amplitude shows as $\mathcal{B}$, "2" indicates the harmonic wave, χ_{eff} is the effective second-order susceptibility as Eq. (29), and relations $k_2 = \sqrt{\mu\epsilon_2}$ and $Z_{\mathrm{r}1/2} = \sqrt{\mu/\varepsilon_{1/2}}$ are used with the definitions $\epsilon_1 \equiv \epsilon(\omega/2)$ and $\epsilon_2 \equiv \epsilon(\omega)$. As shown in Fig. 1(b), $\cos\theta$ is made by $\hat{\mathbf{k}}$ with respect to $\hat{\mathbf{z}}$. Hence, the total efficiency η_{tol} can be estimated as

$$\eta_{\mathrm{tol}} = I_2/I_0 \approx \frac{Z_0 \omega^2}{8c^2 \cos^2\theta} |\chi_{\mathrm{eff}}|^2 l^2 Z_{\mathrm{r}1}^2 Z_{\mathrm{r}2} I_0 \eta_1^2 \eta_{\mathrm{ext}} \tag{57}$$

with the injection efficiency $\eta_1 = I_1/I_0$ (I_0 is the pumping intensity) and the light extraction efficiency η_{ext}, as the energy efficiency of harmonic wave from inside to outside. If we situate the fundamental frequency at the ENZ mode, the high wave impedance $Z_{\mathrm{r}1}$ ($\epsilon_1 \approx 0$) will produce a FC enhancement. The high impedance commonly stops light injection except for the pseudo-Brewster (PB) geometry of a TM fundamental wave (see Ref. [29] and the references therein). At the special PB angle of incidence, the longitudinal field will be amplified according to the BC along the normal electric displacement vector. It can be called ENZ pump, which has been greatly attracted recently [26,27,42–44].

From Eq. (57), of interest is that the match between the harmonic wave and the ENZ mode also promises an enhancement of SHG, even though it only presents a linear growth with $Z_{\mathrm{r}2}$ ($\epsilon_2 \approx 0$), where we call it ENZ output for short. It can be seen in Eq. (56), the efficiency enhancement is due to the enlarged coupling coefficient, an extreme nonlinearity, as is different from the ENZ pump. So, this mode boosts SHG inside the material without discriminating between the TM and TE modes at the doubled frequency. In addition, we find that this kind of SHG owns a PM configuration by nearly counterpropagating pumping, a type of noncollinear PM, which is shown in Fig. 4. The forward fundamental wave $\mathbf{E}_{\mathrm{F1}}$ with $\mathbf{k}_{\mathrm{F1}}$ and the backward one $\mathbf{E}_{\mathrm{B1}}$ with $\mathbf{k}_{\mathrm{B1}}$ will produce three nonlinear polarizations with wave vectors $\mathbf{k}_{\mathrm{p}}^{\mathrm{F}} = 2\mathbf{k}_{\mathrm{F1}}$, $\mathbf{k}_{\mathrm{P}}^{\mathrm{B}} = 2\mathbf{k}_{\mathrm{B1}}$, and $\mathbf{k}_{\mathrm{P}}^{\mathrm{M}} = \mathbf{k}_{\mathrm{F1}} + \mathbf{k}_{\mathrm{B1}}$, through the nonlinear sources of $\boldsymbol{\chi} : \mathbf{E}_{\mathrm{F1}}\mathbf{E}_{\mathrm{F1}}$, $\boldsymbol{\chi} : \mathbf{E}_{\mathrm{B1}}\mathbf{E}_{\mathrm{B1}}$, and the mixing one $\boldsymbol{\chi} : \mathbf{E}_{\mathrm{F1}}\mathbf{E}_{\mathrm{B1}}$, respectively. Here, the mixing wave vector $\mathbf{k}_{\mathrm{P}}^{\mathrm{M}}$ is small due to vector cancellation. It can just be matched by one of the two transmission modes of the harmonic wave, e.g., the forward mode in Fig. 4. Besides, if a negative refractive index results at the pumping wavelength by a normal dispersion, the PM configuration will not be damaged. When $\epsilon_2 \to 0$, the counterpropagating pumping waves will generate nearly phase mismatch-free configuration; i.e., harmonic wave will grow in all possible directions. The special PM ability can get access to a practical conversion efficiency via bulk ENZ materials.

However, for a large impedance mismatch between the material and substrate or air, the TE mode harmonic field generally is hard to be led out without any special treatment, as is easy for the TM mode with a high transmittance at the PB angle of incidence. In addition, they may be guided out by some flux modulation techniques, like the tunneling effect [14]. The extraction of harmonic wave is so subtle

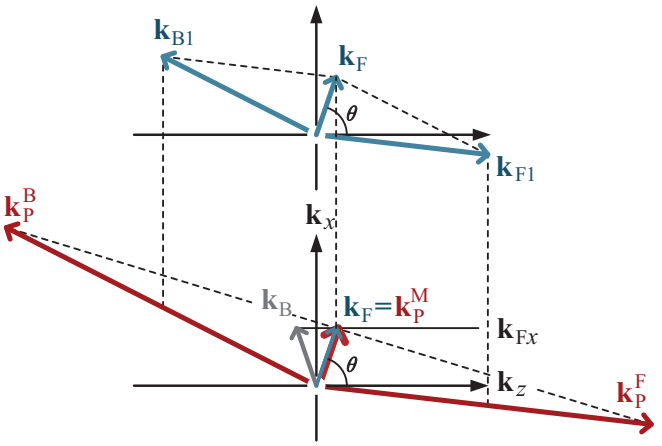

FIG. 4. Schematic of the PM geometry of SHG, shown in the upper diagram, where one harmonic wave vector of two transmission modes, denoted by $\mathbf{k}_F$, is phase matched by the wave vectors $\mathbf{k}_{F1}$ and $\mathbf{k}_{B1}$ from two pumping waves, respectively. The lower diagram demonstrates the three wave vectors of nonlinear polarizations, formed by the two pumping waves.

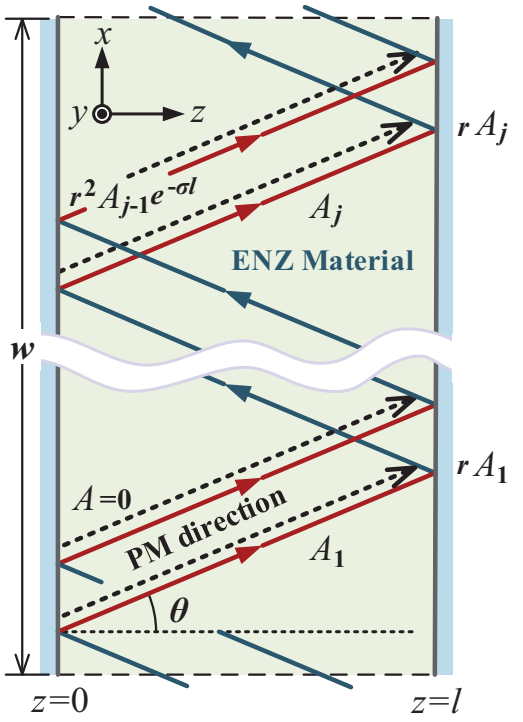

FIG. 5. Planar mirror waveguide with a size w for the SHG of the ENZ output, where the ENZ slab with a thickness $l \ll w$ is sandwiched by the substrate with high refractive index, and the harmonic wave grows in the PM direction but at the same time endures slight energy loss in absorption and transmittance.

and sophisticated that should be examined elsewhere. One interesting case, with which we will deal below, is that the harmonic wave is trapped in a waveguide and undergoes consecutive amplification. Owing to the weak energy loss by the high reflectivity and ENZ property, the waveguide carries a high Q and a much lower threshold. The geometry of light path is demonstrated in Fig. 5. To simplify a later elaboration, we strip off nonessentials to assume that (i) a mirror waveguide shown in Fig. 5 permits guiding the lowest mode by $\cos\theta = \lambda/(2l)$ with the wavelength λ inside the waveguide [77], where the self-consistency condition has a small change if absorption is considered (reflection induces weak changes of phase and then the reflection angle θ), as is ignored here; (ii) the reflection and absorption of pumping waves inside the waveguide are neglected since only a single pass is considered; (iii) the effective length of the waveguide is equal to the beam width of the side pumping waves with the satisfaction of $w \gg \lambda$ and therefore the plane wave theory is applicable; and (iv) the ENZ material is presumed to have the nonvanishing second-order susceptibility χ_{ijk} and to have an isotropic complex $\epsilon_2 = \varepsilon' + i\varepsilon''$ at the harmonic wavelength, which actually induces an anisotropy of refractive index. In such a background, a quantitative estimation can be carried out. We use here a silicon substrate and set the harmonic wavelength at the communication channel: $\lambda_0 = 1.55\,\mu$m (ENZ wavelength) relating to the wavelength λ by $\lambda_0 = n_R\lambda$, where n_R is the real refractive index. Although an undepleted-pump approximation will be succinctly applied below, a pump depletion in demand can be processed.

The generated harmonic wave propagates in a zig-zag path limited in the waveguide and grows the way a parametric amplification does in a resonant cavity. The first generation from the left ($z = 0$) to the right ($z = l$) boundary as in Fig. 5 yields an amplitude $A_1(z)$, which becomes $r^2 A_1(l)\exp(-\sigma l)$ when it returns to the left after a round trip, where r is the amplitude reflectivity and σ is the absorption coefficient along the z direction. From now on, the practical amplitude A in Eq. (30) is used for better understanding. In fact, the backward field radiated by the nonlinear polarization does not rise due to phase mismatch and always goes down to zero at the waveguide boundaries under the self-consistency condition. So, the BFC will not be included in the following derivation. Likewise, the second generating process gives

$$A_2(z) = A_1(z) + r^2 A_1(l)\exp[-\sigma(z+l)], \tag{58}$$

which directly implies

$$\begin{aligned} A_j(z) &= A_1(z) + r^2 A_{j-1}(l)\exp[-\sigma(z+l)] \\ &= A_1(z) + r^2 A_1(l)\exp[-2\sigma l - \sigma(z-l)] \\ &\quad + r^4 A_1(l)\exp[-4\sigma l - \sigma(z-l)] + \cdots \\ &\quad + r^{2(j-1)} A_1(l)\exp[-2(j-1)\sigma l - \sigma(z-l)] \\ &= A_1(z) + A_1(l)\frac{\rho-\rho^j}{1-\rho}\exp[-\sigma(z-l)] \end{aligned} \tag{59}$$

in the jth trip and

$$\rho = \exp[-2(\sigma l - \ln r)]. \tag{60}$$

According to Eq. (25), the amplitude A_1 under PM condition is

$$A_1(z) = \frac{\mu\omega\chi_{\text{eff}}}{c\sigma\tilde{n}_z}\mathcal{B}_1\mathcal{B}_2[1-\exp(-\sigma z)], \tag{61}$$

where $\mathcal{B}_{1/2}$ represents the amplitudes of two pumping waves and $\tilde{n}_z = n_z + i\,n_i$ is the complex refractive index in the z direction. The refractive indices and the absorption coefficient

are complicated compared with a nonabsorbing medium, so they need to be determined with respect to the direction. Certainly, they depend on θ and the complex permittivity ϵ_2. From the condition $\tilde{\mathbf{n}} \cdot \tilde{\mathbf{n}} = \mu(\varepsilon' + i\varepsilon'')$ [Eq. (A2)] and Eq. (27), we can get the following real refractive indices:

$$n_{\mathrm{R}}^2 = n_x^2 + n_z^2 = 2^{-1}\mu(\varepsilon' + \sqrt{\varepsilon'^2 + \varepsilon''^2 \sec^2\theta}), \tag{62a}$$

$$n_i = \sqrt{n_{\mathrm{R}}^2 - \mu\varepsilon'}, \quad n_z = \mu\varepsilon''(2n_i)^{-1}, \quad \sigma = \omega n_i/c. \tag{62b}$$

n_{R}, as an apparent refractive index, determines a real-valued angle of refraction by Snell's law. Since n_x is not greater than $\sqrt{\mu\varepsilon'}$, the angle θ is limited to a maximum

$$\theta_{\max} = \arctan(\sqrt{2\varepsilon'/\varepsilon''}). \tag{63}$$

This will limit the application of the PB angle of incidence for the TM mode to some extent. The above expressions of refractive indices are consistent with those in Ref. [75] (Chap. Optics in Metal). The reflectivity r in optics can be directly applied, but with the complex refractive index $\tilde{n}_{\mathrm{E}} = n_{\mathrm{R}} + in_i \sec\theta$ for the TE mode and $\tilde{n}_{\mathrm{M}} = n_{\mathrm{R}} + in_i \cos\theta$ for the TM mode, respectively [78]. Eventually, the internal conversion efficiency in the waveguide can be expressed as $\eta_{\mathrm{in}} = \eta_0 g$, where η_0 is the basic conversion efficiency, i.e.,

$$\eta_0 = Z_0 Z_{\mathrm{r1}}^2 \left(\frac{\omega\chi_{\mathrm{eff}}\lambda_0}{8c\cos\theta}\right)^2 \frac{2I_{\mathrm{p1}}I_{\mathrm{p2}}}{I_{\mathrm{p1}} + I_{\mathrm{p2}}}\eta_1\eta_2, \tag{64}$$

which represents the conversion efficiency in a $\lambda_0/2$ thick slab with a hypothetic refractive index, equal to that in vacuum. I_{p1} and I_{p2} are the pumping intensities of the two fundamental waves, whose injection efficiencies are η_1 and η_2, respectively. g is the gain and is a product of a single-pass gain g_{s} and a multiple-pass gain g_{mp} with $g = g_{\mathrm{s}}g_{\mathrm{mp}}$, which are

$$g_{\mathrm{s}} = Z_{\mathrm{r2}}\left|\frac{1 - \exp(-\sigma l)}{\tilde{n}_z \sigma l}\right|^2, \quad Z_{\mathrm{r2}} = \mu/n_{\mathrm{R}}, \tag{65a}$$

$$g_{\mathrm{mp}} = \left|\frac{1 - \rho^{j_{\max}}}{1 - \rho}\right|^2, \quad j_{\max} = Wn_{\mathrm{R}}\cot\theta\cos\theta, \tag{65b}$$

where $W = w/\lambda_0$, and $j_{\max}$ is the maximal number of round trips that the waveguide size w can sustain.

B. Simulation and discussion

At $\varepsilon' = 0.01$ with $\mu = 1$, the dependence of the gains g and g_{s} on θ, ε'', and W is computed and related contour plots are shown in Fig. 6. For a better illustration, the integer $j_{\max}$ is treated as a continuous variable and only the harmonic wave with $j_{\max} > 1$ is considered as an effective guided mode. The TE and TM modes each achieve a considerable increase and exhibit a similar pattern in the contour view. The conversion obtains a high gain for small angle propagation which maximizes in an intermediate angle for the TE and in a nearly zero angle for the TM mode. Bouncing in the waveguide at a small angle will greatly increase the effective propagating path inside according to Eq. (65b). That accounts

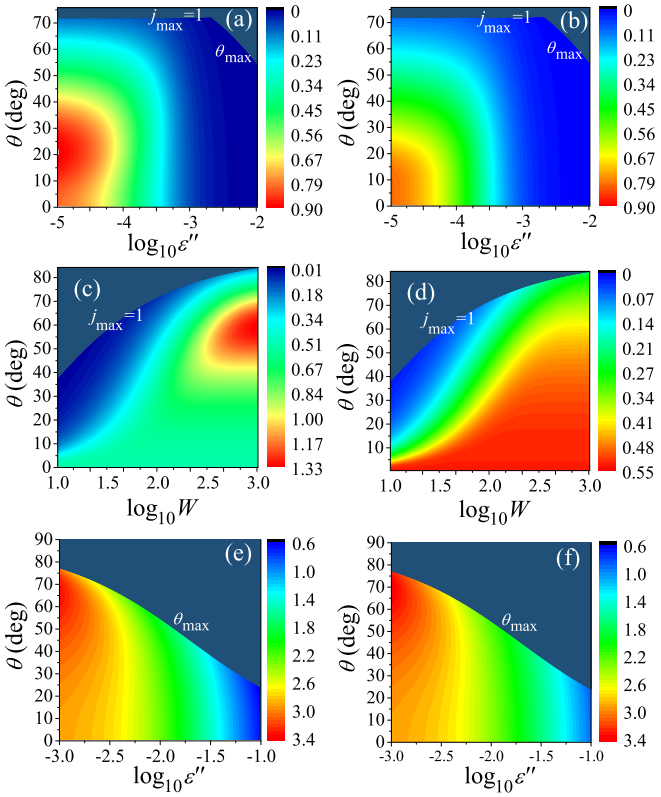

FIG. 6. Contour plots of the gain g scaled by 10^{-5} for TE (a) and TM (b) modes with $\varepsilon' = 10^{-2}$ over θ-ε'' space, and similar plots for TE (c) and TM (d) modes with $W = 100$, $\varepsilon'' = 10^{-4}$ over θ-W space, followed with the plots of the logarithmic single-pass gain $\log_{10} g_{\mathrm{s}}$ in the current model (e) and in the extended old model (f). The considered region is limited by $\theta < \theta_{\max}$ and $j_{\max} > 1$, where the ineffective region is uniformly colored in.

for the angular distribution of the gain. With a rise of the width W, the optimal angle approaches a larger value for the TE mode [see Fig. 6(c)], since it has an increasing reflectivity with θ and coupling with the pump waves. However, the reflectivity in the TM mode decreases when going towards the PB angle (near 90 deg). It is also shown that the optimal gain has an insensitive dependence on θ, so it renders a relaxed angular tolerance which can be used to improve injection efficiency in practice. Especially for the TE mode, it shows a higher gain region over θ-ε''-W space. For small $\varepsilon''/\varepsilon'$ ($\ll$1), the restriction to the gain comes from the loss of transmittance through waveguide boundaries. The single-pass gain g_{s} mainly relies on Z_{r2} and $|\tilde{n}_z| \approx n_{\mathrm{R}}\cos\theta$, i.e.,

$$g_{\mathrm{s}} \approx Z_{\mathrm{r2}}^3 \sec^2\theta/\mu^2. \tag{66}$$

We see that the single-pass gain contributes an enhancement of Z_{r2}^3, in which the coupling coefficient provides Z_{r2}; the other enhancements of Z_{r2}^2 are attributed to the increased waveguide thickness l. From Fig. 6(e), we see that it has an angle-related decrease with a growing ε''. The multiple-pass gain has a weak dependence on the absorption σ at a not very large angle, and is then determined by the reflectivity. In this case, the amplitude

reflectivity r has an approximate form

$$r \approx \begin{cases} -1 + 2\cos\theta\, \tilde{n}_{\mathrm{E}}/n_{\mathrm{si}}, & \text{for TE,} \\ +1 - 2\sec\theta\, \tilde{n}_{\mathrm{M}}/n_{\mathrm{si}}, & \text{for TM,} \end{cases} \tag{67}$$

where n_{si} is the refractive index of silicon at the harmonic wavelength. If the SHG is totally balanced by the multiple reflection and absorption, the harmonic wave will reach a steady distribution, the resonance mode, which is required to meet a condition

$$f \equiv -j_{\max} \ln \rho > 1 \tag{68}$$

with negligible absorption. Therefore, $g_{\mathrm{mp}} \approx 1/|(1-\rho)|^2$, which leads to the following estimation:

$$g \approx \begin{cases} n_{\mathrm{si}}^2 Z_{\mathrm{r2}}^5 \sec^4\theta/(2\mu)^4, & \text{for TE,} \\ n_{\mathrm{si}}^2 Z_{\mathrm{r2}}^5/(2\mu)^4, & \text{for TM.} \end{cases} \tag{69}$$

So g_{mp} contributes another enhancement Z_{r2}^2 due to a further path enlargement inside the waveguide. According to the ENZ setting $\varepsilon' = 0.01$, we have $Z_{\mathrm{r2}} \approx 10$, which gives rise to a nearly 10^5 enhancement. Equation (69) presents an estimation of the maximal gain. The high-gain regions in Figs. 6(a) and 6(b), displayed in red, just fall in the resonance regime. Even though the width factor W does not appear in Eq. (69), it is the driving force to push up the optimal angle in the TE mode and to relax the angular tolerance. W can be effectively enhanced in a sealed waveguide so as to reach the resonant condition. Under a nonresonant condition $f \ll 1, g_{\mathrm{mp}} \approx j_{\max}^2$, which yields

$$g \approx W^2 Z_{\mathrm{r2}} \cot^2\theta. \tag{70}$$

It is evident that the gain is taken over by W and is strongly modulated by $\cot^2\theta$ for both the TE and TM modes. On the other hand, for a large imaginary ϵ_2: $\varepsilon''/\varepsilon'$ (>1), the absorption dominates the loss, leading to $g_{\mathrm{mp}} \approx 1$. The waveguide thereby fails to boost the nonlinear process. The single-pass gain g_{s} represents the total gain

$$g \approx g_{\mathrm{s}} \approx Z_{\mathrm{r2}}^3 \mu^2 \cos^2\theta \left|\pi\left(\epsilon_2 Z_{\mathrm{r2}}^2 - \mu^2\right)\right|^{-2}. \tag{71}$$

It seriously depends on ε'' and gradually loses its advantage.

In this simulation, the anisotropy and BFC are excluded according to the previous simplification. Thus, the original CAE in a weak absorption condition may work well in the resonance mode, in which the absorption should be enough small to meet condition (68). To have access to the oblique incidence case, we ought to put the obliquity factor $\cos\theta$ in the original CAE as an intuitive extension. The only difference, if the exact absorption coefficient and reflectivity are still used, is the complex refractive index $\tilde{n}_z$ in Eq. (61), which will be n_z or $n_{\mathrm{R}}\cos\theta$ in an extended old CAE. They are illustrated in Figs. 6(e) and 6(f) as a demonstration of the single-pass gain. It is shown that they have a similar overall distribution but also denote an observable difference in the high-loss regime. In practice, the original CAE must be used carefully for lossy bulk materials or metal-based metamaterials.

It should be noted that the optimal gain does not coincide with the maximal total conversion efficiency, since there is an angular factor $\sec^2\theta$ in Eq. (64) of η_0 and an undetermined angular relation in the injection efficiency $\eta_1\eta_2$, which may raise the optimal angle. A practical pumping geometry has to be a trade-off between a desired configuration and engineering availability. The analysis of the gains above does not include the difference of the effective susceptibility between the TM and TE modes in harmonics, e.g., a χ_{zzz}-dominated nonlinearity in a metal-based metamaterial may only benefit the SHG of the TM mode. Certainly, the conversion efficiency cannot increase unlimitedly. Temperature variation and Kerr nonlinearity will eventually variate Z_{r2}, in which case, a steady output is expected.

VI. CONCLUSIONS

The general CAEs of the four transmission modes are derived in the plane-wave FC case, together with the integral equations, a generalization of Sipe's integral scheme. There are comparatively three fundamental features in the CAEs. The first is the simultaneous bidirectional FC (FFC and BFC). This is for certain an outcome of abandoning the SVAA. The second is the obliquity effect, which leads inevitably to the noncollinear phase mismatch owing to the space translational symmetry. Besides, the obliquity factor in the denominator accounts for a rise of interaction distance and for a walk-off effect due to the anisotropy. Finally, the absorption, carried by the complex refractive indices, presents transmission properties other than those in the loss-free media. In addition to the CAEs, it is also found that the NR field is negligible in the boundary issue. Therefore, the plausible common practice (the simple BCs) is legitimized and can substitute for the complicated nonlinear BCs. As a general model, it is thereby applicable to a homogenized nonlinear medium even with extreme parameters. For instance, the singular harmonic generation in ENZ materials can be examined in the effective medium approximation. In this paper, the CAEs are applied in a nonlinear ENZ waveguide with a sufficient simplification. It is found that the field-enhanced conversion resides in the mismatch of the wave impedance, and importantly, the near-zero index stands a chance for PM in up-conversion processes, as is very possible for the SHG in the so-called ENZ output case. By a simulation, the SHG is boosted in a mirror reflection waveguide by the resonance in combination with field enhancement, indicative of a maximal gain to the fifth power of the relative wave impedance. It is then expected to acquire further applications in micron-level high-Q photonic cavities.

ACKNOWLEDGMENTS

The authors thank Dr. Guang Zong Dong for helpful suggestions. This work was supported by the Natural Science Foundation of Heilongjiang Province of China under Grants No. F201312 and No. F2016023 and Tianjin Polytechnic University under Grant No. 030713.

APPENDIX A: WAVE VECTORS AND POLARIZATION OF THE TM EIGENMODE

We restrict our discussions in the (x,z) subspace, suitable for the TM eigenmode.The permitted wave constants k_{F} and k_{B} come from an algebraic equation pertinent to a matrix

$$\boldsymbol{M} \equiv \mathbf{kk} - k^2\boldsymbol{I} + k_0^2\boldsymbol{\epsilon}_{\mathrm{t}}, \tag{A1}$$

where $\mathbf{k}$ represents $\mathbf{k}_\mathrm{F}$ or $\mathbf{k}_\mathrm{B}$. Crystal optics requires that the wave vector of $\mathbf{k}$ satisfies

$$\det(\boldsymbol{M}) = 0. \tag{A2}$$

It will lead to the same result as Eq. (14). Expanding the determinant of $\boldsymbol{M}$ shows

$$\det(\boldsymbol{M}) = k_0^4 \det(\boldsymbol{\epsilon}_\mathrm{t}) - k_0^2 \mathbf{k}\cdot\boldsymbol{\epsilon}_\mathrm{t}\cdot\mathbf{k}, \tag{A3}$$

which directly leads to

$$k_0^2 \det(\boldsymbol{\epsilon}_\mathrm{t}) = \mathbf{k}\cdot\boldsymbol{\epsilon}_\mathrm{t}\cdot\mathbf{k} = k_0^{-2}\mathbf{k}\cdot\boldsymbol{M}\cdot\mathbf{k}, \tag{A4}$$

where Eq. (A1) and $\mathbf{k}\cdot(\mathbf{kk} - k^2\boldsymbol{I})\cdot\mathbf{k} = 0$ are used in the second identity. In detail, $\boldsymbol{M}$ can be written as

$$\begin{aligned}\boldsymbol{M}_\mathrm{F/B} &= \begin{bmatrix} k_0^2\epsilon_{11} - k_{\mathrm{F}z/\mathrm{B}z}^2 & k_0^2\epsilon_{13} \pm k_{\mathrm{F}z/\mathrm{B}z}k_x \\ k_0^2\epsilon_{31} \pm k_{\mathrm{F}z/\mathrm{B}z}k_x & k_0^2\epsilon_{33} - k_x^2 \end{bmatrix} \\ &\equiv \begin{bmatrix} \xi_\mathrm{F/B}^2\xi_0^{-1} & \xi_\mathrm{F/B} \\ \xi_\mathrm{F/B} & \xi_0 \end{bmatrix},\end{aligned} \tag{A5}$$

where the plus sign is for k_F, the minus sign is for k_B, and Eq. (A2) is used in the definition. Next, we define two unit vectors in the form of matrix

$$\hat{\mathbf{s}}_\mathrm{F/B} \equiv \pm\frac{\begin{bmatrix}\xi_\mathrm{F/B} \\ \xi_0\end{bmatrix}}{\sqrt{\xi_\mathrm{F/B}^2 + \xi_0^2}}, \quad \hat{\mathbf{a}}_\mathrm{F/B} \equiv \frac{\begin{bmatrix}\xi_0 \\ -\xi_\mathrm{F/B}\end{bmatrix}}{\sqrt{\xi_\mathrm{F/B}^2 + \xi_0^2}}, \tag{A6}$$

which are perpendicular to each other. It is easy to check that the relation $\mathbf{k}_\mathrm{F/B}\cdot\boldsymbol{\epsilon}_\mathrm{t}\cdot\hat{\mathbf{a}}_\mathrm{F/B} = 0$ holds and infers that $\hat{\mathbf{s}}$ is along the Poynting vector $\mathbf{S}$ and $\hat{\mathbf{a}}$ denotes the electric field polarization direction. Unsurprisingly, the two unit vectors just are the same ones used in the foregoing paragraphs. Furthermore, $\boldsymbol{M}_\mathrm{F/B}$ can be rewritten as a dyadic function

$$\boldsymbol{M}_\mathrm{F/B} = \xi_0^{-1}\left(\xi_\mathrm{F/B}^2 + \xi_0^2\right)\hat{\mathbf{s}}_\mathrm{F/B}\hat{\mathbf{s}}_\mathrm{F/B}. \tag{A7}$$

Returning to Eq. (16), G_0 can be expressed as an explicit form associated with polarization directions. From Eqs. (14), we have

$$\begin{aligned} k_{\mathrm{F}z} + k_{\mathrm{B}z} &= \frac{2}{\epsilon_{33}}\sqrt{\xi_0\det(\boldsymbol{\epsilon}_\mathrm{t})} = \frac{2}{k_0^2\epsilon_{33}}\sqrt{\xi_0\mathbf{k}\cdot\boldsymbol{M}\cdot\mathbf{k}} \\ &= \frac{2}{k_0^2\epsilon_{33}}\sqrt{\xi_\mathrm{F/B}^2 + \xi_0^2}(\mathbf{k}_\mathrm{F/B}\cdot\hat{\mathbf{s}}_\mathrm{F/B}), \end{aligned} \tag{A8}$$

where Eqs. (A4), (A6), and (A7) are used. Equation (16) is thus verified. In addition, from Eq. (A6) it is easily found that

$$\hat{\mathbf{a}}_\mathrm{F/B}\cdot\hat{\mathbf{x}} = \pm\hat{\mathbf{s}}_\mathrm{F/B}\cdot\hat{\mathbf{z}}. \tag{A9}$$

With these results, the coefficients in Eqs. (17) are finally determined. Solution (17) seems redundant but can simplify Eq. (20). The definitions in Eqs. (A6) will be complex valued due to a complex wave vector. The normalization $\hat{\mathbf{a}}\cdot\hat{\mathbf{a}} = 1$, $\hat{\mathbf{s}}\cdot\hat{\mathbf{s}} = 1$ still holds in this case. There also is another equivalent definition of $\hat{\mathbf{a}}$ and $\hat{\mathbf{s}}$ by setting $\hat{\mathbf{a}}^*\cdot\hat{\mathbf{a}} = 1$, $\hat{\mathbf{s}}^*\cdot\hat{\mathbf{s}} = 1$.

APPENDIX B: CURL OF GREEN'S FUNCTION ON THE BOUNDARIES

Collecting all the components of the Green's function from Eqs. (9) and (17), it can be rewritten as a dyadic form

$$\boldsymbol{G}(\mathbf{r},\mathbf{r}') = \sum_{n,m}[\mathcal{G}(\mathbf{r},\mathbf{r}')\,\hat{\mathbf{a}}\hat{\mathbf{a}}]_{nm} + \mathcal{G}_\mathrm{NR}\delta(z-z')\hat{\mathbf{z}}\hat{\mathbf{z}} \tag{B1}$$

with $\hat{\mathbf{a}}_{n2} = \hat{\mathbf{a}}_2 = -\hat{\mathbf{y}}$ according to the geometry in Fig. 1 and

$$\mathcal{G}_{nm}(\mathbf{r},\mathbf{r}') \equiv \pm\frac{i\mathrm{Ph}_{nm}(\mathbf{r},\mathbf{r}')}{2(\mathbf{k}\cdot\hat{\mathbf{s}}\hat{\mathbf{s}})_{nm}\cdot\hat{\mathbf{z}}}, \tag{B2a}$$

$$\mathcal{G}_\mathrm{NR}(\mathbf{r},\mathbf{r}') \equiv -\frac{1}{k_0^2\epsilon_{33}}\exp[ik_x(x-x')], \tag{B2b}$$

where the upper sign is for the forward fields and the lower sign is for the backward ones (the same for the expressions below). According to integral formula (20), the electric fields on the boundaries are related to the curl of the dyadic Green's function. We present here a straightforward deduction of the curl step by step, where $0 \leqslant z,z' \leqslant l$ is generally considered without the transition layers. First, the curl of the radiative dyadic fields in Eq. (B1) can be expressed as

$$\begin{aligned}\nabla'\mathcal{G}_{n1}\times\hat{\mathbf{a}}_{n1}\hat{\mathbf{a}}_{n1} &= -[i\mathbf{k}_{n1} \pm \hat{\mathbf{z}}\delta\Theta_\pm^{-1}]\times\hat{\mathbf{a}}_{n1}\hat{\mathbf{a}}_{n1}\mathcal{G}_{n1} \\ &= \mp[i\mathbf{k}_{n1}\cdot\hat{\mathbf{s}}_{n1} + \hat{\mathbf{a}}_{n1}\cdot\hat{\mathbf{x}}\delta\Theta_\pm^{-1}]\hat{\mathbf{y}}\hat{\mathbf{a}}_{n1}\mathcal{G}_{n1},\end{aligned} \tag{B3a}$$

$$\begin{aligned}\nabla'\mathcal{G}_{n2}\times\hat{\mathbf{a}}_2\hat{\mathbf{a}}_2 &= -[i\mathbf{k}_{n2} \pm \hat{\mathbf{z}}\delta\Theta_\pm^{-1}]\times\hat{\mathbf{a}}_2\hat{\mathbf{a}}_2\mathcal{G}_{n2} \\ &= -[ik_2\hat{\mathbf{h}}_{n2}\hat{\mathbf{a}}_2 \pm \hat{\mathbf{x}}\delta\Theta_\pm^{-1}]\mathcal{G}_{n2}\end{aligned} \tag{B3b}$$

with $\Theta_\pm = \Theta(\pm z \mp z')$, $\delta = \delta(z-z')$ and $\hat{\mathbf{h}}_{n2} = \hat{\mathbf{k}}_{n2}\times\hat{\mathbf{a}}_2$, denoting the directions of related magnetic vectors. Second, the curl of the NR field has a simple form as

$$[\nabla'\mathcal{G}_\mathrm{NR}\delta]\times\hat{\mathbf{z}}\hat{\mathbf{z}} = [-ik_x\hat{\mathbf{x}}\delta + \hat{\mathbf{z}}\delta']\times\hat{\mathbf{z}}\hat{\mathbf{z}}\mathcal{G}_\mathrm{NR} = ik_x\delta\hat{\mathbf{y}}\hat{\mathbf{z}}\mathcal{G}_\mathrm{NR}, \tag{B4}$$

where the first prime symbol in δ' indicates a derivative of the Dirac function. Then, an addition of all the coefficients of $\delta(z-z')$ is

$$\begin{aligned}&\hat{\mathbf{y}}(\hat{\mathbf{a}}_{\mathrm{B}1}\cdot\hat{\mathbf{x}}\hat{\mathbf{a}}_{\mathrm{B}1}\mathcal{G}_{\mathrm{B}1} + ik_x\hat{\mathbf{z}}\mathcal{G}_\mathrm{NR} - \hat{\mathbf{a}}_{\mathrm{F}1}\cdot\hat{\mathbf{x}}\hat{\mathbf{a}}_{\mathrm{F}1}\mathcal{G}_{\mathrm{F}1}) \\ &\quad - \hat{\mathbf{x}}\hat{\mathbf{a}}_2(\mathcal{G}_{\mathrm{F}2} - \mathcal{G}_{\mathrm{B}2})\end{aligned} \tag{B5}$$

with a drop of some Heaviside functions. In terms of Eqs. (37) and (38), it is zero at $\mathbf{r} = \mathbf{r}'$. Finally, the curl of the dyadic Green's field can be written as

$$\begin{aligned}\nabla'\times\boldsymbol{G}(\mathbf{r},\mathbf{r}') &= i(\mathcal{G}_{\mathrm{B}1}\mathbf{k}_{\mathrm{B}1}\cdot\hat{\mathbf{s}}_{\mathrm{B}1}\hat{\mathbf{y}}\hat{\mathbf{a}}_{\mathrm{B}1} - k_2\mathcal{G}_{\mathrm{B}2}\hat{\mathbf{h}}_{\mathrm{B}2}\hat{\mathbf{a}}_2) \\ &\quad - i(k_2\mathcal{G}_{\mathrm{F}2}\hat{\mathbf{h}}_{\mathrm{F}2}\hat{\mathbf{a}}_2 + \mathcal{G}_{\mathrm{F}1}\mathbf{k}_{\mathrm{F}1}\cdot\hat{\mathbf{s}}_{\mathrm{F}1}\hat{\mathbf{y}}\hat{\mathbf{a}}_{\mathrm{F}1}).\end{aligned} \tag{B6}$$

It is seen that Eq. (B6) does not present singularity, as expected, which is consistent with the results from the transition layer assumption. By Eqs. (36) and (B6), the expressions on the boundaries in Eq. (20) can be simplified without an effort.

APPENDIX C: INDEPENDENCE OF TRANSMISSION MODES

The propagating equation of the free electromagnetic wave $\mathbf{E}_{nm}$ with the transmission mode nm is

$$(\mathbf{k}\times\mathbf{k}\times\mathbf{E})_{nm}+k_0^2\boldsymbol{\epsilon}\cdot\mathbf{E}_{nm}=0. \tag{C1}$$

When it multiplies the complex conjugate of another free electromagnetic wave $\mathbf{E}_{n'm'}$, Eq. (C1) becomes

$$\mathbf{E}^*_{n'm'}\cdot(\mathbf{k}\times\mathbf{k}\times\mathbf{E})_{nm}+k_0^2\mathbf{E}^*_{n'm'}\cdot\boldsymbol{\epsilon}\cdot\mathbf{E}_{nm}=0. \tag{C2}$$

Similarly, we also have

$$\mathbf{E}_{nm}\cdot(\mathbf{k}\times\mathbf{k}\times\mathbf{E}^*)_{n'm'}+k_0^2\mathbf{E}_{nm}\cdot\boldsymbol{\epsilon}\cdot\mathbf{E}^*_{n'm'}=0. \tag{C3}$$

Subtraction of Eq. (C3) from Eq. (C2) leads to

$$\mathbf{E}^*_{n'm'}\cdot(\mathbf{k}\times\mathbf{k}\times\mathbf{E})_{nm}=\mathbf{E}_{nm}\cdot(\mathbf{k}\times\mathbf{k}\times\mathbf{E}^*)_{n'm'}, \tag{C4}$$

where the symmetry of $\boldsymbol{\epsilon}$ is applied. It can be transformed into

$$(\mathbf{E}^*_{n'm'}\times\mathbf{k}_{nm}\times\mathbf{E}_{nm})\cdot\mathbf{k}_{nm}=(\mathbf{E}_{nm}\times\mathbf{k}_{n'm'}\times\mathbf{E}^*_{n'm'})\cdot\mathbf{k}_{n'm'}. \tag{C5}$$

By use of the identity as follows,

$$(\mathbf{E}^*_{n'm'}\times\mathbf{k}_{nm}\times\mathbf{E}_{nm})\cdot\mathbf{k}_{n'm'}=(\mathbf{E}_{nm}\times\mathbf{k}_{n'm'}\times\mathbf{E}^*_{n'm'})\cdot\mathbf{k}_{nm}, \tag{C6}$$

we get

$$[\mathbf{E}^*_{n'm'}\times(\mathbf{k}\times\mathbf{E})_{nm}+\mathbf{E}_{nm}\times(\mathbf{k}\times\mathbf{E}^*)_{n'm'}]\cdot\mathrm{D}\mathbf{k}=0, \tag{C7}$$

where $\mathrm{D}\mathbf{k}=\mathbf{k}_{nm}-\mathbf{k}_{n'm'}\propto\hat{\mathbf{z}}$. It then becomes Eq. (49) if dropping the same amplitudes and phasors in Eq. (C7).

[1] Y. R. Shen, *Principles of Nonlinear Optics* (John Wiley and Sons, New York, 2003).
[2] R. W. Boyd, *Nonlinear Optics* (Academic Press, New York, 2008).
[3] Y. R. Shen, Ann. Rev. Mater. Sci. **16**, 69 (1986); Annu. Rev. Phys. Chem. **40**, 327 (1989); Nature (London) **337**, 519 (1989).
[4] E. Yablonovitch, Phys. Rev. Lett. **58**, 2059 (1987); S. John, *ibid.* **58**, 2486 (1987).
[5] J. D. Joannopoulos, R. D. Meade, and J. N. Winn, *Photonic Crystals: Molding the Flow of Light* (Princeton University Press, Princeton, NJ, 1995).
[6] M. Scalora, M. J. Bloemer, A. S. Manka, J. P. Dowling, C. M. Bowden, R. Viswanathan, and J. W. Haus, Phys. Rev. A **56**, 3166 (1997).
[7] M. Centini, C. Sibilia, M. Scalora, G. D'Aguanno, M. Bertolotti, M. J. Bloemer, C. M. Bowden, and I. Nefedov, Phys. Rev. E **60**, 4891 (1999).
[8] J. Bravo-Abad, A. Rodriguez, P. Bermel, S. G. Johnson, J. D. Joannopoulos, and M. Soljaccić, Opt. Express **15**, 16161 (2007).
[9] M. Kauranen and A. V. Zayats, Nat. Photon. **6**, 737 (2012).
[10] M. Lapine, I. V. Shadrivov, and Y. S. Kivshar, Rev. Mod. Phys. **86**, 1093 (2014).
[11] J. Butet, P.-F. Brevet, and O. J. F. Martin, ACS Nano **9**, 10545 (2015).
[12] N. Garcia, E. V. Ponizovskaya, and J. Q. Xiao, Appl. Phys. Lett. **80**, 1120 (2002).
[13] A. Alù, N. Engheta, A. Erentok, and R. W. Ziolkowski, IEEE Antennas Propag. Mag. **49**, 23 (2007).
[14] M. Silveirinha and N. Engheta, Phys. Rev. Lett. **97**, 157403 (2006).
[15] B. Edwards, A. Alù, M. E. Young, M. Silveirinha, and N. Engheta, Phys. Rev. Lett. **100**, 033903 (2008).
[16] R. Liu, Q. Cheng, T. Hand, J. J. Mock, T. J. Cui, S. A. Cummer, and D. R. Smith, Phys. Rev. Lett. **100**, 023903 (2008).
[17] B. Harbecke, B. Heinz, and P. Grosse, Appl. Phys. A **38**, 263 (1985).
[18] S. Feng and K. Halterman, Phys. Rev. B **86**, 165103 (2012).
[19] J. Yoon, M. Zhou, Md. A. Badsha, T. Y. Kim, Y. C. Jun, and C. K. Hwangbo, Sci. Rep. **5**, 12788 (2015).
[20] A. Alù and N. Engheta, Phys. Rev. Lett. **103**, 043902 (2009).
[21] A. Alù and N. Engheta, Materials **4**, 141 (2011).
[22] R. Fleury and A. Alù, Phys. Rev. B **87**, 201101 (2013).
[23] E. Liznev, A. Dorofeenko, L. Huizhe, A. Vinogradov, and S. Zouhdi, Appl. Phys. A **100**, 321 (2010).
[24] S. Hrabar, I. Krois, and A. Kiricenko, Metamaterials **4**, 89 (2010).
[25] S. S. Islam, M. R. I. Faruque, and M. T. Islam, Sci. Rep. **6**, 33624 (2016).
[26] A. Ciattoni, arXiv:1103.2864v1.
[27] M. A. Vincenti, D. de Ceglia, A. Ciattoni, and M. Scalora, Phys. Rev. A **84**, 063826 (2011).
[28] C. Argyropoulos, P.-Y. Chen, G. D'Aguanno, N. Engheta, and A. Alù, Phys. Rev. B **85**, 045129 (2012).
[29] D. de Ceglia, S. Campione, M. A. Vincenti, F. Capolino, and M. Scalora, Phys. Rev. B **87**, 155140 (2013).
[30] A. Ciattoni, C. Rizza, A. Marini, A. D. Falco, D. Faccio, and M. Scalora, Laser Photon. Rev. **10**, 517 (2016).
[31] M. Z. Alam, I. De Leon, and R. W. Boyd, Science **352**, 795 (2016).
[32] N. Engheta, Science **340**, 286 (2013).
[33] P. R. West, S. Ishii, G. Naik, N. Emani, V. M. Shalaev, and A. Boltasseva, Laser Photon. Rev. **4**, 795 (2010).
[34] S. Campione, I. Brener, and F. Marquier, Phy. Rev. B **91**, R121408 (2015).
[35] Q. Zhao, J. Zhou, F. Zhang, and D. Lippens, Mater. Today **12**, 60 (2009).
[36] P. Moitra, Y. M. Yang, Z. Anderson, I. I. Kravchenko, D. P. Briggs, and J. Valentine, Nat. Photon. **7**, 791 (2013).
[37] X. Huang, Y. Lai, Z. H. Hang, H. Zheng, and C. T. Chan, Nat. Mater. **10**, 582 (2011).
[38] E. J. R. Vesseur, T. Coenen, H. Caglayan, N. Engheta, and A. Polman, Phys. Rev. Lett. **110**, 013902 (2013).
[39] Y. Li, S. Kita, P. Munoz, O. Reshef, D. I. Vulis, M. Yin, M. Lončar, and E. Mazur, Nat. Photon. **9**, 738 (2015).
[40] J. Valentine, S. Zhang, T. Zentgraf, E. Ulin-Avila1, D. A. Genov, G. Bartal, and X. Zhang, Nature (London) **455**, 376 (2008).
[41] S. Campione, D. de Ceglia, M. A. Vincenti, M. Scalora, and F. Capolino, Phys. Rev. B **87**, 035120 (2013).

[42] T. S. Luk, D. de Ceglia, S. Liu, G. A. Keeler, R. P. Prasankumar, M. A. Vincenti, M. Scalora, M. B. Sinclair, and S. Campione, Appl. Phys. Lett. **106**, 151103 (2015).
[43] A. Capretti, Y. Wang, N. Engheta, and L. D. Negro, ACS Photon. **2**, 1584 (2015).
[44] A. Capretti, Y. Wang, N. Engheta, and L. D. Negro, Opt. Lett. **40**, 1500 (2015).
[45] H. Suchowski, K. O'Brien, Z. J. Wong, A. Salandrino, X. Yin, and X. Zhang, Science **342**, 1223 (2013).
[46] N. Mattiucci, M. J. Bloemer, and G. D'Aguanno, Opt. Express **22**, 6381 (2014).
[47] X. Chen, T. M. Grzegorczyk, and B. I. Wu, J. Pacheco, and J. A. Kong, Phys. Rev. E **70**, 016608 (2004).
[48] D. R. Smith, D. C. Vier, T. Koschny, and C. M. Soukoulis, Phys. Rev. E **71**, 036617 (2005).
[49] Y. Wu, J. Li, Z. Q. Zhang, and C. T. Chan, Phys. Rev. B **74**, 085111 (2006).
[50] I. Tsukerman, J. Opt. Soc. Am. B **28**, 577 (2011).
[51] C. R. Simovski, J. Opt. **13**, 013001 (2011).
[52] J. A. Armstrong, N. Bloembergen, J. Ducuing, and P. S. Pershan, Phys. Rev. **127**, 1918 (1962).
[53] D. A. Kleinman, Phys. Rev. **128**, 1761 (1962).
[54] N. Bloembergen and P. S. Pershan, Phys. Rev. **128**, 606 (1962).
[55] D. S. Bethune, J. Opt. Soc. Am. B **6**, 910 (1989); **8**, 367 (1991).
[56] J. J. Li, Z. Y. Li, and D. Z. Zhang, Phys. Rev. E **75**, 056606 (2007).
[57] H. Li, J. W. Haus, and P. P. Banerjee, J. Opt. Soc. Am. B **32**, 1456 (2015).
[58] P. Guyot-Sionnest, W. Chen, and Y. R. Shen, Phys. Rev. B **33**, 8254 (1986); P. Guyot-Sionnest and Y. R. Shen, *ibid.* **35**, 4420 (1987).
[59] J. E. Sipe, J. Opt. Soc. Am. B **4**, 481 (1987).
[60] N. Hashizume, M. Ohashi, T. Kondo, and R. Ito, J. Opt. Soc. Am. B **12**, 1894 (1995).
[61] A. Liu and O. Keller, Phys. Rev. B **49**, 13616 (1994).
[62] M. P. C. M. Krijn, Opt. Lett. **17**, 163 (1992).
[63] D. W. Berreman, J. Opt. Soc. Am **62**, 502 (1972).
[64] D. R. Smith and D. Schurig, Phys. Rev. Lett. **90**, 077405 (2003).
[65] S. Feng, Phys. Rev. Lett. **108**, 193904 (2012).
[66] L. Sun, S. Feng, and X. Yang, Appl. Phys. Lett. **101**, 241101 (2012).
[67] J. Luo, W. X. Lu, Z. H. Hang, H. Y. Chen, B. Hou, Y. Lai, and C. T. Chan, Phys. Rev. Lett. **112**, 073903 (2014).
[68] J. W. Yoon, G. M. Koh, S. H. Song, and R. Magnusson, Phys. Rev. Lett. **109**, 257402 (2012).
[69] A. Epstein and G. V. Eleftheriades, J. Opt. Soc. Am. B **33**, A31 (2016); Phys. Rev. B **90**, 235127 (2014).
[70] N. Chiotellis and A. Grbic, J. Opt. Soc. Am. B **33**, A51 (2016).
[71] J. Z. Sanborn, C. Hellings, and T. D. Donnelly, J. Opt. Soc. Am. B **20**, 152 (2003).
[72] C. Rodríguez and W. Rudolph, Opt. Lett. **39**, 6042 (2014).
[73] C. Rodríguez and W. Rudolph, Opt. Express **22**, 25984 (2014).
[74] M. A. Vincenti, D. de Ceglia, J. W. Haus, and M. Scalora, Phys. Rev. A **88**, 043812 (2013).
[75] M. Born and E. Wolf, *Principles of Optics*, 7th ed. (Cambridge University Press, Cambridge, UK, 1999).
[76] C. T. Tai, *Dyadic Green's Functions in Electromagnetic Theory*, 2nd ed. (IEEE Press, New York, 1994).
[77] B. E. A. Saleh and M. C. Teich, *Fundamentals of Photonics*, 2nd ed. (John Wiley and Sons, New York, 2007).
[78] Usually, $\tilde{n} \equiv \sqrt{\tilde{\mathbf{n}} \cdot \tilde{\mathbf{n}}} = \sqrt{\mu(\varepsilon' + i\varepsilon'')}$ is taken as the complex refractive index and is used together with a complex counterpart of θ. However, it is convenient here to use the real-valued quantities for the refractive indices and the angle of incidence or refraction.