Second harmonic generation by periodically-structured metal surfaces

W. L. Schaich
Department of Physics, Indiana University, Bloomington, Indiana 47405, USA
(Received 20 August 2008; revised manuscript received 22 October 2008; published 14 November 2008)

We present an approach to calculating second harmonic generation from periodically-structured metal surfaces. It requires the numerical solution of Maxwell's equations at both the first and second harmonic, but the electronic response to these fields is treated by a parametrization scheme developed earlier. Several model systems are considered, including metal patches of various shapes or holes in a metal film. Detailed calculations using the finite-difference time-domain method lead to predictions of the frequency dependence of second harmonic generation for different diffracted beams and polarizations. There is more spectral structure than what is evident in the linear reflection or transmission.

DOI: 10.1103/PhysRevB.78.195416 PACS number(s): 78.67.−n, 78.68.+m, 73.21.−b, 73.22.Lp

I. INTRODUCTION

The optical properties of metallic nanostructures are of great current interest from both fundamental and practical points of view.[1–3] Most of this interest has focused on linear properties, but there is a growing body of experimental studies of nonlinear behaviors.[4–33] A variety of theoretical approaches have also been put forward, usually in the context of a particular class of experiments.[32–46]

In this paper we develop an approach which can be applied to a range of geometric configurations. It is a combination of analytic parametrizations of second harmonic generation (shg) (Refs. 47 and 48) and numerical finite-difference time-domain (fdtd) evaluations.[49] The analytic work builds on the developments of about 20 years ago for the (top) flat surface of a thick metal slab.[50–56] We basically assume that the parametrization scheme for the second-order driving terms still may be used when the flat metal surfaces are not of infinite extent but have edges and corners. The numerical work is needed to determine both the profile of the linear-response fields and the second harmonic fields they produce. The fdtd method has become a standard scheme for calculating the linear behavior of nanostructures,[57–60] particularly those with a spatial scale comparable to the wavelength of the perturbing fields. Its extension to determine nonlinear response has been largely unexplored, and here we describe one way this may be done.

In Sec. II we examine a simple model system that can be solved either analytically or numerically. The results from these two approaches do agree, for appropriate mesh choices in space-time. Then in Sec. III we describe simulations of shg spectra for a sequence of more realistic systems with periodic structures: rectangular patches of metal, rectangular holes in a metal film, and T-shaped patches of metal. Our aim here is not a direct comparison with experiment, but rather a search for insight into what physical features and mechanisms can be important or not. We find that the second harmonic response can yield information that is not obvious in the linear response.

II. TEST PROBLEM

To aid our development of a numerical scheme, we first consider a simple test problem for which the answer is known. The system is the flat interface between semi-infinite jellium metal and vacuum. On one hand, to get any second harmonic generation in this case, the incident light cannot come in along the normal. But on the other hand, obliquely incident light complicates the fdtd calculation. As a compromise we imagine a coherent pair of p-polarized incident light beams, with opposite angles of incidence. The complex-valued, space- and time-dependent fundamental field is then periodic along the surface and can be written as

$$\vec{E}_1 = E_0 \frac{c}{\omega} e^{-i\omega t} \begin{cases} e^{ipx}(iQ \sin Qy, -p \cos Qy, 0) + re^{-ipx}(iQ \sin Qy, p \cos Qy, 0), & x < 0, \\ te^{ip'x}(iQ \sin Qy, -p' \cos Qy, 0), & 0 < x, \end{cases} \tag{1}$$

where ω is the frequency of the light and the metal lies in $x>0$. There is a single plane of incidence and the projection of the wave vector on the surface is $\pm Q = \pm(\omega/c)\sin\theta$, where $\pm\theta$ are the opposing angles of incidence. We align our Cartesian coordinate system with the plane of incidence, using the surface normal to define the x axis, the intersection of the surface with the plane of incidence to define the y axis, and $\hat{z}=\hat{x}\times\hat{y}$ as the third axis. The three triplets of values in Eq. (1) give the relative sizes of components along the $\hat{x}$, $\hat{y}$, and $\hat{z}$ axes for the incident, reflected, and transmitted light. The overall strength of the field is set by E_o, while the Fresnel amplitudes for reflection and transmission are given by

$$r = \frac{\varepsilon p - p'}{\varepsilon p + p'}, \quad t = \frac{2p}{\varepsilon p + p'}, \tag{2}$$

where ε is the metal's dielectric function at frequency ω and the normal components of the wave vector are $p=(\omega/c)\cos\theta$ and $p'=(\omega/c)\sqrt{\varepsilon-\sin^2\theta}$ for outside and inside the metal, respectively.

Next consider the generation of second harmonic fields. These are produced by the driving polarization, which can be divided into surface and bulk types.[54,55] The former are strong but localized near the surface, while the latter are weaker but extend into the bulk of the metal. These different functional forms lead to different ways of treating their effects. For the surface driving polarization $\vec{\mathrm{P}}^{(s)}$, one attributes its integrated (along the surface normal) strength to a δ function placed just outside the metal. The normal component of $\vec{\mathrm{P}}^{(s)}$ is proportional to the square of the normal component of $\vec{\mathrm{E}}_1$ in Eq. (1) for $x=0^+$, while the only other nonzero component of $\vec{\mathrm{P}}^{(s)}$ is along $\hat{y}$ and is proportional to the product of $\hat{x}\cdot\vec{\mathrm{E}}$ and $\hat{y}\cdot\vec{\mathrm{E}}$, again evaluated at $x=0^+$. In Sec. III we discuss the proportionality coefficients in more detail, but here we treat them as free parameters, writing

$$\int dx \hat{x}\cdot\vec{\mathrm{P}}^{(s)} = A(\cos Ky - 1)e^{-i\Omega t}, \tag{3}$$

$$\int dx \hat{y}\cdot\vec{\mathrm{P}}^{(s)} = iB\sin Kye^{-i\Omega t}, \tag{4}$$

where $\Omega=2\omega$ and $K=2Q$ are second harmonic values of frequency and surface wave vector.

The bulk driving polarization $\vec{\mathrm{P}}^{(b)}$ that we use here arises in a jellium picture from the Lorentz (pondermotive) force on free electrons. It is proportional to $\vec{\mathrm{E}}_1\times\vec{\mathrm{B}}_1$, where $\vec{\mathrm{E}}_1$ is the fundamental field of Eq. (1) and $\vec{\mathrm{B}}_1$ is the associated magnetic field. A simple calculation shows that one may write $\vec{\mathrm{P}}^{(b)}$ in $x>0$ as

$$\begin{aligned}\vec{\mathrm{P}}^{(b)} &= \{D[ip'\hat{x}(\cos Ky+1) - Q\hat{y}\sin Ky]e^{2ip'x}\}e^{-i\Omega t} \\ &= \left\{\frac{D}{2}\vec{\nabla}(\cos Ky+1)e^{2ip'x}\right\}e^{-i\Omega t}.\end{aligned} \tag{5}$$

Here D is the parameter (with the same units as A and B) that sets the strength of $\vec{\mathrm{P}}^{(b)}$, while the second line shows that $\vec{\mathrm{P}}^{(b)}$ is longitudinal.

With the driving polarizations defined by Eqs. (3)–(5), one has a straightforward problem to solve for the radiated second harmonic fields.[47,48] The transverse contributions to these can be written as

$$\vec{\mathrm{E}}_{2,\mathrm{tran}} = e^{-i\Omega t}\frac{c}{\Omega}\begin{cases}\alpha_- e^{-iPx}(iK\cos Ky, -P\sin Ky, 0), & x<0,\\ \alpha_+ e^{iP'x}(iK\cos Ky, P'\sin Ky, 0), & 0<x,\end{cases} \tag{6}$$

where $P=(\Omega/c)\cos\theta$ and $P'=(\Omega/c)\sqrt{\mathrm{E}-\sin^2\theta}$, with E as the metal's dielectric function at frequency Ω. We find

$$\alpha_- = 4\pi\frac{\Omega}{c}(\mathrm{E}KA - P'B - DQ)/(P' + \mathrm{E}P), \tag{7}$$

which sets the strength of the second harmonic field radiated back into vacuum. It is this quantity that we try to duplicate with an fdtd calculation. Since α_- is linear in A, B, and D, we can proceed with only one of them being nonzero at a time.

For the fdtd calculation of second harmonic fields, we need to use the driving polarizations in the Ampère-Maxwell equation. Assuming for the moment that E is a real-valued constant, one has

$$\vec{\nabla}\times\vec{\mathrm{H}}_2 = \frac{\mathrm{E}}{c}\frac{\partial\vec{\mathrm{E}}_2}{\partial t} + \frac{4\pi}{c}\frac{\partial\vec{\mathrm{P}}_{\mathrm{nl}}}{\partial t}, \tag{8}$$

where the nonlinear polarization is given by $\vec{\mathrm{P}}_{\mathrm{nl}}=\vec{\mathrm{P}}^{(s)}+\vec{\mathrm{P}}^{(b)}$. Equation (8) converts into an equation for iteration:[49]

$$E_2^{n+1} = E_2^n + \left\{(\nabla\times H_2^{n+1/2}) - 2\,\mathrm{Re}\left[\frac{4\pi}{c}(-i\boldsymbol{\Omega})P_{\mathrm{nl}}^{n+1/2}\right]\right\}c\boldsymbol{\Delta}t/\mathrm{E}, \tag{9}$$

where we have ignored the vector character. The superscripts are discrete labels for the different time steps, separated by Δt. We have evaluated the exact time derivative of P_{nl} at the magnetic time step. The P_{nl} in Eq. (9) are expressions such as in Eqs. (3)–(5) and our convention for producing the physical, real-valued $\partial P_{\mathrm{nl}}/\partial t$ is to take twice the real part (Re) of its complex-valued version.

The different Cartesian components of $\vec{\mathrm{P}}_{\mathrm{nl}}$ are used at the corresponding spatial mesh points for components of $\vec{\mathrm{E}}_2$. This prescription is straightforward for $\vec{\mathrm{P}}^{(b)}$, but for $\vec{\mathrm{P}}^{(s)}$ it requires special care. For $\hat{x}\cdot\vec{\mathrm{P}}^{(s)}$ we use from Eq. (3)

$$\hat{x}\cdot\vec{\mathrm{P}}^{(s)} = \frac{A}{\Delta x}(\cos Ky - 1)e^{-i\Omega t}, \tag{10}$$

where Δx is the spatial step size along $\hat{x}$. Equation (10) is used only in the first layer of x points outside the metal. A similar recipe applies for $\hat{y}\cdot\vec{\mathrm{P}}^{(s)}$.

In our actual calculations E is not a constant but exhibits (Drude) dispersion. We account for this by the method of auxiliary differential equations.[49] The quantities inside the curly brackets in Eq. (9) are not changed by this modification. Another technical feature is that the driving terms are turned on gradually around $t=0$ using

$$e^{-i\boldsymbol{\Omega}t} \to e^{-i\boldsymbol{\Omega}t}\begin{cases}e^{-(t/t_0)^2}, & t<0,\\ 1, & 0<t,\end{cases} \tag{11}$$

where t_0 in the Gaussian factor is large enough to require more than a thousand time steps to "turn on."

In a series of evaluations we used a fundamental wavelength $\lambda=800$ nm and $\theta=53.1°$ so p, Q, and ω/c are in a 3, 4, 5 ratio. The square surface unit cell then has side $2\pi/Q=1$ μm. The metal parameters were those for Al—see Sec. III. A typical result is shown in Fig. 1, which plots the normal component of $\vec{\mathrm{E}}_2$ far outside the metal versus time at the fixed value of $y=0$ (so $\hat{y}\cdot\vec{\mathrm{E}}_2=0$). The analytic and fdtd re-

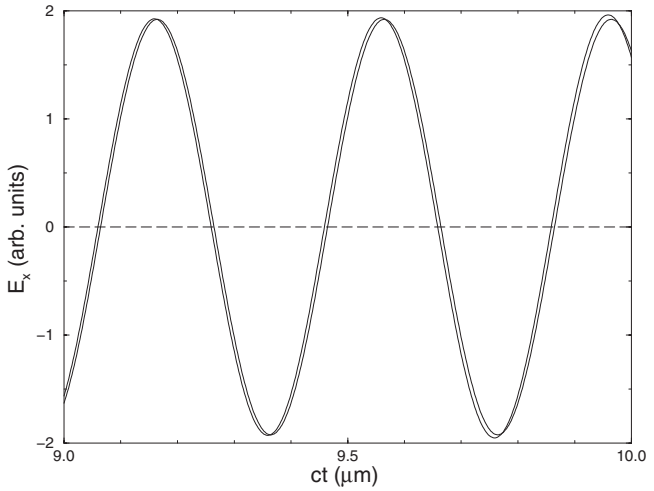

FIG. 1. Comparison of analytic and fdtd calculations of second harmonic fields from a surface driving polarization described by $A \neq 0$ and $B=0=D$. The thick (thin) curve is from the fdtd (analytic) calculation. See text for other parameters used.

sults agree quite well, in both phase and amplitude. Similar agreement was found for other cases, with different field components and/or different nonzero A, B, and D. For all these calculations we used a spatial step size of $\Delta x=\Delta y=\Delta z=4$ nm and a temporal step size of $c\Delta t=2$ nm. The agreement, such as seen in Fig. 1, sets in for $ct \geq 2$ μm. Reducing t_0 in Eq. (11) degrades the agreement. It is also important to choose the spatial mesh points to be consistent with the expected symmetries of the field components; e.g., since $y=0$ is a symmetry line, mesh points should be equally spaced on either side of $y=0$. We conclude from this exercise that our fdtd scheme can, given the driving polarizations, determine the strength of radiated second harmonic fields with a good quantitative accuracy.

III. MODEL CALCULATIONS

Now we turn to more complicated structures. The incident light will come in along the normal to the (average) surface plane and either asymmetries in the structure or detection away from the normal will allow a shg signal to be nonzero. We will consider configurations that can be built with periodic arrays of finite flat surfaces. In this paper we shall ignore for simplicity the effects of any supporting dielectrics. Thus we have only metal-vacuum interfaces, which are the best understood. With this limitation, one can be fairly confident about the size of parameters for $\vec{P}_{nl}$, at least for simple metals such as Al.[47]

The parametrization scheme is based on a jellium picture of the metal. We use, if $\hat{x}$ is along the local surface normal,[47,54–56]

$$\hat{x} \cdot \vec{P}^{(s)} = \frac{a/2}{\Delta x} \frac{e}{8\pi m} \frac{(1-\varepsilon)^2}{\omega_p^2} (\hat{x} \cdot \vec{E}_1)^2, \tag{12}$$

$$\hat{y} \cdot \vec{P}^{(s)} = \frac{b}{\Delta x} \frac{e}{8\pi m} \frac{(1-\varepsilon)^2}{\omega_p^2} (\hat{x} \cdot \vec{E}_1)(\hat{y} \cdot \vec{E}_1), \tag{13}$$

$$\vec{P}^{(b)} = -d \frac{e}{4\pi m} \frac{(1-\varepsilon)(1-E)}{\omega_p^2} \times \left[\vec{E}_1 \times \left(\frac{i\omega}{c} \right) \vec{B}_1 + \frac{\omega}{\omega + (i/\tau)} (\vec{E}_1 \cdot \vec{\nabla}) \vec{E}_1 \right]. \tag{14}$$

Here $e<0$ and m are the charge and mass of the conduction electrons and $\omega_p^2=4\pi n e^2/m$ with n the bulk density is the plasma frequency. The two dielectric functions have the Drude form with a common frequency-independent scattering rate $1/\tau$:

$$\varepsilon = 1 - \frac{\omega_p^2}{\omega(\omega + i/\tau)}, \quad E = 1 - \frac{\omega_p^2}{\boldsymbol{\Omega}(\boldsymbol{\Omega} + i/\tau)}. \tag{15}$$

The fields $\vec{E}_1$ and $\vec{B}_1$ are at the fundamental frequency ω. For $\vec{P}^{(s)}$ they have to be evaluated just inside the metal. The calculation of these fields will be done with fdtd codes. For all the configurations considered here, each local surface normal is along one of our (fixed) Cartesian axes. In general $\vec{P}^{(s)}$ will have all three components nonzero, which means that one needs to supplement Eq. (13) with an equation wherein $\hat{y} \rightarrow \hat{z}$ but the parameter b remains unchanged. As in Sec. II, the equations for $\vec{P}^{(s)}$ describe the strength of "δ-function" terms placed just outside the metal at appropriate (different) lattice points for each component.

Earlier work[47–56] has shown that for jellium $b=-1$ and $d=+1$ for all $\omega < \omega_p/2$, while a is frequency dependent and complex valued. For $\omega \ll \omega_p$ the real part of a dominates and is negative with a magnitude of around 10. In real metals additional contributions to $\vec{P}_{nl}$ may exist[47,48] due, say, to asymmetries in crystal surface structures or interband excitations. We ignore these complications for now. The second term in the square bracket in Eq. (14) is due to the long-ranged contribution of the "convective derivative." This term appears in the equation of motion for the electron velocity field $\vec{v}$ in the form $(\vec{v} \cdot \vec{\nabla})\vec{v}$. One may only keep its bulk contribution since its surface contribution is classically ill defined but is properly included in the quantum-mechanical calculations of the a parameter. For the test problem in Sec. II, the $(\vec{E}_1 \cdot \vec{\nabla})\vec{E}_1$ term is nonzero but produces no radiation. For the more complicated configurations to be considered here, it will produce radiation of a similar strength as the pondermotive term in Eq. (14).

Again we note that we are also ignoring the effect of any supporting dielectrics. *A priori* calculations of how these would affect the a and b parameters have not been done, although experimental measurements do indicate that changes can be significant.[47] The dielectrics also complicate the fdtd calculation of radiated second harmonic fields because the $\vec{P}^{(s)}$ δ functions could not then be placed in vacuum, which would lead to additional screening. The difficulty lies with the normal component of $\vec{P}^{(s)}$ and $\vec{E}_2$ and further clouds the question of what value to use for a.[48]

In our fdtd calculations we suppress from $\vec{P}_{nl}$ the dimensionless factor of

$$eE_0/m\omega_p^2 l = E_0/E_s, \tag{16}$$

where E_0 is the incident electric field amplitude and $l = 1\ \mu$m. The scaling field can be re-expressed as

$$E_s = 4\pi^2\left(\frac{e}{\lambda_p^2}\right)\frac{l}{r_0}, \tag{17}$$

where $\lambda_p = 2\pi c/\omega_p$ is the bulk plasmon wavelength and $r_0 = e^2/mc^2$ is the classical electron radius. For $\lambda_p = 0.1\ \mu$m, Eq. (17) yields $E_s = 2\times 10^{15}$ V/m. The strength of shg is set by a Poynting flux. For a measure independent of the incident flux, one often uses the ratio

$$\Gamma = \langle \hat{x}\cdot\vec{S}_2\rangle/\langle \hat{x}\cdot\vec{S}_1\rangle^2, \tag{18}$$

where $\vec{S}_1$ and $\vec{S}_2$ are the Poynting vectors at the first and second harmonic, respectively, and the angular brackets denote a time average. Since $\langle \hat{x}\cdot\vec{S}\rangle = (c/8\pi)|\vec{E}|^2\cos\theta$ we can write, noting that for an incident beam along the normal $\theta_1 = 0$,

$$\Gamma = \frac{8\pi}{c}|\vec{E}_2|^2\cos\theta_2/|\vec{E}_1|^4 = \frac{8\pi}{c}\left(\frac{e}{m\omega_p^2 l}\right)^2\gamma, \tag{19}$$

where θ_2 is the emission angle of the second harmonic radiation. We calculate the dimensionless γ, but plot it with arbitrary scales. We do note the maximum value of γ in each case. The inverse of the factor preceding γ in Eq. (19) is the flux associated with E_S. Using it one may convert our γ values to absolute shg efficiency magnitudes.

In all the calculations to follow, the metal is treated as Al with $\hbar\omega_p = 12.6$ eV and $1/\omega_p\tau = 0.007$. We also set $a = -10$. For each shg frequency point, we need two separate fdtd time iterations. The first uses an incident pulse plus a subsequent time-Fourier transform to find $\vec{E}_1(\vec{x},\omega)$ and $\vec{B}_1(\vec{x},\omega)$. Using these the $\vec{P}_{nl}$ are calculated from Eqs. (12)–(14). Then the second fdtd iteration is done using the analog of the recipe given in Sec. II [see Eq. (9)] to find $\vec{E}_2$. The mesh in space-time is the same as in Sec. II. Convergence is generally as good as there except near Rayleigh-Wood anomalies, where diffracted beams change from propagating to evanescent, which makes the field vary slowly as one moves away from the surface.

A. Rectangular patches

Our first example is a two-dimensional (2D) square array with period d of rectangular patches, each with dimensions w_x, w_y, and w_z. The experimental system we have in mind is described in Refs. 18 and 26 but it had Au nanoparticles on an indium tin oxide–coated glass substrate. Our patches are "free-standing" Al. We set $d = 480$ nm, $w_x = 20$ nm, and $w_z = 96$ nm, and vary w_y. In Fig. 2 we plot the reflectivity R for several choices of w_y. The incident light is polarized along $\hat{y}$. We are particularly interested in the strong resonance whose location moves steadily with w_y.

The only other evident structures in Fig. 2 are Raleigh-Wood anomalies, which occur near

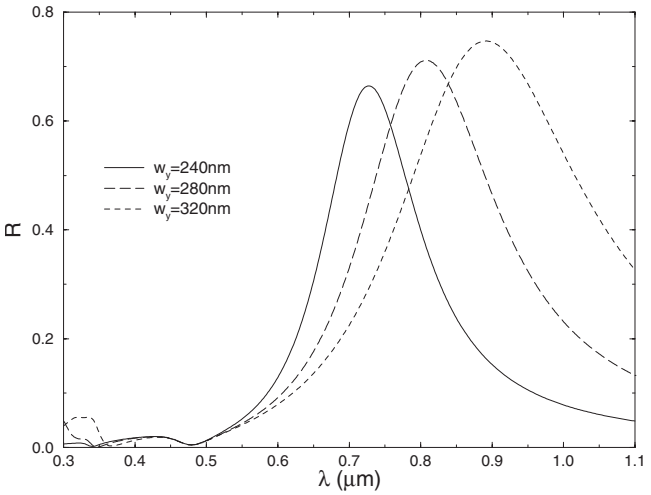

FIG. 2. Normal-incidence reflectivity R versus wavelength λ for a 2D square array of rectangular Al patches. The length of each patch along the direction of polarization is varied for different curves. See text for other parameters.

$$\lambda = d/(j^2 + k^2)^{1/2}, \tag{20}$$

where j and k are integers that define diffracted beams through the 2D reciprocal lattice vectors: $\mathbf{G} = (2\pi/d)(j\hat{y} + k\hat{z})$. Note the structures due to the (1,0) beams near $\lambda = 480$ nm and due to the (1,1) beams near $\lambda = 480/\sqrt{2} = 340$ nm. For the shg calculation we will use $w_y = 280$ nm, so the resonance in R is centered on 808 nm. One might then expect a peak in shg radiation near 400 nm, which for the fundamental light is in the middle of a range of smooth variation in R.

Since our patch system has inversion symmetry, there are no shg beams moving along the normal, i.e., no (0,0) beams. We show in Fig. 3 the shg efficiency, γ of Eq. (19), versus fundamental wavelength for the (1,0) p-polarized reflected beam. We also replot R from Fig. 2 for easier comparison. Both R and γ have been scaled here to have a maximum of unity. The absolute maximum of this γ is $\gamma_p^{(1,0)} = 7.0\times 10^8$. For the alternate s-polarized (1,0) reflected beam, we find a $\gamma_s^{(1,0)}$ that is zero to within our numerical accuracy.

The peaks in the first harmonic R and the second harmonic $\gamma_p^{(1,0)}$ have different locations and shapes. There is no hint in the reflectivity plot as to where the shg peak will occur.

The (1,0) shg beam can propagate to the far field only for fundamental wavelengths less than 960 nm. This constraint accounts for the sharp cutoff of $\gamma^{(1,0)}$ at larger wavelengths. It is in fact difficult to calculate $\gamma^{(1,0)}$ as $\lambda \to 960$ nm because the shg beam fields vary slowly with x and we must move the plane at which $\vec{E}_2$ is used to compute $\gamma^{(1,0)}$ farther out into vacuum. The shg efficiency does have a Rayleigh-Wood anomaly near $\lambda = 960/\sqrt{2} = 680$ nm, but it is not visible on the scale used in Fig. 3.

A rationale for the distinct behavior of γ and R in Fig. 3 can be based on the following premise: the second harmonic driving polarizations $\vec{P}_{nl}$ can excite patch modes that are symmetrically inaccessible within linear response to light at

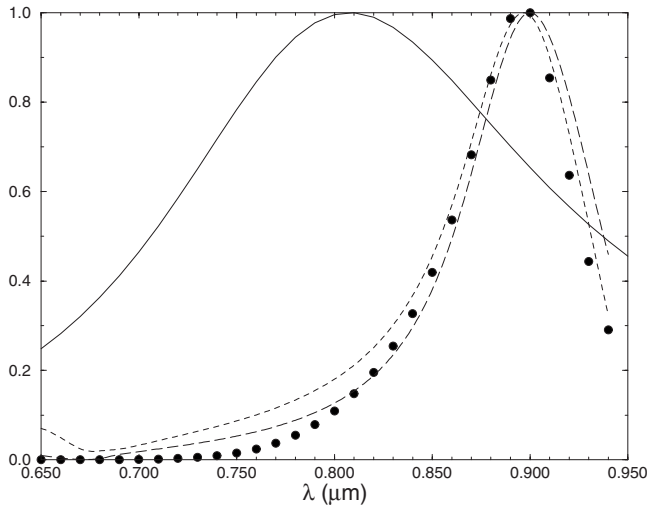

FIG. 3. Reflectivity (solid line) and shg (dots) versus fundamental wavelength λ for the system in Fig. 2 with w_y=280 nm. Both quantities have been scaled to have a maximum of unity. The shg efficiency is for the reflected p-wave (1,0) beam, while the reflectivity is specular, i.e., the (0,0) beam polarized along $\hat{y}$. The dashed curves are estimates of $\gamma_p^{(1,0)}$ based on drastic approximations to $\vec{P}_{nl}$—see text.

normal incidence. To justify this claim, we need to note some facts and to make a few simplifying assumptions. The results for γ are dominated by the surface driving terms in the form of Eq. (12). We have calculated $\gamma_p^{(1,0)}$ vs λ either setting d to zero or setting both b and d to zero. The shape of the plots is essentially the same as in Fig. 3, aside from a change in overall strength. The peak value of $\gamma_p^{(1,0)}$, at λ=900 nm, grows to 8.2×10^8 when $b=0=d$ and grows further to 1.0×10^9 when only $d=0$. The fact that γ is larger when some driving terms are suppressed is a reminder that contributions from a, b, and d can interfere in producing γ. Here we use the fact that a terms dominate to ignore for the sake of a qualitative argument all the others, and further to consider only contributions from the top or bottom of the patch since these surfaces have the greatest area and will produce resonances at the longest wavelengths. For light incident along the normal, polarized along $\hat{y}$, currents will be produced in linear response that flow primarily along $\hat{y}$. This is certainly reasonable for $w_y \gg w_x, w_z$, when the patch becomes a linear antenna. The induced current, as well as the electric field that drives it, must be symmetric about the center of any patch. Placing $y=0=z$ at the center of a patch, the currents near the top or bottom of that patch can be written in the approximate standing-wave form,[61–63]

$$\vec{J}(y,z;t) \approx \hat{y} J_0 \cos(\pi j y/w_y) e^{-i\omega t}, \tag{21}$$

where J_0 is a complex amplitude and j is an odd integer. Even values of j are forbidden by the conditions that $\hat{y}\cdot\vec{J}$ be even in j and vanish at $|y|=w_y/2$. We have calculated profiles of fields and currents within a patch. For $\lambda > 700$ nm they do have the form of Eq. (21) with $j=1$ and J_0 showing a weak resonant peak near 800 nm.

With a current density along $\hat{y}$ and proportional to $\cos(\pi y/w_y)$, the equation of continuity implies that the induced charge density varies as $\sin(\pi y/w_y)$, which in turn produces by Gauss's law a normal component of electric field at the top or bottom surface with the same y dependence. Thus one has at the larger surfaces of a patch (suppressing the time dependence) the following:

$$\hat{x}\cdot\vec{E}_1 \approx E_{x0}\sin(\pi y/w_y),$$

$$\hat{y}\cdot\vec{E}_1 \approx E_{y0}\cos(\pi y/w_y). \tag{22}$$

Referring to Eqs. (12) and (13), we then find

$$\hat{x}\cdot\vec{P}^{(s)} \propto \sin^2(\pi y/w_y) = [1-\cos(2\pi y/w_y)]/2,$$

$$\hat{y}\cdot\vec{P}^{(s)} \propto \cos(\pi y/w_y)\sin(\pi y/w_y) = \sin(2\pi y/w_y)/2. \tag{23}$$

Such driving terms can excite a mode with sinusoidal currents and fields of argument $2\pi y/w_y$. Recalling Eq. (21), we see that the nonlinear couplings lead to a $j=2$ mode from either a or b terms. We believe the sharp resonance in Fig. 3 arises in essence from the $j=2$ mode. Profiles of the second harmonic near fields within a patch are consistent with this interpretation.

To further support our view, we ran our fdtd codes for γ using for $\vec{P}_{nl}$ one or the other line in Eq. (23), evaluated just outside the top of the patch. The proportionality coefficient was independent of λ and adjusted to produce a maximum value for γ of unity. As shown in Fig. 3 these extreme simplifications reproduce quite well the shape of the full calculation of γ vs λ.

We end this subsection with the comment that we also calculated γ for the (0,1) beams. Only the p-wave γ is significant but the maximum value (at λ=830 nm is at $8.3\times10^6 \geq \gamma_p^{(0,1)}$, almost 2 orders of magnitude smaller than the maximum of $\gamma_p^{(1,0)}$.

B. Rectangular holes

Our second example is a (rough) Babinet complement of the first:[64] a 2D square array of period d of holes in an unsupported Al film of thickness h. Each hole has a rectangular cross section $w_y\times w_z$. The motivating experimental system is that of Ref. 25, which had a hole array in a Au film on glass. We set d=480 nm, h=160 nm, and w_y=96 nm, and vary w_z. In Fig. 4 we plot the normal-incidence transmission T for several choices of w_z. The incident polarization is along $\hat{y}$. Our focus is on the longest wavelength resonance, an example of "extraordinary optical transmission,"[65–67] which moves monotonically with w_z.[68,69]

As an aside, we remark that the peaks just above 500 nm are due to a second resonance which is distorted by the Rayleigh-Wood anomaly near 480 nm. This interpretation is supported by further calculations in which we varied either h or d. The former variation moves the resonant peak locations, suggesting that they arise from Fabry-Pérot resonances within the holes.

In Fig. 5 we plot shg efficiencies for two different beams along with the corresponding transmission from Fig. 4.

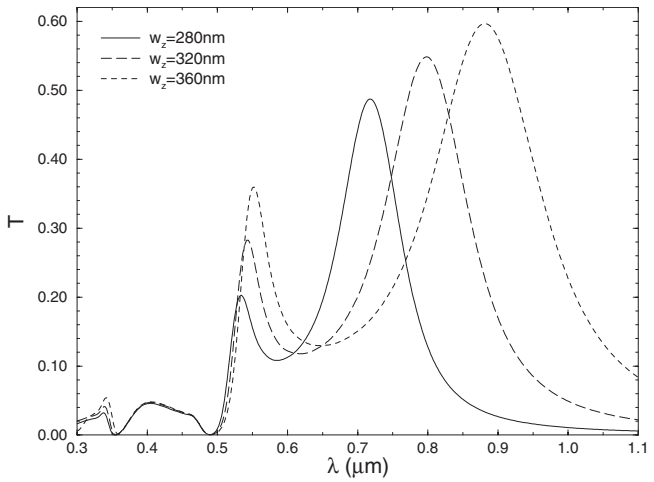

FIG. 4. Normal-incidence transmission T versus wavelength λ for a 2D square array of rectangular holes in Al film. The width of each hole along the direction normal to the polarization is varied for different curves. See text for other parameters.

Again by inversion symmetry there are no shg beams moving along the surface normal. The absolute maxima are $\gamma_p^{(0,1)} \leq 3.2 \times 10^6$ and $\gamma_p^{(1,0)} \leq 1.6 \times 10^6$, while the s-polarized beams are at least 2 orders of magnitude weaker. Comparing to Fig. 3 we again have sharp peaks at locations shifted away from the linear resonance. The shift is here toward shorter wavelengths and the sharp structures lie slightly above the (1,1) Rayleigh-Wood anomaly. We are not sure whether to attribute these structures to the expected anomaly or to excitation of surface plasmons. Such a choice has been controversial. Unlike in Fig. 3, the sharp structures are accompanied by broad resonances close to and with a similar width as the transmission resonance.

Our interpretation of the results in Fig. 5 is that the primary excitations in this system are extended surface plasmons which can be readily scattered by the hole array into different directions. This picture qualitatively holds at both the first and second harmonic energies. Moreover, in contrast to the patch array, no new structure locations arise in the shg although the appearance of structures may be considerably modified—contrast T near 340 nm in Fig. 4 and the γ near 700 nm in Fig. 5.

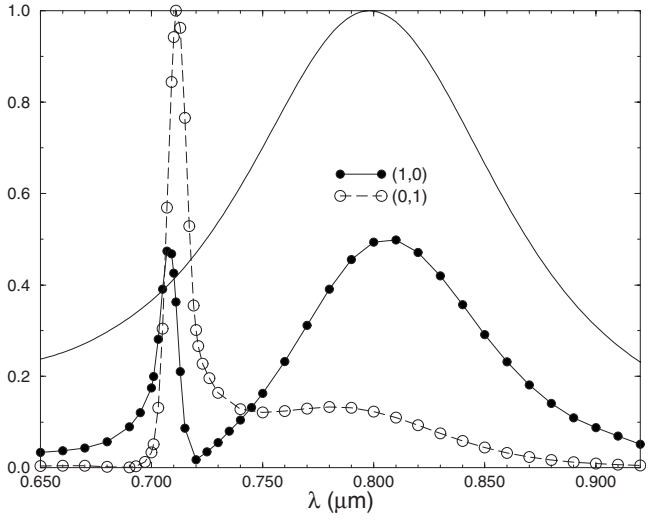

FIG. 5. Transmission (thick line) and shg (circles) versus fundamental wavelength λ for the system in Fig. 4 with $w_z = 320$ nm. All quantities have been scaled to have a maximum of unity except for $\gamma_p^{(1,0)}$, which is on the same scale as $\gamma_p^{(0,1)}$. The shg efficiencies are for the transmitted p-wave beams, while the transmission is for the straight-through (0,0) beam.

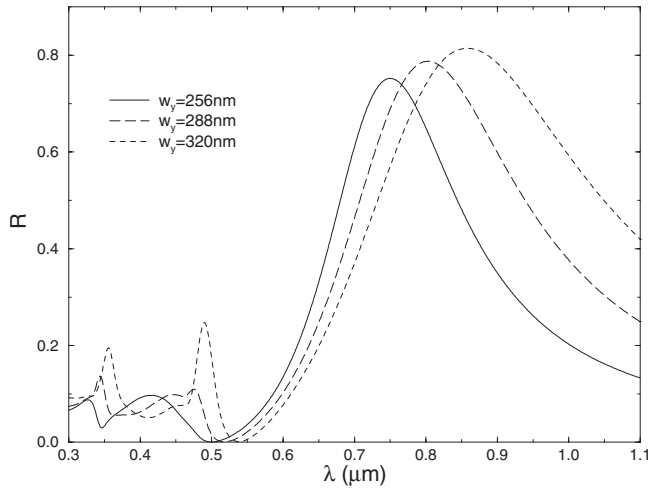

FIG. 6. Normal-incidence reflectivity R versus wavelength λ for a 2D square array of T-shaped Al patches. The direction of polarization is aligned with the height of each T, which is varied for different curves. See text for other parameters.

C. T patches

The last system we consider removes the inversion symmetry present in the first two. It is a collection of T-shaped Al patches arranged in a 2D square array. Each T-shaped patch is described by two adjacent rectangles: $w_{y1} \times w_{z1}$ and $w_{y2} \times w_{z2}$. The height of the T is along $\hat{y}$ and, with $w_{z2} = 2w_{z1}$, the crossbar is along $\hat{z}$. Each patch has thickness w_x and the array period is again d. For a similar experimental system, see Ref.

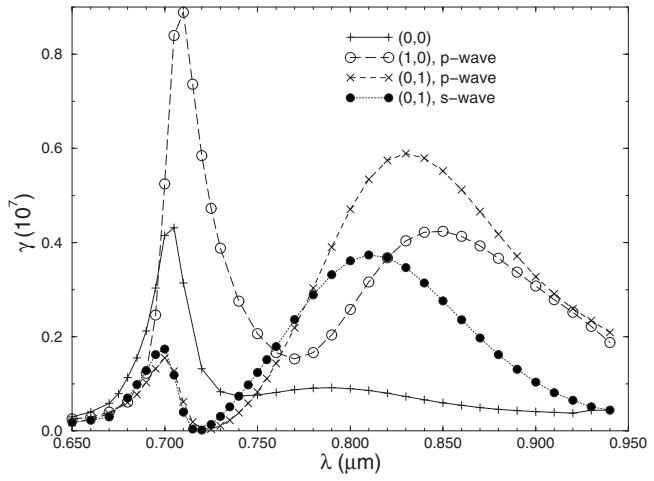

FIG. 7. Various diffracted beams and polarizations of shg versus fundamental wavelength λ for the system in Fig. 6 with $w_y = 288$ nm. The shg efficiencies γ have not been rescaled.

27, which used Au T patches on a silica substrate. We set $d=480$ nm, $w_{z1}=128$ nm, and $w_{z2}=256$ nm, and with $w_{y1}=w_{y2}$ vary $w_y=w_{y1}+w_{y2}$. Figure 6 shows the specular reflectivity at normal incidence with the polarization along $\hat{y}$ for several choices of w_y. We again are interested in the resonance at the longest wavelength, but note the stronger (and more varied) Rayleigh-Wood anomalies compared to those in Fig. 2.

In Fig. 7 we plot shg efficiencies for several beams using the exact scale for γ. Since there are so many curves, we omit redrawing R from Fig. 6. In the absence of inversion symmetry, the (0,0) beam can be nonzero, although it is relatively weak. The polarization of this beam is along $\hat{y}$; the γ for the orthogonal polarization is, like $\gamma_s^{(1,0)}$, at least 3 orders of magnitude weaker. Note, however, that for the (0,1) beams the γ for the p and s polarizations are comparable.

In general terms one has broad peaks in γ near the long wavelength resonance in R plus sharp structures near the Rayleigh-Wood (1,1) point at 680 nm. We do not believe that the sharp structures arise from a broken symmetry by the $\vec{P}_{nl}$, as is the case in Sec. III A, because the T-shaped patches already have no special symmetry along $\hat{y}$, even for the linear fields. This is why so many beams and polarizations have a significant γ in Fig. 7. The location and generally asymmetric shape of the sharp structures suggest that they arise from Rayleigh-Wood anomalies, which apparently are much stronger for the second harmonic than the linear response.

To sum up, we have derived and evaluated a scheme for calculating shg from periodically-structured metals in the form of patches or holes of different sizes but always rectangular shapes. The results show a strong qualitative dependence on geometric configuration. Experimental studies of the spectra of shg would be useful for exhibiting the various spectral features found here.

ACKNOWLEDGMENTS

This work was supported in part by a grant from the NSF-CBET program (Grant No. 0708590).

[1] L. Novotny and B. Hecht, *Principles of Nano-Optics* (Cambridge University Press, Cambridge, 2006).
[2] *Nanophotonics with Surface Plasmons*, edited by V. M. Shalaev and S. Kawata (Elsevier, Amsterdam, 2007).
[3] *Surface Plasmon Nanophotonics*, edited by M. L. Brongersma and P. G. Kirk (Springer, Dordrecht, 2007).
[4] I. I. Smolyaninov, A. V. Zayats, and C. C. Davis, Phys. Rev. B **56**, 9290 (1997).
[5] B. Lamprecht, A. Leitner, and F. R. Aussenegg, Appl. Phys. B: Lasers Opt. **68**, 419 (1999).
[6] A. V. Zayats, T. Kalkbrenner, V. Sandoghdar, and J. Mlynek, Phys. Rev. B **61**, 4545 (2000).
[7] T. Kawazoe, T. Shimizu, and M. Ohtsu, Opt. Lett. **26**, 1687 (2001).
[8] A. Bouhelier, M. Beversluis, A. Hartschuh, and L. Novotny, Phys. Rev. Lett. **90**, 013903 (2003).
[9] A. Nahata, R. A. Linke, T. Ishi, and K. Ohashi, Opt. Lett. **28**, 423 (2003).
[10] C. Anceau, S. Brasselet, J. Zyss, and P. Gadenne, Opt. Lett. **28**, 713 (2003).
[11] S. I. Bozhevolnyi, J. Beermann, and V. Coello, Phys. Rev. Lett. **90**, 197403 (2003).
[12] J. Beermann and S. I. Bozhevolnyi, Phys. Rev. B **69**, 155429 (2004).
[13] B. K. Canfield, S. Kujala, K. Jefimovs, J. Turunen, and M. Kauranen, Opt. Express **12**, 5418 (2004).
[14] B. K. Canfield, S. Kujala, K. Jefimovs, T. Vallius, J. Turunen, and M. Kauranen, J. Opt. A, Pure Appl. Opt. **7**, S110 (2005).
[15] M. Airola, Y. Liu, and S. Blair, J. Opt. A, Pure Appl. Opt. **7**, S118 (2005).
[16] J. Nappa, G. Revillod, I. Russier-Antoine, E. Benichou, C. Jonin, and P.-F. Brevet, Phys. Rev. B **71**, 165407 (2005).
[17] J. Nappa, I. Russier-Antoine, E. Benichou, C. Jonin, and P.-F. Brevet, Chem. Phys. Lett. **415**, 246 (2005).
[18] M. D. McMahon, R. Lopez, R. F. Haglund, E. A. Ray, and P. H. Bunton, Phys. Rev. B **73**, 041401(R) (2006).
[19] M. Breit, S. Malkmus, J. Feldman, and H. U. Danzebrink, Appl. Phys. Lett. **90**, 093114 (2007).
[20] B. K. Canfield, S. Kujala, K. Laiho, K. Jefimovs, J. Turunen, and M. Kauranen, Opt. Express **14**, 950 (2006).
[21] J. Shan, J. I. Dadap, I. Stiopkin, G. A. Reider, and T. F. Heinz, Phys. Rev. A **73**, 023819 (2006).
[22] W. Fan, S. Zhang, N.-C. Panoiu, A. Abdenour, S. Krishna, R. M. Osgood, Jr., K. J. Malloy, and S. R. J. Brueck, Nano Lett. **6**, 1027 (2006).
[23] M. W. Klein, C. Enkrich, M. Wegener, and S. Linden, Science **313**, 502 (2006).
[24] W. Fan, S. Zhang, K. J. Malloy, S. R. J. Brueck, N.-C. Panoiu, and R. M. Osgood, Opt. Express **14**, 9570 (2006).
[25] J. A. H. van Nieuwstadt, M. Sandtke, R. H. Harmsen, F. B. Segerink, J. C. Prangsma, S. Enoch, and L. Kuipers, Phys. Rev. Lett. **97**, 146102 (2006).
[26] M. D. McMahon, D. Ferrara, C. T. Bowie, R. Lopez, and R. F. Haglund, Jr., Appl. Phys. B: Lasers Opt. **87**, 259 (2007).
[27] B. K. Canfield, H. Husu, J. Laukkanen, B. Bai, M. Kuittinen, J. Turunen, and M. Kauranen, Nano Lett. **7**, 1251 (2007).
[28] M. W. Klein, M. Wegener, N. Feth, and S. Linden, Opt. Express **15**, 5238 (2007).
[29] S. Kujala, B. K. Canfield, M. Kauranen, Y. Svirko, and J. Turunen, Phys. Rev. Lett. **98**, 167403 (2007).
[30] T. Xu, X. Jiao, G. P. Zhang, and S. Blair, Opt. Express **15**, 13894 (2007).
[31] M. Zavelani-Rossi, M. Celebrano, P. Biagioni, D. Polli, M. Finazzi, L. Duò, G. Cerullo, M. Labardi, M. Allegrini, J. Grand, and P.-M. Adam, Appl. Phys. Lett. **92**, 093119 (2008).
[32] G. Bachelier, I. Russier-Antoine, E. Benichou, C. Jonin, and P.-F. Brevet, J. Opt. Soc. Am. B **25**, 955 (2008).
[33] N. Feth, S. Linden, M. W. Klein, M. Decker, F. B. P. Niesler, Y. Zeng, W. Hoyer, J. Liu, S. W. Koch, J. V. Molony, and M. Wegener, Opt. Lett. **33**, 1975 (2008).

[34] J. I. Dadap, J. Shan, K. B. Eisenthal, and T. F. Heinz, Phys. Rev. Lett. **83**, 4045 (1999).
[35] S. I. Bozhevolnyi and V. Z. Lozovski, Phys. Rev. B **61**, 11139 (2000).
[36] V. L. Brudny, B. S. Mendoza, and W. L. Mochan, Phys. Rev. B **62**, 11152 (2000).
[37] W. L. Mochan, J. A. Maytorena, B. S. Mendoza, and V. L. Brudny, Phys. Rev. B **68**, 085318 (2003).
[38] N. I. Zheludev and V. I. Emel'yanov, J. Opt. A, Pure Appl. Opt. **6**, 26 (2004).
[39] M. I. Stockman, D. J. Bergman, C. Anceau, S. Brasselet, and J. Zyss, Phys. Rev. Lett. **92**, 057402 (2004).
[40] J. Beermann, S. I. Bozhevolnyi, and V. Coello, Phys. Rev. B **73**, 115408 (2006).
[41] J. I. Dadap, J. Shan, and T. F. Heinz, J. Opt. Soc. Am. B **21**, 1328 (2004).
[42] A. V. Andreev, A. A. Korneev, L. S. Mukina, M. M. Nazarov, I. R. Prudnikov, and A. P. Shkurinov, Phys. Rev. B **74**, 235421 (2006).
[43] B. K. Canfield, S. Kujula, K. Jefimovs, Y. Svirko, J. Turunen, and M. Kauranen, J. Opt. A, Pure Appl. Opt. **8**, S278 (2006).
[44] L. Cao, N. C. Panoiu, and R. M. Osgood, Jr., Phys. Rev. B **75**, 205401 (2007).
[45] M. Finazzi, P. Biagioni, M. Celebrano, and L. Duò, Phys. Rev. B **76**, 125414 (2007).
[46] G. Y. Panasyuk, J. C. Schotland, and V. A. Markel, Phys. Rev. Lett. **100**, 047402 (2008).
[47] A. Liebsch, *Electronic Excitations at Metal Surfaces* (Plenum, New York, 1997).
[48] T. F. Heinz, in *Nonlinear Surface Electromagnetic Phenomena*, edited by H. Ponath and G. Stegeman (Elsevier, Amsterdam, 1991), p. 353.
[49] A. Taflove and S. C. Hagness, *Computational Electrodynamics: The Finite-Difference Time-Domain Method* (Artech House, Boston, 2000).
[50] J. Rudnick and E. A. Stern, Phys. Rev. B **4**, 4274 (1971).
[51] J. Rudnick and E. A. Stern, in *Polaritons*, edited by E. Burstein (Plenum, New York, 1974).
[52] J. E. Sipe, V. C. Y. So, M. Fukui, and G. I. Stegeman, Solid State Commun. **34**, 523 (1980).
[53] J. E. Sipe, V. C. Y. So, M. Fukui, and G. I. Stegeman, Phys. Rev. B **21**, 4389 (1980).
[54] M. Corvi and W. L. Schaich, Phys. Rev. B **33**, 3688 (1986).
[55] W. L. Schaich and A. Liebsch, Phys. Rev. B **37**, 6187 (1988).
[56] A. Liebsch and W. L. Schaich, Phys. Rev. B **40**, 5401 (1989).
[57] D. Amarie, N. D. Rawlinson, W. L. Schaich, B. Dragnea, and S. C. Jacobson, Nano Lett. **5**, 1227 (2005).
[58] T.-D. Onuta, M. Waegele, C. C. DuFort, W. L. Schaich, and B. Dragnea, Nano Lett. **7**, 557 (2007).
[59] A. Hohenau, J. R. Krenn, F. J. Garcia-Vidal, S. G. Rodrigo, L. Martin-Moreno, J. Beermann, and S. I. Bozhevolnyi, J. Opt. A, Pure Appl. Opt. **9**, S366 (2007).
[60] S. G. Rodrigo, F. J. Garcia-Vidal, and L. Martin-Moreno, Phys. Rev. B **77**, 075401 (2008).
[61] W. L. Schaich, G. Schider, J. R. Krenn, A. Leitner, F. R. Aussenegg, I. Puscasu, B. Monacelli, and G. Boreman, Appl. Opt. **42**, 5714 (2003).
[62] G. Schider, J. R. Krenn, A. Hohenau, H. Ditlbacher, A. Leitner, F. R. Aussenegg, W. L. Schaich, I. Puscasu, B. Monacelli, and G. Boreman, Phys. Rev. B **68**, 155427 (2003).
[63] A. Hohenau, J. R. Krenn, G. Schider, H. Ditlbacher, A. Leitner, F. R. Aussenegg, and W. L. Schaich, Europhys. Lett. **69**, 538 (2005).
[64] T. Zentgraf, T. P. Meyrath, A. Seidel, S. Kaiser, H. Giessen, C. Rockstuhl, and F. Lederer, Phys. Rev. B **76**, 033407 (2007).
[65] T. W. Ebbesen, H. L. Lezec, H. F. Ghaemi, T. Thio, and P. A. Wolff, Nature (London) **391**, 667 (1998).
[66] W. L. Barnes, A. Dereux, and T. W. Ebbesen, Nature (London) **424**, 824 (2003).
[67] C. Genet and T. W. Ebbesen, Nature (London) **445**, 39 (2007).
[68] K. J. Klein Koerkamp, S. Enoch, F. B. Segerink, N. F. van Hulst, and L. Kuipers, Phys. Rev. Lett. **92**, 183901 (2004).
[69] K. L. van der Molen, K. J. Klein Koerkamp, S. Enoch, F. B. Segerink, N. F. van Hulst, and L. Kuipers, Phys. Rev. B **72**, 045421 (2005).