Theory of extraordinary light transmission through arrays of subwavelength slits

B. Sturman and E. Podivilov
Institute of Automation and Electrometry, Russian Academy of Sciences, 630090 Novosibirsk, Russia

M. Gorkunov
Institute of Crystallography, Russian Academy of Sciences, Leninskii Prospekt 59, 119333 Moscow, Russia
and Department of Mathematics, University of Strathclyde, 26 Richmond Street, Glasgow G1 1XH, United Kingdom
(Received 3 July 2007; revised manuscript received 18 November 2007; published 7 February 2008)

We propose a self-consistent theory of the extraordinary light transmission through periodic arrays of subwavelength holes in metals. Its basis is an expansion of the light fields in terms of exact eigenmodes—propagating, evanescent, and anomalous—investigated in our recent paper and matching at the interfaces using the exact boundary conditions. An excellent convergence of this expansion has allowed us to decompose the anomalous transmission phenomenon into elementary parts and to investigate the characteristic parametric dependences. Transmission properties of a single interface play a key role in our theory in the subwavelength range. They include the coefficient of energy transmission into the propagating mode and the phases of the reflected and transmitted waves. These key parameters possess remarkable resonant dependences on the wavelength of light; they are sensitive to the size of the holes and rather insensitive to weak losses. The surface-plasmon-related features of the above characteristics are established. Transmission properties of a slab are expressed by the single-interface parameters, the phase incursion for the propagating mode, and the propagating losses.

DOI: 10.1103/PhysRevB.77.075106 PACS number(s): 73.20.Mf, 42.25.Bs, 42.25.Fx, 42.70.Qs

I. INTRODUCTION

Discovery of the extraordinary light transmission (ELT) through subwavelength slits and holes (single and periodically arranged) in metals[1,2] gave rise to an explosion of experimental and theoretical studies (see, e.g., Refs. [3–11] and references therein). These studies are motivated by both the brilliance and fundamental character of this phenomenon and its possible applications employing the ultrastrong light constraint. While many particular features of the ELT are understood, it still remains mysterious in many respects. The great challenge for theory is that the basics for ELT—the macroscopic Maxwell equations—are commonly accepted.

It is clear nowadays that the ideal metal paradigm is generally insufficient for an analysis of the ELT. Refusal of this paradigm in favor of real metals is usually associated with the excitation of the surface plasmons.[3,12] For a real metal, the optical dielectric permittivity $\varepsilon_m=\varepsilon_m'+i\varepsilon_m''$ consists of a negative real part ε_m' and a positive imaginary part ε_m'' responsible for light absorption. The condition for the existence of the surface plasmon on a flat air-metal interface, $\varepsilon_m'<-1$, $\varepsilon_m''\ll|\varepsilon_m'|$, is fulfilled for many metals, including silver and gold.[13,14]

The ELT is often associated with the resonant excitation of the surface plasmons on the opposite faces of a perforated metallic slab. This physical picture implies the resonant Fabry-Pérot enhancement of the light transmission through the slab, at least in the one-dimensional (1D) case where the lowest propagating mode survives in the subwavelength limit, and, simultaneously, an efficient excitation of the plasmon at the input interface.[1,4,5] The mechanism of such a plasmon excitation and the nature of evanescent coupling in the two-dimensional (2D) case, where the propagating modes are usually absent, are far from evident.[15–17] Moreover, arguments against the plasmon mechanism of the ELT are also known in the literature.[6,7]

The controversy in the theoretical studies has serious reasons. Calculations of observable transmission characteristics are based either on complicated numerical methods or on oversimplified analytical models. In the first case, numerous factors affecting the ELT work together and cannot be separated; the possibilities for an analysis of the transmission characteristics on the variable parameters (light wavelength, period of the structure, hole and/or slit size, slab thickness, ε_m', ε_m'', etc.) are restricted; and the essential details of numerical simulations often remain unknown. Simplified models imply noncontrollable assumptions diminishing the practical impact. Furthermore, the term "surface plasmon," which is well defined and transparent for flat dielectric-metal interfaces, becomes less certain for "holey" interfaces. For this reason, the plasmon interpretations of the ELT often have no solid theoretical grounds.

The most physical and controllable calculation methods are based on the knowledge of the eigenmodes for the hole and/or slit structures and the use of modal expansions.[18,19] Surprisingly, the properties of such metal-dielectric waveguiding structures (single holes and photonic crystals) remained underinvestigated and even misrepresented, especially in the actual subwavelength limit. When describing the eigenmode-based calculation methods, it was claimed that the eigenvalues are real in the absence of losses ($\varepsilon_m''=0$) and that an analogy with quantum mechanics holds true.[6,19,20] It turned out, however, that this basic ingredient of theory cannot be applied to metal-dielectric structures. In addition to the well-known propagating and evanescent modes, a sequence of new modes with complex values of the propagating constant exists even for $\varepsilon_m''=0$.[21,22] Missing these anomalous modes makes the set of eigenfunctions incomplete and

leads to a divergence of the calculation procedure.[21]

We have developed a theory of the ELT for periodic arrays of subwavelength slits in metals. It combines analytical and numerical methods and incorporates a number of known ingredients, such as modal expansions, the use of boundary conditions, and Fabry-Perot resonance. However, it differs essentially from the other known theories by the combination of the following key elements.

(1) We employ the exact eigenfunctions and eigenvalues.[22] Taking into account the anomalous modes makes the set of eigenfunctions complete and ensures a quick convergence of our numerical procedure. This makes our numerical tool highly playable.

(2) We use the exact boundary conditions at all metal-dielectric interfaces. In conjunction with the first issue, it makes our numerical results exact within the conventional concept of the local permittivity.

(3) We calculate the mode structure for both $\varepsilon_m'' \neq 0$ and $\varepsilon_m''=0$; i.e., we monitor the influence of the light absorption on the observable characteristics of the ELT.

(4) We carefully investigate the fundamental transmission-reflection properties of a single-interface air–1D photonic crystal. The influence of absorption is very weak here for $\varepsilon_m'' \ll |\varepsilon_m'|$; i.e., the transmission-reflection properties can be treated as lossless. Two parameters are of importance for the ELT in the subwavelength case: the energy transmission coefficient and a combination of the phases of the reflected and transmitted waves. Dependences of these parameters on the wavelength λ, ε_m', and the wall/slit ratio reveal fundamental features and clarify the role of the surface plasmons.

(5) Using the symmetry properties, the transmittance of a perforated slab is expressed by the above single-interface characteristics and the propagation losses. The final relation represents an advanced version of the single-mode model of Ref. 5. It allows us to classify and describe the slab transmittance peaks in a simple manner. The simplified results are supported by rigorous calculations.

In short, we have rectified considerably the known approaches to the ELT and made the theoretical scheme not only accurate but also highly playable. Our findings extend the main known results in the field but, typically, do not contradict them. Specific descriptions of our results can be found below.

While the theory of light transmission via the propagating modes is devoted to the 1D case, it can, to some extent, be applied to the 2D case. This important possibility is considered in Sec. VI.

II. BASIC RELATIONS

Figure 1(a) shows the simplest periodic structure (1D photonic crystal) that is relevant to the ELT. It is uniform along the propagation coordinate z. The optical permittivity $\varepsilon(x)$ takes the values $\varepsilon_m=\varepsilon_m'+i\varepsilon_m''$ and $\varepsilon_d=\varepsilon_d'>0$ in the metal and dielectric parts. A value $\varepsilon_d \neq 1$ accounts for filling the slits with a transparent dielectric. The wall and slit widths are x_m and x_d, and the period of the structure is $x_0=x_d+x_m$. The ELT occurs for the TM polarization when the only nonzero

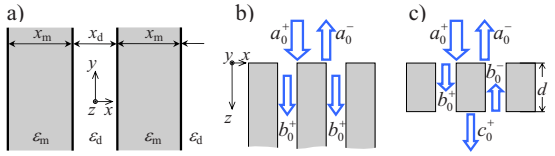

FIG. 1. (Color online) Geometrical schemes for a periodic 1D structure (a), an air-PC interface (b), and slab transmission (c). The permittivities of the dielectric and metal parts are ε_d and ε_m. Only zero-order propagating modes are shown for (b) and (c).

component of the magnetic field amplitude is $H_y \equiv H(x,z)$. To find the eigenmodes, we use the ansatz $H=h(x)\exp(i\beta z+i\kappa x)$, where β is the propagation constant, κ is the Bloch wave vector, and h is an x_0-periodic function of x. Restricting ourselves to the case of normal incidence, [see Figs. 1(b) and 1(c)] we set $\kappa=0$. Then, Maxwell equations are obtained for the eigenfunction $h(x)$,

$$\hat{L}h = \beta^2 h, \quad \hat{L} = \frac{d^2}{dx^2} + \varepsilon k_0^2, \tag{1}$$

where $k_0=2\pi/\lambda$ is the vacuum wave vector. This sets an eigenproblem for the operator $\hat{L}$ with β^2 being the eigenvalue. The nonzero components of the light electric field are expressed by H: $E_x=\beta H/\varepsilon k_0$ and $E_z=(i/\varepsilon k_0)dH/dx$. The quantities H, E_z, and εE_x are continuous at the metal-dielectric interfaces [see Fig. 1(a)] which expresses the exact boundary conditions.

For *lossless dielectric* structures, the operator $\hat{L}$ is Hermitian, the eigenvalues β^2 are real, and an analogy with quantum mechanics is valid. The situation changes when the structure includes *lossless metallic* parts, $\varepsilon_m<0$. To see it, we multiply Eq. (1) for h by h^*/ε (the asterisk means complex conjugation) and integrate over the period x_0. Integrating the left-hand side by parts, using the boundary conditions, and taking the imaginary part, we see that β^2 is real only for $\int \varepsilon^{-1}(x)|h(x)|^2 dx \neq 0$. In the dielectric case, $\varepsilon(x)>0$, this condition is always satisfied. However, in the metal-dielectric case, where $\varepsilon(x)$ is sign alternating, the integral can be zero. Then, the eigenvalue β^2 is complex; i.e., we have an anomalous mode.[22]

The allowed values of β^2 satisfy the known dispersion equation,[11,22,23] which follows from Eq. (1). Each eigenvalue corresponds (within a normalization factor) to a single eigenfunction $h(x)$. The eigenfunctions are either even or odd in x if the point $x=0$ is set at a slit center. Only the even eigenfunctions and the corresponding "even" values of β^2 are necessary for the description of the ELT in the case of normal incidence.

Within a half-period $0<x<x_0/2$, an even eigenfunction can be written as

$$h = \begin{cases} \cos(p_m x_m/2)\cos(p_d x) & (x < x_d/2) \\ \cos(p_d x_d/2)\cos[p_m(x-x_0/2)] & (x_d/2 < x), \end{cases} \tag{2}$$

where $p_{d,m}=(\varepsilon_{d,m}k_0^2-\beta^2)^{1/2}$. The dispersion equation for the determination of the even values of β^2 is[22]

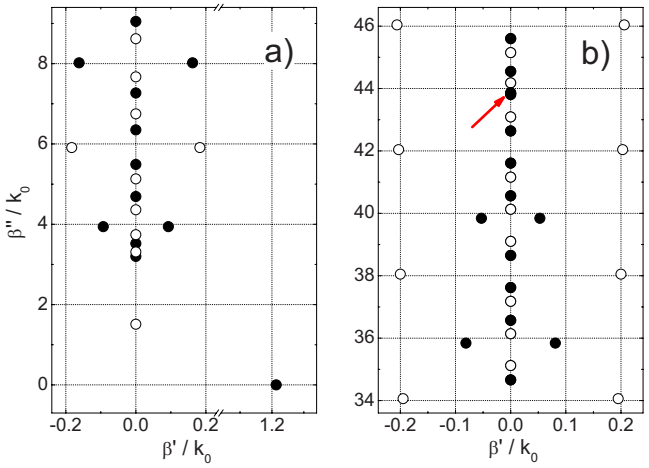

FIG. 2. (Color online) Spectrum of β for $\varepsilon_m=-10$, $x_m=\lambda$, and $x_m/x_d=4$ for modest (a) and large (b) values of β''/k_0. The filled and open dots correspond to the even and odd modes, respectively. The arrow indicates two closely situated evanescent roots originating from the confluence of a pair of anomalous roots.

$$\frac{p_d}{\varepsilon_d}\tan\left(\frac{p_d x_d}{2}\right)+\frac{p_m}{\varepsilon_m}\tan\left(\frac{p_m x_m}{2}\right)=0, \tag{3}$$

where the permittivity of the metal, ε_m, is generally complex.

Each normalized eigenvalue β^2/k_0^2 depends on four dimensionless parameters. It is convenient to choose them as λ/x_0, x_m/x_d, ε_m, and ε_d. Then, the following scaling relation is valid:

$$\beta^2\left(\frac{\lambda}{x_0},\frac{x_m}{x_d},\varepsilon_m,\varepsilon_d\right)=\varepsilon_d\beta^2\left(\frac{\lambda}{\varepsilon_d x_0},\frac{x_m}{x_d},\frac{\varepsilon_m}{\varepsilon_d},1\right). \tag{4}$$

It shows that the relative permittivity $\varepsilon_m/\varepsilon_d$ and the effective period $\sqrt{\varepsilon_d}x_0$ determine the spectrum of the eigenvalues when filling the slits with a transparent dielectric. Since dielectrics (fluids) with the refractive index $\sqrt{\varepsilon_d}>2$ are known and "optofluidics" is a rapidly developing area,[24,25] the relative permittivities $\varepsilon_m/\varepsilon_d\sim-1$ should be accessible. Without loss of generality, we can now restrict ourselves to the case of $\varepsilon_d=1$ (air slits).

The algebraic equation [Eq. (3)] can be solved numerically. The opposite values of the propagation constant $\beta=\pm\sqrt{\beta^2}$ correspond to waves propagating in the $\pm z$ directions. The dots in Fig. 2(a) show the values of $\beta\equiv\beta'+i\beta''$ in the range $\beta''/k_0<9$ for $x_m=4x_d=\lambda$ and $\varepsilon_m=-10$. The last parameter is representative for silver at $\lambda\approx0.5$ μm. The filled/open dots correspond to the even/odd modes. We have a single (even) propagating mode with $\beta/k_0\approx1.2$, a sequence of evanescent modes with $\beta'=0$, and a sequence of anomalous modes with $\beta'\neq0$ and $\beta''/k_0>\sqrt{|\varepsilon_m|}$. The symmetric pairs of anomalous roots correspond to the conjugate values of β^2. About 25% of modes in Fig. 2(a) are anomalous. With weak losses of silver ($\varepsilon_m''\approx0.3$) taken into account, the roots experience small displacements.[22]

With decreasing slit width x_d, the single propagating mode survives and its effective refractive index β/k_0 increases as $1/x_d$. The vertical and horizontal distances between the anomalous roots in Fig. 2 also increase. Increasing wall width x_m weakly influences the propagating mode and the anomalous modes. These modes actually represent the properties of a single slit; in the limit $x_m\to\infty$, they are all $|x|$ localized. The distance between the evanescent roots in the range $\beta''/k_0>\sqrt{\varepsilon_m}$ decreases as $1/x_m$; in the limit $x_m\to\infty$, these roots form a continuous spectrum of a single slit.

With decreasing ratio λ/x_0, the second even propagating mode appears eventually. For $x_m/x_d=4$ and $\varepsilon_m'=-10$, it occurs for $\lambda/x_0<1/2$.

It is remarkable that the sequence of the even anomalous modes is finite for $x_m/x_d<|\varepsilon_m|$ and infinite for $x_m/x_d>|\varepsilon_m|$.[22] The sequence of the odd anomalous modes is always infinite. This is illustrated by Fig. 2(b). Note that the sequence of even anomalous modes remain very long already for $x_m/x_d=5$ and $|\varepsilon_m|=10$.

Changing the spectral behavior of the anomalous modes at $x_m/x_d=|\varepsilon_m|$ has a close relation with the properties of our structure in the macroscopic limit $\lambda\gg2\pi\sqrt{\varepsilon_m}x_0$, where the metal walls are almost transparent. The effective permittivity in this limit is $\varepsilon_{eff}=|\varepsilon_m|x_0/(|\varepsilon_m|x_d-x_m)$.[26] It is positive for $x_m/x_d<|\varepsilon_m|$, negative for $x_m/x_d>|\varepsilon_m|$, and infinite at $x_m/x_d=|\varepsilon_m|$.

III. SINGLE-INTERFACE TRANSMISSION PROPERTIES

Now, we consider the simplest transmission/reflection problem. Let an x-polarized plane wave be incident normally onto our periodic metal-dielectric structure occupying the semispace $z>0$ [see Fig. 1(b)]. It is necessary to determine the transmission and reflection properties of the interface $z=0$. The formulated problem is similar to the basic problem of light transmission through an interface between two media with different optical permittivities, which is described by the Fresnel relations. It is highly useful for an understanding of the physics of ELT.

A. Calculation procedure

In the regions $z<0$ and $z>0$, we use the following exact and explicit modal expansions for H:

$$H=\begin{cases}a_0^+e^{ik_0z}+\displaystyle\sum_{n=0}^{\infty}2a_n^-\cos(nKx)e^{q_nz}, & z<0\\ \displaystyle\sum_{\nu=0}^{\infty}b_\nu^+h^\nu(x)e^{i\beta_\nu z}, & z>0.\end{cases} \tag{5}$$

In the first expansion, a_0^+ is the amplitude of the incident wave, $K=2\pi/x_0$ is the grating vector, and the exponent q_n is $(n^2K^2-k_0^2)^{1/2}$ for $n^2K^2>k_0^2$ and $-i(k_0^2-n^2K^2)^{1/2}$ for $k_0^2>n^2K^2$. Real and imaginary values of q_n correspond to the evanescent and propagating waves in air. With increasing period, new diffraction orders appear in air at $x_0=n\lambda$. Only a single propagating wave (reflected wave) with $n=0$ and $q_0=-ik_0$ is allowed for $\lambda/x_0>1$. In the second expansion, the index ν numerates the even modes in the order of increasing β''. The propagating modes travel and the evanescent and anomalous modes decay in the $+z$ direction. The eigenfunctions and

eigenvalues employed account for the exact boundary conditions at the vertical metal walls. The amplitudes of the modes a_n^- and b_ν^+ are to be found.

Employing the exact boundary conditions at $z=0$, i.e., continuity for H and $\varepsilon^{-1}\partial H/\partial z$, we come to the set of linear algebraic equations for b_ν^+:

$$\sum_{\nu=0}^{\infty} b_\nu^+[h_n^\nu - i\beta_\nu q_n^{-1}(h^\nu/\varepsilon)_n] = 2a_0^+\delta_{n0}, \tag{6}$$

where $h_n^\nu = \langle h^\nu(x)\cos(nKx)\rangle_{x_0}$ and $(h^\nu/\varepsilon)_n = \langle h^\nu(x)\cos(nKx)/\varepsilon(x)\rangle_{x_0}$ are the nth Fourier components of the periodic functions $h^\nu(x)$ and $h^\nu(x)/\varepsilon(x)$; the brackets mean averaging over the period x_0. The Fourier components are easily expressible by β_ν and n using Eq. (2). When the set b_ν^+ is found, the amplitudes a_n^- can be calculated from the following exact relations:

$$2a_0^- = -a_0^+ + \sum_{\nu=0}^{\infty} b_\nu^+ h_0^\nu,$$

$$a_n^- = \sum_{\nu=0}^{\infty} b_\nu^+ h_n^\nu \quad (n \neq 0). \tag{7}$$

A numerical routine with the truncation of set (6) at $\nu_{max}=n_{max}=N\gg 1$ is used to calculate b_ν^+. The procedure converges perfectly well with increasing N when *all eigenmodes*—propagating, evanescent, and anomalous—are taken into account. However, it does not converge when the anomalous modes are ignored: The amplitudes and the transmission/reflection characteristics experience, in this case, strong nonvanishing oscillations with increasing N so that the true characteristics cannot be determined even approximately. This proves (i) that the set of the eigenfunctions is complete with and incomplete without the anomalous modes and (ii) that these modes are crucial for getting true results. Most of the subsequent calculations are performed for the truncation number $N=48$. However, apart from fine features, the results are already the same for $N=20$.

The energy conservation law has been systematically verified during our calculation procedure. It is fulfilled with a very high accuracy ($\lesssim 0.1\,\%$) when the anomalous modes are taken into account.

Our procedure allows us also to describe the dispersion properties of the holey-surface plasmon (HSP) in accordance with the general receipt (see Refs. 27 and 28 and references therein). Represent the solution of set (6) as $b_\nu^+ = D_\nu/D$, where D is the determinant of this set and D_ν is the corresponding cofactor. The determinant D, considered as a function of the complex variable $\zeta=\zeta'+\zeta''$, such that $\zeta'=\lambda/x_0$, possesses a zero point ζ_0 near the real axis. The real part ζ_0' gives the resonant wavelength $\lambda_{HSP}=\zeta_0' x_0$, while the ratio ζ_0'/ζ_0'' gives the quality factor of the HSP. According to the definition, λ_{HSP} is the wavelength of the incident light whose frequency equals the frequency of the plasmon with the wave vector $K=2\pi/x_0$. At $\varepsilon_m''=0$, the nonzero value of ζ_0'' is due to the radiative losses into air and the slits. In the limit $x_d\to 0$, we expect $\zeta_0'=\sqrt{|\varepsilon_m|/(|\varepsilon_m|-1)}$ and $\zeta_0''=0$, i.e., the flat-surface plasmon (FSP).

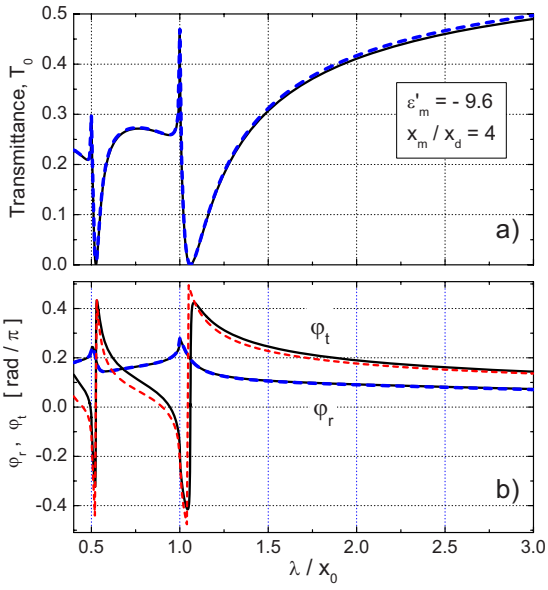

FIG. 3. (Color online) The single-interface transmittance T_0 (a) and phases $\varphi_{t,r}$ (b) versus λ/x_0 for $\varepsilon_m'=-9.6$, $x_m/x_d=4$, and $N=48$; the solid and dotted lines are plotted for $\varepsilon_m''=0.3$ and 0.

B. Results

Consider first the results for the transmittance T_0 defined as the energy-flux fraction transmitted to the single propagating mode at $z=0$. This definition is useful even for $\varepsilon''\neq 0$ when the energy flux for the propagating mode decays as $\exp(-2\beta_0'' z)$. As for the condition for existence of only one propagating mode, it is fulfilled in the subwavelength range, $\lambda/x_0>1$ (which is of our prime interest) and even well beyond this range, unless the walls are thinner than the skin depth.[22] For sufficiently small ratios λ/x_0, the excitation of two or more propagating modes has to be taken into account. The limiting case of independent subwavelength slits, $x_0 \gg \lambda > x_d$ (with many diffraction orders in air involved), is beyond our study.

The solid and dashed lines in Fig. 3(a) show the dependence $T_0(\lambda/x_0)$ in the lossless ($\varepsilon_m''=0$) and lossy ($\varepsilon_m''=0.3$) cases for $\varepsilon_m'=-9.6$, $x_m/x_d=4$, and $N=48$. These lines almost coincide. The influence of light absorption on the transmission properties of the interface air–photonic crystal (PC) is thus very small for $\varepsilon_m'' \ll |\varepsilon_m'|$. Only a single propagating mode is allowed in the shown range of λ/x_0.

The function $T_0(\lambda/x_0)$ possesses very narrow spikelike peaks at $\lambda/x_0=1/s$ ($s=1,2,\ldots$), which correlates with the positions of the Rayleigh-Wood anomalies for reflection gratings[11,29,30] and corresponds to the phase-matching condition $k_0=sK$. For convenience, we will refer to these peaks as to Rayleigh-Wood (RW) peaks/resonances. The strongest peak corresponds to $s=1$. On a larger scale, the peaks are strongly asymmetric. Often, similar features of resonances in complicated systems are attributed to the Fano effect.[31]

Clearly, the shown RW peaks/resonances are precisely pinned to the points where new diffraction orders are opening. This change apparently has as strong an impact on $T_0(\lambda/x_0)$ as the change of the symmetry during the phase transition on the macroscopic properties of the medium. It can also be treated as the consequence of bifurcations at the points $\lambda/x_0=1/s$.

What is the role of the surface plasmon in these spectral anomalies? One might suggest that the resonant value $(\lambda/x_0)_{HSP}$ for the holey-surface plasmon is 1 for the chosen parameters, instead of the flat-surface plasmon value $(\lambda/x_0)_{FSP}=\sqrt{|\varepsilon_m'|/(|\varepsilon_m'|-1)}\simeq 1.06$. However, our analysis of the zero point of the determinant $D(\zeta)$ (see Sec. III A) has shown that it is not the case; the value of $(\lambda/x_0)_{HSP}$ is $\simeq 0.93$ for the parameters of Fig. 3. As for the spectral anomalies of T_0, which practically coincide with those of $|b_0^+|^2=|D_0/D|^2$, they are determined by the sharp spectral dependences of both D and D_0. For the chosen values of x_m/x_d and ε_m, the peaks of $T_0(\lambda/x_0)$ thus *cannot* be attributed to the resonant excitation of the holey-surface plasmon. Generally, the existence of such a surface mode does not ensure its effective resonant excitation.

Returning to Fig. 3(a), we see that the transmittance T_0 turns virtually to zero slightly to the right of each RW peak. For the main peak, it occurs at $\lambda/x_0\simeq 1.06$, which is very close to the FSP resonance $(\lambda/x_0)_{FSP}$. A more careful analysis has shown that the minimum value of T_0 is exactly zero only in the lossless case, $\varepsilon_m''=0$; for $\varepsilon_m''=0.3$, it is very small, $T_0^{min}\lesssim 10^{-4}$. These observations are in line with the results of Refs. 7 and 28.

With further increase of λ/x_0, the transmittance T_0 grows and tends slowly to $\simeq 0.75$. This value corresponds to the continuous-medium limit, i.e., to the Fresnel relations for an interface between media with $\varepsilon=1$ and $\varepsilon_{eff}=|\varepsilon_m|x_0/(|\varepsilon_m|x_d-x_m)\simeq 8.57$.

Another useful characteristic of the interface is the ratio $R_0=|a_0^-|^2/|a_0^+|^2$. In the subwavelength range, when the diffraction orders in air are absent, it gives the reflection coefficient. In this range, we have $R_0=1-T_0$ in the lossless limit, $\varepsilon_m''=0$. With light absorption taken into account, the difference $1-T_0-R_0$ is nonzero. It expresses the energy losses in the skin-deep near-surface layer. As one can expect, these "surface losses" are very weak for $\varepsilon_m''\ll|\varepsilon_m'|$. In particular, we have $1-T_0-R_0\lesssim 0.03$ in the silverlike case, $\varepsilon_m'=-10$, and $\varepsilon_m''=0.3$.

The spectral behavior of the phases of the transmitted and reflected waves, $\varphi_t=\arg(a_0)$ and $\varphi_r=\arg(b_0)$, is also important for what follows. Figure 3(b) gives dependences of these phases for the lossless and lossy cases. In the lossless case, the phase φ_t experiences π jumps at the points where $T_0(\lambda/x_0)$ turns to zero. These jumps occur when the complex amplitude of the propagating mode a_0 crosses the zero point. Except for the singular points, the phase φ_t decreases monotonously with λ/x_0. The most rapid decrease occurs near the peak of T_0. The phase φ_r remains quasiconstant, experiencing only small and sharp bursts at $\lambda/x_0=1,1/2,\ldots$. The influence of weak losses of silver ($\varepsilon_m''=0.3$) on φ_r is practically invisible. The influence on φ_t is small but noticeable for $\lambda/x_0>1$; it is seen as a smoothing of the jumps.

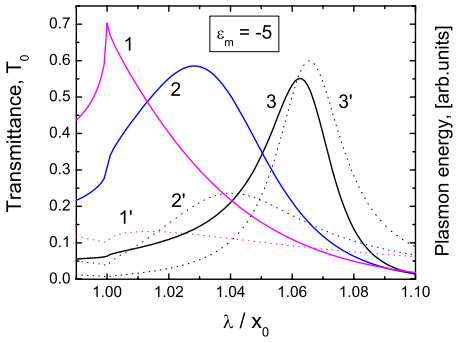

FIG. 4. (Color online) The single-interface transmittance T_0 (solid lines) and plasmon energy $|A_p|^2$ (dotted lines) versus λ/x_0 for $\varepsilon_m=-5$. The pairs of curves $1,1'$, $2,2'$, and $3,3'$ correspond to $x_m/x_d=4$, 10, and 20, respectively. The positions of the FSP and HSP resonances are $(\lambda/x_0)_{FSP}\simeq 1.118$ and $(\lambda/x_0)_{HSP}\simeq 1.063$.

Importantly, the described properties of the peaks of T_0 occur only for $x_m/x_d<|\varepsilon_m'|$, i.e., for $\varepsilon_{eff}>0$. In the opposite case, $x_m/x_d>|\varepsilon_m'|$, the spectral behavior of T_0 is *qualitatively different*. The peaks progressively shift to the right with increasing ratio of $x_m/x_d|\varepsilon_m'|$, have no kinks, and become bell shaped; they cannot be referred to as the RW peaks anymore. All this is illustrated in Fig. 4. One sees also that the height of the main peak decreases slowly with increasing ratio x_m/x_d, so that the peak transmittance T_0^{max} remains very high for pretty narrow slits.

While the position of the peak $(\lambda/x_0)_{max}$ moves toward the FSP resonance $(\lambda/x_0)_{FSP}=\sqrt{|\varepsilon_m|/(|\varepsilon_m|-1)}$ with increasing $x_m/x_d|\varepsilon_m'|$, the position of the minimum transmittance $(\lambda/x_0)_{min}$ remains close to $(\lambda/x_0)_{FSP}$. Our special examination has shown that $(\lambda/x_0)_{FSP}$ is *always smaller* than $(\lambda/x_0)_{min}$. The distance between these spectral points essentially depends on ε_m' and also on the ratio x_m/x_d. This is illustrated by Fig. 5. For $|\varepsilon_m'|>9$, this distance is very small for any reasonable value of x_m/x_d. In the range $|\varepsilon_m'|\lesssim 6$, which is relevant, e.g., to the case of Ni (Ref. 14) or to filling the slits with a transparent dielectric, the distance between $(\lambda/x_0)_{min}$ and $(\lambda/x_0)_{FSP}$ becomes noticeable and increases gradually with decreasing x_m/x_d. Hopefully, these data will finalize the theoretical discussion about the coincidence of $(\lambda/x_0)_{min}$ and $(\lambda/x_0)_{FSP}$ for the basic 1D structure in question.[7,28]

Where is the HSP resonance for $x_m/|\varepsilon_m|x_d>1$? Line 5 in Fig. 5 shows the dependence of $(\lambda/x_0)_{HSP}$ on $|\varepsilon_m|$ for $x_m/x_d=9$. At $|\varepsilon_m|\simeq 7.5$ ($x_m/|\varepsilon_m|x_d\simeq 1.2$), it crosses the line $\lambda/x_0=1$. For smaller $|\varepsilon_m|$, we have $(\lambda/x_0)_{max}\simeq(\lambda/x_0)_{HSP}$. The solid and dashed lines of the inset show the dependences of $(\lambda/x_0)_{HSP}$ and $(\lambda/x_0)_{max}$ on the ratio x_m/x_d for $\varepsilon_m=-5$. They almost coincide for $x_m/|\varepsilon_m|x_d\gtrsim 2$ being far from the FSP resonance, $(\lambda/x_0)_{FSP}\simeq 1.118$. With increasing $|\varepsilon_m|$, the separation between $(\lambda/x_0)_{\max}$ and $(\lambda/x_0)_{HSP}$ grows slightly but remains within the HSP resonance width. Thus, the crossover between the RW and HSP resonances occurs, and the peak of $T_0(\lambda/x_0)$ can be attributed to the *resonant excitation* of the HSP for sufficiently large values of $x_m/|\varepsilon_m|x_d$.

To gain more insight into the plasmon implications, we have calculated the amplitude of the flat-surface plasmon A_p

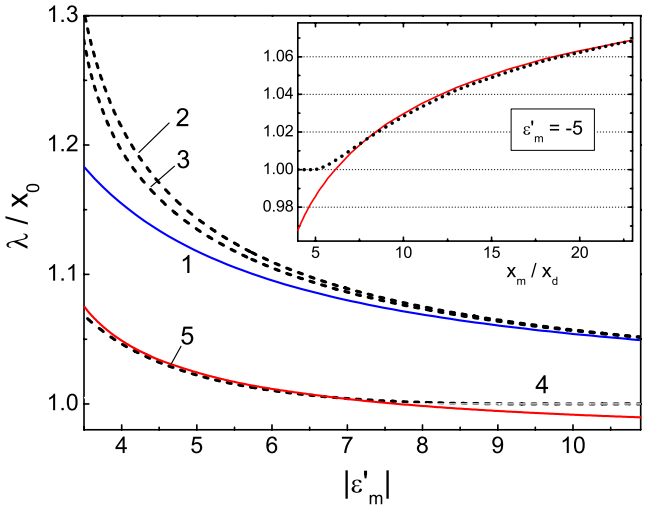

FIG. 5. (Color online) Dependences of $(\lambda/x_0)_{FSP}$ (solid line 1), $(\lambda/x_0)_{min}$ (lines 2 and 3), $(\lambda/x_0)_{max}$ (line 4), and $(\lambda/x_0)_{HSP}$ (solid line 5) on $|\varepsilon_m|$ for the single interface depicted in Fig. 1(b). Line 2 and lines 3, 4, and 5 are plotted for $x_m/x_d=4$ and 9, respectively. In the inset, the solid and dotted lines show $(\lambda/x_0)_{HSP}$ and $(\lambda/x_0)_{max}$ versus x_m/x_d for $\varepsilon_m=-5$.

at a metal part of the interface $z=0$ from the known distribution $H(x,z)$ (see the Appendix for details). A similar idea of calculation has been used in Ref. 8. The dotted lines $1'$, $2'$, and $3'$ in Fig. 4 show the spectral dependences of $|A_p|^2$ for $|\varepsilon_m|=5$ and $x_m/x_d=4$, 10, and 20. For $x_m/x_d=4$ and 10, the spectral maxima of $|A_p|^2$ are uncoupled from the peaks of T_0 and not pronounced. Comparing lines 3 and $3'$, plotted for $x_m/x_d=20$, one sees, however, that the peak of $|A_0|^2$ possesses almost the same position and width as the peak of T_0. This suggests again that for a sufficiently large ratio of $x_m/|\varepsilon_m|x_d$, when the disturbing effect of the slits is small and the HSP is locally (at the metal parts) close to the FSP, the peak of T_0 is due to the plasmon resonance.

The last point to be mentioned is the peculiarities of T_0, which are expected when decreasing $|\varepsilon'_m|$. They are closely related to the so-called corner singularities. As is known, the magnetic field H oscillates infinitely fast and the amplitudes E_x and E_z become infinitely large when approaching the 90° metal corners if ε_m lies in the range $[-3,-1/3]$.[32,33] This interval, as mentioned, is achievable when filling the slits with a suitable dielectric. Taking into account the losses is crucial here to ensure the physical results. While the indicated interval and the relevant complications are beyond the scope of this paper, we can approach the critical value $\varepsilon_m=-3$ from below. Figure 6 shows the dependence of the peak value of $T_0(\lambda/x_0)$ on the ratio x_m/x_d for three real values of ε_m. Here, the optimum ratio is $x_m/x_d\sim|\varepsilon_m|$ and the peak transmittance T_0^{peak} increases with decreasing $|\varepsilon_m|$, which can be considered as a positive effect. The convergence of our calculation procedure worsens for $|\varepsilon_m|\to 3$.

IV. SIMPLIFIED DESCRIPTION OF THE SLAB TRANSMISSION

With the results of Sec. III, it is possible to derive a useful relation for the slab transmittance T_Σ. It fully accounts for

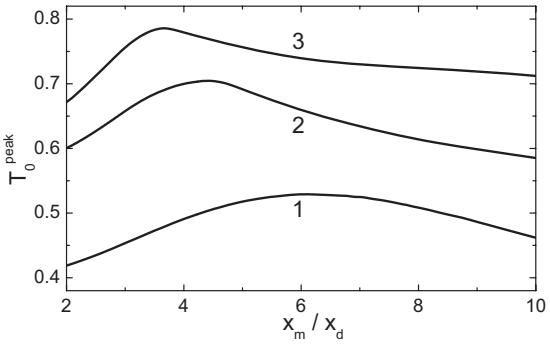

FIG. 6. The peak value of $T_0(\lambda/x_0)$ versus the ratio x_m/x_d; lines 1, 2, and 3 are plotted for $|\varepsilon_m|=9$, 5, and 4, respectively.

the Fabry-Pérot effect for the propagating mode and differs from the single-mode expression of Refs. 5 and 34 by an essential simplification: Instead of two general complex parameters of this expression, we employ only two real parameters with the known properties. The only additional requirement is that the slab thickness d must be considerably larger than the maximum decay distance for the nonpropagating modes. Usually, it is fulfilled for $d\gg\lambda/(2\pi\sqrt{|\varepsilon_m|})$. The outcomes of the simplified relation for T_Σ will be compared with the results of the rigorous numerical calculations.

Let $A_\pm^0$ be the amplitudes of the light waves in air propagating in the $\pm z$ directions at $z=0$, $B_\pm^0$ and $B_\pm^d$ be the amplitudes of the $\pm$ propagating mode of the periodic structure at $z=0$ and d, and $C_\pm^d$ be the amplitudes of the $\pm$ waves in air at $z=d$ (see Fig. 7). The amplitudes are normalized in such a way that their squared absolute values give the corresponding energy fluxes averaged over the period of the structure x_0.

According to our definitions, A_+^0 and B_-^0 are the input amplitudes for the interface $z=0$, while A_-^0 and B_+^0 are the output amplitudes. Since the transmission-reflection problem is linear, we have

$$\begin{pmatrix} A_-^0 \\ B_+^0 \end{pmatrix} = \hat{U}\begin{pmatrix} A_+^0 \\ B_-^0 \end{pmatrix}, \tag{8}$$

where the 2×2 matrix $\hat{U}$ expresses the necessary properties of the input interface. Since the interface $z=d$ is physically the same, the output amplitudes C_+^d and B_-^d are expressed by C_-^d and B_+^d in the same way.

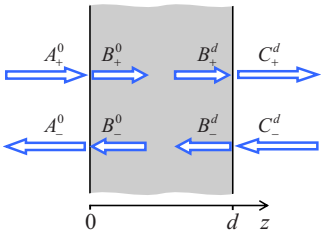

FIG. 7. (Color online) Geometric scheme for the analysis of the slab transmission properties.

Note now the obvious relations for $B_{\pm}^{0,d}$ (see also Fig. 7) $B_{+}^{d}=B_{+}^{0}e^{i\beta_0 d}$ and $B_{-}^{d}=B_{-}^{0}e^{-i\beta_0 d}$, where $\beta_0=+\sqrt{\beta_0^2}$ refers to the forward propagating mode. This parameter is generally complex; i.e., we do not neglect the losses for the propagating mode over the distance d.

Using these relations together with Eq. (8) and setting $A_{-}^{d}=0$, we calculate algebraically the amplitude transmission coefficient C_{+}^{d}/A_{+}^{0},

$$\frac{C_{+}^{d}}{A_{+}^{0}}=\frac{U_{12}U_{21}\exp(i\beta_0 d)}{1-U_{22}^2\exp(2i\beta_0 d)}. \tag{9}$$

General expressions of this kind are well known in the ELT literature.[5,34]

Relation (9) can be simplified if we recall that the influence of losses on transmission/reflection properties of the interface is very weak for $\varepsilon_m'' \ll |\varepsilon_m'|$ (see Sec. III). One can expect that the influence of absorption on the slab transmittance comes mostly from the propagation losses, which are attributed to the factor $\exp(-\beta_0'' d)$ entering Eq. (9), while the $\hat{U}$ matrix can be calculated for $\varepsilon_m''=0$. In other words, we neglect the surface losses as compared to the propagating ones.

Consider the symmetry properties of $\hat{U}$ in the lossless case. Owing to the energy conservation at the interfaces (i.e., owing to the equalities $|A_{+}^{0}|^2+|B_{-}^{0}|^2=|A_{-}^{0}|^2+|B_{+}^{0}|^2$ and $|C_{-}^{d}|^2+|B_{+}^{d}|^2=|C_{+}^{d}|^2+|B_{-}^{d}|^2$), this matrix is unitary, $\hat{U}\hat{U}^{\dagger}=\hat{U}^{\dagger}\hat{U}=\hat{1}$, where the dagger stands for the Hermitian conjugation. Furthermore, the lossless transmission-reflection properties of the interfaces are invariant to the time reversal. This operation means the change of the propagation directions (the input waves become output ones and vice versa) and complex conjugation of the amplitudes. Thus, we have, additionally, $\hat{U}\hat{U}^{*}=\hat{U}^{*}\hat{U}=\hat{1}$.

The general form of the matrix $\hat{U}$, which satisfies both above requirements, is

$$\hat{U}=e^{i\varphi}\begin{pmatrix}\cos\theta e^{i\psi} & \sin\theta\\ \sin\theta & -\cos\theta e^{-i\psi}\end{pmatrix}, \tag{10}$$

where θ, φ, and ψ are three real characteristic parameters. According to Eq. (8) and Fig. 7, the combinations $\cos\theta e^{i(\varphi+\psi)}$ and $\sin\theta e^{i\varphi}$ are the amplitude reflection and transmission coefficients for the input wave A_{+}^{0}. In other words, we have $T_0=\cos^2\theta$, $\varphi_t=\varphi$, and $\varphi_r=\varphi+\psi$. We have expressed thus the matrix $\hat{U}$ by the parameters defined and investigated in Sec. III.

Using Eqs. (9) and (10), we calculate lastly the total transmittance of the slab $T_\Sigma=|C_{+}^{d}|^2/|A_{+}^{0}|^2$,

$$T_\Sigma=\frac{T_0^2\exp(-2\beta_0'' d)}{|1-(1-T_0)\exp(-2\beta_0'' d)\exp(2i\Phi_{FP})|^2}, \tag{11}$$

where the Fabry-Pérot (FP) phase

$$\Phi_{FP}=\beta_0' d+2\varphi_t-\varphi_r. \tag{12}$$

The combination $2\varphi_t-\varphi_r$ is merely the phase shift for the propagating wave during an internal reflection from any of the two slab interfaces. This remark clarifies the physical meaning of the FP phase Φ_{FP}.

The right-hand side of Eq. (11) includes the known characteristics of a single interface (T_0 and $\varphi_{t,r}$) and the products $\beta_0' d$ and $\beta_0'' d$, which can be easily calculated from Eq. (3). The final relation for T_Σ reduces the transmission problem to the determination of the minimum number of parameters. Qualitatively, most of the outcomes of this relation are physically transparent.

In the lossless case, $\beta_0''=0$, the slab transmittance turns to 1 for $\Phi_{FP}=s\pi$ with $s=0,\pm1,\ldots$ This condition corresponds to the FP resonances. Furthermore, we have $T_\Sigma\to1$ for $T_0\to1$ regardless of Φ_{FP}.

As we know, the spectral dependence of the phase φ_t includes a π jump at the zero point of T_0. For Φ_{FP}, it gives a pseudojump 2π. The phase Φ_{FP} can thus be treated as a continuous function of λ/x_0.

Two intuitively clear features come from Eq. (11):

(1) The spectral width of a FP peak is affected by the value of the derivative of Φ_{FP} at the point of maximum; the further is it is from the peak of T_0, the larger is the width.

(2) The smaller T_0 is at a FP resonance, the stronger is the negative influence of the losses on the peak height.

V. SLAB TRANSMISSION RESONANCES

The purpose of this section is twofold. First, we are going to extend our calculation scheme for the slab case to be able to get exact results on the slab transmittance and to verify the field of applicability of the simplified Eq. (11). Second, using the single-interface characteristics of Sec. III, we intend to investigate the spectral dependence $T_\Sigma(\lambda/x_0)$ for different combinations of the characteristic dimensionless parameters x_m/x_d, d/x_0, and ε_m.

A. General calculation scheme

The following eigenmode expansions are valid now in the three discrete regions of z [compare with Eq. (5)]

$$H=\begin{cases}a_0^{+}e^{ik_0 z}+\displaystyle\sum_{n=0}^{\infty}2a_n^{-}\cos(nKx)e^{q_n z}, & z<0\\ \displaystyle\sum_{\nu=0}^{\infty}[b_\nu^{+}e^{i\beta_\nu z}+b_\nu^{-}e^{-i\beta_\nu(z-d)}]h^{\nu}, & [0,d]\\ \displaystyle\sum_{n=0}^{\infty}2c_n^{+}\cos(nKx)e^{-q_n(z-d)}, & z>0.\end{cases} \tag{13}$$

Here, $b_\nu^{\pm}$ are the amplitudes of the eigenmodes propagating in the slab in the $\pm z$ directions and c_n^{+} are the amplitudes of waves (propagating and evanescent) at the output. The other parameters are defined in Sec. III A. For $z>d$, the waves run or decay in the $+z$ direction.

Using the exact boundary conditions at the interfaces $z=0$ and d, we come to the set of linear algebraic relations for determination of $b_\nu^{\pm}$,

$$\sum_{\nu=0}^{\infty}(b_\nu^{+}Q_{\nu n}^{-}+b_\nu^{-}Q_{\nu n}^{+}e^{i\beta_\nu d})=2\delta_{n0},$$

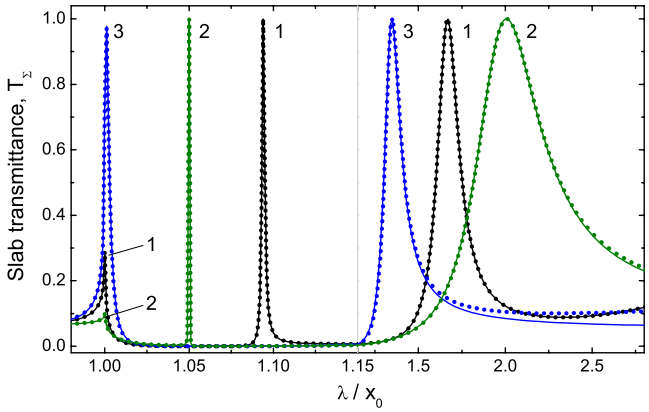

FIG. 8. (Color online) Transmittance T_Σ of the slab depicted in Fig. 1(c) versus λ/x_0 for $\varepsilon_m=-9.6$ and $x_m/x_d=4$. Curves 1, 2, and 3 correspond to $d/x_0=1$, 0.5, and 0.3. The solid/dotted lines are the exact/approximate results. The scale within the range 0.98 to 1.15 is ten times enlarged compared to that for the range of 1.15–2.8.

$$\sum_{\nu=0}^{\infty} (b_\nu^+ Q_{\nu n}^+ e^{i\beta_\nu d} + b_\nu^- Q_{\nu n}^-) = 0, \tag{14}$$

where

$$Q_{\nu n}^\pm = h_n^\nu \pm i\beta_\nu q_n^{-1}(h^\nu/\varepsilon)_n. \tag{15}$$

The output amplitudes in air, c_n, are expressed by $b_\nu^\pm$ using the following exact relations:

$$2c_0^+ = \sum_\nu (b_\nu^+ e^{i\beta_\nu d} + b_\nu^-) h_0^\nu,$$

$$c_n^+ = \sum_\nu (b_\nu^+ e^{i\beta_\nu d} + b_\nu^-) h_n^\nu \quad (n \neq 0). \tag{16}$$

The truncation of set (14) at $\nu_{max}=n_{max}=N$ was used to calculate $b_\nu^\pm$ and determine the slab transmittance $T_\Sigma = |c_0^+|^2/|a_0^+|^2$ with Eq. (16). The procedure converges again perfectly well with increasing N.

Set (14) includes the exponential factors $\exp(i\beta_\nu d)$. The largest of them corresponds to the propagating mode, $\nu=0$; its absolute value is $\exp(-\beta_0'' d)$. Importantly, the other factors ($\nu \geqslant 1$) do not exceed $\exp(-k_0 d\sqrt{|\varepsilon_m'|})$ in absolute value; the larger ν is, the smaller are these factors. For $d \gg \lambda/2\pi\sqrt{|\varepsilon_m'|}$ (the slab thickness is larger than the skin depth) and $k_0\sqrt{|\varepsilon_m'|} \gg \beta_0''$, the largest exponential factor exceeds all the others by orders of magnitude. Keeping in Eq. (14) only this largest factor corresponds to the physical picture of coupling of the opposite faces via a single propagating mode. The validity of this approximation can be checked numerically.

B. Simulation results

Figure 8 shows three representative spectral dependences of T_Σ in the range $\lambda/x_0=(0.98-2.8)$ calculated with our numerical scheme for the lossless case, $\varepsilon_m=-9.6$, $x_m/x_d=4$, and three values of the slab thickness d. Lines 1, 2, and 3 are plotted for $d/x_0=1$, 0.5, and 0.3, respectively. Their solid versions correspond to the exact scheme, whereas for the dotted ones we have neglected the evanescent coupling, i.e., the exponents $\exp(i\beta_\nu d)$ with $\nu \geqslant 1$. To resolve fine spectral features in the vicinity of the RW peaks, we have enlarged the scale within the interval (0.98–1.15); its right border is marked by the vertical line. One sees that for $d/x_0=1$ and 0.5, the exact and approximate dependences practically coincide (including fine features) within the whole subwavelength spectral range $\lambda/x_0 \geqslant 1$. For $d/x_0=0.3$, the difference between the exact and approximate results is noticeable only for $\lambda/x_0 \gtrsim 2$, i.e., far to the right from the RW peak. This difference is not surprising because here the ratio of the slab thickness d to the skin depth is smaller than 3. For $d/x_0 \lesssim 0.1$, the evanescent coupling becomes important in the whole shown range of λ/x_0; the slab transmittance T_Σ is rather small in this case and shows no pronounced spectral peaks. Last, we have found that Eq. (11), with independently calculated single-interface characteristics, gives for $\lambda/x_0 > 1$ *exactly the same results* as the simplified numerical scheme. For $\lambda/x_0 < 1$, as expected, Eq. (11) gives inaccurate results.

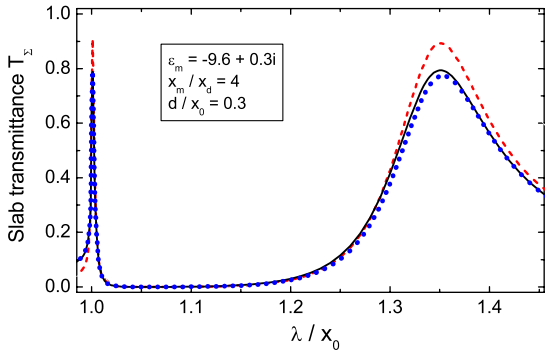

FIG. 9. (Color online) Slab transmittance resonances for $\varepsilon_m'=-9.6$, $\varepsilon_m''=0.3$, $x_m/x_d=4$, and $d/x_0=0.3$. The solid line is the exact numerical result. The dotted line corresponds to the numerical scheme with neglected evanescent terms. The dashed line is plotted using Eq. (11).

The peaks with $T_\Sigma=1$ are the Fabry-Pérot resonances; their positions precisely correspond to the condition $\Phi_{FP}=s\pi$. The farther the FP peak from the zero point of T_0, $\lambda/x_0 \simeq 1.06$, the larger is its width. The FP peak at $\lambda/x_0 \simeq 1.05$ is clearly the narrowest. Three overlapping narrow peaks at $\lambda/x_0=1$ are the RW peaks distorted by the Fabry-Pérot effect; for $d/x_0=0.3$, the phase $\Phi_{FM}(\lambda/x_0=1)$ is occasionally close to 2π.

Consider the influence of light absorption. Figure 9 shows the transmittance resonances for $\varepsilon_m''=0.3$ and $d/x_0=0.3$; the other parameters correspond to Fig. 8. Absorption does not change the position of the resonances but decreases their heights. The evanescent terms are negligible in the whole spectral range and the peak values of T_Σ remain pretty high. Our simplified description (the dashed line) precisely reproduces the position and shape of the resonances and gives $\approx 10\%$ accuracy for the transmittance. The surface losses give thus a minor effect. With increasing d/x_0, they become less important because of the increasing role of the propagation losses.

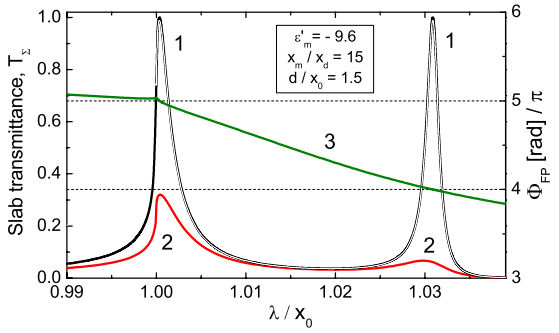

FIG. 10. (Color online) Slab transmittance $T_\Sigma(\lambda/x_0)$ for $\varepsilon_m'=9.6$, $x_m/x_d=15$, and $d/x_0=1.5$; lines 1 and 2 correspond to $\varepsilon_m''=0$ and 0.3, respectively. Line 3 shows the FP phase Φ_{FP}.

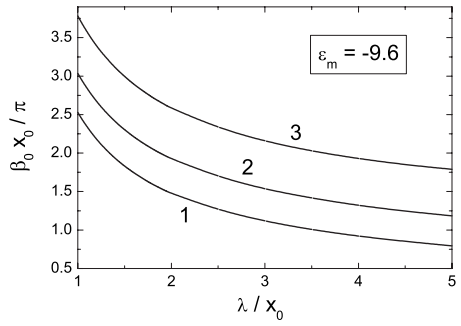

FIG. 11. Dependence of $\beta_0 x_0$ on λ/x_0 for $\varepsilon_m=-9.6$. Lines 1, 2, and 3 are plotted for $x_m/x_d=4$, 10, and 20, respectively.

The influence of light absorption on the ultranarrow FP peaks at $\lambda/x_0\simeq 1.05$ and 1.1 (see Fig. 8) is much stronger. For $\varepsilon_m''=0.3$, the peak values of T_Σ are suppressed by more than 1 order of magnitude. This suppression exemplifies the general rule: The smaller T_0 is at the point of a FP resonance, the stronger is the negative impact of propagation losses.

A FP resonance can be placed, generally, at any position by varying x_m/x_d and d/x_0. Moreover, even two FP resonances can be placed within the narrow spectral peak of T_0, which looks as a split of this peak. Sharp and strong changes of the phase Φ_{FP} in the vicinity of the peak of T_0 are the prerequisites for such a split. This is illustrated by Fig. 10.

Solid line 1 shows a fragment of the spectral dependence of T_Σ calculated exactly or (which is the same for $\lambda/x_0>1$) with the use of Eq. (11) for $\varepsilon_m=-9.6$, very narrow slits, $x_m/x_d=15$, and a large slab thickness, $d/x_0=1.5$. The first FP peak is situated slightly to the left from the maximum of $T_0(\lambda/x_0)$, while the second one is on the right shoulder of the peak of T_0. Both of them are very narrow. Line 3 shows the spectral dependence of Φ_{FP}; the positions of maxima of T_Σ coincide with the positions of the FP resonances. Line 2 is plotted for $\varepsilon_m''=0.3$. Despite a significant suppression owing to the propagation losses, the peak transmittance remains pretty high.

Positions of the FP resonances are determined by the spectral dependence of the phase $\Phi_{FP}=2\varphi_t-\varphi_r+\beta_0'd$ for $\varepsilon_m''=0$. In turn, this dependence is controlled by the parameters x_m/x_d, d/x_0, and $|\varepsilon_m'|$. The combination $2\varphi_t-\varphi_r$ does not depend on d/x_0. It is a decreasing function of λ/x_0, which changes rapidly only in the vicinity of the peak of T_0; this feature is clearly seen in Fig. 3(b). The total drop of $2\varphi_t-\varphi_r$ for $\lambda/x_0>1$ is slightly above π; it grows slowly with increasing x_m/x_d. Only the last contribution to Φ_{FP} depends on (proportional to) the slab thickness. Figure 11 shows the spectral dependence of $\beta_0 x_0$ for three representative values of x_m/x_d. It decreases; the total drop of $\beta_0 x_0/\pi$ is ≈ 1.

As follows from the above observations, the spectral position, width, and height of the FP peaks depend essentially on x_m/x_d and d/x_0. The larger these ratios are, the higher is the spectral density of the FP resonances. The distances between the neighboring resonances and their spectral widths are smallest in the vicinity of the peak of T_0. A detailed quantitative characterization of the FP peaks can be done for a particular experiment, but it is beyond this study.

The case of strong propagation losses, $\beta_0''d\gg 1$, is important in the following respect: The reflectivity of the slab is not affected here by the FP interference. With a high accuracy, the reflection coefficient of the slab is $1-T_0$. Its expected spectral dependence is characterized by a narrow dip in the vicinity of $\lambda/x_0=1$. It is possible thus to measure the predicted spectral dependences of the single-interface characteristics T_0 and φ_r.

VI. DISCUSSION

Our general calculation scheme, based on the modal expansion, uses no simplifying assumptions within the concept of the optical permittivity. It employs the exact eigenfunctions, eigenvalues, and boundary conditions; it meets the known limiting cases and symmetry properties. Therefore, this scheme includes precisely all possible mechanisms of the ELT in the 1D case in question.

The scheme is flexible. It provides all necessary parametric dependences and admits approximations to see the physics behind the ELT. In particular, it allows us to analyze the lossless and lossy cases, to separate the roles of the surface and propagation losses, and to describe the dispersive properties of holey-surface plasmons.

A crucial element of our scheme is inclusion of not only the usual evanescent and propagating modes but also the anomalous eigenmodes with essentially complex values of the propagating constant.[22] Without the anomalous modes, the calculation procedure diverges, while with these modes we have a fast convergence.

By following the physical approach, we aim to reduce the complex ELT phenomenon to the basic elements instead of try to exhaust numerically combinations of the variable parameters or attach ourselves to a particular set of parameters. We pretend thus to give a clear physical picture of the ELT within a wide range of the variable parameters by combining the numerical and analytical tools. Possible simplifications related to the presence of small physical parameters are also important.

Transmission-reflection characteristics of a single interface between air and hole metal are the key to understanding the ELT physics. They are free of the influence of the Fabry-Pérot interference and evanescent coupling inherent in the

slab case. The role of the surface plasmons is maximally clear here.

Our analysis has shown that the physics of the single-interface transmittance is rich, somewhat surprising, but, at the same time, pretty simple. It is greatly controlled by the parameter $x_m/|\varepsilon'_m|x_d$ while small losses ($\varepsilon''_m \ll |\varepsilon'_m|$) are negligible:

(1) For $x_m/|\varepsilon'_m|x_d \leqslant 1$, the dependence of the interface transmittance T_0 on the spectral parameter λ/x_0 is characterized by narrow spikelike RW peaks at $1, 1/2, 1/3, \ldots$. They are due to opening diffraction orders and are uncoupled from the plasmon resonances.

(2) For $x_m/|\varepsilon'_m|x_d > 1$, we have a different quality. The spectral peaks of T_0 are not attached to the RW anomalies and they have no kinks. With increasing x_m/x_d, the main peak moves from $\lambda/x_0=1$ toward the FSP resonance $(\lambda/x_0)_{FSP}=\sqrt{|\varepsilon'_m|/(|\varepsilon'_m|-1)}$.

(3) For $x_m/|\varepsilon'_m|x_d \gtrsim 2$, the spectral position of the main peak of T_0 is very close to the HSP resonance $(\lambda/x_0)_{HSP}$ but still far from $(\lambda/x_0)_{FSP}$, so that the transmittance peak is due to the resonant excitation of the HSP.

Some other features supplement this picture.

(1) The value $x_m/|\varepsilon'_m|x_d=1$, critical for the spectral properties, is also critical in the continuous-medium limit, $\lambda/x_0 \gg 2\pi\sqrt{|\varepsilon'_m|}$. The effective permittivity of our periodic structure, ε_{eff}, is positive for $x_m/|\varepsilon'_m|x_d > 1$ and negative for $x_m/|\varepsilon'_m|x_d < 1$.[26]

(2) The point of the transmittance minimum $(\lambda/x_0)_{min}$ lies usually very close to $(\lambda/x_0)_{FSP}$. This proximity does not signify the negative role of the surface plasmon because the influence of the slits on the dispersive properties of the HSP is indispensable. The distance between $(\lambda/x_0)_{min}$ and $(\lambda/x_0)_{FSP}$ grows with decreasing $|\varepsilon'_m|$ and becomes noticeable for $|\varepsilon'_m|=4-5$. The absence of coincidence of these spectral positions means that there are no general reasons for it.

(3) The peak transmittance T_0^{max} increases and becomes comparable to 1 when the relative permittivity $\varepsilon_m/\varepsilon_d$ approaches the value of -3. The reason is the well-known singularities at the metal corners.[32,33]

While the spectral properties of T_0 are sufficient to understand the physics behind the single-interface resonances, they are insufficient to describe the spectral dependence of the slab transmittance $T_\Sigma(\lambda/x_0)$. To do so, we need to know additionally the spectral dependence of the combination of the phases of the transmitted t and reflected (r) waves $2\varphi_t - \varphi_r$; it represents the phase change during an internal reflection from the interface. The losses affect it very weakly.

Qualitatively, the spectral dependence of $2\varphi_t - \varphi_r$ is similar to that of the phase of an elementary oscillator. The changes are most rapid in the vicinity of the resonance of T_0, and the overall change when passing it is about π. Quantitatively, the dependence of this phase combination on λ/x_0, x_m/x_d, and $|\varepsilon'_m|$ is complicated.

Narrowly viewed, the main outcome of the numerical part is reduced to the calculation of T_0 and $2\varphi_t - \varphi_r$. Unfortunately, to date, there is no known way to provide an adequate analytical description of these fundamental characteristics.

As soon as these two single-interface characteristics are known, the transmittance of a thick slab T_Σ can be calculated straightforwardly for $\lambda \geqslant x_0$ using the simplified single-mode model. The corresponding relation fully accounts for the strong coupling of the opposite interfaces via a single weakly damped propagation mode and Fabry-Pérot resonances for the latter.

Qualitatively, the picture of the slab resonances is simple. Most of them are of the Fabry-Pérot nature. It is most likely to get a FP resonance nearby the spectral peak of T_0, where the phase changes are rapid; such a resonance is very narrow and can be strongly suppressed by the propagating losses. By changing the slab thickness d and the ratio x_m/x_d, it is possible to move a FP resonance far from the peak of T_0; such a resonance is spectrally wide and robust. No direct connection exists between the positions of the FP and plasmon resonances.

The relationship between our findings and the literature data is an important issue to discuss.

Our modal-expansion method is closest to that of Ref. 11. The authors of this paper fixed a complex value of ε_m, saw the anomalous modes (as "an unknown abnormality"), and included them into the numerical scheme. This scheme was then exemplarily used. In the modal expansion of Ref. 6, the anomalous modes were missed; the encountered problems with the convergence are, most probably, due to this fact. A perfect convergence of our method for $\varepsilon''_m=0$ contrasts to the fact that the rigorous coupled-wave analysis fails for lossless metals.[35]

The spectral dependence of the single-interface transmittance was calculated in Refs. 11 and 28. Only a single set of parameters (including a complex value of ε_m) was used in each of them. The authors saw the RW peaks and the minimum transmittance at $(\lambda/x_0)_{FSP}$. The x_m/x_d and ε'_m dependences, the phase changes, and the role of losses were not considered.

A general expression for the slab transmittance, incorporating the FP resonances and known as the single-mode model, was introduced into the ELT theory in Refs. 4 and 5. It works well as a proof of principle. With weakness of the losses taken into account, it can be advanced to gain the prediction power for the positions and heights of the FP transmittance peaks.

A comparison with experiment is beyond the scope of this paper. One recent publication[36] is, however, worthy of mention. Surrounding silver-based slit structures ($x_m/x_d \ll 1$) with a dielectric ($\varepsilon_d \simeq 2.13$), the relative permittivity $\varepsilon_m/\varepsilon_d \simeq -4.37$ was achieved. This means first that the region of small $|\varepsilon_m|/\varepsilon_d$ (compared to Ag and Au), which is expected to be important, is accessible in experiment. Second, spectrally resolved measurements of the slab transmittance[36] elucidated the role of the surface plasmon. For $x_m/x_d \simeq 5.12$, the point of minimum transmittance $(\lambda/x_0)_{min}$ is found to be close to the flat-surface plasmon resonance $(\lambda/x_0)_{FSP}$. Our theory predicts a $\approx 3\%$ excess of $(\lambda/x_0)_{min}$ (see Fig. 5). A careful examination of the experimental points shows a tendency of such a shift, but it cannot be resolved within the experimental accuracy. Increasing accuracy and/or ε_d would allow us to detect the predicted difference. For $|\varepsilon_m|/\varepsilon_d \gtrsim 7$, the difference between $(\lambda/x_0)_{min}$ and $(\lambda/x_0)_{FSP}$ becomes too small to be resolved.

To a certain extent, our results can be applied to the 2D case. The point is that, in contrast to Refs. 16 and 17, a propagating mode can survive even for subwavelength 2D holes.[22,37,38] It occurs when the relative optical permittivity $\varepsilon_m/\varepsilon_d$ lies in the range $[-1,0]$. This range is now exotic for experiment, but it should be accessible with such metals as Ni by filling the holes with a transparent dielectric/fluid.[24,25] Since the propagating modes strongly facilitate the ELT, the indicated range is worthy of close attention. If the surface losses are still weak here, the simplified description of Sec. IV is applicable to analyze the ELT resonances.

Finally, we point out that our modal approach can be generalized for the periodic media consisting of layers with both optical permittivity and permeability negative. Such materials are intensively studied nowadays in connection with the negative refraction phenomenon and unusual transmission properties.[39,40]

VII. SUMMARY

We have proposed an advanced version of the ELT theory for arrays of subwavelength slits in metals. It incorporates the known results and also a big number of original elements; altogether they form a clear physical picture of the effect. The theory clarifies a number of fundamental issues of the ELT phenomenon. More specifically, our findings can be summarized as follows.

(1) The anomalous eigenmodes with essentially complex propagating constants are necessary, in addition to the propagating and evanescent modes, to ensure convergence of the modal expansion in the ELT theory.

(2) Two real characteristics of a single interface between air and perforated metal, the transmittance T_0 and the phase change for the propagating mode, predetermine the slab transmittance. Weak losses, $\varepsilon_m'' \ll |\varepsilon_m'|$, are of minor importance for these characteristics.

(1) The dependence of T_0 on the spectral parameter λ/x_0 is controlled by the ratio $x_m/x_d|\varepsilon_m|$. For $x_m/x_d|\varepsilon_m| \leqslant 1$, the transmittance resonances are due to opening new diffraction orders (the Rayleigh-Wood anomalies at $\lambda/x_0=1,1/2,\ldots$). For $x_m/x_d|\varepsilon_m|>1$, the spectral peaks of T_0 are not the RW resonances. With increasing ratio x_m/x_d, the main peak of $T_0(\lambda/x_0)$ moves to the right, becomes kinkless, and attaches quickly to the holey-surface plasmon resonance $(\lambda/x_0)_{HSP}$. Thus, a crossover between the RW and HSP resonances occurs. The height of the main peak of T_0 approaches 1 for $\varepsilon_m \to -3$ because of the corner singularities.

(2) With weak losses, the slab transmittance T_Σ is expressed explicitly by the above two real single-interface characteristics. The derived relation for T_Σ includes the propagation losses and can be considered as an advanced version of the single-mode model.[5] It describes the positions and heights of the FP peaks in agreement with the direct numerical calculations.

APPENDIX A: CALCULATION OF PLASMON AMPLITUDE

Consider a metallic part of our semi-infinite periodic structure [Fig. 1(b)]. Let it be $x_d/2<x<x_m+x_d/2$. We search for a solution for the magnetic field $H=\Psi(z)\exp(i\gamma x)$, where γ is a new propagation constant and Ψ is the corresponding eigenfunction. Then, we have

$$\frac{d^2\Psi}{dz^2} = (\gamma^2 - \varepsilon k_0^2)\Psi. \tag{A1}$$

The physical solutions must either be localized in $|z|$ or behave as $\exp(iCz)$ with a real parameter C for $z \to \pm\infty$ (radiative modes).

The spectrum of allowed values of γ^2 consists of a single point $\gamma_p^2 = k_0^2\varepsilon_m/(\varepsilon_m+1)$, which corresponds to the surface plasmon, and the range $\gamma^2 < k_0^2$. The wave function for the surface plasmon,

$$\Psi_p = \begin{cases} \exp(k_0 z/\sqrt{|\varepsilon_m|-1}) & (z<0) \\ \exp(-k_0 z|\varepsilon_m|/\sqrt{|\varepsilon_m|-1}) & (z>0), \end{cases} \tag{A2}$$

is localized and real. The functions of the continuous spectrum, $\Psi_\gamma(z)$, possess an oscillating asymptotic behavior for $z \to +\infty$ and/or $z \to -\infty$. Since the values of γ^2 can be negative, the dependence $H_\gamma(x)$ can be exponentially increasing and decreasing. One can see from Eq. (A1) that the following orthogonality relation takes place:

$$\int_{-\infty}^{\infty} \varepsilon^{-1}(z)\Psi_p(z)\Psi_\gamma(z)dz = 0. \tag{A3}$$

Assuming that the set of the eigenfunctions $\{\Psi_p, \Psi_\gamma\}$ is complete, we represent the magnetic field within the interval $x_d/2<x<(x_d+x_0)/2$ as

$$H = (A_p^+ e^{i\gamma_p x} + A_p^- e^{-i\gamma_p x})\Psi_p(z) + \int (A_\gamma^+ e^{i\gamma x} + A_\gamma^- e^{-i\gamma x})\Psi_\gamma(z)d\gamma, \tag{A4}$$

where $A_p^\pm$ are the amplitudes of plasmons propagating in the $\pm x$ directions and $A_\gamma^\pm$ are amplitudes of the eigenfunctions of the continuous spectrum.

The plasmon contribution is obviously even about the wall center so that $A_p^+\exp(i\gamma_p x)+A_p^-\exp(-i\gamma_p x) = A_p\cos[\gamma_p(x-x_0/2)]$. For the plasmon amplitude A_p, we have the following from Eqs. (A3) and (A4):

$$A_p = \int_{-\infty}^{\infty} \varepsilon^{-1}(z)\Psi_p(z)H(x_0/2,z)dz. \tag{A5}$$

The energy of the standing plasmon wave is proportional to $|A_p|^2$. It can be calculated using Eq. (A2) as soon as the function $H(x_0/2,z)$ is known.

[1] T. W. Ebbesen, H. J. Lezec, H. F. Ghaemi, T. Thio, and P. A. Wolff, Nature (London) **391**, 667 (1998).
[2] H. F. Ghaemi, T. Thio, D. E. Grupp, T. W. Ebbesen, and H. J. Lezec, Phys. Rev. B **58**, 6779 (1998).
[3] W. L. Barnes, A. Dereux, and T. W. Ebbesen, Nature (London) **424**, 824 (2003).
[4] J. A. Porto, F. J. Garcia-Vidal, and J. B. Pendry, Phys. Rev. Lett. **83**, 2845 (1999).
[5] F. J. Garcia-Vidal and L. Martin-Moreno, Phys. Rev. B **66**, 155412 (2002).
[6] M. M. J. Treacy, Phys. Rev. B **66**, 195105 (2002).
[7] Q. Cao and Ph. Lalanne, Phys. Rev. Lett. **88**, 057403 (2002).
[8] Ph. Lalanne, J. P. Hugonin, and J. C. Rodier, Phys. Rev. Lett. **95**, 263902 (2005).
[9] P. N. Stavrinou and L. Solymar, Opt. Commun. **206**, 217 (2002).
[10] J. Bravo-Abad, A. Degiron, F. Przybilla, C. Genet, F. J. Garcia-Vidal, L. Martin-Moreno, and T. W. Ebbesen, Nature (London) **2**, 120 (2006).
[11] Y. Xie, A. R. Zakharian, J. V. Moloney, and M. Mansuripur, Opt. Express **14**, 6400 (2006).
[12] H. Raether, *Surface Plasmons* (Springer-Verlag, Berlin, 1998).
[13] P. B. Jonson and R. W. Christy, Phys. Rev. B **6**, 4370 (1972).
[14] *Handbook of Optics*, edited by M. Bass (McGraw-Hill, New York, 1995), Vol. 2.
[15] E. Popov, M. Neviere, S. Enoch, and R. Reinisch, Phys. Rev. B **62**, 16100 (2000).
[16] A. Krishnan, T. Thio, T. J. Kim, H. J. Lozec, T. W. Ebbesen, P. A. Wolff, J. Pendry, L. Martin-Moreno, and F. J. Garsia-Vidal, Opt. Commun. **200**, 1 (2001).
[17] N. Bonod, S. Enoch, L. Li, E. Popov, and M. Neviere, Opt. Express **11**, 482 (2003).
[18] D. M. Whittaker and I. S. Culshaw, Phys. Rev. B **60**, 2610 (1999).
[19] S. G. Tikhodeev, A. L. Yablonskii, E. A. Muljarov, N. A. Gippius, and T. Ishihara, Phys. Rev. B **66**, 045102 (2002).
[20] A. Yariv and P. Yeh, *Optical Waves in Crystals* (Wiley Interscience, New York, 2003), Chap. 11.
[21] B. Sturman, E. Podivilov, and M. Gorkunov, Europhys. Lett. **80**, 24002 (2007).
[22] B. Sturman, E. Podivilov, and M. Gorkunov, Phys. Rev. B **76**, 125104 (2007).
[23] P. Sheng, R. S. Stepleman, and P. N. Sanda, Phys. Rev. B **26**, 2907 (1982).
[24] G. Whitesides and A. Stoock, Phys. Today **54**(6), 42 (2001).
[25] C. Monat, P. Domachuk, and B. J. Eggleton, Nat. Photonics **1**, 106 (2007).
[26] A. Yariv and P. Yeh, J. Opt. Soc. Am. **67**, 423 (1977).
[27] S. Collin, F. Pardo, R. Teissier, and J.-L. Pelouard, Phys. Rev. B **63**, 033107 (2001).
[28] P. Lalanne, C. Sauvan, J. P. Hugonin, J. C. Rodier, and P. Chavel, Phys. Rev. B **68**, 125404 (2003).
[29] R. W. Wood, Philos. Mag. **4**, 396 (1902).
[30] A. Hessel and A. A. Oliner, Appl. Opt. **4**, 1275 (1965).
[31] S. Fan, W. Suh, and J. D. Joannopoulos, J. Opt. Soc. Am. A **20**, 569 (2002).
[32] J. Meixner, IEEE Trans. Antennas Propag. **AP-20**, 442 (1972).
[33] G. R. Hardley, J. Lightwave Technol. **20**, 1219 (2002).
[34] P. Lalanne, J. C. Rodier, and J. P. Hugonin, J. Opt. A, Pure Appl. Opt. **7**, 422 (2005).
[35] P. Lalanne and J.-P. Hugonin, J. Opt. Soc. Am. A **17**, 1033 (2000).
[36] D. Pacifici, H. J. Lezec, H. Atwater, and J. Weiner, arXiv:0708.1886 (unpublished).
[37] P. B. Catrysse, H. Shin, and S. Fan, J. Vac. Sci. Technol. B **23**, 2675 (2005).
[38] H. Shin, P. B. Catrysse, and S. Fan, Phys. Rev. B **72**, 085436 (2005).
[39] J. B. Pendry, Phys. Rev. Lett. **85**, 3966 (2000).
[40] I. V. Shadrivov, A. A. Sukhorukov, and Yu. S. Kivshar, Phys. Rev. Lett. **95**, 193903 (2005).