Single-scattering theory of light diffraction by a circular subwavelength aperture in a finitely conducting screen

Evgeny Popov, Michel Nevière, Anne Sentenac, Nicolas Bonod, Anne-Laure Fehrembach, Jérôme Wenger, Pierre-François Lenne, and Hervé Rigneault

Institut Fresnel, Domaine Universitaire de Saint Jérôme, Université d'Aix Marseille III, CNRS UMR 6133, 13397 Marseille Cedex 20, France

Received June 23, 2006; accepted August 14, 2006;
posted August 23, 2006 (Doc. ID 72319); published January 10, 2007

A perturbation theory based on a single-scattering approximation is developed from the rigorous differential theory of diffraction in cylindrical coordinates. It results in analytical formulas in the inverse space for the field amplitudes providing results that are in quantitative agreement with the results of the rigorous method, in both the near- and far-field regions, when a proper correction to the incident field inside the aperture is made by using the renormalized Born approximation. When working in reflection by a screen having permittivity high in modulus, the method proposes an equivalence with the simple model consisting of the emission by a single magnetic dipole excited inside the pierced layer, emission that is then transferred back into the cladding following the Fresnel's coefficients of transmission from the layer into the cladding. The theory predicts a directivity of the radiation pattern that increases for smaller values of modulus of permittivity, both for dielectrics and metals, thus independently of the possibility of plasmon surface wave excitation along the interface. The theory can take into account such surface wave resonances, as well as the waveguide supported by a dielectric slab, but cannot implicitly recognize the modes carried out by the cylindrical waveguide corresponding to the aperture. This fact limits its domain of validity when used in transmission, although the far- and near-field maps can be reconstructed sufficiently well within a multiplicative factor corresponding to the enhanced transmission due to the excitation of these modes.

OCIS codes: 050.1220, 050.1960.

1. INTRODUCTION

When Bethe[1] developed his theory of light diffraction by a circular aperture in a perfectly conducting screen during the Second World War, he could hardly have imagined what impact it would have after half a century. The almost purely academic interest—as described by Jackson[2]—shown in this problem has received a practical impetus in the new wave that started almost ten years ago with the experimental work of Ebessen *et al.*[3] that demonstrated the possibility of transmission enhancement through arrays of subwavelength holes. This was followed by many theoretical works[4–9] that provided a physical understanding of the phenomenon, explaining it by the combined influence of surface plasmon excitation on the dielectric–metal interface of the array supporting layer and the enhanced transmission due to the waveguide mode existing inside the hollow waveguide formed inside the aperture. These works have invoked an interest in diffraction by a single aperture, accelerated by its importance in chemistry and biology.[10] Using single apertures as nanocavities allows trapping and detecting single molecules and studies of biomembrane segments using luminescence spectroscopy or Raman scattering, provided the light emitted reaches a detectable level. Different solutions of light enhancement inside the nanoholes have already been studied, among which are the propositions to use a coaxial aperture[11] or to introduce a surface corrugation around the aperture in order to resonantly enhance the surface plasmon.[12–14] Recent experimental and theoretical works have already stressed the role played by the surface plasmon excitation and the importance of the polarization effects[15–17] not only in enhancing the field inside the aperture, but also in increasing the directivity of the radiated field (and thus of the detected signal).[18–20]

The theory of Bethe predicts that a small aperture diffracts as if the screen and the aperture were replaced by two emitting electric and magnetic dipoles. In the case of a normally incident linearly polarized plane wave, the aperture remains a single magnetic dipole lying in the plane of the screen and perpendicular to the incident polarization vector. Thus, the diffraction in the plane of polarization (i.e., in the plane perpendicular to the dipole axis) is uniform angularly presenting no directivity, while in the perpendicular plane it follows a simple $\cos^2$ law. Jackson[2] predicts a stronger angular dependence (as discussed in detail in Section 6 of the present paper) for larger apertures, but his predictions for small ones coincide with Bethe's theory, at least concerning the directivity.

However, recent theoretical and experimental works[15,18] indicate that the directivity of the radiation pattern of a single aperture in real metals is larger than the theoretical predictions valid for perfectly conducting screens. Another discrepancy, found only recently,[21] is that Bethe's theory predicts that the diffracted field amplitude grows as the third power of the aperture radius R, faster than the second power given by the Kirchoff ap-

proximation, while numerical results for real metals show that the field inside the aperture grows like R^2 for small radii up to the cutoff of the fundamental mode of the hollow metallic waveguide formed inside the aperture.

These facts indicate the necessity to revisit Bethe's theory in order to obtain a better understanding of the process of light diffraction by small apertures in real-metal screens. Although there exist several numerical methods capable of resolving this problem using general tools like finite-difference time-domain or finite-element methods, or more specialized theories adapted to the aperture geometry,[17,22] a simpler analytical approach providing better physical insight would be welcomed. Several efforts made in this direction[23,24] have provided approximate theories for thick but perfectly conducting screens[24] and relatively large apertures.

The aim of this paper is to fill in the existing gap in the theory for small apertures in finitely conducting screens. On the basis of the rigorous differential method in cylindrical geometry,[22] we propose an approach valid in the first-order approximation with respect to R^2. This approach takes into account only the single-scattering process adapted to real metallic or dielectric screens. The incident wave inside the aperture serves as a source field and is diffracted by the perturbation representing the aperture. This incident wave is considered unchanged in the approximation used and thus the approach is equivalent to the Kirchoff theory and the theory presented by Jackson,[2] but adapted to finitely conducting materials. The resulting equations for the diffracted field represent an approach completely analytical in the inverse space and requiring numerical treatment in order to obtain the field maps in the direct space. Comparison with rigorous numerical results shows very good agreement in the form of near- and far-field distribution obtained using the analytical approximate method, within a common factor of ~ 2.

An improvement of the unperturbed incident field approximation is proposed based on a renormalization[25] of Born approximation by taking into account the singularity of the three-dimensional Green's function tensor[26] due to the self-scattering process, i.e., concerning also the incident field. In the case of a cylindrical scatterer, this renormalization provides the missing factor, which tends toward 2 when the contrast of the optical indices of the layer and the aperture increases.

Another limitation of the analytical method is that owing the single-scattering approximation, it cannot take into account the waveguide mode, evanescent or propagating inside the hollow waveguide formed inside the aperture. Thus, when looking in transmission through a conducting screen, the amplitude of the scattered field is much smaller than the results obtained by the rigorous method, which takes this mode into account. However, the form of the field map obtained by the two methods is quite similar.

Section 6 is devoted to the study of the directivity of the radiation pattern when varying the screen permittivity ε_2. It is demonstrated that while for very strongly conducting screens (or dielectrics with very high permittivity) the angular distribution follows quite well Bethe's predictions (or Jackson's formula for larger radii), smaller moduli of the permittivity lead to larger directivity of the field diffracted in the plane of incident field polarization, thus confirming the previous results. Moreover, the analytical equations provide a simple physical interpretation, enabling us to demonstrate that for permittivities sufficiently large in modulus, the diffraction pattern is equivalent to the diffraction pattern of a magnetic dipole excited by the incident field and emitting inside the plane screen. Its radiation is transferred into the cladding by the refraction law guided by Fresnel coefficients. The stronger the contrast between the screen and the cladding, the smaller the angular variation of these coefficients and thus of the diffracted field. In the limits of $|\varepsilon_2| \to \infty$ it is demonstrated that the diffraction pattern corresponds to the radiation of a magnetic dipole without the screen, confirming Bethe's interpretation. Contrary to the hypothesis made in Ref. 18 that it is the surface plasmon excitation on real-metal screens that increases the directivity of the radiation pattern inside the plane of the incident field polarization, we demonstrate that while this is true in the near-field region, the far-field directivity is increased for both metallic and dielectric interfaces, the latter being unable to support surface or volume guided modes. We show that the increased directivity can be simply explained by the angular variation of Fresnel transmission coefficients.

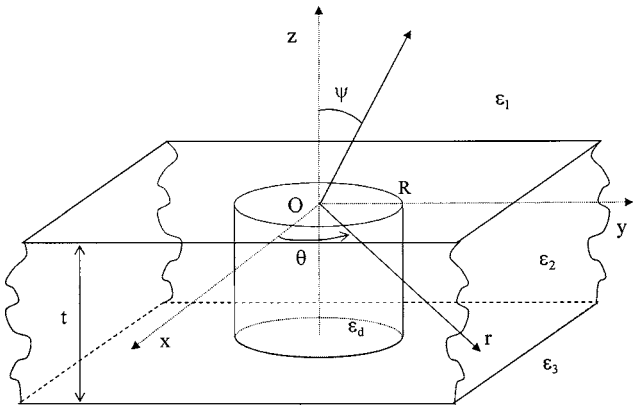

Fig. 1. Schematic representation of a screen with an aperture.

The system under consideration is presented schematically in Fig. 1 together with the coordinate system and some of the notations used below. A monochromatic plane wave linearly polarized in the xOz plane is incident from the cladding. We assume an $\exp(-i\omega t)$ time dependence.

2. GENERAL SOLUTION INSIDE THE INHOMOGENEOUS MEDIUM

Due to the natural 2π periodicity with respect to θ, the electric and magnetic field components can be represented as Fourier series in θ:

$$E_j(r,\theta,z) = \sum_{n=-\infty}^{+\infty} E_{j,n}(r,z)\exp(in\theta),$$

$$H_j(r,\theta,z) = \sum_{n=-\infty}^{+\infty} H_{j,n}(r,z)\exp(in\theta), \quad j = r,\theta,z, \tag{1}$$

so that the Maxwell's equations can be separately written for each field component $E_{j,n}$ and $H_{j,n}$:

$$\frac{\partial E_{\theta,n}}{\partial z} = \frac{in}{r}E_{z,n} - i\omega\mu_0 H_{r,n}, \qquad \frac{\partial H_{\theta,n}}{\partial z} = \frac{in}{r}H_{z,n} + i\omega\varepsilon_0\varepsilon E_{r,n},$$

$$\frac{\partial E_{r,n}}{\partial z} = \frac{\partial E_{z,n}}{\partial r} + i\omega\mu_0 H_{\theta,n}, \qquad \frac{\partial H_{r,n}}{\partial z} = \frac{\partial H_{z,n}}{\partial r} - i\omega\varepsilon_0\varepsilon E_{\theta,n},$$

$$i\omega\mu_0 H_{z,n} = \frac{\partial E_{\theta,n}}{\partial r} + \frac{E_{\theta,n}}{r} - \frac{in}{r}E_{r,n}, \qquad -i\omega\varepsilon_0\varepsilon E_{z,n} = \frac{\partial H_{\theta,n}}{\partial r} + \frac{H_{\theta,n}}{r} - \frac{in}{r}H_{r,n}, \tag{2}$$

where ε_0 is the free-space absolute permittivity and ε is the relative permittivity of the medium.

It is possible to use a Bessel-function basis for the field components by using four spectral amplitudes $b^E_{n,m}(z)$, $b^H_{n,m}(z)$, $c^E_{n,m}(z)$, and $c^H_{n,m}(z)$, as demonstrated in detail in Appendix A:

$$E_{\theta,n}(r,z) = \sum_{m=0}^{\infty} [b^E_{n,m}(z)J_{n+1}(k_m r) + c^E_{n,m}(z)J_{n-1}(k_m r)]\Delta_m, \tag{3}$$

$$E_{r,n}(r,z) = i\sum_m [b^E_{n,m}(z)J_{n+1}(k_m r) - c^E_{n,m}(z)J_{n-1}(k_m r)]\Delta_m, \tag{4}$$

$$k_0^2\varepsilon E_{z,n}(r,z) = i\sum_m [b^H_{n,m}(z) - c^H_{n,m}(z)]J_n(k_m r)k_m\Delta_m, \tag{5}$$

$$\omega\mu_0 H_{\theta,n}(r,z) = \sum_m [b^H_{n,m}(z)J_{n+1}(k_m r) + c^H_{n,m}(z)J_{n-1}(k_m r)]\Delta_m, \tag{6}$$

$$\omega\mu_0 H_{r,n}(r,z) = i\sum [b^H_{n,m}(z)J_{n+1}(k_m r) - c^H_{n,m}(z)J_{n-1}(k_m r)]\Delta_m, \tag{7}$$

$$\omega\mu_0 H_{z,n}(r,z) = -i\sum_m [b^E_{n,m}(z) - c^E_{n,m}(z)]J_n(k_m r)k_m\Delta_m, \tag{8}$$

with k_0 being the free-space wavenumber and $k_m = m\Delta_m$, so that

$$\Delta_m = k_{m+1} - k_m. \tag{9}$$

Assuming the continuity of E_r, E_θ, and $\mathbf{H}$ on the interface between the cladding and the screen pierced by the aperture (which assumption is valid almost everywhere except on the edge $r=R$), Appendix A explains how to obtain the following set of first-order differential equations:

$$\frac{\mathrm{d}}{\mathrm{d}z}b^E_{n,m} = b^H_{n,m} - \frac{k_m^2}{k_0^2\varepsilon_2}(b^H_{n,m} - c^H_{n,m}) - \sum_{m'\neq m}(\varepsilon^{-1})^{n,n}_{m,m'}\frac{k_m k_{m'}}{2k_0^2}(b^H_{n,m'} - c^H_{n,m'}), \tag{10}$$

$$\frac{\mathrm{d}}{\mathrm{d}z}c^E_{n,m} = -c^H_{n,m} - \frac{k_m^2}{k_0^2\varepsilon_2}(b^H_{n,m} - c^H_{n,m}) - \sum_{m'\neq m}(\varepsilon^{-1})^{n,n}_{m,m'}\frac{k_m k_{m'}}{2k_0^2}(b^H_{n,m'} - c^H_{n,m'}), \tag{11}$$

$$\frac{\mathrm{d}}{\mathrm{d}z}b^H_{n,m} = \frac{k_m^2}{2}(b^E_{n,m} - c^E_{n,m}) - k_0^2\varepsilon_2 b^E_{n,m} - k_0^2\sum_{m'\neq m}(\varepsilon)^{n+1,n+1}_{m,m'}b^E_{n,m'}, \tag{12}$$

$$\frac{\mathrm{d}}{\mathrm{d}z}c^H_{n,m} = \frac{k_m^2}{2}(b^E_{n,m} - c^E_{n,m}) + k_0^2\varepsilon_2 c^E_{n,m} + k_0^2\sum_{m'\neq m}(\varepsilon)^{n-1,n-1}_{m,m'}c^E_{n,m'}, \tag{13}$$

where the matrix elements of ε responsible for the diffraction are given by the formulas

$$(\varepsilon)^{n,n}_{m,m'} = k_m\Delta_{m'}\int_0^\infty \varepsilon(r)J_n(k_m r)J_n(k_{m'}r)r\mathrm{d}r. \tag{14}$$

They can be evaluated analytically by taking into account the identity

$$\int_0^\infty J_n(k_{m'}r)J_n(k_m r)r\mathrm{d}r = \frac{\delta(k_{m'} - k_m)}{k_m}, \tag{15}$$

with δ representing the Dirac function, so that

$$\begin{aligned}(\varepsilon)^{n,n}_{m,m'} &= \varepsilon_2\delta_{m,m'} + (\varepsilon_d - \varepsilon_2)k_m\Delta_{m'}\int_0^R J_n(k_m r)J_n(k_{m'}r)r\mathrm{d}r \\ &= \varepsilon_2\delta_{m,m'} + (\varepsilon_d - \varepsilon_2)k_m\Delta_{m'}\frac{R}{k_m^2 - k_{m'}^2}[J_{n+1}(k_m R) \\ &\quad \times J_n(k_{m'}R)k_m - J_n(k_m R)J_{n+1}(k_{m'}R)k_{m'}].\end{aligned} \tag{16}$$

Taking into account that for small arguments $J_n(k_m R) \sim (k_m R)^{|n|}$, it is obvious that the off-diagonal terms of $(\varepsilon)^{n,n}$ responsible for the scattering are then proportional to

$$(\varepsilon)^{n,n}_{m,m'} \sim k_m\Delta_{m'}R^{2(|n|+1)}, \quad m' \neq m. \tag{17}$$

Thus, for small values of R, the most important terms (having the lowest power dependence on R) are the terms with $n=0$, for which the off-diagonal part of Eq. (16) takes the form

$$(\varepsilon_d-\varepsilon_2)k_m\Delta_{m'}\frac{R}{k_m^2-k_{m'}^2}[J_1(k_mR)J_0(k_{m'}R)k_m - J_0(k_mR)J_1(k_{m'}R)k_{m'}]. \quad (18)$$

If the single scattering is predominant, then the larger expression (18) with $m'=i$ (i being the index corresponding to the incident wave), the stronger the scattering into the mth direction, characterized by its k_m radial component of the wave vector. Let us consider at first the simplest, but the most common case: of normal incidence with $k_i=0$.

3. NORMAL INCIDENCE

In normal incidence ($m'\equiv i=0$, so that $k_{m'}=0$), expression (18) is considerably simplified into the form

$$(\varepsilon_d-\varepsilon_2)\Delta_i RJ_1(k_mR), \quad (19)$$

with $\Delta_i\equiv\Delta_0$. This expression has already enabled us to optimize the geometrical parameters of a single aperture or one surrounded by a circular rectangular-groove corrugation, an optimization that was confirmed by numerical results obtained using a rigorous electromagnetic method.[27] This approach can be developed further to study the near-and far-field diffraction by a small-radius aperture, which is the aim of this paper.

The only nonzero Fourier components of the electric field vector of a linearly polarized (say, along the x axis) wave incident normal to the surface are the components with $n=\pm1$, a condition that imposes the same symmetry on the diffracted field. Let us consider at first the set of Eqs. (10)–(13) for $n=1$, preserving only the terms proportional to R^2. Taking into account Eq. (17), the only off-diagonal term that remains is the one in the fourth equation [Eq. (23)]:

$$\frac{\mathrm{d}}{\mathrm{d}z}b_{1,m}^E=b_{1,m}^H-\frac{k_m^2}{2k_0^2\varepsilon_2}(b_{1,m}^H-c_{1,m}^H), \quad (20)$$

$$\frac{\mathrm{d}}{\mathrm{d}z}c_{1,m}^E=-c_{1,m}^H-\frac{k_m^2}{2k_0^2\varepsilon_2}(b_{1,m}^H-c_{1,m}^H), \quad (21)$$

$$\frac{\mathrm{d}}{\mathrm{d}z}b_{1,m}^H=\frac{k_m^2}{2}(b_{1,m}^E-c_{1,m}^E)-k_0^2\varepsilon_2 b_{1,m}^E, \quad (22)$$

$$\frac{\mathrm{d}}{\mathrm{d}z}c_{1,m}^H=\frac{k_m^2}{2}(b_{1,m}^E-c_{1,m}^E)+k_0^2\varepsilon_2 c_{1,m}^E+k_0^2\sum_{m'\neq m}(\varepsilon_d-\varepsilon_2)\times\Delta_{m'}RJ_1(k_mR)c_{1,m'}^E. \quad (23)$$

From here, it is easy to obtain a second-order equation for one of the amplitudes $c_{1,m}^E$:

$$\frac{\mathrm{d}^2}{\mathrm{d}z^2}c_{1,m}^E=-k_{m_z}^2c_{1,m}^E+RJ_1(k_mR)\sum_{m'}\eta_{mm'}c_{1,m'}^E,$$

$$k_{m_z}^2=k_0^2\varepsilon_2-k_m^2 \quad (24)$$

with

$$\eta_{mm'}=\left(1-\frac{k_m^2}{2k_0^2\varepsilon_2}\right)k_0^2(\varepsilon_d-\varepsilon_2)\Delta_{m'}. \quad (25)$$

One of the possible ways to obtain the general solution of Eq. (24) is traced in Appendix B. However, as will become evident later, it is more convenient to work with another set of unknowns, namely, the sum and the difference of the field components:

$$M_m^E=b_{1,m}^E-c_{1,m}^E,\quad P_m^E=b_{1,m}^E+c_{1,m}^E,$$

$$M_m^H=b_{1,m}^H-c_{1,m}^H,\quad P_m^H=b_{1,m}^H+c_{1,m}^H. \quad (26)$$

Taking into account that the only unperturbed (for $R=0$) wave is the one having amplitude $\hat{c}_i^E$, as follows from Eqs. (B10) and (B11), the set of differential equations for these new unknowns becomes, for $m\neq i$,

$$(M_m^E)'=P_m^H, \quad (27)$$

$$(P_m^H)'=-k_{m_z}^2M_m^E-RJ_1(k_mR)\Delta_\varepsilon\hat{c}_i^E, \quad (28)$$

$$(M_m^H)'=-k_0^2\varepsilon_2P_m^E+RJ_1(k_mR)\Delta_\varepsilon\hat{c}_i^E, \quad (29)$$

$$(P_m^E)'=\frac{k_{m_z}^2}{k_0^2\varepsilon_2}M_m^H, \quad (30)$$

with

$$\Delta_\varepsilon=k_0^2(\varepsilon_d-\varepsilon_2),$$

$$\hat{c}_i^E=\Delta_ic_{1,i}^E. \quad (31)$$

For the incident field components $m=i$, the equations are further simplified to take the form corresponding to the unperturbed system:

$$(M_i^E)'=P_i^H,\quad (P_i^H)'=-k_{i_z}^2M_i^E,$$

$$(M_i^H)'=-k_0^2\varepsilon_2P_i^E,\quad (P_i^E)'=\frac{k_{i_z}^2}{k_0^2\varepsilon_2}M_i^H. \quad (32)$$

The solution of these equations is trivial; it represents the incident wave propagating inside the nonpierced layer:

$$M_i^{E\pm}=\mathfrak{M}_i^{E\pm}\exp(\pm ik_{i_z}z),$$

$$P_i^{E\pm}=\mathfrak{P}_i^{E\pm}\exp(\pm ik_{i_z}z), \quad (33)$$

with similar expressions for M_i^H and P_i^H. The amplitudes $\mathfrak{M}_i^{E,H,\pm}$ and $\mathfrak{P}_i^{E,H,\pm}$ are determined using the boundary conditions.

Combining Eqs. (27) and (28), one obtains a second-order inhomogeneous equation for M_m^E:

$$(M_m^E)''=-k_{m_z}^2M_m^E-RJ_1(k_mR)\Delta_\varepsilon\hat{c}_i^E \quad (34)$$

which has a solution in the form (see Appendix B)

$$M_m^E = -\mathfrak{M}_m^{E\pm}\exp(\pm ik_{m_z}z) - RJ_1(k_mR)\frac{\Delta_\varepsilon}{k_{m_z}^2-\gamma_i^2}\hat{C}_i^{\pm}\exp(\pm i\gamma_i z), \tag{35}$$

where $\mathfrak{M}_m^{E\pm}$ are two unknown amplitudes subject to the boundary conditions between the different domains (cladding/pierced layer/substrate), and the "sources" are given by

$$\hat{C}_i^{\pm} = C_i^{\pm}\Delta_i = \Delta_i(\mathfrak{P}_i^{E\pm} - \mathfrak{M}_i^{E\pm})/2. \tag{36}$$

In a similar manner it is possible to obtain the solutions for M_m^H,

$$M_m^H = \mathfrak{M}_m^{H\pm}\exp(\pm ik_{m_z}z) - iRJ_1(k_mR)\frac{\Delta_\varepsilon\gamma_i}{k_{m_z}^2-\gamma_i^2}\hat{C}_i^{\pm}\exp(\pm i\gamma_i z), \tag{37}$$

while the form of P_m^H and P_m^E is obtained from the expressions for M_m^E and M_m^H and Eqs. (27) and (29).

A. Single Interface

Let us consider first a structure consisting of two semi-infinite media, the upper one homogeneous and the lower one pierced with an infinitely long hole, the two separated by a plane interface positioned at $z=0$. The first step is to determine the "source" terms as described in Eq. (36). To this end it is necessary to use the boundary conditions at $z=0$ for $m=i$, which demonstrates the utility of the substitution in Eq. (26). To the extent the r and θ components of the electric and magnetic field vectors are continuous across the interface $z=0$, so also are the amplitudes $b^{E,H}$ and $c^{E,H}$, and thus the amplitudes $M_i^{E,H}$ and $P_i^{E,H}$. As can be observed from Eqs. (27) and (28), the couple (M^E,P^H) forms a couple corresponding to the TE case in Cartesian coordinates, because when $M^H=0$, the z component of the electric field vanishes [refer to Eq. (5)], i.e., the field is a TE wave in cylindrical coordinates. Conversely, the case $M^E=0$ corresponds to a TM wave,because the z component of the magnetic field is zero.

The field above $z=0$ consists, for $m=i$, of two waves, one incident (described in detail in Appendix C and denoted by an upper case index I) and one reflected (upper case index R),

$$M_i^E = \mathfrak{M}_i^{E,I}\exp(-i\alpha_{i_z}z) + \mathfrak{M}_i^{E,R}\exp(i\alpha_{i_z}z), \tag{38}$$

with $\alpha_{i_z}=\sqrt{k_0^2\varepsilon_1-k_i^2}$, so that the continuity of M_i^E and $P_m^H=(M_m^E)'$ at $z=0$ simply gives the Fresnel formulas in TE polarization,

$$\mathfrak{M}_i^{E,I} + \mathfrak{M}_i^{E,R} = \mathfrak{M}_i^{E-}, \tag{39}$$

$$-\alpha_{i_z}\mathfrak{M}_i^{E,I} + \alpha_{i_z}\mathfrak{M}_i^{E,R} = -k_{i_z}\mathfrak{M}_i^{E-}; \tag{40}$$

by taking a sum and a difference of Eq. (39) and Eq. (40), the latter multiplied by α_{i_z} or k_{i_z}, we have

$$\mathfrak{M}_i^{E-} = T_i^{\mathrm{TE}}\mathfrak{M}_i^{E,I}, \quad T_i^{\mathrm{TE}} = \frac{2\alpha_{i_z}}{\alpha_{i_z}+k_{i_z}},$$

$$\mathfrak{M}_i^{E,R} = R_i^{\mathrm{TE}}\mathfrak{M}_i^{E,I}, \quad R_i^{\mathrm{TE}} = \frac{\alpha_{i_z}-k_{i_z}}{\alpha_{i_z}+k_{i_z}}. \tag{41}$$

In a similar manner the continuity of $(M^H,P^E=-M^{H'}/k_0^2\varepsilon)$ gives the Fresnel coefficients in TM polarization:

$$\mathfrak{M}_i^{H-} = T_i^{\mathrm{TM}}\mathfrak{M}_i^{H,I}, \quad T_i^{\mathrm{TM}} = \frac{2\alpha_{i_z}/\varepsilon_1}{\alpha_{i_z}/\varepsilon_1+k_{i_z}/\varepsilon_2},$$

$$\mathfrak{M}_i^{H,\mathcal{R}} = \mathcal{R}_i^{\mathrm{TM}}\mathfrak{M}_i^{H,I}, \quad \mathcal{R}_i^{\mathrm{TM}} = \frac{2\alpha_{i_z}/\varepsilon_1-k_{i_z}/\varepsilon_2}{\alpha_{i_z}/\varepsilon_1+k_{i_z}/\varepsilon_2}. \tag{42}$$

Although trivial, it is necessary to write down these formulas to compare them with the expressions of the scattered field $m\neq i$. The "source" amplitude is obtained through Eq. (36), $\hat{C}_i^-=\Delta_i(\mathfrak{P}_i^{E-}-\mathfrak{M}_i^{E-})/2$, where $\mathfrak{P}_i^{E-}=ik_{i_z}\mathfrak{M}_i^{H-}/k_0^2\varepsilon_2$, as obtained from Eqs. (30) and (33).

The continuity of M_m^E and P_m^H results in the following set, taking into account that there is no incident wave from above when $m\neq i$,

$$\mathfrak{M}_m^{E,R} = \mathfrak{M}_m^{E-} - RJ_1(k_mR)\frac{\Delta_\varepsilon}{k_{m_z}^2-\gamma_i^2}\hat{C}_i^-, \tag{43}$$

$$\alpha_{m_z}\mathfrak{M}_m^{E,R} = -k_{m_z}\mathfrak{M}_m^{E-} + RJ_1(k_mR)\frac{\Delta_\varepsilon\gamma_i}{k_{m_z}^2-\gamma_i^2}\hat{C}_i^-, \tag{44}$$

which gives for the amplitudes scattered in the upper medium

$$\mathfrak{M}_m^{E,R} = -\frac{RJ_1(k_mR)\Delta_\varepsilon\hat{C}_i^-}{(\alpha_{m_z}+k_{m_z})(k_{m_z}+\gamma_i)}, \tag{45}$$

$$\mathfrak{P}_m^{H,R} = i\alpha_{m_z}\mathfrak{M}_m^{E,R} = -\frac{i\alpha_{m_z}RJ_1(k_mR)\Delta_\varepsilon\hat{C}_i^-}{(\alpha_{m_z}+k_{m_z})(k_{m_z}+\gamma_i)}. \tag{46}$$

The continuity of M_m^H and P_m^E gives the second set of boundary conditions,

$$\mathfrak{M}_m^{H,R} = \mathfrak{M}_m^{H-} - iRJ_1(k_mR)\frac{\Delta_\varepsilon\gamma_i}{k_{m_z}^2-\gamma_i^2}\hat{C}_i^-,$$

$$\frac{\alpha_{m_z}}{\varepsilon_1}\mathfrak{M}_m^{H,R} = -\frac{k_{m_z}}{\varepsilon_2}\mathfrak{M}_m^{H-} + i\frac{k_{m_z}^2}{\varepsilon_2}RJ_1(k_mR)\frac{\Delta_\varepsilon}{k_{m_z}^2-\gamma_i^2}\hat{C}_i^-, \tag{48}$$

whence it is easy to determine the scattered amplitudes in reflection:

$$\mathfrak{M}_m^{H,R} = i\frac{k_{m_z}}{\varepsilon_2}\frac{RJ_1(k_mR)\Delta_\varepsilon\hat{C}_i^-}{(\alpha_{m_z}/\varepsilon_1+k_{m_z}/\varepsilon_2)(k_{m_z}+\gamma_i)}, \tag{49}$$

$$\mathfrak{P}_m^{E,R} = -i\frac{\alpha_{m_z}}{k_0^2\varepsilon_1}\mathfrak{M}_m^{H,R} = \frac{\alpha_{m_z}}{k_0^2\varepsilon_1}\frac{k_{m_z}}{\varepsilon_2}\frac{RJ_1(k_mR)\Delta_\varepsilon \hat{C}_i^-}{(\alpha_{m_z}/\varepsilon_1 + k_{m_z}/\varepsilon_2)(k_{m_z}+\gamma_i)}. \tag{50}$$

These expressions can be used to find the coefficients $b_{1,m}^E$, $c_{1,m}^E$, $b_{1,m}^H$, and $c_{1,m}^H$ of the scattered field, and to reconstruct the field vectors by use of Eqs. (3)–(8).

The other part of the solution, corresponding to the term $n=-1$ in the Fourier series, Eq. (1), can be obtained from the term with $n=1$, using the symmetry of the optogeometrical parameters and of the incident field in normal incidence (see Appendix C). In particular, when the incident field is polarized along the x axis, the following relations exist:

$$b_{-n,m}^E = -c_{n,m}^E,$$

$$b_{-n,m}^H = c_{n,m}^H. \tag{51}$$

1. Double Interface

A single pierced layer contains two interfaces with the surrounding semi-infinite layers and allows two sets of waves to propagate downward and upward inside the layer. This will double the number of unknown amplitudes and equations, but as with the reflection and transmission by a homogeneous layer, the use of matrix notations can significantly simplify the formulas.

Instead of Eqs. (39) and (40), one can write a single matrix equation on the upper interface ($z=t$) as

$$\begin{bmatrix} 1 & 1 \\ \alpha_{i_z} & -\alpha_{i_z} \end{bmatrix}\begin{pmatrix} \mathfrak{M}_i^{E,R} \\ \mathfrak{M}_i^{E,I} \end{pmatrix} = \begin{bmatrix} 1 & 1 \\ k_{i_z} & -k_{i_z} \end{bmatrix}\begin{bmatrix} \exp(ik_{i_z}t) & 0 \\ 0 & \exp(-ik_{i_z}t) \end{bmatrix}\begin{pmatrix} \mathfrak{M}_i^{E_+} \\ \mathfrak{M}_i^{E_-} \end{pmatrix}, \tag{52}$$

and on the lower ($z=0$) interface as

$$\begin{pmatrix} 1 \\ -\beta_{i_z} \end{pmatrix}\mathfrak{M}_i^{E,T} = \begin{bmatrix} 1 & 1 \\ k_{i_z} & -k_{i_z} \end{bmatrix}\begin{pmatrix} \mathfrak{M}_i^{E_+} \\ \mathfrak{M}_i^{E_-} \end{pmatrix}, \tag{53}$$

where $\beta_i = \sqrt{k_0^2\varepsilon_3 - k_i^2}$ and the superscript T indicates the field transmitted in the substrate.

As is well known, the last two equations can be used to introduce the transmission matrix of the system:

$$\mathbb{T}_i^{\mathrm{TE}} = \begin{bmatrix} \cos(k_{i_z}t) & \dfrac{i}{k_{i_z}}\sin(k_{i_z}t) \\ ik_{i_z}\sin(k_{i_z}t) & \cos(k_{i_z}t) \end{bmatrix}, \tag{54}$$

which links the field below and the field above the plane homogeneous layer:

$$\begin{bmatrix} 1 & 1 \\ \alpha_{i_z} & -\alpha_{i_z} \end{bmatrix}\begin{pmatrix} \mathfrak{M}_i^{E,R} \\ \mathfrak{M}_i^{E,I} \end{pmatrix} = \mathbb{T}_i^{\mathrm{TE}}\begin{pmatrix} 1 \\ -\beta_{i_z} \end{pmatrix}\mathfrak{M}_i^{E,T}. \tag{55}$$

This set of two equations gives the specular reflected and transmitted amplitudes as functions of the incident one,

$$\begin{pmatrix} \mathfrak{M}_i^{E,R} \\ \mathfrak{M}_i^{E,T} \end{pmatrix} = \left[\begin{array}{c|c} -1 & \\ -\alpha_{i_z} & \mathbb{T}_i^{\mathrm{TE}}\begin{pmatrix} 1 \\ -\beta_{i_z} \end{pmatrix} \end{array}\right]^{-1}\begin{pmatrix} 1 \\ -\alpha_{i_z} \end{pmatrix}\mathfrak{M}_i^{E,I}, \tag{56}$$

where the vertical bar separates the two columns of the matrix. In a similar manner, the TM components can be obtained by using the TM transmission matrix

$$\mathbb{T}_i^{\mathrm{TM}} = \begin{bmatrix} \cos(k_{i_z}t) & i\dfrac{\varepsilon_2}{k_{i_z}}\sin(k_{i_z}t) \\ i\dfrac{k_{i_z}}{\varepsilon_2}\sin(k_{i_z}t) & \cos(k_{i_z}t) \end{bmatrix}, \tag{57}$$

so that

$$\begin{pmatrix} \mathfrak{M}_i^{H,R} \\ \mathfrak{M}_i^{H,T} \end{pmatrix} = \left[\begin{array}{c|c} -1 & \\ -\dfrac{\alpha_{i_z}}{\varepsilon_1} & \mathbb{T}_i^{\mathrm{TM}}\begin{pmatrix} 1 \\ -\dfrac{\beta_{i_z}}{\varepsilon_3} \end{pmatrix} \end{array}\right]^{-1}\begin{pmatrix} 1 \\ -\dfrac{\alpha_{i_z}}{\varepsilon_3} \end{pmatrix}\mathfrak{M}_i^{H,I}. \tag{58}$$

Although trivial, we need these formulas in order to obtain the field components diffracted by the aperture, when $m \neq i$, because the amplitudes $\hat{C}_i^{\pm}$ act as sources for the scattered field, Eqs. (35) and (37). They can be obtained using the following procedure. First, Eq. (53) is used to find $\mathfrak{M}_i^{E\pm}$. A similar equation can be written for $\mathfrak{M}_i^{H\pm}$:

$$\begin{pmatrix} 1 \\ -\dfrac{\beta_{i_z}}{\varepsilon_3} \end{pmatrix}\mathfrak{M}_i^{H,T} = \begin{bmatrix} 1 & 1 \\ \dfrac{k_{i_z}}{\varepsilon_2} & -\dfrac{k_{i_z}}{\varepsilon_2} \end{bmatrix}\begin{pmatrix} \mathfrak{M}_i^{H+} \\ \mathfrak{M}_i^{H-} \end{pmatrix}. \tag{59}$$

Then one of Eq. (32) enables us to obtain $\mathfrak{P}_i^{E\pm}$.

Once the values of $\hat{C}_i^{\pm}$ are known from Eq. (36), we can proceed with the determination of the scattered field with $m \neq i$. The continuity of M_m^E and P_m^H at the upper and the lower interface gives the following matrix relations, using Eqs. (35) and (32):

$$\begin{aligned} \begin{pmatrix} 1 \\ \alpha_{m_z} \end{pmatrix}\mathfrak{M}_m^{E,R} = {} & \begin{pmatrix} 1 & 1 \\ k_{m_z} & -k_{m_z} \end{pmatrix}\begin{pmatrix} \exp(ik_{m_z}t) & 0 \\ 0 & \exp(-ik_{m_z}t) \end{pmatrix} \\ & \times\begin{pmatrix} \mathfrak{M}_m^{E+} \\ \mathfrak{M}_m^{E-} \end{pmatrix} + \frac{RJ_1(k_mR)\Delta_\varepsilon}{k_{m_z}^2 - \gamma_i^2}\begin{pmatrix} 1 & 1 \\ \gamma_{i_z} & -\gamma_{i_z} \end{pmatrix} \\ & \times\begin{pmatrix} \exp(i\gamma_{i_z}t) & 0 \\ 0 & \exp(-i\gamma_{i_z}t) \end{pmatrix}\begin{pmatrix} \hat{C}_i^+ \\ \hat{C}_i^- \end{pmatrix} \end{aligned} \tag{60}$$

$$\begin{aligned} \begin{pmatrix} 1 \\ -\beta_{m_z} \end{pmatrix}\mathfrak{M}_m^{E,T} = {} & \begin{pmatrix} 1 & 1 \\ k_{m_z} & -k_{m_z} \end{pmatrix}\begin{pmatrix} \mathfrak{M}_m^{E+} \\ \mathfrak{M}_m^{E-} \end{pmatrix} \\ & + \frac{RJ_1(k_mR)\Delta_\varepsilon}{k_{m_z}^2 - \gamma_i^2}\begin{pmatrix} 1 & 1 \\ \gamma_{i_z} & -\gamma_{i_z} \end{pmatrix}\begin{pmatrix} \hat{C}_i^+ \\ \hat{C}_i^- \end{pmatrix}. \end{aligned} \tag{61}$$

These are similar to the set of Eqs. (52) and (53), the only difference being in the source terms. It is possible to eliminate the unknown field amplitudes inside the layer, and the result is a system of two algebraic equations for

the reflected and transmitted scattered field amplitudes:

$$\begin{pmatrix}\mathfrak{M}_m^{E,R}\\ \mathfrak{M}_m^{E,T}\end{pmatrix}=\frac{RJ_1(k_mR)\Delta_\varepsilon}{k_{m_z}^2-\gamma_i^2}\left[\begin{array}{c|c}1 & \\ \alpha_{m_z} & \mathbb{T}_m^{\mathrm{TE}}\begin{pmatrix}-1\\ \beta_{m_z}\end{pmatrix}\end{array}\right]^{-1}\times\left\{\begin{bmatrix}1 & 1\\ \gamma_{i_z} & -\gamma_{i_z}\end{bmatrix}\begin{bmatrix}\exp(i\gamma_{i_z}t) & 0\\ 0 & \exp(-i\gamma_{i_z}t)\end{bmatrix}-\mathbb{T}_m^{\mathrm{TE}}\begin{bmatrix}1 & 1\\ \gamma_{i_z} & -\gamma_{i_z}\end{bmatrix}\right\}\begin{pmatrix}\hat{C}_i^+\\ \hat{C}_i^-\end{pmatrix}. \tag{62}$$

The continuity of M_m^H and P_m^E, together with Eqs. (37) and (32), results in a second set of equations corresponding to the TM polarization:

$$\begin{pmatrix}1\\ \dfrac{\alpha_{m_z}}{\varepsilon_1}\end{pmatrix}\mathfrak{M}_m^{H,R}=\begin{bmatrix}1 & 1\\ \dfrac{k_{m_z}}{\varepsilon_2} & -\dfrac{k_{m_z}}{\varepsilon_2}\end{bmatrix}\times\begin{bmatrix}\exp(ik_{m_z}t) & 0\\ 0 & \exp(-ik_{m_z}t)\end{bmatrix}\begin{pmatrix}\mathfrak{M}_m^{H+}\\ \mathfrak{M}_m^{H-}\end{pmatrix}+i\frac{RJ_1(k_mR)\Delta_\varepsilon}{k_{m_z}^2-\gamma_i^2}\begin{bmatrix}\gamma_{i_z} & -\gamma_{i_z}\\ \dfrac{k_{m_z}^2}{\varepsilon_2} & \dfrac{k_{m_z}^2}{\varepsilon_2}\end{bmatrix}\times\begin{bmatrix}\exp(i\gamma_{i_z}t) & 0\\ 0 & \exp(-i\gamma_{i_z}t)\end{bmatrix}\begin{pmatrix}\hat{C}_i^+\\ \hat{C}_i^-\end{pmatrix}, \tag{63}$$

$$\begin{pmatrix}1\\ -\dfrac{\beta_{m_z}}{\varepsilon_3}\end{pmatrix}\mathfrak{M}_m^{H,T}=\begin{bmatrix}1 & 1\\ \dfrac{k_{m_z}}{\varepsilon_2} & -\dfrac{k_{m_z}}{\varepsilon_2}\end{bmatrix}\begin{pmatrix}\mathfrak{M}_m^{H+}\\ \mathfrak{M}_m^{H-}\end{pmatrix}+i\frac{RJ_1(k_mR)\Delta_\varepsilon}{k_{m_z}^2-\gamma_i^2}\begin{bmatrix}\gamma_{i_z} & -\gamma_{i_z}\\ \dfrac{k_{m_z}^2}{\varepsilon_2} & \dfrac{k_{m_z}^2}{\varepsilon_2}\end{bmatrix}\begin{pmatrix}\hat{C}_i^+\\ \hat{C}_i^-\end{pmatrix}, \tag{64}$$

which can be solved in the same way as in TE polarization:

$$\begin{pmatrix}\mathfrak{M}_m^{H,R}\\ \mathfrak{M}_m^{H,T}\end{pmatrix}=i\frac{RJ_1(k_mR)\Delta_\varepsilon}{k_{m_z}^2-\gamma_i^2}\left[\begin{array}{c|c}1 & \\ \dfrac{\alpha_{m_z}}{\varepsilon_1} & \mathbb{T}_m^{\mathrm{TM}}\begin{pmatrix}-1\\ \dfrac{\beta_{m_z}}{\varepsilon_3}\end{pmatrix}\end{array}\right]^{-1}\times\left\{\begin{bmatrix}\gamma_{i_z} & -\gamma_{i_z}\\ \dfrac{k_{m_z}^2}{\varepsilon_2} & \dfrac{k_{m_z}^2}{\varepsilon_2}\end{bmatrix}\begin{bmatrix}\exp(i\gamma_{i_z}t) & 0\\ 0 & \exp(-i\gamma_{i_z}t)\end{bmatrix}-\mathbb{T}_m^{\mathrm{TM}}\begin{bmatrix}\gamma_{i_z} & -\gamma_{i_z}\\ \dfrac{k_{m_z}^2}{\varepsilon_2} & \dfrac{k_{m_z}^2}{\varepsilon_2}\end{bmatrix}\right\}\begin{pmatrix}\hat{C}_i^+\\ \hat{C}_i^-\end{pmatrix}. \tag{65}$$

As can be expected, these equations contain Fabry–Perot resonances in the direction of propagation of the incident wave (γ_i), in the matrices $\mathbb{T}_i^{\mathrm{TE,TM}}$, and in the direction of the mth scattered field due to the exponential terms in $\mathbb{T}_m^{\mathrm{TE,TM}}$. However, as discussed in detail in Section 5, since the approximate method is based on the single-scattering approximation, it does not include the waveguide mode that can propagate inside the hole that behaves as a hollow cylindrical waveguide. Although for small-radius holes this mode is evanescent, its decay constant is smaller than the decay of the wave tunneling through the metallic layer without hole, and thus the mode contributes to increasing the transmission significantly. Future work is required in order to force the approximate method to take into account this waveguide mode. This argument demonstrates the limitations of the approximate methods, whereas the rigorous ones take into account *a priori* all the waves and modes (or at least they are assumed to do this).

4. INCLINED INCIDENCE

The case when $k_i \neq 0$ can be treated in a similar manner. The difference is that the reduced symmetry of the incident wave limits the restriction to have scattered field with −1st and +1st Fourier components only. However, a detailed analysis of Eqs. (10)–(13) using Eq. (17) shows that the terms proportional to powers of R less than or equal to 2 remain only in three cases:

(i) for $n=-1$ in Eq. (12),
(ii) for $n=1$ in Eq. (13),
(iii) for $n=0$ in Eqs. (10) and (11).

A. $n=\pm 1$

The case with $n=1$ differs from the normal incidence only by replacing the term

$$(\varepsilon_d-\varepsilon_2)\Delta_i RJ_1(k_mR) \tag{66}$$

with that presented in Eq. (18),

$$(\varepsilon_d-\varepsilon_2)k_m\Delta_{m'}\frac{R}{k_m^2-k_{m'}^2}[J_1(k_mR)J_0(k_{m'}R)k_m-J_0(k_mR)J_1(k_{m'}R)k_{m'}], \tag{67}$$

in all the formulas obtained in Section (3).

B. $n=0$

The case with $n=0$ will preserve the off-diagonal matrix elements responsible for the scattering in Eqs. (10) and (11):

$$(\varepsilon)_{m,m'}^{0,0}=\varepsilon_2\delta_{mm'}+(\varepsilon_d-\varepsilon_2)k_m\Delta_{m'}\frac{R}{k_m^2-k_{m'}^2}[J_1(k_mR)\times J_0(k_{m'}R)k_m-J_0(k_mR)J_1(k_{m'}R)k_{m'}], \tag{68}$$

so that its inverse (within the terms proportional to R^2) has elements

$$(\varepsilon^{-1})^{0,0}_{m,m'} = \frac{1}{\varepsilon_2}\delta_{mm'} - \frac{(\varepsilon_d - \varepsilon_2)}{\varepsilon_2^2} k_m \Delta_{m'} \frac{R}{k_m^2 - k_{m'}^2} [J_1(k_m R) \times J_0(k_{m'} R) k_m - J_0(k_m R) J_1(k_{m'} R) k_{m'}], \quad (69)$$

i.e., the off-diagonal terms are almost (within the factor $-1/\varepsilon_2^2$) equal to the off-diagonal terms in Eq. (13) for $n = 1$, and thus they can be treated in a similar manner. To this end, let us redefine the unknowns introduced in Eq. (26)

$$M_m^E = b_{0,m}^E - c_{0,m}^E, \quad P_m^E = b_{0,m}^E + c_{0,m}^E,$$

$$M_m^H = b_{0,m}^H - c_{0,m}^H, \quad P_m^H = b_{0,m}^H + c_{0,m}^H. \quad (70)$$

Equations (27)–(30) are changed into the following set:

$$(M_m^E)' = P_m^H, \quad (71)$$

$$(P_m^H)' = -k_{m_z}^2 M_m^E, \quad (72)$$

$$(M_m^H)' = -k_0^2 \varepsilon_2 P_m^E, \quad (73)$$

$$(P_m^E)' = \frac{k_{m_z}^2}{k_0^2 \varepsilon_2} M_m^H + R \mathfrak{G}_{m,i}(R) M_i^H, \quad (74)$$

with

$$\mathfrak{G}_{m,i}(R) = \frac{(\varepsilon_d - \varepsilon_2)}{\varepsilon_2^2} \frac{k_m \Delta_i}{k_m^2 - k_i^2} [J_1(k_m R) J_0(k_i R) k_m - J_0(k_m R) J_1(k_i R) k_i]. \quad (75)$$

The second-order differential equation obtained for M_m^H,

$$(M_m^H)'' = -k_{m_z}^2 M_m^H - k_0^2 \varepsilon_2 R \mathfrak{G}_{m,i}(R) M_i^H, \quad (76)$$

has a form similar to Eq. (34) and its solution is similar in form to Eq. (35):

$$M_m^H = \mathfrak{M}_m^{H\pm} \exp(\pm i k_{m_z} z) - \frac{k_0^2 \varepsilon_2 R \mathfrak{G}_{m,i}(R)}{k_{m_z}^2 - \gamma_i^2} M_i^{H\pm} \exp(\pm i \gamma_i z). \quad (77)$$

As for normal incidence, at first the amplitudes for $m = i$ are evaluated for the incident wave as if the aperture did not exist. Then the boundary conditions, the same as in Section 3, are applied, depending on whether a single interface or a layer is considered, in order to obtain the wave components in reflection (and transmission). Finally, Eqs. (3)–(8) are used to reconstruct the vector-field components.

5. DISCUSSION AND DOMAIN OF VALIDITY

Any theoretical method, numerical or analytical, has its own domain of validity, and in the case of modeling diffraction of light, approximate analytical solutions usually have stronger limitations. This is the case with the method presented in the previous sections. Fortunately, it is possible to compare its results with existing numerical results in order to get some idea of these limitations. Before starting the comparison, let us remember one of the implicit limitations of the method, already discussed in Section 1. Both the rigorous and the analytical method use the same basis (Fourier–Bessel) of functions inside the pierced layer and in the hole, functions which are continuous with respect to r and θ. This fact makes it impossible to extend their validity to an infinitely conducting screen, because inside the screen the electric field is zero and no basis can be used simultaneously inside it and inside the aperture. Thus no direct comparison is possible with the results of Bethe, only an asymptotic one.

Let us first consider a normally incident plane wave linearly polarized along the x direction with unit amplitude of its electric field vector and wavelength $\lambda = 500$ nm. Figure 2 represents the k_m dependence of the normalized field amplitudes $\mathfrak{M}_m^E/k_m$ and $\mathfrak{M}_m^H/k_m$ in the case of a single interface (infinitely profound hole with $R = 10$ nm) between air and aluminum. The calculations were done by using the rigorous numerical method and the analytical expressions presented in Subsection 3.A. Several important features can be observed in the figure. First, inside the optical cone ($k_m < k_0$), the amplitudes are slowly varying functions of k_m, which can be understood by taking into account that, for small R, $J_1(k_m R)/k_m \approx$ const. as a function of k_m, as observed in Eqs. (45) and (49). The coefficients gradually decrease in amplitude in the region of evanescent waves in the cladding. Second, there is a sharp anomaly for the TM amplitudes, corresponding to the excitation of a surface plasmon wave on the metal–air interface. This anomaly can be expected from Eqs. (49) and (50) when their denominator is close to zero, a condition equivalent to the condition of existence of the plasmon surface wave on a plane metallic–dielectric interface. Both the rigorous and the analytical methods predict the same position and type of anomaly. Third, as can be observed, the spectral (in k_m) behavior of the field amplitudes is the same, the only difference between the rigorous and the analytical method lying in the factor of ~2 between the two sets of results. In any case, within this factor, the near- and the far-field pattern of

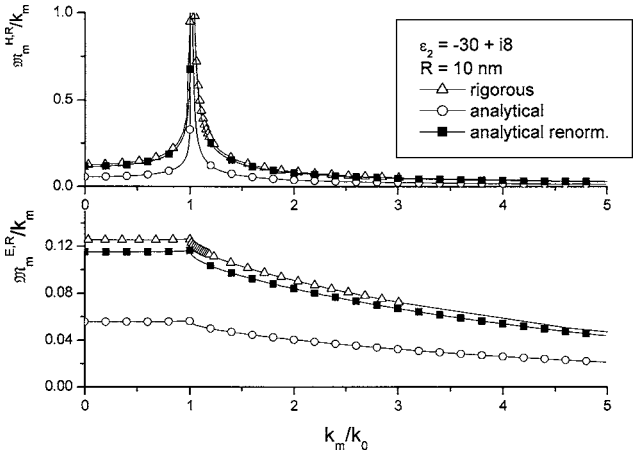

Fig. 2. Spectral amplitudes $\mathfrak{M}_m^{E,R}/k_m$ and $\mathfrak{M}_m^{H,R}/k_m$ in reflection as function of k_m for an aluminum infinitely thick screen, air as cladding and inside the aperture, $R = 10$ nm, wavelength $\lambda = 500$ nm. Open triangles, rigorous results; open circles, analytical results; solid squares, renormalized analytical results.

the field diffracted by the aperture has the same appearance when calculated by either method.

The difference is systematic, as will appear in the next figures, because the local electric field amplitude of the source field (namely, $\hat{C}_i^-$ in the case of a single interface) is approximately twice smaller than the local source field as determined in the Born approximation, presented in Appendix D. This corresponds to the fact that the field inside the aperture is not equal to the field inside the layer as if the aperture were absent. This difference is well known in the Green's function approach and it increases considerably with the optical contrast between the aperture and the unperturbed layer. In the case of an infinitely thin and long cylinder acting as a perturbation, the multiplicative factor applied to the unperturbed field is simply equal to (Ref. 26 and Appendix D)

$$\hat{C}_i^- \to \frac{2\varepsilon_2}{\varepsilon_2 + \varepsilon_d}\hat{C}_i^-, \tag{78}$$

which gives the missing value of 2 when $|\varepsilon_2| \gg |\varepsilon_d|$. As can be observed in Fig. 2, the results of the analytical method renormalized by using relation (78) become much closer to the rigorous results, a tendency observed in the next figures, too.

As an illustration, Fig. 3 presents the x dependence of the amplitude of the electric field of the scattered field (without the field reflected by the plane interface in the absence of aperture). Here the optical index of the substrate is different, but the radius of the aperture is the same as in the previous case. The field map is calculated at a distance of 1 nm above the surface. One can observe a sharp peak at the hole edges, a behavior, well-known now, due to current discontinuities and charge accumulation on the hole walls perpendicular to the incident field polarization. The analytical results follow the behavior of the rigorous ones, the only difference lying in the same factor ~ 2 when the correction due to relation (78) is not taken into account. The comparison for the far-field radiation pattern is made in the Section 6 and the same tendency is observed.

A natural question arises with respect to the assumptions on which the analytical method is based, namely, the range of R in which the terms proportional to R^2 are predominant. There are two important results that give rise to this question. First, Bethe's theory for diffraction by an aperture in a perfectly conducting screen leads to a dependence on the field amplitude $\sim R^3$ rather than R^2. Second, the results in Ref. 18 indicate that the transmission through subwavelength apertures in real metals has behavior different from R^4 (faster) (in intensity). In order to check the limits of validity of the R^2 approximation, Fig. 4 presents the R dependence of the scattered electric field intensity in the cladding (without the fields incident and that specularly reflected by the plane layer) evaluated on the axis of the aperture at 1 nm height. While for values of $R > 30$ nm the rigorous results start to increase more rapidly than the analytical data, the R^4 behavior in intensity (i.e., R^2 in amplitude) for smaller radii is observed quite well in the two curves. The correction obtained by using Eq. (78) brings the analytical results close to the rigorous ones. On the other hand, if we try to take

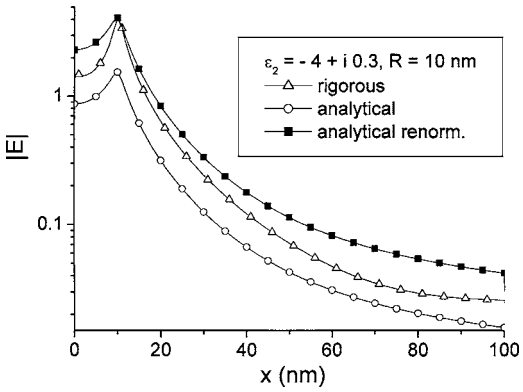

Fig. 3. Amplitude of the diffracted electric field calculated along the x axis at a height $z = 1$ nm. Low-conductivity screen material. The other parameters are as in Fig. 2.

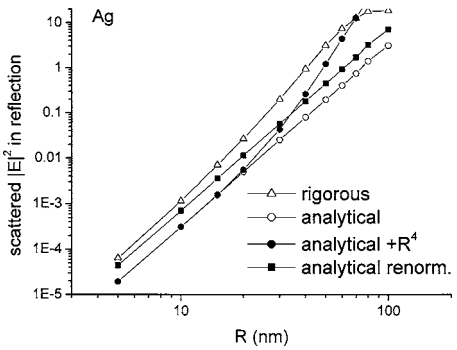

Fig. 4. Electric field intensity just over the center of the opening of the aperture ($z = 1$ nm) as a function of the aperture radius R. Infinitely thick silver screen, air as cladding and inside the aperture, $\lambda = 500$ nm. Open triangles, rigorous results; open circles, analytical results; solid squares, renormalized analytical results; solid circles, analytical results by taking into account terms proportional to both R^2 and R^4.

into account one of the higher-order terms, the curve of solid circles in Fig. 4 shows that its role is negligible below $R = 20$ nm, while for larger R its behavior differs significantly from that of the rigorous and analytical curves. This is an example—quite typical for approximate methods—where every attempt to introduce higher-order terms in order to improve the validity of the method plays the opposite role.

A further discrepancy between the rigorous and the analytical results limits significantly the validity of the latter when used in transmission. As already mentioned, the study presented in Ref. 18 indicates that when considering the transmission through an aperture pierced in a real-metal film, its R dependence differs a lot from the results based on the Kirchhoff approximation, which predicts that for a perfectly conducting layer, the transmitted intensity grows as R^4, showing much more rapid variation for real-metal screens. Figure 5(a) represents the rigorous and the analytical results for a 200 nm thick silver screen (electric field intensity is calculated on the axis of the aperture at a distance of 1 nm below the screen). As expected, the field intensity given by the analytical method is proportional to R^4 both in reflection and transmission, while the rigorous results present a steeper R dependence than in Fig. 4, which fact has found its

explanation[20] in the role played by the fundamental mode supported inside the hollow metallic waveguide formed by the aperture.[24,28] Although this mode is evanescent below its cut-off radius, its decay constant decreases rapidly as R increases, and is smaller than the decay constant of the wave tunneling directly through the layer.

Unfortunately, our analytical method cannot take this mode into account because of the explicit limitations imposed by the fact that only single scattering is taken into account. This can be observed in Eqs. (27)–(30), further on in Eqs. (46)–(50) for a single interface, and in Eqs. (62) and (65) for a layer. The only source of scattering is the field amplitudes that propagate inside the layer in the absence of an aperture. Once diffracted, the scattered field propagates without perturbation and it plays no role as a secondary source. Thus the approximation to take into account only the lowest-order in R terms is equivalent to neglecting higher-order scattering, and within this model additional resonances inside the aperture, different from the unperturbed modes, cannot exist. Since the correct description of the waveguide mode inside the cylindrical aperture requires multiple reflections on the walls of the aperture to be taken into account, it does not exist in this approximation. An alternative is to combine our approach with the modal approach proposed in Ref. 24, but this requires additional work outside the scope of this paper.

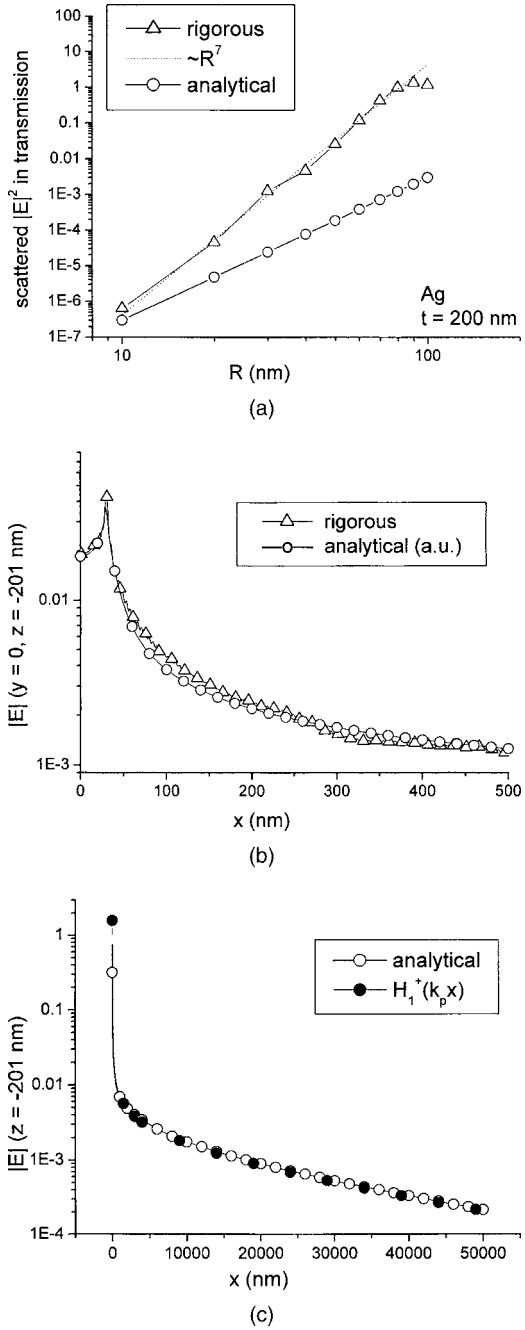

Fig. 5. (a) Electric field intensity after the screen in the center of the aperture ($z=-201$ nm) for a silver screen with thickness $t=200$ nm. Open circles, analytical results; open triangles rigorous results; dotted line $\sim R^7$. (b) Comparison between the rigorous and the analytical (in a.u.) results of the electric field amplitude along the x axis 1 nm below the aperture for $R=30$ nm. (c) Comparison between the analytical results and the Hankel function $H_1^+(k_p x)$, representing the plasmon field traveling away from the aperture.

On the other hand, the unperturbed [existing in the absence of the aperture] plasmon surface waves obtained from the denominator in Eqs. (49) and (50), the Fabry–Perot resonances, and the planar-layer waveguide modes described by the transmission matrices $\mathbb{T}_i^{\mathrm{TE,TM}}$ and $\mathbb{T}_m^{\mathrm{TE,TM}}$ are taken into account even in the single-scattering approach. This provides a necessary condition to correctly represent the field map in the near- and far-field regions in transmission within a multiplicative factor that is due to the different transmission intensity shown in Fig. 5(a). An illustration is presented in Fig. 5(b) that gives the map (in the direction parallel to the incident polarization) of the modulus of the diffracted electric field at a distance of 1 nm below the aperture, which has a 30 nm radius and is pierced in a silver screen 200 nm thick. As can be observed, the approximate method not only represents quite well the details of the diffracted field inside the aperture but also represents the sharp field increase close to the borders of the aperture.

Moreover, when looking far from the aperture but still close to the metal surface, the field behavior is predominantly determined by the plasmon surface wave going away from the aperture, which can be observed in Fig. 5(c), where a comparison is made between the results of the approximate method (open circles) and the simple formula[27]

$$|E(x)| \sim H_1^+(k_p x), \tag{79}$$

where H_1^+ stands for the first-order Hankel function and $k_p=k_0(1.0667+i0.002548)$ is the plasmon propagation constant on the silver–air interface at $\lambda=500$ nm. Without losses, k_p is real and the asymptotic expansion of relation (79) is proportional to $1/\sqrt{k_p x}$; thus the total flux of energy of the plasmon remains constant over the entire metal–dielectric interface, whatever the distance from the aperture. However, for real metals, due to the imaginary part $k_p''>0$ of the plasmon propagation constant, the modulus of the electric field behaves asymptotically for large values of x as $\exp(-k_p'' x)/\sqrt{k_p x}$, i.e., it decreases very rapidly due to absorption.

The absence of higher-order scattering in the approximate method can provide explanation for the systematic difference between the rigorous and the analytical results

[without the correction in eq. (78)] observed in Figs. 2–4. If the higher-order scattering plays a nonnegligible role even in the limit $R \to 0$, any single-scattering approach will give results different from the true ones. And really, as already discussed, the cylindrical waveguide mode excited inside the aperture cannot be described through a single-scattering approach. Another argument supporting the importance of multiple scattering can be found in the fact that when the radius of the aperture is smaller, the scattering points on its walls lie closer to each other, which increases the importance of the second- and higher-order scattering of the field.

To check this hypothesis, we have used the rigorous method as a tool in a numerical experiment. It is possible to modify the code so that only single scattering is taken into account. This is done by using only two different spectral (in the k space) components of the field. The difference from the analytical approach is important, because the rigorous method truncated in this manner is not limited to the lowest terms in R. Thus a set of "single-scatter but all-power series in R" results can be obtained to be compared with the analytical approach containing the lowest power in R approximation. As observed in Fig. 6, the analytical and the extremely truncated set of results coincide for small values of R, which confirms the hypothesis that the single-scatter assumption is responsible for the systematically lower analytical results.

All these studies indicate that the possibility of compensating for the difference between the results of the analytical and the rigorous methods lies outside the limits of the assumptions constituting the base of the analytical method. And indeed, the correction as proposed in relation (78) comes from a completely different approach that takes into account the singularity of the Green's tensor. As already mentioned, the correction factor ≈ 2 in a very large domain of materials, from dielectrics through poorly to highly conducting metals. Figures. 6(a) and 6(b) show the dependence of one of the spectral amplitudes on the real part of ε_2 in a large interval of $\mathrm{Re}(\varepsilon_2)$, positive or negative. The results are obtained using the analytical and the rigorous approach, the "extremely truncated" rigorous method, but the analytical results are also corrected by the factor due to the renormalized Born approximation according to relation (78). One can observe in Fig. 6(b) a spectacular amelioration of the analytical results when this correction is introduced. In the lossy metallic case [Fig. 6(a)], we have a residual difference, probably due to the fact that in real metals the correction factor (78) is not yet sufficient.

Another remark that we should make is that the asymptotic behavior of the analytical expressions when $\mathrm{Re}(\varepsilon_2) \to \pm\infty$ is wrong, because the basis of functions (Fourier–Bessel) used to express the field inside the pierced layer cannot be used when the layer becomes infinitely conducting and thus the electric field inside it vanishes. And indeed, as can be found from the equations in Section 3, the spectral amplitudes decline as $1/\sqrt{\varepsilon_2}$ for large values of ε_2 so that the scattering vanishes within this model. In order to study the perfectly conducting case in our model, it is necessary to include higher-order terms with respect to R, which reminds us that the results of Bethe are correct to predict a behavior proportional to R^3 in amplitude, as far as, according to our model, the terms in R^2 vanish when $\mathrm{Re}(\varepsilon_2) \to \pm\infty$.

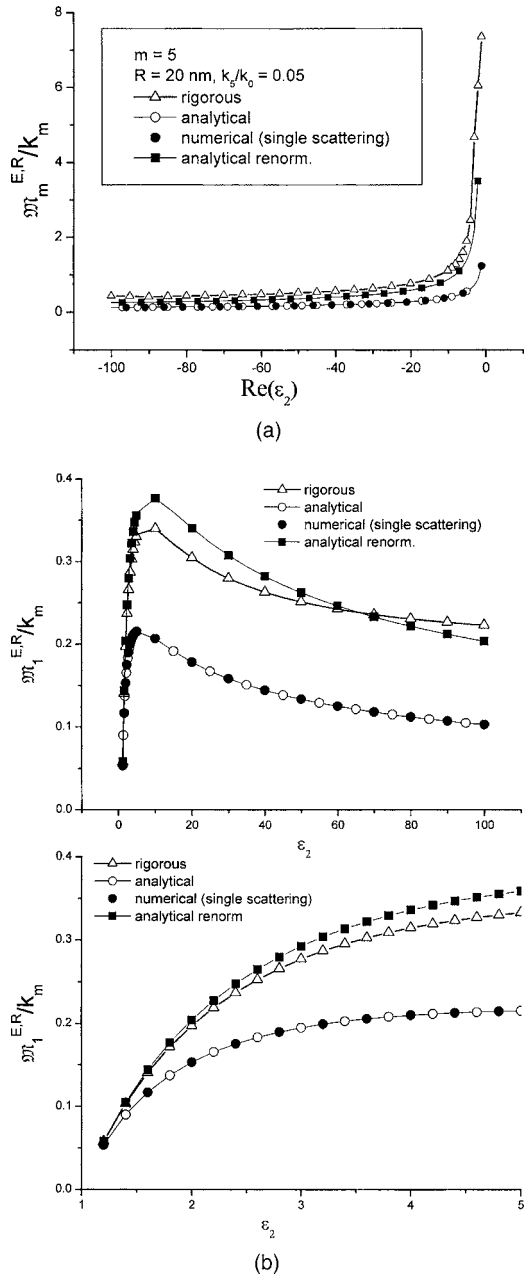

Fig. 6. Spectral amplitude $\mathfrak{M}_m^{E,R}/k_m$ for $k_m/k_0 = 0.05$ calculated in reflection using the rigorous (open triangles), the analytical (open circles), the renormalized analytical (solid squares), and the "extremely truncated" (solid circles) rigorous method, as a function of the real part of permittivity of the screen for $R = 20$ nm and $\lambda = 500$ nm. (a) Metals; (b) dielectrics, with a zoom (bottom panel) close to the origin.

More generally, each perturbative method (including those based on the Born approximation) for which the source of the perturbation is the field inside the unperturbed medium cannot treat materials with infinite permittivity because the latter implies vanishing fields inside; thus any correction factor similar to relation (78) cannot give some nonvanishing contribution. However, as shown in Section 6, the analytical approach permits us to obtain qualitatively (within this undetermined source fac-

tor) the far-field pattern predicted by Bethe in the case of a perfectly conducting screen.

There are several conclusions to be drawn from the results of this section:

1. The analytical results describe well the scattering in reflection by a single aperture, both in the direct and the inverse space, for different dielectric and poorly and highly conducting metal screens.
2. There is a systematic error in the analytical results. This error is due to the single-scattering approximation on which the method is based, which cannot take into account the change of the incident field amplitude induced by the aperture. However, the renormalized Born approximation provides a correction which takes into account the difference between the incident field with and without the aperture, a correction that compensates almost totally the difference between the analytical and the rigorous results in case of dielectrics, and reduces this difference for lossy metals. Within this residual difference, the analytical method correctly predicts the excitation of surface plasmons on the metallic–dielectric interface, or of waveguide modes inside the plane layer if it is a dielectric.
3. Due to the single-scattering approximation, it is not possible to take into account the evanescent mode inside the hollow metallic waveguide formed by the aperture inside the metallic layer. This limitation introduces a significant error in the determination of the field amplitudes in transmission (as observed in Fig. 5), while the analytical method is still able to correctly predict the form of the near- or far-field distributions.

6. FAR-FIELD DIRECTIVITY

The limitations of the analytical method must not be understated, and it has to be used with caution, preferably in parallel with some rigorous method to enable comparison. The development of the electromagnetic grating theories during the entire 20th century has proven this. However, by taking these precautions, it is possible to use the analytical method in order to better understand some numerical and experimental properties of diffraction by apertures, because, although the rigorous numerical methods provide the necessary tools for modeling light diffraction, quite often they serve as a black box that provides results but no physical reasoning and understanding.

One of the interesting problems in light diffraction by small apertures is the angular distribution of the diffracted field, because its directivity is important in increasing the efficiency of detection in biophysical and physicochemical experiments of laser-induced fluorescence and Raman scattering inside nanovolumes, aiming to study single molecules. While the classical theory of Bethe is valid only for infinitely conducting screens, it is sufficient to break down the intuitive expectations from the scalar point sources that the radiation pattern of the diffracted field will be uniformly distributed in the entire half space. Indeed, the prediction, for example, in normal incidence is that the field scattered by the aperture will correspond to the field of a magnetic dipole in the free space positioned at the aperture in the plane of the screen and having direction perpendicular to the incident field polarization. In that simple case, the field (and the radiation pattern) in the plane perpendicular to the dipole and thus containing the incident electric field vector will be uniformly distributed, i.e., not depending on the angle of diffraction, as would be expected from the scalar diffraction. However, in the other plane of diffraction perpendicular to the incident electric field vector, the diffraction pattern is strongly nonuniform and tends to zero at grazing to the screen direction, as can be expected from the field of the magnetic dipole close to its axis. The same behavior can be expected from the theory presented by Jackson,[2] although he insists that it is not valid for small apertures.

Recent studies of the diffraction pattern generated by an aperture in finitely conducting metallic screens indicate that even for small apertures, the angular distribution of the radiation is not flat, either inside the plane of incident field polarization or perpendicular to it.[18] Numerical modeling links this deviation from the predictions of Bethe with the excitation of a surface plasmon wave in the direction lying in the plane of incident field polarization. However, it is not clear how this plasmon can be radiated from the surface apart from the aperture borders in order to increase the angular directivity of the radiation pattern, if the surface is without defects.

To analyze the reasons for this deviation from Bethe's theory, we use both the numerical and the analytical method, first, to study the properties of the radiation pattern and to compare the results of the two methods and second, to try to find the physical reason by simplifying and analyzing the formulas obtained in the previous sections in the case of highly conducting screens. In order to avoid the handicap of the analytical method in transmission, discussed in detail with respect to Fig. 5, the analysis is made in reflection. Then the results of the rigorous method are shown in transmission to confirm the behavior common to the case in reflection.

The following figures represent the radial component of the angular density of the Poynting vector, defined as

$$P_\rho = \frac{1}{2\rho^2}(\mathbf{E} \times \bar{\mathbf{H}}) \cdot \hat{\rho}, \tag{80}$$

where $\rho = \sqrt{r^2+z^2}$ is the distance between the center of the aperture and the observation point and $\hat{\rho}$ is the corresponding unit vector. In the far-field region the values of P_ρ have to be independent of ρ, but they will vary with the azimuthal angle θ and with the polar angle ψ between $\hat{\rho}$ and the z axis (Fig. 1). We consider two planes of diffraction, the first one parallel to the direction of incident wave polarization, having $\theta = 0$, and the second one perpendicular to the first and characterized by $\theta = 90°$.

To obtain the far-field characteristics of the field it is sufficient to go several tens of wavelength away from the aperture. However, in the case of finitely conducting metals, the surface plasmon wave that can be excited on the metal–dielectric interfaces is only slightly attenuated along the surface and can be detected near the surface if the distance from the origin is not sufficiently long. This can be observed in Fig. 7, which shows the polar angle distribution of P_ρ in the plane $\theta = 0$ for three values of ρ. At

shorter distances from the aperture, a significant increase of P_ρ is observed in directions of propagation grazing the surface, with values that gradually decrease with the distance, as follows from the discussion following relation (79). This is why, in what follows, we work at $\rho=50\,\mu$m. On the other hand, Fig. 7 shows that, except for the influence of the plasmon wave, the values of P_ρ as normalized in Eq. (80) do not depend on ρ.

Let us first consider larger apertures, where the theory of Bethe predicts variations of P_ρ in both planes $\theta=0$ and 90°. Figure 8 represents several results for $R=100$ nm. Two different metals are considered, real A1 and an artificial one having permittivity multiplied by 100. In the plane $\theta=0$, the screen of higher conductivity shows an angular dependence that is weaker than for A1 screen (open squares). Both curves are much wider than the angular dependence in the perpendicular plane, $\theta=90°$, which has values varying insignificantly with the conductivity, as observed further on. The theory of Bethe predicts this behavior, but more correct formulas are available for greater radii by Jackson (Ref. 2, p. 492) as

$$P_p \propto \left| \frac{J_1\left(\sin\frac{2\pi}{\lambda}R\sin\psi\right)}{\sin\frac{2\pi}{\lambda}R\sin\psi} \right|^2 (\cos^2\psi+\sin^2\psi\cos^2\theta), \tag{81}$$

where the terms in the second set of parentheses determine the polar angle factor, different in the two planes of observation:

$$(\cos^2\psi+\sin^2\psi\cos^2\theta)=\begin{cases}1, & \theta=0\\ \cos^2\psi, & \theta=90°\end{cases}. \tag{82}$$

While for highly conducting material the three methods [Jackson's Eq. (81), the analytical, and the rigorous method] predict variation of the radiation pattern in the two planes $\theta=0$, 90°, one can observe that the directivity of $P_\rho(\psi)$ increases when $|\varepsilon_2|$ decreases, a result that cannot be obtained from Eq. (81).

Figure 9 shows the variation of the behavior of $P_\rho(\psi)$ for

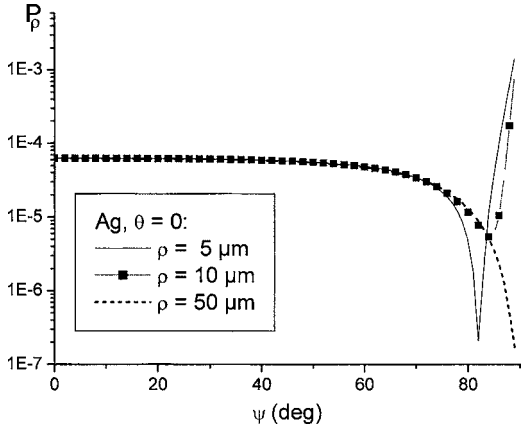

Fig. 7. Polar angle distribution of the power radiated in reflection in a radial direction ρ (see Fig. 1) for three different values of ρ lying in the plane of incident polarization $\theta=0$. Infinitely thick silver screen, $R=10$ nm, and $\lambda=500$ nm.

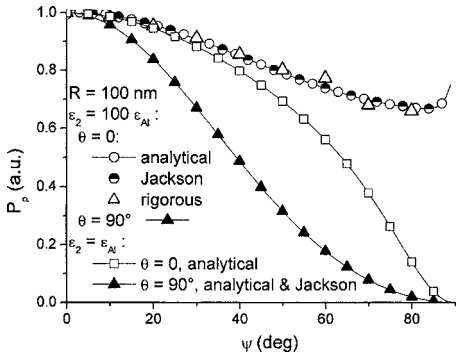

Fig. 8. Variation of the power radiated in reflection with the polar angle ψ in the two planes $\theta=0$, 90° for two different metals and relatively wide aperture, $R=100$ nm, and $\lambda=500$ nm. For $\theta=0$ and very highly conducting material ($\varepsilon_2=100\varepsilon_{A1}$): open circles, analytical results; half-filled circles, Jackson's formula; open triangles, rigorous results. Open squares, real metal (aluminum) and $\theta=0$. Solid triangles, in the plane $\theta=90°$ for both aluminum and very high conductivity.

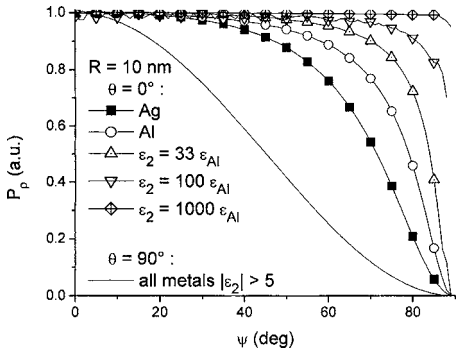

Fig. 9. Similar to Fig. 8 but for much smaller aperture, $R=10$ nm, and screen materials (metals) with values of permittivity increasing in modulus as indicated at bottom of graph. Symbols, $\theta=0$; solid curve $\theta=90°$.

different values of the screen permittivity when the radius of the aperture is reduced to 10 nm. Relation (81) predicts that for these dimensions of the aperture and the wavelength, the curve $P_\rho(\psi)$ will be practically flat when $\theta=0$, while it is evident that even for these small radii, the real metals present a variation that is flattened when $|\varepsilon_2|\to\infty$. In the other plane of observation, $P_\rho(\psi)$ is practically independent of $|\varepsilon_2|$ and its dependence is represented by the $\cos^2\psi$ factor in relation (81).

As already discussed, the stronger directivity in the plane containing the incident electric field vector that exists for finitely conducting metals has been already invoked and explained by the propagation of the existing surface plasmon. However, whereas this is true in the near-field region, where the plasmon "extends" the electromagnetic field along the metallic surface in the direction parallel to the incident electric field vector (Fig. 7), the experimentally observed stronger directivity in the far-field region cannot be due to the surface plasmon, whose field decreases away from the aperture, even in the vicinity of the metallic interface.

In order to confirm or reject the plasmon role in the enhancement of the far-field directivity, we present in Fig. 10 results similar to those given in Fig. 9, but for a dielectric instead of metallic material surrounding the aper-

ture. Similar behavior in the plane $\theta=0$ is obtained, the dependence becomes flatter with the increase of $|\varepsilon_2|$, i.e., for materials with smaller $|\varepsilon_2|$ the directivity is higher, a fact that cannot be explained by a surface wave that is not supported by a single dielectric–dielectric interface, if all the materials are lossless. Contrary to the metallic case, the directivity in the other plane $\theta=90°$ depends on the refractive index and increases with $|\varepsilon_2|$; see Fig. 11.

All this points up that the increased directivity of the radiation pattern in the plane of incident polarization, when finite-conductivity materials (metal or dielectric) are used, needs an explanation different from plasmon surface waves, which play a clear role in the near-field distribution but cannot be invoked in the far-field pattern, all the more in the case when no surface wave can exist. To this end, let us return to the basic equations of the analytical method and try to observe what happens when $|\varepsilon_2|$ becomes much larger than ε_d. In that case, the expressions of $\mathfrak{M}_m^E$ and $\mathfrak{M}_m^H$ in Eqs. (45) and (49) can be easily simplified to

$$\mathfrak{M}_m^{E,R} = T_m^{\mathrm{TE,inv}} \mathfrak{M}_m^{E,D_M}, \tag{83}$$

$$\mathfrak{M}_m^{H,R} = T_m^{\mathrm{TM,inv}} \mathfrak{M}_m^{H,D_M}, \tag{84}$$

where

$$T_m^{\mathrm{TE,inv}} = \frac{2k_{m_z}}{\alpha_{m_z} + k_{m_z}},$$

$$T_m^{\mathrm{TM,inv}} = \frac{2k_{m_z}/\varepsilon_2}{\alpha_{m_z}/\varepsilon_1 + k_{m_z}/\varepsilon_2}, \tag{85}$$

are the Fresnel coefficients for transmission from the substrate into the cladding, and

$$\mathfrak{M}_m^{E,D_M} = -\frac{RJ_1(k_m R)\Delta_\varepsilon \hat{C}_i^-}{2k_{m_z}(k_{m_z} + \gamma_i)}, \tag{86}$$

$$\mathfrak{M}_m^{H,D_M} = i\frac{RJ_1(k_m R)\Delta_\varepsilon \hat{C}_i^-}{2(k_{m_z} + \gamma_i)}, \tag{87}$$

will be shown to be equal to the k_m components of a magnetic dipole lying just below the interface when $|\varepsilon_2| \gg \varepsilon_d$.

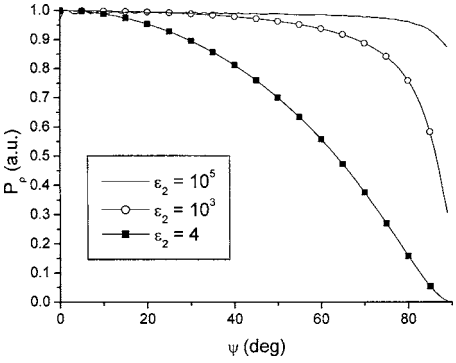

Fig. 10. Variation of P_ρ with polar angle ψ in the plane $\theta=0$ for different dielectric screens with gradually increasing permittivity as given in the inset. $R=10$ nm, $\lambda=500$ nm.

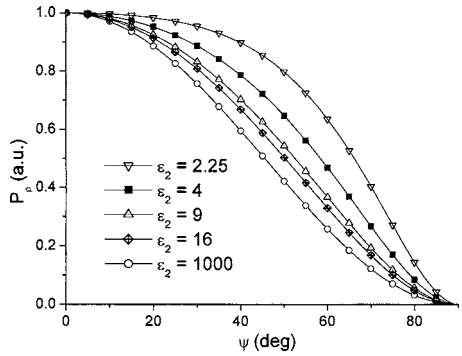

Fig. 11. Same as Fig. 10 but in the $\theta=90°$ plane, perpendicular to the incident wave polarization.

In the far-field zone $k_m \leqslant k_o$, so that $k_{m_z} + \gamma_i \approx 2\sqrt{\varepsilon_2}$, and taking into account that

$$J_1(k_m R) \approx k_m R/2, \tag{88}$$

Eq. (86) takes the form

$$\mathfrak{M}_m^{E,D_M} \approx k_m \frac{\varepsilon_2}{8|\varepsilon_2|} R^2 \hat{C}_1^- = k_m k_2 \frac{Z_2 D_M}{8\pi}, \tag{89}$$

which represents (see Appendix E) the $\mathfrak{M}_m^{E,D_M}$ component of the field radiated by a magnetic dipole oriented in the y direction (i.e., lying in the plane of the screen and perpendicular to the incident field polarization) with a dipole moment equal to

$$D_M = \pi R^2 \frac{\varepsilon_2}{|\varepsilon_2|} \frac{\hat{C}_i^-}{k_2 Z_2}. \tag{90}$$

It can be demonstrated in a similar manner that

$$\mathfrak{M}_m^{H,D_M} \approx -ik_m \frac{k_0^2|\varepsilon_2|}{k_{m_z}} k_2 \frac{Z_2 D_M}{8\pi} \tag{91}$$

represents the $\mathfrak{M}_m^{H,D_M}$ component of the same dipole. The other two field components $\mathfrak{P}_m^E$ and $\mathfrak{P}_m^H$ can be derived from the expressions of $\mathfrak{M}_m^E$ and $\mathfrak{M}_m^H$ using Eqs. (27) and (29).

As already mentioned, these results have the following physical interpretation: the incident field induces a scattered field inside the layer pierced by the aperture. For high values of $|\varepsilon_2|$, the scattered field can be considered as the field of a magnetic dipole having a dipole moment given by Eq. (90) that is proportional to the incident field inside the layer. This incident field is equal to $\hat{C}_i^- = \Delta_i(\mathfrak{P}_i^{E-} - \mathfrak{M}_i^{E-})/2$, following Eq. (36) and corrected by using the Born approximation by the factor given in relation (78). The emission of the dipole in the plane perpendicular to the dipole direction (i.e., the plane containing the incident electric field vector) is uniform angularly. The field emitted by this dipole is transmitted through the layer surface following Eqs. (83)–(85). However, the Fresnel transmission coefficients are angularly dependent, so that the emission in the cladding will depend on the angle of transmission, i.e., even in the plane $\theta=0$ one can expect an angular variation of the emission.

On the other hand, the angular dependence of the Fresnel coefficient decreases with increasing contrast be-

tween the cladding and the layer with the aperture, so that one can expect that with $|\varepsilon_2|$ increasing, the radiation pattern in the plane $\theta=0$ will become less dependent on the polar angle ψ. Indeed, as demonstrated in Fig. 12(a) for metals and in Fig. 12(b) for dielectrics, larger values of $|\varepsilon_2|$ lead to a weaker angular dependence of $\mathfrak{P}_m^{E,R}/k_m$, for example, conversely, when $|\varepsilon_2|$ decreases, the coefficient $\mathfrak{P}_m^{E,R}/k_m$ decreases when $k_m/k_0 \to \sqrt{\varepsilon_1}$, i.e., when $\psi \to 90°$, and this decrease is faster (and thus the directivity stronger) for smaller values of $|\varepsilon_2|$.

By taking the asymptotic expressions of the transmission coefficients in Eqs. (85),

$$\mathrm{T}_m^{\mathrm{TE,inv}} \to 2,$$

$$\mathrm{T}_m^{\mathrm{TM,inv}} \to 2\frac{\varepsilon_1}{\alpha_{m_z}}\frac{k_{m_z}}{\varepsilon_2}, \tag{92}$$

and multiplying them by the dipole field components as given in relations (89) and (91), it is easily found that the components of the field radiated in the cladding are given by

$$\mathfrak{M}_m^{E,R} \simeq \frac{k_m}{4}R^2\frac{\varepsilon_2\hat{C}_i^-}{|\varepsilon_2|}, \qquad \mathfrak{P}_m^{H,R} \simeq i\alpha_{m_z}\frac{k_m}{4}R^2\frac{\varepsilon_2\hat{C}_i^-}{|\varepsilon_2|}, \tag{93}$$

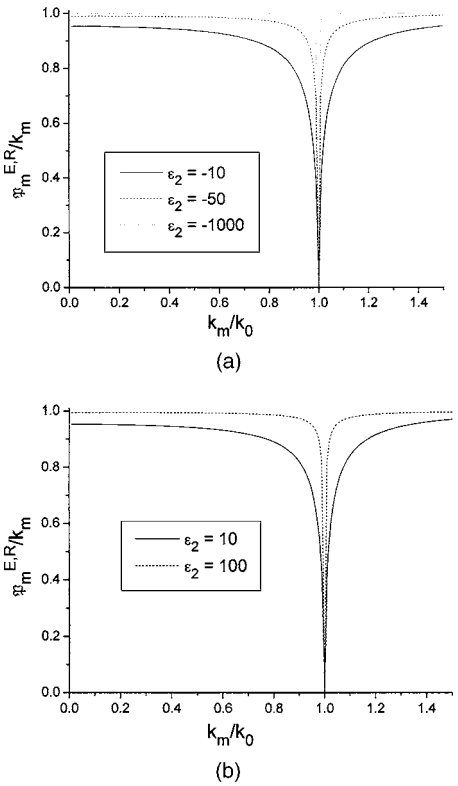

Fig. 12. Variation of the spectral amplitude $\mathfrak{P}_m^{E,R}/k_m$ as a function of k_m for (a) metals, (b) dielectrics for different permittivity of the screen as shown in the insets. $R=10$ nm, $\lambda=500$ nm.

$$\mathfrak{M}_m^{H,R} \approx -i\frac{k_0^2\varepsilon_1}{\alpha_{m_z}}\frac{k_m}{4}R^2\frac{\varepsilon_2\hat{C}_i^-}{|\varepsilon_2|}, \quad \mathfrak{P}_m^{E,R} \simeq -\frac{k_m}{4}R^2\frac{\varepsilon_2\hat{C}_i^-}{|\varepsilon_2|}. \tag{94}$$

These are nothing but the field components of a magnetic dipole emitting in the cladding as if the substrate were absent, having a direction perpendicular to the incident field polarization, and lying in the plane of the layer surface. That is, these expressions represent the equivalence between the scattering by an aperture in an infinitely conducting screen in normal incidence and the radiation pattern of a magnetic dipole. It has to be pointed out that when $|\varepsilon_2| \to \infty$, the source (unperturbed) field has an amplitude $\hat{C}_i^- \sim 1/\sqrt{\varepsilon_2} \to 0$ with or without the Born correction of relation (78), so that another expression is necessary to obtain a nonvanishing scattering. One possibility is the approximation used by Jackson: Assume that the source field is equal to the incident field inside the aperture.

An approximation better than relations (93) and (94) can be obtained by going back to relation (88). Taking it into account, relations (93) and (94) have to be replaced by the following expressions:

$$\mathfrak{M}_m^{E,R} \simeq RJ_1(k_mR)\frac{\varepsilon_2\hat{C}_i^-}{2|\varepsilon_2|},$$

$$\mathfrak{P}_m^{H,R} \simeq i\alpha_{m_z}RJ_1(k_mR)\frac{\varepsilon_2\hat{C}_i^-}{2|\varepsilon_2|}, \tag{95}$$

$$\mathfrak{M}_m^{H,R} \approx -i\frac{k_0^2\varepsilon_1}{\alpha_{m_z}}RJ_1(k_mR)\frac{\varepsilon_2\hat{C}_i^-}{2|\varepsilon_2|},$$

$$\mathfrak{P}_m^{E,R} \simeq -RJ_1(k_mR)\frac{\varepsilon_2\hat{C}_i^-}{2|\varepsilon_2|}, \tag{96}$$

which are close to the expressions presented by Jackson and valid in a much larger interval of radius values.

Let us briefly summarize the conclusions that can be drawn for the diffraction by small apertures:

1. While a scalar wave is diffracted by a small aperture in a manner to produce a uniform angularly scattered field, a linearly polarized vector field is diffracted by an aperture in a finitely conducting metal or nonconducting dielectric screen in a pattern that is nonuniform angularly. This nonuniformity is more pronounced in the plane of diffraction perpendicular to the incident vector polarization.

2. When the permittivity ε_2 of the screen increases in modulus, the scattered field becomes more and more uniform in the plane of the incident electric field vector, and in the limits of $|\varepsilon_2| \to \infty$ the radiation in this plane pattern is completely uniform (a conclusion already drawn from Bethe's theory).

3. With increase of the aperture dimensions, the directivity of the scattered field increases [this conclusion can already be found for perfectly conducting screens; for example, it is contained in Eq. (81), taken from Jackson's book[2]].

APPENDIX A

The Helmholtz equations in cylindrical coordinates for $E_{r,n}$ and $E_{\theta,n}$ are written

$$\Delta E_{\theta,n} - \frac{E_{\theta,n}}{r^2} + \frac{2in}{r}E_{r,n} + k^2 E_{\theta,n} = 0,$$

$$\Delta E_{r,n} - \frac{E_{r,n}}{r^2} - \frac{2in}{r}E_{\theta,n} + k^2 E_{r,n} = 0, \tag{A1}$$

where $k=\omega\sqrt{\mu_0\varepsilon_0\varepsilon}$ is the wavenumber and the Laplacian has the form

$$\Delta = \frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}\frac{\partial^2}{\partial\theta^2} + \frac{\partial^2}{\partial z^2}. \tag{A2}$$

By making two linear combinations of $E_{r,n}$ and $E_{\theta,n}$, it is possible to decouple Eqs. (A1). Defining

$$E_{\mp,n} = E_{\theta,n} \pm iE_{r,n}, \tag{A3}$$

Eqs. (A1) take the form

$$\left[\frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} - \frac{(n\pm 1)^2}{r^2} + k^2 + \frac{\partial^2}{\partial z^2}\right]E_{\pm,n}(r,z) = 0. \tag{A4}$$

Here it is easy to identify the generating equation for Bessel functions of first kind and integer number equal to $n\pm 1$; this enables the writing of the general form for $E_{\pm,n}$ in a Bessel-function basis:

$$E_{-,n} = \int_0^\infty \hat{c}_n^E(k_r,z)J_{n-1}(k_r r)k_r \mathrm{d}k_r = \sum_{m=0}^{\infty} c_{n,m}^E(z)J_{n-1}(k_m r)\Delta_m,$$

$$E_{+,n} = \int_0^\infty \hat{b}_n^E(k_r,z)J_{n+1}(k_r r)k_r \mathrm{d}k_r = \sum_{m=0}^{\infty} b_{n,m}^E(z)J_{n+1}(k_m r)\Delta_m, \tag{A5}$$

where $\{k_m\}$ is the set of discretized values of k_r, $\Delta_m = k_{m+1}-k_m$, and the right-hand side of Eq. (A5) represents the discretized integrals, which is necessary in the numerical treatment.

Equations (3) and (4) appear immediately when taking into account Eqs. (A3) and (A5).

The third line of Eqs. (2) permits obtaining the form of z components of electric and magnetic field vectors in the form given in Eqs. (5) and (8) by using the Bessel-function identities. Let us consider $H_{z,n} = 1/i\omega\mu_0(\partial E_{\theta,n}/\partial r + E_{\theta,n}/r - (in/r)E_{r,n})$. Substituting $E_{\theta,n}$ and $E_{r,n}$ from Eqs. (3) and (4) and taking into account the identities

$$\left(\frac{\mathrm{d}}{\mathrm{d}r} + \frac{n+1}{r}\right)J_{n+1}(k_m r) = k_m J_n(k_m r),$$

$$\left(\frac{\mathrm{d}}{\mathrm{d}r} - \frac{n-1}{r}\right)J_{n-1}(k_m r) = -k_m J_n(k_m r), \tag{A6}$$

one obtains Eq. (8).

A similar treatment can be done for $H_{r,n}$, $H_{\theta,n}$, and $\varepsilon E_{z,n}$ in order to obtain the other three equations in the set (3)–(8). However, in the inhomogeneous region containing the aperture we have $\varepsilon=\varepsilon(r)$, which fact requires special treatment. Using Eq. (5), it is possible to express E_z in a form similar to that of Eqs. (A5):

$$k_0^2\varepsilon(r)E_{z,n} = k_0^2\varepsilon(r)\sum_m e_{n,m}(z)J_n(k_m r)\Delta_m \overset{\text{Eq. (5)}}{=} i\sum_m [b_{n,m}^H(z) - c_{n,m}^H(z)]J_n(k_m r)k_m\Delta_m. \tag{A7}$$

Multiplying the second equivalence by $k_{m'}rJ_n(k_{m'}r)$, integrating over r, and using the orthogonality relations [Eq. (15)], one obtains that

$$k_0^2\sum_m (\varepsilon)_{m',m}^{n,n}e_{n,m}(z) = \frac{i}{\omega}k_{m'}[b_{n,m'}^H(z) - c_{n,m'}^H(z)], \tag{A8}$$

where the matrix element of ε is given in Eq. (16). Inverting the matrix, the Fourier–Bessel coefficients of E_z are

$$e_{n,m}(z) = \frac{i}{k_0^2}\sum_{m'}(\varepsilon^{-1})_{m,m'}^{nn}k_{m'}[b_{n,m'}^H(z) - c_{n,m'}^H(z)]. \tag{A9}$$

A similar expression is obtained for $H_{z,n} = \sum_{m=0}^{\infty}h_{n,m}(z)J_n(k_m r)\Delta_m$ with

$$h_{n,m}(z) = \frac{1}{i\omega\mu_0}[b_{n,m}^H(z) - c_{n,m}^H(z)]. \tag{A10}$$

Combining the first two lines of Eq. (2), replacing $E_{n,z}$ from Eqs. (A7) and (A9), and taking into account the identity

$$\frac{\mathrm{d}}{\mathrm{d}r}J_n(k_m r) = \frac{k_m}{2}[J_{n-1}(k_m r) - J_{n+1}(k_m r)], \tag{A11}$$

one obtains the following set:

$$\frac{\partial}{\partial z}\sum_m c_{n,m}^E(z)J_{n-1}(k_m r)\Delta_m = -\sum_m c_{n,m}^H(z)J_{n-1}(k_m r)\Delta_m - \frac{k_m}{2k_0^2}\sum_{m,m'}(\varepsilon^{n,n})_{m,m'}^{-1}k_{m'}[b_{n,m'}^H(z) - c_{n,m'}^H(z)]J_{n-1}(k_m r), \tag{A12}$$

$$\frac{\partial}{\partial z}\sum_m b_{n,m}^E(z)J_{n+1}(k_m r)\Delta_m = \sum_m b_{n,m}^H(z)J_{n+1}(k_m r)\Delta_m - \frac{k_m}{2k_0^2}\sum_{m,m'}(\varepsilon^{n,n})_{m,m'}^{-1}k_{m'}[b_{n,m'}^H(z) - c_{n,m'}^H(z)]J_{n+1}(k_m r). \tag{A13}$$

Multiplying Eq. (A12) by $k_m rJ_{n-1}(k_m r)$ and Eq. (A13) by $k_m rJ_{n+1}(k_m r)$, integrating them over r, and using the orthogonality relation Eq. (15), one obtains, respectively, Eq. (11) and Eq. (10).

When the first two equations for the z derivatives of **H** in Eq. (2) are combined, the following equations are written for $b_{n,m}^H$ and $c_{n,m}^H$ by using Eqs. (A10) and (A11):

$$\frac{\partial}{\partial z}\sum_m c^H_{n,m}(z)J_{n-1}(k_m r)\Delta_m = \sum_m \frac{k_m^2}{2}[b^E_{n,m}(z) - c^E_{n,m}(z)]J_{n-1}(k_m r)\Delta_m + k_0^2\varepsilon(r)\sum_m c^E_{n,m}(z)J_{n-1}(k_m r)\Delta_m, \tag{A14}$$

$$\frac{\partial}{\partial z}\sum_m b^H_{n,m}(z)J_{n+1}(k_m r)\Delta_m = \sum_m \frac{k_m^2}{2}[b^E_{n,m}(z) - c^E_{n,m}(z)]J_{n+1}(k_m r)\Delta_m - k_0^2\varepsilon(r)\sum_m b^E_{n,m}(z)J_{n+1}(k_m r)\Delta_m. \tag{A15}$$

When multiplying Eq. (A14) by $k_m r J_{n-1}(k_m r)$ and Eq. (A15) by $k_m r J_{n+1}(k_m r)$ and integrating over r, one finds Eq. (13) and Eq. (12), respectively.

APPENDIX B

Let us consider a set of second-order linear differential equations of the form

$$c_m''(z) = -k_{m_z}^2 c_m(z) + RJ_1(k_m R)\sum_n \eta_{mn}c_n(z). \tag{B1}$$

Its general solution can be searched in the form of waves propagating along $\pm z$:

$$c_m(z) = \sum_n C^{\pm}_{mn}\exp(\pm i\gamma_n z), \tag{B2}$$

which leads to the following equation

$$\sum_n C^{\pm}_{mn}\gamma_n^2\exp(\pm i\gamma_n z) = \sum_{p,q}(k^2_{m_z}\delta_{mp} - RJ_1(k_m R)\eta_{mp}) \times C^{\pm}_{pq}\exp(\pm i\gamma_q z). \tag{B3}$$

In so far as this equation has to be valid for each z, one obtains a typical eigenvalue/eigenvector problem for γ_n^2 and $C^{\pm}_{mn}$:

$$C^{\pm}_{mn}\gamma_n^2 = \sum_p (k^2_{m_z}\delta_{mp} - RJ_1(k_m R)\eta_{mp})C^{\pm}_{pn}. \tag{B4}$$

The eigenvalues γ_n^2 of the matrix $(k^2_{m_z}\delta_{mp} - RJ_1(k_m R)\eta_{mp})$ can easily be obtained within the order of R^2; they are simply equal to the diagonal elements of this matrix:

$$\gamma_m^2 \simeq k^2_{m_z} - RJ_1(k_m R)\eta_{mm}. \tag{B5}$$

Then Eq. (B4) results in

$$C^{\pm}_{mn} \simeq \frac{RJ_1(k_m R)}{k^2_{m_z} - \gamma_n^2}\sum_p \eta_{mp}C^{\pm}_{pn}. \tag{B6}$$

On the other hand, the diagonal elements $C^{\pm}_{mm}$ represent the waves that can propagate without aperture [see Eq. (B1) with $R=0$], while the off-diagonal elements correspond to the scattering by the aperture; thus

$$C^{\pm}_{mn} \simeq C^{\pm}_{mm}\delta_{mn} + O(R^2). \tag{B7}$$

In that case, all the terms $RJ_1(k_m R)\eta_{mp}C^{\pm}_{pn}$ in Eq. (B6) will be of the order of R^4, except for the terms with $p=n$. Thus Eq. (B6) leads to the following expression:

$$C^{\pm}_{mn} = \frac{RJ_1(k_m R)\eta_{mn}}{k^2_{m_z} - \gamma_n^2}C^{\pm}_{nn}. \tag{B8}$$

By taking into account Eqs. (B2), (B7), and (B8), the final form of the general solution becomes

$$c_m(z) = C^{\pm}_{mm}\exp(\pm i\gamma_m z) + RJ_1(k_m R)\sum_{n\neq m}\frac{\eta_{mn}}{k^2_{m_z} - \gamma_n^2}C^{\pm}_{nn}\exp(\pm i\gamma_n z). \tag{B9}$$

Let us now consider the case when a single wave (with $m=i$) is incident on the structure. This means that it would be the only propagating wave if the aperture were not existing and thus all the other waves will have amplitudes of the order of $O(R^2)$, i.e., Eq. (B9) can be rewritten separately for the incident wave component ($m=i$) and for the scattered waves ($m\neq i$):

$$c_i(z) = C_i^{\pm}\exp(\pm i\gamma_i z), \tag{B10}$$

$$c_{m\neq i}(z) = \frac{RJ_1(k_m R)\eta_{mi}}{k^2_{m_z} - \gamma_i^2}C_i^{\pm}\exp(\pm i\gamma_i z), \tag{B11}$$

with $C_i^{\pm} \equiv C^{\pm}_{ii}$ being the unperturbed incident field amplitude in the corresponding medium.

APPENDIX C

Let us consider an arbitrary wave vector $\mathbf{k}$ with components $\mathbf{k}_r$ and $\mathbf{k}_z$ in the xOy plane and on the z axis, respectively. If θ_0 is the polar angle of $\mathbf{k}_r$ in the xOy plane, then an incident plane wave is represented in a cylindrical basis in the form

$$\mathbf{E}^i = \mathbf{E}_0\sum_{n=-\infty}^{\infty} i^n J_n(k_r r)\exp[in(\theta-\theta_0)]\exp(-ik_z z). \tag{C1}$$

The form of the amplitude vector $\mathbf{E}_0$ in the cylindrical spatial basis depends on the incident field polarization. For a linearly polarized incident wave, the Cartesian components of $\mathbf{E}_0$ are constant:

$$\mathbf{E}_0 = E_{0x}\hat{\mathbf{x}} + E_{0y}\hat{\mathbf{y}} + E_{0z}\hat{\mathbf{z}}. \tag{C2}$$

Using the expression of the Cartesian unit vectors in cylindrical coordinates, we obtain

$$\mathbf{E}_0 = E_{0x}(\hat{\mathbf{r}}\cos\theta - \hat{\boldsymbol{\theta}}\sin\theta) + E_{0y}(\hat{\mathbf{r}}\sin\theta + \hat{\boldsymbol{\theta}}\cos\theta) + E_{0z}\hat{\boldsymbol{z}}, \tag{C3}$$

$$\mathbf{E}_0 = \hat{\mathbf{r}}\left[\left(\frac{E_{0x}}{2} + \frac{E_{0y}}{2i}\right)\exp(i\theta) + \left(\frac{E_{0x}}{2} - \frac{E_{0y}}{2i}\right)\exp(-i\theta)\right]$$

$$+ \hat{\boldsymbol{\theta}}\left[\left(-\frac{E_{0x}}{2i} + \frac{E_{0y}}{2}\right)\exp(i\theta) + \left(\frac{E_{0x}}{2i} + \frac{E_{0y}}{2}\right)\exp(-i\theta)\right] + \hat{\mathbf{z}}E_{0z}. \tag{C4}$$

From these expressions, one can deduce the expression of the electromagnetic components $E_r^i = \mathbf{E}^i \cdot \hat{\mathbf{r}}$ and $E_\theta^i = \mathbf{E}^i \cdot \hat{\boldsymbol{\theta}}$ of the incident plane wave. For the first

$$E_r^i = \sum_{n=-N}^{N}\left(\frac{E_{0x}}{2} + \frac{E_{0y}}{2i}\right)i^n\exp(-in\theta_0)J_n(k_r r)\exp[i(n+1)\theta] \times\exp(-ik_z z) + \sum_{n=-N}^{N}\left(\frac{E_{0x}}{2} - \frac{E_{0y}}{2i}\right)i^n\exp(-in\theta_0)J_n(k_r r) \times\exp[i(n-1)\theta]\exp(-ik_z z). \tag{C5}$$

A simple translation of the subscript n by ±1 then leads to

$$E_r^i = \sum_{n=-N+1}^{N+1} i\left(\frac{E_{0x}}{2i} - \frac{E_{0y}}{2}\right)i^{n-1}\exp[-i(n-1)\theta_0]J_{n-1}(k_r r) \times\exp(in\theta)\exp(-ik_z z) + \sum_{n=-N-1}^{N-1} i\left(\frac{E_{0x}}{2i} + \frac{E_{0y}}{2}\right)i^{n+1} \times\exp[-i(n+1)\theta_0]J_{n+1}(k_r r)\exp(in\theta)\exp(-ik_z z) \tag{C6}$$

and, in a similar way,

$$E_\theta^i = \sum_{n'=-N+1}^{N+1}\left(-\frac{E_{0x}}{2i} + \frac{E_{0y}}{2}\right)i^{n'-1}\exp[-i(n'-1)\theta_0]J_{n'-1}(k_r r) \times\exp(in'\theta)\exp(-ik_z z) + \sum_{n''=-N-1}^{N-1}\left(\frac{E_{0_x}}{2i} + \frac{E_{0_y}}{2}\right)i^{n''+1} \times\exp[-i(n''+1)\theta_0]J_{n''+1}(k_r r)\exp(in''\theta)\exp(-ik_z z). \tag{C7}$$

By comparing these equations with Eqs. (3) and (4), we obtain the expression of the components of the incident electromagnetic field in a Fourier–Bessel basis in terms of their Cartesian components:

$$b_{n,m}^{E,I}(z=0) = \left(\frac{E_{0_x}}{2i} + \frac{E_{0_y}}{2}\right)\frac{i^{n+1}}{\Delta_m}\exp[-i(n+1)\theta_0], \tag{C8}$$

$$c_{n,m}^{E,I}(z=0) = \left(-\frac{E_{0_x}}{2i} + \frac{E_{0_y}}{2}\right)\frac{i^{n-1}}{\Delta_m}\exp[-i(n-1)\theta_0]. \tag{C9}$$

APPENDIX D

Using notations common to electromagnetism, the electric field $\mathbf{E}(\boldsymbol{\rho})$ scattered by a given optogeometrical system can be described by a vectorial integral equation using the tensorial Green's function of this system $\mathbb{G}(\boldsymbol{\rho}-\boldsymbol{\rho}')$:

$$\mathbf{E}(\boldsymbol{\rho}) = \mathbf{E}_i(\boldsymbol{\rho}) + \int \mathbb{G}(\boldsymbol{\rho}-\boldsymbol{\rho}')\left(\frac{\varepsilon_d}{\varepsilon_2} - 1\right)\mathbf{E}(\boldsymbol{\rho}')d^3\boldsymbol{\rho}', \tag{D1}$$

where ε_2 is the relative permittivity of the unperturbed medium, ε_d of the scatterer, and $\mathbf{E}_i(\boldsymbol{\rho})$ is the incident unperturbed field. As is well known, a three-dimensional Green's function presents a singularity when $\boldsymbol{\rho}=\boldsymbol{\rho}'$ and it can be decomposed into a singular part and a principal value (P_v):

$$\mathbb{G}(\boldsymbol{\rho}-\boldsymbol{\rho}') = \mathbb{L}\delta(\boldsymbol{\rho}-\boldsymbol{\rho}') + P_v\mathbb{G}(\boldsymbol{\rho}-\boldsymbol{\rho}'), \tag{D2}$$

where δ is the Dirac delta function and the tensor $\mathbb{L}$ depends on the shape of the exclusion domain chosen to define the principal value.[25,26] In the case, for example, of a spherical inclusion, it is simply equal to $-\mathbb{I}_{(3)}/3$, where $\mathbb{I}_{(3)}$ is the three-dimensional unit tensor. For a cylindrical cavity much longer than its radius, the singular part is the same as for a small disk and is equal to

$$\mathbb{L} = -\mathbb{I}_{(2)}/2, \tag{D3}$$

where

$$\mathbb{I}_{(2)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \tag{D4}$$

if the cylinder axis lies along the z axis.

This singular part has the physical meaning that it is impossible to neglect the scattered field in the point of scattering. Moreover, this self-scattering is singular and much stronger than the scattering by the other part of the volume, as can be expected due to the Dirac delta function in Eq. (D2), so that the self-scattering part in Eq. (D1) can be expressed in the form

$$\mathbf{E}(\boldsymbol{\rho}) \approx \mathbf{E}_i(\boldsymbol{\rho}) + \mathbb{L}(\varepsilon_d/\varepsilon_2 - 1)\mathbf{E}(\boldsymbol{\rho}), \tag{D5}$$

whence it follows that the "local" field $\mathbf{E}(\boldsymbol{\rho})$ differs from the unperturbed field $\mathbf{E}_i(\boldsymbol{\rho})$,

$$\mathbf{E}(\boldsymbol{\rho}) \approx \left[\mathbb{I}_{(3)} - \mathbb{L}\left(\frac{\varepsilon_d}{\varepsilon_2} - 1\right)\right]^{-1}\mathbf{E}_i(\boldsymbol{\rho}), \tag{D6}$$

by a factor

$$\left[\mathbb{I}_{(3)} - \mathbb{L}\frac{\varepsilon_d - \varepsilon_2}{\varepsilon_2}\right]^{-1} = \begin{bmatrix} \frac{2\varepsilon_2}{\varepsilon_d + \varepsilon_2} & 0 & 0 \\ 0 & \frac{2\varepsilon_2}{\varepsilon_d + \varepsilon_2} & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{D7}$$

where Eqs. (D3) and (D4) are used. This means that for the x and y components of the electric field, it is necessary to introduce a correction factor equal to the term given in relation (78).

APPENDIX E

From the covariant expression of the electric and magnetic fields of a magnetic dipole,

$$\mathbf{E} = -\frac{Zk^2}{4\pi}\hat{\boldsymbol{\rho}}\times\mathbf{D}_M\left(1-\frac{1}{ik\rho}\right)\frac{\exp(-ik\rho)}{\rho},$$

$$\mathbf{H} = -\frac{1}{4\pi}\left\{k^2(\hat{\boldsymbol{\rho}}\times\mathbf{D}_M)\times\hat{\boldsymbol{\rho}} - [3\hat{\boldsymbol{\rho}}(\hat{\boldsymbol{\rho}}\cdot\mathbf{D}_M)-\mathbf{D}_M] \times\left(\frac{1}{\rho^2}-\frac{ik}{\rho}\right)\right\}\frac{\exp(-ik\rho)}{\rho}, \quad \text{(E1)}$$

one can obtain their components in a coordinate system in which the dipole moment is equal to

$$\mathbf{D}_M = D_M\hat{\mathbf{y}}, \quad \text{(E2)}$$

so that, when looking in the xOy plane, the electric field has only one vector component:

$$E_z = -\frac{ZD_Mk^2}{8\pi}\left(1-\frac{1}{ikr}\right)\frac{\exp(-ikr)}{r}[\exp(i\theta)+\exp(-i\theta)]. \quad \text{(E3)}$$

On the other hand, Eq. (5) gives E_z in the Fourier–Bessel basis and, comparing the two expressions, it is easy to identify that

$$-\frac{ZD_Mk^4}{8\pi}\left(1-\frac{1}{ikr}\right)\frac{\exp(-ikr)}{r} = i\sum_m [b^H_{n,m}(0)-c^H_{n,m}(0)]J_n(k_mr)k_m\Delta_m \quad \text{(E4)}$$

by taking into account that in nonmagnetic media $k^2 = k_0^2\varepsilon$. Multiplying the two sides by $rJ_n(k_{m'}r)$, integrating over r, and using the orthogonality in Eq. (15), one obtains that

$$\mathfrak{M}^H_m \equiv b^H_{1,m} - c^H_{1,m} = i\frac{ZD_Mk^4}{8\pi}\int_0^\infty\left(1-\frac{1}{ikr}\right)\times\frac{\exp(-ikr)}{r}J_1(k_mr)r\mathrm{d}r = -i\frac{ZD_Mk^4}{8\pi}\frac{k_m}{kk_{m_z}}, \quad \text{(E5)}$$

where the integral is evaluated using Eqs. (11.4.35–11.4.38) of Ref. 29. After simplification, this expression is equivalent to Eq. (91) when applied to medium 2.

It is impossible in such a simple way to obtain the expression for $\mathfrak{M}^E_m$, because H_z vanishes in the xOy plane. Instead, Appendix A gives another possibility to obtain directly the spectral amplitude $b^H_{1,m}$. To this end, we shall use an equation similar to the second of Eqs. (A5), but written for the magnetic field:

$$H_{\theta,1} - iH_{r,1} = \int_0^\infty \hat{b}^H_1(k_r,z)j_{n+1}(k_rr)k_r\mathrm{d}k_r = \sum_{m=0}^\infty b^H_{1,m}(z)J_2(k_mr)\Delta_m. \quad \text{(E6)}$$

The components of the magnetic field in cylindrical coordinates can be obtained from Eq. (E1) by taking into account that in the xOy plane

$$\hat{\boldsymbol{\rho}}\times\mathbf{D}_M = D_M\cos\theta\hat{\mathbf{z}}, \quad (\hat{\boldsymbol{\rho}}\times\mathbf{D}_M)\times\hat{\boldsymbol{\rho}} = D_M\cos\theta\hat{\boldsymbol{\theta}},$$

$$\hat{\boldsymbol{\rho}}\cdot\mathbf{D}_M = D_M\sin\theta, \quad \text{(E7)}$$

so that

$$H_r = \frac{2D_Mk^2}{4\pi}\left(\frac{1}{k^2r^2}+\frac{1}{ikr}\right)\frac{\exp(ikr)}{r}\frac{1}{2i}[\exp(i\theta)-\exp(-i\theta)], \quad \text{(E8)}$$

$$H_\theta = \frac{D_Mk^2}{4\pi}\left(1-\frac{1}{k^2r^2}-\frac{1}{ikr}\right)\frac{\exp(ikr)}{r}\frac{1}{2}[\exp(i\theta)+\exp(-i\theta)]. \quad \text{(E9)}$$

After applying the orthogonality from Eq. (15) to Eq. (E6), we obtain that

$$b^H_{1,m} = \frac{\omega\mu_0}{2}k_m\int_0^\infty (H_{\theta,1}-iH_{r,1})J_2(k_mr)r\mathrm{d}r = \frac{\omega\mu_0}{2}k_m\frac{D_Mk^2}{8\pi}\int_0^\infty\left(1-\frac{3}{k^2r^2}-\frac{3}{ikr}\right)\times\frac{\exp(ikr)}{r}J_2(k_mr)r\mathrm{d}r = \frac{\omega\mu_0}{2}k_m\frac{D_Mk^2}{8\pi}\left(-i\frac{k_m^2}{k^2k_{m_z}}\right), \quad \text{(E10)}$$

where the integration is made again by using Eqs. (11.4.35–11.4.38) of Ref. 29. Equations (E5) and (E10) give directly

$$\mathfrak{P}^H_m = 2b^H_{1,m} - \mathfrak{M}^H_m = i\frac{ZD_M}{8\pi}kk_mk_{m_z}. \quad \text{(E11)}$$

The other two spectral amplitudes are obtained by using relations (27) and (30), so that

$$\mathfrak{M}^E_m = \frac{1}{ik_{m_z}}\mathfrak{P}^H_m = k_mk\frac{ZD_M}{8\pi},$$

$$\mathfrak{P}^E_m = \frac{ik_{m_z}}{k^2}\mathfrak{M}^H_m = -k_mk\frac{ZD_M}{8\pi}. \quad \text{(E12)}$$

ACKNOWLEDGMENTS

This work is supported by the French Ministry of Research (ACI Nanosciences).

Corresponding author E. Popov's e-mail address is e.popov@fresnel.fr.

REFERENCES

1. H. A. Bethe, "Theory of diffraction by small holes," Phys. Rev. **66**, 163–182 (1944).
2. J. D. Jackson, *Classical Electrodynamics*, 3rd ed. (Wiley, 1998).

3. T. W. Ebbesen, H. J. Lezec, H. F. Ghaemi, T. Thio, and P. A. Wolff, "Extraordinary optical transmission through subwavelength hole arrays," Nature (London) **391**, 667–669 (1998).
4. H. F. Ghaemi, T. Thio, D. E. Grupp, T. W. Ebbesen, and H. J. Lezec, "Surface plasmons enhance optical transmission through sub-wavelength holes," Phys. Rev. B **58**, 6779–6782 (1998).
5. E. Popov, M. Nevière, S. Enoch, and R. Reinisch, "Theory of light transmission through subwavelength periodic hole arrays," Phys. Rev. B **62**, 16100–16108 (2000).
6. S. Enoch, E. Popov, M. Nevière, and R. Reinisch, "Enhanced light transmission by hole arrays," J. Opt. A, Pure Appl. Opt. **4**, S83–S87 (2002).
7. L. Martin-Moreno, F. J. Garcia-Vidal, H. J. Lezec, K. M. Pellerin, T. Thio, J. B. Pendry, and T. W. Ebbesen, "Theory of extraordinary optical transmission through subwavelength hole arrays," Phys. Rev. Lett. **86**, 1114–1117 (2001).
8. L. Salomon, F. Grillot, A. Zayats, and F. de Fornel, "Near-field distribution of optical transmission of periodic subwavelength holes in a metal film," Phys. Rev. Lett. **86**, 1110–1113 (2001).
9. A. Krishman, T. Thio, T. J. Kim, H. J. Lezec, T. W. Ebbesen, P. A. Wolf, J. A. Pendry, L. Martin-Moreno, and F., J. Garcia-Vidal, "Evanescently-coupled surface resonance in surface plasmon enhanced transmission," Opt. Commun. **200**, 1–7 (2001).
10. M. J. Levene, J. Kotach, S. Turner, M. Foquet, H. G. Craighead, and W. W. Webb, "Zero-mode waveguide for single-molecule analysis at high concentrations," Science **209**, 682–686 (2003).
11. A. Moreau, G. Granet, F. I. Baida, and D. Van Labeke, "Light transmission by subwavelength square coaxial aperture arrays in metallic films," Opt. Express **11**, 1131–1136 (2003).
12. F. J. Garcia-Vidal, L. Martin-Moreno, H. J. Lezec, and T. W. Ebbesen, "Focusing light with a single subwavelength aperture flanked by surface corrugations," Appl. Phys. Lett. **83**, 4500–4502 (2003).
13. L. Martin-Moreno, F. J. Garcia-Vidal, H. J. Lezec, A. Degiron, and T. W. Ebbesen, "Theory of highly directional emission from a single subwavelength aperture surrounded by surface corrugations," Phys. Rev. Lett. **90**, 167401 (2003).
14. M. J. Lockyear, A. P. Hibbins, and J. R. Sambles, "Surface-topography-induced enhanced transmission and directivity of microwave radiation through a subwavelength circular metal aperture," Appl. Phys. Lett. **84**, 2040–2042 (2004).
15. A. Degiron, H. J. Lezec, N. Yamamoto, and T. W. Ebbesen, "Optical transmission properties of a single subwavelength aperture in a real metal," Opt. Commun. **239**, 61–66 (2004).
16. L. Yin, V. K. Vlasko-Vlasov, A. Rydh, J. Pearson, U. Welp, S. H. Chang, S. K. Gray, G. C. Schatz, D. B. Brown, and C. W. Kimball, "Surface plasmons at single nanoholes in Au films," Appl. Phys. Lett. **85**, 467–469 (2004).
17. R. Zakharian, M. Mansuripur, and J. V. Moloney "Transmission of light through small elliptical apertures," Opt. Express **12**, 2631–2648 (2004).
18. R. Wanneracher, "Plasmon-supported transmission of light through nanometric holes in metallic thin films," Opt. Commun. **195**, 107–118 (2001).
19. H. J. Lezec, A. Degiron, E. Devaux, R. A. Linke, L. Martin-Moreno, F. J. Garcia-Vidal, and T. W. Ebbesen, "Beaming light from a subwavelength aperture," Science **297**, 820–822 (2002).
20. E. Popov, M. Nevière, A.-L. Fehrembach, and N. Bonod, "Enhanced transmission by structured circular apertures," Appl. Opt. **44**, 6898–6904 (2005).
21. E. Popov, M. Nevière, J. Wenger, H. Rigneault, P.-F. Lenne, P. Chaumet, and N. Bonod, "Field enhancement in single subwavelength apertures," J. Opt. Soc. Am. A **23**, 2342–2348 (2006).
22. N. Bonod, E. Popov, and M. Nevière, "Differential theory of diffraction by finite cylindrical objects," J. Opt. Soc. Am. A **22**, 481–490 (2005).
23. C. J. Bouwkamp, "Diffraction theory," Rep. Prog. Phys. **17**, 35–100 (1954).
24. A. Roberts, "Electromagnetic theory of diffraction by a circular aperture in a thick, perfectly conducting screen," J. Opt. Soc. Am. A **4**, 1970–1983 (1987).
25. Ch.-A. Guérin, P. Mallet, and A. Sentenac, "Effective-medium theory for finite-size aggregates," J. Opt. Soc. Am. A **23**, 349–358 (2006).
26. A Yaghjian, "Electric dyadic Green's function in the source region," Proc. IEEE **68**, 248–263 (1980).
27. E. Popov, M. Nevière, A.-L. Fehrembach, and N. Bonod, "Optimization of plasmon excitation at structured apertures," Appl. Opt. **44**, 6141–6154 (2005).
28. A. Snyder and J. Love, *Optical Waveguide Theory* (Chapman & Hall, London, 1983).
29. M. Abramovitz and I. A. Stegun, eds. *Handbook of Mathematical Functions* [9th printing (1970)] National Bureau of Standards Applied Mathematics Series 55 (US GPD, 1964).