

Minimal model for optical transmission through holey metal films

L Martín-Moreno¹ and F J García-Vidal²

¹ Departamento de Física de la Materia Condensada-ICMA, Universidad de Zaragoza-CSIC, E-50009 Zaragoza, Spain

² Departamento de Física Teórica de la Materia Condensada, Universidad Autónoma de Madrid, E-28049 Madrid, Spain

E-mail: lm@unizar.es

Received 1 February 2008, in final form 15 May 2008

Published 8 July 2008

Online at stacks.iop.org/JPhysCM/20/304214

Abstract

This paper presents a tutorial on the computation of both extraordinary optical transmission and surface electromagnetic modes in holey metal films. Our model consists of a square array of square holes in a perfect conductor. It is shown that considering just the fundamental waveguide mode inside the holes captures the main features of the optical transmission, which allows us to obtain quasi-analytical results. Extraordinary optical transmission is unambiguously linked to the presence of surface electromagnetic modes in the corrugated structure. The particular case of surface electromagnetic modes in a perfect conductor is analyzed, paying attention to different strategies for increasing their confinement to the surface. The use of the energy loss of a charged particle passing close to the surface as a spectroscopic tool for these surface modes is also discussed.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

Ten years ago, Ebbesen *et al* [1] found experimentally that the optical transmission through an array of subwavelength holes drilled in a metal film was, for some resonant wavelengths, orders of magnitude larger than what was expected for a collection of isolated holes. This unexpected result, dubbed extraordinary optical transmission (EOT), has triggered a wealth of experimental and theoretical studies which, in turn, have revealed new phenomena such as: EOT and beaming of light in single apertures flanked by surface corrugations [2–4], the strong influence of the hole shape on transmission properties in both hole arrays [5] and isolated holes [5–8], and interesting nonlinear transmission effects [9–11].

Already the first report on EOT [1] pointed out the close spectral correspondence between transmission peaks and the wavelengths at which surface plasmon polaritons [12] should be excited. Further theoretical work backed up this observation, and explained EOT as being due to the resonant coupling of electromagnetic (EM) surface modes through the evanescent fields inside the hole [13]. Moreover, the mathematical analysis showed that EOT-like behavior should occur whenever two localized modes are, at the same

time, coupled between themselves and weakly coupled to a continuum [13, 14]. This allowed for a Fano type description of the process [15] and pointed to the possibility of extraordinary transmission in other ranges of the EM spectrum, and even for other types of waves. Electromagnetic wave transmission mediated by surface modes has also been found in: perforated metals in both the THz [16, 17] and millimeter regimes [18], doped semiconductors [19], perforated polar semiconductors in the infrared regime [20] and holey slabs in photonic crystals [21, 22]. Additionally, extraordinary transmission of both matter waves (cold atoms) [23] and sound [24, 25] through holey slabs has also been demonstrated.

Interestingly, the calculations showed EOT peaks even when the metal was considered impenetrable to the fields [13]. This idealization, which in the context of metals is usually referred to as the perfect electrical conductor (PEC) approximation, occurs whenever the dielectric constant of the solid, ϵ_M , fulfills $|\epsilon_M| = \infty$.³ The appearance of EOT in perfect conductors was intriguing because, as flat PEC-dielectric interfaces do not support electromagnetic surface

³ Strictly speaking, both a lossy Drude metal in the low frequency limit ($\epsilon = +i\infty$) and a lossless dielectric with $\epsilon = +\infty$ would have the same properties as a PEC.

modes, it looked like a different interpretation from the one involving surface modes was needed in this case. However, *periodically corrugated* PEC surfaces do support surface waves which, in a perforated film, assist the transmission process [26]. The link between extraordinary transmission in real metals and in perfect conductors was more evident after finding that the structured PEC surface could be approximately represented by effective permeabilities which spoof those of a Drude-like metal [26, 27].

Despite the apparent geometrical simplicity of a periodic array of apertures in a metal film, the system is still characterized by a large number of geometrical and material parameters. This complexity leads to the existence of different transmission regimes and the coexistence of different mechanisms. For instance, while surface modes strongly influence the EM transmission in both the microwave and optical regimes, in the latter case also the penetration of EM fields in the ‘vertical’ walls defining the hole has relevant effects. This penetration affects the propagation constants of waveguide modes which, in turn, affect the transmission [28]. However, the existence of EOT in perfect conductors shows that the presence of waveguide modes with ‘modified’ propagation constants is not essential for EOT. Another quantity showing a strong dependence on the different parameters defining the system is the lifetime of surface modes. Surface modes involved in EOT are leaky, i.e., they are coupled to radiative modes. Otherwise, they could not be seen by the incoming field. Therefore, the lifetime of a surface mode depends on both radiation coupling (strongly geometry dependent) and absorption (strongly material dependent). Different transmission regimes occur depending on the relation of this lifetime to both the time needed to build coupled surface modes and the time it takes for the surface mode to cross a finite array.

Many theoretical works have addressed the dependence of EOT phenomena on different parameters [13, 28–34]. However, it is still desirable to have simplified models that neatly provide the main physics. Here we present what, in our opinion, is the simplest possible model able to capture the main aspects of EOT. Our aim is to provide a tutorial presentation, both on the basics of EOT phenomena and on how the different scattering coefficients can be obtained within the modal expansion technique, hoping the exposition will be useful for those interested in the more general results presented in a succinct way elsewhere [13, 35]. We present different formulations of the scattering problem, which allow us to look at the solution from different points of view, therefore improving our physical insight into the relevant mechanisms leading to EOT. Additionally, we give expressions for different scattering coefficients of hole arrays. These expressions have semi-quantitative value in the optical regime for good metals like Au or Ag, but are already good quantitative approximations for wavelengths in the infrared and longer. Moreover, they should have an even larger range of validity for hole arrays in Al [36], which presents a very small skin depth even in the visible.

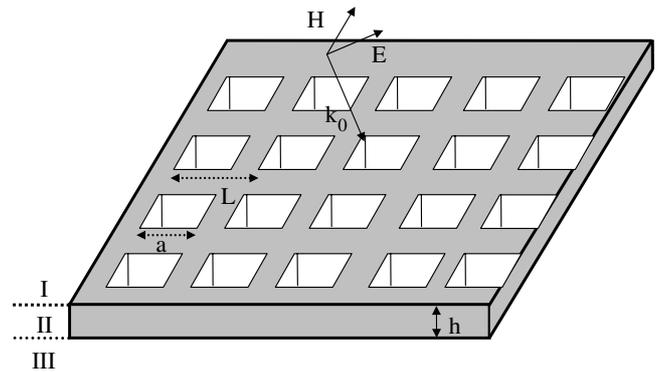


Figure 1. Schematic picture of the hole array considered in this work: a square array of square holes in a metal film of thickness h . The lattice parameter is L and the side of the hole is a . A polarized plane wave is incoming from media I, into the hole array (media II) and is transmitted into media III. The horizontal metal film interfaces are placed at $z = \pm h/2$.

2. Theoretical model

The system under study is an infinite array of holes, placed periodically in a metal film of thickness h . We denote by z the direction normal to the film, and take the metal–dielectric interfaces to be placed at $z = \pm h/2$ (see figure 1). The array is illuminated by a plane wave with given polarization σ_0 and wavevector \mathbf{k}_0 . The structure can be divided into four regions: the superstrate (from where radiation is coming), the substrate (where radiation will be transmitted to), the interior of the hole, and the metal. Each region is characterized by a dielectric constant.

2.1. Approximations and geometry considered in this work

Our model involves two main approximations.

- We treat the metal as a PEC, i.e. $\epsilon_M = -\infty$. The main advantage of neglecting the penetration of the field into the metal is that the fields inside the hole array can then be expressed in terms of the waveguide modes in the hole, which are analytically known for some geometries.
- Only the fundamental waveguide modes inside the hole are considered in the expansion of the wavefield (notice that the fundamental mode is degenerate for some hole shapes). The formalism considering the full expansion on waveguide modes has been presented elsewhere [13, 35]. The single mode approximation leads to quasi-analytical expressions and is quite accurate for subwavelength holes⁴, as the comparison with results obtained with the full expansion will show.

Additionally, we analyze a system with the following restrictions.

- (i) The holes are placed in a square lattice with lattice parameter L .

⁴ In this work we will use the term ‘subwavelength holes’ to denote the wavelength regime in which all modes inside the hole are evanescent.

- (ii) The holes have square cross section, of side a (in this case the fundamental waveguide mode is double degenerate).
- (iii) The incident radiation is a p-polarized plane wave, i.e., for normal incidence the in-plane component of the incident electric field points along the x -direction.

These three conditions allow us a further simplification of the problem: due to symmetry we only have to consider the fundamental waveguide mode that couples directly to the incident field.

In this paper we also consider that:

- (iv) while the dielectric constant inside the holes, ϵ_{hole} , is arbitrary, the dielectric constant of both substrate and superstrate is equal to unity.

We consider this simplest case in order to focus on presenting the main physics and different formalisms within the mode matching technique. The treatment of the general case would require a more cumbersome notation, thus concealing the simplicity of the method. In any case, the scattering coefficients in more general situations, in which restrictions (i)–(iv) are relaxed, can be straightforwardly obtained following the derivations that will be presented in this paper.

2.2. Basis set for modal expansion and notation

In order to compute the transmittance through the structure, we expand the electromagnetic field in terms of the EM eigenmodes in each region, and match the EM field appropriately at the boundaries.

The choice of a good representation for the electromagnetic (EM) field greatly simplifies the calculation. In principle, three components of the EM field are enough for a complete representation: although the EM field comprises six components, the full EM field can be recovered from either the electric or the magnetic field and using Maxwell equations. An even more compact representation is possible when the field is expanded in eigenmodes of the Maxwell equation, as each of these has a well defined propagation constant k_z in each of the pieces of space with translational symmetry along direction z . Then, the equation $\nabla \mathbf{D} = 0$ can be used to relate one of the electric field components to the other two, *except for a sign which can only be known if the direction of propagation of the EM eigenmode is known*. In the representation we choose, we consider as basis for our field representation the x and y components of the electric field, which thus form a bi-vector, and keep track of whether modes propagate in the $+z$ or $-z$ direction.

As well as a compact representation, a convenient notation can greatly simplify the algebra. This can be achieved by using Dirac's notation: we denote by $|\mathbf{E}\rangle$ a bi-vector such that $\langle \mathbf{r}_{\parallel} | \mathbf{E} \rangle = (E_x(\mathbf{r}_{\parallel}), E_y(\mathbf{r}_{\parallel}))^T$, where T denotes transposition. Notice that the z dependence for each mode is not included and must be explicitly stated. With this notation, overlaps between bi-vectors are written in a very economical way as:

$$\langle \mathbf{E} | \mathbf{E}' \rangle = \int_{\text{unit cell}} d\mathbf{r}_{\parallel} \{ E_x^*(\mathbf{r}_{\parallel}) E'_x(\mathbf{r}_{\parallel}) + E_y^*(\mathbf{r}_{\parallel}) E'_y(\mathbf{r}_{\parallel}) \}. \quad (1)$$

An important property, valid for eigenmodes in both free space and waveguides limited by PECs, is that the magnetic field can be related to the electric field bi-vector as: $|\mathbf{H}_{\text{mode}}\rangle = \pm Y |\mathbf{E}_{\text{mode}}\rangle$. In the previous expression Y is the *modal admittance*, which is a proportionality coefficient characteristic of the mode considered, and \mathbf{u}_z is the unitary vector along the $+z$ direction. The \pm sign must be chosen according to whether the wave propagates in the $+$ or $-z$ direction, respectively.

We define $g = \omega/c$, where ω is the frequency of the incoming field and c is the speed of light, and use the cgs system of units (so in a uniform media with dielectric constant ϵ , $\nabla \times \mathbf{E} = \iota g \mathbf{H}$ and $\nabla \times \mathbf{H} = -\iota \epsilon g \mathbf{E}$). The eigenmodes considered in the different regions of space are:

- *Free space*

Plane waves, denoted by $|\mathbf{k}, \sigma\rangle$ according to their wavevector \mathbf{k} and polarization σ (s or p). Their explicit functional forms are:

$$\begin{aligned} \langle \mathbf{r}_{\parallel} | \mathbf{k}, p \rangle &= (k_x, k_y)^T \exp(\iota \mathbf{k}_{\parallel} \mathbf{r}_{\parallel}) / \sqrt{L_x L_y |\mathbf{k}_{\parallel}|^2} \\ \langle \mathbf{r}_{\parallel} | \mathbf{k}, s \rangle &= (-k_y, k_x)^T \exp(\iota \mathbf{k}_{\parallel} \mathbf{r}_{\parallel}) / \sqrt{L_x L_y |\mathbf{k}_{\parallel}|^2}. \end{aligned}$$

At $\mathbf{k}_{\parallel} = 0$ the distinction between s and p modes is arbitrary. Here, we take $|\mathbf{k}_{\parallel} = 0, p\rangle = (1, 0)^T / \sqrt{L_x L_y}$ and $|\mathbf{k}_{\parallel} = 0, s\rangle = (0, 1)^T / \sqrt{L_x L_y}$. These modes are orthonormal when integrated over a unit cell, i.e. $\langle \mathbf{k}, \sigma | \mathbf{k}', \sigma' \rangle = \delta_{\mathbf{k}, \mathbf{k}'} \delta_{\sigma, \sigma'}$, where δ is the Kronecker delta. The corresponding modal admittances are $Y_{\mathbf{k}, p} = g/k_z$ and $Y_{\mathbf{k}, s} = k_z/g$, where $k_z = \sqrt{g^2 - |\mathbf{k}_{\parallel}|^2}$. In order to simplify the notation, we shall not explicitly express that k_z is a function of \mathbf{k}_{\parallel} .

- *Inside the hole at lattice position \mathbf{R}*

The fundamental waveguide mode, denoted by $|0, \mathbf{R}\rangle$. For a square hole, $\langle \mathbf{r}_{\parallel} | 0, \mathbf{R} \rangle = (1, 0)^T \times \sin[q_y(y - R_y + a/2)] / \sqrt{a/2}$, for values of \mathbf{r}_{\parallel} inside the hole (defined on the support $[R_x - a/2, R_x + a/2] \times [R_y - a/2, R_y + a/2]$), and zero otherwise. Here $q_y = \pi/a$. The admittance for this mode is $Y^{\text{H}} = k_z^{\text{H}}/g$, where $k_z^{\text{H}} = \sqrt{\epsilon_{\text{hole}} g^2 - q_y^2}$.

An important point is that, in order to satisfy scattering boundary conditions at $\pm\infty$, the sign of the square root must be chosen so that the imaginary parts of all propagation constants are positive, i.e., $\text{Im}(k_z), \text{Im}(k_z^{\text{H}}) > 0$.

With this choice of basis set, and taking into account Bloch's theorem, the EM fields in different regions of space can be expressed as:

- *Region of reflection, ($z < -h/2$)*

$$|\mathbf{E}_I(z)\rangle = |\mathbf{k}_0 \sigma_0\rangle e^{ik_0 z (z+h/2)} + \sum_{\mathbf{k}\sigma} r_{\mathbf{k}\sigma} |\mathbf{k}\sigma\rangle e^{-ik_z(z+h/2)} \quad (2)$$

$$\begin{aligned} |-\mathbf{u}_z \times \mathbf{H}_I(z)\rangle &= Y_{\mathbf{k}_0, \sigma_0} |\mathbf{k}_0 \sigma_0\rangle e^{ik_0 z (z+h/2)} \\ &- \sum_{\mathbf{k}\sigma} r_{\mathbf{k}\sigma} Y_{\mathbf{k}, \sigma} |\mathbf{k}\sigma\rangle e^{-ik_z(z+h/2)} \end{aligned} \quad (3)$$

where, according to Bloch theorem, sums are over all \mathbf{k} values of the form $\mathbf{k} = \mathbf{k}_0 + \mathbf{K}_R$, \mathbf{K}_R being a vector of the reciprocal lattice, i.e., $\mathbf{K}_R = (2\pi/L)(m\mathbf{u}_x + n\mathbf{u}_y)$ with m and n integers, and \mathbf{u}_x and \mathbf{u}_y being the unitary vectors in the x - and y -direction, respectively.

- Region inside the holes, ($-h/2 < z < h/2$)

$$|\mathbf{E}_{\text{II}}(z)\rangle = \sum_{\mathbf{R}} e^{i\mathbf{k}_0\mathbf{R}} |0, \mathbf{R}\rangle \left\{ A e^{ik_z^{\text{II}}z} + B e^{-ik_z^{\text{II}}z} \right\} \quad (4)$$

$$|-\mathbf{u}_z \times \mathbf{H}_{\text{II}}(z)\rangle = \sum_{\mathbf{R}} e^{i\mathbf{k}_0\mathbf{R}} |0, \mathbf{R}\rangle Y^{\text{II}} \times \left\{ A e^{ik_z^{\text{II}}z} - B e^{-ik_z^{\text{II}}z} \right\}. \quad (5)$$

- Transmission region ($z > h/2$)

$$|\mathbf{E}_{\text{III}}(z)\rangle = \sum_{\mathbf{k}\sigma} t_{\mathbf{k}\sigma} |\mathbf{k}\sigma\rangle e^{ik_z(z-h/2)} \quad (6)$$

$$|-\mathbf{u}_z \times \mathbf{H}_{\text{III}}(z)\rangle = \sum_{\mathbf{k}\sigma} t_{\mathbf{k}\sigma} Y_{\mathbf{k},\sigma} |\mathbf{k}\sigma\rangle e^{ik_z(z-h/2)}.$$

In the previous expressions, $r_{\mathbf{k}\sigma}$, A , B , and $t_{\mathbf{k}\sigma}$ are expansions coefficients to be determined through matching the fields at the different interfaces.

The energy current crossing a unit cell at a given z , $J(z)$, can be computed by integration of the Poynting vector: $J(z) = \frac{1}{2} \text{Re}[\int d\mathbf{r}_{\parallel} \mathbf{u}_z \cdot (\mathbf{E}(\mathbf{r}_{\parallel}, z) \times \mathbf{H}^*(\mathbf{r}_{\parallel}, z))] = \frac{1}{2} \text{Re}[\langle -\mathbf{u}_z \times \mathbf{H}(z) | \mathbf{E}(z) \rangle]$.

Using this expression, it is straightforward to calculate the energy current of the incident field $J_0 = Y_{\mathbf{k}_0, \sigma_0}/2$. The transmittance can be computed in each region as $T_N \equiv J_N/J_0$, where J_N is the current evaluated at values of z in the N th region. We obtain:

$$T_{\text{I}} = 1 - \sum_{\mathbf{k}\sigma} \text{Im}(\iota Y_{\mathbf{k},\sigma}) |r_{\mathbf{k}\sigma}|^2 / Y_{\mathbf{k}_0, \sigma_0} \quad (7)$$

$$T_{\text{II}} = \begin{cases} Y^{\text{II}} (|A|^2 - |B|^2) / Y_{\mathbf{k}_0, \sigma_0}, & \text{if } k_z^{\text{II}} \text{ is real} \\ (2\iota Y^{\text{II}}) \text{Im}(A B^*) / Y_{\mathbf{k}_0, \sigma_0}, & \text{if } k_z^{\text{II}} \text{ is imaginary} \end{cases} \quad (8)$$

$$T_{\text{III}} = \sum_{\mathbf{k}\sigma} \text{Im}(\iota Y_{\mathbf{k},\sigma}) |t_{\mathbf{k}\sigma}|^2 / Y_{\mathbf{k}_0, \sigma_0}. \quad (9)$$

As our model does not include absorption, current conservation implies that all expressions for T should give the same result, providing a useful test for the calculations.

An interesting consequence of equation (8), and the fact that $0 \leq T \leq 1$, is that the condition $|A| \geq |B|$ must be fulfilled for propagating waveguide modes. However, for evanescent modes, current conservation only forces $\text{Im}(A B^*) \leq 0$. Notice also that a single evanescent mode does not carry current: a non-zero current requires both the term decaying as $A \exp(-|k_z^{\text{II}}|z)$ and the ‘reflected’ wave represented by $B \exp(+|k_z^{\text{II}}|z)$.

2.3. Expressions for the field amplitudes and scattering coefficients

The EM field must satisfy the following matching conditions.

- E -field components parallel to the surface must be continuous over the whole surface.
- H -field components parallel to the interfaces must be continuous *only* over the aperture. The H -field is discontinuous at the metal–air interfaces and the value of this discontinuity is not known a priori. Once the H -field has been computed, its discontinuity can be used to compute the electron current running at the metal surface.

The expansion coefficients can be obtained by projection of each matching equation onto the set of modes that span the spatial region over which the equation is defined. This projection procedure provides as many equations as unknown scattering coefficients, even if the considered basis set is not complete (as is the case here, when all diffraction modes enter into the expansion but only the fundamental waveguide mode is taken into account). Therefore, relations arising from continuity of E -field components should be expanded onto $\langle \mathbf{k}, \sigma |$ modes, and the ones related to the H -field should be expanded onto $\langle 0, \mathbf{R} |$.

Matching can be enforced just over one unit cell. If this is satisfied, Bloch’s theorem automatically ensures proper continuity over the whole lattice. We match the fields over the Wigner–Seitz unit cell corresponding to $\mathbf{R} = 0$, and use the simplified notation $|0\rangle \equiv |0, \mathbf{R} = 0\rangle$ and $e_h \equiv \exp(\iota k_z^{\text{II}} h/2)$. This leads to:

$$r_{\mathbf{k}\sigma} = -\delta_{\mathbf{k}, \mathbf{k}_0} \delta_{\sigma, \sigma_0} + (A e_h^{-1} + B e_h) \langle \mathbf{k}, \sigma | 0 \rangle \quad (10)$$

$$-\sum_{\mathbf{k}\sigma} Y_{\mathbf{k},\sigma} r_{\mathbf{k}\sigma} \langle 0 | \mathbf{k}, \sigma \rangle = -Y_{\mathbf{k}_0, \sigma_0} \langle 0 | \mathbf{k}_0, \sigma_0 \rangle + Y^{\text{II}} \times (A e_h^{-1} - B e_h) \quad (11)$$

$$t_{\mathbf{k}\sigma} = (A e_h + B e_h^{-1}) \langle \mathbf{k}, \sigma | 0 \rangle \quad (12)$$

$$\sum_{\mathbf{k}\sigma} Y_{\mathbf{k},\sigma} t_{\mathbf{k}\sigma} \langle 0 | \mathbf{k}, \sigma \rangle = Y^{\text{II}} (A e_h - B e_h^{-1}) \quad (13)$$

where the two first equations of the previous set come from the continuity of fields at $z = -h/2$, and the last two from the matching at $z = h/2$.

The expressions for the overlaps are:

$$\langle \mathbf{k}, \sigma | 0 \rangle = f_{\mathbf{k},\sigma} \sqrt{\frac{a^2}{2L^2}} \text{sinc}[k_x a/2] \left(\text{sinc} \left[\frac{(k_y + q_y)a}{2} \right] + \text{sinc} \left[\frac{(k_y - q_y)a}{2} \right] \right) \quad (14)$$

where $f_{\mathbf{k},p} = k_x/|\mathbf{k}_{\parallel}|$, $f_{\mathbf{k},s} = -k_y/|\mathbf{k}_{\parallel}|$, and $\text{sinc}[x] \equiv \sin[x]/x$.

2.4. Formalism in terms of amplitudes of waveguide modes

The previous set of equations can be solved by eliminating $r_{\mathbf{k}\sigma}$ and $t_{\mathbf{k}\sigma}$ in favor of A and B . After multiplying all previous equations by a factor ι (the convenience of this unusual convention will be apparent in section 2.6), we obtain:

$$\Gamma A + \Delta B = I_0 \quad (15)$$

$$\Delta A + \Gamma B = 0 \quad (16)$$

with

$$I_0 = 2\iota Y_{\mathbf{k}_0, \sigma_0} \langle 0 | \mathbf{k}_0, \sigma_0 \rangle \quad (17)$$

$$G = \iota \sum_{\mathbf{k}\sigma} Y_{\mathbf{k},\sigma} |\langle 0 | \mathbf{k}, \sigma \rangle|^2 \quad (18)$$

$$\Gamma = e_h^{-1} (G + \iota Y^{\text{II}}) \quad (19)$$

$$\Delta = e_h (G - \iota Y^{\text{II}}). \quad (20)$$

Clearly, I_0 is the term that characterizes the illumination of the apertures. The quantity G depends on both the modal

impedances in vacuum and the overlaps between plane waves and the waveguide modes. It can be interpreted as the effective impedance of vacuum, as seen from the fields in the hole. Recall that, as the summation is over wavevectors of the form $\mathbf{k}_0 + \mathbf{K}_R$, $G = G(\mathbf{k}_0)$ depends on the wavevector of the incident field \mathbf{k}_0 . Notice also an important property of G : G diverges whenever one of the diffraction modes in the sum defining G becomes grazing. This is due to the divergence of the modal impedance of p-polarized modes, $Y_{\mathbf{k},p}$, when the wavevector \mathbf{k} is parallel to the surface (i.e. $k_z = 0$). As we will show, the transmission maxima are governed by resonant denominators involving G . The rapid variation of G close to divergencies favors the existence of minima in the denominators, thus linking the different transmission maxima to the deep transmission minima (known in this context as Wood anomalies [1]).

In general G is a complex number. With the convention used, the real part of G arises from the contribution of evanescent modes (for which k_z , and therefore the modal impedance, is a pure complex number), while its imaginary part is due to the radiative modes (characterized by a real k_z). For evanescent modes both e_h and ιY^{II} are real quantities. Therefore, Γ and Δ are complex quantities which, like G , have non-zero imaginary parts if the Bloch combination of waveguide modes couples to radiative modes.

The physical interpretation of the quantities Γ and Δ will be apparent after the solution of the system of equations for A and B . We obtain, from equations (15) and (16):

$$A = \frac{\Gamma}{\Gamma^2 - \Delta^2} I_0; \quad B = \frac{-\Delta}{\Gamma^2 - \Delta^2} I_0. \quad (21)$$

The rest of scattering coefficients, as well as the transmission through the system can be obtained from A and B by using expressions previously derived in this paper.

These equations reveal the possibility of resonant behavior: even for weak illuminations, large fields inside the holes build up when $\Gamma^2 - \Delta^2 \approx 0$. Actually, the condition $\Gamma^2 - \Delta^2 = 0$ marks the existence of surface modes. For $|\mathbf{k}_{0\parallel}| < g$ the coupling to radiation prevents the previous condition from being fulfilled for *real values* of $\mathbf{k}_{0\parallel}$; it still can be fulfilled for complex $\mathbf{k}_{0\parallel}$ signaling the presence of leaky surface modes. For $|\mathbf{k}_{0\parallel}| > g$, and if the conditions are such that $|\mathbf{k}_{0\parallel} + \mathbf{K}_R| > g$ for all reciprocal lattice vectors \mathbf{K}_R , the condition for the existence of surface modes can be fulfilled for real values of $\mathbf{k}_{0\parallel}$, meaning that the surface modes in the system decay in the direction perpendicular to the slab.

In order to gain further insight on the resonant mechanism, let us consider the case of a very thick metal film. For a subwavelength hole k_z is a complex imaginary number, $e_h = \exp(-|k_z|h/2)$ decays exponentially with h , and so does Δ . So, in this limit, Δ is very small and the resonant condition $\Gamma = 0$ marks the existence of two (virtually uncoupled) surface modes, one at each metal-vacuum interface. These two surface modes are coupled for $\Delta \neq 0$ leading to two split resonances, occurring when the condition $\Gamma = \pm\Delta$ is fulfilled. At these resonances $A = \mp B$, so they correspond to the antisymmetric and symmetric combination of surface modes.

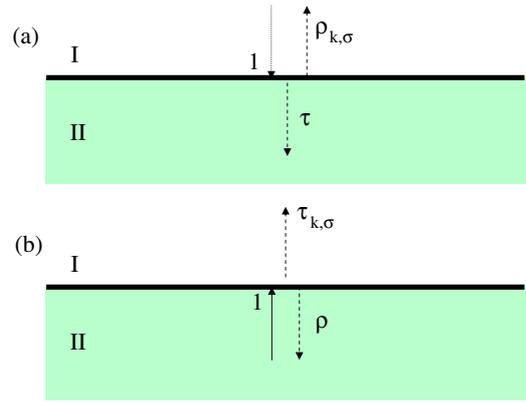


Figure 2. Schematic definition of the different scattering coefficients for EM fields at isolated interfaces dividing two semi-infinite media. (a) Incidence from medium I into medium II, (b) incidence from a waveguide mode in medium II into either medium I or III.

Notice that the expressions above can also be used to study the related system of an array of indentations, i.e. holes closed at the exit side. In that case, continuity of the field at the exit implies $\mathbf{E}_{\parallel} = 0$, leading to $B = -e_h^2 A$. Of course, no transmission through the system is possible in this case, but surface modes still appear at the input interface whenever $\Gamma - e_h^2 \Delta = 0$.

2.5. Multiple scattering formalism

Although the scattering problem was solved in the previous section, additional physical insight is obtained by solving the transmission problem within the multiple scattering formalism. This technique expresses the scattering coefficients of a stratified media in terms of scattering coefficients of isolated interfaces. This is particularly useful when surface modes are present, as they clearly show up in the two-media scattering coefficients.

In the case of a three-layer system, we must solve three two-media scattering problems (see figure 2):

- (i) A plane wave coming from medium I with wavevector \mathbf{k}_0 , impinges onto medium II *which is considered as a semi-infinite medium*. The incoming field has unit amplitude at the interface, being reflected back into the different modes in medium I with amplitude $\rho_{\mathbf{k},\sigma}$, and transmitted into the Bloch combination of waveguide modes with amplitude τ .
- (ii) The field coming from medium II impinges into medium I *which is considered as a semi-infinite medium*. The incoming field is a Bloch combination of waveguide modes, with crystal momentum \mathbf{k}_0 and unit amplitude at the interface, and is reflected back into the Bloch combination of waveguide modes with amplitude ρ and transmitted into the different plane waves in region I with amplitudes $\tau_{\mathbf{k},\sigma}$.
- (iii) As (ii) but for the II–III interface. Obviously, in the case we are considering, in which media I and III have the same dielectric constant, the scattering coefficients for the II–III interface coincide with those for the II–I interface.

Within the multiple scattering technique, the transmission coefficient of the perforated metal film $t_{\mathbf{k},\sigma}$ can be computed as

$$t_{\mathbf{k},\sigma} = \tau e_h^2 \tau_{\mathbf{k},\sigma} + \tau e_h^2 (\rho e_h^2 \rho e_h^2) \tau_{\mathbf{k},\sigma} + \tau e_h^2 (\rho e_h^2 \rho e_h^2)^2 \tau_{\mathbf{k},\sigma} + \dots$$

$$= \tau e_h^2 \left\{ \sum_{n=0}^{\infty} (\rho^2 e_h^4)^n \right\} \tau_{\mathbf{k},\sigma} = \frac{\tau e_h^2 \tau_{\mathbf{k},\sigma}}{D} \quad (22)$$

where $D = 1 - \rho^2 e_h^4$. The origin of this expression is the following: the first term in the series for $t_{\mathbf{k},\sigma}$ represents the direct process of: transmission into the holes, propagation inside them to their exit end and transmission into medium III. The second term is similar but takes into account that, before propagation into medium III, the EM field can: reflect at the exit end back into the holes (factor ρ), propagate inside the holes (factor e_h^2), reflect back at the II–I interface (factor ρ) and propagate again inside the holes to the exit end (factor e_h^2). This intermediate process repeats itself an infinite number of times; each time adding a partial amplitude for the process of transmission into medium III. Fortunately, this infinite series is geometric and can be summed up. The result is that the ‘direct process’ amplitude (the first term in the sum) is ‘renormalized’ by the presence of the denominator D . Large values of $t_{\mathbf{k},\sigma}$ are possible if $1/D$ is large, due to the constructive interference of the waves transmitted in all steps of the multiple scattering process. Similarly, destructive interference of partial waves may lead to small transmission coefficients for the three-media system.

The rest of three-media scattering coefficients can be obtained using a similar reasoning. Their expressions are:

$$r_{\mathbf{k},\sigma} = \rho_{\mathbf{k},\sigma} + \tau e_h^4 \rho \tau_{\mathbf{k},\sigma} / D \quad (23)$$

$$A = \tau e_h / D \quad (24)$$

$$B = \tau e_h^2 \rho e_h / D = \rho e_h^2 A. \quad (25)$$

Two-media scattering coefficients should, in general, be calculated by imposing the proper continuity conditions for the fields across the different interfaces. Here we take advantage that the two-media coefficients can be obtained from equations (15) and (16).

- Coefficients for the I–II interface can be obtained from equations (15) by setting $B = 0$ (in the semi-infinite media there is no reflected wave), and identifying $\tau = A e_h^{-1}$. Therefore

$$\tau = e_h^{-1} I_0 / \Gamma; \quad \rho_{\mathbf{k},\sigma} = -\delta_{\mathbf{k},\mathbf{k}_0} \delta_{\sigma,\sigma_0} + \langle \mathbf{k}, \sigma | 0 \rangle \tau. \quad (26)$$

- Coefficients for the II–III interface are obtained from equations (16) by setting $A = e_h^{-1}$ (the incoming wave must have unit amplitude at $z = +h/2$), and identifying $\rho = B e_h^{-1}$ (recall that the two-media scattering coefficients are defined right at the interface, while A and B are the modal amplitudes at the center of the hole). With this:

$$\rho = -\frac{\Delta}{\Gamma} e_h^{-2} = -\frac{G - \iota Y^{\text{II}}}{G + \iota Y^{\text{II}}}, \quad (27)$$

$$\tau_{\mathbf{k},\sigma} = \frac{2\iota Y^{\text{II}}}{G + \iota Y^{\text{II}}} \langle \mathbf{k}, \sigma | 0 \rangle.$$

Recall that G , Γ and Δ , and therefore all two-media coefficients, depend on the wavevector of the incident plane wave \mathbf{k}_0 .

All two-media scattering coefficients (τ , $\rho_{\mathbf{k},\sigma}$, ρ and $\tau_{\mathbf{k},\sigma}$) present resonances whenever $\Gamma = 0$ which, as we have seen previously, is the mathematical condition for the existence of surface modes. However, three-media scattering coefficients ($r_{\mathbf{k},\sigma}$, A , B , $t_{\mathbf{k},\sigma}$) do not diverge at $\Gamma = 0$; instead they diverge for $D = 0$, which, substituting the expression for ρ is equivalent to $\Gamma^2 - \Delta^2 = 0$. So, as expected, three-media coefficients have information on the *coupled* surface modes.

2.6. Formalism in terms of amplitudes of the electric field at the openings

An alternative reformulation of the scattering problem is obtained by defining the modal amplitudes of the electric field at the input side (E) and output side (E') through $E \equiv e_h^{-1} A + e_h B$, and $E' \equiv -(e_h A + e_h^{-1} B)$.

These quantities satisfy the following system of equations

$$(G - \Sigma)E - G_v E' = I_0 \quad (28)$$

$$-G_v E + (G - \Sigma)E' = 0 \quad (29)$$

where

$$\Sigma \equiv \iota Y^{\text{II}} \frac{e_h^2 + e_h^{-2}}{e_h^2 - e_h^{-2}}, \quad G_v \equiv 2\iota Y^{\text{II}} \frac{1}{e_h^2 - e_h^{-2}}. \quad (30)$$

The case of an array of indentations can also be described by these equations. In this case, the condition $\mathbf{E}_{\parallel}(z = h/2) = 0$ imposes $E' = 0$.

The scattering coefficients and the transmittance (evaluated in different regions) are given in terms of E and E' as

$$r_{\mathbf{k}\sigma} = -\delta_{\mathbf{k},\mathbf{k}_0} \delta_{\sigma,\sigma_0} + \langle \mathbf{k}, \sigma | 0 \rangle E \quad (31)$$

$$t_{\mathbf{k}\sigma} = -\langle \mathbf{k}, \sigma | 0 \rangle E' \quad (32)$$

$$T_{\text{I}} = [2 \text{Re}(\langle \mathbf{k}_0 \sigma_0 | 0 \rangle E) - \text{Im}(G) |E|^2] / Y_{\mathbf{k}_0, \sigma_0} \quad (33)$$

$$T_{\text{II}} = \text{Im}(G_v^* E^* E') / Y_{\mathbf{k}_0, \sigma_0} \quad (34)$$

$$T_{\text{III}} = \text{Im}(G) |E'|^2 / Y_{\mathbf{k}_0, \sigma_0}. \quad (35)$$

These equations have a direct physical interpretation [35].

- E and E' provide the electric field at the surface due to the presence of holes. More precisely, as the expressions for $r_{\mathbf{k}\sigma}$ and $t_{\mathbf{k}\sigma}$ show, E and E' provide the variation of the EM fields in the holey film with respect to those in the uniform slab. Nevertheless, recall that as equations (28) and (29) follow from equations (11) and (13), these equations express the matching of *magnetic fields* at the openings, in terms of the amplitude of the electric field there.
- The expression of EM fields everywhere in terms of those at the opening indicates an alternative interpretation for G : writing $G = \iota \sum_{\mathbf{k}\sigma} Y_{\mathbf{k},\sigma} \langle 0 | \mathbf{k}, \sigma \rangle \langle \mathbf{k}, \sigma | 0 \rangle$, it is apparent that G involves the coupling of the field at the opening in a waveguide mode into plane waves (weighted by $Y_{\mathbf{k},\sigma}$, which acts as a ‘density of plane waves’), and the coupling

back from these into the waveguide mode. Physically this occurs because, according to Huygens principle, each point in the opening acts as an emitter of secondary wavelets. The radiation by a given waveguide mode is the integral of this point emission over all points in the opening, weighted by the modal wavefield. The term GE thus provides how much of this radiation is recollectd by other points in the aperture, onto the waveguide mode under consideration.

- (iii) Σ is ι times the admittance at the aperture (i.e., the quotient between magnetic and electric field components). The term ΣE therefore gives the magnetic field at the opening arising from the presence of an electric field (due to Faraday's and Ampere's law). This 'aperture admittance' differs from the modal admittance by the factor $(e_h^2 + e_h^{-2})/(e_h^2 - e_h^{-2})$, which arises because radiation that couples to a waveguide mode at one end of the hole is reflected back after finding a change in geometry at the other end. Actually, this process repeats itself an infinite number of times, with the field bouncing back and forth inside the hole. The reflection at the bottom changes both \mathbf{E} and \mathbf{H} , and therefore their quotient. Notice that Σ has the same value for both open and 'closed' holes. Of course, the properties of these two configurations are different due to the presence of G_v , which only couples fields at top and bottom *openings* (recall that $E' = 0$ for closed holes).

The formalism in terms of E and E' is reminiscent of Green's Dyadic method. There, the full EM field when some 'scatterer' has been placed in a reference stratified system is solely represented in terms of the electric field inside the scatterer [37]. For a hole array, the reference system is the uniform metallic film, while the 'scatterers' are the collection of holes. In Green's Dyadic formalism, EM fields must be computed in all the volume inside the hole. By contrast, only the fields *at the surface* are needed in the formalism presented in this section, as fields are represented in terms of the waveguide eigenmodes with known propagation along the hole. No attempt has been made here to re-derive the modal expansion expressions from Green's Dyadic formalism, but the expression for G involves some components of Green's Dyadic. Actually, this relation to Green's Dyadic is the origin of the notation of the quantity G , which in the modal expansion formalism appears as a 'weighted admittance'.

An additional advantage of the formalism presented in this section is that the set of equations governing E and E' have an appealing similarity with the tight-binding formalism, used for describing a collection of atoms in terms of localized orbitals. Within this analogy, the 'localized orbitals' are the (Bloch) linear combinations of EM field amplitudes at both entrance and exit hole surfaces. In quantum mechanics, the tight-binding Hamiltonian preserves current, provided the orbital energies are real quantities. Energy losses are often represented by complex values of the orbital energy, with a positive imaginary part. The introduction of the, up to now mysterious, factor ι in the definition of I_0 and G is related to this analogy. With this factor, and if all terms are placed at the left-hand-sides of equations (28) and (29), real quantities (like

Σ in the case of subwavelength holes) represent the energy that remains at the surface. Quantities with positive imaginary part (like G) correspond to energy that leaves the surface (entirely through radiation, in the lossless case we are considering). Finally, pumping energy into the surface is represented by $-I_0$ which, correspondingly, takes negative imaginary values.

Of course, the existence of EM modes also appears in the formalism involving E and E' . Surface modes bound to an isolated interface appear when $G - \Sigma = 0$, and the modes of two coupled interfaces whenever $(G - \Sigma)^2 - G_v^2 = 0$. These equations may not have a solution for a given frequency as, for subwavelength holes, Σ is a real quantity while, in general, G is a complex one. However in the following we will show that, as a function of frequency and angle of incidence, $|G - \Sigma|$ presents deep minima which, if different from zero, correspond to *leaky* modes. Under more restrictive conditions, zeroes of $|G - \Sigma|$ are indeed possible, signaling the existence of truly bound surface modes which, of course, can not be excited by an incoming propagating plane wave (otherwise they would couple back to radiation and become leaky).

All previous expressions have been derived assuming that the metal is a PEC. In what follows we present results obtained with this approximation. Before this, let us mention that this approximation can be relaxed in several ways, through the redefinition of the quantities appearing in the formalism, but maintaining the structure of the theory. First, the penetration of EM fields inside the vertical walls defining the hole can approximately be taken into account by effectively enlarging the hole [14, 38, 39]. This allows for a more accurate representation of both overlaps between waveguide modes and plane waves and, more importantly, of the propagation constant of waveguide modes (notice that the coupling between surface modes depends exponentially on this quantity). Second, the penetration of the EM fields at the top and bottom metal interfaces can be approximately taken into account by the use of the 'surface impedance boundary conditions' (SIBC). Again in this case, the structure of the theory remains unaltered within this refined formalism; only the expressions for the different functions G , G_v , I_0 ... are slightly modified [13, 41].

3. Numerical results

3.1. Transmittance through a hole array

Electromagnetic properties computed within the PEC approximation (for which ϵ_M is frequency independent) remain unaltered if all lengths defining the problem are scaled by a common factor. In what follows we take the lattice parameter as the unit length.

Figure 3 renders the transmittance spectra for a square lattice of holes in a PEC film of thickness $h = 0.2L$. The holes have square cross section, with side $a = 0.4L$. The cutoff wavelength for the fundamental waveguide mode in the hole is $\lambda_c = 2a$ so, for $\lambda > 0.8L$ all waveguide modes in the hole are evanescent. The black curve in figure 3 is the full multimode result, obtained using the method derived in [13]. The red line in figure 3 renders the transmission spectra computed within the single mode approximation (SMA), by considering

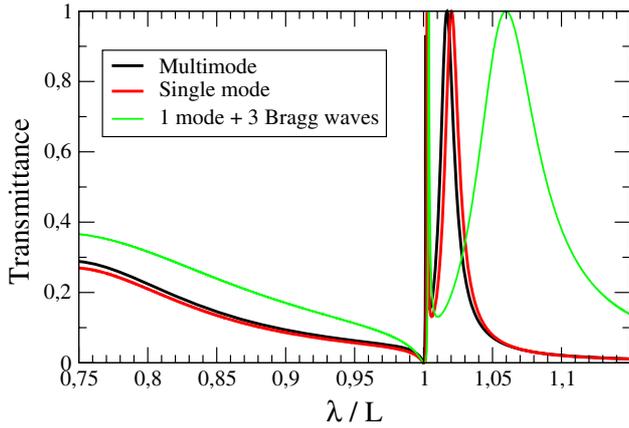


Figure 3. Transmittance through a hole array drilled in a perfect electrical conductor. The side of the hole is $a = 0.4L$ and the film thickness $h = 0.2L$. The black curve is the result from the converged multimode expansion. The red curve is obtained within the single mode approximation discussed in this work. The green curve corresponds to the minimal model described in [13]: single mode where, additionally, only the terms with p-polarization and reciprocal lattice vectors $\mathbf{K}_R = 0, \pm 2\pi/L \mathbf{u}_x$ have been considered in the sum defining G .

as many diffraction modes as needed for convergence in the sum defining G . The green curve was also computed within the single mode approximation, but considering only the terms with p-polarization and reciprocal lattice vectors $\mathbf{K}_R = 0, \pm 2\pi/L \mathbf{u}_x$ in the sum defining G (this is the ‘minimal model’ described in [13]). While clearly the consideration of additional diffraction modes modify the final numerical result, the basics of EOT are already present in this extremely simplified model.

As figure 3 shows, the single mode approximation already provides a very good estimation to the exact transmission curve even when the condition $a/\lambda \ll 1$ is not satisfied (in the considered case $a/\lambda \sim 0.3\text{--}0.5$). The good agreement between the SMA and the exact results stems from the properties of the coupling between the relevant diffraction orders and the waveguide modes: the coupling with the fundamental waveguide mode is stronger, as this mode presents the slower spatial variation. This mechanism takes precedence over the faster decay of higher order waveguide modes inside the hole, thus explaining why the SMA is still a good approximation even for relatively thin films (standard EOT studies in the optical regime are usually done for $h/L \approx 0.3\text{--}0.5$).

Three features are readily visible in figure 3: one very deep minimum at $\lambda = L$ and two transmission peaks. The deep minimum is related to a divergence of G . Recall that G diverges whenever a p-polarized diffraction mode has $k_z = 0$. In this case, the equations previously derived give $A = B = E = E' = 0$, so there is no transmission because the EM field does not enter into the hole. Let us consider first the particular case of normal incidence. Then the diffraction order characterized by the reciprocal lattice vector \mathbf{K}_R has $k_z = 0$ for $|\mathbf{K}_R| = g$. Therefore, for a square lattice, transmission minima occur at $\lambda_{m,n}^{\min} = L/\sqrt{m^2 + n^2}$, where m and n are integers. So, the largest wavelength at which a deep minimum

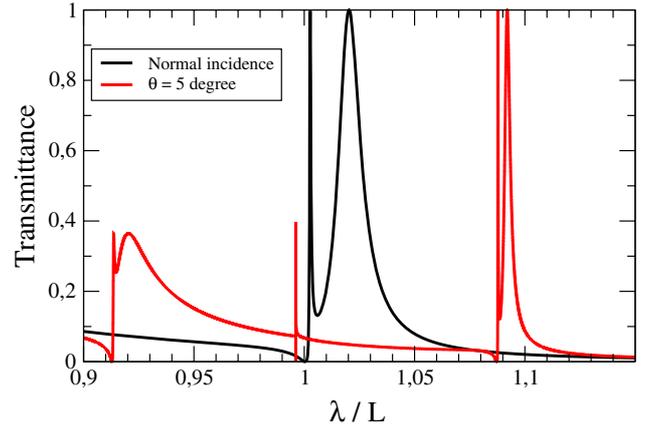


Figure 4. Transmittance for p-polarized light impinging into a hole array, within the single mode and PEC approximations. The side of the hole is $a = 0.4L$ and the film thickness is $h = 0.2L$. The black curve is for normal incidence, while the red curve is for incidence at 5° . In both cases the incoming electric field points along the x -axis.

occurs is $\lambda_{\pm 1,0}^{\min} = \lambda_{0,\pm 1}^{\min} = L$. Actually, this is the result obtained within the PEC approximation. If the penetration of the field into the metal is taken into account by using the SIBC, similar calculations to the ones presented here show that the spectral positions of the transmission minima are more accurately represented by the solutions of the equation $cg = \omega_p(\mathbf{K}_R)$. Here, $\omega_p(\mathbf{k}) = c|\mathbf{k}|\sqrt{(\epsilon_M + 1)/\epsilon_M}$ is the dispersion relation of the surface plasmon polaritons of the *flat* air–metal interface. Therefore, when the dielectric constant of the metal is taken into account, transmission minima are expected at $\lambda_{m,n}^{\min} = L\sqrt{\epsilon_M/(\epsilon_M + 1)}/\sqrt{m^2 + n^2}$ (notice that this is an implicit equation for $\lambda_{m,n}^{\min}$, as ϵ_M is wavelength dependent).

Coming back to the PEC approximation, for the general case of incidence at an angle θ , the condition of grazing diffraction occurs at $|g \sin\theta \mathbf{u}_x + \mathbf{K}_R| = g$. Therefore, at $\theta \neq 0$, the minima associated to $\lambda_{\pm 1,0}^{\min}$ split, appearing now at $\lambda_{\pm 1,0}^{\min} = (1 \pm \sin\theta)L$ and $\lambda_{0,\pm 1}^{\min} = \sqrt{1 - \sin^2\theta}L$. Figure 4 shows such splitting between transmission minima, for a hole array with the same geometrical parameters as in figure 3, but illuminated with a p-polarized plane wave impinging at $\theta = 5^\circ$: the expected minima are clearly seen at wavelengths $\lambda_{-1,0}^{\min}/L = 0.913$, $\lambda_{0,\pm 1}^{\min}/L = 0.996$ and $\lambda_{1,0}^{\min}/L = 1.087$.

Transmission peaks are due to the presence of leaky surface modes of the corrugated metal film. In order to support the previous statement, it is convenient to analyze the properties of the two-media coefficient ρ , that gives the reflection for fields coming from the interior of the hole. Figure 5 renders the real and imaginary parts of ρ , for the cases considered in figure 4. Large values for $\text{Im}(\rho)$ and anomalous behavior for $\text{Re}(\rho)$ occur at spectral positions close to the transmission resonances, both for $\theta = 0^\circ$ and $\theta = 5^\circ$. The spectral dependences of $\text{Re}(\rho)$ and $\text{Im}(\rho)$ are the ones expected, through Kramers–Kronig relations, for causal functions close to localized resonances [40] (a reflected field obviously requires a preexisting incident one, so ρ must satisfy causality). Recall that ρ is a two-media scattering coefficient, so its resonances mark the existence of surface modes bound

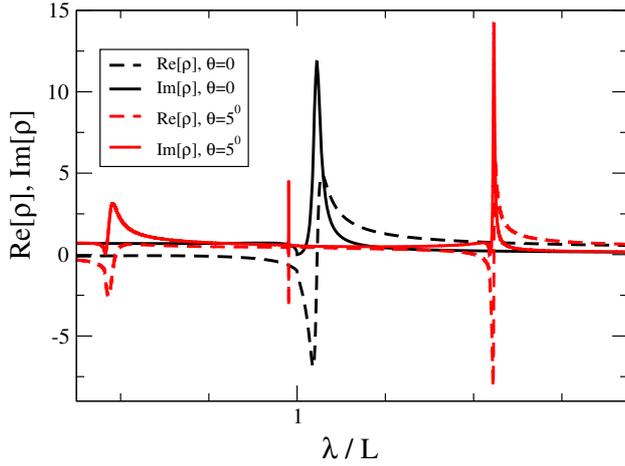


Figure 5. Real (discontinuous lines) and imaginary (continuous lines) parts of ρ , the reflection amplitude for a Bloch combination of waveguide modes impinging into vacuum. The Bloch wavevector forms either 0° with the normal to the surface (black curve) or 5° (red curve). The side of the hole is $a = 0.4L$, and the film thickness is $h = 0.2L$.

to a single interface. The width of the $\text{Im}(\rho(\lambda))$ peak, $\Delta\rho$, is related to the lifetime of the EM surface mode, t_R , which can be estimated from $\Delta\omega t_R \sim 2\pi$ as $t_R \approx t_L(\lambda/L)^2/(\Delta\rho/L)$, where t_L is the time it takes for the surface mode to cross a unit cell.

Figure 5 presents spectral regions with $|\rho| > 1$, implying that the reflected field has larger amplitude than the incident one. This is counter-intuitive, as it seems to indicate that the reflected current is larger than the incident one. However, a similar calculation to that leading to equation (8) shows that, for evanescent modes, current conservation only dictates $\text{Im}(\rho) > 0$, saying nothing about $|\rho|$ (see also discussion at the end of section 2.2).

Obviously, the transmission through a slab involves two interfaces. How do surface modes couple in a metal film? How are they related to the transmittance spectra? The answer to these questions is illustrated in figure 6, which presents results for an array of square holes with $a = 0.4L$, illuminated at normal incidence. The upper panel shows the transmittance spectra for different metal thicknesses, while the middle panel renders the corresponding spectral dependence for both e_h^{-2} (which in the subwavelength regime is always larger than one and increases exponentially with h) and $|\rho|$ (that, being a two-media scattering coefficient, does not depend on metal thickness).

As figure 6 shows, transmittance maxima occur at the wavelengths of minimal distance between the $|\rho|$ and e_h^{-2} curves. This corresponds to the graphical solution for the minima of the ‘renormalization’ denominator D which, as we have previously shown, mark the presence of coupled surface modes. Depending on the metal thickness, two transmission regimes appear for any given resonance of $|\rho|$. For small metal thicknesses (but still larger than 3–4 skin depths, so the metal is optically opaque and the considered model makes sense), the curves for $|\rho|$ and e_h^{-2} cross twice, leading to the presence

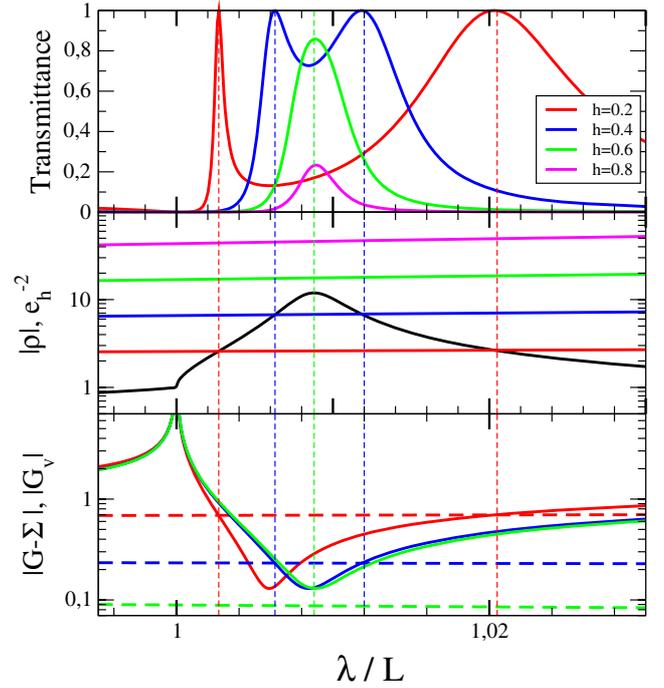


Figure 6. Upper panel: normal-incidence transmittance spectra through a hole array in PEC, for different metal thicknesses, h . The side of the hole is $a = 0.4L$. Middle panel: spectral dependence of $|\rho|$ (black curve) and e_h^{-2} (color lines, for the corresponding metal thicknesses represented in the upper panel). Lower panel: spectral dependence of $|G - \Sigma|$ (continuous lines) and $|G_v|$ (discontinuous lines). The thin vertical discontinuous lines are guides to the eye marking the wavelengths of minimum distance between $|\rho|$ and the different e_h^{-2} curves.

of two transmission maxima which, for a lossless metal reach 100% transmission [13, 32]. This is the regime of strong coupling between surface modes: before being radiated, the EM field stays long enough at the surface to be able to build coupled resonances. For films thick enough, e_h^{-2} is larger than the maximum value of $|\rho|$. In this case, the curves for $|\rho|$ and e_h^{-2} do not cross but there is still one transmission maximum at the wavelength of minimum distance between them, i.e., very approximately at the spectral position of the maximum of $|\rho|$. In this situation, the surface modes of the two surfaces are weakly coupled: the time it would take to build up the resonance (t_{res}) is smaller than the radiation lifetime (t_{rad}) or, in other words, the EM field does not stay long enough in the system to realize that there are two coupled modes. Conversely to what occurs in the strong coupling regime, transmission maxima in the weak coupling regime decay exponentially with h , even in the absence of absorption. Clearly, whether an EOT peak is in the strong coupling or weak coupling regime strongly depends on both the metal thickness (which controls the time for the resonance build-up) and the geometry of the openings (which determines the radiation lifetime). The existence of these two transmission regimes was experimentally confirmed in [42].

The analysis of the existence of coupled leaky modes, and their relation to transmission maxima, can also be done within the formalism involving E and E' . In this case the

resonant condition is expressed in terms of the denominator $(G - \Sigma)^2 - G_v^2$. The lower panel of figure 6 renders the spectral dependence of both $|G - \Sigma|$ and $|G_v|$, for the different metal thicknesses considered in the upper and middle panels of the figure. As the figure shows, there is a bi-univocal correspondence between the spectral position of transmission maxima and the wavelengths of minimum distance between $|G - \Sigma|$ and $|G_v|$. Notice that the resonances appear close to the divergences of G occurring at the wavelengths $\lambda_{m,n}^{\min}$ (both for a PEC and for a real metal).

The previous analysis was done for an infinite hole array and a lossless metal. Going beyond this idealization introduces (at least) another two time scales: the typical time the EM field needs for crossing the finite array and the typical time the EM field can stay in the system before being absorbed. EOT peaks will be largely impaired whenever any of these times is smaller than t_{res} .

3.2. Spoof surface plasmons

As mentioned before, an array of indentations in a PEC is able to bind surface electromagnetic modes. The band structure of these modes can be obtained by finding the solutions of $G - \Sigma = 0$ (or, alternatively, from the spectral position of scattering resonances) as a function of the incident wavevector \mathbf{k}_0 . Complex values of \mathbf{k}_0 must be considered in the search for leaky modes, while truly bound surface modes can only show up for evanescent incident wavefields (with $|\mathbf{k}_{0\parallel}| > g$).

However, the overall form of the band structure can be obtained without the need for numerical computations. For this, it is convenient to consider that the metal has an arbitrarily large (but not infinite) negative dielectric constant. In this case, a flat metal surface supports truly bound surface plasmon polaritons. The presence of a periodic array of small holes can be considered within a perturbative approach. The result is that the dispersion relation of EM surface modes in the corrugated structure will closely follow the one for surface plasmon polaritons, except for \mathbf{k}_0 values lying close to a Brillouin zone boundary, where bands bend in order to accommodate for band gaps. For frequencies above the first band gap, surface modes couple to radiation, thus becoming leaky. However, the band sector below the first band gap still represents a truly bound surface mode. As a result of the band bending caused by the array of holes, the lowest band of surface modes separates from the light line ($\omega = ck$), therefore binding the EM field more strongly to the surface. This reasoning has been discussed before [43], and is at the heart of the whole field of frequency selective surfaces. Notice that this structure of the dispersion relation is based on general arguments, being applicable not only to plasmon polaritons but to all kind of waves in periodic media.

However, in the particular case of the EM field in patterned surfaces, there is an additional mechanism to periodicity-induced band bending for binding the field to the surface. In order to illustrate this point, we compute the bands of a holey metal surface within the PEC approximation, in which case flat surfaces do not support bound plasmon polaritons. In a first approximation, we consider the hole array as a metamaterial

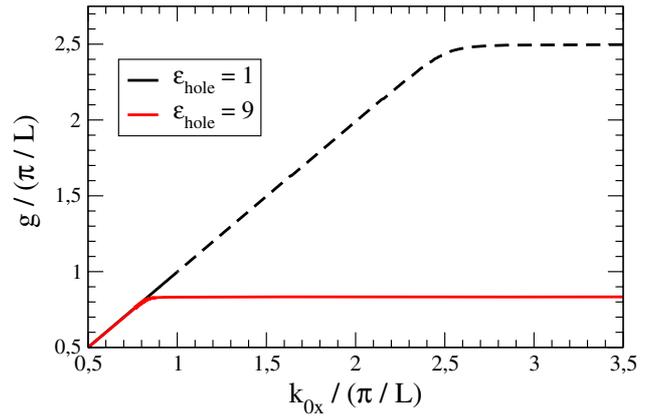


Figure 7. Dispersion relations for spoof surface plasmons in a semi-infinite perforated metal, after matching of the average fields at the surface (see text). The side of the hole is $a = 0.4L$. A medium with dielectric constant $\epsilon = 1$ (black curve), or $\epsilon = 9$ (red curve), fills the holes. The discontinuous line marks the spectral region of leaky modes.

and compute the band structure of surface modes by matching the *average* fields over the surface; the effect of periodicity will be included in a second stage, following a reasoning similar to the one described above. Matching of average fields can be readily done by using the techniques described in this paper (for instance by computing the resonances in ρ), but considering only the term with $\mathbf{K}_R = 0$. Notice that, in this way, all information about the lattice is lost.

Figure 7 renders the band structure along the Γ - X direction for surface modes bound to an isolated holey interface, for both $\epsilon_{\text{hole}} = 1$ (black curve) and $\epsilon_{\text{hole}} = 9$. Notice that bands flatten at certain frequencies, although no diffraction effects have been included yet.

This result can be expressed in the metamaterial language by assigning an effective dielectric constant (ϵ_{eff}) and effective magnetic permeability (μ_{eff}) to the structured surface. In order to find the value of these coefficients, we consider an incoming wave from medium I, impinging into a semi-infinite medium II. The two-media reflection coefficient $\rho_{\mathbf{k}\sigma}$ is given by equation (27). From the expressions for I_0 and Γ and the average value for $G = \iota Y_1 S^2$ (coming from the term with $\mathbf{K}_R = 0$), we obtain:

$$\rho_{\mathbf{k}\sigma} = \frac{S^2 Y_1 - Y^{\text{II}}}{S^2 Y_1 + Y^{\text{II}}} = \frac{Y_1 - Y^{\text{II}}/S^2}{Y_1 + Y^{\text{II}}/S^2} \quad (36)$$

where we have used the shortened notation $Y_1 \equiv Y_{\mathbf{k}\sigma}$ and $S \equiv |\langle \mathbf{k}, \sigma | 0 \rangle|^2$.

This expression resembles the reflection coefficient $\rho_{A \rightarrow B}$ of waves coming from media A (characterized by an admittance Y_A) into media B (with admittance Y_B). For a flat interface, $\rho_{A \rightarrow B} = (Y_A - Y_B)/(Y_A + Y_B)$. Therefore, equation (36) states that the field coming from medium I ‘sees’ the structured metal surface as if it were a uniform medium with (effective) admittance $Y_{\text{eff}} \equiv \sqrt{\epsilon_{\text{eff}}/\mu_{\text{eff}}} = Y^{\text{II}}/S^2$.

The effective admittance involves a combination of both ϵ_{eff} and μ_{eff} , so these coefficients cannot yet be extracted. We obtain a second equation involving ϵ_{eff} and μ_{eff} by enforcing

the EM fields to decay in the effective media as they do in the structured surface, i.e., by imposing $k_z^{\text{eff}} = k_z^{\text{II}}$, which implies $g\sqrt{\epsilon_{\text{eff}}\mu_{\text{eff}}} = gY^{\text{II}}$.

Finally, we get $\epsilon_{\text{eff}} = (Y^{\text{II}})^2/S^2$, and $\mu_{\text{eff}} = S^2$. The expression for ϵ_{eff} is of the form:

$$\epsilon_{\text{eff}}(\omega) = \epsilon_{\text{hole}} S^{-2} \left(1 - \frac{\omega_{\text{p}}^2}{\omega^2} \right). \quad (37)$$

This functional form for $\epsilon_{\text{eff}}(\omega)$ is similar to Drude's expression for the dielectric constant of a metal. Therefore, it can be said that a corrugated PEC surface spoofs a flat surface of a *real* conductor, characterized by a (geometry dependent) 'effective plasma frequency' $\omega_{\text{p}} = (c/\sqrt{\epsilon_{\text{hole}}})\pi/a$, which coincides with the cutoff frequency of the hole.

This is the result derived in [26]. Actually, the system is anisotropic, so care must be taken about the different components of the effective dielectric constant tensor (see [26]). The anisotropy is also responsible for the fact that the flat region of the dispersion curve for *spoof surface plasmons* appears at $\epsilon_{\text{eff}} = 0$, whereas the dispersion relation for surface plasmons bounded to the interface between two *isotropic* media flattens at $\epsilon = -1$.

Periodicity has two effects on the spoof plasmon band structure: it opens gaps at wavevectors \mathbf{k} close to Brillouin zone boundaries and couples bands with $|\mathbf{k}| > \pi/L$ to radiative modes. This is illustrated in figure 7, where the discontinuous lines represent modes that become leaky when diffraction effects are considered.

Notice that spoof surface plasmons can still be expected if the holes are not arranged on a crystalline two-dimensional lattice. In this case, however, spoof surface plasmons will be leaky surface modes for all frequencies.

Spoof surface plasmons close to the flat region of the dispersion curve are interesting because they are strongly bound to the surface. As previously said, this occurs for wavelengths close to the cutoff of the hole, i.e., $\lambda \approx 2\sqrt{\epsilon_{\text{hole}}}a$. More generally, all plane-wave components of any truly bound surface mode must be evanescent, which translates into $|\mathbf{k}_{\parallel} + \mathbf{K}_{\text{R}}| > g$ for all \mathbf{K}_{R} , where \mathbf{k}_{\parallel} is the in-plane component of \mathbf{k} . For $k_y = 0$, the previous condition requires $k_x - 2\pi/L < -g$, and $k_x > g$. Therefore, truly bound modes can only exist for $\lambda > 2L$, independently of how tightly bound they are. Together with the obvious condition that a should be smaller than L , this implies that strongly bound surface modes can only exist in square lattices of square arrays if the holes are filled with a material with $\epsilon_{\text{hole}} > 1$. Figure 7 renders such a case: all surface modes depicted by the red curve (for which $\epsilon_{\text{hole}} = 9$) are truly bound to the surface, even if diffraction modes are considered.

Up to here in this section, we have considered only the case of an isolated interface and subwavelength holes. Interesting phenomena appear when we consider a metal film, and/or modes inside the hole are not evanescent. This last situation resembles the slit array case [27], when the effective surface admittance can be further engineered, by playing with the depth of the hole. Here it is even possible to make the corrugated metal surface mimic a dielectric, in which case

bound modes are better pictured by waveguide modes in the 'effective' dielectric film than by surface plasmons.

Surface EM modes still appear if higher order waveguide and diffraction modes are included in the calculation, and their dispersion relation still flattens at $\omega = \omega_{\text{p}}$, given by the hole cutoff. However, strong confinement only occurs for frequencies much closer to ω_{p} than what the effective parameter expression predicts [44]. Also, the couplings between waveguide modes and plane waves depend on the wavevector direction of the latter [44]. In order to take this effect into account, one must go beyond a local theory, limiting the applicability of the 'metamaterial' effective parameters. Still, the effective parameter concept is useful for guiding our intuition in different circumstances, for instance in the pursuit of spoof surface plasmon controlled propagation through modulation of the underlying periodicity.

At frequencies above the cutoff frequency of the hole, spoof surface plasmons are leaky and can be detected in transmission [18] or reflection experiments [45]. But below the cutoff frequency they can not be accessed by radiation. One possible way to excite and detect truly bound spoof surface plasmons is to use the energy loss of a charged particle passing close to the surface [46]. In the next section we describe how such a problem could be analyzed within the present model.

3.3. Energy loss of a charged particle passing close to a holey surface

Consider a particle with charge q , moving with constant speed v at a distance d from a holey metal surface (so, with our definitions, at $z = d - h/2$). Assuming that the energy loss into the surface is much smaller than the particle's kinetic energy, the calculation of the power loss spectra can be translated into that of the reflection coefficients of the EM fields associated to the particle.

Let us briefly sketch the method described in [47]. First, the fields associated to the particle in its rest reference frame are Lorentz transformed onto the surface reference frame (in which the particle moves with $\beta = v/c$). Decomposing this field into plane waves, the parallel components of the electric field can be expressed as:

$$\mathbf{E}_{\parallel}(\mathbf{r}, t) = C \int dg \mathbf{E}_{\parallel}(\mathbf{r}, g) \exp(-t c g t) \quad (38)$$

where $\omega = c g$, C is a proportionality constant (important for the actual value of the power loss but not for its spectral dependence), and:

$$\begin{aligned} \mathbf{E}_{\parallel}(\mathbf{r}, g) = & \int dk_y (k_x(1 - \beta^2), k_y)^{\text{T}} \exp(i\mathbf{k}_{\parallel}\mathbf{r}_{\parallel}) \\ & \times \exp(i\mathbf{k}_z|z - d + h/2|)/k_z. \end{aligned} \quad (39)$$

In the previous expression, for each frequency the value of k_x must satisfy $k_x = g/\beta$. As $k_x > g$, all plane waves in equation (39) are evanescent, which reflects the well-known result that an isolated charged particle moving at constant speed does not radiate energy.

The projection of this wavefield into the basis set of plane waves considered in this paper gives:

$$|E_{\parallel}(z, g)\rangle = \sum_{\sigma} \int dk_y C_{\mathbf{k},\sigma} |\mathbf{k}, \sigma\rangle \exp(i\mathbf{k}_z |z - d + h/2|) \quad (40)$$

with $C_{\mathbf{k},p} = (k_x^2(1 - \beta^2) + k_y^2)/(k_z k_{\parallel})$, and $C_{\mathbf{k},s} = (k_x k_y \beta^2)/(k_z k_{\parallel})$.

These are the fields attached to the isolated moving particle. In order to calculate the EM fields everywhere in space, the formalism described in this paper can be applied, considering each of the plane waves in the integral describing $|E_{\parallel}(\mathbf{r}, g)\rangle$ as an ‘incident field’.

Once this is done, the power loss spectra can be obtained by computing the energy current across two surfaces with constant z : one between the particle and $z = -\infty$ (which would give the energy radiated away into vacuum), and another one at z between the particle and the surface (giving the energy radiated into the surface). Part of this energy loss is due to modes that become radiative, after picking up a reciprocal lattice wavevector. Another part of the energy loss is due to evanescent modes: the presence of an evanescent reflected wave takes energy away from the particle. A straightforward computation gives that this contribution is proportional to

$$\text{Power loss} \sim |C_{\mathbf{k},\sigma}|^2 \text{Re}(iY_{\mathbf{k},\sigma}) \text{Im}(r_{\mathbf{k}_{\parallel},\sigma \rightarrow \mathbf{k}_{\parallel},\sigma}) \times \exp(-2 \text{Im}(k_z)). \quad (41)$$

We have explicitly included in $r_{\mathbf{k}_{\parallel},\sigma \rightarrow \mathbf{k}_{\parallel},\sigma}$ the dependence on both incident and reflected wavevector, in order to stress that only the ‘specular’ reflection coefficient is required. The fact that this reflection coefficient diverges when surface modes are excited makes energy loss an interesting technique for the study of spoof plasmons.

In practice, our model is too simple to obtain realistic values for the energy loss. Recall that we have only considered in this paper one of the degenerate fundamental waveguide modes. Electric fields associated to the particle also have y components and their consideration requires a multimode expansion. Therefore, no attempt has been made to estimate the energy loss in this paper.

4. Summary

Our paper contains a tutorial presentation for the modal expansion formalism, applied to the optical properties of a holey metal film. We have developed the simplest possible model able to capture the phenomenon of extraordinary optical transmission through hole arrays, by considering the metal as a perfect conductor and using a single mode approximation. We re-derive several results known in the field within this simple model which, we believe, makes the relevant physical mechanism very transparent. We show that EOT is intrinsically related to the presence of leaky surface EM modes. Finally, we study the properties of these surface modes and derive a formalism for analyzing them through the energy loss of a charged particle passing close to the surface.

Acknowledgments

The authors would like to thank Professor J B Pendry for introducing them to the fascinating field of electromagnetic phenomena in structured media, and for the many exciting discussions and fruitful collaborations enjoyed with him ever since.

Financial support from the STREP ‘Plasmon Enhanced Photonics’ (IST-FP6-034506) and the Spanish ‘Ministerio de Educacion y Ciencia’ under contract MAT2005-06608-C02 is also acknowledged.

References

- [1] Ebbesen T W, Lezec H J, Ghaemi H F, Thio T and Wolff P A 1998 *Nature* **391** 667
- [2] Lezec H J, Degiron A, Devaux E, Linke R A, Martín-Moreno L, García-Vidal F J and Ebbesen T W 2002 *Science* **297** 820
- [3] Martín-Moreno L, García-Vidal F J, Lezec H J and Ebbesen T W 2003 *Phys. Rev. Lett.* **90** 167401
- [4] García-Vidal F J, Lezec H J, Ebbesen T W and Martín-Moreno L 2003 *Phys. Rev. Lett.* **90** 213901
- [5] Klein Koerkamp K J, Enoch S, Segerink F B, van Hulst N F and Kuipers L 2004 *Phys. Rev. Lett.* **92** 183901
- [6] Degiron A, Lezec H J, Yamamoto N and Ebbesen T W 2004 *Opt. Commun.* **239** 61
- [7] van der Molen K L, Klein Koerkamp K J, Enoch S, Segerink F B, van Hulst N F and Kuipers L 2005 *Phys. Rev. B* **72** 045421
- [8] García-Vidal F J, Moreno E, Porto J A and Martín-Moreno L 2005 *Phys. Rev. Lett.* **95** 103901
- [9] Porto J A, Martín-Moreno L and García-Vidal F J 2004 *Phys. Rev. B* **70** 081402(R)
- [10] Wurtz G A, Pollard R and Zayats A V 2006 *Phys. Rev. Lett.* **97** 057402
- [11] van Nieuwstadt J A, Sandtke M, Harmsen R H, Segerink F B, Prangma J C, Enoch S and Kuipers L 2006 *Phys. Rev. Lett.* **97** 146102
- [12] Barnes W L, Dereux A and Ebbesen T W 2003 *Nature* **424** 824 and references therein
- [13] Martín-Moreno L, García-Vidal F J, Lezec H J, Pellerin K M, Thio T, Pendry J B and Ebbesen T W 2001 *Phys. Rev. Lett.* **86** 1114
- [14] Martín-Moreno L and García-Vidal F J 2004 *Opt. Express* **12** 3619
- [15] Genet C, van Exter M P and Woerdman J P 2003 *Opt. Commun.* **225** 331
- [16] Qu D, Grischkowsky D and Zhang W 2004 *Opt. Lett.* **29** 896
- [17] Ye Y-H and Zhang J Y 2004 *Appl. Phys. Lett.* **84** 2977
- [18] Beruete M, Sorolla M, Campillo I, Dolado J S, Martín-Moreno L, Bravo-Abad J and García-Vidal F J 2004 *Opt. Lett.* **29** 2500
- [19] Rivas J, Schotsch C, Haring Bolívar P and Kurz H 2003 *Phys. Rev. B* **68** 201306
- [20] Marquier F, Joulain K and Greffet J J 2004 *Opt. Lett.* **29** 2178
- [21] Kramper P, Agio M, Soukoulis C M, Birner A, Müller F, Wehrspohn R B, Gosele U and Sandoghdar V 2004 *Phys. Rev. Lett.* **92** 113903
- [22] Moreno E, García-Vidal F J and Martín-Moreno L 2004 *Phys. Rev. B* **69** 121402(R)
- [23] Moreno E, Fernández-Dominguez A I, Cirac J I, García-Vidal F J and Martín-Moreno L 2005 *Phys. Rev. Lett.* **95** 170406
- [24] Hou B, Mei J, Ke M, Wen W, Liu Z, Shi J and Sheng P 2007 *Phys. Rev. B* **76** 054303

- [25] Christensen J, Fernandez-Dominguez A, deLeon-Perez F, Martin-Moreno L and Garcia-Vidal F J 2007 *Nat. Phys.* **3** 851
- [26] Pendry J B, Martin-Moreno L and García-Vidal F J 2004 *Science* **305** 847
- [27] Garcia-Vidal F J, Martin-Moreno L and Pendry J B 2005 *J. Opt. A: Pure Appl. Opt.* **7** S97–101
- [28] Popov E, Neviere M, Enoch S and Reinisch R 2000 *Phys. Rev. B* **62** 16100
- [29] Salomon L, Grillot F, Zayats A V and de Fornel F 2001 *Phys. Rev. Lett.* **86** 1110
- [30] Baida F I and Van Labeke D 2003 *Phys. Rev. B* **67** 155314
- [31] Sarrazin M and Vigneron J P 2005 *Phys. Rev. B* **71** 075404
- [32] García de Abajo F J, Gómez-Medina R and Sáenz J J 2005 *Phys. Rev. E* **72** 0166608
- [33] Kats A V, Nesterov M L and Nikitin A Yu 2007 *Phys. Rev. B* **76** 045413
- [34] Liu H and Lalanne P 2008 *Nature* **452** 728
- [35] Bravo-Abad J, Garcia-Vidal F J and Martin-Moreno L 2004 *Phys. Rev. Lett.* **93** 227401
- [36] Ekinci Y, Solak H H and David C 2007 *Opt. Lett.* **32** 172
- [37] Martin O J, Girard C and Dereux A 1995 *Phys. Rev. Lett.* **74** 526
- [38] Gordon R and Brolo A G 2005 *Opt. Express* **13** 1933
- [39] Rodrigo S G, Garcia-Vidal F J and Martin-Moreno L 2008 *Phys. Rev. B* **77** 075401
- [40] See for instance Landau L D and Lifshitz E M 1984 *Electrodynamics of Continuous Media* 2nd edn (Oxford: Pergamon)
- [41] Lopez-Tejiera F, Garcia-Vidal F J and Martin-Moreno L 2005 *Phys. Rev. B* **72** 161405
- [42] Degiron A, Lezec H J, Barnes W L and Ebbesen T W 2002 *Appl. Phys. Lett.* **81** 4327
- [43] Ulrich R and Tacke M 1973 *Appl. Phys. Lett.* **22** 251
- [44] Garcia de Abajo F J and Saenz J J 2005 *Phys. Rev. Lett.* **95** 233901
- [45] Hibbins A P, Evans B R and Sambles J R 2005 *Science* **308** 670
- [46] Young-Min S, Jin-Kyu S, Kyu-Ha J, Jong-Hyo W, Anurag S and Gun-Sik P 2007 *Phys. Rev. Lett.* **99** 147402
- [47] Pendry J B and Martin-Moreno L 1994 *Phys. Rev. B* **50** 5062