In the diffraction shadow: Norton waves versus surface plasmon polaritons in the optical region

2009 New J. Phys. 11 123020

(http://iopscience.iop.org/1367-2630/11/12/123020)

New Journal of Physics
The open-access journal for physics

In the diffraction shadow: Norton waves versus surface plasmon polaritons in the optical region

A Yu Nikitin[1,2], Sergio G Rodrigo[1], F J García-Vidal[3] and L Martín-Moreno[1,4]

[1] Instituto de Ciencia de Materiales de Aragón and Departamento de Física de la Materia Condensada, CSIC-Universidad de Zaragoza, E-50009, Zaragoza, Spain
[2] A Ya Usikov Institute for Radiophysics and Electronics, Ukrainian Academy of Sciences, 61085 Kharkov, Ukraine
[3] Departamento de Física Teórica de la Materia Condensada, Universidad Autónoma de Madrid, E-28049 Madrid, Spain
E-mail: lmm@unizar.es

New Journal of Physics **11** (2009) 123020 (15pp)
Received 7 October 2009
Published 16 December 2009
Online at http://www.njp.org/
doi:10.1088/1367-2630/11/12/123020

Abstract. Surface electromagnetic modes supported by metal surfaces have a great potential for use in miniaturized detectors and optical circuits. For many applications, these modes are excited locally. In the optical regime, surface plasmon polaritons (SPPs) have been thought to dominate the fields at the surface, beyond a transition region comprising 3–4 wavelengths from the source. In this work, we demonstrate that at sufficiently long distances SPPs are not the main contribution to the field. Instead, for all metals, a different type of wave prevails, which we term Norton waves (NWs) for their resemblance to those found in the radio-wave regime at the surface of the Earth. Our results show that NWs are stronger at the surface than SPPs at distances larger than 6–9 SPP absorption lengths, the precise value depending on wavelength and metal. Moreover, NWs decay more slowly than SPPs in the direction normal to the surface.

[4] Author to whom any correspondence should be addressed.

The confinement of the electromagnetic (EM) field associated with surface plasmon polaritons (SPPs), and their intrinsic speed, make them very interesting candidates for use in photonics [1, 2]. Due to this, the study of the EM fields radiated by localized sources (like defects [3], nano-gratings or apertures [4]) placed on a surface has received a renewed interest in the last decade. This is an old problem, which came to prominence in the early 1900s due to its possible relevance to the transmission of radio signals. The seminal works of Zenneck [5] and Sommerfeld [6] unveiled the existence of surface waves running along the Earth, which can be considered as a lossy dielectric. The interest in these works waned after the realization that radio transmission does not occur via the exponentially damped surface modes, but through reflection at the ionosphere. Nevertheless, Norton subsequently showed that, in the long distance limit, radio waves decay algebraically at the surface [7]. This result triggered a debate on the range of validity of Zenneck–Sommerfeld and Norton waves (NWs) in the radio regime that has persisted to the present day (see [8] for more details and a historical account). Recently, advances in nanofabrication have allowed the scaling down of old radio devices into the optical regime [9]. Metallic surfaces are especially interesting because they support SPPs, which are surface EM modes strongly confined to the plane. The analysis of the surface EM fields created by a localized source in a metal surface has revealed the existence of a near-field region, extending for 3–4 wavelengths, where the field presents a complex dependence [10, 11]. SPPs have been thought to dominate the EM field beyond this region. In this work, we show that, irrespective of the metal considered, the long-distance asymptotic limit of the EM field at the metal surface is not the SPP but a different type of wave, which we denote as NWs due to their resemblance to those found in dielectric surfaces. We show the range of validity of SPPs and NWs and the distance and field amplitude after which the latter dominate.

Although we will show later how the obtained results apply to dipole sources, let us concentrate first on the EM fields emerging from a subwavelength slit, placed in an optically thick metal film. The film is back-illuminated by normal-incident p-polarized light with wavelength λ (i.e. the wavevector in vacuum is $g = 2\pi/\lambda$). The frequency-dependent dielectric constant of the metal is ϵ_{m}. This system has been chosen for analytical simplicity (the full EM field can be expressed in terms of the magnetic field along the slit axes, $H_y(X, Z)$) and, also because it is a configuration that has been amply studied both theoretically [12]–[17] and experimentally [18]–[21]. Figure 1 is a snapshot of the radiated $H_y(X, Z)$ (computed with the finite difference time domain (FDTD) method) for slit width of $A = 100\,\mathrm{nm}$ and $\lambda = 540\,\mathrm{nm}$, for both a perfect electrical conductor (PEC, characterized by $|\epsilon_{\mathrm{m}}| = \infty$) and Au [22][5]. The choice of metal and wavelength is motivated for proof-of-principle purposes on the existence of NWs, but we will show later on that our results are applicable to other metals and frequency ranges. Our treatment fully takes into account the vectorial nature of the EM fields and, therefore, goes beyond the scalar approximations considered in other works [17, 20].

It is apparent from figure 1 that the effect of a finite ϵ_{m} is to strongly modify the radiation pattern close to the metal surface. Although we will provide expressions for the field everywhere, our main focus will be to characterize the fields within the diffraction shadow, which loosely speaking is the region where radiation from a slit in a real metal is strongly reduced with respect to the PEC case.

[5] All dielectric constants used are taken from the fit of experimental results in Drude–Lorentz terms, see [22] and references therein.

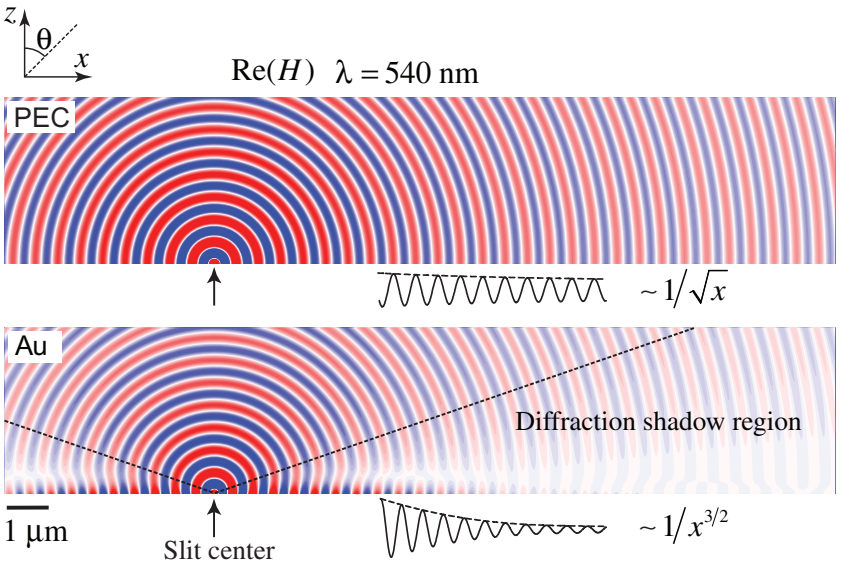

Figure 1. Snapshot of the magnetic field radiated by a subwavelength slit in an optically thick metal film, back-illuminated by p-polarized light. In the top panel, the metal is treated as a PEC, while the metal in the lower panel is Au. The wavelength is $\lambda = 540\,\text{nm}$.

The Green's dyadic method is more suitable for an analytical study of this problem. Within this method, the field radiated at the point $\mathbf{R} = (X, Z)$ by a slit of width A is (see appendix A for the justification of this expression and its validation with numerical calculations)

$$H(x, z) \approx \sqrt{\varepsilon_{\text{m}} - 1} \int_{-a/2}^{a/2} G(x - x', z)\, E_x(x', z = -\delta)\, \text{d}x', \tag{1}$$

where the Green's function $G(x, z)$ is the magnetic field generated by a dipolar source with the electric field pointing along the x-direction, placed at the metal interface, and δ is the skin depth for the metal. In this expression and throughout the paper, all distances denoted by lower case letters are expressed in dimensionless units as $x = gX$, $z = gZ$ and $a = gA$. Alternatively, given that the fundamental waveguide mode inside the slit is constant in the x-direction, $G(x, z)$ can be seen as the magnetic field radiated by an infinitesimally thin slit. The angular spectrum representation of this function is

$$G(x, z) = \int_{-\infty}^{\infty} D(q)\, \text{e}^{\text{i}qx + \text{i}q_z z}\, \text{d}q, \tag{2}$$

where q is the x-component of the wavevector (in units of g), $D(q) = q_{zm}/[2\pi\,(\epsilon_m q_z + q_{zm})]$, $q_z = \sqrt{1 - q^2}$ and $q_{zm} = \sqrt{\epsilon_m - q^2}$.[6]

The solution to this integral is not known in the closed form. Fortunately, there are mathematical methods (see [23]) for extracting its long-distance asymptotic expression, $G_{\text{asymp}}(x, z)$. The rigorous calculation for $G_{\text{asymp}}(x, z)$ is provided in appendix B and, additionally, a simplified derivation will be given later on. But before going into the

[6] The sign in the square root must be taken so that $\text{Im}(q_z) \geqslant 0$, in order to satisfy the radiation condition.

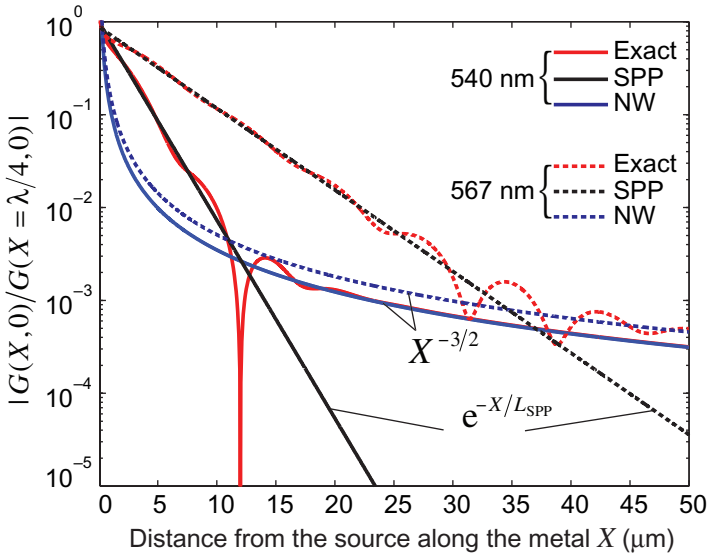

Figure 2. The magnetic field at the metal surface radiated by a horizontal dipole as a function of distance. The dependences are presented for Au at two wavelengths: 540 nm (continuous curves) and 567 nm (discontinuous curves). The figure shows the exact result (red) and the SPP (black) and NW (blue) contributions.

mathematical details, let us now concentrate on the fields at the metal surface and give the result obtained:

$$G_{\text{asymp}}(x,0) = G_{\text{SPP}}(x,0) + G_{\text{NW}}(x,0). \tag{3}$$

In this expression, $G_{\text{SPP}}(x,0)$ is the SPP contribution

$$G_{\text{SPP}}(x,0) = 2\pi \mathrm{i}\, C_{\text{p}}\, \mathrm{e}^{\mathrm{i}q_{\text{p}}x}, \tag{4}$$

where $q_{\text{p}} = \sqrt{\varepsilon_{\text{m}}/(1+\varepsilon_{\text{m}})}$ is the SPP momentum, and $C_{\text{p}} = q_{\text{p}}^3/[2\pi(\varepsilon_{\text{m}}-1)]$ is the residue of $D(q)$ at q_{p}.

The second term is

$$G_{\text{NW}}(x,0) = \frac{\mathrm{e}^{\mathrm{i}x+\mathrm{i}\pi/4}}{\sqrt{2\pi}} \frac{\varepsilon_{\text{m}}}{\sqrt{\varepsilon_{\text{m}}-1}} x^{-3/2}. \tag{5}$$

As will be shown later, this term is the two-dimensional (2D) optical analogue in metal surfaces of the NW [7] found in the study of the radio-wave radiation of point dipoles on lossy dielectric interfaces. Dimensionality accounts for the difference between the decay laws: $x^{-3/2}$ (2D dipoles) and the x^{-2} (3D dipoles).

The validity of equation (3) and the competition between SPPs and NWs is illustrated in figure 2, which shows the magnetic field at the surface radiated by an infinitesimally thin subwavelength slit, for Au at two different wavelengths. In each case this figure shows the exact result (computed numerically from equation (1)) and the SPP and NW contributions. For the cases considered in this figure, the asymptotic result given by equation (3) is virtually indistinguishable from the exact result even for $X \approx 3\,\mu$m and it is not represented. The field is mainly SPP-like at the shorter distances, while NW dominates at sufficiently long distances from the source. Note that the relative phase of the NW and the SPP contributions at the distance

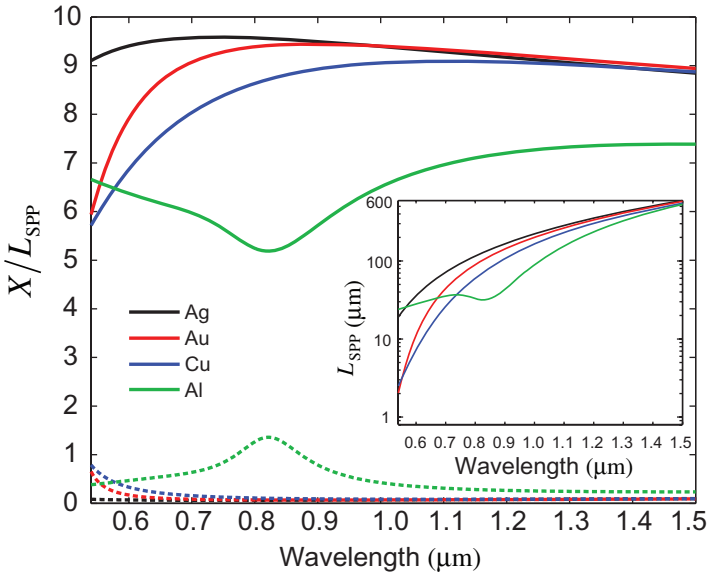

Figure 3. Spectral dependence of the crossover between SPP and NW at the metal surface, for different metals. The continuous lines represent x_{NW}, defined as the distance at which the amplitude of the NW is larger than the SPP one. The discontinuous lines represent $x_a(0.1)$ the minimum distance at which G_{asymp} gives a relative error of 10% with respect to the exact result. The inset shows the spectral dependence SPP absorption length for the considered metals.

where their modulus are equal changes with wavelength. Hence, their destructive interference may lead to the cancellation of the field (as in Au, at $\lambda = 540\,\text{nm}$ at $X \approx 12\,\mu\text{m}$) or, if the cancellation is not complete, to the appearance of small oscillations in the total field amplitude (as in the presented case of Au at $\lambda = 567\,\text{nm}$). It is worth noting that similar oscillations were found in scanning near field optical microscope experiments in Au [20], but their origin was unknown.

Beyond the particular examples presented in figure 2, the expression given by equation (3) is a good approximation for the field at the surface, at sufficiently long distances from the source. In order to quantify this statement, we define $x_a(\beta)$ as the minimum distance such that $|(G(x_a, 0) - G_{\text{asymp}}(x_a, 0))/G(x_a, 0)| < \beta$. The discontinuous lines in figure 3 show the spectral dependence of the optical and telecom regimes of $x_a(0.1)$ for different metals, in units of the corresponding SPP absorption length. Given that the NW decays algebraically and the SPP exponentially with distance, at sufficiently large distances the NW is the main contribution to the field at the surface, for all metals and all wavelengths. The crossover from NW to SPP is represented in figure 3, which shows the spectral dependence of the distance at which the NW contribution is larger than the SPP one, x_{NW}, for different metals. This distance strongly depends on the dielectric permittivity of the metal, being smaller for very lossy metals, as Cu and Au in the region of inter-band transition (close to $\lambda = 500\,\text{nm}$ for both metals).

Undoubtedly, the existence of NWs in metal surfaces has passed unnoticed up to now due to their small amplitude. In order to characterize how much the field has decayed when the NW takes over, we consider the ratio $|G(X_{\text{NW}}, 0)/G(X = \lambda/4, 0)|$ (the distance $X = \lambda/4$ has been arbitrarily chosen to give a representative reference in the near-field). In the optical regime,

when the NW takes over the field has decayed by a factor ranging from 10^{-2} for lossy metals (like Ni) to 10^{-3}–10^{-4} for Ag. Therefore, the NW is not a good channel for sending information along the surface. Nevertheless, and given that estimations of the field at the surface far away from the source based on the decay of SPPs may be orders of magnitude wrong, NWs may have to be taken into account for precise analysis or design of experiments.

In order to show the origin of NWs, and their relation to other waves discussed in the literature as creeping waves (CWs) and SPPs, let us concentrate on the physical interpretation of the field radiated by the slit. Additionally, this will lead to a poor man's (yet correct) derivation of some of the main results. It is clear from equation (2) that a slit excites the whole range of diffraction modes (both radiative and evanescent) with an amplitude given by $D(q)$, which can loosely speaking be understood as the density of EM modes with a given wavevector q at the slit position[7]. The standard treatment of $G(x, z)$ in the far-field relies on the observation that, although all modes are always present, their contributions cancel out due to destructive interference whenever the phase $\Phi = \mathbf{qr}$ changes rapidly. Thus, only the region in q-space where the phase presents an extremum contributes to the far field. For a given point (x, z) (or (r, θ) in polar coordinates with θ defined as the angle from the normal to the surface), the extremum occurs at the condition $(q/q_z)_{\min} = x/z$, i.e. $q_{\min} = \sin\theta$. Expanding the integrand around this extremum leads to the 'ray-optics' (RO) contribution

$$G_{\mathrm{RO}}(r, \theta) = \sqrt{\frac{2\pi}{r}}\, \mathrm{e}^{\mathrm{i}r - \mathrm{i}\pi/4}\, \cos\theta\, D(\sin\theta)$$
$$= \frac{\mathrm{e}^{\mathrm{i}r - \mathrm{i}\pi/4}}{\sqrt{2\pi r}}\, \frac{\cos\theta \sqrt{\epsilon_{\mathrm{m}} - \sin^2\theta}}{\epsilon_{\mathrm{m}} \cos\theta + \sqrt{\epsilon_{\mathrm{m}} - \sin^2\theta}}. \tag{6}$$

This analytical result reproduces what was observed in figure 1: the magnetic field radiated by an infinitesimally thin slit in a PEC is isotropic, but the pattern in a real metal is strongly modified close to the surface, for angles such that $\cos\theta \lesssim 1/\sqrt{|\epsilon_{\mathrm{m}}|}$.

However, right at the surface the derivative of the phase $\Phi = qx$ never cancels and the saddle point approximation outlined above cannot be directly applied. As the integral of the product of a smooth and rapidly oscillating function is very small, only the parts of the angular spectrum where $D(q)$ changes rapidly in the scale of $2\pi/x$ will give a net contribution to the integral. For very small x, all the 'density of states' contribute. As x increases, the smooth long-q region of $D(q)$ is progressively canceled out in the integral, which is eventually dominated by the strong (and rapid) contribution from the pole in $D(q)$. The contribution of this pole gives the SPP field. Note that, in a lossy metal, the density of states associated with the plasmon pole has a finite width, which causes the exponential decrease of the SPP amplitude with distance (characterized by the SPP propagation length $l_{\mathrm{SPP}} = \mathrm{Im}(q_{\mathrm{p}})^{-1}$).

The previous argument explains why the field at the surface is not the SPP for all distances and is expected to have a complex dependence with x. Recently, Lalanne and Hugonin [11] have assigned the term 'CW' to the difference between the exact field and the approximation given by the SPP pole. Their numerical studies, restricted to distances to the source of a few wavelengths, have shown that the CW is a damped surface wave which, along the surface, oscillates with the free-space wavevector and decays faster than the SPP. A point to note is that despite the $\mathrm{e}^{\mathrm{i}x}$ dependence, the CW arises from the whole angular spectrum, not only from regions close to $q = 1$.

[7] Actually the density of EM modes with wavevector q at the surface is related to $Im(D(q))$.

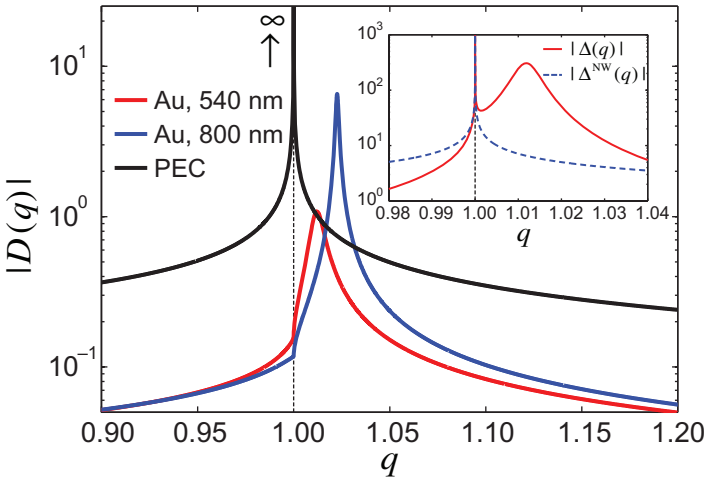

Figure 4. The modulus of $D(q)$, defined in equation (2). The blue and red curves are for the Au surface (at $\lambda = 800$ and 540 nm, respectively), whereas the green curve is for the PEC ($\xi = 0$). The inset shows the same for the function $\Delta(q)$ defined in the text.

However, the SPP pole is not the sharpest feature of $D(q)$: the derivative of $D(q)$ diverges at the branch point $q_z = 0$. This is illustrated in figure 4, which shows that $|D(q)|$ has a kink at $q = 1$. The contribution to the integral from this kink is expected to be small but, as the kink cannot be characterized by a typical width in q-space, it is not as strongly suppressed as the SPP contribution when integrated with an oscillatory function. In order to show that the kink originates the NW, it is convenient to integrate by parts $G(x, 0)$. Then, from equation (2) we obtain $G(x, 0) = (\mathrm{i}/x) \int_{-\infty}^{\infty} \Delta(q)\, \mathrm{e}^{\mathrm{i}qx}\, \mathrm{d}q$, with $\Delta(q) = q\, G'(q)$. This representation has the advantage that the kink in $D(q)$ transforms into a square root singularity (see inset in figure 4). The contribution close to $q_z = 0$ can be retrieved by keeping the singularity but setting $q_z = 0$ everywhere else, this is, by defining $\Delta^{\mathrm{NW}}(q) = (\varepsilon_{\mathrm{m}}/\left[2\pi\sqrt{\varepsilon_{\mathrm{m}} - 1}\right])\,(1/q_z)$. The inset to figure 4 shows the comparison between $\Delta(q)$ and $\Delta^{\mathrm{NW}}(q)$, for a representative case. Of course, $\Delta^{\mathrm{NW}}(q)$ is only a good approximation to $\Delta(q)$ close to $q = 1$, so its use for integration over the whole angular spectrum could seem unjustified. However for very large x this is valid, as only the region close to $q = 1$ contributes. With this, $G_{\mathrm{NW}}(x, 0) = (\mathrm{i}/x) \int_{-\infty}^{\infty} \Delta^{\mathrm{NW}}(q)\, \mathrm{e}^{\mathrm{i}qx}\, \mathrm{d}q = \mathrm{i}(\epsilon_{\mathrm{m}}/\left[2\sqrt{\varepsilon_{\mathrm{m}} - 1}\right]) H_0^{(1)}(x)/x$. Recalling that this expression is only valid for large x, we substitute $H_0^{(1)}(x)$ for its asymptotic value and obtain the result in equation (5). The motivation for our terminology on this type of wave is that the asymptotic term found by Norton, for the case of radio waves emitted by a dipole in a dielectric [7], also decays algebraically and originates from the angular spectrum close to $q = 1$. The SPP contribution can also be extracted from the previous representation by expanding Δ_q close to $q = q_{\mathrm{p}}$. As the NW and the SPP arise from different parts of the angular spectrum, their fields can be directly added up, leading to equation (3).

The relevance of NWs with respect to SPPs increases when we move away from the surface. Since for $z = 0$ the NW originates from q-values close to the light line, its decay with distance to the surface is expected to be slower than the exponential decay of SPPs. In order to obtain the dependence of the NW on both x and z, we have carried out the asymptotic analysis

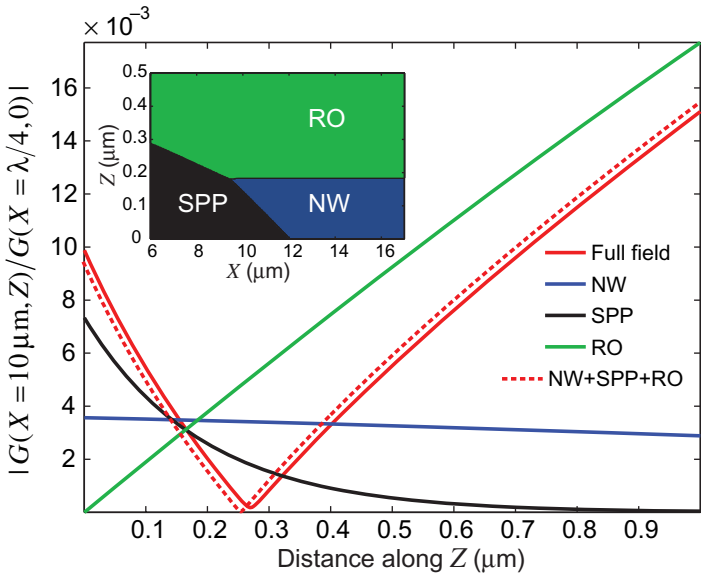

Figure 5. The magnetic field amplitude close to the metal surface for Au at $\lambda = 540\,\text{nm}$. The main figure shows the dependence with Z for fixed $X = 10\,\mu\text{m}$. The exact result $G(X, Z)$ (red curve) and the asymptotic one, $G_{\text{asymp}}(X, Z)$ (red dashed), are presented together with the contribution from the SPP (black curve), RO (green curve) and NW (blue line). The inset shows the regions in the X–Z-plane, where the different terms dominate.

of Green's function given by equation (1), using the general method described in [23]. The derivation and full asymptotic form can be found in appendix B. The result up to terms of the order $r^{-1/2}$ was already presented in [16], where it was shown that the long-distance asymptotic expressions are excellent approximations even for distances as small as $x = 1$ (i.e. $X = \lambda/(2\pi)$). However, the asymptotic expression given by [16] misses some contributions of the order $r^{-3/2}$, so it cannot be used for the problem discussed in this paper. We find that the expression for $G(x, z)$ in the far-field ($r \gg 1$) is (see appendix B):

$$G_{\text{asymp}}(x, z) = G_{\text{SPP}}(x, z) + G_{\text{RO}}(x, z) + G_{\text{NW}}(x, z), \tag{7}$$

where $G_{\text{SPP}}(x, z) = G_{\text{SPP}}(x, 0)\,\mathrm{e}^{\mathrm{i}q_{\text{pz}}z}$, with $q_{\text{pz}} = -1/\sqrt{(1+\varepsilon_{\text{m}})}$, $G_{\text{RO}}(x, z)$ is given by equation (6), and $G_{\text{NW}}(x, z)$, defined as the term that goes as $r^{-3/2}$, is given by

$$G_{\text{NW}}(x, z) = \frac{\mathrm{e}^{\mathrm{i}r+\mathrm{i}\pi/4}}{\sqrt{2\pi}} \left\{ \frac{\mathrm{d}^2}{\mathrm{d}\phi^2} \left[\frac{-\pi D(\sin\phi)\cos(\phi)}{\cos((\phi-\theta)/2)} \right] \right\}_{\phi=\theta} r^{-3/2}. \tag{8}$$

The difference between the exact and asymptotic expressions $\Delta G(x, z) = G(x, z) - G_{\text{asymp}}(x, z)$ decays as $r^{-5/2}$.

Figure 5 presents the comparison between the z-dependence of the exact $G(x, z)$ and $G_{\text{asymp}}(x, z)$ at $X = 10\,\mu m$. These results show that the asymptotic expression is very accurate. Also that, as expected, the NW contribution decays with distance to the surface much more slowly than the SPP one. In both chosen examples, the SPP dominates right at the surface, but the NW takes over at a finite distance from it. However, at sufficiently large z the RO contribution always prevails. The inset to figure 5 shows, for Au at $\lambda = 540\,\text{nm}$, which of the

three 'asymptotic' terms (SPP, RO and NW) dominates in the X–Z-plane. This inset shows that the NW is the largest contribution over a 'stripe' close to the surface. As the region where the NW is dominant satisfies $z \ll x$, the full expression for the NW given by equation (8) can be approximated by (for $|\varepsilon_m| \gg 1$)

$$|G_{\mathrm{NW}}(x, z)| \simeq |G_{\mathrm{NW}}(x, 0)| \frac{1}{|1 + \sqrt{\varepsilon_{\mathrm{m}}} z/x|^3}. \tag{9}$$

Note that the algebraic decay of the NW with z reflects that this wave arises from the interference of its constituent components, and not from a pole in the angular spectrum $D(q)$. The comparison of the expression given by equation (9) with that of the RO contribution allows for an estimation of the distance to the surface at which the crossover between NW and RO occurs, z_{NW}. We obtain $z_{\mathrm{NW}} = |\sqrt{\varepsilon_{\mathrm{m}}}|/|1 + \sqrt{\varepsilon_{\mathrm{m}}} z/x|^3$ which, if $z_{\mathrm{NW}}/x \ll 1/|\sqrt{\varepsilon_m}|$, implies $z_{\mathrm{NW}} \approx |\sqrt{\varepsilon_{\mathrm{m}}}|$ or $Z_{\mathrm{NW}} \approx \lambda|\sqrt{\varepsilon_{\mathrm{m}}}|/(2\pi)$.

To summarize, we have shown that, in the asymptotic limit of long distances to the source, the SPPs are not the main channel for EM fields at the metal surface. Instead, after a few SPP absorption lengths, NWs take over. This occurs for any metal and any frequency range. NWs decay much more slowly than SPPs both along the surface (as $x^{-3/2}$ for 2D dipoles) and along the perpendicular direction.

Acknowledgments

We acknowledge support from the Spanish Ministry of Science and Innovation under projects MAT2008-06609-C02 and CSD2007-046-Nanolight.es.

Note added in proof. During the review process of this manuscript, the $x^{-3/2}$ asymptotic behavior of the fields *at the metal surface* was reported by Dai and Soukoulis [26] and Lalanne *et al* [27].

Appendix A. Field in the vacuum half-space using Green's dyadic

Consider a p-polarized EM wave (with wavelength λ and wavevector $g = 2\pi/\lambda$) incident onto a metal film with a subwavelength slit. The metal film is optically thick and extends from $Z = -W$ (where the EM field impinges) to $Z = 0$ (the exit side). The dielectric constant of the metal is ε_{m}. The slit has width A and we set the origin of the x-axis at the center of the slit.

According to the Lippmann–Shwinger integral equation [24], the electric field at any point at exit side of the film ($z > 0$) is given by the following integral relation:

$$\mathbf{E}(\mathbf{R}) = \mathbf{E}_0(\mathbf{R}) + g^2 \int_V \mathrm{d}\mathbf{R}'\, \Delta\varepsilon(\mathbf{R}')\, \hat{G}^E(\mathbf{R}, \mathbf{R}')\, \mathbf{E}(\mathbf{R}'), \tag{A.1}$$

where $\mathbf{E}_0(\mathbf{R})$ is the solution without the slit, $\Delta\varepsilon(\mathbf{R}) = 1 - \varepsilon_{\mathrm{m}}$ in the volume occupied by the slit, V, and zero everywhere else.

In the case of an optically thick film, the field $\mathbf{E}_0(\mathbf{R})$ can be neglected at the exit side and the dyadic $\hat{G}^E(\mathbf{R}, \mathbf{R}')$ can be approximated by the one corresponding to a single metal–vacuum interface. In order to obtain the magnetic field from equation (A.1), we use the Maxwell equation $\mathbf{H} = (-\mathrm{i}/g)\nabla_{\mathbf{R}} \times \mathbf{E}$ and arrive at

$$\mathbf{H}(\mathbf{r}) = \int_V \mathrm{d}\mathbf{r}'\, \Delta\varepsilon(\mathbf{r}')\, \hat{G}^H(\mathbf{r}, \mathbf{r}')\, \mathbf{E}(\mathbf{r}'), \tag{A.2}$$

where we have passed to dimensionless distances $\mathbf{r} = g\mathbf{R}$. The dyadic $\hat{G}^H$ connects the magnetic field outside of the slit with the electric field inside the slit. For the considered 2D geometry, where only p-polarized waves are involved, the magnetic field $\mathbf{H}$ points along the y-direction. Assuming that the electric field inside the slit mainly points along the x-direction, only the yz element of the dyadic $\hat{G}^H$ needs to be computed. We denote this element by $\hat{G}^H_{yx}(x, z; x', z')$. Following [25], we find that

$$\hat{G}^H_{yx}(x, z; x', z') = \frac{\mathrm{i}}{2\pi} \int \mathrm{d}q \frac{q_{zm}}{\varepsilon_\mathrm{m} q_z + q_{zm}} \mathrm{e}^{\mathrm{i}q(x-x')+\mathrm{i}q_z z - \mathrm{i}q_{zm} z'} \tag{A.3}$$

where q is the dimensionless x-component of the wavevector, $q = k_x/g$, $q_z = \sqrt{1-q^2}$ and $q_z = \sqrt{\epsilon_\mathrm{m} - q^2}$.

The integrand contains the exponential factor $\mathrm{e}^{-\mathrm{i}q_{zm}z'}$, which decays at the distance of a skin depth $\delta = 1/\mathrm{Im}(q_{zm})$, which is of the order of a few tens of nm in the optical regime. Therefore, the integration limits in z' can be extended to $[-\infty, 0]$. Moreover, the variation of Green's dyadic is much faster than that of the electric field inside the slit, hence the electric field inside the slit can be approximated by its value at the distance $z = -\delta$ (this is obtained as the average distance to the surface, weighted by the exponential decay of the field). An additional advantage of using the field at a short distance inside the slit is that the numerical problems related to the treatment of corners are eliminated.

The integration over z' can be performed in the following way:

$$\int_\mathrm{slit} \mathrm{e}^{-\mathrm{i}q_{zm}z'} \mathrm{d}z' \simeq \int_{-\infty}^{0} \mathrm{e}^{-\mathrm{i}q_{zm}z'} \mathrm{d}z' = \frac{\mathrm{i}}{q_{zm}}. \tag{A.4}$$

We then obtain

$$H_y(\mathbf{r}) = (\varepsilon_\mathrm{m} - 1) \int_{-a/2}^{a/2} \mathrm{d}x' G_\mathrm{slit}(x - x', z) E_{x'}(x', z = -\delta), \tag{A.5}$$

with

$$G_\mathrm{slit}(x, z) = \frac{1}{2\pi} \int \mathrm{d}q \frac{1}{\varepsilon_\mathrm{m} q_z + q_{zm}} \mathrm{e}^{\mathrm{i}qx + \mathrm{i}q_z z}. \tag{A.6}$$

Since we are interested in the field close to the surface and away from the source (which, as we will show, arises from the region in the angular spectrum close to $q = 1$) and for optical frequencies (where ε_m is large), this result can be related to the field radiated by a dipolar source $G(x, y)$ at the metal surface:

$$H_y(x, z) \simeq \sqrt{\varepsilon_\mathrm{m} - 1} \int_{-a/2}^{a/2} \mathrm{d}x' G(x - x', z) E_{x'}(x', z = -\delta), \tag{A.7}$$

where

$$G(x, z) = \frac{1}{2\pi} \int \mathrm{d}q \frac{q_{zm}}{\varepsilon_\mathrm{m} q_z + q_{zm}} \mathrm{e}^{\mathrm{i}qx + \mathrm{i}q_z z}. \tag{A.8}$$

The relation between equation (A.5) and equation (A.7) can be easily seen by noting that $q_{zm} \approx \sqrt{\varepsilon_m - 1}$ in the region of interest.

While equation (A.7) involves an additional approximation, we have preferred to work with Green's dyadic for a dipole source (rather than with G_slit) due to its wider applicability to other problems, involving metal surfaces with, for instance, inlaid metals, dielectric or metal

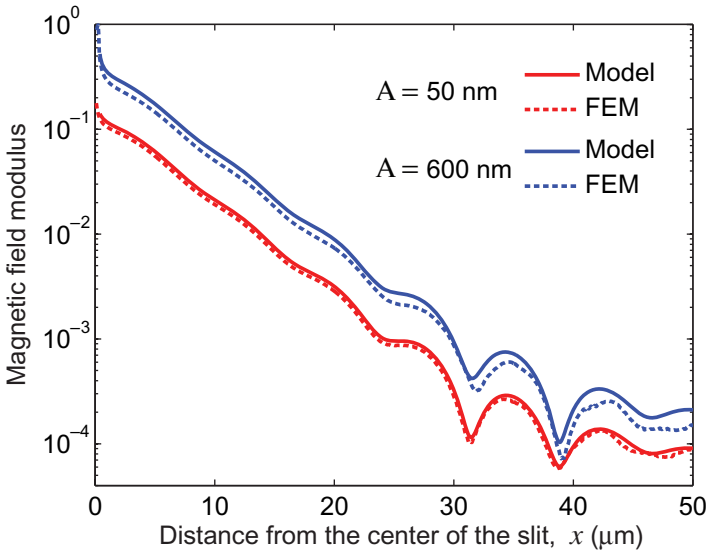

Figure A1. Comparison between the FEM and the analytic computation for the magnetic field on the gold–vacuum interface emerging from a slit. The slit widths considered are 50 and 600 nm, and the wavelength is 567 nm. The dashed lines are for the FEM calculations, while the continuous lines correspond to the approximate result using equation (A.7).

protrusions, or lines of fluorescent molecules. In any case, the differences for the slit case are minimal and the methods described in this paper could be straightforwardly applied to G_{slit}.

In order to validate the expression (A.7), we first performed full-vectorial computations using the finite element method (FEM) of the magnetic field emerging from slits with thicknesses 50 and 600 nm. The wavelength was chosen to be 576 nm and the metal of the film is gold. Then we computed the integral given by equation (A.7) extracting the electric field inside the slit at $z = -\delta$ from the FEM calculations. The function $G(x, z)$ was substituted by its asymptotic value, see appendix B. The comparison is shown in figure A1.

These results clearly show that equation (A.7) is very accurate for subwavelength slits ($A \ll \lambda$) and even provide a good approximation for $A \sim \lambda$.

Appendix B. Asymptotic behavior of the field: a steepest descent method

Consider Green's function for a dipole placed at the metal surface given by equation (A.8). In this section, we sketch the asymptotic analysis performed, which has been done following the general method described in [23] for treating Sommerfeld integrals.

In this method, the integrand is first prolonged into the complex q-plane. Subsequently, the following changes of variable are performed $q = \sin\phi$ and $s = \sqrt{2}\mathrm{e}^{\mathrm{i}(\pi/4)} \sin((\phi - \theta)/2)$. In polar coordinates ($x = r\sin\theta$ and $y = r\cos\theta$), the integral takes the following form:

$$G = \mathrm{e}^{\mathrm{i}r} \int_C \mathrm{d}s\, \Phi(s) \mathrm{e}^{-rs^2}, \tag{B.1}$$

with

$$\Phi(s) = \frac{1}{2\pi} \frac{\sqrt{2}\mathrm{e}^{-\mathrm{i}\pi/4}}{\cos[\frac{\phi(s)-\theta}{2}]} \frac{\cos[\phi(s)]\sqrt{\epsilon_m - \sin^2[\phi(s)]}}{\epsilon_m \cos[\phi(s)] + \sqrt{\epsilon_m - \sin^2[\phi(s)]}}, \tag{B.2}$$

where the contour C in the s-plane corresponds to the real axis in the initial plane q.

Then the integrand is separated into singular and non-singular parts

$$\Phi(s) = \frac{C_\mathrm{p}}{s - s_\mathrm{p}} + \Phi_0(s), \tag{B.3}$$

with

$$\Phi_0(s) = \frac{\Phi(s)(s - s_\mathrm{p}) - C_\mathrm{p}}{s - s_\mathrm{p}}, \tag{B.4}$$

where C_p is the residue given by $C_\mathrm{p} = \frac{\epsilon\sqrt{\epsilon}}{2\pi(\epsilon^2-1)\sqrt{1+\epsilon}}$, and the position of the pole in the complex plane s is $s_\mathrm{p} = \sqrt{2}e^{\mathrm{i}\pi/4} \sin\big((\phi_\mathrm{p} - \theta)/2\big)$, where $\phi_\mathrm{p} = \arccos\Big(-1/\sqrt{\epsilon+1}\Big)$. Then, after deforming the integral path to the steepest descent one (real axis in the plane s), the singular part yields the complementary error function $\mathrm{i}\pi C_\mathrm{p}\mathrm{e}^{r(\mathrm{i}-s_\mathrm{p}^2)}\mathrm{erfc}(-\mathrm{i}s_\mathrm{p}\sqrt{r})$ with the argument being the square root of the 'numerical distance' introduced by Sommerfeld. The non-singular part of the integral can be expanded in the Taylor series close to the saddle point $s = 0$ providing the infinite sum of the integrals of the Gaussian type. The final result reads

$$G = \mathrm{i}\pi C_\mathrm{p}\mathrm{e}^{\mathrm{i}r\mathbf{q}_\mathrm{p}}\mathrm{erfc}(-\mathrm{i}s_\mathrm{p}\sqrt{r}) + \mathrm{e}^{\mathrm{i}r} \sum_{n\in\mathrm{even}} \frac{\Gamma(\frac{n+1}{2})}{n! r^{(n+1)/2}} \frac{\mathrm{d}^n \Phi_0}{\mathrm{d}s^n}|_{s=0}. \tag{B.5}$$

Ung and Sheng [16] reported a similar expression for G, but containing only the error-function term and the $n = 0$ term in the sum. That expression is correct up to order $r^{-1/2}$ but at the surface, the term going as $x^{-1/2}$ vanishes. In order to obtain the correct $r^{-3/2}$ behavior (and therefore, to represent properly the NW) both the error function and the $n = 2$ term in the sum must be retained.

The function Φ_0 is composed of two parts: Φ and $C_\mathrm{p}/(s_\mathrm{p} - s)$. Note that the part of the sum in (B.5) coming from the term $C_\mathrm{p}/(s_\mathrm{p} - s)$ coincides, up to a sign, with the asymptotic expansion of the complementary error function for large arguments without the first term (the first term coincides with the residue contribution into the integral). This expansion reads

$$\mathrm{erfc}(-\mathrm{i}s_\mathrm{p}\sqrt{r}) = 2\Theta(\theta - \theta_\mathrm{p}) + \mathrm{i}\frac{\mathrm{e}^{r s_\mathrm{p}^2}}{\pi} \sum_{n\in\mathrm{even}} \frac{\Gamma(\frac{n+1}{2})}{(s_\mathrm{p}^2 r)^{(n+1)/2}}, \tag{B.6}$$

where Θ is the Heaviside step function and θ_p is the angle defining the diffraction shadow, $\theta_\mathrm{p} = \mathrm{Re}(\phi_\mathrm{p}) - \arccos(1/\cosh[\mathrm{Im}(\phi_p)])$. This is the critical angle such that for $\theta > \theta_\mathrm{p}$ we have $\mathrm{Im}(s_\mathrm{p}) < 0$ and the initial transformation of the integration path into the steepest-decent one leads to the crossing of the pole, so that the residue must be taken into account (see figures B1 and B2). An example of the critical angle limiting the diffraction shadow is represented in figure 1 by dashed lines.

Retaining the residue contribution, G_SPP, and the first two terms G_RO (proportional to $r^{-1/2}$) and G_NW (proportional to $r^{-3/2}$) in the sum coming from Φ we obtain an expression that is correct in the far-field, up to terms $\Delta G = O(r^{-5/2})$. The result can be rewritten as

$$G = G_\mathrm{SPP} + G_\mathrm{RO} + G_\mathrm{NW} + \Delta G. \tag{B.7}$$

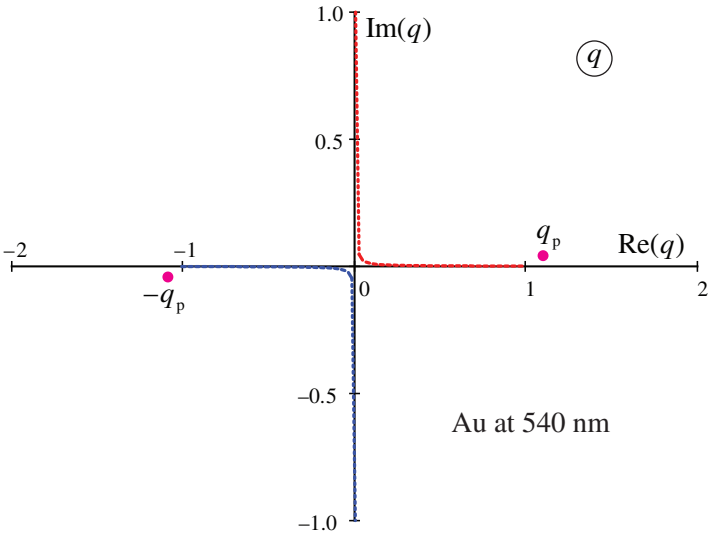

Figure B1. The poles $q = \pm q_p$ and the branch cuts $\mathrm{Im}(q_z) = 0$ with infinitesimally small absorption of the vacuum in the complex plane q. The poles correspond to the dots. For $x > 0$, only the pole $q = q_p$ contributes to the result.

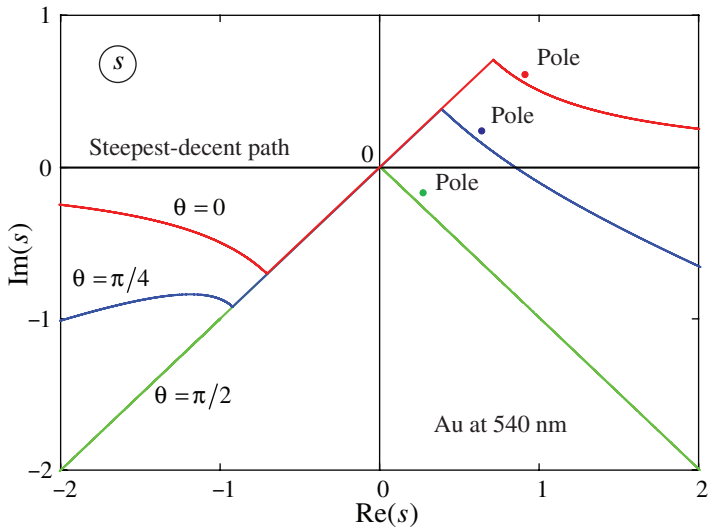

Figure B2. The contours of the integration in the complex plane s for different angles θ. The positions of the poles s_p are denoted by dots. When the pole is crossed during the transformation of the initial path to the steepest-decent one, the residue must be taken into account.

The plasmonic term is

$$G_{\mathrm{SPP}} = 2\pi \mathrm{i} C_p \mathrm{e}^{\mathrm{i}\mathbf{r}\mathbf{q}_p}, \tag{B.8}$$

the 'RO' term is

$$G_{\mathrm{RO}} = \frac{\mathrm{e}^{\mathrm{i}r - \mathrm{i}\pi/4}}{\sqrt{2\pi r}} \frac{\cos\theta\sqrt{\epsilon_m - \sin^2\theta}}{\epsilon_m \cos\theta + \sqrt{\epsilon_m - \sin^2\theta}}, \tag{B.9}$$

the NW is given by

$$G_{\mathrm{NW}} = \frac{1}{2}\frac{\mathrm{e}^{\mathrm{i}r-\mathrm{i}(3\pi/4)}}{r\sqrt{2\pi r}}\frac{\mathrm{d}^2}{\mathrm{d}\phi^2}\left[\frac{1}{\cos(\frac{\phi-\theta}{2})}\frac{\cos\phi\sqrt{\epsilon-\sin^2\phi}}{\epsilon\cos\phi+\sqrt{\epsilon-\sin^2\phi}}\right]_{\phi=\theta}, \tag{B.10}$$

and the rest is

$$\begin{aligned}\Delta G &= \mathrm{i}\pi C_{\mathrm{p}}\mathrm{e}^{r(\mathrm{i}-s_{\mathrm{p}}^2)}[\mathrm{erfc}(-\mathrm{i}s_{\mathrm{p}}\sqrt{r})-2] \\ &\quad + \mathrm{e}^{\mathrm{i}r}\sum_{n\in\mathrm{even}}\frac{\Gamma(\frac{1+n}{2})}{n!r^{(1+n)/2}}\left(\sigma_n\frac{\mathrm{d}^n\Phi(s)}{\mathrm{d}s^n}-\frac{\mathrm{d}^n}{\mathrm{d}s^n}\frac{C_{\mathrm{p}}}{s-s_{\mathrm{p}}}\right)_{s=0}, \\ \sigma_n &= \begin{cases}1, & n\geqslant 4,\\ 0, & n<4.\end{cases}\end{aligned} \tag{B.11}$$

References

[1] Barnes W L, Dereux A and Ebbesen T W 2003 Surface plasmon subwavelength optics *Nature* **424** 824–30
[2] Maier S A 2007 *Plasmonics: Fundamentals and Applications* (Berlin: Springer)
[3] Krenn J R *et al* 1999 Direct observation of localized surface plasmon coupling *Phys. Rev.* B **60** 5029
[4] Genet C and Ebbesen T W 2007 Light in tiny holes *Nature* **445** 7123
[5] Zenneck J 1907 Über die fortpflanzung ebener elektromagnetischer wellen mngs einer ebenen leiterflache und ihre beziehung zur drahtlosen telegraphie *Ann. Phys.* **23** 846
[6] Sommerfeld A N 1909 Über die ausbreltung der wellen in der drahtlosen telegraphie *Ann. Phys.* **28** 665–736
[7] Norton K A 1936 The propagation of radio waves over the surface of the earth and in the upper atmosphere *Proc. IRE* **24** 1367–87
[8] Collin R E 2004 Hertzian dipole radiation over a lossy earth or sea: some early and late 20th century controversies *IEEE Antennas Propag. Mag.* **46** 64–79
[9] Ebbesen T W, Genet C and Bozhevolnyi S I 2008 Surface-plasmon circuitry *Phys. Today* **61** 44
[10] López-Tejeira F, García-Vidal F J and Martín-Moreno L 2005 Scattering of surface plasmons by one-dimensional periodic nanoindented surfaces *Phys. Rev.* B **72** 161405
[11] Lalanne P and Hugonin J P 2006 Interaction between optical nano-objects at metallo-dielectric interfaces *Nat. Phys.* **2** 551–6
[12] Lalanne P, Hugonin J P and Rodier J C 2005 Theory of surface plasmon generation at nanoslit apertures *Phys. Rev. Lett.* **95** 263902
[13] Schouten H F *et al* 2005 Plasmon-assisted two-slit transmission: Young's experiment revisited *Phys. Rev. Lett.* **94** 053901
[14] Leveque G, Martin O J F and Weiner J 2007 Transient behavior of surface plasmon polaritons scattered at a subwavelength groove *Phys. Rev.* B **76** 155418
[15] Besbes M *et al* 2007 Numerical analysis of a slit-groove diffraction problem *J. Eur. Opt. Soc.* **2** 07022
[16] Ung B and Sheng Y 2008 Optical surface waves over metallo-dielectric nanostructures: Sommerfeld integrals revisited *Opt. Express* **16** 9073–86
[17] Gay G *et al* 2006 The optical response of nanostructured surfaces and the composite diffracted evanescent wave model *Nat. Phys.* **2** 262
[18] López-Tejeira F *et al* 2007 Efficient unidirectional nanoslit couplers for surface plasmons *Nat. Phys.* **3** 324
[19] Aigouy L *et al* 2007 Near-field analysis of surface waves launched at nanoslit apertures *Phys. Rev. Lett.* **98** 153902
[20] Anikeyev V, Temnov V V, Woggon U, Devaux E and Ebbesen T W 2008 Propagation oscillations in the near-field response of traveling surface waves launched by metallic nanoapertures *Appl. Phys.* B **93** 171–6

[21] Laluet J Y, Drezet A, Genet C and Ebbesen T W 2008 Generation of surface plasmons at single subwavelength slits: from slit to ridge plasmon *New J. Phys.* **10** 105014
[22] Rodrigo S G, García-Vidal F J and Martín-Moreno L 2008 Influence of material properties on extraordinary optical transmission through hole arrays *Phys. Rev.* B **77** 075401
[23] Felsen L P and Marcuvitz N 1994 *Radiation and Scattering of Waves* (Piscataway, NJ: IEEE) chapter 4
[24] Martin O J F, Girard C and Dereux A 1995 Generalized field propagator for electromagnetic scattering and light confinement *Phys. Rev. Lett.* **74** 526
[25] Novotny L and Hecht B 2006 *Principles of Nano-Optics* (New York: Cambridge University Press)
[26] Dai W and Soukoulis C M 2009 *Phys. Rev.* B **80** 155407
[27] Lalanne P *et al* 2009 *Surf. Sci. Rep.* **64** 453