STRUCTURED SURFACES AS OPTICAL METAMATERIALS

Edited by

ALEXEI A. MARADUDIN
University of California, Irvine
1

Physics of extraordinary transmission through subwavelength hole arrays

EVGENY POPOV AND NICOLAS BONOD

1.1 A brief reminder of the history of grating anomalies and plasmon surface waves

The recent history of the research and development around plasmon surface waves that was initiated by the work published in *Nature* in 1998 by Ebbesen *et al.* [1] looks like a ten-fold compressed version of studies initiated more than a century ago by Robert Wood with his discovery of anomalies in the efficiency of metallic diffraction gratings, now known as Wood's anomalies [2]. In 1902, R. Wood wrote: “I was astounded to find that under certain conditions, the drop from maximum illumination to minimum, a drop certainly from 10 to 1, occurred within a range of wavelengths not greater than the distance between the sodium lines,” an observation that marked the discovery of grating anomalies.

The first period of the search for their explanation is marked by the attempt of Lord Rayleigh [3, 4] to link Wood’s anomalies to the redistribution of the energy due to the passing-off (cut-off) of higher diffraction orders of the grating (transfer from propagating into evanescent type). As pointed out by Maystre [5], his prediction was all the more remarkable as the author first ignored the groove frequency of the grating used by Wood, and thus could not verify this assumption with experimental data.

It took more than 30 years for the second period of experimental and theoretical studies to establish another explanation of Wood’s anomalies. In 1941, Fano [6] was the first to distinguish between two types of anomaly: (i) an edge anomaly, with a sharp behavior connected with the passing-off of a higher diffraction order, and (ii) an anomaly, generally consisting of a minimum and a maximum in the efficiency, which appears in a much broader interval. The second type of anomaly was described by Fano as a resonance one, linked with the excitation of a guided (leaky) wave along the grating surface. Hessel and Oliner [7] published a

pioneering paper that shows for the first time, using a theory based on an analysis of electromagnetic scattering from a generic model of a periodic structure yielding a simple closed form solution, that Wood's anomalies are of two types: one due to branch point singularities that correspond physically to the onset of a new propagating spectral order (first indicated by Lord Rayleigh), and the other due to pole singularities that correspond to the condition of resonance for leaky surface waves guided by the structure. In addition, Hessel and Oliner developed a so-called phenomenological approach to the resonant anomalies that permitted describing the anomaly by a very small number of physical parameters: a pole of the scattering matrix, a zero of the diffracted amplitudes, and smoothly varying coefficients.

The third period that continues even today contains studies in three different directions. First are the grating manufacturers and instrumental optics users for whom it is strongly advisable to avoid anomalies because of their devastating effects for spectral instrument performance. Second, the excitation of surface plasmons can lead to a total absorption of the incident light, a phenomenon predicted and observed by Hutley and Maystre [8] in a single polarization for classical gratings with one-dimensional (1D) periodicity, with the magnetic field vector parallel to the groove direction (TM, transverse magnetic, polarization). By the use of crossed gratings with two-dimensional periodicity, it is possible to obtain total light absorption in unpolarized light [9, 10]. In both cases, it was necessary to use gratings with subwavelength periods that can support only the specular (zeroth) reflected order.

The total light absorption by metallic gratings evidenced that the coupling between the incident light and metals can be strongly enhanced by the excitation of surface plasmons, and this effect opened the way to many applications based on the strongly enhanced light–matter interaction. Third, as the electromagnetic field is localized in the vicinity of the metallic surface, light absorption leads to very strong optical intensities at the surface. As any surface presents natural roughness that can excite the surface wave, the field enhancement obtained under specific conditions was sufficient to provide a proper physical explanation of the surface enhanced Raman scattering (SERS) effect [11]. The same effect is also used to enhance otherwise weak nonlinear phenomena [12]. Biosensors based on surface plasmons are highly dependent on the refractive index of the surrounding medium. Binding or adsorption of molecules onto the metallic surface induces a change of the local refractive index of the dielectric medium, so that such biosensors can be called refractometric sensors [13–17].

### 1.2 Generalities of the surface waves on a single interface

Before discussing in detail the historical development of the studies of enhanced light transmission through arrays of holes in a metallic screen, let us introduce several notations and basic principles. Let us consider a plane metal–dielectric interface in the $xz$ plane that separates two nonmagnetic media with relative dielectric permittivities $\varepsilon_1$ and $\varepsilon_2$. In TM (transverse magnetic) polarization and incidence in the $xOy$ plane, the two components of the electric and magnetic field that are parallel to the interface, and thus continuous across it, are:

$$
\omega \mu H_z = \exp(ik_x x) \left[ \exp(-ik_{1y} y) + r \exp(ik_{1y} y) \right]
$$

$$
E_x = \frac{k_y}{k_0} \frac{1}{\varepsilon_1} \exp(ik_x x) \left[ \exp(-ik_{1y} y) - r \exp(ik_{1y} y) \right]
$$

(1.1)

in the cladding and

$$
\omega \mu H_z = t \exp(ik_x x) \exp(-ik_{2y} y)
$$

$$
E_x = \frac{k_y}{k_0} \frac{1}{\varepsilon_2} \exp(ik_x x) \exp(-ik_{2y} y)
$$

(1.2)

in the substrate. The first terms in the brackets in Eqs. (1.1) correspond to the incident wave, and the second terms correspond to the reflected wave, with $r$ the reflection coefficient for the magnetic field amplitude. The transmission coefficient is denoted by $t$. Note that $k_x$ is the $x$-component of the incident wavevector, and that $k_{1y}$ and $k_{2y}$ are the $y$-components of the wavevectors in the cladding and in the substrate:

$$
k_{jy} = \sqrt{k_0^2 \varepsilon_j - k_x^2}, \quad j = 1, 2,
$$

(1.3)

with $k_0$ the free-space wavenumber. The Fresnel reflection coefficients depend on the polarization and have the following form for transverse electric (TE) polarization:

$$
r_{TE} = \frac{k_{1y} - k_{2y}}{k_{1y} + k_{2y}}
$$

(1.4)

and for TM polarization:

$$
r_{TM} = \frac{k_{1y} / \varepsilon_1 - k_{2y} / \varepsilon_2}{k_{1y} / \varepsilon_1 + k_{2y} / \varepsilon_2}
$$

(1.5)

It is well-known that $r_{TE}$ has neither a pole nor a zero. In contrast, when both $\varepsilon_1$ and $\varepsilon_2$ are real and positive, there is a zero of $r_{TM}$ called the Brewster effect. There also exists a pole (a zero of the denominator) if one of the media is a dielectric and the other a metal, a pole that corresponds to a surface wave that can propagate along the interface. When expressed in terms of the wavevector component parallel to the interface, the solution has the same form for the Brewster effect and the pole,

$$
k_y = k_0 \sqrt{\varepsilon_1 \varepsilon_2} / (\varepsilon_1 + \varepsilon_2),
$$

(1.6)
due to the ambiguity of the choice in Eq. (1.3) of the sign of the square root for complex arguments. Indeed, combining Eqs. (1.3) and (1.6), we obtain the classical form of the Brewster angle in the incident medium: \( \tan \theta_i = k_x / k_\parallel = \sqrt{\varepsilon_2 / \varepsilon_1} \).

When the second medium is a metal with the real part of \( \varepsilon_2 \) negative and smaller than \(-1\), the real part of \( k_\parallel \) in Eq. (1.6) is greater than the wavenumber \( k_\parallel \sqrt{\varepsilon_1} \) in the upper medium; i.e., the wave is evanescent in the cladding (and inside the metal), with increasing distance from the interface, representing a surface wave with a propagation constant equal to \( k_\parallel \), a solution that we shall note as \( k_g \), the index \( g \) standing for "guided," and its normalized propagation constant will be denoted as \( \sigma_g = k_g / k_0 \). As \( \sigma_g \) always has a non-zero imaginary part due to the absorption losses in the metal, the surface wave decays as it propagates. Quite often the negative permittivity is due to the collective oscillations of the free electron plasma in the metal, which gives the names surface plasmon or plasmon surface wave (PSW) to these surface waves. As an incident electric field creates polarization states of the plasma, some authors call this wave a surface plasmon polariton (SPP). In the case of a polar crystal/vacuum interface, the corresponding surface waves represent surface phonon polaritons. They all have common properties from an electromagnetic point of view, although the background solid state physics can be quite different. As they represent a zero of the denominator of \( r_{TM} \), they are solutions of the homogeneous problem—a scattered field with zero incident field—and thus represent proper (eigen) modes of the system.

When considering an idealized presentation of a perfectly conducting metal with \( \varepsilon_2 \to -\infty \), the propagation constant in Eq. (1.6) becomes equal to the vacuum wavenumber, and thus the solution represents a plane wave propagating parallel to the interface inside the cladding, with its electric field vector perpendicular to the surface; i.e., the solution is not localized to the surface.

When \( \sigma_g \) is greater than \( n_1 \), such a wave cannot be excited with an incident plane propagating wave. The excitation of the surface wave is possible through the Kretschmann configuration [13]: the surface plasmon is excited on the lower surface of a metallic layer having on its upper surface a prism with refractive index higher than the index of the substrate in order to match the horizontal component of the incident wavevector to the real part of the PSW wavenumber on the lower interface. The surface plasmon is then coupled to the incident light by tunneling through the metallic layer. A surface plasmon can also be excited in a prism coupler in the Otto configuration. In that case, the metallic film is coated on a glass substrate. The strength of excitation depends on the distance between the prism and the metallic layer. In both cases, the reflection of light is strongly attenuated when the surface plasmon is coupled to the incident wave, and the angle of incidence where the absorption is maximum depends on the refractive index of the dielectric medium surrounding the metallic layer.

Much more efficient coupling occurs when gratings are used with a periodicity that serves as a generator of wavevectors parallel to the surface. An incident plane wave generates an infinite number of diffraction orders. In the case of 1D periodicity in the \( x \)-direction with period \( d \), the wavevector component of each diffraction order \( m \) is given by the grating equation:

\[
 k_{1x,m} = k_{1x,0} + m \frac{2\pi}{d}, \tag{1.7}
\]

where \( k_{1x,0} \) is equal to the \( x \)-component of the incident wavevector and \( m \) is an integer. If, for a certain value of \( m \), \( k_{1x,m} \) is close to \( k_g \), a surface wave can be excited. A simplified notation leads to the condition of excitation:

\[
 \text{Re}(\sigma_g) = \sin \theta_i + \frac{m^2}{d} \tag{1.8}
\]

if the upper interface is air (more precisely, a vacuum).

The coupling of the incident wave to the surface wave (mode) is reciprocal; i.e., the surface wave can be radiated into propagating diffraction orders in the cladding following Eq. (1.7). This phenomenon is called leakage and the surface wave becomes a leaky one, which leads to an increase of the imaginary part of \( \sigma_g \).

Another important feature that is not obvious from Eq. (1.8) is that the real part of the propagation constant, as well as the electromagnetic field distribution of the surface wave characteristics, are modified by the presence of surface corrugation. Another possibility to excite a PSW realizes itself in SERS, where the surface roughness scatters the incident plane wave into waves with different \( k_x \), and, in particular, with \( k_x > k_0 \), with part of the incident energy coupled to the PSW.

### 1.3 Extraordinary transmission and its first explanations

Just as Robert Wood 95 years earlier was astounded by his experimental observation, Thomas Ebbesen and his collaborators found it quite surprising to observe that when light tries to pass through an array of holes of subwavelength dimensions in an optically thick (opaque) metallic sheet (Fig. 1.1(a)), and whose entire area is much smaller than the total illuminated surface, there are spectral regions with anomalously high transmission (Fig. 1.1(b)) compared with the predictions of classical diffraction theory. The surprise was so great that it prevented the publication of the results for almost ten years from their first observation [18] of the effect in the NEC laboratories.

As with Wood’s anomalies, it is possible to separate the studies on this extraordinary transmission into three much shorter and more dynamic periods. In 1998, in contrast to the situation at the start of the twentieth century, electromagnetic theories
of gratings (1D or 2D) were largely developed. The understanding of the role of the PSW (and surface waves in general) in grating anomalies, field enhancement, etc., was much deeper, and the number of scientists working in the field incomparably larger. The end of the Cold War moved large human resources from defense microelectronics, solid state and high energy physics into optics, creating neologisms like photonic crystals, photonics, metamaterials, etc., causing, for instance, the change of *Optics News* into *Optics and Photonic News*. In addition, production resources such as optical photolithography, focused ion-beam and laser-beam writing and etching, became more available in optics laboratories, which permitted constructing metals and dielectric media at the nanometer scale and developing photonic devices able to control light at the subwavelength scale. Already in the original paper [1], the authors clearly indicated that the transmission enhancement appears at spectral positions closely given by PSW excitation by a bi-periodic structure, but they were not satisfied by this qualitative explication. The publication of these results by Ebbesen and coworkers in *Nature* immediately strongly impacted the newly enlarged optical community that was closely interested in photonics, to find a new but similar interest in plasmon surface waves, giving birth to another neologism, plasmonics.

The first numerical results were so close to the experimental enhancement that doubts were generated whether the transmission increase was so extraordinary. The main characteristic of this first period (see, for example, refs. [19] and [20]) was the use by theoreticians of gratings with 1D periodicity made of periodic slits in a metallic screen (Fig. 1.2(a)). The results were quite nice, with a clearly visible flow of the electromagnetic field inside the slits, so that some authors started to indicate the decisive role of another wave — a vertical plasmon wave that propagates inside the slits of the metal–dielectric interface — that is responsible for the enhanced transmission, acting simultaneously with the grating-induced resonances of the horizontally propagating PSW, excited at spectral positions given by Eq. (1.8). Figure 1.2(b) represents such a vertical wave which propagates inside the slits as in free space for perfectly conducting walls. For metals with a finite conductivity, this wave represents a hybrid wave that can propagate in the vertical direction formed by two coupled plasmons on the slit walls [21].
The finite conductivity of the metal changes the boundary conditions on the vertical walls, but even for very narrow slits and a silver grating, the TM electromagnetic field preserves its propagating nature, as seen in Fig. 1.2(c).

These idyllic conclusions were common for the first period of about two to three years following 1998. The main problem was that they were not applicable to the geometry involved in the initial experiment made by Ebbesen et al., where the slits were replaced by small holes. In fact, the enhanced transmission through slit metallic gratings (or gratings having similar grooves) in TM polarization had been known for quite a long time and resulted in commercially available wire-grating polarizers (see, for example, [10]). Such gratings (as represented in Fig. 1.2(a)) with subwavelength periods small enough to support just the specular reflected and transmitted orders, reflect incident light of TE polarization almost completely, while light of TM polarization can be transmitted almost totally for a proper choice of grating parameters. The reason lies in the existence of a waveguide mode inside each slit, which in TM polarization has no cut-off wavelength. Let us consider a slit, neglecting the absorption losses inside the metal walls. As in the case of the plane horizontal interface between a lossless metal and a dielectric in TM polarization, a vertical interface also supports a wave of plane-wave type inside the dielectric propagating parallel to the interface with a magnetic field vector parallel to the interface (Fig. 1.2(b)). The same wave can propagate inside slits with lossless walls, as it satisfies the boundary conditions on both walls, whatever the width of the slit. The mode is characterized by a real propagation constant in the vertical direction (neglecting absorption losses, as assumed). On the upper and lower interfaces (Fig. 1.2(a)), the mode is reflected backwards in the slit, and is partially transferred into propagating waves in the cladding and in the substrate, representing a Fabry–Perot resonator. If the system is symmetrical (the same optical index of the substrate and the cladding, as in wire polarizers), the Fabry–Perot resonance maxima can reach 100% in transmission. In contrast, in TE polarization the corresponding slit mode has a cut-off, because the electric field vector is parallel to the slit walls. The electric field vanishes on both walls and satisfies the following relation (Fig. 1.3(a)):

$$E_z \sim \sin \left( \frac{\pi x}{w} \right),$$  \hspace{1cm} (1.9)

so that the $y$-component of its wavevector, given by

$$q_y = \sqrt{k_0^2 - \frac{\pi^2}{w^2}} = \pi \frac{4}{\lambda_l} - \frac{1}{w^2},$$ \hspace{1cm} (1.10)

becomes imaginary for small widths $w$; i.e., the mode is evanescent in the vertical $y$-direction if the slit width is smaller than $\lambda/2$. The smaller the width, the faster the exponential decrease of the mode amplitude inside the slit depth. The narrower the slit and the thicker the metal layer, the smaller the amount of transmitted energy, and thus the polarizing properties of the device.

When finite conductivity is taken into account, the cut-off width is slightly smaller for finitely conducting walls than the $\lambda/2$ value obtained from Eq. (1.10). In addition, very narrow slits absorb an electromagnetic field, and the transverse variation of the field becomes very weak, as can be observed in Fig. 1.3(b) for silver walls and a 40 nm slit width.

In contrast to what happens with slits, holes with a finite width in both directions of their cross-section do not support modes without cut-off; i.e., below a given width of the hole, the field of the modes inside is always evanescently decreasing. Although obvious, these considerations were not taken into account in the first modelizations, when the hole array was replaced by periodic slits. However, these
first works were relatively easily carried out from theoretical and computational points of view, and the large number of studies based on this assumption attracted substantial interest to metallic gratings and surface plasmons.

1.4. The role of the evanescent mode

The second period of the studies of extraordinary transmission started with the first rigorous electromagnetic modeling of the array of holes with finite subwavelength cross-section dimensions [22]. The numerical results are similar to the experimental observations (Fig. 1.4).

The grating period in both directions is the same and equal to 0.9 μm, the cladding is air, and the substrate is glass. The metal is silver 0.2 μm thick, and the holes have a square cross-section with a width of 0.25 μm, much below the cut-off dimensions for the spectral interval under study. Two peaks are clearly distinguished, the shorter-wavelength one, lying around 1 μm, corresponds to the excitation of PSW on the upper air–silver interface. The long-wavelength peak is due to the excitation of PSW on the lower glass–silver interface.

For an infinitely conducting metal, the fundamental TE mode of the hollow square waveguide formed inside each hole has a propagation constant of the order of \( q_y \approx 111 \mu m^{-1} \), which corresponds to a decay constant in the \( y \)-direction, \( \gamma = \text{Im}(q_y)/k_0 \approx 2.5 \). When the finite conductivity of the metal is taken into account, the electromagnetic field of the mode penetrates inside the hole walls, so that its propagation constant changes, and the decay constant becomes smaller and equal to \( \gamma \approx 2.1 \) at a wavelength of 1.4 μm, a value more than four times smaller than the imaginary part of the refractive index of the silver film, \( n_{Ag} = 0.1 + i \times 9.4 \), in the IR. When compared to the direct tunneling through the metal layer, with a decay constant given by \( \text{Im}(n_{Ag}) \), the decay of the fundamental mode, even though it has an exponentially decreasing amplitude, within the film thickness of 0.2 μm is 20 000 times smaller than without the perforations.

Numerical simulations can easily confirm these conclusions, because one can smoothly vary the layer thickness quite easily numerically. Figure 1.5 presents the transmission under the same conditions (\( \lambda = 1.38 \mu m \)) as a function of the metal layer thickness. As can be observed, the dependence for \( h < 0.6 \mu m \) is linear on a semi-logarithmic scale, and the slope close to 20 μm^{-1} corresponds quite well to 2q_y, because the intensity decreases as the square of the amplitude.

In parallel with the rigorous electromagnetic study of the enhanced transmission through hole arrays, several authors have presented approximate models that not only provide an easier physical understanding of the various interactions, but also in some cases predicted new phenomena. In the early 2000s, a simple model that we describe below was developed due to the understanding of the role of the evanescent waveguide mode inside the hole [24, 25]. Let us consider an interface between air and a perforated metal with infinite thickness, in order to eliminate the role of the lower interface. When the array period(s) is small enough to avoid propagation of higher diffracted orders in the cladding, and when the hole cross-section of
width \( w \) is small in comparison with the period and thus with the wavelength, the predominant fields in TM incident polarization are the incident and specularly reflected fields, as given in Eqs. (1.1). Inside the rectangular hole with perfectly conducting walls, the transverse electric fundamental mode is parallel to the \( x \)-axis, it does not depend on \( z \), and it must vanish at \( x = 0 \) and \( w \), as in Eq. (1.9), but \( \rho \)-perturbing the \( z \)-and \( x \)-axes, because the TE mode with electric field oriented in the \( z \)-direction cannot be excited with an incident electric field oriented in the \( x \)-\( y \) plane. If the amplitude of \( H_z \) is denoted by \( r \), the \( x \)- and \( z \)-components are given by

\[
\omega \mu_0 H_z = re^{-i\phi_3} \sin \left( \frac{\pi z}{w} \right),
\]

\[
E_x = \begin{cases} \frac{r}{q} e^{-i\phi_3} \sin \left( \frac{\pi z}{w} \right), & 0 \leq x, z \leq w \\ 0, & x, z \not\in [0, w]. \end{cases}
\] (1.11)

so that the boundary conditions require that the tangential electric field components vanish on the walls at \( x = 0 \) and \( w \), and at \( z = 0 \) and \( w \).

The requirement of the continuity of the tangential \((x \text{ and } z)\) components of the electromagnetic field at \( y = 0 \) links the system of Eqs. (1.1) and (1.11) and yields the two equations:

\[
e^{ik_x x} (1 + r) = \tau \sin \left( \frac{\pi z}{w} \right),
\]

\[
\frac{k_x}{k_0} e^{ik_x x} (1 - r) = \begin{cases} \frac{\tau}{q} \sin \left( \frac{\pi z}{w} \right), & x, z \in [0, w] \\ 0, & x, z \not\in [0, w]. \end{cases}
\] (1.13)

The first equation must be satisfied within the aperture opening, because the magnetic field inside the metal is not known. The basic modal functions of the modes in the apertures are \( \sin(m \pi z/w) \) and \( \cos(p \pi x/w) \), where \( m \) and \( p \) are integers. If we multiply Eq. (1.12) by \( \sin(\pi z/w) \) and integrate over the aperture with respect to \( z \) and \( x \), the result is as follows

\[
\frac{2}{\pi} (1 + r) \left( \sin \left( \frac{k_y w}{2} \right) \right) = \frac{\tau}{2}.
\] (1.14)

The second boundary condition must be valid within the entire cell, with basic modal functions in \( x \) and \( y \) being the exponential functions. By multiplying the equation by \( \exp(-i k_y x) \) and integrating from 0 to \( d \) in \( x \) and \( z \), we obtain a similar relation:

\[
\frac{k_x}{k_0} d^2 (1 - r) = \frac{\tau}{q} \frac{2 \alpha^2}{\pi} \left( \sin \left( \frac{k_y w}{2} \right) \right) \left( \frac{k_y w}{2} \right).
\] (1.15)

Figure 1.6. Reflection coefficient \( r \) from Eq. (1.16) as a function of the \( x \)-component of the incident wave wavevector \( k_x/k_0 \).

It is straightforward to obtain the reflection coefficient, which has a form similar to Eq. (1.5):

\[
r = \frac{k_x d y / k_0 - I^2}{k_x d y / k_0 + I^2},
\] (1.16)

where

\[
l = \frac{w}{d} \frac{2 \sqrt{2}}{\pi} \left( \sin \left( \frac{k_y w}{2} \right) \right) \left( \frac{k_y w}{2} \right)
\]

represents the coupling integral between the electromagnetic field in the cladding and in the apertures. Assuming that the aperture size is much smaller than the wavelength, the dependence of \( I \) on the angle of incidence can be neglected. In the lossless case and for evanescent modes (imaginary \( q_x \)), the modulus of \( r \) is equal to unity, as for a planar metal-air interface without losses.

The situation can change if we consider incidence with \( k_x > k_0 \), as is the case with the PSW. With \( k_x \) becoming imaginary, it is possible to find a zero of the denominator of \( r \) in Eq. (1.16) representing also a resonance of the mode amplitude inside the aperture, because the pole of the reflection coefficient is a pole in transmission too. It must be stressed that this resonance can be excited only with evanescent incident waves, as observed in Fig. 1.6, where the reflection coefficient \( r \) is plotted as a function of \( k_x/k_0 \), with \( d = 1 \, \mu m \), \( w = 0.2 \, \mu m \), wavelength \( \lambda = 1.3 \, \mu m \), and the refractive index inside the aperture \( n = 1 \). A sharp peak can be observed for an evanescent incident wave. The periodicity of the array can provide such waves through the grating equation, and a PSW excited by the grating will
serve as a strong source to excite the evanescent mode resonantly through the zero of the denominator of Eq. (1.16).

Let us see, from the example provided by the original paper [11], how this could happen. If the ratio between the wavelength and the hole side dimensions is approximately equal to $1.4/0.2 = 7$, the value $q_y/k_0$ is equal to $1.35$, as given by Eq. (1.10). The normalized propagation constant of a PSW on a highly conducting surface is slightly greater than unity, $k_x/k_0 \simeq 1.01$, whereas $k_y/k_0 \simeq 0.01$. The product $k_y q_y / k_0^2 \simeq -0.034$. With $\alpha/d = 0.2$, the value of $I^2$ in Eq. (1.16) is close to 0.032, so that $r = 33$, indicating the existence of a resonance.

In the case of a single aperture, light can be coupled to a localized surface plasmon, as discussed in Section 1.6, that is able to excite further a waveguide mode. However, this effect is much weaker than when using the periodicity of the hole array.

Of course, the model discussed here is quite simplified. The interaction between the PSW and the evanescent mode changes the PSW propagation constant, as can be observed in Fig. 1.7 for the case of the perfectly conducting model represented by Eq. (1.16). A larger aperture width leads to a stronger interaction between the mode and the incident wave, and a shift of the resonance to longer wavenumbers is observed, with a cut-off width equal to $\lambda/2$, above which $q_y$ becomes real.

In practice, the mode properties are also modified by the finite conductivity of the metal walls, and its cut-off wavelength is increased whereas its cut-off hole dimension is reduced. In addition, higher diffraction orders will play a role in the grating scattering. However, such a model serves well to unveil the physics of the process. Another, more sophisticated, model is proposed in ref. [26] based on the assumption of small aperture diameter. It shows that the amplitude of the electric field inside small apertures grows linearly with the aperture diameter. For larger apertures this approximation is no longer valid, but rigorous theoretical calculations and measurement show that a maximum of the electromagnetic field intensity is obtained when the hole dimensions are just below the cut-off of the fundamental mode. This can be understood by taking into account the fact that the real part of the propagation constant $q_y$ of the waveguide mode at its cut-off is almost zero, as if the field were accumulated at the entrance of the aperture. As a consequence, the field inside the single aperture can be enhanced several times when compared to its value in free space, which leads to enhanced transmission through the waveguide mode inside the hole. An example well confirmed by fluorescent measurements [27] is given in Fig. 1.8(a), which presents the mean electric field.
intensity $I_S$ as measured just below the aperture entrance and defined as

$$I_S = \frac{1}{\pi R^2} \int \frac{|E|^2}{r^2} dS,$$

where the integral is taken over the cross-section of the aperture. For small radii its dependence on the radius is quadratic, and a seven-fold enhancement of it when compared to the field inside the solvent without apertures is observed close to the cut-off of the fundamental mode. Figure 1.8(b) presents the real and imaginary parts of the normalized mode constant $\gamma = q_y / k_0$.

1.5 Enhanced Fabry–Perot resonances through evanescent modes

A direct consequence of the fact that a plasmon surface wave propagating on the horizontal surface can resonantly interact with the vertical evanescent mode inside a hole is discussed in ref. [24]. If the model presented in the preceding section is extended to a layer with two metallic surfaces, a perforated layer, there will be additional reflection of the mode when it reaches the lower surface. Multiple reflection leads to Fabry–Perot resonances, which, however, are quite weak when evanescent waves are used instead of propagating waves. Indeed, a textbook formula is applicable to both propagating and evanescent modes. The transmission of a Fabry–Perot resonator is proportional to a denominator that contains the mode propagation constant $q_y$, the layer thickness $h$ (waveguide length), and the product of the reflection coefficients of the mode on the upper ($r_+$) and lower ($r_-$) surfaces:

$$T \sim \frac{1}{1 - r_+ r_- \exp(2i q_y h)}.$$

The usual case of evanescent waves with imaginary $q_y$ cannot ensure strong resonances because the reflection coefficients are smaller than unity in modulus. However, this is not the case discussed in the preceding section. We have observed the possibility of having a ten-fold or more increase in the coefficient of reflection $r$, although we have considered the reflection from the cladding into the cladding, rather than from inside the hole backwards to the hole. A similar analysis shows that $r_+$ and $r_-$ are enhanced in a similar manner to compensate the exponential decay due to $\exp(2i q_y h)$, leading to enhanced transmission through evanescent-wave Fabry–Perot resonances.

1.6 What resonance predominates?

As shown in Sections 1.4 and 1.5, the evanescent fundamental TE mode plays a crucial role in the transmission mechanism. However, this role alone cannot explain the sharp spectral and angular variations of the transmission, because its wavevector component $k_z / k_0$ is much larger than unity. Indeed, in the example considered at the end of Section 1.4, the propagation constant in the vertical direction, $q_y / k_0 = 1.35$, corresponds to a value of $k_z / k_0 = 3.5$ in air or 3.67 in glass. Such a mode is quite difficult to excite by an incident plane wave, even with the help of the grating periodicity.

On the other hand, we have observed in Section 1.4 that the interaction between the incident plane wave and the fundamental TE$_{11}$ mode inside each aperture can have a resonance close to $k_z / k_0$ slightly greater than unity (in air), provided that both the incident wave and the mode are evanescent. On the other hand, the excitation of the PSW on a plane surface also requires evanescent incident fields with $k_z / k_0 > 1$. A natural question that arises is whether there is a difference between these resonances, or whether they represent two physical interpretations of the same resonance. If there is a real difference obtained through a complete diffraction analysis (and not by a simplified model), this means that the excitation of the evanescent mode on the aperture opening would create a new surface resonance on the interface between the perforated metal and the dielectric cladding, addition to the PSW. A resonance of the reflection coefficient means the existence of a scattered field without the existence of an incident one. In addition, as far as $k_z / k_0 > 1$, this diffracted field will be evanescent in the cladding, i.e. localized close to the interface, in the same manner as for the PSW. As a consequence of this hypothesis, it would be possible to excite both resonances with the help of the grating periodicity, and the excitation would occur, in general, at different wavelengths and angular conditions. However, both experimental and numerical results do not show such a double resonance. Another possibility is that the two resonances exist, and that they represent different phenomena, but the values of $k_z$ at which they appear coincide, as for the case of phase-assisted second-harmonic generation in periodically corrugated (or poled) dielectric waveguides with guide mode waves excited at both the fundamental and the harmonic frequencies [28]. However, if such a case appears, the resonant response will be observed as a double Lorentzian (with two identical poles). Moreover, a rigorous numerical analysis carried out in ref. [23] shows that the transmission intensity $T$ of the structure analyzed experimentally in ref. [1] can be extremely well approximated by a single pole of the transmission amplitude:

$$T = \left| \frac{k_z - k_z^*}{k_z - k_z^*} \right|^2 \sim \frac{\lambda - \lambda^*}{\lambda - \lambda^*},$$

where $\rho$ stands for a pole, and $z$ for zero.

Although, in general, there are four surface waves excited in normal incidence, those propagating in the $+x$, $-x$, $+z$, and $-z$-directions, they are all coupled
by the grating to form four different standing waves, symmetric or antisymmetric with respect to the origin. There is only one solution symmetric with respect to the change of sign of the \( z \)-axis and antisymmetric with respect to the sign of the \( x \)-axis, having the same symmetry as the incident wave is polarized along the \( x \)-axis. The other three solutions cannot be excited in normal incidence by an incident plane wave.

The Lorentzian resonance response is deformed by the existence of a zero, a fact that is well known in grating theories, and is necessary in order to limit the value of \( T \) and to compensate the pole when the modulation due to the grating (the hole dimensions in our case) tends to zero, because no anomaly is observed without the grating. As demonstrated for classical one-dimensional gratings [29], the zero \( \lambda_0^2 \) can become real for a certain value of the grating depth, which leads to a total light absorption in TM polarization. The comparison between the numerical values obtained by rigorous electromagnetic theory and the results obtained using Eq. (1.19) are given in Fig. 1.9, which clearly demonstrates the existence of only a single pole (the square arises because it is the intensity that is being calculated).

An additional argument (presented in ref. [23]) that the resonance responsible for the enhanced transmission is due to the PSW is its trajectory in the complex \( \lambda \)-plane when the aperture is shrunk to zero. The resonant wavelength gradually tends to the resonant wavelength of the PSW on the unperturbed plane metal–glass interface, as observed in Fig. 1.10, in the same manner as the pole of Eq.(1.16) for a perfectly conducting metal tends towards the wavenumber of the cladding; see Fig. 1.7.

The presence of a hole modifies the constant of propagation of the PSW due to scattering, to the radiation losses through the grating periodicity, and to the interaction with the waveguide evanescent mode. That is the reason why the analytical formula given by Eq. (1.6) for the complex value of \( \lambda_0 \) in the case of an unperturbed flat surface is no longer valid in the case of a periodic array of holes. The red-shift of the pole explains why the transmission maximum appears at longer wavelengths than those predicted by Eq. (1.6). The zero \( \lambda_0^2 \) (see the caption of Fig. 1.9) is almost real, leading to the almost zero transmission when \( \lambda = \lambda_0 \) observed in Fig. 1.4. In addition, the real part of \( \lambda_0^2 \) almost coincides with the initial position of the pole, when \( w = 0 \) (Fig. 1.10), which explains why the transmission drops to zero at a wavelength corresponding to PSW excitation at a flat metallic surface without holes, a fact that has put in question the role of the PSW in the enhanced transmission. The correct explanation comes from a proper understanding of the red-shift of the pole observed in Fig. 1.10, presenting Fano-type anomalies in transmission [30, 31].

A direct proof that the resonance of the mode reflection coefficient, discussed in Section 1.4 as a consequence of Eq. (1.16), is the same as the PSW on the perforated structure, can be found using a more sophisticated model, which analyzes the scattering by a single aperture in a real metal in the approximation of a very small aperture diameter. The excitation of the waveguide mode in such small
cross-section waveguides can be represented by a single magnetic dipole, as in Bethe’s paper [32], but buried inside the metallic layer [26]. Its coupling to the outside radiation is achieved through the classical Fresnel coefficients of transmission and reflection at the dielectric–metal interface, which presents a pole, known as the PSW. When the hole diameter is larger, the coupling is stronger, as expected, and one always finds in the spatial spectrum of the scattered light the signature of the PSW pole, as observed in Fig. 1.11 for the \( k_x \) decomposition of \( E_y \) on the entrance surface of a circular aperture (250 nm diameter) in a silver sheet (thickness 200 nm) illuminated normally with TM polarized light (wavelength 500 nm) from air. The position of the resonance anomaly coincides with the normalized constant of propagation of the PSW on the planar surface. When a hole array is considered, this spatial component can be enhanced by the periodicity under the conditions given by the grating equation, Eq. (1.8).

1.7 Nonplasmonic contributions

The role of PSW in producing the minima and maxima observed in the spectral features of the transmission through subwavelength holes is now clearly established. However, the question remains as to whether the PSW is the only propagating wave along the metallic surface launched by subwavelength holes. This question has opened the third period of research since the discovery of Ebbesen and co-workers. Let us recall that the first period (1998–2000) was dedicated to the study of enhanced transmission through subwavelength slits. The second period (2000–2004) was marked by the development of rigorous numerical methods able to tackle arrays of subwavelength apertures. Numerical and experimental studies permitted understanding the role of the shape and size of a hole. This period corresponds also to the emergence of the complementary thematic of light diffraction by single subwavelength apertures and its applications in biosensing, single-molecule analysis, membrane analysis, etc.

During the entire decade, however, there have always been studies that contested the role of the surface plasmon in the enhanced transmission (see, for example, the discussion after Fig. 1.10). In 2004, another model, called the composite diffracted evanescent wave (CDEW) model, was introduced into the theory of light diffraction by subwavelength indentations [34]. The CDEW model predicts that the summation of all diffracted inhomogeneous waves by a subwavelength indentation results in a propagating wave along the metallic surface, with a propagation constant \( k_0 \) and an amplitude decaying with increasing distance from an indentation as \( 1/r \). This paper attempted to explain the minima and maxima observed in the transmission of light by an array of subwavelength holes with the use of the CDEW model only and by fully neglecting the role of the PSW. It is now established that PSWs are a key component of the mechanism of enhanced transmission. However, this study attracted the attention of researchers to other kinds of electromagnetic waves propagating close to the surface. The fundamental problem that arose was the determination of the properties of waves launched by a subwavelength indentation when illuminated by an incident plane wave. It was then necessary to simplify considerably the device under study to determine quantitatively the properties of this wave. Gay et al. experimentally studied the interference between light transmitted through a slit and waves launched by a neighboring nanoslit as a function of the distance between the two neighboring slits [35]. They explained their result with the use of the CDEW model, but rapidly this formalism has been strongly debated [36, 37]. This model has been therefore replaced by the so-called quasi-cylindrical wave (QCW) model.

The existence of this wave is better understood on the basis of the experiment made by Aigouy et al. [38]. In 2007, they studied experimentally the near field distribution created by a slit-doublet on a metallic sheet. When the slits were illuminated in TM polarization, they observed the presence of an interference pattern between the two slits generated by two counter-propagating waves (Fig. 1.12). The interference pattern has been thoroughly studied at different distances from...
the slit, and Aigouy et al. confirmed in the near field the observations made by Gay et al. in the far field: a rapid decrease of the amplitude of the waves close to the slits followed by a persistent wave farther from the slit, i.e. in the central part of the doublet. The fringes observed in the central part are well fitted by the interference of two counter-propagating PSWs only. The persistent surface wave was then identified with PSWs and the rapidly decaying wave was attributed to QCWs (see the inset in Fig. 1.12). Both waves propagate on the metallic surface, the PSW with a propagation constant $k_{SP}$ slightly higher than $k_0$, and the QCW with a propagation constant equal to $k_0$. Thorough experimental, numerical, and theoretical studies revealed that QCWs present two decay rates with respect to the distance $\rho$ from the line source along the surface, different from the decay rate obtained with the CDEW model. It has been established that the QCW decays as $\rho^{-1/2}$ and $\rho^{-3/2}$. An analytical expression of the QCW can be obtained in the case of a line source located at a metal–dielectric interface. It is the product of an Erf-like envelope function of $\rho$, and a term proportional to $\rho^{-3/2}$ with a wavevector parallel to the interface. The two decay rates observed experimentally are the two asymptotes of the Erf-like envelope: very close to the slit the decay rate of the QCW is $\rho^{-1/2}$; farther away, the envelope vanishes and the rate tends towards $\rho^{-3/2}$ [39]. A numerical study [33] made in 2005 of light diffracted by a single circular aperture on a metallic sheet in a direction parallel to the surface had already shown that the electric field decay as a function of the distance $\rho$ from the hole is determined mainly by two contributions: the PSW and a spherical wave decaying as $r^{-1}$, which represents the field radiated by a single dipole, as predicted by Bethe’s model. Along the surface, the propagation constant of this spherical wave is equal to the free space wavevector $k_0$ (in vacuum or air). If we consider a slit, or a chain of holes, as discussed in the next paragraph, they radiate like a line antenna, with an electric field decreasing as $\rho^{-1/2}$, a fact that provides a simple explanation of the QCW.

In 2008, Liu and Lalanne published a microscopic analysis of the enhanced light transmission based on the scattering of PSW on an array of subwavelength holes [40]. This analysis permitted tackling separately the contribution of the PSW and quantifying its role in the energy transmitted. Microscopic here means that the holes are considered individually as single scatterers. More precisely, the authors considered a one-dimensional linear chain of holes, and they listed the elementary scattering coefficients when the PSWs interact with the chain: PSWs can be reflected (with reflection coefficient $P$), transmitted (transmission coefficient $T$) on the interface plane, or they can be scattered into outgoing plane waves and modes in the holes (with coefficient $\alpha$). The coefficient of reflection $R_\lambda$ of the Bloch mode supported by the array on the front surface of the holes can be written in the case of normal incidence as

$$R_\lambda = R + \frac{2\alpha^2}{u^{-1} - (P + T)},$$  (1.20)

where $R$ is the coefficient of reflection of the Bloch mode of a single hole chain, $u = \exp(i k_0 d)$, and $d$ is the distance between the hole chains. This analytical expression takes into account the excitation of a PSW by a linear chain of holes and the scattering of this PSW by the infinity of other linear chains of holes.

In addition, the authors of ref. [40] were able to account for the contribution of the QCW wave. Not surprisingly, in conditions close to those of the initial experiment, this model reveals all the spectral features of light transmission by a 2D array of subwavelength holes. Maxima and minima are well represented, and a comparison with a rigorous numerical study confirms that their frequencies are well predicted by the pure PSW model. However, this comparison also reveals that this model does not predict the exact portion of light transmitted. More precisely, the pure PSW model predicts about half of the total light transmitted. This fact demonstrates that PSWs are fully involved in the transmission of light through 2D arrays of subwavelength holes but that QCWs are responsible for the other half of the light energy transmitted. The amount of light energy transmitted by PSWs
diminishes for lower frequencies, where QCWs are predominant. Liu and Lalanne extracted the PSW contribution from the total field and they represented the real part of the $y$-component of the magnetic field along the $x$-axis. They confirmed the fact that the field scattered by subwavelength holes is not a pure PSW mode but that it also contains a QCW contribution. The two waves are excited with a very moderate phase difference, their phase velocities $k_x$ and $k_y$ are very close, and their contributions interfere constructively. The radiative decay of the QCW scales as $\rho^{-1/2}$, so that it is much faster than the PSW (as observed on the slit model in Fig. 1.12). However, close to the linear chain, the QCW contributes equally with the PSW to the total light transmitted.

1.8 Conclusions

We are now able to understand quite well the phenomenon of enhanced light transmission through an array of subwavelength apertures in a metallic screen. The scattering of the incident field on the aperture boundaries creates waves with a variety of wavevector directions, in particular the surface plasmon along the upper boundary. When the array periodicity corresponds to the resonant conditions of PSW excitation, as given by the grating equation, the PSW wave is enhanced significantly, which creates a local field enhancement on the metal surface. This enhancement leads to a resonant excitation of the fundamental waveguide mode inside each aperture, due to the fact that the transmission coefficient for its excitation has a resonance corresponding to the PSW on the interface. Of course, the propagation constant of the PSW is modified by the interaction, which can be used to create similar surface waves in the microwave domain. At normal incidence, the PSW forms a standing wave on the entrance interface, so that, at both sides of each aperture, the electromagnetic energy flows towards that opening, somehow as if the incident intensity is gathered inside the aperture, as seen in Fig. 1.13. In addition to the PSW, each hole generates a radiated field, a part of which propagates in grazing directions to the surface, according to ref. [39] and [40], and can reach the neighboring holes to enhance the field there.

The fact that the fundamental mode is resonantly enhanced in this process explains why the field transmitted by the mode is not negligible at the exit side, even if the mode is evanescent for a small hole cross-section. At the exit side, the mode excites both the PSW and a radiated field, and the radiation from each hole interferes to form the transmitted zeroth diffracted order. If the system is symmetrical with identical substrate and cladding, the resonant conditions for excitation of the PSW on the entry interface are the same as for a resonant emission of the PSW on the exit interface into the zeroth transmission order, which enhances the transmission even more.

![Figure 1.13. Schematic view of the Poynting vector flow on the upper surface and in the vicinity of the hole entrance.](image)

In the case when the periodicity is not suitable to excite the PSW on the entry interface, another resonance can be observed when a PSW is excited on the lower interface. The waveguide mode inside the holes acts to enhance the tunneling of light from the entry to the exit surface. When the grating period is suitably chosen, the periodicity can add in phase the PSW generated on each exit aperture, enhancing the PSW. In addition, the same periodicity also enhances the radiation of the PSW into the substrate, which explains the existence of the transmission peak close to 1.4 $\mu$m in the experiment of Ebbesen et al. Another way of explaining this case is to use the reciprocity theorem, which implies in particular that the transmission into the substrate when light is incident from the cladding is the same in the reciprocal case of transmission into the cladding when light is incident from the substrate side. The latter provides suitable conditions for the excitation of the PSW on the entry side with all the resulting transmission enhancement, as explained above.

Other factors can modify the system response and contribute to its quantitative understanding. For example, the hole form modifies the polarization response. If the dimensions are close to the cut-off, the field inside the aperture is enhanced as the mode propagation constant in the vertical direction approaches zero.

References


