



Nano-optics of surface plasmon polaritons

Anatoly V. Zayats^{a,*}, Igor I. Smolyaninov^b, Alexei A. Maradudin^c

^a*School of Mathematics and Physics, The Queen's University of Belfast, Belfast BT7 1NN, UK*

^b*Department of Electrical and Computer Engineering, University of Maryland, College Park, MD 20742, USA*

^c*Department of Physics and Astronomy and Institute for Surface and Interface Science, University of California, Irvine, CA 92697, USA*

Accepted 1 November 2004

editor: D.L. Mills

Abstract

A surface plasmon polariton (SPP) is an electromagnetic excitation existing on the surface of a good metal. It is an intrinsically two-dimensional excitation whose electromagnetic field decays exponentially with distance from the surface. In the past, it was possible to study only the (far-field) scattered light produced by the interaction of surface polaritons with surface features. Only with the development of scanning near-field optical microscopy did it become possible to measure the surface polariton field directly in close proximity to the surface where the SPP exists. Here we overview the near-field studies of surface polaritons on randomly rough and nanostructured surfaces, theoretical models of SPP interaction with surface features, and SPP applications in novel photonic technologies. Surface polariton scattering, interference, backscattering, and localization will be discussed, as well as concepts of surface polariton optics and polaritonic crystals. Surface polaritons are finding an ever increasing number of applications in traditional domains of surface characterization and sensors as well as in newly emerging nano-photonic and optoelectronic technologies.

© 2004 Elsevier B.V. All rights reserved.

PACS: 73.20.Mf; 78.66.–w; 07.79.Fc

Keywords: Surface plasmon polaritons; Scanning near-field optical microscopy; Nano-optics

* Corresponding author.

E-mail address: a.zayats@qub.ac.uk (A.V. Zayats).

Contents

1. Introduction	133
1.1. Surface electromagnetic waves at a dielectric–metal interface	134
1.1.1. Surface plasmon polaritons	135
1.1.2. Surface plasmons	138
1.1.3. A metal film on a dielectric substrate	139
1.1.4. Localized surface plasmons	142
1.1.5. Optical excitation of surface plasmon polaritons	144
1.2. Scanning near-field optical microscopy	146
2. Surface polariton scattering	149
2.1. Single scattering regime: the SPP interaction with individual defects	151
2.1.1. The interaction of a surface plasmon polariton with a single surface defect	151
2.1.2. The propagation of a surface plasmon polariton near an index step	163
2.1.3. The interaction of a surface plasmon polariton with a subsurface defect	169
2.1.4. Near-field imaging of the SPP interaction with a single surface defect	175
2.1.5. Interference of surface plasmon polariton beams	179
2.2. Multiple scattering regime: backscattering of surface plasmon polaritons	185
2.2.1. Enhanced backscattering of surface polaritons on a randomly rough surface	186
2.2.2. Direct observation of SPP backscattering with SNOM	192
2.3. Multiple scattering regime: surface polaritons localization on rough surfaces	196
3. Two-dimensional optics of surface polaritons	202
3.1. The interaction of surface plasmon polaritons with material and/or geometrical discontinuities in their propagation path	202
3.1.1. The impedance boundary condition	203
3.1.2. Modal expansions (discrete)	205
3.1.3. Modal expansions (continuous)	211
3.2. Elements of SPP optics	214
4. Near-field microscopy of surface plasmon polaritons and characterization of metal surfaces	218
4.1. Characterization of metallic nanostructures via local excitation of surface polaritons	219
4.2. Imaging of thin film interfaces with air–metal and glass–metal surface polaritons	221
4.3. Fractal properties of rough surfaces and SPP scattering	222
5. Surface polaritons on periodically structured surfaces	226
5.1. Surface polaritonic crystals	228
5.1.1. One-dimensional periodic surface profile	228
5.1.2. Two-dimensional periodic surface profile	234
5.2. Imaging of the SPP Bloch waves	237
5.3. Waveguiding in line-defects of a SPP crystal	239
6. Enhanced optical properties mediated by surface plasmon polaritons	243
6.1. Surface plasmon polaritons and the enhanced backscattering of light from a randomly rough metal surface	243
6.2. Surface plasmon polaritons and the enhanced optical transmission of periodically structured films	257
6.2.1. Optical transmission of a metal film with a periodic hole array: a near field view	257
6.2.2. The origin of the enhanced transmission through periodically nanostructured metal films	262
6.2.3. Polarization effects in the light transmission through a polaritonic crystal	265
6.2.4. Light-controlled transmission of polaritonic crystals	269
7. Nonlinear optics of surface plasmon polaritons	274
7.1. Second harmonic generation in reflection from rough metal surfaces	275
7.1.1. Second harmonic generation from metallic diffraction gratings	281
7.1.2. Second harmonic generation at randomly rough metal surfaces	286
7.2. Near-field second-harmonic generation at metal surfaces	291

7.3. Near-field SHG from metallic diffraction gratings	296
8. SPP interrogation using other scanning probe microscopy techniques	298
8.1. Influence of surface polaritons on the tunneling current of STM	298
8.2. Light emission induced by the electron tunneling in STM: a role of surface plasmon polaritons	304
9. Conclusion	306
Acknowledgements	307
References	308

1. Introduction

In its simplest form a surface plasmon polariton (SPP) is an electromagnetic excitation that propagates in a wave like fashion along the planar interface between a metal and a dielectric medium, often vacuum, and whose amplitude decays exponentially with increasing distance into each medium from the interface [1–3]. Thus, a SPP is a surface electromagnetic wave, whose electromagnetic field is confined to the near vicinity of the dielectric–metal interface. This confinement leads to an enhancement of the electromagnetic field at the interface, resulting in an extraordinary sensitivity of SPPs to surface conditions. This sensitivity is extensively used for studying adsorbates on a surface, surface roughness, and related phenomena. Surface plasmon polariton-based devices exploiting this sensitivity are widely used in chemo- and bio-sensors [4]. The enhancement of the electromagnetic field at the interface is responsible for surface-enhanced optical phenomena such as Raman scattering, second harmonic generation (SHG), fluorescence, etc. [1,5].

The intrinsically two-dimensional nature of SPPs provides significant flexibility in engineering SPP-based all-optical integrated circuits needed for optical communications and optical computing [6,7]. The relative ease of manipulating SPPs on a surface opens an opportunity for their application to photonics and optoelectronics for scaling down optical and electronic devices to nanometric dimensions. Most importantly, active photonic elements based on nonlinear surface plasmon polariton optics, which allow controlling optical properties with light, are much easier to realize with suitably patterned metal surfaces, due to the SPP-related electromagnetic field enhancement near a metal surface [7].

Since the electromagnetic field of a SPP decays exponentially with distance from a surface, it cannot be observed in conventional (far-field) experiments unless the SPP is transformed into light, e.g., by its interaction with surface inhomogeneities. In the past, it was possible to measure only the (far-field) scattered light produced by the interaction of a SPP with surface features, and then to derive the information on the surface polariton behavior from further assumptions about the generally unknown scattering processes [1,2]. It was possible to change in one or another way the average defect structure of a surface and to judge the SPP behavior on the surface from the changes in the angular spectrum and intensity of the scattered light. The microscopic structure of the surface was largely unknown in such experiments. Even if a surface topography is visualized by electron microscopy or later, by scanning atomic force and electron tunneling microscopy, the optical signal measured in the far-field is averaged over an ensemble of surface features due to diffraction effects, and it is therefore possible only to correlate the behavior of the scattered light with the average surface topography. The problem of local measurement of the SPP field close to a surface was addressed using a photosensitive polymer layer covering the metal surface under investigation [8]. The electromagnetic field distribution was recorded on the photoresist and, after it was developed, examined with an electron microscope. However, the polymer layer introduces significant

perturbations of the SPP resonant conditions and the SPP field over the metal surface, which limits the applicability and possibilities of this approach.

Only with the development of scanning probe techniques did it become possible to study SPPs locally on a surface. In particular, scanning near-field optical microscopy (SNOM) [9] provides an opportunity to probe the SPP field directly over the surface where the SPP exists, with a resolution in the nanometer range. The implementation of SNOM has led to a breakthrough in surface polariton studies [10–18]. Surface plasmon polariton scattering, interference, backscattering and localization have been visualized and investigated directly on the surface [16–20]. The idea of two-dimensional surface polariton optics has been proposed and realized experimentally, resulting in the development of optical elements for surface polaritons that allow manipulating and directing SPP beams in the same way as optical beams are directed in three dimensions [17]. Very recently a new class of two-dimensional photonic crystals based on surface plasmon polaritonic band-gap structures has been created on a metal surface [21–26].

In this article we present an overview of the theory of the behavior of SPP on structured surfaces, and of recent advances in experimental surface plasmon polariton studies. SNOM has allowed visualizing SPP related processes on a surface, correlating the SPP behavior to surface structure, and confronting the experimental studies with model calculations. Scattering of surface plasmon polaritons by surface roughness is one of the subjects that can be significantly complemented with local studies of the electromagnetic field distribution over a surface provided by scanning near-field optical microscopy. These will be discussed in Section 2. Imaging of SPP scattering from individual surface defects as well as from ensembles of defects will also be considered. Different types of SPP behavior on rough surfaces, such as scattering, reflection, interference, backscattering, and localization will be analyzed, and related theoretical models will be presented. A detailed theoretical description of the interaction of SPP with surface features of different topology, with surface and subsurface defects, as well as with topographical and index-step defects will be discussed. This interaction is important for an understanding of near-field SPP images and the development of SPP optics. Examples of the elements of two-dimensional SPP optics will be described, and possible applications will be discussed, in Section 3. Section 4 is devoted to the applications of near-field studies of SPPs on rough surfaces to the characterization of metal surfaces, including SNOM measurements of the dielectric constants of metallic micro- and nanostructures, imaging of inner interfaces of thin metal films, and fractal properties of surfaces. The effect of periodically structured surfaces on SPP behavior will be discussed in Section 5, where properties of 1D and 2D surface polariton band-gap structures will be considered. The dependence of “bulk” optical properties of rough and periodically nanostructured metallic films on SPP effects will be described in Section 6, where polarization and nonlinear control of optical properties of nanostructures will be discussed. In Section 7, the influence of SPP on nonlinear optical processes occurring on a metal surface will be considered on the example of second-harmonic generation at rough metal surfaces. Finally, in Section 8, applications of scanning tunneling microscopy in surface polariton measurements are presented, and associated mechanisms are considered that are related to the interplay between surface polaritons and localized surface electromagnetic resonances.

1.1. Surface electromagnetic waves at a dielectric–metal interface

Although the emphasis in this article is on the interaction of surface plasmon polaritons with rough and nanostructured surfaces and films, it is useful to preface the discussion of these interactions with a

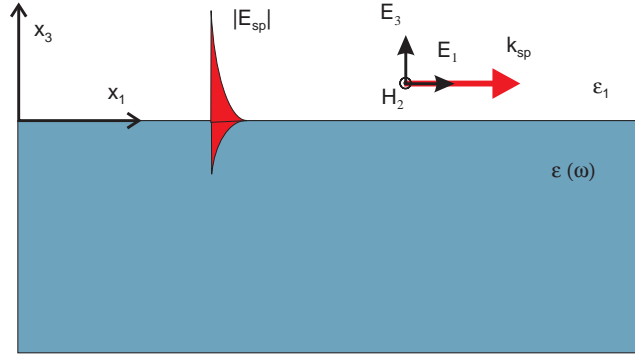


Fig. 1. Surface plasmon polariton at a dielectric–metal interface.

brief summary of those properties of surface plasmon polaritons on smooth surfaces and in smooth films that will be needed in what follows. In this way the results presented in this review can be compared or contrasted with the better known results for a planar surface, and in the case of weak surface roughness can serve as the basis of perturbative treatments of these interactions.

1.1.1. Surface plasmon polaritons

The electromagnetic field of a surface plasmon polariton at a dielectric–metal interface is obtained from the solution of Maxwell’s equations in each medium, and the associated boundary conditions. The latter express the continuity of the tangential components of the electric and magnetic fields across the interface, and the vanishing of these fields infinitely far from the interface. To introduce the main parameters characterizing surface plasmon polaritons, let us consider a system consisting of a dielectric material, characterized by an isotropic, real, positive dielectric constant ϵ_1 , in the half-space $x_3 > 0$, and a metal, characterized by an isotropic, frequency-dependent, complex dielectric function $\epsilon(\omega) = \epsilon_1(\omega) + i\epsilon_2(\omega)$, in the half-space $x_3 < 0$ (Fig. 1).

We first consider a p-polarized (transverse magnetic or TM) wave in this structure that propagates in the x_1 -direction. (Due to the optical isotropy of the two media there is no loss of generality in choosing this direction of propagation.) In a wave of this polarization the magnetic vector is perpendicular to the plane of incidence—the plane defined by the direction of propagation and the normal to the surface. The solutions of Maxwell’s equations that are wavelike in the x_1 -direction and whose amplitudes decay exponentially with increasing distance into each medium from the interface $x_3 = 0$ can be written as

$$\mathbf{H}^>(\mathbf{x}; t) = (0, A, 0)e^{ikx_1 - k_3^{(1)}x_3 - i\omega t} \quad (1.1a)$$

$$\mathbf{E}^>(\mathbf{x}; t) = -A \frac{c}{i\omega\epsilon_1} (k_3^{(1)}, 0, ik)e^{ikx_1 - k_3^{(1)}x_3 - i\omega t} \quad (1.1b)$$

in the region $x_3 > 0$, and as

$$\mathbf{H}^<(\mathbf{x}; t) = (0, B, 0)e^{ikx_1 + k_3^{(m)}x_3 - i\omega t} \quad (1.2a)$$

$$\mathbf{E}^<(\mathbf{x}; t) = -B \frac{c}{i\omega\epsilon(\omega)} (-k_3^{(m)}, 0, ik)e^{ikx_1 + k_3^{(m)}x_3 - i\omega t} \quad (1.2b)$$

in the region $x_3 < 0$, where $k_3^{(1,m)}$ determine the decay of the electromagnetic field with increasing distance from the surface:

$$k_3^{(1)} = (k^2 - \epsilon_1(\omega/c)^2)^{1/2} , \quad (1.3a)$$

$$k_3^{(m)} = (k^2 - \epsilon(\omega)(\omega/c)^2)^{1/2} . \quad (1.3b)$$

The real parts of $k_3^{(1)}$ and $k_3^{(m)}$ must be positive in order that Eqs. (1.1) and (1.2) describe an electromagnetic wave localized to the dielectric–metal interface at $x_3 = 0$. The boundary conditions at the plane $x_3 = 0$ yield the pair of equations

$$A = B, \quad A \frac{k_3^{(1)}}{\epsilon_1} = -B \frac{k_3^{(m)}}{\epsilon(\omega)} . \quad (1.4)$$

The condition that this system of equations have a nontrivial solution provides the dispersion relation connecting the frequency ω of the p-polarized wave and its wavenumber k ,

$$\frac{k_3^{(m)}}{k_3^{(1)}} = -\frac{\epsilon(\omega)}{\epsilon_1} . \quad (1.5)$$

If, for the moment, we assume that the dielectric function of the metal, $\epsilon(\omega)$ is real, then $k_3^{(1)}$ and $k_3^{(m)}$ must be real and positive in order that Eqs. (1.1) and (1.2) describe a surface electromagnetic wave. It follows, therefore, from Eq. (1.5) that $\epsilon(\omega)$ must be negative for this surface electromagnetic wave to exist. In this frequency range the wave vector of a volume electromagnetic wave is pure imaginary. Thus such surface electromagnetic waves exist only in the frequency range where volume electromagnetic waves cannot propagate in a metal.

The surface electromagnetic wave whose field vectors are given by Eqs. (1.1) and (1.2) with $A = B$, and whose frequency is obtained from the dispersion relation, Eq. (1.5) is called a *surface plasmon polariton*. Physically, one can understand it as photons coupled to collective excitations of conduction electrons near a metal surface.

If we square both sides of Eq. (1.5), and rearrange terms, we obtain an explicit expression for the wavenumber k_{sp} of the surface plasmon polariton as a function of its frequency ω ,

$$k_{\text{sp}} = \frac{\omega}{c} \left[\frac{\epsilon_1 \epsilon(\omega)}{\epsilon_1 + \epsilon(\omega)} \right]^{1/2} , \quad (1.6)$$

a relation that is valid even if $\epsilon(\omega)$ is complex.

To illustrate the preceding results let us assume for the dielectric function of the metal the too often used, and never completely valid, free electron expression

$$\epsilon(\omega) = 1 - \frac{\omega_p^2}{\omega^2} , \quad (1.7)$$

where ω_p is the frequency of bulk longitudinal electron excitations, the plasma frequency. The corresponding dispersion curve is shown in Fig. 2. The requirement that $k_3^{(1)}$ be real and positive in order that the electromagnetic field of the surface plasmon polariton decay exponentially into the dielectric medium in contact with the metal has the consequence that the SPP dispersion curve must lie to the

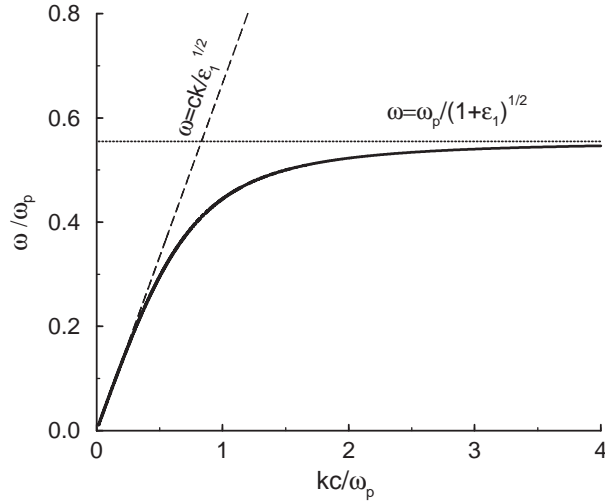


Fig. 2. Dispersion of SPP at a dielectric–metal interface. Also plotted is the dispersion of light in the dielectric medium, and the corresponding surface plasmon frequency. The parameters assumed are $\omega_p = 11.9989 \times 10^{15} \text{ s}^{-1}$ (silver) and $\epsilon_1 = 2.25$.

right of the dispersion curve of light in the dielectric medium, $\omega = ck/\epsilon_1^{1/2}$, the so-called dielectric light line. Consequently, the SPP cannot radiate light into the dielectric medium, and cannot be excited with conventional illumination from the adjacent dielectric.

Due to ohmic losses in the metal, characterized by the imaginary part of the dielectric function of the metal $\epsilon_2(\omega) (> 0)$, the energy carried by a SPP decays exponentially as the SPP propagates along the planar dielectric–metal interface. The $1/e$ decay length, called the energy propagation length L_{sp} is determined by the imaginary part of the SPP wavenumber, Eq. (1.6),

$$\frac{1}{L_{\text{sp}}} = 2 \text{Im } k_{\text{sp}} = \frac{\omega}{c} \frac{\epsilon_1^{3/2} \epsilon_2(\omega)}{|\epsilon_1(\omega)|^{1/2} (|\epsilon_1(\omega)| - \epsilon_1)^{3/2}}, \quad (1.8)$$

to the lowest nonzero order in $\epsilon_2(\omega)$. In writing Eq. (1.8) we have explicitly taken into account that the real part of $\epsilon(\omega)$, $\epsilon_1(\omega)$, is negative in the frequency range in which the SPP exists.

A surface plasmon polariton has both transverse and longitudinal electromagnetic field components. The magnetic field of a SPP is parallel to the surface and perpendicular to its direction of propagation. The electric field of a SPP has a component that is parallel to its direction of propagation and a component perpendicular to the surface. The ratio of the transverse and longitudinal components of the electric field in the dielectric medium is

$$\frac{E_3^>}{E_1^>} = i \frac{k_{\text{sp}}}{k_3^{(1)}} = i \left(\frac{-\epsilon(\omega)}{\epsilon_1} \right)^{1/2} = i \left(\frac{\omega_p^2 - \omega^2}{\epsilon_1 \omega^2} \right)^{1/2}, \quad (1.9)$$

where the last expression is obtained when the free electron expression (1.7) is assumed for $\epsilon(\omega)$. The transverse component is dominant in the electric field of SPPs with small wavenumbers and low frequencies (photon-like excitations) for which the dispersion curve is close to the dielectric light line. Only for

very large wave numbers are the transverse and longitudinal components comparable. Indeed, they are equal at the frequency given by $\omega_p/(\epsilon_1 + 1)^{1/2}$.

We now consider an s-polarized (transverse electric or TE) wave in the structure depicted in Fig. 1. In a wave of this polarization it is the electric vector that is perpendicular to the plane of incidence. The solutions of Maxwell's equations that are wavelike in the x_1 -direction and are localized to the interface can be written as

$$\mathbf{E}^>(\mathbf{x}; t) = (0, A, 0)e^{ikx_1 - k_3^{(1)}x_3 - i\omega t} \quad (1.10a)$$

$$\mathbf{H}^>(\mathbf{x}; t) = A \frac{c}{i\omega} (k_3^{(1)}, 0, ik)e^{ikx_1 - k_3^{(1)}x_3 - i\omega t} \quad (1.10b)$$

in the region $x_3 > 0$, and

$$\mathbf{E}^<(\mathbf{x}; t) = (0, B, 0)e^{ikx_1 + k_3^{(m)}x_3 - i\omega t} \quad (1.11a)$$

$$\mathbf{H}^<(\mathbf{x}; t) = B \frac{c}{i\omega} (-k_3^{(m)}, 0, ik)e^{ikx_1 + k_3^{(m)}x_3 - i\omega t} , \quad (1.11b)$$

in the region $x_3 < 0$. The continuity of the tangential components of the electric and magnetic fields across the interface $x_3 = 0$ yields the pair of equations

$$A = B, \quad A \frac{c}{i\omega} k_3^{(1)} = -B \frac{c}{i\omega} k_3^{(m)} , \quad (1.12)$$

which can be combined into the single equation

$$(k_3^{(1)} + k_3^{(m)}) A = 0 . \quad (1.13)$$

It is not difficult to show from Eq. (1.13) that because the real parts of $k_3^{(1)}$ and $k_3^{(m)}$ are both required to be positive in order that we have an electromagnetic wave localized to the dielectric–metal interface $x_3 = 0$, the only solution of this equation is $A = 0$, so that $B = 0$. Thus an s-polarized surface plasmon polariton cannot exist in the structure depicted in Fig. 1.

1.1.2. Surface plasmons

We have indicated in the opening paragraph of the Introduction that surface plasmon polaritons are solutions of Maxwell's equations in which the effects of retardation—the finiteness of the speed of light—are taken into account. An important subclass of surface plasmon polaritons are *surface plasmons*. These can be viewed as the limiting case of surface plasmon polaritons when the speed of light is allowed to become infinitely large. Alternatively, and equivalently, they are obtained from solutions of Laplace's equation for a scalar potential that propagate in a wavelike fashion along a planar dielectric–metal interface, and whose amplitudes decay exponentially with increasing distance from the interface into each medium. They are therefore electrostatic surface waves. One can think of them as related to non-propagating collective vibrations of the electron plasma near the metal surface.

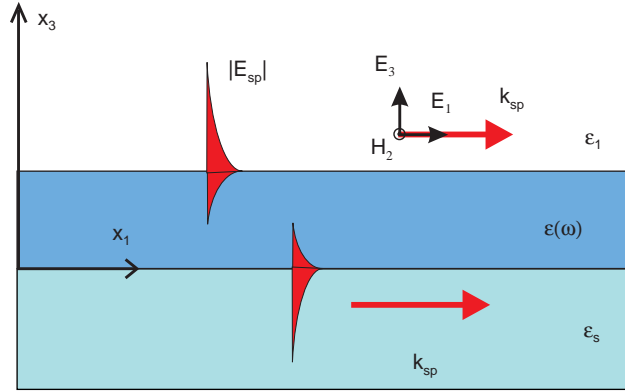


Fig. 3. Surface plasmon polaritons on metal film interfaces.

The solution of Laplace’s equation that is wavelike in the x_1 -direction, and decays exponentially as $|x_3| \rightarrow \infty$, is

$$\phi(\mathbf{x}; t) = A e^{ikx_1 - |k|x_3 - i\omega t}, \quad x_3 > 0, \quad (1.14a)$$

$$= B e^{ikx_1 + |k|x_3 - i\omega t}, \quad x_3 < 0. \quad (1.14b)$$

We must require that $\phi(\mathbf{x}; t)$ be continuous across the plane $x_3 = 0$. This ensures the continuity of the tangential components of the corresponding electric field $\mathbf{E}(\mathbf{x}; t) = -\nabla\phi(\mathbf{x}; t)$ across this plane. This condition yields the result $A = B$. We also have to require that $\epsilon\partial\phi(\mathbf{x}; t)/\partial x_3$ be continuous across the plane $x_3 = 0$. This condition is equivalent to the requirement that the normal component of the electric displacement be continuous across this plane. The condition whose satisfaction allows this to be done is

$$\epsilon(\omega) + \epsilon_1 = 0. \quad (1.15)$$

The surface excitation whose potential is given by Eq. (1.14) with $A = B$, and whose frequency is obtained from Eq. (1.15) is called a *surface plasmon*.

We see from Eq. (1.15) that $\epsilon(\omega)$ has to be negative in order that a surface plasmon exist. Moreover, if we compare Eqs. (1.6) and (1.15) we see that the frequency of the surface plasmon is also the limiting frequency of the surface plasmon polariton as $k \rightarrow \infty$. If the metal is characterized by the free electron form of its dielectric function Eq. (1.7), the frequency of the surface plasmon is given by

$$\omega_{sp} = \frac{\omega_p}{\sqrt{\epsilon_1 + 1}}. \quad (1.16)$$

1.1.3. A metal film on a dielectric substrate

Surface plasmon polaritons on the surfaces of metal films possess interesting properties that surface plasmon polaritons on the surface of a semi-infinite metal do not possess. Thus, let us consider the system consisting of vacuum ($\epsilon_1 = 1$) in the region $x_3 > d$, a metal film characterized by a complex, frequency-dependent, dielectric function $\epsilon(\omega)$ in the region $0 < x_3 < d$, and a dielectric medium characterized by a real, positive dielectric constant ϵ_s in the region $x_3 < 0$ (Fig. 3). The single nonzero

component of the magnetic field in this system that describes a p-polarized electromagnetic wave propagating in the x_1 -direction and localized to each of its two interfaces, can be written in the following forms:

$$H_2(\mathbf{x}; t) = A e^{ikx_1 - k_3^{(1)}x_3 - i\omega t}, \quad x_3 \geq d \quad (1.17a)$$

$$= e^{ikx_1} \left[B e^{k_3^{(m)}x_3} + C e^{-k_3^{(m)}x_3} \right] e^{-i\omega t}, \quad 0 \leq x_3 \leq d \quad (1.17b)$$

$$= D e^{ikx_1 + k_3^{(s)}x_3 - i\omega t}, \quad x_3 \leq 0, \quad (1.17b)$$

where

$$k_3^{(1)} = [k^2 - \epsilon_1(\omega/c)^2]^{1/2}, \quad k_3^{(s)} = [k^2 - \epsilon_s(\omega/c)^2]^{1/2}, \quad (1.18)$$

while $k_3^{(m)}$ has been defined in Eq. (1.3b). For this wave to be localized at each of the two interfaces, the real parts of the decay constants $k_3^{(1)}$, $k_3^{(m)}$, $k_3^{(s)}$ must be positive.

The boundary conditions satisfied by $H_2(\mathbf{x}; t)$ at each interface are

(i)

$$H_2(\mathbf{x}; t) \text{ continuous}; \quad (1.19a)$$

(ii)

$$\frac{1}{\epsilon} \frac{\partial}{\partial x_3} H_2(\mathbf{x}; t) \text{ continuous}. \quad (1.19b)$$

When these conditions are applied to the solutions given by Eqs. (1.17), a set of homogeneous linear equations for the coefficients A, B, C, D is obtained. The solvability condition for this system is the dispersion relation for the surface plasmon polaritons in this structure:

$$\left[\frac{\epsilon(\omega)}{\epsilon_1} \frac{k_3^{(1)}}{k_3^{(m)}} + 1 \right] \left[\frac{\epsilon(\omega)}{\epsilon_s} \frac{k_3^{(s)}}{k_3^{(m)}} + 1 \right] = \left[\frac{\epsilon(\omega)}{\epsilon_1} \frac{k_3^{(1)}}{k_3^{(m)}} - 1 \right] \left[\frac{\epsilon(\omega)}{\epsilon_s} \frac{k_3^{(s)}}{k_3^{(m)}} - 1 \right] e^{-2k_3^{(m)}d}. \quad (1.20)$$

We see from Eq. (1.20) that in the limit as $d \rightarrow \infty$ it yields the pair of equations

$$\frac{\epsilon(\omega)}{\epsilon_1} \frac{k_3^{(1)}}{k_3^{(m)}} + 1 = 0, \quad \frac{\epsilon(\omega)}{\epsilon_s} \frac{k_3^{(s)}}{k_3^{(m)}} + 1 = 0. \quad (1.21)$$

These are recognized to be the dispersion relations for surface plasmon polaritons at the vacuum–metal and metal-substrate interfaces, respectively. For finite values of d the electromagnetic interaction between the two interfaces cannot be neglected, the surface plasmon polariton at one interface “feels” the existence of the surface plasmon polariton at the other interface, a “repulsion of levels” occurs, and the dispersion curves for the surface plasmon polariton localized at each of the two interfaces become distorted by the interaction of these waves [27]. This is depicted in Fig. 4 for a metal film characterized by the free electron dielectric function (1.7) on a glass substrate.

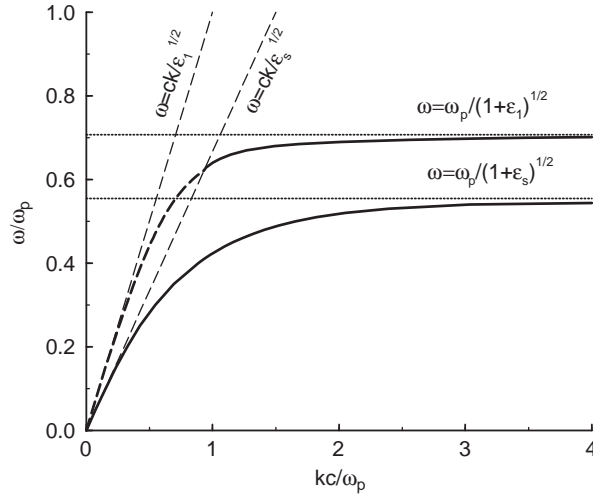


Fig. 4. Dispersion of SPPs at the interfaces of a metal film with a vacuum cladding and a dielectric substrate. Also plotted are the dispersions of light in the vacuum and the dielectric substrate, and the corresponding surface plasmon frequencies. The dashed part of the upper branch of the SPP dispersion curve shows the frequency range where this SPP, while being localized at the vacuum–metal film interface, leaks into the substrate. The parameters assumed are $d = 30$ nm, $\epsilon_1 = 1$, $\epsilon_s = 2.25$, and $\omega_p = 11.9989 \times 10^{15} \text{ s}^{-1}$ (silver).

An interesting special case of this structure is the one where both superstrate and substrate are vacuum, so that $\epsilon_1 = \epsilon_s = 1$ and $k_3^{(s)} = k_3^{(1)}$. In this case Eq. (1.20) can be rearranged into the pair of equations

$$\epsilon(\omega) \frac{k_3^{(1)}}{k_3^{(m)}} = -\coth \frac{1}{2} k_3^{(m)} d, \quad (1.22)$$

$$\epsilon(\omega) \frac{k_3^{(1)}}{k_3^{(m)}} = -\tanh \frac{1}{2} k_3^{(m)} d. \quad (1.23)$$

Equations (1.22) and (1.23) correspond to modes in which the tangential component of the electric field (E_1) is an even and an odd function of $(x_3 - d/2)$, respectively. For large wavenumbers the frequencies of the modes defined by Eqs. (1.22) and (1.23) are

$$\omega_- = \frac{\omega_p}{\sqrt{2}} [1 - e^{-|k|d}]^{1/2} \quad (1.24)$$

and

$$\omega_+ = \frac{\omega_p}{\sqrt{2}} [1 + e^{-|k|d}]^{1/2} \quad (1.25)$$

respectively, when the free electron form (1.7) for $\epsilon(\omega)$ is assumed.

For such thin metal films that the electromagnetic interaction between the two surface plasmon polaritons cannot be neglected, the imaginary part of the wavenumber determining the propagation length of the high frequency mode (the solution of Eq. (1.23)) decreases as the square of the film thickness d .

According to Eq. (1.8) this leads to a very long propagation length of such a SPP mode, which is called a long-range SPP. In turn, this leads to a large field enhancement associated with this type of mode. In contrast, the imaginary part of the wavenumber of the low frequency mode (the solution of Eq. (1.22)) increases as the thickness d of the film is decreased, leading to a propagation length that decreases with decreasing d .

The physical reason for the significantly longer propagation length of the mode whose dispersion relation is given by Eq. (1.23) compared with the propagation length of the mode whose dispersion relation is given by Eq. (1.22) is that the tangential component of the electric field (E_1) of the former mode vanishes at the midplane of the film $x_3 = d/2$, while it is a maximum at that plane for the latter mode. The mode with the smaller fraction of its electric field inside the film has the longer propagation length because it interacts more weakly with the dissipative mechanisms in the film.

1.1.4. Localized surface plasmons

In addition to surface plasmon polaritons and surface plasmons at a planar dielectric–metal interface, localized surface electromagnetic excitations can exist in other geometries, such as metallic particles or voids of various topologies. Such surface excitations in bounded geometries are called localized surface plasmons (LSPs) [3,28]. The frequency of an LSP can be determined in the non-retarded (electrostatic) approximation by solving Laplace’s equation with suitable boundary conditions. The electrostatic approximation, which neglects the effects of retardation, is valid if the characteristic size a of a system is small compared to the wavelength λ corresponding to the LSP frequency, $a \ll \lambda$.

For example, in the case of a metallic sphere of radius R , centered at the origin, and embedded in a medium of dielectric constant ϵ_0 , the solution of Laplace’s equation for the electrostatic potential $\phi^<(r, \theta, \phi)$ inside the sphere that is finite at the origin can be written in the form

$$\phi^<(r, \theta, \phi) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} a_{\ell m} r^{\ell} Y_{\ell m}(\theta, \phi), \quad 0 \leq r \leq R, \quad (1.26)$$

where $Y_{\ell m}(\theta, \phi)$ is a spherical harmonic. Similarly, the solution of Laplace’s equation for the electrostatic potential $\phi^>(r, \theta, \phi)$ outside the sphere that vanishes at infinity can be written as

$$\phi^>(r, \theta, \phi) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} b_{\ell m} \frac{1}{r^{\ell+1}} Y_{\ell m}(\theta, \phi), \quad r \geq R. \quad (1.27)$$

The imposition of the boundary conditions at the surface of the sphere, namely the continuity of ϕ and of $\epsilon \partial \phi / \partial r$, yields the dispersion relation for the frequencies of the LSPs,

$$\frac{\epsilon(\omega)}{\epsilon_0} + \frac{\ell + 1}{\ell} = 0. \quad (1.28)$$

For a metal described by the free electron form of the dielectric function, Eq. (1.7), the solutions of Eq. (1.28) are given by

$$\omega_{\ell} = \omega_p \left[\frac{\ell}{\epsilon_0(\ell + 1) + \ell} \right]^{1/2}, \quad \ell = 1, 2, 3, \dots \quad (1.29)$$

For small spheres, only the dipole-active excitation ($\ell = 1$) is important. As the size of the sphere increases, the contributions of higher multipoles become more and more significant, with the result that in the limit

of a very large sphere ($\ell \rightarrow \infty$) the LSP frequency approaches that of a surface plasmon at a planar dielectric–metal interface, $\omega_\infty = \omega_p/(\epsilon_0 + 1)^{1/2}$. For an ellipsoidal particle, in general there are three sets of localized surface plasmon frequencies related to the three axes of an ellipsoid [28]. The three frequencies of the dipole-active LSPs in small ellipsoids can be calculated from $\epsilon(\omega)/\epsilon_0 = 1 - 1/L_i$, where L_i is the so-called depolarization factor depending only on the ellipsoid shape ($i = 1, 2, 3$ denotes the coordinate axes linked to the ellipsoid axes) [28]. In the case of a sphere $L_i = 1/3$, and Eq. (1.28) holds. The range of frequencies where localized surface plasmon modes exist is bounded from above by the bulk plasma frequency of the metal, i.e. it is the frequency range in which the dielectric function of the metal is negative.

It should be emphasized that the surface plasmon polaritons considered in the preceding sections are a different type of excitation than localized surface plasmons. A surface plasmon polariton has the dispersion relation $\omega_{sp} = \omega(k_{sp})$ given by Eq. (1.6) and is a propagating surface mode. In contrast, localized surface plasmons (also called electromagnetic surface shape resonances) are confined to curved metal objects. They are characterized by discrete, complex frequencies, which depend on the size and shape of the object to which the surface plasmon is confined, and by its dielectric function. Localized surface plasmons can be resonantly excited with light of appropriate frequency (and polarization) irrespective of the wave vector of the exciting light. Therefore, they also effectively decay with the emission of light. In contrast, an SPP mode can be excited only if both the frequency and wavevector of the exciting light match the frequency and wavevector of the SPP.

Localized surface plasmons can be also assigned to features on a metal surface. As will be discussed below, LSP resonances play a significant role in the behavior of SPP on rough surfaces if their frequency is close to the SPP frequency. LSPs can decay into surface plasmon polaritons and, in turn, can be excited by SPP. This results in a significant enhancement of SPP scattering by surface defects if the frequency of SPP and the resonant frequency of LSP are close to each other [29]. On the other hand, SPP excitation is an efficient channel of LSP decay on a metal surface [30].

Since a LSP is confined to a particle (or a curved surface), it results in a significant electromagnetic field enhancement at small metallic particles due to the small volume of the LSP mode [31]. This effect contributes to numerous phenomena such as light emission from STM tunnel junctions, enhanced scattering, surface enhanced Raman scattering, and second-harmonic generation, and finds applications in active photonic elements and apertureless scanning near-field microscopy [5,31–37].

In the case of voids in the bulk of a metal, the corresponding localized surface plasmon frequencies can be found by replacing $\epsilon(\omega)$ by $1/\epsilon(\omega)$ in the corresponding Dirichlet problem. The localized surface plasmon frequencies of a particle and a void of the same shape are related to each other by

$$\omega_{\text{particle}}^2 + \omega_{\text{void}}^2 = \omega_p^2 . \quad (1.30)$$

Thus, the resonances associated with surface plasmons localized on voids in a metal can be estimated from known LSP frequencies on metal particles.

The spectrum of localized surface plasmons associated with an ensemble of metallic particles (or voids) is determined by the interaction between the individual LSP resonances (cf. the modes of a thin metal film). The resulting spectrum and the magnitude of the electromagnetic field enhancement depend significantly on the shape and size of the individual particles and on the distance between them. In such ensembles of interacting small metallic particles a very strong electromagnetic field enhancement can be observed [35]. A similar but simpler situation occurs in the case of a metallic particle situated above a

metal surface [33,34]. Localized surface plasmons associated with this geometry are sometimes called gap plasmons.

Another interesting type of surface plasmon excitations sharing properties of both LSP and surface plasmon polaritons occurs in long metallic cylinders (or cylindrical channels) which have two significantly different characteristic sizes in different directions (usually, the diameter $a \ll \lambda$, so that the electrostatic approximation can be used, and the length $H \gg \lambda$, so that retarded modes should be invoked). Such cylindrical surface plasmons have a discrete spectrum of resonances asymptotically approaching the surface plasmon frequency for high mode numbers [3]. For infinitely long cylinders, the spectrum of wavevectors along the cylinder axis is continuous. Physically one can imagine these excitations as surface modes with a spiral trajectory on the surface of a cylinder. In general, the dispersion of these surface modes has both radiative and nonradiative branches for different frequencies, except for the lowest purely nonradiative mode.

1.1.5. Optical excitation of surface plasmon polaritons

In order to excite a surface plasmon polariton by p-polarized light incident on a planar metal surface from the adjacent dielectric medium, the frequency of the incident light must equal the frequency of the SPP, and the component of the wave vector of the incident light parallel to the surface, $\sqrt{\epsilon_1}(\omega/c) \sin \theta$ where θ is the angle of incidence, must equal the wavenumber of the SPP $(\omega/c)[\epsilon_1\epsilon(\omega)/(\epsilon(\omega) + \epsilon_1)]^{1/2}$. The first condition is easily satisfied. However, as is seen from the SPP dispersion relation, the SPP wavenumber is larger than the magnitude of the wave vector of the light in the adjacent dielectric medium. Thus, light illuminating a metal surface through that medium cannot directly couple to surface plasmon polaritons. Therefore, special experimental arrangements have been designed to provide the necessary wavevector conservation. The photon and SPP wavevectors can be matched by using either photon tunneling in the total internal reflection geometry (Kretschmann and Otto configurations) or diffraction effects (Fig. 5).

In the Kretschmann configuration (Fig. 5a), a metal film is illuminated through a dielectric prism at an angle of incidence greater than the critical angle for total internal reflection [38]. The wavevector of light is increased in the optically more dense medium. At the angle of incidence θ at which the in-plane component of the photon wavevector in the prism coincides with the SPP wavevector at the air–metal interface, resonant light tunneling through the metal film occurs, and light is coupled to surface polaritons:

$$k_{\text{sp}} = \frac{\omega}{c} \sqrt{\epsilon_{\text{prism}}} \sin \theta . \quad (1.31)$$

Under these resonant conditions, a sharp minimum is observed in the reflectivity from the prism–metal interface as light can be coupled to SPPs with almost 100% efficiency. Since the SPP field is concentrated close to a metal surface, it is significantly enhanced at the surface. This gives, for example, an intensity enhancement by more than two orders of magnitude for a 60 nm thick silver film illuminated with red light [2]. Comparing the SPP intensity to the intensity of the light transmitted at normal incidence, the enhancement is more than three orders of magnitude. With the increase of the metal film thickness the efficiency of the SPP excitation (and the field enhancement) decreases as the tunneling distance increases. The SPP on the interface between the prism and the metal cannot be excited in this geometry as the wavevector of the SPP at this interface is greater than the photon wavevector in the prism for all angles of incidence. To be able to excite SPP on the internal metal interface, an additional dielectric layer with a refractive index smaller than the one of the prism should be deposited between the prism and the metal film (Fig. 5b). In such a two-layer geometry, the photon tunneling through this

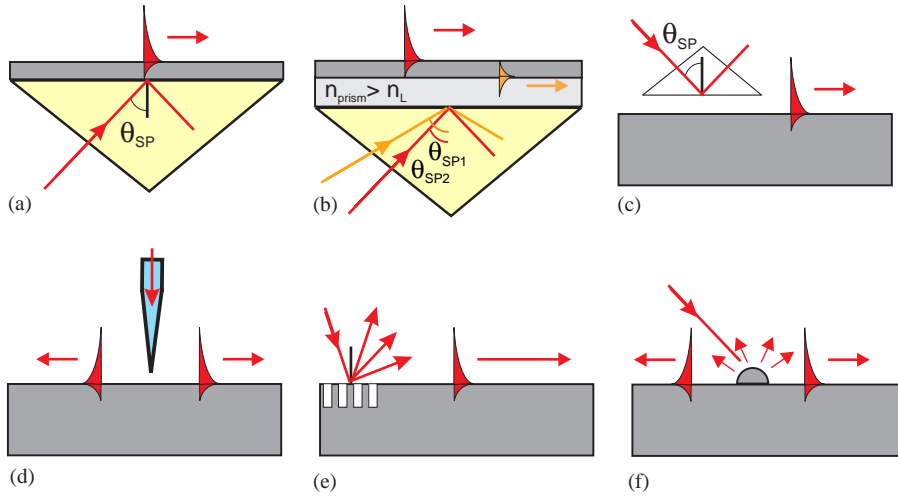


Fig. 5. SPP excitation configurations: (a) Kretschmann geometry, (b) two-layer Kretschmann geometry, (c) Otto geometry, (d) excitation with a SNOM probe, (e) diffraction on a grating, and (f) diffraction on surface features.

additional dielectric layer can provide resonant excitation of SPP on the inner interface. Thus, both SPP modes (on the surface and the interface) can be excited in such a configuration, at different angles of illumination.

For thick metal films (or surfaces of bulk metal), for which the Kretschmann configuration cannot be used, SPP can be excited in the Otto configuration (Fig. 5c). Here, the prism where the total internal reflection occurs is placed close to the metal surface, so that photon tunneling occurs through the air gap between the prism and the surface [39]. The resonant conditions are analogous to the ones in the Kretschmann configuration (Eq. (1.31)).

Another way to provide the wavevector conservation for SPP excitation is to use diffraction effects. If a diffraction grating is created on part of an otherwise smooth metal film, components of the diffracted light whose wavevectors coincide with the SPP wavevector will be coupled to surface polaritons (Fig. 5e). Diffraction on a periodic structure provides the wavevector conservation and coupling to surface polaritons [2]:

$$\mathbf{k}_{\text{SPP}} = \frac{\omega}{c} n_s \sin \theta \mathbf{u}_{12} \delta_p \pm p \frac{2\pi}{D} \mathbf{u}_1 \pm q \frac{2\pi}{D} \mathbf{u}_2, \quad (1.32)$$

where $\delta_p = 1$ for p-polarized incident light (with respect to a surface plane) and 0 for s-polarized light, \mathbf{u}_{12} is the unit vector in the direction of the in-plane component of the wavevector of the incident light, n_s is the refractive index of the medium through which the film is illuminated, \mathbf{u}_1 and \mathbf{u}_2 are the unit lattice vectors of a periodic structure, D is its period (assumed to be the same in both the x_1 - and x_2 -directions), and p and q are integer numbers corresponding to the different propagation directions of the excited SPPs. Such an excitation configuration can provide efficient coupling to both air–metal and substrate–metal SPP modes of a metal film if the film thickness and grating profile depth are suitably related.

On a randomly rough surface, the SPP excitation conditions can be achieved without any special arrangements. This is possible since in the near-field region diffracted components of light with all wavevectors are present [1,2,40]. Thus, SPP can be excited on conventionally illuminated rough surfaces.

The problem with random roughness is the ill-defined SPP excitation conditions resulting in a low efficiency of light-to-SPP coupling. This is a nonresonant excitation, and there is a strong presence of the reflected exciting light close to the surface. As with a diffraction grating, depending on the metal film thickness and the depth of the defect, SPP can be excited on both interfaces of a film. Such nonresonant SPP excitation processes result in a complex field distribution over the surface due to the interference of the SPPs excited on different interfaces of the film and the non-coupled illuminating light [40].

Yet another approach to the optical excitation of SPP is through diffraction of light from surface features (Fig. 5f). An example of SPP excitation by the diffraction of p-polarized light from the two edges of a metal strip of finite width on the base of a glass prism through which the light is incident is presented in Fig. 6. This is the Kretschmann configuration, Fig. 5a, with the difference that the metal film (strip) covers only a portion of the base of the prism. The SPP are excited at both edges of the metal strip on both of its interfaces, and propagate in the direction of the illuminating light (from the left edge) as well as in the opposite direction (from the right edge). As no light coupling into SPPs takes place at a smooth part of the surface under nonresonant conditions, there is a strong presence of the evanescent field of the exciting light over the metal. Mutual interference of different field components results in the observed near-field distribution [40]. In contrast, under resonant excitation conditions, the air–metal surface polaritons are excited not only close to the edges of the strip but all over the metal. At the same time, the glass–metal SPPs are excited only due to diffraction on the metal discontinuities. The interference pattern between different SPP modes is clearly visible. Under resonant excitation, virtually all of the incident light is coupled into SPPs leading to a much higher intensity over the film than under nonresonant excitation.

The application of SNOM to SPP studies introduced a new technique that provides the possibility to excite SPP locally at a given place on a surface [19]. Using illumination through a SNOM fiber tip (Fig. 5d), circular SPP waves can be locally launched at the surface. Ideologically, this configuration can be treated as either a diffraction or a tunneling mechanism of SPP excitation. In the former case, one can consider a near-field coupling of the light diffracted on the subwavelength aperture of the fiber tip into SPPs. In the latter case, one can describe the process as photon tunneling taking place from the fiber tip to the metal surface, as mainly the evanescent field components are generated at the aperture. This technique is somewhat analogous to the Otto configuration but allows local (with sub-wavelength precision) excitation of SPPs at the position of the SNOM tip.

Due to the longitudinal nature of the surface plasmon polariton field, in all excitation configurations discussed above, the exciting light should have a component of its electric field perpendicular to the metal surface or parallel to the propagation direction of the excited SPPs (Fig. 6).

1.2. *Scanning near-field optical microscopy*

Only with the development of scanning probe techniques did it become possible to study SPP properties directly at the surface along which the SPP propagates with resolution at the nanometer scale. The first scanning probe technique applied to the investigation of SPPs was scanning tunneling microscopy (STM), relying on the detection of the additional tunneling current induced by surface polaritons [41–44] or the far-field scattered light due to the local SPP interaction with a STM tip [45]. In another approach, a dielectric SiN probe of an atomic force microscope (AFM) was used instead of a STM tip [46,47]. It was realized later that although such approaches provide in the first approximation information on the SPP field at the place of a probe tip, the metal or silicon tips introduce significant perturbations in the SPP field

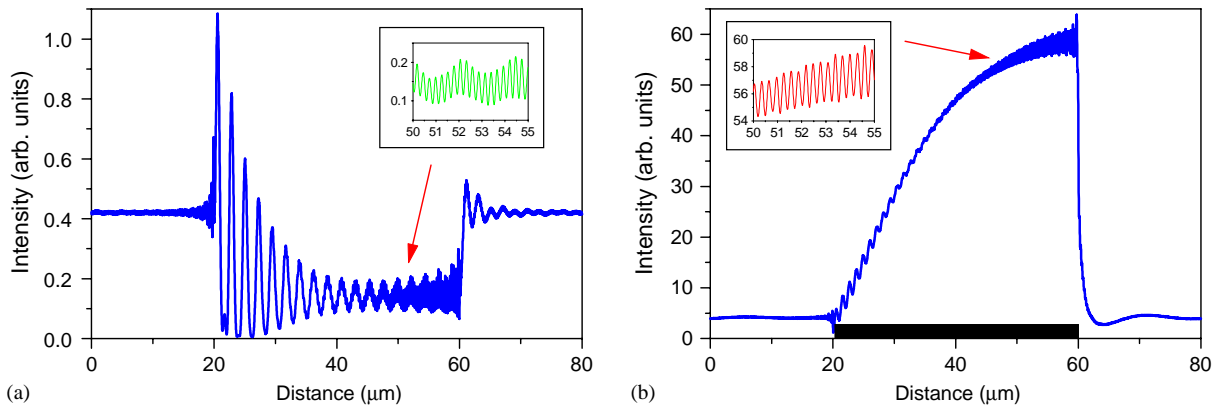


Fig. 6. Calculated distributions of the electromagnetic field intensity at the distance of 50 nm over a gold strip on a glass substrate illuminated in the Kretschmann configuration for different angles of incidence: (a) 65° (nonresonant excitation) and (b) $\theta_{\text{SP}} = 45.7^\circ$ (resonant excitation). The strip is situated between 20 and 60 nm as shown in (b). (Ref. [40].)

due to the field enhancement effects related to localized surface plasmons and the lightning-rod effect (geometrical field enhancement at a highly curved surface) at the tip–surface junction [48,49]. These effects, depending in a complex nonlinear manner on the tip and surface defect topology and size, as well as on their mutual position and orientation, prevent the direct measurements of the local SPP field on a surface.

Scanning near-field optical microscopy with uncoated optical fiber tips provides the possibility to probe the surface polariton field directly over a surface [9]. This is done by scanning the sample with a small aperture formed at the extremity of a sharpened glass fiber (Fig. 7). The low dielectric constant of glass (compared to silicon or a metal) and a distance between a probe tip and a surface that is more than two orders of magnitude larger than in STM ensure that the SPP field is perturbed as little as possible.

The fiber tip is usually scanned at a fixed distance above the local sample surface within a few tens of nanometers. This provides simultaneous topographical and optical imaging and allows to directly correlate the optical field distribution over the surface to its topography. To achieve tip–surface distance control during SNOM measurements, a variant of atomic force microscopy is usually implemented, called shear-force microscopy [50,51]. Here, the fiber tip is vibrated laterally and the amplitude of the vibration is monitored optically or electrically. As the tip–sample distance decreases, the interaction between a tip and a surface reduces the vibration amplitude of the tip. This provides the necessary signals to drive a control system that keeps the tip–surface distance constant. Two other SNOM operation modes have also been used sometimes for SPP studies. In the so-called constant intensity mode, the tip is scanned above a surface so that the optical signal is maintained constant. Thus a profile of a surface of constant intensity is recorded. In the constant distance mode, the tip is scanned at the fixed distance from the average surface without following the local topography variations. Although sometimes useful, these two SNOM modes provide only optical information that in many cases is difficult to correlate with the unknown topography of a surface.

It has been traditional to overcoat a fiber tip with a metal layer and to leave an aperture of 50–100 nm size at the tip apex [52]. This delineates in an apparently clear way the region where evanescent fields can couple to and from a surface. However, studies of metal surfaces using metal coated fiber tips are prone to

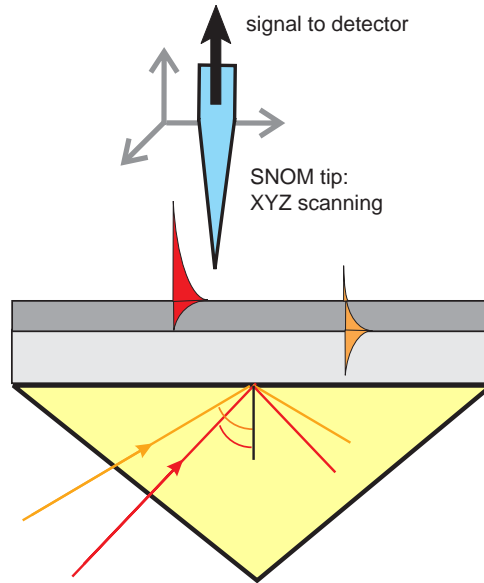


Fig. 7. Schematic of a SNOM experimental set-up for studying surface polaritons.

problems related to a strong tip–surface interaction that can significantly modify the local electromagnetic field. The use of metal coated fibers in the near-field proximity to a metal surface results in a significant perturbation of the electromagnetic field due to the tip–surface interaction, as in the case of the STM and AFM based optical measurements discussed above. In this case, the detected signal is related to LSP and the dielectric properties of both tip and surface, rather than to the SPP field above a surface. Moreover, reflection of SPP from the tip in addition to surface defects, creates another type of artifact depending on the relative position of the SNOM tip with respect to surface features.

In contrast, bare (uncoated) fiber tips introduce a much smaller perturbation in the measured electromagnetic field. Uncoated fiber tips with a relatively low refractive index can be considered as nonperturbative probes. The signal detected with such a probe will be closely proportional to the near-field intensity [53–55]. The deviation from this behavior will increase with the increase of the tip dielectric constant. It has generally been assumed that light leakage from bare tapers, which should occur in the far field when the mode-field diameter becomes commensurate with the tip diameter, would lead the spatial resolution to be $\sim \lambda/3$. However, this neglects the predominance of near-field coupling that occurs when a bare fiber tip is used to collect light close to a surface [56]. The resolution of the SPP mapping obtained with an uncoated fiber tip is related to the gradient of the SPP evanescent field decay above a metal surface. The resolution of SNOM mapping of the SPP field with uncoated fiber tips routinely reaches about 100 nm at the 633 nm wavelength of the detecting light [57]. This is significantly better than the resolution obtained with uncoated fiber tips in reflection/transmission measurements where propagating field components are dominating.

The efficiency of the evanescent field detection is related to the opening angle of the probe tip and, from a practical point of view, its quality (roughness and inhomogeneities of the tip apex). Generally, for the ideal conical tip, the electric field components polarized perpendicularly to the SNOM tip axis

(parallel to the surface plane) are more strongly coupled to the fiber tip than the field with the polarization parallel to the probe tip. The SPP field has components both in the direction perpendicular to and in the surface plane. The former field component is dominant for the SPP wave vectors close to the light wave vector (the contribution of the in-plane polarization increases for large wave vectors). The detected signal related to the field component normal to the surface increases for small tip opening angles as the fourth power of the angle [53]. Discrimination of the propagating waves above a surface, which results from the SPP scattering out of the surface plane, against the evanescent field of the excited and scattered (in the surface plane) SPPs is also much more efficient with probe tips with wider opening angles. The propagating (in air) field contains components parallel to the surface plane, and the detected signal related to these components increases with the cone angle as the square of the angle of tip opening (for small opening angles) [53]. Consequently, the relative contribution of the perpendicular (to the surface plane) field components in the detected signal increases with the cone angle, implying an increase of the relative efficiency of the SPP-related signal compared to scattered light with the increase of the angle of a tip opening. It should be noted that both orthogonally polarized SPP field components can be efficiently coupled to the fiber tip due to scattering and depolarization effects on the inhomogeneities of the tip apex, which are often present in real experiments.

A typical SNOM apparatus for SPP measurements is shown in Fig. 7. In this set-up an uncoated fiber tip is used to detect the evanescent field of the resonantly excited surface polaritons above the sample surface. For example, in the case of a two-layer structure at which two SPP modes can be excited at different angles of incidence, the angle of incidence can be tuned to excite the air–metal SPP on the surface of a film, and the topography and the field distribution can be recorded with a SNOM. Then, the angle of incidence can be changed to excite the glass–metal SPP on the inner interface, and the corresponding field distribution can be recorded. In this way the SPP field over a surface can be directly correlated to its topographical structure measured simultaneously by shear-force feedback.

2. Surface polariton scattering

Surface polariton scattering by surface features is a basis for SPP applications. It plays a crucial role in determining optical properties of rough and nanostructured metal films, and is a key for understanding all SPP related processes. The SPP interaction with a surface defect can be described by three processes, namely, scattering of SPP into SPP propagating in another direction (or SPP reflection), propagation of SPP through the defect region in the same direction as the incoming SPP (SPP transmission), and scattering of SPP into light (Fig. 8). In the past, using conventional far-field measurements, it has only been possible to study the latter process. All investigations of surface polaritons have been based on such far-field observations of the intensity and angular distribution of the light scattered due to the interaction of SPP with surface features [1]. It was possible to modify the surface structure in one or another (generally unknown) way and to follow the changes in the scattered light behavior without knowledge of the SPP scattering processes on a surface. With the development of SNOM these in-plane scattering processes can be directly studied by measuring the electromagnetic field distribution related to the surface polaritons in the near-field region over a surface.

The interplay between in-plane and out-of-plane scattering processes determines the character of the electromagnetic field distribution over a surface and, therefore, surface optical properties and the field enhancement. The scattering processes depend on the size, geometrical shape, and the dielectric constant

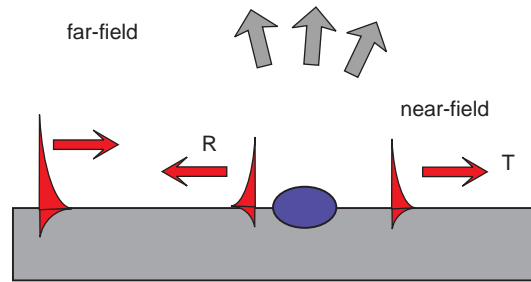


Fig. 8. Schematic of the SPP scattering processes on a topographical surface feature and/or a dielectric constant variation.

of the surface features (Section 2.1). The surface polariton scattering is a complex phenomenon which in most cases (except for very simple geometries) requires tedious numerical modeling. Similar to all wave scattering processes, the SPP scattering depends on the relation between a defect size and the SPP wavelength. Nevertheless, since a surface polariton is an intrinsically two-dimensional excitation, two characteristic defect sizes should be considered: one characterizing the lateral (in-plane) size of the defect and another characterizing the defect height (out-of-plane size). To understand SPP scattering processes, these characteristic sizes of scatterers should be related to respective in-plane and out-of-plane characteristics of the surface polariton wave: SPP wavelength and SPP field extension from the surface, respectively. In addition, the gradient of the topography variations should be taken into account. Naturally, for a slowly varying topography the SPP scattering will be less important. Thus, the efficiency of the in-plane and out-of-plane SPP scattering by surface features depends in a nontrivial way on the lateral size, height (or depth) of a defect through their relations to the SPP wavelength, and on the SPP field extension length above (or below) a surface.

For the in-plane SPP scattering on a randomly rough surface that can be considered as a defect ensemble, another parameter is of importance, namely, the characteristic distance between the defects l . The ratio of l and the SPP propagation length L_{SP} is crucial for determining the nature of the scattering: a single scattering regime for $L_{SP} \lesssim l$ and a multiple scattering regime for $L_{SP} \gg l$.

The surface polariton behavior in the different regimes of scattering by different types of surface defects has been studied with SNOM [17–20,58–64]. To illustrate different possible types of scattering, we consider gold and silver films with different types of defect structures providing a wide range of ratios between SPP and defect parameters. We begin with the investigations of the surface polariton scattering by individual defects artificially fabricated with a near-field lithography on smooth silver films. The single scattering regime has also been studied on randomly rough gold films with the defect structure on two distinctively different size scales: large scale defects can be treated as individual defects in the single scattering regime. Various combinations of the defect parameters and SPP parameters on gold and silver surfaces allowed to identify several different SPP scattering processes on a surface, such as shadow-like scattering, coherent 2D forward and backward scattering, and SPP reflection followed by the interference of reflected and incident SPPs. The multiple scattering regime has been achieved on both gold and silver films with small scale roughness. The multiple beam interference leads in this case to the effects related to a defect ensemble, such as SPP localization and backscattering. Additional features of SPP interaction with surface defects arise when the feature size corresponds to LSP resonance at the excitation wavelength, leading to resonant LSP-to-SPP and vice versa transformations and light confinement at surface defects.

Different types of SPP interaction with surface features open up the possibility to manipulate and direct the SPP beams on a surface in the same manner as light beams are manipulated in 3D. Various types of elements of SPP optics can be created employing appropriate surface structures such as mirrors, lenses, prisms, resonators, etc. This allows to develop 2D optics of surface polaritons that can find numerous applications in nanophotonics and related technologies.

2.1. Single scattering regime: the SPP interaction with individual defects

The electromagnetic field above a surface is determined by SPP scattering processes on surface features. In the absence of a surface structure (perfectly flat surface), the measured intensity distribution would be constant. The SPP scattering leads to generally complex intensity variations across a surface due to the interference of scattered surface polaritons. An understanding of SPP scattering processes on topography variations, subsurface defects, and refractive index variations is important for the interpretation of near-field images as well as for applications of SNOM imaging for characterization of metal surfaces and thin films. In this section we present a general theory of the SPP scattering by one- and two-dimensional surface defects and discuss SNOM measurements of SPP scattering on individual surface defects.

2.1.1. The interaction of a surface plasmon polariton with a single surface defect

When a surface plasmon polariton propagating along an otherwise planar vacuum–metal interface impinges on a surface defect that is localized in one or more spatial directions, it is partially scattered into other surface plasmon polaritons, and partially into volume electromagnetic waves in the vacuum region above the metal propagating away from the interface. The scattering of surface plasmon polaritons by surface defects has been studied theoretically both in the case of one-dimensional defects (grooves and ridges) and in the case of point defects. In this section we consider both of these cases in turn.

a. One-dimensional surface defects. The scattering of surface plasmon polaritons by one-dimensional surface defects is conveniently carried out on the basis of an impedance boundary condition [65,66]. Thus, let us consider a system consisting of vacuum in the region $x_3 > \zeta(x_1)$, and a metal characterized by an isotropic, frequency-dependent, dielectric function $\epsilon(\omega)$ in the region $x_3 < \zeta(x_1)$. The surface profile function $\zeta(x_1)$ is assumed to be a single-valued function of x_1 that is differentiable as many times as is necessary, and is sensibly nonzero only for $|x_1| < R$. A p-polarized surface plasmon polariton of frequency ω is assumed to be propagating across this vacuum–metal interface in the x_1 -direction. Its sagittal plane is the x_1x_3 -plane. In the vacuum region $x_3 > \zeta(x_1)$ the single nonzero component of the magnetic field has the form $H_2(x_1, x_3; t) = H_2^>(x_1, x_3|\omega) \exp(-i\omega t)$, where the amplitude $H_2^>(x_1, x_3|\omega)$ satisfies the Helmholtz equation

$$\left(\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_3^2} + \frac{\omega^2}{c^2} \right) H_2^>(x_1, x_3|\omega) = 0. \quad (2.1)$$

The solution of this equation is written as the sum of an incident surface plasmon polariton and a scattered field,

$$H_2^>(x_1, x_3|\omega) = \exp[ik(\omega)x_1 - \beta_0(\omega)x_3] + \int_{-\infty}^{\infty} \frac{dq}{2\pi} R(q, \omega) \exp[iqx_1 + i\alpha_0(q)x_3], \quad (2.2)$$

where

$$k(\omega) = k_1(\omega) + ik_2(\omega) = \frac{\omega}{c} \left(1 - \frac{1}{\epsilon(\omega)} \right)^{1/2}, \quad (2.3)$$

$$\beta_0(\omega) = \left[k(\omega)^2 - \frac{\omega^2}{c^2} \right]^{1/2} = \frac{\omega}{c} [-\epsilon(\omega)]^{-1/2}, \quad (2.4)$$

and

$$\alpha_0(q) = \left(\frac{\omega^2}{c^2} - q^2 \right)^{1/2}, \quad |q| < \frac{\omega}{c}, \quad (2.5a)$$

$$= i \left(q^2 - \frac{\omega^2}{c^2} \right)^{1/2}, \quad |q| > \frac{\omega}{c}. \quad (2.5b)$$

In writing Eq. (2.2) we have assumed a time dependence of $\exp(-i\omega t)$, but have not indicated it explicitly. The superscript $>$ in Eq. (2.2) denotes the vacuum region. We note that the components of the wave vector of the surface plasmon polariton in the vacuum region $k(\omega)$ and $\beta_0(\omega)$ have been written in the forms that apply in the limit that $|\epsilon(\omega)| \gg 1$. This is due to the fact that the exact boundary condition satisfied by $H_2^>(x_1, x_3|\omega)$ at the interface $x_3 = \zeta(x_1)$ is replaced by the impedance boundary condition [67]

$$\left. \frac{\partial}{\partial x_3} H_2^>(x_1, x_3|\omega) \right|_{x_3=0} = -\frac{\omega}{c} \frac{1 + s(x_1)}{[-\epsilon(\omega)]^{1/2}} H_2^>(x_1, x_3|\omega) \Big|_{x_3=0} \quad (2.6)$$

on the plane $x_3 = 0$, and this replacement is valid when $|\epsilon(\omega)| \gg 1$. The surface impedance function $s(x_1)$ is related to the actual surface profile function $\zeta(x_1)$ by [68]

$$s(x_1) = \frac{1 - \epsilon(\omega)}{d(\omega)\epsilon(\omega)} [1 - d^2(\omega)D^2]^{1/2} \zeta(x_1) + O(\zeta^2), \quad (2.7)$$

where $d(\omega) = (c/\omega)[- \epsilon(\omega)]^{-1/2}$ is the optical skin depth of the metal at the frequency ω , and $D \equiv d/dx_1$. Equations (2.6) and (2.7) have been shown to yield accurate quantitative results in calculations of grating-induced surface plasmon polariton–photon coupling [65]. When the solution (2.2) is substituted into the boundary condition (2.6), the resulting integral equation for the scattering amplitude $R(q, \omega)$ can be written in the form

$$R(q, \omega) = G_0(q, \omega)V(q|k(\omega)) + G_0(q, \omega) \int_{-\infty}^{\infty} \frac{dp}{2\pi} V(q|p)R(p, \omega), \quad (2.8)$$

where

$$G_0(q, \omega) = \frac{i\epsilon(\omega)}{\epsilon(\omega)\alpha_0(q) + i(\omega/c)[- \epsilon(\omega)]^{1/2}} \quad (2.9)$$

is the surface plasmon polariton Green's function at the unperturbed interface ($s(x_1) \equiv 0$). It should be noted that $G_0(q, \omega)$ has simple poles at the wavenumbers of the surface plasmon polariton, $q = \pm k(\omega)$. The scattering potential $V(q|k)$ in Eq. (2.8) is given by

$$V(q|p) = \beta_0(\omega)\hat{s}(q - p), \quad (2.10)$$

with

$$\hat{s}(Q) = \int_{-\infty}^{\infty} dx_1 s(x_1) \exp(-iQx_1) . \quad (2.11)$$

Equation (2.8) can be rewritten in a more convenient form if we write

$$R(q, \omega) = G_0(q, \omega)T(q, \omega) . \quad (2.12)$$

The equation for $T(q, \omega)$ is then

$$T(q, \omega) = V(q|k(\omega)) + \int_{-\infty}^{\infty} \frac{dp}{2\pi} V(q|p)G_0(p, \omega)T(p, \omega) . \quad (2.13)$$

Once Eq. (2.13) has been solved for $T(p, \omega)$, it follows from Eqs. (2.2) and (2.12) that the amplitude of the transmitted surface plasmon polariton in the region $x_1 \gg R$ is

$$t(\omega) = 1 + iC(\omega)T(k(\omega), \omega) , \quad (2.14)$$

while the amplitude of the reflected surface plasmon polariton in the region $x_1 \ll -R$ is

$$r(\omega) = iC(\omega)T(-k(\omega), \omega) , \quad (2.15)$$

where

$$C(\omega) = \frac{\omega}{c} \frac{1}{[-\epsilon(\omega)]^{1/2}k(\omega)} \quad (2.16)$$

is the residue of $G_0(q, \omega)$ at the poles it possesses at $q = \pm k(\omega)$. The corresponding surface plasmon polariton reflection and transmission coefficients are then

$$R(\omega) = |r(\omega)|^2 , \quad (2.17a)$$

$$T(\omega) = |t(\omega)|^2 . \quad (2.17b)$$

The total power carried away from the surface in the form of volume electromagnetic waves propagating in the volume region above it, per unit length of the system along the x_2 -axis is

$$P_{sc} = \int_{-\infty}^{\infty} dx_1 \langle S_3^{sc} \rangle = \frac{c^2}{8\pi\omega} \int_{-\frac{\omega}{c}}^{\frac{\omega}{c}} \frac{dq}{2\pi} \alpha_0(q) |R(q, \omega)|^2 , \quad (2.18)$$

where $\langle S_3^{sc} \rangle$ is the time average of the 3-component of the Poynting vector of the scattered field. The expression given by Eq. (2.18) must be normalized by the power carried by the incident surface plasmon polariton per unit length of the system along the x_2 -axis:

$$P_{inc} = \int_0^{\infty} dx_3 \langle S_1^{in} \rangle = \frac{c^2 k(\omega)}{16\pi\omega\beta_0(\omega)} , \quad (2.19)$$

where $\langle S_1^{in} \rangle$ is the time average of the 1-component of the Poynting vector of the incident field. Thus, the total, normalized scattered power $S(\omega)$ is given by

$$S(\omega) = \frac{P_{sc}}{P_{in}} = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} d\theta_s \frac{\partial R}{\partial \theta_s} , \quad (2.20)$$

where

$$\frac{\partial R}{\partial \theta_s} = \frac{1}{2\pi} \frac{\beta_0(\omega)}{2k(\omega)} \left(\frac{\omega}{c}\right)^2 \cos^2 \theta_s \left| R\left(\frac{\omega}{c} \cos \theta_s, \omega\right) \right|^2 \quad (2.21)$$

is the differential reflection coefficient, namely $(\partial R / \partial \theta_s) d\theta_s$ is the fraction of the incident flux that is scattered into the angular region of width $d\theta_s$ about the scattering direction θ_s , where the scattering angle θ_s is measured clockwise from the x_3 -axis.

Equation (2.13) has been solved numerically for the case of a surface profile function given by the Gaussian function $\zeta(x_1) = h \exp(-x_1^2/a^2)$ [65,66]. Thus, the surface defect is a ridge if h is positive, and a groove if h is negative. The Fourier transform of the impedance function $s(x_1)$ defined by Eq. (2.11) becomes

$$\hat{S}(Q) = \frac{1 - \epsilon(\omega)}{d(\omega)\epsilon(\omega)} [1 + d^2(\omega)Q^2]^{1/2} \hat{\zeta}(Q), \quad (2.22)$$

where

$$\begin{aligned} \hat{\zeta}(Q) &= \int_{-\infty}^{\infty} dx_1 \exp(-iQx_1) \zeta(x_1) \\ &= \pi^{1/2} h a \exp\left(-\frac{1}{4} Q^2 a^2\right). \end{aligned} \quad (2.23)$$

In the numerical calculations carried out in Refs. [65,66] the square root entering Eq. (2.22) was replaced by unity, an approximation that had virtually no effect on the results obtained for the values of ω , h , and a used in these calculations.

In these calculations it was further assumed that the dielectric function of the metal was real, and had the simple free electron form $\epsilon(\omega) = 1 - (\omega_p^2/\omega^2)$, where ω_p is the plasma frequency of the conduction electrons. This is a realistic assumption because the mean free path of surface plasmon polaritons whose wavelength is that of a He–Ne laser, $\lambda = 632.8$ nm, is more than $20 \mu\text{m}$, which is much larger than the linear dimensions of the defects we consider, and increases as the wavelength increases. However, to define how the poles of $G_0(p, \omega)$ at $p = \pm k(\omega)$ were to be treated in evaluating the integral in Eq. (2.13), $\epsilon(\omega)$ was given an infinitesimal positive imaginary part that was then set equal to zero after the contribution from the poles had been calculated.

The integral equation (2.13) was solved by converting it into a matrix equation. This was done in two steps. First, the infinite limits of integration were replaced by large but finite limits. The p -axis was then divided into a set of equally spaced points at which a numerical quadrature scheme was employed. This discretization mesh was chosen in such a way that $p = \pm k(\omega)$ were always on this mesh, as required by Eqs. (2.14) and (2.15). In addition, the discretization mesh was not uniform: the density of points around the poles at $p = \pm k(\omega)$ was considerably larger ($\Delta p \approx 10^{-4}(\omega/c)$) than it was in the radiative region $|p| \leq \omega/c$ or in the nonradiative region away from the poles ($\Delta p \approx 10^{-2}(\omega/c)$). The number N of p points used in the numerical calculations depended not only on the accuracy required to sample the pole regions, but also on the form of the surface defect, which enters the calculations through its Fourier transform in Eq. (2.11). The value of N used in these calculations ranged from $N = 2600$ for narrow defects up to $N = 4000$ for the widest defects. The convergence of the numerical results with increasing N was checked in the most unfavorable cases.

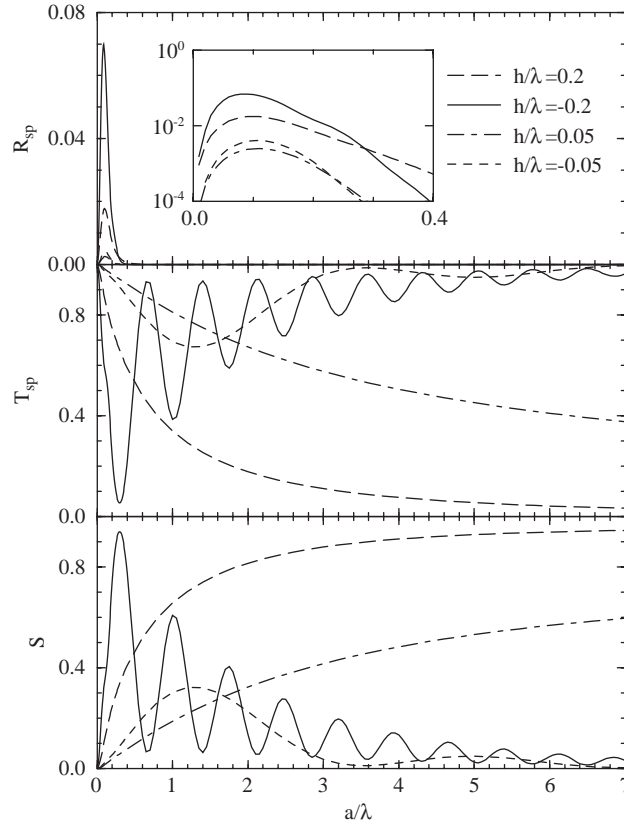


Fig. 9. Surface plasmon polariton reflection and transmission coefficients $R(\omega)$ and $T(\omega)$, respectively, and the total normalized scattered power $S(\omega)$ as functions of the Gaussian surface defect half-width: $\hbar\omega = 1.96$ eV ($\lambda = 632.8$ nm) and $\epsilon(\omega) = -17.2$. Long dashed curve: $h = 0.2\lambda$; solid curve: $h = -0.2\lambda$; dot-dashed curve: $h = 0.05\lambda$; dashed curve: $h = -0.05\lambda$. The inset shows the reflection coefficient in a semi-logarithmic scale for narrow defects. (After Ref. [66].)

In the numerical calculations it was assumed that the frequency of the incoming surface plasmon polariton was given by $\hbar\omega = 1.96$ eV, which corresponds to the vacuum wavelength $\lambda = 632.8$ nm of the light used in PSTM experiments [13,15,58,59] to excite surface plasmon polaritons on silver surfaces ($\epsilon(\omega) = -17.2$).

In Fig. 9 we have plotted the dependence on a/λ of the coefficients $R(\omega)$, $T(\omega)$, and $S(\omega)$ for Gaussian ridges and grooves of amplitudes $h/\lambda = 0.05$ and 0.2 . The conclusions drawn from these results can be summarized as follows. First, the reflection of surface plasmon polaritons is significant only for very narrow surface defects, $a < \lambda/5$, for either ridges or grooves. In fact, there is an optimum defect width for which $R(\omega)$ is maximal [65]. For the Gaussian profile it is given by $a_{opt} = (2^{1/2}k(\omega))^{-1}$, which equals 69 nm for the silver surface and surface plasmon polariton frequency assumed. This prediction is confirmed by the numerical results presented in Fig. 9, which show that $a_{opt}/\lambda \cong 0.1$, irrespective of the defect height for which Eq. (2.7) is valid. The maximum surface plasmon polariton reflectivity, however, increases with the defect height, and is slightly larger for grooves than for ridges. Such defects are called *plasmon mirrors* [58,59,65]. As the width of the defect increases, ridges and grooves begin to behave differently, except

for the fact that they both reflect surface plasmon polaritons negligibly. On the one hand the transmission of a surface plasmon polariton through a ridge decreases monotonically at the expense of radiation. The decrease is more rapid the higher the ridge is. On the other hand, the transmission of a surface plasmon polariton through a groove and its conversion into radiation exhibit oscillatory dependences as the defect width increases: the radiation (surface plasmon polariton transmission) increases (decreases), passes through a maximum (minimum), and then tends asymptotically to 0(1). The period of the oscillations, the width of the groove that yields the maximum radiation, and the value of this maximum all depend on the depth of the groove. Both ridges and grooves may behave as *light emitters*, i.e. as defects that display a high surface plasmon polariton–light conversion efficiency, for an appropriate (and distinct) range of defect parameters. Ridges increasingly radiate more light at near grazing scattering angles at the expense of surface plasmon polariton transmission as the defect width increases. Values of S greater than 0.9 are observed in Fig. 9 for ridges with $h/\lambda = 0.2$ and $a/\lambda \geq 3.6$. Qualitatively, the fact that the metal ridge enters the vacuum half-space seems to favor the surface plasmon polariton–photon coupling, i.e. it plays the role of a *launching platform*. A value of $S = 0.94$ is observed in the scattering of a surface plasmon polariton by a groove with $h/\lambda = 0.2$ and $a/\lambda = 0.3$ (Fig. 9), but the surface plasmon polariton–photon coupling rapidly decreases in an oscillatory fashion as a/λ increases. The range of defect widths for which high coupling efficiencies are encountered is therefore far more restrictive for grooves than for ridges. Only sufficiently deep grooves ($h/\lambda \geq 0.2$) are capable of producing radiation efficiencies $S \geq 0.9$, and only for a narrow range of widths.

b. Two-dimensional surface defects. The only calculation to date of the scattering of a surface plasmon polariton by a two-dimensional surface defect was carried out by Shchegrov et al. [29] for the case where the surface defect had circular symmetry. The system studied in this case consists of vacuum in the region $x_3 > \zeta(\mathbf{x}_{\parallel})$, where $\mathbf{x}_{\parallel} = (x_1, x_2, 0)$, and a metal in the region $x_3 < \zeta(\mathbf{x}_{\parallel})$ characterized by an isotropic, frequency-dependent dielectric function $\epsilon(\omega)$. The surface profile function is assumed to be a single-valued function of \mathbf{x}_{\parallel} , which essentially vanishes for $|\mathbf{x}_{\parallel}|$ larger than some characteristic length R . Moreover, $\zeta(\mathbf{x}_{\parallel})$ is assumed to be a circularly symmetric function of \mathbf{x}_{\parallel} , i.e. it depends on \mathbf{x}_{\parallel} only through its magnitude x_{\parallel} . This assumption greatly simplifies the calculation without sacrificing a great deal of generality.

We assume that a surface plasmon polariton of frequency ω propagating in the $+x_1$ -direction from $x_1 = -\infty$ is scattered by the surface defect. The total electric field in this system has the form $\mathbf{E}(\mathbf{x}; t) = \mathbf{E}(\mathbf{x}|\omega) \exp(-i\omega t)$, where the amplitude function $\mathbf{E}(\mathbf{x}|\omega)$ in the vacuum region $x_3 > \zeta(\mathbf{x}_{\parallel})$ is given by

$$\begin{aligned} \mathbf{E}^>(\mathbf{x}|\omega) = & \frac{c}{\omega} [i\hat{\mathbf{x}}_1 \beta_0(\omega) - \hat{\mathbf{x}}_3 k_{\parallel}(\omega)] \exp[ik_{\parallel}(\omega)x_1 - \beta_0(\omega)x_3] \\ & + \int \frac{d^2 q_{\parallel}}{(2\pi)^2} \left\{ \frac{c}{\omega} [i\hat{\mathbf{q}}_{\parallel} \beta_0(q_{\parallel}) - \hat{\mathbf{x}}_3 q_{\parallel}] A_p(\mathbf{q}_{\parallel}) + (\hat{\mathbf{x}}_3 \times \hat{\mathbf{q}}_{\parallel}) A_s(\mathbf{q}_{\parallel}) \right\} \\ & \times \exp[i\mathbf{q}_{\parallel} \cdot \mathbf{x}_{\parallel} - \beta_0(q_{\parallel})x_3] . \end{aligned} \quad (2.24)$$

The first term on the right hand side of Eq. (2.24) describes the field of the incident surface plasmon polariton. The wave vector of this incident wave $\mathbf{k}_{\parallel} = (k_{\parallel}(\omega), 0, 0)$ is directed along the x_1 -axis, and its magnitude is

$$k_{\parallel}(\omega) = \frac{\omega}{c} \left(\frac{\epsilon(\omega)}{\epsilon(\omega) + 1} \right)^{1/2} . \quad (2.25)$$

The function $\beta_0(\omega)$ is the inverse decay length of the electromagnetic field of the surface plasmon polariton in the vacuum region, and is given by

$$\beta_0(\omega) = \frac{\omega}{c} \left(\frac{-1}{\epsilon(\omega) + 1} \right)^{1/2}. \quad (2.26)$$

A caret over a vector indicates that it is a unit vector. The second term on the right hand side of Eq. (2.24) represents the scattered field, where $A_p(\mathbf{q}_{\parallel})$ and $A_s(\mathbf{q}_{\parallel})$ are the scattering amplitudes of its p- and s-polarized components with respect to the plane of scattering. The vector \mathbf{q}_{\parallel} in Eq. (2.24) is $\mathbf{q}_{\parallel} = (q_1, q_2, 0)$, and

$$\beta_0(q_{\parallel}) = \left(q_{\parallel}^2 - \frac{\omega^2}{c^2} \right)^{1/2} \quad q_{\parallel} > \frac{\omega}{c} \quad (2.27a)$$

$$= -i \left(\frac{\omega^2}{c^2} - q_{\parallel}^2 \right)^{1/2} \quad q_{\parallel} < \frac{\omega}{c}. \quad (2.27b)$$

An expression analogous to Eq. (2.24) can be written for $\mathbf{E}^<(\mathbf{x}_{\parallel}, \omega)$ in the metal region $x_3 < \zeta(\mathbf{x}_{\parallel})$, and then used with Eq. (2.24) in satisfying the boundary conditions at $x_3 = \zeta(\mathbf{x}_{\parallel})$. However, it has been shown by Brown et al. [69] that the field in the scattering medium can be eliminated from the problem. The amplitudes $A_{p,s}(\mathbf{q}_{\parallel})$ then satisfy a 2×2 matrix integral equation called the *reduced Rayleigh equation*, which, in the present case, can be written in the form

$$f_i(p_{\parallel}) A_i(\mathbf{p}_{\parallel}) + \sum_{j=p,s} \int \frac{d^2 q_{\parallel}}{(2\pi)^2} g_{ij}(\mathbf{p}_{\parallel} | \mathbf{q}_{\parallel}) A_j(\mathbf{q}_{\parallel}) = -g_{ip}(\mathbf{p}_{\parallel} | \mathbf{k}_{\parallel}), \quad i = p, s, \quad (2.28)$$

where

$$f_p(p_{\parallel}) = \frac{\epsilon(\omega)\beta_0(p_{\parallel}) + \beta(p_{\parallel})}{1 - \epsilon(\omega)}, \quad (2.29a)$$

$$f_s(p_{\parallel}) = \frac{\beta_0(p_{\parallel}) + \beta(p_{\parallel})}{1 - \epsilon(\omega)}, \quad (2.29b)$$

$$g_{pp}(\mathbf{p}_{\parallel} | \mathbf{q}_{\parallel}) = J(\mathbf{p}_{\parallel} | \mathbf{q}_{\parallel}) [p_{\parallel} q_{\parallel} - \beta(p_{\parallel}) \hat{\mathbf{p}}_{\parallel} \cdot \hat{\mathbf{q}}_{\parallel} \beta_0(q_{\parallel})], \quad (2.30a)$$

$$g_{ps}(\mathbf{p}_{\parallel} | \mathbf{q}_{\parallel}) = -i J(\mathbf{p}_{\parallel} | \mathbf{q}_{\parallel}) (\omega/c) \beta(p_{\parallel}) (\hat{\mathbf{p}}_{\parallel} \times \hat{\mathbf{q}}_{\parallel})_3, \quad (2.30b)$$

$$g_{sp}(\mathbf{p}_{\parallel} | \mathbf{q}_{\parallel}) = i J(\mathbf{p}_{\parallel} | \mathbf{q}_{\parallel}) (\omega/c) (\hat{\mathbf{p}}_{\parallel} \times \hat{\mathbf{q}}_{\parallel})_3 \beta_0(q_{\parallel}), \quad (2.30c)$$

$$g_{ss}(\mathbf{p}_{\parallel} | \mathbf{q}_{\parallel}) = J(\mathbf{p}_{\parallel} | \mathbf{q}_{\parallel}) (\omega/c)^2 \hat{\mathbf{p}}_{\parallel} \cdot \hat{\mathbf{q}}_{\parallel}, \quad (2.30d)$$

with

$$J(\mathbf{p}_{\parallel} | \mathbf{q}_{\parallel}) = \int d^2 x_{\parallel} \exp[-i(\mathbf{p}_{\parallel} - \mathbf{q}_{\parallel}) \cdot \mathbf{x}_{\parallel}] \frac{\exp[\beta(p_{\parallel}) - \beta_0(q_{\parallel})] \zeta(\mathbf{x}_{\parallel}) - 1}{\beta(p_{\parallel}) - \beta_0(q_{\parallel})}, \quad (2.31)$$

where $\beta(q_{\parallel}) = [q_{\parallel}^2 - \epsilon(\omega)(\omega/c)^2]^{1/2}$ is the inverse decay length of the electromagnetic field inside the metal. Within the Rayleigh hypothesis [70] Eq. (2.28) is the exact equation for the scattering amplitudes $A_{p,s}(\mathbf{q}_{\parallel})$ for an arbitrary surface profile function $\zeta(\mathbf{x}_{\parallel})$.

At this point we exploit the circular symmetry of the surface defect. In this case the expression for $J(\mathbf{p}_{\parallel}|\mathbf{q}_{\parallel})$ can be rewritten in the form

$$J(\mathbf{p}_{\parallel}|\mathbf{q}_{\parallel}) = \frac{1}{\beta(p_{\parallel}) - \beta_0(q_{\parallel})} \int_0^{\infty} dx_{\parallel} x_{\parallel} [\exp(\beta(p_{\parallel}) - \beta_0(q_{\parallel})) - 1] \\ \times \int_{-\pi}^{\pi} d\theta_x \exp[-ip_{\parallel} x_{\parallel} \cos(\phi_p - \phi_x)] \exp[iq_{\parallel} x_{\parallel} \cos(\phi_q - \phi_x)], \quad (2.32)$$

where ϕ_p, ϕ_q, ϕ_x are the azimuthal angles of the vectors $\mathbf{p}_{\parallel}, \mathbf{q}_{\parallel}, \mathbf{x}_{\parallel}$, respectively. With the aid of the expansion

$$\exp(iz \cos \theta) = \sum_{m=-\infty}^{\infty} i^m \exp(im\theta) J_m(z), \quad (2.33)$$

where $J_m(z)$ is a Bessel function, we obtain the result

$$J(\mathbf{p}_{\parallel}|\mathbf{q}_{\parallel}) = \sum_{m=-\infty}^{\infty} N_m(p_{\parallel}|q_{\parallel}) \exp[im(\phi_p - \phi_q)], \quad (2.34)$$

where

$$N_m(p_{\parallel}|q_{\parallel}) = \frac{2\pi}{\beta(p_{\parallel}) - \beta_0(q_{\parallel})} \int_0^{\infty} dx_{\parallel} x_{\parallel} \{\exp[\beta(p_{\parallel}) - \beta_0(q_{\parallel})] \zeta(x_{\parallel}) - 1\} \\ \times J_m(p_{\parallel} x_{\parallel}) J_m(q_{\parallel} x_{\parallel}). \quad (2.35)$$

If we then expand $A_i(\mathbf{q}_{\parallel})$ in the form

$$A_i(\mathbf{q}_{\parallel}) = \sum_{n=-\infty}^{\infty} A_i^{(m)}(q_{\parallel}) \exp(im\phi_q), \quad (2.36)$$

the substitution of this expansion together with Eq. (2.34) into Eq. (2.28) yields the following set of coupled one-dimensional integral equations for the amplitudes $\{A_i^{(m)}(q_{\parallel})\}$:

$$f_p(p_{\parallel}) A_p^{(m)}(p_{\parallel}) + \frac{1}{2\pi} \int_0^{\infty} dq_{\parallel} q_{\parallel} [h_{pp}^{(m)}(p_{\parallel}|q_{\parallel}) A_p^{(m)}(q_{\parallel}) + h_{ps}^{(m)}(p_{\parallel}|q_{\parallel}) A_s^{(m)}(q_{\parallel})] \\ = -h_{pp}^{(m)}(p_{\parallel}|k_{\parallel}) \quad (2.37a)$$

$$f_s(p_{\parallel}) A_s^{(m)}(p_{\parallel}) + \frac{1}{2\pi} \int_0^{\infty} dq_{\parallel} q_{\parallel} [h_{sp}^{(m)}(p_{\parallel}|q_{\parallel}) A_p^{(m)}(q_{\parallel}) + h_{ss}^{(m)}(p_{\parallel}|q_{\parallel}) A_s^{(m)}(q_{\parallel})] \\ = -h_{sp}^{(m)}(p_{\parallel}|k_{\parallel}), \quad (2.37b)$$

where

$$h_{pp}^{(m)}(p_{\parallel}|q_{\parallel}) = p_{\parallel} q_{\parallel} N_m(p_{\parallel}|q_{\parallel}) - \frac{1}{2} \beta(p_{\parallel}) \beta_0(q_{\parallel}) [N_{m-1}(p_{\parallel}|q_{\parallel}) + N_{m+1}(p_{\parallel}|q_{\parallel})] \quad (2.38a)$$

$$h_{ps}^{(m)}(p_{\parallel}|q_{\parallel}) = \frac{1}{2} \frac{\omega}{c} \beta(p_{\parallel}) [N_{m-1}(p_{\parallel}|q_{\parallel}) - N_{m+1}(p_{\parallel}|q_{\parallel})] \quad (2.38b)$$

$$h_{sp}^{(m)}(p_{\parallel}|q_{\parallel}) = -\frac{1}{2} \frac{\omega}{c} \beta_0(q_{\parallel}) [N_{m-1}(p_{\parallel}|q_{\parallel}) - N_{m+1}(p_{\parallel}|q_{\parallel})] \quad (2.38c)$$

$$h_{ss}^{(m)}(p_{\parallel}|q_{\parallel}) = \frac{1}{2} \frac{\omega^2}{c^2} [N_{m-1}(p_{\parallel}|q_{\parallel}) + N_{m+1}(p_{\parallel}|q_{\parallel})] . \quad (2.38d)$$

Since the unperturbed surface (without the defect) supports surface plasmon polaritons of p polarization, the scattering amplitude $A_p(\mathbf{q}_{\parallel})$, and hence the $\{A_p^{(m)}(q_{\parallel})\}$, must have a pole at $q_{\parallel} = k_{\parallel}(\omega) + i\eta$, where the positive imaginary infinitesimal is added to ensure that the scattered waves are outgoing. Therefore, in solving Eqs. (2.37) it is convenient to seek $A_p^{(m)}(q_{\parallel})$ in the form

$$A_p^{(m)}(q_{\parallel}) = \frac{a_p^{(m)}(q_{\parallel})}{f_p(q_{\parallel})} , \quad (2.39)$$

where the coefficients $\{a_p^{(m)}(q_{\parallel})\}$ are smoothly varying functions of q_{\parallel} . Thus, Eqs. (2.37) become

$$\begin{aligned} a^{(m)}(p_{\parallel}) + \frac{1}{2\pi} \int_0^{\infty} dq_{\parallel} q_{\parallel} \left[\frac{h_{pp}^{(m)}(p_{\parallel}|q_{\parallel})}{f_p(q_{\parallel})} a_p^{(m)}(q_{\parallel}) + h_{ps}^{(m)}(p_{\parallel}|q_{\parallel}) A_s^{(m)}(q_{\parallel}) \right] \\ = -h_{pp}^{(m)}(p_{\parallel}|k_{\parallel}) , \end{aligned} \quad (2.40a)$$

$$\begin{aligned} f_s(p_{\parallel}) A_s^{(m)}(p_{\parallel}) + \frac{1}{2\pi} \int_0^{\infty} dq_{\parallel} q_{\parallel} \left[\frac{h_{sp}^{(m)}(p_{\parallel}|q_{\parallel})}{f_p(q_{\parallel})} a_p^{(m)}(q_{\parallel}) + h_{ps}^{(m)}(p_{\parallel}|q_{\parallel}) A_s^{(m)}(q_{\parallel}) \right] \\ = -h_{sp}^{(m)}(p_{\parallel}|k_{\parallel}) . \end{aligned} \quad (2.40b)$$

The scattered field is given by the second term on the right hand side of Eq. (2.24):

$$\begin{aligned} \mathbf{E}^>(\mathbf{x}|\omega)_{sc} = \int \frac{d^2q_{\parallel}}{(2\pi)^2} \left\{ \frac{c}{\omega} [i\hat{\mathbf{q}}_{\parallel} \beta_0(q_{\parallel}) - \hat{\mathbf{x}}_3 q_{\parallel}] A_p(\mathbf{q}_{\parallel}) \right. \\ \left. + (\hat{\mathbf{x}}_3 \times \hat{\mathbf{q}}_{\parallel}) A_s(\mathbf{q}_{\parallel}) \right\} \exp[i\mathbf{q}_{\parallel} \cdot \mathbf{x}_{\parallel} - \beta_0(q_{\parallel}) x_3] . \end{aligned} \quad (2.41)$$

After determining $A_{p,s}(\mathbf{q}_{\parallel})$ by the use of Eqs. (2.36), (2.39), and (2.40) we can apply the method of stationary phase, described in Ref. [71] to calculate the scattered field in the far field:

$$\mathbf{E}_{vac}^>(\mathbf{x}|\omega)_{sc} = -\frac{i\omega}{2\pi c} \cos \theta_x \frac{e^{i\omega/cx}}{x} \left\{ \hat{\mathbf{e}}_p A_p \left(\hat{\mathbf{x}}_{\parallel} \frac{\omega}{c} \sin \theta_x \right) + \hat{\mathbf{e}}_s A_s \left(\hat{\mathbf{x}}_{\parallel} \frac{\omega}{c} \sin \theta_x \right) \right\} , \quad (2.42)$$

where $\hat{\mathbf{e}}_p = (\cos \theta_x \cos \phi_x, \cos \theta_x \sin \phi_x, -\sin \theta_x)$ and $\hat{\mathbf{e}}_s = (-\sin \phi_x, \cos \phi_x, 0)$ are polarization vectors, and the point of observation in the far field is defined by the vector $\mathbf{x} = x(\sin \theta_x \cos \phi_x, \sin \theta_x \sin \phi_x, \cos \theta_x)$. The electric field of the surface plasmon polariton excited by the scattering of the incident surface plasmon polariton by the defect, is given by 2π times the residue at the pole of the integrand in Eq. (2.41) that occurs at $q_{\parallel} = k_{\parallel}(\omega) > \omega/c$. It can be calculated by the approach described in Ref. [71], and has the form of a cylindrical wave in the far field ($|\mathbf{x}_{\parallel}| \rightarrow \infty$):

$$\mathbf{E}_{sp}^>(\mathbf{x}|\omega) = \frac{e^{ik_{\parallel}(\omega)x_{\parallel} - \beta_0(\omega)x_3 + i\pi/4}}{(2\pi k_{\parallel}(\omega)x_{\parallel})^{1/2}} \frac{c\beta(\omega)}{\omega} \frac{i\hat{\mathbf{x}}_{\parallel} \beta_0(\omega) - \hat{\mathbf{x}}_3 k_{\parallel}(\omega)}{\epsilon(\omega) + 1} a_p(\hat{\mathbf{x}}_{\parallel} k_{\parallel}(\omega)) , \quad (2.43)$$

where $\beta(\omega) = (\omega/c)[- \epsilon^2(\omega)/(\epsilon(\omega) + 1)]^{1/2}$, and

$$a_p(\mathbf{q}_{\parallel}) = \sum_{m=-\infty}^{\infty} a_p^{(m)}(q_{\parallel}) \exp(im\phi_q) . \quad (2.44)$$

From the results given by Eqs. (2.42) and (2.43) we can calculate differential cross sections, measured in units of length, for the scattering of the incident surface plasmon polariton into volume electromagnetic waves in the vacuum and into other surface waves. They are defined by

$$\sigma_{\text{vac}}(\theta_s, \phi_s) = \frac{P_{\text{vac}}(\theta_x, \phi_x)}{P_{\text{inc}}} \quad (2.45a)$$

$$\sigma_{\text{sp}}(\phi_s) = \frac{P_{\text{sp}}(\phi_x)}{P_{\text{inc}}} . \quad (2.45b)$$

In these expressions $P_{\text{vac}}(\theta_x, \phi_x)$ is the power scattered into the vacuum in the direction (θ_x, ϕ_x) . It is obtained from the total time-averaged flux scattered into the vacuum P_{vac} by

$$P_{\text{vac}} = \int_0^{\pi/2} d\theta_x \sin \theta_x \int_{-\pi}^{\pi} d\phi_x x^2 \text{Re } S_r^c(\mathbf{x}|\omega)_{\text{vac}} \quad (2.46a)$$

$$\equiv \int_0^{\pi/2} d\theta_x \sin \theta_x \int_{-\pi}^{\pi} d\phi_x P_{\text{vac}}(\theta_x, \phi_x) , \quad (2.46b)$$

where $S_r^c(\mathbf{x}|\omega)_{\text{vac}}$ is the radial component of the complex Poynting vector of the field scattered into the vacuum, and calculated on the basis of Eq. (2.42). Similarly, $P_{\text{sp}}(\phi_x)$ is the power scattered into the surface waves in the direction $\hat{\mathbf{x}}_{\parallel}$, and is obtained from the total time-averaged flux scattered into surface waves P_{sp} by

$$P_{\text{sp}} = \int_0^{\infty} dx_3 \int_{-\pi}^{\pi} d\phi_x x_{\parallel} \text{Re } S_r^c(\mathbf{x}|\omega)_{\text{sp}} \quad (2.47a)$$

$$\equiv \int_{-\pi}^{\pi} d\phi_x P_{\text{sp}}(\phi_x) , \quad (2.47b)$$

where $S_r^c(\mathbf{x}|\omega)_{\text{sp}}$ is the radial component of the complex Poynting vector of the field scattered into other surface plasmon polaritons and is calculated on the basis of Eq. (2.43). Finally, P_{inc} is the total time-averaged incident flux, which is given by

$$P_{\text{inc}} = \int_{-\frac{1}{2}L_2}^{\frac{1}{2}L_2} dx_2 \int_0^{\infty} dx_3 \text{Re } S_1^c(\mathbf{x}|\omega)_{\text{inc}} \quad (2.48a)$$

$$= L_2 \frac{c^2}{16\pi\omega} [-\epsilon(\omega)]^{1/2} , \quad (2.48b)$$

where L_2 is the length of the surface in the x_2 -direction, and $S_1^c(\mathbf{x}|\omega)_{\text{inc}}$ is the 1-component of the complex Poynting vector of the incident field, which is calculated on the basis of the first term on the right hand side of Eq. (2.24).

The calculations of the scattering of a surface plasmon polariton by a two-dimensional surface defect carried out in Ref. [29] were based on the assumption of a circularly symmetric surface profile function of Gaussian form, $\zeta(x_{\parallel}) = A \exp(-x_{\parallel}^2/R^2)$. In this case the use of the result [72]

$$\int_0^{\infty} e^{-a^2x^2} J_{\nu}(bx) J_{\nu}(cx) x \, dx = \frac{1}{2a^2} e^{-(b^2+c^2)/4a^2} I_{\nu} \left(\frac{bc}{2a^2} \right), \quad (2.49b)$$

where $I_{\nu}(z)$ is a modified Bessel function, yields the following expansion for $N_m(p_{\parallel}|q_{\parallel})$:

$$N_m(p_{\parallel}|q_{\parallel}) = \pi A R^2 \sum_{n=1}^{\infty} \frac{[(\beta(p_{\parallel}) - \beta_0(q_{\parallel}))A]^{n-1}}{n \cdot n!} e^{-(p_{\parallel}^2+q_{\parallel}^2)R^2/2n} I_n \left(\frac{p_{\parallel}q_{\parallel}R^2}{2n} \right). \quad (2.50)$$

The solution of the resulting system of integral equations for $a_p^{(m)}(q_{\parallel})$ and $A_s^{(m)}(q_{\parallel})$ was carried out numerically by converting it into a system of matrix equations. The infinite range of integration was replaced by a finite range $(0, Q_{\max})$. It was found that values of Q_{\max} of the order of $8/R$ to $12/R$ sufficed to yield accurate values of the integrals. This range was then broken up into the five intervals $(0, \omega/c)$, $(\omega/c, k_{\parallel}(\omega) - \delta)$, $(k_{\parallel}(\omega) - \delta, k_{\parallel}(\omega) + \delta)$, $(k_{\parallel}(\omega) + \delta, 2k_{\parallel}(\omega) - \delta)$, and $(2k_{\parallel}(\omega) - \delta, Q_{\max})$. Each of these intervals was subdivided into equally spaced intervals, at the midpoints of which (abscissas) the integrands in Eqs. (2.40) were evaluated and then multiplied by the width of the corresponding interval. The number of abscissas in each interval was varied to reflect how slowly or rapidly the integrand was changing in that interval. A total of about 150 abscissas was found to be sufficient to obtain accurate results. A sum over the contributions from all of these intervals replaced the integrals in Eqs. (2.40). When the variable p_{\parallel} was given the values of the midpoints of the same intervals, a pair of coupled matrix equations for $a_p^{(m)}(q_{\parallel})$ and $A_s^{(m)}(q_{\parallel})$ was obtained. This pair of equations was solved for $m=0, \pm 1, \pm 2, \pm 3, \pm 4, \pm 5$.

The pole at $q_{\parallel} = k_{\parallel}(\omega)$ in the integrand of each of Eqs. (2.40) due to the presence of $f_p^{-1}(q_{\parallel})$ was dealt with very carefully. The contribution from the delta function was calculated analytically (the point $q_{\parallel} = k_{\text{sp}}(\omega)$ was one of the abscissas in the numerical integration over q_{\parallel}). The contribution from the principal value integrals was dealt with by excluding a symmetric interval of width 2δ about the point $q_{\parallel} = k_{\parallel}(\omega)$, where δ ranged from $10^{-6}k_{\parallel}(\omega)$ to $10^{-3}k_{\parallel}(\omega)$, depending on the values of the remaining parameters.

The calculations in Ref. [29] were carried out for a Gaussian indentation defined by $A = -0.05 \mu\text{m}$ and $R = 0.25 \mu\text{m}$ on a silver surface with $\epsilon(\omega) = -17.8$. These values correspond approximately to one of the cases studied experimentally by Smolyaninov et al. [58]. The value of $\epsilon(\omega)$ corresponds approximately to the vacuum wavelength of the light emitted by a He–Ne laser, $\lambda = 632.8 \text{ nm}$.

In Fig. 10 a contour plot of $\sigma_{\text{vac}}(\theta_s, \phi_s)$ is presented. The maximum of the scattered intensity occurs at $\theta_s = 28^\circ$, $\phi_s = 0^\circ$. The total cross section for the waves scattered into the vacuum, $\sigma_{\text{vac}}^{(\text{tot})}$, in this case is small, only $3.7 \times 10^{-3} \mu\text{m}$. This is not surprising for such a shallow defect. The result for $\sigma_{\text{sp}}^{(\text{tot})}$ is of the same order of magnitude, namely $2.6 \times 10^{-3} \mu\text{m}$. The angular dependence of $\sigma_{\text{sp}}(\phi_s)$ is depicted in Fig. 11. It is seen that the scattering of surface plasmon polaritons in the forward and backward directions is suppressed. The main portion of the scattered energy goes into two surface plasmon polariton beams separated by approximately 70° . This result is even better illustrated in Fig. 12, which shows the field intensity $|\mathbf{E}^>(\mathbf{x}|\omega)|^2$ at 5 nm above the surface profile ($x_3 = \zeta(\mathbf{x}_{\parallel}) + 5 \text{ nm}$), which corresponds to the quantity measured in Ref. [58]. The results shown in Figs. 11 and 12 are in agreement with the corresponding experimental result of Smolyaninov et al. [58], especially the shadow behind the defect.

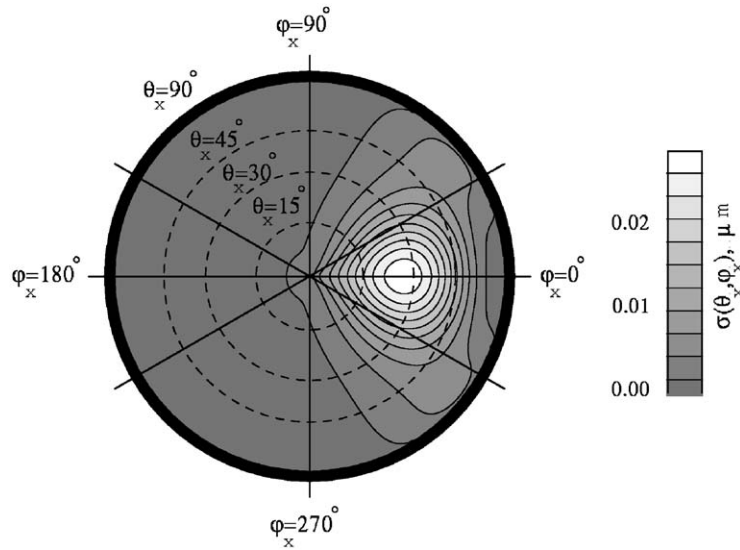


Fig. 10. A contour plot of $\sigma_{\text{vac}}(\theta_s, \phi_s)$. The concentric circles are the lines of constant θ_s , with $\theta_s = 0^\circ$ at the center, $\theta_s = 90^\circ$ at the border. The azimuthal angle ϕ_s varies from 0° to 360° . $A = -0.05 \mu\text{m}$, $R = 0.25 \mu\text{m}$. (Ref. [29].)

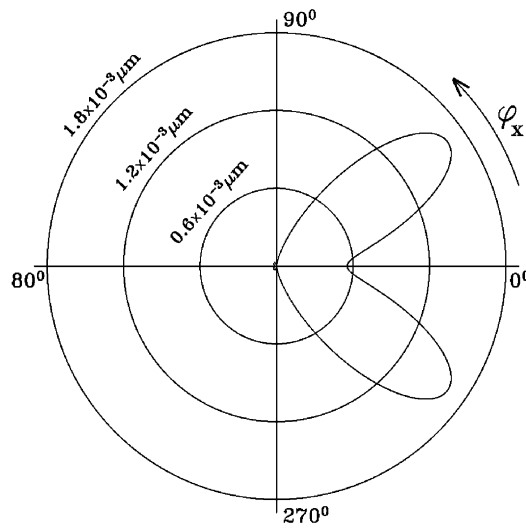


Fig. 11. A polar plot of $\sigma_{\text{sp}}(\phi_s)$ for $A = -0.05 \mu\text{m}$ and $R = 0.25 \mu\text{m}$. (Ref. [29].)

The fact that some features that are present in Fig. 12 (e.g., additional weaker diffraction maxima) are not seen in the experimental image is due to the fact that the tip used in measuring the intensity in Ref. [58] was not a point detector in both the vertical and lateral directions. Its actual lateral resolution was about $0.1\text{--}0.2 \mu\text{m}$. Therefore, a comparison between theory and experiment can be made only in an average sense. Finally, both $\sigma_{\text{vac}}^{(\text{tot})}$ and $\sigma_{\text{sp}}^{(\text{tot})}$ were calculated as functions of the frequency of the incident surface

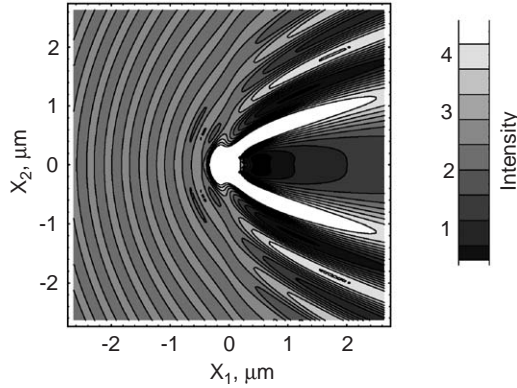


Fig. 12. The field intensity $|\mathbf{E}^>(\mathbf{x}|\omega)|^2$ as a function of \mathbf{x}_{\parallel} for $x_3 = \zeta(x_{\parallel}) + 5 \mu\text{m}$. $A = -0.05 \mu\text{m}$, $R = 0.25 \mu\text{m}$.

plasmon polariton in the range $0 < \omega < \omega_p$. This calculation was prompted by the fact that a localized surface defect of the kind considered in this section supports electromagnetic surface shape resonances (localized surface plasmons) [73]. These are excitations that are spatially localized in the vicinity of the defect, and are characterized by discrete, complex frequencies, whose imaginary part arises from their decay into volume and surface electromagnetic waves. The actual values of these frequencies depend on the shape of the defect. The resonances with the longest lifetimes exist in the region of ω in which $\epsilon(\omega)$ is slightly smaller than -1 . When the frequency of the incident surface plasmon polariton matches the real part of the frequency of one of these resonances, an enhancement of $\sigma_{\text{sp}}^{(\text{tot})}$ is expected. Such an enhancement was observed in the results of Ref. [29], at the resonance frequency $\omega = 0.687\omega_p$. It has not yet been observed experimentally.

2.1.2. The propagation of a surface plasmon polariton near an index step

In the scattering of light from a weakly rough surface the surface profile function and the dielectric contrast always enter the expression for the scattered field in one and the same combination. This makes the separation of the roughness and the dielectric properties of surface or subsurface defects a very difficult problem, even for modern near-field optical microscopy. When the surface structure is created by low-contrast coatings, the problem becomes even more difficult, since the scattering from such surfaces is very weak. Until now, obtaining and interpreting images of low-contrast samples, often encountered in biophysical studies, remain challenging problems, especially when a high lateral resolution is required.

For solving these problems the surface plasmon polariton microscopy technique offers an extremely high contrast sensitivity without loss of spatial resolution. In this technique surface plasmon polaritons are used as illumination instead of light. The strong enhancement of the surface plasmon polariton field at the surface, and an extremely high sensitivity of the surface plasmon polariton dispersion law to optical properties of the interface, make them a unique probe of the properties of surfaces and ultrathin films. When a surface plasmon polariton propagating along a surface interacts with a surface or subsurface defect, it is scattered by the defect, and in the region of the defect it changes its wavenumbers [74–78]. The simultaneous propagation of electromagnetic waves with different wavenumbers, for example the scattered volume waves and the surface plasmon polaritons, leads to the appearance of a complicated interference pattern in the intensity in the far field. This interference pattern can be inverted to provide

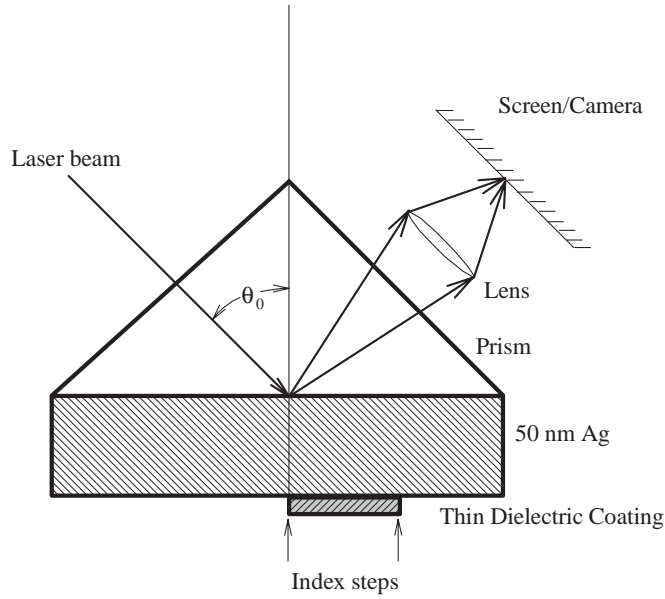


Fig. 13. The scattering system constituting a surface plasmon polariton microscope.

the dielectric contrast and lateral dimensions of the defect. The great advantage of this technique over near-field techniques is that the measurements can be made in the far field.

Surface plasmon polariton microscopy has been used by several authors in experimental studies of optical properties of laterally inhomogeneous thin dielectric films deposited on a planar metal surface that supports surface plasmon polaritons [79–84]. However, the theoretical work used in interpreting the experimental results [82,83] was of a somewhat ad hoc nature. In this section we outline a systematic theoretical analysis [85] of the scattering of p-polarized light from the structure that has been used in the experiments described in Refs. [79–84].

The physical system we study is depicted in Fig. 13. It consists of a prism, characterized by the dielectric constant ϵ_0 , in the region $x_3 > D$; a metal film, characterized by a complex, frequency-dependent, dielectric function $\epsilon_1(\omega)$ in the region $0 < x_3 < D$; a thin dielectric film, characterized by the dielectric constant ϵ_2 in the region $\zeta(x_1) < x_3 < 0$; and vacuum ($\epsilon_3 = 1$) in the region $x_3 < \zeta(x_1)$.

This structure is illuminated by p-polarized light of frequency ω , whose plane of incidence is the x_1x_3 -plane. The medium of incidence is the prism in the region $x_3 > D$. The single nonzero component of the magnetic field in this system has the following form in each of the layers comprising this structure: $x_3 > D$

$$H_2(x_1, x_3|\omega) = \exp i[kx_1 - \alpha_0(k)(x_3 - D)] + \int_{-\infty}^{\infty} \frac{dq}{2\pi} R(q|k) \exp i[qx_1 + \alpha_0(q)(x_3 - D)]; \quad (2.51)$$

$0 < x_3 < D$

$$H_2(x_1, x_3|\omega) = \int_{-\infty}^{\infty} \frac{dq}{2\pi} \exp(iqx_1) [T_1(q|k) \exp(i\alpha_1(q)x_3) + T_2(q|k) \exp(-i\alpha_1(q)x_3)]; \quad (2.52)$$

$$\zeta(x_1) < x_3 < 0$$

$$H_2(x_1, x_3|\omega) = \int_{-\infty}^{\infty} \frac{dq}{2\pi} \exp(iqx_1)[T_3(q|k) \exp(i\alpha_2(q)x_3) + T_4(q|k) \exp(-i\alpha_2(q)x_3)] ; \quad (2.53)$$

$$x_3 < \zeta(x_1)$$

$$H_2(x_1, x_3|\omega) = \int_{-\infty}^{\infty} \frac{dq}{2\pi} T(q|k) \exp i[qx_1 - \alpha_3(q)x_3] , \quad (2.54)$$

where

$$\alpha_i(q) = [\epsilon_i(\omega/c)^2 - q^2]^{1/2}, \quad \text{Re } \alpha_i(q) > 0, \text{Im } \alpha_i(q) > 0, \quad i = 0, 1, 2, 3 . \quad (2.55)$$

From the boundary conditions at the interface $x_3 = D$, at the interface $x_3 = 0$, and at the interface $x_3 = \zeta(x_1)$, namely the continuity of the tangential components of the magnetic and electric fields across each interface, it is possible to obtain an integral equation for the scattering amplitude $R(q|k)$ alone. It can be written in the form [85]

$$\int_{-\infty}^{\infty} \frac{dq}{2\pi} M(p|q)R(q|k) = -N(p|k) , \quad (2.56)$$

where

$$M(p|q) = \left\{ \begin{aligned} & \frac{pq + \alpha_3(p)\alpha_2(q)}{\alpha_3(p) - \alpha_2(q)} [1 + r_{21}(q)r_{10}(q)e^{2i\alpha_1(q)D}] I(\alpha_3(p) - \alpha_2(q)|p - q) \\ & + \frac{pq - \alpha_3(p)\alpha_2(q)}{\alpha_3(p) + \alpha_2(q)} [r_{21}(q) + r_{10}(q)e^{2i\alpha_1(q)D}] I(\alpha_3(p) + \alpha_2(q)|p - q) \end{aligned} \right\} \\ \times \frac{e^{-i\alpha_1(q)D}}{t_{21}(q)t_{10}(q)} \quad (2.57a)$$

$$N(p|k) = \left\{ \begin{aligned} & \frac{pk + \alpha_3(p)\alpha_2(k)}{\alpha_3(p) - \alpha_2(k)} [r_{10}(k) + r_{21}(k)e^{2i\alpha_1(k)D}] I(\alpha_3(p) - \alpha_2(k)|p - k) \\ & + \frac{pk - \alpha_3(p)\alpha_2(k)}{\alpha_3(p) + \alpha_2(k)} [r_{21}(k)r_{10}(k) + e^{2i\alpha_1(k)D}] I(\alpha_3(p) + \alpha_2(k)|p - k) \end{aligned} \right\} \\ \times \frac{e^{-i\alpha_1(k)D}}{t_{21}(k)t_{10}(k)} . \quad (2.57b)$$

In writing Eqs. (2.57) we have introduced the functions

$$I(\gamma|Q) = \int_{-\infty}^{\infty} dx_1 \exp(-iQx_1) \exp[-i\gamma\zeta(x_1)] , \quad (2.58)$$

$$r_{ij}(q) = \frac{\Delta_{ij}(q)}{d_{ij}(q)}, \quad t_{ij}(q) = \frac{2\alpha_i(q)}{\epsilon_i d_{ij}(q)} , \quad (2.59)$$

with

$$\Delta_{ij}(q) = \frac{\alpha_i(q)}{\epsilon_i} - \frac{\alpha_j(q)}{\epsilon_j}, \quad d_{ij}(q) = \frac{\alpha_i(q)}{\epsilon_i} + \frac{\alpha_j(q)}{\epsilon_j} . \quad (2.60)$$

Once the scattering amplitude $R(q|k)$ is known, the transmission amplitude $T(q|k)$ can be obtained in terms of it according to [85]

$$\begin{aligned}
2\epsilon_2\alpha_3(p)T(p|k) = & (\epsilon_3 - \epsilon_2) \left\{ \frac{pk - \alpha_3(p)\alpha_2(k)}{\alpha_3(p) + \alpha_2(k)} [r_{01}(k) + r_{21}(k)e^{2i\alpha_1(k)D}] \right. \\
& \times I(-\alpha_3(p) - \alpha_2(k)|p - k) + \frac{pk + \alpha_3(p)\alpha_2(k)}{\alpha_3(p) - \alpha_2(k)} [r_{21}(k)r_{10}(k) + e^{2i\alpha_1(k)D}] \\
& \times I(-\alpha_3(p) + \alpha_2(k)|p - k) \left. \right\} \frac{e^{-i\alpha_1(k)D}}{t_{21}(k)t_{10}(k)} \\
& + (\epsilon_3 - \epsilon_2) \int_{-\infty}^{\infty} \frac{dq}{2\pi} \left\{ \frac{pq - \alpha_3(p)\alpha_2(q)}{\alpha_3(p) + \alpha_2(q)} [1 + r_{21}(q)r_{10}(q)e^{2i\alpha_1(q)D}] \right. \\
& \times I(-\alpha_3(p) - \alpha_2(q)|p - q) + \frac{pq + \alpha_3(p)\alpha_2(q)}{\alpha_3(p) - \alpha_2(q)} [r_{21}(q) + r_{10}(q)e^{-2i\alpha_1(q)D}] \\
& \times I(-\alpha_3(p) + \alpha_2(q)|p - q) \left. \right\} e^{-i\alpha_1(q)D} \frac{R(q|k)}{t_{21}(q)t_{10}(q)}. \quad (2.61)
\end{aligned}$$

Eq. (2.56) can be solved numerically [86]. This is done by replacing the infinite range of integration by a finite range $(-Q, Q)$. The integral over q is then replaced by a sum through the use of a numerical quadrature scheme. When p and k assume the values of the abscissas used in this quadrature scheme, a matrix equation for determining $R(q|k)$ is obtained, which can be solved by a standard linear equation solver algorithm.

In the case that the dielectric film is thin, so that $\sqrt{\epsilon_1}(\omega/c)d \ll 1$, Eq. (2.56) can be solved analytically [85]. In this case the function $I(\gamma|Q)$ can be expanded in powers of the surface profile function, and only the first two terms kept, with the result

$$I(\gamma|Q) = 2\pi\delta(Q) - i\gamma\hat{\zeta}(Q), \quad (2.62)$$

where $\hat{\zeta}(Q)$ is the Fourier transform of the surface profile function,

$$\hat{\zeta}(Q) = \int_{-\infty}^{\infty} dx_1 \zeta(x_1) \exp(-iQx_1). \quad (2.63)$$

If the solution of Eq. (2.56) is sought in the form

$$R(q|k) = 2\pi\delta(q - k)R_0(k) - \frac{t_{10}(q)e^{i\alpha_1(q)D}}{D(q)}\Phi(q|k) - \frac{t_{10}(k)e^{i\alpha_1(k)D}}{D(k)}\frac{2\alpha_0(k)}{\epsilon_0}, \quad (2.64)$$

where $R_0(k)$ is the Fresnel reflection coefficient of the prism–metal film–vacuum system,

$$R_0(k) = -\frac{r_{10}(k) + r_{31}(k)e^{2i\alpha_1(k)D}}{1 + r_{31}(k)r_{10}(k)e^{2i\alpha_1(k)D}}, \quad (2.65)$$

and

$$D(q) = d_{31}(q)[1 + r_{31}(q)r_{10}(q)e^{2i\alpha_1(q)D}], \quad (2.66)$$

then the equation satisfied by $\Phi(q|k)$ is

$$\begin{aligned} \Phi(p|k) = & \left(\frac{1}{\epsilon_3} - \frac{1}{\epsilon_2} \right) [pk - \alpha_3(p)\alpha_3(k) \frac{\epsilon_2}{\epsilon_3}] \hat{\zeta}(p-k) \\ & + \left(\frac{1}{\epsilon_3} - \frac{1}{\epsilon_2} \right) \int_{-\infty}^{\infty} \frac{dq}{2\pi} \frac{\hat{\zeta}(p-q)}{D(q)} \left\{ pq[1 + r_{10}(q)e^{-2i\alpha_1(q)D}] \right. \\ & \left. + \alpha_3(p)\alpha_1(q) \frac{\epsilon_2}{\epsilon_1} [1 - r_{01}(q)e^{2i\alpha_1(q)D}] \right\} \Phi(q|k) . \end{aligned} \quad (2.67)$$

When the dielectric film covers half the surface of the metal film, the surface profile function has the form $\zeta(x_1) = -d\theta(x_1)$, where $\theta(x)$ is the Heaviside unit step function, and the Fourier transform of this surface profile function is

$$\hat{\zeta}(Q) = -\frac{d}{i(Q - i\eta)} , \quad (2.68)$$

where η is a positive infinitesimal. If the film covers a part of the metal surface that is of length L , the surface profile function is $\zeta(x_1) = -d\theta(x_1)\theta(L - x_1)$, and its Fourier transform is

$$\hat{\zeta}(Q) = d \frac{e^{-iQL} - 1}{iQ} . \quad (2.69)$$

When $\hat{\zeta}(Q)$ is given by Eq. (2.68), Eq. (2.67) can be solved by the Wiener–Hopf technique [87]; when it is given by Eq. (2.69), Eq. (2.67) can be solved by the modified Wiener–Hopf technique [87].

To aid in the interpretation of the results obtained by the approach described above, we note that the structure employed is the standard Kretschmann attenuated total reflection configuration widely used in experiments for the excitation of surface plasmon polaritons at a metal–vacuum interface. In carrying out numerical calculations based on this structure parameters of the system have been chosen that are similar to those used in the experiment of Ref. [83]: the dielectric constant of the prism is $\epsilon_0 = 2.3$; the dielectric function of the metal film (silver) depends on the wavelength; the dielectric constant of the thin dielectric coating (SiO_2) is $\epsilon_2 = 2.1 + i0.02$; the thickness of the metal film is $D = 50.2\text{nm}$; and the thickness of the dielectric coating is $d = 30\text{nm}$.

In the absence of the dielectric film the Kretschmann configuration supports two surface electromagnetic waves. The thickness of the metal film is usually chosen such that the surface waves are decoupled, and are localized in the vicinity of the prism–metal and metal–vacuum interfaces. The surface plasmon polariton associated with the prism–metal interface is nonradiative, i.e. its field decays exponentially into the prism and the metal film. The surface plasmon associated with the metal–vacuum interface is a leaky wave, whose field decays exponentially into the vacuum, and is radiated into the prism. The dispersion curves of the system are depicted in Fig. 14 by the solid curves. When a thin dielectric film is deposited on the metal surface the dispersion curve of the surface wave associated with the metal–vacuum interface shifts, while that of the surface wave associated with the prism–metal interface practically does not feel the presence of the film. The dispersion curves of the surface polaritons in the system with a thin dielectric film on the metal surface are shown by dashed lines in Fig. 14.

In Fig. 15 we present a plot of the intensity of the scattered light, $|H_2(x_1, x_3|\omega)|^2$, as a function of the distance along the surface for incident light of wavelength $\lambda = 600\text{nm}$, in the case where the dielectric film of length $L = 200\lambda$ covers part of the metal surface, $0 < x_1 < 200\lambda$. In this figure the intensity of the

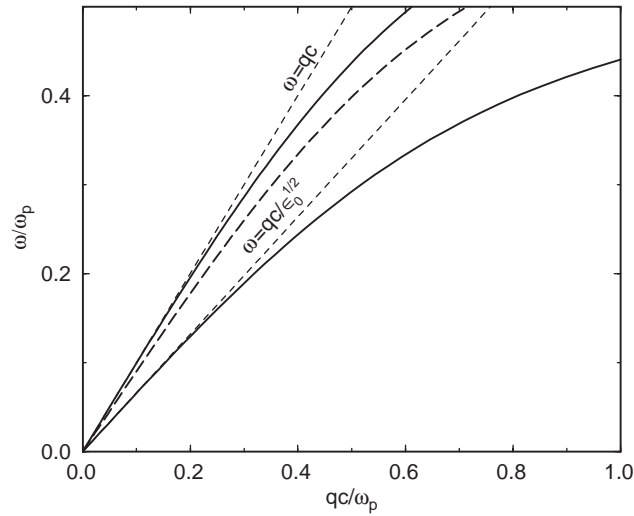


Fig. 14. The dispersion curves of the surface plasmon polaritons supported by the system studied in this section. The higher frequency solid curve is the dispersion curve of the surface plasmon polariton associated with the metal–vacuum interface in the absence of the dielectric film; the lower frequency solid curve is the dispersion curve of the surface plasmon polariton associated with the metal–vacuum interface in the absence of the dielectric film; the dashed curve is the dispersion curve of the surface plasmon polariton associated with the metal–vacuum interface in the presence of the dielectric film. (Ref. [85].)

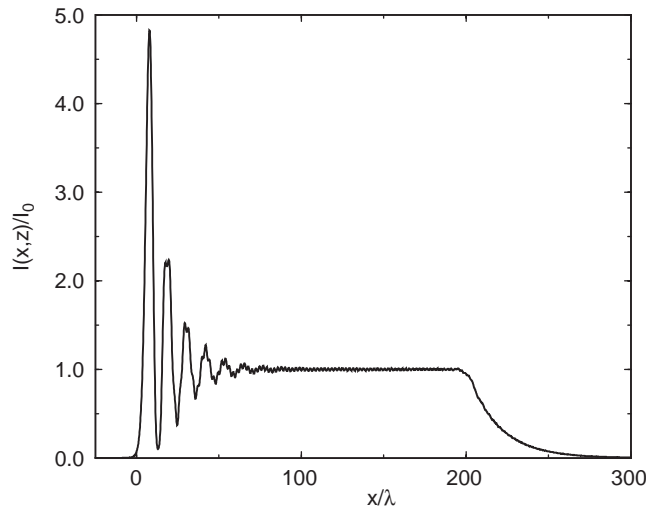


Fig. 15. The normalized intensity of the scattered light in the case where the dielectric film covers part of the metal surface of length $L = 200\lambda$, $\lambda = 600\text{nm}$, and $\theta_0 = 43.18^\circ$. (Ref. [86].)

scattered light is calculated from the results of a numerical solution [86] of Eq. (2.56) as a function of x_1/λ for a fixed value of x_3 , namely $x_3 = 120\lambda$. In these calculations the angle of incidence $\theta_0 = 43.18^\circ$ is the optimal angle for the excitation of surface plasmon polaritons on a clean metal–vacuum interface. The intensity of the scattered light was normalized by the reflectivity of the system when the dielectric film

totally covers the metal surface. As can be seen from Fig. 15 the intensity of the scattered light changes abruptly at the positions of the edges of the index step, showing that when the width L of the dielectric film is large, $L > \lambda$, it can easily be determined from plots of the intensity of the scattered light. Results of other calculations (not shown) indicate that even when the width of the dielectric film is as small as 5λ – 10λ , it can be determined in this way.

As can be seen from Fig. 15, oscillations of several different periods and different amplitudes can be observed in the intensity of the scattered light. In the region $x_1 < 0$ the short period oscillations are due to the interference of the incident surface plasmon polariton on the clean metal–vacuum interface with the reflected surface plasmon polariton. Their period is given by $T = \lambda/(2n_0)$, where n_0 is the index of refraction of surface plasmon polaritons on a clean metal surface, and is defined by $k = n_0(\omega/c)$, where in the Kretschmann geometry the wavenumber k coincides with the tangential component of the wave vector of the incident light, $k = \sqrt{\epsilon_0}(\omega/c) \sin \theta_0$. The oscillations of long period in the region $x_1 < 0$ are due to the interference between the reflected surface plasmon polariton and the volume (cylindrical) electromagnetic wave diffracted from the edge of the index step at $x_1 = 0$. Their period is given by $T = \lambda/(\sqrt{\epsilon_0} - n_0)$, and their amplitude is small, since the reflection of the surface plasmon polaritons is weak.

In the region to the right of the index step, $x_1 > L$, only short period oscillations are observed. They are due to the interference of the specularly reflected volume electromagnetic wave, which is in fact the incident surface plasmon polariton, with the volume (cylindrical) electromagnetic wave diffracted from the edge of the index step at $x_1 = L$. Their period is given by $T = \lambda/[\sqrt{\epsilon_0}(1 - \sin \theta_0)]$.

In the region of the x_1 -axis covered by the dielectric film, $0 < x_1 < L$, both short period and long period oscillations of the intensity are observed. The long period oscillations are due to the interference of the transmitted surface plasmon polariton at the metal film–dielectric film interface, which is excited due to diffraction at the edge $x_1 = 0$ of the dielectric film, and the specularly reflected volume electromagnetic wave. Their period is $T = \lambda/(n_1 - n_0)$, where n_1 is the index of refraction of surface plasmon polaritons at the metal film dielectric film interface, $k_{sp} = n_1(\omega/c)$. The oscillations of small period are due to the interference of the scattered volume (cylindrical) waves in the prism and the specularly reflected electromagnetic wave. Their period is given by $T = \lambda/(n_c - n_1)$, where $n_c \cong \sqrt{\epsilon_0}[1 - (\epsilon_2 d/\lambda)^2]$.

Thus, the surface plasmon polariton microscope can determine the spatial limits of a dielectric film on a metal surface. Moreover, from measurements at two different wavelengths the thickness and dielectric constant of the dielectric film can be determined. It has been found that the sensitivity of this method is better the thinner the dielectric film and the smaller the dielectric contrast between the film and vacuum [86]. Finally, the fact that the surface plasmon polariton microscope yields data in the far field makes it simpler to work with than a near-field microscope.

2.1.3. The interaction of a surface plasmon polariton with a subsurface defect

The theoretical study of the scattering of a surface wave by a subsurface defect has a long history. Its origins are in the work of Dean [88] and Ursell [89,90], who showed that a surface wave incident normally on a rigid cylinder of circular cross section submerged in an incompressible fluid of infinite depth passes over the cylinder with a change of phase but without a change of amplitude, and experiences no reflection, whatever the frequency of the incident wave, for any values of the radius and depth of the cylinder.

This work stimulated several subsequent papers [91–93] in which the effects of changing the cross section of the cylinder, of increasing the number of cylinders, of only partially submerging the cylinder, and

of making the depth of the fluid finite, on the scattering of a surface wave were investigated. The results of these investigations in turn raised the question of whether an effect analogous to the absence of reflection from a cylinder of circular cross section in a fluid of infinite depth exists for surface electromagnetic waves. The first study of this problem was carried out by Cullen [94], who considered the case of a plane p-polarized surface electromagnetic wave propagating over a corrugated metal surface, and incident normally on an infinitely long cylinder of elliptic cross section, whose cross section dimensions are small compared to the wavelength of the surface wave. By the use of an impedance boundary condition at the metal surface he showed that if the elliptic cylinder is perfectly conducting, and if the ratio of the semi-major and semi-minor axes of the cross section of the cylinder is suitably chosen, there exists one specific frequency of the surface wave at which the reflected wave vanishes.

This work was followed by several papers [95–97] in which generalizations of Cullen’s work were presented. These investigations, like Cullen’s were largely analytic in nature, and were based on simplifying approximations such as the use of impedance boundary conditions. With the increasingly widespread use of computers such approximations are no longer needed, and more realistic situations can be studied. In this section we outline a numerical approach to the problem of the scattering of a surface plasmon polariton from a subsurface defect [78].

The system we consider consists of a linear dielectric medium characterized by an isotropic dielectric constant ϵ_1 in the region $x_3 > 0$, and a linear dielectric medium characterized by an isotropic dielectric constant ϵ_3 in the region $x_3 < 0$. We assume that the interface $x_3 = 0$ separating these two media supports a p-polarized surface electromagnetic wave. We will be concerned with the situation in which the region $x_3 > 0$ is a lossless dielectric, while the region $x_3 < 0$ is a lossless metal. The dielectric constants ϵ_1 and ϵ_2 are, therefore real, and the conditions ensuring the existence of the surface wave are $\epsilon_1\epsilon_3 < 0$ and $\epsilon_1 + \epsilon_3 < 0$. We now introduce into the region $x_3 < 0$ a cylindrical defect of finite extent in the x_1 - and x_3 -directions, whose generators are parallel to the x_2 -axis. The resulting system is, therefore, invariant in the x_2 -direction. For simplicity we also assume that this cylinder does not intersect the plane $x_3 = 0$. The defect is characterized by an isotropic dielectric constant ϵ_2 , which is also assumed to be real. Consequently, the dielectric constant of the system is now position dependent, and can be written in the form

$$\epsilon(x_1, x_3) = \epsilon_1\theta(x_3) + \theta(-x_3)[\epsilon_3 + (\epsilon_2 - \epsilon_3)S(x_1, x_3)] , \quad (2.70)$$

where $\theta(x)$ is the Heaviside unit step function, and $S(x_1, x_3)$ is the characteristic function of the cylinder’s cross section, i.e. $S(x_1, x_3) = 1$ if the point (x_1, x_3) lies within the cross section of the cylinder, and $S(x_1, x_3) = 0$ if it lies outside it.

If we write a typical component of the electromagnetic field in this system in the form

$$A_i(x_1, x_3; t) = A_i(x_1, x_3|\omega) \exp(-i\omega t) , \quad (2.71)$$

the Maxwell equations for the two nonzero components $E_1(x_1, x_3|\omega)$ and $E_3(x_1, x_3|\omega)$ of the electric vector of a p-polarized electromagnetic field, whose plane of incidence is the x_1x_3 -plane, become

$$\begin{pmatrix} \frac{\partial^2}{\partial x_3^2} + \epsilon(x_3)\frac{\omega^2}{c^2} & -\frac{\partial^2}{\partial x_1\partial x_3} \\ -\frac{\partial^2}{\partial x_1\partial x_3} & \frac{\partial^2}{\partial x_1^2} + \epsilon(x_3)\frac{\omega^2}{c^2} \end{pmatrix} \begin{pmatrix} E_1(x_1, x_3|\omega) \\ E_3(x_1, x_3|\omega) \end{pmatrix} \\ = -\frac{\omega^2}{c^2}\theta(-x_3)[(\epsilon_2 - \epsilon_3)S(x_1, x_3)] \begin{pmatrix} E_1(x_1, x_3|\omega) \\ E_3(x_1, x_3|\omega) \end{pmatrix} , \quad (2.72)$$

where $\epsilon(x_3) = \epsilon_1\theta(x_3) + \epsilon_3\theta(-x_3)$. The solutions of Eq. (2.72) are subject to the boundary conditions

$$E_1|_{x_3=0+} = E_1|_{x_3=0-}, \quad \epsilon_1 E_3|_{x_3=0+} = \epsilon_3 E_3|_{x_3=0-} \quad (2.73)$$

at the interface $x_3 = 0$, and to vanishing boundary conditions as $x_3 \rightarrow \pm\infty$.

The partial differential equation (2.72) and associated boundary conditions can be transformed into an integral equation

$$E_i(x_1, x_3|\omega) = E_i^{(0)}(x_1, x_3|\omega) + 4\pi \frac{\omega^2}{c^2} (\epsilon_2 - \epsilon_3) \sum_{j=1,3} \int_{-\infty}^{\infty} dx'_1 \int_{-\infty}^0 dx'_3 \\ \times G_{ij}(x_1, x_3|x'_1, x'_3) S(x'_1, x'_3) E_j(x'_1, x'_3|\omega), \quad i = 1, 3, \quad (2.74)$$

where $E_i^{(0)}(x_1, x_3|\omega)$ is the solution of the homogeneous form of Eq. (2.72) and the boundary conditions (2.73), and represents the incident surface plasmon polariton:

$$E_1^{(0)}(x_1, x_3|\omega) = E^{(0)} e^{ik(\omega)x_1 - \beta_1(\omega)x_3}, \quad x_3 > 0 \quad (2.75a)$$

$$= E^{(0)} e^{ik(\omega)x_1 + \beta_3(\omega)x_3}, \quad x_3 < 0, \quad (2.75b)$$

$$E_3^{(0)}(x_1, x_3|\omega) = \frac{ik(\omega)}{\beta_1(\omega)} E^{(0)} e^{ik(\omega)x_1 - \beta_1(\omega)x_3}, \quad x_3 > 0 \quad (2.76a)$$

$$= -\frac{ik(\omega)}{\beta_3(\omega)} E^{(0)} e^{ik(\omega)x_1 + \beta_3(\omega)x_3} \quad x_3 < 0, \quad (2.76b)$$

with

$$k(\omega) = \frac{\omega}{c} \left(\frac{\epsilon_1 \epsilon_3}{\epsilon_1 + \epsilon_3} \right)^{1/2}, \quad (2.77a)$$

$$\beta_1(\omega) = \frac{\omega}{c} \left(\frac{-\epsilon_1^2}{\epsilon_1 + \epsilon_3} \right)^{1/2}, \quad (2.77b)$$

$$\beta_3(\omega) = \frac{\omega}{c} \left(\frac{-\epsilon_3^2}{\epsilon_1 + \epsilon_3} \right)^{1/2}. \quad (2.77c)$$

The functions $G_{ij}(x_1, x_3|x'_1, x'_3)$ ($i, j = 1, 3$) are the elements of the Green's tensor for the differential operator on the left hand side of Eq. (2.72). They satisfy the differential equation

$$\begin{pmatrix} \frac{\partial^2}{\partial x_3^2} + \epsilon(x_3) \frac{\omega^2}{c^2} & -\frac{\partial^2}{\partial x_1 \partial x_3} \\ -\frac{\partial^2}{\partial x_1 \partial x_3} & \frac{\partial^2}{\partial x_1^2} + \epsilon(x_3) \frac{\omega^2}{c^2} \end{pmatrix} \begin{pmatrix} G_{11}(x_1, x_3|x'_1, x'_3) & G_{13}(x_1, x_3|x'_1, x'_3) \\ G_{31}(x_1, x_3|x'_1, x'_3) & G_{33}(x_1, x_3|x'_1, x'_3) \end{pmatrix} \\ = -4\pi\theta(-x'_3)\delta(x_1 - x'_1)\delta(x_3 - x'_3) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad (2.78)$$

subject to the boundary conditions

$$G_{1j}(x_1, x_3|x'_1, x'_3)|_{x_3=0+} = G_{1j}(x_1, x_3|x'_1, x'_3)|_{x_3=0-}, \quad j = 1, 3, \quad x'_3 < 0 \quad (2.79a)$$

$$\epsilon_1 G_{3j}(x_1, x_3|x'_1, x'_3)|_{x_3=0+} = \epsilon_3 G_{3j}(x_1, x_3|x'_1, x'_3)|_{x_3=0-}, \quad j = 1, 3, \quad x'_3 < 0, \quad (2.79b)$$

and either outgoing wave or vanishing boundary conditions at $x_3 = \pm\infty$. They can be represented in the form

$$G_{ij}(x_1, x_3|x'_1, x'_3) = \int_{-\infty}^{\infty} \frac{dk}{2\pi} g_{ij}(k|x_3, x'_3) \exp[ik(x_1 - x'_1)]. \quad (2.80)$$

Although x'_3 is required to be negative because the cylindrical defect is in the region $x'_3 < 0$, x_3 can be either positive or negative. If we denote by $g_{ij}^>(k|x_3, x'_3)$ the Fourier coefficient when $x_3 > 0$, $x'_3 < 0$, and by $g_{ij}^<(k|x_3, x'_3)$ the Fourier coefficient when $x_3 < 0$, $x'_3 < 0$, we find that these functions are given by

$$g_{11}^>(k|x_3, x'_3) = \frac{4\pi i}{\epsilon_3(\omega^2/c^2)} \frac{\alpha_1(k)\alpha_3(k)}{\epsilon_1\alpha_3(k) + \epsilon_3\alpha_1(k)} e^{i\alpha_1(k)x_3 - i\alpha_3(k)x'_3}, \quad (2.81a)$$

$$g_{13}^>(k|x_3, x'_3) = -\frac{4\pi i}{(\omega^2/c^2)} \frac{k\alpha_1(k)}{\epsilon_1\alpha_3(k) + \epsilon_3\alpha_1(k)} e^{i\alpha_1(k)x_3 - i\alpha_3(k)x'_3}, \quad (2.81b)$$

$$g_{31}^>(k|x_3, x'_3) = -\frac{4\pi i}{(\omega^2/c^2)} \frac{k\alpha_3(k)}{\epsilon_1\alpha_3(k) + \epsilon_3\alpha_1(k)} e^{i\alpha_1(k)x_3 - i\alpha_3(k)x'_3}, \quad (2.81c)$$

$$g_{33}^>(k|x_3, x'_3) = \frac{4\pi i}{(\omega^2/c^2)} \frac{k^2}{\epsilon_1\alpha_3(k) + \epsilon_3\alpha_1(k)} e^{i\alpha_1(k)x_3 - i\alpha_3(k)x'_3}, \quad (2.81d)$$

where

$$\alpha_i(k) = (k^2 - \epsilon_i(\omega)(\omega^2/c^2))^{1/2} \quad \text{Re } \alpha_i(k) > 0, \quad \text{Im } \alpha_i(k) < 0, \quad (2.82)$$

and

$$g_{11}^<(k|x_3, x'_3) = \frac{2\pi i\alpha_3(k)}{\epsilon_3(\omega^2/c^2)} e^{i\alpha_3(k)|x_3 - x'_3|} - \frac{2\pi i\alpha_3(k)}{\epsilon_3(\omega^2/c^2)} \frac{\epsilon_1\alpha_3(k) - \epsilon_3\alpha_1(k)}{\epsilon_1\alpha_3(k) + \epsilon_3\alpha_1(k)} e^{-i\alpha_3(k)(x_3 + x'_3)}, \quad (2.83a)$$

$$g_{13}^<(k|x_3, x'_3) = -\frac{2\pi ik}{\epsilon_3(\omega^2/c^2)} e^{i\alpha_3(k)|x_3 - x'_3|} \text{sgn}(x_3 - x'_3) - \frac{2\pi ik}{\epsilon_3(\omega^2/c^2)} \frac{\epsilon_1\alpha_3(k) - \epsilon_3\alpha_1(k)}{\epsilon_1\alpha_3(k) + \epsilon_3\alpha_1(k)} e^{-i\alpha_3(k)(x_3 + x'_3)}, \quad (2.83b)$$

$$g_{31}^<(k|x_3, x'_3) = -\frac{2\pi ik}{\epsilon_3(\omega^2/c^2)} e^{i\alpha_3(k)|x_3 - x'_3|} \text{sgn}(x_3 - x'_3) - \frac{2\pi ik}{\epsilon_3(\omega^2/c^2)} \frac{\epsilon_1\alpha_3(k) - \epsilon_3\alpha_1(k)}{\epsilon_1\alpha_3(k) + \epsilon_3\alpha_1(k)} e^{-i\alpha_3(k)(x_3 + x'_3)}, \quad (2.83c)$$

$$g_{33}^{\leq}(k|x_3, x'_3) = -\frac{4\pi}{\epsilon_3(\omega^2/c^2)}\delta(x_3 - x'_3) + \frac{2\pi ik^2}{\alpha_3(k)\epsilon_3(\omega^2/c^2)}e^{i\alpha_3(k)|x_3-x'_3|} - \frac{2\pi ik^2}{\alpha_3(k)\epsilon_3(\omega^2/c^2)}\frac{\epsilon_1\alpha_3(k) - \epsilon_3\alpha_1(k)}{\epsilon_1\alpha_3(k) + \epsilon_3\alpha_1(k)}e^{-i\alpha_3(k)(x_3+x'_3)}. \quad (2.83d)$$

We note that in the case of interest here, namely where ϵ_1 is real and positive and $\epsilon_3(\omega)$ is real and negative, each of the functions $g_{\alpha\beta}^{\geq}(k|x_3, x'_3)$ and $g_{\alpha\beta}^{\leq}(k|x_3, x'_3)$ ($\alpha, \beta = 1, 3$) has a simple pole at $k = \pm k(\omega)$, where $k(\omega)$, the wavenumber of the surface plasmon polariton of frequency ω supported by the dielectric–metal interface, is the solution of $\epsilon_1\alpha_3(k) + \epsilon_3\alpha_1(k) = 0$, and is given by (Eq. (2.77a)) $k(\omega) = (\omega/c)\{\epsilon_1\epsilon_3(\omega)/[\epsilon_1 + \epsilon_3(\omega)]\}^{1/2}$.

The integrations over x'_1 and x'_3 in Eq. (2.74) are carried out over the cross section of the cylindrical defect. Consequently, the integral equation (2.74) can be solved by the method of moments [98]. In this approach the cross section of the defect is mapped into a set of rectangular cells of dimension Δx and Δz parallel to the x_1 - and x_3 -axis, respectively, which are small enough that the electric field can be assumed to be constant within each cell. In this way we obtain the pair of coupled matrix equations [99]

$$E_1^i = E_{01}^i + M_{11}^{ij}E_1^j + M_{13}^{ij}E_3^j, \quad (2.84a)$$

$$E_3^i = E_{03}^i + M_{31}^{ij}E_1^j + M_{33}^{ij}E_3^j, \quad (2.84b)$$

where repeated superscripts in the same term are to be summed over. In Eqs. (2.84) \mathbf{E}^i and \mathbf{E}_0^i are the total and incident fields at the center (x_{1i}, x_{3i}) of the i th cell, the cells i and j are restricted to the cross section of the defect cylinder, and the matrix elements $M_{\alpha\beta}^{ij}$ are given by

$$M_{\alpha\beta}^{ij} = 4\pi\frac{\omega^2}{c^2}(\epsilon_2 - \epsilon_3) \int_{x_{1j}-\frac{1}{2}\Delta x}^{x_{1j}+\frac{1}{2}\Delta x} dx'_1 \int_{x_{3j}-\frac{1}{2}\Delta z}^{x_{3j}+\frac{1}{2}\Delta z} dx'_3 G_{\alpha\beta}(x_{1i}, x'_1|x_{3i}, x'_3). \quad (2.85)$$

Care must be taken in the numerical evaluation of the integrals in Eq. (2.85) because of the singularity of $G_{\alpha\beta}(x_1, x_3|x'_1, x'_3)$ when $(x_1, x_3) = (x'_1, x'_3)$, and because of the pole that $g_{\alpha\beta}^{\leq}(k|x_3, x'_3)$ possesses at the wavenumber of the surface plasmon polariton of frequency ω . A way of carrying out these integrations is described in detail in Ref. [99].

The solution of Eqs. (2.84) gives the components of the total electric field within the cross section of the defect cylinder. Once they are known, the components of the total field at any point outside this region can be obtained from Eqs. (2.84) by allowing the point (x_{1i}, x_{3i}) to be any point outside the region, while the sum over j still runs over the cells within this region.

The surface plasmon polariton reflection and transmission coefficients can be calculated in the following way [78,99]. For the calculation of the reflection coefficient R the incident field and the surface plasmon polariton contribution to the scattered field $\mathbf{E}_s(x_1, x_3|\omega)$ are determined at the same point (x_{1i}, x_{3i}) in the metal in front of the defect cylinder, and R is then obtained from [78]

$$R = \left| \frac{E_{s1}(x_{1i}, x_{3i}|\omega)}{E_{01}(x_{1i}, x_{3i}|\omega)} \right|^2. \quad (2.86)$$

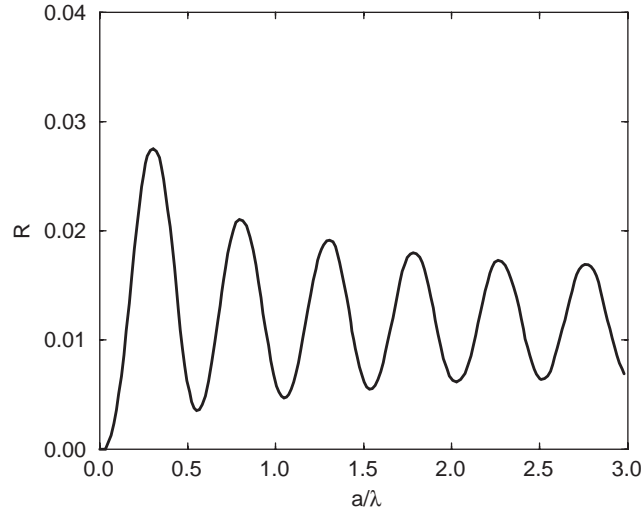


Fig. 16. Reflection coefficient for a system consisting of a dielectric cylinder totally embedded in a metal. $\epsilon_1 = 1$, $\epsilon_2 = 16$, $\epsilon_3 = -3.472$. $b = \lambda/20$, $h = \lambda/20$, $\Delta x = \Delta z = \lambda/40$. (Ref. [78].)

Similarly, the incident field and the total field are calculated at the same point (x_{1i}, x_{3i}) in the metal behind the defect cylinder, and the transmission coefficient T is then obtained from [78]

$$T = \left| \frac{E_1(x_{1i}, x_{3i}|\omega)}{E_{01}(x_{1i}, x_{3i}|\omega)} \right|^2. \quad (2.87)$$

The fraction S of the power in the incident surface plasmon polariton that is converted into volume electromagnetic waves in the lossless dielectric medium can also be calculated on the basis of results obtained in Ref. [78].

In Fig. 16 is plotted the surface plasmon polariton reflection coefficient for a system consisting of a dielectric cylinder of rectangular cross section embedded in a semi-infinite metal. The edge of the cylinder parallel to the x_1 -axis is denoted by a ; the edge of the cylinder parallel to the x_3 -axis is $b = \lambda/20$; and the dielectric constant of the cylinder is $\epsilon_2 = 16$. The top of the cylinder is at a depth $h = \lambda/20$. The dielectric constant $\epsilon_1 = 1$ (vacuum), and the dielectric constant of the metal is $\epsilon_3 = -3.472$. In the numerical calculations leading to Fig. 16 it was assumed that $\Delta x = \Delta z = \lambda/40$. The reflection coefficient R is plotted as a function of a/λ in Fig. 16. It is seen to be an oscillating function of this variable. The oscillations arise from the incident surface plasmon exciting a second surface polariton in the rectangular dielectric cylinder, whose wavenumber $k_{sp}^{(2)}$ must be almost equal to that of the single surface polariton at the same frequency in the four layer system obtained when the length a of the dielectric particle is allowed to become infinite. This second polariton will be reflected from the right edge of the rectangular dielectric cylinder, giving rise to an oscillating dependence of the reflection coefficient R on the length a of the cylinder with a period $\Delta a_R = 2\pi/2k_{sp}^{(2)}$.

The numerically implemented volume integral equation approach to the scattering of a surface plasmon polariton by a subsurface defect in a metal enables the reflection and transmission coefficients of the surface wave to be calculated, together with the fraction of the power in the incident wave that is converted

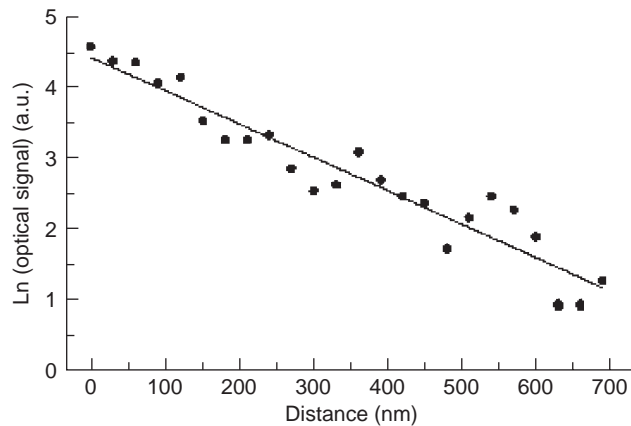


Fig. 17. Dependence of the optical signal on the tip–surface distance measured with a smooth silver film: (dots) experimental data and (line) linear fit. (Ref. [59].)

into volume electromagnetic waves in the vacuum region above the metal, and in the near field. Knowledge of the dispersion curves of surface plasmon polaritons in the corresponding multilayer system (i.e. the system where the defect of rectangular cross section has an infinite length) plays a key role in understanding the nature of the scattering process. The reflection and transmission coefficients as functions of the length of the defect display an interference structure that yields information about the wavenumbers of the surface waves, and thus about the dielectric constant of the defect. This result might be useful for determining the dielectric constant of a waveguide whose geometry is known.

2.1.4. Near-field imaging of the SPP interaction with a single surface defect

Detailed studies of the surface polariton interaction with individual surface defects and elementary scattering processes on defects of different sizes have been performed using artificially structured silver films [58,59]. The surface polaritons have been resonantly excited on a silver surface in the Kretschmann configuration. The quality of the initial silver films [59] was good enough to provide the homogeneous electromagnetic field distribution over the surface related to the resonant SPP excitation. The optical signal dependence on the distance between the tip and the silver surface exhibited an exponential decrease with distance $\text{Ln}(I) = -2k_3(i)x_3$ superimposed on a small background (Fig. 17). This background signal is due to the scattering of SPPs into light over the entire illuminated film area. The measured value of the SPP field extension above a surface ($k_3(i) = 1/416 \text{ nm}^{-1}$) is consistent with the angle of SPP excitation ($k_3(i) = 1/412 \text{ nm}^{-1}$). Individual defects have been created on a surface using laser ablation through an uncoated tapered fiber tip [100]. By changing the power of the ablating UV laser beam, defects of different sizes can be fabricated in a silver film. Scattering of SPP by such defects is described below for different sizes of the defects.

The optical near-field distribution (Fig. 18) around the large bumps ($\sim 100 \text{ nm}$ height and $\sim 2 \mu\text{m}$ radius) exhibits strong maxima at the places of the bumps. In addition, narrow SPP beams of about $1 \mu\text{m}$ width scattered from the bumps in the direction of the exciting light can be seen. The far-field images of such beams taken at the distance of about $7 \mu\text{m}$ from the surface is completely dark on the same halftone scale. A number of images recorded at intermediate distances have shown lateral spreading and exponential attenuation of the beam brightness with distance from the surface (Fig. 18c,d). (Please note that the

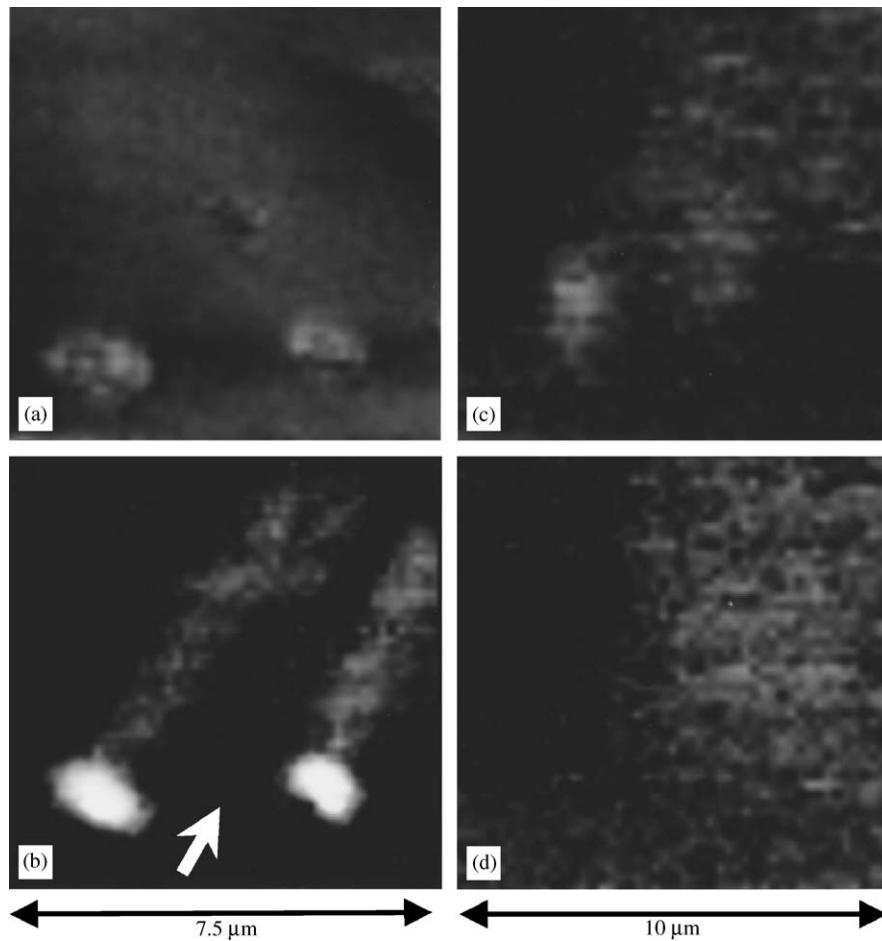


Fig. 18. Topography (a) and optical near-field intensity distribution (b) around large bumps on a silver surface. Far-field images (c) and (d) have been obtained at distances of 1 and 4 μm from the surface, respectively. The gray-scale of the topographical image corresponds to 140 nm. The direction of the SPP excitation is shown by the arrow. (Ref. [59].)

contrast of all images is internal and not relative to one another.) The distance dependencies confirm that these beams are related to surface polaritons. The appearance of such forward scattered SPP beams is rather common near surface features of larger sizes. If the defect has a complex shape as in Fig. 19, the SPP beam could arise from some small region of the defect. The cross section of the beam along its length shows the exponential decay of the SPP intensity with the distance from the defect consistent with the SPP propagation length (Fig. 19d). Some surface defects give rise not only to forward scattered SPP beams but also to, generally weaker, SPP beams scattered in the backward direction (Fig. 20). Thus, the scattering of surface polaritons into surface polaritons takes place at this kind of defect. It is important to note that the gradient of the topography variation is quite low in the case of the bumps discussed. In addition, despite their significant height, the extension length of the SPP field in air is noticeably larger than the defect height in this case. These two reasons result in quite a low efficiency of the SPP scattering on such surface features.

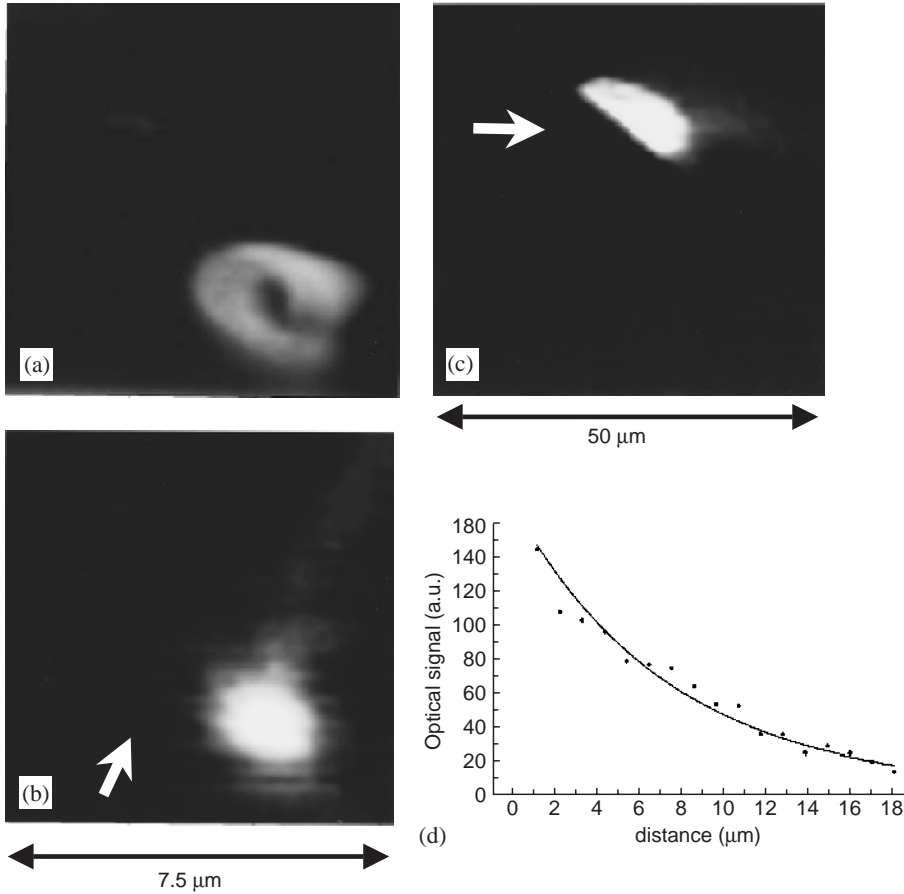


Fig. 19. Forward-scattering of SPPs by surface defects of complex shapes: topography (a) and optical near-field intensity distribution (b) around a defect on a silver surface; (c) the near-field intensity distribution around a large defect. (d) cross section of the optical near-field image along a forward scattered SPP beam as shown in (c). The gray-scale of the topographical image corresponds to 400 nm. The direction of the SPP excitation and cross section is indicated by the arrow. (Ref. [59].)

The effect of the scattered SPP beam appearance in the forward and backward directions can be understood directly from the Huygens–Fresnel principle in a two-dimensional geometry considering a surface defect as a set of secondary SPP sources. For example, considering a set of 10 secondary sources of SPP cylindrical waves equally spaced along a straight line, the resulting SPP intensity distribution over a surface can be found by a coherent superposition of the scattered SPPs from individual sources:

$$I \sim \left(\sum_n \frac{e^{i(kr_n + \phi_n)}}{r_n^{1/2}} \right)^2, \quad (2.88)$$

where r_n is the distance between the point of observation and the n th secondary source. This SPP intensity distribution is shown in Fig. 21(a) in the case where the phases of the secondary SPP sources vary linearly with the source position n . An SPP beam similar to the one observed in Fig. 18(b) is clearly

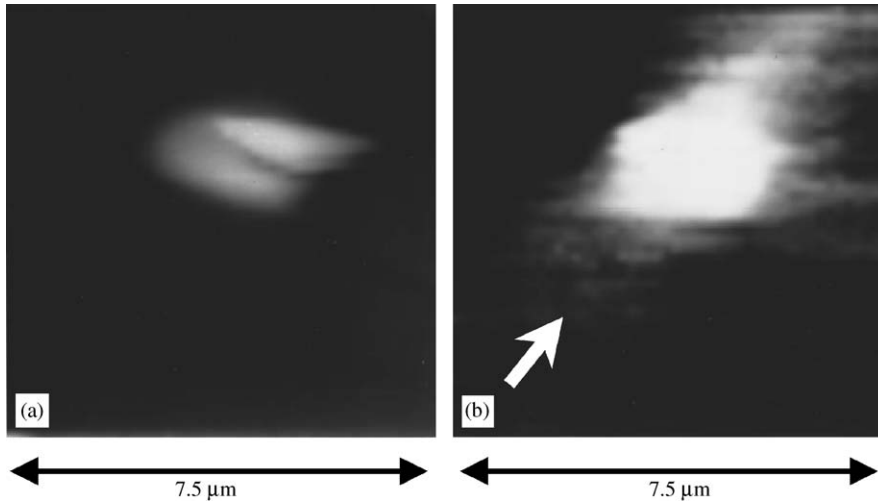


Fig. 20. Topography (a) and optical near-field intensity distribution (b) around a surface defect exhibiting both forward and backward scattered SPPs. The gray-scale of the topographical image corresponds to 400 nm. The direction of the SPP excitation is shown by the arrow. (Ref. [59].)

visible. The primary reason for the appearance of this SPP beam is the weak $\sim r^{-1/2}$ distance dependence of the electromagnetic field of cylindrical SPP waves from the point sources. For comparison, the intensity distribution related to the same source but calculated using the 3D distance dependence ($\sim r^{-1}$) for SPP scattering in light is shown in Fig. 21(b): no forward scattered beam appears in this case. The phase shift between the different secondary SPP waves arises in the experiment due to the complex shape of the surface defects resulting from the “angular” illumination of the different parts of the defect by the incident plane SPP wave. Fig. 22 presents modeling of the data shown in Fig. 18. Two sets of secondary SPP sources representing two surface defects are positioned at different distances from each other. The resulting intensity distribution (Fig. 22c) demonstrates a very close resemblance to the experiment (Fig. 18b). Two surface defects can be positioned within a few microns from each other giving rise to distinctly separate scattered SPP beams.

SPP scattering on some small surface defects (~ 50 nm height and $\lesssim 1$ μm radius) has a completely different nature (Figs. 23 and 24). Bright near-field maxima are not seen around these defects anymore. Instead, prominent shadows are observed behind the defects (in the direction of the SPP propagation). For different defects the shadows have somewhat different but always large angular widths. Such angular width cannot be accounted for by diffraction. The diffraction angle can be easily estimated from the uncertainty principle: $\phi \sim \delta k_{\text{sp}}/k_{\text{sp}} \sim \lambda_{\text{sp}}/(2\pi d)$, where d is the size of a scatterer. A micrometer size scatterer should give a diffraction angle of the order of 5° . This value is consistent with the small visible width of the SPP beams scattered from the larger defects (Fig. 18b). The detailed theory of SPP scattering by circularly-symmetric defects has provided an explanation of the shadow-like effect (Section 2.1.1).

Defects of similar size but different topology scatter SPP differently. In Fig. 24 one can see three defects of about the same lateral size: two of them are shallow bumps, the third is a pit of about the same vertical size. The shallow bumps seem not to affect the SPP propagation at all: as their height is smaller than the height of the large bumps discussed above, the SPP scattering efficiency is further reduced. On the other

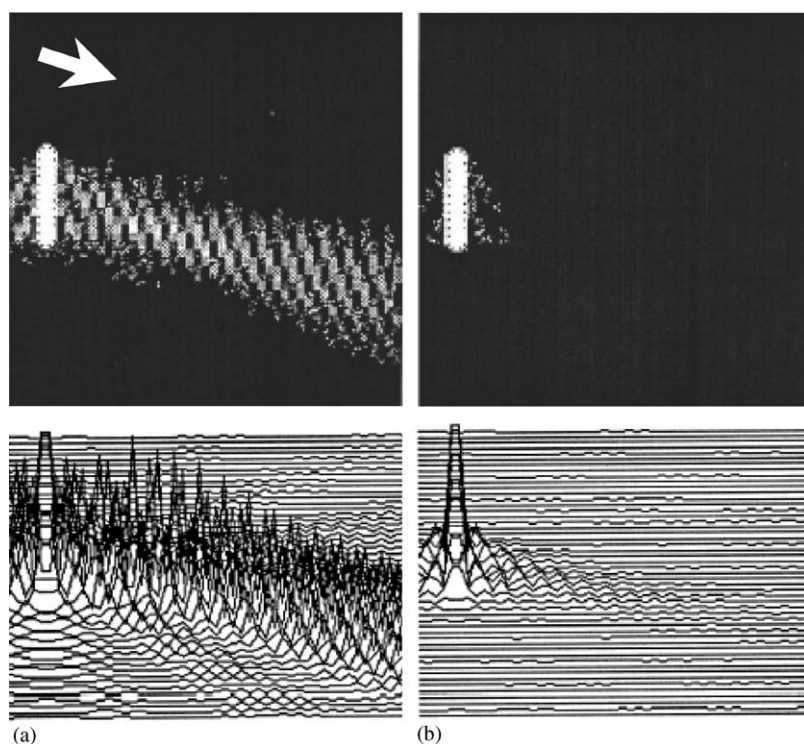


Fig. 21. Near-field intensity distributions ($10 \times 10 \mu\text{m}^2$) calculated for the SPP scattering on the defect modelled by 10 coherent secondary sources equally spaced along a $2 \mu\text{m}$ line: (a) SPP scattering into SPP and (b) SPP scattering into light. A pronounced forward-scattered SPP beam appears in the former case due to two-dimensional scattering. The direction of the incident plane SPP wave is shown by the arrow. (Ref. [59].)

hand, the pit gives a distinct angular shadow. This difference in the efficiency of the SPP scattering is related to the SPP field extension from the surface into the metal. In the case of a silver film, the SPP decay length in the metal is more than 10 times shorter ($\sim 30 \text{ nm}$) than the one in the air. Thus, surface polaritons have a much stronger interaction with small pits than with bumps of the same vertical size. This agrees well with the theoretical predictions.

2.1.5. Interference of surface plasmon polariton beams

Significantly stronger SPP scattering into SPPs on surface bumps has been observed with gold films [17]. The different relation between the defect sizes and SPP parameters (shorter SPP wavelength and an almost two times smaller field extension length above the surface) than in the case of silver films results in a higher efficiency of the in-plane SPP scattering on defects of the same size. The presence of strongly scattered SPPs leads to pronounced fringe patterns due to the interference of the incident and scattered surface polaritons. The relatively smooth gold films used in these experiments had, in addition to the small-scale roughness, a set of micron-sized (in lateral direction) bumps sparsely placed on a surface at distances of several tens of microns between them. The large-scale defects dominate the SPP scattering, providing the realization of the single-scattering regime similar to that produced by the individual surface defects on the silver film described above.

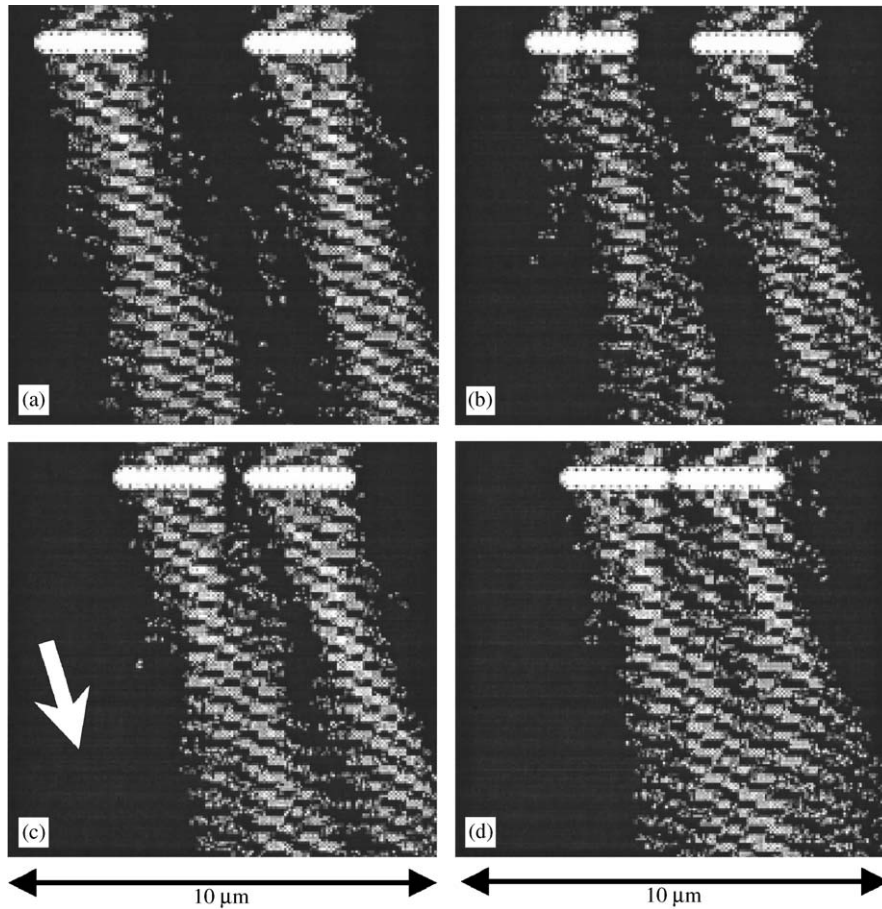


Fig. 22. Near-field intensity distributions calculated for SPP scattering on two surface defects of $2\ \mu\text{m}$ size (as in Fig. 21) positioned at different distances from each other. The distance between defect centers is (a) $5\ \mu\text{m}$, (b) $4\ \mu\text{m}$, (c) $3\ \mu\text{m}$, and (d) $2.5\ \mu\text{m}$. The direction of the incident plane SPP wave is shown by the arrow. (Ref. [59].)

This type of SPP scattering has been studied using a two-layer structure that allows exciting two SPP modes, i.e. a fast SPP (FSPP) on the gold–air interface and a slow SPP (SSPP) on the gold–fluoride (internal) interface. The excitation of both SPPs results in the pronounced minima in the angular dependence of the reflectivity of p-polarized light (Fig. 25). The propagation length of the SPP has been estimated to be about $8\ \mu\text{m}$. Thus, for large scale roughness the single scattering regime conditions $L_{\text{SP}} \sim l$ are satisfied for both fast and slow surface polaritons. The ratio between the maximum value of the SPP field (at the interface) and the value of the same SPP field at the other interface has been calculated to be about 6 for both SPPs. This ratio is large enough for detecting the optical signal from the SPP confined to the internal interface but sufficiently small to neglect the interaction between the different SPP modes.

Topographical images of the surface reveal a relatively smooth external interface with average surface variation of about $10\ \text{nm}$ (small-scale roughness). In addition, widely spaced micron-sized bumps with $100\text{--}200\ \text{nm}$ heights have been observed (Fig. 26 a). The surface polaritons propagating in the vertical

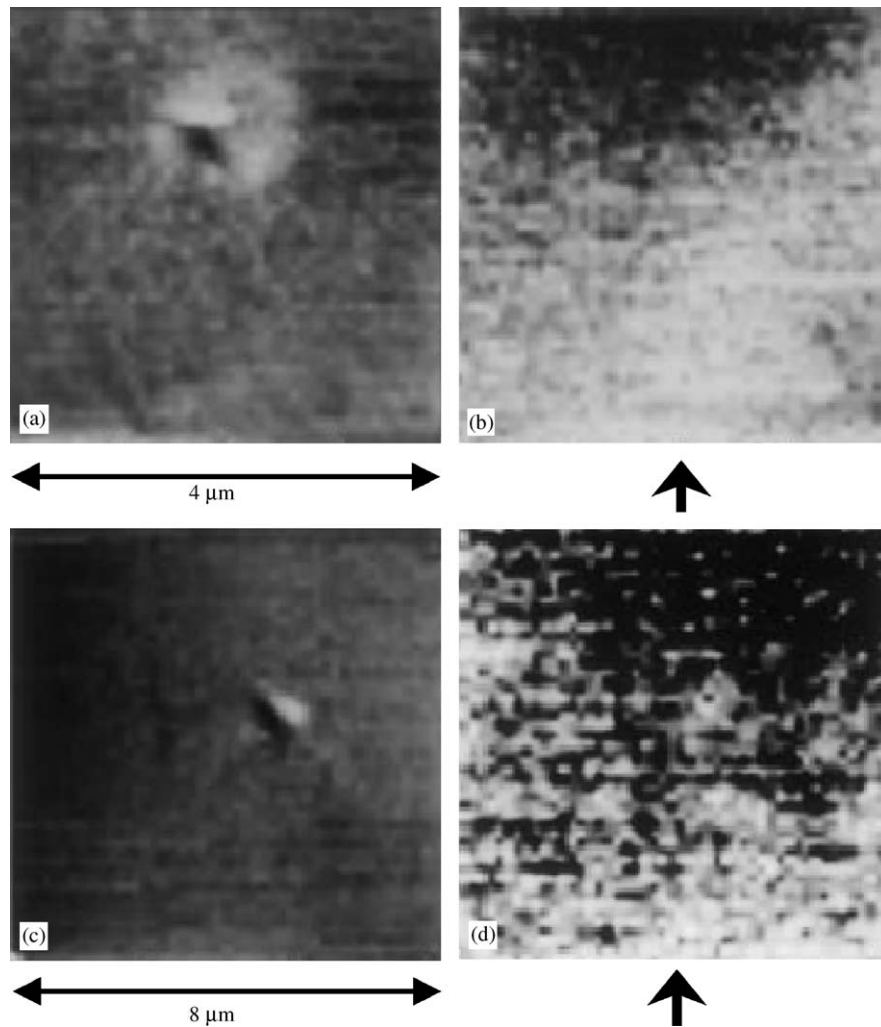


Fig. 23. Topography (a,c) and optical near-field intensity distributions (b,d) around small surface features on a silver surface. The gray-scale of the topographical images corresponds to 100 nm. The direction of the SPP excitation is shown by the arrow. Shadows are seen behind the defects. (Ref. [59].)

direction from bottom to top with respect to the images have been resonantly excited in the Kretschmann geometry. The near-field optical images obtained with the surface polariton excited on an air–metal interface exhibit a well pronounced interference pattern related to the interference between the surface polaritons resonantly excited on the surface, and the scattered surface polaritons (Fig. 26b). The average optical signal was about 20 times smaller if the angle of incidence was out of resonance by about 2° or if the tip was moved $1 \mu\text{m}$ away from the surface. This indicates the relatively low efficiency of the SPP scattering into light by such defects, while the in-plane scattering processes are important judging from the visibility of the SPP interference pattern recorded in the near-field. The far-field optical image has also exhibited an interference pattern (Fig. 26c), but its period was considerably larger, corresponding to

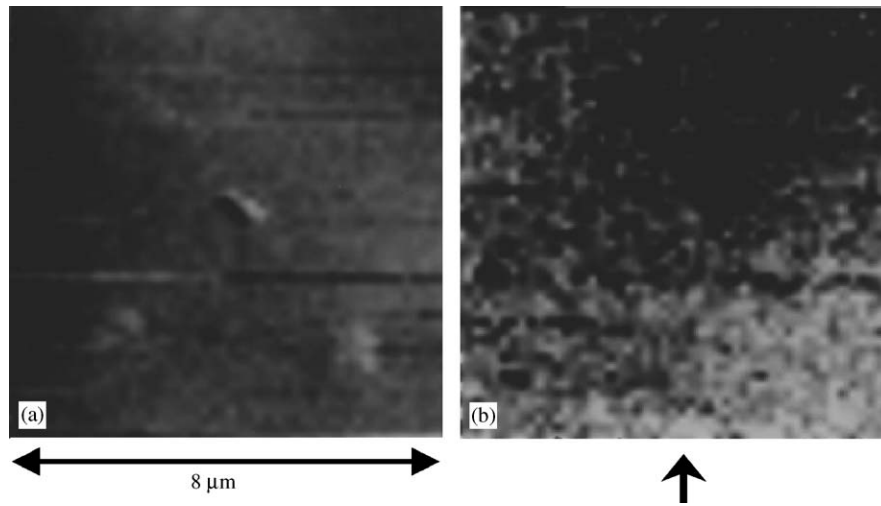


Fig. 24. Topography (a) and optical near-field intensity distribution (b) around small bumps and craters on a surface. The gray-scale of the topographical image corresponds to 90 nm. The direction of the SPP excitation is shown by the arrow. (Ref. [59].)

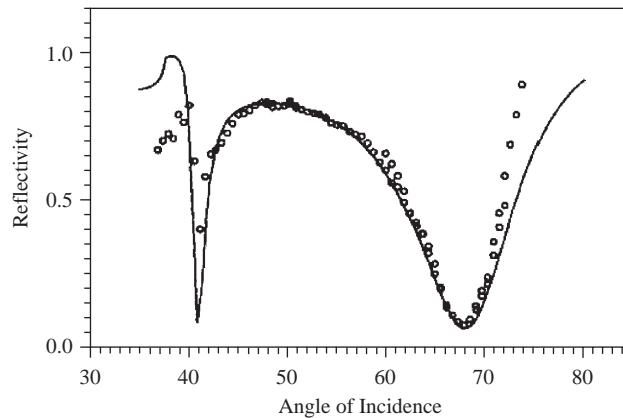


Fig. 25. Angular excitation spectra of surface polaritons measured (circles) and calculated (solid lines) for the two-layer gold structure. A smooth gold film (thickness 60 nm) is deposited on a magnesium fluoride (thickness 170 nm, $n = 1.28$) coated glass prism ($n = 1.64$). (Ref. [17].)

the wavelength of light propagating in air (Fig. 27). The interference patterns can be used to determine directly the SPP wavelength. Taking into account the propagation direction of the surface polaritons, the FSPP wavelength can be estimated to be about 550 nm in good agreement with the value calculated from the resonant angle of FSPP excitation ($\lambda_{\text{sp}} \approx 590$ nm).

Two near-field optical images generated in turn due to fast and slow SPP modes approximately at the same place on a surface are shown in Fig. 28. The average optical signal detected using the SSPP propagating along the internal gold-fluoride interface was about 20 times smaller than that due to the gold-air SPP, which agrees with estimates of the SPP field penetration through the gold film. The surface

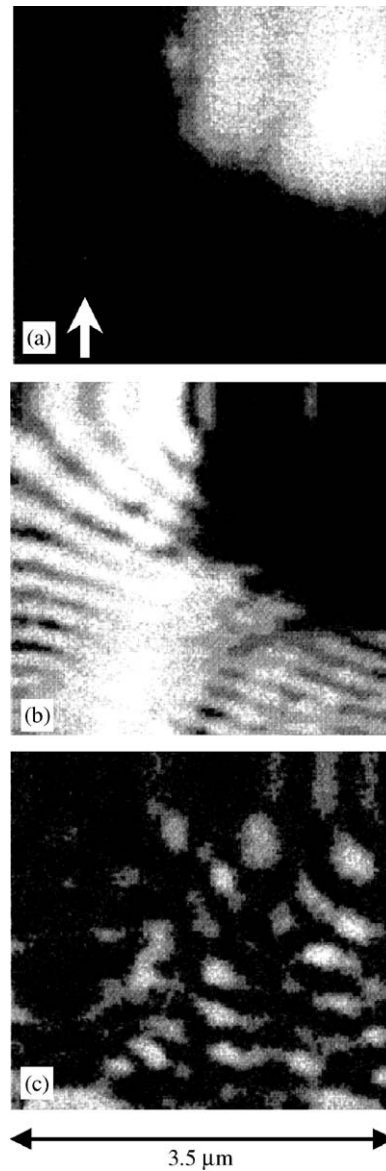


Fig. 26. Topography (a), near-field (b) and far-field (c) intensity distributions over the smooth gold surface. The gray-scale of the topographical image corresponds to 140 nm. The optical images were obtained at different tip–surface distances: ~ 5 nm (b) and $12 \mu\text{m}$ (c) with the air–metal surface polariton being resonantly excited. The optical images are presented in the different gray-scales corresponding to 10–600 pW (b) and 0–20 pW (c). The SPP excitation direction is shown by the arrow. (Ref. [17].)

bump (Fig. 26a) apparently scatters both surface polaritons since it gives rise to the interference patterns observed in both images. The periods of these interference fringes are distinctly different (Fig. 29) as should be expected due to the difference in the wavelengths of the SPPs: $\lambda_{\text{sp}} \approx 590$ nm for the air–metal SPP and $\lambda_{\text{sp}} \approx 420$ nm for the fluoride–metal SPP.

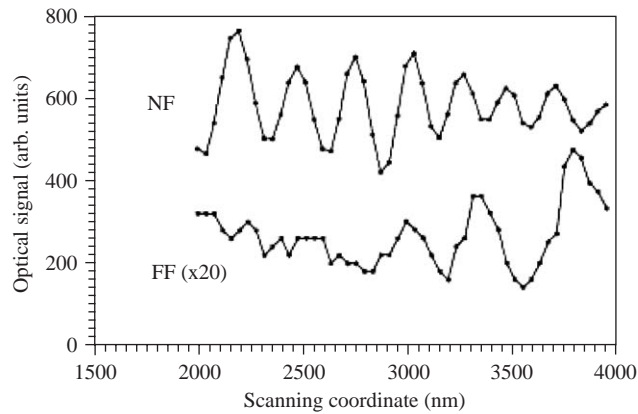


Fig. 27. The cross sections of the near-field (NF) and far-field (FF) optical images along the direction of SPP excitation indicated by the arrow in Fig. 26. (Ref. [17].)

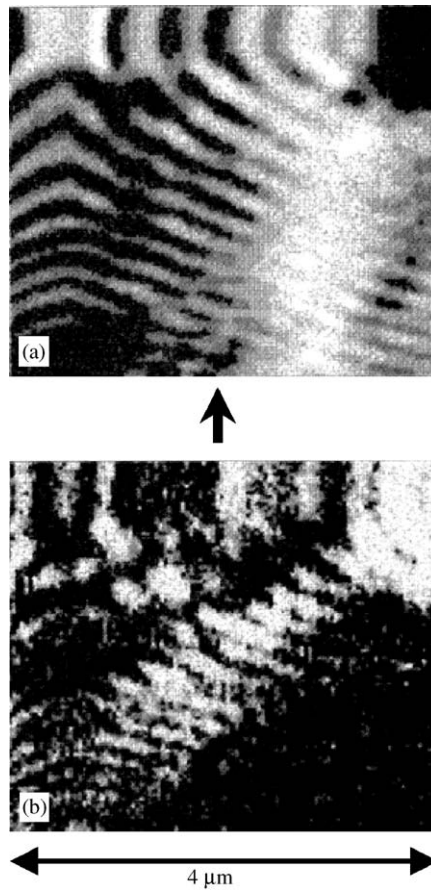


Fig. 28. The near-field intensity distributions obtained at nearly the same place as those in Fig. 26 related to (a) the air-gold SPP and (b) the fluoride-gold SPP. The optical images are presented in different scales corresponding to 10–600 pW (a) and 0–30 pW (b). The SPP excitation direction is shown by the arrow. (Ref. [17].)

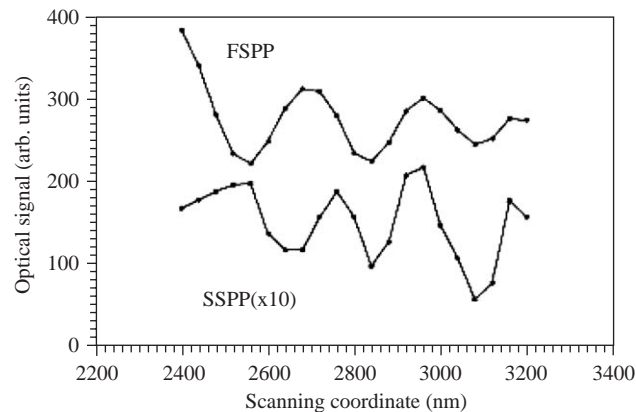


Fig. 29. The cross sections of the near-field optical images obtained with the air–gold (FSP) and fluoride–gold (SSPP) surface polaritons along the direction of SPP excitation indicated by the arrow in Fig. 28. (Ref. [17].)

Thus, the electromagnetic field distribution over such smooth metal surfaces with a large-scale defect structure ($l \sim L_{SP}$) is determined by single scattering processes on dominating defects resulting in the distinctive interference patterns related to the in-plane scattering of the surface polaritons. The air–metal SPP does not penetrate into the surface defect region while the fluoride–metal SPP clearly goes through the defect (Fig. 28a,b). Judging from the distribution of the electromagnetic field in the images one may presume that the gold film is relatively thinner in the bump area, which can be the case if the bump is related to the fluoride layer roughness. Estimations of the amplitude reflection coefficient for this particular defect gives values of about 0.2 and 0.4 for the air–metal and fluoride–metal SPP, respectively, which is much greater than the reflection coefficient for the defects in the silver films discussed in the preceding section.

2.2. Multiple scattering regime: backscattering of surface plasmon polaritons

If a metal surface exhibits a defect structure with distances between the defects smaller than the SPP propagation length, multiple scattering effects dominate the SPP field distribution. In the multiple-scattering regime the scattered SPP can be successively scattered again and again. This can lead to a complex optical field distribution due to the interference of multiple SPP beams resulting from the SPP scattered by different defects. For a small number of defects, such interference patterns are easy to predict and understand [19,59,60]. Nevertheless, with the increase of the number of defects contributing to the interference, the pattern becomes more complex. To ensure a sizeable contribution from multiple scattering, a single scattering event should not lead to strong losses due to scattering into light, while providing reasonably strong in-plane elastic scattering of SPPs. Thus, the requirements on the number of defects within the SPP propagation length and the SPP scattering efficiency can be satisfied for a defect ensemble consisting of defects with sizes which are smaller than but comparable to the SPP wavelength whose heights (or depth) are comparable to the SPP field extension length.

Multiple scattering effects followed by multiple beam interference of scattered waves can result in weak localization (also known as backscattering) or strong (Anderson type) localization of surface waves [101,102].

2.2.1. Enhanced backscattering of surface polaritons on a randomly rough surface

The enhanced backscattering of a quantum or classical wave from a disordered medium is the presence of a well-defined peak in the retroreflection direction in the angular dependence of the intensity of the scattered waves. It is caused by the coherent interference of each multiple-scattering sequence with its reciprocal partner when the wave vectors of the incident and scattered waves are equal in magnitude but are oppositely directed. The importance of these interference processes for scattering into the retroreflection direction seems to have been appreciated first by Watson [103] in the context of a theory of the multiple scattering of electromagnetic waves from an underdense plasma.

The enhanced backscattering of any wave that has its origin in interference between multiply scattered waves has come to be called *weak localization*. This name originated in earlier theoretical studies of electrical conduction in disordered materials, in which it was discovered that when the mean free path between consecutive collisions of an electron with the impurities becomes shorter than its wavelength, this transport process is affected by coherent effects in the electronic wave function that are not taken into account in the standard Boltzmann transport equation for this process. These coherent effects produce an enhanced probability for an electron to return to its origin [104,105]. The enhanced probability of return to the origin reduces the diffusion constant of the electron and, consequently, the electrical conductivity of the disordered material. The enhanced backscattering of the electron, and the decrease in its diffusion constant that it causes, are called weak localization [106]. When the concentration of the impurities becomes very dense, and the scattering from each impurity is very strong, electrical conduction will vanish at absolute zero temperature. The vanishing of electrical conduction under these conditions is called *strong localization* or *Anderson localization* [107]. Weak localization is therefore a precursor to strong localization.

The first experimental observation of the enhanced backscattering effect was by Kuga and Ishimaru [108], in a study of the scattering of light from an aqueous suspension of polystyrene microspheres. It was interpreted at the same time as due to the coherent interference of reciprocal, multiply scattered light paths by Tsang and Ishimaru [109]. The connection of this enhanced backscattering with weak localization, however, was first pointed out approximately a year later by van Albada and Lagendijk [110] and Wolf and Maret [111].

The enhanced backscattering of light from a weakly rough random metal surface was first predicted by McGurn et al. [112]. It arises from the scattering sequences in which p-polarized incident light excites a surface plasmon polariton through the surface roughness, the surface plasmon polariton propagates along the surface and is scattered several times by the roughness, and at the final scattering event is converted back into volume electromagnetic waves. Each such sequence and its reciprocal partner, in which the light and the surface plasmon polariton interact with the surface roughness at the same points but in the reverse order, interfere constructively if the wave vectors of the initial and final volume electromagnetic waves are oppositely directed. The enhanced backscattering of light by this mechanism was subsequently observed experimentally by West and O'Donnell [113], whose experimental results were corroborated by theoretical calculations carried out by Maradudin et al. [114].

In view of the nature of the enhanced backscattering effect, it would seem that it should be expected to occur when a surface plasmon polariton propagates across a weakly rough two-dimensional random surface. This expectation is indeed fulfilled [115]. Thus, let us consider the system consisting of vacuum in the region $x_3 > \zeta(\mathbf{x}_{\parallel})$ and a metal, characterized by an isotropic, frequency-dependent, complex dielectric function $\epsilon(\omega) = \epsilon_1(\omega) + i\epsilon_2(\omega)$. We will be interested in the frequency range in which $\epsilon_1(\omega)$ is negative, as this is the range in which surface plasmon polaritons exist. We assume that the surface profile function

$\zeta(\mathbf{x}_{\parallel})$ is a single-valued function of \mathbf{x}_{\parallel} that is differentiable with respect to x_1 and x_2 as many times as is necessary, and constitutes a zero-mean, stationary, isotropic, Gaussian random process defined by

$$\langle \zeta(\mathbf{x}_{\parallel}) \zeta(\mathbf{x}'_{\parallel}) \rangle = \delta^2 W(|\mathbf{x}_{\parallel} - \mathbf{x}'_{\parallel}|) . \quad (2.89)$$

The angle brackets in Eq. (2.89) denotes an average over the ensemble of realizations of $\zeta(\mathbf{x}_{\parallel})$, and $\delta = \langle \zeta^2(\mathbf{x}_{\parallel}) \rangle^{1/2}$ is the rms height of the surface. In the present discussion it will be assumed that the surface height autocorrelation function $W(|\mathbf{x}_{\parallel}|)$ has the Gaussian form $W(|\mathbf{x}_{\parallel}|) = \exp(-x_{\parallel}^2/a^2)$, where a is the transverse correlation length of the surface roughness.

It is also convenient to introduce the Fourier integral representation of $\zeta(\mathbf{x}_{\parallel})$:

$$\zeta(\mathbf{x}_{\parallel}) = \int \frac{d^2 k_{\parallel}}{(2\pi)^2} \hat{\zeta}(\mathbf{k}_{\parallel}) \exp(i\mathbf{k}_{\parallel} \cdot \mathbf{x}_{\parallel}) . \quad (2.90)$$

The Fourier coefficient $\hat{\zeta}(\mathbf{k}_{\parallel})$ is also a zero-mean Gaussian random process defined by

$$\langle \hat{\zeta}(\mathbf{k}_{\parallel}) \hat{\zeta}(\mathbf{k}'_{\parallel}) \rangle = (2\pi)^2 \delta(\mathbf{k}_{\parallel} + \mathbf{k}'_{\parallel}) \delta^2 g(|\mathbf{k}_{\parallel}|) , \quad (2.91)$$

where

$$g(|\mathbf{k}_{\parallel}|) = \int d^2 x_{\parallel} W(|\mathbf{x}_{\parallel}|) \exp(-i\mathbf{k}_{\parallel} \cdot \mathbf{x}_{\parallel}) \quad (2.92)$$

is the power spectrum of the surface roughness. For the Gaussian form of the surface height autocorrelation function $W(|\mathbf{x}_{\parallel}|)$, the power spectrum also has a Gaussian form, $g(|\mathbf{k}_{\parallel}|) = \pi a^2 \exp(-a^2 k_{\parallel}^2/4)$.

The electric field in the vacuum region $x_3 > \zeta(\mathbf{x}_{\parallel})$ is represented as the sum of an incident and a scattered field of frequency ω , $\mathbf{E}^>(\mathbf{x}; t) = [\mathbf{E}^>(\mathbf{x}|\omega)_{\text{inc}} + \mathbf{E}^>(\mathbf{x}|\omega)_{\text{sc}}] \exp(-i\omega t)$. The incident field is assumed to have the form

$$\mathbf{E}^>(\mathbf{x}|\omega)_{\text{inc}} = \int \frac{d^2 k_{\parallel}}{(2\pi)^2} \left\{ -\frac{c}{\omega} [\hat{\mathbf{k}}_{\parallel} \alpha_0(k_{\parallel}) + \hat{\mathbf{x}}_3 k_{\parallel}] B_p(\mathbf{k}_{\parallel}) + (\hat{\mathbf{x}}_3 \times \hat{\mathbf{k}}_{\parallel}) B_s(\mathbf{k}_{\parallel}) \right\} \\ \times \exp[i\mathbf{k}_{\parallel} \cdot \mathbf{x}_{\parallel} - i\alpha_0(k_{\parallel})x_3] , \quad (2.93)$$

where $\alpha_0(k_{\parallel}) = [(\omega/c)^2 - k_{\parallel}^2]^{1/2}$, with $\text{Re } \alpha_0(k_{\parallel}) > 0$, $\text{Im } \alpha_0(k_{\parallel}) > 0$. The coefficients $B_p(\mathbf{k}_{\parallel})$ and $B_s(\mathbf{k}_{\parallel})$ will be defined below. The electric vector of the scattered field is

$$\mathbf{E}^>(\mathbf{x}|\omega)_{\text{sc}} = \int \frac{d^2 q_{\parallel}}{(2\pi)^2} \left\{ \frac{c}{\omega} [\hat{\mathbf{q}}_{\parallel} \alpha_0(q_{\parallel}) - \hat{\mathbf{x}}_3 q_{\parallel}] A_p(\mathbf{q}_{\parallel}) + (\hat{\mathbf{x}}_3 \times \hat{\mathbf{q}}_{\parallel}) A_s(\mathbf{q}_{\parallel}) \right\} \\ \times \exp[i\mathbf{q}_{\parallel} \cdot \mathbf{x}_{\parallel} + i\alpha_0(q_{\parallel})x_3] . \quad (2.94)$$

We also need the magnetic vector of the scattered field,

$$\mathbf{H}^>(\mathbf{x}|\omega)_{\text{sc}} = \int \frac{d^2 q_{\parallel}}{(2\pi)^2} \left\{ (\hat{\mathbf{x}}_3 \times \hat{\mathbf{q}}_{\parallel}) A_p(\mathbf{q}_{\parallel}) - \frac{c}{\omega} (\hat{\mathbf{q}}_{\parallel} \alpha_0(q_{\parallel}) - \hat{\mathbf{x}}_3 q_{\parallel}) A_s(\mathbf{q}_{\parallel}) \right\} \\ \times \exp[i\mathbf{q}_{\parallel} \cdot \mathbf{x}_{\parallel} + i\alpha_0(q_{\parallel})x_3] . \quad (2.95)$$

We wish to calculate the angular dependence of the incoherent (diffuse) component of the intensity of the scattered field in the vacuum only, since the energy flow associated with the scattered field in the

metal is not accessible experimentally. Thus, we consider the complex Poynting vector of the diffusely scattered field,

$$\langle \mathbf{S}^c \rangle_{\text{diff}} = \frac{c}{8\pi} \langle (\mathbf{Q}\mathbf{E}^>(\mathbf{x}|\omega)_{\text{sc}}) \times (\mathbf{Q}\mathbf{H}^>(\mathbf{x}|\omega)_{\text{sc}})^* \rangle, \quad (2.96)$$

where the operator Q produces the fluctuating part of any random process to which it is applied, e.g., $\mathbf{Q}\mathbf{E}^>(\mathbf{x}|\omega)_{\text{sc}} = \mathbf{E}^>(\mathbf{x}|\omega)_{\text{sc}} - \langle \mathbf{E}(\mathbf{x}|\omega)_{\text{sc}} \rangle$. From the results given by Eqs. (2.94) and (2.95) we see that for calculating $\langle \mathbf{S}^c \rangle_{\text{diff}}$ we need to calculate averages of the type of $\langle (QA_i(\mathbf{q}_{\parallel})) (QA_j^*(\mathbf{q}'_{\parallel})) \rangle$ with $i, j = p, s$.

The amplitudes $A_{p,s}(\mathbf{q}_{\parallel})$ are related to the amplitudes $B_{p,s}(\mathbf{k}_{\parallel})$ by

$$A_i(\mathbf{q}_{\parallel}) = \sum_{j=p,s} \int \frac{d^2k_{\parallel}}{(2\pi)^2} R_{ij}(\mathbf{q}_{\parallel}|\mathbf{k}_{\parallel}) B_j(\mathbf{k}_{\parallel}), \quad i = p, s, \quad (2.97)$$

where the matrix of scattering amplitudes $\overleftrightarrow{R}(\mathbf{q}_{\parallel}|\mathbf{k}_{\parallel})$ is related to the matrix $\overleftrightarrow{G}(\mathbf{q}_{\parallel}|\mathbf{k}_{\parallel})$ of the exact Green's functions for surface plasmon polaritons on a rough surface by [115]

$$\overleftrightarrow{R}(\mathbf{q}_{\parallel}|\mathbf{k}_{\parallel}) = -(2\pi)^2 \delta(\mathbf{q}_{\parallel} - \mathbf{k}_{\parallel}) \overleftrightarrow{I} - 2i \overleftrightarrow{G}(\mathbf{q}_{\parallel}|\mathbf{k}_{\parallel}) \alpha_0(k_{\parallel}), \quad (2.98)$$

where \overleftrightarrow{I} is the 2×2 unit matrix.

The Green's function $\overleftrightarrow{G}(\mathbf{q}_{\parallel}|\mathbf{k}_{\parallel})$ is the solution of the Dyson equation [69]

$$\overleftrightarrow{G}(\mathbf{q}_{\parallel}|\mathbf{k}_{\parallel}) = (2\pi)^2 \delta(\mathbf{q}_{\parallel} - \mathbf{k}_{\parallel}) \overleftrightarrow{G}^{(0)}(k_{\parallel}) + \overleftrightarrow{G}^{(0)}(q_{\parallel}) \int \frac{d^2p_{\parallel}}{(2\pi)^2} \overleftrightarrow{V}(\mathbf{q}_{\parallel}|\mathbf{p}_{\parallel}) \overleftrightarrow{G}(\mathbf{p}_{\parallel}|\mathbf{k}_{\parallel}), \quad (2.99)$$

where $\overleftrightarrow{G}^{(0)}(k_{\parallel})$ is the Green's function for surface plasmon polaritons on a planar metal surface,

$$\overleftrightarrow{G}^{(0)}(k_{\parallel}) = \begin{pmatrix} \frac{i\epsilon(\omega)}{\epsilon(\omega)\alpha_0(k_{\parallel}) + \alpha(k_{\parallel})} & 0 \\ 0 & \frac{i}{\alpha_0(k_{\parallel}) + \alpha(k_{\parallel})} \end{pmatrix}, \quad (2.100)$$

and $V(\mathbf{q}_{\parallel}|\mathbf{k}_{\parallel})$ is a scattering potential which, to the leading order in $\zeta(\mathbf{x}_{\parallel})$ is given by [69]

$$\overleftrightarrow{V}(\mathbf{q}_{\parallel}|\mathbf{k}_{\parallel}) = \frac{\epsilon(\omega) - 1}{\epsilon(\omega)} \hat{\zeta}(\mathbf{q}_{\parallel} - \mathbf{k}_{\parallel}) \begin{pmatrix} \frac{\epsilon(\omega)q_{\parallel}k_{\parallel} - \alpha(q_{\parallel})\hat{\mathbf{q}}_{\parallel} \cdot \hat{\mathbf{k}}_{\parallel} \alpha(k_{\parallel})}{\epsilon(\omega)} & -\frac{\omega}{c} \alpha(q_{\parallel}) (\hat{\mathbf{q}}_{\parallel} \times \hat{\mathbf{k}}_{\parallel})_3 \\ -\frac{\omega}{c} (\hat{\mathbf{q}}_{\parallel} \times \hat{\mathbf{k}}_{\parallel})_3 \alpha(k_{\parallel}) & \epsilon(\omega) \frac{\omega^2}{c^2} \hat{\mathbf{q}}_{\parallel} \cdot \hat{\mathbf{k}}_{\parallel} \end{pmatrix}. \quad (2.101)$$

The averaged Green's function $\langle \overleftrightarrow{G}(\mathbf{q}_{\parallel}|\mathbf{k}_{\parallel}) \rangle$ is diagonal in its arguments $(\mathbf{q}_{\parallel}, \mathbf{k}_{\parallel})$ due to the stationarity of the random process $\zeta(\mathbf{x}_{\parallel})$, and is also diagonal in its polarization indices (i, j) as a consequence of the conservation of angular momentum,

$$\langle G_{ij}(\mathbf{q}_{\parallel}|\mathbf{k}_{\parallel}) \rangle = (2\pi)^2 \delta(\mathbf{q}_{\parallel} - \mathbf{k}_{\parallel}) \delta_{ij} G_i(k_{\parallel}). \quad (2.102)$$

The fact that $G_i(k_{\parallel})$ depends on the wave vector \mathbf{k}_{\parallel} only through its magnitude is due to the isotropy of the random process $\zeta(\mathbf{x}_{\parallel})$.

The reduced Green's function $G_i(k_{\parallel})$ has the form

$$G_i(k_{\parallel}) = \frac{1}{G_i^0(k_{\parallel})^{-1} - M_i(k_{\parallel})}, \quad i = p, s, \quad (2.103)$$

where $M_i(k_{\parallel})$ is an averaged self energy defined by

$$\langle M_{ij}(\mathbf{q}_{\parallel}|\mathbf{k}_{\parallel}) \rangle = (2\pi)^2 \delta(\mathbf{q}_{\parallel} - \mathbf{k}_{\parallel}) \delta_{ij} M_i(k_{\parallel}), \quad (2.104)$$

and $\overleftrightarrow{M}(\mathbf{q}_{\parallel}|\mathbf{k}_{\parallel})$ is the solution of

$$\overleftrightarrow{M}(\mathbf{q}_{\parallel}|\mathbf{k}_{\parallel}) = \overleftrightarrow{V}(\mathbf{q}_{\parallel}|\mathbf{k}_{\parallel}) + \int \frac{d^2 p_{\parallel}}{(2\pi)^2} \overleftrightarrow{M}(\mathbf{q}_{\parallel}|\mathbf{p}_{\parallel}) \overleftrightarrow{G}(p_{\parallel}) (\overleftrightarrow{V}(\mathbf{p}_{\parallel}|\mathbf{k}_{\parallel}) - \langle \overleftrightarrow{M}(\mathbf{p}_{\parallel}|\mathbf{k}_{\parallel}) \rangle). \quad (2.105)$$

The ij element of the Green's function $\overleftrightarrow{G}(\mathbf{q}_{\parallel}|\mathbf{k}_{\parallel})$ can then be expressed in terms of the diagonal elements of the averaged Green's function by [116]

$$G_{ij}(\mathbf{q}_{\parallel}|\mathbf{k}_{\parallel}) = (2\pi)^2 \delta(\mathbf{q}_{\parallel} - \mathbf{k}_{\parallel}) \delta_{ij} G_i(k_{\parallel}) + G_i(q_{\parallel}) t_{ij}(\mathbf{q}_{\parallel}|\mathbf{k}_{\parallel}) G_j(k_{\parallel}), \quad (2.106)$$

where the matrix $\overleftrightarrow{t}(\mathbf{q}_{\parallel}|\mathbf{k}_{\parallel})$ is the solution of

$$\overleftrightarrow{t}(\mathbf{q}_{\parallel}|\mathbf{k}_{\parallel}) = \overleftrightarrow{v}(\mathbf{q}_{\parallel}|\mathbf{k}_{\parallel}) + \int \frac{d^2 p_{\parallel}}{(2\pi)^2} \overleftrightarrow{v}(\mathbf{q}_{\parallel}|\mathbf{p}_{\parallel}) \overleftrightarrow{G}(p_{\parallel}) \overleftrightarrow{t}(\mathbf{p}_{\parallel}|\mathbf{k}_{\parallel}). \quad (2.107)$$

with

$$\overleftrightarrow{v}(\mathbf{q}_{\parallel}|\mathbf{k}_{\parallel}) = \overleftrightarrow{V}(\mathbf{q}_{\parallel}|\mathbf{k}_{\parallel}) - \langle \overleftrightarrow{M}(\mathbf{q}_{\parallel}|\mathbf{k}_{\parallel}) \rangle. \quad (2.108)$$

It follows from Eqs. (2.102) and (2.106) that $\overleftrightarrow{t}(\mathbf{q}_{\parallel}|\mathbf{k}_{\parallel})$ possesses the useful property that $\langle \overleftrightarrow{t}(\mathbf{q}_{\parallel}|\mathbf{k}_{\parallel}) \rangle = 0$.

When the preceding results are combined with Eq. (2.97) we obtain for the fluctuating part of $A_i(\mathbf{q}_{\parallel})$

$$\begin{aligned} \mathcal{Q}A_i(\mathbf{q}_{\parallel}) &= \sum_{j=p,s} \int \frac{d^2 k_{\parallel}}{(2\pi)^2} (-2i\alpha_0(k_{\parallel})) \mathcal{Q}G_{ij}(\mathbf{q}_{\parallel}|\mathbf{k}_{\parallel}) B_j(\mathbf{k}_{\parallel}) \\ &= G_i(q_{\parallel}) \sum_{j=p,s} \int \frac{d^2 k_{\parallel}}{(2\pi)^2} t_{ij}(\mathbf{q}_{\parallel}|\mathbf{k}_{\parallel}) (-2i\alpha_0(k_{\parallel})) G_j(k_{\parallel}) B_j(\mathbf{k}_{\parallel}), \end{aligned} \quad (2.109)$$

where we have used the result that $\mathcal{Q}t_{ij}(\mathbf{q}_{\parallel}|\mathbf{k}_{\parallel}) = t_{ij}(\mathbf{q}_{\parallel}|\mathbf{k}_{\parallel})$. Because only p-polarized surface plasmon polaritons exist on a weakly rough random surface [117], in the sum over initial states in Eq. (2.109) we choose the term with $j = p$ and $\mathbf{k}_{\parallel} = \hat{\mathbf{x}}_1 K_{\parallel}(\omega) = \mathbf{K}_{\parallel}(\omega)$, where the wavenumber of the surface plasmon polariton of frequency ω on the random surface is given by the real part of the root $K_{\parallel}(\omega) + i\Delta_{\text{tot}}(\omega)$ of the equation $[G_p(k_{\parallel})]^{-1} = 0$. The function $\Delta_{\text{tot}}(\omega)$ gives the total amplitude damping rate of a surface plasmon polariton due to ohmic losses and to its roughness-induced conversion into volume electromagnetic waves in the vacuum and into other surface plasmon polaritons. Thus we choose the amplitude $B_p(\mathbf{k}_{\parallel})$ to satisfy the condition

$$(-2i\alpha_0(k_{\parallel})) G_p(k_{\parallel}) B_p(\mathbf{k}_{\parallel}) = (2\pi)^2 \delta(\mathbf{k}_{\parallel} - \mathbf{K}_{\parallel}(\omega)). \quad (2.110)$$

As a consequence, $QA_i(\mathbf{q}_{\parallel})$ takes the simple form

$$QA_i(\mathbf{q}_{\parallel}) = G_i(q_{\parallel})t_{ip}(\mathbf{q}_{\parallel}|\mathbf{K}_{\parallel}), \quad i = p, s. \quad (2.111)$$

To obtain the angular dependence of the energy flow associated with the scattering of surface plasmon polaritons into other surface plasmon polaritons, we substitute Eq. (2.111) into the expression for $\langle \mathbf{S}^c \rangle_{\text{diff}}$ given by Eqs. (2.94)–(2.96), namely

$$\begin{aligned} \langle \mathbf{S}^c \rangle_{\text{diff}} &= \frac{c}{8\pi} \int \frac{d^2 q_{\parallel}}{(2\pi)^2} \int \frac{d^2 q'_{\parallel}}{(2\pi)^2} \frac{c q_{\parallel}}{\omega} (\hat{\mathbf{q}}'_{\parallel}) \exp[i(\mathbf{q}_{\parallel} - \mathbf{q}'_{\parallel}) \cdot \mathbf{x}_{\parallel}] \\ &\quad \times \exp[i(\alpha_0(q_{\parallel}) - \alpha_0^*(q'_{\parallel}))] \langle (QA_p(\mathbf{q}_{\parallel})) (QA_p^*(\mathbf{q}'_{\parallel})) \rangle. \end{aligned} \quad (2.112)$$

In obtaining this expression we have again used the result that only p-polarized surface plasmon polaritons can exist on a weakly rough random surface, and have retained only that part of the energy flow that is parallel to the mean surface. The use of Eq. (2.111) yields the result

$$\langle (QA_p(\mathbf{q}_{\parallel})) (QA_p^*(\mathbf{q}'_{\parallel})) \rangle = (2\pi)^2 \delta(\mathbf{q}_{\parallel} - \mathbf{q}'_{\parallel}) |G_p(q_{\parallel})|^2 \tau_{pp}(\mathbf{q}_{\parallel}|\mathbf{K}_{\parallel}), \quad (2.113)$$

where we have used the result that

$$\langle t_{pp}(\mathbf{q}_{\parallel}|\mathbf{K}_{\parallel}) t_{pp}^*(\mathbf{q}'_{\parallel}|\mathbf{K}_{\parallel}) \rangle = (2\pi)^2 \delta(\mathbf{q}_{\parallel} - \mathbf{q}'_{\parallel}) \tau_{pp}(\mathbf{q}_{\parallel}|\mathbf{K}_{\parallel}), \quad (2.114)$$

which defines the function $\tau_{pp}(\mathbf{q}_{\parallel}|\mathbf{K}_{\parallel})$. Consequently, Eq. (2.112) becomes

$$\langle \mathbf{S}^c \rangle_{\text{diff}} = \frac{c}{8\pi} \int \frac{d^2 q_{\parallel}}{(2\pi)^2} \hat{\mathbf{q}}_{\parallel} \frac{c q_{\parallel}}{\omega} |G_p(q_{\parallel})|^2 \tau_{pp}(\mathbf{q}_{\parallel}|\mathbf{K}_{\parallel}) \exp[-2\text{Re} \beta_0(q_{\parallel})x_3], \quad (2.115)$$

where $\beta_0(q_{\parallel}) = -i\alpha_0(q_{\parallel}) = [q_{\parallel}^2 - (\omega/c)^2]^{1/2}$. The total power carried by the scattered surface plasmon polaritons is obtained by integrating Eq. (27) on x_3 from zero to infinity:

$$\int_0^{\infty} dx_3 \langle \mathbf{S}^c \rangle_{\text{diff}} = \frac{c}{8\pi} \int \frac{d^2 q_{\parallel}}{(2\pi)^2} \hat{\mathbf{q}}_{\parallel} \frac{c q_{\parallel}}{\omega} |G_p(q_{\parallel})|^2 \frac{\tau_{pp}(\mathbf{q}_{\parallel}|\mathbf{K}_{\parallel})}{2\text{Re} \beta_0(q_{\parallel})}. \quad (2.116)$$

In the vicinity of $q_{\parallel} = K_{\parallel}$, the p-polarization Green's function $G_p(q_{\parallel})$ has the form [116]

$$G_p(q_{\parallel}) = \frac{C_p}{q_{\parallel} - K_{\parallel} - i\Delta_{\text{tot}}}, \quad (2.117)$$

where the residue C_p is given by $C_p = (-\epsilon_1(\omega))^{3/2} / [\epsilon_1^2(\omega) - 1]$. In the limit that $\Delta_{\text{tot}} \ll K_{\parallel}$, the approximation

$$|G_p(q_{\parallel})|^2 = \frac{\pi C_p^2}{\Delta_{\text{tot}}} \delta(p_{\parallel} - K_{\parallel}) \quad (2.118)$$

is valid, and we obtain

$$\int_0^{\infty} dx_3 \langle \mathbf{S}^c \rangle_{\text{diff}} = \frac{c^2 C_p^2 K_{\parallel}^2}{64\pi^2 \Delta_{\text{tot}} \text{Re} \beta_0(K_{\parallel}) \omega} \int_{-\pi}^{\pi} d\phi (\cos \phi, \sin \phi) \tau_{pp}(\mathbf{Q}_{\parallel}|\mathbf{K}_{\parallel}), \quad (2.119)$$

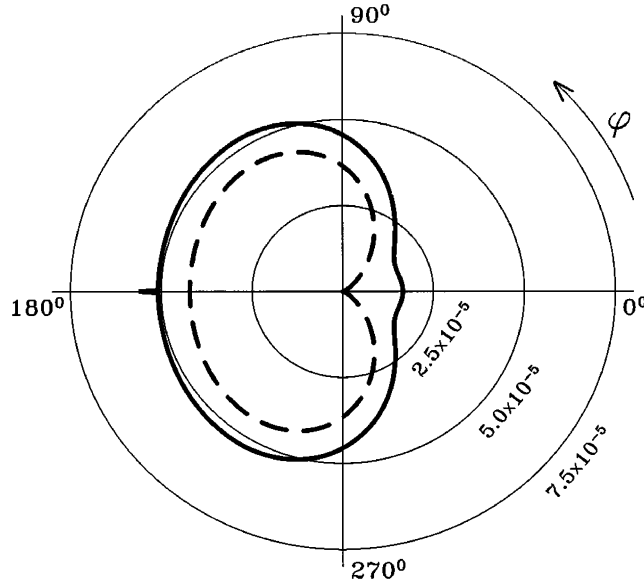


Fig. 30. A polar plot of $\tau(\phi)_{\text{sp}}$ for a random silver surface, characterized by $\delta = 0.075 \mu\text{m}$ and $a = 2.5 \mu\text{m}$, at the vacuum wavelength $\lambda_0 = 10 \mu\text{m}$. The dashed curve shows the second-order contribution in δ (single scattering), and the solid curve gives the sum of the contributions of the second and fourth orders in δ . (Ref. [115].)

where $\mathbf{Q}_{\parallel} = K_{\parallel}(\cos \phi, \sin \phi)$. The magnitude of the energy flow in the vacuum carried by the diffuse part of the surface plasmon polariton field scattered into the angular interval $(\phi, \phi + d\phi)$ is then given by

$$\langle P(\phi)_{\text{sp}} \rangle_{\text{diff}} = \frac{c^2 C_p^2 K_{\parallel}^2}{64\pi^2 \Delta_{\text{tot}} \text{Re} \beta_0(K_{\parallel}) \omega} \tau_{pp}(\mathbf{Q}_{\parallel} | \mathbf{K}_{\parallel}). \quad (2.120)$$

All of the angular dependence of this function resides in $\tau_{pp}(\mathbf{Q}_{\parallel} | \mathbf{K}_{\parallel})$.

The function $\tau_{pp}(\mathbf{Q}_{\parallel} | \mathbf{K}_{\parallel})$ was calculated by Shchegrov [115] to fourth order in the surface profile function, which is the lowest order in which enhanced backscattering effects occur [118], and which gives the dominant contribution to these effects. The result was used to calculate $\tau(\phi)_{\text{sp}} = \tau_{pp}(\mathbf{Q}_{\parallel} | \mathbf{K}_{\parallel})$ for a weakly rough silver surface. The dielectric function of silver was represented by the simple, free electron, form $\epsilon(\omega) = 1 - \omega_p^2 / [\omega(\omega + i\gamma)]$, with $\omega_p = 7.2 \times 10^4 \text{ cm}^{-1}$ and $\gamma = 225 \text{ cm}^{-1}$ [119]. The roughness parameters were chosen to be $\delta = 0.075 \mu\text{m}$ and $a = 2.5 \mu\text{m}$. The frequency ω of the incident surface plasmon polariton corresponded to a vacuum wavelength $\lambda_0 = 2\pi c / \omega = 10 \mu\text{m}$, lying in the infrared. The surface plasmon polariton attenuation length $\ell_{\text{tot}} = 1/2\Delta_{\text{tot}}$ at this frequency was calculated to be $\ell_{\text{tot}} = 0.38 \text{ cm}$. A polar plot of $\tau(\phi)_{\text{sp}}$ calculated to fourth order in δ is presented in Fig. 30, together with the contribution from single-scattering processes alone, which is of $O(\delta^2)$. Although the single scattering contribution is the dominant one in Fig. 30, the weaker double-scattering contribution, which is of $O(\delta^4)$, contributes a peak to $\tau(\phi)_{\text{sp}}$ in the retroreflection direction. This statement is contradicted by the next section.

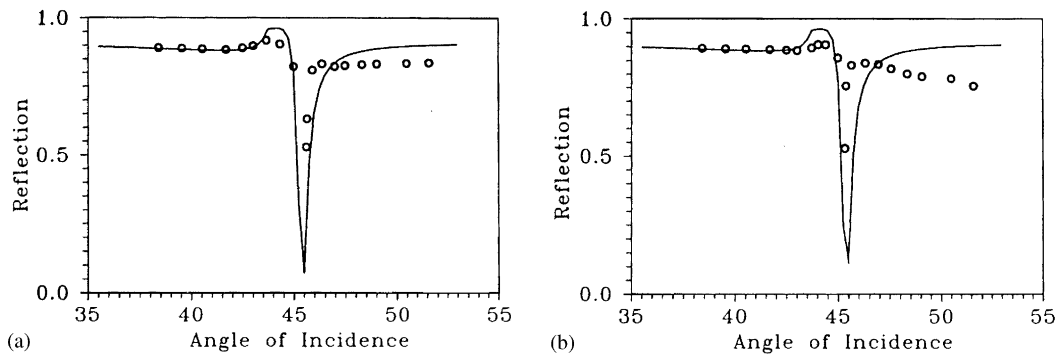


Fig. 31. Angular dependencies of the reflected (p-polarized) light intensity measured (circles) and calculated (solid lines) for silver films exhibiting different roughness. (Ref. [20].)

2.2.2. Direct observation of SPP backscattering with SNOM

Enhanced backscattering of light from random media has been intensively studied experimentally (see, e.g., Refs. [102,113,120] and references therein). Backscattering enhancement arising from the SPP excitation on a rough metal surface has been observed in the diffusely scattered light [113]. This enhancement has been found to result from the coherent interference of scattered contributions (propagating in air) from counterpropagating SPPs [120]. The enhanced backscattering of SPPs in a surface plane has also been observed indirectly by detecting a sharp peak in the angular dependence of the efficiency of second harmonic generation in the direction perpendicular to the sample surface observed in transmission through a rough metal film in the Kretschmann attenuated total reflection geometry [121] (see also the discussion in Section 7). The occurrence of such a peak is a fingerprint of the enhanced backscattering of SPPs, since SHG in the normal direction is related to the nonlinear interaction between counterpropagating (i.e. between the excited and backscattered) SPPs at the fundamental frequency [122].

The in-plane SPP backscattering can be directly imaged using scanning near-field microscopy. Two slightly different weakly rough surfaces of silver films similar to those used in the SHG experiments have been studied [20]. Angular dependencies of the total reflection from both films exhibit similar behavior with well pronounced resonance minima (Fig. 31). The SPP wavelength and propagation length have been estimated to be 610 nm and 25 μm , respectively, for both films.

Topographical images of the silver film surface of the sample A show a relatively smooth surface with rarely spaced submicron-sized bumps of 40–80 nm heights (Fig. 32). The average distance between these scatterers is about 8 μm . The near-field measurements of the SPP field at the silver–air interface of this sample exhibit a well pronounced resonance behavior, which was expected from the far-field measurements (Fig. 31). The average optical signal was about 25 times smaller if the angle of incidence was out of resonance by $\sim 2^\circ$ or if the fiber tip was moved a few microns away from the surface. Therefore, the contribution of propagating in air components of the scattered field in the detected signal can be disregarded. If a large scale scatterer is present in the field of view, the near-field optical image (with the SPP being resonantly excited) shows an interference pattern due to interference between the excited and scattered SPPs (Fig. 32b), which is similar to the behavior for the smooth gold films (Fig. 26). However, as the SPP propagation length is about 3 times larger for the silver film than that for the gold film, the interference pattern recorded with the silver film reveals also the presence of SPPs

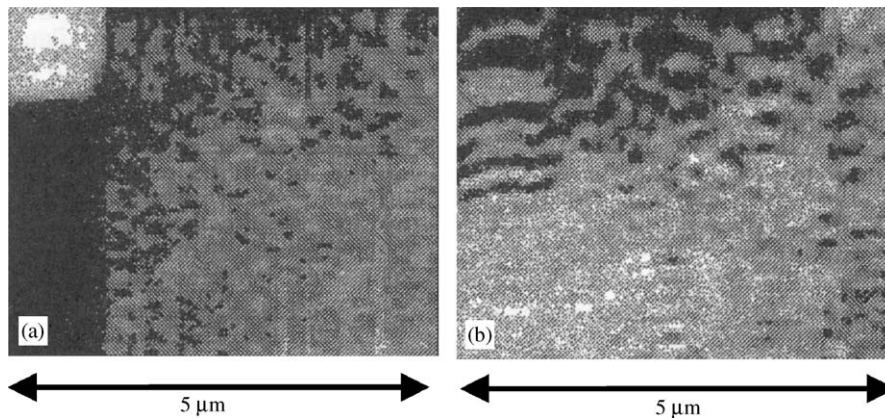


Fig. 32. Topography (a) and corresponding near-field intensity distribution (b) obtained with the silver film A. The gray-scale of the topographical image corresponds to 73 nm. The scale of the optical image corresponds to 0–2 nW. SPPs are resonantly excited upwards in the vertical direction. (Ref. [20].)

scattered from other scatterers that are not seen in the topographical image. Consequently, $L_{\text{SP}} > l$ for the silver film whereas $L_{\text{SP}} \sim l$ for the gold film of similar topography. Thus, one can expect to observe interference effects related to the regime of multiple scattering of SPPs propagating along the silver film surface.

Weak localization of SPPs manifests itself in the formation of the backscattered SPP due to interference effects in multiple scattering of SPPs in the surface plane. The physical origin of this two-dimensional phenomenon is evidently the same as that of the enhanced backscattering in three dimensions: the waves which travel along the same light path in opposite directions will always have the same phase and interfere constructively only in the direction of pure backscattering [102]. Apparently, formation of the backscattered SPP should be most pronounced relatively far away from individual scatterers, where the SPPs scattered along different routes are of the same magnitude. In three dimensions, this situation corresponds to the detection of scattered waves outside of a medium with random scatterers. The presence of the backscattered SPP can be deduced from, in general, a complicated interference pattern by extracting the specific pattern corresponding to the interference between the excited and backscattered SPPs. Since the excited SPP propagates upwards in the images, the appearance of horizontally oriented fringes with a period equal to half of the SPP wavelength in the near-field optical image can be regarded as evidence of the presence of the backscattered SPP.

Topographical and near-field optical images of the film A which exhibited a smooth surface topography (with the maximum depth being less than 20 nm) without well defined scatterers, clearly contain such interference fringes (Figs. 33). The Fourier transform of the optical image exhibits a number of bright spots along the central horizontal line of the FFT plane corresponding to the signal variations between successive scans (the scanning direction is vertical in the images) as well as two very bright spots situated symmetrically along the central vertical line (Fig. 33d). These spots correspond to the periodic variation in the detected optical signal with the period of ~ 305 nm, the same as for the horizontal fringes seen on the processed optical image (Fig. 33c) demonstrating the presence of the backscattered SPPs on the surface under consideration. The period of the interference fringes corresponds well to the expected value of $\lambda_{\text{sp}}/2$ within the accuracy of the measurements.

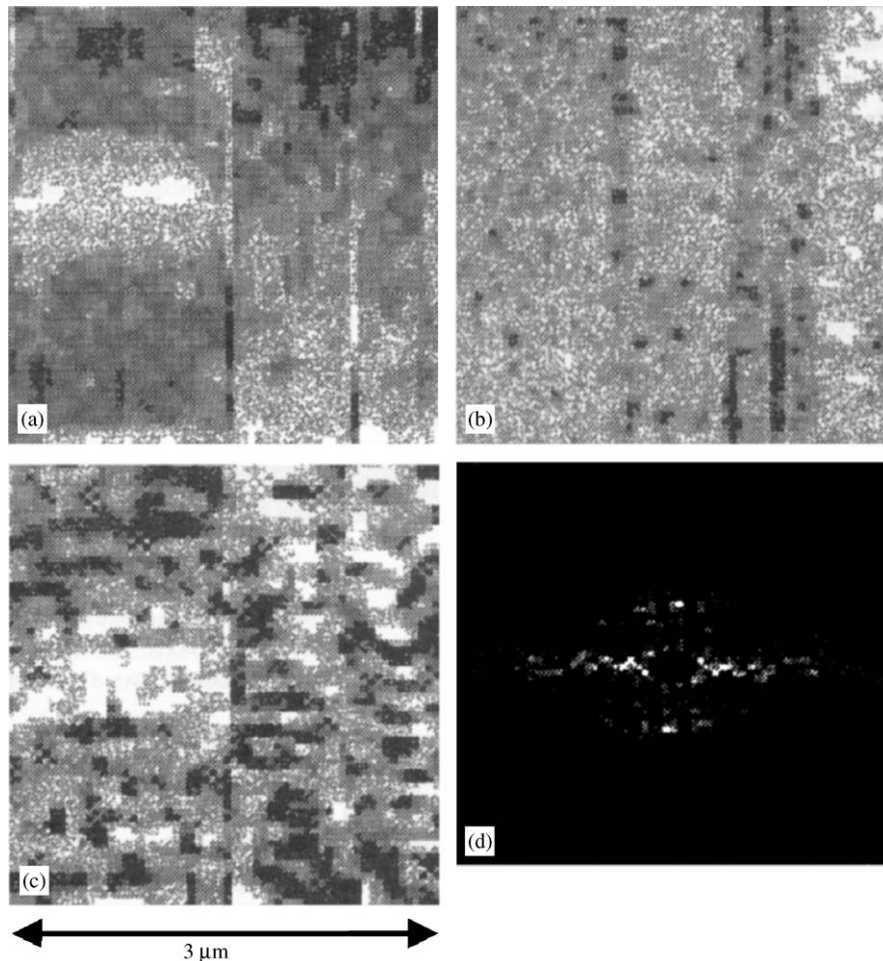


Fig. 33. Topography (a) and corresponding near-field intensity distributions (b,c) obtained with the silver film A. The gray-scale of the topographical image corresponds to 18 nm. The scale of the optical image corresponds to 0.5–2 nW. SPPs are resonantly excited upwards in the vertical direction. The optical image (c) is processed by averaging and filtering low spatial frequencies out. (d) The Fourier transform of the image (b) with the central part being omitted. (Ref. [20].)

Topographical images of the silver film surface of the sample B reveal a surface topography similar to that of the film A (cf. Figs. 32a and 34a,b) but with more closely situated surface bumps. The average distance l between the scatterers was estimated to be about $5 \mu\text{m}$, which is almost two times smaller than for the film A. The decrease of l has been expected due to different fabrication conditions of the films. The near-field measurements of the SPP field over the surface of the sample B exhibit behavior similar to that observed with the film A. However, the near-field optical signal under the resonant SPP excitation was about 5 times smaller than the signal detected with the film A, which is related to the increase of the SPP scattering.

The near-field optical images obtained with the sample B are similar to those recorded with the previous sample, but the interference effects related to multiple scattering of SPPs are much more pronounced.

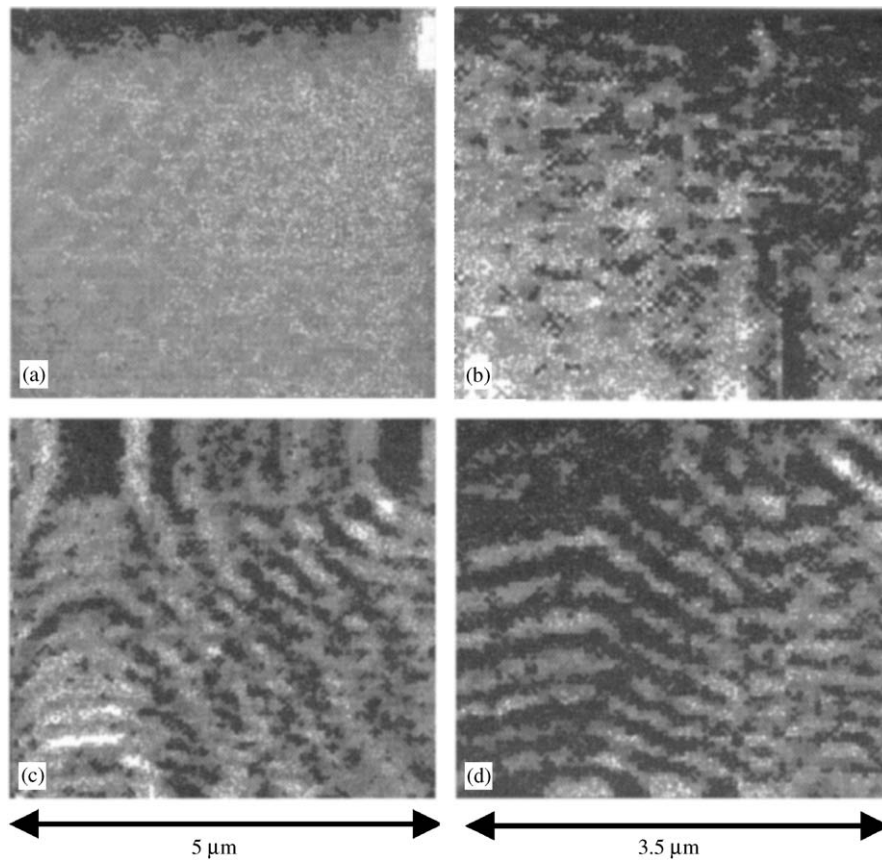


Fig. 34. Topography (a,b) and near-field intensity distributions (c,d) obtained in different places of the silver film B. The gray-scale of the topographical images correspond to 57 (a) and 16 nm (b). The scale of the optical images corresponds to 0–0.2 nW. SPPs are resonantly excited upwards in the vertical direction. (Ref. [20].)

Even in the presence of a scatterer in the field of view, the optical image showed a complicated interference pattern indicating the presence of several scattered SPPs (Fig. 34c). One can already see the appearance of the characteristic horizontal fringes associated with the presence of the backscattered SPP. The difference in the optical images of two surface regions with similar topography but of different films (cf. Figs. 32 and 34) is apparently related to the difference in the average distance l between surface scatterers for these films: multiple scattering of SPPs is more developed in the case of smaller l (i.e. with the film B) resulting in more pronounced interference effects. One should expect to find a similar enhancement of contrast in the optical images recorded in the absence of scatterers in the field of view. The horizontal interference fringes due to backscattered SPPs are clearly seen (Fig. 34d).

Thus, the multiple scattering of SPPs in a surface plane can result for certain parameters of surface roughness in the surface polariton backscattering leading to the appearance of interference fringes oriented perpendicularly to the direction of SPP propagation. Such fringes have appeared more pronounced and clearly visible for the film with larger roughness. This agrees well with the conclusions of the aforementioned SHG experiments [121] with the silver films analogous to those described here: the efficiency

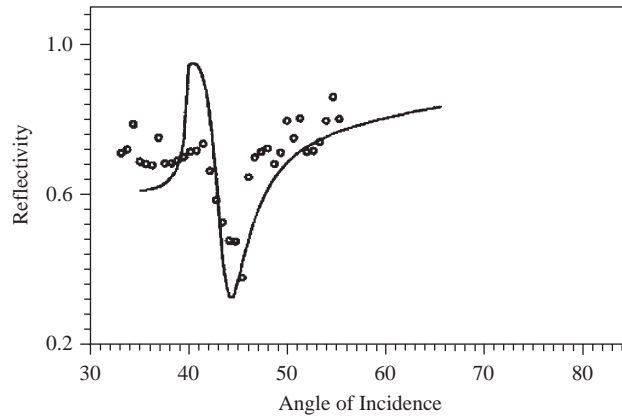


Fig. 35. Angular excitation spectra of surface polaritons measured (circles) and calculated (solid lines) for the rough gold film. (Ref. [18].)

of SHG in the direction perpendicular to the surface has been found to be significantly larger for films with larger roughness confirming the stronger backscattering effect.

2.3. Multiple scattering regime: surface polaritons localization on rough surfaces

If the parameters of the multiple scattering regime are chosen appropriately, the strong (Anderson type [107]) localization of surface polaritons can be achieved. However, it is rather difficult to modify surface roughness in a controllable way on such distance scales. There is a better chance to achieve this by choosing not only a different surface structure but also a different metal providing a different SPP propagation length and SPP scattering efficiency. For the smooth gold films discussed above the SPP propagation length was smaller than the distance between scatterers. Thus, the SPP behavior observed with those gold films was related mainly to single-scattering events. For the silver films, the weak localization regime was realized in multiple scattering: the SPP propagation length is greater than the distance between defects ($L_{\text{SP}} > l$) but not yet enough for the strong localization to occur ($L_{\text{SP}} \gg l$). If the distance between scatterers is much smaller than the SPP propagation length, the stronger regime of multiple scattering of SPPs can develop, and the (strong) Anderson type SPP localization is observed.

Anderson localization of SPP has been studied with surface polaritons on a relatively rough gold film supporting the air–gold SPP of about 570 nm wavelength (Fig. 35) [17,18]. The SPP propagation length can be estimated from the dielectric constant of the film to be about 8 μm . By comparing the angular excitation spectra of the smooth and rough films (Figs. 25 and 35) one can immediately suggest a significant difference in quality of these gold surfaces. Consequently, the detected near-field optical signal over this rough film is on average about 30 times lower than the one due to the air–gold SPP of the smooth gold film. Topographical images of the film surface reveal the typical island structure of the film consisting of ~ 20 nm-high bumps(or pits) with various submicron (down to 100 nm) sizes in the surface plane (Fig. 36 a). The optical signal obtained with the SPP being resonantly excited is much smaller than in the case of a smooth gold surface, but the optical images also exhibited some extremely bright spots (Fig. 36b). Typically, such a spot had a size of ~ 250 nm, and the peak signal in it was up to 10 times larger than the average level (Fig. 37). If the angle of incidence was out of resonance by a few

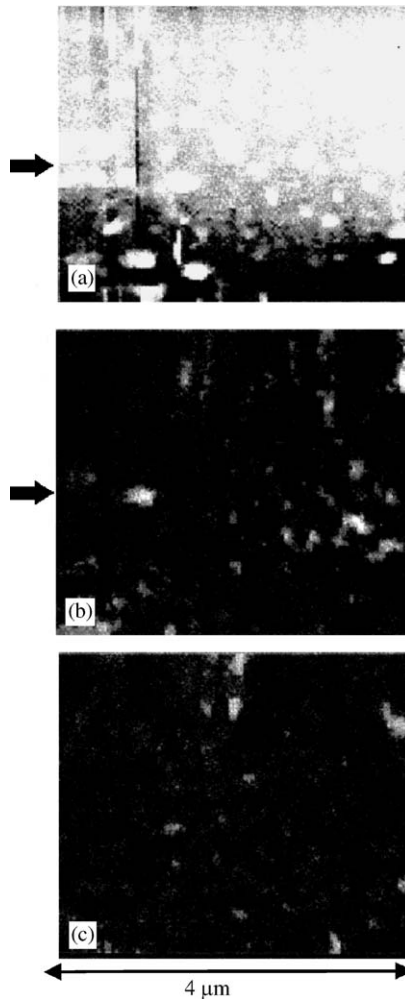


Fig. 36. Topography (a) and near-field intensity distributions (b,c) over the rough gold film. The gray-scale of the topographical image corresponds to 40 nm. The optical images were obtained at the same place under (b) resonant and (c) nonresonant ($\Delta\theta \approx 3^\circ$) conditions of SPP excitation. The optical images are presented in a common scale corresponding to 0–30 pW. The exciting light is incident upwards in the vertical direction. The arrows indicate the direction of the cross sections shown in Fig. 37. (Ref. [18].)

degrees, then the average optical signal was only two times smaller, but the previously observed bright spots disappeared completely, while similar bright spots arose at other places of the surface (Fig. 36c). The dependencies of the optical images on the tip–surface distance were also very different than that in the case of the smooth gold film. It appeared that after moving the fiber tip about 1 μm away from the surface the average optical signal decreases by about half, and it is nearly independent of the tip–surface separation, whereas the bright spots were present only in the near-field images (Fig. 38). One can notice that the observed bright spots are almost round and their locations are not correlated with the local surface topography: the position of a peak optical signal can correspond to a surface pit (Figs. 36 and 39a,b), to a surface bump (Fig. 39e,f), or to some intermediate position on the surface slope (Fig. 39c,d).

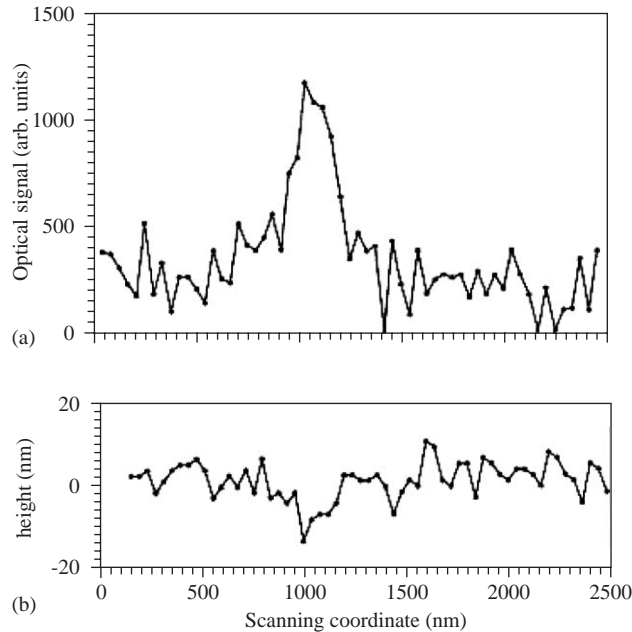


Fig. 37. The cross sections of the near-field intensity (a) and topography (b) obtained from the images in Fig. 36 along the directions indicated by the arrows. (Ref. [18].)

The gold film under consideration reveals the same character of the topographical structure in different places on the surface with obviously different topographical details. This gives the possibility to study details of the SPP localization on a different surface topography. For different topographical structures with the same characteristic parameters as above (5–100 nm defect height and 50–100 nm lateral size) the localization spots (smaller than 250 nm in diameter) are observed in the near-field optical distributions (Fig. 38). At larger distances from the surface the bright spots' sizes are increased with a simultaneous redistribution of the brightness, so that the bright spots quickly disappear (Fig. 38c,d). It should be noted that an approximately two-fold increase of spot sizes has been observed when increasing the tip–surface distance up to 500 nm. After that the size changes are less significant.

Detailed studies of the angular dependencies of the near-field images (Fig. 40c,d) show that the main difference observed for different angles of incidence is the average level of the optical signal, which most probably is related to the efficiency of the SPP coupling. The localization spots are seen on all images, but their brightness varied strongly with the angle of incidence. In general, an angular variation of about 2° was enough to extinguish or to lighten up a spot. This behavior is demonstrated by the optical images obtained for excitation angles near the SPP resonance (Fig. 40). (The bright spot in Fig. 40b is remarkable because it has the size of about 150 nm and a signal enhancement by a factor more than 7.) Such an angular dependence of the optical images can be expected for two reasons: (i) the SPP can be excited at a sufficiently rough surface in a wide range of the angles of incidence as the SPP resonance is broadened and (ii) the localization is an interference phenomenon, which means that the phase distribution (along the surface) of the scattered SPP is very important. The observed angular width of 2° , within which an individual bright spot is pronounced, can be used to estimate the size of the surface area within which

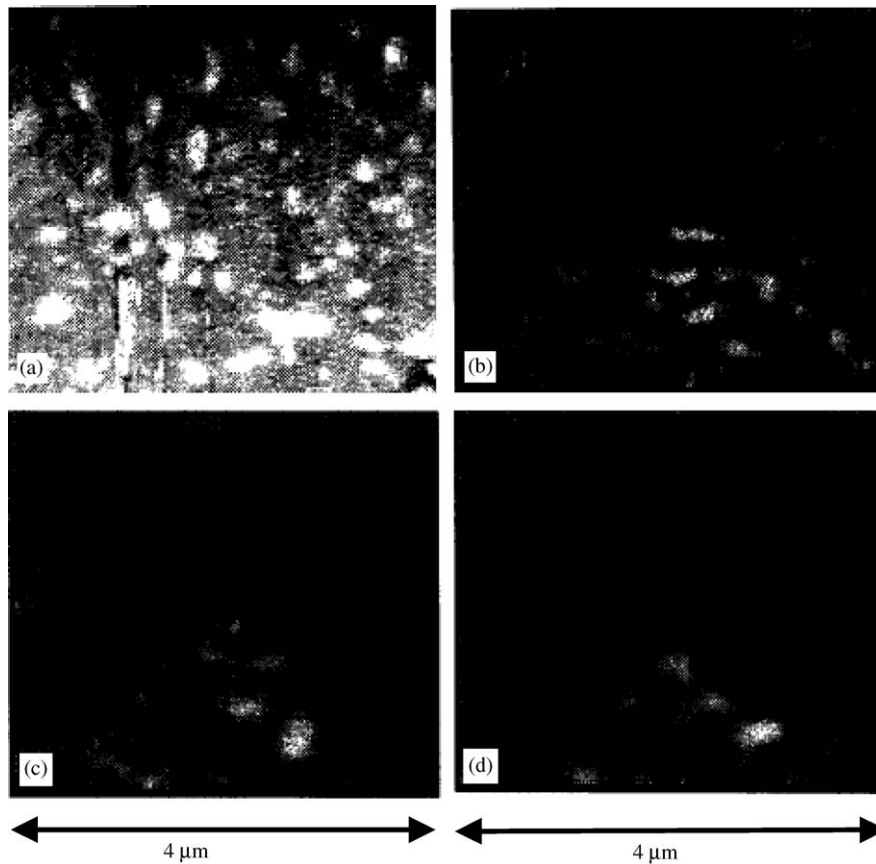


Fig. 38. Topography (a), near-field (b) and far-field (c, d) intensity distributions over the same area of the rough gold surface. The gray-scale of the topographical image corresponds to 87 nm. The optical images were obtained under resonant conditions of SPP excitation at different tip–surface distances: 5 nm (b), 0.5 μm (c) and 1.2 μm (d). The optical images are presented in different scales corresponding to (b) 0–50 and (c,d) 0–35 pW. The exciting light is incident upwards in the vertical direction. (Ref. [18].)

SPP scattering contributes to the formation of the localization spot. This has been found to be about 10 μm , which is about the same as the SPP propagation length.

The angular and distance dependencies of the optical images indicate strong interaction of the SPP with surface roughness that results in the situation that, on average, only a half of the optical signal measured near the surface is related to the SPP field intensity. Another part, which is likely to be due to the direct light scattering, should not depend much on either the tip–surface distance or the angle of incidence. Since the near-field optical images have been obtained with a constant tip–surface distance, the contribution of propagating light in the overall optical signal is known from the angular and distance dependencies of the optical images: $\sim 10\%$ on average. This contribution is about 10 times smaller than the peak signal values at bright spots. The propagating waves cannot form a spot image with a size smaller than the diffraction limit ($\lambda/2 \approx 316 \text{ nm}$). Therefore, the observed spatially localized enhancement is indeed related to surface polariton localization caused by surface roughness.

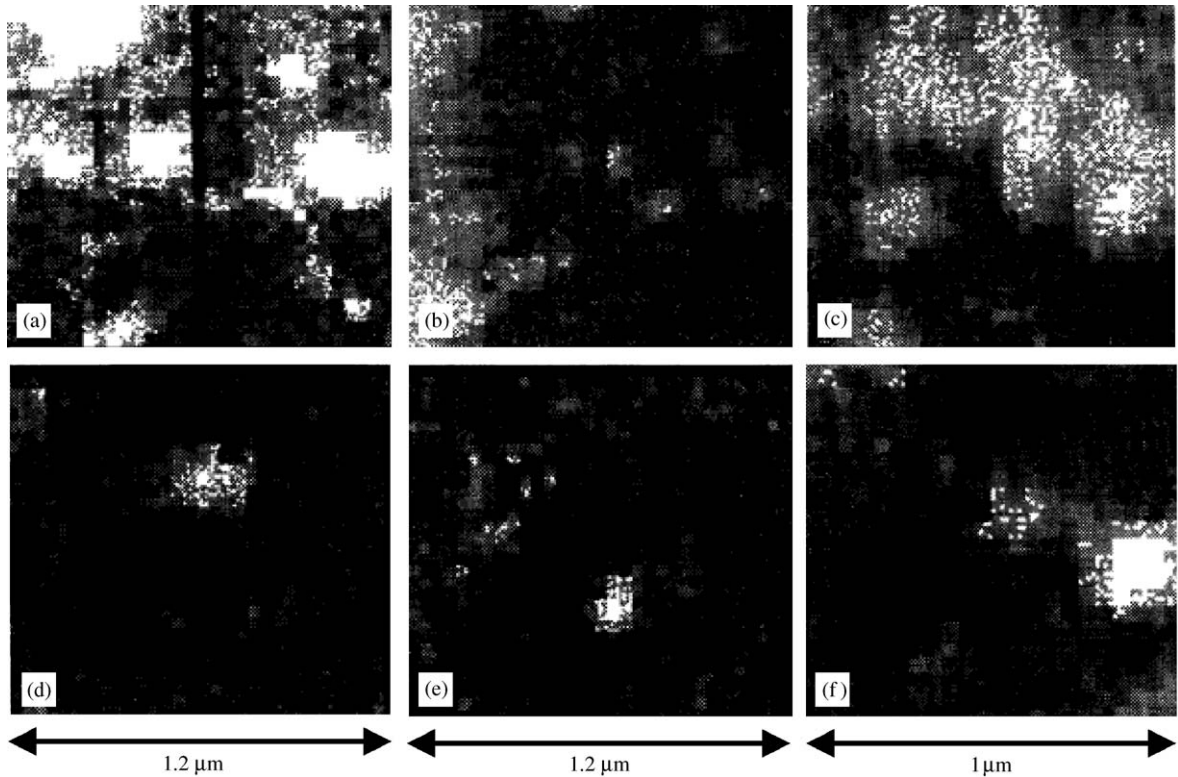


Fig. 39. Topography (top) and corresponding near-field intensity distributions (bottom) at different regions of the rough gold surface. The gray-scale of the topographical image corresponds to (a) 31, (b) 50, and (c) 27 nm. The optical images are presented in a common scale corresponding to 0–50 pW. The exciting light is incident upwards in the vertical direction. (Ref. [18].)

These observations are consistent with the theoretical results on SPP localization on randomly rough surfaces [115,123]. Strong (Anderson type) localization effects such as electron localization observed in the conductivity of metals at low temperatures [124], or light localization in strongly scattering media [101] can occur in media with strong disorder, so that interference of multiply scattered waves makes their propagation impossible leading instead to localization. Localization, which is associated with multiple scattering, can take place only if the propagation length is sufficiently long compared to the mean free path due to elastic scattering and satisfies the Ioffe–Regel criterion $2\pi l_{sc}/\lambda \sim 1$, where l_{sc} is the mean free path due to elastic scattering and λ is the wavelength. Moreover, localization in the surface plane (two-dimensional localization) is much easier to achieve than the localization of light in three dimensions. In the absence of absorption, the wave localization in two-dimensions will take place with any degree of disorder [106,125,126]. In view of this, localization of the surface polaritons in the surface plane should be expected if the Ioffe–Regel criterion is satisfied, viz., $l \sim \lambda_{sp}$, and losses associated with the SPP propagation, such as the out-of-plane SPP scattering into light, and Ohmic losses are sufficiently small $l \ll L_{SP}$ to provide significant multiple-scattering effects. These conditions are satisfied for the rough gold films discussed above.

It should be noted that the Anderson type localization of surface polaritons described above is ideologically different from the case of localized surface plasmons. Localized surface plasmons are surface

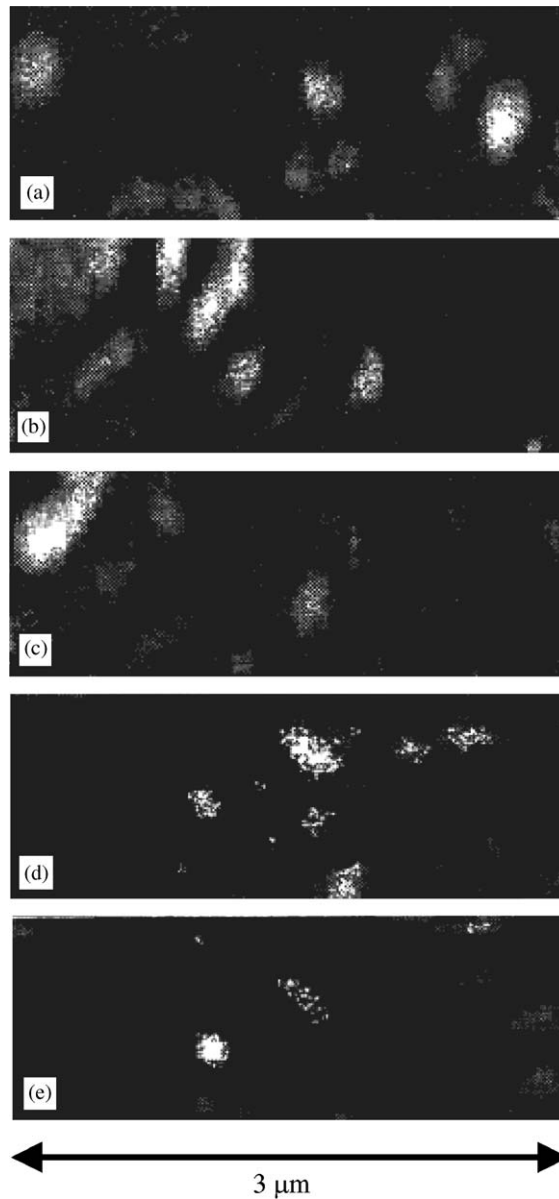


Fig. 40. Near-field intensity distributions obtained at the same area of the rough gold surface at different angles of incidence $\theta = 43^\circ$ (a), 45° (b), 47° (c), 50° (d) and 52° (e). The gray-scale of the optical images corresponds to (a,b,c,d) 0–40 and (e) 0–70 pW. The exciting light is incident upwards in the vertical direction. (Ref. [18].)

modes confined to an individual metal particle or to a defect of a metal surface whose geometrical parameters satisfy certain resonant conditions as discussed in Section 1.1.2. The excitation of LSPs leads to the electromagnetic field confinement at metal particles or features on a metal surface, while SPP localization is caused by interference effects of multiply scattered SPP. This difference was clearly demonstrated in

the SNOM imaging of the electromagnetic field above artificial 2D metallic nanostructures with which either LSP or Anderson localization can be observed for different parameters [64]. The situation might be different in the 1D case for which the contribution of localized surface plasmons on random one-dimensional defects was calculated to be more important than the Anderson localization of SPP [127]. This can be related to the different behavior of the SPP wave incident from the smooth surface (and thus significant SPP scattering effects on the boundary with the rough part of a surface) in contrast to the SPP excitation all over the rough surface and/or the specifics of the SPP scattering in the 1D case compared to the 2D geometry.

3. Two-dimensional optics of surface polaritons

The observed surface polariton behavior in a surface plane and the SPP interaction processes with surface features analogous to the processes involving light and optical elements can be generalised to introduce surface polariton optics. One can think about manipulating SPP on a surface in the same way as light is manipulated and directed in 3D with optical elements. The advantage of the SPP approach is the intrinsic two-dimensional nature of surface polaritons. Such two-dimensional quasi-optical circuits can be indispensable when the optical schemes need to be scaled down to submicron dimensions. The most obvious applications of the 2D SPP optics can be envisaged in the area of optical communications and optical computing for directing and routing optical signals between the elements of all-optical integrated circuits. Such SPP optical elements can also find application in chemical and bio-sensing for parallel processing of different SPP based sensor channels. The SPP optics approach can also allow the development of efficient local optical addressing of nanoscopic objects such as single molecules and quantum dots.

In the past, the main disadvantage of SPP optics was thought to be a relatively short SPP propagation length along a surface. For example, for silver films the propagation length is of the order of several tens of microns in the visible spectral range, reaching hundreds of microns in the near-infrared spectral range. Nowadays, however, a propagation length of several microns could be sufficient for carrying a signal in optical circuits which themselves are of microscale dimensions. In addition, surface polaritons on thin metal films in symmetrical surroundings exhibit a very long propagation range compared to SPP on the surface of the same semi-infinite metal due to the interaction between SPPs on different interfaces [27,128]. Such long-range surface polaritons can have a propagation length exceeding centimeters in the near-infrared spectral range.

Theoretical considerations of 2D optical elements for surface polariton waves have been known for a long time. Interaction of SPPs with topography steps and topographyless refraction index steps in a dielectric overlayer were considered as a way to create SPP mirrors. Fabry–Perot resonators for SPPs were theoretically described [75,129]. After the first successful attempts of SNOM imaging of SPP fields, the concept of SPP optics was introduced experimentally [17] followed by a continuously growing number of realizations of various SPP optical elements.

3.1. The interaction of surface plasmon polaritons with material and/or geometrical discontinuities in their propagation path

The surfaces on which the propagation of surface plasmon polaritons has been studied up to now in this article have terminated substrates that were homogeneous in their material properties along the normals

to them, and were homogeneous in the direction of propagation. In this section we consider the interaction of surface plasmon polaritons with material and/or geometrical discontinuities in their propagation path, such as steps on the surface or discontinuous changes in the dielectric function of the substrate.

The study of such interactions is of interest both for basic physics reasons and for possible applications of the results in the design of all-optical, integrated optics devices. Since the electromagnetic field of a surface plasmon polariton is confined to the near vicinity of the surface supporting it, it can be used as an experimental probe of the properties of the transition layer between vacuum and the bulk of the substrate or of thin films deposited on the substrate. For example, when a thin dielectric film is deposited on a metallic substrate, and the frequency-dependent dielectric function of the film possesses a pole at the frequency ω_0 of an electric-dipole-excitation (vibration) that falls in the frequency range in which surface polaritons at the metal/vacuum interface can exist, a second surface polariton can exist in this structure [130]. If the film covers only the portion $x_1 > 0$ of the metal surface and a surface polariton at the metal/vacuum interface is incident on the edge of the film from the region $x_1 < 0$, a portion of the energy in the incident surface polariton is converted into the two surface polaritons that can exist in the region $x_1 > 0$, a portion is converted into a reflected surface polariton on the vacuum/metal interface in the region $x_1 < 0$, and a portion is converted into volume waves in the vacuum above the structure. For frequencies of the incident surface polariton near ω_0 the conversion of the surface polariton into volume waves is enhanced [131], which effect can be used to obtain information about the vibrational spectrum of the film.

The refraction of guided wave polaritons incident on the interface between two waveguides or the end face of a waveguide is also of interest in guided wave technology. With the recent interest in bistability and similar all-optical operations in integrated optics structures it has been realized that the guided wave end face reflectivity determines the finesse of a guided wave cavity, and hence the critical power required for switching.

During the past decade several groups have studied the interaction of surface polaritons with material and/or geometrical discontinuities in their propagation path by a variety of theoretical approaches.

3.1.1. The impedance boundary condition

In the work of Agranovich and his colleagues the diffraction of a surface plasmon polariton incident normally on a semi-infinite dielectric film, or on a dielectric film of finite length, coating a metal surface was studied by means of an impedance boundary condition on the magnetic field in the vacuum region. This approach is valid when the thickness of the dielectric film d is small compared to the wavelength λ of the surface polariton along the surface, and when the frequency of the surface polariton is sufficiently low that the real part of the dielectric function of the metal substrate is negative and large in magnitude. The form that this boundary condition takes is

$$\left[\frac{\partial}{\partial x_3} + \beta_0(\omega) \right] H_2^>(x_1, x_3|\omega) + \frac{\partial}{\partial x_1} \left[\mu(\omega; x_1) \frac{\partial H_2^>(x_1, x_3|\omega)}{\partial x_1} \right] = 0 \quad (3.1)$$

at the plane $x_3 = 0$. In this equation $H_2^>(x_1, x_3|\omega)$ is the single, nonzero component of the total magnetic field in the vacuum region, $x_3 > 0$, when we take the plane of incidence to be the x_1x_3 -plane. The function $\beta_0(\omega) = (\omega/c)(-\epsilon(\omega))^{-1/2}$ where $\epsilon(\omega)$ is the dielectric function of the metal substrate, and $\mu(\omega; x_1) = d(\epsilon_d^{-1}(\omega; x_1) - 1)$, where d is the thickness of the dielectric film and $\epsilon_d(\omega; x_1)$ is its dielectric function. The x_1 -dependence of the latter arises from the circumstance that the film covers only a finite or semi-infinite segment of the x_1 -axis (see below).

The boundary condition (3.1) has been used in the study of the diffraction of surface plasmon polaritons from the edge of a semi-infinite dielectric film on a metal substrate [131], and from the edge of a dielectric film of finite length on a metal substrate [75]. When this boundary condition is used, only the electromagnetic field of the surface polariton in the vacuum region needs to be calculated. All of the effects of the film and metallic substrate are included in the boundary condition (3.1).

As an example, we consider the case of a surface polariton incident from the region $x_1 < 0$ onto a dielectric step that covers the metal surface in the region $0 < x_1 < L$. In this case

$$\mu(\omega; x_1) = d(\epsilon_d^{-1}(\omega) - 1)\theta(x_1)\theta(L - x_1), \quad (3.2)$$

where $\theta(x_1)$ is the Heaviside unit step function. The field amplitude $H_2^>(x_1, x_3|\omega)$ of a surface polariton in the vacuum region $x_3 > 0$ can be written as the sum of an incident and a diffracted wave,

$$H_2^>(x_1, x_3|\omega) = e^{ik(\omega)x_1 - \beta_0(\omega)x_3} + \int_{-\infty}^{\infty} \frac{dq}{2\pi} \frac{qF(q)}{\beta_0(q, \omega) - \beta_0(\omega)} e^{iqx_1 - \beta_0(q, \omega)x_3}, \quad (3.3)$$

where the wavenumber $k(\omega)$ of the incident wave is given by

$$k(\omega) = \frac{\omega}{c} \left(1 - \frac{1}{\epsilon(\omega)} \right)^{1/2}, \quad (3.4)$$

which is the form obtained from the impedance boundary condition (3.1) at a vacuum/metal interface ($\mu(\omega; x_1) \equiv 0$), together with the relation $-k^2 + \beta_0^2 + \frac{\omega^2}{c^2} = 0$. The same relation yields the result that $\beta_0(q, \omega) = (q^2 - (\omega/c)^2)^{1/2}$, with $\text{Re } \beta_0(q, \omega) > 0$, $\text{Im } \beta_0(q, \omega) < 0$. When we substitute Eq. (3.3) into the boundary condition (3.1) we find that the scattering amplitude $F(q)$ satisfies the integral equation

$$F(p) + \int_{-\infty}^{\infty} \frac{dq}{2\pi} \hat{\mu}(p - q) \frac{q^2 F(q)}{\beta_0(q, \omega) - \beta_0(\omega)} = k(\omega) \hat{\mu}(p + k(\omega)), \quad (3.5a)$$

where

$$\hat{\mu}(Q) = \int_{-\infty}^{\infty} dx_1 \mu(x_1) e^{iQx_1} = d(\epsilon_d^{-1}(\omega) - 1) \frac{e^{iQL} - 1}{iQ}. \quad (3.5b)$$

The integral equation (3.5) has been solved [75] by the factorization method [132]. The transmission coefficient for surface polaritons T has been calculated as a function of L from the result, and is plotted in Fig. 41 for the case of a silver film on an aluminum substrate. In this structure two surface polaritons of frequency ω can exist with different wavenumbers $k_1(\omega)$ and $k_2(\omega)$, with $k_2(\omega) > k_1(\omega)$ when ω is lower than a certain critical frequency. The oscillations of T of the smaller period $\Delta L_2 = \pi/k_2$ are due to the multiple reflection of the additional surface polariton (of wavenumber k_2) from the edges of the film. The oscillations of the larger period $\Delta L_{12} = \pi/(k_2 - k_1)$ are due to the interference of the two surface polaritons as they propagate in the region of the film, $0 < x_1 < L$. The experimental study of such interference phenomena may be useful for the determination of the optical properties of very thin films, down to monolayer coverage.

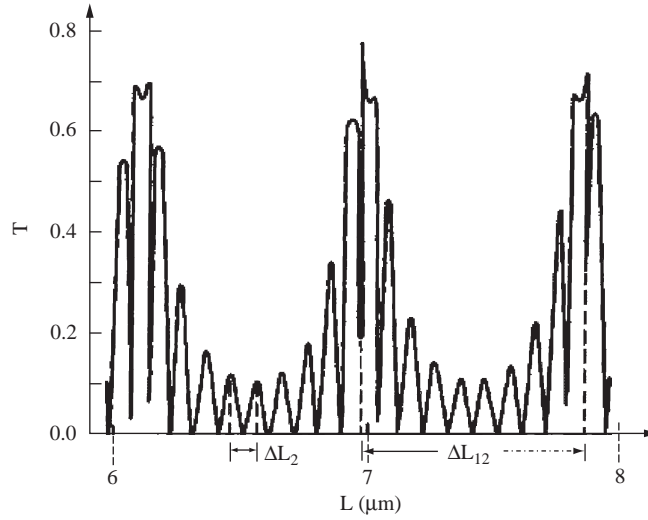


Fig. 41. The surface plasmon polariton transmission coefficient T as a function of the length of the film L in the case that an aluminum surface ($\hbar\Omega_p = 15.8$ eV) is covered by a silver film ($\hbar\omega_p = 3.8$ eV). $d = 20$ Å, $\omega/\omega_p = 0.95$. (Ref. [75].)

The work of Agranovich and his colleagues on the refraction of surface and guided wave polaritons by material and/or geometrical discontinuities in their propagation path has been summarized in a review article [133].

3.1.2. Modal expansions (discrete)

A different approach to the problem of the refraction of a surface or guided wave polariton by material and/or geometrical discontinuities in its path has been taken by Maradudin et al. [77] and their colleagues. In this work the interaction volume is limited by placing two shorting planes above and below the surface(s) on which propagation takes place. These planes are perfect conductors that force the tangential components of the electric field to vanish on their surfaces. The resulting structure is a (closed) waveguide. The electromagnetic modes of this waveguide are discrete, and provide a convenient basis in which to expand the electromagnetic field in each segment of the x_1 -axis (the direction of propagation) defined by the refracting system.

Thus, in the study of the refraction of a surface polariton by a vertical interface between two different surface active dielectric media, characterized by isotropic, frequency-dependent, dielectric functions $\epsilon_1(\omega)$ and $\epsilon_2(\omega)$, which are both negative at the frequency ω of the surface polariton, the open structure depicted in Fig. 42(a) is replaced by the closed waveguide structure depicted in Fig. 42(b). The latter structure consists of two closed waveguides of the type depicted in Fig. 43 joined at the plane $x_1 = 0$. The waveguide modes in the structure of Fig. 43 that describe a p-polarized wave propagating in the x_1 -direction are

$$H_2(x_1, x_3|\omega) = \frac{\cosh \alpha_0(d - x_3)}{\cosh \alpha_0 d} e^{i\beta x_1}, \quad 0 \leq x_3 \leq d, \quad (3.6a)$$

$$= \frac{\cosh \alpha(d + x_3)}{\cosh \alpha d} e^{i\beta x_1}, \quad -d \leq x_3 \leq 0, \quad (3.6b)$$

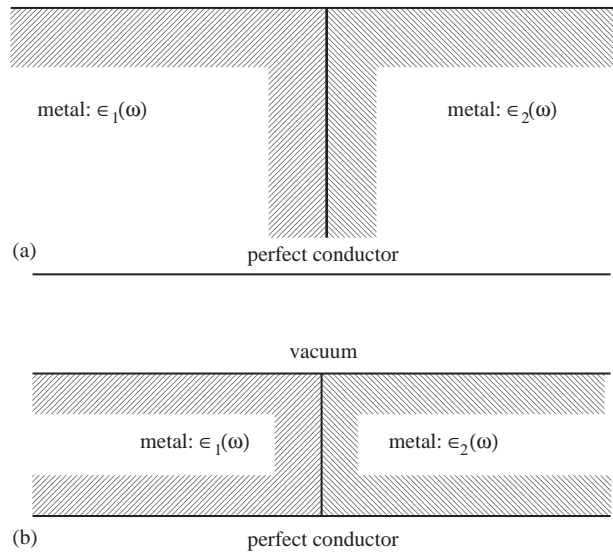


Fig. 42. (a) An open structure consisting of a metal characterized by a dielectric function $\epsilon_1(\omega)$ in contact across a vertical interface with a second metal characterized by a dielectric function $\epsilon_2(\omega)$, when both metals have a common interface with vacuum. (b) The closed waveguide structure that replaces the open structure depicted in (a).

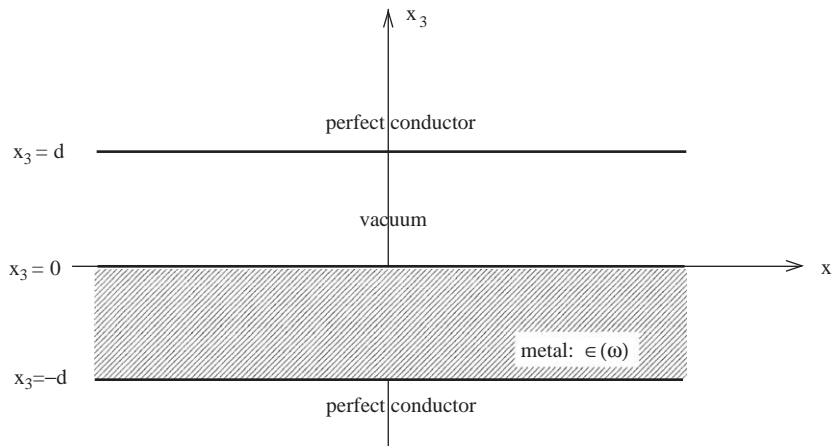


Fig. 43. A closed waveguide of the kind used in creating the structure depicted in Fig. 42(b).

$$E_1(x_1, x_3|\omega) = \frac{ic\alpha_0}{\omega} \frac{\sinh \alpha_0(d - x_3)}{\cosh \alpha_0 d} e^{i\beta x_1}, \quad 0 \leq x_3 \leq d, \quad (3.7a)$$

$$= -\frac{ic\alpha}{\omega\epsilon(\omega)} \frac{\sinh \alpha(d + x_3)}{\cosh \alpha d} e^{i\beta x_1}, \quad -d \leq x_3 \leq 0, \quad (3.7b)$$

$$E_3(x_1, x_3|\omega) = \frac{c\beta}{\omega} \frac{\cosh \alpha_0(d - x_3)}{\cosh \alpha_0 d} e^{i\beta x_1}, \quad 0 \leq x_3 \leq d, \quad (3.8a)$$

$$= -\frac{c\beta}{\omega\epsilon(\omega)} \frac{\cosh \alpha(d + x_3)}{\cosh \alpha d} e^{i\beta x_1}, \quad -d \leq x_3 \leq 0, \quad (3.8b)$$

where

$$\alpha_0(\beta, \omega) = \left(\beta^2 - \frac{\omega^2}{c^2} \right)^{1/2}, \quad \text{Re } \alpha_0 > 0, \text{Im } \alpha_0 < 0, \quad (3.9a)$$

$$\alpha(\beta, \omega) = \left(\beta^2 - \epsilon(\omega) \frac{\omega^2}{c^2} \right)^{1/2}, \quad \text{Re } \alpha > 0, \text{Im } \alpha < 0. \quad (3.9b)$$

For a given value of the frequency ω the allowed values of β for which the modes (3.6)–(3.8) are defined are the solutions of the dispersion relation

$$\epsilon(\omega) \frac{\alpha_0(\beta, \omega)}{\alpha(\beta, \omega)} = -\frac{\tanh \alpha(\beta, \omega)d}{\tanh \alpha_0(\beta, \omega)d}. \quad (3.10)$$

The solutions of this equation are discrete and are labeled by an index $m = 0, 1, 2, \dots$. They are complex in general, $\beta = \beta_R + i\beta_I$, where $\beta_R > 0, \beta_I > 0$, for a wave that propagates in the $+x_1$ -direction, or decays exponentially with increasing x_1 . The modes are ordered according to decreasing β_R with $\beta_I = 0$ (which corresponds to propagating modes) until $\beta_R = 0$. They are then ordered according to increasing β_I (which corresponds to evanescent modes). The value of β corresponding to the surface polariton is real, and is denoted by β_0 .

The remaining solutions of Eq. (3.10) describe modes of several types. These include modes which, in the present closed waveguide structure, are standing waves in the coordinate x_3 normal to the surface of propagation in the region $0 < x_3 < d$ of the waveguide filled with a dielectric medium whose dielectric constant is positive. Hence in these modes energy is being returned to the surface of propagation from the region of the perfectly conducting planes as well as flowing away from it. A similar necessity of including incoming as well as outgoing waves in obtaining the normal modes of open waveguides has been noted by Shevchenko [134]. Finally, the electromagnetic modes of our waveguide include some that decay exponentially in either the $+x_1$ - or $-x_1$ -directions.

If it is assumed that the surface polariton is incident on the interface $x_1 = 0$ from the region $x_1 < 0$, the components of the electromagnetic field in each of the regions $x_1 < 0$ and $x_1 > 0$ can be expanded in the corresponding waveguide modes. Thus, in the region $x_1 < 0$ we have

$$H_2^<(\mathbf{x}; t) = \left\{ H_2^{(10)}(x_1, x_3|\omega) + r \bar{H}_2^{(10)}(x_1, x_3|\omega) + \sum_{m(>0)} R_m \bar{H}_2^{(1m)}(x_1, x_3|\omega) \right\} e^{-i\omega t} \quad (3.11a)$$

$$E_1^<(\mathbf{x}; t) = \left\{ E_1^{(10)}(x_1, x_3|\omega) + r \bar{E}_1^{(10)}(x_1, x_3|\omega) + \sum_{m(>0)} R_m \bar{E}_1^{(1m)}(x_1, x_3|\omega) \right\} e^{-i\omega t} \quad (3.11b)$$

$$E_3^<(\mathbf{x}; t) = \left\{ E_3^{(10)}(x_1, x_3|\omega) + r \bar{E}_3^{(10)}(x_1, x_3|\omega) + \sum_{m(>0)} R_m \bar{E}_3^{(1m)}(x_1, x_3|\omega) \right\} e^{-i\omega t}, \quad (3.11c)$$

while in the region $x_1 > 0$ we have

$$H_2^>(\mathbf{x}; t) = \left\{ tH_2^{(20)}(x_1, x_3|\omega) + \sum_{m(>0)} T_m H_2^{(2m)}(x_1, x_3|\omega) \right\} e^{-i\omega t} \quad (3.12a)$$

$$E_1^>(\mathbf{x}; t) = \left\{ tE_1^{(20)}(x_1, x_3|\omega) + \sum_{m(>0)} T_m E_1^{(2m)}(x_1, x_3|\omega) \right\} e^{-i\omega t} \quad (3.12b)$$

$$E_3^>(\mathbf{x}; t) = \left\{ tE_3^{(20)}(x_1, x_3|\omega) + \sum_{m(>0)} T_m E_3^{(2m)}(x_1, x_3|\omega) \right\} e^{-i\omega t} . \quad (3.12c)$$

In these expansions the waveguide modes have been labeled by a double index (im), where $i = 1, 2$, indicates which dielectric function, $\epsilon_1(\omega)$ or $\epsilon_2(\omega)$, respectively, appears in the dispersion relation (3.10), while m labels the solution for β for a given $\epsilon(\omega)$. In addition, modes for which β is replaced by $-\beta$, corresponding to reflected waves, are denoted by a bar over the symbols of the corresponding field components.

Thus, the first term on the right hand side of each of Eqs. (3.11a)–(3.11c) represents the incident surface polariton; the second term represents the reflected surface polariton; while the third term represents all the reflected modes other than the surface polariton. Similarly, the first term on the right hand side of each of Eqs. (3.12a)–(3.12c) represents the transmitted surface polariton, while the second represents all the transmitted modes other than the surface polariton.

The coefficients in these expansions are obtained from the Maxwell boundary conditions at the interface $x_1 = 0$, which require the continuity of the tangential components of the magnetic and electric field, viz. of $H_2(\mathbf{x}; t)$ and $E_3(\mathbf{x}; t)$, across this interface. The equations expressing these conditions are

$$\begin{aligned} H_2^{(10)}(0, x_3|\omega) + r\bar{H}_2^{(10)}(0, x_3|\omega) + \sum_{m(>0)} R_m \bar{H}_2^{(1m)}(0, x_3|\omega) \\ = tH_2^{(20)}(0, x_3|\omega) + \sum_{m(>0)} T_m H_2^{(2m)}(0, x_3|\omega), \quad -d < x_3 < d, \end{aligned} \quad (3.13a)$$

and

$$\begin{aligned} E_3^{(10)}(0, x_3|\omega) + r\bar{E}_3^{(10)}(0, x_3|\omega) + \sum_{m(>0)} R_m \bar{E}_3^{(1m)}(0, x_3|\omega) \\ = tE_3^{(20)}(0, x_3|\omega) + \sum_{m(>0)} T_m E_3^{(2m)}(0, x_3|\omega), \quad -d < x_3 < d. \end{aligned} \quad (3.13b)$$

The mathematical necessity for including waveguide modes other than those corresponding to the incident, transmitted, and reflected surface polaritons in solving the problem of the refraction of a surface polariton by a transverse discontinuity is due to the impossibility of satisfying the boundary conditions at the transverse boundaries of the refracting system by the use of the electromagnetic fields of the surface polaritons on the two sides of the boundaries alone. This is caused by the different decay lengths of the surface polaritons in the directions normal to the plane of propagation. An admixture of the remaining

types of waveguide modes is needed to satisfy these boundary conditions. Physically, the presence of volume modes in the expansion of the electromagnetic field in each region of the refracting system corresponds to the conversion of a portion of the energy carried by the incident surface polariton into volume modes when the surface polariton strikes a transverse discontinuity.

The method generally used to solve a system of equations of this type was the collocation method. In this method these equations are required to be satisfied at a discrete set of N equally spaced values of x_3 in the interval $(-d, d)$ along the x_3 -axis. This requirement leads to a set of $2N$ equations in the case of Eqs. (3.13). The first N modes are used in the expansion of the field components on each side of the interface. Thus, a set of $2N$ unknowns has to be solved for the coefficients, r , t , R_m , and T_m . The number of points N is increased until convergence of the solution is achieved. The convergence criterion used was that the energy in the system be conserved, which can be formulated in the following way.

The time averaged power flow in the x_1 -direction per unit width in the x_2 -direction in any waveguide mode is given by

$$P_1 = \frac{c}{8\pi} \operatorname{Re} \int_{-d}^d dx_3 E_3(x_1, x_3|\omega) H_2^*(x_1, x_3|\omega) . \quad (3.14)$$

This quantity is nonzero only for those modes for which the corresponding β is real. This result can be used to obtain the reflection and transmission coefficients for the surface polariton. These are obtained by normalizing the power flow per unit width in the reflected and transmitted surface polaritons by that in the incident surface polariton. In this way we obtain the results

$$\mathcal{R}_{\text{sp}} = |r|^2 \left| \frac{\int_{-d}^d dx_3 \bar{E}_3^{(10)}(x_1, x_3|\omega) \bar{H}_2^{(10)}(x_1, x_3|\omega)^*}{\int_{-d}^d dx_3 E_3^{(10)}(x_1, x_3|\omega) H_2^{(10)}(x_1, x_3|\omega)^*} \right| = |r|^2 , \quad (3.15a)$$

$$\mathcal{T}_{\text{sp}} = |t|^2 \left| \frac{\int_{-d}^d dx_3 E_3^{(20)}(x_1, x_3|\omega) H_2^{(20)}(x_1, x_3|\omega)^*}{\int_{-d}^d dx_3 E_3^{(10)}(x_1, x_3|\omega) H_2^{(10)}(x_1, x_3|\omega)^*} \right| . \quad (3.15b)$$

The efficiencies of converting the incident surface polariton into reflected and transmitted volume modes are defined in the same way:

$$\mathcal{R}_m = |R_m|^2 \left| \frac{\int_{-d}^d dx_3 \bar{E}_3^{(1m)}(x_1, x_3|\omega) \bar{H}_2^{(1m)}(x_1, x_3|\omega)^*}{\int_{-d}^d dx_3 E_3^{(10)}(x_1, x_3|\omega) H_2^{(10)}(x_1, x_3|\omega)^*} \right| , \quad (3.16a)$$

$$\mathcal{T}_m = |T_m|^2 \left| \frac{\int_{-d}^d dx_3 E_3^{(2m)}(x_1, x_3|\omega) H_2^{(2m)}(x_1, x_3|\omega)^*}{\int_{-d}^d dx_3 E_3^{(10)}(x_1, x_3|\omega) H_2^{(10)}(x_1, x_3|\omega)^*} \right| . \quad (3.16b)$$

It should be kept in mind that \mathcal{R}_m and \mathcal{T}_m are nonzero only for those modes on either side of the interface $x_1 = 0$ for which the corresponding β_m is real.

Since we are dealing with lossless dielectric media the requirement of energy conservation is that all of the energy in the incident surface polariton is converted into the energy carried by the reflected and transmitted surface polaritons and by the reflected and transmitted volume modes. This condition can be expressed as

$$\mathcal{R}_{\text{sp}} + \mathcal{T}_{\text{sp}} + \sum_m \mathcal{R}_m + \sum_m \mathcal{T}_m = 1 . \quad (3.17)$$

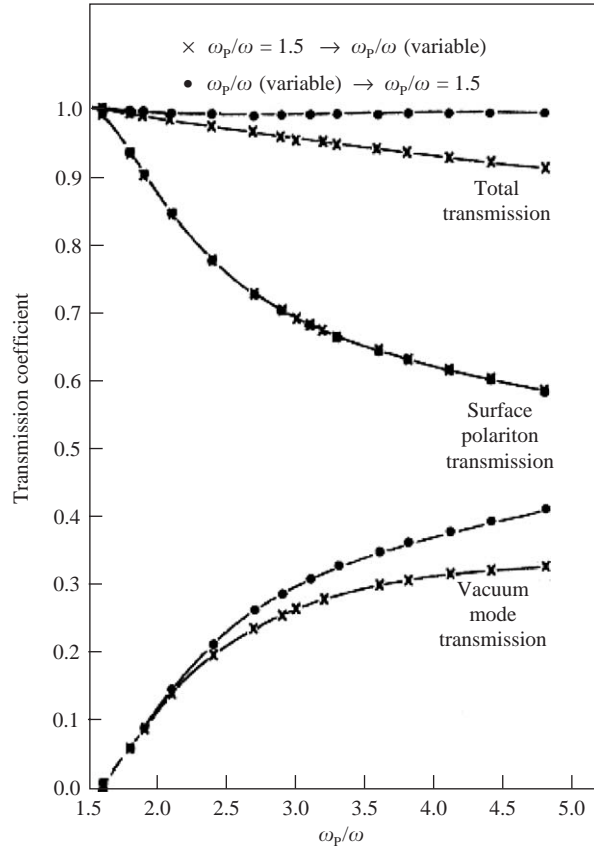


Fig. 44. The energy transmission coefficients for surface plasmon polaritons for $d = d_0 = 1$ in units of $2\pi c/\omega$ Ref. [135].

This condition could be satisfied with an error of less than 1% by making N sufficiently large, e.g., 50 to 80, depending on the specific structure studied.

The surface plasmon polariton transmission and reflection coefficients have been calculated for the structure depicted in Fig. 42(b) [135], and the transmission coefficient is plotted in Fig. 44. The dielectric function $\epsilon(\omega)$ for each metal was assumed to have the simple free electron form, $\epsilon(\omega) = 1 - (\omega_p^2/\omega^2)$, with a different plasma frequency ω_p characterizing the two metals. Two situations were considered. In the first it assumed that the frequency ω of the surface polariton incident on the interface $x_1 = 0$ is such that $\omega_p/\omega = 1.5$ where ω_p is the plasma frequency of the metal in the region $x_1 < 0$. The plasma frequency of the metal in the region $x_1 > 0$ is assumed to be variable and to increase in such a way that ω_p/ω increases from 1.5 to 5 while ω is kept fixed. It is seen from Fig. 44 that the transmission coefficient of the surface polariton decreases as the mismatch $\Delta[\omega_p/\omega]$ between the surface polariton fields on both sides of the transverse discontinuity increases. In the second case considered the surface polariton of frequency ω was incident on the interface from the region $x_1 < 0$, in which the plasma frequency ω_p was variable, so that the ratio ω_p/ω increased from 1.5 to 5. The plasma frequency of the metal in the region $x_1 > 0$ was kept constant in such a way that the ratio ω_p/ω was fixed at a value of 1.5. In this case the transmission coefficient remains constant at nearly unity as the value of ω_p/ω the region of incidence

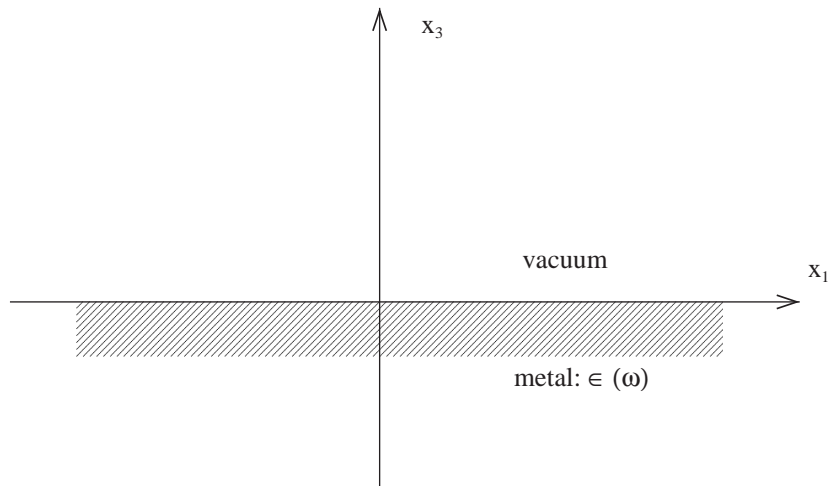


Fig. 45. A system consisting of a semi-infinite vacuum separated by a planar interface from a semi-infinite metal.

increases. Perhaps the most interesting conclusion obtained from the results depicted in Fig. 44 is the lack of reciprocity of the transmission process. That is, given a fixed set of material parameters, the surface polariton transmission coefficient depends on which side of the transverse interface the surface polariton is incident from. The reason for this nonreciprocity is the generation of volume fields and nonpropagating modes that accompanies the transmission phenomenon.

The method outlined in this subsection has been applied to the study of the interaction of a surface plasmon polariton with a variety of material and/or geometrical discontinuities in its propagation path, both when the surface polariton is incident normally on the discontinuity, and when it is incident obliquely on it. This work is summarized in the review article [77].

3.1.3. Modal expansions (continuous)

Although transmission and reflection properties of surface polaritons at transverse interfaces are well described by the results of theoretical studies of these phenomena based on waveguide modes in closed waveguides, such as were described in Section 3.1.2, those of the volume modes are not. This is due to the fact that the perfectly conducting planes that define the waveguide can be placed far enough from the plane of propagation that the surface polariton, which is localized to the vicinity of this plane, becomes insensitive to their presence. On the other hand, the volume waves, due to their radiative nature, always feel the presence of the perfectly conducting planes. A distortion of the radiation pattern, i.e. the angular distribution of the intensity of the electromagnetic radiation in the vacuum region caused by the conversion of some of the energy in the incident surface wave into radiative modes, therefore results.

This deficiency of the method described in Section 3.1.2 can be overcome if the fields on both sides of a transverse boundary are expanded in terms of the modes of an open waveguide. This has been done, albeit in an approximate fashion, by Voronko et al. [136]. In the case of p-polarized surface polaritons propagating on the structure depicted in Fig. 42(a), their procedure reduces to the following. The solution of Maxwell's equations for the surface polariton propagating in the two layer system shown in Fig. 45

can be written as

$$\mathbf{H}^s(\mathbf{x}; t) = (0, H_2^s(x_3), 0)e^{ik(\omega)x_1 - i\omega t} \quad (3.18a)$$

$$\mathbf{E}^s(\mathbf{x}; t) = (E_1^s(x_3), 0, E_3^s(x_3))e^{ik(\omega)x_1 - i\omega t}, \quad (3.18b)$$

where

$$H_2^s(x_3) = H_0 e^{-\beta_0(\omega)x_3}, \quad x_3 > 0 \quad (3.19a)$$

$$= H_0 e^{\beta(\omega)x_3}, \quad x_3 < 0, \quad (3.19b)$$

with

$$k(\omega) = \frac{\omega}{c} \left(\frac{|\epsilon(\omega)|}{|\epsilon(\omega)| - 1} \right)^{1/2}, \quad (3.20a)$$

$$\beta_0(\omega) = \frac{\omega}{c} \frac{1}{(|\epsilon(\omega)| - 1)^{1/2}}, \quad \beta(\omega) = \frac{\omega}{c} \frac{|\epsilon(\omega)|}{(|\epsilon(\omega)| - 1)^{1/2}}, \quad (3.20b)$$

while the components $E_1^s(x_3)$ and $E_3^s(x_3)$ can be obtained from

$$E_1^s(x_3) = -\frac{ic}{\omega\epsilon} \frac{\partial H_2^s(x_3)}{\partial x_3}, \quad (3.21a)$$

$$E_3^s(x_3) = -\frac{ck(\omega)}{\omega\epsilon} H_2^s(x_3). \quad (3.21b)$$

In Eqs. (3.21) ϵ denotes the dielectric function of the medium in which the electric field is being calculated. In a similar fashion the solution of Maxwell's equations representing radiative modes in the system depicted in Fig. 45 can be written as

$$\mathbf{H}^{k_3}(\mathbf{x}; t) = (0, H_2^{k_3}(x_3), 0)e^{ik_1x_1 - i\omega t}, \quad (3.22a)$$

$$\mathbf{E}^{k_3}(\mathbf{x}; t) = (E_1^{k_3}(x_3), 0, E_3^{k_3}(x_3))e^{ik_1x_1 - i\omega t}, \quad (3.22b)$$

where

$$H_2^{k_3}(x_3) = B(k_3) \left[\cos k_3x_3 + \frac{\beta}{\epsilon(\omega)k_3} \sin k_3x_3 \right], \quad x_3 > 0 \quad (3.23a)$$

$$= B(k_3)e^{\beta x_3}, \quad x_3 < 0 \quad (3.23b)$$

with

$$k_1^2 = \frac{\omega^2}{c^2} - k_3^2, \quad (3.24a)$$

$$\beta^2 = k_1^2 - \epsilon(\omega) \frac{\omega^2}{c^2} = (1 - \epsilon(\omega)) \frac{\omega^2}{c^2} - k_3^2, \quad (3.24b)$$

while

$$E_1^{k_3}(x_3) = \frac{-ic}{\omega\epsilon} \frac{dH_2^{k_3}(x_3)}{dx_3} \quad (3.25a)$$

$$E_3^{k_3}(x_3) = -\frac{ck_1}{\omega\epsilon} H_2^{k_3}(x_3) . \quad (3.25b)$$

The coefficients H_0 and $B(k_3)$ are defined by the orthogonality conditions

$$\int_{-\infty}^{\infty} dx_3 H_2^s(x_3) E_3^s(x_3) = -\frac{1}{\omega} , \quad (3.26a)$$

$$\int_{-\infty}^{\infty} dx_3 H_2^{k'_3}(x_3) E_3^{k_3}(x_3)^* = -\frac{1}{\omega} \delta(k'_3 - k_3) , \quad (3.26b)$$

and have the values

$$H_0 = \left[\frac{2\beta_0(\omega)}{ck(\omega)} \frac{\epsilon^2(\omega)}{\epsilon^2(\omega) - 1} \right]^{1/2} , \quad (3.27)$$

$$B(k_3) = \left[\frac{2}{\pi} \frac{k_3^2}{ck_1} \frac{1}{k_3^2 + (\beta^2/\epsilon^2(\omega))} \right]^{1/2} . \quad (3.28)$$

The continuity of H_2 and E_3 at $x_1 = 0$ can be expressed through the equations

$$\begin{aligned} H_{12}^s(x_3) + R\bar{H}_{12}^s + \int_0^{m_1} dk_3 R(k_3) \bar{H}_{12}^{k_3}(x_3) \\ = T H_{22}^s(x_3) + \int_0^{m_2} dk_3 T(k_3) H_{22}^{k_3}(x_3) \end{aligned} \quad (3.29a)$$

$$\begin{aligned} E_{13}^s(x_3) + R\bar{E}_{13}^s + \int_0^{m_1} dk_3 R(k_3) \bar{E}_{13}^{k_3}(x_3) \\ = T E_{23}^s(x_3) + \int_0^{m_2} dk_3 T(k_3) E_{23}^{k_3}(x_3) , \end{aligned} \quad (3.29b)$$

which have to be satisfied for all values of x_3 in the range $(-\infty, \infty)$. In these equations the overhead bar denotes a mode in which $k(\omega)$ is replaced by $-k(\omega)$ or k_1 is replaced by $-k_1$. In the double subscript (i, j) to each of the field components in these equations the first index denotes the medium in which the component is being calculated, while the second denotes the Cartesian component of the corresponding field. The upper limits m_1 and m_2 on the integrals over k_3 are obtained from the requirement that the decay constant β in each medium be positive, $k_3 \leq m = (\omega/c)(1 - \epsilon(\omega))^{1/2}$.

Equations (3.29) were solved by Voronko et al. [136] by the use of the orthogonality conditions (3.26) together with the approximation of neglecting the interaction of reflected and transmitted radiative modes with different transverse wavenumbers k_3 . (The latter approximation can be avoided at the expense of a good deal of computational effort if the reflection and transmission amplitudes of the radiative modes $R(k_3)$ and $T(k_3)$, are expanded in a complete set of orthonormal functions in the intervals $0 \leq k_3 \leq m_1$, and $0 \leq k_3 \leq m_2$, respectively.) In this way analytic expressions were obtained for the surface polariton

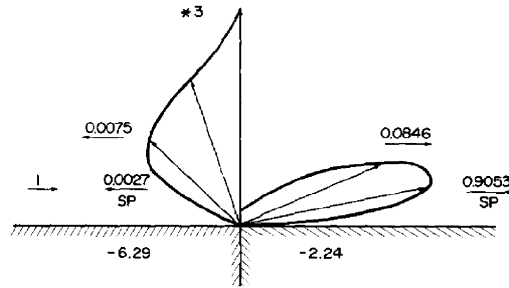


Fig. 46. The angular dependence of the energy scattered into the vacuum region when a surface plasmon polariton is incident from a metal with $\omega_p/\omega = 2.7$ on a metal with $\omega_p/\omega = 1.8$. The surface plasmon polariton transmission coefficient is 0.9053; its reflection coefficient is 0.0027 Ref. [136].

reflection and transmission amplitudes R and T , and for the radiative mode reflection and transmission amplitudes $R(k_3)$ and $T(k_3)$. From the latter the angular distribution of the intensity of the light scattered into the vacuum above the metal surface could be calculated. A typical result of these calculations is presented in Fig. 46.

Despite the attention that has been directed at the problem of the refraction of surface plasmon polaritons by material and/or geometrical discontinuities in their propagation path, and the variety of the methods that have been used in studying such problems, it seems fair to say that an approach capable of yielding numerically accurate results without the necessity of making simplifying assumptions and approximations remains elusive. The development of such an approach is a challenge to physicists working in this field.

3.2. Elements of SPP optics

Various elements are needed for applications in surface polariton optics which can efficiently perform their function (in-plane SPP manipulation) while introducing as small as possible out-of-plane losses. On the basis of the SPP behavior discussed in the previous Sections, such elements can be especially designed with appropriate assemblies of surface features of different shapes and sizes. Both topographical effects and the effects related to variations of the dielectric constant can be used to optimise the element properties. Prototypes of many elements of SPP optics have already been demonstrated. Applying a direct analogy to the element base of conventional optics, SPP mirrors, resonators, prisms, polaritonic crystals, waveguides, etc. have been proposed [17,25,59,74,75,129,137].

It is relatively easy to introduce an effective refractive index for the interaction of a SPP with a surface feature like a shallow crater or bump that scatters a SPP wave. The dispersion relation for the SPP in the air–metal–glass system can be written as [59,138]

$$\omega(k_{\text{sp}}) = ck_{\text{sp}}((\epsilon_2 + 1)/\epsilon_2)^{1/2}(1 + Ae^{-2k_2d_2}) , \quad (3.30)$$

where

$$A = 2\epsilon_2 \frac{(-\epsilon_1\epsilon_2 + \epsilon_2 - \epsilon_1)^{1/2} + i\epsilon_1}{(\epsilon_2^2 - 1)((-\epsilon_1\epsilon_2 + \epsilon_2 - \epsilon_1)^{1/2} + i\epsilon_1)} . \quad (3.31)$$

Here, $k_i^2 = k_{\text{sp}}^2 - (\omega/c)^2 \epsilon_i$ is the normal component of a wavevector. The index $i = 1, 2$ denotes the glass and the metal film, respectively. In the case of a silver film [59], if its thickness is large enough ($e^{-2k_2 d_2} \ll 1$), this coefficient can be estimated to be $A \sim 0.1$. Thus, the SPP phase velocity depends on the thickness of a metal film (Eq. (3.30)). The SPPs are faster in the areas where the film is thinner. The effective refractive index of a shallow defect with respect to SPP interaction can be then written as

$$n = v_0/v = 1 + A(e^{-2k_2 d_0} - e^{-2k_2 d}) , \quad (3.32)$$

where v is the SPP phase velocity in the region of a surface feature of thickness d and v_0 is the SPP phase velocity on a smooth metal film of the thickness d_0 . Numerical analysis of this expression shows that in the experiments with silver films described in Section 2.1.4 ($d_0 = 80$ nm) the effective refractive index n varies within the range 0.89–1.0001. For shallow bumps $n > 1$ and the changes in n are negligible, while for shallow craters the refractive index $n < 1$ and changes substantially. The SPP refractive index behavior is related to the fact that the SPP field extension in a metal is almost 10 times shorter than in air, thus “subsurface” defects interact with surface polaritons more strongly than do shallow bumps. This result agrees qualitatively with the observations in Fig. 24: shallow bumps have a much smaller effect on SPP propagation than shallow craters.

Similar effective refractive indices could be assigned to areas of the metal film where the composition of a film is modified or there is an additional overlayer on a metal surface. In some cases, the presence of an overlayer could lead to the appearance of additional SPP modes, [139] but in many cases the presence of a few molecular monolayers on the surface of a metal film leads only to a shift in the SPP phase velocity [140]. In terms of an effective refractive index, this shift can be expressed as

$$n = 1 + ((\epsilon_a - 1)/\epsilon_a (-\epsilon_m - 1)^{1/2}) 2\pi d_a / \lambda_{\text{sp}} , \quad (3.33)$$

where ϵ_a is the dielectric constant and d_a is the thickness of an adsorbed layer (here it is assumed that $\epsilon_a \ll |\epsilon_m|$). For example, four monolayers of cadmium arachidate give a shift in the SPP phase velocity corresponding to $n = 1.033$ [141]. This effect can be used to create simple optical elements for the SPP field redistribution such as lenses or prisms. A triangular region of the silver film covered with an overlayer would act as a prism on a propagating SPP beam. One could estimate the dispersion of such a prism by taking into account the dispersion of the metal ϵ_m and dielectric ϵ_a . This has the same order of magnitude as the dispersion of any bulk optical prism.

The most elementary optical element is probably a mirror. SPP reflection from finite as well as from semicontinuous topographical barriers on a surface has been described theoretically [133]. The reflection of SPP from a topography-free surface region with variations of the dielectric constant has also been studied numerically [129]. The strong in-plane reflection properties of surface defects makes them good candidates for applications as SPP mirrors if they are appropriately fabricated on a surface to provide the required directionality. Different types of SPP scattering such as presented in Figs. 20 and 26 can be used for creating a SPP mirror. The two surface features can be positioned within a few microns from each other giving rise to distinctly separate SPP beams (Figs. 18 and 22). The SPP beam from one surface feature can be directed onto another, etc. One can envisage numerous applications of such structures in sensor arrays as each SPP beam could probe the surface chemistry in a very localized region.

Another approach to SPP mirrors employs multiple scatterers of smaller lateral size to reduce the out-of-plane scattering into light. The appropriate arrangement of such scatterers can result in the efficient and controllable change of the scattered SPP direction due to interference effects of the SPPs scattered

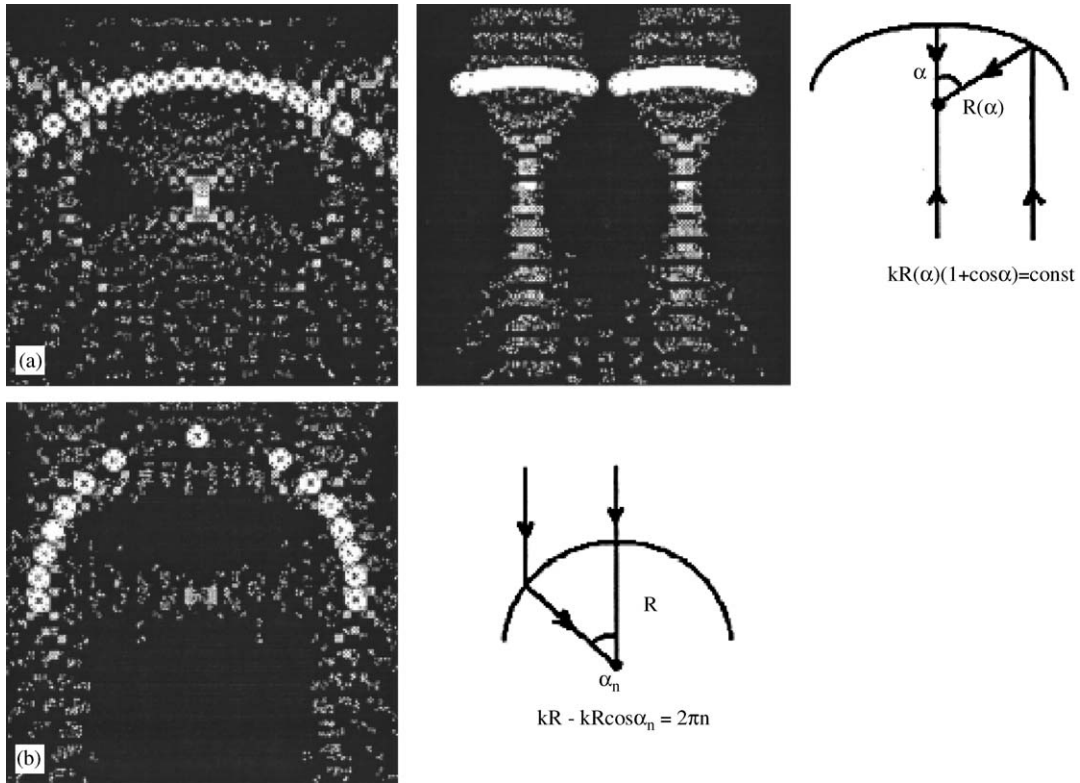


Fig. 47. SPP intensity distributions produced by (a) one and two focusing SPP mirrors (left and middle images, respectively) and (b) SPP lens. Image size is $10 \times 10 \mu\text{m}^2$. Geometries showing arrangements of individual defects in the SPP mirror and lens are also presented. Direction of the incident plane SPP wave is shown by the arrows. (Ref. [59].)

by individual defects. A set of defects can be used for this purpose to create Bragg-type reflectors for surface polaritons [24].

In addition to flat mirrors, one could think of focusing mirrors or lenses consisting of surface defects similar to those in Fig. 18. Some examples of the modeling of such SPP optical elements are shown in Fig. 47. A curved defect described in polar coordinates by $R(\alpha) = 2F/(1 + \cos \alpha)$ would represent a focusing mirror with focal length F (Fig. 47a). The positions of the point sources that direct reflected secondary SPP waves to be exactly in phase at some particular focal point, is given by the equation $kR(\alpha) + \phi(\alpha) = \text{const}$, where $\phi(\alpha) = kR(\alpha) \cos(\alpha)$. A surface object shown in Fig. 47a (middle), which consists of two focusing mirrors positioned next to each other should have two focal points. Another possibility is to create a SPP focusing lens using a zone plate principle. Fig. 47b shows the intensity distribution produced by such an element. The defects are positioned on a circle of radius R in order to compensate for the distance dependence of the field in a cylindrical scattered SPP wave. The angular position of the n th SPP source is given by the expression $kR - kR \cos(\alpha_n) = 2\pi n$. All these optical elements could find numerous applications in integrated optics.

With the availability of SPP mirrors, Fabry–Perot type resonators for surface polaritons can be introduced. Theoretically this type of resonators for SPP has already been described [75,129]. The proposed

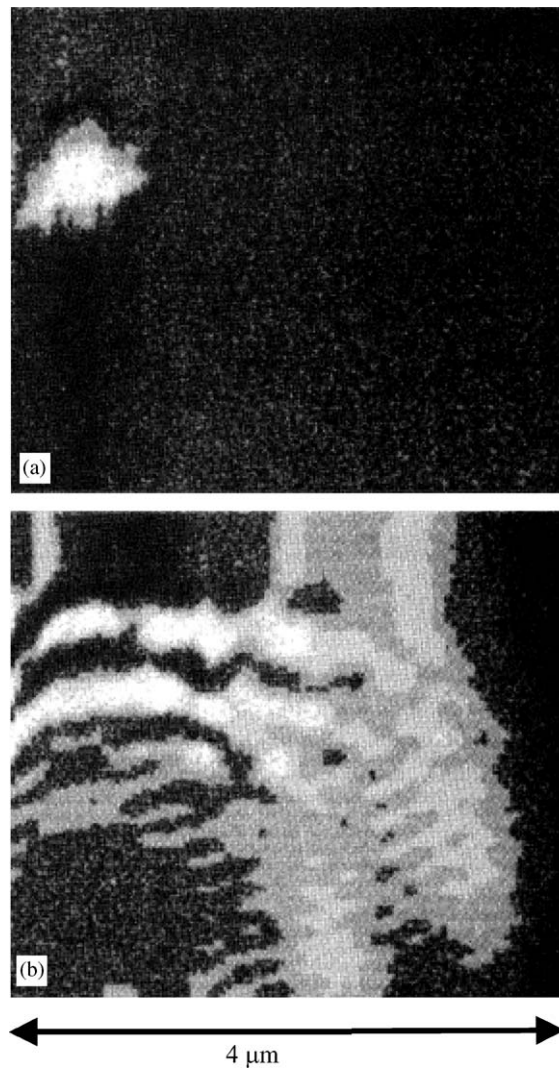


Fig. 48. Topography (a) and near-field intensity distribution (b) close to the defect playing the role of a resonator for surface polaritons. The gray-scale of the topographical image corresponds to 115 nm. The field distribution observed in the middle and right-bottom part of the optical image is related to the scattering on the subsurface defect situated in the right-bottom corner and not visible in the topography (see Ref. [17]). The SPP wave propagates upwards in the vertical direction. (Ref. [17].)

resonator can be created by two types of metal with different dielectric constants, while the surface topography across the resonator is kept smooth. Such an arrangement having a length smaller than the SPP propagation length allows efficient formation of SPP standing waves over the structure. Recent studies of metallic pads and stripes on a dielectric surface have shown that resonator-type behavior is possible to achieve with such structures [40]. In this case the reflection of SPPs occurs on the topographical structure rather than on the dielectric contrast variation. Appropriately sized circularly symmetric defects can also act as cavities for SPP creating desirable field distributions [17]. One example of such cavity is presented in Fig. 48. The topographical image reveals a relatively small bump on the surface, whose lateral size and

height are $\sim 0.5 \times 0.5 \mu\text{m}^2$ and 100 nm, respectively. This bump results in SPP scattering that leads in turn to an interference pattern. In contrast to the previously observed interference related to larger defects (Fig. 26), in this case the SPP propagates along the bump surface and scatters on the bump borders, thus forming a compact standing-wave field distribution around the bump. This can serve, at least in the first approximation, as an example of a microresonator (microcavity) for SPPs.

In analogy to photonic crystals, surface polaritonic crystals can be introduced [23] relying on periodically structured metal films influencing propagation of SPPs on the surface. One-dimensional behavior of the surface polaritons on metallic diffraction gratings has been extensively studied revealing pronounced band-gap effects [2,21]. Similar to photonic crystals, they attract enormous interest due to their waveguiding properties using specially introduced defect structures in the regular crystal lattice [25].

In addition to the discussed approach to 2D optics based on surface plasmon polaritons on nanostructured surfaces, alternative approaches have been proposed based on the guiding of SPP waves by grooves in a metal film [142] or by finite width metal stripes [143–147]. These techniques can provide a high degree of the electromagnetic field confinement and delivery to a localized place allowing to address locally single molecules. However, the bend-losses of this kind of “SPP wires”, which has not yet been studied in detail, seem to be one of the disadvantages for their photonic applications.

Another approach to light confinement and manipulation relies on discrete metallic nanostructures that are chains of closely spaced metallic nanospheres [143,148]. These can be either a simple linear chain or more complex, e.g., Y-shaped structures for optical signal splitting or combining. In the simple model, the coupling of the illuminating light into LSP on one sphere can be considered followed by LSP migration along the chain by tunneling from one sphere to another. The exact description of such discrete nanostructures should involve fully retarded electromagnetic modes of a metallic chain.

Thus, the use of modern nanofabrication techniques provides the possibility for precise engineering of surface structures of required sizes and configurations, so that eventually all elements of surface polariton optics are within reach.

4. Near-field microscopy of surface plasmon polaritons and characterization of metal surfaces

Surface characterization is a very important area of scientific efforts in the past decades. The main information required to characterise a surface is the dielectric constant of a film and surface roughness parameters. It will not be an exaggeration to say that the SPP studies have been mainly driven in the past by their surface sensitivity. Due to the SPP field localization at the metal–dielectric interface, SPP behavior is extremely sensitive to the state of the surface and has been used for the characterization of surface roughness as well as for studying optical properties related to roughness of metal films. Numerous surface polariton applications can be found in the area of chemo- and bio-sensors as even a slight (submonolayer) modification of a metal surface results in a measurable change of the SPP resonant conditions and allows monitoring the adsorption processes on a surface, and thus the environmental conditions to which the surface is exposed.

The measurements of the resonant conditions of SPP excitation on a surface have been used for determining the real and imaginary parts of the dielectric constant of a thin film [2]. Despite excellent differential sensitivity (to the changes of surface conditions) the absolute accuracy of the dielectric

constant measurements of metal films with surface polaritons is compromised by their sensitivity to surface roughness. Different films of the same material or even different places of a surface can exhibit different roughness that may lead to slightly different SPP resonant conditions and as a result, to ambiguous data derived from the SPP measurements. Since scanning near-field microscopy has been commissioned for surface polariton studies, the local characterization of metal surfaces is extensively developing using SNOM measurements of local SPP parameters.

In an approach analogous to conventional surface characterization using the SPP resonance measurements, the local SPP parameters such as a wavelength, the field extension length from and the propagation length along a surface are measured with SNOM followed by the calculation from these measurements of the dielectric constant of a metal film [63]. In addition to providing data on the local dielectric constant, this approach significantly simplifies the numerical treatments of the experimental data, as up to three different experimental parameters can be used for the derivation of the two parts of the dielectric constant. In the conventional treatments, an additional assumption (e.g., Kramers–Krönig relations) should be considered, or fitting of the angular dependencies must be performed to derive a dielectric constant.

To derive surface roughness parameters, far-field measurements of SPP scattering into light on rough surfaces has been widely employed [2,149]. The data on the far-field light scattering in principle can allow estimations of roughness parameters if the dielectric constant variations are known (or vice versa), but a mathematical treatment of the inverse scattering problem is extremely difficult. Near-field microscopy can be used for direct characterization of the surface roughness via observation of the surface polariton scattering in the plane of a surface. Although STM, AFM, or shear-force measurements can provide precise knowledge of the surface topography, optical measurements are indispensable for the studies of inner (bulk) defects in thin films, defects associated with dielectric constant variations, and defects on a buried (glass–metal) interface of a film.

4.1. Characterization of metallic nanostructures via local excitation of surface polaritons

A special consideration of dielectric constant measurements is needed in the case of micro- and nanostructures on a surface. Such structures become increasingly important in modern technology. Although the SPP resonance measurements are very sensitive to the surface structure, if the lateral size of a structure is much smaller than the size of the illuminating spot, the relative influence on the SPP resonance position will be negligible compared to the dominating reflection from an unmodified surface. Thus, conventional SPP characterization techniques are difficult to apply in such cases. Neither SPP resonance measurements nor scattering measurements may be sensitive enough if the lateral size of the structure is small.

By using near-field studies of the light distribution over a surface with metallic micro- or nanostructures the dielectric constant of metal can be directly determined via measurements of the SPP wavelength and propagation length [40]. For example, in the case of a microscopic metal stripe placed on a glass surface, the surface polariton can be excited on the metal surface as well as on the glass–metal interface due to diffraction effects at the edges of the metal eventually at any angle of illumination (Section 1.1.3). The mutual interference of excited SPP modes and light will lead to a pronounced near-field interference pattern over the metal surface (Fig. 6). SNOM measurements of this interference pattern can be used for deriving the dielectric constant of a metal, similar to the far-field surface plasmon polariton microscopy technique described in Section 2.1.2.

The near-field images of the intensity distribution over the gold stripe measured with a SNOM working in a constant intensity mode reveal interference fringes related to the different locally excited SPP modes

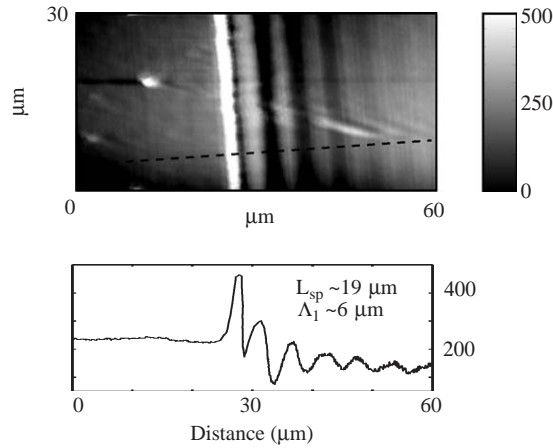


Fig. 49. Image of the electromagnetic field distribution over a gold stripe on a glass substrate (height $a=45$ nm, length $L=400$ μm) measured with a SNOM operating in a constant intensity mode for the nonresonant angle of incidence of 52° . The cross section is taken in the direction perpendicular to the interference fringes. (Ref. [40].)

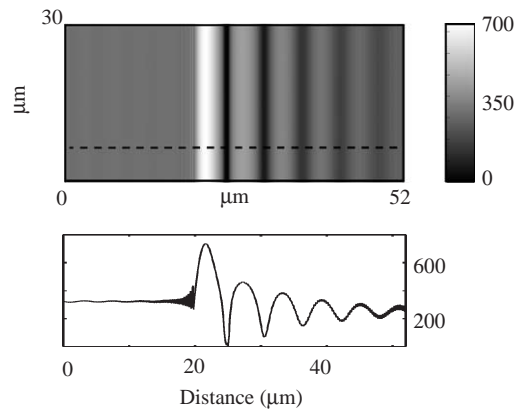


Fig. 50. Calculated image and cross section of the electromagnetic field profile in a constant intensity mode ($I = 0.08$) for the experimental parameters as in Fig. 49. (Ref. [40].)

propagating on the stripe (Fig. 49). The measurements of the interference pattern over the metal provide information on the interference period, which is directly related to the SPP wavelength, as well as on the lateral decay of the interference pattern from the edge of the metal, which is determined by the SPP decay length. Knowing these two surface polariton parameters, the real and imaginary part of the dielectric constant of a metal can be easily recalculated (Eqs. (1.6) and (1.8)).

The constant intensity SNOM image at $I = 0.08$ (assuming the incident light intensity to be $I_0 = 1$) simulated for the experimental parameters is shown in Fig. 50 (cf. Fig. 6). The period of the main fringes close to the left edge of the metal stripe ($\Lambda_1 \approx 6$ μm) corresponds to the interference involving the air–metal SPP propagating in the direction of the exciting light (Fig. 51). The interference related to the

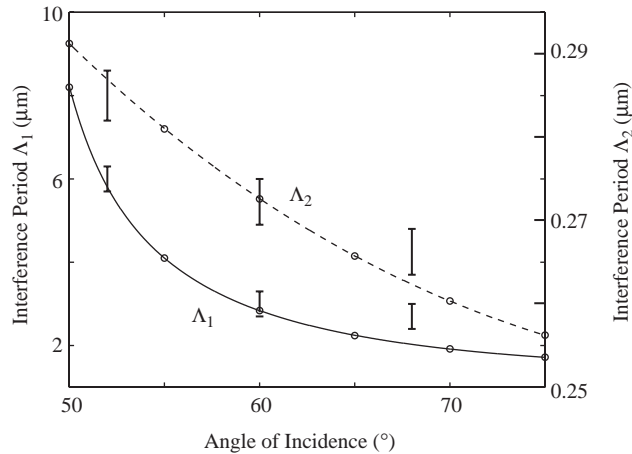


Fig. 51. Angular dependencies of the interference periods Λ_1 and Λ_2 related to the SPP excited at different edges of the metal stripe: (circles) theory, (bars) experiment. (Ref. [40].)

glass–metal SPPs and exciting light results in the distortion of the main interference pattern close to the left edge of the metal. Since the propagation length of the glass–metal SPP is almost half as long as that of the air–metal SPP, this distortion tends to disappear with distance from the edge. The short-period oscillations (Λ_2) related to the counterpropagating surface polaritons and exciting light are also visible close to the right metal edge. The interference period (Fig. 51) corresponds to the real part of the dielectric constant $\text{Re } \epsilon \approx -13.2$ at the wavelength of the exciting light $\lambda = 633 \text{ nm}$. An SPP propagation length of about $19 \mu\text{m}$ has been estimated from the experimental image. This gives the imaginary part of the dielectric constant of the metal $\text{Im } \epsilon \approx 0.83$.

Thus, both the periodicity of the interference pattern over the metal and the decay length of the oscillation amplitude can be used for determining the dielectric constant of metal microstructures. The value of the lateral decay of the interference amplitude gives information on the imaginary part of the dielectric constant of metal, while the real part of the dielectric constant can be deduced from the periodicity of the interference pattern.

4.2. Imaging of thin film interfaces with air–metal and glass–metal surface polaritons

In addition to determination of the dielectric constant of metal films, characterization of the surface and interface roughness can be done with SNOM measurements of surface polariton field distributions. As has been shown in Section 2, the scattering of SPP on defects results in a distinctive electromagnetic field distribution over a surface, which can be imaged and employed for roughness characterization. This is not crucial for the characterization of the air–metal surface which is open to inspection with STM or AFM techniques. Nevertheless, the characterization of the inner (buried) interface of a film is a more challenging problem.

If a metal film is not too thick, the fiber tip can probe both the air–metal and glass–metal SPP fields, either of which is most sensitive to the structure of the interface on which they propagate. By recording three images at the same place, namely the topographical image of the external interface obtained by the

shear-force technique, and the two near-field optical images generated in turn due to the two SPP modes, one has the possibility to study the interaction of the different surface polaritons with the surface features on the interface to which they are confined as well as on the other interface. This allows in principle to deduce the topography of the internal interface. Though this is not an easy problem, one can presume that the reconstruction of the internal surface relief might be performed at least in two extreme cases: for slowly varying (adiabatic) surface profiles and for subwavelength features (similar to pointlike scatterers) placed far away from each other. In the first case the SPP scattering can be neglected and the detected optical signal due to the SSP (propagating along the internal interface) is more or less directly related to the local film thickness. In the second case, coupling between scatterers can be disregarded and the power and phase of the scattered light is related to the size and nature (bump or pit) of the scatterer. In both cases the reconstruction of the internal interface seems feasible after some calibration procedure. Even if there is no possibility to excite the glass–metal SPP field resonantly, the investigation of the correlation between the surface topography and optical images obtained with the air–metal SPP can provide indications concerning the structure of the internal interface and/or the “bulk” defect structure of a thin film, since the air–metal SPP interacts with the defects under the film surface due to the penetration of the SPP field into the metal. The strength of this interaction obviously depends on the depth at which the defects are situated.

By analyzing the optical images and topography of the surface, some conclusion on the nature of the surface defects can be obtained [17]. For example, comparing the field distribution related to air–metal and glass–metal SPP above the bump (Fig. 28) one may presume that the gold film is relatively thick in the area corresponding to the low right corner of the image and thin at the bump area (upper right corner): the air–metal SPP field is very low over the defect (strongly scattered) while the fluoride–metal SPP field is significantly stronger in the defect area. Thus, the bump on the surface is probably related to the roughness (a bump) of the fluoride layer resulting in a bump at the surface of the gold film. In this case, the defect interacts strongly with both the air–metal and glass–metal SPPs leading to distinctive interference patterns observed with both SPPs.

Another example of the imaging of the internal interface with surface polaritons is presented in Fig. 52. The topography of an external interface at this place of a film reveals a relatively small bump, which scatters the air–metal SPP. The glass–metal SPP propagation along the internal interface was, however, not much influenced by this bump (Fig. 52c). At the same time, the image obtained with this SPP shows more than a tenfold increase of the near-field intensity at a place (low right corner in Fig. 52c) that does not exhibit any specific features, either in the external interface topography or in the optical image generated by the air–metal SPP. One may reasonably assume that there is a strong scatterer at the internal interface and it is probably related to a pit (not a bump) of the fluoride layer, otherwise it would have affected also the propagation of the air–metal SPP along the external interface.

4.3. *Fractal properties of rough surfaces and SPP scattering*

A standard description of random surface roughness can be achieved via determining the average size of the surface features and the correlation length, which is related to the mean distance between the defects assuming one or another statistical description of the roughness [150,151]. A complementary characterization of surface roughness can be provided by considering a surface with random roughness as an example of fractal (in the surface plane) structure [152]. Such two-dimensional fractal structures are sometimes called self-affine structures to distinguish them from 3D fractals. SNOM measurements

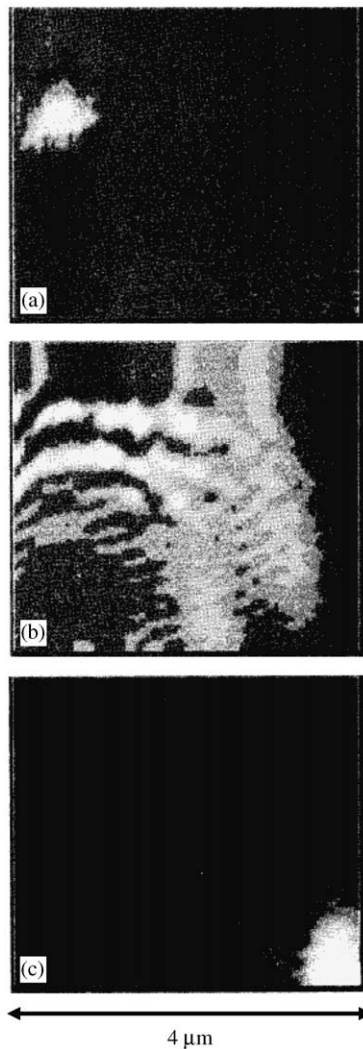


Fig. 52. Topography (a) and near-field optical images (b,c) of a relatively smooth gold film. The gray-scale of the topographical image corresponds to 115 nm. The optical images were taken at the same place with the air–metal (b) and glass–metal (c) surface polaritons being in turn resonantly excited. The optical images are presented in different gray-scales corresponding to 20–500 pW (b) and 0–50 pW (c). (Ref. [17].)

of SPP scattering provide also knowledge of the surface topography and provide for the development of a straightforward procedure for calculating the normalized fractal dimension of a surface and relating the different regimes of the SPP scattering to the difference in fractality of the studied surfaces [153].

The different regimes of SPP scattering are usually discussed using conventional characteristics of surface roughness. The topographical images obtained together with optical images can be used to deduce parameters of the defect ensemble. We shall discuss this on the example of the smooth and rough gold films with which the SPP interference and localization have been observed, respectively (Section 2).

Disregarding large bumps, which are separated by more than $5\ \mu\text{m}$, the root-mean-square roughness amplitude (RMS roughness) and the maximum peak-to-valley distance (MPV distance) were determined to be about 5 and 40 nm, respectively, for the smooth gold films at which the SPP interference has been observed (Section 2.1.5). For the rough gold films revealing the SPP localization (Section 2.3), the corresponding values of the RMS roughness and MPV distance were found to be about 10 and 120 nm, respectively. It is clearly seen from the measurements of the RMS roughness and MPV distance that the second film surface is rougher than the first one. The penetration depth of the SPP field in the gold film can be estimated to be about $\sim 35\ \text{nm}$, which is about the MPV distance (disregarding large bumps) for the first type of films and 3 times smaller than the MPV distance for the second type of films. This explains the pronounced resonance character of the SPP excitation and the absence of SPP scattering (except for that by large bumps) for the smooth gold films, as well as the weakly resonant excitation and the strong scattering of SPPs for the rough gold films. In both cases, the RMS roughness is about 10 times smaller than the corresponding MPV distance, which is a rather typical ratio for randomly rough surfaces. However, one can hardly deduce from the aforementioned characteristics of surface roughness that SPPs should be not only scattered but also strongly localized in the case of the rough surface. Moreover, with the large bumps taken into consideration the MPV distance of the first film becomes of the order of 140 nm, and the two surfaces come to look similar to each other while exhibiting distinctly different regimes of SPP scattering.

This seeming contradiction is related to the circumstance that, while useful in many instances, the conventional roughness characteristics fail to distinguish certain spatial frequency differences in surface profiles [154]. The fractal characterization of surface roughness can provide additional information in order to estimate the possibility of strong localization of SPPs. It is known that surfaces with random roughness should exhibit a fractal structure, at least in an intermediate range of sizes, and that the fractality can result in the localization of light [155].

In order to characterize the fractality of the film surface, the surface area S can be calculated as a function of the discretization step d for various experimentally measured surface profiles. A fractal structure should exhibit the dependence $S \sim d^{2-D}$, where D is the fractal (Hausdorff) dimension [156]. The calculated dependencies $\ln[S(\ln d)]$ showed the expected linear behavior, but the corresponding value of D appeared to be rather small, typically $D \sim 2.01$ ($D = 2$ is expected for a perfectly flat surface). Nevertheless, for the characterization of the SPP behavior, it is more appropriate to calculate the fractal dimension D from normalized dependencies, in which the dimensions in the surface plane are normalized by the SPP wavelength ($\lambda_{\text{sp}} \approx 570\ \text{nm}$ in this case) and the surface elevation is normalized by the SPP penetration depth ($\sim 35\ \text{nm}$). The corresponding dependencies calculated for four different regions of the rough gold film (typical topography is shown in Fig. 38) are shown in Fig. 53. The choice of the minimum discretization step ($d_0 = 40\ \text{nm}$) was dictated by the spatial resolution of the shear force technique used for topography measurements. The obtained dependencies $\ln[S(\ln d)]$ show similar and approximately linear behavior in the range of sizes 40–640 nm. Differences in their slopes as well as deviations from linearity can be explained by the fact that the area ($4 \times 4\ \mu\text{m}$) of the topographical images provided by the experimental setup is too small to ensure a representative set of surface topography data. In order to overcome this restriction, these four surface regions were artificially connected into one, which was then used to calculate the dependence $\ln[S(\ln d)]$ presented in Fig. 54. The resulting dependence is linear in the range of sizes 80–640 nm with the slope corresponding to the fractal dimension $D \approx 2.26$. The bend in the dependence for large values of d is evidently related to the absence of micron-sized surface features on the sample surface. The corresponding dependence $\ln[S(\ln d)]$ calculated for the smooth

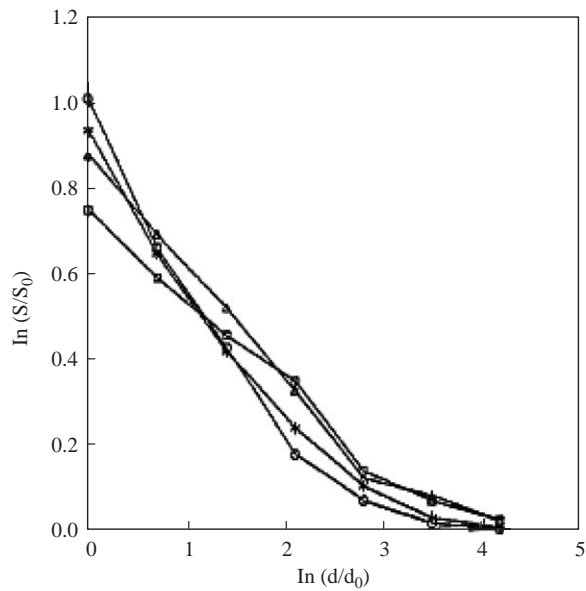


Fig. 53. Normalized surface area S as a function of the discretization step d for four different regions of the surface of the rough gold film (Section 2.3). S_0 is the area of the corresponding flat surface, and $d_0 = 40$ nm is the minimum discretization step. (Ref. [153].)

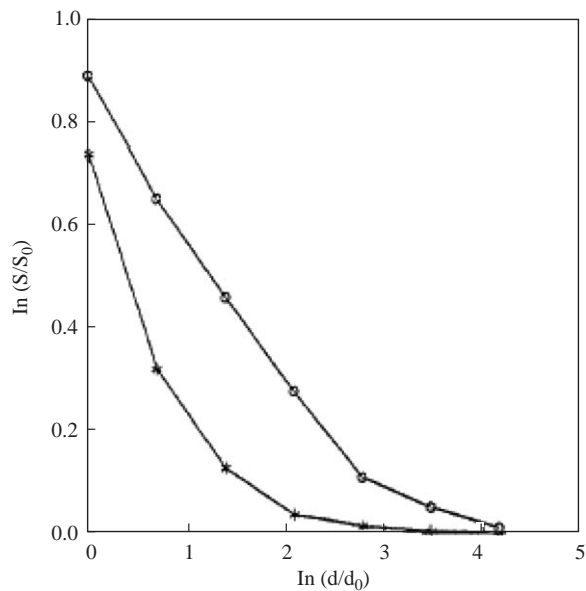


Fig. 54. Normalized surface area S as a function of the discretization step d for the surface area composed of four different regions of the surface of the rough gold film (circles) discussed in Section 2.3 and the smooth surface (large bumps excluded) of the smooth gold film (stars) discussed in Section 2.1.5. The values of S_0 and d_0 are the same as in Fig. 53. (Ref. [153].)

region of the smooth gold film surface (Fig. 26) is also shown in Fig. 54. It is clearly seen that this dependence is very different from those for the rough gold film and essentially nonlinear, suggesting that this surface cannot be viewed as fractal structures, at least on the scale $d > 80$ nm. It is also seen that this smooth gold surface can be considered as a (nearly) perfectly flat one starting from the discretization step of 320 nm. Since the SPP wavelength on this surface is about 590 nm, small-scale nonuniformities of the film do not influence the SPP propagation, and one observes the SPP scattering only by micron-sized bumps (Fig. 26). For rough gold films, the normalized surface area S is noticeably different from the flat surface area S_0 even for the discretization step of 1.28 μm (Fig. 54), indicating the strong SPP scattering.

Thus, the difference in the SPP behavior at relatively smooth and rough surfaces can be related to the difference in fractality of surfaces. The rough surfaces, which result in the strong localization of SPPs, have a fractal structure with a normalized fractal dimension $D \approx 2.26$ in the spatial range 80–640 nm. At the same time the smooth surfaces, which support well-pronounced propagating SPPs, cannot be viewed as fractal structures for discretization steps exceeding 80 nm. A surface should exhibit a fractal structure in a sufficiently large range of sizes around the SPP wavelength to be able to exhibit SPP localization.

5. Surface polaritons on periodically structured surfaces

The optical properties of spatially periodic dielectric structures called photonic crystals have recently attracted much interest [157,158]. The analogy between electromagnetic wave propagation in periodic structures and electron wave propagation in real crystals has proven to be a valuable one. Electron waves traveling in the periodic potential of a crystal lattice are arranged into energy bands separated by gaps in which propagating states are prohibited [159]. Analogous band gaps exist when electromagnetic waves propagate in a periodic dielectric structure with a period comparable to the wavelength. These frequency gaps are referred to as “photonic band gaps”. Photonic crystals can have a profound impact in many areas of pure and applied physics. For example, optical modes, spontaneous emission, and zero-point vacuum fluctuations are all absent inside a photonic bandgap. From a practical point of view, photonic crystals can be designed to transmit or reflect light within a specific range of frequencies, and their properties may be tunable by modification of the periodicity or refraction index. Moreover, when a set of defects is introduced in the periodic structure, waveguiding can be achieved along such defects that can guide light around sharp corners.

A periodic arrangement of defects on a metal–dielectric interface should exhibit the properties of a one- or two-dimensional “polaritonic crystal” when the periodicity of the surface structure is comparable with the wavelength of the SPPs propagating on the interface. A complicated system of band gaps has indeed been calculated and measured using different metallic gratings and bi-gratings [21–23,26,160–163]. As has been discussed in Section 3, shallow surface features can be assigned an effective refractive index describing the SPP interaction with the features. Thus, a complete analogy can be achieved in this case with 2D photonic crystals.

The simplest realization of polaritonic crystals is a metallic diffraction grating. Diffraction gratings are widely used for the excitation of surface polaritons on smooth metal films. In this case, surface polaritons are excited on a diffraction grating but propagate on a smooth interface, thus the SPP dispersion relations given by Eq. (1.6) (Section 1.1.3) remain valid. In the case of a periodically structured film, surface polaritons are not only excited by, but also propagate on a periodic structure. The SPP propagation on

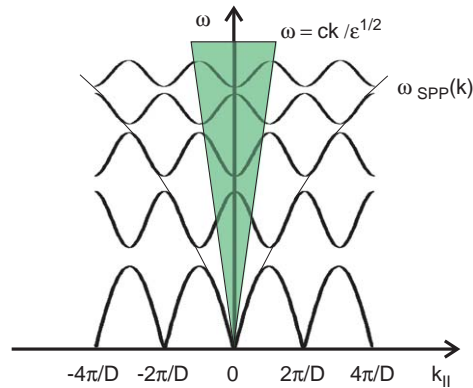


Fig. 55. Schematic of the SPP dispersion on a periodically structured surface (period D) demonstrating SPP band-gap formation. SPP dispersion on a smooth surface is shown by the thin solid line.

a periodic structure results in changes of the dispersion relations due to interaction with periodic surface features. This interaction leads to the scattering of SPPs into SPPs as well as to scattering of SPPs into light. The increased radiative losses due to the latter process, naturally, lead to the broadening and spectral shift of the SPP resonance [2]. The former process (SPP-into-SPP scattering) followed by the multiple SPP beam interference of the SPP scattered by periodic features leads to the appearance of the SPP band-gaps (Fig. 55) and SPP Bloch waves on a periodically structured surface [164]. Only the surface polaritons satisfying the Bloch waves condition can be excited and propagate on such a surface. The SPP dispersion on a metallic grating can be numerically calculated by finding the poles of the scattering matrix in the frequency–wavevector plane [23,162].

One and two-dimensional surface polaritonic crystals have recently been intensively studied [21–26,160–163]. The band-gap effects, SPP dispersion relations, excitation conditions and their dependence on grating parameters such as period, groove size, and depth have been calculated. Numerous applications of SPP crystals in optical devices for decoupling light from light-emitting structures, in resonant absorbers and reflectors, etc. have been suggested and realized. The role of one-dimensional SPP crystals formed by metallic gratings has been studied theoretically and experimentally in the context of surface enhanced Raman scattering, second-harmonic generation, as well as quantum electrodynamics effects such as spontaneous emission lifetime modification [165]. Manifestation of the SPP band-gap effects in these processes is related to suppression of SPP modes at the frequencies inside the SPP band-gap, modification of the optical density of states and the related field enhancement. SPP waveguiding on a nanostructured metal surface has been very recently demonstrated [25]. This effect has later resulted in the possibility to split and combine SPP waves using line defects in SPP crystals [166].

The SPP Bloch waves can be excited in a polaritonic crystal at a number of resonant angles at which the SPP dispersion relation (Eq. (1.6), Section 1.1.1) on a periodic structure matches light diffraction from a periodic surface (Eq. (1.32), Section 1.1.3). Thus, if a periodically structured surface is illuminated even without any special arrangements, eventually for any angle of incidence the SPP excitation can be achieved at certain illuminating light wavelengths. At the same time, the excitation and propagation of SPP at other wavelengths will be prohibited. This can be used to control the SPP excitation on a structured surface except at the places where nanostructuring is different (or the surface is smooth), thus

providing a possibility for SPP waveguiding and manipulation, multiplexing/demultiplexing, and other optical processes, which can be achieved in conventional photonic crystals.

A fundamental difference between photonic and polaritonic crystals is a different electromagnetic field distribution close to the surface. In contrast to photonic crystals, electromagnetic field enhancement takes place at a metal surface on which a SPP is excited. Such enhancement effects are absent in photonic crystals. The field enhancement results in a much stronger nonlinear response of metallic structures since it is related to the local field strength [36,167,168]. This can significantly facilitate implementation of active elements of integrated optical circuits that rely on optical nonlinear effects in SPP crystals [7,168].

5.1. *Surface polaritonic crystals*

It is a characteristic of the propagation of a quantum or classical wave in a periodic structure in any number of spatial dimensions that the dispersion relation that relates the frequency of the wave to the wave vector characterizing its propagation possesses an infinite number of branches. These branches form bands that are separated by frequency gaps at points of symmetry in the corresponding Brillouin zone. In some cases an absolute gap occurs, namely a frequency range in which no waves can propagate that exists for all values of the wave vector in the Brillouin zone, and gives rise to a gap in the density of states of the waves propagating through the structure.

Consequently, when a surface plasmon polariton propagates across a metal surface whose surface profile function is a periodic function in a single direction on the surface (a classical grating) or in two noncollinear directions (a bigrating), it is expected that the resulting dispersion relation will also possess an infinite number of branches that form bands that possess gaps at points of symmetry in the corresponding first Brillouin zone. In this section we show that this is indeed the case, and present examples of such surface polaritonic band structures.

5.1.1. *One-dimensional periodic surface profile*

The overwhelming majority of the theoretical [21–23,150,169–172] and experimental [173–181] studies of the propagation of surface plasmon polaritons across a classical grating have been carried out in the case that the sagittal plane is perpendicular to the generators of the surface. Comparatively few studies have been carried out in the case where the sagittal plane is not perpendicular to the generators of the surface. We refer to the latter case as the conical propagation of surface plasmon polaritons across a grating, by analogy with the conical scattering of light from a classical grating [182]. The first theoretical study of the conical propagation of a surface plasmon polariton across a one-dimensional periodically corrugated surface was carried out by Mills [183]. In this work the periodic corrugations of the surface were caused by the passage of a Rayleigh surface acoustic wave along the surface, and the calculations were carried out by a coupled mode approach. One of the main results of Mills' analysis was an expression for the width of the gap in the surface plasmon polariton dispersion curve between the two lowest frequency branches as a function of the angle between the direction of the surface plasmon polariton propagation and the generators of the surface. More recently, Seshadri [184] also used a coupled mode approach to study the dependence of the dispersion relation in the vicinity of a band gap of a surface plasmon polariton propagating across a classical metallic grating on the angle the wave vector of the surface plasmon polariton makes with respect to the generators of the surface. When differences in notation are taken into account and a misprint in [183] is corrected, his result agrees with that of Mills [183], and both of these

results agree with the result of Barnes et al. [23] in the case of surface plasmon polariton propagation normal to the generators of the grating surface.

The problem of obtaining the dispersion curve of a surface plasmon polariton propagating across a classical metallic grating when its two-dimensional wave vector is not normal to the generators of the surface is of interest because the extra degree of freedom provided by having a component of its wave vector parallel to the generators of the surface provides a means of varying the position and width of the gap in the surface plasmon polariton dispersion curve that occurs between the two lowest frequency branches at the boundary of the first Brillouin zone. This could be useful, for example, in the design of filters for surface wave propagation.

Because the more commonly studied propagation of a surface plasmon polariton across a classical grating when its wave vector is normal to the generators of the surface is a particular case of the conical propagation of the surface wave, in this section we study the latter, more general, case, and present results for the former case in doing so.

The physical system we consider initially consists of vacuum in the region $x_3 > \zeta(\mathbf{x}_{\parallel})$, and a metal characterized by a real, isotropic, frequency-dependent dielectric function $\epsilon(\omega)$ in the region $x_3 < \zeta(\mathbf{x}_{\parallel})$. The surface profile function $\zeta(\mathbf{x}_{\parallel})$ is assumed to be a single-valued function of \mathbf{x}_{\parallel} that is differentiable with respect to x_1 and x_2 as many times as is necessary. The dielectric function $\epsilon(\omega)$ is assumed to be negative and to satisfy the condition $\epsilon(\omega) < -1$ in some frequency range. It is within this frequency range that surface plasmon polaritons exist.

We are concerned here with the normal modes of free oscillations of the electric field in this system, not with the scattering of a volume or surface electromagnetic wave from it. Thus, we are concerned with a homogeneous problem rather than with a scattering problem. The electric field amplitude $\mathbf{E}^>(\mathbf{x}|\omega)$ in the vacuum region $x_3 > \zeta(\mathbf{x}_{\parallel})$ is therefore given by the second term on the right hand side of Eq. (2.24):

$$\begin{aligned} E^>(\mathbf{x}|\omega) = & \int \frac{dq_{\parallel}}{(2\pi)^2} \left\{ \frac{c}{\omega} [\mathbf{i}\hat{\mathbf{q}}_{\parallel}\beta_0(q_{\parallel}) - \hat{\mathbf{x}}_3 q_{\parallel}] A_p(\mathbf{q}_{\parallel}) + (\hat{\mathbf{x}}_3 \times \hat{\mathbf{q}}_{\parallel}) A_s(\mathbf{q}_{\parallel}) \right\} \\ & \times \exp[\mathbf{i}\mathbf{q}_{\parallel} \cdot \mathbf{x}_{\parallel} - \hat{\mathbf{x}}_3 \beta_0(q_{\parallel}) x_3] . \end{aligned} \quad (5.1)$$

The amplitudes $A_p(\mathbf{q}_{\parallel})$ and $A_s(\mathbf{q}_{\parallel})$ now satisfy the homogeneous form of the reduced Rayleigh equation, Eq. (2.28), which we rewrite in the following form [185]:

$$\begin{aligned} & \int \frac{dq_{\parallel}}{(2\pi)^2} \frac{K(\beta(p_{\parallel}) - \beta_0(q_{\parallel})) |\mathbf{p}_{\parallel} - \mathbf{q}_{\parallel}|}{\beta(p_{\parallel}) - \beta_0(q_{\parallel})} \\ & \times \begin{pmatrix} p_{\parallel} q_{\parallel} - \beta(p_{\parallel}) \hat{\mathbf{p}}_{\parallel} \cdot \hat{\mathbf{q}}_{\parallel} \beta_0(q_{\parallel}) & -\mathbf{i}(\omega/c) \beta(p_{\parallel}) (\hat{\mathbf{p}}_{\parallel} \times \hat{\mathbf{q}}_{\parallel})_3 \\ \mathbf{i}(\omega/c) (\hat{\mathbf{p}}_{\parallel} \times \hat{\mathbf{q}}_{\parallel})_3 \beta_0(q_{\parallel}) & (\omega/c)^2 \hat{\mathbf{p}}_{\parallel} \cdot \hat{\mathbf{q}}_{\parallel} \end{pmatrix} \begin{pmatrix} A_p(\mathbf{q}_{\parallel}) \\ A_s(\mathbf{q}_{\parallel}) \end{pmatrix} = 0 , \end{aligned} \quad (5.2)$$

where

$$K(\gamma|\mathbf{Q}_{\parallel}) = \int d^2x_{\parallel} \exp(-\mathbf{i}\mathbf{Q}_{\parallel} \cdot \mathbf{x}_{\parallel}) \exp(\gamma\zeta(\mathbf{x}_{\parallel})) . \quad (5.3)$$

The relation between $K(\beta(p_{\parallel}) - \beta_0(q_{\parallel})) |\mathbf{p}_{\parallel} - \mathbf{q}_{\parallel}|$ and the function $J(\mathbf{p}_{\parallel}|\mathbf{q}_{\parallel})$ entering Eqs. (2.30) is

$$K(\beta(p_{\parallel}) - \beta_0(q_{\parallel})) |\mathbf{p}_{\parallel} - \mathbf{q}_{\parallel}| = (2\pi)^2 \delta(\mathbf{p}_{\parallel} - \mathbf{q}_{\parallel}) + (\beta(p_{\parallel}) - \beta_0(q_{\parallel})) J(\mathbf{p}_{\parallel}|\mathbf{q}_{\parallel}) . \quad (5.4)$$

We now specialize to the case of interest, namely the case in which the surface profile function $\zeta(\mathbf{x}_{\parallel})$ is a periodic function of x_1 with period a , and is independent of the coordinate x_2 :

$$\zeta(x_1 + a) = \zeta(x_1) . \quad (5.5)$$

In this case the function $K(\gamma|\mathbf{Q}_{\parallel})$ in Eq. (5.2) becomes

$$\begin{aligned} K(\gamma|\mathbf{Q}_{\parallel}) &= \int_{-\infty}^{\infty} dx_1 \exp(-iQ_1 x_1) \exp(\gamma\zeta(x_1)) \int_{-\infty}^{\infty} dx_2 \exp(-iQ_2 x_2) . \\ &= 2\pi\delta(Q_2) \sum_{n=-\infty}^{\infty} \int_{(n-\frac{1}{2}a)}^{(n+\frac{1}{2}a)} dx_1 \exp(-iQ_1 x_1) \exp(\gamma\zeta(x_1)) \\ &= 2\pi\delta(Q_2) \sum_{n=-\infty}^{\infty} \exp(-iQ_1 n a) \int_{-\frac{1}{2}a}^{\frac{1}{2}a} dx \exp(-iQ_1 x) \exp(\gamma\zeta(x_1)) \\ &= 2\pi\delta(Q_2) \sum_{m=-\infty}^{\infty} 2\pi\delta(Q_1 - (2\pi m/a)) \tilde{K}_m(\gamma) , \end{aligned} \quad (5.6)$$

where

$$\tilde{K}_m(\gamma) = \frac{1}{a} \int_{-\frac{1}{2}a}^{\frac{1}{2}a} dx_1 \exp(-i(2\pi m/a)x_1) \exp(\gamma\zeta(x_1)) . \quad (5.7)$$

In order that the electric field in the vacuum region, Eq. (5.1) satisfy the Bloch–Floquet theorem, we set

$$A_{p,s}(\mathbf{q}_{\parallel}) = \sum_{n=-\infty}^{\infty} 2\pi(q_1 - k_1 - (2\pi n/a)) 2\pi\delta(q_2 - k_2) a_{p,s}^{(n)}(\mathbf{k}_{\parallel}) , \quad (5.8)$$

where $\mathbf{k}_{\parallel} = (k_1, k_2)$ is the two-dimensional wave vector of the surface plasmon polariton.

When the results given by Eqs. (5.6) and (5.8) are substituted into Eq. (5.2), we obtain as the equation satisfied by $a_{p,s}^{(n)}(\mathbf{k}_{\parallel})$

$$\begin{aligned} &\sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \frac{(2\pi)^2 \delta(p_1 - K_m) \delta(p_2 - k_2)}{(K_m^2 + k_2^2)^{1/2} (K_n^2 + k_2^2)^{1/2}} \\ &\times \frac{\tilde{K}_{m-n}([K_m^2 + k_2^2 - \epsilon(\omega)\omega^2/c^2]^{1/2} - [K_n^2 + k_2^2 - \omega^2/c^2]^{1/2})}{[K_m^2 + k_2^2 - \epsilon(\omega)\omega^2/c^2]^{1/2} - [K_n^2 + k_2^2 - \omega^2/c^2]^{1/2}} \\ &\times \left(\begin{array}{l} (K_m^2 + k_2^2)(K_n^2 + k_2^2) - [K_m^2 + k_2^2 - \omega^2/c^2]^{1/2} (K_m K_n + k_2^2) [K_n^2 + k_2^2 - \omega^2/c^2]^{1/2} \\ i(\omega/c) [K_m^2 + k_2^2 - \omega^2/c^2]^{1/2} k_2 (K_m - K_n) \\ -i(\omega/c) [K_m^2 + k_2^2 - \epsilon(\omega)\omega^2/c^2]^{1/2} k_2 (K_m - K_n) \\ (\omega/c)^2 (K_m K_n + k_2^2) \end{array} \right) \begin{pmatrix} a_p^{(n)}(\mathbf{k}_{\parallel}) \\ a_s^{(n)}(\mathbf{k}_{\parallel}) \end{pmatrix} = 0 , \end{aligned} \quad (5.9)$$

where, to simplify the notation, we have defined $K_m = k_1 + (2\pi m/a)$. On equating to zero the coefficient of each delta function, we obtain the equation for $a_p^{(n)}(\mathbf{k}_{\parallel})$ and $a_s^{(n)}(\mathbf{k}_{\parallel})$ in the form

$$\sum_{n=-\infty}^{\infty} \frac{1}{(K_m^2 + k_2^2)^{1/2}(K_n^2 + k_2^2)^{1/2}} \frac{\tilde{K}_{m-n}([K_m^2 + k_2^2 - \epsilon(\omega)(\omega/c)^2]^{1/2} - [K_n^2 + k_2^2 - (\omega/c)^2]^{1/2})}{[K_m^2 + k_2^2 - \epsilon(\omega)(\omega/c)^2]^{1/2} - [K_n^2 + k_2^2 - (\omega/c)^2]^{1/2}} \times \left(\begin{aligned} & (K_m^2 + k_2^2)(K_n^2 + k_2^2) - [K_m^2 + k_2^2 - \epsilon(\omega)(\omega/c)^2]^{1/2}(K_m K_n + k_2^2)[K_n^2 + k_2^2 - (\omega/c)^2]^{1/2} \\ & i(\omega/c)(K_m - K_n)k_2[K_n^2 + k_2^2 - (\omega/c)^2]^{1/2} \\ & -i(\omega/c)[K_m^2 + K_n^2 - \epsilon(\omega)(\omega/c)^2]^{1/2}k_2(K_m - K_n) \end{aligned} \right) \left(\begin{aligned} & a_p^{(n)}(\mathbf{k}_{\parallel}) \\ & a_s^{(n)}(\mathbf{k}_{\parallel}) \end{aligned} \right) = 0$$

$$m = 0, \pm 1, \pm 2, \dots \quad (5.10)$$

The dispersion relation for the conical propagation of surface plasmon polaritons is obtained by equating to zero the determinant of the matrix of the coefficients in this system of equations.

If we set $k_2 = 0$ in Eq. (5.10), so that the surface plasmon polariton is propagating normal to the generators of the surface, the system of equations decouples into a system of equations for the $\{a_p^{(n)}(\mathbf{k}_{\parallel})\}$ alone and a system of equations for the $\{a_s^{(n)}(\mathbf{k}_{\parallel})\}$ alone:

$$\sum_{n=-\infty}^{\infty} \frac{K_m K_n}{|K_m| |K_n|} \frac{\tilde{K}_{m-n}([K_m^2 - \epsilon(\omega)(\omega/c)^2]^{1/2} - [K_n^2 - (\omega/c)^2]^{1/2})}{[K_m^2 - \epsilon(\omega)(\omega/c)^2]^{1/2} - [K_n^2 - (\omega/c)^2]^{1/2}}, \quad m = 0, \pm 1, \pm 2, \dots, \quad ,$$

$$\times \{K_m K_n - [K_m^2 - \epsilon(\omega)(\omega/c)^2]^{1/2}[K_n^2 - (\omega/c)^2]^{1/2}\} a_p^{(n)}(\mathbf{k}_{\parallel}) = 0 \quad (5.11a)$$

$$\sum_{n=-\infty}^{\infty} \frac{K_m K_n}{|K_m| |K_n|} \frac{\tilde{K}_{m-n}([K_m^2 - \epsilon(\omega)(\omega/c)^2]^{1/2} - [K_n^2 - (\omega/c)^2]^{1/2})}{[K_m^2 - \epsilon(\omega)(\omega/c)^2]^{1/2} - [K_n^2 - (\omega/c)^2]^{1/2}} a_s^{(n)}(\mathbf{k}_{\parallel}) = 0, \quad ,$$

$$m = 0, \pm 1, \pm 2, \dots \quad (5.11b)$$

The solvability condition for each system of equations is that the determinant of the matrix of coefficients in it vanishes. The system of equations (5.11a) yields a computationally convenient dispersion relation for surface plasmon polaritons propagating normal to the generators of the surface. The second system of equations, Eq. (5.11b), does not appear to have any solutions in the nonradiative region of the (ω, k_1) -plane, i.e. s-polarized surface plasmon polaritons do not exist at a periodically corrugated vacuum–metal interface, just as they do not exist at a planar vacuum–metal interface.

The solutions $\omega(\mathbf{k}_{\parallel})$ of the dispersion relation obtained from Eq. (5.10) possess three general properties. The first is that for a fixed value of k_2 $\omega(\mathbf{k}_{\parallel})$ is a periodic function of k_1 with period $2\pi/a$, $\omega(k_1 + (2\pi/a), k_2) = \omega(k_1, k_2)$. The second is that $\omega(\mathbf{k}_{\parallel})$ is an even function of \mathbf{k}_{\parallel} , $\omega(-\mathbf{k}_{\parallel}) = \omega(\mathbf{k}_{\parallel})$. The third is that for a given value of k_2 $\omega(\mathbf{k}_{\parallel})$ is an even function of k_1 , $\omega(-k_1, k_2) = \omega(k_1, k_2)$. These three properties of $\omega(\mathbf{k}_{\parallel})$ have the consequence that all the distinct solutions of the dispersion relation are obtained if the wavenumber k_1 is restricted to the interval $0 \leq k_1 \leq \pi/a$, while k_2 varies throughout the range $0 \leq k_2 \leq \infty$.

From Eq. (2.27) it is straightforward to show that if k_1 and k_2 are real, and $(k_1^2 + k_2^2)^{1/2}$ is larger than ω/c , $\beta_0(K_n, k_2)$ is purely real and positive for all n when ω is also real. From Eqs. (5.1) and (5.8) it then follows that the electric field in the vacuum region in this case tends to zero as $x_3 \rightarrow \infty$. At the same time, since $\epsilon(\omega)$ is assumed to be real and negative, it follows that $\beta(K_n, k_2)$ is purely real and positive for all n when ω is also real. This means that the electric field in the metal also tends to zero

as $x_3 \rightarrow -\infty$. Thus, the Bloch-type surface plasmon polaritons that are true eigenmodes of the electric field in the presence of the periodically modulated metal surface for a given value of k_2 are found only in the region of the (ω, k_1) -plane bounded from the left by the dispersion curve of volume electromagnetic waves in the vacuum, $\omega = c(k_1^2 + k_2^2)^{1/2}$, and from the right by the boundary of the first Brillouin zone, $k_1 = \pi/a$. This region is called the nonradiative region of the (ω, k_1) -plane. It is in this region that the solutions of the dispersion relation obtained from Eq. (5.10) will be sought.

In determining the dispersion curves the infinite determinant formed from the coefficients in Eq. (5.10) is replaced by the determinant of a $2(2N + 1) \times 2(2N + 1)$ matrix obtained by restricting m and n to run from $-N$ to N . The zeros of the determinant were found by fixing k_1 and k_2 , increasing ω from 0 to ω_p in small increments, and looking for changes in the sign of the determinant. The convergence of the solutions found was tested by increasing N and seeing if they approached stable limiting values. Values of $N = 5$ proved sufficient to obtain solutions that changed by less than 0.1% as N was increased to $N + 1$.

In the numerical calculations the relations $k_1 = k_{\parallel} \cos \theta$, $k_2 = k_{\parallel} \sin \theta$ were used, where θ is the angle between the surface plasmon polariton wave vector and the x_1 -axis, i.e. the normal to the generators of the surface. The angle θ was varied in the interval $0 \leq \theta \leq \pi/2$, and k_{\parallel} was restricted by the condition $0 \leq k_{\parallel} \leq \pi/(a \cos \theta)$, i.e. $0 \leq k_1 \leq \pi/a$. The vacuum light line is given by $\omega = ck_{\parallel}$.

In Fig. 56 we have plotted the dispersion curves obtained in this fashion as functions of k_{\parallel} for four increasing values of θ for surface plasmon polaritons propagating across a silver surface defined by the symmetric sawtooth profile

$$\zeta(x_1) = \begin{cases} H + \frac{4H}{a}x_1, & -\frac{a}{2} \leq x_1 \leq 0, \\ H - \frac{4H}{a}x_1, & 0 \leq x_1 \leq \frac{a}{2}, \end{cases} \quad (5.12)$$

for $|x_1| \leq a/2$. The values of H and a assumed are $H = 0.005 \mu\text{m}$ and $a = 0.2 \mu\text{m}$. The Fourier coefficient $\tilde{K}_m(\gamma)$ for this surface profile function is given by

$$\tilde{K}_m(\gamma) = \begin{cases} \frac{4H\gamma \sinh H\gamma}{4H^2\gamma^2 + \pi^2m^2} & m \text{ even}, \\ \frac{4H\gamma \cosh H\gamma}{4H^2\gamma^2 + \pi^2m^2} & m \text{ odd}. \end{cases} \quad (5.13)$$

The dielectric function of silver is assumed to have the simple free electron metal form $\epsilon(\omega) = 1 - (\omega_p/\omega)^2$, where a value of $\hbar\omega_p = 3.78 \text{ eV}$ appropriate for silver is assumed [140]. Only the three lowest frequency branches have been plotted.

Strictly speaking, according to a theorem of Millar [186], the Rayleigh hypothesis is not valid for this nonanalytic surface profile. Nevertheless, earlier calculations of the dispersion curves of surface plasmon polaritons propagating normal to the generators of this surface showed that convergent results can be obtained if $H/a \lesssim 0.2$ and N is not too large [22]. The latter restriction is due to the fact that if a finite number of terms is kept in the Fourier series for $\zeta(x_1)$, the result is an analytic surface profile function for which the Rayleigh hypothesis is valid for sufficiently small H/a . However, as the number of terms in the Fourier series is increased, it approaches more and more closely a nonanalytic surface profile function, the range of values of H/a for which the Rayleigh hypothesis is valid becomes smaller and smaller [187], and in the limit as $N \rightarrow \infty$ it vanishes.

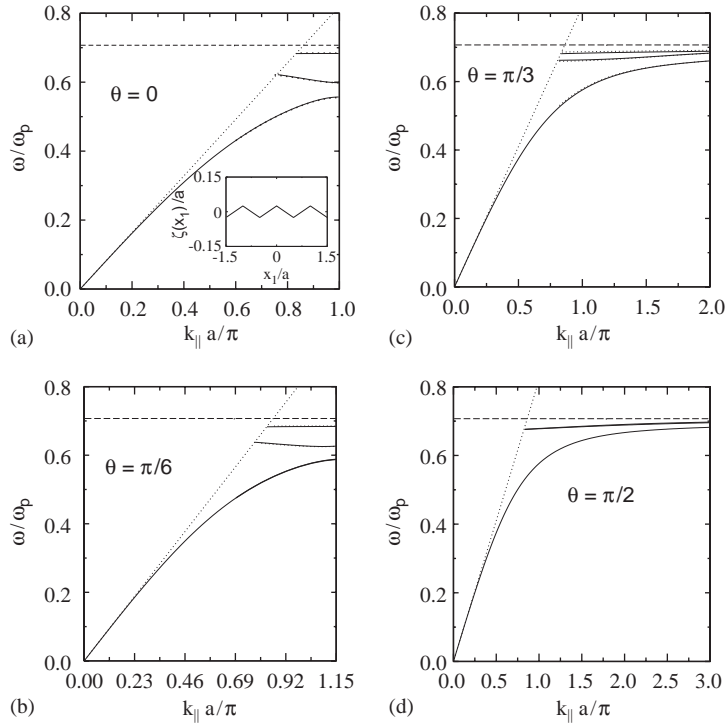


Fig. 56. Dispersion curves for the three lowest frequency branches as functions of k_{\parallel} for surface plasmon polaritons propagating across the symmetric sawtooth grating defined by Eq. (5.11) with $H = 0.005 \mu\text{m}$ and $a = 0.2 \mu\text{m}$, for four values of θ . The inset to Fig. 56(a) shows three periods of the surface profile function. The horizontal dashed curve is at the frequency $\omega = \omega_p/\sqrt{2}$, the maximum frequency of a surface plasmon polariton at a planar vacuum–metal interface, and the dotted curve is the vacuum light line $\omega = ck_{\parallel}$. (Ref. [188].)

In the results presented in Fig. 56 the case $\theta = 0$ corresponds to the propagation of the surface plasmon polariton normal to the generators of the surface. As the angle θ increases, the frequencies of all three branches plotted increase monotonically, the frequency of the center of the gap between the two lowest frequency branches at the boundary of the Brillouin zone, $k_{\parallel} = \pi/(a \cos \theta)$, increases while the width of the gap decreases, and vanishes at $\theta = \pi/2$. At this value of θ the surface plasmon polariton is propagating parallel to the ridges and grooves of the grating, and the boundary of the first Brillouin zone has moved off to infinity. The dispersion curves in this case have therefore been plotted for k_{\parallel} out to only $3\pi/a$. It is seen that for $\theta = \pi/2$, as k_{\parallel} tends to infinity all branches of the dispersion curve tend to the frequency of $\omega_p/\sqrt{2}$, which is the limiting frequency of a surface plasmon polariton at a planar vacuum–metal interface as $k_{\parallel} \rightarrow \infty$. This is because as k_{\parallel} tends to infinity and the wavelength of the surface plasmon polariton becomes smaller than any linear dimension characterizing the surface profile function, the surface plasmon polariton sees a surface that is locally planar, and all branches of its dispersion curve tend to the limiting frequency it has on that surface.

In Figs. 57a and b are plotted the dependence of the width of the gap and the frequency of the center of the gap on θ , respectively. These plots confirm the trends present in Figs. 56a–d. The structure on the former curve for values of θ greater than approximately 67° is caused by the repulsion of the second and

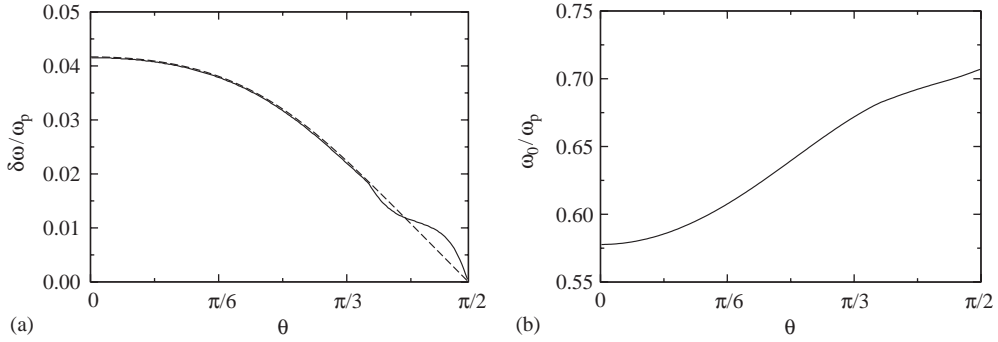


Fig. 57. (a) The width of the gap between the two lowest frequency branches at the boundary of the first Brillouin zone for surface plasmon polaritons propagating across the symmetric sawtooth grating defined by Eq. (5.11) with $H = 0.005 \mu\text{m}$ and $a = 0.2 \mu\text{m}$, as a function of θ . (b) The frequency of the center of this gap as a function of θ . (Ref. [188].)

third branches of the dispersion relation which, as functions of k_{\parallel} , approach each other and would cross inside the first Brillouin zone for values of θ increasing from 67° were it not for off-diagonal elements of the coefficient matrix in Eq. (5.10) that lift this degeneracy [188]. The width of the gap for θ in this angular range is then determined by the separation between the lowest frequency branch and the branch that was the third branch for θ smaller than 67° but becomes the second branch for θ greater than 67° . The crossover from the second to the third branch as defining the high frequency edge of the gap that occurs for $\theta \gtrsim 67^{\circ}$ gives rise to the structure present in the curve plotted in Fig. 57a.

Thus, the present approach allows the dispersion curves of surface plasmon polaritons on a classical metallic grating to be calculated nonperturbatively for any angle between the wave vector of the surface wave and the generators of the surface.

5.1.2. Two-dimensional periodic surface profile

A convenient starting point for the determination of the dispersion relation for surface plasmon polaritons on a two-dimensional periodic metal surface is the reduced Rayleigh equation (5.2). Let us now assume that the surface profile function $\zeta(\mathbf{x}_{\parallel})$ is a doubly periodic function of \mathbf{x}_{\parallel} ,

$$\zeta(\mathbf{x}_{\parallel} + \mathbf{x}_{\parallel}(\ell)) = \zeta(\mathbf{x}_{\parallel}), \quad (5.14)$$

where

$$\mathbf{x}_{\parallel}(\ell) = \ell_1 \mathbf{a}_1 + \ell_2 \mathbf{a}_2. \quad (5.15)$$

In Eq. (5.15) \mathbf{a}_1 and \mathbf{a}_2 are two noncollinear primitive translation vectors of a two-dimensional Bravais lattice, while ℓ_1 and ℓ_2 are positive and negative integers and zero, which we denote collectively by ℓ . The area of the primitive unit cell of the two-dimensional lattice defined by Eq. (5.15) is $a_c = |\mathbf{a}_1 \times \mathbf{a}_2|$.

We also require the lattice reciprocal to the one defined by Eq. (5.15), whose sites are given by

$$\mathbf{G}_{\parallel}(h) = h_1 \mathbf{b}_1 + h_2 \mathbf{b}_2, \quad (5.16)$$

where the primitive translation vectors \mathbf{b}_1 and \mathbf{b}_2 satisfy the conditions

$$\mathbf{a}_i \cdot \mathbf{b}_j = 2\pi\delta_{ij} , \tag{5.17}$$

and h_1 and h_2 are positive and negative integers and zero, which we denote collectively by h .

For a surface of the type defined by Eq. (5.14) the function $K(\gamma|\mathbf{Q}_{\parallel})$ becomes

$$K(\gamma|\mathbf{Q}_{\parallel}) = (2\pi)^2 \sum_{\mathbf{G}_{\parallel}} \delta(\mathbf{Q}_{\parallel} - \mathbf{G}_{\parallel}) \frac{1}{a_c} \int_{a_c} d^2x_{\parallel} \exp(-i\mathbf{G}_{\parallel} \cdot \mathbf{x}_{\parallel}) \exp(\gamma\zeta(\mathbf{x}_{\parallel})) . \tag{5.18}$$

In obtaining this expression we have used the result

$$\sum_{\ell} \exp[-i(\mathbf{q}_{\parallel} - \mathbf{k}_{\parallel}) \cdot \mathbf{x}_{\parallel}(\ell)] = \frac{(2\pi)^2}{a_c} \sum_{\mathbf{G}_{\parallel}} \delta(\mathbf{q}_{\parallel} - \mathbf{k}_{\parallel} - \mathbf{G}_{\parallel}) . \tag{5.19}$$

We also introduce the expansions

$$A_{p,s}(\mathbf{q}_{\parallel}) = (2\pi)^2 \sum_{\mathbf{G}_{\parallel}} \delta(\mathbf{q}_{\parallel} - \mathbf{k}_{\parallel} - \mathbf{G}_{\parallel}) a_{\mathbf{k}_{\parallel}}^{(p,s)}(\mathbf{G}_{\parallel}) , \tag{5.20}$$

where \mathbf{k}_{\parallel} is the two-dimensional wave vector of the surface plasmon polariton, in order that the electric field in the vacuum region satisfy the Bloch–Floquet theorem. When the results given by Eqs. (5.18) and (5.20) are substituted into Eq. (5.2), we obtain as the equation for the coefficients $\{a_{\mathbf{k}_{\parallel}}^{(p,s)}(\mathbf{G}_{\parallel})\}$ [185]

$$\sum_{\mathbf{G}'_{\parallel}} \frac{\tilde{K}(\beta(K_{\parallel}) - \beta_0(K'_{\parallel})|\mathbf{G}_{\parallel} - \mathbf{G}'_{\parallel})}{\beta(K_{\parallel}) - \beta_0(K'_{\parallel})} \times \begin{pmatrix} K_{\parallel}K'_{\parallel} - \beta(K_{\parallel})\hat{\mathbf{K}}_{\parallel} \cdot \hat{\mathbf{K}}'_{\parallel} \beta_0(K'_{\parallel}) & -i(\omega/c)\beta(K_{\parallel})(\hat{\mathbf{K}}_{\parallel} \times \hat{\mathbf{K}}'_{\parallel})_3 \\ i(\omega/c)(\hat{\mathbf{K}}_{\parallel} \times \hat{\mathbf{K}}'_{\parallel})_3 \beta_0(K'_{\parallel}) & (\omega/c)^2 \hat{\mathbf{K}}_{\parallel} \cdot \hat{\mathbf{K}}'_{\parallel} \end{pmatrix} \begin{pmatrix} a_{\mathbf{k}_{\parallel}}^{(p)}(\mathbf{G}'_{\parallel}) \\ a_{\mathbf{k}_{\parallel}}^{(s)}(\mathbf{G}'_{\parallel}) \end{pmatrix} = 0 , \tag{5.21}$$

where

$$\tilde{K}(\gamma|\mathbf{G}_{\parallel}) = \frac{1}{a_c} \int_{a_c} d^2x_{\parallel} \exp(-i\mathbf{G}_{\parallel} \cdot \mathbf{x}_{\parallel}) \exp(\gamma\zeta(\mathbf{x}_{\parallel})) \tag{5.22}$$

and, to simplify the notation, we have introduced the vectors $\mathbf{K}_{\parallel} = \mathbf{k}_{\parallel} + \mathbf{G}_{\parallel}$ and $\mathbf{K}'_{\parallel} = \mathbf{k}_{\parallel} + \mathbf{G}'_{\parallel}$ in writing Eq. (5.21). The dispersion relation for surface plasmon polaritons on a bigrating is then obtained by equating to zero the determinant of the matrix of the coefficients in Eq. (5.21).

The solutions of this equation are even functions of \mathbf{k}_{\parallel} , $\omega_s(-\mathbf{k}_{\parallel}) = \omega_s(\mathbf{k}_{\parallel})$, where s labels the solutions for a given \mathbf{k}_{\parallel} in the order of increasing magnitude. They are also periodic functions of \mathbf{k}_{\parallel} with a period that is the first Brillouin zone of the bigrating, $\omega_s(\mathbf{k}_{\parallel} + \mathbf{G}_{\parallel}) = \omega_s(\mathbf{k}_{\parallel})$. The solutions can then be sought for values of \mathbf{k}_{\parallel} inside the first Brillouin zone, and inside the nonradiative region defined by $|\mathbf{k}_{\parallel}| > (\omega/c)$.

In Fig. 58 we present dispersion curves along the symmetry directions in the first Brillouin zone for surface plasmon polaritons propagating across a triangular lattice of hemi-ellipsoids of revolution on the otherwise planar surface of a metal characterized by a simple, free electron dielectric function $\epsilon(\omega) = 1 - (\omega_p^2/\omega^2)$ [189]. The primitive translation vectors of the lattice are

$$\mathbf{a}_1 = (a, 0), \quad \mathbf{a}_2 = (a/2, \sqrt{3}a/2) , \tag{5.23}$$

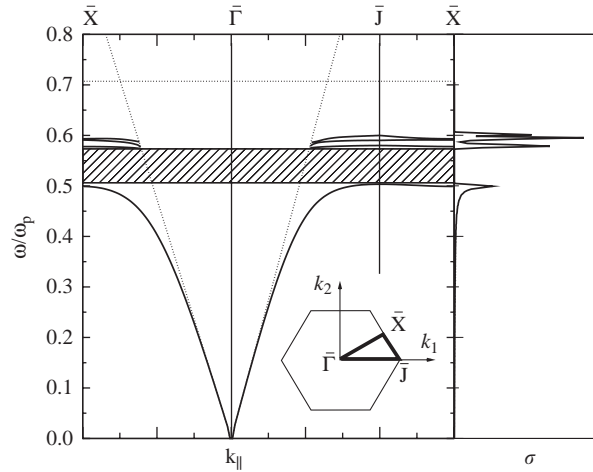


Fig. 58. The polaritonic band structure for a triangular lattice of hemi-ellipsoids of revolution, defined by Eqs. (5.14), (5.22), and (5.24), on the otherwise planar surface of silver, with $H = 0.4$, $R = 0.075 \mu\text{m}$ and $a = 0.2 \mu\text{m}$. Only the four lowest frequency branches are plotted. The inset depicts the two-dimensional first Brillouin zone for this structure, and the symmetry directions along which these branches are plotted. (Ref. [189].)

so that $a_c = \sqrt{3}a^2/2$, while

$$\mathbf{b}_1 = (2\pi/a)(1, -\sqrt{3}/3), \quad \mathbf{b}_2 = (2\pi/a)(0, 2\sqrt{3}/3) . \tag{5.24}$$

The surface profile function within the two-dimensional Wigner–Seitz cell of the triangular lattice is given by

$$\begin{aligned} \zeta(\mathbf{x}_{\parallel}) &= H(R^2 - x_{\parallel}^2)^{1/2}, \quad 0 \leq x_{\parallel} \leq R , \\ &= 0, \quad x_{\parallel} \geq R , \end{aligned} \tag{5.25}$$

where $R < a$. The function $\tilde{K}(\gamma|\mathbf{G}_{\parallel})$ in this case becomes

$$\tilde{K}(\gamma|\mathbf{G}_{\parallel}) = \frac{2\pi R^2}{a_c} \sum_{n=0}^{\infty} \frac{(\gamma H R)^n}{n!} \frac{2^{\frac{n}{2}} \Gamma(\frac{n}{2} + 1)}{(G_{\parallel} R)^{\frac{n}{2} + 1}} J_{\frac{n}{2} + 1}(G_{\parallel} R) . \tag{5.26}$$

When $\mathbf{G}_{\parallel} = 0$, we have that

$$\tilde{K}(\gamma|\mathbf{0}) = \frac{2\pi R^2}{a_c} \sum_{n=0}^{\infty} \frac{(\gamma H R)^n}{(n + 2)n!} . \tag{5.27}$$

The values of the parameters used in obtaining the results plotted in this figure were $H = 0.4$, $R = 0.075 \mu\text{m}$, and $a = 0.2 \mu\text{m}$. The 113 reciprocal lattice vectors closest to and including the origin were used in constructing the determinant whose zeros are the frequencies $\omega_s(\mathbf{k}_{\parallel})$ for each value of \mathbf{k}_{\parallel} . Only the four lowest frequency branches of the dispersion relation in the nonradiative region have been plotted. It is seen that an absolute band gap exists between the two lowest frequency branches (Fig. 58).

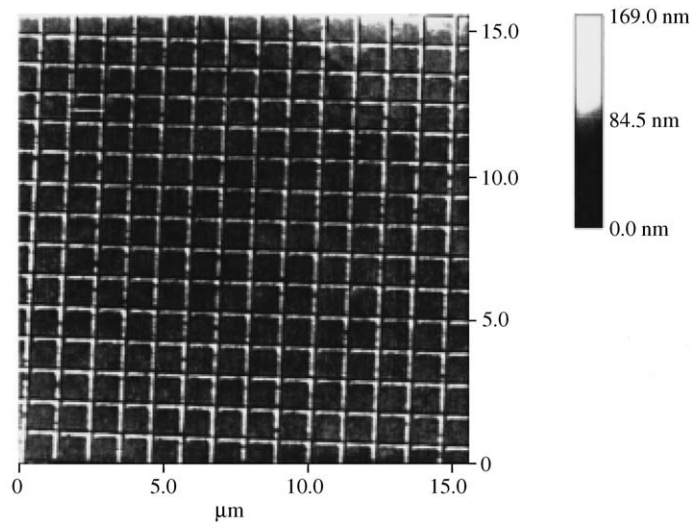


Fig. 59. Topography of the bigrating forming a surface plasmon polaritonic crystal measured with an atomic force microscope. (Ref. [190].)

Thus, the homogeneous form of the reduced Rayleigh equation for the electric field in the vacuum above a two-dimensional rough surface is an effective approach to the calculations of the dispersion relation for surface plasmon polaritons on a metallic bigrating.

5.2. Imaging of the SPP Bloch waves

Near-field optical images of the electromagnetic mode structure in two-dimensional SPP band-gap structures have recently been observed experimentally [190]. Interference patterns of the SPP Bloch waves in polaritonic crystals and the electromagnetic field formation close to the defects of the periodic structure have been studied. The imaging of SPP Bloch waves has been performed using the surface polaritonic crystal formed by a gold coated two-dimensional rectangular grating with a period of about $1\ \mu\text{m}$ (Fig. 59). The structure has been illuminated with a collimated laser beam that may be idealized as an excitation by a delta function in momentum space. The SNOM working in the transmission mode has been used to measure directly the local electromagnetic mode structure on the surface of a SPP crystal in coordinate space.

The angular dependence of the near-field optical signal measured at around one of the resonant angles (Eq. (1.32)) clearly reveals the resonant conditions at which the surface polariton Bloch waves are excited (Fig. 60). The images of the intensity distribution related to the SPP Bloch waves over the surface are shown in Fig. 61. The structure of the optical images is much more complex than the one of the topographical image as seen from the Fourier analysis (Fig. 61d). The observed Fourier spectral peaks correspond to different periodicities of the SPP interference pattern from the lattice period to larger periods.

The intensities of different spectral peaks of the Fourier spectrum contain information on the parameters of the SPP interaction with the surface features forming a polaritonic crystal. The Bloch theorem can be used to extract this information. The physical value measured in the experiment is the local electromagnetic field intensity over the surface of the SPP crystal. If a SPP Bloch wave propagates in the crystal along,

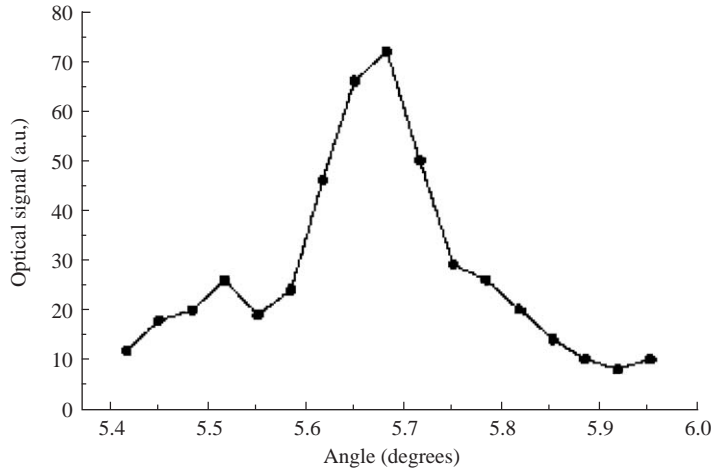


Fig. 60. Angular dependence of near-field optical signal near one of the SPP Bloch wave excitation resonances. (Ref. [190].)

e.g., the x_1 -direction ($p \neq 0, q = 0$), its field distribution can be written as

$$E = \sum_n A_n e^{i(k_{sp}^0 x_1 - \omega t)} e^{i2\pi n x_1 / D}, \quad (5.28)$$

where D is the period of the crystal lattice in the direction of propagation, n is an integer, and k_{sp}^0 is the SPP wavevector in the irreducible Brillouin zone. Thus, the intensity distribution imaged with a SNOM is

$$I = 1/2 \sum_{mn} A_m A_n \cos(2\pi(n - m)x_1 / D). \quad (5.29)$$

Thus the observed interference pattern is related to the SPP Bloch modes excited on the bigrating. Measurements of the intensities of spectral peaks at spatial frequencies $2\pi n / D$ in the Fourier spectra can allow recovering the coefficients A_n of the SPP Bloch waves which are related to the efficiency of their interaction with periodic surface features. For example, the Fourier spectrum of the image Fig. 61c shows that the brightest spectral peak corresponds to a spatial period of $3.5 \mu\text{m}$ (Fig. 61d). This value is approximately equal to half of the largest wavelength of the SPP Bloch wave propagating on the surface of the polaritonic crystal ($\lambda_{sp} = 7.26 \mu\text{m}$) corresponding to $(p, q) = (-2, 0)$ or $(p, q) = (0, -2)$ according to Eq. (1.32).

Although the surface quality of the polaritonic crystal under consideration is very good (Fig. 59), nevertheless some surface defects can be seen in the topography. In addition to the periodic surface features, which contribute to the SPP scattering, this results in the appearance of complex interference patterns in the vicinity of defects of the periodic structure. To enhance these effects and investigate the field distribution close to surface defects, artificial defects have been created in the SPP crystal (Fig. 62). A shadow-like scattering (cf. Section 2.1) is seen just behind the defect. A set of interference fringes with a period of about $3 \mu\text{m}$ related to the SPP Bloch wave scattering on this defect are clearly visible. Thus, there is a possibility to use such a structure for studying 2D cavity electrodynamic effects related to defects of the surface periodic structure, or to use appropriately shaped and sized defects for guiding SPP on a surface.

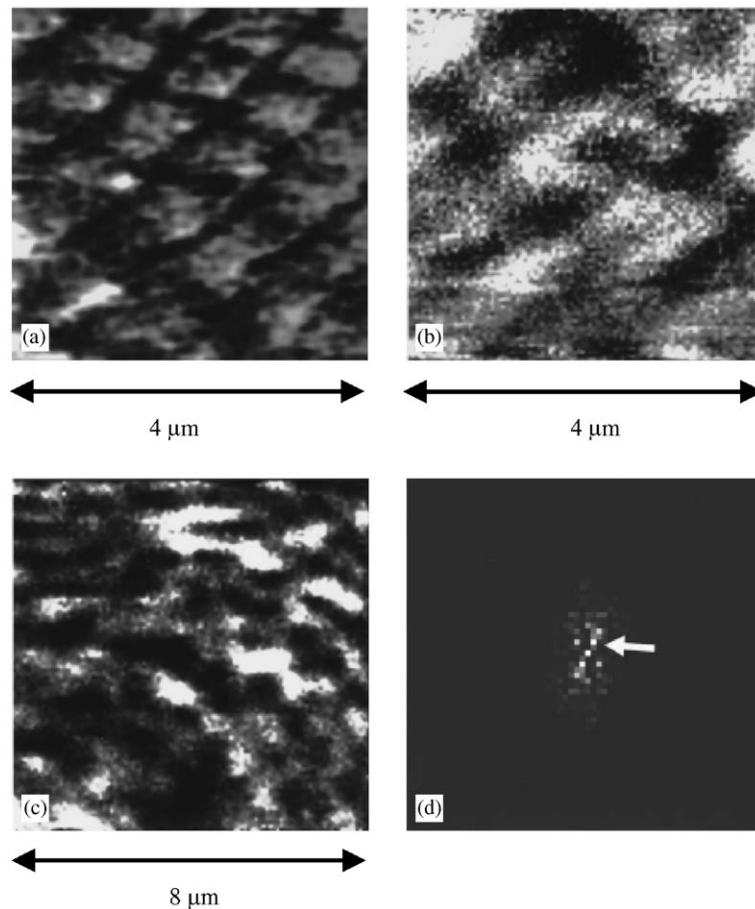


Fig. 61. Topography (a) and optical near-field distributions (b,c) over the SPP crystal measured under the resonant excitation of the SPP Bloch wave. (d) Fourier spectrum of the image (c). The spectral peak corresponding to the spatial period of about $3.5 \mu\text{m}$ is shown by the arrow. (Ref. [190].)

5.3. Waveguiding in line-defects of a SPP crystal

Due to the different dispersions of SPPs on a smooth surface and on a SPP crystal, the SPP wave propagating on the smooth part of a surface will not propagate through the SPP crystal if its frequency falls into a SPP crystal band-gap. This effect is the basis for SPP waveguiding in line-defects of SPP crystals. If an SPP is excited in a defect (or is coupled to a defect from a smooth surface) it propagates in the direction of the defect, since propagation outside the defect (in the SPP crystal) is prohibited. Thus, efficient waveguiding can be realized along straight and bent line defects as well as splitting and combining SPP waves in Y-shaped defects in SPP crystals, in analogy to two-dimensional photonic crystals based on dielectric nanostructures. Waveguiding in the defects of planar dielectric structures suffers from significant losses due to out-of-plane scattering. The scattering of SPP into light during guiding in SPP crystals is also present. Nevertheless, the intrinsic binding of the SPP wave to the surface

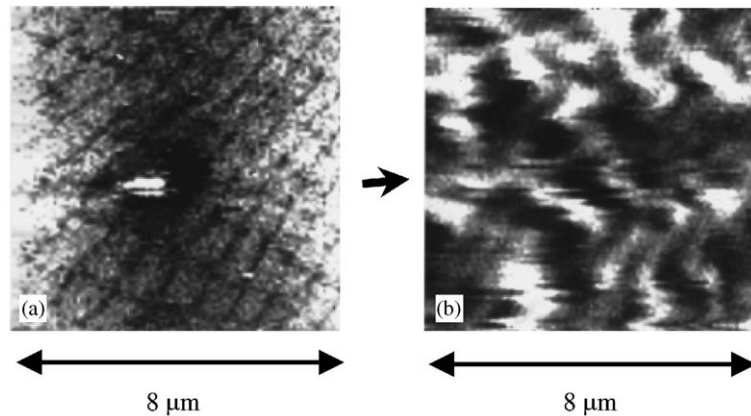


Fig. 62. Topography (a) and optical near-field distribution (b) around an artificially created surface defect in the SPP crystal (cf. Fig. 48). The direction of the SPP propagation is shown by the arrow. (Ref. [190].)

might provide lower bend-losses than 2D dielectric photonic band-gap structures [166,191]. In contrast to photonic crystals, using properties of angular and polarization dependencies of the SPP excitation in one or another direction on the surface, the waveguiding modes propagating in different directions of the line-defects can be excited by direct illumination of a SPP crystal [192].

The important property of photonic crystals is the so-called superprism effect, which relies on the ability of a photonic crystal of finite length to direct the transmitted light in drastically different angular directions depending on small changes of the light wavelength [193–195]. This effect arises from the very strong photon dispersion $d\mathbf{k}/d\omega$ near the edges of the photonic Brillouin zone, and thus can be artificially engineered for a given wavelength and required propagation direction. Even negative refraction has been recently predicted in appropriately shaped Brillouin zones in photonic crystals [196]. In SPP crystals, a conventional superprism effect (wavelength dependent direction of SPP Bloch wave propagation) can also be easily realized [24] with very high dispersion, since in polaritonic crystals it is possible to engineer flat SPP bands by choosing appropriate geometrical parameters of the surface features [197]. Negative refraction is also present in SPP crystals since some branches provide $dk_{sp}/d\omega < 0$ (Fig. 55).

Due to a strong dependence of SPP Bloch waves excitation on the polarization of the illuminating light (Eq. 1.32), an effect analogous to a (wavelength dependent) superprism effect can be considered in the polarization domain [192]. It manifests itself in a polarization dependence of the SPP mode excitation in a waveguide formed by line-defects in SPP crystals.

On a periodically structured surface, the SPP excitation takes place when the wavevector of diffracted light matches the wavevector (absolute value and direction) of the SPP Bloch mode of the SPP crystal. On the other hand, the wavelength at which the SPP is excited by the same periodic structure on a smooth surface always falls in the band-gap of the SPP crystal corresponding to the same direction of SPP propagation (Fig. 55). For this reason waveguiding of SPPs along line-defects in a periodically structured surfaces is possible [25].

SPP dispersion in a waveguide generally depends on the width and other geometrical details of the waveguide. Nevertheless, for the estimations of the excitation conditions a waveguiding mode can be

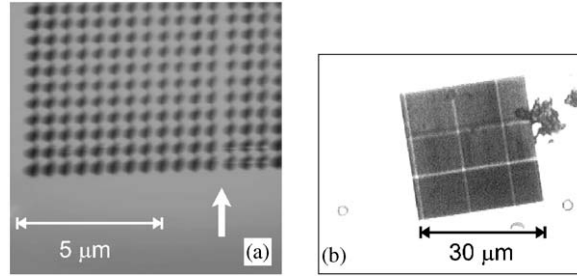


Fig. 63. Topography (a) and optical image (b) of the surface polaritonic band-gap structure with orthogonal line-defects. The square array of holes (periodicity $D = 600$ nm, diameter $d = 300$ nm) was fabricated using a focused-ion-beam microscope in a 50 nm gold film. The optical image is recorded in reflection using white light illumination. (Ref. [192].)

considered to be formed (in a simple ray-model picture) by SPP on a smooth surface. In addition to the vectorial equation (1.32), polarization selection rules should be satisfied since a surface polariton has both transverse and longitudinal electromagnetic field components and, therefore, the illuminating light should have the appropriate polarization to provide an electric field component perpendicular to a surface ($E_3 \sim \sin \psi \sin \theta \neq 0$) or parallel to the direction of SPP propagation ($\mathbf{E} \cdot \mathbf{k}_{\text{sp}} \neq 0$). Here, $E_1 \sim \sin \psi \cos \theta$ and $E_2 \sim \cos \psi \cos \theta$, where ψ is the angle between the magnetic field vector \mathbf{H} of the incident light and the plane of incidence: $\psi = \pi/2$ for p-polarized light and $\psi = 0$ for s-polarized light. SPP modes excited by a periodic structure can couple into a waveguide formed by a line-defect in the SPP crystal under the conditions providing dispersion matching in the direction of the defect orientation. These resonant conditions of the waveguiding mode excitation depend strongly on the polarization of the incident light. Thus, in addition to wavelength sensitivity, the light coupling into differently oriented SPP waveguides should exhibit a strong polarization dependence.

This effect of polarization sensitive coupling and routing of surface plasmon polaritons has been demonstrated in waveguides formed by orthogonal line-defects in SPP crystals (Fig. 63). It was observed under direct illumination of the nanostructured metal surface and does not require special illumination conditions for SPP excitation (such as illumination in total internal reflection).

For pure p-polarized light incident approximately along the holes rows (along the horizontal direction of the recorded images), e.g., $\mathbf{u}_{12} \approx \mathbf{u}_1$, an angle of incidence of about 40° is close to the angle of the SPP Bloch wave excitation on the nanostructure in the ($p = \pm 1, q = \pm 1$) direction (the Γ - M direction of the Brillouin zone). For this reason, the overall transmission of the nanostructured metal is high [24]. At the same time, the resonant excitation of SPPs on a smooth surface is achieved in the direction almost perpendicular to the plane of incidence (Eq. (1.32)). These resonant conditions are about the same as for s-polarized light, viz. ($p = 0, q = \pm 1$) corresponding to the Γ - X direction of the Brillouin zone. This direction corresponds to the orientation of the vertical line-defects in the structure. Thus, the image obtained with p-polarized light is bright with horizontal waveguides appearing to be dark (since there is no coupling to the waveguides in the horizontal direction) and hardly distinguishable vertical waveguides (there is some coupling in the vertical direction but the visibility is low because of the background transmission of the nanostructure). For pure s-polarized incident light, the overall transmission is smaller with some visibility of vertical waveguiding. The resonant excitation of the waveguiding mode in that direction at this wavelength occurs at an intermediate polarization angle ψ (Figs. 64 b–d). At some

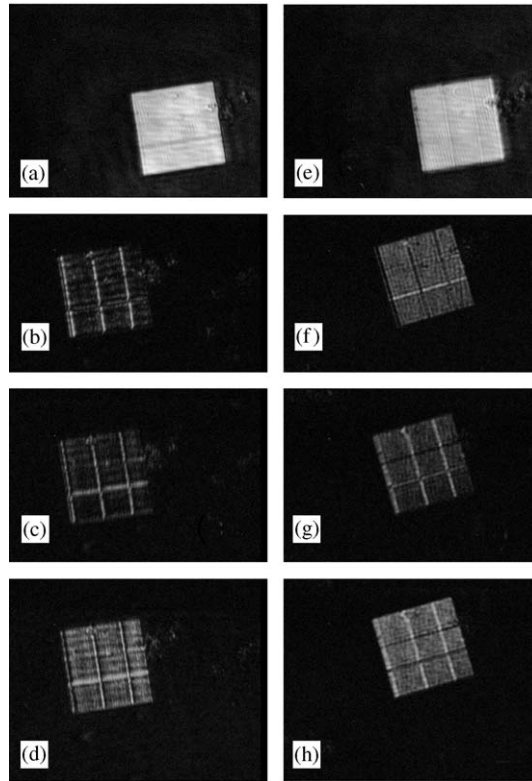


Fig. 64. The electromagnetic intensity distributions over the SPP crystal with line defects obtained at the angle of incidence of about (a–e) 40° and (f–h) 5° for different polarizations of the incident light: (a) p-polarization ($\psi = 90^\circ$), (b) $\psi = 32^\circ$, (c) $\psi = 30^\circ$, (d) $\psi = 26^\circ$, (e) s-polarization ($\psi = 0^\circ$) and (f) $\psi = 50^\circ$, (g) $\psi = 58^\circ$, and (h) $\psi = 66^\circ$. The optical contrast is internal for each image. (Ref. [192].)

polarization of the incident light, both waveguiding modes in mutually perpendicular directions can be simultaneously excited. However, the “horizontal” waveguiding mode cannot be excited alone at these conditions without the mode of the vertical line-defect (cf. Eq. (1.32)). The variation of the polarization angle required to switch off the horizontal mode is about 2° , showing a very high polarization sensitivity of the light coupling into the defect waveguiding modes.

To achieve complete switching between vertical and horizontal waveguides, the angle of incidence should be close to normal ($\theta \approx 0$) in order to suppress the SPP excitation in mutually perpendicular directions along holes rows for some polarizations of the illuminating light (Eq. (1.32) and the polarization selection rules). Close to normal incidence, only a very weak polarization dependence of the intensity of light transmitted through the hole array is observed, as should be expected. Under such conditions, the excitation of SPP modes which can couple to a line defect oriented vertically can be significantly suppressed at certain polarizations (Fig. 64f). In this case switching between horizontal and vertical waveguides, as well as the excitation of both waveguiding modes with about equal efficiency can be achieved. However, the range of polarization angles required for switching-off one of the waveguides is larger (about 4°) than in the case of oblique incidence discussed above. Directional

routing in mutually perpendicular directions can be achieved within a 10° interval of the polarization angles.

Moreover, it should be possible to achieve all-optical control of this wavelength- and polarization-dependent effect by employing nonlinear optical effects. This can be realized even at very low light intensities due to the electromagnetic field enhancement related to surface plasmons [36,167,168]. These effects can find numerous applications in surface plasmon polariton based photonics.

6. Enhanced optical properties mediated by surface plasmon polaritons

The cooperative effect of many features on a randomly rough surface or in a polaritonic crystal can influence profoundly not only in-plane propagation of SPPs on a nanostructured film but also conventional (3D) optical properties, such as transmission and reflection of light, as they will be determined by the surface polariton behavior on the film interfaces. The wavevector conservation needed to achieve SPP excitation by the incident light is provided by diffraction on surface roughness or a periodic structure. As all surfaces are rough to some extent, it is surface polaritons that determine many optical properties of rough and nanostructured surfaces and films. The widely known example of this is the so-called extraordinary enhanced optical transmission through a periodic hole array in a metal film [198], which in many cases can be explained by resonant light tunneling involving SPP states formed on one (or both) film interfaces due to the presence of a periodic structure [24,199–202].

6.1. *Surface plasmon polaritons and the enhanced backscattering of light from a randomly rough metal surface*

A striking manifestation of surface plasmon polaritons is found in the weak localization effects to which they give rise in the scattering of electromagnetic waves from, and in their transmission through, weakly rough random metal surfaces. These effects include enhanced backscattering [112], namely the presence of a well-defined peak in the retroreflection direction in the angular dependence of the intensity of the light scattered from such surfaces; enhanced transmission [203]—the presence of a well-defined peak in the antispecular direction in the angular dependence of the light transmitted through a thin metal film with a weakly rough illuminated surface; satellite peaks [204], which are subsidiary peaks on both sides of the enhanced backscattering and transmission peaks in the scattering or transmission of light from free-standing or supported metallic or dielectric films that support two or more surface or guided waves; and memory effect and reciprocal memory effect peaks in the envelopes of angular intensity correlation functions of light scattered from weakly rough metal and dielectric surfaces [205]. These effects are all multiple-scattering effects, and all have been observed and studied experimentally [206].

From among these effects the one that has been studied most intensively is enhanced backscattering, and we will focus our attention on it in this section. It arises in the following way [118]. The incident light strikes the rough surface of a metal and through the roughness excites a surface plasmon polariton. The surface plasmon polariton thus excited then propagates along the surface and is scattered from the roughness $s-1$ times before being scattered back into the vacuum as a volume electromagnetic wave propagating away from the surface at the s th scattering event. In general, all such s -order scattering sequences are incoherent due to the random phase introduced by the roughness of the surface. However, any given sequence has a reciprocal partner in which the light and the surface plasmon polariton are

scattered from the same points on the surface, but in the reverse order. The waves emerging into the vacuum after travelling along these two optical paths interfere constructively, and give a contribution to the intensity of the scattered light. When the wave vectors of the incident and scattered light are oppositely directed these two scattering sequences are coherent, and it is therefore necessary to add the amplitudes for them when calculating their contribution to the intensity of the scattered light. Thus, if A_d and A_r are the amplitudes of the direct and reciprocal scattering sequences, the contribution to the intensity of the scattered light from them is therefore $|A_d + A_r|^2$. As the scattering angle departs from the retroreflection direction, a random phase difference with nonzero mean and increasing variance develops between the two scattering paths. As a result, they are no longer coherent, so that only their intensities add, and their contribution to the intensity of the scattered light becomes $|A_d|^2 + |A_r|^2$. Thus, within a narrow angular range about the retroreflection direction, the intensity of the scattered light can become twice as large as the intensity of the background in this direction, when or if the amplitudes A_d and A_r are equal, due to the cross terms in $|A_r + A_d|^2$. We will see below that in the scattering of light from a randomly rough surface the scattering potential is reciprocal, the amplitudes A_d and A_r are equal, and their interference is constructive, leading to the appearance of the enhanced backscattering peak.

The contribution of the single-scattering processes has to be subtracted in obtaining this factor of two enhancement because it is not subject to coherent backscattering.

The earliest calculations of the enhanced backscattering of light from a randomly rough surface were carried out for the scattering of p-polarized light from a weakly rough one-dimensional random metal surface [112]. They were carried out by many-body perturbation theory. Subsequently, they were also carried out by small-amplitude perturbation theory [118]. Similarly, the first calculations of the enhanced backscattering of light from a two-dimensional randomly weakly rough metal surface were carried out by many-body perturbation theory [207] and then subsequently by small-amplitude perturbation theory [208].

Although there are some interesting polarization effects occurring in the scattering of light from two-dimensional random surfaces, since in-plane and out-of-plane co- and cross-polarized scattering can occur in this case, the role played by the multiple scattering of surface plasmon polaritons in the formation of the enhanced backscattering peak is already clearly and most simply seen in the scattering of p-polarized light from a one-dimensional weakly rough random metal surface when the plane of incidence is perpendicular to the generators of the surface. It is this case that we consider here.

Thus, the physical system that we consider here consists of vacuum in the region $x_3 > \zeta(x_1)$ and a metal, characterized by an isotropic, complex, frequency-dependent dielectric function $\epsilon(\omega) = \epsilon_1(\omega) + i\epsilon_2(\omega)$, in the region $x_3 < \zeta(x_1)$. The surface profile function $\zeta(x_1)$ is assumed to be a single-valued function of x_1 that is differentiable and constitutes a zero-mean, stationary Gaussian random process defined by

$$\langle \zeta(x_1)\zeta(x'_1) \rangle = \delta^2 W(|x_1 - x'_1|) . \quad (6.1)$$

The angle brackets here and in all that follows denote an average over the ensemble of realizations of the surface profile function, and $\delta = \langle \zeta^2(x_1) \rangle^{1/2}$ is the rms height of the surface.

It is convenient to introduce the Fourier integral representation of the surface profile function,

$$\zeta(x_1) = \int_{-\infty}^{\infty} \frac{dk}{2\pi} \hat{\zeta}(k) \exp(ikx_1) . \quad (6.2)$$

The Fourier coefficient $\hat{\zeta}(k)$ also constitutes a zero mean Gaussian random process. It is defined by

$$\langle \hat{\zeta}(k)\hat{\zeta}(k') \rangle = 2\pi\delta(k+k')\delta^2 g(|k|) , \quad (6.3)$$

where $g(|k|)$, the power spectrum of the surface roughness, is defined in terms of the surface height autocorrelation function $W(|x_1|)$ by

$$g(|k|) = \int_{-\infty}^{\infty} dx_1 W(|x_1|) \exp(-ikx_1) . \quad (6.4)$$

Two forms of $W(|x_1|)$, and hence of $g(|k|)$, have been assumed in theoretical studies of the enhanced backscattering of light from one-dimensional random surfaces. They are the Gaussian form

$$W(|x_1|) = \exp(-x_1^2/a^2) \quad (6.5a)$$

$$g(|k|) = \sqrt{\pi}a \exp(-a^2k^2/4) , \quad (6.5b)$$

and the West–O'Donnell [113] form

$$W(|x_1|) = \frac{\sin k_{\max}x_1 - \sin k_{\min}x_1}{(k_{\max} - k_{\min})x_1} \quad (6.6a)$$

$$g(|k|) = \frac{\pi}{k_{\max} - k_{\min}} [\theta(k_{\max} - k)\theta(k - k_{\min}) + \theta(k_{\max} + k)\theta(-k - k_{\min})] , \quad (6.6b)$$

where $\theta(z)$ is the Heaviside unit step function. In the case of the latter power spectrum, if we define

$$k_{\min} = k_{\text{sp}}(\omega) - (\omega/c) \sin \theta_m , \quad k_{\max} = k_{\text{sp}}(\omega) + (\omega/c) \sin \theta_m , \quad (6.7)$$

where $k_{\text{sp}}(\omega) = (\omega/c) \text{Re}\{\epsilon(\omega)/[\epsilon(\omega) + 1]\}^{1/2}$ is the wave number of the surface plasmon polariton of frequency ω supported by a planar vacuum–metal interface, the significance of the angle θ_m is that if p-polarized light of frequency ω is incident on the metal surface from the vacuum side with an angle of incidence θ_0 (measured counterclockwise from the x_3 -axis) that lies in the interval $(-\theta_m, \theta_m)$, it can excite forward and backward propagating surface plasmon polaritons of wavenumber $k_{\text{sp}}(\omega)$ through the surface roughness. At the same time, the surface plasmon polaritons excited in this way can be converted back into volume electromagnetic waves in the vacuum, through their interaction with the surface roughness, if the scattering angle θ_s (measured clockwise from the x_3 -axis) lies in the interval $(-\theta_m, \theta_m)$. Thus, the use of the power spectrum (6.6b) strongly enhances multiple-scattering processes in which surface plasmon polaritons play the role of intermediate states.

The surface $x_3 = \zeta(x_1)$ is illuminated from the vacuum side by a p-polarized plane wave of frequency ω whose plane of incidence is the x_1x_3 -plane. It is only a wave of this polarization that can excite a surface plasmon polariton of this frequency. The single nonzero component of its magnetic vector is given by

$$H_2(x_1, x_3|\omega)_{\text{inc}} = \exp[ikx_1 - i\alpha_0(k)x_3] , \quad (6.8)$$

where a time-dependence of the form $\exp(-i\omega t)$ has been assumed but not indicate explicitly, and $\alpha_0(k) = [(\omega/c)^2 - k^2]^{1/2}$, with $\text{Re} \alpha_0(k) > 0$, $\text{Im} \alpha_0(k) > 0$. The scattered field in the region $x_3 > \zeta(x_1)_{\max}$ is then a superposition of outgoing waves

$$H_2(x_1, x_3|\omega)_{\text{sc}} = \int_{-\infty}^{\infty} \frac{dq}{2\pi} R(q|k) \exp[iqx_1 + i\alpha_0(q)x_3] . \quad (6.9)$$

The differential reflection coefficient $\partial R/\partial\theta_s$ is defined in such a way that $(\partial R/\partial\theta_s) d\theta_s$ is the fraction of the total time-averaged incident flux that is scattered into the angular interval $(\theta_s, \theta_s + d\theta_s)$. It is given in terms of the scattering amplitude $R(q|k)$ by [112]

$$\frac{\partial R}{\partial\theta_s} = \frac{1}{L_1} \frac{\omega}{2\pi c} \frac{\cos^2\theta_s}{\cos\theta_0} |R(q|k)|^2, \quad (6.10)$$

where L_1 is the length of the x_1 -axis covered by the random surface. The wavenumbers k and q in this expression are to be expressed in terms of the angles of incidence and scattering by

$$k = (\omega/c) \sin\theta_0, \quad q = (\omega/c) \sin\theta_s. \quad (6.11)$$

Since we are dealing with scattering from a random surface, it is the average of $\partial R/\partial\theta_s$ over the ensemble of realizations of the surface profile function that we must calculate. More specifically, since we are interested in the intensity of the scattered light in the vicinity of the retroreflection direction $\theta_s = -\theta_0$, it is the contribution to the ensemble average of $\partial R/\partial\theta_s$ from the light that has been scattered incoherently (diffusely) that we wish to calculate. It is given by [112]

$$\left\langle \frac{\partial R}{\partial\theta_s} \right\rangle_{\text{incoh}} = \frac{1}{L_1} \frac{\omega}{2\pi c} \frac{\cos^2\theta_s}{\cos\theta_0} [\langle |R(q|k)|^2 \rangle - \langle R(q|k) \rangle \langle R(q|k) \rangle]. \quad (6.12)$$

The scattering amplitude $R(q|k)$ satisfies the reduced Rayleigh equation [209]

$$\int_{-\infty}^{\infty} \frac{dq}{2\pi} M(p|q) R(q|k) = -N(p|k), \quad (6.13)$$

where

$$M(p|q) = \frac{I(\alpha(p) - \alpha_0(q)|p - q)}{\alpha(p) - \alpha_0(q)} [pq + \alpha(q)\alpha_0(q)], \quad (6.14a)$$

$$N(p|k) = \frac{I(\alpha(p) + \alpha_0(k)|p - k)}{\alpha(p) + \alpha_0(k)} [pk - \alpha(p)\alpha_0(k)], \quad (6.14b)$$

and $\alpha(k) = [\epsilon(\omega)(\omega/c)^2 - k^2]^{1/2}$ with $\text{Re } \alpha(k) > 0$, $\text{Im } \alpha(k) > 0$. The function $I(\gamma|Q)$ appearing in Eqs. (6.14) is defined by

$$I(\gamma|Q) = \int_{-\infty}^{\infty} dx_1 \exp(-iQx_1) \exp(-i\gamma\zeta(x_1)). \quad (6.15)$$

We seek the solution of Eq. (6.13) in the form

$$R(q|k) = 2\pi\delta(q - k)R_0(k) - 2iG_0(q)T(q|k)G_0(k)\alpha_0(k), \quad (6.16)$$

where $R_0(k)$ is the Fresnel coefficient for the scattering of p-polarized light, incident from the vacuum side, from a planar vacuum–metal interface,

$$R_0(k) = \frac{\epsilon(\omega)\alpha_0(k) - \alpha(k)}{\epsilon(\omega)\alpha_0(k) + \alpha(k)}, \quad (6.17)$$

and

$$G_0(k) = \frac{i\epsilon(\omega)}{\epsilon(\omega)\alpha_0(k) + \alpha(k)} \quad (6.18)$$

is a Green's function for a surface plasmon polariton at a planar vacuum–metal interface. The transition matrix $T(q|k)$ is required to satisfy the equation

$$T(q|k) = V(q|k) + \int_{-\infty}^{\infty} \frac{dp}{2\pi} T(q|p)G_0(p)V(p|k) . \quad (6.19)$$

Equations (6.13), (6.16), and (6.19) define the scattering potential $V(q|k)$ that appears in Eq. (6.19). From these equations we find that this potential is the solution of the equation [209]

$$\int_{-\infty}^{\infty} \frac{dq}{2\pi} [M(p|q) - N(p|q)] \frac{V(q|k)}{2i\alpha_0(q)} = \{M(p|k)[\epsilon(\omega)\alpha_0(k) - \alpha(k)] + N(p|k)[\epsilon(\omega)\alpha_0(k) + \alpha(k)]\} \frac{1}{2\epsilon(\omega)\alpha_0(k)} , \quad (6.20)$$

where $M(p|q)$ and $N(p|q)$ have been defined in Eqs. (6.14). It can be shown [209] that the solution of this equation is reciprocal, $V(q|k) = V(-k|-q)$.

It is convenient to introduce the Green's function for a surface plasmon polariton at a one-dimensional randomly rough vacuum–metal interface, $G(q|k)$, as a solution of the equation

$$G(q|k) = 2\pi\delta(q - k)G_0(k) + G_0(q) \int_{-\infty}^{\infty} \frac{dp}{2\pi} V(q|p)G(p|k) \quad (6.21a)$$

$$= 2\pi\delta(q - k)G_0(k) + G_0(q)T(q|k)G_0(k) . \quad (6.21b)$$

On combining Eqs. (6.16) and (6.21b), and using the definitions (6.17) and (6.18), we obtain the useful relation

$$R(q|k) = -2\pi\delta(q - k) - 2iG(q|k)\alpha_0(k) . \quad (6.22)$$

The result given by Eq. (6.22) enables us to rewrite Eq. (6.12) in the form

$$\left\langle \frac{\partial R}{\partial \theta_s} \right\rangle_{\text{incoh}} = \frac{1}{L_1} \frac{2}{\pi} \left(\frac{\omega}{c} \right)^2 \cos^2 \theta_s \cos \theta_0 [\langle |G(q|k)|^2 \rangle - \langle |G(q|k)| \rangle^2] . \quad (6.23)$$

This is a convenient expression for $\langle \partial R / \partial \theta_s \rangle_{\text{incoh}}$ because $\langle |G(q|k)|^2 \rangle - |\langle G(q|k) \rangle|^2$ can be calculated from the solution of the Bethe–Salpeter equation [116]

$$\begin{aligned} \langle G(q|k)G^*(q'|k') \rangle &= \langle G(q|k) \rangle \langle G^*(q'|k') \rangle \\ &+ \int_{-\infty}^{\infty} \frac{dr}{2\pi} \int_{-\infty}^{\infty} \frac{dr'}{2\pi} \int_{-\infty}^{\infty} \frac{ds}{2\pi} \int_{-\infty}^{\infty} \frac{ds'}{2\pi} \langle G(q|r) \rangle \langle G^*(q'|r') \rangle \\ &\times \langle \Gamma(r, r'|s, s') \rangle \langle G(s|k)G^*(s'|k') \rangle, \end{aligned} \quad (6.24)$$

where $\langle \Gamma(q, q'|k, k') \rangle$ is the irreducible vertex function. The averaged Green's function $\langle G(q|k) \rangle$ is diagonal in q and k ,

$$\langle G(q|k) \rangle = 2\pi\delta(q - k)G(k), \quad (6.25)$$

a result that follows from the stationarity of the surface profile function $\zeta(x_1)$. $G(k)$ is related to the unperturbed Green's function $G_0(k)$ by

$$G(k) = \frac{1}{G_0(k)^{-1} - M(k)}, \quad (6.26)$$

where $M(k)$ is a proper self-energy that is obtained from the pair of equations [116]

$$\langle M(q|k) \rangle = 2\pi\delta(q - k)M(k), \quad (6.27a)$$

$$M(q|k) = V(q|k) + \int_{-\infty}^{\infty} \frac{dp}{2\pi} \int_{-\infty}^{\infty} \frac{dr}{2\pi} M(q|p) \langle G(p|r) \rangle [V(r|k) - \langle M(r|k) \rangle]. \quad (6.27b)$$

If we make use of Eq. (6.25), Eq. (6.24) becomes

$$\begin{aligned} \langle G(q|k)G^*(q'|k') \rangle &= 2\pi\delta(q - k)2\pi\delta(q' - k')G(q)G^*(q') + G(q)G^*(q') \int_{-\infty}^{\infty} \frac{ds}{2\pi} \int_{-\infty}^{\infty} \frac{ds'}{2\pi} \\ &\times \langle \Gamma(q, q'|s, s') \rangle \langle G(s|k)G^*(s'|k') \rangle. \end{aligned} \quad (6.28)$$

We now set $q = q'$, and obtain

$$\begin{aligned} \langle G(q|k)G^*(q|k') \rangle &= 2\pi\delta(q - k)2\pi\delta(q - k')|G(q)|^2 \\ &+ |G(q)|^2 \int_{-\infty}^{\infty} \frac{ds}{2\pi} \int_{-\infty}^{\infty} \frac{ds'}{2\pi} \langle \Gamma(q, q|s, s') \rangle \langle G(s|k)G^*(s'|k') \rangle. \end{aligned} \quad (6.29)$$

However, due to the stationarity of the surface profile function, $\langle \Gamma(q, q|s, s') \rangle$ must be diagonal in s and s' ,

$$\langle \Gamma(q, q|s, s') \rangle = 2\pi\delta(s - s')U(q|s). \quad (6.30)$$

Consequently, on setting $k = k'$, Eq. (6.29) becomes

$$\langle |G(q|k)|^2 \rangle = L_1 2\pi\delta(q - k)|G(q)|^2 + |G(q)|^2 \int_{-\infty}^{\infty} \frac{ds}{2\pi} U(q|s) \langle |G(s|k)|^2 \rangle, \quad (6.31)$$

where we have used the result that $[2\pi\delta(q - k)]^2 = L_1 2\pi\delta(q - k)$. The solution of this equation can be written formally as

$$\langle |G(q|k)|^2 \rangle = L_1 2\pi\delta(q - k)|G(q)|^2 + L_1 |G(q)|^2 R(q|k) |G(k)|^2, \quad (6.32)$$

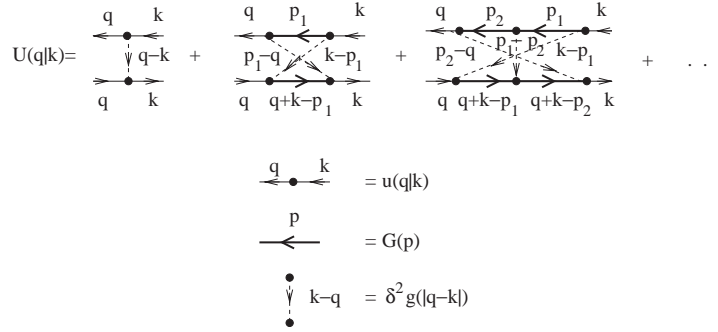


Fig. 65. The first three maximally crossed diagrams that contribute to the irreducible vertex function $U(q|k)$, together with the analytic equivalents of the components of these diagrams.

where $R(q|k)$ is the reducible vertex function. It is related to the irreducible vertex function $U(q|k)$ by

$$R(q|k) = U(q|k) + \int_{-\infty}^{\infty} \frac{ds}{2\pi} U(q|s) |G(s)|^2 R(s|k) . \tag{6.33}$$

On combining Eqs. (6.23) and (6.32) we find that the mean differential reflection coefficient $\langle \partial R / \partial \theta_s \rangle_{\text{incoh}}$ is given by

$$\left\langle \frac{\partial R}{\partial \theta_s} \right\rangle_{\text{incoh}} = \frac{2}{\pi} \left(\frac{\omega}{c} \right)^3 \cos^2 \theta_s \cos \theta_0 |G(q)|^2 R(q|k) |G(k)|^2 , \tag{6.34}$$

with $q = (\omega/c) \sin \theta_s$, $k = (\omega/c) \sin \theta_0$.

We approximate the irreducible vertex function $U(q|k)$ by the sum of the contributions from all maximally-crossed diagrams because, as we shall see, they describe the phase coherent multiple-scattering processes that give rise to enhanced backscattering. In calculating these contributions we make the small roughness approximation, which consists of approximating the scattering potential $V(q|k)$ by the solution of Eq. (6.20) that is of first order in the surface profile function $\zeta(x_1)$,

$$V(q|k) \cong u(q|k) \hat{\zeta}(q - k) , \tag{6.35}$$

where

$$\begin{aligned} u(q|k) &= \frac{\epsilon(\omega) - 1}{\epsilon^2(\omega)} [\epsilon(\omega) q k - \alpha(q) \alpha(k)] \\ &= u(k|q) = u(-q| -k) = u(-k| -q) . \end{aligned} \tag{6.36}$$

The results we obtain are therefore limited to weakly rough surfaces.

For what follows it is convenient, although not necessary, to neglect the small imaginary part of $\epsilon(\omega)$ in calculating $u(q|k)$, so that $u(q|k)$ becomes real. (For example, for silver at a wavelength $\lambda = 457.9$ nm, $\epsilon(\omega) = -7.5 + i0.24$ [210], so only a small quantitative error is made as a result of this neglect.)

The first three maximally-crossed diagrams contributing to $U(q|k)$ are depicted in Fig. 65, together with the analytic equivalents of the components constituting these diagrams. It should be noted that reversing the direction of the arrow on a heavy black line produces the complex conjugate of the corresponding averaged Green's function $G(p)$. The incoming and outgoing lines labeled by k and q do not give a

contribution to $U(q|k)$, but serve only to help fix the arguments of the vertex functions $\{u(q|k)\}$ that enter the contributions to $U(q|k)$ represented diagrammatically in Fig. 65. Due to the stationarity of the surface profile function, the sum of the wavenumbers entering a vertex must equal the sum of the wave numbers leaving that vertex. Finally, the variables p_1, p_2, \dots , are integrated with volume elements $dp_1/2\pi, dp_2/2\pi, \dots$.

By following these rules we obtain the following expression for $U(q|k)$:

$$\begin{aligned}
U(q|k) = & [u(q|k)]^2 \delta^2 g(|q-k|) + \int_{-\infty}^{\infty} \frac{dp_1}{2\pi} u(p_1|k) u(q|p_1) u(q|Q-p_1) u(Q-p_1|k) \\
& \times G(p_1) G^*(Q-p_1) \delta^2 g(|k-p_1|) \delta^2 g(|p_1-q|) \\
& + \int_{-\infty}^{\infty} \frac{dp_1}{2\pi} \int_{-\infty}^{\infty} \frac{dp_2}{2\pi} u(p_1|k) u(p_2|p_1) u(q|p_2) \\
& \times u(q|Q-p_1) u(Q-p_1|Q-p_2) u(Q-p_2|k) \\
& \times G(p_1) G^*(Q-p_1) G(p_2) G^*(Q-p_2) \delta^2 g(|k-p_1|) \delta^2 g(|p_1-p_2|) \delta^2 g(|p_2-q|) \\
& + \int_{-\infty}^{\infty} \frac{dp_1}{2\pi} \int_{-\infty}^{\infty} \frac{dp_2}{2\pi} \int_{-\infty}^{\infty} \frac{dp_3}{2\pi} u(p_1|k) u(p_2|p_1) u(p_3|p_2) u(q|p_3) u(q|Q-p_1) \\
& \times u(Q-p_1|Q-p_2) u(Q-p_2|Q-p_3) u(Q-p_3|k) \\
& \times G(p_1) G^*(Q-p_1) G(p_2) G^*(Q-p_2) \\
& \times G(p_3) G^*(Q-p_3) \delta^2 g(|k-p_1|) \delta^2 g(|p_1-p_2|) \delta^2 g(|p_2-p_3|) \\
& \times \delta^2 g(|p_3-q|) + \dots, \tag{6.37}
\end{aligned}$$

where, to simplify the notation, we have defined $Q \equiv q+k$. Because we are particularly interested in the phenomenon of enhanced backscattering, which occurs when $q = -k$, we can simplify the expansion (6.37) by setting $Q = 0$ everywhere that this does not give rise to a singularity, namely in the vertex functions $\{u(q|k)\}$. With this simplification Eq. (6.37) can be rearranged into the following expression:

$$U(q|k) = [u(q|k)]^2 \delta^2 g(|q-k|) + A(q|k), \tag{6.38}$$

where

$$A(q|k) = \int_{-\infty}^{\infty} \frac{dp_1}{2\pi} \int_{-\infty}^{\infty} \frac{dp_2}{2\pi} u(q|-p_1) W(q|p_2) K(p_2) t(p_2|p_1) W(p_1|k) u(-p_2|k), \tag{6.39}$$

and we have introduced the definitions

$$W(q|k) \equiv u(q|k) \delta^2 g(|q-k|) \tag{6.40a}$$

$$K(q) \equiv G(q) G^*(Q-q). \tag{6.40b}$$

The function $t(p_2|p_1)$ entering Eq. (6.39) is the solution of the integral equation

$$t(p_2|p_1) = 2\pi \delta(p_2-p_1) + \int_{-\infty}^{\infty} \frac{dp_3}{2\pi} X(p_2|p_3) K(p_3) t(p_3|p_1), \tag{6.41}$$

where

$$X(q|k) = [u(q|k)]^2 \delta^2 g(|q-k|). \tag{6.42}$$

To solve Eq. (6.41) we will use the pole approximation for the averaged Green's function $G(p)$. To obtain this approximation we first note that the unperturbed Green's function $G_0(p)$ given by Eq. (6.18) can be written in the form

$$\begin{aligned} G_0(p) &= \frac{i\epsilon(\omega)}{1 - \epsilon^2(\omega)} \frac{\epsilon(\omega)\alpha_0(p) - \alpha(p)}{p^2 - \frac{\omega^2}{c^2} \frac{\epsilon(\omega)}{\epsilon(\omega) + 1}} \\ &\cong \frac{C(\omega)}{p - k_{\text{sp}}(\omega) - i\Delta_\epsilon(\omega)} - \frac{C(\omega)}{p + k_{\text{sp}}(\omega) + i\Delta_\epsilon(\omega)}, \end{aligned} \quad (6.43)$$

where to lowest order in $\epsilon_2(\omega)$

$$k_{\text{sp}}(\omega) = \frac{\omega}{c} \left(\frac{|\epsilon_1(\omega)|}{|\epsilon_1(\omega)| - 1} \right)^{1/2} \quad (6.44)$$

is the wavenumber of a surface plasmon polariton of frequency ω at a planar vacuum–metal interface, while

$$\Delta_\epsilon(\omega) = \frac{1}{2} \frac{\omega}{c} \frac{\epsilon_2(\omega)}{|\epsilon_1(\omega)|^{1/2} (|\epsilon_1(\omega)| - 1)^{3/2}} \quad (6.45)$$

is the decay rate of the surface plasmon polariton due to ohmic losses as it propagates along the interface. The coefficient $C(\omega)$ is the residue of $G_0(p)$ at the poles it possesses at $p = \pm[k_{\text{sp}}(\omega) + i\Delta_\epsilon(\omega)]$,

$$C(\omega) = \frac{|\epsilon_1(\omega)|^{3/2}}{\epsilon_1^2(\omega) - 1}. \quad (6.46)$$

The pole approximation to $G(p)$ is now obtained quite simply. From Eqs. (6.26) and (6.43) we find that

$$G(p) \cong \frac{C(\omega)}{p - k_{\text{sp}}(\omega) - i\Delta(\omega)} - \frac{C(\omega)}{p + k_{\text{sp}}(\omega) + i\Delta(\omega)}, \quad (6.47)$$

where we have defined the total damping rate of the surface plasmon polariton $\Delta(\omega)$ by

$$\Delta(\omega) = \Delta_\epsilon(\omega) + \Delta_{\text{sp}}(\omega), \quad (6.48)$$

with

$$\Delta_{\text{sp}}(\omega) = C(\omega) \text{Im} M(k_{\text{sp}}(\omega)). \quad (6.49)$$

The function $\Delta_{\text{sp}}(\omega)$ is the decay rate of the surface plasmon polariton due to its roughness-induced conversion into volume electromagnetic waves in the vacuum above the metal surface and into other surface plasmon polaritons. In obtaining Eq. (6.47) we have neglected the small renormalization of the wavenumber $k_{\text{sp}}(\omega)$ arising from the real part of the proper self-energy $M(k_{\text{sp}}(\omega))$, but have taken into account that $\Delta_{\text{sp}}(\omega)$ can be comparable to, or larger than, $\Delta_\epsilon(\omega)$. The small renormalization of the residue $C(\omega)$ due to the surface roughness has also been neglected.

In the small roughness approximation, the decay rate $\Delta_{\text{sp}}(\omega)$ is given by (see Eqs. (6.49) and (6.27))

$$\begin{aligned}
 \Delta_{\text{sp}}(\omega) &= C(\omega) \text{Im} \frac{1}{L_1} \int_{-\infty}^{\infty} \frac{dp}{2\pi} [u(k_{\text{sp}}(\omega)|p)]^2 (\hat{\zeta}(k_{\text{sp}}(\omega) - p) \hat{\zeta}(p - k_{\text{sp}}(\omega))) G(p) \\
 &= C(\omega) \text{Im} \delta^2 \int_{-\infty}^{\infty} \frac{dp}{2\pi} [u(k_{\text{sp}}(\omega)|p)]^2 g(|k_{\text{sp}}(\omega) - p|) \\
 &\quad \times C(\omega) \left[\frac{1}{p - k_{\text{sp}}(\omega) - i\Delta(\omega)} - \frac{1}{p + k_{\text{sp}}(\omega) + i\Delta(\omega)} \right] \\
 &\cong \frac{1}{2} C^2(\omega) \delta^2 \int_{-\infty}^{\infty} dp [u(k_{\text{sp}}(\omega)|p)]^2 g(|k_{\text{sp}}(\omega) - p|) \\
 &\quad \times [\delta(p - k_{\text{sp}}(\omega)) + \delta(p + k_{\text{sp}}(\omega))] \\
 &= \frac{1}{2} C^2(\omega) [u(k_{\text{sp}}(\omega)| - k_{\text{sp}}(\omega))]^2 \delta^2 g(2k_{\text{sp}}(\omega)) .
 \end{aligned} \tag{6.50}$$

In obtaining this result we have used the result that

$$u(k_{\text{sp}}|k_{\text{sp}}) = u(-k_{\text{sp}}| - k_{\text{sp}}) = 0 , \tag{6.51}$$

in view of the dispersion relation for surface plasmon polaritons, $p^2 = (\omega/c)^2 \{\epsilon(\omega)/[\epsilon(\omega) + 1]\}$.

We can now proceed to the solution of Eq. (6.41). By the use of the result that

$$\begin{aligned}
 K(p) &= G(p)G^*(Q - p) = \frac{C^2}{2i\Delta - Q} \frac{2i\Delta}{(p - k_{\text{sp}})^2 + \Delta^2} + \frac{C^2}{2i\Delta + Q} \frac{2i\Delta}{(p + k_{\text{sp}})^2 + \Delta^2} \\
 &\cong \frac{2\pi i C^2}{2i\Delta - Q} \delta(p - k_{\text{sp}}) + \frac{2\pi i C^2}{2i\Delta + Q} \delta(p + k_{\text{sp}}) ,
 \end{aligned} \tag{6.52}$$

which is valid in the limit of small Δ , we can rewrite Eq. (6.41) in the form

$$\begin{aligned}
 t(p_2|p_1) &= 2\pi\delta(p_2 - p_1) + \frac{iC^2}{2i\Delta - Q} X(p_2|k_{\text{sp}})t(k_{\text{sp}}|p_1) \\
 &\quad + \frac{iC^2}{2i\Delta + Q} X(p_2| - k_{\text{sp}})t(-k_{\text{sp}}|p_1),
 \end{aligned} \tag{6.53}$$

where we have also used Eq. (6.51). By setting $p_2 = k_{\text{sp}}$ and $p_2 = -k_{\text{sp}}$ in Eq. (6.53) we obtain the results

$$t(k_{\text{sp}}|p_1) = \frac{Q^2 + 4\Delta^2}{Q^2 + 4(\Delta^2 - \Delta_{\text{sp}}^2)} \left\{ 2\pi\delta(p_1 - k_{\text{sp}}) + \frac{2i\Delta_{\text{sp}}}{2i\Delta + Q} 2\pi\delta(p_1 + k_{\text{sp}}) \right\} \tag{6.54a}$$

$$t(-k_{\text{sp}}|p_1) = \frac{Q^2 + 4\Delta^2}{Q^2 + 4(\Delta^2 - \Delta_{\text{sp}}^2)} \left\{ 2\pi\delta(p_1 + k_{\text{sp}}) + \frac{2i\Delta_{\text{sp}}}{2i\Delta - Q} 2\pi\delta(p_1 - k_{\text{sp}}) \right\} , \tag{6.54b}$$

where we have used the results (see Eqs. (6.42) and (6.50)) that

$$X(k_{\text{sp}}| - k_{\text{sp}}) = X(-k_{\text{sp}}|k_{\text{sp}}) = \frac{2\Delta_{\text{sp}}}{C^2} . \tag{6.55}$$

On substituting Eq. (6.53) into Eq. (6.39) we find that $A(q|k)$ becomes

$$A(q|k) = \int_{-\infty}^{\infty} \frac{dp_1}{2\pi} u(q|p_1) W(q|k_{\text{sp}}) \frac{iC^2}{2i\Delta - Q} t(k_{\text{sp}}|p_1) W(p_1|k) u(-k_{\text{sp}}|k) \\ + \int_{-\infty}^{\infty} \frac{dp_1}{2\pi} u(q|p_1) W(q|k_{\text{sp}}) \frac{iC^2}{2i\Delta + Q} t(-k_{\text{sp}}|p_1) W(p_1|k) u(k_{\text{sp}}|k) . \quad (6.56)$$

Substitution of Eqs. (6.54) into Eq. (6.56) leads, after a good deal of algebra, to the result

$$A(q|k) = \frac{A(q|k)}{Q^2 + 4(\Delta^2 - \Delta_{\text{sp}}^2)} \quad (6.57)$$

where

$$A(q|k) = 2C^2 \Delta [X(q|k_{\text{sp}}) X(k_{\text{sp}}|k) + X(q|k_{\text{sp}}) X(-k_{\text{sp}}|k)] \\ + 2C^2 \Delta_{\text{sp}} [X(q|k_{\text{sp}}) X(-k_{\text{sp}}|k) + X(q|k_{\text{sp}}) X(k_{\text{sp}}|k)] \\ = A(-k|q) . \quad (6.58)$$

Thus, our approximation to the irreducible vertex function $U(q|k)$ is finally

$$U(q|k) = X(q|k) + \frac{A(q|k)}{(q+k)^2 + 4\Gamma^2} , \quad (6.59)$$

where

$$\Gamma = [\Delta^2 - \Delta_{\text{sp}}^2]^{1/2} = [\Delta_{\epsilon}(\Delta_{\epsilon} + 2\Delta_{\text{sp}})]^{1/2} . \quad (6.60)$$

The result given by Eq. (6.59) has to be substituted into Eq. (6.33), which is then solved for $R(q|k)$ by iteration. However, in each of the integral terms of the iterative solution only the contribution associated with the first term on the right hand side of Eq. (6.59) is kept, and all contributions that contain the second term are omitted. The reason for this is that $|G(s)|^2$, according to Eq. (6.52) with $Q = 0$, is large only for $s \cong \pm k_{\text{sp}}$, i.e. for wave numbers outside the radiative region $(-\omega/c, \omega/c)$. On the other hand, $A(q|k)$ is large only for $q \cong -k$. Both q and k are required to be inside the radiative region in our calculation of $\langle \partial R / \partial \theta_s \rangle_{\text{incoh}}$, Eq. (6.34). Thus, the terms containing $A(q|k)$ in all the integral terms in the iterative solution of Eq. (6.33) give only a small correction to the result given by omitting them, either because they are off resonance ($q \neq -k$), or else are multiplied by higher powers of δ^2/a^2 (in the case of the Gaussian power spectrum (6.5b)) than the first when they are resonant.

In this way we obtain the result that

$$R(q|k) = \frac{A(q|k)}{(q+k)^2 + 4\Gamma^2} + X(q|k) + \int_{-\infty}^{\infty} \frac{ds_1}{2\pi} X(q|s_1) |G(s_1)|^2 X(s_1|k) \\ + \int_{-\infty}^{\infty} \frac{ds_1}{2\pi} \int_{-\infty}^{\infty} \frac{ds_2}{2\pi} X(q|s_1) |G(s_1)|^2 X(s_1|s_2) |G(s_2)|^2 X(s_2|k) + \dots \\ \equiv \frac{A(q|k)}{(q+k)^2 + 4\Gamma^2} + \sum (q|k) , \quad (6.61)$$

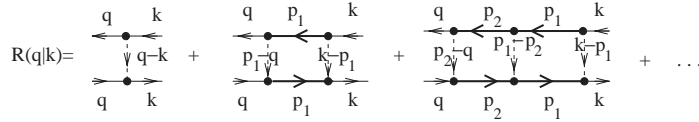


Fig. 66. The first three ladder diagrams that contribute to the reducible vertex function $R(q|k)$.

where $\sum(q|k)$ is the solution of the integral equation

$$\sum(q|k) = X(q|k) + \int_{-\infty}^{\infty} \frac{ds}{2\pi} X(q|s) |G(s)|^2 \sum(s|k) . \tag{6.62}$$

The function $\sum(q|k)$ is readily seen to be the sum of the contributions associated with the ladder diagrams depicted in Fig. 66.

The solution of Eq. (6.62) is carried out in exactly the same way as Eq. (6.41) was solved, with the aid of Eqs. (6.52), and (6.55). The result is

$$\sum(q|k) = X(q|k) + \frac{A(q|k)}{4\Gamma^2} . \tag{6.63}$$

Consequently, our result for the reducible vertex function $R(q|k)$ takes the form

$$R(q|k) = X(q|k) + \frac{A(q|k)}{4\Gamma^2} + \frac{A(q|k)}{(q+k)^2 + 4\Gamma^2} . \tag{6.64}$$

When we combine this result with Eq. (6.34) we obtain the result that the contribution to the mean differential reflection coefficient from the light that has been scattered incoherently is

$$\left\langle \frac{\partial R}{\partial \theta_s} \right\rangle_{\text{incoh}} = \frac{2}{\pi} \left(\frac{\omega}{c} \right)^3 \cos^2 \theta_s \cos \theta_0 |G(q)|^2 \left\{ X(q|k) + \frac{A(q|k)}{4\Gamma^2} + \frac{A(q|k)}{(q+k)^2 + 4\Gamma^2} \right\} |G(k)|^2 , \tag{6.65}$$

where $k = (\omega/c) \sin \theta_0$ and $q = (\omega/c) \sin \theta_s$.

The first term on the right hand side of Eq. (6.65) is the contribution from single-scattering processes; the second is the contribution from the ladder diagrams; and the third is the contribution from the maximally crossed diagrams. When $q = -k$, the latter two terms are equal for scattering into the retroreflection direction, i.e. when $\theta_s = -\theta_0$. This is the reason why the contribution from the ladder diagrams has been included in the result for $R(q|k)$. The conclusion that the result expressed by Eq. (6.65) predicts enhanced backscattering is clearly seen from the presence of the denominator $(q+k)^2 + 4\Gamma^2$ in the last term on the right hand side of Eq. (6.65), which therefore has a peak at $q = -k$, whose angular width at half maximum is $4c\Gamma/(\omega \cos \theta_0)$, and hence increases with increasing angle of incidence. This peak is the manifestation of enhanced backscattering. We see that if the contribution to $\langle \partial R / \partial \theta_s \rangle_{\text{incoh}}$ from single-scattering is subtracted off, the peak in the retroreflection direction is twice as high as the height of the background at its position. The origin of this peak in the contribution to $\langle \partial R / \partial \theta_s \rangle_{\text{incoh}}$ from the maximally crossed diagrams indicates that it is associated with the cross-terms in the squared modulus of the sum of the amplitudes of the direct and reciprocal scattering sequences arising in a calculation of the intensity of the scattered light.

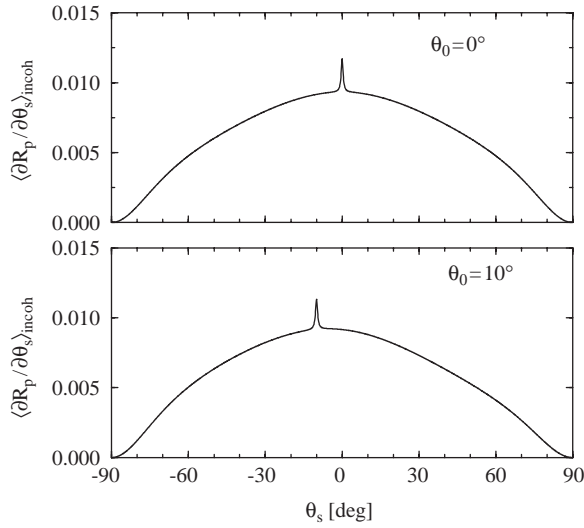


Fig. 67. The contribution to the mean differential reflection coefficient from the incoherent component of the scattered light when p-polarized light of wavelength $\lambda=457.9$ nm is incident on a one-dimensional randomly rough silver surface ($\epsilon(\omega)=-7.5+i0.24$) characterized by the Gaussian power spectrum (6.5b) with $\delta=5$ nm and $a=100$ nm. (a) $\theta_0=0^\circ$; (b) $\theta_0=10^\circ$. These results are calculated on the basis of Eq. (6.66).

We note that the result for $\langle \partial R / \partial \theta_s \rangle_{\text{incoh}}$ given by Eq. (6.65) differs in some details from the corresponding result in the original paper by McGurn et al. [112], but numerical results obtained by the use of Eq. (6.65) are in very good quantitative agreement with those obtained by the use of the expressions from that earlier work.

In Fig. 67 we present plots of $\langle \partial R / \partial \theta_s \rangle_{\text{incoh}}$ obtained on the basis of Eq. (6.65) for the scattering of p-polarized light of wavelength $\lambda=457.9$ nm from a one-dimensional, random, silver surface ($\epsilon(\omega)=-7.5+i0.24$), characterized by the Gaussian power spectrum (6.5b) with roughness parameters $\delta=5$ nm and $a=100$ nm. For each angle of incidence a well-defined peak is present in the retroreflection direction, whose width increases with increasing angle of incidence. Enhanced backscattering of light from a one-dimensional randomly rough metal surface characterized by a Gaussian power spectrum has yet to be observed experimentally.

The only experimental evidence for the enhanced backscattering of light from a weakly rough, one-dimensional, random surface has been provided by West and O'Donnell [113]. These authors fabricated a one-dimensional randomly rough gold surface that was characterized by the power spectrum given by Eq. (6.6b). The values of k_{min} and k_{max} were determined from Eqs. (6.7) with $\theta_{\text{max}}=13.5^\circ$, and $k_{\text{sp}}(\omega)$ the wavenumber of surface plasmon polaritons at the frequency ω corresponding to a wavelength $\lambda=612$ nm ($\epsilon(\omega)=-9.00+i1.29$ [211]). It followed that $k_{\text{min}}=8.48 \times 10^{-3} \text{ nm}^{-1}$ and $k_{\text{max}}=13.26 \times 10^{-3} \text{ nm}^{-1}$. The rms height of the surface was $\delta=10.9$ nm.

In Fig. 68 we have plotted the experimental data of West and O'Donnell for the contribution to the mean differential reflection coefficient from the incoherent component of the scattered light, for three different angles of incidence (Fig. 3 of Ref. [113]), namely $\theta_0=0^\circ$, 10° and 18° . We have also plotted the corresponding theoretical results calculated on the basis of Eq. (6.65). It is seen from these results that

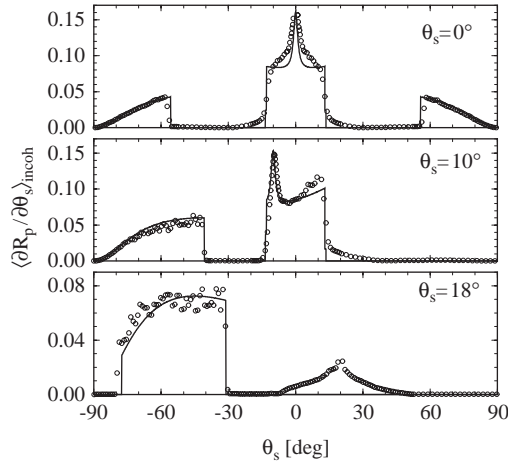


Fig. 68. The contribution to the mean differential reflection coefficient from the incoherent component of the scattered light when p-polarized light of wavelength $\lambda = 612$ nm is incident on a one-dimensional randomly rough gold surface ($\epsilon(\omega) = -9.00 + i1.29$) characterized by the West–O’Donnell power spectrum (6.66) with $\delta = 10.9$ nm, $k_{\min} = 8.48 \times 10^{-3} \text{ nm}^{-1}$, $k_{\max} = 13.26 \times 10^{-3} \text{ nm}^{-1}$, and $\theta_{\max} = 13.5^\circ$. The angles of incidence are as noted. (o) experimental data of Ref. [113], (solid line) results calculated on the basis of Eq. (6.65). (After Ref. [212].)

the theoretical curves are in good qualitative and quantitative agreement with the experimental data, all the more so since there are no adjustable parameters in the theory. An enhanced backscattering peak is seen in the results for angles of incidence $\theta_0 = 0^\circ$ and 10° , whereas none is seen for $\theta_0 = 18^\circ$. Since the angle $\theta_{\max} = 13.5^\circ$, the absence of an enhanced backscattering peak for $\theta_0 = 18^\circ$ is due to the inability of the incident light to excite surface plasmon polaritons for $\theta_0 > \theta_{\max}$ on account of the power spectrum of the surface roughness.

Since the single-scattering contribution to $\langle \partial R / \partial \theta_s \rangle_{\text{incoh}}$ is proportional to $g(|q - k|) = g((\omega/c) |\sin \theta_s - \sin \theta_0|)$, it follows from the fact that the power spectrum given by Eq. (6.6b) vanishes for $|q - k| < k_{\min}$, that single-scattering processes make no contribution to $\langle \partial R / \partial \theta_s \rangle_{\text{incoh}}$ for scattering angles in the interval

$$-\frac{ck_{\min}}{\omega} + \sin \theta_0 < \sin \theta_s < \frac{ck_{\min}}{\omega} + \sin \theta_0. \quad (6.66)$$

For $\theta_0 = 0^\circ$, this condition translates into the statement that single-scattering processes do not contribute to $\langle \partial R / \partial \theta_s \rangle_{\text{incoh}}$ for $-55.69^\circ < \theta_s < 55.69^\circ$; for $\theta_0 = 10^\circ$, they do not contribute when $-40.7^\circ < \theta_s < 88.42^\circ$; and for $\theta_0 = 18^\circ$, they do not contribute when $-31.13^\circ < \theta_s < 90^\circ$. Within these angular intervals $\langle \partial R / \partial \theta_s \rangle_{\text{incoh}}$ is determined only by multiple-scattering processes. Consequently, from the argument given in the introduction to this section, and the result given by Eq. (6.65), we expect that the height of the enhanced backscattering peak should be twice as high as the background at its position. This is indeed the case in the theoretical results presented in Fig. 68, and very nearly the case in the corresponding experimental results. Outside the interval defined by Eq. (6.66), $\langle \partial R / \partial \theta_s \rangle_{\text{incoh}}$ is determined by both single- and multiple-scattering processes. Moreover, because the surface plasmon polaritons excited by the incident light when $|\theta_0| < \theta_{\max}$ can be converted into volume electromagnetic waves in the vacuum only when $|\theta_s| < \theta_{\max}$, we expect that the multiple-scattering contribution to $\langle \partial R / \partial \theta_s \rangle_{\text{incoh}}$ will be nonzero only for $|\theta_s| < \theta_{\max}$, independent of the value of θ_0 . This expectation is confirmed by the theoretical results for

all three angles of incidence, and in the experimental data for $\theta_0 = 0^\circ$ and 10° . However, it is seen that there is a nonzero contribution to $\langle \partial R / \partial \theta_s \rangle_{\text{incoh}}$ in the interval $|\theta_s| < 13.5^\circ$ for $\theta_0 = 18^\circ$, in which case no surface polaritons are excited. This result is due to the fact that the measured power spectrum $g(|k|)$ of the surface fabricated by West and O'Donnell was not identically zero for $|k| < k_{\text{min}}$ [113].

The good agreement between the experimental data and theoretical results based on the coherent interference of multiply scattered surface plasmon polaritons, excited through the roughness of the surface, with their reciprocal partners, support the conclusion of West and O'Donnell that their results confirm the surface polariton mechanism for the enhanced backscattering of light from weakly rough one-dimensional random metal surfaces predicted in Ref. [112].

It should also be noted that West and O'Donnell found no evidence of enhanced backscattering in their results for the scattering of s-polarized light from the same surfaces used in the experiments with p-polarized light. This is consistent with the conclusions of the theoretical work presented in Refs. [112,118], and reflects the fact that a planar vacuum–metal interface does not support s-polarized surface electromagnetic waves [213]. For large-amplitude, large-slope, one-dimensional randomly rough surfaces, enhanced backscattering is predicted when they are illuminated by *both* p- and s-polarized light [17,214]. However, the mechanism responsible for it in this case is not associated with surface plasmon polaritons [215], and hence lies outside the scope of this article.

6.2. Surface plasmon polaritons and the enhanced optical transmission of periodically structured films

Surface polaritons on periodically structured films and related SPP band-gap effects not only affect the electromagnetic near-field close to the structure but also modify optical properties of periodically nanostructured films since they are determined by the behavior of SPP on a structured surface. Recent experimental observations of the enhanced optical transmission of metal films with periodic subwavelength holes has given rise to a considerable interest in the optical properties of such metallic structures [198,216,217]. Not only transmission but also reflection and absorption by the film are determined by SPP properties on the periodically structured interfaces of the film.

6.2.1. Optical transmission of a metal film with a periodic hole array: a near field view

The optical transmission of a subwavelength aperture in a nontransparent screen is very low and is proportional to the fourth power of the ratio of its diameter and the light wavelength [218]. However, if a metal film is perforated with a periodic array of such holes, the optical transmission can be significantly enhanced at certain wavelengths [198]. If it is normalized to the total area of the illuminated holes, the transmission coefficient corresponds to an enhancement of up to three orders of magnitude compared to the transmission of the same number of individual holes. This enhancement depends on the array geometry (hole shape and size and periodicity), light wavelength, angle of incidence, as well as on the material of a film.

The spectral dependencies of the enhanced transmission and the role of the resonant excitation of surface polaritons in the enhancement have been studied in Ref. [24]. The transmission spectra obtained for different parameters of a square array of holes are presented in Fig. 69 together with the transmission of a continuous silver film of the same thickness. The transmission of a continuous film increases as the wavelength of light decreases towards the bulk plasmon wavelength ($\lambda_p \approx 360$ nm) at which a metal film is significantly transparent. The spectral dependence of the transmission of a hole array has a complex behavior revealing enhanced as well as suppressed transmission compared to that of the continuous film.

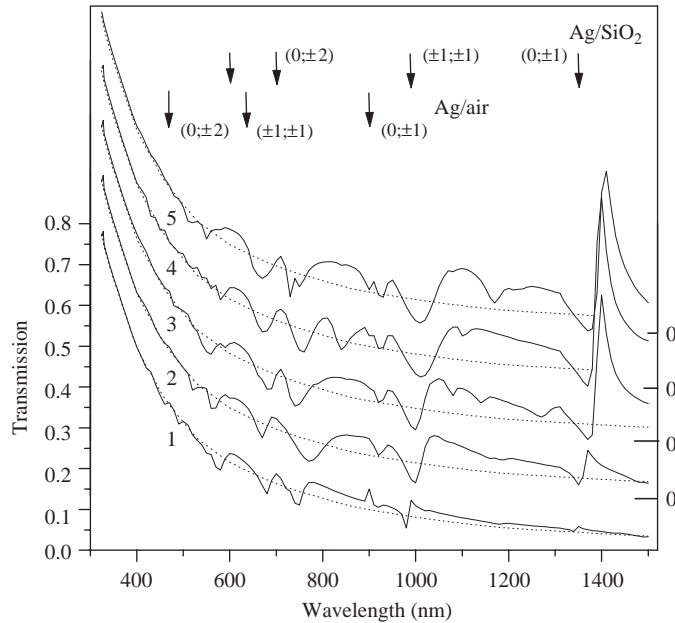


Fig. 69. Normal incidence transmission spectra of a continuous silver film (dotted lines) and a film perforated with a periodic hole array (solid lines). Periodicity $D = 900$ nm, hole diameters (1–5) 150, 225, 250, 275, and 300 nm, respectively. Please note that the curves (2–5) are shifted with respect to the y -axis scale. The film thickness is 20 nm. The excitation spectra of surface polaritons on smooth Ag interfaces are indicated by arrows with corresponding (p, q) -parameters. (Ref. [24].)

The magnitude and spectral dependence of the transmission depends strongly on the array parameters. For the 150 nm size holes the observed enhancement is relatively low (Fig. 69). The enhancement increases with the hole diameter, and its spectral position shifts towards the longer wavelengths. The increase of the array periodicity leads to a blue shift of the enhanced transmission spectra but is less important for the magnitude of the transmission.

The crucial difference between a smooth and a periodically structured metal film is that when the latter is illuminated even at normal incidence surface plasmon polaritons can be excited on both interfaces of the film under certain conditions, as has been discussed above. In the case of a perforated metal film, the same periodic structure forms two polaritonic crystals on the two interfaces of the film. If a film is sufficiently thick, the coupling of the SPP excited on the different interfaces is sufficiently weak, and two sets of SPP Bloch waves can be excited on the different interfaces of the film. The spectra of the resonant SPP excitation on smooth air–metal and glass–metal interfaces of an Ag film (not SPP Bloch waves) calculated with Eq. (1.32) (Section 1.1.3) are shown in Fig. 69. One can see a good agreement in the spectral position of the SPP resonances with the enhanced transmission of the 150 nm hole array, but for the holes of 300 nm diameter the transmission spectra are significantly red-shifted.

The spectral ranges of the enhanced transmission correspond to the excitation of surface polaritons on a metal film with periodic holes, which are SPP Bloch waves, as has been discussed above. The dependence of the transmission enhancement on the periodicity is determined by the SPP Bloch wave conditions on the periodic structure. Because of the weaker scattering of SPPs by smaller holes, the formation of SPP Bloch waves is not as efficient as in the case of larger holes. Thus, both a shift of the

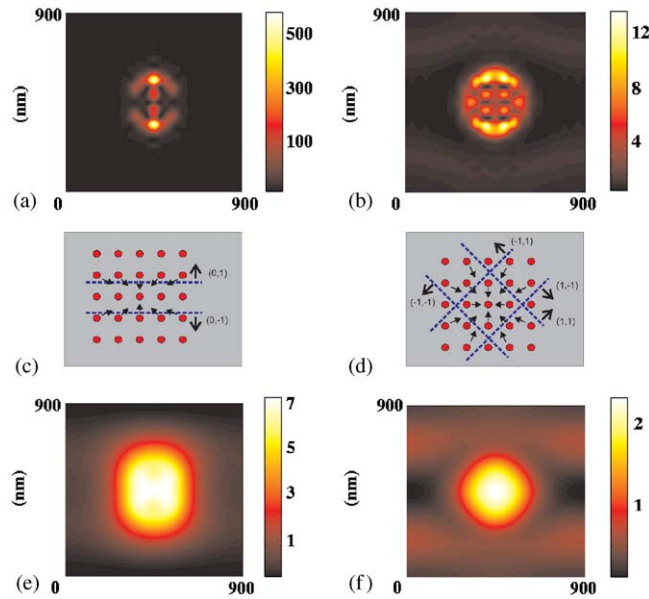


Fig. 70. Intensity distributions of the transmitted light over a unit cell of a periodic array at the wavelength of 1420 nm corresponding to the SPP Bloch wave with $p = 0$, $q = \pm 1$ (a,e) and 800 nm corresponding to $p = \pm 1$, $q = \pm 1$ (b,f) at the distance of 15 nm (a,b) and 100 nm (e,f) from the surface. Schematic of the SPP Bragg reflection demonstrating a build-up of the near-field distribution of the transmitted light (c,d). The hole array periodicity is 900 nm, a hole diameter is 300 nm, and the silver film thickness is 20 nm. Polarization of the incident light is along a vertical axis of the images. (Ref. [24].)

SPP resonances on the structured film from the SPP resonances of a smooth film and the transmission enhancement are smaller in the case of small holes. However, SPP into SPP scattering and scattering of SPPs into light (radiation damping) become more and more important with the increase of the hole diameter resulting in a spectral shift from the position of SPP resonances on a smooth film and an increase of the transmission enhancement for larger holes, as has been shown by exact numerical modeling (Fig. 69).

To understand the mechanisms of the transmitted wave formation, the near-field distribution of transmitted light above a surface of the array has been studied at different wavelengths corresponding to the enhanced as well as suppressed transmission. The near-field distribution of the transmitted light over a surface has been calculated, revealing a fine structure related to the electromagnetic coupling between holes. The symmetry of the observed near-field distribution corresponds to the symmetry of the Brillouin zone of the hole lattice. The near-field images of the intensity distribution have been calculated for the wavelengths of $\lambda = 800$ and 1420 nm, which correspond to the SPP Bloch wave excitation (Fig. 70), as well as for the wavelengths of $\lambda = 633$ nm at which the transmission is suppressed, and $\lambda = 330$ nm at which no surface polaritons exist on a silver film (Fig. 72). In the latter case, the transmitted intensity distribution is related to the hole itself. In the near-field, additional lobes of the field distribution are observed in the direction of polarization of the exciting light, which are probably related to the localized electronic oscillations close to the edge of a hole excited by the incident light. This intensity distribution can be obtained from continuity conditions of the electromagnetic field for an individual hole. The intensity of the transmitted light exhibits a very weak distance dependence indicating that the propagating

components of the diffracted field are dominant in the near-field at this wavelength, while the contribution of the evanescent components is negligible.

The transmitted light distribution at the wavelengths of the SPP Bloch waves is far more complex in the near-field, revealing a rich fine structure at and around the holes (Fig. 70). The lowest energy surface polariton that can be launched by the array under consideration corresponds to the $(p, q) = (0, \pm 1)$ modes at a glass–metal interface ($\lambda_{\text{SP}} \approx 940$ nm) excited with 1420 nm wavelength light. Two other modes $(\pm 1, 0)$ at this wavelength are forbidden due to the polarization of the exciting light. Under illumination at normal incidence, the two excited SPP modes correspond to standing SPP Bloch waves that are excited at the edge of the Brillouin zone. They are formed by SPPs in a vertical direction undergoing scattering/reflection by the holes in the film. The reflected SPPs produce a multiple beam interference pattern of the symmetry determined by the orientation of the hole rows playing the role of a diffraction grating for the SPPs or, in analogy with a crystalline lattice of solids, a two-dimensional Bragg reflector. The hole rows of different orientations with respect to the SPP propagation direction reflect the SPPs in a given direction, determined by a diffraction order, with different efficiencies. The efficiency of reflection in different directions corresponds to the corresponding components of the Fourier spectrum of a periodic structure. In addition, a surface polariton decays during propagation due to losses in metal. These factors result in the intensity variation of the observed interference maxima related to different reflection directions.

Two strongly localized spots of the highest intensity correspond to the reflection from the hole rows perpendicular to the SPP propagation direction (zero-order diffraction). The SPP reflection in this direction is most efficient. In the localization spots an intensity enhancement of up to 500 times is observed. The size of the localization is about 40 nm suggesting a good quality factor of the structure with respect to the SPP reflection processes. The SPP propagation length along a smooth metal/glass surface at this wavelength is about 300 μm , which corresponds to more than 100 rows of the hole array efficiently contributing to the SPP Bloch wave formation.

Different surface polariton modes propagate in the different directions determined by the (p, q) -parameters which in turn correspond to different directions in the Brillouin zone. For this reason, the symmetry of the near-field distribution of the transmitted light is different at different wavelengths as it is related to the different SPP Bloch modes. For example, the SPP modes with $(p, q) = (\pm 1, \pm 1)$ on an air–metal interface ($\lambda_{\text{SP}} \approx 780$ nm) are excited with light of 800 nm wavelength. Despite the complex lateral variations of the near-field intensity (Fig. 70) its fine structure corresponds to the symmetry of the Bragg grating orientations. As in the case discussed above, the maximum intensity is observed for the spots corresponding to zero-order diffraction.

Although the maximum intensity of the transmitted light is observed in the vicinity of the holes in the metal due to the field configuration of the SPP Bloch modes, the logarithmic representation of the same near-field distribution (Fig. 71) clearly shows that the transmitted light gets through the metal film itself: photon tunneling via the SPP modes takes place all over the metal film where the SPP field exists [200,202,219].

In contrast to the transmission without SPP excitation, the field at the wavelengths of the enhanced transmission conditions naturally has strong evanescent components. With the increase of the distance from a surface the localization pattern is smoothed out but the overall symmetry of the light distribution above a hole is still determined by the symmetry of the scattered SPP modes (Fig. 70).

Light of wavelength 633 nm falls in the SPP band gap of the periodic structure under consideration (Fig. 72). Nevertheless, as the band gap is narrow, low-efficiency nonresonant excitation of the SPP

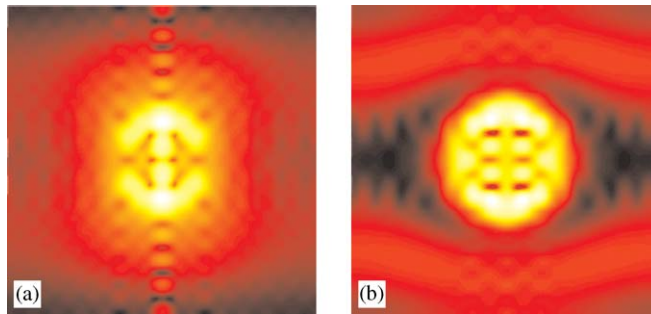


Fig. 71. Log-scale representation of the near-field intensity distributions presented in Fig. 70a,b. (After Refs. [24,219].)

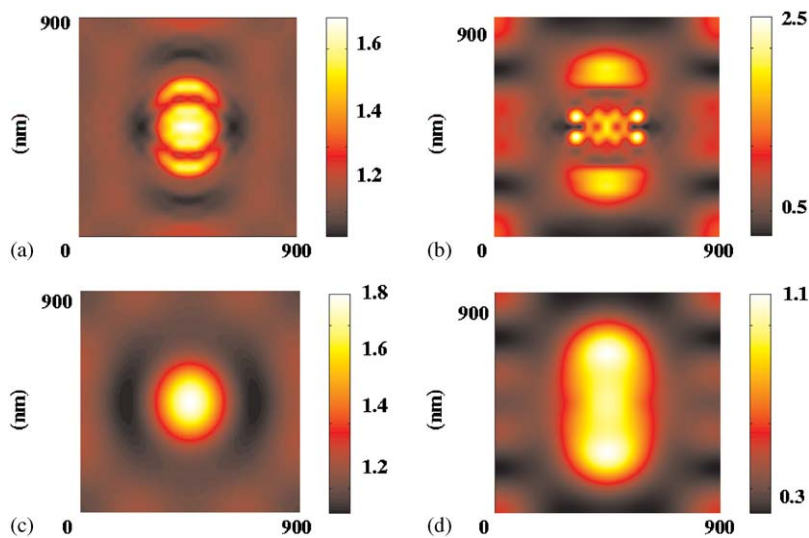


Fig. 72. Intensity distributions of the transmitted light over a unit cell of a periodic array at the wavelength of 330 nm at which no SPP modes are allowed (a,c) and 633 nm at which no SPP Bloch waves are excited (b,d) at the distance of 15 nm (a,b) and 100 nm (c,d) from a surface. All other parameters as in Fig. 70. (Ref. [24].)

modes still takes place, leading to a very complex field distribution over a surface. In addition to the field localization close to the hole edges (as in the case without SPP excitation), the structure related to the excitation of nearby situated SPP modes is also visible. The near-field intensity is much lower in this case than at the SPP Bloch wave wavelengths.

Numerical modeling of the optical transmission through an array of periodic subwavelength holes has shown that the enhanced transmission is related to photon tunneling via SPP Bloch modes on periodically structured films. Calculated spectra are in a good agreement with the experimental data [198] describing spectral dependencies of the enhanced transmission. However, for different geometries of the film, depending on the shape and size of the features as well as on the film thickness, other mechanisms can contribute to the transmission enhancement as will be discussed below.

6.2.2. *The origin of the enhanced transmission through periodically nanostructured metal films*

Since the first experimental observations of the enhanced transmission, the role of surface plasmon polaritons has been assumed but the detailed picture of the transmission enhancement is still being debated. Different models have recently considered a number of mechanisms such as resonant transfer of surface polaritons excited on the inner (glass–metal) interface of a film to the upper (air–metal) interface, penetration of the SPPs excited on the inner interface to the surface through the holes, localization of the electromagnetic field inside the holes, the excitation of SPP Bloch waves discussed above and their subsequent tunneling through a film, etc. [24,199,220–226]. All these theories heavily rely on thorough numerical simulations of the electromagnetic field propagation through a structure in one or another approximation, and not surprisingly most of them give about the same spectral dependencies, which are in relatively good agreement with the experiment. However, the physical interpretation of the numerical results is focused, in most cases, on a one at a time feature of this complex process, resulting in a fragmented and obscure picture of the transmission enhancement origin.

As has been shown above, surface polariton excitation itself results in a significant field enhancement in the near-field proximity to a surface (Section 1). If a smooth metal film is illuminated through a prism, under a certain angle of incidence the SPP excitation conditions are satisfied, and light is resonantly transformed into surface polaritons on the air–metal interface leading to a very low reflection from a film. It should be noted that in this configuration the SPP mode on a smooth glass–air interface cannot be excited under any circumstances. Thus the incident light is directly coupled to surface polaritons on the upper interface. The process of the photons transfer to the opposite side of a metal film is analogous to the well-known effect of resonant tunneling through a potential barrier. Such resonant coupling leads to the intensity enhancement over the film surface by up to 200 and 40 times at the 600 nm wavelength for silver and gold films, respectively [2]. The magnitude of the enhancement depends on the thickness of the film, and there is an optimal thickness at which the enhancement reaches its maximum, when the internal (Ohmic) losses equal the losses due to SPP radiation back into the prism. For thicker and thinner films the enhancement will be smaller.

The SPP electromagnetic field decays exponentially with distance from the surface. Thus in the conventional (far-field) measurements no transmission can be observed despite the significant enhancement of the transmitted field close to the surface. This transmission enhancement has been measured and studied indirectly by the observation of surface enhanced optical phenomena and, more recently, directly with scanning near-field microscopy [15,17]. Thus, even for smooth films the transmission enhancement occurs under the SPP resonant conditions, but to detect it a scatterer should be introduced to convert a surface polariton into a propagating lightwave.

The simplest realization of this is the use of a (randomly) rough film. In this case, coherent scattering of SPPs by surface defects gives rise to the far-field transmitted light. Moreover, a rough surface allows the SPPs to be excited without any special arrangements, as the diffraction of light on surface defects will provide coupling to the SPP modes on both the air–metal and glass–metal interfaces, as has been discussed in Section 1. Thus, even conventionally illuminated rough films can exhibit enhanced transmission. Indeed, the enhanced transmission through a free standing rough metal film due to multiple scattering of the excited SPPs has been predicted theoretically and observed experimentally [203,227,228]. It has also been shown that the coupling between the SPPs on the two surfaces of a film, which is significant if this film is sufficiently thin, results in additional side peaks in the angular distribution of the intensity of the transmitted light in addition to the main peak of the enhanced transmission of thick (no coupling) films. The problem with random roughness is the ill-defined SPP excitation conditions, which result in the low

efficiency of light-into-SPP coupling, broad SPP resonances due to strong scattering and, as a result, a low transmission enhancement.

To overcome this problem and achieve well-defined SPP excitation resonances, efficient light-into-SPP coupling as well as SPP-into-light scattering, periodically structured surfaces should be considered. Diffraction on a periodic structure provides the wavevector conservation and coupling of illuminating light into SPP Bloch waves on a structured surface. As has been shown above, only the surface polaritons satisfying the Bloch waves condition can be excited and subsequently contribute to transmitted light originating from coherent scattering of the SPPs by the array of periodic scatterers. As different SPP Bloch waves propagate in different directions (Eq. (1.32)), they result in the specific near-field distribution of the transmitted light above the surface that has been clearly shown: the near-field distribution has a different symmetry for different SPP Bloch modes corresponding to the different wavelengths of the enhanced transmission [24]. The variation of the structure periodicity leads to the changes of the resonant conditions and thus modification of the enhanced transmission spectrum. The scatterer size has a smaller effect on the spectrum as it influences the resonant conditions indirectly via the efficiency of SPP scattering. Thus, the use of a periodic array of scatterers (instead of random scatterers) allows the efficient excitation of SPPs and at the same their efficient scattering into light. The excitation conditions strongly depend on the parameters of the periodic structure, allowing the efficient control of the transmission spectrum via the control of the SPP band-gap. The resonant light tunneling through a metallic film via states of SPP Bloch waves is responsible for the enhanced transmission. In this case, no holes through a film are required for the transmission enhancement [200–202,226].

Let us consider now qualitatively the scattering of SPPs into light on periodic arrangements of surface features. This is important for understanding the transmitted field formation over a film. Empirically the scattering of SPP into light by surface features can be considered as the excitation of virtual dipoles (related to a surface feature) followed by re-radiation. Such scattering by a periodic system of dipoles must take into account electromagnetic coupling between the dipoles. In periodic arrays of scatterers, the external electric field is coupled with the collective dipole excitations of the array. This results in a coherent scattering of SPPs into light by the dipole array leading to an N^2 dependence of the scattering intensity on the number N of dipoles within a coherence area. For incoherent scattering the intensity is proportional to the number of individual dipoles N contributing to the scattering. The increase of the transmission with the number of scatterers has recently been observed experimentally by the use of films which contain, in addition to a periodic hole array, periodically arranged small protrusions on the surface [229]. In this arrangement, the excitation of SPP Bloch waves (the spectrum of the transmission) is determined by the holes, which provide strong in-plane scattering, and the SPP band-gap formation, while the array of small protrusions (contributing less to the in-plane scattering of SPPs than the holes) provides additional out-of-plane scattering of surface polaritons. Thus, the introduction of additional weak scatterers which can contribute to scattering while not affecting significantly the SPP band-gap formation can be beneficial for the transmission enhancement. Weak surface roughness can play a somewhat analogous role and be significant for out-of-plane scattering, but collective effects in scattering will be lost in this case.

(a) *Photon tunneling via the SPP Bloch modes.* As one can see, the considerations above do not involve any specific type of scatterers (holes or protrusions) on a surface as long as they provide SPP-into-SPP and SPP-into-light scattering. The transmission enhancement can be explained by the resonant tunneling of light from one side of the film to another via the SPP Bloch states on at least one of the film interfaces (tunneling photons into SPP states). Additional transmission enhancement can take place through a film both of whose interfaces are periodically nanostructured. This is the situation when two SPP Bloch

modes are formed at opposite interfaces at the same frequency (tunneling between SPP states) [200]. The simplest realization of this can be the hole structure in a film discussed in the previous Section. In this case, light is first coupled to the SPP Bloch states on one of the interfaces and then tunnels into the SPP Bloch states on the opposite interface. This can occur if some of the resonant states of the SPP crystal formed on one film interface happen to be close to the resonances of the SPP crystal formed on the other film interface. It should be noted that a full energy exchange between tunnel-coupled states, and thus the enhancement magnitude, requires a significant interaction time (depending on the coupling strength) and, therefore, depends on the lifetime of the SPP states involved: one can imagine that the transmitted energy is first built-up in the resonant SPP state [200]. In the absence of Ohmic losses (infinite SPP lifetime, if the radiative losses are small) the transmission magnitude reaches $T = 1$ for a film in a symmetric environment for which similar Bloch states are involved in the tunneling and provide the best lateral overlap of electromagnetic fields on opposite interfaces [200–202]. For thin metal films, for which the interaction between the SPPs on opposite film interfaces is significant, the film SPP modes should be considered instead of independent surface modes. The transmission in this case depends in a complex manner on the interplay between the symmetric and antisymmetric film SPP modes and the symmetries of the SPP Bloch modes involved [202,226,230]. This may result in suppression of the transmission at some frequencies at which the crossing of the SPP Bloch mode branches of different symmetries related to opposite film interfaces takes place [202,230]. Strong absorption is observed at these frequencies.

At normal incidence, only even branches of the SPP Bloch modes are radiative and contribute to the transmission [200]. At oblique incidence, however, both even and odd branches are radiative, leading to two transmission (as well reflection and absorption) peaks instead of the one observed at normal incidence [230]. The tunneling efficiency depends on the densities of states of the SPP modes involved, which are highest at the edges of the Brillouin zones ($d\omega/dk_{\text{SPP}}=0$), i.e., at normal incidence (Fig. 55). Not only transmission is enhanced under resonant conditions, but also absorption in a metal nanostructure. This paradox is resolved taking into account the decrease of the reflection from the nanostructured film [202,226,230]. If a metal film is in asymmetrical surroundings (e.g., a metal film on a substrate), the transmission does not depend on which side is illuminated. However, the absorption may be different upon illumination through the substrate or from another side [226]. This is related to the fact that the absorption is determined mainly by the SPP modes of the illuminated interfaces, while SPP modes on both interfaces participate in the transmission.

(b) *Additional mechanisms of the enhanced transmission.* The other mechanisms of the enhanced transmission through a hole array proposed in the literature and mentioned above require the excitation either of surface polaritons which, as has been shown, can only be SPP Bloch waves on a periodic structure, or of various kinds of localized surface plasmons. Tunneling between LSP states can also contribute to the transmission enhancement. The contribution of these mechanisms to the overall transmission depends on the details of the periodic structure, such as film thickness, scatterer size and type, etc. The specifics of the light transmission through periodic slits in a metal film should be noted. In contrast to the holes, whose cross sections are connected domains, electromagnetic guided modes without a cut-off frequency can exist in slits [222,231]. That is why 1D treatments of the transmission through periodic slits are not always applicable to a hole array.

Depending on the details of the periodic surface structure, and primarily on the size of surface features and the thickness of a film, different dominating transmission mechanisms can be considered. This is important for periodic structures with very small or very large (compared to the period) holes or slits. In the latter case, when the filling factor is large, the total area of the holes is larger than the area of

the remaining metal surface. For such a metallic structure, surface plasmon polaritons cannot be valid electromagnetic eigenmodes. In this case, (interacting) LSP related to the remaining metal features can contribute to the transmission. In the opposite case of very small holes or narrow slits, the SPP interaction with surface features is weak as it depends on their size. This means that for a given SPP propagation length on a smooth surface, the reflection/scattering may be too weak to form a SPP band-gap. (In a periodic system without losses, a band-gap structure is formed always independently of the strength of the interaction.) In this extreme case, other mechanisms such as light tunneling via different kinds of surface plasmons localized in the holes or grooves can be important [223,225]. In the other specific case of dielectric channels in a thick metal film (well defined cylindricality of the channels $h \gg d_h$), it has been suggested that cylindrical surface plasmon modes in the channels contribute to the transmission, in addition to the SPP related tunneling [7,168].

So far we have considered surface plasmon polaritons on a periodically structured surface, which are the combined excitations related to surface plasmons and photons. In dielectrics, when the light wavelength is close to the resonances related to optical phonons, analogous excitations can exist on a surface due to phonon–photon coupling, namely, surface phonon polaritons [1]. They have the same properties as surface plasmon polaritons and, therefore, one should expect analogous properties of surface polaritonic crystals in the appropriate spectral range. Thus, enhanced transmission should be expected for periodically structured dielectric films corresponding to the Bloch waves in surface phonon polaritonic band-gap structures [219].

The optical transmission due to light tunneling via surface polariton states provides a possibility for the efficient control of the spectrum and magnitude of the transmission, and makes periodically nanostructured films very promising for applications in novel photonic and optoelectronic devices. One example of such applications in active photonic devices that allow controlling light with light is discussed in the next Section. Surface polaritonic crystals provide the possibility of controlling not only the spectrum and spatial structure of the transmitted light, but also the temporal delay of the transmitted light pulses, that is important for certain applications in optical communication [232].

6.2.3. Polarization effects in the light transmission through a polaritonic crystal

The transmission spectrum of a metal film with a square array of circular features does not depend on the polarization of the incident light at normal incidence. However, by introducing asymmetry into the surface structure it is possible to achieve polarization control of the transmission, which can be advantageous for possible applications of the metallic nanostructures [233].

To understand the polarization sensitivity of the SPP-related enhanced transmission, the spectrum of the SPP Bloch waves on a periodically structured film should be considered. A rigorous consideration of the symmetry properties of the eigenstates of a periodic structure with an asymmetrical lattice basis would require an approach analogous to the linear combination of atomic orbitals (LCAO) method or the Wannier functions conventionally used in solid state physics [159]. Nevertheless, it will be instructive to consider a simplified model using the SPP Bloch modes.

The field of SPP modes on a periodically structured surface can be described by

$$E(x_1, x_2) = U_{\mathbf{k}_{\text{sp}}}(\xi, \eta) \exp[i(k_{\text{sp}}^{(1)} x_1 + k_{\text{sp}}^{(2)} x_2)] , \quad (6.67)$$

where $U_{\mathbf{k}_{\text{sp}}}(\xi, \eta)$ possesses the periodicity of the array and is the SPP Bloch function and \mathbf{k}_{sp} is the wave vector of the Bloch wave. Here the coordinate systems (ξ, η) of the basis and (x_1, x_2) of the lattice are not

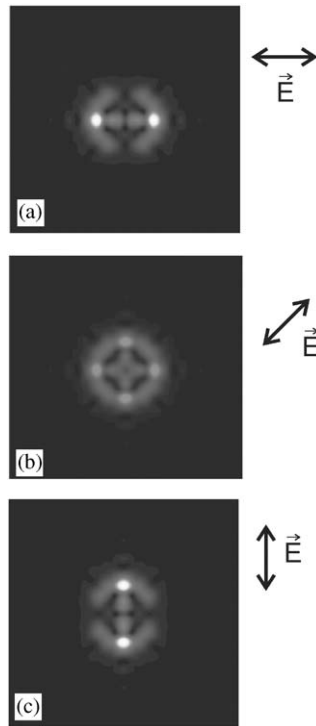


Fig. 73. The near-field distribution of the transmitted light over a unit cell of a hole array as in Fig. 70(a) at the transmission peak corresponding to the SPP modes in the Γ - X direction of the Brillouin zone ($\lambda = 1420$ nm) for different polarizations of the incident light (indicated by arrows). (After Refs. [24,219].)

generally independent, and are introduced to simplify symmetry considerations. At normal incidence of light, the allowed SPP wavevectors on the periodically structured surface are determined by

$$\mathbf{k}_{\text{spp}} = \pm p \frac{2\pi}{D} \mathbf{u}_1 \pm q \frac{2\pi}{D} \mathbf{u}_2, \quad (6.68)$$

where \mathbf{u}_1 and \mathbf{u}_2 are the unit reciprocal lattice vectors of the periodic structure, D is its periodicity (assumed to be the same in both the x_1 - and x_2 -directions), and p and q are integer numbers corresponding to the different directions of the SPP Brillouin zone. At normal incidence, SPP can be excited if the electric field of the incident light has a component in the direction of SPP propagation: $(\mathbf{E} \cdot \mathbf{k}_{\text{spp}}) \neq 0$. Thus, the polarization dependence of the coupling efficiency in different directions of the Brillouin zone is proportional to $|p \sin \phi + q \cos \phi| / (p^2 + q^2)^{1/2}$, where ϕ is the polarization azimuth angle with respect to the x_2 axis of the lattice (Fig. 74).

It is clear from this analysis that for a square lattice with basis elements of circular symmetry no polarization dependencies of the SPP excitation and, thus, of the enhanced transmission, can be expected at normal incidence. This leads to the situation that the near-field distributions of the transmitted light are different for different incident light polarizations (Fig. 73) but the transmission spectrum is the same (Fig. 69). However, in the case of an elliptical basis of the lattice (Fig. 74), polarization dependencies of the transmission spectrum are significant.

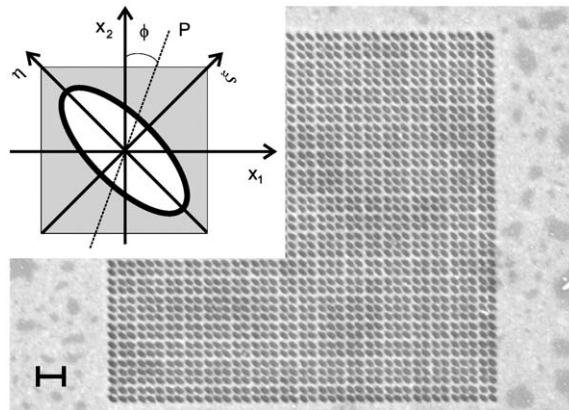


Fig. 74. The image of the periodic array of elliptical holes in a gold film. The scale bar is $2\ \mu\text{m}$. A schematic of the primitive cell of the lattice is shown in the inset. (Ref. [233].)

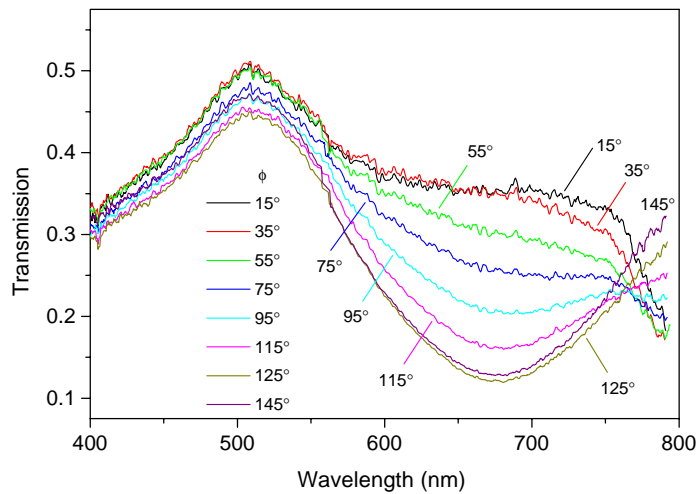


Fig. 75. Normal incidence transmission spectra of an array of elliptical holes for different polarizations of the incident light without polarization analysis of the transmitted light. The polarization angle ϕ is measured with respect to the y -axis (Fig. 74). (Ref. [233].)

The transmission spectra of the structure obtained at normal incidence without polarization analysis of the transmitted light are shown in Fig. 75 for different polarization states of the incident light with respect to the orientation of the array. For incident light polarized parallel to the principal lattice axes (x_1 or x_2) the transmission spectrum is observed with broad features in the 450–550 and 750–800 nm spectral ranges. With the increase of the polarization azimuth ϕ up to about 20–25° the transmission increases in the 550–750 nm wavelength range, where the spectrum becomes rather flat. With further increase of ϕ when the light is polarized along the short principal axis of the elliptical holes, the transmission in the long-wavelength spectral range (750–800 nm) is significantly suppressed. For the polarization azimuth

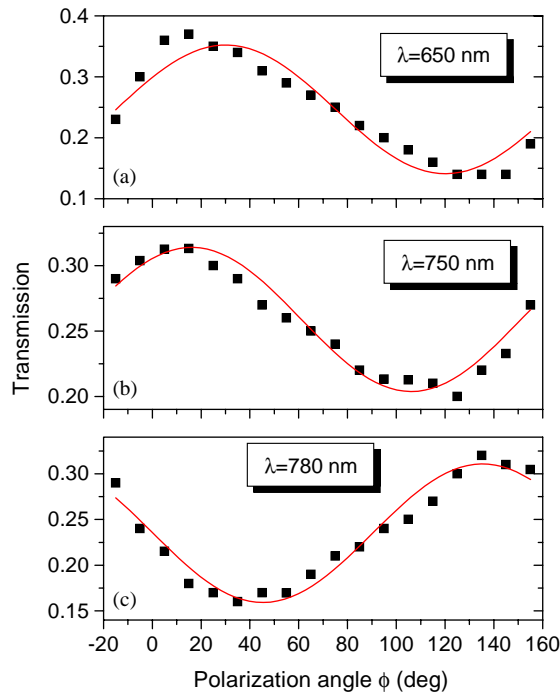


Fig. 76. Polarization dependencies of transmission at the wavelengths indicated in the panels. Polarization angle ϕ is measured with respect to y -axis (Fig. 74). (Ref. [233].)

along the long principal axis of the ellipse, the transmission in the far red part of the spectrum is enhanced, but becomes smaller at around 600–700 nm. Thus, with rotation of the polarization of the incident light, the intensity of the spectral components of the transmitted light exhibit complex, oscillatory behavior different in different spectral ranges (Fig. 76). The relatively well-defined long-wavelength band at around 780–800 nm has a transmission minimum for the polarization of light parallel to the short principal axis of an ellipse, and a maximum for the polarization along the long principal axis. For all wavelengths, the dependencies show a 2-fold symmetry corresponding to the symmetry of the array: a square lattice with basis elements of a 2-fold rotational symmetry.

Taking into account different scattering properties of the ellipse in different directions, one can understand this by considering the mixing of the SPP Bloch states of the square lattice due to the lower symmetry of the basis elements. Elliptical holes with different sizes in the different directions modify the Brillouin zone structure by introducing size-dependent anisotropic scattering. This leads to the reduction of the Brillouin zone symmetry, showing a 2-fold rotation axis instead of the 4-fold axis appropriate for a square lattice, effectively creating a medium with a pronounced linear birefringence in dichroism. Linearly polarized waves with polarization azimuth along the main axes of the elliptical basis will be the polarization eigenstates of the structure. As a result of the anisotropic SPP Brillouin zone, the transmission spectrum of the structure is dependent on the polarization of the incident light.

Mixing and re-excitation of surface plasmon resonances in different directions of the Brillouin zone are responsible for anisotropic retardation and eventually for the ellipticity of the polarization state of

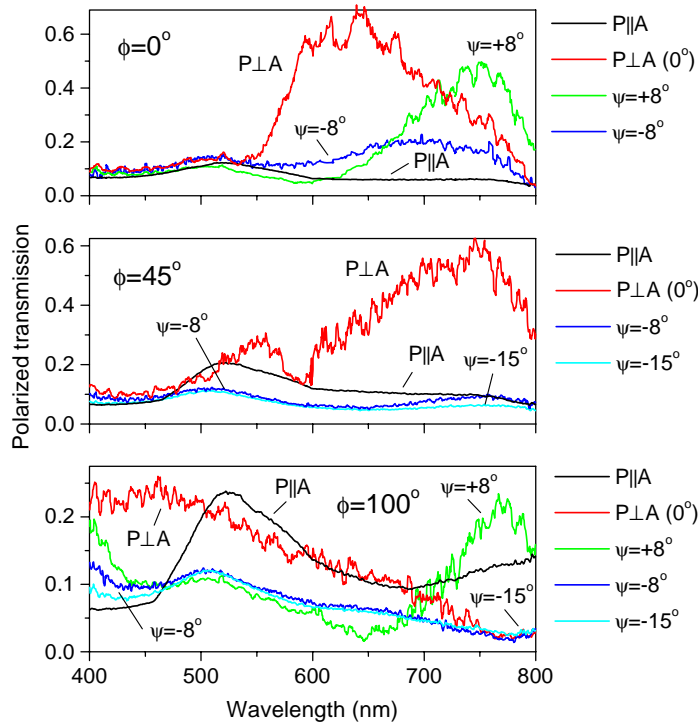


Fig. 77. Spectra of polarized transmission of the array measured for the incident light polarization (a) $\phi = 0^\circ$ and (b) 45° , (c) 100° . The spectra of the light transmitted without polarization state change ($P_{\parallel A}$), with orthogonal polarization ($P_{\perp A}$, $\psi = 0^\circ$), and close to orthogonal polarization $\psi = \pm 8^\circ$ and 15° , are shown. (Ref. [233].)

the transmitted light. Polarized transmission spectra show that the changes of the polarization state of the transmitted light are strongly wavelength dependent (Fig. 77). This allows identification of transmission resonances contributing to the broad transmission spectrum. Several resonances can be identified by the different degree of ellipticity at around 750, 650, and 550 nm. Thus, by controlling the polarization state of the incident light and exercising polarization selection of the transmitted light, the transmission spectrum of the elliptical hole array can be tuned at a specific resonant band corresponding to one or another SPP mode of the system, quasi-continuously from blue to red wavelengths of the visible spectrum [233].

6.2.4. Light-controlled transmission of polaritonic crystals

Since surface polaritons are extremely sensitive to the refractive index of an adjacent dielectric medium (Eq. (1.6)), one can expect that by controlling this refractive index, an effective control of the SPP-mediated light transmission through periodically structured metal films can be realized [7,168,201]. The changes of the dielectric properties of the material in contact with polaritonic crystals lead to the modification of the SPP Bloch state spectrum, and thus, to changes in the resonant conditions of light tunneling. One can imagine several choices for an active dielectric medium such as electro-optic and ferroelectric materials whose dielectric properties depend on an applied electric field, magnetic materials which can be controlled with a magnetic field, as well as nonlinear optical materials that can be controlled with

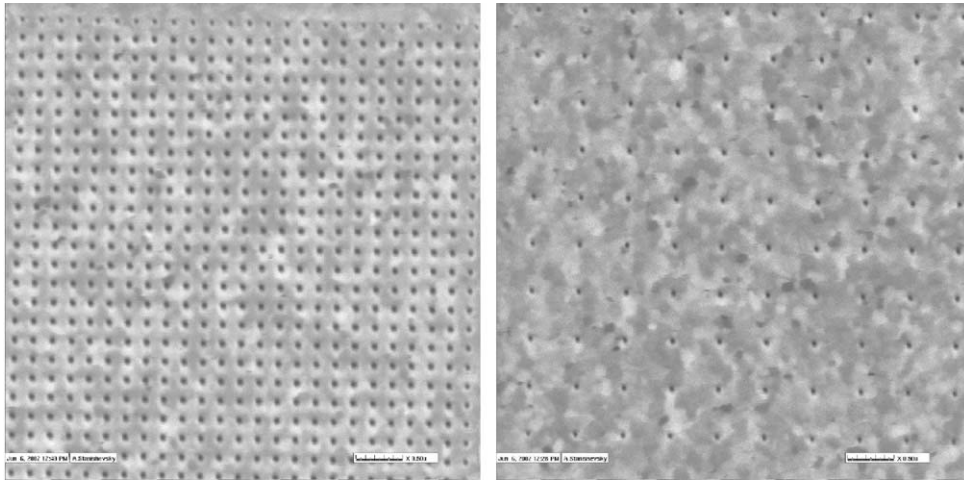


Fig. 78. Focused ion beam milled arrays of cylindrical channels in a 400 nm thick free standing gold film. The scale bar is 500 nm. (Ref. [168].)

light. The latter is especially important for the development of all-optical active devices. This effect has recently been observed in light transmission through a periodic array of subwavelength holes (cylindrical channels) in a free standing metal film embedded in a nonlinear material: the transmitted light intensity was controlled with light of a different wavelength [168].

The advantage of a surface polaritonic crystal for the development of this kind of all-optical active elements is two-fold: a high sensitivity of SPP resonance to minor modifications of surroundings, and the electromagnetic field enhancement related to surface plasmon modes that can allow achieving these modifications at a low illuminating light intensity. The local electric field E_L induces changes in the dielectric constant ϵ of the nonlinear optical material close to a surface:

$$\epsilon_i = \epsilon_i^{(0)} + 4\pi\chi_i^{(3)}|E_L|^2, \quad (6.69)$$

where $\epsilon_i^{(0)}$ and $\chi_i^{(3)}$ are the linear dielectric constant and the third-order nonlinear susceptibility of the optical material adjacent to the metal surface, respectively. As a consequence of dielectric constant change, the SPP modes may experience a noticeable frequency shift, thus modifying the resonant tunneling conditions and transmittance of the nanostructure.

In the experiments on all-optical control of SPP crystal transmission, square arrays of nanometric cylindrical holes have been used in a freestanding 400 nm thick gold membrane (Fig. 78). The membrane was embedded in a poly-3-butoxy-carbonyl-methyl-urethane (3BCMU) polydiacetylene thin film. BCMU polydiacetylene hold the current record for the largest fast nonresonant optical $\chi^{(3)}$ nonlinearity [234]. The samples were illuminated at an angle of incidence of approximately 45° , so that the projection of the illuminating light wavevector onto the surface is in the direction approximately corresponding to the direction along the hole rows (Γ - X direction of the Brillouin zone). The transmitted light was collected through a fiber tip positioned over the array using shear-force distance regulation.

Modulation of the transmission of the array of 20 nm holes at the signal light wavelength ($\lambda = 633$ nm) induced by modulation of p-polarized control light ($\lambda = 488$ nm) illuminating the same area was directly

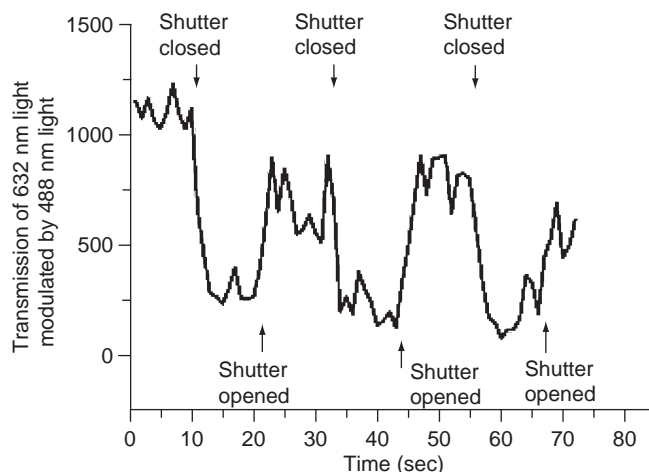


Fig. 79. Modulation of the 20 nm holes array transmission at 633 nm induced by modulation of the 488 nm control light. The time behavior of the observed switching is determined by the time constant of the measurements (~ 3 s) needed for the signal integration. The shutter of the control light was closed and opened three times during the experiment. (Ref. [168].)

observed (Fig. 79). At the same time, for the arrays with larger holes (~ 100 nm diameter) no significant variations of the transmission were observed for the given wavelengths and intensities of the signal and control light. The polarization properties of the signal light passing through the array of smallest holes are strikingly different from the polarization properties of the transmitted red light measured in the absence of control (blue) light illumination. In the absence of modulating blue light, the transmission of p-polarized red light is higher than the transmission of s-polarized red light by approximately a factor of 3. This observation is consistent with the theoretical model of the transmission of subwavelength hole arrays relying on the excitation of surface polaritons on both film interfaces, since at the oblique incidence the light of different polarizations interacts with different SPP resonances (as was discussed in the previous sections). This ratio of the intensities of p- and s-polarized transmitted light has been observed to be rather insensitive to the hole diameter in the range of hole sizes studied, as is expected since the spectrum of the SPP excitations depends mainly on the periodicity of the structure.

However, under modulation with p-polarized control light, the ratio of intensities of p- and s-polarized transmitted signal light increases to approximately a factor of 20. Polarization measurements offer the best distinction between the nonlinear optical mixing effects of interest and possible thermal effects such as thermal expansion of the holes, heating of the filling material, etc., which could interfere with the nonlinear optical effects. The thermal effects on the transmission through hole arrays has been observed at much larger than usual illumination intensities of the modulating blue light. However, the polarization dependence of the modulated signal light obtained under such conditions is very weak (Fig. 80 c): the ratio of transmission of p- to s-polarized light falls to approximately a factor of 1.5 in the thermally modulated red light transmission.

The strong polarization dependence of the modulated transmission of the signal light can be explained taking into account the properties of surface polariton Bloch waves on the periodically perforated gold film, the properties of surface plasmons in cylindrical channels, and the properties of the $\chi^{(3)}$ nonlinear susceptibility tensor of 3BCMU polydiacetylene. To understand the induced transmission behavior, let

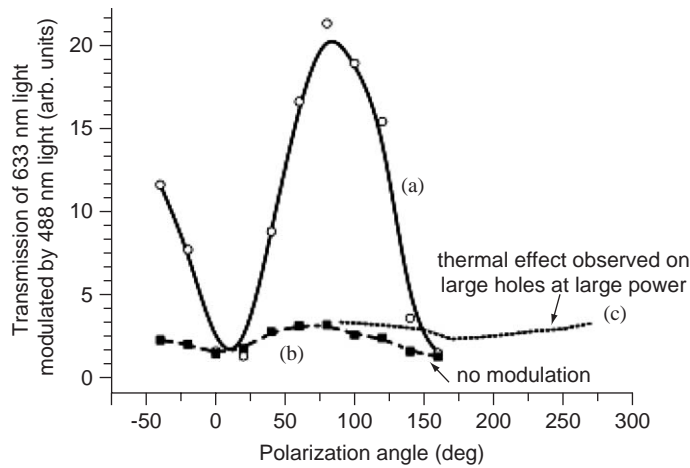


Fig. 80. (a) Polarization dependence of the modulation of the transmission through a 20 nm holes array at 633 nm induced by modulation of p-polarized 488 nm light. (b) Polarization dependence of the transmission through the same array at 633 nm without illumination with 488 nm light. (c) Polarization dependence of the thermal modulation of 633 nm light transmission. (Ref. [168].)

us consider the spectrum of electromagnetic excitations in our system. The metal film is sufficiently thick to neglect the interaction between surface polaritons excited on different interfaces. Therefore, two independent sets of resonances can be considered (see Section 2). Since one side of the metal film is covered with a polymer ($\epsilon \approx 1.7$) while the other is in contact with Cr/Si₃N₄ ($\epsilon \approx 2.0$), the SPP resonant conditions will be slightly different on different interfaces of the film. The filling factor of the structures under consideration is rather small ($f = (d/D)^2$), so that the modifications of the SPP dispersion on a smooth interface are also expected to be small. This can be used to estimate the resonant frequencies of the SPP Bloch waves excitation.

For p-polarized illuminating light and the angle of incidence of 45° used in the experiment, the SPP Bloch waves corresponding to the 3rd and 4th Brillouin zones in the Γ - X direction ($p = \pm 1, \pm 2, \dots; q = 0$) are excited at wavelengths close to $\lambda = 633$ nm on the polymer-metal interface (Fig. 81). The SPP states related to the illuminated (Si₃N₄) interface will be red-shifted with respect to this frequency. At the same time for s-polarized light the resonant SPP frequencies on the metal-polymer interface are far from the signal light frequency, while the SPP Bloch waves from the 4th Brillouin zone of the illuminated interface are relatively close to it. This SPP is about two times weaker than the SPP on a pure gold surface due to the presence of a Cr layer, which has strong losses at this wavelength ($\text{Im } \epsilon \approx 30$). An analogous picture can be constructed in the Γ - M direction of the SPP Brillouin zone, but the related resonant frequencies $(p, q) = (\pm 1, \pm 1)$ will lie between the frequencies determined by $(p, q) = (\pm 1, 0)$ and $(p, q) = (\pm 2, 0)$, and therefore, far from the signal wavelength. However, these resonances can become important for other angles of incidence.

To complete the picture, the resonances associated with cylindrical channels in a metal film should be considered [3]. For thick films with well defined cylindricality of the channels ($d \ll h$), the spectrum of surface electromagnetic excitations in channels has a rather complicated structure with both radiative and nonradiative modes present. Individual cylindrical channels will have a discrete spectrum of resonances

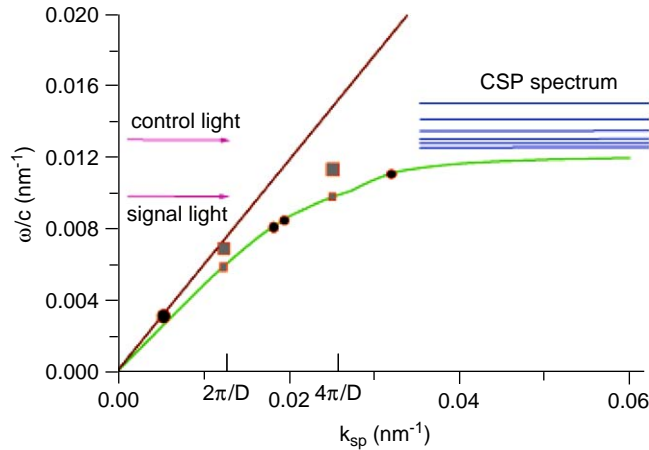


Fig. 81. The SPP mode spectrum of the gold nanostructure embedded in the polymer and schematic of the cylindrical surface plasmon (CSP) states in the polymer-filled channels at large wave vectors. (squares) SPP resonances excited with p-polarized light, (circles) SPP resonances excited with s-polarized light at the angle of incidence of 45° , exciting light is incident along the hole rows. Positions of the signal and control light frequencies are shown by the arrows.

asymptotically approaching the surface plasmon frequency ($\epsilon_m = -\epsilon$) from the high frequency side (Section 2). For polymer-filled channels these modes overlap the frequency of the control blue light for large quantum numbers $s \gg 1$ (Fig. 81). The interaction between channels in an array can additionally broaden these resonances, leading to minibands in the cylindrical surface plasmon spectrum [235]. Thus, a quasi-continuous spectrum of the states related to the cylindrical surface plasmons can be expected in the spectral range of the control light. For infinitely long cylindrical channels the spectrum of wavevectors along the cylinder axis h_z is continuous. Real (nonradiative) surface modes cannot be excited directly by light, but at the frequency corresponding to the control light the very long wavevector SPP can be excited on the periodically perforated polymer–metal interface, which then can be coupled to cylindrical surface plasmons.

In the absence of control light, the transmission of red light takes place via resonant tunneling through the states of SPP Bloch waves on the polymer–metal interface. Changing the polarization of the incident light results in shifts of the SPP resonances and, hence, to a variation of the transmission. Being confined to the polymer-coated interface, the SPP modes responsible for this transmission are very sensitive to the dielectric constant of the polymer, since changes of the dielectric constant modify the SPP resonant conditions and the transmission coefficient. Control (blue) light coupled to cylindrical surface plasmons either via their radiative part or via surface polaritons results in local changes of the dielectric constant of the polymer due to the third-order Kerr nonlinearity. The local electromagnetic field is strongly enhanced in and around the channel due to cylindrical surface plasmon excitation because of the small volume of these surface modes ($E_L \sim 1/d$).

The polarization properties of the polymer molecules themselves may play a significant role in the observed increase of the p to s ratio in the modulated transmission of red light. The third-order nonlinearity of 3BCMUPolydiacetylene is contributed mainly by the π -electrons in the backbone of the polymer [234]. Each straight segment of polymer backbone could be treated as an identical one-dimensional rod-like

chromophore. At the microscopic level the $\chi^{(3)}$ tensor of 3BCMU is dominated by only one component: $\chi_{ssss}^{(3)}$, where s is the direction of the polymer chain. This fact has been verified in measurements of different components of the macroscopic $\chi^{(3)}$ tensor of spin-coated thin polymer 3BCMU films, where polymer backbones have random in-plane orientations, determined by a flat substrate [234]. In our case, the preferential direction for the long polymer backbones inside narrow cylindrical channels ought to be the direction along the channel. The nonlinear optical mixing of interest is determined by the components $D_l(\omega_2)$ of the optical field inside the channels at the wavelength of signal light λ_2 :

$$D_l(\omega_2) = \chi_{ijkl}^{(3)} E_i(\omega_1) E_j(\omega_1) E_k(\omega_2) , \quad (6.70)$$

where ω_1 and ω_2 are the frequencies of the control and signal light, respectively. If the $zzzz$ -components (along the channel direction) of $\chi_{ijkl}^{(3)}$ dominate the third-order susceptibility, the modulated red light will be strongly p polarized. At the same time, at the frequency of the control light ω_1 the electromagnetic field in the cylindrical channels is dominated by the plasmon modes with large wavevectors along the channel, for which the optical field oscillations have a significant longitudinal component that is along the channel direction. Thus, one should expect enhanced nonlinear optical mixing to occur while the electromagnetic field is traveling through small cylindrical holes.

This nonlinear effect in optical transmission of periodic hole array can find numerous applications in active photonic devices and in a new class of “gated” photon tunneling devices for massive parallel all-optical signal processing operating at very low control light intensities.

7. Nonlinear optics of surface plasmon polaritons

The role of surface plasmons in the surface enhanced nonlinear optical phenomena on a surface has been studied for a long time [1,5]. Such phenomena are related to the electromagnetic field enhancement that occurs on a metal surface. As has been discussed above, the SPP electromagnetic field that is confined to a surface can significantly exceed the field of the incident light. In addition to this, further enhancement can take place if SPP localization (Anderson type) occurs, as has been shown in Section 2, or localized surface plasmons are excited at surface features. The field enhancement effects are especially important for nonlinear optical phenomena that depend on the excitation intensity in a nonlinear manner. The most important examples of these are surface enhanced Raman scattering, second-harmonic generation (SHG), coherent anti-Stokes Raman scattering, Kerr nonlinearity, etc.

The SPP-related enhancement of nonlinear optical processes occurs when the electromagnetic field of the exciting or generated signal (or both) are coupled into surface plasmon modes on a metal–dielectric interface. In the case of the coupling of the excitation field, the effect occurs due to a strong SPP field close to the interface. The enhancement of the generated field occurs in a similar way. Nevertheless, in this case the field needs to be decoupled into light to observe it in the far-field. The decoupling can occur on surface defects, or special arrangements such as a diffraction grating or a prism can be employed. The nonlinear excitation of SPPs can take place due to nonlinear mixing of light at a metal surface under ordinary illumination. Generation of light due to nonlinear processes involving SPPs has been also studied as well as nonlinear optical processes due to the mixing of SPPs and light [236].

An example of such processes, SHG enhancement due to SPP, can be observed in the case of SHG excitation in the Kretschmann geometry so that the fundamental light is coupled into SPP on a metal surface [236]. If, in addition, the surface possesses roughness that can significantly modify the SPP behavior, not

only the SHG intensity but also its angular distribution can be significantly modified [237,238]. In the case of the surfaces at which SPP backscattering can occur, the related SH wave can be generated in the direction exactly perpendicular to the surface due to the momentum conservation of two counterpropagating SPPs which produce a SH photon [121]. SPP localization effects play an important role in diffuse SH generation and can arise from the SPP localization at both the fundamental and SH frequencies [122].

Even under ordinary illumination, SPP can be excited on a rough surface directly, or via excitation of localized plasmons following their decay into SPP. Thus, in many cases the SPP related effects are important in all SHG studies at rough surfaces. Depending on the defect structure of rough or periodic surfaces and the details of the SHG excitation such as polarization and angle of incidence, different SHG mechanisms have been observed in investigations of near-field distributions of the SH over a surface related either to SPP Anderson localization (SPP excitation and interaction with an ensemble of surface defects) or to the excitation of localized surface plasmons on individual surface features [239–244].

7.1. *Second harmonic generation in reflection from rough metal surfaces*

The second harmonic generation of light in reflection from a metal surface [245,246] has attracted considerable interest since the experimental discovery by Chen et al. [247] that SHG at a silver–air interface can be enhanced by a factor of $\sim 10^4$ by roughening the surface. The mechanism responsible for this enhancement was assumed to be the roughness-induced excitation of a surface plasmon polariton by the incident light, which in turn enhances the nonlinear source term in the Maxwell equations responsible for the reflected light at twice the frequency of the incident light.

A quantitative theory of this effect was difficult to construct because a randomly rough surface profile is difficult to treat except in the small roughness limit. Consequently, in the earliest theories of the second harmonic generation of light in reflection from a randomly rough metal surface a deterministic periodic surface profile provided by a one-dimensional grating was assumed as a model for a randomly rough surface. Such a surface also allows the incident light to excite surface plasmon polaritons and thereby to enhance the intensity of the second harmonic generation in reflection from a sinusoidal metal grating surface.

Thus, Agarwal and Jha [248] using small-amplitude perturbation theory, in which the scattered electromagnetic field was expanded in powers of the surface profile function through terms of second order, calculated the second harmonic generation from a grating, and discussed the role of surface plasmon polaritons in its enhancement. However, in their calculations Agarwal and Jha assumed that the tangential component of the electric field at the harmonic frequency is continuous across the vacuum–metal interface, whereas, as we will see below, it is in fact discontinuous across this interface.

In a subsequent paper Farias and Maradudin [249] presented a nonperturbative theory of second-harmonic generation in reflection from a grating. In their approach the single nonzero component of the magnetic field in the system satisfied a Helmholtz equation both at the fundamental frequency ω and at the harmonic frequency 2ω . The nonlinear polarization entered the problem only through the boundary conditions on the magnetic field at the harmonic frequency. The method of reduced Rayleigh equations [170,250] was used in obtaining the magnetic field at the fundamental frequency and at the harmonic frequency. An advantage of this approach is that only the magnetic fields directly required are calculated, which reduces the size of the matrix equations that need to be solved by a factor of two. Thus, in the determination of the field at the fundamental frequency only the field in the metal is calculated, because it is only this field that enters the boundary conditions at the harmonic frequency. Similarly, in determining

the field at the harmonic frequency only the field in the vacuum was calculated, because it is only this field that determines the intensity of second harmonic generation. In the vicinity of the frequency at which the incident electromagnetic field couples to surface plasmon polaritons on a grating, an enhancement of the second harmonic intensity of the order of 10^4 over its value for a flat surface was obtained for a sinusoidal grating on silver.

In an alternative approach to the study of second harmonic generation in reflection from a metallic grating by Reinisch and Neviere [251], the surface profile function also was not treated perturbatively: these authors used the differential method [252], in which the Maxwell equations are integrated numerically across the selvedge region, viz. the region between the minimum and maximum values of the surface profile function. With this method it is possible to study more strongly corrugated surfaces than is possible with the method of the reduced Rayleigh equation. However, in the calculations of Reinisch and Neviere the nonlinear polarization appears both as a source term in the Maxwell equations for the second harmonic fields and in the boundary conditions they satisfy, in contrast with the approach of Farias and Maradudin [249], where it appears only in the boundary conditions for the second harmonic fields.

Interest in second harmonic generation in reflection from rough metal surfaces has shifted in recent years from second harmonic generation at one-dimensional periodically corrugated surfaces to second harmonic generation at one-dimensional randomly rough metal surfaces. This is due to growing interest in the broader area of interference effects occurring in the multiple scattering of electromagnetic waves from, and their transmission through, randomly rough metal surfaces, and the related enhanced backscattering [112], enhanced transmission [203], and satellite peak phenomena [253]. It has been expected that nonlinear optical interactions at a randomly rough metal surface should also give rise to new features in the interference of multiply scattered electromagnetic waves.

In the first theoretical study of the phenomenon McGurn et al. [122] predicted on the basis of a perturbative calculation that enhanced second harmonic generation of light at a weakly rough, clean, metal surface should occur not only in the retroreflection direction, but also in the direction normal to the mean surface. The peak predicted in the direction normal to the mean surface was attributed to interference effects in the multiple scattering of surface plasmon polaritons of the fundamental frequency, excited by the incident light through the roughness of the surface, while the peak predicted in the retroreflection direction was attributed to interference effects in the multiple scattering of surface plasmon polaritons of the harmonic frequency.

This work stimulated several experimental studies of second harmonic generation in the scattering of light from random metal surfaces [237,254–258], and enhanced second harmonic generation peaks in the direction normal to the mean surface and in the retroreflection direction were observed [237,254–256,258]. In these experiments, however, the scattering system was not a clean random interface between vacuum and a semi-infinite metal: to amplify the second harmonic signal a random interface with a dielectric or vacuum of a thin metal film deposited on the planar base of a dielectric prism through which the light was incident (the Kretschmann attenuated total reflection geometry [259]) was used.

Experimental studies of multiple-scattering effects in the second harmonic generation of light scattered from a clean, one-dimensional vacuum–metal interface were carried out by O'Donnell and his colleagues [238,260,261]. These experiments were carried out with characterized, specially prepared one-dimensional silver surfaces, and the scattering data were absolutely normalized, which permitted quantitative comparisons with theoretical work. It was found that for both weakly [238,260] and strongly [261] rough surfaces a dip is present in the retroreflection direction in the angular dependence of the

intensity of the scattered second harmonic light instead of the peak that occurs in scattering at the fundamental frequency. This result was in agreement with the results of rigorous numerical calculation results of second harmonic generation from such surfaces carried out by Leyva-Lucero et al. [262,263]. However, no peak or dip in the direction normal to the mean surface was observed in the experiments of Refs. [238,260,261].

In this section we outline theoretical approaches to the second harmonic generation of light in reflection from clean, one-dimensional, periodically corrugated and randomly rough metal surfaces, since these are the only kinds of surfaces for which multiple-scattering effects can be calculated readily at the present time, and for which experimental results exist against which theoretical results can be compared. Theoretical treatments of second harmonic generation from randomly rough metal films in the Kretschmann ATR geometry of the type employed in the experiments of Refs. [256–258] can be found in Refs. [264,265], and the reader is referred to these papers for the details of the calculations. Second harmonic generation from nonlinear thin films, both of whose surfaces are one-dimensional randomly rough surfaces, has been studied theoretically by Enoch [266]. In this work the nonlinearity enters the Maxwell equations at the harmonic frequency as source terms rather than in the boundary conditions.

The physical system studied here consists of vacuum in the region $x_3 > \zeta(x_1)$ (region I), and a nonlinear medium (metal) in the region $x_3 < \zeta(x_1)$ (region II). A p-polarized electromagnetic field of frequency ω is incident on the vacuum–metal interface $x_3 = \zeta(x_1)$ from the vacuum side. The plane of incidence is assumed to be the x_1x_3 -plane, so that it is perpendicular to the generators of the interface. The surface profile function $\zeta(x_1)$ is assumed to be a single-valued function of x_1 that is differentiable as many times as is necessary. In Section 7.1.1 it will be assumed to be a periodic function of x_1 , while in Section 7.1.2 it will be assumed to constitute a zero-mean stationary, Gaussian random process.

The analysis in the latter two sections will be based on the well known fact that homogeneous and isotropic metals possess inversion symmetry, so that the dipole contribution to the bulk nonlinear polarization is absent ($\chi^{(2)} = 0$). The presence of the surface breaks the inversion symmetry, and since both the electromagnetic fields and material constants vary rapidly at the surface, their gradients give rise to the optical nonlinearity of the surface. The second harmonic radiation we are interested in is generated in a vacuum–metal interface layer that has a finite thickness on the microscopic scale. Consequently, the nonlinear polarization can be taken into account though the boundary conditions for the second harmonic fields.

In the following analysis we will neglect the small contribution to the nonlinearity coming from the bulk of the metal and the possible anisotropy of the material constants. In this formulation of the problem both the fundamental and harmonic fields satisfy Helmholtz equations above and below the interface.

In p-polarization the magnetic vector in our system has the form

$$\mathbf{H}(\mathbf{x}; t) = (0, H_2(x_1, x_3; t), 0) , \quad (7.1)$$

while the electric vector is given by

$$\mathbf{E}(\mathbf{x}; t) = (E_1(x_1, x_3; t), 0, E_3(x_1, x_3; t)) . \quad (7.2)$$

Since $\mathbf{H}(\mathbf{x}; t)$ has only a single nonzero component, it is convenient to work with it.

In order to obtain an equation for the harmonic component of the magnetic field, we expand all fields in the usual way [245],

$$H_2(x_1, x_3; t) = H_2(x_1, x_3|\omega) \exp(-i\omega t) + H_2(x_1, x_3|2\omega) \exp(-i2\omega t) + \dots \quad (7.3a)$$

$$E_1(x_1, x_3; t) = E_1(x_1, x_3|\omega) \exp(-i\omega t) + E_1(x_1, x_3|2\omega) \exp(-i2\omega t) + \dots \quad (7.3b)$$

$$E_3(x_1, x_3; t) = E_3(x_1, x_3|\omega) \exp(-i\omega t) + E_3(x_1, x_3|2\omega) \exp(-i2\omega t) + \dots \quad (7.3c)$$

The equations for the amplitude of the magnetic field at the fundamental frequency ω are then

$$\left(\frac{\partial}{\partial x_1^2} + \frac{\partial^2}{\partial x_3^2} + \frac{\omega^2}{c^2} \right) H_2^{(I)}(x_1, x_3|\omega) = 0, \quad x_3 > \zeta(x_1), \quad (7.4a)$$

$$\left(\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_3^2} + \epsilon(\omega) \frac{\omega^2}{c^2} \right) H_2^{(II)}(x_1, x_3|\omega) = 0, \quad x_3 < \zeta(x_1), \quad (7.4b)$$

where $\epsilon(\omega)$ is the frequency-dependent dielectric function of the metal, while the equations for the amplitude of the magnetic field at the harmonic frequency 2ω become

$$\left(\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_3^2} + 4 \frac{\omega^2}{c^2} \right) H_2^{(I)}(x_1, x_3|2\omega) = 0, \quad x_3 > \zeta(x_1), \quad (7.5a)$$

$$\left(\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_3^2} + 4\epsilon(2\omega) \frac{\omega^2}{c^2} \right) H_2^{(II)}(x_1, x_3|2\omega) = 0, \quad x_3 < \zeta(x_1). \quad (7.5b)$$

The boundary conditions at the surface $x_3 = \zeta(x_1)$ satisfied by the magnetic field at the fundamental frequency express the continuity of the tangential components of the magnetic and electric fields across this interface:

$$H_2^{(I)}(x_1, x_3|\omega) = H_2^{(II)}(x_1, x_3|\omega), \quad (7.6a)$$

$$\frac{\partial}{\partial n} H_2^{(I)}(x_1, x_3|\omega) = \frac{1}{\epsilon(\omega)} \frac{\partial}{\partial n} H_2^{(II)}(x_1, x_3|\omega), \quad (7.6b)$$

where $\partial/\partial n$ is the derivative along the normal to the interface directed from the metal into the vacuum, and is given by

$$\frac{\partial}{\partial n} = \{1 + [\zeta'(x_1)]^2\}^{-1/2} \left(-\zeta'(x_1) \frac{\partial}{\partial x_1} + \frac{\partial}{\partial x_3} \right). \quad (7.7)$$

The nonlinear boundary conditions for the harmonic fields are obtained by integrating Maxwell's equations for them across the interface layer, and then passing to the limit of a vanishing thickness of the layer. In carrying out this calculation it is convenient to introduce a local coordinate system (x, y, z) with its origin at each point of the surface $x_3 = \zeta(x_1)$ that is defined by the unit vectors $\{\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}\}$, where $\hat{\mathbf{x}} = (1, 0, \zeta'(x_1))/\phi(x_1)$, $\hat{\mathbf{y}} = \hat{\mathbf{x}}_2$, and $\hat{\mathbf{z}} = (-\zeta'(x_1), 0, 1)/\phi(x_1)$ with $\phi(x_1) = [1 + (\zeta'(x_1))^2]^{1/2}$. $\hat{\mathbf{x}}$ and $\hat{\mathbf{z}}$ are unit vectors tangent and normal to the interface in the plane perpendicular to its generators. On integrating the tangential component of the equation $\nabla \times \mathbf{H} = -(2i\omega/c)\mathbf{D}$, namely $\partial \mathbf{H}_t / \partial z = (2i\omega/c)$

$\hat{\mathbf{z}} \times \mathbf{D}_t + \nabla_t H_z$, along a contour around the vacuum–metal interface, we obtain with the use of the relation $\mathbf{D} = \epsilon(2\omega, z)\mathbf{E} + 4\pi\mathbf{P}^{\text{NL}}$, the first nonlinear boundary condition

$$\mathbf{H}_t^{(\text{I})}(x|2\omega) - \mathbf{H}_t^{(\text{II})}(x|2\omega) = \frac{2i\omega}{c}\hat{\mathbf{z}} \times \mathbf{P}_t^s(x|2\omega), \quad (7.8)$$

where $\mathbf{P}_t^s(x|2\omega)$ is the tangential component of the nonlinear surface polarization, and is given by

$$\mathbf{P}_t^s(x|2\omega) = \lim_{\eta \rightarrow 0} \int_{-\eta}^{\eta} dz [4\pi\mathbf{P}_t^{\text{NL}}(x, z|2\omega)], \quad (7.9)$$

while $\mathbf{P}_t^{\text{NL}}(x, z|2\omega)$ is the tangential component of the nonlinear polarization vector. The subscript t denotes quantities that are tangent to the surface, and the subscript z denotes quantities that are normal to the surface.

The components of the second-order surface susceptibility tensor $\chi_{ijk}^{(2)}$ are the constants of proportionality that relate the amplitudes of the components of the nonlinear surface polarization to the amplitudes of the electric field at the surface at the fundamental frequency, according to

$$P_t^s(x|2\omega) = \sum_{j,k} \chi_{ijk}^{(2)} E_j^{(\text{I})}(x|\omega) E_k^{(\text{I})}(x|\omega). \quad (7.10)$$

Since we are dealing with a centrosymmetric medium, it is found that the tangential component of the nonlinear surface polarization is

$$\mathbf{P}_t^s(x|2\omega) = \chi_{ttz} E_t^{(\text{I})}(x|\omega) E_z^{(\text{I})}(x|\omega). \quad (7.11)$$

The boundary condition (7.8) can then be written as

$$\mathbf{H}_t^{(\text{I})}(x|2\omega) - \mathbf{H}_t^{(\text{II})}(x|2\omega) = \frac{2\omega}{c} \chi_{ttz} \hat{\mathbf{z}} \times \mathbf{E}_t^{(\text{I})}(x|\omega) E_z^{(\text{I})}(x|\omega). \quad (7.12)$$

In a similar fashion, by integrating the tangential component of the equation $\nabla \times \mathbf{E} = (2i\omega/c)\mathbf{H}$, namely $\partial \mathbf{E}_t / \partial z = -(2i\omega/c)\hat{\mathbf{z}} \times \mathbf{H}_t + \nabla_t E_z$, along a contour around the vacuum–metal interface, we obtain the second nonlinear boundary condition,

$$\mathbf{E}_t^{(\text{I})}(x|2\omega) - \mathbf{E}_t^{(\text{II})}(x|2\omega) = -\frac{\partial}{\partial \mathbf{t}} P_z^s(x|2\omega), \quad (7.13)$$

where $\partial/\partial \mathbf{t} = (\partial/\partial x, \partial/\partial y, 0)$, and where the normal component of the surface nonlinear polarization $P_z^s(x|2\omega)$ is defined by

$$P_z^s(x|2\omega) = \lim_{\eta \rightarrow 0} \int_{-\eta}^{\eta} dz \left[\frac{4\pi P_z^{\text{NL}}(x, z|2\omega)}{\epsilon(2\omega, z)} \right], \quad (7.14)$$

with $\epsilon(2\omega, z > \eta) = 1$ in vacuum, and $\epsilon(2\omega, z < -\eta) = \epsilon(2\omega)$ in the metal.

The use of Eq. (7.10) and the symmetry properties of the medium yields the result that

$$P_z^s(x|2\omega) = \chi_{zzz} \left[E_z^{(\text{I})}(x|\omega) \right]^2 + \chi_{ztt} \mathbf{E}_t^{(\text{I})}(x|\omega) \cdot \mathbf{E}_t^{(\text{I})}(x|\omega). \quad (7.15)$$

The second nonlinear boundary condition (7.13) then takes the form

$$\mathbf{E}_t^{(I)}(x|2\omega) - \mathbf{E}_t^{(II)}(x|2\omega) = -\frac{\partial}{\partial \mathbf{t}} \left\{ \chi_{zzz} \left[E_z^{(I)}(x|\omega) \right]^2 + \chi_{ztt} \mathbf{E}_t^{(I)}(x|\omega) \cdot \mathbf{E}^{(I)}(x|\omega) \right\} . \quad (7.16)$$

The solution of the linear problem and the nonlinear surface susceptibilities both enter the nonlinear boundary conditions required for determining the second harmonic field. Several models of the nonlinear surface polarization have been proposed in the literature [245,267–273] in the context of the second harmonic generation of light in reflection from metal surfaces, leading to different nonlinear constants and to different notations. The μ parameters of Agranovich and Darmanyan [270] may be identified with the surface nonlinear susceptibilities of other authors [274], and are related to the dimensionless frequency-dependent parameters $a(\omega)$ and $b(\omega)$ of Rudnick and Stern [267]. One finds that [275]

$$\mu_1 = \chi_{zzz} = -\frac{1}{64\pi^2 n_e e} \left(\frac{\epsilon(\omega) - 1}{\epsilon(\omega)} \right)^2 a(\omega) , \quad (7.17a)$$

$$\mu_2 = \chi_{ztt} , \quad (7.17b)$$

$$\mu_3 = \chi_{ttz} = -\frac{1}{32\pi^2 n_e e} \left(\frac{\epsilon(\omega) - 1}{\epsilon(\omega)} \right)^2 b(\omega) , \quad (7.17c)$$

where e is the magnitude of the electron charge and n_e is the bulk electron number density. The nonlinear constants obtained on the basis of the free electron model [245,271] are often employed in such calculations because they are given by simple expressions

$$\chi_{zzz} = -\frac{2}{3} \beta \left[\frac{(\epsilon(\omega) - 1)(\epsilon(\omega) - 3)}{2\epsilon^2(\omega)} - \frac{2}{3} \ln \left(\frac{\epsilon(\omega)}{\epsilon(2\omega)} \right) \right] , \quad (7.18a)$$

$$\chi_{ztt} = 0 , \quad (7.18b)$$

$$\chi_{ttz} = \beta \left(\frac{\epsilon(\omega) - 1}{\epsilon(\omega)} \right) , \quad (7.18c)$$

where $\beta = e/(8\pi m\omega^2)$ and m is the electron mass. These expressions for the nonlinear constants coincide with those obtained by Mendoza and Mochán [274].

From the boundary conditions (7.12) and (7.16) it can be shown that, in general, the second harmonic field contains both s- and p-polarized components. However, in our case of a one-dimensional rough surface and a plane of incidence that is perpendicular to the generators of the surface, pure s- and p-polarized incident fields generate only p-polarized light at the harmonic frequency. Consequently, if anisotropy of the scattering medium and contributions from the bulk are neglected, an s-polarized second harmonic field can be generated only by a mixture of s- and p-polarized incident fields. The generation of the harmonic field is known to be more efficient in the case of a p-polarized incident field [276], and it is for this reason that we have assumed this polarization for this field.

The fields entering the boundary conditions (7.12) and (7.16) are written in the local coordinate system (x, y, z) . On returning to the laboratory coordinate system (x_1, x_2, x_3) , and expressing these fields in terms of the single nonzero component of the magnetic field in the system, the nonlinear boundary conditions

finally become [275]

$$H^{(I)}(x_1|2\omega) - H^{(II)}(x_1|2\omega) = \frac{2ic}{\omega} \frac{\chi_{ttz}}{\phi^2(x_1)} L^{(I)}(x_1|\omega) \frac{d}{dx_1} H^{(I)}(x_1|\omega) \equiv A(x_1|2\omega) \quad (7.19a)$$

$$\begin{aligned} &L^{(I)}(x_1|2\omega) - \frac{1}{\epsilon(2\omega)} L^{(II)}(x_1|2\omega) \\ &= \frac{2ic}{\omega} \frac{d}{dx_1} \left\{ \frac{1}{\phi^2(x_1)} \left[\chi_{zzz} \left(\frac{d}{dx_1} H^{(I)}(x_1|\omega) \right)^2 + \chi_{ztt} \left(L^{(I)}(x_1|\omega) \right)^2 \right] \right\} \equiv B(x_1|2\omega) , \end{aligned} \quad (7.19b)$$

where we have introduced the source functions

$$H^{(I,II)}(x_1|\Omega) = H_2^{(I,II)}(x_1, x_3|\Omega)|_{x_3=\zeta(x_1)} \quad (7.20a)$$

$$L^{(I,II)}(x_1|\Omega) = \frac{\partial}{\partial N} H_2^{(I,II)}(x_1, x_3|\Omega) \Big|_{x_3=\zeta(x_1)} , \quad (7.20b)$$

Ω is either ω or 2ω , $\phi(x_1) = \{1 + [\zeta'(x_1)]^2\}^{1/2}$, and

$$\frac{\partial}{\partial N} = -\zeta'(x_1) \frac{\partial}{\partial x_1} + \frac{\partial}{\partial x_3} \quad (7.21)$$

is the nonnormalized derivative along the normal to the surface at each point.

We now turn to an application of the results of this section in turn to the second harmonic generation of light in reflection from periodic and random one-dimensional metal surfaces.

7.1.1. Second harmonic generation from metallic diffraction gratings

In the case that the surface profile function $\zeta(x_1)$ is a periodic function of x_1 with a period a , $\zeta(x_1 + a) = \zeta(x_1)$, we are concerned with second harmonic generation in reflection from a classical metallic grating. In this case the solution of Eqs. (7.4) that satisfy the boundary conditions at infinity can be written in forms that also satisfy the Bloch–Floquet condition in the forms

$$\begin{aligned} H_2^{(I)}(x_1, x_3|\omega) &= \exp[ikx_1 - i\alpha_0(k, \omega)x_3] \\ &+ \sum_{p=-\infty}^{\infty} A_p(k) \exp[ik_p x_1 + i\alpha_0(k_p, \omega)x_3], \quad x_3 > \zeta(x_1)_{\max} , \end{aligned} \quad (7.22a)$$

$$H_2^{(II)}(x_1, x_3|\omega) = \sum_{p=-\infty}^{\infty} B_p(k) \exp[ik_p x_1 - i\alpha(k_p, \omega)x_3], \quad x_3 < \zeta(x_1)_{\min} , \quad (7.22b)$$

where

$$\alpha_0(k_p, \omega) = [(\omega/c)^2 - k_p^2]^{1/2}, \quad |k_p| < \omega/c \quad (7.23a)$$

$$= i[k_p^2 - (\omega/c)^2]^{1/2}, \quad |k_p| > \omega/c , \quad (7.23b)$$

$$\alpha(k_p, \omega) = [\epsilon(\omega)(\omega/c)^2 - k_p^2]^{1/2} \quad \text{Re } \alpha(k_p) > 0, \quad \text{Im } \alpha(k_p) > 0, \quad (7.24)$$

$$k_p = k + (2\pi p/a), \quad p = 0, \pm 1, \pm 2, \dots \quad (7.25)$$

and

$$k = (\omega/c) \sin \theta_0,$$

with θ_0 the angle of incidence.

The Rayleigh hypothesis [277] is the assumption that the solutions given by Eqs. (7.22) can be continued in to the interface $x_3 = \zeta(x_1)$ itself and used to satisfy the boundary conditions (7.6). If we make this assumption, we obtain a pair of coupled equations for determining the $\{A_p(k)\}$ and $\{B_p(k)\}$:

$$\begin{aligned} \sum_{p=-\infty}^{\infty} \{-A_p(k) \exp[ik_p x_1 + i\alpha_0(k_p, \omega)\zeta(x_1)] + B_p(k) \exp[ik_p x_1 - i\alpha(k_p, \omega)\zeta(x_1)]\} \\ = \exp[ik x_1 - i\alpha_0(k)\zeta(x_1)] \end{aligned} \quad (7.26a)$$

$$\begin{aligned} \sum_{p=-\infty}^{\infty} \{-A_p(k)[- \zeta'(x_1)ik_p + i\alpha_0(k_p, \omega)] \exp[ik_p x_1 + i\alpha_0(k_p, \omega)\zeta(x_1)]\} \\ + \frac{1}{\epsilon(\omega)} B_p(k)[- \zeta'(x_1)ik_p - i\alpha(k_p, \omega)] \exp[ik_p x_1 - i\alpha(k_p, \omega)\zeta(x_1)] \\ = -[\zeta'(x_1)ik + i\alpha_0(k)] \exp[ik x_1 - i\alpha_0(k, \omega)\zeta(x_1)]. \end{aligned} \quad (7.26b)$$

Simpler expressions are obtained for the nonlinear boundary conditions (7.19) in the present case if they are expressed in terms of the fundamental field in the metal (because the latter has no incident field). Therefore the coefficients $\{A_p(k)\}$ are eliminated from Eqs. (7.26) to yield an equation for the $\{B_p(k)\}$ alone. This is done by first multiplying Eq. (7.26a) by $[-\zeta'(x)1]ik_r - i\alpha(k_r, \omega)] \exp[-ik_r x_1 + i\alpha_0(k_r, \omega)\zeta(x_1)]$ and Eq. (7.26b) by $\exp[-ik_r x_1 + i\alpha_0(k_r, \omega)\zeta(x_1)]$; integrating each of the resulting equations with respect to x_1 over the interval $(-a/2, a/2)$; and finally adding the two equations so obtained. The result can be written in the form [170,250]

$$\begin{aligned} \sum_{p=-\infty}^{\infty} \frac{\alpha_0(k_r, \omega)\alpha(k_p, \omega) + k_r k_p}{\alpha_0(k_r, \omega) - \alpha(k_p, \omega)} \mathcal{J}_{r-p}(\alpha_0(k_r, \omega) - \alpha(k_p, \omega)) B_p(k) \\ = \delta_{r0} \frac{2\epsilon(\omega)\alpha_0(k, \omega)}{1 - \epsilon(\omega)}, \end{aligned} \quad (7.27)$$

where

$$\mathcal{J}_m(\gamma) = \frac{1}{a} \int_{-a/2}^{a/2} dx_1 \exp[-i(2\pi m/a)x_1 + i\gamma\zeta(x_1)]. \quad (7.28)$$

The solutions of Eqs. (7.5) that satisfy the boundary conditions at infinity and the Bloch–Floquet condition can be written as

$$H_2^{(I)}(x_1, x_3|2\omega) = \sum_{p=-\infty}^{\infty} A_p(2k) \exp[ik_p^{(2)}x_1 + i\alpha_0(k_p^{(2)}, 2\omega)x_3], \quad x_3 > \zeta(x_1)_{\max}, \quad (7.29a)$$

$$H_2^{(II)}(x_1, x_3|2\omega) = \sum_{p=-\infty}^{\infty} B_p(2k) \exp[ik_p^{(2)}x_1 - i\alpha(k_p^{(2)}, 2\omega)x_3], \quad x_3 < \zeta(x_1)_{\min}, \quad (7.29b)$$

where

$$k_p^{(2)} = 2k + (2\pi p/a), \quad p = 0, \pm 1, \pm 2, \dots \quad (7.30)$$

$$\alpha_0(k_p^{(2)}, 2\omega) = [4(\omega/c)^2 - (k_p^{(2)})^2]^{1/2}, \quad |k_p^{(2)}| < 2\omega/c \quad (7.31a)$$

$$= i[(k_p^{(2)})^2 - 4(\omega/c)^2]^{1/2}, \quad |k_p^{(2)}| > 2\omega/c, \quad (7.31b)$$

$$\alpha(k_p^{(2)}, 2\omega) = [4\epsilon(2\omega)(\omega/c)^2 - (k_p^{(2)})^2]^{1/2}, \quad \text{Re } \alpha(k_p^{(2)}, 2\omega) > 0, \quad \text{Im } \alpha(k_p^{(2)}, 2\omega) > 0. \quad (7.32)$$

When the solutions (7.29) are used to satisfy the nonlinear boundary conditions (7.19), we obtain the pair of equations

$$\begin{aligned} & \sum_{p=-\infty}^{\infty} \{A_p(2k) \exp[i\alpha_0(k_p^{(2)}, 2\omega)\zeta(x_1)] - B_p(2k) \exp[-i\alpha(k_p^{(2)}, 2\omega)\zeta(x_1)]\} \exp(ik_p^{(2)}x_1) \\ & = A(x_1|2\omega), \end{aligned} \quad (7.33a)$$

$$\begin{aligned} & \sum_{p=-\infty}^{\infty} \{A_p(2k)[- \zeta'(x_1)k_p^{(2)} + \alpha_0(k_p^{(2)}, 2\omega)] \exp[i\alpha_0(k_p^{(2)}, 2\omega)\zeta(x_1)] \\ & - \frac{B_p(2k)}{\epsilon(2\omega)} [- \zeta'(x_1)k_p^{(2)} - \alpha(k_p^{(2)}, 2\omega)] \exp[-i\alpha(k_p^{(2)}, 2\omega)\zeta(x_1)]\} \exp(ik_p^{(2)}x_1) \\ & = -iB(x_1|2\omega). \end{aligned} \quad (7.33b)$$

The coefficients $\{A_p(2k)\}$ give the amplitudes of the field at the harmonic frequency that is diffracted back into the vacuum. The coefficients $\{B_p(2k)\}$ can be eliminated from the pair of equations (7.29) to yield an equation for the $\{A_p(2k)\}$ alone. We do this by first multiplying Eq. (7.33a) by $[-\zeta'(x_1)k_r^{(2)} + \alpha(k_r^{(2)}, 2\omega)] \exp[-ik_r^{(2)}x_1 - i\alpha(k_r^{(2)}, 2\omega)\zeta(x_1)]$, then multiplying Eq. (7.33b) by $\epsilon(2\omega) \exp[-ik_r^{(2)}x_1 - i\alpha(k_r^{(2)}, 2\omega)\zeta(x_1)]$, adding the resulting pair of equations, and finally integrating the result with respect to x_1 over the interval $(-a/2, a/2)$. In this way we obtain

$$\begin{aligned} & \sum_{p=-\infty}^{\infty} \frac{\alpha(k_r^{(2)}, 2\omega)\alpha_0(k_p^{(2)}, 2\omega) + k_r^{(2)}k_p^{(2)}}{\alpha(k_r^{(2)}, 2\omega) - \alpha_0(k_p^{(2)}, 2\omega)} \mathcal{J}_{r-p}(\alpha_0(k_p^{(2)}, 2\omega) - \alpha(k_r^{(2)}, 2\omega))A_p(2k) \\ & = \frac{1}{\epsilon(2\omega) - 1} \int_{-a/2}^{a/2} dx_1 \{A(x_1|2\omega)[- \zeta'(x_1)k_r^{(2)} + \alpha(k_r^{(2)}, 2\omega)] - i\epsilon(2\omega)B(x_1|2\omega)\} \\ & \quad \times \exp[-ik_r^{(2)}x_1 - i\alpha(k_r^{(2)}, 2\omega)\zeta(x_1)]. \end{aligned} \quad (7.34)$$

By the use of the boundary conditions (7.6) and the definitions (7.20), the source functions $A(x_1|2\omega)$ and $B(x_1|2\omega)$ are now given by

$$\begin{aligned}
 A(x_1|2\omega) &= \frac{2ic}{\omega\epsilon(\omega)} \frac{\chi_{ttz}}{\phi^2(x_1)} L^{(II)}(x_1|\omega) \frac{d}{dx_1} H^{(II)}(x_1|\omega) \\
 &= \frac{2ic}{\omega\epsilon(\omega)} \frac{\chi_{ttz}}{\phi^2(x_1)} \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} B_m(k) B_n(k) [-\zeta'(x_1) ik_m - i\alpha(k_m, \omega)] \\
 &\quad \times [ik_n - i\alpha(k_n, \omega)\zeta'(x_1)] \exp[i(k_m + k_n)x_1 - i(\alpha(k_m, \omega) + \alpha(k_n, \omega))\zeta(x_1)] \quad (7.35a)
 \end{aligned}$$

$$\begin{aligned}
 B(x_1|2\omega) &= \frac{2ic}{\omega} \frac{d}{dx_1} \left\{ \frac{1}{\phi^2(x_1)} \left[\chi_{zzz} \left(\frac{d}{dx_1} H^{(II)}(x_1|\omega) \right)^2 + \frac{\chi_{ztt}}{\epsilon^2(\omega)} \left(L^{(II)}(x_1|\omega) \right)^2 \right] \right\} \\
 &= \frac{2ic}{\omega} \frac{d}{dx_1} \left\{ \frac{\chi_{zzz}}{\phi^2(x_1)} \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} B_m(k) B_n(k) \right. \\
 &\quad \times [ik_m - i\alpha(k_m, \omega)\zeta'(x_1)][ik_n - i\alpha(k_n, \omega)\zeta'(x_1)] \\
 &\quad \times \exp[i(k_m + k_n)x_1 - i(\alpha(k_m, \omega) + \alpha(k_n, \omega))\zeta(x_1)] \\
 &\quad + \frac{\chi_{ztt}}{\epsilon^2(\omega)\phi^2(x_1)} \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} B_m(k) B_n(k) \\
 &\quad \times [-\zeta'(x_1) ik_m - i\alpha(k_m, \omega)][-\zeta'(x_1) ik_n - i\alpha(k_n, \omega)] \\
 &\quad \left. \times \exp[i(k_m + k_n)x_1 - i(\alpha(k_m, \omega) + \alpha(k_n, \omega))\zeta(x_1)] \right\} . \quad (7.35b)
 \end{aligned}$$

When these expressions for $A(x_1|2\omega)$ and $B(x_1|2\omega)$ are substituted into the right hand side of Eq. (7.34) the resulting equation has the form

$$\begin{aligned}
 &\sum_{p=-\infty}^{\infty} \frac{\alpha(k_r^{(2)}, 2\omega)\alpha_0(k_p^{(2)}, 2\omega) + k_r^{(2)}k_p^{(2)}}{\alpha(k_r^{(2)}, 2\omega) - \alpha_0(k_p^{(2)}, 2\omega)} \mathcal{J}_{r-p} \left(\alpha_0(k_p^{(2)}, 2\omega) - \alpha(k_r^{(2)}, 2\omega) \right) A_p(2k) \\
 &= \frac{1}{\epsilon(2\omega) - 1} \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} C_{r mn} B_m(k) B_n(k), \quad r = 0, \pm 1, \pm 2, \dots \quad (7.36)
 \end{aligned}$$

The expression for the coefficient $C_{r mn}$ is cumbersome and will not be given here.

The efficiency of the n th order beam diffracted from the grating at the frequency 2ω was defined by Farias and Maradudin [249] as the total time-averaged flux in the beam crossing a plane $x_3 = \text{constant}$ in the $+x_3$ -direction divided by the total time-averaged flux in the incident beam crossing the same plane in the $-x_3$ -direction. If this efficiency is normalized by the corresponding one for the planar surface, the result was called the intensity of the n th order diffracted beam at frequency 2ω :

$$I_n = \frac{\alpha_0(k_n^{(2)}, 2\omega) |R_n(2k)|^2}{\alpha_0(k_0^{(2)}, 2\omega) |R_f(2k)|^2}, \quad (7.37)$$

where $R_f(2k)$ is the amplitude of the field diffracted from a planar surface. It is given by the solution of Eq. (7.36) in the case that $\zeta(x_1) \equiv 0$.

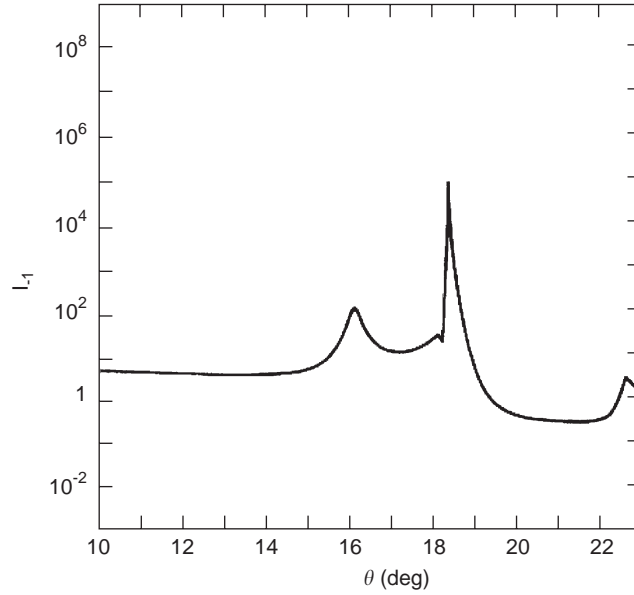


Fig. 82. The intensity I_{-1} as a function of the angle of incidence θ_0 , for the diffraction of p-polarized light from a silver surface defined by $x_3 = \zeta(x_1) = \zeta_0 \cos(2\pi x_1/a)$ with $\zeta_0 = 120 \text{ \AA}$ and $a = 8000 \text{ \AA}$. The frequency of the incident light corresponds to an energy $\hbar\omega = 1.17 \text{ eV}$. The dielectric constants at the fundamental and harmonic frequencies are $\epsilon(\omega) = -56.248 + i0.6$ and $\epsilon(2\omega) = -11.557 + i0.367$.

The intensity I_{-1} has been calculated as a function of the angle of incidence θ_0 by Farias and Maradudin [249] for a silver grating defined by a surface profile function $\zeta(x_1) = \zeta_0 \cos(2\pi x_1/a)$, where $\zeta_0 = 120 \text{ \AA}$ and $a = 8000 \text{ \AA}$. It is illuminated by p-polarized light whose frequency corresponds to $\hbar\omega = 1.17 \text{ eV}$. The dielectric functions of silver at the fundamental and harmonic frequencies are $\epsilon(\omega) = (-56.248 + i0.6)$ and $\epsilon(2\omega) = (-11.557 + i0.367)$. The result is presented in Fig. 82.

In their theory Farias and Maradudin used nonlinear boundary conditions of the type given by Eqs. (7.19) and (7.35), but with a different form for the source function $B(x_1|2\omega)$ associated with the inclusion of a contribution from the bulk nonlinear polarization and the assumption of the free electron model. Nevertheless, the features observed in the result plotted in Fig. 82 can be interpreted on the basis of the theory outlined here. An enhancement in I_{-1} is observed at an angle of incidence $\theta_0 = 18.39^\circ$. At this angle, at the fundamental frequency ω , $k_{-1} \cong -k_{\text{sp}}(\omega)$. Thus, at this angle the incident light excites a surface plasmon polariton with frequency ω on the grating surface, and enhances the amplitude $B_{-1}(k)$. The propagation constant $\alpha_0(k_{-1}, \omega)$ at this angle of incidence is pure imaginary, which means that $B_{-1}(k)$ corresponds to a field localized to the surface. However, the corresponding $\alpha_0(k_{-1}^{(2)}, 2\omega)$ is real, and the second harmonic beam associated with $A_{-1}(2k)$ is an outgoing wave. Since, from Eqs. (7.36) and (7.37) the intensity I_{-1} is approximately proportional to $|B_{-1}(k)|^4$, an increase in the magnitude of $B_{-1}(k)$ leads to an enhancement of I_{-1} . A peak in I_{-1} is also seen at $\theta_0 = 16.15^\circ$. At this angle of incidence $k_{-2}^{(2)} \cong -k_{\text{sp}}(2\omega)$. Thus at this angle of incidence the incident light excites a surface plasmon polariton of frequency 2ω on the grating surface and enhances the amplitude $A_{-2}(2k)$. However, the propagation constant $\alpha_0(k_{-2}^{(2)}, 2\omega)$ in this case is imaginary, so that the amplitude $A_{-2}(2k)$ is that of a field

localized to the surface, and the diffracted intensity I_{-2} cannot be observed in the far field. However, the excitation of the surface plasmon polariton at the frequency 2ω is associated with a decrease of the determinant of the matrix of the coefficients of the $\{A_p(2k)\}$ on the left hand side of Eq. (7.36), because the vanishing of this determinant is the dispersion relation for the surface plasmon polariton at the frequency 2ω . Consequently, not only $A_{-2}(2k)$ increases, but $A_{-1}(2k)$ also can increase, and in this way an enhancement of I_{-1} occurs at this angle of incidence. Finally, an enhancement of I_{-1} is also observed at an angle of incidence $\theta_0 = 22.6^\circ$. At this angle $k_1^{(2)} \cong k_{sp}(2\omega)$, but $\alpha_0(k_1^{(2)}, 2\omega)$ is imaginary and I_1 corresponds to an evanescent wave. However, as in the case of the angle of incidence $\theta_0 = 16.15^\circ$, enhancements of the intensities of other modes associated with outgoing waves occurs. The enhancement of I_{-1} at $\theta_0 = 16.15^\circ$ due to the increase of I_{-2} is larger than the enhancement of I_{-1} at $\theta_0 = 22.6^\circ$ due to the increase of I_{-1} , because the mode (-2) is closer to the mode (-1) than to the mode (1) .

7.1.2. Second harmonic generation at randomly rough metal surfaces

In the case that the surface of a semi-infinite metal is a one-dimensional random surface defined by the equation $x_3 = \zeta(x_1)$, it is assumed that the surface profile function $\zeta(x_1)$ is at least twice differentiable, and constitutes a zero-mean, stationary Gaussian random process, defined by

$$\langle \zeta(x_1)\zeta(x'_1) \rangle = \delta^2 W(|x_1 - x'_1|) . \tag{7.38}$$

The angle brackets in Eq. (7.38) denote an average over the ensemble of realizations of the surface profile function, and $\delta = \langle \zeta^2(x_1) \rangle^{1/2}$ is the rms height of the surface.

The power spectrum of the surface roughness $g(|k|)$ is the Fourier transform of the surface height autocorrelation function $W(|x_1|)$,

$$g(|k|) = \int_{-\infty}^{\infty} dx_1 W(|x_1|) \exp(-ikx_1) . \tag{7.39}$$

In the numerical calculations of the efficiency of second harmonic generation in reflection from such a metal surface, the form of the power spectrum to be used will be

$$g(|k|) = \frac{\pi}{k_{\max} - k_{\min}} [\theta(k - k_{\min})\theta(k_{\max} - k) + \theta(-k - k_{\min})\theta(k_{\max} + k)], \tag{7.40}$$

where $\theta(z)$ is the Heaviside unit step function, $k_{\min} < k_{sp}(\Omega) < k_{\max}$, $k_{sp}(\Omega) = \text{Re}[(\Omega/c)[\epsilon(\Omega)/(\epsilon(\Omega) + 1)]^{1/2}$ is the wavenumber of a surface plasmon polariton of frequency Ω , and Ω stands for ω or 2ω . Surfaces characterized by a power spectrum of this type have been used in recent experimental studies of the scattering of light from weakly rough random metal surfaces [113,238,260].

Because it can deal with both weakly rough and strongly rough surfaces, the most successful theoretical approach to the study of multiple-scattering effects in the second harmonic generation of light in reflection from randomly rough surfaces is that of a Monte Carlo computer simulation approach. In order to implement this approach the Green's functions for an infinite vacuum region and for an infinite metal at each of the frequencies ω and 2ω are required. These satisfy the equations

$$\left(\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_3^2} + \frac{\Omega^2}{c^2} \right) G_0^{(\Omega)}(x_1, x_3|x'_1, x'_3) = -4\pi\delta(x_1 - x'_1)\delta(x_3 - x'_3) \tag{7.41}$$

and

$$\left(\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_3^2} + \epsilon(\Omega) \frac{\Omega^2}{c^2} \right) G_\epsilon^{(\Omega)}(x_1, x_3|x'_1, x'_3) = -4\pi\delta(x_1 - x'_1)\delta(x_3 - x'_3) , \quad (7.42)$$

respectively, where Ω stands for ω or 2ω , and outgoing wave or vanishing boundary conditions at infinity. Explicit expressions for these Green's functions are

$$\begin{aligned} G_0^{(\Omega)}(x_1, x_3|x'_1, x'_3) &= i\pi H_0^{(1)} \left(\frac{\Omega}{c} [(x_1 - x'_1)^2 + (x_3 - x'_3)^2]^{1/2} \right) \\ &= \int_{-\infty}^{\infty} \frac{dq}{2\pi} \frac{2\pi i}{\alpha_0(q, \Omega)} \exp[iq(x_1 - x'_1) + i\alpha_0(q, \Omega)|x_3 - x'_3|] \\ G_\epsilon^{(\Omega)}(x_1, x_3|x'_1, x'_3) &= i\pi H_0^{(1)} \left(\sqrt{\epsilon(\Omega)} \frac{\Omega}{c} [(x_1 - x'_1)^2 + (x_3 - x'_3)^2]^{1/2} \right) \\ &= \int_{-\infty}^{\infty} \frac{dq}{2\pi} \frac{2\pi}{\beta(q, \Omega)} \exp[iq(x_1 - x'_1) - \beta(q, \Omega)|x_3 - x'_3|] \end{aligned} \quad (7.43)$$

where $H_0^{(1)}(z)$ is a Hankel function,

$$\begin{aligned} \alpha_0(q, \Omega) &= \left(\frac{\Omega^2}{c^2} - q^2 \right)^{1/2}, \quad |q| < \frac{\Omega}{c} \\ &= i \left(q^2 - \frac{\Omega^2}{c^2} \right)^{1/2}, \quad |q| > \frac{\Omega}{c}, \end{aligned} \quad (7.44)$$

and

$$\beta(q, \Omega) = \left(q^2 - \epsilon(\Omega) \frac{\Omega^2}{c^2} \right)^{1/2}, \quad \text{Re } \beta(q, \Omega) > 0, \quad \text{Im } \beta(q, \Omega) < 0 . \quad (7.45)$$

Since the fields of both frequencies ω and 2ω satisfy Helmholtz equations in the vacuum and in the metal, by applying Green's second integral identity in the plane [278] to these regions, we obtain the pair of equations

$$\begin{aligned} \theta(x_3 - \zeta(x_1))H_2^{(I)}(x_1, x_3|\omega) &= H_2^{(I)}(x_1, x_3|\omega)_{\text{inc}} \\ &+ \frac{1}{4\pi} \int_{-\infty}^{\infty} dx'_1 \left\{ \left[\frac{\partial}{\partial N'} G_0^{(\omega)}(x_1, x_3|x'_1, x'_3) \right] H_2^{(I)}(x'_1, x'_3|\omega) \right. \\ &\left. - G_0^{(\omega)}(x_1, x_3|x'_1, x'_3) \frac{\partial}{\partial N'} H_2^{(I)}(x'_1, x'_3|\omega) \right\}_{x'_3=\zeta(x'_1)} \end{aligned} \quad (7.46a)$$

$$\begin{aligned} \theta(\zeta(x_1) - x_3)H_2^{(II)}(x_1, x_3|\omega) &= \frac{-1}{4\pi} \int_{-\infty}^{\infty} dx'_1 \left\{ \left[\frac{\partial}{\partial N'} G_\epsilon^{(\omega)}(x_1, x_3|x'_1, x'_3) \right] H_2^{(II)}(x'_1, x'_3|\omega) \right. \\ &\left. - G_\epsilon^{(\omega)}(x_1, x_3|x'_1, x'_3) \frac{\partial}{\partial N'} H_2^{(II)}(x'_1, x'_3|\omega) \right\}_{x'_3=\zeta(x'_1)} \end{aligned} \quad (7.46b)$$

for the field at the fundamental frequency, where $H_2^{(I)}(x_1, x_3|\omega)_{\text{inc}}$ is the incident field, and the pair of equations

$$\begin{aligned} \theta(x_3 - \zeta(x_1))H_2^{(I)}(x_1, x_3|2\omega) &= \frac{1}{4\pi} \int_{-\infty}^{\infty} dx'_1 \left\{ \left[\frac{\partial}{\partial N'} G_0^{(2\omega)}(x_1, x_3|x'_1, x'_3) \right] H^{(I)}(x'_1, x'_3|2\omega) \right. \\ &\quad \left. - G_0^{(2\omega)}(x_1, x_3|x'_1, x'_3) \frac{\partial}{\partial N'} H_2^{(I)}(x'_1, x'_3|2\omega) \right\}_{x'_3=\zeta(x'_1)} \end{aligned} \tag{7.47a}$$

$$\begin{aligned} 0 &= -\frac{1}{4\pi} \int_{-\infty}^{\infty} dx'_1 \left\{ \left[\frac{\partial}{\partial N'} G_{\epsilon}^{(2\omega)}(x_1, x_3|x'_1, x'_3) \right] H_2^{(II)}(x'_1, x'_3|2\omega) \right. \\ &\quad \left. - G_{\epsilon}^{(2\omega)}(x_1, x_3|x'_1, x'_3) \frac{\partial}{\partial N'} H_2^{(II)}(x'_1, x'_3|2\omega) \right\}_{x'_3=\zeta(x'_1)} \end{aligned} \tag{7.47b}$$

for the field at the harmonic frequency.

If we set the point of observation on the surface in each of these equations by setting $x_3 = \zeta(x_1) + \eta$, where η is a positive infinitesimal, and use the linear and nonlinear boundary conditions, we obtain the equations satisfied by the source functions $H^{(I)}(x_1|\Omega)$ and $L^{(I)}(x_1|\Omega)$:

$$\begin{aligned} H^{(I)}(x_1|\omega) &= H^{(I)}(x_1|\omega)_{\text{inc}} + \frac{1}{4\pi} \int_{-\infty}^{\infty} dx'_1 \left\{ \left[\frac{\partial}{\partial N'} G_0^{(\omega)}(x_1, x_3|x'_1, x'_3) \right]_{\substack{x_3=\zeta(x_1)+\eta \\ x'_3=\zeta(x'_1)}} \right. \\ &\quad \left. \times H^{(I)}(x'_1|\omega) - \left[G_0^{(\omega)}(x_1, x_3|x'_1, x'_3) \right]_{\substack{x_3=\zeta(x_1)+\eta \\ x'_3=\zeta(x'_1)}} L^{(I)}(x'_1|\omega) \right\} \end{aligned} \tag{7.48a}$$

$$\begin{aligned} 0 &= -\frac{1}{4\pi} \int_{-\infty}^{\infty} dx'_1 \left\{ \left[\frac{\partial}{\partial N'} G_{\epsilon}^{(\omega)}(x_1, x_3|x'_1, x'_3) \right]_{\substack{x_3=\zeta(x_1)+\eta \\ x'_3=\zeta(x'_1)}} H^{(I)}(x'_1|\omega) \right. \\ &\quad \left. - \epsilon(\omega) \left[G_{\epsilon}^{(\omega)}(x_1, x_3|x'_1, x'_3) \right]_{\substack{x_3=\zeta(x_1)+\eta \\ x'_3=\zeta(x'_1)}} L^{(I)}(x'_1|\omega) \right\} \end{aligned} \tag{7.48b}$$

and

$$\begin{aligned} H^{(I)}(x_1|2\omega) &= \frac{1}{4\pi} \int_{-\infty}^{\infty} dx'_1 \left\{ \left[\frac{\partial}{\partial N'} G_0^{(2\omega)}(x_1, x_3|x'_1, x'_3) \right]_{\substack{x_3=\zeta(x_1)+\eta \\ x'_3=\zeta(x'_1)}} H^{(I)}(x'_1|2\omega) \right. \\ &\quad \left. - \left[G_0^{(2\omega)}(x_1, x_3|x'_1, x'_3) \right]_{\substack{x_3=\zeta(x_1)+\eta \\ x'_3=\zeta(x'_1)}} L^{(I)}(x'_1|2\omega) \right\} \end{aligned} \tag{7.49a}$$

$$\begin{aligned} Q^{(I)}(x_1|2\omega) = \frac{1}{4\pi} \int_{-\infty}^{\infty} dx'_1 \left\{ \left[\frac{\partial}{\partial N'} G_{\epsilon}^{(2\omega)}(x_1, x_3|x'_1, x'_3) \right]_{\substack{x_3=\zeta(x_1)+\eta \\ x'_3=\zeta(x'_1)}} H^{(I)}(x'_1|2\omega) \right. \\ \left. - \epsilon(2\omega) \left[G_{\epsilon}^{(2\omega)}(x_1, x_3|x'_1, x'_3) \right]_{\substack{x_3=\zeta(x_1)+\eta \\ x'_3=\zeta(x'_1)}} L^{(I)}(x'_1|2\omega) \right\} , \end{aligned} \quad (7.49b)$$

where

$$\begin{aligned} Q^{(I)}(x_1|2\omega) = \frac{1}{4\pi} \int_{-\infty}^{\infty} dx'_1 \left\{ \left[\frac{\partial}{\partial N'} G_{\epsilon}^{(2\omega)}(x_1, x_3|x'_1, x'_3) \right]_{\substack{x_3=\zeta(x_1)+\eta \\ x'_3=\zeta(x'_1)}} A(x'_1|2\omega) \right. \\ \left. - \epsilon(2\omega) \left[G_{\epsilon}^{(2\omega)}(x_1, x_3|x'_1, x'_3) \right]_{\substack{x_3=\zeta(x_1)+\eta \\ x'_3=\zeta(x'_1)}} B(x'_1|2\omega) \right\} . \end{aligned} \quad (7.50)$$

These equations are solved numerically by converting them into matrix equations, by truncating the region of integration from $(-\infty, \infty)$ to $(-L/2, L/2)$, and using a numerical quadrature scheme [279].

From Eq. (7.47a) we find that the scattered field in the far-field in the vacuum region is given by

$$\begin{aligned} H_2^{(I)}(x_1, x_3|2\omega)_{sc} = \frac{1}{4\pi} \int_{-\infty}^{\infty} dx'_1 \left\{ \left[\frac{\partial}{\partial N'} G_0^{(2\omega)}(x_1, x_3|x'_1, x'_3) \right]_{x'_3=\zeta(x'_1)} H^{(I)}(x'_1|2\omega) \right. \\ \left. - \left[G_0^{(2\omega)}(x_1, x_3|x'_1, x'_3) \right]_{x'_3=\zeta(x'_1)} L^{(I)}(x'_1|2\omega) \right\} \\ = \int_{-\infty}^{\infty} \frac{dq}{2\pi} R(q, 2\omega) \exp[iqx_1 + i\alpha_0(q, 2\omega)x_3] , \end{aligned} \quad (7.51)$$

where

$$\begin{aligned} R(q, 2\omega) = \frac{i}{2\alpha_0(q, 2\omega)} \int_{-\infty}^{\infty} dx_1 \exp[-iqx_1 - i\alpha_0(q, 2\omega)\zeta(x_1)] \\ \times \left\{ i[q\zeta'(x_1) - \alpha_0(q, 2\omega)]H^{(I)}(x_1|2\omega) - L^{(I)}(x_1|2\omega) \right\} . \end{aligned} \quad (7.52)$$

It is convenient to normalize the scattered power in such a way that it does not depend on the incident power and the illuminated area. We therefore define the normalized second harmonic scattered power as

$$I_{sc}^{(2\omega)} = \frac{P_{sc}^{(2\omega)}}{[P_{inc}^{(\omega)}]^2} S , \quad (7.53)$$

where S is the illuminated area. If we choose a plane wave of amplitude \mathcal{H}_0 for the incident field, the normalized incoherent component of the scattered light may be written as

$$\left\langle I_{sc}^{(2\omega)}(\theta_s) \right\rangle_{incoh} = \frac{\langle |r(\theta_s|2\omega)|^2 \rangle - |\langle r(\theta_s|2\omega) \rangle|^2}{2\omega L \cos \theta_0 |\mathcal{H}_0|^4} \quad (7.54)$$

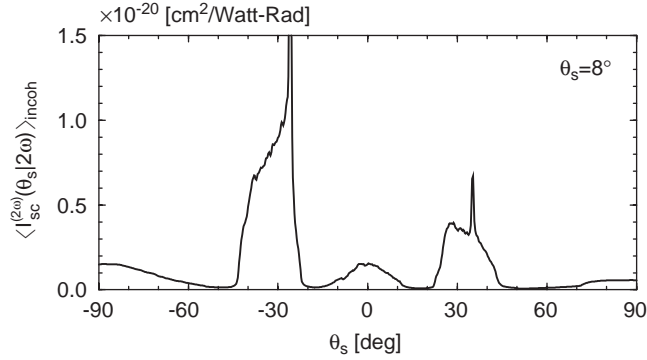


Fig. 83. The incoherent component of the mean normalized second harmonic intensity as a function of the scattering angle θ_s for the scattering of p-polarized light from a one dimensional random silver surface characterized by a West–O’Donnell power spectrum centered at $k_{sp}(\omega)$ with $\theta_{max} = 15^\circ$ and $\delta = 28.3$ nm. The angle of incidence is $\theta_0 = 8^\circ$, and the fundamental wavelength is $\lambda = 1.064$ μm . The dielectric constants are $\epsilon(\omega) = -56.25 + i0.60$ and $\epsilon(2\omega) = -11.56 + i0.37$. The results from $N_p = 2000$ realizations were averaged to obtain this figure. The vertical lines indicate the normal and backscattering directions.

where L is the length of the surface, and

$$r(\theta_s | 2\omega) = \int_{-\infty}^{\infty} dx_1 \left\{ i \frac{2\omega}{c} [\zeta'(x_1) \sin \theta_s - \cos \theta_s] H^{(1)}(x_1 | 2\omega) - L^{(1)}(x_1 | 2\omega) \right\} \exp \left\{ -i \frac{2\omega}{c} [x_1 \sin \theta_s + \zeta(x_1) \cos \theta_s] \right\} . \quad (7.55)$$

The angle brackets in Eq. (7.54) denote an average over the results obtained for N_p (≈ 1000) realizations of the surface profile function, which are generated by standard numerical procedures [280]. θ_s is the scattering angle, and θ_0 is the angle of incidence.

To illustrate some of the results that have been obtained by the approach described in this subsection, we present scattering distributions at the harmonic frequency for a shallow one-dimensional random silver surface defined by a power spectrum that enhances the excitation of surface plasmon polaritons at the fundamental frequency. It is a West–O’Donnell power spectrum, Eq. (7.40), that is centered at the surface plasmon polariton wavenumber $k_{sp}(\omega)$. The values of $k_{min}^{(1)}$ and $k_{max}^{(1)}$ are given by

$$\begin{aligned} k_{min}^{(1)} &= k_{sp}(\omega) - (\omega/c) \sin \theta_{max} , \\ k_{max}^{(1)} &= k_{sp}(\omega) + (\omega/c) \sin \theta_{max} , \end{aligned} \quad (7.56)$$

with $\theta_{max} = 15^\circ$. The rms height of this surface is $\delta = 28.3$ nm. The significance of the angle θ_{max} is that if the angle of incidence θ_0 is in the interval $(-\theta_{max}, \theta_{max})$ the incident light can excite surface plasmon polaritons with wavenumbers $\pm k_{sp}(\omega)$. Similarly, if the scattering angle θ_s is in the same interval, the surface plasmon polaritons so excited will be coupled to volume electromagnetic waves in the vacuum, propagating away from the surface. Computer simulation results for the mean normalized second harmonic scattered intensity for this surface are presented in Fig. 83.

The fact that when the angle of incidence is smaller than 15° surface plasmon polaritons are launched in both the $+x_1$ - and $-x_1$ -directions means that a relatively intense electromagnetic field is created on

the surface. For this surface the main mechanisms for the generation of the second harmonic field are the nonlinear mixing of a surface plasmon polariton and the incident field, and the nonlinear mixing of two surface plasmon polaritons at the fundamental frequency, traveling in the same or opposite directions.

The components parallel to the surface of the wave vectors involved in the nonlinear mixing must satisfy well-known momentum selection rules, which yield the directions at which the second harmonic light should appear. In the case of the second harmonic scattered light arising from the nonlinear mixing of the incident field and a surface plasmon polariton, the wavenumber conservation law reads

$$k_{\text{inc}}(\omega) \pm k_{\text{sp}}(\omega) = k_{\text{sc}}(2\omega) , \quad (7.57)$$

where $k_{\text{inc}}(\omega) = (\omega/c) \sin \theta_0$ and $k_{\text{sc}}(\omega) = (\omega/c) \sin \theta_s$. Consequently, for an angle of incidence $\theta_0 = 8^\circ$, a relatively large second harmonic intensity is expected at $\theta_s = -26^\circ$ and 35° . Bands of light at these angles are evident in Fig. 83.

Two surface plasmon polaritons at the fundamental frequency propagating in the same direction produce, through nonlinear mixing, a surface plasmon polariton at the harmonic frequency propagating in the same direction. This mechanism produces the two broad distributions around $\pm(30\text{--}40^\circ)$ in Fig. 83. On a planar surface two surface plasmon polaritons propagating in opposite directions produce no far-field radiation at the harmonic frequency. It has been argued [122], however, that for a rough surface they can produce a feature in the direction normal to the mean surface. In this respect, we note that in the second harmonic distribution plotted in Fig. 83 there are no peaks in the backscattering and normal directions.

The result presented in Fig. 83 can be compared directly with the result presented in Fig. 8 of the paper by O'Donnell et al. [238]. The main features are reproduced, and the intensity levels are the same. There are only subtle differences between the experimental and theoretical distributions.

7.2. Near-field second-harmonic generation at metal surfaces

Second-harmonic generation (SHG) from smooth and rough metal surfaces has been extensively studied in the far-field. SHG was used in the past to study electron dynamics in metals, interactions in a correlated electron gas, Fermi surfaces in different metals, etc. Surface SHG in reflection obeys a set of selection rules. It can be excited with p-polarized light leading to p- or s-polarized SH light generated in the direction of the specularly reflected exciting light. Second-harmonic generation is known to be surface sensitive on an atomic scale (especially when the nonlinearity itself is due to the presence of the surface). Studies of the nonlinear optical response are advantageous for understanding the relationship between optical properties and the morphology (roughness, defects, impurities, adsorbates) of metallic surfaces [149,281–283].

The consideration of roughness leads to several phenomena affecting SHG, such as relaxation of polarization selection rules and an enhancement of the electromagnetic field at surface defects. Less obvious is the influence of weak and strong light localization effects on SHG from rough surfaces that results in significant changes of the angular spectrum of SHG observed in the far-field region [122,238]. Even more pronounced changes should be expected in the near-field region close to the surface. In analyzing near-field SHG measurements one should consider the substantial influence of evanescent fundamental and second-harmonic fields. Phase-matching conditions and polarization selection rules are meaningless in the near-field of a rough surface. Both are defined with respect to the plane of incidence determined by the direction of propagation and the normal to the surface. There is no direction of propagation for an evanescent wave. A mean normal direction depends on the scale of observation and close to a surface

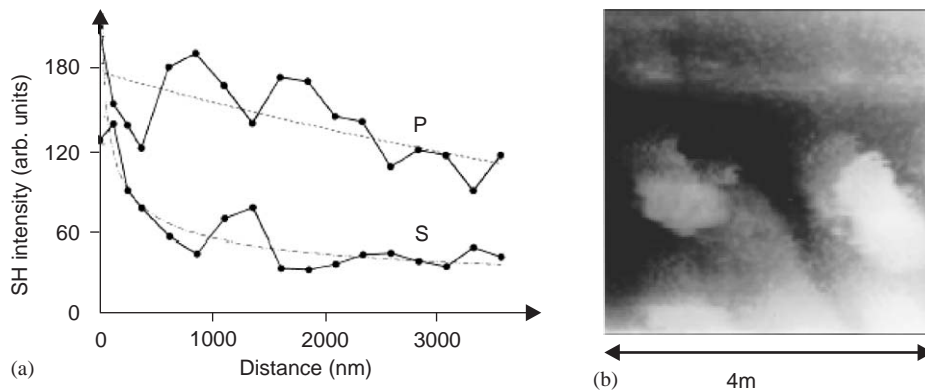


Fig. 84. (a) Distance dependence of the SH intensity for s- and p-polarizations of the fundamental light measured at the center of the surface region shown in (b). Dashed lines are power-law dependencies drawn as a guide to the eye. (Ref. [239].)

feature the polarization of light which is determined with respect to a mean surface plane will be different compared to a local surface topography. The field enhancement effects at the fundamental as well as SH frequencies can arise due to the lightning-rod effect at highly curved surfaces as well as to localized surface plasmon excitation at the surface features [282,31]. Less obvious is the strong influence of SPP localization effects on the SHG from rough surfaces that result in significant changes of the angular spectrum of SHG observed in the far-field region [122,237,261]. The imaging of the SH field distribution over a surface provides direct information on the contributions of different effects to the SHG.

In order to determine the contribution of near-field processes to the SH generation, the dependence of the second-harmonic signal on the tip-sample distance for s- and p-polarized exciting light was studied (Fig. 84). A strong decrease in the signal was observed at about 500 nm from the surface. This behavior is especially pronounced for s-polarized excitation. The initial drop in the signal is followed by a number of oscillations for both excitation polarizations: each maximum for p-polarized excitation corresponds to a minimum for s-polarized excitation and vice versa. This behavior results from the electromagnetic interaction due to multiple reflections of fundamental and SH light between the glass tip and silver film surface. The phase shift between the distance dependencies can be explained by the π phase shift between p- and s-polarized fundamental waves reflected between the tip and metal at angles larger than the Brewster angle. The ratio of the SH signals induced by p- and s-polarized light changes substantially in the near-field compared to the far-field region. This ratio falls from about 3 at a distance larger than 3 μm from the surface to approximately 1 near the surface. This substantial increase in the s-polarized light induced evanescent SH field in the vicinity of a rough surface is a nontrivial near-field phenomenon, which is a consequence of the relaxation of the polarization selection rules and phase matching conditions in the near field.

In conventional experiments on surface SHG one has the possibility to measure only an intensity and an angular distribution of the SHG originated from a relatively large surface area (determined by the size of the illuminating spot) and, therefore, averaged over a large number of defects without an exact knowledge of the surface morphology. Before the introduction of near-field optical techniques the only experimental approach to SH characterization of rough surfaces was to change in one way or another the distribution of defect sizes and shapes and then compare the measured tendencies of the average SH enhancement with

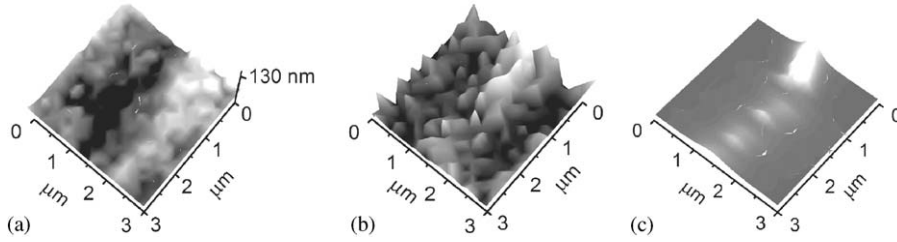


Fig. 85. (a) Topography; (b) measured; (c) calculated SH images of a silver surface with a defect supporting a localized surface plasmon. The fundamental light ($\lambda = 1064$ nm) is s-polarized. (Ref. [239].)

model predictions obtained for an average random defect ensemble. In an ideal experiment one would like to know exactly the size and shape of defects contributing to SHG. At the same time, the lateral distribution of the electromagnetic field over a surface is not uniform and depends on the surface defect structure: the local field intensity can vary by several orders of magnitude on a scale smaller than half a wavelength along the surface [35]. Therefore, investigations of the averaged optical response in many cases do not result in an understanding of the underlying physics. The use of near-field measurements provides an opportunity to study SHG from specifically chosen defects before the SH contributions from individual defects are averaged in the far-field zone.

Studies of SHG from different types of individual surface defects have revealed that the main contrast mechanism for near-field SH imaging is topography variations. Such behavior is observed for different kinds of surface defects such as grooves, pits, or protrusions on a metal surface [239,241–243]. In all these cases the enhancement of SH generation induced by s-polarized excitation occurs near the places on the surface where inhomogeneities are present. This effect has the same origin as the growth of the ratio of s- to p-polarized light induced SH signals near the rough surface discussed above. The polarization of the exciting light is important for achieving one kind of field enhancement or another. However, in the case of p-polarized excitation, the SHG related to surface inhomogeneities is observed in the background SHG due to nonlinear processes in the electron gas close to a smooth surface.

Let us consider several examples of SHG enhancement on individual surface defects. In the case of the surface shown in Fig. 85 which is essentially flat with a narrow deep groove, the SH image looks like a “negative” of the topographical image. The SH signal measured in the flat regions is much smaller than the signal measured near the groove [239]. In the case of protrusions on a metal surface (Fig. 86), up to a 10-fold local enhancement of the SHG at defects with micrometer-size lateral dimensions has been observed, while the average SHG enhancement was negligible (of about a factor of 1.2) [243].

In a general form, the enhancement of SHG excited and detected at a surface defect can be described as

$$I_{\text{SH}} \sim |L(2\omega)L^2(\omega)|^2, \quad (7.58)$$

where $L(2\omega)$ and $L(\omega)$ are the field enhancement factors at the fundamental and SH wavelengths, respectively [31]. The local field enhancement factor $L = L_{\text{LR}}L_{\text{LSP}}$ takes into account the lightning-rod effect (L_{LR}) due to the geometrical electromagnetic field enhancement at a highly curved surface and localized surface plasmons excitation (L_{LSP}). For both enhancement mechanisms the orientation of the light polarization with respect to a defect is important.

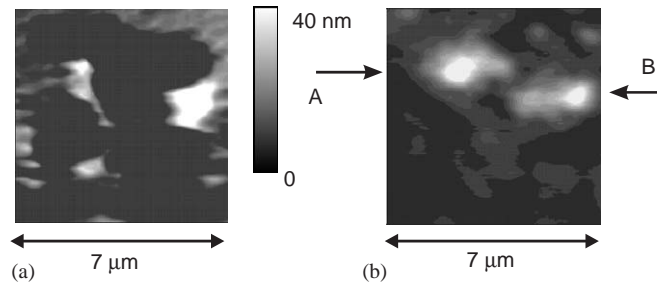


Fig. 86. (a) Topography and (b) near-field SH intensity over a gold film with large defects. The fundamental wavelength is 790 nm. Arrows indicate the directions of the cross sections shown in Fig. 87. (Ref. [243].)

For the defect in Fig. 85a the main SHG enhancement effect is related to surface plasmons localized in the groove due to the variation of the film thickness. The LSP spectrum and the related field distribution can be relatively easily calculated using the model of a pit of Lorentzian shape [30,241]. For the parameters of the groove measured from the topographical image, the localized plasmon resonance is inferred to be close to the SH photon energy, and the $L_{\text{LSP}}(2\omega)$ term in Eq. (7.58) dominates. The SH intensity distribution calculated in the model of the groove-related surface plasmon using the measured topography reproduces the experimentally observed image very well. An increase of the groove width results in a decrease of the observed SH intensity and vice versa (only topography associated with the pit was used for the model calculations; other topographical features were neglected in the modeling).

The dipolar localized surface plasmons related to the protrusions on the gold surface in Fig. 86 are estimated to be at about 0.9 eV, which is much lower than the energy of fundamental ($\hbar\omega \approx 1.57$ eV) and second-harmonic photons. In this case, mainly the lightning-rod effect, L_{LR} , is important. For estimation of this geometrical field enhancement the surface defects can be considered as particles of a conical shape. The enhancement close to the apex of such particles can be estimated in the electrostatic model as

$$E \sim E_3^{(0)} 1/r^{1-\nu} . \quad (7.59)$$

Here, $E_3^{(0)}$ is the component of the electric field parallel to the cone axis near the planar surface far from the defect, r is the distance to the cone apex, and ν is the first root of the equation $P_\nu(\cos(\theta_0)) = 0$, where P_ν and θ_0 are the Legendre function of the first kind and the cone opening angle, respectively. Depending on the cone angle, the parameter ν is in the range 0–1 for an ideal metal [284]. The corrections due to a finite dielectric constant of a material and surroundings allow $\nu \geq -0.5$ to be achieved [285]. This model gives fairly good agreement with the experimentally measured SH intensity distributions for the parameters of the defects obtained from the topographic image (Fig. 87). The SHG from smaller but sharper surface defects is dominating. The finite resolution of the SNOM leads to averaging of the signal over the excitation area and, therefore, to a subsequent reduction of the experimentally observed enhancement at a cone apex.

The sizes of individual protrusions are in this case too large to provide LSP resonances close to fundamental or SH wavelengths. Nevertheless, when the polarization of the excitation light is parallel to the axis between neighboring defects, the SHG intensity is enhanced due to the possible LSP resonance in the gap between these defects (Fig. 86). These defects are hardly visible in the SHG image for the orthogonal polarization, for which LSP cannot be not excited [243].

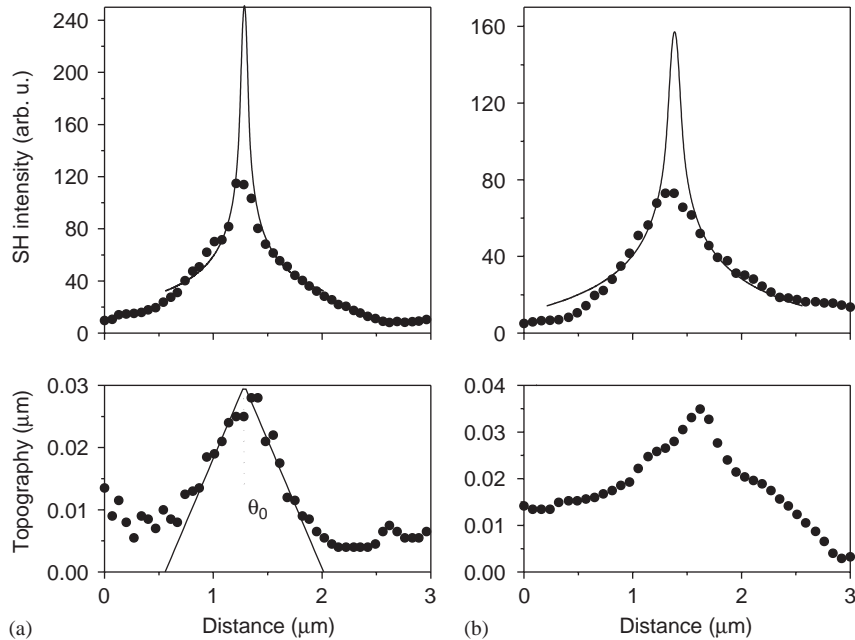


Fig. 87. Cross sections of the topography and SH intensity related to the defects indicated by the arrows in Fig. 86: (symbols) experimental data, (solid lines) calculations for the lightning-rod effect model. (Ref. [243].)

Both SPPs and localized plasmons at SH frequencies can be effectively generated at rough metal surfaces due to nonlinear wave-mixing processes, which provide the necessary momentum conservation needed for SPPs to be excited [236]. No momentum conservation rules apply for the excitation of localized plasmons. The contribution of such LSPs in the far-field SH light arises either due to direct radiation into light or due to the decay of LSP into propagating surface polaritons and their subsequent scattering into bulk light by other surface inhomogeneities which gives rise to diffuse SHG from rough surfaces. The latter process of the LSP decay in many cases is the predominant mechanism of the radiative damping leading to additional LSP resonance broadening and subsequent reduction of the related field enhancement (Section 2.1).

Generally, p-polarized light induced near-field SH images are much more complex due to the interplay between SHG processes related to surface inhomogeneities and surface plasmon polariton related effects. For example, in the case of a typical randomly rough surface (Fig. 88) with a number of defects of various sizes, larger defects have different but uniform brightness in the SH image (e.g., on the right hand side of the images). However, above the surface area with many small defects, the SH intensity distribution exhibits the presence of small bright spots not directly related to the defects (e.g., tops of the images). The width of the spots is about 240 nm, which is about half a wavelength of the SH light. These spots have been identified as the result of surface plasmon polariton localization on randomly rough surfaces [239,240] similar to those observed in the linear SNOM images of rough metal films exhibiting a similar topographic structure (Section 2.3).

In fact, second-harmonic generation at a rough surface and SPP excitation are closely related phenomena since every surface defect may act as the source of a surface wave. Localization effects in the SH generation from rough metal surfaces have been predicted [122]. It has been shown that multiple scattering

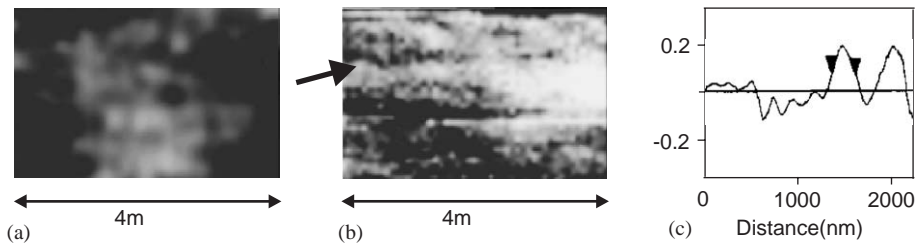


Fig. 88. (a) Topography and (b) near-field SH intensity distribution above an arbitrary rough silver surface obtained with p-polarized fundamental light. (c) The cross section of the SH image along the line indicated by the arrow. (Ref. [239].)

leading to light localization gives a considerable contribution to SHG resulting in its local enhancement. Both fundamental and second-harmonic SPP-related localization can contribute to the effect, giving rise to components of SH light propagating perpendicularly to the average sample surface and in the reverse direction to the fundamental wave propagation, respectively. This has been observed indirectly by measuring weak changes in the far-field angular distribution of the SH light [121,237,261]. The images observed in the near-field provide direct evidence of the SPP localization processes in SHG from randomly rough surfaces. Further SNOM studies have enabled identification of the spectral dependencies of this SPP localization phenomenon in SHG [240].

7.3. Near-field SHG from metallic diffraction gratings

In addition to arbitrary rough surfaces, surface enhanced optical phenomena have been widely studied on metallic gratings. Just as in the case of the enhanced linear optical transmission, gratings provide well defined resonant conditions for SPP excitation, and therefore for the field enhancement related to the SPP field. Gratings have been especially widely used for modifications of the spontaneous emission rate of molecules [165] and for studies of SHG [286–288].

Studies of the electromagnetic field distribution over sinusoidal gratings have revealed that the magnitude of the field enhancement over a grating is close to but somewhat larger than the value of the enhancement at smooth surfaces at which SPP are excited in the Kretschmann configuration [2,289]. The highest enhancement does not always correspond to the maxima in the grating profile. Both the magnitude of the field enhancement and the field distribution depend on the depth of the grating. In the case of deep gratings, similar to the single groove in a metal film discussed above, surface plasmons can be localized in a groove providing a strong field enhancement. In contrast, the field enhancement on periodic structures which occurs due to SPP Bloch waves has more complex lateral variations. In the latter case, the maxima of the field distribution can be achieved at different places of the grating both at minima and maxima of the topographical relief that is determined by the type of the excited SPP Bloch mode that can be excited (depending on the angle of incidence) in different Brillouin zones [180]. This has been recently observed in the far-field images of SHG on metallic gratings [290].

The field enhancement effects result in a corresponding enhancement of SHG from metallic gratings. In this case the enhancement of both the fundamental and second-harmonic fields should be considered. However in many cases the main effect is SPP excitation at the fundamental wavelength. In addition to the field enhancement a grating can provide the phase-matching conditions required for efficient SHG leading to selective enhancement of SHG observed in the direction of the diffraction orders of the

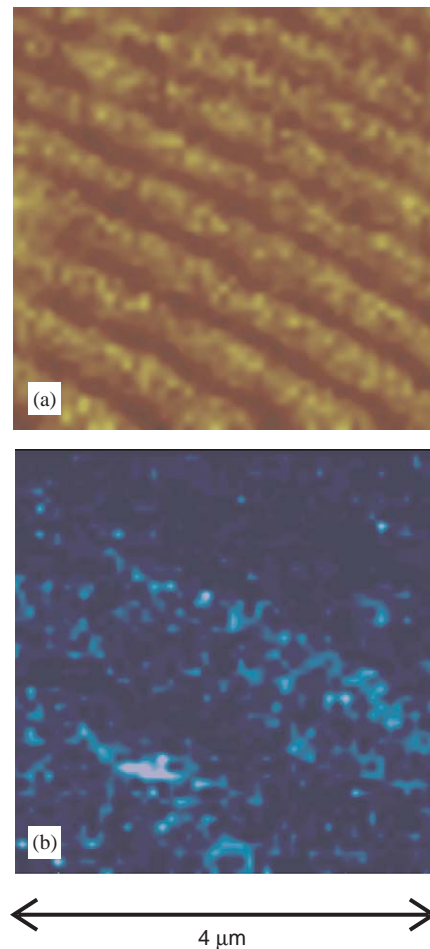


Fig. 89. (a) Topography and (b) near-field SH intensity distribution over an aluminum diffraction grating. The fundamental light ($\lambda = 1064$ nm) is p-polarized with respect to the grating plane. The gray-scale of the topographical image corresponds to 150 nm. (Ref. [242].)

fundamental light [287,288]. The enhancement magnitude depends strongly on the depth of the grooves, and is very sensitive to its local variations. The field enhancement effects at the small-scale roughness of the diffraction grating overlaying the periodic profile has also been found to be very important for the SHG enhancement in the frequency range corresponding to the LSP excitation [288].

Comparison of the simultaneously measured topographical images and SHG distribution can allow exact determination of the places where SHG is originated at a grating and where SHG enhancement takes place. The near-field distribution of SHG has been experimentally studied using an aluminum echellete [242]. A diffraction grating of 1800 grooves/mm has been used with a profile consisting of triangular grooves. The SHG has been excited with p-polarized exciting light. The grating has been oriented in such a way that the grooves were either parallel or perpendicular to the plane of incidence.

The near-field distribution of SH intensity reveals a more complex structure than the grating topography (Fig. 89). In addition to the regular structure related to the grating grooves, the images of the SH light

distribution are overlaid with fine structure related to the defect structure of a grating surface. In the near-field region close to the grooves, the additional roughness seems to play a determining role and the SH light distribution reflects the topographical features that are smaller and produce a stronger local field enhancement. In contrast to an arbitrary rough metal surface, a grating supports only SPPs having certain wavevectors. This leads to a different structure of the SH near-field on a grating compared to a rough surface under p-polarized excitation. If no resonant SPP excitation takes place, the enhancement of SHG on the defects of the relief instead of the surface polariton related effects dominates. Under the excitation close to the SPP resonance, both SPP Bloch waves and surface roughness contribute to the SHG enhancement. The near-field images demonstrate that the field enhancement due to localized surface plasmons at the small-scale defects of a grating can be stronger than the effects related to SPPs on a diffraction grating.

8. SPP interrogation using other scanning probe microscopy techniques

Despite the fact that the first near-field studies of surface polaritons have been performed with a scanning tunneling microscope [41,42,45], SNOM has almost completely ousted STM from SPP investigations. The reason is the strong interaction of an STM tip and the SPP field resulting in a significant perturbation of the surface polariton behavior and, as a result, a complex relation of the signal to the actual electromagnetic field over a surface. Nevertheless, STM studies provide additional important information on SPP behavior in a confined space and electrostatics at the nanoscale. Since an STM tip is scanned much closer to the surface, only at a distance of a few Ångströms, an STM allows one to study the effects of a strong electromagnetic interaction between a metal tip and a metal surface, such as localized surface plasmons at the tip–surface junction and related strong light confinement and field enhancement. Additional effects of the interaction of a surface polariton with localized surface plasmon resonances analogous to those mentioned in Section 2.1 can be investigated arising from the interplay between SPP and LSP when their resonant frequencies are close to each other. The interaction between an STM tip and a surface modifies the LSP resonances of surface defects that allows tuning them into a desired spectral range [49].

8.1. Influence of surface polaritons on the tunneling current of STM

A strong electromagnetic field can significantly modify electronic processes related to a surface of a metal film, particularly electron tunneling effects [291]. The field enhancement due to surface polariton excitation provides the possibility of observing such electromagnetic field induced effects. In turn, this gives an opportunity to study SPPs by observing the induced electronic processes. The scanning tunneling microscopy technique has been used to investigate the interaction of surface plasmon polaritons with tunneling electrons [41–43]. An additional SPP-induced tunneling current was observed. It was assumed that this current originates due to a rectification of the SPP electromagnetic field caused by the nonlinearity of the current–voltage characteristic of the tunnel junction. The distribution of the SPP induced current over the silver film surface was also measured, and a correlation between this distribution and the topography of the film was found. An influence of surface polaritons, excited on the external and internal interfaces of a gold film, on the STM tunneling current has been also investigated [44]. Images of the SPP induced signal distribution over the film interface taken with different SPP modes have been compared in order to obtain information on the structure of the internal interface. Possible mechanisms which might give

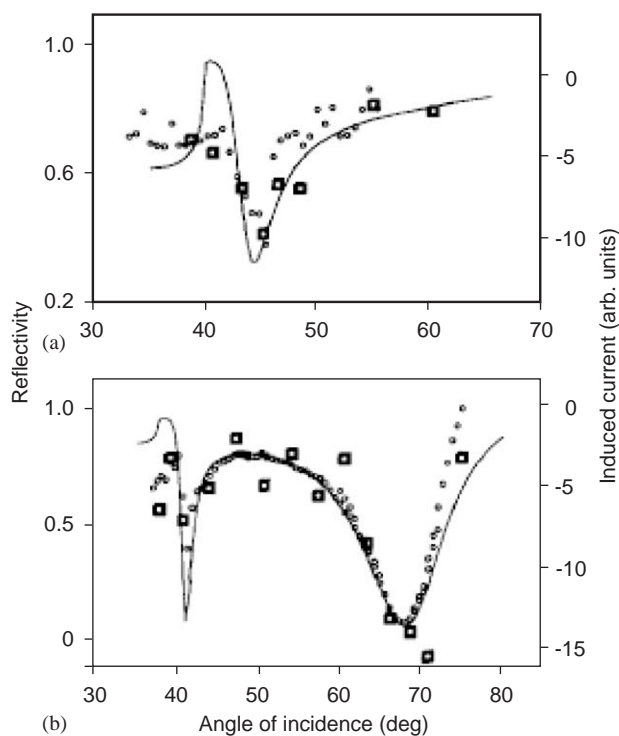


Fig. 90. Angular excitation spectra of the surface polaritons (circles) and angular dependencies of the induced tunneling current (squares) for the Au/glass (a) and Au/fluoride/glass (b) structures. The solid line is the reflectivity calculated with a gold film refractive index $n = 0.13 + 2.7i$. (Ref. [44].)

rise to the SPP induced change in the tunneling current have been discussed. It was shown that the nonlinearity of the current–voltage characteristic of a tunnel junction and the corrections stemming from the electromagnetic rectification of the SPP field give rise to induced tunneling currents of the same order of magnitude.

The angular dependencies of the induced tunneling current and the reflection are shown in Fig. 90 for two samples: a rough gold/glass surface (sample A) and the smooth surface of a gold/fluoride/glass structure (sample B). It is obvious from the measurements that the light induced tunneling is significant only close to the angle of resonant excitation of surface polaritons. The angular dependence of the induced tunneling current follows the SPP excitation spectra. Both surface and interface SPP modes are visible in the tunneling current recorded with the two-layer structure (sample B).

The angular dependent light induced images recorded from the same place of the sample B are illustrated in Fig. 92. In the topographical image one can see three large regions of the film having different heights. The region close to the right edge of the image is probably contaminated since the associated tunneling current is quite noisy. This strong noise in turn leads to an angle independent background signal of the induced current. Even though the tunneling current from the region in the bottom of the topographical image is much less noisy, it was not possible to detect a SPP induced signal in this region. In the top-left region of the image a prominent angular dependent SPP-induced tunneling current is observed at the angles of incidence at which the air–metal or fluoride–metal SPP are excited.

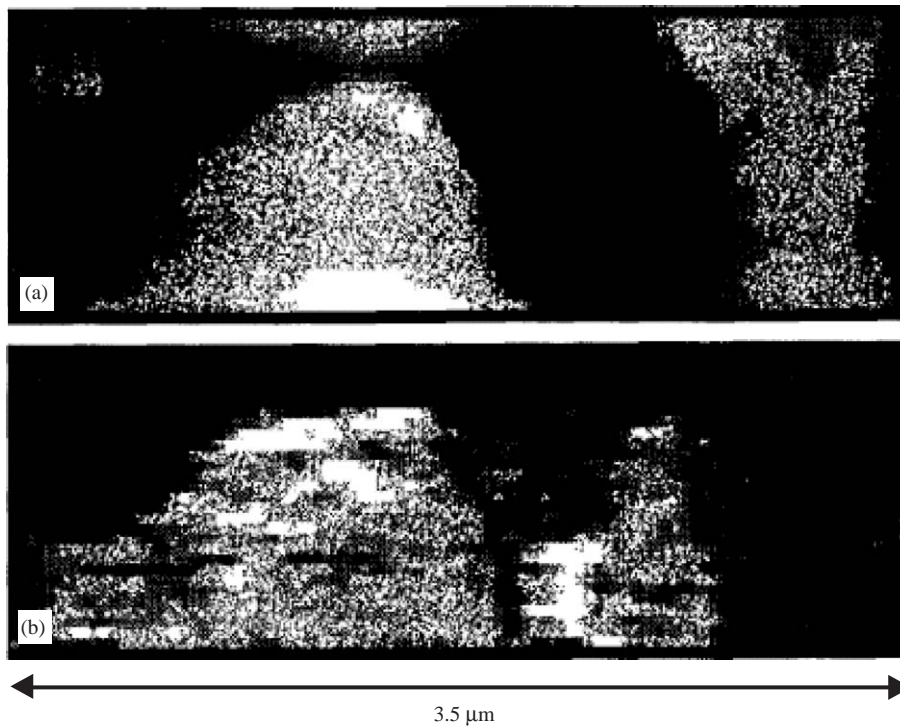


Fig. 91. (a) Topography and (b) the corresponding SPP-induced tunneling current obtained with the sample A under the resonant SPP excitation conditions (cf. Fig. 90a). The gray-scale of the topographical image corresponds to 120 nm. (Ref. [44].)

Detailed studies show that the SPP-induced current variations over a metal surface are related to the local topography. However, the topography is reflected in the SPP-induced images in a nonlinear manner. The topography of the film A (Fig. 91) is typical for a rather inhomogeneous micron-size island structure. The SPP-induced image reveals similar structure and is apparently caused by the topographical variations. The gold film in the two-layers structure (sample B) is flatter than the one in sample A. This results in a less pronounced lateral structure of the air–metal SPP-induced image (Fig. 93). At the same time, a smooth surface facilitates observations of the structure of the internal interface in the fluoride–metal SPP induced image due to the reduction of the background caused by the external surface topography. (As discussed in Section 2, both fast and slow SPPs can be scattered by the surface and internal defects, thus the SPP images induced by them reflect the total defect structure of the film.) The image recorded with the fast SPP is more related to the surface topography of the film than the (slow) fluoride–metal SPP-induced image. In the latter one, the prominent border between the dark and bright regions, which is not correlated with the surface topography, is clearly seen in the top part of the image. The sharpness of the border has been estimated from the edge profile in the SSP image to be equal of about 30 nm (Fig. 93, inset). This sharpness can be considered as an indicator of the resolution of this imaging technique.

The spatial variation of the induced tunneling current can be understood in terms of the lateral distribution of the electromagnetic field over a surface. Comparison of the field distributions detected with an STM and those measured with a SNOM using an uncoated fiber tip (cf. the images in Section 2 obtained

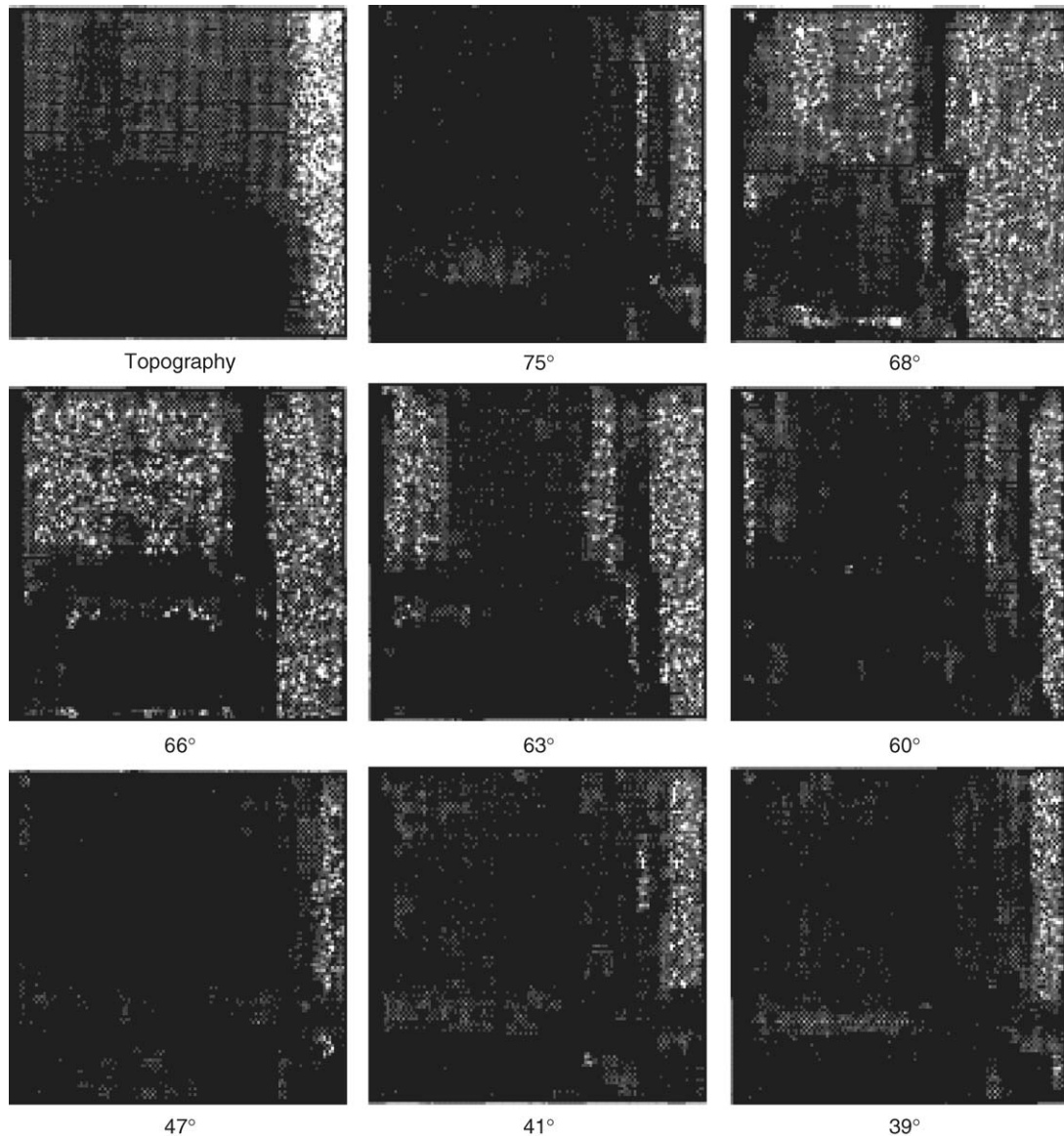


Fig. 92. $1.2 \times 1.2 \mu\text{m}^2$ images of topography and the SPP-induced tunneling current recorded at different angles of incidence with the sample B (cf. Fig. 90b). (Ref. [44].)

for the same type of gold films) reveals distinctly different imaging mechanisms of these techniques. The SNOM images exhibit the SPP interference on the smooth gold films and the SPP localization on the rough gold films, while the SPP-induced STM images directly (although nonlinearly) reflect the surface topography. The signal detected with an uncoated fiber tip is proportional to the intensity distribution of the SPP over a surface. Such a tip made of glass perturbs the SPP field relatively slightly, being scanned at the distance of approximately 10 nm from a surface (shear-force feedback). At the same time, an STM

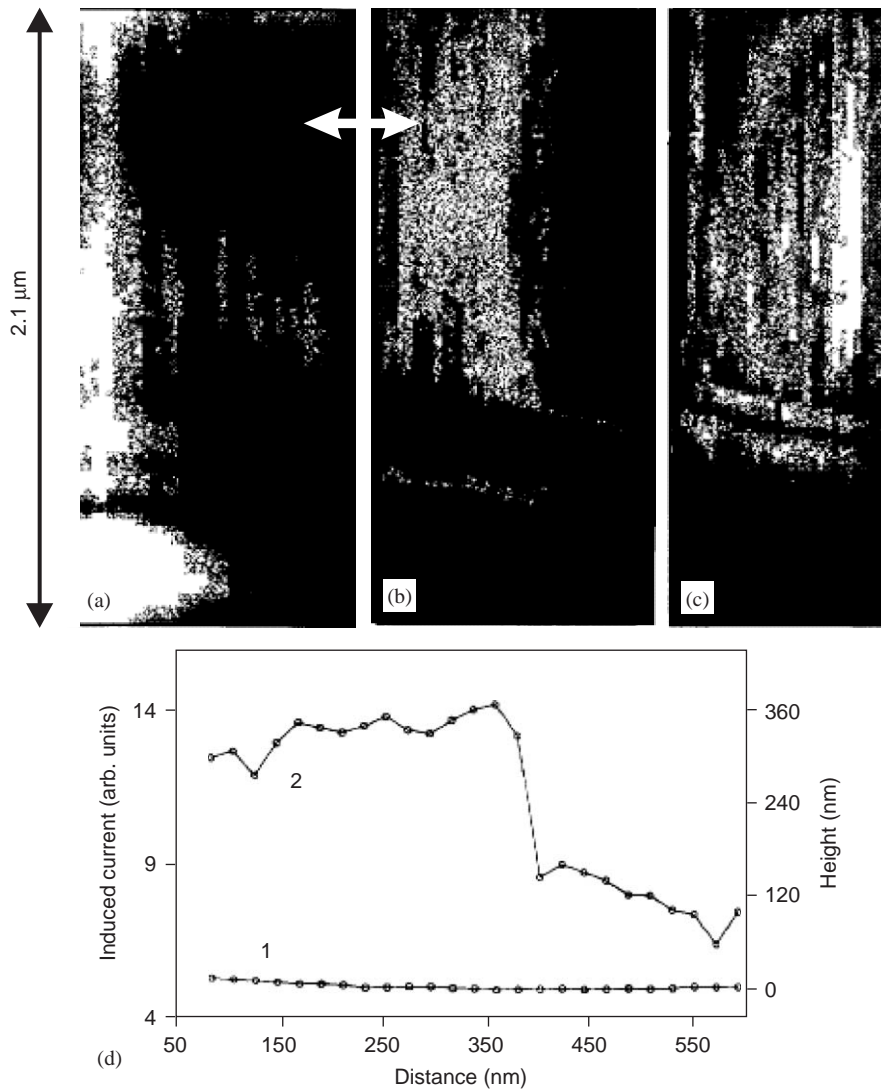


Fig. 93. (a) Topography and (b,c) the induced tunneling current variations obtained with the fluoride–metal (a) and air–metal (b) SPPs (sample B). The gray-scale of the topographical image corresponds to 20 nm. (c) The cross sections of the topography (1) and the fluoride–metal SPP induced image (2) in the direction indicated by the arrow. (Ref. [44].)

tip being metallic has a strong interaction with the SPP field. In addition, with STM feedback, the tip is scanned much closer to a surface (in our case, about 4 Å from it). At such distances the electromagnetic interaction between a metallic tip and a metal surface results in the formation of localized surface plasmons at the tip–surface junction [33,292,293]. Such localized plasmons can lead to a strong enhancement of the electromagnetic field which, in fact, is responsible for the induced tunneling current. Since localization parameters, and thus the field enhancement, are very sensitive to the local topography and depend on it in a complex manner, the induced tunneling current variations, although related to the surface topography

variations, do not directly correspond to it. The light-induced STM image reflects not only the spatial distribution of SPPs, but mainly the properties of a nanometric resonator formed between the tip and the sample. This is generally true for all scanning probe techniques using metal (or metal coated) probes.

Thus, surface polaritons play an intermediate role in light-induced tunneling, the resonant coupling of the illuminating light into SPPs leads to an electromagnetic field enhancement due to the SPP field confinement to a surface followed by the excitation of the localized surface plasmon at the tip–surface junction, providing a further field enhancement.

A separate question needing thorough consideration is the mechanism of the electromagnetic field induced tunneling. Several effects may contribute to the induced tunneling current [41,42]. Some of the effects, such as thermal expansion of the sample and tip due to light absorption, thermovoltage effects, and photoemission, were estimated to be small in the Kretschmann geometry. In this configuration only a small part of a tip is exposed to the light, compared to the experiments in which the tunneling junction is illuminated by the light from the tip side. One of the main contributions that has been considered is the rectification of the electromagnetic field by the nonlinearity of the tunneling current–voltage dependence. It was assumed that the induced current arises from the rectification of the high frequency field via the nonlinear static current–voltage characteristic of a tunnel junction [41], i.e.

$$\Delta I = \frac{1}{4} \Delta V^2 \frac{d^2 I_t}{dV_t^2}, \quad (8.1)$$

where ΔI is the induced current, ΔV is the electromagnetic field induced voltage, and V_t is the tunneling voltage. The experimental data in Refs. [41,43] were explained in this model by setting the induced voltage ΔV to be about 500 mV. This gives an estimate of the electric field in the tunnel junction to be $E \sim 1$ V/nm. Such a large amplitude can be accepted if there is a strong field enhancement of several orders of magnitude in the tip–surface region. Such field enhancement can be achieved with localized surface plasmon excitation in the tip–surface junction [33]. Nonetheless, in these estimates the static current–voltage characteristic of a tunnel junction has been used for the electromagnetic field at optical frequencies as the actual current–voltage characteristic at such frequencies is not known.

Another mechanism of light-induced tunneling has been proposed which does not involve unknown current–voltage dependencies [44]. This mechanism is similar to noise assisted tunneling and, more specifically, to tunneling through a periodically driven potential barrier [294]. The problem of electron motion through the tunnel junction in the presence of a high frequency electromagnetic field is closely related to the problem of the Kapitza pendulum: a pendulum whose point of support oscillates with a high frequency [295]. The effective potential energy for such a pendulum is

$$U_{\text{eff}} = U + \delta U = U + \frac{f^2}{2m\omega^2}, \quad (8.2)$$

where U is the unperturbed potential in the absence of the high frequency force $f \cos(\omega t)$, and m is the mass of the pendulum. Even the vertically upward position of the pendulum can be stable in such an effective potential. Qualitatively, one would expect that the effective potential for an electron in an electromagnetic field is determined by a similar expression

$$U_{\text{eff}}(\mathbf{r}) = U(\mathbf{r}) + \delta U(\mathbf{r}) = U(\mathbf{r}) + \frac{e^2 E_p^2(\mathbf{r})}{2m\omega^2}, \quad (8.3)$$

where $U(\mathbf{r})$ is the potential configuration for tunneling electrons, e is the electron charge, and $E_p(\mathbf{r})$ is the localized plasmon field. This result can be understood easily from a quantum mechanical treatment. If in the Schrödinger equation

$$i\hbar \frac{\partial \psi}{\partial t} = \left(\frac{(\mathbf{p} - e\mathbf{A})^2}{2m} + U(\mathbf{r}) \right) \psi \quad (8.4)$$

for the electron wave function ψ one considers \mathbf{A} to be the vector potential of the localized plasmon field and $U(\mathbf{r})$ to be the potential configuration, the same correction $\delta U(\mathbf{r})$ to the effective potential arises from the DC part of the term $e^2 \mathbf{A}^2 / 2m$ taking into account that $\mathbf{E}_p = -\partial \mathbf{A} / \partial t$. Using the same value as above for the field inside the tunnel junction, $E_p \sim 1 \text{ V/nm}$, this model gives the correction to the tunneling potential $\delta U \sim 10 \text{ mV}$, a value that is only one order of magnitude smaller than the voltage assumed to explain the induced current with nonlinear current–voltage characteristics [41]. Since this is only a rough estimation of the localized surface plasmon field E_p , which enters quadratically in the induced potential δU , one can assume that the effect discussed might be responsible for the observed tunneling current. Under the experimental conditions in Ref. [44], similar estimates of the field needed to explain the induced tunneling current give the value $E_p \sim 0.3 \text{ V/nm}$ in the model of a periodically driven potential barrier. Taking into account the power of the illuminating light and the experimental geometry, a total electromagnetic field enhancement of the order of 10^5 is required for the detection of the observed changes in the tunneling current. Such values of the field enhancement are typical for the surface enhanced Raman scattering experiments and can be achieved with localized surface plasmons.

8.2. Light emission induced by the electron tunneling in STM: a role of surface plasmon polaritons

In addition to optical excitation of surface plasmon polaritons, they can also be excited by an electron beam directed onto a surface [2]. Such an approach allows excitation of SPPs of a wide range of wavelengths including the short-wavelength SPPs ($k_{\text{sp}} \gg k_{\text{photon}}$) which are difficult to launch effectively with light. The excitation of SPPs by an electron beam, their scattering into light by surface defects, and their radiation into a prism have been studied [296]. A similar situation can be realized in a scanning tunneling electron microscope in which a sharp metal tip brought close to a surface acts as a source of electrons tunneling into a metal. With STM it might be possible to achieve a local excitation of SPPs by electrons. In contrast to the SPP excitation by an electron beam, with a STM, however, the tip–surface interaction in many cases might initially result in the excitation of localized surface plasmons in the tunnel gap [297].

As was discussed above, the images of the SPP-induced tunneling current contain mixed information about the SPP field distribution and the properties of localized surface plasmons in the tip–surface gap. Local mapping of localized surface plasmons can provide valuable information on surface chemistry, properties of adsorbed molecules, etc. The studies of the light emission induced by the electron tunneling in STM provide direct probing of inelastic processes in the tunnel gap.

In the first experiments on the STM induced light emission it has been shown that the STM tunnel gap between two metals emits light when a voltage is applied between the tip and the sample. Initially, this effect has been observed in an ultra high vacuum [32,298], but a similar effect under ambient conditions has been detected as well [299–304]. It has been demonstrated that a variety of STM operation modes

including isochromat spectroscopy, optical emission spectroscopy, and photon mapping can be used simultaneously with topographical measurements. A combination of several imaging modes can give additional information on local surface properties. The ability of this technique to provide the chemical surface data on the nanometer scale has been shown in the studies of light emission from small metallic particles on semiconductor substrates [305,306]. In different experiments with different types of surfaces (metal, semiconductor, or adsorbed molecular layer on a surface), the nature of the STM induced light emission can be different due to the different processes by which the tunneling electrons interact with a surface. The STM induced light emission from metal surfaces is believed to involve localized surface plasmons which are excited in the gap between the tip and the sample by the tunneling electrons. The decay of this LSP into light either directly or via excitation of surface polaritons, which then are reradiated into light results in the observed light emission.

The quantum efficiency of the STM induced light emission from metal surfaces can be rather high (of the order of 10^{-3} – 10^{-4}). At the same time, the small currents in the STM experiments result in a total number of detected photons of the order of 10^3 – 10^4 photons/s.

The typical STM induced photon map of a freshly prepared gold film under ambient conditions is presented in Fig. 94. This image was taken at a tunneling voltage $V_t = 1.8$ V by the use of a gold tip [307]. The number of detected photons varies with the tunneling current. The averaged count rate of about 10^3 photons per second was detected at the tunneling current of 5 nA. The light emission intensity varies from one flat surface terrace to another. The monatomic steps between terraces appeared as dark regions on the light emission maps. Similar effects have also been detected for the Cu(1 1 1) film surface under UHV conditions [305].

The spectrum of light emission from the tunnel gap scales with the mean plasma frequency of the STM tip–sample system both in air and in vacuum [299,300,305]. Under ambient conditions, however, the much stronger dependence of the tip–sample distance on the tunneling voltage results in quite different shapes of the dependencies $L(V)$ of the light intensity on the applied voltage [308]. Under UHV conditions these dependencies look almost the same for any pair of metals used as a tip and a sample [305]. On the other hand, in air each spectral peak of the STM induced light emission corresponds to a distinct peak in the $L(V)$ dependence [299,300]. Figure 95 presents $L(V)$ curves measured for different pairs of metals. These spectral peaks can be explained in terms of different localized surface plasmon resonances in the gap between the tip and the surface [33,297].

The exact theoretical description of the experimental situation is somewhat difficult since a correct theory of inelastic electron tunneling must take into account the spatial dispersion of the dielectric function of the electron gas and, hence, nonlocal effects. Even in the simple plane geometry corrections stemming from the spatial dispersion effects can not be neglected [309]. Nevertheless, it was established that plasma excitations with the characteristic wavevectors $k_p \sim 1 \text{ nm}^{-1}$ do play a main role in the STM induced light emission from metal surfaces. The inelastic electron tunneling seems to be responsible for inducing these collective electron excitations [310,311]. It is not clear whether these plasma excitations are localized surface plasmons or surface plasmon polaritons in the tail of the SPP dispersion relation. In the latter case, the SPP can be scattered into photons leading to the observed light emission. In the former case, LSP can decay into light either directly or via the excitation of SPPs. The localized surface plasmon decay into SPP has proven to be an effective channel for the resonance damping of LSP [30]. The distribution of the electron tunneling induced light over molecular crystals deposited on a metal surface (such as CuPc/Au structures) can be interpreted as the Bragg diffraction of the excited surface polaritons on the lattice of a 2D molecular crystal [307].

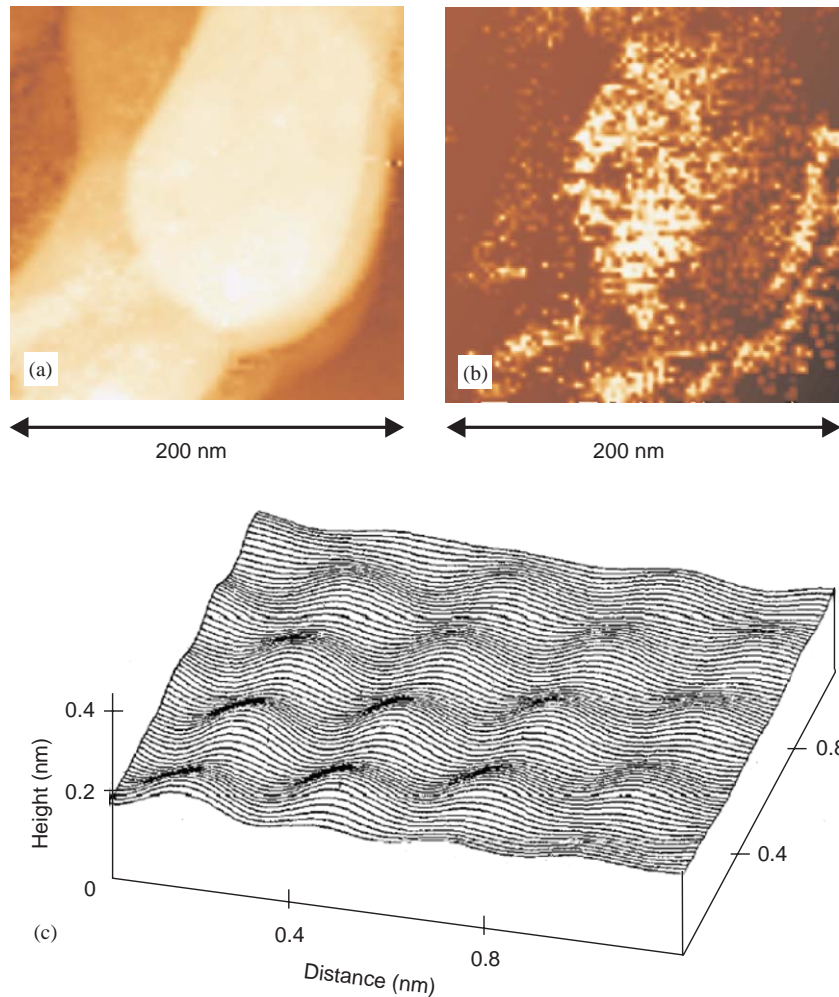


Fig. 94. (a) Topography and (b) STM-induced light emission image of a gold surface. The gray-scale of the topographic image corresponds to 3 nm. The atomically resolved image (c) is taken at the flat terrace as seen in (a). (Ref. [307].)

Molecular [307,312] and even atomic [313] resolution has been demonstrated in the electron tunneling induced light emission experiments. This suggests that optical spectroscopic analysis of the emission on the scale of individual molecules can be achieved. This possibility opens exciting prospects of chemical surface analysis with molecular-scale resolution.

9. Conclusion

In this review we have tried to present and generalise theoretical and experimental results of near-field investigations of surface plasmon polaritons. Direct interrogation of SPP behavior on a surface has allowed researchers to obtain insight into SPP-related processes, to make direct comparison to theoretical

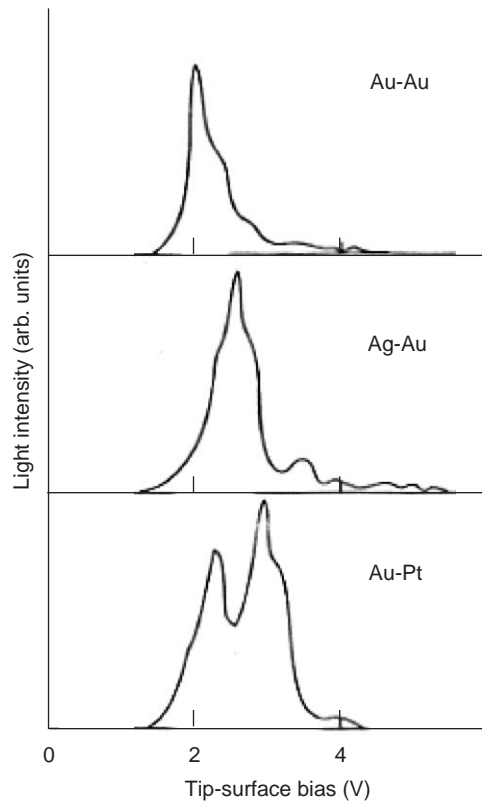


Fig. 95. Dependencies of the emitted light intensity on the tunneling voltage for *Au–Au*, *Ag–Au*, and *Au–Pt* tunnel junctions. The first symbol in the pairs corresponds to the tip material. (Ref. [299].)

models, and in turn to develop new models and concepts. Such concepts and the combination of SNOM and SPP techniques have already led to a number of qualitatively new applications and novel photonic technologies which can play a major role in the future of optical communications and optical computing. They will be indispensable in many areas of future nanotechnologies.

The term “near-field” reflects the fact that we deal with surface excitations whose electromagnetic field decays exponentially from a surface. Only with the development of near-field optical microscopy did it become possible to study these surface excitations directly. This led to the development of surface plasmon polariton optics with eventually all components with functions similar to conventional 3D optical elements. The approach to routing and manipulation of optical signals using surface plasmon polaritons can provide significant advantages in all-optical integration due to their intrinsic 2D nature. Moreover, these excitations result in a strong electromagnetic field enhancement that facilitates implementation of low-light intensity nonlinear effects in all-optical active photonic elements.

Acknowledgements

We would like to thank all our colleagues and friends who in one or another way contributed to the progress of the work presented in this review, among them are W. Atia, S.I. Bozhevolnyi, S.A. Darmanyany,

C.C. Davis, W. Dickson, J. Elliott, F. de Fornel, D. Gérard, F. Grillot, A. Güngör, O. Keller, T. A. Leskova, J. Mait, D. Mazzoni, L. Salomon, A. Stanishevsky, N.I. Zheludev. We are indebted to V.M. Agranovich for numerous fruitful discussions.

References

- [1] V.M. Agranovich, D.L. Mills (Eds.), *Surface Polaritons*, North-Holland, Amsterdam, 1982.
- [2] H. Raether, *Surface Plasmons*, Springer-Verlag, Berlin, 1988.
- [3] A.D. Boardman (Ed.), *Electromagnetic Surface Modes*, John Wiley & Sons, New York, 1982.
- [4] G. Boisde, A. Harmer, *Chemical and Biochemical Sensing with Optical Fibers and Waveguides*, Artech House, Boston, 1996.
- [5] H.-E. Ponath, G.I. Stegeman (Eds.), *Nonlinear Surface Electromagnetic Phenomena*, North-Holland, Amsterdam, 1991.
- [6] W.L. Barnes, A. Dereux, T.W. Ebbesen, *Nature* 424 (2003) 824.
- [7] A.V. Zayats, I.I. Smolyaninov, *J. Opt. A: Pure Appl. Opt.* 5 (2003) S16.
- [8] F. Keilmann, *Proc. SPIE* 1029 (1988) 50.
- [9] D.W. Pohl, D. Courjon (Eds.), *Near Field Optics*, Kluwer, The Netherlands, 1993.
- [10] R.C. Reddick, R.J. Warmack, T.L. Ferrel, *Phys. Rev. B* 39 (1989) 767.
- [11] D. Courjon, K. Sarayeddine, M. Spajer, *Opt. Commun.* 71 (1989) 23.
- [12] F. de Fornel, J.P. Goudonnet, L. Salomon, E. Lesniewska, *Proc. SPIE* 1139 (1989) 77.
- [13] O. Marti, H. Bielefeldt, B. Hecht, S. Herminhaus, P. Leiderer, J. Mlynek, *Opt. Commun.* 96 (1993) 225.
- [14] P.M. Adam, L. Salomon, F. de Fornel, J.P. Goudonnet, *Phys. Rev. B* 48 (1993) 2680.
- [15] P. Dawson, F. de Fornel, J.-P. Goudonnet, *Phys. Rev. Lett.* 72 (1994) 2927.
- [16] D.P. Tsai, J. Kovasc, Z. Wang, M. Moskovits, V.M. Shalaev, J.S. Suh, R. Botet, *Phys. Rev. Lett.* 72 (1994) 4149.
- [17] S.I. Bozhevolnyi, I.I. Smolyaninov, A.V. Zayats, *Phys. Rev. B* 51 (1995) 17916.
- [18] S.I. Bozhevolnyi, B. Vohnsen, I.I. Smolyaninov, A.V. Zayats, *Opt. Commun.* 117 (1995) 417.
- [19] B. Hecht, H. Bielefeldt, L. Novotny, Y. Inouye, D.W. Pohl, *Phys. Rev. Lett.* 77 (1996) 1889.
- [20] S.I. Bozhevolnyi, B. Vohnsen, A.V. Zayats, in: M. Nieto-Vesperinas, N. García (Eds.), *Optics at the Nanometer Scale*, Kluwer Academic Publ., Dordrecht, 1996, p. 163.
- [21] N.E. Glass, M. Weber, D.L. Mills, *Phys. Rev. B* 29 (1984) 6548.
- [22] B. Laks, D.L. Mills, A.A. Maradudin, *Phys. Rev. B* 29 (1981) 4965.
- [23] W.L. Barnes, T.W. Preist, S.C. Kitson, J.R. Sambles, *Phys. Rev. B* 54 (1996) 6227.
- [24] L. Salomon, F. Grillot, F. de Fornel, A.V. Zayats, *Phys. Rev. Lett.* 86 (2001) 1110.
- [25] S.I. Bozhevolnyi, J. Erland, K. Leosson, P.M.W. Skovgaard, J.M. Hvam, *Phys. Rev. Lett.* 86 (2001) 3008.
- [26] M. Kretschmann, A.A. Maradudin, *Phys. Rev. B* 66 (2002) 245408.
- [27] R.H. Ritchie, *Phys. Rev.* 106 (1957) 874.
- [28] U. Kreibig, M. Vollmer, *Optical Properties of Metal Clusters*, Springer, Berlin, 1995.
- [29] A.V. Shchegrov, I.V. Novikov, A.A. Maradudin, *Phys. Rev. Lett.* 78 (1997) 4269.
- [30] V.M. Agranovich, V.E. Kravtsov, T.A. Leskova, *Solid State Commun.* 47 (1983) 925.
- [31] M. Moskovits, *Rev. Mod. Phys.* 57 (1985) 783.
- [32] S. Ushioda, Y. Uehara, M. Kuwahara, *Appl. Surf. Sci.* 60 & 61 (1992) 448.
- [33] A.G. Malshukov, *Phys. Rep.* 194 (1990) 343.
- [34] D.L. Mills, *Phys. Rev. B* 65 (2002) 125419.
- [35] V.M. Shalaev, *Nonlinear Optics of Random Media*, Springer, Berlin, 2000.
- [36] I.I. Smolyaninov, A.V. Zayats, A. Güngör, C.C. Davis, *Phys. Rev. Lett.* 88 (2002) 187402.
- [37] A.V. Zayats, V. Sandoghdar, *Opt. Commun.* 178 (2000) 245.
- [38] E. Kretschmann, H. Raether, *Z. Naturforsch. A* 23 (1968) 2135.
- [39] A. Otto, *Z. Physik* 216 (1968) 398.
- [40] L. Salomon, G. Bassou, J.P. Dufour, F. de Fornel, A.V. Zayats, *Phys. Rev. B* 65 (2002) 125409.
- [41] R. Möller, U. Albrecht, J. Boneberg, B. Koslowski, P. Leiderer, K. Dransfeld, *J. Vac. Sci. Technol. B* 9 (1991) 506.
- [42] N. Kroo, J.P. Thost, M. Völcker, W. Krieger, H. Walther, *Europhys. Lett.* 15 (1991) 289.

- [43] C. Baur, B. Koslowski, R. Möller, K. Dransfeld, in: D.W. Pohl, D. Courjon (Eds.), *Near Field Optics*, Kluwer, The Netherlands, 1993, p. 325.
- [44] I.I. Smolyaninov, A.V. Zayats, O. Keller, *Phys. Lett. A* 200 (1995) 438.
- [45] M. Specht, J.D. Pedarnig, W.M. Heckl, T.W. Hansch, *Phys. Rev. Lett.* 68 (1992) 476.
- [46] R.B.G. de Hollander, N.F. van Hulst, R.P.H. Kooyman, *Ultramicroscopy* 57 (1995) 263.
- [47] Y.-K. Kim, J.B. Ketterson, D.J. Morgan, *Opt. Lett.* 21 (1996) 165.
- [48] R.W. Rendell, D.J. Scalapino, *Phys. Rev. B* 24 (1981) 3276.
- [49] A.V. Zayats, *Opt. Commun.* 161 (1999) 156.
- [50] E. Betzig, P.L. Finn, J.S. Weiner, *Appl. Phys. Lett.* 60 (1992) 2484.
- [51] R. Toledo-Crow, P.C. Yang, Y. Chen, M. Vaez-Iravani, *Appl. Phys. Lett.* 60 (1992) 2957.
- [52] E. Betzig, J.K. Trautman, T.D. Harris, J.S. Weiner, R.L. Kostelak, *Science* 251 (1991) 1468.
- [53] D. van Labeke, D. Barchiesi, *J. Opt. Soc. Am. A* 10 (1993) 2193.
- [54] S.I. Bozhevolnyi, *J. Microsc.* 202 (2002) 313.
- [55] R. Carminati, J.-J. Greffet, *Opt. Commun.* 116 (1995) 316.
- [56] J.-J. Greffet, R. Carminati, *Prog. Surf. Sci.* 56 (1997) 133.
- [57] S. Bozhevolnyi, B. Vohnsen, *J. Opt. Soc. Am. B* 14 (1997) 1656.
- [58] I.I. Smolyaninov, D.L. Mazzoni, C.C. Davis, *Phys. Rev. Lett.* 77 (1996) 3877.
- [59] I.I. Smolyaninov, D.L. Mazzoni, J. Mait, C.C. Davis, *Phys. Rev. B* 56 (1997) 1601.
- [60] S.I. Bozhevolnyi, F.A. Pudonin, *Phys. Rev. Lett.* 78 (1997) 2823.
- [61] S.I. Bozhevolnyi, V.A. Markel, V. Coello, W. Kim, V.M. Shalaev, *Phys. Rev. B* 58 (1998) 11441.
- [62] A. Bouhelier, Th. Huser, J.M. Freyland, H.-J. Güntherodt, D.W. Pohl, *J. Microsc.* 194 (1999) 571.
- [63] P. Dawson, B.A.F. Puygranier, J.-P. Goudonnet, *Phys. Rev. B* 63 (2001) 205410.
- [64] D. Peyrade, R. Quidant, J.C. Weeber, A. Dereux, G. Leveque, C. Girard, Y. Chen, *Europhys. Lett.* 56 (2001) 517.
- [65] J.A. Sánchez-Gil, *Appl. Phys. Lett.* 73 (1998) 3509.
- [66] J.A. Sánchez-Gil, A.A. Maradudin, *Phys. Rev. B* 60 (1999) 8359.
- [67] J.A. Sánchez-Gil, *Phys. Rev. B* 53 (1996) 10317.
- [68] A.A. Maradudin, in: M.P. Das (Ed.), *Topics in Condensed Matter Physics*, Nova, New York, 1994, p. 33.
- [69] G.C. Brown, V. Celli, M. Haller, A. Marvin, *Surf. Sci.* 136 (1984) 381.
- [70] Lord Rayleigh, *The Theory of Sound*, vol. II, second ed., London, Macmillan, 1896, pp. 89, 297–311.
- [71] D.L. Mills, *Phys. Rev. B* 12 (1975) 4036.
- [72] I.S. Gradshteyn, I.M. Ryzhik, *Tables of Integrals, Sums, Series, and Products*, State Publishing House for Physico-Mathematical Literature, Moscow, 1963, p. 732, formula 6.623.2.
- [73] A.A. Maradudin, W.M. Visscher, *Z. Phys. B* 60 (1985) 215.
- [74] Z. Schlesinger, A.J. Sievers, *Appl. Phys. Lett.* 36 (1980) 409.
- [75] T.A. Leskova, N.I. Gapotchenko, *Solid State Commun.* 53 (1985) 351.
- [76] V.M. Agranovich, T.A. Leskova, *Prog. Surf. Sci.* 29 (1987) 169.
- [77] A.A. Maradudin, R.F. Wallis, G.I. Stegeman, *Prog. Surf. Sci.* 33 (1990) 171.
- [78] F. Pincemin, A.A. Maradudin, A.D. Boardman, J.-J. Greffet, *Phys. Rev. B* 50 (1994) 15261.
- [79] B. Rothenhäusler, W. Knoll, *Surf. Sci.* 191 (1987) 585.
- [80] B. Rothenhäusler, W. Knoll, *Appl. Phys. Lett.* 51 (1987).
- [81] B. Rothenhäusler, W. Knoll, *Nature* 332 (1988) 615.
- [82] B. Rothenhäusler, W. Knoll, *Appl. Phys. Lett.* 52 (1988) 1554.
- [83] B. Rothenhäusler, W. Knoll, *J. Opt. Soc. Am. B* 5 (1988) 1401.
- [84] C.E.H. Berger, R.P.H. Kooyman, J. Greve, *Opt. Commun.* 167 (1999) 183.
- [85] T.A. Leskova, A.A. Maradudin, W. Zierau, *Proc. SPIE* 4100 (2000) 1.
- [86] T.A. Leskova, A.A. Maradudin, W. Zierau, Paper B1-8, Abstracts of the National Radio Science Meeting, 9–12 January, 2002, Boulder, CO, p. 38.
- [87] R. Mitra, S.W. Lee, *Analytical Techniques in the Theory of Waveguides*, MacMillan, New York, 1971.
- [88] W.R. Dean, *Proc. Camb. Philos. Soc.* 44 (1948) 483.
- [89] F. Ursell, *Proc. Camb. Philos. Soc.* 46 (1950) 141.
- [90] F. Ursell, *Proc. Camb. Philos. Soc.* 46 (1950) 153.
- [91] H. Levine, *J. Math. Phys.* 6 (1965) 1231.

- [92] A.M.J. Davis, M.J. Hood, *SIAM J. Appl. Math.* 31 (1976) 16.
- [93] A.M.J. Davis, F.G. Leppington, *Quart. J. Mech. Appl. Math.* 32 (1978) 19.
- [94] A.L. Cullen, *Electron. Lett.* 11 (1975) 479.
- [95] A.M.J. Davis, F.G. Leppington, *Proc. Roy. Soc. London Ser. A* 353 (1977) 55.
- [96] A.M.J. Davis, F.G. Leppington, *Proc. Roy. Soc. London Ser. A* 358 (1977) 243.
- [97] A.M.J. Davis, *Proc. Roy. Soc. London Ser. A* 391 (1984) 181.
- [98] R.F. Harrington, *Field Computation by Moment Methods*, Macmillan, New York, 1968.
- [99] F. Pincemin, A. Sentenac, J.-J. Greffet, *J. Opt. Soc. Am. A* 11 (1994) 1117.
- [100] I.I. Smolyaninov, D.L. Mazzoni, C.C. Davis, *Appl. Phys. Lett.* 67 (1995) 3859–3861.
- [101] P. Sheng, *Introduction to Wave Scattering, Localization, and Mesoscopic Phenomena*, Academic Press, New York, 1995.
- [102] M.P. van Albada, M.B. van der Mark, A. Langendijk, in: P. Sheng (Ed.), *Scattering and Localization of Classical Waves in Random Media*, World Scientific, Singapore, 1990, p. 97.
- [103] K.M. Watson, *J. Math. Phys.* 10 (1969) 688.
- [104] G. Bergmann, *Phys. Rev. B* 28 (1983) 2914.
- [105] D.E. Khmel'nitskii, *Physica* 126 B (1984) 235.
- [106] E. Abrahams, P.W. Anderson, D.C. Licciardello, I.V. Ramakrishnan, *Phys. Rev. Lett.* 42 (1979) 673.
- [107] P.W. Anderson, *Phys. Rev.* 109 (1958) 1493.
- [108] Y. Kuga, A. Ishimaru, *J. Opt. Soc. Am. A* 1 (1984) 831.
- [109] L. Tsang, A. Ishimaru, *J. Opt. Soc. Am. A* 1 (1984) 836 (A2 (1985) 2187).
- [110] M.P. van Albada, A. Lagendijk, *Phys. Rev. Lett.* 55 (1985) 2692.
- [111] P.E. Wolf, G. Maret, *Phys. Rev. Lett.* 55 (1985) 2696.
- [112] A.R. McGurn, A.A. Maradudin, V. Celli, *Phys. Rev. B* 31 (1985) 4866.
- [113] C.S. West, K.A. O'Donnell, *J. Opt. Soc. Am. A* 12 (1995) 390.
- [114] A.A. Maradudin, A.R. McGurn, E.R. Méndez, *J. Opt. Soc. Am. A* 12 (1995) 2500.
- [115] A.V. Shchegrov, *Phys. Rev. B* 57 (1998) 4132.
- [116] G.C. Brown, V. Celli, M. Haller, A.A. Maradudin, A. Marvin, *Phys. Rev. B* 31 (1985) 4993.
- [117] A.A. Maradudin, W. Zierau, *Phys. Rev. B* 14 (1976) 484.
- [118] A.A. Maradudin, E.R. Méndez, *Appl. Opt.* 32 (1993) 3335.
- [119] Z. Schlesinger, A.J. Sievers, *Solid State Commun.* 43 (1982) 671 A.J. Sievers, Z. Schlesinger, *J. Phys. Colloq.* 44 (1983) C10.
- [120] T.R. Michel, M.E. Knotts, K.A. O'Donnell, *J. Opt. Soc. Am. A* 12 (1995) 548.
- [121] O.A. Aktsipetrov, V.N. Golovkina, O.I. Kapusta, T.A. Leskova, N.N. Novikova, *Phys. Lett. A* 170 (1991) 231.
- [122] A.R. McGurn, T.A. Leskova, V.M. Agranovich, *Phys. Rev. B* 44 (1991) 11441.
- [123] K. Arya, Z.B. Su, J.L. Birman, *Phys. Rev. Lett.* 54 (1985) 1559.
- [124] P.A. Lee, T.V. Ramakrishnan, *Rev. Mod. Phys.* 57 (1985) 287.
- [125] S. John, H. Sompolinsky, M.J. Stephen, *Phys. Rev. B* 27 (1983) 5592.
- [126] S. John, *Phys. Rev. Lett.* 53 (1984) 2169.
- [127] J.A. Sánchez-Gil, *Phys. Rev. B* 68 (2003) 113410.
- [128] J.J. Burke, G.I. Stegeman, T. Tamir, *Phys. Rev. B* 33 (1986) 5186.
- [129] T.A. Leskova, *Solid State Commun.* 50 (1984) 869.
- [130] V.M. Agranovich, *Zh. Eksp. Teor. Fiz.* 77 (1979) 1124 [*Sov. Phys.—JETP* 50 (1979) 567].
- [131] V.M. Agranovich, V.E. Kravtsov, T.A. Leskova, *Zh. Eksp. Teor. Fiz.* 81 (1981) 1828 [*Sov. Phys.—JETP* 54 (1981) 968].
- [132] L.A. Vainshtein, *The Factorization Method in Diffraction*, Soviet Radio, Moscow, 1966.
- [133] V.M. Agranovich, T.A. Leskova, *Prog. Surf. Sci.* 29 (1988) 169.
- [134] V.V. Shevchenko, *Continuous Transitions in Open Waveguides*, Golem, Boulder, Colorado, 1971.
- [135] G.I. Stegeman, A.A. Maradudin, T.S. Rahman, *Solid State Commun.* 38 (1981) 477 *Phys. Rev. B* 23 (1981) 2576.
- [136] A.I. Voronko, L.G. Klimova, G.N. Shkerdin, *Solid State Commun.* 61 (1987) 361.
- [137] G.I. Stegeman, A.A. Maradudin, T.P. Shen, R.F. Wallis, *Phys. Rev. B* 29 (1984) 6530.
- [138] C.A. Ward, K. Bhasin, R.J. Bell, R.W. Alexander, I. Tyler, *J. Chem. Phys.* 62 (1975) 1674.
- [139] V.M. Agranovich, *Sov. Phys. JETP* 50 (1979) 567.
- [140] F. Abelés, T. Lopez-Rios, in: V.M. Agranovich, D.L. Mills (Eds.), *Surface Polaritons*, North-Holland, Amsterdam, 1982, p. 252.

- [141] J.G. Gordon, J.D. Swalen, *Opt. Commun.* 22 (1977) 374.
- [142] I.V. Novikov, A.A. Maradudin, *Phys. Rev. B* 66 (2002) 035403.
- [143] J.R. Krenn, M. Salerno, N. Felidj, B. Lamprecht, G. Schider, A. Leitner, F.R. Aussenegg, J.C. Weeber, A. Dereux, J.P. Goudonnet, *J. Microsc.* 202 (2000) 122.
- [144] J.C. Weeber, J.R. Krenn, A. Dereux, B. Lamprecht, Y. Lacroute, J.P. Goudonnet, *Phys. Rev. B* 64 (2001) 045411.
- [145] J.C. Weeber, Y. Lacroute, A. Dereux, *Phys. Rev. B* 68 (2003) 115401.
- [146] P. Berini, *Phys. Rev. B* 61 (2000) 10484.
- [147] T. Nikolajsen, K. Leosson, I. Salakhutdinov, S.I. Bozhevolnyi, *Appl. Phys. Lett.* 82 (2003) 668.
- [148] M. Quinten, A. Leitner, J.R. Krenn, F.R. Aussenegg, *Opt. Lett.* 23 (1998) 1331.
- [149] K.A. O'Donnell, in: A. Consortini (Ed.), *Current Trends in Optics*, ICO Series, 1995.
- [150] A.A. Maradudin, in: V.M. Agranovich, D.L. Mills (Eds.), *Surface Polaritons*, North-Holland, Amsterdam, 1982, p. 405.
- [151] H. Raether, in: V.M. Agranovich, D.L. Mills (Eds.), *Surface Polaritons*, North-Holland, Amsterdam, 1982, p. 331.
- [152] B.B. Mandelbrot, *The Fractal Geometry of Nature*, Freeman, San Francisco, 1982.
- [153] S.B. Bozhevolnyi, A.V. Zayats, B. Vohnsen, I.I. Smolyaninov, *Surf. Sci.* 356 (1996) 268.
- [154] L. Spanos, E.A. Irene, *J. Vac. Sci. Technol. A* 12 (1994) 2646.
- [155] V.A. Markel, L.S. Muratov, M.I. Stockman, T.F. George, *Phys. Rev. B* 43 (1991) 8183.
- [156] K.J. Falconer, *Fractal Geometry: Mathematical Foundations and Applications*, Wiley, Chichester, 1990.
- [157] J.D. Joannopoulos, R.D. Meade, J.N. Winn, *Photonic Crystals*, Princeton University Press, Princeton, 1995.
- [158] A.A. Maradudin, A.R. McGurn, in: C.M. Soukoulis (Ed.), *Photonic Band Gaps and Localization*, Plenum Press, New York, 1993, p. 247.
- [159] C. Kittel, *Introduction to Solid State Physics*, Wiley, New York, 1976.
- [160] R.A. Watts, J.B. Harris, A.P. Hibbins, T.W. Preist, J.R. Sambles, *J. Mod. Opt.* 43 (1996) 1351.
- [161] V. Celli, P. Tran, A.A. Maradudin, D.L. Mills, *Phys. Rev. B* 37 (1988) 9089.
- [162] W.-C. Tan, T.W. Preist, J.R. Sambles, N.P. Wanstall, *Phys. Rev. B* 59 (1999) 12661.
- [163] W.L. Barnes, S.C. Kitson, T.W. Preist, J.R. Sambles, *J. Opt. Soc. Am. A* 14 (1997) 1654.
- [164] S.C. Kitson, W.L. Barnes, J.R. Sambles, *Phys. Rev. Lett.* 77 (1996) 2670.
- [165] W.L. Barnes, *J. Mod. Opt.* 45 (1998) 661.
- [166] S.I. Bozhevolnyi, V.S. Volkov, K. Leosson, A. Boltasseva, *Appl. Phys. Lett.* 79 (2001) 1076.
- [167] I.I. Smolyaninov, A.V. Zayats, C.C. Davis, *Appl. Phys. Lett.* 81 (2002) 3314.
- [168] I.I. Smolyaninov, A.V. Zayats, A. Stanishevsky, C.C. Davis, *Phys. Rev. B* 66 (2002) 205414.
- [169] P.M. van den Berg, J.C.M. Borburgh, *Appl. Phys.* 3 (1974) 55.
- [170] F. Toigo, A. Marvin, V. Celli, N.R. Hill, *Phys. Rev. B* 15 (1977) 5618.
- [171] S.R. Seshadri, *J. Appl. Phys.* 57 (1985) 4874.
- [172] E.P. DaSilva, G.A. Farias, A.A. Maradudin, *J. Opt. Soc. Am. A* 4 (1987) 2022.
- [173] R.H. Ritchie, E. Arakawa, J.J. Cowan, *Phys. Rev. Lett.* 21 (1968) 1530.
- [174] W. Knoll, M.R. Philpott, J.D. Swalen, A. Girlando, *J. Chem. Phys.* 77 (1982) 2254.
- [175] Y.J. Chen, E.S. Koteles, R.J. Seymour, G.J. Sonek, I.M. Ballantyne, *Solid State Commun.* 46 (1983) 95.
- [176] D. Heitmann, N. Kroo, C. Schulz, Z. Szentirmay, *Phys. Rev. B* 35 (1987) 2660.
- [177] H. Raether, *Surface Plasmons*, Springer, Berlin, 1988, p. 113.
- [178] B. Fischer, T.M. Fischer, W. Knoll, *J. Appl. Phys.* 75 (1994) 1577.
- [179] H. Lochbihler, *Phys. Rev. B* 50 (1994) 4795.
- [180] W.L. Barnes, T.W. Preist, S.C. Kitson, J.R. Sambles, N.P.K. Cotter, D.J. Nash, *Phys. Rev. B* 51 (1995) 11164.
- [181] S.C. Kitson, W.L. Barnes, G.W. Bradberry, J.R. Sambles, *J. Appl. Phys.* 79 (1996) 7383.
- [182] D. Maystre, in: E. Wolf (Ed.), *Progress in Optics*, vol. XXI, North-Holland, Amsterdam, 1984, p. 3.
- [183] D.L. Mills, *Phys. Rev. B* 15 (1997) 3097.
- [184] S.R. Seshadri, *J. Appl. Phys.* 58 (1985) 1733.
- [185] A.A. Maradudin, in: R.F. Wallis, G.I. Stegeman (Eds.), *Electromagnetic Surface Excitations*, Springer, Berlin, 198657 (Surface, North-Holland, Amsterdam).
- [186] R.F. Millar, *Proc. Camb. Phil. Soc.* 65 (1969) 773.
- [187] P.M. van den Berg, J.T. Fokkema, *J. Opt. Soc. Am.* 69 (1979) 27.
- [188] M. Kretschmann, T.A. Leskova, A.A. Maradudin, *Opt. Commun.* 215 (2003) 205.
- [189] M. Kretschmann, A.A. Maradudin, *Phys. Rev. B* 66 (2002) 245408.

- [190] I.I. Smolyaninov, W. Atia, C.C. Davis, *Phys. Rev. B* 59 (1999) 2454.
- [191] S.I. Bozhevolnyi, V.S. Volkov, K. Leosson, *Opt. Commun.* 196 (2001) 41.
- [192] A.V. Zayats, I.I. Smolyaninov, W. Dickson, C.C. Davis, *Appl. Phys. Lett.* 82 (2003) 4438.
- [193] H. Kosaka, T. Kawashima, A. Tomita, M. Notomi, T. Tamamura, T. Sato, S. Kawakami, *Phys. Rev. B* 58 (1998) R10096.
- [194] H. Kosaka, T. Kawashima, A. Tomita, M. Notomi, T. Tamamura, T. Sato, S. Kawakami, *Appl. Phys. Lett.* 74 (1999) 1370.
- [195] K.B. Chung, S.W. Hong, *Appl. Phys. Lett.* 81 (2002) 1549.
- [196] C. Luo, S.G. Johnson, J.D. Joannopoulos, *Appl. Phys. Lett.* 81 (2002) 2352.
- [197] A.P. Hibbins, J.R. Sambles, C.R. Lawrence, *Appl. Phys. Lett.* 80 (2002) 2410.
- [198] T.W. Ebbesen, J. Lezec, H.F. Ghaemi, T. Thio, P.A. Wolff, *Nature (London)* 391 (1998) 667.
- [199] L. Martin-Moreno, F.J. García-Vidal, H.J. Lezec, K.M. Pellerin, T. Thio, J.B. Pendry, T.W. Ebbesen, *Phys. Rev. Lett.* 86 (2001) 1114.
- [200] S.A. Darmanyan, A.V. Zayats, *Phys. Rev. B* 67 (2003) 035424.
- [201] A.M. Dykhne, A.K. Sarychev, V.M. Shalaev, *Phys. Rev. B* 67 (2003) 195402.
- [202] S.A. Darmanyan, M. Nevière, A.V. Zayats, *Phys. Rev. B* 70 (2004) 075103.
- [203] A.R. McGurn, A.A. Maradudin, *Opt. Commun.* 72 (1989) 279.
- [204] V. Freilikher, M. Pustilnik, I. Yurkevich, A.A. Maradudin, *Opt. Commun.* 110 (1994) 263.
- [205] V. Malyskin, A.R. McGurn, T.A. Leskova, A.A. Maradudin, M. Nieto-Vesperinas, *Opt. Lett.* 22 (1997) 946.
- [206] Recent surveys of theoretical and experimental studies of rough surface scattering are: A. A. Maradudin and E. R. Méndez, in: E.R. Pike, P. Sabatier (Eds.), *Scattering*, Academic Press, New York, 2002, p. 864; A.V. Shchegrov, A.A. Maradudin, E.R. Méndez, in: E. Wolf (Ed.), *Progress in Optics*, Elsevier Science Publishers, Amsterdam, 2004, p. 117.
- [207] A.R. McGurn, A.A. Maradudin, *J. Opt. Soc. Am. B* 4 (1987) 910.
- [208] A.R. McGurn, A.A. Maradudin, *Waves in Random Media* 6 (1996) 251.
- [209] G.C. Brown, V. Celli, M. Coopersmith, M. Haller, *Surf. Sci.* 129 (1983) 507.
- [210] P.W. Johnson, R.W. Christy, *Phys. Rev. B* 6 (1972) 4370.
- [211] This value of $\epsilon(\omega)$ was obtained by West and O'Donnell by interpolation from the data in E. Palik, *Handbook of Optical Constants of Solids*, Academic Press, New York, 1985, p. 294, see Ref. 10 of Ref. [114].
- [212] A.A. Maradudin, A.R. McGurn, E.R. Méndez, *J. Opt. Soc. Am. A* 12 (1995) 2500.
- [213] E. Burstein, A. Hartstein, J. Schoenwald, A.A. Maradudin, D.L. Mills, R.F. Wallis, in: E. Burstein, F. de Martini (Eds.), *Polaritons*, Pergamon, London, 1974, p. 89.
- [214] A.A. Maradudin, E.R. Méndez, T. Michel, *Opt. Lett.* 14 (1989) 151.
- [215] A.A. Maradudin, E.R. Méndez, T. Michel, in: J.C. Dainty, M. Nieto-Vesperinas (Eds.), *Scattering from Volumes and Surfaces*, North-Holland, Amsterdam, 1990, p. 157.
- [216] H.F. Ghaemi, T. Thio, D.E. Grupp, T.W. Ebbesen, J. Lezec, *Phys. Rev. B* 58 (1998) 6779.
- [217] T.J. Kim, T. Thio, T.W. Ebbesen, D.E. Grupp, H.J. Lezec, *Opt. Lett.* 24 (1999) 256.
- [218] H. Bethe, *Phys. Rev.* 66 (1944) 163.
- [219] A.V. Zayats, L. Salomon, F. de Fornel, *J. Microsc.* 210 (2003) 344.
- [220] J.A. Porto, F.J. García-Vidal, J.B. Pendry, *Phys. Rev. Lett.* 83 (1999) 2845.
- [221] U. Schröter, D. Heitmann, *Phys. Rev. B* 58 (1998) 15419.
- [222] E. Popov, M. Nevière, S. Enoch, R. Reinisch, *Phys. Rev. B* 62 (2000) 16100.
- [223] W.-C. Tan, T.W. Preist, R.J. Sambles, *Phys. Rev. B* 62 (2000) 11134.
- [224] A. Krishnan, T. Thio, T.J. Kim, H.J. Lezec, T.W. Ebbesen, P.A. Wolf, J. Pendry, L. Martin-Moreno, F.J. García-Vidal, *Opt. Commun.* 200 (2001) 1.
- [225] A.K. Sarychev, V.A. Podolskiy, A.M. Dykhne, V.M. Shalaev, *IEEE J. Quant. Electr.* 38 (2002) 956.
- [226] D. Gerard, L. Salomon, F. de Fornel, A.V. Zayats, *Phys. Rev. B* 69 (2004) 113405.
- [227] J.A. Sánchez-Gil, A.A. Maradudin, J.Q. Lu, V.D. Freilikher, *Phys. Rev. B* 51 (1995) 17100.
- [228] Z.-H. Gu, R.S. Dummer, A.A. Maradudin, A.R. McGurn, E.R. Méndez, *Appl. Opt.* 30 (1991) 4094.
- [229] D.E. Grupp, H.J. Lezec, T. Thio, T.W. Ebbesen, *Adv. Mater.* 11 (1999) 860.
- [230] D. Gerard, L. Salomon, F. de Fornel, A.V. Zayats, *Opt. Express* 12 (2004) 3652.
- [231] R.E. Collin, *Field Theory of Guided Waves*, second ed., IEEE, New York, 1991.
- [232] A. Dogariu, T. Thio, L.J. Wang, T.W. Ebbesen, H.J. Lezec, *Opt. Lett.* 26 (2001) 450.
- [233] J. Elliott, I.I. Smolyaninov, N.I. Zheludev, A.V. Zayats, *Opt. Lett.* 29 (2004) 1414.

- [234] K. Yang, W. Kim, A. Jain, J. Kumar, S. Tripathy, *Opt. Commun.* 164 (1999) 203.
- [235] V. Kuzmiak, A.A. Maradudin, F. Pincemin, *Phys. Rev. B* 50 (1994) 16835.
- [236] Y.R. Shen, F. de Martini, in: V.M. Agranovich, D.L. Mills (Eds.), *Surface Polaritons*, North-Holland, Amsterdam, 1982, p. 629.
- [237] Y. Wang, H.J. Simon, *Phys. Rev. B* 47 (1993) 13695.
- [238] K.A. O'Donnell, R. Torre, C.S. West, *Phys. Rev. B* 55 (1997) 7985.
- [239] I. Smolyaninov, A. Zayats, C. Davis, *Phys. Rev. B* 56 (1997) 9290.
- [240] S.I. Bozhevolnyi, J. Beermann, V. Coello, *Phys. Rev. Lett.* 90 (2003) 197403.
- [241] A. Zayats, I. Smolyaninov, C. Davis, *Opt. Commun.* 169 (1999) 93.
- [242] A.V. Zayats, I.I. Smolyaninov, C.C. Davis, *Proc. SPIE* 3732 (1999) 81.
- [243] A.V. Zayats, T. Kalkbrenner, V. Sandoghdar, J. Mlynek, *Phys. Rev. B* 61 (2000) 4545.
- [244] A.V. Zayats, I.I. Smolyaninov, *Phil. Trans. Roy. Soc. A* 362 (2004) 843.
- [245] N. Bloembergen, R.K. Chang, S.S. Jha, C.H. Lee, *Phys. Rev.* 174 (1968) 813.
- [246] C.C. Wang, A.N. Duminski, *Phys. Rev. Lett.* 20 (1968) 668 21 (1968) 266(E)..
- [247] C.K. Chen, A.R.B. de Castro, Y.R. Shen, *Phys. Rev. Lett.* 46 (1981) 145.
- [248] G.S. Agarwal, S.S. Jha, *Phys. Rev. B* 26 (1982) 482.
- [249] G.A. Farias, A.A. Maradudin, *Phys. Rev. B* 30 (1984) 3002.
- [250] A.A. Maradudin, *J. Opt. Soc. Am.* 73 (1983) 759.
- [251] R. Reinisch, M. Neviere, *Phys. Rev. B* 28 (1983) 1870.
- [252] See, for example, P. Vincent, in: R. Petit (Ed.), *Electromagnetic Theory of Gratings*, Springer, New York, 1980, p. 101.
- [253] J.A. Sánchez-Gil, A.A. Maradudin, Jun Q. Lu, V.D. Freilikher, M. Pustilnik, I. Yurkevich, *Phys. Rev. B* 50 (1994) 15353.
- [254] X. Wang, H.J. Simon, *Opt. Lett.* 16 (1991) 1475.
- [255] H.J. Simon, Y. Wang, L.B. Zhou, Z. Chan, *Opt. Lett.* 17 (1992) 1268.
- [256] O.A. Aktsipetrov, V.N. Golovkina, O.I. Kapusta, T.A. Leskova, N.N. Novikova, *Phys. Lett. A* 170 (1992) 231.
- [257] L. Kuang, H.J. Simon, *Phys. Lett. A* 197 (1995) 257.
- [258] S.I. Bozhevolnyi, K. Pedersen, *Surf. Sci.* 377–379 (1997) 384.
- [259] E. Kretschmann, *Z. Physik* 241 (1971) 313.
- [260] K.A. O'Donnell, R. Torre, C.S. West, *Opt. Lett.* 21 (1996) 1738.
- [261] K.A. O'Donnell, R. Torre, *Opt. Commun.* 138 (1997) 341.
- [262] M. Leyva-Lucero, E.R. Méndez, T.A. Leskova, A.A. Maradudin, Jun Q. Lu, *Opt. Lett.* 21 (1996) 1809.
- [263] M. Leyva-Lucero, E.R. Méndez, T.A. Leskova, A.A. Maradudin, *Opt. Commun.* 161 (1999) 79.
- [264] T.A. Leskova, M. Leyva-Lucero, E.R. Méndez, A.A. Maradudin, I.V. Novikov, *Opt. Commun.* 183 (2000) 529.
- [265] I.V. Novikov, A.A. Maradudin, T.A. Leskova, E.R. Méndez, M. Leyva-Lucero, *Waves in Random Media* 11 (2001) 1.
- [266] S. Enoch, *Opt. Commun.* 148 (1998) 137.
- [267] J. Rudnick, E.A. Stern, *Phys. Rev. B* 4 (1971) 4274.
- [268] J.E. Sipe, V.C.Y. So, M. Fukui, G.I. Stegeman, *Phys. Rev. B* 21 (1980) 4389.
- [269] J.E. Sipe, G.I. Stegeman, in: V.M. Agranovich, D.L. Mills (Eds.), *Surface Polaritons*, North-Holland, Amsterdam, 1982, p. 661.
- [270] V.M. Agranovich, S.A. Darmanyany, *JETP Lett.* 35 (1982) 80.
- [271] D. Maystre, M. Neviere, *Appl. Phys. A* 39 (1986) 115.
- [272] M. Corvi, W.L. Schaich, *Phys. Rev. B* 33 (1986) 3688.
- [273] J.A. Maytorena, W. Luis Mochán, B.S. Mendoza, *Phys. Rev. B* 51 (1995) 2556.
- [274] B.S. Mendoza, W.L. Mochán, *Phys. Rev. B* 53 (1996) 4999.
- [275] A.V. Shchegrov, A.A. Maradudin, E.R. Méndez, in: E. Wolf (Ed.), *Progress in Optics*, vol. 46, Elsevier, Amsterdam, 2004, p. 117.
- [276] See, for example, Fig. 3 of F. Brown, R.E. Parks, A.M. Sleeper, *Phys. Rev. Lett.* 14 (1965) 1029.
- [277] Lord Rayleigh, *The Theory of Sound*, vol. II, MacMillan, London, 1885, pp. 89, 297.
- [278] A. E. Danese, *Advanced Calculus*, vol. I, Allyn and Bacon, Boston, 1965, p. 123.
- [279] A.A. Maradudin, T. Michel, A.R. McGurn, E.R. Méndez, *Ann. Phys. (N.Y.)* 203 (1990) 255.
- [280] V. Freilikher, E. Kanziiper, A.A. Maradudin, *Phys. Rep.* 288 (1997) 127 Appendix A.
- [281] J.F. McGilp, *Progr. Surf. Sci.* 49 (1995) 1.
- [282] G.T. Boyd, Th. Rasing, J.R.R. Leite, Y.R. Shen, *Phys. Rev. B* 30 (1984) 519.

- [283] T.F. Heinz, M.M.T. Loy, W.A. Thompson, *Phys. Rev. Lett.* 54 (1985) 63.
- [284] L.D. Landau, E.M. Lifshits, *Electrodynamics of Continuous Media*, Pergamon, London, 1984.
- [285] Y. Kawata, C. Xu, W. Denk, *J. Appl. Phys.* 85 (1999) 1294.
- [286] H.J. Simon, Z. Chen, *Phys. Rev. B* 39 (1989) 3077.
- [287] A.C.R. Pipino, G.C. Schatz, R.P. van Duyne, *Phys. Rev. B* 49 (1994) 8320.
- [288] A.C.R. Pipino, R.P. van Duyne, G.C. Schatz, *Phys. Rev. B* 53 (1996) 4162.
- [289] N.E. Glass, A.A. Maradudin, V. Celli, *Phys. Rev. B* 27 (1983) 5150.
- [290] H. Sano, M. Kume, H. Nakagawa, G. Mizutani, *J. Appl. Phys.* 91 (2002) 4229.
- [291] S. Grafstrom, *J. Appl. Phys.* 91 (2002) 1717 *Europhys. Lett.* 15 (1991) 289.
- [292] U.Ch. Fischer, D.W. Pohl, *Phys. Rev. Lett.* 62 (1989) 458.
- [293] A.V. Zayats, V. Sandoghdar, *J. Microsc.* 202 (2001) 94.
- [294] J. Lehmann, P. Reiman, P. Hänggi, *Phys. Rev. Lett.* 84 (2000) 1639.
- [295] L.D. Landau, E.M. Lifshitz, *Mechanics*, Pergamon, Oxford, 1960.
- [296] H. Raether, *Excitation of Plasmons and Interband Transitions by Electrons*, Springer, Berlin, 1980.
- [297] P. Johansson, R. Monreal, P. Apell, *Phys. Rev. B* 42 (1990) 9210.
- [298] J.H. Coombs, J.K. Gimzewski, B. Reihl, J.K. Sass, R.R. Schlittler, *J. Microsc.* 152 (1988) 325.
- [299] I.I. Smolyaninov, M.S. Khaikin, V.S. Edelman, *Phys. Lett. A* 149 (1990) 410.
- [300] I.I. Smolyaninov, V.S. Edelman, V.V. Zavyalov, *Phys. Lett. A* 158 (1991) 337.
- [301] K. Takeuchi, Y. Uehara, S. Ushioda, S. Morita, *J. Vac. Sci. Technol. B* 9 (1991) 557.
- [302] V. Sivel, R. Coratger, F. Ajustron, J. Beauvillain, *Phys. Rev. B* 45 (1992) 8634.
- [303] N. Venkateswaran, K. Sattler, J. Xhie, M. Ge, *Surf. Sci.* 274 (1992) 199.
- [304] M.J. Gallagher, S. Howells, L. Yi, D. Sarid, *Surf. Sci.* 278 (1992) 270.
- [305] R. Berndt, J.K. Gimzewski, *Phys. Rev. B* 48 (1993) 4746.
- [306] I.I. Smolyaninov, E.V. Moskovets, *Phys. Lett. A* 165 (1992) 252.
- [307] I.I. Smolyaninov, *Surf. Sci.* 364 (1996) 79.
- [308] I.I. Smolyaninov, *Phys. Lett. A* 181 (1993) 175.
- [309] F. Forstmann, H. Stenschke, *Phys. Rev. B* 17 (1978) 1489.
- [310] B.N.J. Persson, A. Baratoff, *Phys. Rev. Lett.* 68 (1992) 3224.
- [311] R. Berndt, A. Baratoff, J.K. Gimzewski, in: R.J. Behm, N. Garcia, H. Rohrer (Eds.), *Scanning Tunneling Microscopy and Related Methods*, Kluwer, Dordrecht, 1990, pp. 269–280.
- [312] R. Berndt, R. Gaisch, J.K. Gimzewski, B. Reihl, R.R. Schlittler, W.D. Schneider, M. Tschudy, *Science* 262 (1993) 1425.
- [313] R. Berndt, R. Gaisch, W.D. Schneider, J.K. Gimzewski, B. Reihl, R.R. Schlittler, M. Tschudy, *Phys. Rev. Lett.* 74 (1995) 102.