Plasmon optics of structured silver films

A. Bouhelier,* Th. Huser,† H. Tamaru,‡ H.-J. Güntherodt, and D. W. Pohl
Institute of Physics, University of Basel, Klingelbergstrasse 82, CH-4056 Basel, Switzerland

Fadi I. Baida and D. Van Labeke
Laboratoire d'Optique PM Duffieux, CNRS UMR 6603, Institut des Microtechniques de Franche-Comté, FR0067 CNRS, Université de Franche-Comté, F-25030 Besançon Cedex, France

(Received 29 February 2000; revised manuscript received 22 June 2000; published 23 March 2001)

Excitation and propagation of surface plasmons in silver films structured with narrow straight grooves were visualized in unprecedented detail by means of near-field optical techniques. Reflectivity, transmissivity, and scattering loss of the grooves are demonstrated. Their values are determined semiquantitatively. The surface plasmon attenuation are found to be dominated by material and surface/interface imperfections.

DOI: 10.1103/PhysRevB.63.155404 PACS number(s): 73.20.Mf, 42.25.Hz, 42.30.Va, 78.20.Ci

I. INTRODUCTION

The electric properties of metals are far from being ideal at optical frequencies. The concept of the gas of conduction electrons, however, is still valid and instrumental for the understanding of plasmon phenomena, i.e., electron density waves. Such waves may propagate over sizable distances like sound waves in real gases. At metal surfaces and metal/dielectric interfaces, a particular mode, the ''surface plasmon'' (SP) causes a variety of optical effects, among them a strong field enhancement at the interface.[1,2]

The intensity decay length of a plane SP wave in a perfect planar metal film between two perfect dielectric media defines its *intrinsic* decay length $L_{int}=1/2k''$, which is a measure of the ''ideality'' of the electron gas. k'' is defined as the imaginary part of the complex surface plasmon wave vector $k_{SP}=k'+ik''$. The value of k' in silver is about 5% larger than k_0, the wave vector of the light wave with the same frequency ω as the surface plasmon. The size of k'' is mainly influenced by the imaginary part ε'' of the dielectric constant $\varepsilon=\varepsilon'+i\varepsilon''$ of the metal. *Intrinsic* losses are caused by inelastic scattering of conduction electrons, scattering of electrons at interfaces, and leakage radiation (LR) (radiation damping[1]).

LR is emitted from the interface between the metal film and the higher refractive index dielectric medium (the substrate, glass in the present experiment). With respect to the normal of the interface, LR is radiated at a characteristic angle of inclination θ_{SP}.[1] At this angle, the LR in the substrate is phase matched to the SP wave, satisfying $k_{SP}=nk_0\sin\theta_{SP}$ where k_{SP} and nk_0 are the wave vectors of the SP and LR, respectively, n being the refractive index of the substrate. For glass with $n=1.5$, $\theta_{SP}\simeq 44°$, $\theta_{crit}=41.8°$. Hence $\theta_{SP}>\theta_{crit}$, the critical angle of the substrate/air interface.

Inhomogeneities of the metal film and/or substrate cause additional losses in SP propagation. These *extrinsic* losses can be reduced by improving the quality of the film and substrate. Note that the commonly used literature values of $\mathrm{Im}(\varepsilon)$ are based on reflectivity/transmissivity measurements, and hence influenced by extrinsic contributions.

Knowledge of the truly intrinsic losses would allow predictions on the ultimate SP decay length. A large SP decay length opens the possibility of manipulating SP beams, like light beams. However, SP functional elements such as mirrors, filters, diffraction gratings, waveguides, and modulators would have to be developed for this purpose. This might be achieved in the simplest case by structures such as grooves engraved into a SP-supporting metal film.

The parameters governing SP propagation in structured metal films are generally unknown; knowledge so far is restricted to a few theoretical investigations.[3–5] Information hence has to be gathered on quite an elementary level, starting with structures that can be readily implemented and easily understood. It then will be the designer's task to define structures with optimized SP properties and to assess their chances as potential replacements of ''genuinely'' optic elements.

The efficiency of electro-optic (e-o) surface plasmon modulators, for instance, is predicted to exceed that of conventional ones by far.[6–8] Knowledge of $\varepsilon(\omega)$, in particular its imaginary part, is hence of both theoretical and practical interest. The established values[9–11] however, are quite vague about $\varepsilon''(\omega)$. The dielectric function is usually determined from the reflectivity of a metal surface, frequently in combination with transmissivity data of a thin film of that metal. The uncertainty in ε'' is quite large in general (up to 100%).[10] This is a consequence of the relatively large areas of metal surface illuminated in those experiments. The resulting data are thus averages that include the effects of all bulk and interface imperfections of the respective areas.

Methods that allow study of SP propagation in such unperturbed areas may yield more precise data on the intrinsic optical properties of metal surfaces. Scanning near-field optical microscopy (SNOM) provides such an opportunity: The SNOM probe acts as a SP point source that can be positioned precisely over a selected spot of a (metal) sample surface. The propagation of the SP can be imaged with high lateral resolution by means of a conventional optical microscope (COM) equipped with a large numerical aperture (NA) immersion objective; variation of the total SP excitation with position is obtained from ''forbidden light'' SNOM images.[12]

The early studies with this method[14] showed that each

inhomogeneity acts as a scattering center. The scattering strength is correlated with the size of the center but a quantitative description has not been possible so far. The reason is the random and ill-defined nature of the scattering centers. Qualitative observation of scattering centers was reported by a number of authors using different methods.[13–17]

In the present paper, we report experimental observations of SP propagation in thin silver films structured by well-defined narrow grooves. Grooves were chosen since they are among the most elementary structures that can be implemented in a metal film. More complex artificial structures comprising, e.g., variations of thickness and composition as well as additional overlayers can be investigated in the same way. The narrow grooves were arranged parallel and orthogonal to each other as described briefly in Ref. 18. The present paper gives a complete account of these investigations. Specifically, evidence is given for SP reflection and transmission by the grooves. Reflectivity and transmissivity are estimated. As a by-product we found that the SP decay length is strongly influenced by film inhomogeneities which tend to mask intrinsic damping effects. Our experimental investigation was complemented by a theoretical and computational study which will be published elsewhere.[19]

II. EXPERIMENT

The two experimental setups consist of a standard aperture SNOM head, a sample stage, and two different arrangements for detection. The setups are the same as in Refs. 12 and 14 except for slight modifications. We therefore restrict ourselves to a short description here. The essential parts of the two setups are sketched in Fig. 1.

The SNOM probes were aluminum-coated, pulled optical fiber tips with sub-100-nm diameter apertures at the apex. They were mounted on a standard piezotube/tuning fork combination used for distance and position control. Rough lateral positioning was possible with a mechanical x/y stage. Linearly polarized light from a HeNe or an Ar^{2+} laser was used for excitation.

Silver films, 60–90 nm in thickness, deposited on thin microscope cover slides were chosen as samples. Sets of narrow grooves were cut into certain regions of the films by means of a focused ion beam microscope. On sample 1, all the grooves were arranged parallel to each other, forming two different patterns: (a) single lines at 40 μm separation, consisting of 2 μm long sections whose width increases stepwise from 200 to 2000 nm, and (b) triplets of parallel, 200 nm wide grooves at 2 μm distance, the distance between triplets being 40 μm. On sample 2, 200 nm wide grooves formed squares with 2.2 μm side length. The samples were mounted on a piezoelectric x/y stage. Scan images could be generated either by tip or by sample scanning.

The observation of SP excitation and propagation is based on the detection of LR. At the film/glass interface, the LR intensity at a certain lateral position on the glass side of the metal/glass interface is proportional to that of the SP on the metal side at the same lateral position. Also the *azimuthal* direction equals that of the SP at each point of the interface.

The detection of LR requires optical elements in contact with the bottom part of the sample substrate to avoid total internal reflection. This is achieved in the first detection scheme, Fig. 1(a), by means of an oil immersion objective ($\times 63$, NA$=1.4$) in optical contact with the sample substrate. The objective is part of an inverted classical optical microscope focused on the metal/glass interface. When the SNOM head is in operating position, the microscope image shows the lateral distribution of the LR at the interface with high magnification. The distribution of SP was videotaped with a charge-coupled device (CCD) camera while the sample with the film structures of interest was scanned under the SNOM probe. The CCD is placed in the image plane of the metal/glass interface. Each pixel element hence captures the flux of LR emerging from the corresponding spot in the object plane. The latter is a direct measure of the SP excitation at that spot.

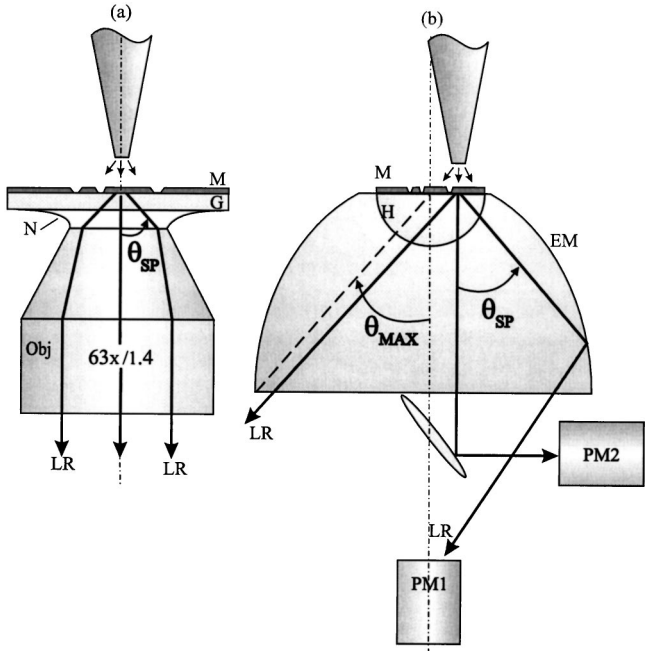

FIG. 1. Experimental schemes: The surface plasmon (SP) is excited by SNOM tip in a structured silver film M on glass substrate G. Leakage radiation (LR) emitted into G at angle θ_{SP} is captured in both arrangements. (a) Direct imaging of LR intensity distribution at the M/G interface by a conventional optical microscope: Immersion contact N enables microscope objective Obj to capture the LR although $\theta_{SP} > \theta_{crit}$. (b) Imaging of SP excitation by a ''misaligned'' forbidden light SNOM. LR emitted toward the right into glass hemisphere H is collected by an ellipsoidal mirror (EM) and focused onto photomultiplier PM1. Slight displacement of the tip from the optical axis prevents LR emitted toward the left from reaching PM1 ($\theta_{SP} < \theta_{max}$). Directly transmitted light (''allowed light'') is recorded by photomultiplier PM2.

The direct SP imaging capability of the SNOM/COM combination is unique and specific for this scheme: We do not know of any other experiment providing local SP information in similar detail. It may be worth mentioning that similar imaging should also be possible with classical SP excitation.[20,21]

In the second detection scheme Fig. 1(b), optical contact with a hemispherical lens avoids total internal reflection at

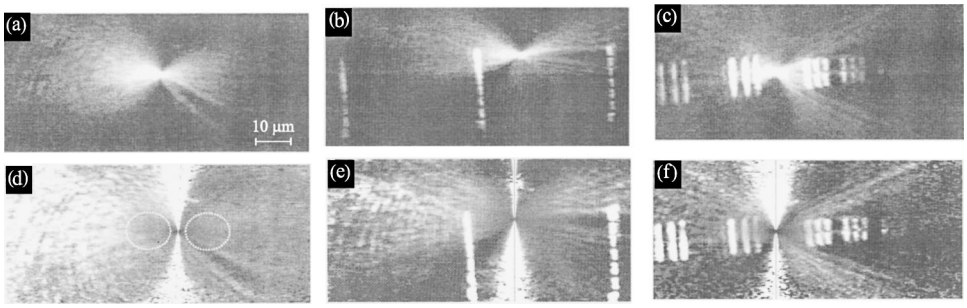

FIG. 2. (a)–(c) Microscope images of sample 1, illuminated by SNOM probe in an unperturbed region, to the right of a nearby groove, and near the triplets of grooves, respectively (NA=1.4, ×63, $\lambda_0=632$ nm, polarization normal to grooves); (d)–(f) same images multiplied by $\rho/\cos^2\psi$ to compensate the intensity loss due to spreading and the dipole characteristics. Note the plateaus of constant gray scale values at intermediate intensities in (d)–(f) and the interference fringes in (c) and (f).

the exit plane of the sample substrate. The hemisphere is part of a ''forbidden'' (and ''allowed'') light SNOM, sketched and discussed in Refs. 22 and 23. Radiation from the sample is selected by wave vector rather than by location of the origin. ''Forbidden'' light emitted at $\theta > \theta_{crit}$ with respect to the normal on the film surface is focused on photomultiplier PM1 by an ellipsoidal mirror segment. ''Allowed'' light emitted or scattered at $\theta < \theta_{crit} - \Delta\theta$ is collected by a centered objective and directed onto a second photomultiplier PM2 by an appropriate, tilted mirror. The residual ''allowed'' range $\theta_{crit} - \Delta\theta < \theta < \theta_{crit}$ remains undetected for instrumental reasons ($\Delta\theta \approx 3°$).

The present forbidden light SNOM differs from the previous one[12,14] in the possibility of selecting certain azimuthal directions of forbidden light for detection. This is achieved by controlled displacement of the SNOM probe and sample with respect to the axis of the mirror. The displacement and its effect are sketched in Fig. 1(b).

A 2 mm displacement is sufficient to suppress the LR emitted toward the left hand side since the cutoff angle θ_{max} of the mirror is only slightly smaller than θ_{SP}. Hence only LR directed toward the right hand side (rhs), corresponding to azimuth $\psi \simeq -45°$ to $+45°$, was recorded by PM1.

The PM signals are proportional to the *total flux into the respective acceptance angles* of the two PM's. The forbidden light signal is dominated by the flux of LR emitted from the whole area of the sample into the selected azimuthal directions. Hence it is a measure of SP energy stored in the radiation lobes pointing in the selected direction. In addition, light scattered from sample inhomogeneities into forbidden directions is recorded but this is much weaker than that of the LR in general. The allowed light signal represents the directly transmitted light together with contributions from the SP scattered into the allowed directions by the grooves and film inhomogeneities.

III. RESULTS AND DISCUSSION: LOCAL IMAGING BY COM

Figure 2 depicts COM images of the relevant parts of sample 1. The surface-plasmon-excited areas shine brightly due to the captured LR. The intensity of the large lobes was found to increase exponentially during final approach of the SNOM probe to the silver film. The (z) distance dependence equals that reported before,[14] confirming the excitation of SP's. The center of excitation at the position of the SNOM probe is overexposed in general due to the limited dynamic range of the CCD camera. The SP excitation intensity was adjusted to a level beyond the maximum CCD response near the source point in order to image the decay of the SP over a decently wide area. The central range was intentionally excluded from comparison with theory since the variation of SP intensity with position is expected to be more complex there than farther away from the source. This is a consequence of the finite size of the SNOM probe.

A. Unstructured silver film

The following images are representative of the large number of SP patterns videotaped with the COM images of the first detection scheme. For supplementary information, a number of short video sequences may be found on the Basel authors' home page.[24] They show the SP pattern and its variation while the SNOM probe is scanned over the sample surface. Some of the features of relevance are more obvious in the motion pictures of the video than in still images.

Figure 2(a) depicts the SP pattern in a film area without artificial structures. Figure 3(a) shows the pattern in a different orientation, achieved by rotating the polarization of the exciting light. The two-lobe radiation characteristic reported before[14] is seen in these figures, too, indicating that we are indeed looking at SP excitation. The lobe pattern has some

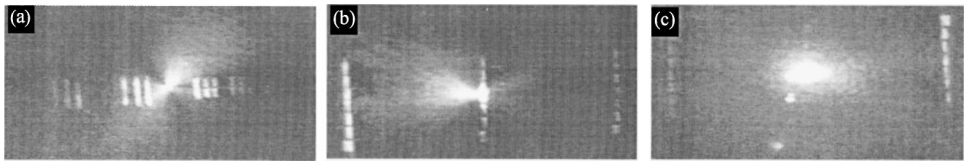

FIG. 3. (a)–(c) Microscope images, same settings and sample as in Fig. 2. (a) Rotation of polarization results in same polarization of the SP lobes; (b) SNOM probe positioned to the left of a nearby groove and (c) right on top of that groove.

irregularities. The intensities in the two lobes differ by about 30% and a dark stripe disturbs the lower part of the rhs lobe. The irregularities are caused by imperfections of the SNOM tip. Unfortunately, it is difficult to obtain more perfect SNOM probes with present day manufacturing techniques.

The abrupt localized changes of intensity observed in Fig. 2(a) are probably caused by SP scattering at film inhomogeneities such as grain boundaries or hillocks standing on the film surface. The stepwise attenuation at scattering centers is particularly obvious when the SNOM probe is moved along the silver film. The video record shows that the positions of the spots remain fixed but their ''shadows'' change direction in accordance with the position and relative orientation of the source.[24]

To gain a better understanding of the emission pattern in Fig. 2(a), we tried to find a mathematical description of the SP intensity distribution $I(\rho,\psi)$. The coordinates ρ and ψ are the distance from the source and the azimuth with respect to the central axis of the rhs lobe, respectively. The simplest angular representation of a two-lobe pattern is $\cos^2\psi$. For radial decay of intensity, we expect a $1/\rho$ dependence due to spreading in two dimensions. In addition, damping, resulting from intrinsic losses and, potentially, extrinsic ones is to be accounted for by an exponential decay term (decay constant α).

Intrinsic losses are caused by inelastic electron scattering and generation of LR. Extrinsic losses are caused by inhomogeneities (SP scattering at film inhomogeneities smaller than the resolution limit of the COM). When averaging over larger areas as in the second detection scheme (cf. Sec. IV), the discrete steps in intensity seen in Fig. 2 may also be approximated by an exponential decay and added to the extrinsic losses. This contribution will be called ''extrinsic loss'' since its value depends on the perfection of the metal film and its substrate. With the first detection scheme, it is possible to recognize optically resolvable sources (ORS's) of extrinsic losses, i.e., sources separated from each other by distances larger than the resolution limit of the COM.

In view of the above considerations, we compare the experimental intensity distribution with the expected functional dependence

$$I_{r,l}(\rho,\psi)=\frac{F^0_{r,l}}{\rho}e^{-2\alpha\rho}\cos^2\psi, \tag{1}$$

where $F^0_{r,l}$ is a measure of the rhs and lhs lobe intensities, respectively. α is the *average* SP amplitude decay constant, not to be confused with the *intrinsic* damping constant (k'') mentioned before. ρ and ψ are the radial and azimuthal coordinates, respectively. $\psi=0$ is defined as the central direction of the rhs SP lobe.

For comparison, it is convenient to multiply the experimental intensity $I_{r,l}(\rho,\psi)$ by $\rho/\cos^2\psi$. This will eliminate the effect of spreading and of azimuthal variation if $I_{r,l}(\rho,\psi)$ really obeys Eq. (1). The result of the multiplication is depicted in Figs. 2(d)–2(f). Three zones can be distinguished in each image: (i) a bright area next to the origin and the *y* axis caused by overexposure and division by $\cos\psi\rightarrow 0$, (ii) a pair of areas with almost uniform gray level [surrounded by a dashed line in Fig. 2(d)], and (iii) a far-out area of irregular intensity essentially representing noise amplified by multiplication by large values of ρ. Comparison of the experimental data with Eq. (1) is restricted to zone (ii). The almost uniform gray level indicates that the result of the multiplication is nearly constant (this means that damping losses are very low) except for the dark shadows, which originate at the inhomogeneities (the origin is obvious in the movie showing the SP pattern while the tip is scanned along the metal film[24]). The plateaus in gray scale are consistent with Eq. (1) if damping was negligibly small. The good agreement with Eq. (1) means that the SP source, i.e., the SNOM probe, may be modeled as a SP point dipole[19,25,26] with respect to the SP far field. For instance, Eq. (1) is obtained by squaring and adding the expressions for the in-plane Green's function of a SP dipole in Ref. 26. Note that only the SP far field distribution is imaged with our technique. Figure 3(a) shows the SP pattern in a different orientation. Rotation is achieved by rotating the polarization of the exciting light. Detailed analysis in the most perfect areas of Figs. 2(a)–2(c) shows that the damping length in fact must be larger than 50 μm, which implies that also $L_{\mathrm{intr}}=1/(2k'')>50$ μm. This is a *lower* limit value. It fixes the *upper* limit of ε'' at 0.3 which is close to the mean values cited in the literature.[9–11]

The result also shows that SP damping values obtained by averaging over large areas will be due to extrinsic losses at ORS's in general. SNOM excitation combined with COM detection allows us to distinguish between ORS-type extrinsic and other contributions to SP losses. The method can be improved and extended in various ways by more detailed and better tailored measurements.

B. Single grooves

Next, a set of single grooves was positioned in the scan range of the SNOM tip. The lobes were adjusted normal to the grooves in order to facilitate the interpretation of experimental results (see below). Figures 2(b), 2(e), and 3(b) depict situations where one of the lobes is near one of the grooves. It is seen in Figs. 2(b) and 2(e) that the lobe is only weakly attenuated by this obstacle, indicating high SP transmissivity. The lower part, after passing the groove, can be compared directly with the upper part, which does not interact with the groove. The width of the lobe is large enough not to be dominated by diffraction at the groove edge. The slits themselves shine brightly even at considerable distance from the SNOM probe where the SP lobes are not visible any more. Conversion of SP radiation into scattered light by the groove is hence much more efficient than conversion into LR in the unperturbed areas. Similar behavior was found recently in the investigation of metal films with holes instead of slits. The finding was attributed to SP enhanced hole transmissivity and SP launching inside a hole[27–29] but essentially refers to the same phenomenon. In Figs. 2(b) and 2(e), the probe is located to the right of the groove and the asymmetry is more or less compensated by its finite transmissivity. In Fig. 3(b) the probe is moved to the left of the groove. Asymmetry and finite transmission now add up, resulting in drastically different lobe intensities. Comparison of the two images allows

us to estimate the transmissivity to ≈90%. Figure 3(c) shows the SNOM probe right on top of the groove. The SP intensity is greatly reduced as expected while the intensity of the directly transmitted radiation is increased.

Reflection is not immediately obvious in still images of the type of Figs. 2 and 3. The video sequence generated while the grooves were scanned under the tip, however, allows recognition of interference fringes and their motion with tip position.[24] This clearly proves the existence of SP reflection by the groove and coherent superposition of direct and reflected radiation. The shape of the fringes allows the assumption that they belong to a family of confocal hyperbolas, which would be typical for interference between a point source and its mirror image. The reflectivity estimated from the visibility, i.e., the depth of modulation of the fringes, is 3–6 %, a large uncertainty resulting from the random intensity fluctuations at the film imperfections and from overexposure.

C. Arrays of grooves

In Figs. 2(c) and 2(f), the SP is launched between the triplets of grooves. Attenuated transmission and reflection are now quite obvious, the signature of the latter being extended interference fringes outside the range of the grooves. The triplets apparently act as surface plasmon multilayer mirrors. The visibility of the fringes yields a reflectivity of $(9\pm4)\%$. On the basis of our previous results, the reflectivity of such an array might reach up to 30% for optimum groove separation. The latter depends on the phase shift upon reflection, which we do not know yet. The low reflectivity in the present arrangement allows us to ignore effects of multiple reflections between adjacent triplets of grooves, which form a Fabry-Pérot interferometer in principle.

IV. RESULTS AND DISCUSSION: SNOM IMAGING

A. General

Each pixel of a SNOM image represents the total light flux received by the detector. Without SP excitation, all the light emerges from a narrow area next to the SNOM probe, providing the high resolution images SNOM is known for. The allowed light signal, dominated by directly transmitted light, satisfies this condition to a large extent. The present forbidden light signals, however, are dominated by LR, which is emitted from an area roughly 20 μm by 40 μm in size. High lateral resolution cannot be expected under these circumstances except for local variations of SP excitation efficiency. This may occur when the probe is above a film imperfection. Aside from those, spatial variations in intensity have to be interpreted as variations of the LR flux into the selected angular range.

Sample 2 and a fresh SNOM probe with fairly symmetric emission characteristics were used for this part of our investigation (sample 1 was not available because of limited lifetime). The misaligned SP source was oriented normal to one side of the square [arrow in Figs. 4(b) and 4(d); $\psi=0$] and LR emitted to the left ($\psi=\pi$) is suppressed.

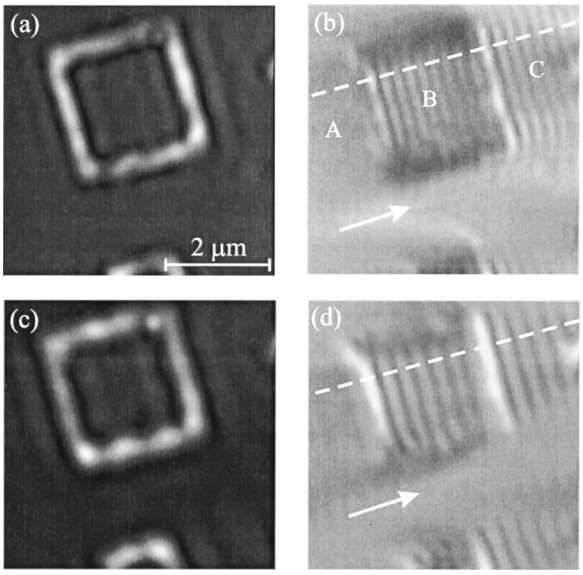

FIG. 4. ''Allowed'' (a),(c) and ''forbidden'' (b),(d) light SNOM images of a square of 0.2 μm wide grooves. (a),(b) $\lambda_0=632$ nm, (c),(d) $\lambda_0=488$ nm. The arrows define the direction $\psi=0$. The light flux profiles measured along the dashed lines are presented in Fig. 5 below.

B. Examples of ''allowed'' and ''forbidden'' light SNOM images

The *allowed light images* of Figs. 4(a) and 4(c) depict the square with high resolution for both 488 and 632 nm excitation. The grooves can be clearly recognized. The remaining area is almost dark. The light intensity inside the grooves and in their vicinity is slightly modulated with period λ_{SP}. Periodicity and position indicate that interference between light directly transmitted through the sample and light generated by SP scattering at one of the grooves may be (and in fact is, cf. Ref. 19) the source of modulation.

In the *forbidden light images*, Figs. 4(b) and 4(d), the squares can be recognized also but the sides are blurred and surrounded by regions of varying and modulated brightness. The period of modulation is approximately $\lambda_{SP}/2$. This is most obvious inside the 2 μm$\times$2 μm size square with eight and six bright fringes for 488 and 632 nm excitation, respectively. The regions of reduced brightness (*A*) at the top left and of undulated brightness inside (*B*) and to the bottom right (*C*) hand side of the square will be discussed here in some detail.

C. Discussion

The brightness variations seen in regions *A*, *B*, and *C* are puzzling at first glance. Why is the SNOM signal, i.e., the integral over the LR in the direction of the rhs lobe, reduced in front of the square over a rather large distance? Why does this same integral undergo undulations in regions *B* and *C*, and why do these undulations decay so rapidly? The decay length, in fact, is about 0.2 μm, in gross disagreement with our previous finding of $1/(2\alpha)\approx10$ μm.

To gain some understanding of these features, the SP intensity distribution near the grooves has to be determined

and integrated within appropriate limits to obtain the detected flux. This is a mathematically demanding problem that requires numerical computation. Baida *et al.*[19] present computer simulations for a sample very similar to sample 2. The results in fact reproduce the experimental images in surprising detail, including even minor features in both the forbidden (LR) and allowed (scattered) light images.

The present discussion will be restricted to a more qualitative interpretation of the forbidden light images. For simplicity, we assume that the SP distribution is imaged perfectly onto the detector by the captured LR and that the grooves are of infinite length, i.e., we replace the square by a pair of parallel grooves at distance $d=2$ μm. The grooves will be considered low reflectivity semitransparent surface plasmon mirrors. The SP lobes are oriented normal to the grooves and assumed to obey Eq. (1). The coordinate system is translated, however, from the tip to the relevant groove and with $F_r^0=F_l^0=F^0$. The decay rate α is considered to be small such that both $\alpha \ll |k_{\mathrm{SP}}|$ and $\alpha \ll 1/x_s$, where x_s is the distance between SNOM probe and groove. It turns out that all the essential features in regions *A*, *B*, and *C* can be reproduced with these assumptions. Groove transmissivity (T), reflectivity (R), and scattering loss (S) can be estimated from fits to the experimental data. The main steps of the flux calculation under these simplified conditions are presented in the following paragraphs; a few intermediate steps are described in the Appendix.

When the probe is positioned in region *A*, to the left of the lhs groove at distance $x_s<0$, the signal at PM1 is generated by the rhs lobe alone at $\theta=\theta_{\mathrm{SP}}$. The lobe extends beyond the grooves into regions *B* and *C* where its intensity is attenuated by T and T^2, respectively. Upon approach to the groove (x_s tends to 0), the flux continuously decreases from Φ^0, the flux emitted by the unperturbed film, to a value close to $T^2\Phi^0$ at $x_s=0$:

$$\Phi^A(x_s)\approx\Phi^0[1-e^{2\alpha x_s}+Te^{2\alpha x_s}(1-e^{-2\alpha d}) + T^2e^{2\alpha(x_s-d)}] \quad (x_s<0). \tag{2}$$

When the probe is positioned in region *B*, $0<x_s<d$, the lhs lobe is partially reflected by the lhs groove. This results in two-point source-type interference between the lobes. The resulting pattern consists of an unmodulated and a modulated part. The intereference fringes have the shape of confocal hyperbolas (cf. Sec. III), the number of fringes equaling the modulo of $2k'x_s$.

The superimposed rhs and reflected lhs lobes extend beyond region *B* into region *C* where their intensities are reduced by *T*. The effect increases for $x_s\rightarrow d$, resulting in a correspondingly reduced flux upon approach of the probe to the rhs groove. Integration of the unmodulated and the modulated parts (cf. Appendix) yields the flux components

$$\Phi^B_{\mathrm{UM}}(x_s)\approx\Phi^0(1+Re^{-2\alpha x_s})[1-(1-T)e^{2\alpha(x_s-d)}] \quad (0<x_s<d), \tag{3a}$$

$$\Phi^B_{\mathrm{MOD}}(x_s)\approx\frac{4\pi}{\pi+1}\Phi^0\sqrt{R}J_0(2k'x_s)e^{-2\alpha x_s} \times[1-(1-T)e^{2\alpha(x_s-d)}]. \tag{3b}$$

J_0 is the zeroth order Bessel J function. The argument $2k'x_s$ implies that Φ^B_{MOD} undergoes a rapidly decaying oscillation when the probe is moved away from the lhs groove. The physical origin of the modulation is the different weight of the interference fringes in the process of integration. Specifically, it makes a difference whether the center of the interference pattern is dark or bright. Hence the rapid decay of undulation seen in Figs. 4(b) and 4(d) is not a sign of SP damping but the characteristic of $J_0(2k'x_s)$. Mathematically, Eq. (3a) is a good approximation to the true flux integral for $(k'x_s)^2\gg 1$; for small $k'x_s$, the flux is overestimated.

In region *C*, the lhs lobe is retroreflected at both grooves. The rapid decay of the J_0 function, however, allows one to ignore the modulated part caused by reflection from the lhs groove. The unmodulated part is increased by the unmodulated flux from the second reflected beam with the approximate result

$$\Phi^C_{\mathrm{UM}}(x_s)\approx\Phi^0[1+Re^{-2\alpha(x_s-d)}(1+T^2e^{-4\alpha d})] \quad (x_s>d), \tag{4a}$$

$$\Phi^C_{\mathrm{MOD}}(x_s)\approx\frac{4\pi}{\pi+1}\Phi^0\sqrt{R}J_0[2k'(x_s-d)]e^{-2\alpha(x_s-d)}. \tag{4b}$$

The modulated parts in regions *B* and *C* are particularly convenient for data fitting since the amplitude ratio is insensitive to offset variations. For low reflectivity and small *d*, the ratio of the modulation amplitudes for different probe positions directly provides the SP transmissivity of the groove, independent of the absolute value of flux.

The qualitative agreement of the above predictions with the forbidden light SNOM image Figs. 4(b) and 4(d) is obvious although serious additional modulation effects can be seen. This reflects the small size of the squares in our sample. The complicated modulation pattern in fact is well reproduced in the numerical simulation.[19] But even the present simplified approach allows for quantitative fits to the experimental data along appropriately chosen paths. Two examples, one for red (632 nm) and the other for blue-green (488 nm) light excitation, are shown in Fig. 5. Good matches were obtained with the same values of the four fit parameters T, R, α, and Φ^0 in all three regions (A,B,C). The numerical values for 632 (488) nm excitation of the relevant first three parameters were $T=0.9$ (0.9), $R=0.05$ (0.06), and $2\alpha=0.09$ (0.11) $\mu\mathrm{m}^{-1}$. The scattering loss and intensity decay length are thus $S=1-T-R=0.05$ (0.04) and $L_\alpha=1/(2\alpha)=11$ (9) μm, respectively. It should be recalled that L_α is not the intrinsic damping length but the value including all the losses due to the imperfections in the integration area of the detector.

Large error bars have to be attached to the numerical fit parameters in view of the assumptions made and the strong variations of signal level in Figs. 4(b) and 4(d). The slight differences in the values for 632 nm and 488 nm hence are

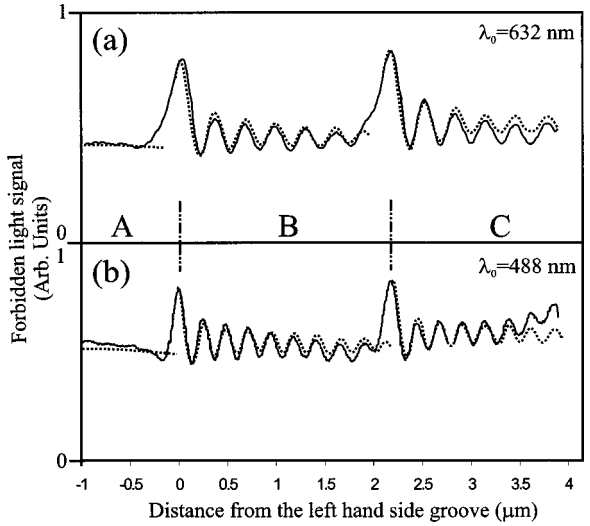

FIG. 5. Experimental (solid lines) and fitted (dashed lines) ''forbidden light'' flux profiles along the dashed lines in Fig. 4. (a) $\lambda_0 = 632$ nm, (b) $\lambda_0 = 488$ nm.

irrelevant. The order of magnitude of T, R, S, and L_α, however, should be correct since various further fit procedures that we made provided similar results. The value for the reflectivity is also consistent with previous theoretical considerations[5] and with Baida *et al.*'s numerical simulation.[19]

V. CONCLUSIONS AND SUMMARY

The SNOM probe provides a point source for SP studies. It can be placed next to any selected structure on a metal surface. This allows investigations, in particular, of SP propagation on a local scale. The present investigations were restricted to thin metal films but bulk metal surfaces can be excited in the same way. Both COM and SNOM imaging schemes can be used for the observation of SP propagation in thin metal films, based on the detection of LR. COM with immersion contact between objective and sample visualizes the propagation of the SP to a level of detail that is limited only by the resolving power of the microscope. Extremely small features on top or inside the metal film can be recognized.

The results of such local SP imaging indicate that the main source of SP decay is imperfections, not intrinsic damping. This implies that there is room for improvement of metal mirrors and waveguides by elimination of defects. Metallic waveguides will provide increased propagation lengths, which is a precondition for the application of SP's in integrated optics.

Each pixel value in the forbidden light SNOM image of a metal film is the integral over the radiation from the sample area excited by the surface plasmon. Reflection at a groove results in characteristic undulations of the flux with position of the probe. Three decay lengths play a role in the present experiment: (i) intrinsic, including losses due to LR ($L_{\mathrm{intr}} > 50\,\mu$m), (ii) average, including losses due to imperfections ($L_\alpha \sim 10\ \mu$m), and (iii), in the SNOM imaging mode, the short decay length of the amplitude oscillations of the Bessel J function, which roughly equals λ_{SP}.

Reflectivity, transmissivity, and scattering loss of narrow grooves in a metal film can be estimated from suitable COM and SNOM images. The values found in the present investigation represent the first experimental estimate of these surface plasmon optical parameters to our knowledge. It was also demonstrated that bundles of several such grooves provide higher reflectivity, apparently acting as multilayer mirrors. The information obtained with COM and forbidden light SNOM imaging is complementary since the first provides lateral, the second angular resolution. Correspondingly, the first is better suited for the investigation of microscopic details, the second for properties averaged over the extension of the surface plasmon lobes.

ACKNOWLEDGMENTS

The authors profited from the help of J. Brugger, H. Heinzelmann, J. M. Freyland, and T. Wrase and from numerous discussions with R. Eggers, J. Toquant, and K. Miyano. The project was supported in part by TMR Project No. FMRX-CT98-0242 (BBW No. 97-0558), and by Japan Science and Technology Corporation (JST) through the CREST program.

APPENDIX

Single groove

When the probe is at the left of the groove at distance $x_s < 0$, Eq. (1) has to be integrated separately over the areas to the left and to the right. The position of the groove in polar coordinates ρ and ψ is $|x_s|/\cos\psi$. This is the upper (lower) limit of integration over ρ which is followed by integration over ψ between $-\pi/2$ and $+\pi/2$ (estimated from the size of PM1). The result is Eq. (2).

When the probe is positioned to the right of the groove ($x_s > 0$), the flux Φ^{right} results from the superposition of the rhs surface plasmon lobe with the reflected lhs lobe. The interference between the two is elegantly described in terms of prolate elliptic coordinates $-1 < \xi < +1$ and $1 < \eta < \infty$. The SP point source and mirror image reside at the foci, $\xi = \pm 1$, and the groove defines the symmetry axis $\xi = 0$. The length scale is normalized to x_s. The confocal hyperbolas $\xi = \text{const}$ are lines of constant phase difference. The confocal ellipses $\eta = \text{const}$ asymptotically approach the normalized circles $\rho/x_s = \text{const}$. The dimensionless wave vector is $\kappa = \kappa' + i\kappa'' = x_s(k' + ik'')$.

The intensity of the SP radiation emitted by the source (1) and its mirror image (-1) alone obey the dipole characteristic of Eq. (1) which here becomes [subscript SP omitted; parameters K, L, M defined in Eqs. (A5a)–(A5c) below]

$$I_{\pm 1}(\xi,\eta) = \begin{Bmatrix} 1 \\ R \end{Bmatrix} \times F^0 e^{-\kappa(\eta \mp \xi)} K_{\pm 1}(\xi,\eta). \qquad \text{(A1)}$$

The interference term has the form

$$I_{(1-1)}(\xi,\eta) = 2\sqrt{I_1 I_{-1}}\cos\kappa'\xi. \qquad \text{(A2)}$$

$I_1+I_{(-1)}$ and $I_{(1-1)}$ are integrated from $\xi=1/\sqrt{2}$ to 1 and $\eta=1$ to ∞ with appropriate consideration of the metric coefficients. The integrals Φ_{UM}, Φ_{MOD} representing the unmodulated and the modulated parts of Φ^{right}, respectively, may be expressed by the sums

$$\Phi_{\mathrm{UM}}=[\mathrm{UM}_{1,1}+\mathrm{UM}_{1,-1}+R(\mathrm{UM}_{-1,1}+\mathrm{UM}_{-1,-1})]\times 2/(\pi+1), \tag{A3a}$$

$$\Phi_{\mathrm{MOD}}=\sqrt{R}(\mathrm{MOD}_1+\mathrm{MOD}_{-1})\times 4/(\pi+1) \tag{A3b}$$

with

$$\mathrm{UM}_{\mu\nu}=2\int_1^{\infty}\int_{1/\sqrt{2}}^{1} e^{-2\kappa''(\eta-\nu\xi)}K_{\mu}(\xi,\eta)\,M_{\nu}(\xi,\eta)\,d\xi d\eta, \tag{A4a}$$

$$\mathrm{MOD}_{\nu}=2\int_1^{\infty}\int_{1/\sqrt{2}}^{1} e^{-2\kappa''(\eta-\nu\xi)}\cos(2\kappa'\xi)\,L(\xi,\eta)\times M_{\nu}(\xi,\eta)\,d\xi d\eta, \tag{A4b}$$

and $\mu=\pm 1$, $\nu=\pm 1$; and further

$$K_{\mu}(\xi,\eta)=\frac{(\xi\eta-\mu)^2}{(\eta-\mu\xi)^3}\rightarrow \xi^2/\eta, \tag{A5a}$$

$$L(\xi,\eta)=\frac{\xi^2\eta^2-1}{(\eta^2-\xi^2)^{3/2}}\rightarrow \xi^2/\eta, \tag{A5b}$$

$$M_{\nu}(\xi,\eta)=\left[\frac{\eta^2-1}{1-\xi^2}\right]^{\nu/2}\rightarrow[\eta/\sqrt{1-\xi^2}]^{\nu}. \tag{A5c}$$

Most of the LR contributing to the detected flux emerges far away from the SP source, i.e., $\eta\gg 1$, since the experimental $\kappa''=\alpha x_s\ll 1$. We therefore replace the expressions for K, L, and M by the terms to the right of the arrows in Eqs. (A5a)–(A5c). $\eta\gg 1$ also implies $M_1\gg M_{(-1)}$; hence the latter will be ignored, too. The integrals are readily factorized with these assumptions. The phase-sensitive factor $\cos 2\kappa'$ in Eq. (A4b) is retained.

Mathematically, $\eta\gg 1$ is equivalent to the asymptotic transition from elliptic to polar coordinates and allows for integration in closed form. The ξ integral of Eq. (A5b) has the analytic solution $\pi/2[J_0(2\kappa')-J_1(2\kappa')/(2\kappa')]$ for the limits $\xi=0,1$. J_0 and J_1 are the zeroth and first order Bessel J functions. A minor error only is introduced if the actual limits of integration $(1/\sqrt{2},1)$ are replaced by (0,1) and if the contribution of J_1 is omitted (fails for very small x_s). The term $2/(\pi+1)$ is a normalization factor. The final result is

$$\Phi^{\mathrm{right}}_{\mathrm{UM}}(x_s)\approx\Phi^0(1+Re^{-2\alpha x_s})\quad (x_s>0), \tag{A6a}$$

$$\Phi^{\mathrm{right}}_{\mathrm{MOD}}(x_s)\approx\frac{4\pi}{\pi+1}\Phi^0\sqrt{R}J_0(2k'x_s)e^{-2\alpha x_s}. \tag{A6b}$$

Pair of grooves

In region *A*, the lobes are divided into three parts, whose weights scale with 1, T, and T^2. In region *B*, both the rhs and the retroreflected lhs lobes are divided into two parts which are separated by the rhs groove at $x=d$. In region *C*, the weights of the two reflected lobes scale with R and RT^2.

*Electronic address: Alexandre.Bouhelier@unibas.ch
†Present address: Department of Chemistry and Material Sciences, Lawrence Livermore National Laboratory, Livermore CA 94550.
‡Permanent address: Department of Applied Physics, The University of Tokyo, Tokyo 113-8656, Japan.

[1] H. Raether, *Surface Plasmons*, Vol. 111 of *Springer Tracts in Modern Physics*, edited by G. Höhler (Springer-Verlag, Berlin, 1988).
[2] H. Raether, *Excitation of Plasmons and Interband Transitions by Electrons*, Vol. 88 of *Springer Tracts in Modern Physics*, edited by G. Höhler (Springer-Verlag, Berlin, 1980).
[3] R.F. Wallis, A.A. Maradudin, and G.I. Stegeman, Appl. Phys. Lett. **42**, 764 (1983).
[4] A.V. Shchegrov, I.V. Novikov, and A.A. Maradudin, Phys. Rev. Lett. **78**, 4269 (1997).
[5] J.A. Sánchez-Gil, Appl. Phys. Lett. **73**, 3509 (1998).
[6] J. Takahara, S. Yamagishi, H. Taki, A. Morimoto, and T. Kobayashi, Opt. Lett. **22**, 475 (1997).
[7] J. Takahara, S. Yamagishi, and T. Kobayashi (unpublished).
[8] T. Yamauchi, H. Yoshimura, J. Takahara, H. Murata, and T. Kobayashi (unpublished).
[9] M.M. Dujardin and M.L. Theye, J. Phys. Chem. Solids **32**, 2033 (1971).
[10] P.B. Johnson and R.W. Christy, Phys. Rev. B **6**, 4370 (1972).
[11] U. Schroeder, Surf. Sci. **102**, 118 (1981).
[12] B. Hecht, H. Bielefeldt, D.W. Pohl, L. Novotny, and H. Heinzelmann, J. Appl. Phys. **84**, 5873 (1998).
[13] P. Dawson, F. de Fornel, and J.-P. Goudonnet, Phys. Rev. Lett. **72**, 2927 (1994).
[14] B. Hecht, H. Bielefeldt, L. Novotny, Y. Inouye, and D.W. Pohl, Phys. Rev. Lett. **77**, 1889 (1996).
[15] S.I. Bozhevolnyi and F.A. Pudonin, Phys. Rev. Lett. **78**, 2823 (1997).
[16] I.I. Smolyaninov, D.L. Mazzoni, J. Mait, and C.C. Davis, Phys. Rev. B **56**, 1601 (1997).
[17] J.R. Krenn, A. Dereux, J.C. Weeber, E. Bourillot, Y. Lacroute, J.P. Goudonnet, G. Schider, W. Gotschy, A. Leitner, F. R. Aussnegg, and C. Girard, Phys. Rev. Lett. **82**, 2590 (1999).
[18] A. Bouhelier, Th. Huser, J.M. Freyland, H.-J. Güntherodt, and D.W. Pohl, J. Microsc. **194**, 571 (1999).
[19] F. Baida, D. Van Labeke, A. Bouhelier, Th. Huser, and D.W. Pohl, J. Opt. Soc. Am. A (to be published).
[20] E. Kretschmann and H. Raether, Z. Naturforsch. A **23**, 2135 (1968).
[21] A. Otto, Z. Angew. Phys. **27**, 207 (1968).
[22] D. Barchiesi and D. Van Labeke, Ultramicroscopy **61**, 17 (1995).

[23] B. Hecht, D.W. Pohl, H. Heinzelmann, and L. Novotny, Ultramicroscopy **61**, 99 (1995).
[24] http://www.monet.physik.unibas.ch/~bouhelie/video. A digital copy of the film can be ordered upon request.
[25] L. Novotny, B. Hecht, and D.W. Pohl, J. Appl. Phys. **81**, 1798 (1997).
[26] O. Martin and N.B. Piller, Phys. Rev. E **58**, 3909 (1998).
[27] T.W. Ebbesen, H.J. Lezec, H.F. Ghaemi, T. Thio, and P.A. Wolff, Nature (London) **391**, 667 (1998).
[28] H.F. Ghaemi, T. Thio, D.E. Grupp, T.W. Ebbesen, and H.J. Lezec, Phys. Rev. B **58**, 6779 (1998).
[29] C. Soennichsen, A.C. Duch, G. Steininger, M. Koch, and G. von Plessen, Appl. Phys. Lett. **76**, 140 (2000).