

Influence of detection conditions on near-field optical imaging

B. Hecht,^{a)} H. Bielefeldt,^{b)} and D. W. Pohl

Zurich Research Laboratory, IBM Research Division, CH-8803 Rüschlikon, Switzerland

L. Novotny

Pacific Northwest National Laboratory, P.O. Box 999, Richland, Washington 99352

H. Heinzelmann

Universität Basel, Institut für Physik, CH-4056 Basel, Switzerland

(Received 26 February 1998; accepted for publication 24 August 1998)

The process of image formation in transmission mode scanning near-field optical microscopy is analyzed both theoretically and experimentally. Changes in the dielectric and topographic properties of the sample influence not only the total transmitted intensity, but also its angular distribution in the far field. This opens up an *additional* source of optical information about the sample. Some of this additional information is retrieved by separate but simultaneous detection of the radiation emitted at angles smaller (allowed light) and larger (forbidden light) than the critical angle of total internal reflection, respectively. Different experimental setups and their respective advantages are discussed. High resolution, constant height mode optical images of test structures are compared with theoretical predictions. Forbidden-light optical images frequently provide enhanced resolution and/or contrast as compared to allowed light images. For small *phase* objects, in contrast to amplitude objects, a contrast reversal between forbidden and allowed light images is observed. © 1998 American Institute of Physics. [S0021-8979(98)02923-5]

I. INTRODUCTION

One of the most common realizations of near-field optical (NFO) microscopy is the transmission scanning near-field optical microscope (SNOM) with an aperture probe.¹⁻⁵ A metal-coated fiber tip with a small aperture at the apex usually serves as a light source illuminating a sample spot of subwavelength dimensions.⁴ For image generation, the optical probe is raster scanned over the sample while maintaining a gap width smaller than the aperture diameter and much smaller than the wavelength of illumination. The transmitted light is detected in the far field for each probe position.

Local variations in the optical properties of a sample manifest themselves as changes in the total transmitted light flux in the first place. This effect has been mainly used for NFO imaging so far. It is, however, not the only manifestation of locally changing optical properties of a sample: The angular distribution of the transmitted intensity, i.e., the far-field emission *pattern* is also influenced by the local dielectric and topographic sample properties. The information encoded in these angular variations of the far-field emission pattern is left unattended in standard SNOM. The transmitted light is simply collected within a certain cone, given by the numerical aperture of the detection optics and detected integrally only.

The most complete SNOM detection scheme would be a two-dimensional record of the whole far-field emission pat-

tern for each image pixel. A very large amount of data would have to be processed and analyzed in this case. However, the division of the far-field emission pattern into the *allowed* and *forbidden* angular regimes already provides significant additional information.⁶⁻¹⁰

Some differences between allowed and forbidden light imaging will be discussed in the first part of this paper. The discussion will be based on a consideration of the information transfer from the near field into the far field and the resulting contrast and resolution. In the second part of the paper, several SNOM setups are described which permit simultaneous detection of forbidden and allowed light. These microscopes are called tunneling near-field optical microscopes (TNOM). Test images obtained with the most frequently used instrument confirm and illustrate the predicted properties of forbidden light imaging. The present paper thus completes the series of our previous publications⁶⁻¹⁰ centered around the concept of "forbidden light NFO."

II. PRINCIPLE OF OPERATION

A. Near-field-far-field information transfer

Light emission from nanosources in close proximity to a plane interface^{11,12} has been studied and simulations of imaging properties in the TNOM geometry¹³ have been performed using the fact that the emission from an optical probe, i.e., a spatially confined light source can be conveniently described in terms of the "angular spectrum representation."^{14,15} The treatment presented here, also based on the angular spectrum representation of a nanosource, can give some direct insight into the differences between allowed and forbidden light imaging.

^{a)}Address for correspondence: Physical Chemistry Department, Swiss Federal Institute of Technology, CH-8092 Zürich, Switzerland; Electronic mail: hecht@phys.chem.ethz.ch

^{b)}Present address: Experimentalphysik VI, Universität Augsburg, D-86135 Augsburg, Germany.

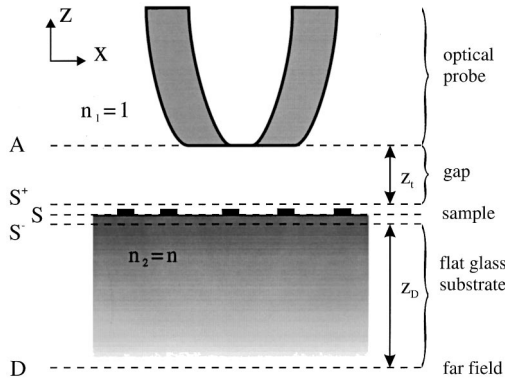


FIG. 1. Geometry of the interaction zone in aperture SNOM: A aperture plane=reference plane; S^+ , S , S^- top, average, and bottom sample planes, respectively; D plane of detection (in the far field); z_i , z_D gap width and distance between sample and detector, respectively; x , z Cartesian coordinates; n_1 , n_2 refractive indices.

The confined field of the optical probe is decomposed into a spectrum of plane propagating and evanescent excitations. The decomposition comprises a two-dimensional Fourier transform of the confined field in an arbitrarily chosen reference plane.^{14,15} Propagation of the probe field normal to the reference plane through stratified layers of different media can then be readily treated in terms of its propagating and evanescent components.

For the present problem, we choose the exit plane A of the small aperture (radius a) at the end of a metal-coated tapered glass fiber as reference plane. It is assumed to be oriented parallel to the nearby substrate surface and normal to the z axis of the Cartesian coordinate system (Fig. 1). The field distribution in plane A in the presence of the plane substrate is roughly known from a number of calculations.^{16,17} Its most characteristic feature is a spatial extension roughly equal to the diameter a of the aperture. Hence, in the Fourier representation in the aperture plane A the amplitudes

$$\mathbf{E}^A(\mathbf{k}) = \int d\mathbf{r} \mathbf{E}^A(\mathbf{r}) e^{i\mathbf{k}\mathbf{r}} \quad (1)$$

are sizeable for $|\mathbf{k}|$ up to values $\approx 2\pi/a$ which typically is 5–20 times larger than $k_0 = 2\pi/\lambda$, the wave vector of the probe light in free space. The field amplitudes in the plane of the substrate surface S^+ at distance z_t from plane A are

$$\mathbf{E}_0^S(\mathbf{k}) = \mathbf{E}^A(\mathbf{k}) \cdot e^{ik_z z_t}. \quad (2)$$

The propagator from A to S^+ , $e^{ik_z z_t}$, is either oscillating or exponentially decaying depending on

$$k_z = (k_0^2 - |\mathbf{k}|^2)^{1/2} \quad (3)$$

being real or imaginary.

We assume the sample to be located as a thin layer in the plane S . In the experiments discussed below, for instance, the sample structure is prepared on a flat glass slide which is attached to the substrate by an index-matching liquid. For a simple mathematical description, the sample is considered to

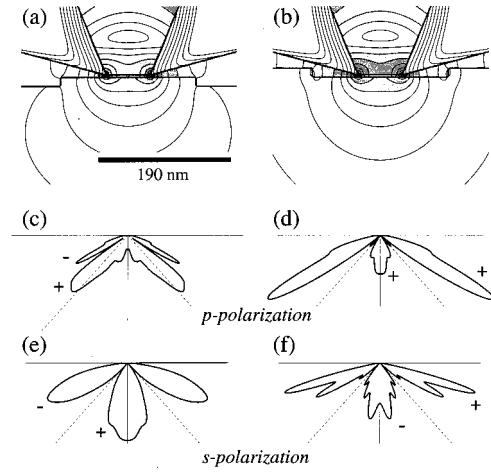


FIG. 2. Computed influence of the sample topography on the near- and far-field emission patterns of a two-dimensional (2D) SNOM: (a),(b) contour plot of electric energy density $|\mathbf{E}|^2$ for p polarization when the optical probe is over a terrace (a) and over a valley (b) of the sample, $\lambda = 488$ nm. (c),(e); (d),(f) polar plots of the difference in the far-field intensities in configurations (a) and (b), respectively, with reference to a flat substrate for p polarization (c),(e) and s polarization (d),(f). The + and - signs at the lobes denote positive and negative differences. The dotted lines indicate the critical angle θ_c .

include the empty space between the various structures within the thin layer between the top and bottom sample planes S^+ and S^- (cf. Fig. 1). Variations in topography hence are reduced to variations of material composition and consequently the dielectric constant. Sufficiently thin samples may be represented by the (Fourier transformed) lateral variation of the dielectric constant, averaged over the thickness of the sample layer, $\epsilon^S(\mathbf{k})$.

Aiming at a semiquantitative discussion only, we restrict our considerations to sample structures which consist of small dielectric and/or very small metallic structures. Such samples are optically “passive,”^{13,18} i.e., the amplitudes of the scattered radiation $\mathbf{E}_1^{S^+}$ and $\mathbf{E}_1^{S^-}$ right above and below S , respectively, are small compared to the respective amplitudes of the unperturbed incident radiation $\mathbf{E}_0^S(\mathbf{k})$. As a consequence, the scattered fields are obtained to first order approximation by convolutions of this unperturbed incident field with the appropriate scattering tensors $\mathbf{S}^\pm(\mathbf{k})$.

The scattering tensors can be readily determined from $\epsilon^S(\mathbf{k})$; for the present purpose it is sufficient, however, to recognize that their spectra extend over roughly the same k vector range as the spectrum of the sample dielectric constant. Its width in k space corresponds to the (inverse) minimum size of the sample structures.

The scattered radiation rides on the large background of the incident radiation. We therefore write

$$\mathbf{E}^{S^\pm}(\mathbf{k}) = \mathbf{E}_0^S(\mathbf{k}) + \mathbf{E}_1^{S^\pm}(\mathbf{k})$$

$$\mathbf{E}_1^{S^\pm}(\mathbf{k}) = \int d\mathbf{k}' \mathbf{S}^\pm(\mathbf{k}') \cdot \mathbf{E}_0^S(\mathbf{k} - \mathbf{k}'). \quad (4)$$

$\mathbf{E}_1^{S^+}(\mathbf{k})$ and $\mathbf{E}_1^{S^-}(\mathbf{k})$ are the relevant fields for the study of image formation in reflection and transmission mode SNOM,

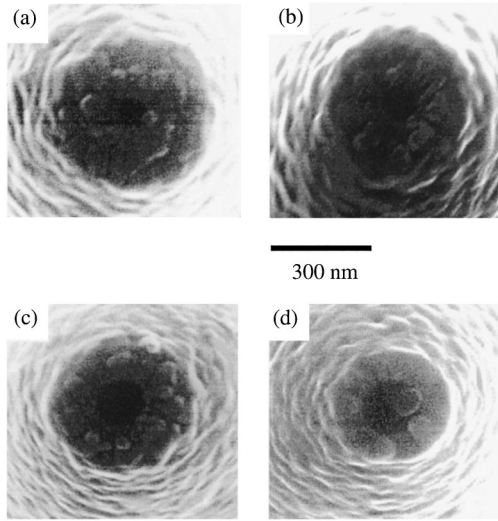


FIG. 3. Scanning electron microscope images of optical probe apertures with apparent sizes between 80 and 150 nm that were obtained by changing the pulling parameters.

respectively. Our present considerations are restricted to the latter mode. Specifically, the detector plane D is assumed to be integrated in the substrate material. This is equivalent to using either a hemispherical substrate or a substrate in optical contact with a high numerical aperture (NA) objective, see Sec. III B, and simplifies the further discussion. The wave amplitudes in the detector plane are obtained by applying the propagator corresponding to the distance z_D between S and D

$$\mathbf{E}^D(\mathbf{k}) = \mathbf{E}_0^D(\mathbf{k}) + \mathbf{E}_1^D(\mathbf{k})$$

$$\mathbf{E}_1^D(\mathbf{k}) = e^{ik_z z_D} \mathbf{E}_1^{S^-}(\mathbf{k}) \quad (5)$$

with $\mathbf{E}_0^D(\mathbf{k}) = \mathbf{E}_0^{S^-}(\mathbf{k}) e^{ik_z z_D}$ being the signal that would be received without a sample from a plane substrate.

B. Forbidden and allowed light gapwidth dependence

Inside the substrate, the range in k space of propagating modes increases in proportion to the refractive index n :

$$k_z = [n k_0^2 - |\mathbf{k}|^2]^{1/2}. \quad (6)$$

If $k_0 z_D \gg 1$ which is usually the case in a SNOM, then only propagating modes contribute to the detected signal, i.e., modes with lateral wave vector $|\mathbf{k}|_{\max} = k_{\max} < n k_0$. The corresponding direction of propagation is

$$\theta = \arcsin(|\mathbf{k}|/n k_0). \quad (7)$$

Waves are directed into the ‘‘allowed’’ angular regime $0^\circ \leq \theta \leq \theta_c = \arcsin 1/n$ for

$$0 \leq |\mathbf{k}| \leq k_0. \quad (8)$$

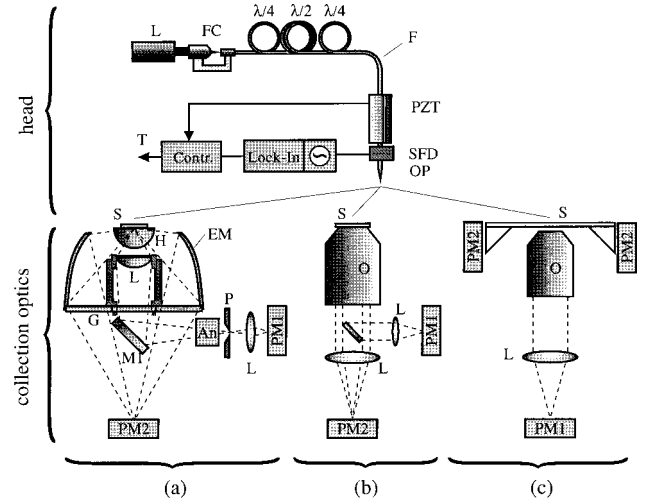


FIG. 4. TNOM configurations: Head: (L) laser; (FC) fiber coupler; ($\lambda/2$ and $\lambda/4$) half and quarter wave fiber loops; (F) single-mode optical fiber; (PZT) scan piezo; (SFD) shearforce detector; (OP) optical probe; (T) topographic signal. Collection optics: (a) ‘‘TNOM-2’’-setup: (S) sample; (H) hemisphere; (EM) elliptical mirror; (L) lens; (M1) mirror; (An) polarization analyzer configuration described in the text; (P) pinhole; (PM1, PM2) side-on photomultipliers for allowed and forbidden light, respectively; (G) circular glass window. (b) Inverted microscope setup: ‘‘TNOM-3’’; (O) microscope objective. (c) Waveguide prism coupler setup: ‘‘TNOM-4.’’

The background contribution, $\mathbf{E}_0^D(\mathbf{k})$, in this angular regime is caused by modes in the gap region which are *propagating*. ‘‘Allowed’’ light therefore is detected for any gap width z_t .

The background contribution of waves propagating at an angle $\theta_c \leq \theta \leq \pi/2$ and a lateral wave vector of

$$k_0 \leq |\mathbf{k}| \leq n k_0, \quad (9)$$

respectively, emerges from modes which are *evanescent* in the gap region. The background intensity therefore decays exponentially with increasing gap width, similar to the decay of evanescent waves generated by total internal reflection. This is the classically ‘‘forbidden’’ angular regime since light emitted by a far-field source can only propagate in the substrate at angles up to θ_c . Selective detection of the ‘‘forbidden light’’ therefore has the advantage that its background is unaffected by stray light, for instance light leaking out from imperfections on the shaft of the NFO probe.

C. Resolution

The field $\mathbf{E}_1^D(\mathbf{k}) = e^{ik_z z_D} \mathbf{E}_1^{S^-}(\mathbf{k})$ contains relevant data about the sample structure. Information about a particular Fourier component $\mathbf{S}^-(\mathbf{k}')$ is conveyed to the far field only if at least one of the components $\mathbf{S}(\mathbf{k}') \cdot \mathbf{E}_0^{S^+}(\mathbf{k} - \mathbf{k}')$ contributing to $\mathbf{E}_1^{S^-}(\mathbf{k})$ can reach the detector. In order to resolve a sample structure with a maximum spatial wave vector $|\mathbf{k}'| = k_{\max}^S$ it is necessary that

$$k_{\max}^S \leq k_{\max}^P + k_{\max}^{XD}. \quad (10)$$

Here k_{\max}^P is the maximum lateral wave vector of the probe field in the sample plane S^+ and $k_{\max}^{XD} \leq n k_0$ is the maximum

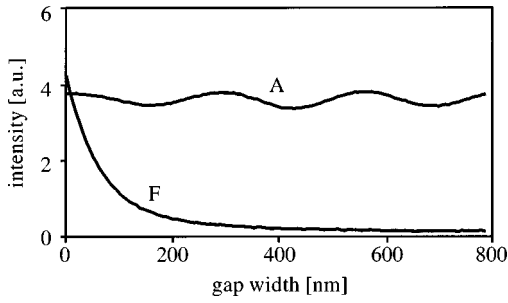


FIG. 5. Approach curves obtained with the TNOM-2 setup for a small aperture optical probe illuminated by $\lambda = 514$ nm light and a plane dielectric interface (BK7, $n=1.51$) as sample: (A) allowed light, (F) forbidden light.

lateral wave vector of plane waves arriving at the detector. Obviously, in order to optimize the resolving power, k_{\max}^{XD} should be chosen as large as possible. In standard SNOM which detects “allowed” light only ($X=A$), the maximum receiving angle is the critical angle θ_c corresponding to $k_{\max}^{AD} = k_0$.

The NFO microscopes used in our experiments [Figs. 4(a), 4(b) and 4(c)] receive and detect the “forbidden” light signals ($X=F$) in addition to the allowed ones. The maximum k vector of the forbidden light is $k_{\max}^{FD} = nk_0$, corresponding to $\theta = \pi/2$. The setup of TNOM-2⁸⁻¹⁰ detects all this radiation by means of its elliptical collection mirror.

The immersion oil objective with $NA=1.3$ of the inverted microscope setup, TNOM-3 [Fig. 4(b)], provides $k_{\max}^{FD} = NA \cdot k_0$ and a maximum receiving angle of $\theta = \arcsin NA/n \approx 61^\circ$ only. A look at the angular dependence of the forbidden radiation,⁷ however, indicates that the major contributions in fact stem from the angular range near θ_c . The resolving power of TNOM-3 therefore is comparable to that of TNOM-2. The combination with a classical microscope, however, is better suited for localization and alignment of the interesting sample features.

The above values of k_{\max}^{XD} represent absolute limits since the SNOM detectors integrate over the whole allowed and forbidden angular range, respectively:

$$k_{\min}^{XD} \leq |\mathbf{k}| \leq k_{\max}^{XD}. \quad (11)$$

Here $k_{\min}^{XD} = 0$ or k_0 , for $X=A$ or $X=F$, respectively.

Since the far field intensity in general peaks in forward direction and behind the critical angle (see Fig. 6 and Refs. 7 and 19), one has to assume that the effective k_{\max}^{XD} is a fraction $\gamma^X \leq 1$ of k_{\max}^{XD} only. The resolution limits of the “allowed” and “forbidden” images hence are

$$\begin{aligned} k_{\max}^{S,A} &= k_{\max}^P + \gamma^A k_0, \\ k_{\max}^{S,F} &= k_{\max}^P + \gamma^F n k_0. \end{aligned} \quad (12)$$

It is seen from Eq. 12 that forbidden light detection can improve the resolution over allowed light detection by a factor

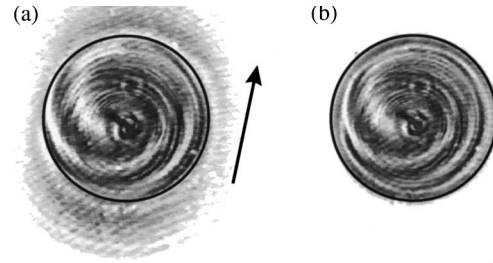


FIG. 6. Far-field emission patterns of an optical probe with an aperture of 80 nm diameter and linear polarized illumination at $\lambda = 633$ nm over a plane dielectric interface ($n=1.51$): (a) shearforce contact, gap width ≈ 10 nm (the arrow indicates the direction of polarization); (b) gap width ≥ 600 nm. The patterns were recorded with TNOM-3 [Fig. 4(b)], slightly defocused, with a video camera as detector. The black circle indicates the border between forbidden and allowed light.

$$\begin{aligned} \Omega &= (1 + n\gamma^F\alpha)/(1 + \gamma^A\alpha) \\ &\approx 1 + \alpha(n\gamma^F - \gamma^A) \end{aligned} \quad (13)$$

with $\alpha = k_0/k_{\max}^P$. Assuming $\alpha \approx 0.1$, $\gamma^A \approx 0.3$ and $\gamma^F \approx 0.85$, one arrives at a resolution improvement of 20%. Employment of a high refractive index substrate made, e.g., of heavy flint glass or of SrTiO_3 can further increase the forbidden light resolution capability. The improvement is considerable for apertures with diameters around 100 nm. Further advantages of forbidden light imaging are a better suppression of stray light and the soft high pass filtering effect due to the limitation of k^D to values larger than k_0 .

D. Contrast

To get an idea of the magnitude of the contrast achievable with forbidden and allowed light imaging, the amplitudes $\mathbf{E}_{\text{tot}}^D(\mathbf{k})$ were calculated for a specific sample structure by means of the “multiple-multipole” method (MMP).²⁰ The test sample is a bare glass grating with a periodicity of 380 nm, consisting of 10 nm high, 190 nm wide terraces, separated by valleys of the same width.⁸⁻¹⁰

The optical probe is a slit of 50 nm width, formed by an aluminum film on the side walls of a glass wedge. The wedge is oriented parallel to the grooves of the grating which reduces the problem of solving Maxwell’s equations to two dimensions. This simplifies the computational effort without loss of the essential features of the NFO imaging process.¹⁹

Figures 2(a) and 2(b) show the computed energy densities $|\mathbf{E}|^2$ for p -polarized excitation with the probe over the center of a terrace and of a valley, respectively. The radiation is more concentrated in the forward direction in the first case than in the second, where it also spreads sideways. A similar behavior is found for the transmission of the s -polarized light. These different spreading properties remind us of classical focusing/defocusing by convex/concave lenses, respectively.

The impression conveyed by Figs. 2(a) and 2(b) is confirmed by the far-field polar plots for p and s polarization in Figs. 2(c), 2(d) and 2(e), 2(f), respectively. The plots represent the differences of the radiation patterns with and without

the sample. The intensity differences in the allowed and forbidden light signals are given by the areas and the signs of the lobes within the intervals $[-\theta_c, \theta_c]$ and $[\pm\theta_c, \pm\pi/2]$, respectively. The forbidden light intensity obviously increases when the tip moves from terrace [see Figs. 2(c) and 2(e)] to valley [see Figs. 2(d) and 2(f)] position. The opposite is true for the allowed light signal. As a consequence, the valleys will show up bright in the forbidden image but dark in the allowed one; the opposite is true for the terraces.

It turns out that this contrast reversal is characteristic for small (or edges of large) dielectric topographic features and small *phase objects* in general. They can be distinguished from *amplitude objects* in this way because the latter were found to provide the same contrast in both allowed and forbidden light images. Experimental results concerning both amplitude and phase objects have been presented in Refs. 8–10. This behavior also implies that if both allowed and forbidden light are focused onto the same detector, the achieved contrast may be decreased rather than increased because of an unfavorable phase.

III. EXPERIMENT

Detection of light behind the critical angle of TIR in a NFO microscope has been proposed for the first time by Fischer *et al.*²¹ The setups used here to realize forbidden light NFO imaging²² are the *tunnel near-field optical microscopes* TNOM-2^{7–10} and the new TNOM-3 which was briefly described in a recent publication.²³ TNOM-1 was used for the first exploratory experiments only;^{6,7} TNOM-4 so far underwent some preliminary, successful tests. We feel that a more detailed description of the instruments, in particular TNOM-3 is required in view of the numerous results obtained with the TNOM technique.^{8–10,23}

A. probe head

The probe heads of the different TNOMs (Fig. 4 head) are identical and interchangeable. They resemble to a great extent the classical SNOM arrangements.^{2,4} An Ar⁺ or HeNe laser (L) is used as a light source in all experiments. It is coupled into a single mode fiber (F) by means of a fiber coupler (FC). The output of the laser is attenuated to 10–20 mW. This leads to an intensity of several mW in the single-mode fiber. The polarization of the light in the fiber can be adjusted by means of a “fiber loop polarization controller.”²⁴ At the far end of the single-mode fiber a metal-coated tip with an aperture of 80–100 nm at the apex serves as near-field optical probe (OP). The optical probes are produced by heating, pulling, and subsequent coating with 100 nm of aluminum.⁴ The coating quality and aperture size for each individual tip is characterized in a scanning electron microscope. Figure 3 shows examples of tip apex shapes observed. Only tips with small and circular aperture were used in our experiments. A point of concern is the irregular grainy structure of the coating at the apex which seems to be an inherent property of Al films. Its elimination would greatly enhance the reproducibility of the NFO probes and the quality of the optical images.²⁵

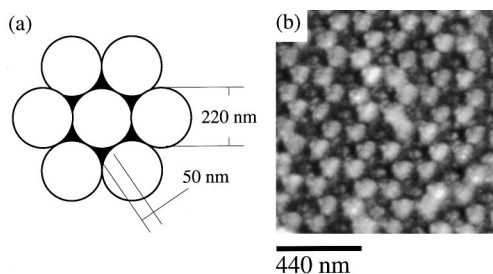


FIG. 7. Metal island film: (a) Sketch of the metal patches arranged in hexagons in the interstices of the densely packed 220 nm latex spheres. (b) Typical shearforce image of such a structure.

Suitable optical probes are mounted onto a piezo tube with a scan range of about 6 μm . The shear force technique can be used for gap width control.^{26,27} A highly sensitive fiber interferometer (SFD) which is attached to the piezo tube (PZT) allows one to detect resonant lateral vibrations of the probe. For vibration amplitudes of 1 nm and a bandwidth of 10 kHz a signal to noise ratio of 5–10 can be achieved in the detected vibration signal. This leads to a topographic signal with a z noise of about 1 \AA .

The head also comprises a means for the mechanical rough and fine approach of the optical probe to the sample.

B. Collection optics

Sample substrates (S) usually are plane-parallel glass slides. Forbidden light is totally reflected at the backside of the slide and therefore lost unless special precautions are taken.

1. TNOM-2

In this arrangement the substrate carrier is a glass hemisphere in optical contact with the sample substrate slide. When properly aligned, all radiation emerging from the NFO probe hits the (hemi-)spherical surface at normal incidence, providing an undistorted projection of its angular distribution in a detector plane beneath the hemisphere.

The forbidden radiation is refocused efficiently by means of a confocally arranged segment of an ellipsoidal mirror. The segment is limited by two planes perpendicular to the main axis of the ellipsoid intersecting at one focus and at a position determined by the critical angle of hemisphere flat surface, respectively. A side-on photomultiplier (PM2), located at the second focus of the ellipsoid, provides the forbidden light signal.

The allowed light is collected with an aspheric condenser lens (L1) with $\text{NA}=0.5$ corresponding to a maximum acceptance angle of 30° . The angular range between 30° and the critical angle $\theta_c=41.5^\circ$ is shadowed by the lens holder. The lens holder is mounted into a circular glass window (G) to avoid distortion of the forbidden light. The object distance is chosen such that an image of the optical probe is generated at a distance of 125 mm behind the sample plane. A 500 μm diam pinhole (P) in the image plane cuts off most of the stray

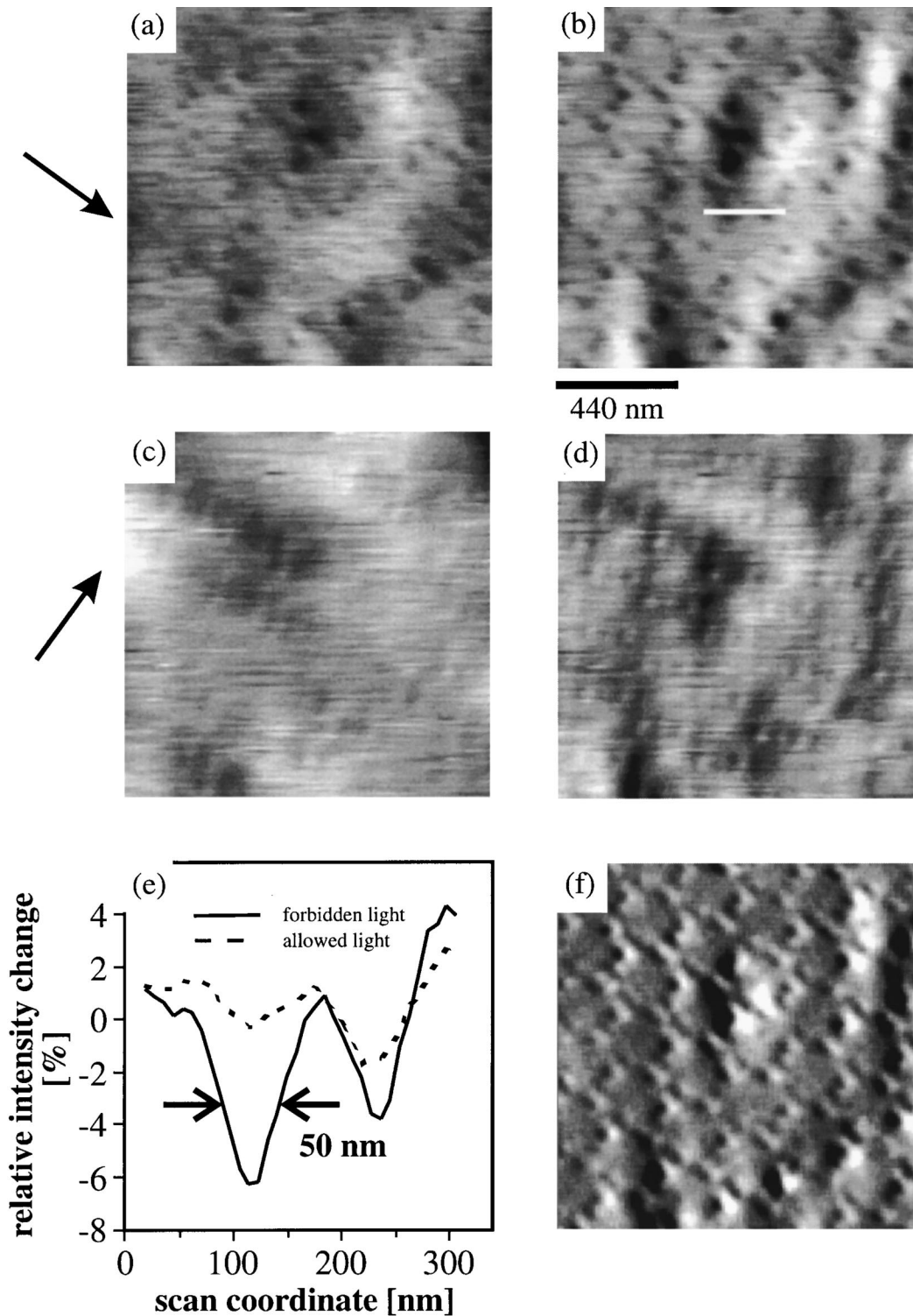


FIG. 8. Metal island film imaged with a small aperture: allowed light (a),(c) and forbidden light images (b),(d) were recorded in constant height mode. The polarization denoted by the arrows was rotated by $\approx 90^\circ$ between the first and the second row. (f) High pass filtered version of the forbidden light image (b); (e) cut along the white line in (b) showing the relative intensity changes of forbidden and allowed light, respectively. The absolute average intensity in allowed light was about twice the forbidden light average intensity. The wavelength used for illumination was 633 nm. All images show raw data.

light from various sources such as light from imperfections in the coating of the optical probe far away from the tip apex. The pinhole is chosen large enough on the other hand, to tolerate small deviations from ideal alignment without cutoff of allowed light from the the tip apex region. The box labeled An in Fig. 4(a) is a polarization analyzer combined

with a mirror oriented 45° out of the paper plane. The arrangement compensates for the elliptic polarization introduced by the reflection at the first metallic mirror (M1). The allowed radiation is focused onto another photomultiplier (PM1) which is of the same type as PM2 to allow a quantitative comparison of the respective signals.

In contrast to high NA objectives, which can accept only very thin samples, the TNOM-2 arrangement also admits thick samples. In this case the height of the hemisphere must be reduced to compensate for the finite thickness of the sample.

The limitations of this setup are given by the unavoidable deviations from ideal alignment of the optical probe with respect to H and hence to EM. Ray tracing shows, however, that the resulting deviations from spot size and spot position do not critically depend on the misalignment parameters. Specifically, the size of the active area of PM2 comfortably accounts even for the maximum possible deviation.

2. TNOM-3

Many applications require conventional microscopic inspection of the sample in order to position the tip in regions of interest.²³ The TNOM-3 is therefore based on an inverted microscope (Zeiss Axiovert 10) that utilizes an immersion oil objective with NA=1.3 for light detection [see Fig. 4(b)]. This corresponds to an acceptance angle $\leq 61^\circ$ for the transmitted light. This value is large enough for forbidden light recording since the major contributions stem from the angular range near θ_c .⁷

The objective is infinity corrected, i.e., light emitted by the optical probe is collimated into a parallel beam behind the objective if its apex is in focus. This greatly facilitates the separation of forbidden and allowed light because the allowed light can be reflected out of the center of the collimated beam by a mirror of suitable diameter as shown in Fig. 4(b). The setup also allows one to image the angular intensity distribution very easily by placing a suitable image sensor in the collimated beam (see Fig. 6).

A prerequisite for unperturbed collimation is the fixed relative alignment of objective and apex of the optical probe. The sample therefore has to be scanned in this configuration. An *x-y*-bimorph scanner of the type first introduced by Murralt *et al.*²⁸ turned out to be particularly suitable for this purpose. The scanner allows for a flat design, large scan range, and low voltage operation. The symmetric arrangement of the bimorph elements makes it insensitive to temperature drifts during operation.

3. Still another configuration ("TNOM-4")

In this particularly simple configuration a standard object carrier is used as a sample substrate [see Fig. 4(c)]. The totally reflected radiation propagates sideways in a zigzag path until it arrives at the sidewalls of the sample substrate. There the light can be coupled out by appropriate prisms or gratings and directed onto detectors (PM2). Alternatively, the detectors may be brought directly in optical contact with the object carrier near the side walls.²⁹

A favorable arrangement may utilize circular substrates with the sample at the center; detectors arranged around the circumference will provide information on the azimuthal distribution of the probe light which in turn allows conclusions on the state of polarization and its change by the object.

First experimental tests indicated a performance comparable to TNOM-2 and TNOM-3 as can be expected.

C. Influence of the gap width

1. Approach curves

The gap width dependence of allowed (*A*) and forbidden light (*F*) intensities for an ≈ 80 nm aperture optical probe illuminated by $\lambda=514$ nm light and a plane dielectric interface (BK7, $n=1.51$) as a sample is shown in Fig. 5. The $1/e$ decay length is approximately 70 nm for the forbidden light intensity (*F*). This corresponds to the decay length of an evanescent wave generated by total internal reflection (TIR) of a plane wave and an angle of incidence of $\approx 53^\circ$ as it is used in a scanning tunneling optical microscope.^{30,31} In our experiment the optical probe generates such evanescent waves which, by inversion of the TIR light path, are converted into forbidden radiation. The angle of 53° for the average direction of the forbidden light is in qualitative agreement with the measured intensity/angle relation in Ref. 7 and numerical results of Figs. 2(c)–2(f) as well as of Ref. 19. The allowed light (*A*) shows interference undulations with a periodicity of $\lambda/2$.^{2,8,32}

At shear force contact the intensity ratio between forbidden and allowed light depends strongly on (i) the aperture diameter and (ii) the effective distance between the aperture and the average sample surface. This latter is limited, e.g., by protrusions on the rim of the aperture [cf. Fig. 3(c)] which prevents complete closure of the gap. Thus the intensity ratio at shear force contact provides a simple criterion for the NFO quality of an optical probe: A "good" tip with aperture diameter ≤ 100 nm and small protrusions on the aperture rim provides intensity ratios $I_{\text{forbidden}}/I_{\text{allowed}} \geq 1$ (see Fig. 5). The ratio increases with further reduction of the aperture diameter or size of protrusions.

2. Emission patterns

Figure 6 shows the far-field emission pattern for small (a) and large (b) gap widths recorded with TNOM-3 on a flat substrate. The black circle indicates the border between forbidden and allowed light. Its radius corresponds to the critical angle of TIR. The two lobes outside the circle in Fig. 6(a) represent the contribution of the forbidden light to the far-field pattern. They do not show up in Fig. 6(b) as expected from the approach curves of Fig. 5. The spiral structures visible inside the circles of allowed radiation are caused by interference of light within the detection optics and are irrelevant for the present discussion.

The forbidden light lobes are oriented along the direction of polarization (*electric* field, see arrow in Fig. 6). This is to be expected for illumination with linearly polarized light^{16,17} and a manifestation of the *magnetic* dipole character of a small aperture in a *metallic* screen.

D. Constant height imaging

All optical images presented and discussed here are recorded in constant height mode (CHM) in order to avoid the

so-called z -motion artifact.²⁵ In order to still maintain a constant small gap width during scanning a number of alignment steps have to be made:

(1) The plane raster scanned by the optical probe is oriented parallel to the mean sample surface plane. This is achieved by compensating the tilt in the corresponding constant gap width mode (CGM) image with the help of an electronic amplifier matrix which adds adjustable fractions of the x - and y -voltage ramps to the z -control voltage of the piezo scanner.

(2) The probe is retracted mechanically. This is compensated by the feedback loop via an elongation of the piezo tube. The retraction is stopped when $\approx 9/10$ of full elongation of the tube piezo are reached. The mechanical retraction might introduce a drift of the scanned area, therefore another CGM image might be necessary at this point as a cross check.

(3) A (negative) offset voltage is added to the z -piezo preamplifier that would further retract the optical probe. To compensate, however, the feedback loop raises the output voltage until the preamplifier is saturated. Any further increase in offset voltage beyond that point translates to a corresponding increase in gap width.

(4) The optical probe can now be positioned at the desired height, e.g., just above the highest elevation within the CGM image by changing the offset voltage and thus the piezo elongation. If shear force contact still occurs during scanning, the feedback loop becomes activated and retracts the optical probe as usual. In this way the tip is protected against crashes that might occur, e.g., because of thermal drift or protruding topographic features within the scan range.

E. Imaging tests

1. Metal island film

The test sample is a 15 nm thick aluminum island film produced by means of the latex sphere shadow mask technique.^{33,34} The structure is sketched in Fig. 7(a). The objects of interest are the metal patches at the interstices between the spheres. They are formed by evaporation and subsequent dissolution of the spheres. To characterize their size, we propose to choose the altitude of the inscribable equilateral triangle as a standard. The length of this altitude is $h = (\sqrt{3} - 3/2)d \approx 0.23d$ where d is the sphere radius. For the $d = 220$ nm latex spheres used for the present test sample, the altitude hence is 50 nm. The nearest-neighbor distance between metal patches is $d/\sqrt{3} = 127$ nm.

The calculated distance between the patches is in agreement with the shear force image [Fig. 7(b)] of the sample obtained after step (2) of the CHM alignment procedure. The shear force image represents a convolution of a rather blunt protrusion on the aperture rim (see Fig. 3) with the real sample topography. Therefore the triangular shape of the metal patches, which was confirmed independently by scanning electron microscopy, is not visible.

The series of CHM optical images (Fig. 8) was obtained with the same tip as used for Fig. 7(b) and at the same

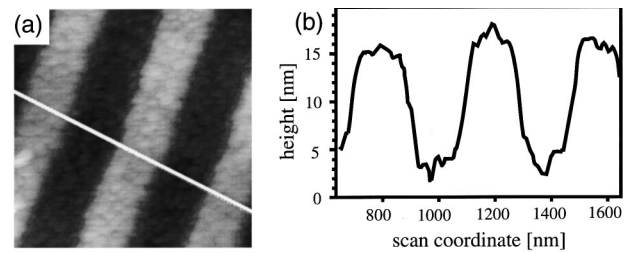


FIG. 9. (a) Tapping mode AFM image of the grating test structure. The line cut (b) is taken along the white line in the scan image (a).

location on the aluminum island film. The light emitted by the optical probe was linearly polarized (arrows in Fig. 8) in two orthogonal directions. The extinction ratio measured for the allowed light was about 15. Figures 8(a) and 8(c) show the respective allowed and Figs. 8(b) and 8(d) show the forbidden light images. Figure 8(e) presents a cut through two neighboring dots at the position indicated by the white bar in Fig. 8(b). Figure 8(f) is a high pass-filtered representation of Fig. 8(b). Remember that CHM images are of purely optical nature.

The optical scan images show a distinct fine structure. The hexagonal pattern corresponding to the metal patches is in particular very well visible in the forbidden light images. The allowed light images are less well resolved; complete hexagons can hardly be found. The structures in the images in the upper row of Fig. 8 show systematic distortions. In particular, the patches are elongated in the direction of polarization. This might be attributed to the polar concentration of the electric energy on opposite sides of the rim of a small

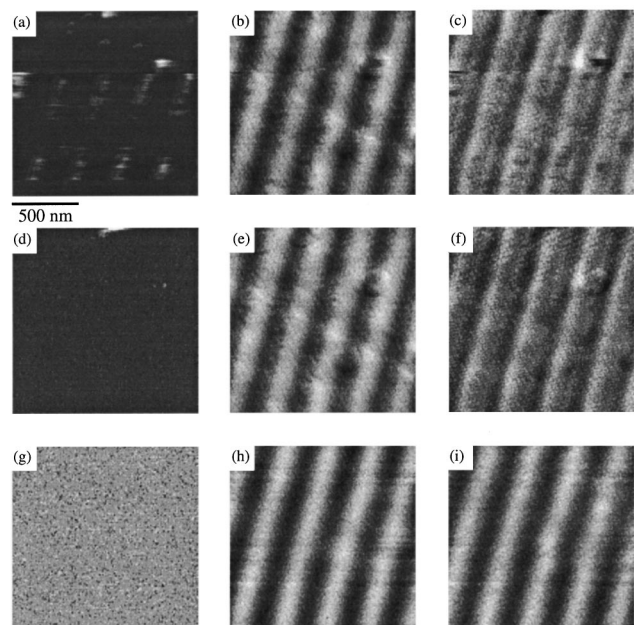


FIG. 10. Series of grating images in constant height mode; increasing gap width (a)–(c) ≈ 0 nm, (d)–(f) few nm, (g)–(i) ≈ 100 nm. Columns: residual topography, allowed light, forbidden light.

aperture^{16,17,35} which is the near-field counterpart of the far-field lobes discussed in Sec. III C 2. This rather sharp double peak structure is washed out a few nanometers behind the aperture plane, forming an approximately elliptic spot. The direction of strongest confinement hence corresponds to the direction perpendicular to the polarization in agreement with the observed elongation.

Rotation of the polarization by 90° leads to the images in the second row of Figs. 8(c)–8(d). The contrast is strongly reduced now. The hexagonal arrangement of the 50 nm dots is still visible in forbidden light, whereas the allowed light image shows a very weak modulation only. Elongations along the direction of polarization are weakly visible in the forbidden light image [Fig. 8(d)].

The optical resolution in both forbidden and the allowed light images [Figs. 8(a), 8(b) and 8(f), respectively] can be estimated from the line cut [Fig. 8(e)] to be at least 50 nm. The relative intensity changes of forbidden and allowed light differ by a factor of up to 6, indicative of the higher contrast achievable with forbidden light.

2. Topographic grating

A grating structure (period 383 nm, step height 8 nm) etched into a thin glass substrate was chosen as the second test object. A tapping mode atomic force microscope (AFM) picture of the structure is shown in Fig. 9.

Figure 10 shows a collection of CHM images of the grating. The depicted area is not the same as in Fig. 9. The gap width was increased from ≈ 0 nm in Figs. 10(a)–10(c) to 5 nm in Figs. 10(d)–10(f) and finally to 100 nm in Figs. 10(g)–10(i).

In the first row [Figs. 10(a)–10(c)] the feedback loop still is marginally active retracting the tip by less than 1 nm as it moves over the highest elevations of the grating. This allows one to correlate these topographic features to the observed optical structures. The bumps in Fig. 10(a) correlate with the dark spots in the corresponding positions in the forbidden light optical images [Figs. 10(c) and 10(f)] as expected from Fig. 2. The correlation with the bright spots in the allowed light images Figs. 10(b) and 10(e) is less obvious. The contrast reversal with regard to the grating lines in allowed and forbidden light images, on the other hand, can unambiguously be confirmed by comparison of the columns of allowed and forbidden light images, respectively, see Sec. II D. The resolution decrease with increasing gap width is documented in Figs. 10(g)–10(i). Only the fundamental periodicity of the grating can still be resolved. The optical signature of the topographic feature in Fig. 10(c), for example, is no longer visible in Fig. 10(i). The resolution in the forbidden light image is slightly better than in the allowed light image as more of fine structure is still visible along the grating lines in Fig. 10(i) than in the respective allowed light image.

IV. SUMMARY

The separate but simultaneous detection of forbidden and allowed light provides information that could not have been extracted or even found with standard detection

schemes. Theoretical considerations show that the division into forbidden and allowed light is natural, following from a different propagation behavior of the optical modes in the gap between optical probe and sample. Forbidden light originates from evanescent modes, whereas allowed light has its roots in the propagating modes traversing the gap. Theoretical arguments were presented concerning a higher resolution achievable with forbidden light as compared to allowed light. High resolution imaging in constant height mode was used to confirm theoretical predictions for resolution and contrast excluding the risk of topographical artifacts. For (topographic) phase objects, contrast reversal between forbidden and allowed light was found.

ACKNOWLEDGMENTS

The authors wish to thank H. Bach, B. Curtis, C. Durkan, G. Eggers, L. Fleury, H. J. Güntherodt, T. Huser, Y. Inouye, T. Lacoste, and B. Vohnsen. This work was supported in part by the SPP “OPTIQUE” of the Board of the Swiss Federal Institutes of Technology and by the European program “Human Capital and Mobility.”

- ¹D. Pohl, W. Denk, and M. Lanz, *Appl. Phys. Lett.* **44**, 651 (1984).
- ²U. Dürig, D. Pohl, and F. Rohner, *J. Appl. Phys.* **59**, 3318 (1986).
- ³E. Betzig, M. Isaacson, and A. Lewis, *Appl. Phys. Lett.* **51**, 2088 (1987).
- ⁴E. Betzig and J. Trautman, *Science* **257**, 189 (1992).
- ⁵M. Paesler and P. Moyer, *Near Field Optics Theory, Instrumentation, and Applications* (Wiley, New York, 1996).
- ⁶H. Heinzelmann, B. Hecht, L. Novotny, and D. Pohl, *J. Microsc.* **177**, 115 (1994).
- ⁷B. Hecht, H. Heinzelmann, and D. Pohl, *Ultramicroscopy* **57**, 228 (1995).
- ⁸B. Hecht, D. Pohl, H. Heinzelmann, and L. Novotny, *NATO Series E: Photons and Local Probes*, edited by O. Marti and R. Möller (Kluwer Academic, Dordrecht, 1995), Vol. 300, p. 93.
- ⁹B. Hecht, D. Pohl, H. Heinzelmann, and L. Novotny, *Proc. SPIE* **2535**, 61 (1995).
- ¹⁰B. Hecht, D. Pohl, H. Heinzelmann, and L. Novotny, *Ultramicroscopy* **61**, 99 (1995).
- ¹¹D. van Labeke, D. Barchiesi, and F. Baida, *J. Opt. Soc. Am. A* **12**, 695 (1995).
- ¹²D. Barchiesi and D. van Labeke, *Ultramicroscopy* **61**, 17 (1995).
- ¹³D. van Labeke, F. Baida, D. Barchiesi, and D. Courjon, *Opt. Commun.* **114**, 470 (1995).
- ¹⁴M. Nieto-Vesperinas, *Scattering and Diffraction in Physical Optics, Pure and Applied Optics* (Wiley, Chichester, 1991).
- ¹⁵J. W. Goodman, *Introduction to Fourier Optics, Physical and Quantum Electronics Series* (McGraw-Hill, New York, 1968).
- ¹⁶L. Novotny and D. Pohl, *NATO Series E: Photons and Local Probes*, edited by O. Marti and R. Möller (Kluwer Academic, Dordrecht, 1995), Vol. 300, p. 21.
- ¹⁷L. Novotny, D. Pohl, and B. Hecht, *Opt. Lett.* **20**, 970 (1995).
- ¹⁸R. Carminati, A. Madrazo, M. Nieto-Vesperinas, and J.-J. Greffet, *J. Appl. Phys.* **82**, 501 (1997).
- ¹⁹L. Novotny, D. Pohl, and P. Regli, *J. Opt. Soc. Am. A* **11**, 1768 (1994).
- ²⁰C. Hafner, *The Generalized Multiple Multipole Technique for Computational Electromagnetics* (Artech, Boston, 1990).
- ²¹U. Fischer and M. Zapletal, *Ultramicroscopy* **42–44**, 393 (1992).
- ²²B. Hecht, PhD thesis, University of Basel, 1996.
- ²³B. Hecht, H. Bielefeldt, L. Novotny, Y. Inouye, and D. W. Pohl, *Phys. Rev. Lett.* **77**, 1889 (1996).
- ²⁴H. C. Lefevre, *Electron. Lett.* **16**, 778 (1980).
- ²⁵B. Hecht, H. Bielefeldt, Y. Inouye, D. W. Pohl, and L. Novotny, *J. Appl. Phys.* **81**, 2492 (1997).
- ²⁶R. Toledo-Crow, P. Yang, Y. Chen, and M. Vaez-Iravani, *Appl. Phys. Lett.* **60**, 2957 (1992).

- ²⁷E. Betzig, P. Finn, and S. Weiner, *Appl. Phys. Lett.* **60**, 2484 (1992).
- ²⁸P. Muralt, D. W. Pohl, and W. Denk, *IBM J. Res. Dev.* **30**, 443 (1986).
- ²⁹B. Hecht, H. Heinzelmann, L. Novotny, and D. W. Pohl, *Europ. Patent Application*, Internat. Publication No. WO 95/10060 (1993).
- ³⁰D. Courjon, K. Sarayeddine, and M. Spajer, *Opt. Commun.* **71**, 23 (1989).
- ³¹R. Reddick, R. Warmack, and T. Ferrell, *Phys. Rev. B* **39**, 767 (1989).
- ³²H. Heinzelmann *et al.*, *Opt. Eng.* **34**, 2441 (1995).
- ³³U. C. Fischer and H. P. Zingsheim, *J. Vac. Sci. Technol.* **19**, 881 (1981).
- ³⁴H. W. Deckman and J. H. Dunsmuir, *Appl. Phys. Lett.* **41**, 377 (1982).
- ³⁵C. J. Bouwkamp, *Philips Res. Rep.* **5**, 321 (1950).