



ELSEVIER

Ultramicroscopy 61 (1995) 1–9

ultramicroscopy

# Light confinement in scanning near-field optical microscopy

Lukas Novotny<sup>a,\*</sup>, Dieter W. Pohl<sup>b</sup>, Bert Hecht<sup>b</sup>

<sup>a</sup> Swiss Federal Institute of Technology, ETH Zurich, CH-8092 Zurich, Switzerland

<sup>b</sup> IBM Research Division, Zurich Research Laboratory, CH-8803 Rüschlikon, Switzerland

Received 7 July 1995; accepted 7 August 1995

## Abstract

Light propagation and light confinement in scanning near-field optical microscopy (SNOM) are studied by means of the multiple multipole method (MMP). Helmholtz' vector wave equation is solved in three dimensions for four different near-field optical probes: a bare glass tip, a metal-clad aperture probe, an "entirely coated" tip and a novel plasmon probe. The latter two are capable of producing a near-field light spot of less than 20 nm (FWHM) extension. The spots have a single intensity peak at the centre, in contrast to the field distribution behind an aperture, which is doubly peaked. The effect of various parameters on light throughput and confinement is investigated for these probes.

## 1. Introduction

The two most popular concepts in scanning near-field optical microscopy (SNOM) are *aperture* SNOM [1,2] and *photon scanning tunnelling microscopy* (PSTM or "STOM") [3,4]. In *aperture* SNOM, the light-emitting probe consists of a metal-clad, sharply-pointed transparent structure, most commonly a tapered optical fibre. The front end of the probe is left uncoated to form a narrow aperture. The light emitted from the aperture illuminates the sample of interest in close proximity. The probe is raster-scanned over the sample, and for each position the corresponding far-field radiation is recorded. In PSTM/STOM the transparent sample is illuminated with far-field radiation in the total internal reflection (TIR) configuration, such that the generated evanescent wave propagates along the sample surface. A

dielectric, completely transparent probe scans the evanescent field in this case. It is still an open question which of these probes provides the better resolution and contrast.

In principle, the radiation transferred from the optical probe to the sample or vice versa must be kept laterally as confined as possible in order to retrieve optical information of the sample surface with high resolution. The light flux through the confined zone, on the other hand, should be as large as possible for reasons of efficiency. Optical field confinement within dimensions that are considerably smaller than the wavelength can only be obtained only with the help of correspondingly small material structures. This goal can be achieved in two different ways that, to some extent, are complementary: (i) by forcing light through small openings or (ii) by exciting light in a piece of material with subwavelength dimensions.

*Aperture* SNOM is a typical example for the light-forcing configuration. Local light sources can

\* Corresponding author. Fax: +41 1 6321198; E-mail: novotny@ifh.ee.ethz.ch

be realized by excitation of fluorescence [5] or by second harmonic generation (SHG) [6]. These processes allow the clear distinction between point light source and the exciting radiation, but unfortunately suffer from bleaching (fluorescence) or low conversion efficiency (SHG). Another realization of a local light source may be seen in the Mie Scattering of light by individual small scattering centres [7], in particular if these are excited at the plasmon resonance frequency [8,9]. The scattered radiation can be discriminated against the exciting radiation by its direction. This process is less selective than in the first case, but allows high intensity and stability.

The spatial ( $k$ -vector) spectrum of a highly confined light field is much broader than that of a diffraction-limited field distribution as it contains strong evanescent components. The long-wavelength part of these components can be converted into propagating waves by approaching a dielectric half-space to the light source [10]. They are refracted into directions beyond the critical angle (*forbidden light*) [11,12]. A criterion for light confinement can therefore be formulated on the basis of the ratio between forbidden and allowed light intensity when the light source is in the proximity of a flat dielectric substrate.

In the present paper we concentrate on “forcing” configurations. We study and compare light confinement and transmission properties of (a) a bare glass tip as used in PSTM/STOM, (b) a metal-coated aperture probe as used in conventional aperture-SNOM, (c) an “entirely coated” tip [13], a modification of (b) that has several advantageous features, and (d) a novel “plasmon tip” that represents a combination of the “forcing” and “local excitation” concepts. The field distributions of these four configurations are determined by means of the multiple multipole method (MMP), which has already been successfully used in previous near-field optical studies.

## 2. The multiple multipole method

Near-field optics requires the solution of the full formalism of Maxwell’s equations, because both near and far fields are relevant and the objects can range in size from subwavelength to several wavelengths.

For complex structures numerical methods have to be applied as analytical methods are limited to simple configurations. The multiple multipole method [14] is a semi-analytical approach to solve Maxwell’s equations in multiple connected homogeneous, isotropic and linear media. The method has already been applied to different near-field optical problems [12,13,15].

In MMP the field inside each domain is described by a series expansion of known analytical solutions of Maxwell’s equations:

$$\text{field} \approx \sum_n c_n \text{solution}_n, \quad (1)$$

where  $c_n$  are unknown parameters. MMP mainly uses multipoles with different origins as basis functions of the series expansion. Other solutions such as plane waves or waveguide modes may also be introduced. To determine the unknown parameters in the series expansion, the boundary conditions have to be satisfied in discrete points on the interfaces between adjacent domains. As more conditions than unknown parameters are used, the problem leads to an overdetermined system of equations that is solved in the least-squares sense. The error in the individual boundary points is also calculated and represents a measure for the quality of the solution. If the result is not accurate enough, additional or more appropriate basis functions can be included in the series expansion. Note that Maxwell’s equations are exactly fulfilled inside the domains. Being given by an analytical series expansion, the field can readily be evaluated in any desired point. The computational advantage of MMP is that only the boundaries, not the domains themselves, need to be discretized. The choice of suitable sets of basis functions is the most difficult task in MMP as no optimum can be defined in a unique way. Prior knowledge about the solution allows the definition of appropriate basis functions in a heuristic manner. A cylindrical structure, for example, will be expanded in cylindrical waveguide modes and not in plane waves. In the problem presented here, mainly ring-multipoles were used [16].

## 3. Results

In this section we will discuss the shape, field distributions, and properties of the different probes.

All four probes consist of a cylindrical part and a tapered end piece, forming a pointed "tip". Fig. 1 shows the foremost part of the probes, which is of greatest interest here. The figure shows, besides the tip shape, contour plots of the square modulus of the electric field  $|E|^2$ , because at optical frequencies,

only the electric field is responsible for the interaction with matter (fluorescence, polarization, etc.).

The analytically known  $HE_{11}$  waveguide mode [17], incident from the infinite cylindrical structure and polarized in the  $x$ -direction, excites the fields in the tips. The contours are discontinuous in the plane

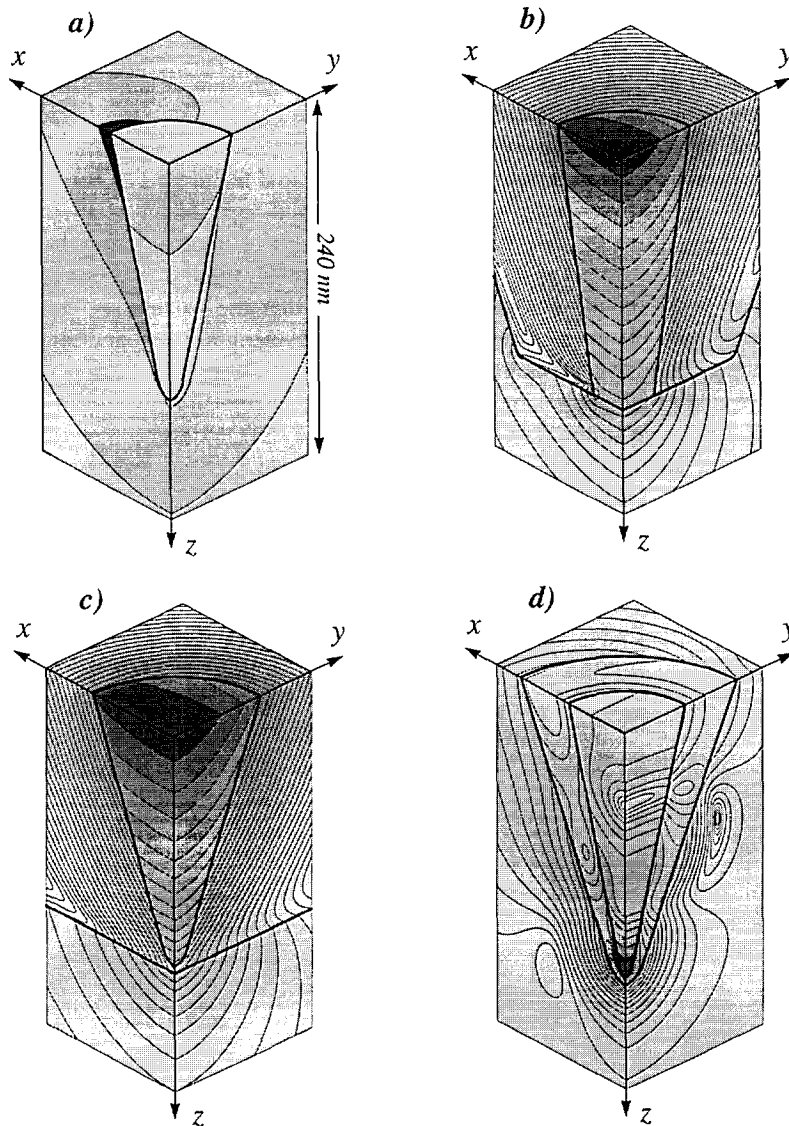


Fig. 1. Contour lines of constant  $|E|^2$  on three perpendicular planes through the centre of the probes (factor of 2 between successive lines). The exciting  $HE_{11}$  mode is polarized along the  $x$ -direction. (a) Dielectric probe:  $\lambda = 488$  nm,  $\epsilon_s = 2.16$ ; (b) aperture probe:  $\lambda = 488$  nm,  $\epsilon_s = 2.16$ ,  $\epsilon_{A1} = -34.5 + i8.5$ ; (c) "entirely coated" probe:  $\lambda = 488$  nm,  $\epsilon_s = 2.16$ ,  $\epsilon_{A1} = -34.5 + i8.5$ ; (d) plasmon probe:  $\lambda = 380$  nm,  $\epsilon_s = 2.16$ ,  $\epsilon_{A2} = -3.39 + i0.19$ .

of polarization, ( $y = 0$ ), as the electric fields have a net component perpendicular to the boundaries. In the perpendicular plane ( $x = 0$ ), on the other hand, the electric fields are always parallel to the boundaries, leading to continuous contour lines. In the following, the details of Fig. 1 will be discussed separately for each type of probe. For the dielectric and the *aperture* probe Fig. 2 shows the decay of the fields along the forward direction. The near fields immediately in front of the probes are plotted in Fig. 3, and the corresponding field confinement is shown in Fig. 4 for the three major directions. Fig. 5 compares light confinement versus light throughput for the “entirely coated” probe, and Fig. 6 shows the spectral field enhancement of the plasmon probe.

### 3.1. The dielectric probe

The purely dielectric probe is a homogeneous glass rod with a pointed end. It is a waveguide structure analogous to the dielectric fibre, but the lower-index cladding is neglected. This simplification, however, becomes irrelevant in the low-diameter tapered region near the apex, where the waveguide fields extend into the surrounding ambient

(air). The tapered, conical part of the probe may be imagined as a series of cylinders with decreasing diameter. At each intersection, the  $HE_{11}$  field distribution adapts to the distribution appropriate for the next slimmer section. This is possible without limit because the fundamental mode  $HE_{11}$  has no cutoff [18]. With each step, however, a part of the radiation is reflected, and the transmitted  $HE_{11}$  mode becomes less confined as the fields extend more and more into the surrounding medium (air). One may hence expect high throughput but little confinement of the transmitted radiation at the level of the tip apex.

The calculation, see Fig. 2a, qualitatively supports the expected behaviour but reveals some interesting additional features: The superposition of incident and reflected light leads to an intensity maximum at a diameter of approximately half the wavelength. Further down, the light penetrates the sides of the probe such that the tip end is in an intensity minimum; subwavelength light confinement is achieved with this configuration only in a subtractive sense.

The consequences on image formation in PSTM/STOM are not obvious, but it is clear that objects located on the sample surface within a distance from the tip of, say, one wavelength strongly

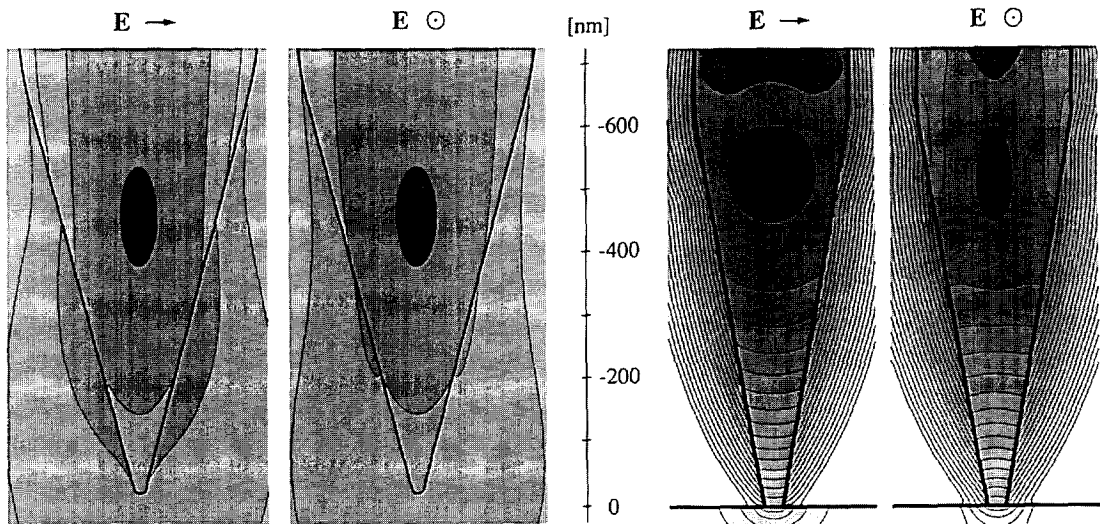


Fig. 2. Contours of constant power density on two perpendicular planes through the centre of the probes (factor of 3 between successive lines). The fields are excited by the  $HE_{11}$  mode (polarization indicated by horizontal arrows) incident from the upper cylindrical part. Left: dielectric probe:  $\lambda = 488$  nm,  $\epsilon_s = 2.16$ ; right: *aperture* probe;  $\lambda = 488$  nm,  $\epsilon_s = 2.16$ ,  $\epsilon_{A1} = -34.5 + i8.5$ . For the dielectric probe a larger taper angle is used so as to have the relevant features on the same scale.

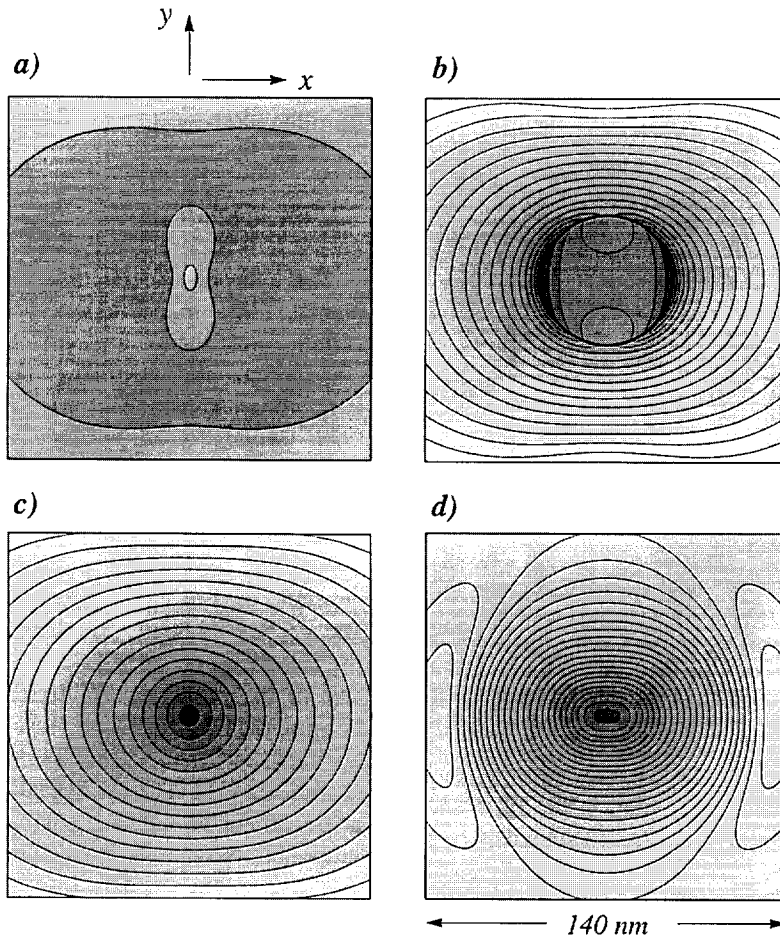


Fig. 3. Near fields evaluated on a plane 1 nm in front of the probes: (a) dielectric probe, (b) *aperture* probe, (c) "entirely coated" probe, (d) plasmon probe. Contours of constant  $|E|^2$  (factor of  $2^{1/2}$  between successive lines).

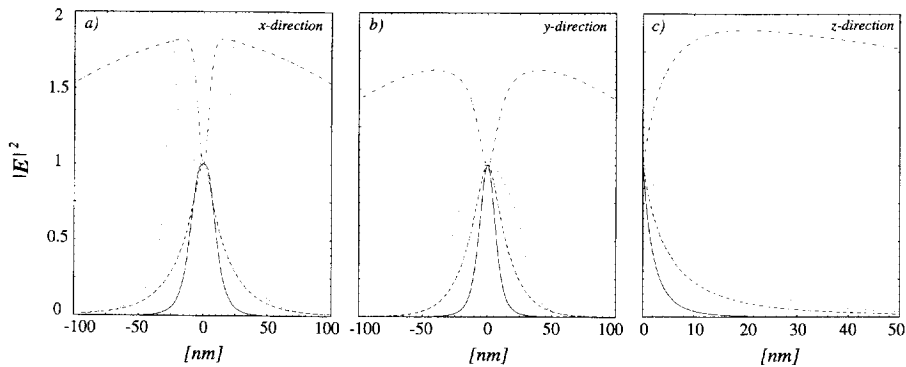


Fig. 4. (a,b) Decay along the  $x$  and  $y$ -direction of Fig. 3, and (c) decay along the forward direction. Dielectric probe (dash-dotted lines), *aperture* probe (dotted lines), "entirely coated" probe (dashed lines) and plasmon probe (solid lines). A laterally confined field is equivalent to a highly concentrated field in three dimensions. Note that the evaluation at distances 1 nm from the probe is arbitrary and that evaluation immediately in front of the probe would lead to an even better lateral confinement.

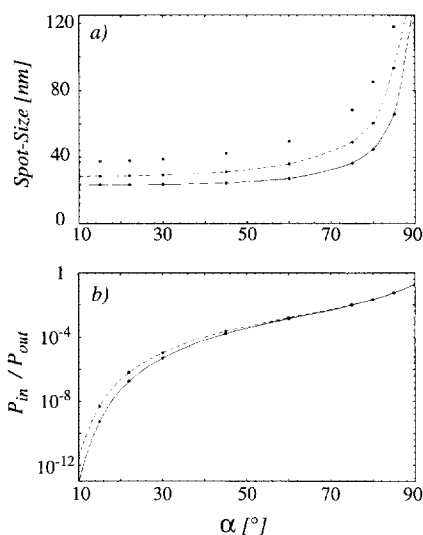


Fig. 5. (a) Spot size of the "entirely coated" probe as a function of taper angle. Tip radius 10 nm. Residual metal thickness in front of tip:  $D = 3$  nm (solid line),  $D = 5$  nm (dashed line) and  $D = 10$  nm (dotted line). (b) Power transmission of the "entirely coated" probe as a function of taper angle. Residual metal thickness in front of tip: 3 nm. Tip radius:  $R = 5$  nm (solid line) and  $R = 10$  nm (dashed line). From Ref. [13].

couple to the radiation propagating in the fibre and may therefore contribute to the recorded signals (see Fig. 4).

The near field of the dielectric probe is evaluated in Fig. 3a on a plane that is 1 nm in front of the tip. The shadow in front of the tip is clearly visible. In

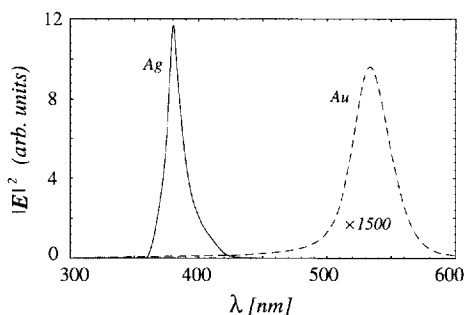


Fig. 6. (a) Field enhancement spectrum evaluated 1 nm in front of the plasmon probe tip. Tip radius is 5 nm and residual metal thickness 5 nm. The curve for the Au-coated probe is scaled by a factor of 1500. For the Ag plasmon probe the electric field intensity at resonance is a factor of  $5 \times 10^4$  higher compared to  $\lambda = 500$  nm. Interpolated dielectric constants for Au and Ag taken from Ref. [21].

the plane of polarization, the field extends over a larger distance than in the plane perpendicular to it. This is due to surface polarization charges resulting from the large electric field components perpendicular to the boundaries.

### 3.2. The aperture probe

For a metal-clad circular waveguide the fundamental  $HE_{11}$  mode decays very rapidly below a certain cutoff radius. Owing to finite damping of the cladding material, no "hard" cutoff exists, i.e. there is no abrupt transition from pure propagating modes (*real propagation constant*) to pure evanescent modes (*imaginary propagation constant*) as is the case for lossless waveguides [17].

In the present example we used an excitation wavelength of  $\lambda = 488$  nm, for which the dielectric constants of core and aluminum cladding are  $\epsilon = 2.16$  and  $\epsilon_{Al} = -34.5 + i8.5$ , respectively. Fig. 2b shows the fast power decay towards the aperture. The intensity in the probe decays faster than exponentially because the core radius decreases on approaching the aperture and causes the modes to become more evanescent [15]. The aperture diameter and the taper angle cannot be reduced arbitrarily because a minimum light throughput must be ensured for the detection process. As a consequence, the low transmission becomes the main limiting factor for the improvement of resolution in SNOM.

Some of the incident power ( $\sim 30\%$ ) is reflected back, leading to a standing wave pattern at the basis of the probe. To the sides of the core the field also penetrates into the aluminum cladding, where a considerable amount of power is dissipated. The skin depth of  $\sim 6.5$  nm is a lower limit for the field confinement and thus also for the ultimate resolution that can be achieved by conventional *aperture-SNOM*.

Fig. 3b presents the near field of a 50 nm *aperture-SNOM* probe evaluated in a plane 1 nm in front of the aperture plane. In the plane of polarization the fields are strongly enhanced at the edges of the aperture owing to the high curvature of the metal surface at the rims (tip effect, lightning-rod effect). The resulting near field of the *aperture-SNOM* probe therefore consists of two spots, and its extent is given by the aperture diameter.

The metal cladding needs to have a minimum thickness in order to attenuate the fields sufficiently in the lateral direction. For too thin a cladding, the fields transmitted from the core through the cladding can couple with external surface modes [15]. In analogy to cylindrical waveguides, they have almost no attenuation [17]. Therefore, most of the energy associated with these modes propagates with small losses on the outer surface of the cladding towards the aperture plane. For too thin a cladding, it can happen that the light from the surface of the cladding is stronger than the light emitted by the aperture, resulting results in a spatially extended light source.

### 3.3. The “entirely coated” probe

As an alternative to the *aperture*-SNOM probe, we recently proposed the “entirely coated” probe [13]. Instead of there being a physical hole in the coating as aperture, the metal cladding covers the entire transparent tip, but its thickness is strongly reduced at the apex region (Fig. 1c). As for this configuration the metal surface is highly curved only in front of the tip, near the field has one maximum that is centred on the symmetry axis of the probe (Fig. 3c). The field distributions in the probe as well as the power decay towards the tip of the probe remain the same as for the *aperture*-SNOM probe. The spot size and the transmissivity of the “entirely coated” SNOM probe were investigated as a function of taper angle ( $= 1/2$  cone angle) and residual metal thickness in front of the tip. The spot size is defined as the average between the FWHM of  $|E|^2$  along the  $x$ - and  $y$ -directions, 1 nm in front of the probe. Surprisingly it was found that the spot size remains almost constant over a large range of taper angles. As seen in Fig. 5, a taper angle of  $45^\circ$  allows a transmissivity that is nine orders higher than for a probe with a taper angle of  $10^\circ$ . Thus, the spot sizes for the two different taper angles are almost the same. Note that Fig. 5 is also valid for the *aperture*-SNOM probe of a given aperture diameter if the abscissae are scaled correspondingly.

The existence of one single spot in the near field of the “entirely coated” SNOM probe opens the possibility of controlled particle trapping. Dielectric particles are dominated by gradient forces  $F \approx \nabla|E|^2$  and by forces caused by radiation pressure

$F \approx |E \times H^*|$  [19]. Gradient forces attract particles towards high-intensity zones, whereas radiation pressure repels the particles from high-intensity zones. In the present configuration the gradient force is able to trap a particle in a very confined region; the light pressure, on the other hand, may repel the particle from the immediate surface of the probe.

### 3.4. The plasmon probe

Scanning near-field optical microscopy using *localized* surface plasmons was introduced by Fischer and Pohl [8]. Silva and Schultz used a similar setup to image of magnetic domains [9]. In both configurations a single metal particle (silver or gold) is used to scatter light locally at the particle’s surface plasmon frequency. The *plasmon tip* introduced here is based on a different scheme, viz., the “entirely coated” tip. For this geometry, plasmons can be excited with visible light if the cladding material is silver (Ag) or gold (Au), i.e. a metal with low electronic interband transitions. Especially for Ag the resonances are only slightly damped.

The plasmon resonance of the nearly hyperboloidal structure is expected to be weakly localized at the apex, which resembles a sphere, with wings extending into the conical shaft that have propagating character. Such a plasmon might be excited by radiation propagating in the dielectric core towards the apex if the parameters are chosen appropriately. It may be accompanied by strong fields extending into the exterior of the tip. The resonance condition depends sensitively on the dielectric constants of the metal,  $\epsilon_m$  and of the dielectric tip,  $\epsilon_s$ . Because of the complicated geometry, it is not obvious which values will provide optimum conditions. Therefore, the field distributions have to be computed for various values. As starting point, the condition for the fundamental cylindrical plasmon resonance,  $\epsilon_m(\omega) = -1$ , or for that of the small sphere,  $\epsilon_m(\omega) = -2$  [20], may be chosen.

For our computations, we first chose a tip with a thick layer of Ag at the apex, assuming that such a geometry would favour the localization. It turned out, however, that this configuration radiates to the sides rather than into the forward direction, leaving a shadow zone in front of the tip. A more favourable situation was found for a silver film that was very

thin at the apex, as in the case of the “entirely coated” probe. For this configuration a plasmon mode could be excited whose field is concentrated at the inner surface of the coating. Proper choice of the wavelength and thus of  $\epsilon_m$  produces a strong maximum at the apex of the dielectric core that penetrates the film and produces a strong, highly confined field on the outer side of the tip. The silver film of the probe depicted in Fig. 1d has a thickness of 5 nm; the radius of curvature of the dielectric core apex is also 5 nm.

Fig. 3d shows the near field on a plane 1 nm in front of the probe at a wavelength of  $\lambda = 380$  nm. At this wavelength the dielectric constants of core and Ag cladding are  $\epsilon_s = 2.16$  and  $\epsilon_{Ag} = -3.39 + i0.19$  [21], respectively. Comparison with the corresponding fields of the other probes shows best field confinement. It is limited only by tip radius and residual metal thickness. At resonance ( $\lambda = 380$  nm) the electric field intensity ( $|E|^2$ ) reaches a value that is almost five orders of magnitude higher than at  $\lambda = 500$  nm (Fig. 6). However, as we used bulk material constants and did not account for mean free path limitation [22], the true enhancement will be somewhat smaller. For an Au cladding the resonance occurs at  $\lambda = 533$  nm but is less pronounced. Note that a core material with higher dielectric constant would shift the plasmon resonance towards the red. Therefore a spectral adjustment of the resonance peak can be established by appropriate selection of the core material.

The presence of a dielectric substrate and objects on its surface also influence the position of the resonance. Thus, the resonance shift provides additional information about the immediate neighbourhood of the tip which may be considered the particular strength of plasmon SNOM. The extremely high field gradients in front of the tip and the existence of a single intensity maximum have again potential for particle trapping.

#### 4. Conclusions

As to be expected, the different probes provide vastly different near-field distributions. The dielectric probe shows high throughput but no subwave-

length light confinement. The *aperture* probe is able to localize light on an area limited by the aperture diameter. However, the resulting near field is not uniform, and small taper angles and small aperture diameters lead to very low throughput, which limits the resolution. As shown for the “entirely coated” probe, which provides a single near-field spot, a larger taper angle ( $\leq 45^\circ$ ) results in a much higher transmissivity and does not affect the size of the near-field spot. One single spot is able to trap dielectric particles and therefore opens the door for local sample modifications. The highest field confinement and the largest field gradients are achieved by a novel plasmon probe. It is possible to excite the plasmon resonance at the foremost part of the probe. The resonance thereby occurs at the core-cladding interface. The field confinement is limited by the tip diameter and the residual metal thickness in front of the dielectric tip.

Imaging the same sample with all of the probes presented here would lead to different images. It is therefore difficult to investigate near-field optical imaging mechanisms as long as there are different possible probe configurations. The highest field-confinement can be achieved by a single dipole, which is the smallest radiating electromagnetic unit (Green’s function). The question about the ultimate resolution in SNOM must therefore be answered by a discrete dipole model that also accounts for the presence of the dielectric substrate.

#### Acknowledgements

We thank Ch. Hafner, P. Regli and O.J.F. Martin for many illuminating discussions. The work was supported in part by the Swiss National Science Foundation.

#### References

- [1] D.W. Pohl, W. Denk and M. Lanz, Appl. Phys. Lett. 44 (1984) 651.
- [2] E. Betzig and J.K. Trautman, Science 257 (1992) 189.
- [3] R.C. Reddick, R.J. Warmack and T.L. Ferrell, Phys. Rev. B 39 (1989) 767.
- [4] D. Courjon, K. Sarayedine and M. Spajer, Opt. Commun. 71 (1989) 23.

- [5] K. Lieberman, S. Harush, A. Lewis and R. Kopelman, *Science* 247 (1990) 59.
- [6] H.M. Hertz, L. Malmqvist, L. Rosengren and K. Ljungberg, *Ultramicroscopy* 57 (1995) 309.
- [7] P. Gleyzes, A.C. Boccara and R. Bachelot, *Ultramicroscopy* 57 (1995) 318.
- [8] U. Ch. Fischer and D.W. Pohl, *Phys. Rev. Lett.* 62 (1989) 458.
- [9] T.J. Silva, S. Schultz and D. Weller, *Appl. Phys. Lett.* 65 (1994) 658.
- [10] W. Lukosz and R.E. Kunz, *J. Opt. Soc. Am.* 67 (1977) 1607.
- [11] B. Hecht, D.W. Pohl, H. Heinzelmann and L. Novotny, "Tunnel" near-field optical microscopy TNOM-2, in: *Photons and Local Probes*, Eds. O. Marti and R. Möller, NATO ASI Series E (Kluwer, Dordrecht, 1995) p. 93.
- [12] L. Novotny, D.W. Pohl and P. Regli, *J. Opt. Soc. Am. A* 11 (1994) 1768.
- [13] L. Novotny, D.W. Pohl and B. Hecht, *Opt. Lett.* 20 (1995) 970.
- [14] Ch. Hafner, *The Generalized Multiple Multipole Technique for Computational Electromagnetics* (Artech, Boston, MA, 1990).
- [15] L. Novotny and D.W. Pohl, Light propagation in scanning near-field optical microscopy, in: *Photons and Local Probes*, Eds. O. Marti and R. Möller, NATO ASI Series E (Kluwer, Dordrecht, 1995) p. 21.
- [16] J. Zheng, *Conf. Proc. ACES* 7 (1991) 170.
- [17] L. Novotny and Ch. Hafner, *Phys. Rev. E* 50 (1994) 4094.
- [18] D. Marcuse, *Light Transmission Optics* (Krieger, Malabar, 1989).
- [19] O. Marti and V. Balykin, Light forces on dielectric particles and atoms, in: *Near Field Optics*, Eds. D.W. Pohl and D. Courjon, NATO ASI Series E 242 (Kluwer, Dordrecht, 1993) p. 121.
- [20] C.F. Bohren and D.R. Huffman, *Absorption and Scattering of Light by Small Particles* (Wiley, New York, 1983).
- [21] P.B. Johnson and R.W. Christy, *Phys. Rev. B* 6 (1972) 4370.
- [22] U. Kreibitz, *J. Phys. F* 4 (1974) 999.