Chapter 6

Near-field optical probes

Optical near-fields are created near material boundaries. More general, the concentration of optical fields to dimensions smaller than the diffraction limit relies on inhomogeneities in an otherwise empty space. The understanding, construction, and optimization of light confining structures is one of the key topics in nano-optics. It is desirable that the actual light confining structure, the optical probe, can be positioned and manipulated. This requires some kind of mechanical support. Because of the evanescent nature of optical near-fields, strong light confinement can only be guaranteed near the material boundaries of the probe. With increasing distance from the probe the fields spread out and the field strength decays rapidly. Near-field optical probes, such as laser-irradiated metal tips, are the key components of near-field optical microscopes discussed in the previous chapter. No matter whether the probe is used as a local illuminator, a local collector, or both, the optical spatial resolution solely depends on the confinement of the optical energy at the apex of the probe.

This chapter discusses light propagation and light confinement in different probes used in near-field optical microscopy. Where applicable we study fundamental properties using electromagnetic simulations (see chapter ??) and provide an overview of current methods used for the fabrication of optical probes. We hope to provide the basic knowledge to develop a clear sense of the potentials and the technical limitations of such probes. The most common optical probes are 1) uncoated fiber probes, 2) aperture probes, 3) pointed metal and semiconductor probes, and 4) nano emitters, such as single molecules or nano crystals. The reciprocity theorem of electromagnetism states that a signal remains unchanged upon exchange of source and detector [1]. Therefore, it suffices to investigate a given probe only in one mode of operation. In the majority of applications it is undesirable to expose the sample surface on a large scale due to the risk of photo damage or long-range interference effects complicating image-reconstruction. Therefore, we will preferentially consider the local illumination configuration.
6.1 Dielectric probes

Dielectric, i.e. transparent tips are an important class of near-field optical probes and in addition are the key components for the fabrication of more complicated probes, e.g. aperture probes. Transparent tips can be produced by tapering of optical fibers yielding conical tips, by suitable breaking of glass slides to produce tetrahedral tips, by polymer molding processes, or by silicon (nitride or oxide) microfabrication techniques. Tips at the end of glass fibers have the distinct advantage that the coupling of light into the taper region can be done easily by exciting the guided modes in the fiber at the far fiber end. Microfabricated or molded tips can be mounted at the end of cleaved fibers. In the following we will discuss the most important methods that can be used to create sharp dielectric tips.

6.1.1 Tapered optical fibers

Tapering of optical fibers can be done by chemical etching, or by local heating and subsequent pulling. Here we compare the results of different etching and pulling techniques and discuss their respective features, advantages and disadvantages.

Etching

Chemical etching of glass fibers is very attractive because it has the potential for batch fabrication of a larger number of identical tips. Initially, etching of glass fibers was done using Turner’s method [2, 3]. Here, fibers with their plastic coating stripped off are dipped into a 40 % HF solution. A thin overlayer of an organic solvent is usually added to (i) control the height of the meniscus of the HF forming at the glass fiber by competing wetting properties of HF and overlayer and (ii) to prevent dangerous

Figure 6.1: Sketch of the Turner etching method. The meniscus height of the 40 % HF solution is expected to decrease as the diameter of the fiber decreases during etching. The process should terminate if the tip is formed. For more details see [2].
vapors to escape from the etching vessel. By using different organic overlayers the opening angle of the resulting conical tapers can be changed [3]. Large taper angles are of interest because, as we will see, they result in high-throughput optical probes. Taper formation in the Turner method takes place because the height of the meniscus is a function of the diameter of the remaining cylindrical fiber. The initial meniscus height depends on the type of organic overlayer. Since the fiber diameter shrinks during etching, the meniscus height is reduced such preventing higher parts of the fiber from being etched further. Finally, if the fiber diameter becomes close to zero the etching process in principle should be terminating itself. The time evolution of the process is sketched in Fig. 6.1.

This sounds quite attractive, but the method has some important drawbacks: (i) The process is not really self-terminating. Diffusion of the small $HF$ molecules into

![Figure 6.2: Schematic view of the time evolution of the tube etching process. The insets show in situ video frames of the etching process. Cleaved fibers are dipped into a 40 % $HF$ solution with an organic overlayer ($p$-xylene or iso-octane). The etching proceeds along different pathways wether or not the polymer fiber cladding is permeable for $HF$. In the case of a nontransparent cladding the tip forms at the end of the fiber and keeps its shape while shortening inside the tube. In the second case the tip forms at the meniscus between $HF$ and organic overlayer. From [4] without permission.](image)
the organic solvent overlayer degrade the tip if it is not removed immediately after it has just finished. (ii) The surface of the conical taper is usually rather rough. This roughness most probably is due to the fact that the meniscus of HF does not move continuously and smoothly during etching but rather jumps from one stable position to the next. This results in a faceted, rather rough surface structure which can pose problems in later processing steps, e.g. resulting in a mediocre opacity of metal coatings.

This roughness problem can be overcome by applying the so-called tube etching method [4]. Here, the fibers are dipped into the HF solution with an organic solvent overlayer (p-xylene or iso-octane) without stripping off their plastic coating. The plastic coatings of standard optical fibers are chemically stable against HF. Depending on the type of fiber, HF can diffuse through the plastic coating or not. Fig. 6.2 schematically shows the progress of the etching process for (a) impermeable and (b) for permeable cladding, respectively. The insets show photographs of the etched fibers in situ. Both types of claddings result in different pathways for tip formation. For more details the reader is referred to the original publications [4].

Fig. 6.3 shows typical results for fiber tips etched by the different techniques. Note the difference in roughness between Turner and tube-etched tips.

Besides the Turner and the tube-etching method there are a number of other etching methods that result in sharp tips. A prominent method was introduced [5] based on dipping cleaved fibers into a buffered HF solution consisting of a mixture with volume ratio $NH_4F:HF:H_2O=X:1:1$, where $X$ denotes a variable volume. In

Figure 6.3: Etched fiber tips. Left: Turner’s method. Right: Tube-etched tip. The upper panels show optical images taken with a conventional microscope. The lower panel shows higher-resolution scanning electron micrographs of the surface roughness of the tips sputtered with 3 nm platinum at 77 K. From [4] without permission.
6.1. DIELECTRIC PROBES

general mixtures with $X_1$ are used. The opening angle of the tips monotonously decreased for increasing $X$ and tends to a stationary value for $X_6$. The magnitude of the stationary angle depends strongly on the $Ge$ concentration. It varies between $100^\circ$ and $20^\circ$ for doping ratios of 3.6 and 23 mol%, respectively. The method relies on the fact that in such a solution $Ge$ rich parts of optical fibers are etched at a lower rate. Since the core of suitable fibers is doped with $Ge$ it starts protruding from an otherwise flat fiber. Fig. 6.4 shows the typical shape of fiber tips created by Ohtsu’s method. The fiber is flat apart from a short and sharp protrusion sitting on the fiber core. For the method to work, the $Ge$ concentration in the core has to have suitable profile which is not the case for all types of standard commercial single mode fibers. More involved techniques have been applied to even achieve tapers with discontinuous opening angles, so called multiple tapers [6].

Heating and Pulling

Another successful method to produce tapered optical fibers is local heating of an intact fiber and subsequently pulling it apart. The technology used here was originally developed for electrophysiology studies of cells using the patch clamp technique. The patch clamp technique was developed in the 70th by Erwin Neher and Bert Sakmann [7] at the Max Planck Institute for Biophysical Chemistry in Göttingen, Germany. In 1991 they were awarded the Nobel price in medicine for this discovery. Micropipettes are produced from quartz capillaries by local heating and pulling. The shape and the apex diameter of heat pulled pipettes depends strongly on all kind of parameters involved in the heating and pulling, like pulling speed profile, size of the heated area, and on the heating time profile.

For applications in nano-optics, as mentioned before, tapered optical fibers should exhibit a short and robust taper region with a large opening angle at the apex. In order to achieve this goal, the length of the heated area of the fiber should be smaller or equal to the fiber diameter. In order to achieve a symmetric tip shape,

Figure 6.4: Scanning electron microscopy images of fiber tips etched with Ohtsu’s method. Scale bars from left to right: 10, 5, 0.5 $\mu$m. From http://phya.snu.ac.kr/~sk_eah/sk_eah/tip_etching.htm without permission.
the temperature distribution in the glass should have cylindrical symmetry. Also, heating of the glass should be moderate because a certain minimum viscosity of the glass before pulling is necessary to achieve short enough tips. A too small viscosity leads to the formation of thin filaments upon pulling. In many labs CO₂ lasers at a wavelength of 10.6 µm are used to heat the glass which at this wavelength is a very efficient absorber. Alternatively, a perforated heating foil or a heating coil can be used. Fig. 6.5 shows a typical setup for heating and pulling of fibers. There exist commercial pipette pullers (e.g., Sutter Instruments) that are used to pull optical fibers since they provide control over magnitude and timing of all relevant process parameters. A rather detailed study on how to adapt a pipette puller for fiber pulling is found in [8].

Close inspection of fiber tips by scanning electron microscopy reveals that pulled tips tend to show a flat plateau at the apex. The diameter of the plateau is a function of the pulling parameters. A probable explanation for the occurrence of the plateau is that there is a brittle rupture once the diameter of the glass filament becomes very small and cooling is very effective. This would imply that the diameter of the plateau should scale with the heating energy applied to the fiber. This was actually observed. Fig. 6.6 shows a series of pulled tips with decreasing heating power. There is also a distinct correlation between opening angle and supplied heating energy. The angle becomes the larger the less heating energy is supplied. Unfortunately, concomitantly also the diameter of the flat facet at the apex increases as can be seen in the insets of Fig. 6.6.

It is important to note that tapers created by etching and by pulling are not completely identical. Some groups report problems with pulled tips when polarization of light is an issue. There seems to be some kind of stress relaxation over time that creates time-dependent polarization behavior of pulled tips [9]. Also, for pulled tips

![Figure 6.5: Sketch of a typical setup for pulling of optical fibers using a CO₂ laser. The laser is focused onto the fiber. For heating, a laser pulse of some milliseconds is applied. The pulling starts after the laser pulse and follows a distinct velocity profile. See e.g. [8] for details.](image-url)
the refractive index profile in the taper is changed since both the fiber core and the cladding are affected by the heating and pulling. For etched tips the fiber core is unaffected as long as the diameter of the taper is still larger than the core diameter. In pulled fibers, in contrast to etched fibers, the thinning of the core can lead to unfavorable mode distortions while the light propagates towards the tip apex. The lower index coating becomes irrelevant in the low-diameter tapered region near the apex, where the waveguide fields extend into the surrounding ambient (air). On the other hand, the tapers of pulled fibers show very little surface roughness which is favorable for subsequent processing, e.g. metal coating.

While the shape of tapered fibers can be accurately determined in scanning electron microscopes, the optical properties, e.g. the effective optical diameter, are more difficult to assess experimentally in a standard way. Here we want to point the interested reader to a method that relies on imaging a pattern of standing evanescent waves [10]. By comparing the measured with the expected fringe contrast using a simple model for the tip’s collection function, one can obtain an estimation for the effective optical diameter of a given tip (see problem 6.1). It is found that for pulled glass fiber tips this value is about 160 nm.

6.1.2 Tetrahedral tips

Tetrahedral tips [11] are produced by cleaving a rectangular slab of glass twice at an angle. Fig. 6.7 schematically shows a resulting fragment with triangular cross section. The fragments can be produced from 170 μm thick cover slips, so that the overall size

![Figure 6.6: Scanning electron microscopy images of pulled glass fibers sputtered with 20 nm of gold. The insets show magnifications of the respective tip apex. There is a trend that the shorter the tip and the larger the opening angle is, the more pronounced a plateau occurs at the apex. This plateau defines the smallest possible aperture that can be achieved after metal coating.](image-url)
of the fragment is rather small. In order to couple in light that is focused to the
tip (marked by the circle in Fig. 6.7) a coupling prism has to be used. A particular
feature of tetrahedral tips is that they are not rotationally symmetric which after
metal coating and aperture formation can lead to interesting field distributions due
to the breaking of the rotational symmetry [12].

6.2 Light propagation in a conical dielectric probe

Dielectric tips can be regarded as homogeneous glass rods with a conical pointed end.
The analytically known $HE_{11}$ waveguide mode, incident from the infinite cylindrical
glass rod and polarized in the $x$-direction, excites the field in the conical probe. For
weakly guiding fibers, the modes are usually designated as $LP$ (linearly polarized). In
this case, the fundamental $LP_{01}$ mode corresponds to the $HE_{11}$ mode. The tapered,
conical part of the probe may be imagined as a series of disks with decreasing diam-
eters and infinitesimal thicknesses. 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 [13]. With each step,
however, part of the radiation is reflected, and the transmitted $HE_{11}$ mode becomes
less confined as the field extends more and more into the surrounding medium (air).
One hence expects high throughput but poor confinement for this type of probe.

The calculated field distribution in Fig. 6.8 qualitatively supports the expected
behavior 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 internal wavelength. Further down the cone, the light penetrates the sides of the
probe such that at the tip apex there is in an intensity minimum; subwavelength light
confinement is achieved with this configuration only in a subtractive sense. Thus, the
fiber probe is not a local illumination source and one can expect that the best field

Figure 6.7: Tetrahedral tip created by cleaving a rectangular slab of glass
twice at an angle. The actual tip is marked by the circle. For details see [11].
confinement is on the order of \( \approx \lambda/(2n_{\text{tip}}) \), with \( n_{\text{tip}} \) being the refractive index of the fiber.

If the field in a plane in front of the probe is transformed into the spectral \((k)\) domain, it is found that evanescent field components are confined to the probe tip, whereas plane wave components are spread over larger distances. Evanescent field components localized to the very end of the fiber probe can be selectively probed by using a high index dielectric substrate which transforms evanescent field components into plane waves propagating in the substrate at angles \( \alpha > \alpha_c \) (forbidden light), where \( \alpha_c \) is the critical angle of total internal reflection. As a consequence, forbidden light contains information on a confined region close to the fiber tip and therefore leads to improved resolution. This finding was experimentally confirmed by Hecht et al. by recording forbidden light and allowed light separately [14]. In general, the spatial \((k\text{-vector})\) spectrum of a highly confined light field is much broader than that of a diffraction-limited field distribution as it contains strong evanescent components. Evanescent components that are transformed into propagating waves in the substrate decay as

\[
e^{iz\sqrt{k_1^2-k_2^2\sin^2(\alpha)}}, \tag{6.1}
\]

where \( k_1 = k_o \) and \( k_2 = n k_o \) are the wavenumbers of the upper medium and the substrate, respectively. It follows, that the larger the refractive index of the substrate \( n \) is, the faster the decay of the exponential term 6.1 will be. Thus, for high \( n \), forbidden light contains information on spatially better confined fields, leading to

![Figure 6.8: Contours of constant power density on two perpendicular planes through the center of a dielectric probe (factor of 3 between adjacent lines). The fields are excited by the \( H\overline{E}_{11} \) mode (polarization indicated by horizontal arrows) incident from the upper cylindrical part. \( \lambda = 488 \text{ nm}, \varepsilon = 2.16 \).](image)
higher resolution.

To understand the efficiency of the fiber probe in the collection mode we simply apply time-reversal to the illumination mode configuration. The essence is as follows: in illumination mode, the $H_{E_{11}}$ mode propagating in the fiber is converted into radiation near the end of the tip. The radiation field can be decomposed into plane waves propagating into various directions with different magnitudes and polarizations (angular spectrum). Reversing the propagation direction of each plane wave will excite a $H_{E_{11}}$ mode in the fiber probe with the same magnitude as used in the illumination mode. Hence, at first glance it seems that no high resolution can be achieved with a fiber probe in collection mode. However, as long as the fields to be probed are purely evanescent, such as along a waveguide structure, the fiber probe will collect only the evanescent modes available and the recorded images will represent the local field distribution. But if the sample contains scatterers which convert the evanescent modes into propagating modes then there is a good chance that the measured signal is dominated by radiation that is coupled into the probe along the tip shaft and image interpretation becomes difficult. Therefore, the fiber probe turns out to be an unfavorable near-field probe for radiating structures.

Resolutions beyond the diffraction limit were still reported by groups using the fiber probe for both illumination and detection (see chapter ?? and e.g. [15, 9]). Although the reported resolutions are lower than those achieved by aperture SNOM, the experiments show that resolution can be further increased by passing light twice through the fiber probe.

![Figure 6.9: Cartoon of the successive cutoff of guided modes and exponential decay of the fields towards the aperture in a tapered, metal-coated waveguide.](image)

Adapted from [16].
6.3 Aperture probes

Probes based on metal-coated dielectrics with a transparent spot at the apex are often referred to as aperture probes. The metal coating basically prevents the fields from leaking through the sides of the probe. The most common example is a tapered optical fiber coated with a metal, most often aluminum. In order to understand the light propagation in such a probe we note that it can be viewed as a hollow metal waveguide filled with a dielectric. Towards the probe apex, the diameter of the waveguide is constantly decreasing. The mode structure in a tapered hollow waveguide changes as a function of the characteristic dimension of the dielectric core [17]. For a cylindrical waveguide this is the diameter. For larger diameters of the dielectric core there exist a number of guided modes in the waveguide. These run into cutoff one after the other as the diameter decreases while approaching closer to the apex. Finally, at a well-defined diameter even the last guided mode runs into cutoff. For smaller diameters of the dielectric core the energy in the core decays exponentially towards the apex because the propagation constants of all modes become purely imaginary. This situation is visualized in a cartoon in Fig. 6.9. The mode cut-off is essentially the reason for the low light throughput of aperture probes. The low light throughput of metal-coated dielectric waveguides is the price for the superior light confinement. Fig. 6.10 shows a comparison of the fields of the fiber probe and the aperture probe obtained from

\[
\lambda = 488 \text{ nm}, \quad \varepsilon_{\text{core}} = 2.16, \quad \varepsilon_{\text{coat}} = -34.5 + i \ 8.5. \quad \text{The exciting } HE_{11} \text{ mode is polarized along the } x \text{ direction.}
\]
an electromagnetic simulation. In both figures the contours are discontinuous in the plane of polarization, \((y = 0)\), as the electric fields have a net component perpendicular to the boundaries. While the dielectric probe shows very low field confinement, the aperture probe suffers from very low throughput. For the latter, approximately one third of the incident light is reflected and two thirds are dissipated.

This behavior determines some of the design goals and limitations of aperture probes. (i) The larger the opening angle of the tapered structure is, and the higher the refractive index of the dielectric core, the better the light transmission of the probe will be. This is because the final cutoff approaches closer to the probe apex [18]. (ii) in the region of cutoff, the energy is partly dissipated in the metal layer. This can result in a significant heating of the metal coating in this region which as a consequence might be destroyed. The maximum power that can be sent down such a probe is therefore limited. Improving the heat dissipation in the relevant region or increasing the thermal stability of the coating can increase this destruction threshold [19]. These effects will be analyzed in some detail in the following section.

### 6.3.1 Power transmission through aperture probes

Fig. 6.11 shows the calculated power density inside an aperture probe. The probe is excited by the analytically known cylindrical \(HE_{11}\) waveguide mode at a wavelength of \(\lambda = 488nm\). At this wavelength the dielectric constants of the dielectric core and the aluminum coating are \(\epsilon_{\text{core}} = 2.16\) and \(\epsilon_{\text{coat}} = -34.5 + 8.5i\), respectively.* The corresponding skin depth is 6.5\(nm\). The core has a diameter of 250 \(nm\) at the upper cylindrical part and a half cone angle of 10° at the taper.

In the cylindrical part the \(HE_{11}\) mode is still in the propagating regime, i.e. its propagation constant has a negligibly small imaginary part. As the core radius becomes smaller, the modes of the tapered part become evanescent and the field decays extremely fast, faster than exponentially, towards the aperture. Since here roughly one third of the incident power is reflected backwards this leads to standing wave pattern at the upper part of the probe. To the sides of the core the field penetrates into the aluminum coating where roughly two thirds of the incident power are dissipated into heat. In order not to destroy the metal coating, the input power has to be kept below a certain threshold.

The fast power decay inside the aperture probe can be well explained by a mode

\[
\epsilon(\omega) = 1 - \frac{\omega_p^2}{\omega^2 + i\gamma \omega},
\]

(6.2)

where a plasma frequency of \(\omega_p = 15.565 \text{ eV}/\hbar\) and a damping constant \(\gamma = 0.608 \text{ eV}/\hbar\) yield a good approximation for the dielectric function [17].

---

*The complex dielectric function of aluminium for visible wavelength can be well described by a plasma dispersion law (see chapter ??),
matching analysis. In this approach, the tapered part of the probe is subdivided into small cylindrical waveguide pieces as shown in Fig. 6.12. For a lossy waveguide the propagation constant $k_z$ of any mode is usually written as

$$k_z = \beta + i\alpha,$$  \hspace{1cm} (6.3)

where $\beta$ is the phase constant and $\alpha$ the attenuation constant. According to waveguide theory, the power loss in the $n^{th}$ waveguide section is

$$P_{loss}(n\, dz) = P(n\, dz) \left(1 - e^{-2\alpha_{11}(n\, dz)\, dz}\right)$$  \hspace{1cm} (6.4)

where $P(n\, dz)$ is the incident power and $\alpha_{11}(n\, dz)$ the attenuation constant of the $HE_{11}$ mode in the $n^{th}$ waveguide section. $\alpha_{11}$ depends on the diameter of the waveguide section, on the wavelength and on the material properties. A more detailed discussion on lossy waveguide modes can be found in Ref. [20]. Summing Eq. (6.4) over all waveguide sections, using

$$P([n+1]\, dz) = P(n\, dz) - P_{loss}(n\, dz),$$  \hspace{1cm} (6.5)

and taking the limit $dz \rightarrow 0$ we obtain the power distribution

$$P(z) = P(z_0) e^{-2\int_{z_0}^{z} \alpha_{11}(z)\, dz}.$$  \hspace{1cm} (6.6)

Figure 6.11: Contours of constant power density on two perpendicular planes through the center of an infinitely coated aperture probe (factor of 3 between adjacent lines). The field is excited by the $HE_{11}$ mode incident from the cylindrical part. The calculation is based on the MMP method.
This formula is compared in Fig. 6.13 with the computationally determined power along the probe axis (curve a). The power in the probe can also be plotted against the core diameter \( D \) using the geometrical relationship

\[
z = -\frac{D - D_a}{2 \tan \delta},
\]

where \( \delta \) is the 1/2-cone angle and \( D_a \) the diameter of the aperture. Note that \( z_o \leq z \leq 0 \) for the coordinates chosen in Fig. 6.12. The asymptotic values of \( P(z) \) are indicated by curves d and e which describe the decay of the \( HE_{11} \) mode in the cylindrical part of the aperture probe and the decay of a wave inside bulk aluminum, respectively. Since the presence of the aperture has almost no influence on \( P(z) \) the curve may be applied in good agreement to any \( D_a \). The power transmission of aperture probes with \( D_a = 100\,\text{nm}, 50\,\text{nm} \) and \( 20\,\text{nm} \) therefore is \( \approx 10^{-3}, 10^{-6} \) and \( 2 \times 10^{-12} \), respectively. The steep decay of the transmission curve (c.f. Fig. 6.13) indicates that in the chosen configuration (especially for the chosen cone angle) it is very difficult to decrease the aperture size considerably below \( 50 - 100\,\text{nm} \), the diameters of most commonly used aperture probes.

For an aperture probe with a thick (infinite) coating, Fig.6.14 shows \( \alpha \) and \( \beta \) for the \( HE_{11} \) mode as a function of \( z \) and \( D \). The transition from the propagating to the evanescent region occurs at \( D \approx 160\,\text{nm} \). The agreement of the power decay obtained by the MMP computation and the power decay obtained by Eq. (6.6) is dependent on the lower integration limit \( z_o \). Excellent fits are obtained if \( z_o \) is chosen to be in the evanescent region of the \( HE_{11} \) mode where \( \alpha_{11}(z) \) is well described by an exponential function

\[
\alpha_{11}(D) = \text{Im}\{n_{coat} \} k_o e^{-AD}
\]

where \( n_{coat} \) is the index of refraction of the metal coating, \( k_o = \frac{2\pi}{\lambda} \) the propagation constant in free space and \( A \) a constant which is determined to be \( 0.016\,\text{nm}^{-1} \) in the

![Figure 6.12](image.png)

Figure 6.12: Mode matching approximation for the power \( P(z) \) in the aperture probe. In each waveguide section the attenuation of the \( HE_{11} \) mode is calculated analytically. The contributions of all sections are added and the limit \( dz \to 0 \) is applied.
present example. If Eq. (6.8) is inserted into Eq. (6.6) and the integration in the exponent is carried out, we arrive at

\[ P(z) = P(z_0) \exp[a - b(e^{2Az\tan\delta})] \]  

(6.9)

with the two constants

\[ a = \frac{\text{Im}\{n_{\text{coat}}\} k_0}{A \tan \delta} e^{-A D_o}, \quad b = \frac{\text{Im}\{n_{\text{coat}}\} k_0}{A \tan \delta} e^{-A D_o}, \]

where \( D_o \) is the core diameter at \( z = z_0 \). According to Eq. (6.9) the power transmission is higher for larger \( \delta \). However, at the same time more power penetrates the edges of the aperture leading to a larger effective aperture width. The analysis above is valid for a \( \delta \) which is not too large since reflections in the probe were neglected. This explains also the deviation of curve \( \text{b} \) in Fig. 6.13 where \( z_o \) was chosen to be in the propagating region of the probe.

The outlined mode-matching analysis can be simplified if a perfectly conducting metal coating is assumed. In this case, the propagation constant \( k_z \) of the lowest order \( TE_{11} \) mode can be calculated as

\[ k_z(D) = \sqrt{\varepsilon_{\text{core}} k_o^2 - (3.68236/D)^2}, \]  

(6.10)

with \( \varepsilon_{\text{core}} \) being the dielectric constant of the core. For large core diameters \( D \) the propagation constant is real and the \( TE_{11} \) mode propagates without attenuation.

Figure 6.13: Power decay in an infinitely coated aperture probe as a function of distance from the aperture \( z \) and of the core diameter \( D \). a: Computed decay, b: mode matching approximation with \( z_o = -600 \text{nm} \), c: mode matching approximation with \( z_o = -400 \text{nm} \), d: decay of the \( HE_{11} \) mode in the cylindrical part of the probe, e: decay of a wave inside bulk aluminum. The vertical line indicates the transition from the cylindrical to the tapered part of the probe.
However, for diameters $D < 0.586\lambda/\sqrt{\varepsilon}$ the propagation constant becomes purely imaginary and the waveguide mode decays exponentially in direction of $z$. Therefore, in the attenuated regime, we can write

$$\alpha_{11}(D) = \sqrt{(3.68236/D)^2 - \varepsilon_{\text{core}}/k_o^2},$$

which can be inserted into Eq. (6.6). A similar analysis has been carried out by Knoll and Keilmann for a perfectly conducting aperture probe with a square cross-section [21].

The throughput of the aperture probe depends also strongly on the taper angle. As the 1/2-cone angle $\delta$ is increased the spot size will decrease because more and more radiation penetrates through the edges of the aperture. Surprisingly, the spot size remains almost constant over a large range of $\delta$ and increases rapidly for $\delta > 50^\circ$ [22]. However, as shown in Fig. 6.15 the power transmission behaves very differently. A strong variation is observed in the range between $10^\circ$ and $30^\circ$. The datapoints in the figure are calculated by three-dimensional MMP calculations for a probe with aperture diameter of 20 nm and excitation at $\lambda = 488 nm$. The solid line on the other hand is calculated according to mode matching theory, i.e. by using Eqs. (6.6)- (6.9). The analysis leads to

$$\frac{P_{\text{out}}}{P_{\text{in}}} \propto e^{-B \cot \delta},$$

with $B$ being a constant. While the above theory leads to a value of $B = 3.1$ the best fit to the numerical results is found for $B = 3.6$. Fig. 6.15 shows that the agreement

![Figure 6.14: Attenuation constant $\alpha_{11}$ and phase constant $\beta_{11}$ of the cylindrical $HE_{11}$ mode as a function of the core diameter $D$. $z$ is the corresponding distance from the aperture. The vertical line indicates the transition from the cylindrical to the tapered part of the aperture probe. From Ref. [20].](image-url)
is excellent for $10^\circ < \delta < 50^\circ$. The deviation above $50^\circ$ is mainly due to the neglect of reflections in the mode matching model. Changing the taper angle from $10^\circ$ to $45^\circ$ increases the power throughput by 9 (nine) orders of magnitude while the spot-size remains almost unaffected. Thus, methods which produce sharp fiber tips with large taper angles are of utmost importance.

### 6.3.2 Field distribution near small apertures

To understand light-matter interactions near an aperture probe we need a model for the field distribution near subwavelength-sized apertures. In classical optics, the Kirchhoff approximation is often applied to study the diffraction of light by an aperture in an infinitely thin, perfectly conducting screen. The Kirchhoff approximation assumes that the field inside the aperture is the same as the excitation field in the absence of the aperture. Of course, this assumption fails near the edges of the aperture, and consequently the Kirchhoff approximation becomes inaccurate for small apertures. For an aperture considerably smaller than the wavelength of the exciting radiation it is natural to consider the fields in the electrostatic limit. Unfortunately, for a wave at normal incidence the fields in the electrostatic limit become identical zero because the exciting electric field consisting of a superposition of incident and reflected waves disappears at the surface of the metal screen. Therefore, the electric field has to be calculated by using a first-order perturbative approach. On the other hand, it is possible to solve the magnetostatic problem.

![Figure 6.15: Dependence of power transmission on taper angle (δ = 1/2 cone angle). The aperture diameter is 20nm and the wavelength $\lambda = 488nm$. Changing the taper angle from $10^\circ$ to $45^\circ$ increases the power throughput by nine orders of magnitude. MMP calculation (points) and according to Eq. (6.12) with a value of $B = 3.6$ (solid line).](image)
In 1944 Bethe has derived an analytical solution for the electromagnetic field near a small aperture [23]. He also showed that in the far field the emission of the aperture is equal to the radiation of a magnetic and an electric dipole located at the center of the aperture. The electric dipole is only excited if the exciting plane wave is incident from an oblique angle. In 1950 Bouwkamp revealed that the electric field derived by Bethe is discontinuous in the hole, contrary to what is required by the boundary conditions [24].

To derive the correct solution, Bouwkamp first calculates the solution for a disk and then uses Babinet’s principle to obtain the magnetic currents for the case of the aperture. The solution is derived from an integral equation containing the current distribution function on the disk as an unknown function. The integral equation is then solved using a series expansion method and making use of the singularity condition at the rim of the disk. This condition states that the electric field component tangential to the edge of the disk must vanish as the square root of the distance from it. Furthermore, the electric field component normal to the edge must become infinite as the inverse square root of the distance from the edge. This boundary condition has been already used before by Sommerfeld in the study of diffraction by a semi-infinite metal plate. An alternative approach for solving the fields near a small disk can be found in Ref. [25].

Babinet’s principle is equivalent to replacing the electric currents and charges induced in the metal screen by magnetic currents and charges located in the aperture. The magnetic surface current density \( K \) and magnetic charge density \( \eta \) in the aperture give rise to a magnetic vector potential \( A^{(m)} \) and a magnetic scalar potential \( \Phi^{(m)} \) as

\[
A^{(m)} = \varepsilon_o \int K e^{i k R} \frac{4 \pi R}{4 \pi R} dS, \quad \Phi^{(m)} = \frac{1}{\mu_o} \int \eta e^{i k R} \frac{4 \pi R}{4 \pi R} dS, \quad (6.13)
\]

where \( R = |r - r'| \) denotes the distance between the source point \( r' \) and the field point \( r \), and the integration runs over the surface of the aperture. Similar to the electric case, \( A^{(m)} \) and \( \Phi^{(m)} \) are related to the electric and magnetic fields as

\[
E = \frac{1}{\varepsilon_o} \nabla \times A^{(m)}, \quad H = i \omega A^{(m)} - \nabla \Phi^{(m)} \approx -\nabla \Phi^{(m)}. \quad (6.14)
\]

In what follows, we neglect the first term in the expression for \( H \) because it is proportional to \( k = \omega / c \) and therefore negligible in the limit of a small aperture \( a (ka \ll 1) \).

To solve for \( A^{(m)} \) and \( \Phi^{(m)} \) it is convenient to introduce oblate-spheroidal coordinates \( r = (u, v, \varphi) \) defined by

\[
z = av, \quad x = a \sqrt{(1 - u^2)(1 + v^2)} \cos \varphi, \quad y = a \sqrt{(1 - u^2)(1 + v^2)} \sin \varphi, \quad (6.15)
\]

where \( 0 \leq u \leq 1, -\infty \leq v \leq \infty, 0 \leq \varphi \leq 2\pi \). The surfaces \( v = 0 \) and \( u = 0 \) correspond
6.3. APERTURE PROBES

Plane wave at normal incidence

For a plane wave at normal incidence, the Laplace equation $\nabla^2 \Phi^{(m)} = 0$ yields the solution

$$\Phi^{(m)} = -H_o \frac{2a}{\pi} P_1^1(u)Q_1^1(iv) \sin \varphi,$$

where $P_n^m$ and $Q_n^m$ are associated Legendre functions of the first and second kind, respectively [26], and $E_o$ and $H_o = E_o \sqrt{\varepsilon_o/\mu_o}$ are the magnitudes of the electric and magnetic field of the incident plane wave polarized in x-direction ($\varphi = 0$). The solution for the magnetic vector potential $A^{(m)}$ is much more difficult to derive since it cannot be calculated statically. The expression derived by Bouwkamp reads as

$$A_x^{(m)} = -\varepsilon_o E_o \frac{ka^2}{36\pi} P_2^2(u)Q_2^2(iv) \sin 2\varphi,$$

$$A_y^{(m)} = \varepsilon_o E_o \frac{ka^2}{36\pi} [ -48Q_0(iv) + 24P_2(u)Q_2(iv) + P_2^2(u)Q_2^2(iv) \cos 2\varphi ] ,$$

and is different from Bethe’s previous calculation.

The electric and magnetic fields are now easily derived by substituting $\Phi^{(m)}$ and $A^{(m)}$ into Eq. (6.14). The electric field becomes

$$E_x/E_o = ikz - \frac{2}{\pi} ikau \left[ 1 + v \arctan v + \frac{1}{3} \frac{u^2+v^2}{u^2+v^2(1+v^2)} \right],$$

$$E_y/E_o = -\frac{4ikxu}{3\pi a(u^2+v^2)(1+v^2)^2},$$

$$E_z/E_o = -\frac{4ikxv}{3\pi(a^2+v^2)(1+v^2)},$$

(6.18)

and the magnetic field turns out to be

$$H_x/H_o = -\frac{4xyv}{\pi a^2(u^2+v^2)(1+v^2)^2},$$

$$H_y/H_o = 1 - \frac{2}{\pi} \left[ \arctan v + \frac{v}{u^2+v^2} + \frac{v(x^2-y^2)}{\pi a^2(u^2+v^2)(1+v^2)^2} \right],$$

$$H_z/H_o = -\frac{4ayu}{\pi a^2(u^2+v^2)(1+v^2)},$$

(6.19)

By evaluating the electric and magnetic fields on the metal screen it is straightforward to solve for the electric charge density $\sigma$ and the electric surface current density $I$ as

$$\sigma(\rho, \phi) = \varepsilon_o E_o \frac{8i}{3} ka \frac{a/\rho}{\sqrt{\rho^2/a^2 - 1}} \cos \phi$$
\[ \mathbf{I}(\rho, \phi) = \frac{H_o \mathbf{n}_\rho}{\pi^2} \left[ \arctan(\sqrt{\rho^2/a^2} - 1) + \frac{a}{\rho} \sqrt{1 - a^2/\rho^2} \right] \cos \phi - (6.20) \]

\[ \frac{H_o \mathbf{n}_\phi}{\pi^2} \left[ \arctan(\sqrt{\rho^2/a^2} - 1) + \frac{1 + a^2/\rho^2}{\sqrt{\rho^2/a^2 - 1}} \right] \sin \phi . \]

Here, a point on the metal screen is defined by the polar coordinates \((\rho, \phi)\) and \(\mathbf{n}_\rho, \mathbf{n}_\phi\) are the radial and azimuthal unit vectors, respectively. It is important to notice that the current density is independent of the parameter \(ka\) indicating that it is equal to the magnetostatic current for which \(\nabla \cdot \mathbf{I} = 0\). On the other hand, the charge density is proportional to \(ka\) and therefore cannot be derived from electrostatic considerations.

At the edge of the aperture \((\rho = a)\) the component of the current normal to the edge vanishes whereas the tangential component of the current and the charge density becomes infinitely large.

The fields determined above are only valid in the vicinity of the aperture, i.e. within a distance \(R \ll a\). To derive expressions for the fields at larger distance one can calculate the spatial spectrum of the fields in the aperture plane and then use the angular spectrum representation to propagate the fields [27]. However, as shown in problem ?? this approach does not correctly reproduce the far fields because the near-field is only correct up to order \(ka\) whereas the far field requires orders up to \((ka)^3\). Bouwkamp calculates the fields in the aperture up to order \((ka)^5\) [28]. These fields are sufficiently accurate to be used in an angular spectrum representation valid from near-field to farfield.

Bethe and Bouwkamp show that the farfield of a small aperture is equivalent to the farfield of a radiating magnetic dipole located in the aperture and with axis along the negative \(y\)-direction, i.e. opposite to the magnetic field vector of the incident plane wave. The magnetic dipole moment \(\mathbf{m}\) turns out to be

\[ \mathbf{m} = -\frac{8}{3} a_o^3 H_o . \]

It scales with the third power of \(a_o\) indicating that the aperture behaves like a three dimensional polarizable object.

**Plane wave at arbitrary incidence**

Bouwkamp derives the fields for a small disk irradiated by a plane wave with arbitrary incidence [28]. Using Babinet’s principle it is straightforward to translate the solution to the case of an aperture. It turns out that the farfield is no longer equivalent to the radiation of a magnetic dipole alone. Instead, the electric field induces also an electric dipole oriented perpendicular to the plane of the aperture and antiparallel to the driving field component. Thus, the farfield of a small aperture irradiated by an
arbitrary plane wave is given by the radiation of an electric dipole and a magnetic
dipole with the following moments [23]

\[ \mathbf{\mu} = -\frac{4}{3} \varepsilon_0 a_o^3 [\mathbf{E}_o \cdot \mathbf{n}_z] \mathbf{n}_z, \quad \mathbf{m} = -\frac{8}{3} a_o^3 [\mathbf{n}_z \times (\mathbf{E}_o \times \mathbf{n}_z)] , \quad (6.22) \]

with \( \mathbf{n}_z \) being the unit vector normal to the plane of the aperture pointing in direction
of propagation.

**Bethe-Bouwkamp theory applied to aperture probes**

Fig. 6.16 compares the near-fields behind the aperture probe and the ideal aperture. The fields look very similar at first glance but there are significant differences. The field of the ideal aperture is singular at the edges in the plane of polarization and zero along the \( y \)-axis outside the aperture. This is not the case for an aperture probe with metal coating of finite conductivity. The Bouwkamp approximation further shows higher confinement of the fields and much higher field gradients which would lead if they were real, for instance, to larger forces exerted on particles next to the aperture. Notice, that the infinitely conducting and infinitely thin screen as used in the Bethe-Bouwkamp theory is a strong idealization. At optical frequencies, the best metals have skin-depths of \( 6 - 10 \) nm which will enlarge the effective aperture size and smooth out the singular fields at the edges. Furthermore, any realistic metal screen will have a thickness of at least \( \lambda/4 \). The exciting field of the aperture is therefore given by the waveguide mode in the hole and not by a plane wave.

![Figure 6.16](image-url)

Figure 6.16: Comparison between Bouwkamp’s solution (left) and the fields in front of an aperture probe with aluminum coating (\( \lambda = 488 \) nm) calculated by the MMP method (right). Contours of constant \( |\mathbf{E}|^2 \) (factor of 2 between adjacent lines). The incident polarization is along the \( x \)-axis.
An ideal aperture radiates as a coherent superposition of a magnetic and an electric dipole [23]. In the case of an ideal aperture illuminated by a plane wave at normal incidence the electric dipole is not excited. However, the fields in the aperture of a realistic probe are determined by the exciting waveguide mode. A metal coating with finite conductivity gives always rise to an exciting electric field with a net forward component in the plane of the aperture. One therefore might think that a vertical dipole moment must be introduced. However, since such a combination of dipoles leads to an asymmetric farfield it is not a suitable approximation. Also, the magnetic dipole alone gives no satisfactory correspondence with the radiation of the aperture SNOM probe. Obermüller et al. propose an electric and a magnetic dipole which both lie in the plane of the aperture and which are perpendicular to each other [29]. This configuration fulfills the symmetry requirements for the farfield radiation and good agreement with experimental measurements.

Figure 6.17: Contours of constant $|\mathbf{E}|^2$ on three perpendicular planes near the foremost end of an aperture probe (factor of 2 between successive lines). The arrows indicate the time averaged Poynting vector. The incident polarization is in the plane $y=0$. The transmission through the probe is increased when a dielectric substrate ($\varepsilon = 2.25$) is brought close.
6.3.3 Near-fields of Aperture Probes

Fig. 6.17 shows the fields in the aperture region of an aperture probe in vacuum and above a dielectric substrate. The coating is tapered towards the aperture and the final thickness is 70\text{nm}. The aperture diameter is chosen to be 50\text{nm}. In the plane of polarization (y=0) a field enhancement at the edges of the coating is observed which is due to the large field components perpendicular to the boundaries and the high curvature of the geometry (lightning rod effect). In the plane perpendicular to the plane of polarization (x=0) the electric field is always parallel to the boundaries leading to continuous contour lines.

Part of the field penetrates the edges of the aperture into the metal thereby increasing the effective width of the aperture. When a dielectric substrate is approached towards the aperture the power transmission through the probe increases. This can be seen in Fig. 6.17 by comparing the contour lines in the probe. Part of the emitted field is scattered around the probe and couples to external surface modes propagating backwards along the coating surface.

External surface modes can also be excited in the forward direction by the field transmitted from the core through the coating. In analogy to cylindrical waveguides they have almost no attenuation [20]. Most of the energy associated with these modes propagates therefore towards the aperture plane. If the coating is chosen to be too thin it may happen that the light from the surface of the coating is stronger than the light emitted by the aperture. In this case the field is strongly enhanced at the outer edges of the coating leading to the field pattern shown in Fig. 6.18c. To avoid such an unfavorable situation a sufficiently thick coating has to be chosen. A tapered

Figure 6.18: Contours of constant $|\mathbf{E}|^2$ (factor of $3^{1/2}$ between successive lines) in the aperture planes of three aperture probes with different coating thicknesses. Left: Infinite coating. Middle: Finite coating, the field is dominated by the flux emitted by the aperture. Right: Finite coating, the field is dominated by the flux from the outside coating surface.
coating could be a reasonable way to reduce the coating thickness near the aperture. It has to be emphasized, that surface modes cannot be excited by illumination from outside since they possess propagation constants which are larger than the propagation constant of free propagating light similar to surface plasmons (see chapter ??).

The Bethe-Bouwkamp theory has been used by various authors to approximate the near-field of aperture probes. Single-molecule experiments have shown a good qualitative agreement [30] and are the perfect tool to analyze the field distribution of a given aperture.

### 6.3.4 Enhancement of transmission and directionality

Ebbesen and co-workers have demonstrated that the transmission through a metal screen with subwavelength-sized holes can be drastically increased if a periodic arrangement of holes is used [31]. The effect originates from the constructive interference of scattered fields at the irradiated surface of the metal screen and thus depends strongly on the excitation wavelength. The periodic arrangement of holes increases the energy density on the surface of the metal screen through the creation of standing surface waves. However, the enhanced transmission relies on an illumination area that is much larger than that of a diffraction limited spot.

The enhanced transmission in a periodically perforated metal screen was first ascribed to the creation and interference of surface plasmons until it was pointed out that the same effect persists in an ideal metal that does not support any surface modes. The debate was relieved by realizing that a periodically perforated ideal metal acts as an effective medium supporting surface modes that 'mimick' surface plasmons encountered on noble metal surfaces [32]. Thus, even though an ideal metal cannot support any 'bound' surface modes, it is the periodic arrangement of holes which helps the ideal metal to act as a noble metal. Within the effective medium framework, Pendry and co-workers derived the following dispersion relation for a perforated metal screen [32]

\[
k_{\parallel}(\omega) = \frac{\omega}{c} \sqrt{1 + \frac{64a^4}{\pi^4d^4} \frac{\omega^2}{\omega_{pl}^2 - \omega^2}}.
\]

(6.23)

Here, \(k_{\parallel}\) represents the propagation constant along the surface of the perforated metal screen, \(c\) is the vacuum speed of light, \(a\) is the hole diameter, and \(d\) is the hole spacing. The plasma frequency \(\omega_{pl}\) of the effective medium is defined as

\[
\omega_{pl} = \frac{\pi c}{a\sqrt{\varepsilon\mu}}
\]

(6.24)

with \(\varepsilon\) and \(\mu\) being the material constants of the material filling the holes. Eq. (6.23) is similar to the familiar dispersion relation of surface plasmons supported by a Drude metal (see Chapter ??). However, while for a Drude metal the plasmon resonance
\( k_{||} \to \infty \) occurs at a lower frequency than the plasma frequency, the plasmon resonance for the perforated metal screen is identical with the plasma frequency \( \omega_{pl} \). The interesting outcome is that it is possible to simulate real surface plasmons by a perforated metal screen and that the dispersion relation can be tailored by the hole size and the hole periodicity. Notice that the periodicity of the holes implies a periodicity of \( 2\pi/d \) in the dispersion relation similar to the theory of photonic crystals or the electronic theory of semiconductors. This property is not reflected in Eq. (6.23) and it implies that it is impossible to reach the surface plasmon resonance.

In similar experiments, Lezec and co-workers have used a single aperture with a concentric microfabricated grating to delocalize the radiation in the near-zone of the aperture [33]. This delocalization leads to either an increased transmission or improved directionality of the emitted radiation. To better understand this effect, we note that the theory of Bethe and Bouwkamp predicts that the light emerging from a small irradiated aperture propagates in all directions. The smaller the aperture the stronger the divergence of radiation will be. A significant portion of the electromagnetic energy does not propagate and stays 'attached' to the back-surface of the aperture. This energy never reaches a distant observer (c.f. Fig. 6.19a). With the help of a concentric grating, Lezec and co-workers convert the non-propagating near-field into propagating fields which can be seen by a distant observer (c.f. Fig. 6.19b). Because the grating at the exit plane artificially increases the radiating area it also destroys the light confinement in the near-field which is not suitable for applications in near-field optical microscopy. However, light throughput can be strongly increased by placing the grating on the front-side of the aperture.

![Figure 6.19: Improving the directionality of light emission by a grating fabricated on the exit side of a small aperture. (a) Without the grating radiation diffracts into all directions. (b) The grating delocalizes the near-field and converts it into directional radiation.](image-url)
6.4 Fabrication of aperture probes

In order to create aperture probes [34] in the laboratory, the transparent tapered structure that forms the basis of the optical probe has to be coated with a nontransparent layer everywhere but at the tip apex. Usually metals are chosen due to their strong reflection for visible wavelengths. Among all metals, aluminum has the smallest skin depth in the visible spectrum. Fig. 6.20 shows the transmission and reflection of various metal thin films as a function of the film thickness. It is easy to see from

![Graph showing transmission and reflection of thin films as a function of film thickness for various metals.](image)

Figure 6.20: Transmission and reflection of thin films as a function of the film thickness for various metals. (a) Transmission vs. film thickness. Measurements were performed at a wavelength of 550±5 nm for Ag, Au, Cu, Ga, In, Mn, Pd, Al, Co, Cr, Fe, Pt, Ti, and Sb and at a wavelength of 503±5 nm for Ni, Pb, Sn and using white light for Bi and Te. The films were thermally evaporated at a pressure of 1×10⁻⁵ Torr at a rate of ≈100 nm/min and then tempered at a few hundred degrees Celsius in vacuum. From [35] without permission.
these plots that aluminum (Al) shows the best performance. Coating of dielectric tips with aluminum can be done e.g. by thermal evaporation, electron-beam (e-beam) assisted evaporation or by sputtering. Thermal and e-beam evaporation have the advantage of being a directed process. Certain areas of a sample can be excluded from being coated by exploiting shadowing effects. Sputtering, on the other hand is an isotropic process. All surfaces even of complex bodies will be coated at the same time. The formation of apertures at the apex of fiber tips can be accomplished by exploiting the shadowing effect supported by thermal and e-beam evaporation. In this process, the tips are positioned and oriented such that the stream of metal vapor hits the tip under an angle slightly from behind. At the same time the tips are rotating. The deposition rate of metal at the tip apex is much smaller than on the sides which leads to the self-aligned formation of an aperture at the apex as is illustrated in Fig. 6.21. Evaporation and sputtering suffer from the tendency of aluminum to form rather large grains. These grains have a typical size of about 0.1-1µm and can be observed when imaging coated tip structures using a focused-ion beam apparatus. Fig. 6.22 (a) shows an optical probe coated with aluminum. The enhanced visibility of grains in the focused ion beam microscope is caused by ion-channelling effects in grain boundaries (see e.g. [36]). The grain formation in aluminum films is unfavorable for two reasons: (i) leakage of light at grain boundaries and related imperfections can occur, which interferes with the weak wanted emission at the apex. (ii) The optical apertures are rather ill-defined since the aperture size is usually smaller than the average grain size. Grains also prevent the actual optical aperture from approaching close to the sample because of protruding grains. This can strongly degrade the resolution that can be achieved with a given aperture probe even if the aperture seems to be very small by inspection with the SEM. The latter effect is illustrated in Fig. 6.22 (b) and (c). E-beam evaporation produces often smoother aluminum coatings compared

Figure 6.21: Self-aligned formation of an aperture by thermal evaporation. The evaporation takes place under an angle slightly from behind while the tip is rotating. Adapted from [16].
with thermal evaporation.

The small amount of light that is emitted by a near-field aperture is a limiting factor experiments. Therefore one is tempted to just increase the input power at the fiber far end. However, aperture probes can be destroyed by too strong illumination. This happens because of the pronounced energy dissipation in the metal coating which, as a consequence, is strongly heated. Temperature measurements along a taper of aluminum coated fiber probes have been performed (see e.g. [37]) that showed that the strongest heating occurs far away from the tip in the upper part of the taper. Here temperatures of several hundred degrees Celsius can be reached for input powers up to 10 mW. For larger input powers usually the aluminum coating breaks down leading to a strong increase of light emission from the structure. Breakdown usually happens either by straightforward melting of the aluminum layer or by fracture and a subsequent rolling up of the metal sheets due to internal stress. This is illustrated by the Fig. 6.23 which shows a tip that was exposed to high energy light pulses [19]. Using additional adhesion layers or multilayer coatings were shown to improve the destruction threshold by up to a factor of two [19]. It should be pointed out, however, that the far-field transmission of an aperture probe does not take into account the enhanced near field close to the aperture. With this in mind, a low far-field transmission might still provide enough energy density at the aperture to perform certain tasks, such as polymerization of a photo resist, or excitation of single emitters.

Figure 6.22: Grains and apertures in aluminum coated optical probes. (a) Image of a aluminum-coated optical probe recorded in a focused ion beam apparatus. The enhanced visibility of grains is caused by ion-channelling effects in grain boundaries (see e.g. [36]). The aperture is well defined because the apex was cut off by the focused ion beam. From Twente webpage without permission. (b),(c) Scanning electron microscope image of a pristine aperture with large grains. From [16] without permission. Scale bars are 300 nm.
6.4. Aperture formation by focused ion beam milling

The availability of high-resolution focused ion beams open new possibilities for micromachining with nanometer-scale resolution [38]. Current focused ion beam (FIB) machines operate with liquid metal sources. To ensure a constant supply of ions for the beam, a tungsten coil with a tip [38] is wetted with Gallium or Indium which is then field ionized and accelerated. Using conventional electromagnetic lenses as in SEM such an ion beam can be focused down to a diameter of 10 nm. At an ion flux of $\sim 11$ pA at 30 kV, aluminum can be locally removed. The ablated material can be chemically analyzed using mass spectrometry [38]. At much lower ion flux (1 pA), or with an electron beam, the micromachined structure can be inspected with nearly negligible material ablation.

The standard procedure of probe processing by FIB is to cut conventional aluminum-coated probes by slicing them perpendicular to the optical axis [39]. Depending on where the cut is performed, either an existing aperture can be smoothed and improved by removing protruding grains or a closed tip can be opened to any desired aperture radius. An example of the result of such micromachining is shown in Fig. 6.22 (a). FIB-treated probes show superior performance since no grains prevent the probe from coming very close to the sample. This is a prerequisite to exploit the full confinement of the optical near-field. Also the field enhancement in the optical near field that strongly decays with increasing gapwidth can be exploited to a much larger extent using smooth probes. This is of major importance when high instantaneous intensities are needed at the sample as is the case, e.g. for imaging single fluorescent molecules [39] which were used to record the field distribution at the apex of such probes (see chapter ??). It was found that the optical near-field distribution could be recorded reproducibly and that it very much resembles the fields of a Bethe/Bouwkamp aperture [39]. For conventional non-smoothed apertures such patterns were observed very

Figure 6.23: Destruction of an aperture probe by excessive input of light. From [19] without permission.
rarely, maybe only once, e.g. in 1993 by Betzig & Chichester [40] and could not be reproduced before the advent of FIB treated optical probes. One challenge that is encountered when using FIB milled apertures is the adjustment of the aperture plane parallel to the sample surface. Typically, the lateral size of the probe is up to 1\( \mu m \) and, to ensure high resolution, it has to be placed as close as 5-10 nm to the sample surface.

It can be expected that the use of FIB techniques in near-field optics will be further extended as the next generation of FIB machines becomes available to a larger number of researchers. Micromachining of prototype structures at the apex of tips that are more complex than simple apertures can lead to improved probe structures with very high field confinement and strong enhancement (see section [?]).

### 6.4.2 Electrochemical opening and closing of apertures

FIB is a fascinating and simple possibility to micromachine structures at length scales suitable for near-field optics. However, it is a rather expensive technique. Significantly less expensive procedures have been put forth for the reliable fabrication of aperture probes. Here, we discuss two alternative electrochemical processes.

Electrochemistry is usually performed in liquid environments and this poses a problem in applications of micromachining. In presence of a liquid, in general large areas are wetted and nanometer scale material processing can not be achieved. However, there exist solid electrolytes that show significant transport of metal ions in the solid phase. Such electrolytes were used to perform controlled all solid-state electrolysis (CASSE). A prominent electrolyte is amorphous silver metaphosphateiodide (AgPO3:AgI), chosen from a variety of known solid electrolytes [41] for its high ionic conductivity, optical transparency, and ease of fabrication [42]. The aperture forma-

![Figure 6.24: Aperture at the apex of an optical probe created using the CASSE technique [42]. Note the small diameter of the aperture (dark region in the center) and the nicely smooth endface. Image courtesy of J. Toquant and D.W. Pohl.](image)
tion is induced by approaching a fully silver-covered tapered transparent tip towards the solid electrolyte. A voltage (∼100 mV) is applied between the tip and a thin silver electrode beneath the electrolyte. The tip usually has to be approached beyond the point of shearforce contact in order to achieve a current flow. Once a current is established it is kept constant via a feedback loop while the shearforce feedback is switched off. An additional feedback loop is used to terminate the process as soon as the light emission from the probe reaches a predefined value. Fig. 6.24 shows the result of such an experiment.

Another electrochemical method that is actually a light induced corrosion process was introduced in Ref. [43]. In this approach, an aperture is produced in the metal layer at the probe apex by a simple, one-step, low-power, laser-thermal oxidation process in water. The apex of a tip is locally heated due to the absorption of light from an evanescent field created by total internal reflection at a glass/water interface. Due to the heating, the passivation layer that normally covers aluminum is dissolved in an aqueous environment. The loading force acting on the probe has to be set high enough to ensure contact between the tip and the glass substrate during the complete corrosion process. Fig. 6.25 shows a typical result obtained for a laser intensity of 2.5 mW/µm² at the interface and an incidence angle of ∼62°. The aperture is formed within the first 5 s of tip exposure to the evanescent field.

Figure 6.25: Aperture formation by laser-thermal oxidation. SEM image of an aluminum coated AFM cantilever whose tip apex was exposed for 10 s to a 488 nm laser beam at 2.5 mW/µm². The silicon nitride tip can be seen protruding from the otherwise flat end face of the tip. Adapted from [43] without permission.
6.4.3 Aperture punching

Aperture punching, or in other words, the opening of a small aperture at the apex of a completely metal coated dielectric tip by plastic deformation of the metal near the apex, was the method that was used by the pioneers of near-field optics to produce apertures of small size and high flatness [44]. This method was later adapted by other groups [45, 12], because it has distinct advantages: (i) The aperture is created in situ, i.e. an initially opaque tip is mounted to the microscope and is opened up by inducing a slight contact to the sample. If the sample surface is flat, then the rim of the aperture will be flat as well and, equally important, completely parallel to the sample. The minimum gapwidth that can be achieved by approaching the tip to the sample is therefore very small allowing for high-resolution optical imaging. (ii) Similar as for the CASSE method, the aperture size can be controlled by monitoring the far-field intensity recorded from the apex region during pressing. Fig. 6.26 shows the results of pounding an etched optical fiber sputtered with 200 nm of gold. A circular aperture with a flat rim can be observed. An example for a punched aperture in aluminum can be found e.g. in [12].

6.4.4 Microfabricated probes

Because the production of individual probes is tedious and not always easily reproducible in different labs it would be much more desirable to fabricate standardized probes in large batches, e.g. using established silicon micromachining techniques. This would yield large amounts of probes with equal properties, like aperture size and shape and thus also transmission. There have been several ideas and attempts to produce such probes based on standard AFM cantilever technology. A clear problem in such a concept is the delivery of light to the actual optical probe. It seems a good idea to integrate a waveguide into the cantilever [46]. This, however, complicates the overall

Figure 6.26: Scanning electron micrographs of (a) a side view and (b) an overhead view of an aperture with a diameter of 100 nm produced by aperture punching. Adapted from [45] without permission.
design of such a lever and adds additional problems. As a consequence, most developments deal with the microfabrication of aperture tips only. Such tips can then be bonded to fibers or they can be integrated into a cantilever.

Fig. 6.27 summarizes some work that is aimed at the fabrication of hybrid probes. They combine the advantages of fibers in delivering light from a remote location with the reproducibility of microfabrication.

In a study of a prototype probe, Krogmeier and Dunn have modified commercial cantilevers by FIB micromachining [50]. They have attached a high-refractive-index glass sphere to a standard AFM cantilever (see Fig. 6.28, left panel). This glass sphere was then shaped into a pyramidal tip with controllable opening angle by focused ion beam milling (see Fig. 6.28, right panel). After coating the whole structure with aluminum, an aperture with controlled size was opened also by FIB milling. This type of work is a good example for the strength of FIB milling to produce unique prototype structures in nano-optics. The use of a high refractive index material and a large opening angle pushes the mode cutoff towards the tip and thus increases the

---

Figure 6.27: Microfabricated probes based on optical fibers. (a) Microfabricated photoplastic probe attached to the end of a single mode fiber. From [47] without permission. (b) Hybrid optical fiber probe. From [48] without permission. (c) Another example adapted from [49] without permission.
transmission of light through the tip (c.f. section 6.3.1).

Batch fabrication of cantilever-based optical probes was realized by Eckert et al. \cite{51}. They succeeded in fabricating quartz tips on silicon cantilevers that were coated with aluminum. The use of high-index quartz material in the tip increases the transmission compared to a previous hollow pyramid designs \cite{52}. Interestingly the probes were transmissive even though the aluminum layer covered the tip completely. Fig. 6.29 shows an SEM image of the whole structure. The inset is a TEM close-up of the completely covered probe tip apex. Despite the total coverage, Eckert et al. were able to observe light emission from this tip. An optical resolution of $\sim 30$ nm was demonstrated by using single molecules as probes. The recorded patterns hint to a field enhancement effect.

The cantilevered probes discussed so far can be classified as ‘passive’ probes because they influence the propagation of light but not its generation. An ‘active’ probe is one that directly converts near-field optical intensity into an electrical signal or is itself a subwavelength source of light driven by an electric current. In the following we discuss two exemplary approaches that were used to realize active detection probes. To combine the high lateral resolution of AFM with near-field optical measurements, the use of microfabricated piezoresistive cantilevers as miniaturized photosensitive elements was proposed. The silicon-based sensors consist of a $p$-doped piezoresistive path, which also includes the tip. The resistance of the piezoresistive path can be changed either by pressure on the lever or by light. For combined optical and topographical measurements, an evanescent field above a suitable sample should be created by TIR. Because the AFM tip is the only part of the cantilever that is exposed to the evanescent field, the tip can be used as a near-field optical probe. In \cite{53} it was shown that it is possible to extract the exponential decay of the evanescent field from combined force/optical measurements. To decouple optical and topographical information, the intensity of the evanescent field is modulated and the optical signal is measured by lock-in techniques.

Figure 6.28: AFM cantilever modified by focused ion beam milling. For details see text. From \cite{50} without permission.
Another approach relies on the fact that silicon microstructures are in principle compatible with integrated electronic functionality. Standard n-doped silicon cantilevers, e.g., can be equipped with a Schottky diode (a semiconductor-metal junction) at the tip apex by evaporating a thin layer of metal as indicated in Fig. 6.30 [54]. Such probes are well suited to detect optical near-fields if the scattering background is kept low e.g. by evanescent illumination.

In order to create an active light emitting probe, a laser source can be directly integrated behind an optical probe. Fig. 6.31 shows two structures where this has been realized. In (a), a diode laser is covered with a multilayer metal coating in which a small whole is drilled by FIB milling. The structure is intended for use in an optical recording head [55]. In later work, a vertical cavity surface emitting laser (VCSEL) emitting at 980 nm was integrated on to a Gallium-Arsenide (GaAs) cantilever [56].

An important goal in near-field optical microscopy will be reached when spatial resolutions of \( \approx 10 \) nm can be reached on a routine bases. 10 nm is the length-scale of quantum confinement in semiconductor nanostructures and the size of proteins in biological membranes. However, as discussed before, the poor power transmission through aperture probes is a limiting factor for high resolution near-field optical imaging. The aperture diameter cannot be arbitrarily reduced because a minimum light level has to be guaranteed in order to keep the signal-to-noise ratio above a reasonable level. The problem cannot be overcome by increasing the input power arbitrarily because of thermal heating of the metal coating. Therefore, it is necessary to explore alternative ways to achieve nanometer-scale light confinement. For example, from waveguide theory it is known that a waveguide consisting of two parallel, isolated, metallic structures has no cutoff. The transmission through the aperture probe could therefore be increased by cutting the metal coating along two lines in direction of the probe axis. The field becomes then mainly localized near the resulting slits. In order

Figure 6.29: SEM image of a silicon cantilever with integrated aluminum-coated quartz tip. The inset shows a TEM image of the tip. The tip is completely covered with 60 nm of aluminum, still it is transmissive for light. From [51] without permission.
to have only one near-field source, the end of the probe has to be obliquely cut such that only one of the two slits forms the foremost part of the probe. Another probe that has been proposed for overcoming the low throughput problem is the coaxial probe consisting of two concentric isolated metal structures or the bow-tie antenna [57]. An overview on alternative probe structures is given in the following sections.

6.5 Optical antennas: tips, scatterers, and bow-ties

In essence, the design of an optical near-field probe is a classical antenna problem. In case of a receiver, electromagnetic energy has to be channelled to the near-field zone of the antenna. Vice versa, the energy has to be released from the near-field zone if the antenna is operated as a sender. An antenna is a device that establishes efficient coupling between the near field and the far field by use of impedance matching. Although antenna theory has been developed for the radio-frequency and the microwave range of the electromagnetic spectrum it holds great promise for inspiring new concepts in the optical frequency range [58]. Field enhancement is a natural phenomenon in antenna theory. It occurs because an antenna concentrates electromagnetic energy into a tight space thereby generating a zone of high energy density. In the context of near-field optics one would like to use this property to create a highly confined light source. A simple type of antenna, tough not necessarily efficient in view of impedance matching, is a pointed tip acting as a lightning-rod antenna. It is encountered, for example, on roofs for attracting lightnings. In near-field optics, a sharply pointed, laser-irradiated metal tip proved to be a powerful near-field probe.

Figure 6.30: Top and side view of an n-doped silicon cantilever with tip. Two successive metal evaporation processes create a Schottky diode at the tip apex. From [54] without permission.
6.5.1 Solid metal tips

Near-field optical microscopy based on local field-enhancement has been proposed already in 1985, even before the invention of atomic force microscopy [59]. Since then various related implementations have been demonstrated most of them using a sharp vibrating tip to locally scatter the near-field at the sample surface. Homodyne or heterodyne detection using lock-in techniques is commonly applied to discriminate the small scattered signal from the tip apex against the background from a diffraction limited illumination area.

It has been shown that under certain conditions a scattering object can also act as a local light source [59, 60]. As discussed before, this light source is established by the field enhancement effect which has similar origins than the lightning-rod effect in electrostatics. Thus, instead of using an object to scatter the sample's near-field, the object is used to provide a local near-field excitation source to record a local spectroscopic response of the sample. This approach enables simultaneous spectral and sub-diffraction spatial measurements, but it depends sensitively on the magnitude of the field enhancement factor [61]. The latter is a function of wavelength,

Figure 6.31: Small-aperture laser diode. From [55] without permission. VCSEL integrated into a GaAs cantilever. From [56] without permission.
material, geometry and polarization of the exciting light field. Although theoretical investigations have led to an inconsistent spread of values for the field enhancement factor, these studies are consistent with respect to polarization conditions and local field distributions.

Fig. 6.32 shows the field distribution near a sharp gold tip in water irradiated by two different monochromatic plane-wave excitations. The wavelength of the illuminating light is \( \lambda = 810 \text{ nm} \). The dielectric constants of tip and water were taken to be \( \varepsilon = -24.9 + 1.57i \) and \( \varepsilon = 1.77 \), respectively. In Fig. 6.32a, a plane-wave is incident from the bottom with the polarization perpendicular to the tip axis, whereas in Fig. 6.32b the tip is illuminated from the side with the polarization parallel to the tip axis. A striking difference is seen for the two different polarizations: in Fig. 6.32 (b), the intensity near the tip end is strongly increased over the illuminating intensity, whereas no enhancement beneath the tip exists in Fig. 6.32 (a). This result suggests that it is crucial to have a large component of the excitation field along the axial direction to obtain a high field enhancement. Calculations of platinum and tungsten tips show lower enhancements, whereas the field beneath a dielectric tip is reduced compared to the excitation field (c.f. Section 6.2).

Fig. 6.33 shows the induced surface charge density for the two situations shown in Fig. 6.32. The incident light drives the free electrons in the metal along the direction

Figure 6.32: Near-field of a gold tip (5nm tip radius) in water illuminated by two different monochromatic waves at \( \lambda = 810 \text{ nm} \). Direction and polarization of the incident wave is indicated by the \( \mathbf{k} \) and \( \mathbf{E} \) vectors. The figures show contours of \( E^2 \) (factor of 2 between successive lines). The field in b) is almost rotationally symmetric in the vicinity of the tip.
of polarization. While the charge density is zero inside the metal at any instant of time \( \nabla \cdot \mathbf{E} = 0 \), charges accumulate on the surface of the metal. When the incident polarization is perpendicular to the tip axis [Fig. 6.32a], diametrically opposed points on the tip surface have opposite charges. As a consequence, the foremost end of the tip remains uncharged. On the other hand, when the incident polarization is parallel to the tip axis [Fig. 6.32b], the induced surface charge density is rotationally symmetric and has the highest amplitude at the end of the tip. In both cases the surface charges form an oscillating standing wave (surface plasmons) with wavelengths shorter than the wavelength of the illuminating light indicating that it is essential to include retardation in the analysis.

The magnitude of the field enhancement factor is crucial for imaging applications. The direct illumination of the sample surface gives rise to a far-field background signal. If we consider an optical interaction that is based on a n-th order nonlinear process and assume that only the sample surface is active, then the far-field background will be proportional to

\[
S_{ff} \sim A I_o^n \text{,} \quad (6.25)
\]

where \( A \) is the illuminated surface area and \( I_o \) is the laser intensity. The signal that we wish to detect and investigate (near-field signal) is excited by the enhanced field at the tip. If we designate the enhancement factor for the electric field intensity \( (E^2) \)

Figure 6.33: Induced surface charge density corresponding to Fig. 6.32 (a) (left) and Fig. 6.32 (b) (right). The surface charges form a standing wave in each case. In (a) the surface charge wave has a node at the end of the tip, whereas in (b) there is a large surface charge accumulation at the foremost part, responsible for the field enhancement.
by $f$ then the near-field signal of interest is proportional to

$$S_{nf} \sim a \left( f I_o \right)^n,$$

(6.26)

where $a$ is a reduced area given by the tip size. If we require that the signal be stronger than the background ($S_{nf}/S_{ff} > 1$) and use realistic numbers for the areas [$a = (10nm)^2$, $A = (500nm)^2$] then we find that an enhancement factor of

$$f > \sqrt{2500}$$

(6.27)

is required. For a first order process ($n = 1$) such as scattering or fluorescence an enhancement factor of 3 to 4 orders of magnitude is required which is beyond the calculated values. Therefore, it is necessary to involve higher order nonlinear processes. For a second order nonlinear process the required enhancement factor is only 50. This is the reason why the first experiments have been performed with two-photon excitation [60]. To maximize the field enhancement various alternative probe shapes and materials have been proposed. It has been determined that finite sized elongated shapes exhibit very low radiation damping and therefore provide very high enhancement factors [62, 63]. Even stronger enhancement is found for tetrahedral shapes [60].

It is found that no matter what the magnitude of the enhancement factor is, the field in the vicinity of a sharp tip can be quite accurately described by the fields of an effective dipole $\mathbf{p}(\omega)$ located at the center of the tip apex (c.f. Fig. ??) and with the magnitude

$$\mathbf{p}(\omega) = \begin{bmatrix} \alpha_\perp & 0 & 0 \\ 0 & \alpha_\perp & 0 \\ 0 & 0 & \alpha_\parallel \end{bmatrix} \mathbf{E}_o(\omega)$$

(6.28)

where the $z$-axis coincides with the tip axis. $\mathbf{E}_o$ is the exciting electric field in the absence of the tip. $\alpha_\perp$ and $\alpha_\parallel$ denote the transverse and longitudinal polarizabilities defined by

$$\alpha_\perp(\omega) = 4\pi \varepsilon \omega^2 \frac{\varepsilon(\omega) - 1}{\varepsilon(\omega) + 2}$$

(6.29)

and

$$\alpha_\parallel(\omega) = 2\pi \varepsilon \omega^2 f_e(\omega),$$

(6.30)

respectively. Here, $\varepsilon$ denotes the bulk dielectric constant of the tip, $r_o$ the tip radius, and $f_e$ the complex field enhancement factor. For a wavelength of $\lambda = 830$nm, a gold tip with $\varepsilon = -24.9 + 1.57i$ and a tip radius of $r_o = 10nm$, numerical calculations based on the MMP method lead to $f_e = -7.8 + 17.1i$. While $\alpha_\perp$ is identical to the polarizability of a small sphere, $\alpha_\parallel$ arises from the requirement that the magnitude of the field produced by $\mathbf{p}(\omega)$ at the surface of the tip is equal to the computationally determined field which we set equal to $f_e\mathbf{E}_o$. The value for $f_e$ was derived by using an excitation localized at the end of the tip. Therefore, diffraction effects along the
tip shaft are insignificant. Once the tip dipole is determined, the electric field $E$ in the vicinity of the tip is calculated as

$$E(r, \omega) = E_0(r, \omega) + \frac{1}{\varepsilon_0} \frac{\omega^2}{c^2} \mathbf{G}(r, r_o, \omega)p(\omega),$$  \hspace{1cm} (6.31)

where $r_o$ specifies the origin of $p$ and $\mathbf{G}$ is the dyadic Green’s function.

In fluorescence studies, the enhanced field is used to locally excite the sample under investigation to a higher electronic state or band. Image formation is based on the subsequent fluorescence emission. However, the fluorescence can be quenched by the presence of the probe, i.e. the excitation energy can be transferred to the probe and be dissipated through various channels into heat [64] (c.f. Problem ??). Thus, there is a competition between field enhancement and fluorescence quenching. The nonradiative energy transfer rate from molecule to the tip depends on the inverse sixth power of the distance $d$ between tip and molecule, similar to the case of Förster energy transfer (see chapter ??). However, the excitation rate of the molecule also depends on $d^{-6}$ since it is proportional to the square of the electric near-field. Hence,

Figure 6.34: Comparison of the near-fields of a metal tip and a metal sphere. (a,b) Excitation with an on-axis, focused ($NA = 1.4$) Gaussian beam. (c,d) Excitation with an on-axis, focused Hermite-Gaussian $(1,0)$ beam. The strong field enhancement in (c) is due to the longitudinal field of the excitation beam. The cross-sections are evaluated in a plane $1\text{ nm}$ beneath the tip. The results indicate that the field distribution near the tip is well approximated by the dipole fields of a small sphere. However, the field strength for longitudinal excitation (c) is much stronger compared with the field strength of an irradiated sphere (d). While in (a,b) the fields are in-phase, they are $155^\circ$ out of phase in (c,d).
for small distances from the tip there should be no distance dependence of the fluorescence rate of a single molecule. Of course, this is a rough estimate and more accurate calculations are needed to understand the interplay between field enhancement and quenching.

It has been shown that metal tips are a source of second-harmonic radiation and of broadband luminescence if excited with ultrashort laser pulses. The local second harmonic generation has been used as a localized photon source for near-field absorption studies \[65\]. Whereas second-harmonic generation is an instantaneous effect, the lifetime of the tip’s broadband luminescence has been measured to be shorter than \(4\,\text{ps}\) \[66\].

**Fabrication of solid metal tips**

Fabrication procedures for sharp metal tips have been mainly established in the context of field ion microscopy \[67\] and scanning tunnelling microscopy (STM) (see e.g. \[68\]). The actual geometrical shape of the tip is not so important for applications in STM on flat samples as long as there is a foremost atom and there is sufficient conductivity along the tip shaft. On the other hand, in optical applications one also cares about the tip’s *mesoscopic* structure, i.e. its roughness, cone angle, radius of curvature, and crystallinity. Not all etching techniques yield tips of sufficient ”optical” quality. Therefore, focused ion beam milling can be an alternative to produce very well-defined tips \[69\].

In electrochemical etching, a metal wire is dipped into the etching solution and a voltage is applied between wire and a counter-electrode immersed into the solution. The surface tension of the solution forms a meniscus around the wire. Etching proceeds most rapidly at the meniscus. After the wire is etched through, the immersed lower portion of the wire drops down into the supporting vessel. By this time, a tip

![Figure 6.35: Schematic of an ac etching circuit for gold tips. The etching voltage is automatically switched off after drop-off. The circuit works also for other tip materials if HCl is replaced with a suitable etching solution. See text for details.](image)
has been formed at both ends, at the rigidly supported upper portion of the wire and the lower portion which dropped down. By the time of drop-off, the upper tip is still in contact with the solution because of meniscus formation. Therefore, if the etching voltage is not switched off immediately after drop-off, etching will proceed on the upper tip and the sharpness of the tip will be affected. Hence, it is crucial to switch off the etching voltage as soon as drop-off has occurred.

Various electronic schemes have been introduced to control the drop-off event. Most of them use \textit{dc} etching voltages. However, it is observed that for certain materials \textit{dc} etching produces relatively rough tip surfaces. Especially for gold and silver, \textit{ac} etching is favorable. A schematic for the fabrication of sharp gold tips is shown in Fig. 6.35. A function generator provides a periodic voltage overlayed with a certain offset. The voltage is sent through an analog switch and applied to a gold wire that is vertically dipped into a solution of hydrochloric acid (HCl) and centered into a circular counter-electrode (Pt) placed just below the surface of the solution. The counter-electrode, held on virtual ground, directs the etching current to a current-to-voltage converter. The resulting voltage is averaged by an \textit{rms}-converter and then compared with an adjustable threshold voltage by means of a comparator. At the beginning of the etching process, the diameter of the wire and thus the etching current are at a maximum. With ongoing time, the diameter of the wire and the current decrease. The diameter of the wire decreases more rapidly at the meniscus giving rise to tip formation. When the diameter at the meniscus becomes small enough, the lower portion of the tip drops off and the etching current decreases abruptly. Consequently, the \textit{rms}-voltage at the input of the comparator drops below the preset voltage threshold and the output of the comparator opens the analog switch thereby interrupting the etching process. Because of the \textit{rms} conversion, the circuit cannot respond faster than the time of $2\times 10^{-10}$ periods of the waveform provided by the function generator. It turns out that the speed of the circuit is not the limiting factor.

Figure 6.36: (a) Gold tip etched from a gold wire according to the method described in the text. Radius of curvature at the tip is about 10 nm.
factor for achieving good tip quality. The waveform, threshold voltage, concentration of $HCl$, depth of counter-electrode, and length of wire are factors which are much more important. These factors vary from setup to setup and have to be determined empirically. With a good set of parameters one can achieve tip diameters of less than 20nm with a yield of 50%. An example of such a tip is shown in Fig. 6.36. This particular tip has a radius of curvature of about 10 nm.

It has to be stressed, that the fabricated tips are not mono-crystalline, i.e. the metal atoms do not have a periodic arrangement throughout the tip volume. Instead, the tip consists of an arrangement of crystalline grains with sometimes varying lattice configurations. The origin of this grain formation lies in the fabrication process of the original metal wire and is known since the early days of field ion microscopy. Because of grain formation it is only a rough approximation to describe the tip’s electromagnetic properties by a macroscopic dielectric function $\varepsilon(\omega)$. In fact, it is commonly observed that the field enhancement factor is much weaker than predicted by calculations and that it shows high variability from tip to tip. This observation is likely to be related to the grain-structure of the tips. A quantitative comparison of theory and experiment and the assessment of nonlocal effects demands the development of single-crystal metal tips.

To reduce the background signal associated with exposure of the sample to the irradiating laser beam, Frey et al. pioneered the so-called tip-on-aperture (TOA) probe. In this approach, a mini tip is directly grown on the end face of an aperture probe. The fabrication principle is sketched in Fig. 6.37. The end face of a completed

![Figure 6.37: Tip on aperture geometry: (a), (b) Sketch of the fabrication process including mini tip formation and subsequent metal coating. (c),(d) SEM images of the resulting structure corresponding to the sketches in (a) and (b). (e) shows the fluorescence response of a fluorescent bead under a mini tip. The strong vertical confinement is indicative of a high lateral resolution. Adapted from [70] without permission.](image)
aperture probe is exposed to a focused electron beam in a standard scanning electron microscope (SEM). The electron beam gives rise to growth of a carbon tip at the location of exposure [Fig. 6.37 (a),(c)]. After the growth of this 'contamination-tip', the probe is metallized by thin layers of chromium and aluminum by evaporation under an angle as sketched in Fig. 6.37 (b). This results in a closing of the aperture apart from a slit [Fig. 6.37 (d)] that originates from shadowing by the tip. The mini tip can be excited through the narrow slit by simply coupling light into the other fiber end. Fig. 6.37 (e) shows the fluorescence response of such a probe using a fluorescent bead as test object. The strong confinement in z-direction holds promise for very high resolution near-field optical imaging. The TOA approach is also favorable from a perspective of tip alignment and drift. An externally irradiated metal tip has to be kept within the irradiated area and long-term drift requires readjustments. It can be expected that the TOA configuration will become more widely used in future near-field applications.

### 6.5.2 Particle-plasmon probes

The dynamics of a free electron gas in a finite-sized geometry is characterized by distinct modes known as surface plasmon resonances (see chapter ?? for more details). These resonances are accompanied by enhanced electromagnetic fields. The explicit application of surface plasmons in the context of near-field optical microscopy has been put forth by different groups. Among the various schemes is the original proposal by

![Figure 6.38: Particle plasmon probe. (a) A polystyrene bead on a flat glass substrate is covered with a 20 nm gold layer and illuminated in Kretschmann configuration. The scattering of the protrusion is recorded as a sample is approached from the other side. (b) Recorded scattering intensity versus particle-surface distance for both p- and s-polarization. (c) Image recorded in constant height mode using electron tunnelling feedback. Adapted from [71] without permission.](image-url)
Wessel [59] and the field-enhancing metal tips discussed in the previous chapter.

An elegant demonstration of the principle of a plasmon probe was the experiment by Fischer and Pohl in 1989 [71]. It is schematically shown in Fig. 6.38 (a). A 20 nm thick gold film covers polystyrene beads that are adsorbed on a gold coated glass substrate. Kretschmann-type illumination is used (see chapter ??) to launch surface plasmons on the gold film. The surface plasmon scattering from a selected protrusion [indicated in Fig. 6.38 (a)] is recorded as a function of the distance between the scatterer and an approaching glass surface [Fig. 6.38 (c)]. The main effect of the distance variation is that the mean dielectric constant of the environment can be tuned, which leads to a shift of the resonance condition for the particle plasmon supported by the protrusion. A peak is observed for p-polarized excitation and for small separations which is indicative for a surface plasmon resonance. The peak is absent for s-polarization which reinforces the surfaces plasmon interpretation. It is evident that the existence of the resonance peak can be used for near-field optical imaging in reflection, i.e. backscattered light is strongly sensitive to local dielectric variations near the protrusion. Fig. 6.38 shows that the technique is able to resolve metal patches on glass with high resolution. A similar approach was adapted later to image magnetic domains on opaque materials [72]. Also, the gold coated polystyrene particles, later called nanoshells, found applications in diverse sensing applications as demonstrated in the work of Halas et al. [73].

Another example of a plasmon tip is presented in Fig. 6.39 (a)-(c). A sharpened fiber, created by Ohtsu’s etching procedure (see Fig. 6.4), is overcoated with a ∼30 nm gold layer. It is then dipped into an evanescent wave created by a p-polarized unfocused beam of a tunable dye laser (see Fig. 6.39d) at a glass-air interface. The gap between tip and glass surface is adjusted to ∼5 nm. The excess of light that is picked

![Figure 6.39: Plasmon fiber probe. (a)-(c) SEM images of the optical probe.
(d) Scanning tunneling optical microscopy setup used in the study. (e) Wavelength dependence of the light picked up by the plasmon probe. The inset shows a comparison to a bare fiber probe. Adapted from [74] without permission.](image-url)
up by the plasmon probe as compared to a bare probe is plotted in Fig. 6.39(e) as a function of the excitation wavelength. The peak at about 590 nm is attributed to the excitation of a particle plasmon at the tip apex. A strong dependence of the resonance on the gap is also reported [74].

In Ref. [74], the shape of the metal particle at the tip apex that supports the plasmon is not well defined which hinders a quantitative comparison with theoretical predictions. A more controlled approach is the attachment of a well-defined spherical or elliptical metal nanoparticle at the apex of a dielectric tip or an aperture probe. In the latter case, it is desirable that the particle is positioned into the center of the aperture in order to minimize the coupling with the metal coating at the rims. Both ideas have been realized. Fig. 6.40 (a) shows an SEM image of a chemically grafted 60 nm gold particle inside the aperture of an aperture probe. The spectral response of such a structure, i.e. the ratio of light transmission with and without particle shows a peak probably caused by the excitation of a particle plasmon which results in enhanced emission from the aperture/particle system [75]. Fig. 6.40 (b) shows a gold particle that was chemically grafted onto a purely dielectric tip along with its scattering spectrum. The spectrum can be well fitted using Mie theory for a

![Figure 6.40: Particle plasmon tips: (a) Chemically grafted gold particle (diameter ~60 nm) in the aperture (diameter ~200 nm) of an aluminum-coated silica tip. Adapted from [75] without permission. (b) Chemically grafted 100 nm gold particle at the apex of a dielectric fiber probe. (c) Spectrum of the light scattered off the particle tip in (b). (b) and (c) adapted from [76] without permission.](image-url)
subwavelength scatterer.

6.5.3 Bowtie antenna probes

An optical antenna is a metal nano structure with characteristic dimensions matched to an integer multiple of half the wavelength of the light it is to interact with. As discussed earlier, the primary purpose of an antenna is to provide efficient coupling between far-field and near-field by means of impedance matching. The near-field zone, called the 'feed gap', is the location where the emitter or receiver resides. Antenna theory has been primarily developed for the electromagnetic radiation in the radio frequency regime. On basis of scale invariance of Maxwell’s equations one would expect that antenna concepts can simply be scaled down to the optical regime. However, material constants change dramatically between microwave and optical frequencies. While in the microwave regime metals can still be considered as ideal conductors, this assumption is no longer legitimate in the optical regime. Optical antennas have to fight with losses and deal with collective electron resonances, i.e. surface plasmons. The latter do not occur in the traditional antenna regime. The exploitation of surface plasmon resonances in the design of optical antennas holds great promise for compensation of material losses. Although the design of optical antennas is likely to be

![Figure 6.41: The bowtie antenna. (a) Geometrical outline of the antenna. (b) Experimental setup used to demonstrate the performance at microwave frequencies. (c), (d) Measured intensity distribution without and with the antenna placed in front of the waveguide. Adapted from [77] without permission.](image)
inspired by the developments in the radio-frequency or microwave regime, it can be expected that the presence of new physical phenomena will demand the exploration of new geometries and material compositions [58].

The bowtie antenna is an antenna with almost perfect impedance matching. In the context of near-field optics it was introduced in 1997 together with a microwave proof-of-principle experiment [77]. These experiments clearly demonstrate that sub-wavelength confinement of electromagnetic radiation can be achieved using a bowtie antenna structure. Furthermore, it has been shown that due to optimized impedance matching the delivery of energy to the near-zone (throughput) is very efficient. The fabrication of optical bowtie antennas is being pursued by different groups. In order to serve as a near-field optical probe, the bowtie needs to be fabricated on the sides of a sharp dielectric tip such as an AFM probe.

6.6 Conclusion

This chapter provided an overview of the types of probes used in near-field optics. We discussed the diversity of probes in terms of the variety of near-field and far-field illumination and detection schemes. Besides the theoretical background necessary to understand and to correctly apply the respective probe structures we have also discussed fabrication procedures and possible problems that might arise during applications. This chapter is not intended to be complete as the development of new probe concepts and designs is a very active field and new innovations can be expected in the years to come. Also, many more probe structures and fabrication procedures can be found in the literature. We tried, however, to pick the most important and representative work to provide a concise overview.
Problems

**Problem 6.1** Calculate the intensity distribution in a standing evanescent wave above a glass-air interface created by spatially overlapping two evanescent waves of the same intensity, polarization and antiparallel in-plane wave vector. Take a line profile perpendicular to the interference fringes and calculate the convolution with a Gaussian of a given halfwidth. How does the halfwidth influence the fringe visibility.

**Problem 6.2** Calculate the difference in transmission through an aluminum coated aperture probe and an aperture probe with an infinitely conducting coating. Assume an aperture diameter of 100 nm and a taper angle of \( \delta = 10^\circ \).

**Problem 6.3** Apply Babinet’s principle to derive the fields near an ideally conducting disk. Use Bouwkamp’s solution and state the fields in the plane of the disk.

**Problem 6.4** Calculate second harmonic generation at a laser illuminated metal tip. Assume that the fields near the tip are given by Eqs. 6.28-6.31 and that second harmonic generation at the tip originates from a local surface nonlinear polarizability \( \chi_s^{(2)} \). The nonlinear surface polarization is determined by the field \( E_n \) normal to the surface of the tip as

\[
P_n(r', 2\omega) = \chi_{nnn}^{s}(-2\omega; \omega, \omega) E_n^{(\text{vac})}(r', \omega)E_n^{(\text{vac})}(r', \omega),
\]

where the index \( n \) denotes the surface normal, \( r' \) a point on the surface of the tip and the superscript ‘\( \text{vac} \)’ indicates that the fields are evaluated on the vacuum side of the surface. The field at the second-harmonic frequency generated by \( P_s \) is calculated as

\[
E(r, 2\omega) = \frac{1}{\varepsilon_0} \frac{(2\omega)^2}{c^2} \int_{\text{surface}} \mathbf{\tilde{G}} (r, r', 2\omega) \mathbf{P}_s (r', 2\omega) \, d^3r'.
\]

Consider only the near-field of \( \mathbf{\tilde{G}} \) and assume a semi-spherical integration surface. Determine an effective tip dipole oscillating at the second-harmonic frequency.
References


54 REFERENCES


