Near field microscopy and near field optics

1994 Rep. Prog. Phys. 57 989

(http://iopscience.iop.org/0034-4885/57/10/002)

Rep. Prog. Phys. 57 (1994) 989–1028. Printed in the UK

Near field microscopy and near field optics

D Courjon and C Bainier

Laboratoire d'Optique P M Duffieux, URA 214 CNRS, Université de Franche-Comté, 25030-Besançon, France

Abstract

The early eighties have experienced a revolution in the perception of physical phenomena. This revolution is the birth of a new generation of imaging systems based on the detection of non-radiating fields. The near field optical microscope is the latest of this family. Like its prestigious brothers, the STM and the AFM, it allows one to see the physical world with new eyes. The objective of this article is to provide an overview concerning the physical mechanisms and paradoxes taking place in non-radiative detection. We will first explain the connection between Heisenberg uncertainty relations and the Rayleigh Criterion. The fundamental role of evanescent fields will be pointed out through the plane wave expansion. Finally a catalogue of the different configurations currently in use will be given and illustrated with some experimental results.

This review was received in May 1994.

0034-4885/94/100989+40$59.50

Contents

1. Introduction

The early eighties have experienced a revolution in the domain of the perception of physical phenomena. This revolution is the birth of a new generation of imaging systems based on the detection and the use of non-radiating fields.

The most famous representative of this family is the electron scanning tunnelling microscope (STM) which was invented in 1981 in the Research Center of IBM Zürich by Binnig and Rohrer [1]. The success of this new tool overshadowed the more discrete birth of the optical near field microscope (SNOM) which was proposed around 1983 at the same place [2] and at Cornell University in the USA [3].

The two techniques are fundamentally similar: the STM is based on the detection of tunnelling electrons and the SNOM detects tunnelling photons. We will see in the following that the comparison stops there. Photons have specific properties (no mass, no electric charge, large wavelength, easy polarization change, propagation in air and in many dielectric materials, etc). This specificity of optical radiation makes the near field microscope a new tool able to complete harmoniously the family of local probe based microscopes.

2. Near field detection or the art of detecting non-radiating fields

Until then photons and electrons have been used like projectiles. Emitted from a source, they are projected onto the object-target and captured after reflection by a suitable detector (the eyes of an observer or a camera). Since the trajectory and the number of reflected particles depend on the object properties, information about the object characteristics is carried away by the particle beam. Its projection onto a target bears the generic name of 'image'. It is obvious that the aim is to ensure a relationship as close as possible between the object and its image (see figure 1). This problem which

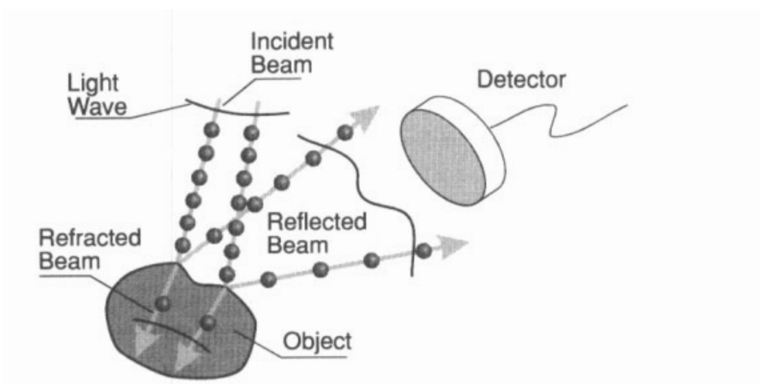


Figure 1. Schematic of the interaction between an object and a light beam. In first approximation, the light beam can be considered as projectiles launched against a target (the object) and then reflected towards the detector. This interpretation is primitive but provides the basis for the understanding of the notion of image.

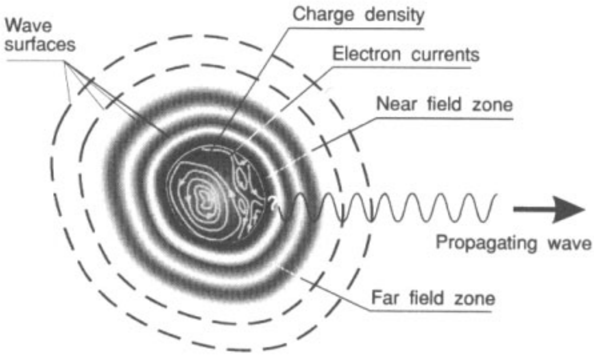


Figure 2. Field emitted from an object. The electron currents (in the case of conducting materials) and the charge densities inside the object induce an electromagnetic field radiating from the surface. Far away from the surface the field has the well known structure of propagating waves. Very close to the object (the region of the question mark), the field has a more complex structure since it is composed of propagating and non-radiating components.

is probably one of the oldest problems in physics is a real dilemma. Object and image are physically very different: the object is generally a three-dimensional bulk of matter. The image is often a two-dimensional projection of a quantity connected to the object structure. Because of the two dimensions of the recording media, this quantity is generally an intensity since detectors are only sensitive to intensities. Here is the first fundamental problem: comparing quantities belonging to different classes of physical beings is meaningless. However, we can avoid the problem by replacing the object itself by the light field associated with it. In this case, it is possible to compare the object field and the image field or, more often, the object field intensity and the corresponding intensity in the image plane. The first question to answer is: what is the relationship between object structure and light field over an object? The answer is given by Maxwell equations linking the electron currents and more generally the electron densities inside the matter to the external electromagnetic field. In figure 2 the emission of light by a self-luminous object has been represented. The oscillating charges and currents induce some electromagnetic field variations able to propagate from the object surface to the free space. By applying the principle of continuity it seems logical to admit that the spatial field distribution very close to the object will reproduce the density and current distributions on the object surface. Since the latter can vary over exceedingly small distances (in any case, smaller than the wavelength) we can suppose that the field very close to the object will vary over such small distances. This deduction seems to be in contradiction with the observation and analysis of electromagnetic fields. Indeed we know that the smallest details which can be detected are always larger than a half wavelength. This discrepancy is due to the fact that until recently, all observations, analyses and measures have been made far away from the object (at least, a few wavelengths distant). So, two field regions must be distinguished: the first one a few nanometres from the surface is called the near field zone. The second one is referred to as the far field zone. It extends from the near field zone to infinity. They are the fields which are detected in every conventional device such as microscopes, telescopes and more generally lenses. Concerning the near field, it was shown one century ago that its structure is not simple. It contains both components able to propagate and components confined over the surface and damping outside. The latter are non-homogeneous waves whose properties are intimately connected to the object material just beyond the surface. They exist because of the presence of matter and thus cannot exist in free space. They bear the name of evanescent waves, a terminology probably introduced by Newton himself when he discovered total internal reflection two centuries ago.

The previous description involving propagating and non-propagating components in the near field zone is somewhat misleading. It does not mean that it is possible to separate physically those two kinds of components. In fact the non-propagating components exist because the propagating components do and vice versa. This non-separability can be explained in a few ways, evoking for example the analyticity of the electromagnetic signal. Another pragmatic and straightforward explanation could consist of noting that the non-radiating terms, necessarily correspond to an oscillating energy confined near the surface. Because photons cannot be stored as electrons, this energy must escape from the surface leading to the propagating fields. As a consequence, if we perturb the non-radiating components, the far field will be modified. The problem of separability of field components has fed several debates in near field optics for a long time.

A typical structure of such non-propagating fields is the following one:

$$U(x, y, z, t) = A(x, y, z) \times \exp -\mathrm{j}(k_x x + k_y y) \times \exp(-\alpha z) \times \exp \mathrm{j}(\omega t) \quad (1)$$

where A is the amplitude of the field at the point (xyz), $\exp -\mathrm{j}(k_x x + k_y y)$ corresponds to the propagation term of the wave in the plane (xy); $\exp(-\alpha z)$ expresses the decrease of the field along the z-axis. The coefficient α depends upon the material properties and upon its spatial structure. It has been shown that the smaller the details, the higher the coefficient α and consequently, the stronger the field confinement over the surface. Finally, $\exp(\mathrm{j}\omega t)$ expresses the time dependence of the field. Physically the field propagates along the (xy) direction and vanishes along z, oscillating at the frequency of the light. The light beam is thus unable to propagate, it is confined over the object surface. We will refer to this as a non-radiative field in the following even though the notion of non-radiative fields is generally associated to purely static fields.

Practically, for metrologists, the existence of non-radiating components has disastrous consequences because a part of the information contained in the object which is transferred in the non-radiating field, *will not* be transmitted to the detector. What is worse is that the non-radiating components contain the information about the sub-wavelength details of the object. As an example, if the light source emits in the visible range ($\lambda = 550$ nm) the periodic features whose period is smaller than about 300 nanometres will never be imaged. This analysis keeps its validity for electrons but in this case the de Broglie wavelength is exceedingly small (between a few angströms and a few nanometres). The loss of information about the subwavelength details is not so severe. We thus understand the success of the scanning electron microscope facing the classical optical microscope.

The impossibility of getting a subwavelength resolution in optics has been known for a long time. It was expressed in terms of diffraction theory more than one century ago by Ernst Abbe, a German physicist who predicted the existence of a resolution limit. Later it was reformulated by Lord Rayleigh in the following very concise form:

$$r \geqslant \frac{1.22\lambda}{2n \sin(\theta)}. \quad (2)$$

It stipulates that two point objects will be seen separately only and only if the distance r between the two points is greater than a quantity connected to the wavelength λ of the light beam, to the index n of the medium and to θ, the semi-angle of aperture of the objective used for collecting and focusing the light beam onto the detector.

This relation is known as the Rayleigh criterion.

The study of the above inequality shows that the only ways for increasing the resolution (i.e. decreasing r) consists of:

(1) choosing very short wavelengths (UV, x radiation in the case of electromagnetic fields or more efficiently, propagating electrons);

(2) working in very high index materials for increasing n. It is the principle of the immersion microscopy which was given by Amici in the middle of the XIXth century;

(3) increasing the aperture angle of the microscope.

All these methods are well known and except for the use of electrons instead of photons where the gain in resolution is dramatic, the other solutions lead to a slight resolution improvement only. Let us notice that this criterion has been established assuming propagating waves. If we can detect non-radiating fields we can expect to circumvent the Rayleigh criterion and to break down the diffraction barrier definitely.

It is clear that this principle of detection is philosophically very new because scientists are accustomed to working with propagating waves. Most physical phenomena have been thus deduced from far field observation. It is the case of Maxwell equations (although they can predict the existence of non-radiating fields).

The necessity to detect non-accessible fields implies an upsetting of the notion of detection itself: instead of placing the collector far away from the source (or the object), the collector must be brought to a distance smaller than half a wavelength from the object. Roughly speaking, we can say that the collector captures the field before it propagates.

We will see that the consequences of such a procedure are often unexpected and lead to new unusual problems that have to be solved.

First, the collector of light must be placed and driven a nanometric distance far from the object without touching it. This needs the use of suitable actuators such as piezoelectric motors. This is not a real problem because, thanks to the development of the STM, these tools are easily available and rather cheap. Another problem specific to photons deals with the type of collector to be used. Because of the exceedingly small distance between specimen and probe, no imaging system can take place. The only solution consists of using a point-like collector able (1) to collect the light locally and (2) to convert it into electric current or to re-emit it in the free space or through a suitable light guide toward a photodiode or a photomultiplier. But since the collector has to be as small as possible for limiting the spatial integration, the signal collector cannot be (up to now) a photon–electron converter. The only way is a simple passive light collector such as the extremity of a tapered fibre.

Finally because of local detection, no image can be directly obtained. In order to generate an image-like structure, the collector must scan the object surface the same way the electron spot scans the TV screen.

3. The detection of non-radiating fields: the optical tunnelling effect

The basis of the near field microscope is now established. However, until now, we have not talked about the technique for capturing the non-radiating components of the field.

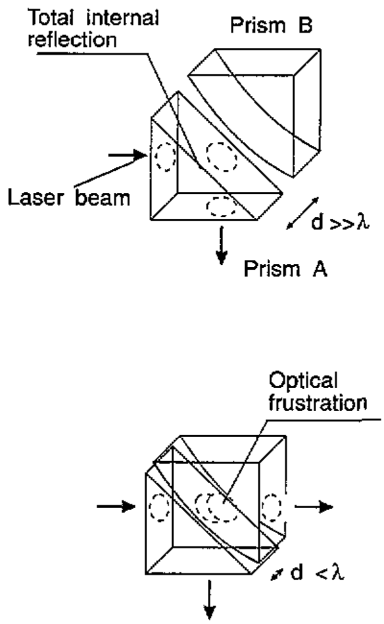


Figure 3. The famous experiment of Newton. A light beam is projected onto a prism. As expected, the beam is internally and totally reflected on the larger side of the prism. If a second prism is brought to the first one, no effect is detected unless the distance between the two prisms becomes smaller than a fraction of a micron. The light beam then seems to be captured by the second prism, frustrating the total reflection. The beam intensity transmitted through the second prism depends exponentially on the distance d.

Since the non-propagating components have the same structure as evanescent waves, the only way consists of detecting them by optical tunnelling whose principle is described below.

Let us consider the famous experiment of Isaac Newton realized three centuries ago. Interested in chromatic effect of light, he noticed that a beam could be totally reflected inside a prism for a certain incidence angle although the prism surface was not metallized. Starting from this observation, he tried to 'frustrate' the total reflection by placing a second prism in contact with the first one as shown in figure 3. In order to observe the behaviour of the field in the (xy) plane, he used a slightly convex surface prism. He thus expected to see the light passing just where the two prisms were in contact. Instead, he noted that the transmission surface between the prisms was larger than the contact region. It was for him a strange phenomenon: it seemed possible to perturb the total reflection by introducing an optical element in a region where there was no visible radiation. He suggested the light particles should be carried away by their velocity through the surface before being reattracted by the matter. The explanation was wrong but not meaningless.

Today we can explain this phenomenon in term of continuity of boundary conditions over the prism surface: since the field exists inside the prism (just under the surface) the field *must* exist outside the prism (just above the surface). Such a field propagates along the surface and must vanish in the perpendicular direction. Obviously, Newton's experiment can be generalized to devices other than prisms. Consequently, if a suitable dielectric material is immersed in an evanescent field, this one will be converted into propagating field in respect to the continuity conditions at the interface. This effect is the optical or photon tunnelling effect. It can be explained classically from Maxwell equations without invoking any quantum mechanism.

4. The existence of evanescent fields in the vicinity of objects characterized by very fine structure

To prove that a near field microscope can work, the previous demonstration is not sufficient. The frustration mechanism mentioned above implies macroscopic interactions between prism and field. In near field microscopy, the very tip of the probe can be exceedingly small (a few nanometres radius). The Newton approach involving large objects cannot be applied. We have to take into account the diffraction in the conversion process.

Various ways can be proposed for describing the interaction mechanism between a diffracting object and an evanescent field. The simplest one consists of assuming that the very tip of the probe behaves as a dipole, that is an elementary point scatterer. When this dipole is placed in the non-radiating field, it is excited and consequently generates an electromagnetic field containing both propagating and non-propagating components. Only the propagating components will be detected by a remote photon-electron converter as sketched in figure 4. This approach can be generalized to extended objects by noting that: *a light beam impinging on a limited object will be always converted into propagating and evanescent field.* The incident field can be either propagative or evanescent. This remarkable assumption is derived from a theoretical work due to Wolf and Nieto-Vesperinas [4].

To understand this theorem let us first define what a limted object is. A limited object is a material structure which presents a sharp discontinuity. In terms of spatial frequency, its associated spatial Fourier spectrum is infinite. Consequently it contains all the spatial frequencies from zero to infinity. A good example is a hole pierced in an opaque screen, a little ball, a dust particle, etc. An extended object such as the rough surface of a piece of glass can be then considered as the juxtaposition of small limited grains with sharp edges. We can thus apply the Wolf–Nieto theorem with such an object and conclude that: *a light beam impinging on an object characterized by a fine structure (with details smaller than $\lambda/2$) will be converted into propagating components able to propagate towards the detector and evanescent components confined on the surface. The first ones are associated to the low spatial frequencies of the object whereas the second ones are connected to their high frequencies.* This last point will be demonstrated further.

All the philosophy of near field microscopy is summarized in this theorem: (1) a high frequency object generates evanescent waves when illuminated either with a propagating or evanescent wave; (2) the resulting evanescent fields do not obey the Rayleigh criterion, they can exhibit strong local variations over distances much smaller than one wavelength; (3) by reciprocity, this undetectable high frequency local field can be, in its turn, converted into propagating field by means of a small limited object which will convert the evanescent field into evanescent and propagating field; (4) the latter is then directed towards a suitable remote detector. It is interesting to note that the conversion evanescent field-propagating one is linear: the detected field is proportional to the Poynting vector at a given point in the evanescent field. The propagating field will then reproduce faithfully the local and sharp variations of the evanescent

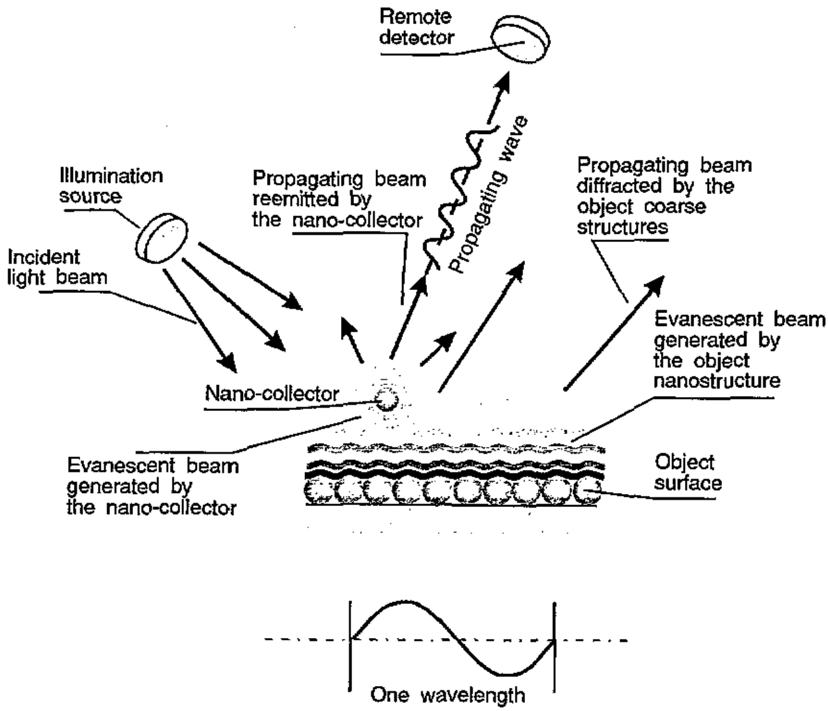


Figure 4. Sketch of near field detection. Step 1: generation of the object near field by the illumination process. A light source illuminates an object represented as composed of discrete components. These components are excited by the incident field and re-emit light. The waves associated to the reflected beam are composed of evanescent waves confined on the object surface and of propagating waves. If the periodic structures of the object are smaller than the wavelength (it is the case of the figure), the reflected field, far away from the object, does not contain any information on the fine structure of the object. Step 2: detection of the near field. For detecting the subwavelength object information, a small scattering centre (the nano-collector) is brought close enough to the object surface. The near field lying on the surface will excite the scattering centre which will re-emit light. The re-emitted light is again composed of evanescent waves (non-detectable) and propagating ones which can propagate far away to the remote detector.

field. To generate a 2D image, we will scan the surface with a small limited object (in fact the very tip of a tapered fibre) as explained previously.

Finally, we see that near field microscopy is the result of a series of conversions: conversion of the incident beam into an evanescent one by the object structure itself and conversion of the evanescent field into a propagating one by the nanocollector (see figure 4).

If we compare the principle of the Newton experiment with the detection mechanism described above, we can see the perfect correlation of the two processes. In short, diffraction as well as refraction mechanisms are responsible for the conversion of propagating fields into evanescent ones and *vice versa*. By definition such a conversion is a tunnelling effect applied to light waves. It is interesting to notice that the optical or wave tunnelling effect does not need any particle description to be understandable. It can be easily deduced from Maxwell equations and has not the magic character of the tunnelling electron which, in spite of its mass, is able to go through a material barrier.

5. The proliferation of acronyms hiding a common principle

Finally every near field detection involving conversion of evanescent fields into propagating ones belongs to the family of tunnelling microscopy. It is thus rather confusing and disconcerting for the beginner to choose between the plethora of acronyms describing near field microscopes. Not less than 5 acronyms are used presently. Depending on your location (in Europe or in the USA) you can use SNOM for scanning near field optical microscopy or NSOM by inverting two letters. If you illuminate your object by means of an incident evanescent field you will use STOM, an acronym for scanning tunnelling optical microscope (proposed by myself in 1988) or PSTM for photon scanning tunnelling microscopy, although neither particle character nor quantum mechanics are used to understand the principle. Finally if emphasis is put on evanescent wave detection you will use EFOM for evanescent field optical microscopy. Unfortunately, such discriminations are still maintained although they are artificial and without scientific basis. To soften my comment, I will admit that some technological differences could explain this strange situation. Now the best choice seems to be SNOM or NSOM (it will be difficult to separate the partisans of the two acronyms). This designation put emphasis on the notion of near field which deals with both evanescent and propagating fields. It is thus more general than tunnelling microscopy implying the detection of evanescent waves only. It is obvious that the nanocollector is only sensitive to energy and cannot discriminate evanescent waves from propagating ones.

6. Common structure of every near field microscope

The structure of the near field microscope clearly appears: it looks like a STM where the metallic exceedingly sharp needle is replaced by a specific tip able to emit or collect photons. The first function of the tip is its capability to diffract the light beam (this property is easily fulfilled). Depending on the developed technique, it can be used to transmit the collected photons to a remote detector. The image is then generated by moving the tip along the axes x and y over the surface of the object. This scanning is ensured by means of piezo ceramics driven by a simple personal computer. To limit the noise, the incident light beam emitted from a laser is often temporally modulated. A lock-in amplifier only amplifies the output signal which is modulated at the same frequency as the laser. This procedure allows one to decrease the noise dramatically. Unfortunately, it slows the detection. The collected photon flux is then directed to a photomultiplier where it is amplified and converted into electric current. The near field microscope and its illumination possibilities are described in figure 5.

Building such a microscope is simple in theory: we will see that some specific properties distinguish electron tunnelling from near field microscopy. First, photons propagate easily over large distances. Consequently, they will be easily reflected on remote defects even outside the observation field. They will thus interfere with each other leading to field variations having only a remote relationship with the object topography. The main consequence of such a property will be the difficulty, and even the impossibility, finding a distance control method of the tip position over the sample. This point will be discussed further. Finally, because the near field is composed of a complex mixture of propagating and non-propagating components, it does not decrease monotonically while removing the tip from the surface. The beautiful evanescent wave decay we

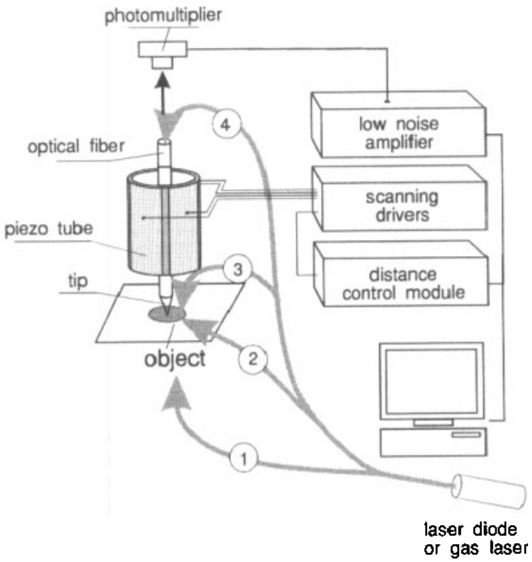


Figure 5. Scheme of the near field optical microscope. Very similar to the STM, it is mainly composed of a scanning and displacement stage, an electronic system and the nano-collector or emittor. The scanning system is usually a piezo-tube able by bending under a suitable voltage to generate the three x, y, z motions. The scanning area ranges between a few nanometres and about 100 microns (or more). The nano-detector is generally a tapered fibre whose other extremity is connected to the detector. This one can be a low noise photovoltaic cell or a photomultiplier. The illumination can be realized in transmission in collection mode (1), in transmission in illumination mode (4), in internal reflection in collection mode (2), in external reflection in collection mode (3) or even in collection and illumination mode (the fibre illuminates the sample and collects the reflected light). The electronic devices are composed of low noise amplifiers (for amplifying the very weak optical signals coming from the PM) of scanning drivers allowing us to explore and scan the surface and finally of a distance module control, the role of which is to keep constant the distance between tip and sample (for example).

observed on a perfectly flat surface is only a particular case, which is rarely met in analysing complex structures.

Among the differences with electron microscopy we can mention the possibility of focusing the photons (photons are bosons and can be bunched and directed easily). The direction of polarization can be modified at will. Finally, the interaction light–matter is generally soft and the object after analysing is generally not modified (except actinic materials such as photoresists).

The sources used are generally gas lasers or diode lasers, the wavelength ranging from UV to IR. The main experiments until now have been carried out in the visible range. Let us notice that nothing prohibits us from using spatially non-coherent sources. However because of propagation from the source to the sample, the light field on the object will be necessarily partially coherent (the light coming from the stars is coherent even though a star is a highly incoherent source). We can thus expect the photons to be strongly correlated over distances smaller than a wavelength that is over distances of the same order as the scanned field. Except for spectroscopy, the interest of using incoherent sources like filament lamps is thus limited. Finally let us notice that some attempts are made to use x-rays generated from a synchrotron source.

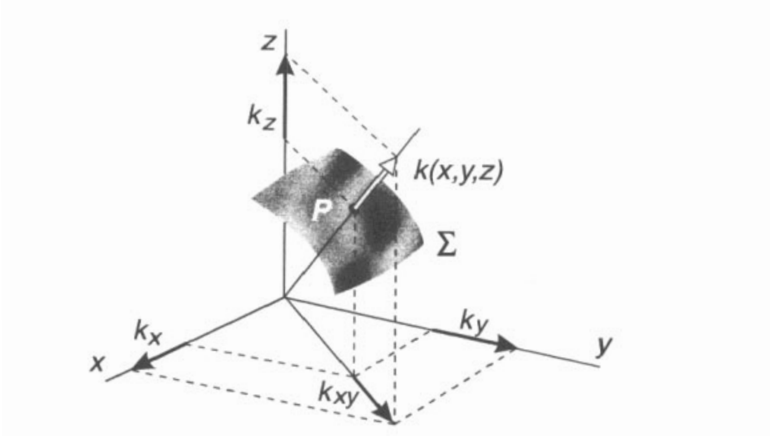


Figure 6. Illustration of the Heisenberg principle. The diagram represents a piece of field exhibiting amplitude variations (in grey). At a given point P we can determine the field charactristics which reduce to the amplitude at the point P and the propagation vector passing by P. From the Heisenberg principle, we determine the smallest field variation we can expect.

7. Heisenberg principle and near field detection

If such a technological approach does not suffer any controversy, we can wonder about the relation between near field super-resolution and the Heisenberg uncertainty principle.

Let us then consider a given field distribution $U(xyz)$ at a point $P(xyz)$ [5]. It can be seen as the image of a remote object (see figure 6).

This field can be described by its amplitude at this point and by a certain vector $k(k_x, k_y, k_z)$ called the propagation vector, originated at the point P and pointing in the direction of propagation. Its strength is simply connected to the wavelength of the light in the vacuum by the relation

$$|k| = 2\pi n/\lambda \qquad (3)$$

where n is the index of the medium inside which the wave propagates.

Because the field encodes the object information, it varies locally so that two contiguous points P_1 and P_2 do not usually have the same intensity. The question raising then is: what is the smallest distance separating two discernible points? In other words: what is the highest resolution we can expect in analysing the light field without any technological limitation due to the optical system? To answer, let us call Δx, Δy, Δz the uncertainty on the measure of the position of the point P and Δk_x, Δk_y, Δk_z the uncertainty of the measure of the propagation vector of the photon. If we limit our analysis to the x direction (it holds for y obviously), the Heisenberg principle stipulates that [5, 6]

$$\Delta x \times \Delta k_x \geqslant 1 \Rightarrow 1/\Delta k_x = 1/2k_x \qquad (4)$$

by noting that the range of variation of k_x is $\Delta k_x = 2k_{x\mathrm{Max}}$ where $k_{x\mathrm{Max}}$ is the largest value that k_x can take.

From this relation we see that Δx can only take very small values when k_x takes large values. Physically it means that if a field varies rapidly (Δx small) the light is strongly scattered (Δk_x gets large). this consequence is well known in the optics world: a strongly perturbed field is always strongly scattered and diffracted.

Now the right question is what is the largest value we can expect for the k_x component?

Here is the point of divergence between Heisenberg principle and Abbe theory. From a classical point of view, k_x, k_y and k_z are real quantities and k_x is merely the projection of the $\boldsymbol{k}$ vector on the x axis. It can be written:

$$k_x = |k| \sin(\theta) \tag{5}$$

where θ is the projection angle of $\boldsymbol{k}$ on the x axis. k_x is thus always smaller than the modulus of the $\boldsymbol{k}$ vector (see figure 6).

Relation (4) can then be rewritten:

$$\Delta x \geqslant \lambda/2n \sin(\theta) \geqslant \lambda/2n \tag{6}$$

that is simply the Abbe limit (in the case of one-dimensional approach).

Because n can never be larger than 2 or 3, the smallest values of Δx are a fraction of λ. Now let us assume that the k-components can take complex values. In this case, the only condition to fulfil for k_x is

$$|k_x| = \sqrt{|k|^2 - (k_z)^2 + (k_y)^2} \tag{7}$$

this equation simply expresses that the modulus of a vector is equal to the sum of the squares of its components.

If k_z and k_y can take complex values, $|k_x|$ is no longer limited to $|k|$. It can take large values if and only if one of the two other components are imaginary. This result is the key of subwavelength resolution in optics. The most classical case of this situation is when k_x and k_y are real and k_z is complex. We find again the definition of evanescent fields described by equation (1).

Heisenberg uncertainty relations confirm that an electromagnetic field can vary over distances much smaller than one wavelength with the only restriction that such a field will be characterized by a bidimensional $\boldsymbol{k}$ vector in the real space and will be bound to the concerned object. The Heisenberg principle can be seen as a generalization of the Rayleigh criterion. The latter only concerns propagating fields whereas the Heisenberg uncertainty principle does not make any restriction about the kind of field were dealing with.

Finally, three conclusions can be deduced from this analysis. (1) Nanometre variations can exist in a light beam. (2) They can only be observed by detecting non-radiating fields. (3) As a consequence, nanometric light sources can be imagined and built. However such sources will necessarily be partially non-radiative. The latter point is particularly important for metrologists having in mind to use nanosources in data writing (such as optical microlithography). The temptation to force the light beam through small holes for generating propagating collimated beam is a beautiful dream but unfortunately, not realistic.

8. An example of localized subwavelength source: the dipole

A well known example of such nano-sources is the dipole whose size is much smaller than a wavelength.

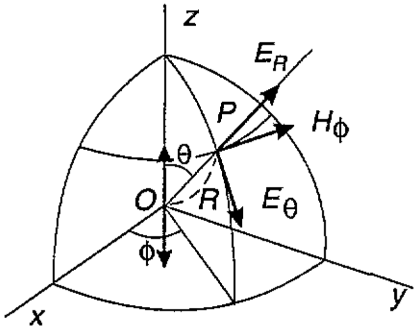


Figure 7. Field emission of a dipole located in O and oscillating in the z direction. The field at a given point P reduces to the three components E_R, E_θ and H_ϕ. The double arrow in grey indicates the dipole oscillation direction (along the z axis).

Let us then consider figure 7. A dipole is located at the point O and oscillates in the z direction. We plan to determine the electric field at the point P situated at a distance R from the point O. If $[p]$ is an oscillating retarded function of time describing the field propagation at the point P and if $[p']$ and $[p'']$ are its first derivatives, the components of the fields E and H can be written [5]:

$$E_R = 2\{[p]/R^3 + [p']/cR^2\}\cos(\theta)$$
$$E_\theta = \{[p]/R^3 + [p']/cR^2 + [p'']/c^2R\}\sin(\theta) \tag{8}$$
$$H\phi = \{[p']/cR^2 + [p'']/c^2R\}\sin(\theta).$$

From equations (8) it appears that the structure of the field varies strongly when the point P comes near the surface, that is for values of R such as:

$$R \ll cp/p' \qquad R \ll cp'/p''. \tag{9}$$

Assuming that p can be written:

$$[p(t)] = p_0\cos(\omega t - R/c) \tag{10}$$

where ω is the light pulsation, inequalities (9) lead to the condition

$$R \ll \lambda/2\pi \tag{11}$$

where λ is the wavelength associated to the field.

As an example, let us assume a dipole emitting at a wavelength of 632 nanometres. For detecting the near field, R must be much smaller than 100 nanometres.

Now in order to point out the non-radiative characteristics of the near field, let us determine the amount of energy crossing a spherical surface surrounding the point O. It is simply given by the Poynting vector over this surface. In the usual case of far field detection ($R \gg \lambda/2\pi$), equations (8) reduces to:

$$E_R = 0 \qquad E_\theta = (p''/c^2R)\sin(\theta) \qquad H_\phi = (p''/c^2R)\sin(\theta) = E_\theta. \tag{12}$$

The magnitude of the Poynting vector is then:

$$S = (c/4\pi)|E_\theta H_\phi| = (p''^2/4\pi c^3R^2)\sin^2(\theta). \tag{13}$$

After integrating over the surface, the amount of energy crossing it is:

$$W=\int S\sin(\theta)\,2\pi R^2\,d\theta=p_0^2\omega^4/3c^3=\text{constant}. \quad (14)$$

This non-zero quantity is constant and consequently does not depend on R. Since the energy crossing any sphere surrounding the dipole is constant whatever the distance of detection, the energy radiates through the spherical surface. The field is thus purely radiative. Now what happens in the near zone? Although the previous approach can be generalized by taking into account all the terms in the field equations, it is simpler to notice that the radiative energy is necessarily the same in the near field and in the far field (the energy can neither appear nor disappear when going out from the dipole). It means that the terms we have neglected in the far field approximation do not carry any energy. They are only terms that oscillate at the light frequency and vanish in inverse power of R^2 and R^3 when moving away from the dipole.

Remark. We could be surprised to see that the previous formalism does not introduce explicitly the evanescent waves. It is obviously possible to reformalize the dipole emission in terms of propagating and evanescent waves. We have chosen the classical approach because it is probably simpler to understand.

A question subsists: if the dipole emission can be described without taking into account the near field non-radiating terms, why not omit them? In fact, their presence ensures the confinement of the field emitted by the point-like dipole. If the non-radiating terms are neglected, the dipole will be perceived as a small sphere of about $\lambda/2$ diameter (more exactly a diffraction pattern known as an Airy pattern after projection onto a screen). This is precisely what happens when we observe a dipole far away from the dipole centre. Such a result is in perfect agreement with the Rayleigh criterion and confirms that no purely radiative nano-source can exist. A subwavelength source is always partially non-radiative.

9. Generalization to more complicated objects

The analysis can be easily extended to more complicated objects by assuming that extended objects are composed of a combination of elementary dipoles. In this case, instead of generalizing the previous analysis by integrating over all the space containing the dipoles, a simpler way consists of developing a continuous and macroscopic approach derived from the Fourier techniques.

9.1. Plane wave expansion technique

Any wave can be described by its amplitude $U(x,y)_{z=z_P}$ in a given plane ($z=z_P$). Unfortunately, such a description does not inform us about the propagation of the wave. However if this field distribution is Fourier analysed, each Fourier component can be identified as a plane waves travelling in a given direction. This description is strictly equivalent to the direct space analysis, but it is more suitable for describing the propagation of diffracted beams in the whole space.

We thus propose to analyse the perturbation of the field going through a given object, by using a Fourier analysis method called the plane wave expansion method.

This method consists of expanding any field distribution into elementary plane waves, each of them being characterized by a diffraction direction (i.e. its direction cosines α, β, γ) and an amplitude. As an example, if the incident beam is a plane wave falling normally onto the surface of a flat grating, the plane wave expansion of the transmitted field will be composed of a set of elementary discrete plane waves corresponding to the diffraction directions. The interest of such an expansion is that it is rigourous and suffers no artefact. It is in fact a simple mathematical tool for describing a field distribution differently.

Following Goodman [7], it is then possible to express the angular distribution of light in the plane ($z=z_P$) in terms of the angular distribution in the plane ($z=0$).

A given plane wave at the distance z from the object can be then expressed in the form

$$A(\alpha/\lambda, \beta/\lambda, z)=A_0(\alpha/\lambda, \beta/\lambda)\exp[\mathrm{J}(2\pi/\lambda)\sqrt{1-\alpha^2-\beta^2}z] \tag{15}$$

where $A_0(\alpha/\lambda, \beta/\lambda)$ is the angular spectrum of the field $U(x,y)_{z=0}$.

Since α, β, γ, are the components of a unit vector associated to the propagation direction of the plane wave, they always verify the following equality.

$$\alpha^2+\beta^2+\gamma^2=1. \tag{16}$$

Relation (15) expresses that every angular field distribution at a given distance from an object or a source emitting light can be described by the product of a term connected to the angular spectrum of the light distribution in the object plane and a phase term due to the propagation from the plane $z=0$ to the plane $z=z_P$.

Equation (16) expresses the same normalization condition as equation (7) in the paragraph dealing with the Heisenberg uncertainty principle.

As an example, let us then consider the field transmitted through a grating characterized by the spatial frequencies u, v. The direction cosines of the field (α, β, γ), are Fourier conjugates of the space coordinates of the field. They are thus connected to the spatial frequences u, v of the grating by the relations

$$\alpha=\lambda u \qquad \beta=\lambda v \qquad \gamma=\sqrt{1-\lambda^2(u^2+v^2)} \tag{17}$$

We can consider two cases: (1) the grating period is larger than the wavelength. It is the classical case of utilization of diffraction gratings; (2) the grating period is smaller than the wavelength.

(1) the grating period is larger than the wavelength
In this case

$$(u^2+v^2)<1/\lambda^2 \qquad \Rightarrow\alpha^2+\beta^2<1 \qquad \gamma^2>0 \tag{18}$$

leading to $\alpha\in R$, $\beta\in R$, $\gamma\in R$ where R is the real space.
Since α, β, γ are real quantities, relation (15) takes the form

$$A(\alpha/\lambda, \beta/\lambda, z)=A_0(\alpha/\lambda, \beta/\lambda)\exp \mathrm{j}\varphi(z) \tag{19}$$

where $\varphi(z)$ is a phase delay depending on z. It is due to the propagation from the object plane to the detection one. The waves are simply diffracted by the grating and propagate away.

(2) the grating period is smaller than the wavelength
In this case, the normalization condition becomes

$$(u^2+v^2)>1/\lambda^2 \qquad \Rightarrow \alpha^2+\beta^2>1 \qquad \gamma^2<0 \tag{20}$$

leading to $\alpha \in R$, $\beta \in R$, $\gamma \in C$ where C is the complex space.
In this case, the three direction cosines of the propagation vector are no longer real. The angular field structure can be rewritten.

$$A(\alpha/\lambda, \beta/\lambda, z)=A_0(\alpha/\lambda, \beta/\lambda)\exp[-(2\pi/\lambda)z\sqrt{\alpha^2+\beta^2-1}] \tag{21}$$

where the exponential argument is no longer imaginary but real. When z increases, the exponential term becomes negligible as well as $A(\alpha/\lambda, \beta/\lambda, z)$. Relation (21) is simply the angular spectrum of an evanescent wave. Close to the object plane it tends towards $A(\alpha/\lambda, \beta/\lambda)$, far away it tends towards zero. The resulting field does not propagate although in the object vicinity its structure is not that of a conventional plane wave.

As an example, assuming that $\lambda=632$ nm, $u=5000$ lines/mm and $v=0$ (the grating lines are parallel to the y direction). In this situation, the amplitude of the evanescent wave will drop by a factor of 2 for an observation plane located at $z=23$ nm from the grating surface. For a grating period $u=10\,000$ lines per mm, z reduces to 11 nm. It consequently appears that the damping of the evanescent field becomes proportional to the spatial frequency of the grating for periods much smaller than the wavelength.

As a comparison, let us recall that the evanescent wave generated by total internal reflection (Newton's experiment) depends on the incident angle and on the index of the prism. For an incidence angle of nearly 90 degrees and a material index of 2, the distance z corresponding to an amplitude damping of 2 is $z=64$ nm. Even for these extreme conditions, the damping of the evanescent field is weak.

Here is one of the fundamental differences between the properties of evanescent waves generated by refraction or by diffraction. In the latter case the confinement of the evanescent field can be only a few nanometres. The main consequence is that the detection of such evanescent waves will need to put the detector as close as possible to the object. In short, the higher the spatial frequencies to detect, the smaller the distance of exploration. This point is the key of all near field microscopies.

Now we can generalize this approach to any object, because it is well known from Fourier analysis that any object can be considered as the superimposition of a finite or infinite number of elementary amplitude gratings with given amplitude and phase delays.

By applying the theorem of superposition, we can deduce that the surface of the object will be 'recovered' by a complex association of evanescent fields propagating along the surface and characterized by different damping parameters. After a few nanometres, the field is redistributed and converted into classical propagating waves. A 10 000 lines/mm grating is seen by the remote observer as a simple piece of perfectly polished glass. The far field does not contain any information about the subwavelength features of the object.

We retrieve the previous conclusions about high and low spatial field frequencies. The fine structures of the electromagnetic field are due to local variations of its non-radiating components whereas the low frequencies deal with conventional propagating terms.

Finally, from the Heisenberg principle and Fourier analysis, we arrive at the strange conclusion that a light field can locally vary over distances much smaller than one wavelength. Unfortunately, such variations vanish at distances smaller than a wavelength. It is thus not possible to retrieve the subwavelength details of the object, encoded in the light beam, in the far field detection.

As an anecdote, Toraldo di Francia a well known physicist published in 1949 a paper in the review *Nuovo Cimento* dealing with the experimental observation of evanescent waves generated by diffraction on a grating in the microwave range [8]. It is thus a well known phenomenon in optics. The observation of these non-radiating field is recent because of the lack of available positioning devices some decades ago.

10. Nano-aperture against nano-antenna

It is now clear that the high spatial frequencies in the object are coded in the non-radiating part of the near field. The only way for detecting them will consist of bringing close to the surface a suitable component able to convert the non radiating terms into propagating ones. From simple dimensional considerations, a lens cannot be used for collecting the light. Moreover a lens obeys macroscopic optical laws. The smallest distance between object and collectors of typically a few nanometres is not compatible with the smallest focal length we are able to get presently. Another solution could be a nanometric photovoltaic converter. Although some promising attempts have been made, it is not mature yet. As explained previously, the only way at the moment, is the use of a scattering centre the size of which will be small enough to avoid integration effects. If such an element is used as nano-collector, it will convert the non-radiating components into propagating field by the optical tunnelling effect (see again figure 4). On the contrary, if it is used as nano-emittor, they are the non-radiating components of the so created nanosource which will scan the sample and that will be converted into propagating components able to propagate towards the remote detector. By reciprocity, the two configurations are theoretically equivalent, however, it is almost impossible to reproduce the same experimental conditions in the two modes. Generally the illuminated area in illumination mode is smaller than that in collection mode, because it is directly connected to the sunmicron emitting part of the tip. On the contrary, in collection mode the illumination area is defined by the aperture of the microscope objective which is used for focusing the light beam over the sample.

The nano-collector (or nano-emittor) is probably the main and crucial element in near field microscopy. Its realization is not yet standardized. The opinions are shared in the scientific community about its fabrication and even about the mechanisms of its working.

Presently, two techniques are in competition: the first one is inherited from the early works in near field microscopy [9]. It uses the concept of nano-aperture; that is the use of an exceedingly small hole pierced in an opaque screen and playing the role of a scanning element. This idea is old, it was proposed 66 years ago by Synge in a premonitory paper published in the *Philosophical Magazine* [10]. As an anecdote, it seems that Einstein himself strongly encouraged Synge to publish what was at that time a pure speculation. In this early paper, Synge suggested to scan a very thin biological sample by means of a very small hole illuminated strongly and brought very close to the sample, that is typically at a distance of about 10 nanometres. The light transmitted through the sample is then detected far away by means of photo-voltaic cell.

This speculative work is the basis of all the techniques using nano-sources or nano-collectors.

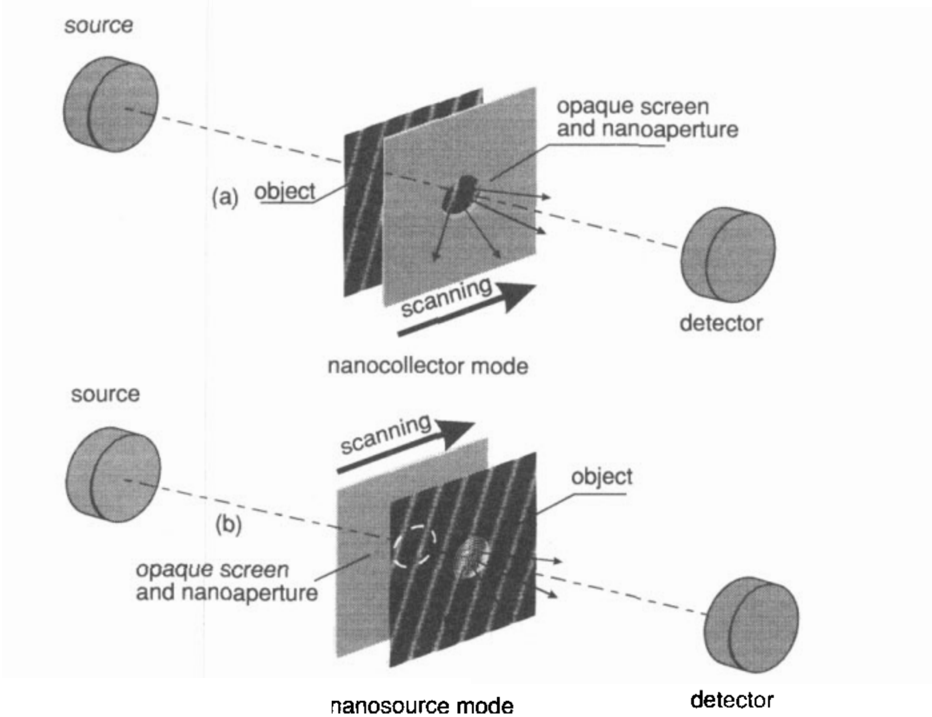


Figure 8. *Basis of near field detection.* In the first case, the source illuminates the object and a small hole in an opaque screen scans the object plane at a distance of a few nanometres (transmission collection mode). In the second case, the scanning screen is placed between the source and the object. The hole in the screen plays the role of a nano-source scanning the object surface (transmission illumination mode). In both cases the detector is set far away from the object.

Note that the pinhole used for scanning the object can be placed either before or behind the object. If it is placed before it plays the role of nano-emittor scanning the surface. If it is placed behind the object, it plays the role of a nano-collector sampling the field transmitted through the object (see figure 8).

The second technique is inherited from the works on the emission of electromagnetic waves by scattering centres and more generally by subwavelength antennas. In simple terms, a very sharp-pointed dielectric tip placed in an electromagnetic field will locally perturb the field. The result of the perturbation will be a re-emission of light that can be detected far away from the illuminated sample. Note that in this case, the tip can be purely dielectric without any metallic coating. It can even be in semi-conducting or metallic material.

Now the two ways are considered. A larger number of applications use the nano-aperture technique. In the near future the two methods will proably meet because from simple consideration such as Babinet's theorem, a small hole and its 'negative'—that is a small opaque spot—lead to similar diffraction patterns.

Nevertheless, each approach has its partisans defending their position with many good arguments, as we will discuss below.

10.1. Technique of the nano-aperture

The basic principle consists of piercing a very small hole in a sufficiently thin opaque screen. The hole being necessarily smaller than the wavelength, it is impossible to use optical mechanisms. An elegant solution was proposed by Fischer some years ago [11]. It consisted of depositing a suspension of polystyrene spheres in water on a suitable clean glass substrate. The glass was then metallized with a very thin gold or aluminium layer. The sample was then immersed in a suitable solvent bath able to dissolve the polystyrene. Because of the thinness of the coating, the metal surrounding the spheres get removed, leading to quasi perfectly circular holes whose diameter was defined by the size of the beads. Because of the metal deposition technique it was possible to ensure a thickness of less than 30 nanometres. Moreover the beads can be very well calibrated and well defined diameters from 50 nm to 1 μm are easily available.

The main drawback of the technique was the impossibility of monotoring the sphere position during the deposition step. Finally the second drawback inherent to the aperture technique was the large flat surface of the screen increasing the risk of contact between probe and object. Nevertheless the technique worked and gave the first significant near field images.

Another solution proposed by Pohl *et al* at IBM Research Division in Zürich [12] consisted of realizing a tapered quartz rod by mechanically grinding. The rough resulting surface was flattened by means of a polishing suspension directed in a thin jet onto the tip. Finally an argon laser was used for the final sharpening. The metal was classically deposited in a vacuum chamber. For generating the hole at the apex, the tip was pressed against a flat part of the sample. A deformation and a perforation of the metal layer resulted, leading to a nanosource of a few nanometres. Apart from the complexity of the fabrication of the tip, the main drawback was the non-reproducibility of the resulting tips. However, until now the results of the group of Zürich have not been surpassed in spatial resolution.

At the same time another technique inherited from the microelectrode fabrication technology was developed at the Cornell University in America by Betzig *et al* [13]. It consisted of heating and simultaneously pulling a capillary until it breaks. The result was a tapered microcapillary whose outer part at the very tip did not exceed 500 nanometres and the aperture could be as small as 100 nanometres. The nanoprobe was then metallized with aluminium leading to a perfectly circular aperture at the very tip.

The technique is easy to use, because the pulling machines have existed for a long time. Moreover, the nano aperture is well defined and perfectly circular. The main problem is the transmission of light through the tip. The strong index step between air (or vacuum) and glass leads to a tremendous loss of light during the propagation to the detector.

Among the more exotic possibilities, some attempts have been made to pierce silicon nitride tips generally used in AFM [14], by means of ion bombarding. The technique works fine; unfortunately it needs a sophisticated environment to succeed.

Now, the main technique comes from the fibre technology. The principle consists of heating and pulling a monomode fibre as proposed some years earlier [15]. Because of traction, the fibre will break at the level of the waste leading to an almost flat breaking surface. The tip is then metallized with aluminium. To avoid covering the extremity, the fibre is tilted in the vacuum chamber as indicated in figure 9, which represents the different steps of tip fabrication. The method benefited from the development of the micropipette pullers. It is easy to set by choosing suitably the initial conditions (temperature, velocity of pulling strengths exerted on the fibre) to generate any shape from simple conical to pseudoparaboloidal ones, all is possible in theory at least.

Such a nanometric device can be used either as nano-emittor or nano-collector. Its main drawback is the large surface of the extremity due to the metallic coating plus

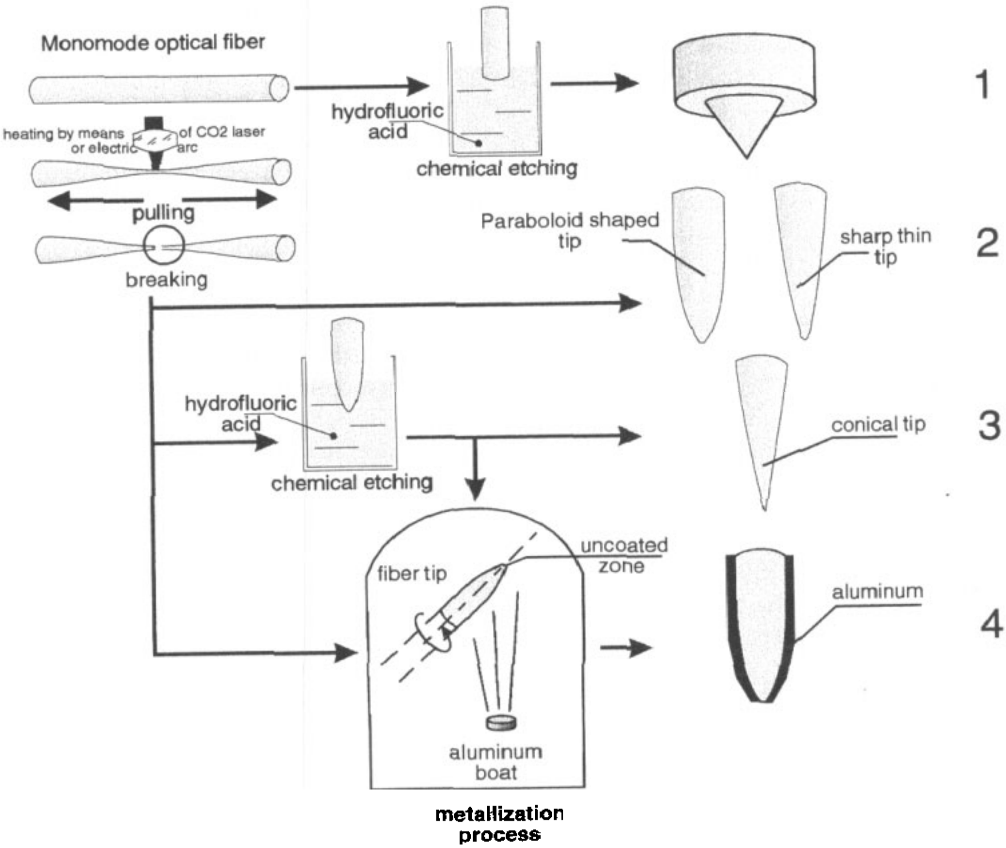


Figure 9. Fabrication of the tapered fibres. Two basic techniques are currently developed. The first one consists of chemically etching a monomode fibre in a buffered hydrofluoric bath. The second one is based on the heating and pulling of a fibre in order to get a tapered extremity. The two techniques can be combined in various ways.

the fibre cross section. In return, the polarization can be rather easily kept constant, if the tapering is weak. This situation is known as adiabatic propagation conditions.

Finally, the presence of metal can be seen as an advantage because of the role played by the magnetic field [16] in the image formation.

10.2. Technique of the nano-antenna

10.2.1. Nano-antenna used as nano-emittor or nano-collector. A second school of thought appeared in the middle of the eighties. Probably inspired by the problems of field perturbation by scattering centres, two groups simultaneously proposed to detect evanescent fields by means of single mode tapered fibres [17, 18]. The basic idea was simple: by immersing a very sharp dielectric tip into an evanescent field generated by total internal reflection, the tip apex would be excited by the evanescent field in spite of the fact that such a field cannot propagate. A small amount of propagating light would then be transmitted by the fibre to a suitable detector such as a photomultiplier.

The same idea was proposed at the same time by one of the group [19] in order to realize a reflection microscope. It consisted of injecting a light beam in a monomode fibre whose extremity was tapered. The light emitted by the apex would interact with the surface of an object and would be partially collected back by the same apex.

The technique works theoretically because it is merely the transposition in the optical domain of the antenna emission widely used in radio. Nobody expresses doubt about the possibility of detecting remote radio stations emitting in the long wavelength range with an antenna of a few centimetres. Moreover, it is also well known that the detected field varies strongly when moving the antenna on the ground. As an example, when entering a tunnel, the signal intensity decays over a few metres only. In this example, the resolution gain is better than 100 (λ/r). The difference lies in the antenna scale. It seems from experience that the shape and the nature of the material constituting the antenna is crucial. The interest of the technique is the simplicity of the making of the tip in comparison with the aperture technique.

The first step of the tip fabrication is similar to that of the nano-aperture. The only difference is the non-metallization and the search for very sharp tips. The tip is often chemically etched for reaching ultimate nanometre size. The unavoidable protrusions play probably the role of antenna itself.

Technically, several methods have been proposed in recent years. The simplest one consists of etching the tip in a hydrofluoric acid bath chemically. This way allows one to get a very sharp apex (see again figure 9 for fabrication procedure). Unfortunately, the final result strongly depends on parameters such as the temperature of the bath, the concentration of the acid or of the buffers etc. Quasi-perfect cone shaped tips have been obtained since 1990 [20, 21]. However, the images obtained with these tips are not better resolved than those obtained with other methods. Another technique is the heating and pulling as previously described. Three kinds of heating sources have been used. The simplest one derived from the micropipette pulling machines is a tungsten filament surrounding the fibre. Such a device suffers several drawbacks: the temperature is limited by the melting point of the filament, metal or oxide can deposit during the heating phase, finally the heated zone cannot be modified. The other way consists of using a CO_2 focused laser beam. The solution is versatile, the focusing point can be adjusted at will, the temperature can be modulated along the fibre by moving the focusing point. The main drawback is the use of a rather expensive and hazardous source. Finally the last solution consists of using a commercial arc fibre splicer used in a particular way: the electric arc generated between two tungsten knives heats the fibre locally and strongly. A mechanical puller is then added in order to separate the two fibre segments as fast as possible. The tapered fibre is used as is or is immersed in a hydrofluoric bath in order to reduce its size to reach nanometre range. Such a system is very simple, unfortunately for obscure reasons commercial apparatuses are very expensive and do not encourage the use of this method.

Finally, whatever the technique used, it is possible to realize very sharp apexes from a few nanometres to 100 nanometres diameter. The real problem is that we do not know until now what is the best shape for ensuring maximum conversion and collection efficiency.

10.2.2. Nano-antenna used as near field perturbing system. A probably promising way to be explored is the use of the nano-antenna for perturbing the near field on the object surface locally.

The principle consists of setting a sharp tip in the field generated by the object under test. This sharp tip will perturb the object field and the global scattering of light from the object will be affected. Because the modification of the overall scattering field will be very weak, it will be necessary to introduce suitable techniques such as spatial filtering or tip dithering for increasing the signal to noise ratio.

The main interest of such techniques is the versatility. It is no longer necessary to use transparent guiding materials for fabricating the tip. By using tungsten tips it will be possible to get an exceedingly small apex. Finally the principle is independent of the illumination technique. We will see in the following part that such a principle has been applied with success to external reflection and total reflection microscopes.

10.3. Collecting the field diffracted by the very tip

In the previous analysis we have focused our discussion on the capability of detecting the light in the near zone. However from experiments, if we assume a laser power of 1 mW entering the probe (in the case of a nano-emittor), the power density in the near zone is smaller than 1 μW dropping down to 1 nW after passing through the apex. It is thus clear that the energy waste is due to the propagation in the tapered part of the fibre. So the question to ask is: 'what shape of tip has to be chosen for ensuring an optimal transfer from the guiding part of the fibre to the tip?' Some authors propose to transpose the techniques developed in solar concentration in the domain of near field. It has been shown that some specific shapes can limit dramatically the energy losses when focusing the sun rays [22]. This approach is interesting and although it is not the only way for increasing the resolution, it shows that some simple shapes such as the well-known conical one are accompanied by a tremendous loss of light. Parabola shaped tips seem more efficient. Such shapes are rather easy to realize by setting the parameters of the puller (temperature, exerted force, velocity of traction, etc).

11. Scanning technique of the probe

Let us recall that the distance control between tip and sample is a key parameter in local probe microscopy. One of the techniques used in electron scanning tunnelling microscopy (STM) consists of maintaining the tip at a constant distance from the surface by means of an electronic feedback loop. The loop performs tunnelling current measurement and tip readjustment several times per second. This control is realized by driving the piezo-actuator with a voltage related to the detected tunnelling current. Assuming isotropic material, the z-displacement will be then strongly connected to the object topography.

At the present time, most STM scanning is performed in constant current mode. However, the STM can also work in constant altitude mode. In this case, the tip extension is not changed during the scan.

These two scanning modes have been transposed in near field microscopy (see figure 10(*a*) and (*b*)). Unfortunately, in near field optics, the relationship connecting topography and detected intensity is complex: the field does not increase monotonically when the tip approaches the surface. The complexity of the relation is due to the fact that the near field is a mixture of propagating and evanescent components. The result of such a combination is a field whose variations depend (1) on the topography, (2) on the nature of the material and finally on the distance of analysis. Except in the case

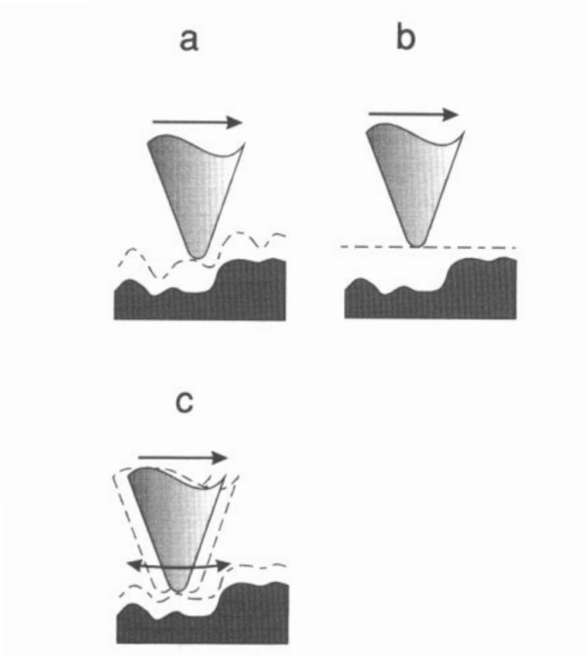


Figure 10. Procedure of surface scanning. In (*a*) the tip follows the isodensity lines of the field. This mode is known as constant intensity mode. In (*b*), the tip moves in a given plane (constant altitude mode). Even if this mode is not satisfying it is often used because of its simplicity of use. The case (*c*) is known as constant distance mode. It is very similar to the equivalent mode in STM. The distance is kept approximately constant by vibrating the tip and measuring the amplitude of vibration. Because the latter depends on the distance between the tip and the sample, it is rather easy to minotor this distance precisely.

of very small topography variations, [23] the distance control by means of the detected signal itself is highly hazardous. It is hazardous because the resulting information is far away from the topography and overall, because the tip following a virtual topography, can crash on the surface.

Recently, a new control technique has been proposed by two American groups [24, 25]. It consists of controlling the distance by measuring the lateral force exerted on the tip by the surface (see figure 10(*c*)). When the tip is close enough to the surface (few nanometres), the long range forces such as capillarity or viscosity will brake the *xy* motion of the tip. This effect is generally so weak that it cannot be noticed during scanning. However, if the tip vibrates slightly, at the resonance frequency, the damping of the vibration can be easily observed and measured. This is the basis of the new 'shear force controlled' microscopes. Two different techniques have been proposed almost simultaneously. The first one is derived from chopping measurement [24]. Assuming nano-aperture emittor configuration, the image of the illuminated area of the sample is projected on a pinhole by means of an auxiliary lens (see figure 11(*g*)). Because of lateral vibration of the nano-aperture, the pinhole will chop the light beam periodically. A photomultiplier will then convert the light signal into electric current whose constant part (direct current) will be proportional to the near field signal (see comments about Synge device) and the oscillating part after rectification will inform on the modulation amplitude of the hole. If the hole approaches the surface, the resulting damping will reduce the amplitude of modulation and consequently the alternative component of the electric current. The system works fine, the only drawback is the possible coupling between the two signals emanating from the small light source (the pinhole must be perfectly centred in order that the oscillating term average equals zero). The second technique [25] uses an auxiliary laser beam. This one is split into two beams with two different polarizations. They are then focused on two points A and B of the fibre as indicated in figure 11(*h*). When the tip oscillates the amplitude and the phase at A and B will be different. The technique consists of measuring the phase difference by means of a Nomarski-like interferometer (heterodyne interferometer). The method is very precise but the use of an interferometer does not simplify the set-up. However, excellent results have been obtained with such a method.

Finally, the previous technique has been simplified as shown in figure 11(*j*) [26]. In this configuration the laser beam is projected onto the fibre. The resulting diffraction pattern is directed to a photodiode slightly set off. Because of oscillation, the diffraction pattern will scan the photodiode area leading to an alternative photocurrent. After removing the background the rectified oscillating term will be used for monitoring the z position of the tip. To introduce the tip dithering, in all the methods we simply add a piezo-element vibrating the tip laterally over a few nanometres. The main drawback of such a technique is the loss of resolution introduced by the lateral oscillation. Moreover, the nature of the shear force is not well known. It is probably the combination of van der Waals forces, capillarity forces, viscosity forces, etc. The shear force topography will then depend on the material properties. Moreover, the optical image is probably polluted by the shear force control.

Anyway, the method works rather well and it recently allowed the AT&T group to provide very good near field images.

Presently, almost all the groups working in near field microscopy have adopted this tip position control.

Another technique must be mentioned. Although it follows the same aim, it is radically different: instead of removing the tip from the sample, it consists of putting the tip directly on the surface. For avoiding or at least limiting the risk of scratches, the tip is associated to a microcantilever whose stiffness is small enough for limiting such risks. The simplest way thus consists of using microcantilevers commercialized for AFM use. With such a technique, the tip follows the true topography and gives information on the field on the surface. At the same time, it is possible to get the topography of the sample.

The technique has been only used in total internal configuration (figure 11(*i*)). Two ways of detection have been explored. The first one consists of combining a classical AFM detection system for measuring the topography [27]. An auxiliary lens projects the magnified image of the tip onto a small hole ensuring a spatial filtering. The resulting signal is proportional to the light field on the sample. The other solution [14] consists of placing a cleaved fibre near the reflecting surface of the cantilever. A light beam launched in the fibre will then be reflecting on the cantilever leading to interference between the latter and the fibre extremity. When the tip scans the surface, the distance fibre–cantilever varies and the reflected intensity will be modulated by the interference pattern. By demodulating it is possible to retrieve the topography of the sample. The fibre plays the role of collector and spatial filter for the light beam transmitted through the tip.

The two techniques are almost equivalent. Such a procedure could give information directly from the field on the surface. However, the probable intercorrelation between the two types of data must be taken into account. During scanning, the cantilever can

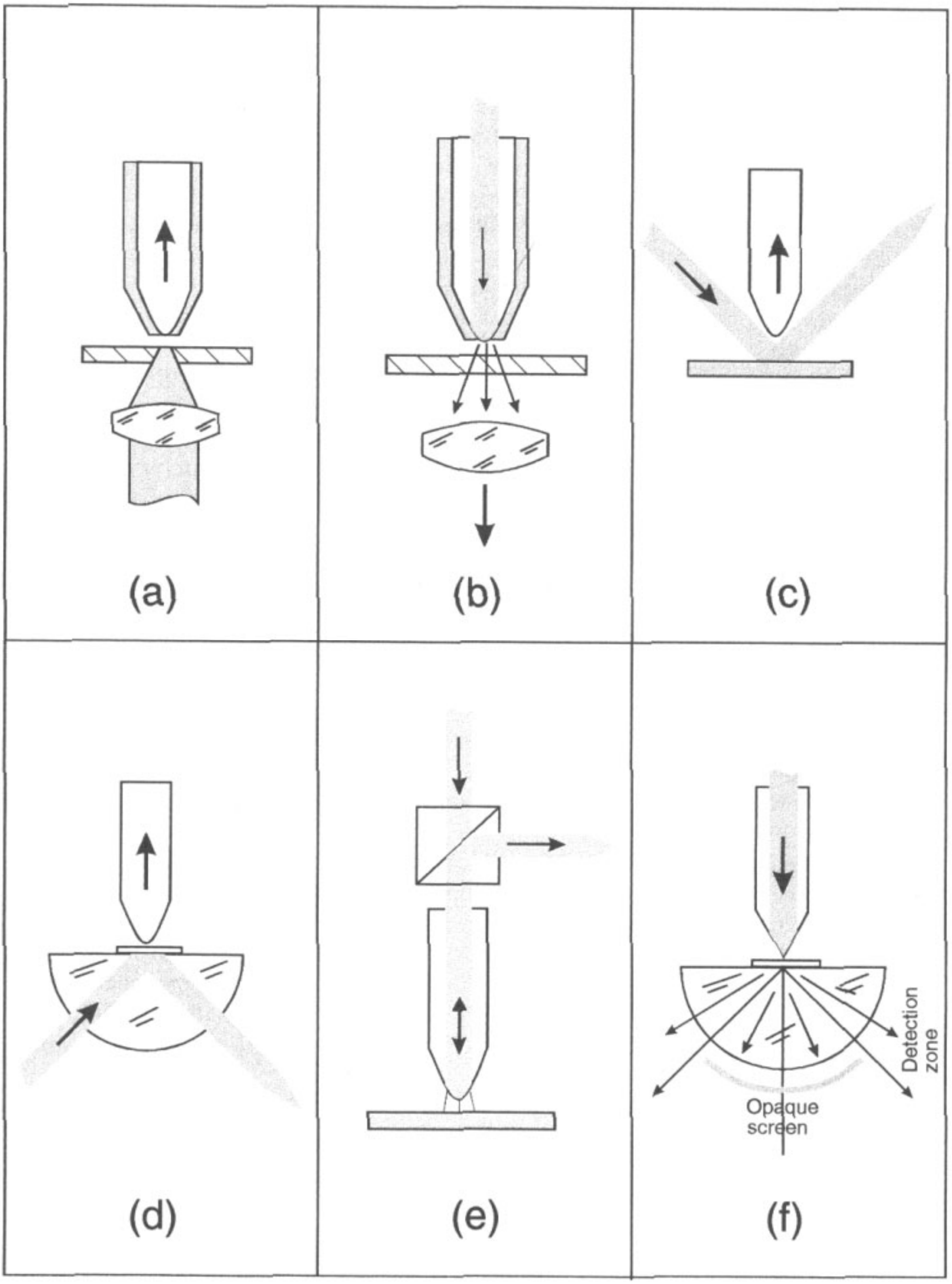
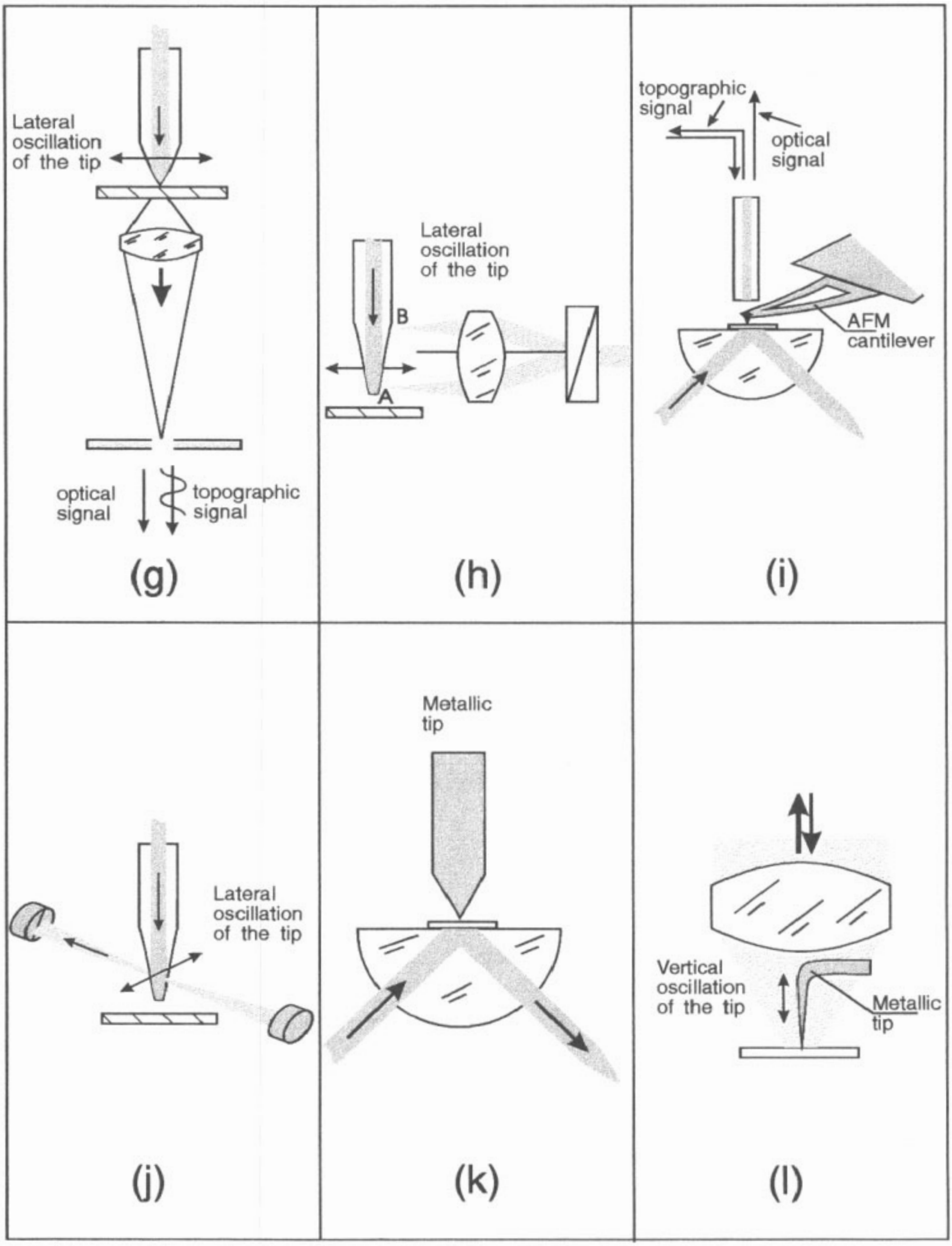


Figure 11. Different SNOM configurations. (*a*) Transmission collection mode. The tip is generally metallized, the very tip being free of metal for ensuring the transmission of light through the nano-aperture so created. (*b*) Transmission illumination mode. In this case the tip plays the role of nano-source. (*c*) External reflection collection mode. The tip collects the object asperities. (*d*) Internal reflection collection mode. The illumination beam is totally reflected inside a piece of glass. The tip, generally uncoated, collects the resulting evanescent waves lying on the surface. (*e*) External illumination-collection mode. It is a combination of (*a*) and (*b*). (*f*) Internal reflection illumination mode. It is an inverted STOM (configuration (*d*)). The light is detected beyond the critical angle. (*g*) Transmission illumination mode + shear force control. The tip vibrates laterally in the vicinity of the object. The beam transmitted by a pinhole is separated into two components: a constant voltage signal proportional to the near field on the object, and an oscillating signal whose amplitude depends on the distance between tip and object. After rectification, this signal is used for monitoring the tip position.



(*h*) Similarly to (*g*), the tip oscillates laterally. The oscillation amplitude is measured by projecting two beams on the fibre and detecting the phase variations between them. (*i*) Internal reflection collection mode + AFM control. It could be called contact mode optical microscopy because the tip touches the surface. The new field is detected through the pyramidal hollow tip. The topography is detected by classical AFM detection. (*j*) Reflection or transmission mode + shear force control. The simplest way to control the tip position. The lateral oscillation of the tip is detected by measuring the diffraction (or the shadow) of a light beam impinging on the fibre. A single or two quadrant photovoltaic cell detects the time variations of the resulting field. (*k*) Internal reflection perturbation mode. A completely opaque tip perturbs the evanescent field on the object surface. (*l*) External reflection perturbation mode. Similar to (*k*), this system works in reflection. A very thin metallic tip oscillates vertically in the vicinity of the object. The resulting new field perturbation is detected by means of a classical microscope objective.

bend and the field transmitted by the tip can be deflected giving artefacts in the near field detection process.

Preliminary results have proved the interest of such an approach even if routine works are still unrealistic.

12. Illumination procedure

The second important feature is the illumination mode. We can imagine microscopes working in transmission or reflection mode, where the tip will be used either in collection mode or in illumination mode. Theoretically, because of reciprocity, the collection or emission modes are physically equivalent. From an experimental point of view, it is generally not the case. This point will be discussed further.

The widely used technique proposed as soon as 1983 is the transmission illumination technique whose principle is schematized in figure 11(*b*). Belonging to the same family, the total internal reflection microscope which was born in 1988 uses an evanescent field for illuminating the sample. Because of the peculiar geometry it can be considered as the intermediate step towards reflection microscopy. Concerning reflection microscopy, three techniques have been proposed. The first one is based on a modification of the transmission microscope using nano-apertures, the second one uses the same fibre for illuminating the sample and collecting the light at the same time. Finally the third one is based on the measure of the scattered field by a tip.

Many configurations have been derived from these basic schemes. We will describe the merits and drawbacks of each one in the following.

12.1. Transmission microscopy

As mentioned, transmission microscopy is the oldest technique and probably the simplest one (figure 11(*b*)). It was developed first at IBM Zurich Research laboratory. It is basically very similar to STM, in which the metallic thin tip is replaced by a partially metallized quartz tip. This is probably the main element of the set-up. As already explained the authors started from a quartz rod which was cut and polished to the shape and then covered with an opaque metallic coating. The latter was cold-deformed at the apex by pressing against the sample surface. This action will lead to a tiny weak spot of light. Under these conditions apertures with a diameter of 10 nanometres were created. Unfortunately, such a method was not reliable enough and the shape of the nanoaperture was not reproducible at will. The sample was mounted on a combination of a conventional xy coarse positioning system and a bimorph driven $50 \times 50\ \mu m$ stage for raster scanning. Distance control was ensured by measuring the tunnel current between the (metallic) tip and the sample assumed to be sufficiently conducting too [12]. The light transmitted through the aperture and the sample was directed on to a photomultiplier by means of an auxilliary microscope objective. The optical signal was picked up from the image plane on top of the microscope with a glass fibre allowing one to discriminate the signal from the stray light. The first images obtained by this configuration are not obsolete today although progress has been made in simplifying the set-up. Almost at the same time, at Cornell University, Betzig *et al* developed another transmission configuration. The tip was formed by the end of a micropipette which was produced by heating and pulling a glass capillary. The technique of fabricating such a micropipette is well known by biologists who use them as microelectrodes for exciting muscles or cells or as a capillary for injecting or drawing up cell components. The resulting inner and outer diameters of such microcapillaries decrease from the former value to a few nanometres. With practice, it is possible to get inner diameters of about 50 nanometres. After realization, the tip is metallized. The quasi-perfectly circular symmetry of the technique and the possibility of generating a true nano-aperture (the metallic coating leaves the opening of the capillary free of metal) are the two significant advantages in Pohl's technique. Unfortunately, the optical coupling was bad and some hints have been proposed for example by filling the capillary with a suitable liquid able to match the index of propagation. It consists of illuminating a very thin slice of sample by means of a large aperture microscope objective. The resulting transmitted field is then detected in the near field zone by means of a nano-aperture collector. Because of the problems of coupling mismatching inside the capillary, the latter has been replaced by a cone shaped fibre the fabrication of which has been described previously. By changing the heating and pulling conditions it is possible to generate various shapes. As an example, if the diameter decreases slowly enough (compared with the length of the cone), it is possible to ensure adiabaticity in the transmission through the fibre. It is thus possible to keep the polarization of the near field and thus to map the polarization of the near field along the object.

Such microscopes can be easily inverted, the aperture playing then the role of a nanoemittor (figure 11(*a*)).

With resolution of about 50 nanometres, it is probably the most mature microscope at the moment. Its redhibitory defect is the illumination mode imposing first transparent materials. It needs particular sample preparation methods.

12.2. Total internal reflection microscope

Born at the same time in Europe and in the USA [17, 18], the total internal reflection microscope is still controversed (figure 11(*d*)). For some people (we belong to them), it is a simple version of the transmission microscope. For others, it is a new concept because of the evanescent illumination mode. Anyway, its interest is the ability to analyse transparent surfaces and very low profile objects. Moreover, the use of an incident evanescent field allows one to control the tip position with a rather good precision. Finally the stray light is strongly limited (for the same reason). As in the previous microscope, it is possible to choose the polarization of the incident beam. In this sense it is a good analytical tool because of the different parameters which can be studied [28, 29]. Its main drawback is the oblique illumination introducing a strong anisotropy in the case of an object with topography variations of several hundred nanometres. The set-up is composed of an element ensuring total internal reflection. It can be a prism, a guiding glass plate, a cylindrical lens or even a hemispherical lens. A slightly focused laser beam is projected in the component in order that it reflects totally. The object, for example a thin glass plate with topography variations, is put on the flat surface of the element. It will then perturb the incident evanescent field which will be converted into propagating components and secondary evanescent fields.

The tip, generally a monomode fibre suitably tapered as explained previously, is then immersed in the evanescent field.

Here is a first point of controversy concerning the word 'perturb'. Strictly speaking, an electromagnetic field impinging on an object is converted into new electromagnetic fields rather than perturbed. This conversion is composed of two steps: the first one is the induction of electric charges and current densities in the matter, the second one is the re-emission of an electromagnetic field by the charges and by the induced electric currents. The latter depending on the excitation and on the nature of the material, the re-emitted light will code the information about the lighted object. This analysis holds whatever the type of incident waves (propagating or evanescent ones). Consequently, immersing an object in an evanescent field will not be equivalent to generating a kind of imprint of the object inside the evanescent field. This one will be strongly modified and the relationship between topography and detected field will be generally very difficult to determine.

Note that, recently, the system has been used in illumination mode as shown in figure 11(f) [43].

12.3. Reflection microscopy

12.3.1. Oblique illumination and nano-collection. The solution that is often mentioned consists of using the transmission configuration and illuminating the sample from the top as indicated figure 11(c). The technique has not proved its viability at the moment because of the mismatching between illumination beam and collector system.

12.3.2. Illumination and detection with the same tip. The second way, which seems to be more promising is to use the same fibre as the nano-emitter and nano-collector (figure 11(e)). Such a technique is not free from inconveniences: first the amount of light transmitted by the very tip (the only region allowing high resolution) and back reflected inside the fibre is exceedingly small. Moreover it is still a point of controversy. For certain authors, the signal to noise ratio is so weak that only low frequencies can be detected. It is true that from the first results that low frequencies (around ω/c) have been imaged. It is thus often necessary to use synchronous detection for increasing the S/N ratio. Paradoxically only a few laboratories have invested in this way, probably because of the common opinion that the signal is too weak to be detectable. However, both theory and experiment seem to contradict this position. By exploiting resonances it is probably possible to gain both in resolution and in signal to noise ratio [30].

12.3.3. Nano-emission and isotropic detection. In this configuration the tip is used as nano-emitter and the scattered light over the object surface is detected by means of a suitable catadioptric system, for example a cassegrain telescope [31]. It is a probable promising solution.

12.3.4. Perturbation of the field by means of a nano-antenna. This new concept is based on the fact that the near field lying on the object surface can be locally perturbed by a small scattering centre. The latter will be excited and the resulting coupling will slightly modify the field in the vicinity of the scatterer. If the scatterer oscillates the resulting perturbation could be detected by extracting the alternative component in the scattered field. The principle is interesting because it is no longer necessary to use dielectric tips. Tungsten tips like those which are used in STM can be employed. Moreover, it is possible to benefit from the high polarizability of metals.

From this basic idea, several devices have already been developed. The first one is a transposition of the STOM/PSTM. In this configuration, as explained previously, a light beam is internally reflected inside a prism. The novelty brought to the basic configuration is to detect the reflected beam rather than the light frustrated by the tip (see figure 11(*e*)). For increasing the signal to noise ratio, a spatial filter is added. When the tip (in metal) is brought very close to the sample, the field perturbation will affect both the evanescent field and the reflected one. By measuring the intensity variation of the latter, it is possible to detect the topography of the sample. If the object itself is metallic, the plasmon resonance will dramatically enhance the perturbation. Finally it is possible by measuring the electron tunnelling effect between tip and sample to monitor the tip position with high precision. Some interesting and unsurpassed resolutions have been obtained in the group of Münich in Germany [32].

Another more recent device is derived from conventional microscopy. Let us assume a microscope objective focusing a light beam onto an object If a very thin metal oscillating needle is set between object and objective, the magnified image of the sample will be composed of a low resolution image locally perturbed by the oscillating tip. The detection of the alternative component in the detected field will be connected to the high resolution details in the object. The first results obtained confirm the validity of the approach. It is probably an interesting alternative to more conventional techniques [33].

13. Applications

What is the reason for developing near field microscopy? The straightforward argument is the ability of near field detection to break the diffraction barrier and to lead to resolutions never attained with usual optical tools. This argument is justified although highly resolved images have rarely been published. In this sense we have not assisted the fantastic progress we observed some years ago in electron scanning tunnelling microscopy. The reasons for the difficulty in reaching ultimate resolutions are probably the problem of the scanning plane and the signal to noise ratio. Concerning the first point, because of the non-trivial relationship between topography and field distribution, it is generally not possible to use the optical signal itself for controlling the distance between tip and sample. Such a method works only for exceedingly weak topography objects (only a few nanometres variation, see figure 12). In the general case only the constant altitude scanning mode works. Unfortunately, in such a case, it is highly improbable to be during the whole scanning in the near field region (see figure 10). It means that only small areas will be well resolved. Such an effect has been observed very often whatever the type of near field configuration which is developed. The solution of controlling the tip–sample distance by means of an auxiliary force detection control, is the last resort. Although it works well and maintains the tip at a short distance from the surface, the detected signal is not a true optical signal. It is always the mixture of optical effects and force artefacts.

As an example, let us assume that the object to analyse is a perfectly flat surface composed of two different regions. We suppose that the two regions are optically similar (same index) and mechanically different (difference of hardness or of vapour pressure). When scanning such a surface, the tip will perceive the difference and will move vertically when passing from one region to another. A false topography will result and the optical signal will lead to the same variation confusing the observer. This effect is certainly

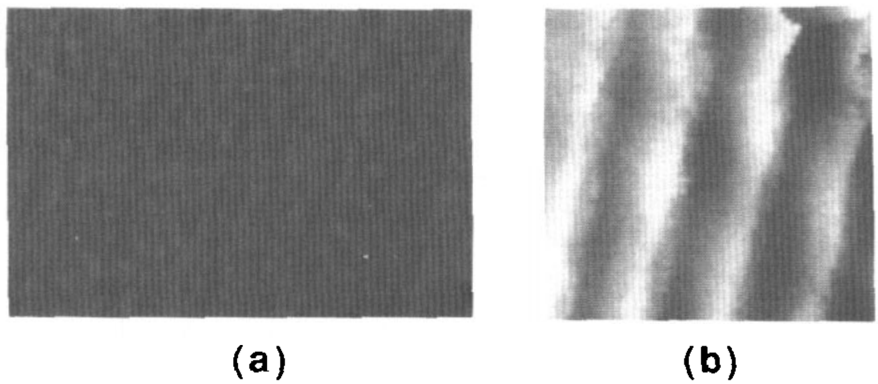


Figure 12. Image of a grating obtained by internal reflection (configuration (*d*) figure 11(*a*)). The grating is 3 nanometres high. (*a*) is the SEM image (electron scanning reflection microscope), (*b*) is the near field optical image. In this particular case, the topography is exceedingly weak. The optical image has been obtained in constant intensity mode.

very common in hybrid microscopy. As we previously said, the microscope works and gives rather well resolved images. However it will never become an analytical tool as its brother the STOM/PSTM does.

The second potentiality offered by near field optics is the ease of changing the wavelength.

A simple interference filter can be used for selecting a wavelength or rejecting such and such chromatic band. The first client for this facility is local spectroscopy. This domain will probably be one of the principal activities in near field.

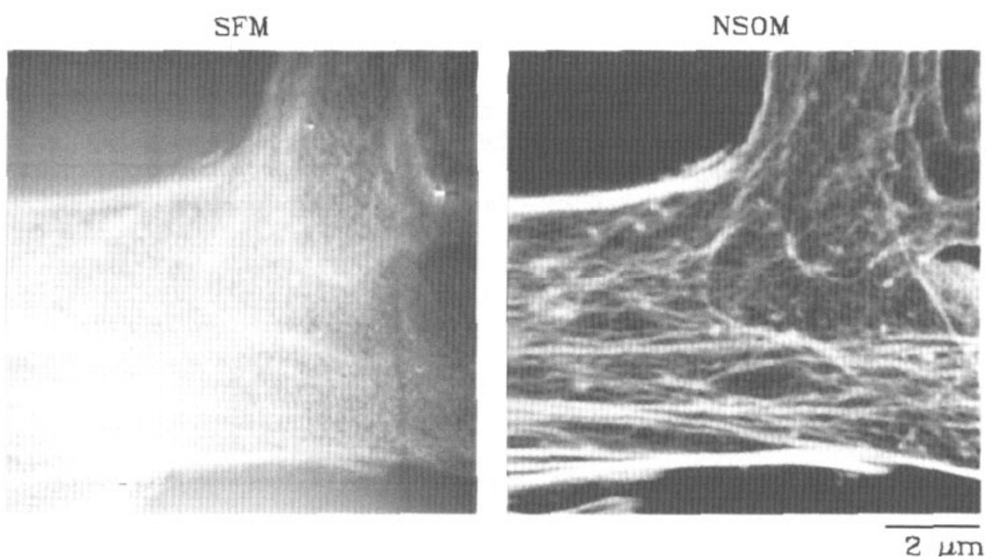


Figure 13. Image of a single lamellipodium from a mouse fibroblast cell. Image (*a*) is a shear force topographic image whereas (*b*) is the near field fluorescent image of the same region. We note the good signal to noise ratio in the last case and the lateral resolution which is clearly superior to confocal microscopy. These images have been obtained with the configuration (*g*) described in figure 11(*b*). (By courtesy of E. Betzig [41]).

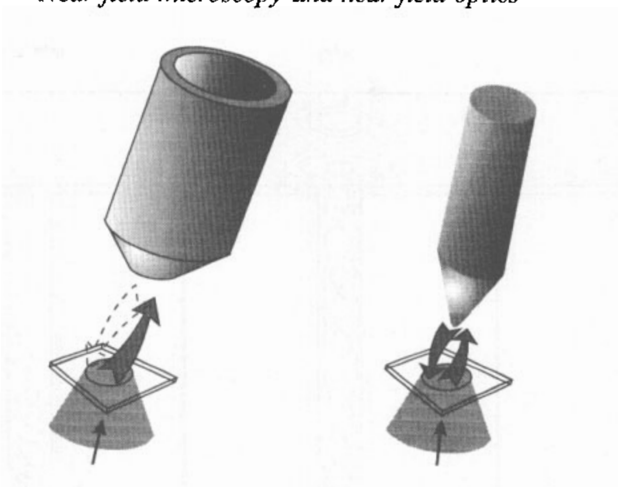


Figure 14. The fundamental difference between microscope objective collection (*a*) and tip collection (*b*). In the first case, the signal propagates from the sample to the objective. Almost no light is back-reflected from the objective lens to the object. The communication between object and collector is thus one way. In near field (case (b)) the light does not propagate, the interaction between sample and tip is mainly non-radiative. The communication between tip and sample is perfectly reciprocal.

It is also very simple toggling from one polarization to another. This can be performed by means of a half-wavelength plate and if we worry about the risk of perturbing the system during mechanical handling of the plate, it is possible to use an electro-optical device instead, such as a Kerr cell for rotating the polarization.

If the capacity of changing the wavelength opens the door to local spectroscopy and fluorescence, toggling polarization can offer a new way in storage data techniques. These various possibilities will be described in more detail in the following.

Another application of near field microscopy is simply the study of near field behaviour. The near field microscope is a fantastic tool not only for experimentalists but for theoreticians too. The behaviour of evanescent waves is not well known because, in essence, they cannot be detected. However, if we use a macroscopic collector (the second prism in Newton's experiment), the latter will strongly perturb the evanescent field detection and the resulting propagating field will not reproduce faithfully the evanescent field distribution on the object surface. The interest of using a very small collector such as the tapered extremity of a fibre clearly appears: the perturbation will be small enough to respect the evanescent field behaviour. Recent images of near field propagation of plasmons over a metallic surface confirm the exactness of this analysis.

A consequence of the deepening of our knowledge in near field microscopy is near field optics or non-radiative optics. It will be soon a new opportunity in integrated optics, since using non-propagating fields allows one to circumvent the conditions of propagation which impose structure dimensions larger than half a wavelength.

If the first meeting specifically devoted to near field optics [34] showed the rapid growth of near field techniques, the second one in October 1993 [35] confirmed the explosion of the domain in many different directions. In instrumentation, improvements of the basic configurations have been described; as it is mentioned above, hybrid techniques combining force microscopy and near field microscopy are now widely used. Similarly, low temperature microscopes, infra red microscopes, etc, begin to run in the

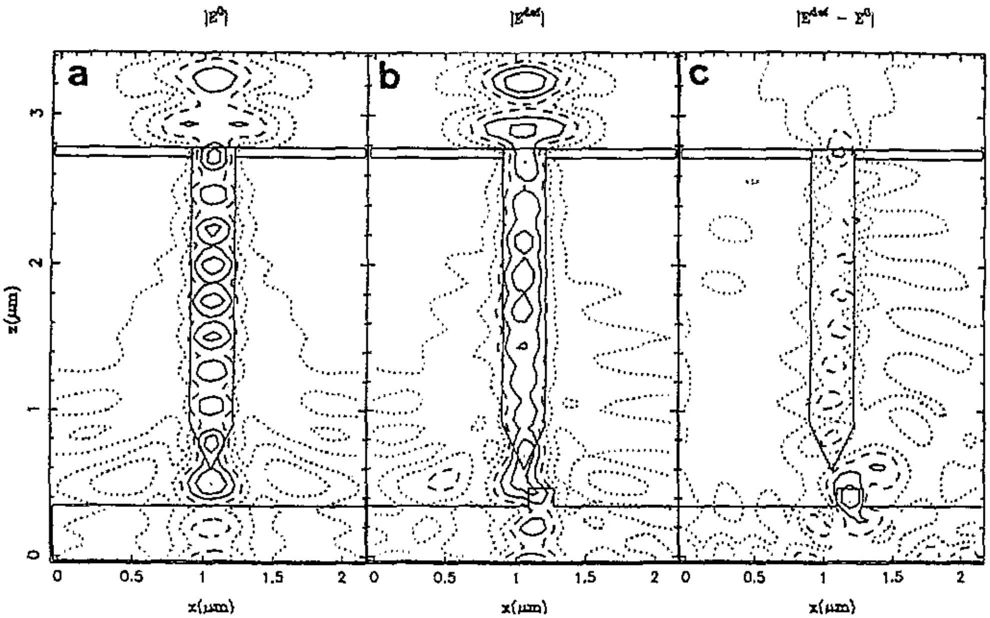


Figure 15. Example of numerical simulation. It shows the near field distribution inside a two-dimensional SNOM junction. This simulation is based on the parallel resolution of Lippman–Schwinger and Dyson equations. (By courtesy of C Girard [42] and A Dereux.) (*a*) A perfectly plane surface is placed in front of the tip. (*b*) A small defect is introduced on the surface. (*c*) Amplitude of the difference between configurations (*b*) and (*a*).

laboratories. But the most significant advances are undoubtedly in spectroscopy and fluorescence. The first experiments which have been carried out consisted of illuminating a sample at a given wavelength and detecting the electromagnetic field at a different wavelength, the object for these first attempts being for example small polystyrene spheres or other fluorescent material. The aim of such attempts is first of all to determine the fluorescence in the subwavelength domain. Many applications can be imagined from biological markers for tracking a given biological component with a high spatial resolution to chemistry, semiconductor technology, etc (see figure 13). Some physical data can be deduced from such experiments. It is well known that the fluorescence of a molecule or a bulk of matter can be quenched by bringing an object very close to the fluorescent material. Such a property is well known but has been observed only for macroscopic objects. It is particularly interesting from a simple scientific point of view to study the fluorescence quenching in the nanometre range. The first sigificant attempts have been made recently by a group in Munich [36]. These works open the door to the unknown domain of local spectroscopy. Very recently, in a paper published in *Science*, a group at AT&T has imaged single molecules of carbocyanine dye molecules [37]. These first images of molecules analysed by near field optics are dramatic even if the resolution does not reach that of STM. But it is definitely the first step towards a new form of optics.

13.1. Notion of detection, localization and imaging

One of the most subjective data in imaging is the notion of image quality and more scientifically the notion of resolution.

About detection, we know that some detectors are able to detect elementary particles such as molecules and even atoms. The photomultiplier is a well known example. It is sensitive enough to detect the presence of a single emissive particle but cannot localize it. If we combine such an array of highly sensitive detector with an imaging system, it is then possible to detect the emission of single particles and at the same time to localize them with accuracy; even though the images of each single particle extend over distances larger then $\lambda/2$. Based on this principle, in the sixties, scientists developed a new kind of optical microscopes called ultramicroscopes because of their ability to detect and to image exceedingly small and weak light emission from viruses or other features. The name chosen for these microscopes was somewhat misleading because it pointed out the capability of such devices to surpass a certain barrier. In fact they behave as cassegrainian telescopes collecting by means of very open and low absorption reflecting systems the maximum of available photons. As explained above, the interest of such microscopes is their capacity of localizing single particles with a high precision. Indeed, if we *a priori* know that the observation scene only contains single particles, by determining the energy maximum in every detected spot, we can define with accuracy the distance between the particles without imaging them. Consequently we can follow bacteria or virus motions even though the moving spot does not contain any information about the object itself.

The last step in the exploration of terminology is the notion of imaging.

Imaging means the generation of a facsimile of the object distribution. It obviously includes the two previous conditions i.e. detection and localization. In astronomy, the presence of a star is first detected, then localized and sometimes imaged.

Experience shows that localization is often confused with imaging.

Two complementary ways can be used for proving that the output signal can be considered as an image. The first one is the study of its Fourier spectrum. Assuming no scanning artefacts, if the spectrum varies when changing the object, there is a real chance to retrieve in the image plane object information. A second way consists of imaging well known structures such as two points separated by a distance d. When the two points are seen individually, the system behaves as an imaging system. On the contrary, if a single spot is observed, the system has detected the presence of the two points without discriminating each individual point. This test is probably the most serious one for proving that a given resolution is reached. Unfortunately it is not systematically used by experimentalists.

13.2. Quenching light by near field detection

The near field microscope used for this experiment is based on the nano-antenna principle and detection by near field perturbation [36]. Let us consider a green He–Ne laser beam ($\lambda = 543$ nm) focused on a glass plate. Ruby microcrystals deposited on the glass plate will be optically excited and will emit fluorescence at a larger wavelength ($\lambda = 694$ nm). The fluorescence light is collected by the same collector and directed after filtering through a dichroic filter to a cooled photomultiplier. Simultaneously, an AFM (atomic force microscope) allows one to image the topography of the dust particles. For ensuring the fixing of the microcrystallites on the glass the latter is heated until its melting point. It is then possible to measure both the topography of the sample and at the same time the effect of the presence of the pyramidal nitride tip on the fluorescence of the ruby particles. The risk of confusing quenching and simple frustration of light by the tip (leading to a similar result) has been avoided by comparing the images of two particular states of ruby particles. The results obtained by the scientists of Munich show differences in the images confirming the validity of the technique.

In this experiment the fluorescence emanates from the sample itself. It is possible to imagine a fluorescent tip as suggested some years ago by Lewis *et al.* They proposed to fill a microcapillary with a fluorescent liquid excited by means of a UV pump beam. Some attempts are currently developed by fixing some fluorescent molecules at the very tip of a glass fibre and using the quenching effect when approaching the tip.

13.3. Observing single molecules by local detection

As mentioned above, the group of AT&T published an interesting work dealing with the observation of individual carbocyanine dye molecules in a sub-monolayer spread [37]. They claimed they were able to localize with a precision of about $\lambda/50$ some molecules with a sensitivity of at least 0.005 molecule. Moreover, it was possible to determine the orientation of each molecular dipole.

Let us see the experimental set-up in more details. It is based on the use of a nano-aperture playing the role of nano-source (see the techniques of nano-aperture described previously). The beam emitted by a gas laser is injected into an optical fibre suitably metallized for generating a nano-aperture. The sample was prepared by spreading a dilute solution of carbocyanine dye across a cover slip previously covered with a 30 nm thick PMMA, given an area of about 23 molecules. The sample was then imaged by using an hybrid SNOM microscope combining distance control by shear force and optical detection as shown in figure 11(*g*). The fluorescence of the dye molecule was then detected with suitable low noise photodetectors. The results show without ambiguity some bright spots corrsponding to single molecules (or at least to a very small number of molecules). Following the authors it was then possible by changing the polarization to determine the orientation of each molecular dipole. Even if these results have to be confirmed, they show the power of near field microscopy, its capacity of working with very weak signal and it was then possible to localize some single molecules. The interest of such an experiment is the possibility of analysing the polarization effects of the light beam over the molecules.

14. A few words about theory

Because of the near field image complexity, the help of theoreticians is absolutely necessary for understanding the image formation. Before describing briefly the tools developed in modelling let us first sum up the problem of information transfer in optical microscopy.

After the last world war thanks to the works of scientists such as Hopkins or P M Duffieux, we discovered that the transfer of information through any imaging systems reduced to a simple product of a function characterizing the object and a function characterizing the apparatus. The first one is connected to the spatial frequency distribution in the object whereas the second one expresses a transfer coefficient for every spatial frequency of the object. Generally this coefficient is nearly 1 for low object frequencies (the latter are well transmitted), and it drops to 0 for high frequencies. It is then possible to determine a cut-off frequency beyond which no frequencies can be transmitted. This apparatus function is called the transfer function and whatever the system and the illumination technique, it is perfectly defined and unique. In other terms it is possible by knowing the object structure and the transfer function to predict with a high precision the image intensity distribution. It is nothing but the transposition to optics of the temporal filtering technique holding in electronics.

The possibility of defining a single and unique transfer function results from the fact that the object does not interact with the collector (here the microscope objective). In other words, whatever the microscope, the light emitted by the object to be imaged is not (or at least slightly) perturbed by the objective. This property can be easily understood: because of propagation, the light goes from the object to the objective. The amount of light going from the objective to the object is exceedingly weak. By using a mathematical terminology, we can say that the information transfer in a conventional microscope is essentially asymmetric: information goes from object to detector. When modifying the collector–detector system we modify the filtering effect of the system, but the relation object–image stays linear. Such a property is very useful for microscopists, because it allows one to predict the kind of image we can expect with accuracy.

What about near field microscopy? Following the same processes of thought, some attempts have been made to transpose the powerful notion of transfer function in near field optics.

Unfortunately, the notion of collection–detection is completely different in this case (see figure 14). First, the tip does not image the sample, it collects for a given position *the near field intensity. Moreover, do not forget that the significant features in the near* field are the non-radiative components. The latter do not propagate, they are converted *into propagating ones by optical tunnelling. Consequently, this kind of detection is* based on the perturbation of the field to be detected. In other words, on the contrary to far field detecton which does not modify the incoming field, near field detection destroys the field to be analysed. This paradoxical situation has a dramatic consequence: the image is the mixture of sample and tip information. Roughly speaking, the tip images the sample and the sample images the tip. As an example, if the tip is larger than the object detail to be analysed, the resulting image will be more correlated to the tip profile rather than to the sample topography. This rhedibitory drawback is well known in STM and AFM.

In this kind of situation, it seems to be strictly impossible to define an intrinsic transfer function. The latter depends upon the tip structure, the tip distance and the object characteristics. However some attempts have been made in a specific case and the information contained in the so called *apparent or pseudo-transfer function* is interesting for the metrologists despite the restriction mentioned above. Such apparent transfer functions must be handled carefully keeping constantly in mind that they have sense for a given situation and can never be extended to any object.

The consequence of the failure is that the solution of the inverse problem is highly improbable in the future. Let us recall that the inverse problem solution consists in trying to retrieve from the image data the object features. It is well known in astronomy where strongly blurred images from the sky are numerically processed to restitute the star characteristics. Near field microscopy allows us to break down the diffraction barrier. In return, the collector totally interacts with the objects to analyse. It is the *price to pay for the gain in resolution.*

15. Solving the direct problem: the modelling of the interaction object–light–tip

Before trying to retrieve the object features from the image data, one can try to define a model of interaction between the object and the light field in the first step and between the resulting field and the collector in the second step.

Two classes of mathematical methods have been carried out [38]. The first consists of starting from Maxwell equations by assuming continuity conditions over the various surfaces interacting with the field. These techniques are known as macroscopic and continuous techniques. Their main advantage is to keep track of the physical phenomena taking place because they are essentially analytical.

The second category of techniques called discrete techniques consist of describing the matter as a set of elementary components (some dipoles for example). The problem to solve is then to find the right transmission function between each object element and each tip element. This transmission function is generally a Green function. The technique seems simple. Unfortunately, to ensuring a sufficient precision in the description of the matter, we have to take into account a large number of discrete elements. As an example, if we assume that the object is composed of 100 by 100 elements and the tip contains 100 elements, we must consider the 10 thousand object elements individually which will interact with the hundred elements of the tip (and also with each other). For modelling more faithfully the physical reality of the matter, you have to increase the number of independent elements and overall accept to work in three dimensions. In these conditions a supercomputer will quickly become necessary and unavoidable.

A significant effort is presently being made between two laboratories in the framework of the European network [39, 40]. Using discrete methods the implied scientists are developing powerful software able to run on small UNIX stations. Such software could be spread out in the SNOM community allowing the scientists to predict with a rather good precision the behaviour of the near field on the object under test and overall the capacity of collecting of the tip as shown in figure 15.

The software works fine in two dimensions, it is currently generalized to 3D.

16. Conclusion: near field optics, the second step after near field microscopy

When a new tool is invented in physics, the first application is always the observation. It has been the case for optical microscopy, classical electron microscopy, acoustic microscopy STM . . . . After this passive period, the tool is used for understanding the physical mechanisms taking place. Spectroscopy is then the following step. Finally, when the phenomena are understood they are applied and used in applications far away from the simple observation. It is typically the case of near field microscopy. Because we know that the near field can vary rapidly over smaller distances than the wavelength, we can imagine a new optics involving interference between evanescent (non-radiating) fields. It is not meaningful to generate such high frequency fringe pattern undetectable by far field diffraction. Such subwavelength pattern could be used in nanometre range optical processors. If experiments verify these hypotheses, it could be possible to ensure transfer of information by non-radiative transfer over very small distances and inside subwavelength structures. The integrated optics could benefit this subwavelength optics.

References

[1] Binnig G and Rohrer H 1982 Scanning tunneling microscopy *Helv. Phys. Acta* 55 726–35

[2] Dürig U, Pohl D W and Rohrer H 1986 Near-field optical-scanning microscopy *J. Appl. Phys.* 59 3318–27

[3] Betzig E, Lewis A, Harootunian A, Isaacson M and Kratschmer E 1986 Near-field scanning optical microscopy (NSOM) *Biophys. J.* 49 269–79

[4] Wolf E and Nieto-Vesperinas M 1985 Analycity of the angular spectrum amplitude of scattered fields and some of its consequences *J. Opt. Soc. Am.* **2** 886–9

[5] Born M and Wolf E 1975 *Principles of Optics* (Oxford: Pergamon Ps) pp 81–4

[6] Vigoureux J M and Courjon D 1992 Detection of non-radiative fields in light of Heisenberg uncertainty principle and Rayleigh criterion, *Appl. Opt.* **31** 3170–7

[7] Goodman J W 1968 *Introduction to Fourier Optics* (London: McGraw-Hill) pp 48–55

[8] Schaeffner M and Toraldo di Francia G 1949 Microonde evanescenti generate per diffrazione *Nuovo Cimento* **VI** (2) 1–6

[9] Pohl D W 1991 Scanning near-field Optical microscopy, *Advances in Optical and Electron Microscopy* ed C J R Sheppard and T Mulvey **12** (London: Academic) p 12

[10] Synge E H 1928 Suggested method for extending microscopic resolution into the ultra-microscopic region *Phil. Mag.* **6** 356–362

[11] Fischer U C 1986 Submicrometer aperture in a thin metal film as a probe of its microenvironment through enhanced light scattering and fluorescence *J. Opt. Soc. Am.* **3** 1239–44

[12] Pohl D W, Denk W and Lanz M 1984 Optical sthetoscopy: image recording with resolution λ20 *Appl. Phys. Lett.* **44** 651–3

[13] Betzig E, Lewis A, Harootunian A, Isaacson M and Kratschmer E 1986 Near-field scanning optical microscopy (NSOM) *Biophys.* **49** 269–79

[14] Baida F, Courjon D and Tribillon G 1992 Combination of a fiber and a silicon nitride tip as a bifunctional detector; first results and perspectives *Proc. NATO* **242** 71–8

[15] Spajer M, Courjon D, Sarayeddine K, Jalocha A and Vigoureux J M 1991 Microscopie en champ proche par réflexion *J. Physique* **1** 1–12

[16] Van Labeke D and Barchiesi D 1993 Probes for scanning tunneling optical microscopy: a theoretical comparison *J. Opt. Soc. Am.* A **10** 2193–201

[17] Courjon D, Sarayeddine K and Spajer M 1989 Scanning tunneling optical microscopy, *Opt. Commun.* **71** 23–8

[18] Reddick R C, Warmack R J and Ferrell T L 1989 New form of scanning optical microscopy *Phys. Rev.* B **39** 767–70

[19] Courjon D, Vigoureux J M, Spajer M, Sarayeddine K D and Leblanc S 1990 External and internal reflection near field microscopy: experiments and results *Appl. Opt.* **29** 3734–40

[20] Courjon D, Spajer M, Jalocha A and Leblanc S Near-field microscopy and optical tunneling detection Scanned Probe Microscopy, AIP Conference Proceedings 241 ed: K. Wikramasinghe, Santa Barbara 1991

[21] Ohtsu M, Jiang S, Pangaribuan T and Kozuma M 1993 Nanometer resolution photon STM and single atom manipulation *Near Field Optics* ed D W Pohl and D Courjon (Dordrecht: Kluwer **242** 131–9

[22] Paesler M 1994 private communication; Paesler M and Moyer P 1994 to be presented at *OSA meeting October, Dallas, TX*

[23] Courjon D, Bainier C and Spajer M 1992 Imaging of submicron index variations by scanning optical tunneling *J. Vac. Sci. Technol.* **B 10** 2436–9

[24] Betzig A, Finn P L and Weiner J S 1992 Combined shear force and near-field scanning optical microscopy *Appl. Phys. Lett.* **60** 2484–6

[25] Toledo-Crow R, Yang P C, Chen Y and Vaez-Iravani M 1992 Near-field differential scanning optical microscope with atomic force regulation *Appl. Phys. Lett.* **60** 2957–9

[26] Bielefeldt H, Horsh I, Krausch G, Lux-Steine M, Mlynek J and Marti O 1994 Reflection scanning near-field optical microscopy and spectroscopy of opaque samples, *Appl. Phys.* A in press

[27] Van Hulst N F, Moers M H P, Noordman O F J, Faulkner T, Segerink F B, Van der Werf K O, de Grooth B G and Bölger B 1992 Operation of a scanning near field optical microscope in reflection in combination with a scanning force microscope *Scanning probe microsc.* **1639** 36–43

[28] Piednoir A, Creuzet F, Licoppe C, de Fornel F 1993 First specifications of a PSTM working in the infrared *Near Field Optics* ed D W Pohl and D Courjon (Dordrecht: Kluwer) **242** 309–16

[29] Paesler M A, Moyer P J and Jahncke C J 1990 Analytical photon scanning tunneling microscopy *Phys. Rev.* B **42** 6750–3

[30] Keller O, Xiao M and Bozhevolnyi S 1993 Configurational resonances in optical near-field microscopy: a rigorous point-dipole approach *Surf. Sci.* **280** 217–30

[31] Ghaemi H, Cates C and Goldberg B 1994 Low temperature near field spectroscopy and microscopy *2nd Conf. on Near Field Optics. Ultramicrosc.* in press

[32] Pedarnig J P, Specht M, Heckl W M and Hänsch T W 1993 Scanning plasmon near-field microscope *Near Field Optics* ed D W Pohl and D Courjon (Dordrecht: Kluwer) **242** 273–80

[33] Gleyzes P, Bachelot R and Boccara C 1994 Near field optical microscopy using a metallic vibrating tip *2nd Conf. on Near Field Optics. Ultramicrosc.* in press

[34] *Proc. NFO1: Near Field Optics* 1993 ed D W Pohl and D Courjon (Dordrecht: Kluwer)

[35] *Proc. NFO2: 2nd Conf. on Near Field Optics. Ultramicrosc.* in press

[36] Specht M, Pedarnig J D and Hänsch T W 1994 Fluorescence quench microscope *Ultramicrosc.* to be published

[37] Betzig E and Chichester R 1994 Single molecules observed by near field scanning optical microscopy *Science* **262** 1422–5

[38] Van Labeke D and Barchiesi D 1993 Theoretical problems in scanning near-field optical microscopy *Near Field Optics* ed D W Pohl and D Courjon (Dordrecht: Kluwer) **242** 157–78

[39] Martin O, Dereux A and Girard C 1994 Iterative scheme for computing exactly the total field propagating in dielectric structures of arbitrary shape *J. Opt. Soc. Am.* **11** 1073–80

[40] Girard C and Dereux A 1994 Optical spectroscopy of a surface at the nanometer scale: theoretical study in real space *Phys. Rev.* B **49** 11344–51

[41] Betzig E, Chichester R J, Lanni F and Taylor D L 1993 Near-field fluorescence imaging of cytoskeletal actin Bioimaging **1** 129–35

[42] Castiaux A, Dereux A, Vigneron J P, Girard C and Martin O J F 1994 Electromagnetic fields in two dimensional models of near field optical microscope tips *Proc. 2nd Conf. on Near Field Optics. Ultramicrosc.* in press

[43] Heinzelmann H, Hecht B, Novotny L and Pohl D W 1994 Scanning near field optical microscopy *Appl. Phys.* A **58** in press