Pergamon

Progress in Surface Science, Vol. 56, No. 3, pp. 133–237, 1997

0079-6816/98 $19.00

PII: S0079-6816(98)00004-5

IMAGE FORMATION IN NEAR-FIELD OPTICS

JEAN-JACQUES GREFFET* and RÉMI CARMINATI†

Laboratoire d'Energétique Macroscopique et Moléculaire, Combustion.
Ecole Centrale Paris, Centre National de la Recherche Scientifique,
92295 Châtenay-Malabry Cedex, France

Abstract

An overview is presented of the image formation theory in near-field optical microscopy. The emphasis is placed on the basic concepts and the understanding of the images. We briefly recall the general principles used in near-field optics to break the resolution limit. Since some of the concepts widely used in optics become meaningless in near field, a brief critical review of basic concepts is given. A theory of scattering of electromagnetic waves by inhomogeneous surfaces is then presented. For objects much smaller than the wavelength, a closed-form expression of the scattered field is derived, which provides a link between the near field and the structure of the sample. The different set-ups and their imaging capabilities are analysed. A general relationship between the signal and the induced currents in the sample is derived by means of the reciprocity theorem. The set-ups are compared and an equivalence between illumination and collection mode is proven. It is shown that, when multiple scattering between the sample and the rest of the system can be neglected, an impulse response can be defined for the three different types of set-ups : illumination mode, collection mode and apertureless. The importance of coherence in the near field is studied. Finally, the influence of the different control modes (constant height, constant intensity, constant tip-sample distance) is analysed and the existence of artifacts is discussed.

*E-mail : jjg@em2c.ecp.fr
†E-mail : remi@em2c.ecp.fr

Contents

1. Introduction 136

2. Basic Concepts I 138

A. Resolution limit 138

B. Resolution limit and evanescent waves 139

C. Resolution limit and Heisenberg uncertainty principle 140

3. How to Break the Fundamental Limit ? 142

A. Synge's idea 142

B. Near-field detection 144

C. Apertureless techniques 145

D. Role of evanescent waves 146

E. Interpretation of images 147

4. Basic Concepts II 147

A. Dipole radiation, near field and electrostatic limit 148

B. Optical index and reflection factors are meaningless in near field 148

C. Polarization effects and electric dipole 150

D. Concept of phase in near field 154

5. Near Field Scattered by an Inhomogeneous Sample 156

A. Formulation of the problem 157

B. Equivalent surface profile 159

C. Expression in Fourier space 161

D. Intensity versus equivalent surface profile 161

E. Transition from near field to far field 163

F. Polarization effects 165

6. Equivalence of the Illumination and Collection Modes of the SNOM 166

A. Reciprocity theorem 167

B. Reciprocity with extended incoherent sources and extended optical detectors 168

C. Application to SNOM set-ups 170

D. Role of coherence 171

E. Conclusion 172

7. Image Formation 172

A. Collection mode 173

B. Illumination mode 181

C. Apertureless set-up 185

D. Conclusion 190

8. Influence of Coherence 191

A. Near field and coherence 192

B. Transfer function and coherence 198

9. Operating Modes and Artifacts 200

A. Three operating modes 200

B. Equivalence of constant-height and constant-intensity images 201

C. Constant-distance mode and z-motion artifacts 206

10. Concluding Remarks 215

Appendices 216

References 227

Acronyms

AFM	Atomic-Force Microscope or Microscopy
NFO	Near-Field Optics
NSOM	Near-field Scanning Optical Microscope or Microscopy
PSTM	Photon Scanning Tunneling Microscope or Microscopy
SNOM	Scanning Near-field Optical Microscope or Microscopy
STM	Scanning Tunneling Microscope or Microscopy
STOM	Scanning Tunneling Optical Microscope or Microscopy
TE, TM	Transverse Electric, Transverse Magnetic
TNOM	Tunneling Near-field Optical Microscope or Microscopy

1. Introduction

In the mid eighties, the recently developed scanning tunneling microscope (STM) had just opened the way to microscopy at the nanometer scale, and most of the interest of the scientific community was concentrated on this new technique. Nevertheless, taking advantage of the advances in nanotechnology initiated by scanning tunneling microscopy, two different groups recorded optical images with subwavelength resolution [1, 2, 3, 4]. A historical survey of the first steps in scanning near-field optical microscopy can be found in the introduction of ref.[5].

At the beginning of the nineties, some new techniques appeared, leading to optical images with a resolution beyond the diffraction limit. We will briefly review the different techniques in §3. A more comprehensive presentation can be found in the proceedings of the near-field optics conferences [5, 6, 7, 8, 9] and in a few review articles [10, 11, 12, 13]. The scanning near-field optical microscope (SNOM) became a more common instrument, and was used in many applications :

- To image surfaces at a nanometer scale, using all the various sources of contrast provided by *optical* microscopy [14].

- For the detection and measurement of confined electromagnetic fields, such as surface plasmon polaritons [15, 16, 17, 18, 19, 20, 21, 22], localized surface plasmons polaritons [23, 24, 25, 26], guided waves [27], modes of cavity semiconductor lasers [28] or resonant modes in a spherical microcavity [29].

- For a local spectroscopy of surfaces [30, 31, 32], or a local collection of the emission of fluorescent samples [33, 34, 35, 36, 37], opening the way to single-molecule detection [38, 39].

- For the modification of surface properties [30], nanowriting [40], ignition of local photochemical processes [41] or modification of magneto-optics domains [42, 43], with an application in high-density data storage.

At the same time, the first theoretical studies were presented. The basic principles of near-field optics were analyzed by Vigoureux *et al.* [44, 45, 46]. The first numerical studies were reported by Girard and Courjon, based on a coupled-dipoles method [47], and by van Labeke

and Barchiesi, based on a pertubative method [48]. A state of the art of the theory of near-field optics in 1993 can be found in ref.[49]. A more recent review article presents an overview of the theories developed for the modelling of near-field experiments [50].

Whatever the application, near-field optical microscopy is based on the interaction of an electromagnetic field with nanometric structures. In order to understand the images, it is necessary to understand the basic mechanisms of the image formation. The problem is *a priori* extremely difficult. One has to calculate the electromagnetic field in confined subwavelength geometries, with low symmetry, and in which multiple scattering can take place. The ray optics and scalar physical optics approximations are no longer valid. One has to use a full vectorial electromagnetic formalism. The problem is even more complicated because some useful concepts in classical microscopy are meaningless in near-field optics. For example, we shall see later that the concept of reflection factor does not make sense, when one deals with nanometric structures. In fact, the basic mechanism is not reflection, but scattering. Thus, one has to change a few habits, and replace some common concepts in classical microscopy by new ones. Here lies one of the major goal of the theoretical analysis. Understanding near-field images cannot be done only by numerical simulations. The work presented in the present review is oriented in this direction. Since a complete survey of near-field issues and numerical techniques was recently published [50], we shall not repeat it, but instead focus on approximations and analytical work. Indeed, analytical theories are needed in order to identify basic mechanisms, and also to introduce some general tools and concepts. For example, we shall see how a general relationship between the signal and the sample can be derived for each type of near-field set-up, using reciprocity theorem. We shall also see under what condition the concept of impulse response (or transfer function) can be introduced. Such relationships and tools are necessary in order to compare the potential capabilities of the different techniques currently available.

The review is organised as follows. In §2 to §4, we review the basic concepts of near-field optics, and the different techniques that are used to produce images with subwavelength resolution. In §5, we study the near field scattered by an inhomogeneous surface. We essentially solve a scattering problem, and put forward the role of topography and dielectric contrast. In §6, we use the reciprocity theorem of electromagnetism to demonstrate the equivalence between the illumination and collection modes of the SNOM. In §7, starting also with the reciprocity

theorem, we derive a relation between the signal in each technique and the sample structure. We also examine the conditions under which an impulse response may be introduced to describe the image formation. In §8, we study the role of the coherence of the illuminating field on the image formation. We give in this section a description of the speckle pattern in the near field. Finally, we study in §9 the influence of the operating mode used to control the tip-sample distance, with a particular attention to the presence of artifacts. In §10, we present our concluding remarks.

2. Basic Concepts I

A. Resolution limit

The resolution in classical optical microscopy has been limited for two centuries by a fundamental limit, called the *diffraction limit*. Near-field optical microscopy was introduced ten years ago as a technique providing optical images with a resolution beyond this limit. Before discussing how to overcome the diffraction limit, let us examine its fundamental origin. To do so, we use two different points of view. Consider a classical optical microscope using an objective to collect the light coming from a sample. For the sake of simplicity, we assume that the sample is a perfectly conducting sinusoidal grating, with a period d, illuminated by an s-polarized plane wave. The objective is located at a distance which is typically a few millimeters (i.e., thousands of wavelengths), as represented schematically in Fig. 1.

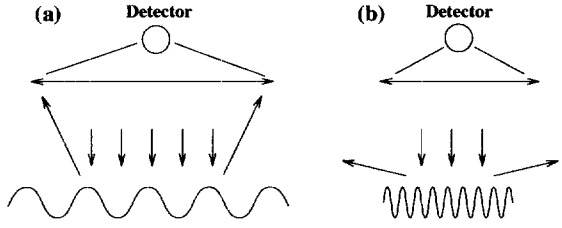

Fig. 1: *Classical detection of the light scattered by a grating. (a) Diffracted orders are detected by the objective. (b) When the period decreases, the diffracted orders are scattered at grazing angles and finally disappear. Whatever the angle of incidence, the objective of the microscope cannot detect scattered waves if the period d is smaller than $\lambda/2$.*

The question of resolution asks : " What is the smallest structure that can be seen by the microscope ?" We consider, as a test sample, a sinusoidal grating with period d. The question becomes : "Is there any information on the surface in the diffracted light ?" In reply, we simply use the grating law :

$$\frac{2\pi}{\lambda}\sin(\theta_s) = \frac{2\pi}{\lambda}\sin(\theta_i) + n\frac{2\pi}{d}\ , \tag{2.1}$$

where λ is the wavelength, θ_i and θ_s the angles of incidence and detection, respectively, and n is an integer. For a period $d < \lambda/2$, no value of θ_s and θ_i can be found to satisfy (2.1), which means that no diffracted order can propagate away from the surface. Thus, *no information about the surface can reach the objective of the microscope.* We see that there is a fundamental limit of the conventional microscope : a grating with a period smaller than $\lambda/2$ cannot diffract light towards the objective. This does not mean that the field has no rapid variation. In fact, the field has to satisfy the boundary conditions at the interface (it has to vanish in the case of a perfect conductor). If $d < \lambda/2$, the field has subwavelength lateral variations close to the surface. But these rapid variations cannot propagate away from the surface, and reach the objective. They are localized close to the sample. These variations are accounted for by the so-called *evanescent waves.*

B. Resolution limit and evanescent waves

To discuss the root of the fundamental limit of resolution more precisely, we now use a formal description of the field. The electric field in a plane $z = z_0$ above the highest point of the sample surface can be represented by its angular spectrum, or plane-wave expansion (a temporal dependence $\exp(-i\omega t)$ is assumed and omitted throughout the text) :

$$\mathbf{E}(\mathbf{r}_{||}, z_0) = \int \mathbf{e}(\mathbf{k}_{||}) \exp(i\mathbf{k}_{||}\cdot\mathbf{r}_{||} + i\gamma z_0)\, d\mathbf{k}_{||}\ , \tag{2.2}$$

where $\mathbf{r}_{||} = (x, y)$, $\mathbf{k}_{||} = (k_x, k_y)$, $k_0 = \omega/c$ and $\gamma = (k_0^2 - \mathbf{k}_{||}^2)^{1/2}$, with the determination $\mathrm{Re}(\gamma) > 0$, $\mathrm{Im}(\gamma) > 0$. This expression of the field is established in Appendix A.

In (2.2), the field is represented as a superposition of plane waves, each of them having a complex amplitude $\mathbf{e}(\mathbf{k}_{||})$ and a wave vector $\mathbf{k} = (\mathbf{k}_{||}, \gamma)$. In terms of Fourier analysis, $\mathbf{e}(\mathbf{k}_{||})\exp(i\gamma z_0)$ is the Fourier transform of the field in the plane $z = z_0$. For low spatial frequencies $|\mathbf{k}_{||}| \leq k_0$ (corresponding to the smooth lateral variations of the field in the plane

$z = z_0$), γ is real and the exponential is only a phase factor. The corresponding plane waves are *homogeneous*, and propagate away from the surface, in the direction defined by the wave vector $\mathbf{k}$. For high spatial frequencies $|\mathbf{k}_{||}| > k_0$ (corresponding to the subwavelength lateral variations of the field in the plane $z = z_0$), γ is purely imaginary and the exponential is an attenuation factor $\exp[-\mathrm{Im}(\gamma)z_0]$. The corresponding plane waves are *evanescent*. They propagate along the xy plane, and are attenuated exponentially in the z direction. Thus, they disappear at large distance, and cannot be detected by the objective of a classical microscope. For spatial frequencies much larger than k_0, $\mathrm{Im}(\gamma) \sim |\mathbf{k}_{||}|$ and the attenuation factor is $\exp(-|\mathbf{k}_{||}|z_0)$. Thus $1/z_0$ *appears as a cutoff frequency.* Similarly, for a given spatial frequency $\mathbf{k}_{||}$, $1/\mathbf{k}_{||}$ gives the decay length of the associated evanescent wave. For instance, if we try to observe the near field produced by a sinusoidal grating with a period of $\lambda/10$, the decay length is roughly $\lambda/20\pi$! If we use a He-Ne laser source to illuminate the grating, the decay length is 10 nm for the amplitude !

This comment suggests a first definition for the "near field". The near field is not only the region where the far-field approximation is not valid. In the context of superresolution, the near field is the region where the evanescent waves contribute significantly to the field. As a rule of thumb, *to detect an object with a resolution d, the tip-surface distance should be smaller than* $d/2\pi$, which defines the extension of the near field.

C. Resolution limit and Heisenberg uncertainty principle

A parallel can be established between the diffraction limit and Heisenberg uncertainty principle. Nevertheless, the obtention of subwavelength resolution by detection of the near field *is not* in contradiction with this principle. A nice discussion of this topic can be found in a paper by Vigoureux and Courjon [46].

Let us consider Heisenberg uncertainty principle for position and momentum, for a one dimensional problem. The wave function $\Psi(x)$ of a quantum particle can be written in a momentum representation :

$$\Psi(x) = \frac{1}{\sqrt{2\pi\hbar}} \int \check{\Psi}(p) \exp(i\frac{p}{\hbar}x)\, dp \tag{2.3}$$

where p is the momentum along the x-axis. The probability density of the particle position

is given by $|\Psi(x)|^2$, whereas the probability density of the particle momentum is $|\tilde{\Psi}(p)|^2$. The uncertainty Δx on the position is the width of $\Psi(x)$, the uncertainty Δp on the momentum being the width of $\tilde{\Psi}(p)$. Heisenberg uncertainty principle states that :

$$\Delta x \, \Delta p \geq h \ . \tag{2.4}$$

Equation (2.3) has the form of a Fourier transform, and (2.4) expresses a general relationship between the width of a function and the width of its Fourier transform. The same kind of relationship may be obtained for the electromagnetic field.

A good starting point is the angular spectrum (2.2). Considering the angular spectrum as a two-dimensional Fourier transform in the xy plane, the widths Δx and Δy of the field in direct space and the widths Δk_x and Δk_y of its Fourier transform are related by :

$$\begin{aligned} \Delta x \, \Delta k_x &\geq 2\pi \ , \\ \Delta y \, \Delta k_y &\geq 2\pi \ . \end{aligned} \tag{2.5}$$

Relations (2.5) find a simple interpretation : in order to improve the lateral resolution (by minimizing Δx and Δy), one has to increase the uncertainty on the direction of propagation of the light that is collected, which amounts to increasing the aperture of the objective.

In classical microscopy, one only detects the homogeneous waves, corresponding to a real value of the wave vector γ in the z direction. The maximum spatial frequency $\mathbf{k}_{\|}$ that can be detected is such that $|\mathbf{k}_{\|}| = k_0 \sin(\theta_{max})$, where $\sin(\theta_{max})$ is the numerical aperture NA of the objective. Using (2.5), one obtains the resolution limit in the lateral directions :

$$\begin{aligned} \Delta x &\geq \lambda/(2NA) \ , \\ \Delta y &\geq \lambda/(2NA) \ . \end{aligned} \tag{2.6}$$

We recover the diffraction limit $\lambda/2$, that would be obtained by an objective with an arbitrary large aperture, placed in a vacuum ($NA = 1$). Equations (2.6) are in agreement with Heisenberg uncertainty principle, and the existence of a diffraction limit is often presented as a consequence of this principle. A question arises immediately : is the idea of subwavelength resolution in contradiction with the uncertainty principle ?

Assume that we are able to detect the evanescent waves of the angular spectrum, corresponding of a purely imaginary value of γ. Then, the widths Δk_x and Δk_y may become as

large as desired (and larger than $2k_0$). Thus, the lateral resolution described by $\Delta x = 2\pi/\Delta k_x$ and $\Delta y = 2\pi/\Delta k_y$ may become as small as possible (in particular, smaller than $\lambda/2$). This possibility of increasing arbitrarily the lateral resolution depends on the ability of collecting the light corresponding to imaginary values of γ. As we have seen before, this light is contained in the evanescent waves of the angular spectrum, and remains localized close to the sample surface. Thus, the widths Δk_x and Δk_y that are actually achievable will depend on the distance of detection. Nevertheless, as soon as a detection of the evanescent waves is possible, the idea of subwavelength resolution is not in contradiction with Heisenberg uncertainty principle [46].

3. How to Break the Fundamental Limit ?

A. Synge's idea

In a (now famous) paper of 1928, Synge suggested a method to obtain optical images beyond the diffraction limit [51]. The idea is conceptually simple. If one realizes a light source with dimensions much smaller than the wavelength and brings this source close to the sample, a subwavelength spot will illuminate the surface. Scanning the source (and thus the light spot) along the surface and recording the transmitted (or reflected) light produces an image with subwavelength resolution [see Fig. 2(a)]. In principle, one realizes such a source by creating a very small aperture in an opaque screen.

An interesting discussion on superresolution was presented by McCutchen in 1967 [52], who put forward the technical difficulties associated with the detection of evanescent fields. The first experimental application of Synge's idea was reported in 1972 by Ash and Nicholls, and realized in the microwave range, for obvious experimental reasons [53]. In the mid eighties, Massey presented a theory and an experiment of an aperture near-field microscope in the microwave range [54], and pointed out that, with the technical progress made by the recent development of the Scanning Tunneling Microscope (STM), handling a source at nanometric distance from the sample was possible. Thus, the same experiment with visible light could be realized. Almost at the same time, the first experiments in the visible range were reported by Pohl *et al.* [1] and Lewis *et al.* [2].

One of the major difficulty in this technique is the production of a source with sufficient

output power. Another difficulty is to keep the nanosource as close as possible to the sample surface. Thus, the use of a small hole in an almost flat screen [2, 55] had to be replaced by the use of a metal-coated tip with an aperture [1, 3, 4].

Two major advances have been reported in the early nineties. Nowadays, the nanosource is usually an aluminum coated, adiabatically tapered, single mode optical fiber [30]. The control of the distance between the aperture and the sample has been solved by the introduction of a distance regulation mechanism. One can use an electronic tunneling current, if the sample is metalized [1]. The shear-force regulation has been largely used since its introduction by Toledo-Craw *et al.* [56] and Betzig *et al.* [57]. The basic idea is to induce a lateral oscillation of the tip. By detecting the amplitude of the oscillations, while approaching the surface, short-range interactions between the surface and the tip become detectable. This provides a means of avoiding contact between the tip and the surface, and enables the tip to be kept at a short distance during the scan, generally estimated to be less than 10 nm. The introduction of this technique has solved, to a large extent, the experimental difficulties associated with the scanning. It was a breakthrough and the starting point of a large development of the near-field community. Yet, the exact mechanism of the interaction is not well understood, so that it is not obvious that this regulation is really able to provide a constant-distance regulation [58]. Moreover, the shear-force regulation may be sensitive to the optical properties of the sample [59].

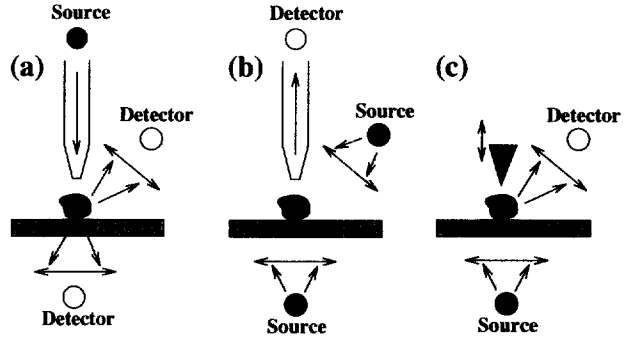

Fig. 2: *Different types of near-field optical microscopes. (a) Illumination mode. (b) Collection mode. (c) Apertureless.*

Recently, a new concept of efficient illuminating tip has been demonstrated. It is based on the propagation of light along one edge of a tetrahedral tip. The light is scattered at the corner and produces a very localized spot [60, 61]. This technique seems to be an interesting alternative to the aperture probe.

In the following, we denote all the techniques using a local illumination of the sample and a far-field detection by *illumination-mode* SNOM. Examples of such techniques are the illumination-mode SNOM initially proposed [1, 2, 3, 4], the TNOM, in which the transmitted light is collected above the critical angle of the substrate [62] or the SNOM using the tetrahedral tip [60].

B. Near-field detection

An alternative to the local illumination is the *local detection*. Having the reciprocity theorem in mind, it appears natural to try to exchange the position of the detectors and sources [see Fig. 2(b)]. In view of the discussion in §2, this could be the most natural way of reaching superresolution : one collects the light in the near field, in order to get the information contained in the evanescent waves. Nevertheless, this technique was developed historically a few years after the illumination-mode technique.

The first experiments were reported by Betzig *et al.* [63], Vaez-Iravani *et al.* [64] and Buckland *et al.* [65]. The idea was to reverse the light path in the illumination-mode SNOM. The sample was illuminated through a classical microscope objective, the detection of the near field being done by the tip of the optical fiber.

A slightly different approach to the superresolution was proposed by different groups in the late eighties [66, 67, 68]. The basic idea was to create an instrument by analogy with the STM. In STM, a metallic tip is scanned along a conducting sample. The tunneling current depends exponentially on the tip-sample distance. Keeping the tunneling current constant during the scan amounts to keeping the tip-sample distance constant. Recording the position of the tip enables the atomic topography of the sample to be obtained. In order to produce a field with an exponential decay, the basic idea was to illuminate the sample by total internal reflection. Indeed, an evanescent wave is then produced and can be detected by bringing an optical fiber with a small radius of curvature close to the surface. Such a tip can be produced by chemical

etching for instance. In view of the analogy with STM, this instrument was called Photon Scanning Tunneling Microscope (PSTM) [66, 67] or Scanning Tunneling Optical Microscope (STOM) [68]. An advantage of this set-up is that the regulation can be done by using the optical signal itself. Indeed, by bringing the tip close to the surface, the signal increases exponentially when the tip penetrates the evanescent field. The image is then produced by *the position of the tip while keeping the optical signal constant.* Nevertheless, despite its original historic development, the PSTM/STOM is a particular type of collection-mode instrument. The only peculiarity is the illuminating field, which is an evanescent wave created by total internal reflection

Another collection-mode technique has been proposed recently. The idea is to realize a nanometer-size optoelectronic detector, that can scan the near field [69, 70].

In the following, we refer to all those techniques, using a far-field illumination of the sample and a local detection of the near field, as *collection-mode* SNOM. Note that we include the PSTM/STOM in this category.

C. Apertureless techniques

A completely different type of instrument has been proposed more recently by Specht *et al.* [71, 72], Kawata *et al.* [73, 74], Gleyzes *et al.* [75], van Hulst *et al.* [76] and Wickramasinghe *et al.* [77]. The common point of these experiments is that they do not use an optical fiber to illuminate the sample or collect the near field. Indeed both the source and the detector are in the far field. The basic idea is to bring a small scattering tip into the near field, the sample being illuminated from the far field. The tip is illuminated by the near field produced by the sample, which contains subwavelength variations, and one detects *in the far field* the light scattered by the tip [see Fig. 2(c)]. Thus, *the tip acts as a small probe, which allows the near-field variations to be transferred into the far field.* By scanning the tip along the surface, one is able to record the near-field variations. In general, one uses a vertical dithering of the tip and a lock-in detection, in order to discriminate the signal from the background.

This technique cannot be simply classified as a collection-mode technique. Indeed, when the tip is brought very close to the surface, and illuminated with a suitable polarization, the light scattered by the tip also illuminates the sample. Under these conditions, the tip also acts

as a nanosource. This subject will be analysed in detail in §7.

All these techniques, using a scattering tip, are usually refered to as *apertureless* techniques. This name was given in opposition to the classical illumination-mode SNOM, which generally uses coated fiber tip with an aperture.

D. Role of evanescent waves

In §2, we introduced the concept of evanescent waves, which is a very useful concept in understanding the ability of the techniques described above to produce superresolution.

The superresolution capability of the illumination-mode SNOM comes from the high localization of the source. The field exiting the aperture is localized in a subwavelength area. When approaching the aperture to the sample, one illuminates the sample with evanescent waves. The closer is the tip, the more important is their contribution, and the more confined is the illumination. The sample, which contains subwavelength variations in topography and/or in optical properties, scatters this illuminating field. During the scattering process, some of the evanescent waves are converted into homogeneous (or propagating) waves, which are detected in the far field.

Concerning the collection-mode scheme, the role of evanescent waves is more easily understood in light of the discussion in §2. The light scattered by the sample (under far-field illumination) contains evanescent waves. These evanescent waves are collected by the tip of an optical fiber. In fact, the scattering process at the tip realizes a coupling between the near field and the propagating modes inside the fiber. Once again, some evanescent waves are converted into propagating waves (the modes) that are detected in the far field. The role of evanescent waves in the field scattered by the sample should not be confused with the possible evanescent wave in the illuminating field, as used in the PSTM/STOM configuration. One should realize that, from a theoretical point of view, the illumination by an evanescent wave is not necessary. The main reason for using an evanescent wave is to avoid a large background signal. Yet, *subwavelength information on the sample is only due to evanescent waves scattered by the sample.* These evanescent waves exist whatever the illuminating field.

The apertureless technique may work as a collection-mode instrument, or both as a collection-mode and an illumination-mode SNOM, depending on the tip-sample distance and the polariza-

tion of the incident light. We shall examine this point in §7. In the first case, some evanescent waves are created by the sample. These evanescent waves are converted into propagating waves by the tip. In the other case, a second mechanism has to be added. The tip also scatters the incident light, and creates evanescent waves, which illuminate the sample. The scattering process in the sample converts part of these illuminating evanescent waves into propagating waves, which also contribute to the signal.

E. Interpretation of images

Experimental images with subwavelength features have been obtained with all the above techniques. A major problem, in the field of scanning imaging, is to be able to understand the *contrast mechanisms*. On one hand, there is a real signal delivered by the detector that depends on the position of the tip (there is a contrast). On the other hand, there is a sample characterized by its topography and its optical properties (complex index, optical spectrum, Raman spectrum, fluorescence spectrum), an incident field on the sample characterized by its spatial and temporal coherence, its polarization and its power, and finally a detection system. The image is sensitive to all these parameters, so it is crucial to understand the effect of all of them on the image. We refer to this as the *image-formation process*, whose understanding is essential to improve the experimental set-ups, to compare their relative merits, and to learn how to read an image.

4. Basic Concepts II

The aim of this section is to review a number of standard concepts that will be useful in the analysis of the contrast mechanisms in near-field optics. It is important to know a number of basic results, in order to understand the features of the images. For instance, it is known that modifying the polarization affects drastically the images in near-field optics. Most of these polarization effects can be understood by having in mind the properties of the field created by an electrostatic dipole. The concept of phase is not trivial in the near field. Indeed, the meaning of "phase" in the *electrostatic limit* deserves a careful examination. It can also be discussed in simple terms using the dipole model. Finally, we point out that the impulse response, that will

be introduced later to account for the image formation process, is essentially the field radiated by a dipole.

A. Dipole radiation, near field and electrostatic limit

Let us consider a point dipole $\mathbf{p}$ situated at the origin of the coordinates, which radiates at a frequency ω. The expression of the field radiated by this dipole for $r > 0$ is :

$$\mathbf{E}(\mathbf{r}) = -\frac{k_0^2}{\epsilon_0} \overleftrightarrow{\mathbf{G}}_0 \cdot \mathbf{p} \ , \tag{4.1}$$

where $k_0 = \omega/c$, and the dyadic $\overleftrightarrow{\mathbf{G}}_0$ is given by

$$\overleftrightarrow{\mathbf{G}}_0 (\mathbf{r}) = \frac{\exp(ik_0 r)}{4\pi r} \left[\overleftrightarrow{\mathbf{I}} - \mathbf{u}_r \mathbf{u}_r - \left(\frac{i}{k_0 r} + \frac{1}{k_0^2 r^2} \right) (\overleftrightarrow{\mathbf{I}} - 3\mathbf{u}_r \cdot \mathbf{u}_r) \right] \ , \tag{4.2}$$

$\overleftrightarrow{\mathbf{I}}$ being the unit dyadic. Let us stress that $\overleftrightarrow{\mathbf{G}}_0$ is not the Green's dyadic, since it is not defined for r=0. Equation (4.2) is valid only for $r > 0$. By contrast, the Green's dyadic is defined for *any* value of r. Its expression is given in Appendix B, where we discuss the presence of a singularity at the origin and its implications. Equation (4.2) suggests a new definition of the near field. In near field, we are interested in variations of the field on length scales which are much smaller than the wavelength. It is seen in (4.2) that the leading terms are given by the terms varying as $1/r^3$. If one retains only these terms, the expression of the field coincides with that of an electrostatic dipole, which does not contain any retardation effects. We now have an alternative definition of the near-field region : *the region where the electrostatic contribution is the leading contribution.*

B. Optical index and reflection factors are meaningless in near field

An important consequence of the fact that the electrostatic limit is well suited to the description of the near field is that *the concepts of index and reflection factors become meaningless.* Indeed, the index is useful when describing the phase velocity and the attenuation of a wave. Obviously, this concept is not relevant when dealing with subwavelength distances. The index is also important when dealing with reflection, since it explicitly appears in the expression of the Fresnel factors. But the concept of reflection factor is also meaningless. Indeed, a small particle with a subwavelength size does not *reflect* but *scatters* light. When dealing with a surface larger

than the wavelength, each element of the surface scatters light and interferences between all the surface elements produces the *reflected field*, which is impossible for subwavelength structure. Thus, the intuitive concept of local Fresnel reflectivity is meaningless. Yet, both the index and the reflectivity have to be replaced by appropriate concepts. *The fundamental mechanism is scattering.* The basic concept is therefore the polarisability for a single molecule and the susceptibility (or equivalently the dielectric constant) for a mesoscopic sample.

The concept of local reflectivity can be misleading when trying to analyse the near-field signal. To illustrate this statement, we have plotted in Fig. 3 the square modulus of the electric near field (intensity), calculated [78, 79] at a constant height $z_0 = 4$ nm above a ridge of rectangular section, deposited on a glass substrate. It is seen that the intensity is smaller above a silver ridge (material with high reflectivity) than above a glass ridge (material with low reflectivity). Note that the physical explanation is simple in this case : in s polarization, the field is continuous at the ridge boundaries, and for a metallic ridge, the field vanishes inside. Another interesting feature in Fig. 3 is the contrast, which is much more important for the silver ridge, having the highest dielectric constant ϵ.

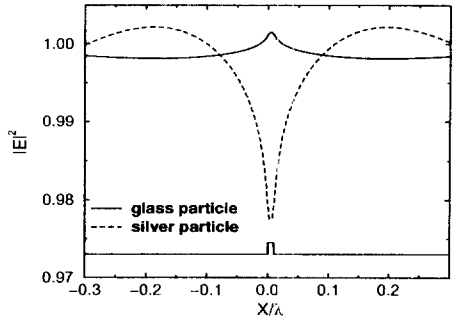

Fig. 3: *The concept of "local reflectivity" breaks down in near-field optics.* $|\mathbf{E}|^2$ *above a small ridge deposited on a flat glass substrate. Illumination in reflection by an s-polarized plane wave* ($\lambda = 633$ *nm) at nornal incidence. The ridge has a height height* $h = 2$ *nm and a width* $w = 6$ *nm. Solid line : glass particle (dielectric constant* $\epsilon = 2.25$*). Contrast enhanced by a factor 2. Dashed line : silver particle* ($\epsilon = -18 + i0.5$)*. An infinite silver surface has a higher reflectivity than an infinite glass surface. Nevertheless, the near-field intensity is smaller above the silver ridge.*

The presence of some nanometric structures will be the source of scattered light, which will interfere with the illuminating source to produce a modulation. It is important to realize that the result of this interaction cannot be anticipated on the basis of usual common sense. Indeed, the interference depends on the difference of phase. Since the phase depends on the components of the vectorial field (see Eq. 4.2), it is clear that one cannot assume that a small metallic structure will reflect more than the surrounding dielectric substrate.

We conclude that the concepts of local reflectivity and optical index are not appropriate to the analysis of the optical near field. Yet, an analogue of the reflectivity can be used in some cases. For instance, the field produced by an electrostatic dipole close to an interface separating two semi-infinite media is the sum of two terms [80]. The contribution of the image dipole can be seen as a reflection term. Note, however, that the *reflectivity* term is written in terms of dielectric constant and not with the index.

C. Polarization effects and electric dipole

Let us consider a point dipole with a polarizability α, illuminated by an incident wave with frequency ω. The dipole moment is given by $\mathbf{p} = \alpha(\omega)\mathbf{E}_{inc}$. When this dipole is a mesoscopic particle with radius a and dielectric constant $\epsilon(\omega)$, its polarizability can be estimated by the well-known formula

$$\alpha(\omega) = 4\pi\epsilon_0 a^3 \frac{\epsilon(\omega) - 1}{\epsilon(\omega) + 2} \tag{4.3}$$

The total field is the sum of the incident field and the field scattered by the dipole, denoted $\mathbf{E}_s$. It has been often assumed that the detected signal is the square modulus of the electric field, also called *intensity*. We will discuss, in detail, the link between the signal and the electric field in §7, which is actually more involved, but, for the sake of simplicity, we deal with the near-field intensity in this section :

$$I(\mathbf{r}) = |\mathbf{E}_{inc} + \mathbf{E}_s|^2 \simeq |\mathbf{E}_{inc}|^2 + 2\mathrm{Re}\{\mathbf{E}^*_{inc} \cdot \mathbf{E}_s\} \tag{4.4}$$

where Re denotes the real part and * denotes the complex conjugate. The structure of the intensity pattern produced by the dipole is given by the last term in (4.4). The result is plotted along a plane $z = z_0$ ($z_0 = 30$ nm) in Fig. 4, for two different orientations of the dipole. The calculation is performed in the *electrostatic limit*. In fact, we represent the real part of the dot

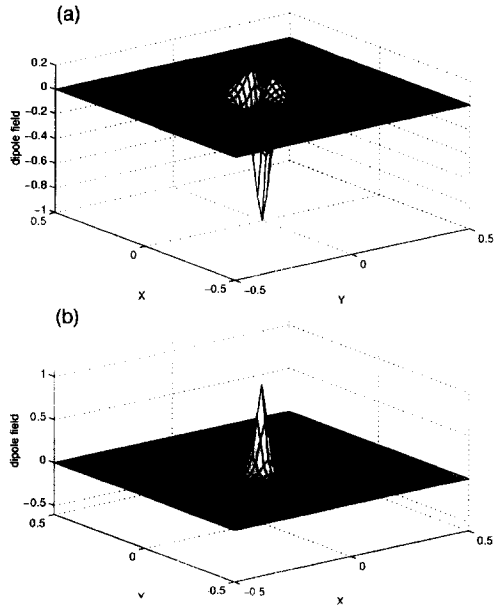

Fig. 4: *Near-field intensity produced by a point dipole in the electrostatic approximation (the intensity of the total field is represented, namely, the sum of the incident field and the field radiated by the dipole). (a) dipole parallel to the x-axis. (b) dipole perpendicular to the observation plane $z = z_0$. $z_0 = 30$ nm.*

product of the complex conjugate of the incident field and the field radiated by the electrostatic dipole. In Fig. 4(a), the dipole is parallel to the observation plane whereas in Fig. 4(b), the dipole is perpendicular to the plane. Let us first emphasize that this scattering pattern is very similar to the scattering pattern obtained for a small object close to an interface. For instance, we show in Fig. 5 the scattering pattern produced by a small glass cube with size $a = 6$ nm, deposited on a flat substrate and illuminated in total internal reflection. The calculation is performed through first-order perturbation theory (see §5). The observation distance is $z_0 = 30$ nm. Similar patterns have been computed by exact numerical simulations by Girard and Courjon[47] and Martin *et al.* [81].

Measurements have been reported recently by Weeber *et al.*[82] showing good agreement with theory. Upon comparison of Fig. 4 and Fig. 5, the first remark is that the image is

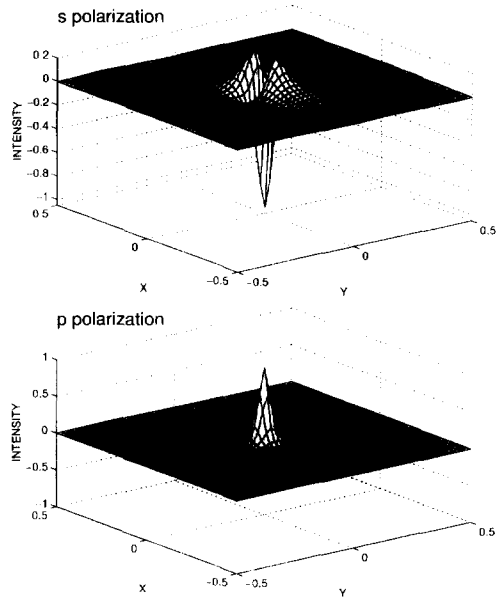

Fig. 5: *Near-field intensity (total field) scattered by a small glass cube (size $a = 6$ nm) deposited on a flat substrate. Illumination in total internal reflection ($\lambda = 600$ nm). The intensity is computed in a plane $z = z_0$, with $z_0 = 30$ nm.*

extremly sensitive to the polarization. A second feature is that the resemblance between the field scattered by the dipole and the field scattered by a small square is evident. We conclude that the polarization effects observed on small structures are a manifestation of the anisotropy of the field of an electrostatic dipole. It is therefore interesting to examine, in detail, the characteristics of the near-field intensity along a plane located at a distance z_0 above the dipole.

We observe in Figs. 4 and 5, several well-known features in near-field optical microscopy. First, upon comparison of Fig. 4(a) and Fig. 4(b), it is seen that the same object, namely, a point dipole, produces either a positive or negative contrast. This is related to the so-called *contrast inversion* obtained by changing the polarization, a result that has been very often reported [50]. Second, we observe in Fig. 4(b) that the image created by a point-like structure shows a typical camel shape for s-polarization. Thus, the link between the image and the object

is non-local. Thirdly, it is remarkable to note that the width is polarization dependent. Again, we retrieve a property that has been often observed in near-field : the resolution depends on the polarization.

We have mentionned a number of features that have been observed in near field. These features seems largely surprising, at first glance, because they do not have a counterpart in far-field imaging. Yet, they are related to elementary properties of the field of a dipole as we shall see. In order to discuss these properties, it will prove convenient to rewrite the tensor $\overleftrightarrow{\mathbf{G}}_0$ in a slightly different way. Let us denote x and y the coordinates in the observation plane and z_0 the distance between the plane and the dipole. For convenience, we note $r = \sqrt{x^2 + y^2 + z_0^2}$ the distance between the observation point and the dipole and $\rho = \sqrt{x^2 + y^2}$.

As seen from Eqs. (4.2) to (4.4), the intensity varies like the G_0^{zz} (G_0^{xx}) component of the tensor $\overleftrightarrow{\mathbf{G}}_0$ for an incident field linearly polarized along the z axis (x axis). From (4.2), it is easy to write explicitly these components. Keeping only the leading terms varying like $1/r^3$, we find :

$$-k_0^2 G_0^{zz} = \frac{\exp(ikr)}{z_0^3}\left[\frac{1}{1+\rho^2/z_0^2}\right]^{5/2}\frac{\rho^2 - 2z^2}{z_0^2}, \tag{4.5}$$

$$-k_0^2 G_0^{xx} = \frac{\exp(ikr)}{z_0^3}\left[\frac{1}{1+\rho^2/z_0^2}\right]^{5/2}\frac{y^2 - 2x^2 + z_0^2}{z_0^2}. \tag{4.6}$$

These elements of the tensor $\overleftrightarrow{\mathbf{G}}_0$ are the product of three terms. The first term gives the magnitude and varies like $1/z_0^3$ as expected. More interestingly, the second term provides the *width of the peak given by a Lorentzian* $(1 + \frac{\rho^2}{z_0^2})^{-1}$. Therefore, *the width of the peak is simply given by the plane-dipole distance* z_0. Finally, the third term accounts for the different shapes of the peaks. The G_0^{zz} term is rotationally invariant, whereas the G_0^{xx} term describes the peaks and dips seen in Fig. 4b. Note also that the signs are different at the center of the peak ($x = y = \rho = 0$) accounting for the so-called inverse contrast.

Coming back to the issue of imaging, it is clear that the intensity pattern produced by a dipole can be viewed as an impulse response. A good image is obtained if the impulse response is a symmetric and localized function. If it is not symmetric, then distorsions are introduced in the image. If the width increases, the resolution decays. As a result, it turns out that *the near-field intensity is localized around the dipole provided that one is able to detect the component of the field perpendicular to the scanning plane*, which is extremly important to optimize both

resolution and simplicity of the link between the image and the structure. Let us assume that the detection is done by bringing some point-like particle in the near field. Such a passive particle is assumed to be polarized by the near field and to scatter light in the far field. What is then the best experimental configuration ? We have just learnt that the best image is given by the vertical component of the field. This component induces a vertical component on the probe dipole, which, in turn, scatters light in the far field in the horizontal direction (i.e., perpendicularly to the induced probe dipole). Thus, *choosing a detection angle in the far-field amounts to selecting a polarization component in the near-field* ! The angle of detection is an analyser, as discussed in refs.[83, 84].

D. Concept of phase in near field

The issue of phase is of interest in the context of superresolution, since it has been argued that the phase could yield superresolution, even in the far-field [85, 86]. In an extensive discussion, Totzeck *et al.* showed that there is no possibility of having superresolution in the far field [87]. Indeed, the concept of cut-off frequency, introduced in §2, applies to the complex field, namely, to both its phase and its amplitude.

Recently, interferometric measurements have provided a way to record the phase of the near field, in the microwave regime [87, 88], and with visible light [77, 89]. These new kinds of near-field optical measurements are promising, since a nanometric resolution was obtained with the set-up of ref.[77, 90]. Thus, studying the phase in the near field might be interesting in order to understand these new experiments, and some theoretical studies have been reported [91, 92].

Phase conjugation of optical near fields has also been demonstrated [93], and has motivated some theoretical studies [94, 95]. This is another aspect of the phase properties of the near field.

(i) Phase and electrostatic limit

At first glance, introducing a phase in the near field seems to be contradictory to the statement that the near field is correctly described in the electrostatic limit. Indeed, in the electrostatic limit, the field is always a real quantity and the concept of delay is meaningless. To understand how the phase may come into play, we examine again (4.1)-(4.3). It is seen that two quantities

are complex : the polarizability, if the material of the small sphere is lossy, and the term varying as $1/r^2$. Thus, the dipole field is indeed complex and, therefore, we can define its amplitude and its phase.

Whether it is interesting to image the phase or the amplitude depends on the specific problem considered. A discussion of the potential resolution, that can be obtained by detecting the phase, can be found in refs.[91, 92]. Yet, it is clear that the rapidly varying terms are complex, so that the phase does certainly contain subwavelength information. In addition, it is seen in (4.2) that there is a relationship between the phase and the polarization. Indeed, the phase is not the same for all the components of the field. Therefore, imaging the phase rather than the modulus may be a means of choosing the near-field polarization, a point of importance, as we have just discussed. In fact, strong near-field polarization effects in the phase were put forward in ref. [92], leading to *phase-confinement effects*, similar to those observed in the intensity.

Another point deserves to be emphasized. When looking at the complex amplitude of the field scattered by a dipole, we adress the phase of the *scattered field.* In some cases, one is interested in the phase of the total field, namely, the sum of the incident and scattered field. It is important to clearly distinguish between these two quantities. Some experiments might produce a measurement of the phase of the scattered field, while other experiments might measure the phase of the total field.

(ii) Holograms in near field

To conclude this section, we address the issue of interference in near-field optics. The basic concept of interference stems from the angular spectrum. The field is represented as a sum of homogeneous (propagating) and inhomogeneous (evanescent) plane waves. The wave vectors have real components along the xy plane, and complex components along the z-axis. When looking at the field in the xy plane, we expect to observe interferences between the evanescent waves. Note that this is the essence of the angular spectrum.

Direct measurements of interferences have been reported recently by means of near-field microscopes operating in the collection mode [96, 89]. If interferences are possible, then, holography is also possible. Let us first show that holograms appear very naturally in near-field

optical microscopy. For instance, a collection-mode scheme, such as the PSTM/STOM, is based on the local detection of the near field. For the sake of simplicity, let us assume that the detected signal is the near-field intensity. Thus, (4.4) shows that the signal is basically the interference pattern between the illuminating wave and the scattered field. As discussed in refs.[97, 98], *the signal recorded when using a PSTM/STOM is a hologram of the object*, a fundamental remark. If both phase and amplitude of the field are recorded, all the information on the field is available. In principle, by inverse propagation, we can reconstruct the field at the sample position, and the sample properties. Numerical simulations of reconstruction have been reported [98, 99], showing a large potential for correction of the bias introduced by the scattering at the sample and the filtering associated with propagation. Experimental realization of near-field holograms have been reported [100]. We see that phase and interferences play a fundamental role in near-field imaging. Actually, interferences are important in the image formation mechanism of any SNOM instruments, as will be discussed in §7.

5. Near Field Scattered by an Inhomogeneous Sample

In this section, we describe the scattering of an incident field by a sample with both variations in topography and dielectric constant. We solve a pure problem of scattering by an inhomogeneous surface, without taking into account the presence of a SNOM probe. On the one hand, the description of the near field (or its intensity) is of interest in certain applications, as reflection of atoms on evanescent-wave mirrors [101, 102, 103, 104], study of surface waves, such as surface plasmon polaritons [15, 16, 17, 18, 19, 20, 21, 22], localized surface plasmons polaritons [23, 24, 25, 26] or guided waves [27]. On the other hand, studying the basic scattering mechanisms in the near field can be helpful in the analysis of the SNOM images, and a lot of theoretical work has dealt with a description of the near field itself, in the absence of the SNOM probe [50].

The near-field distribution around the sample results from a scattering problem, and depends on both the topography and the dielectric contrast of the sample. Dielectric contrast as a source of contrast for the SNOM images was demonstrated experimentally by various authors in PSTM/STOM [105, 106, 107], in illumination-mode SNOM [14, 64], in reflection SNOM [108]

and in apertureless SNOM [109, 110, 90]. Some comparisons between optical and topographical images were presented [64, 108], but some clear limits appeared : no quantitative relationship was put forward, and the correlations between the two kind of images seemed to depend on the technique that was used. In addition, a same object has completely different images when the illumination conditions, such as coherence, polarization and angle of incidence, are changed [111, 112]. The role of the dielectric contrast in forming the images was studied theoretically by numerical simulation [113, 114, 115, 81]. The importance of spectral resonance as a source of purely optical information has also been put forward [116]. Our aim is to give a closed-form expression of the near field scattered by the sample, so that the influence of the illumination parameters becomes transparent. In addition, we show that the field is influenced by topography and dielectric contrast through a coupled term. The model is based on a perturbation treatment of near-field scattering by an inhomogeneous sample [115]. Similar models have been developed in far-field scattering [117].

A. Formulation of the problem

(i) Geometry

We consider here the most general sample, described in Fig. 6.

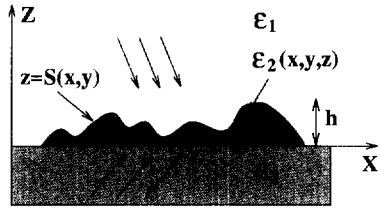

Fig. 6: *Scattering geometry.*

The lower half-space $z < 0$ (substrate) is filled with a homogeneous and isotropic material described by its dielectric constant ϵ_3. The upper half-space $z > S(x, y)$ is also homogeneous and isotropic, with a dielectric constant ϵ_1. The region $0 \leq z \leq S(x, y)$ is the real sample. It consists of a layer deposited on the substrate. The material is also isotropic, but inhomogeneous. It is described by a dielectric constant $\epsilon_2(x, y, z)$ and a surface profile (or topography) $z = S(x, y)$.

The system is illuminated from above (reflection geometry) or below (transmission geometry) by a monochromatic field of frequency ω, which can be a plane wave or any kind of beam. Our goal is to calculate the near-field distribution in the upper medium, and to relate this field to the properties of the sample [$S(x,y)$ and $\epsilon_2(x,y,z)$].

(ii) Exact integral equation for the field

From Maxwell's equations, in a Green-function formalism, an exact integral equation for the electric field can be derived [118] (see Appendix C for details) :

$$\mathbf{E}(\mathbf{r}) = \mathbf{E}^{(0)}(\mathbf{r}) + k_0^2 \int [\epsilon_2(\mathbf{r}') - \epsilon_1] \overleftrightarrow{\mathbf{G}}(\mathbf{r}_{||} - \mathbf{r}'_{||}, z, z') \, \mathbf{E}(\mathbf{r}') \, d\mathbf{r}' \,, \tag{5.1}$$

where $\mathbf{r} = (x,y,z)$ and $\mathbf{r}_{||} = (x,y)$. The expression for the Green dyadic $\overleftrightarrow{\mathbf{G}}$, which accounts for the presence of the substrate, is given in Appendix B. $\mathbf{E}^{(0)}$ is the field that would exist without the sample, for the flat suface between the substrate and the upper medium (illuminating field). The integral describes the scattered field, which contains the information of the sample. It is extended to the volume of the sample, namely, to all (x,y,z), such that $0 \leq z \leq S(x,y)$.

Equation (5.1) has the form of a Lippmann-Schwinger equation, and describes all the multiple interactions inside the sample, and between the sample and the substrate. For complex geometries, as used in NFO, it can, in general, only be solved numerically, and various schemes have been proposed [78, 79, 119, 120, 118]. Nevertheless, a lot of information may be obtained from a first-order approximation of this equation [115].

(iii) First-order expression

To derive an explicit expression of the scattered field, we make two simplifications : firstly, we use the first Born approximation ; secondly, we approximate the Green dyadic and the illuminating field by their values at $z' = 0$. We thus assume that their variations along the z-axis are small, in the range $0 \leq z \leq h$, where $h = \sup|S(\mathbf{r}_{||})|$, so that only terms to first order in h/λ are retained [121, 122, 115]. First-order approximations have been used by different authors in NFO [48, 123, 124, 97, 112, 125]. They lead to simple analytic expressions, that are useful for a physical discussion of the image formation. Roughly speaking, they are valid when $h \ll \lambda$ and/or the dieletric contrast $\Delta\epsilon = \sup|\epsilon_2(\mathbf{r}) - \epsilon_1|$ is low, which is usually the

case with the samples used in NFO. A precise study of the range of validity of the first-order approximation in the near field has been provided elsewhere [112, 115]. One obtains :

$$\mathbf{E}(\mathbf{r}) = \mathbf{E}^{(0)}(\mathbf{r}) + k_0^2 \int \overleftrightarrow{\mathbf{G}}(\mathbf{r}_{||} - \mathbf{r}'_{||}, z, 0)\mathbf{E}^{(0)}(\mathbf{r}'_{||}, 0)\,\langle \Delta\epsilon(\mathbf{r}'_{||})\rangle\, S(\mathbf{r}'_{||})\, d\mathbf{r}'_{||}\ , \tag{5.2}$$

where

$$\langle \Delta\epsilon(\mathbf{r}'_{||})\rangle = \frac{1}{S(\mathbf{r}'_{||})} \int_0^{S(\mathbf{r}'_{||})} [\epsilon_2(\mathbf{r}') - \epsilon_1] dz'\ . \tag{5.3}$$

Note that (5.2) describes the scattered field in terms of a *surface* integral in the xy plane. If one defines an equivalent *surface* current $\mathbf{j}_s = -i\omega\epsilon_0\langle\Delta\epsilon(\mathbf{r}'_{||})\rangle\mathbf{E}^{(0)}(\mathbf{r}'_{||}, 0)S(\mathbf{r}'_{||})$, the scattered field is equal to the field radiated by $\mathbf{j}_s$ [126, 127].

B. Equivalent surface profile

Instead of defining an equivalent surface current, we can define an equivalent surface profile S_{eq} by [99]

$$S_{eq}(\mathbf{r}'_{||}) = \frac{1}{\epsilon_3 - \epsilon_1} \int_0^{S(\mathbf{r}'_{||})} [\epsilon_2(\mathbf{r}') - \epsilon_1] dz'\ . \tag{5.4}$$

In the case of a homogeneous sample [$\epsilon_2(\mathbf{r}) = \epsilon_3$ for all $\mathbf{r}$], $S_{eq}(\mathbf{r}_{||})$ reduces to the true topographic profile $S(\mathbf{r}_{||})$. Using this new concept, (5.2) can be rewritten :

$$\mathbf{E}(\mathbf{r}) = \mathbf{E}^{(0)}(\mathbf{r}) + k_0^2(\epsilon_3 - \epsilon_1) \int \overleftrightarrow{\mathbf{G}}(\mathbf{r}_{||} - \mathbf{r}'_{||}, z, 0)\mathbf{E}^{(0)}(\mathbf{r}'_{||}, 0) S_{eq}(\mathbf{r}'_{||})\, d\mathbf{r}'_{||}\ . \tag{5.5}$$

Equation (5.5) has a simple interpretation : in the framework of first-order perturbation theory, the near field scattered by an inhomogeneous sample with variations in both topography and dielectric constant is the same as that scattered by a homogeneous sample of equivalent surface profile S_{eq}.

Equation (5.5) explicitely describes the influence of both topography and dielectric-constant variations on the near field. Three remarks can be made :

- The near field is sensitive to the *equivalent* surface profile, which is a coupled term between topography and dielectric contrast, which can be understood as an averaged dielectric contrast along the z-direction [see also (5.2)].

- Two different samples, with the same equivalent suface profile, produce the same near field, as is illustrated below by numerical simulation.

- For a given sample, a single measurement of the near field would not enable the topography to be distinguished from the dielectric contrast. In order to extract the dielectric contrast, it is necessary to measure the topography by some other means, and to solve the difficult inverse problem, as was examined in ref.[99]. Another possibility would be to perform a tomography.

(i) Numerical illustration

In order to assess the prediction of the perturbative approximation, we present numerical calculations based on a rigorous solution of the Lippman-Schwinger equation as described, for instance, in ref. [79]. Figure 7 displays the near-field intensity ($|\mathbf{E}|^2$) produced by different samples, with different dielectric contrasts and different heights, but with the same equivalent surface profile. It is clearly seen that the curves are identical. Thus, these different objects produce the same near field.

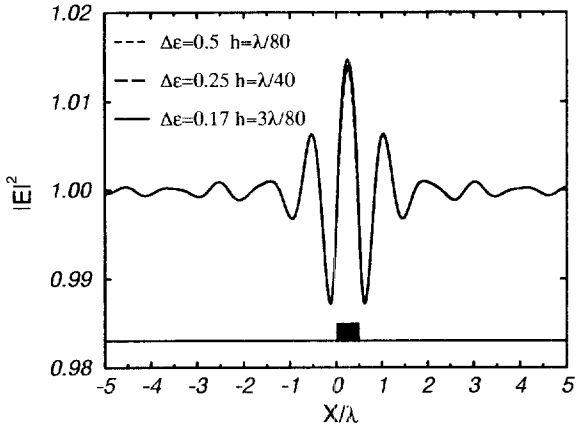

Fig. 7: *Square modulus of the electric field along a line at a height $z_0 = 40$ nm above a 2D surface patch of height h and dielectric contrast $\Delta\epsilon = \epsilon_2 - \epsilon_3$. The surface is illuminated in transmission at normal incidence ($\lambda = 633$ nm). The position of the patch is indicated at the bottom of the figure.*

C. Expression in Fourier space

We provide here a first-order expression of the field in Fourier space. For the sake of simplicity, we assume that the sample is illuminated in transmission by an incident monochromatic plane wave of amplitude $\mathbf{e}_{inc}$ and wave vector $\mathbf{k}_{||}^{inc}$ along the xy plane. The illuminating field $\mathbf{E}^{(0)}$ for $z = 0$ is :

$$\mathbf{E}^{(0)}(\mathbf{r}_{||}, 0) = \overleftrightarrow{\mathbf{T}}\,(\mathbf{k}_{||}^{inc})\,\mathbf{e}_{inc}\ , \tag{5.6}$$

where $\overleftrightarrow{\mathbf{T}}\,(\mathbf{k}_{||}^{inc})$ is a dyadic containing Fresnel transmission factors at the interface $z = 0$.

In the context of scattering from rough surfaces, several methods have been used to derive a first-order expression of the scattered field. One can use the Rayleigh hypothesis in order to write the boundary conditions accross the surface [128, 129, 130, 131]. Let us stress that the use of the Rayleigh hypothesis, in the context of perturbative solutions, *does not* limit the validity of the solution. In fact, the perturbative solutions obtained from the extinction theorem (without the Rayleigh hypothesis) yields the same result [132, 133, 134].

Another possibility is to start with (5.5), and introduce the Fourier decomposition of the Green dyadic [121, 122, 115]. Whatever the method, we obtain :

$$\mathbf{E}(\mathbf{r}) = \mathbf{E}^{(0)}(\mathbf{r}) + \int \tilde{S}_{eq}(\mathbf{k}_{||}^{inc} - \mathbf{k}_{||})\ \overleftrightarrow{\mathbf{L}}\,(\mathbf{k}_{||}, \mathbf{k}_{||}^{inc})\,\mathbf{e}_{inc}\,\exp(i\mathbf{k}_{||}\cdot\mathbf{r}_{||} + i\gamma z)\,d\mathbf{k}_{||}\ , \tag{5.7}$$

where $\tilde{S}_{eq}(\mathbf{k}_{||})$ is the spectrum of the equivalent surface profile defined by :

$$\tilde{S}_{eq}(\mathbf{k}_{||}) = \int S_{eq}(\mathbf{r}_{||})\exp(i\mathbf{k}_{||}\cdot\mathbf{r}_{||})\,d\mathbf{r}_{||}\ . \tag{5.8}$$

Note that (5.7) is identical to the first-order equations previously obtained with homogeneous rough surfaces [121, 122, 117, 131]. The use of the equivalent surface profile has allowed us to extend the result to inhomogeneous surfaces. The dyadic $\overleftrightarrow{\mathbf{L}}\,(\mathbf{k}_{||}, \mathbf{k}_{||}^{inc})$ in (5.7) connects the components of the incident field to that of the scattered field, and describes all the polarization effects. Its analytical expression is given in Appendix E. The same kind of expression in Fourier space, with a homogeneous sample, was obtained by van Labeke and Barchiesi [48, 135].

D. Intensity versus equivalent surface profile

We now turn to the description of the near-field intensity, defined as the square modulus of the electric field. We will see in §7 that the detected signal in SNOM is, in general, more complicated

than $|\mathbf{E}|^2$, except in particular situations [136, 137, 125]. Nevertheless, the knowledge of $|\mathbf{E}|^2$ is often useful for a first analysis of the SNOM images. It can also give some physical insight into the expected results of a given experiment. Other applications also justify the development of simple techniques to describe $|\mathbf{E}|^2$. For example, the knowledge of $|\mathbf{E}|^2$ is necessary in order to calculate the potential felt by atoms being reflected by optical evanescent fields [101, 102, 103, 104].

(i) Hologram structure

If we rewrite (5.7) in the form $\mathbf{E}(\mathbf{r}) = \mathbf{E}^{(0)}(\mathbf{r}) + \mathbf{E}^{(1)}(\mathbf{r})$, the intensity is :

$$I(\mathbf{r}) = \left|\mathbf{E}^{(0)}(\mathbf{r}) + \mathbf{E}^{(1)}(\mathbf{r})\right|^2 \simeq \left|\mathbf{E}^{(0)}(\mathbf{r})\right|^2 + 2\mathrm{Re}\left[\mathbf{E}^{(0)*}(\mathbf{r}) \cdot \mathbf{E}^{(1)}(\mathbf{r})\right] , \tag{5.9}$$

where we have kept the zero and first-order terms only. The first term on the left-hand side is the intensity of the illuminating field, $I^{(0)}$ When the illuminating field is a plane wave, $I^{(0)}$ is a function of z only. The second term, $I^{(1)}$, is an interference term between the illuminating field and the scattered field. Note that due to this interference, both the *amplitude* and the *phase* of the scattered field $\mathbf{E}^{(1)}$ are encoded in the intensity. Thus, the intensity has the structure of a *hologram*, $\mathbf{E}^{(0)}$ being the reference field [97, 98].

(ii) Impulse response

Equations (5.7) and (5.9) show that there is a linear relationship between the intensity and the equivalent surface profile. Let us assume that the illumination is done in transmission by a monochromatic plane wave of frequency ω and wave vector $\mathbf{k}_{\|}^{inc}$, and that the intensity is measured in a plane at a constant height z_0. One obtains from (5.7) and (5.9) :

$$I(\mathbf{r}_{\|}, z_0) = I^{(0)}(z_0) + \int H_c(\mathbf{r}_{\|} - \mathbf{r}'_{\|}, \mathbf{k}_{\|}^{inc}, \omega, z_0)\, S_{eq}(\mathbf{r}'_{\|})\, d\mathbf{r}'_{\|} . \tag{5.10}$$

H_c is an *impulse response*, which does not depend on the sample. Its Fourier transform (or transfer function)

$$\tilde{H}_c(\mathbf{k}_{\|}, \mathbf{k}_{\|}^{inc}, \omega, z_0) = \int H_c(\mathbf{r}_{\|}, \mathbf{k}_{\|}^{inc}, \omega, z_0)\, \exp(i\mathbf{k}_{\|} \cdot \mathbf{r}_{\|})\, d\mathbf{r}_{\|} . \tag{5.11}$$

has an analytical expression, given in Appendix E. We stress that the existence of the impulse response was not postulated, but arises naturally from the analysis through first-order perturbation theory (linearity), and from the fact that the illuminating intensity $I^{(0)}$ only depends on z (translational invariance in the xy plane).

The impulse response describes the non-local relationship between the intensity in a plane $z = z_0$ and the equivalent surface profile, and describes completely the scattering process. In particular, it contains the filtering and polarization effects, as illustrated in the following section. Note that the illumination by a monochromatic plane wave is fully coherent. Thus, H_c is an impulse response for coherent illumination. An impulse response for partially-coherent illumination will be introduced in §8.

E. Transition from near field to far field

We now use the impulse response to describe the filtering effect, when the distance z_0 from the substrate to the observation point increases. We show in Fig. 8 the impulse response H_c for an illumination in transmission at normal incidence, and for three distances of observation : (a) $z_0 = 0.05\lambda$, (b) $z_0 = 0.1\lambda$, (c) $z_0 = \lambda$. The substrate (medium $z < 0$) has a dielectric constant $\epsilon_3 = 2.25$. The upper medium is a vacuum ($\epsilon_1 = 1$). The incident field is polarized along Ox. All images in Fig. 8 have been normalized by their peak value.

Let us recall the meaning of the impulse response. If a small object is deposited on the substrate, the convolution product of the impulse response by the (equivalent) surface profile of the object gives the distribution of the intensity $I^{(1)}$ in the plane $z = z_0$. Figure 8 shows that the impulse response strongly depends on the observation distance z_0. Close to the substrate (Fig. 8a), the impulse response is anisotropic, with regions in positive and negative contrast. When z_0 increases, the impulse response becomes isotropic and looses its rapid variations.

The impulse response is also useful in discussing the resolution issue. Its width is a measure of the achievable resolution at a given height. Figure 8 shows that the width of the impulse response increases dramatically with z_0. Note that the scales on the three graphs are different. The width of the impulse response is $\simeq 0.1\lambda$ for $z_0 = 0.05\lambda$, while it is $\simeq \lambda$ for $z_0 = \lambda$. In this case (Fig. 8c), no variations on a scale smaller than λ is visible in the impulse response. This is already the far zone, where resolution is limited by diffraction.

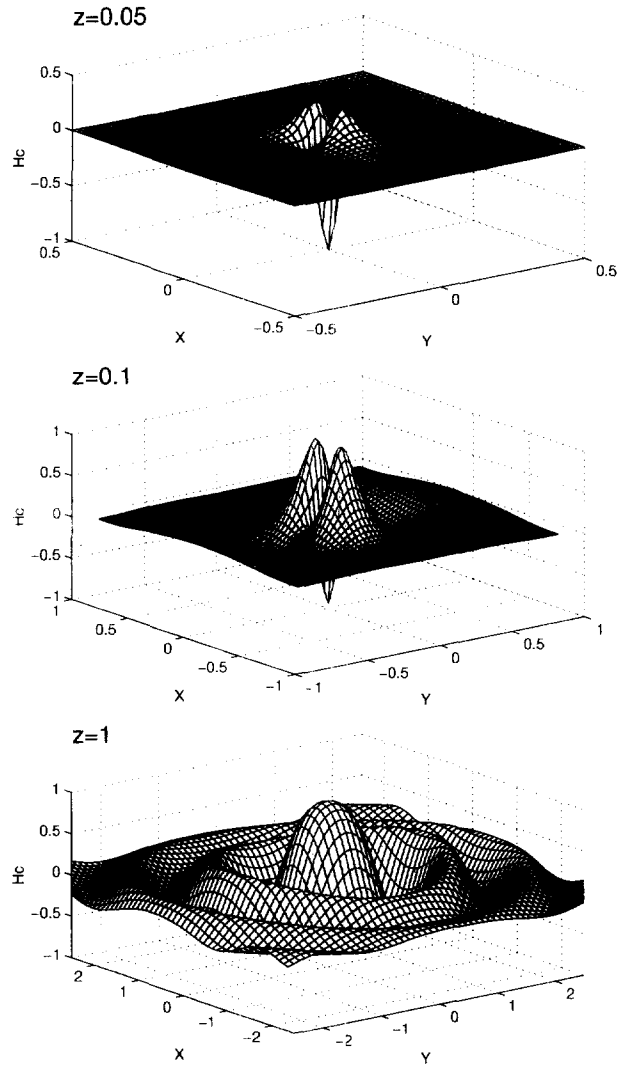

Fig. 8: *Impulse response versus the observation distance* z_0. *Normal incidence. The incident field is polarized along* Ox. *(a)* $z_0 = 0.05\lambda$, *(b)* $z_0 = 0.1\lambda$. *(c)* $z_0 = \lambda$.

F. Polarization effects

Several authors have put forward strong polarizations effects on 3D samples [135, 138, 139, 120, 82]. In particular, it has been shown that the intensity pattern closely followed the surface profile of small objects, when the illuminating field was p-polarized [139, 120, 82]. We discussed this observation in §4, and showed that its physical origin was the anisotropy of the electrostatic dipole. The impulse response, introduced above, is a convenient tool to describe the polarization effects, as we now proceed to show.

Figure 9 represents the impulse response calculated over a glass substrate ($\epsilon_3 = 2.25$) at a distance $z_0 = 0.05\lambda$. The illumination is done in transmission. The plane of incidence is yz. The angle of incidence is $\theta_i = 45^o$, which corresponds to an illumination in total internal reflection. In Fig. 9(a), the incident field is polarized along (Ox) (s-polarization). In Fig. 9(b), the incident electric field is in the plane of incidence (p polarization).

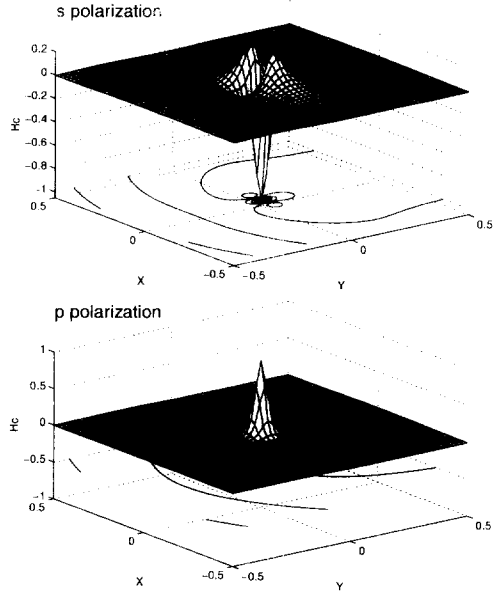

Fig. 9: *Impulse response for $\theta_i = 45^o$. $z_0 = 0.05\lambda$, plane of incidence $y - z$. (a) electric field along Ox, (b) electric field in the plane of incidence.*

The impulse response strongly depends on the polarization. In s-polarization, the impulse response is similar to that calculated at normal incidence (see Fig. 8a). It is anisotropic and presents regions in positive and negative contrast. In p-polarization, the impulse response is narrower and isotropic. Its convolution product with the sample profile gives rise to an intensity pattern which follows the sample profile.

From a physical point of view, the impulse response is the distribution of intensity created by a point-like object deposited on the substrate, and illuminated by a plane wave. Close to the object, the main contribution stems from the r^{-3} term of the dipole radiation, as discussed in §4. Note the similarity between the impulse response (Fig. 8) and the radiation of an electrostatic dipole (Fig. 4). As pointed out previously, the impulse response presents the same features as the images of small objects calculated in ref.[139] and measured in ref.[82]. In particular, the two peaks on each side of the impulse response in s-polarization are present in the images of small objects [139, 82].

6. Equivalence of the Illumination and Collection Modes of the SNOM

We compare here the illumination and collection modes of the SNOM. Because the development of each mode has its own history, and its own (sometimes intuitive) motivations, some advantages or hindrances of a particular mode are sometimes put forward. Nevertheless, no general answer to the question "are the illumination and collection modes fundamentally different or not ? " has been given so far.

A recent analysis by Méndez *et al.* [140] showed that, in principle, the illumination- and collection-mode instruments are identical, as far as their imaging properties and their potential resolution are concerned. We will closely follow this original analysis here. We shall show that the differences that have sometimes been reported are only due to the use of different instrumentation and operating conditions. Using the reciprocity theorem of electromagnetism, we shall demonstrate that, for a given collection-mode set-up, there exists an equivalent illumination-mode set-up, in the sense that it produces the same images. Let us stress that the use of reciprocity theorem is not straightforward, because optical detectors are sensitive to energy.

Thus, the coherence of the field has to be taken into account properly in the analysis. We also stress that applying the reciprocity theorem is more subtle than simply reversing the path of the light. In particular, reciprocity is not equivalent to time-reversal invariance (a simple argument is that reciprocity holds for lossy media).

A. Reciprocity theorem

Let us consider a scatterer, made of one or several bodies, occupying a region of space of boundary Γ (Fig. 10). The scatterer is assumed to be linear, with frequency-dependent constitutive tensors $\overleftrightarrow{\epsilon}(\mathbf{r},\omega)$ and $\overleftrightarrow{\mu}(\mathbf{r},\omega)$. The tensors $\overleftrightarrow{\epsilon}$ and $\overleftrightarrow{\mu}$ are assumed to be *symmetric*, which means that the scatterer may be inhomogeneous and anisotropic, but not optically active.

In a first situation [Fig. 10(a)], let $\mathbf{E}(\mathbf{r}_1,\mathbf{r}_0,\omega)$ be the complex amplitude of the electric field created at a point $\mathbf{r}_1$ by a dipole source of amplitude $\mathbf{A}_0(\omega)$ placed at a point $\mathbf{r}_0$, and radiating at a frequency ω. Similarly [Fig. 10(b)], let $\mathbf{E}(\mathbf{r}_0,\mathbf{r}_1,\omega)$ be the electric field created at $\mathbf{r}_0$ by a dipole source of amplitude $\mathbf{A}_1(\omega)$ placed at $\mathbf{r}_1$ and radiating at the same frequency ω. The reciprocity theorem, derived in Appendix D, states that [141] :

$$\mathbf{A}_0(\omega)\cdot\mathbf{E}(\mathbf{r}_0,\mathbf{r}_1,\omega)=\mathbf{A}_1(\omega)\cdot\mathbf{E}(\mathbf{r}_1,\mathbf{r}_0,\omega)\ . \tag{6.1}$$

It is the basis of many results presented in this review.

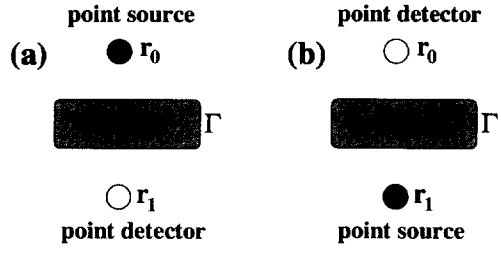

Fig. 10: *Geometry considered for the application of the reciprocity theorem, with a point source and a point detector.*

B. Reciprocity with extended incoherent sources and extended optical detectors

For reasons that will become clear later, we now extend the reciprocity theorem to the case of an extended incoherent source and an extended optical detector. Optical detectors respond to the time-averaged energy flux carried by the field. We consider detectors that are insensitive to the polarization of the light. We will also assume that they are *Lambertian detectors*, in the sense that the only dependence of the signal on the direction of incidence is through the cross-section they project in that direction. In the linear regime, the signal is the product of the incident intensity and a quantity $\mathcal{D}$, which characterizes the efficiency of the detector. If the detector is placed in the far field of the boundary Γ, the incident intensity is simply the square modulus of the electric field.

We consider planar, incoherent, polychromatic and unpolarized Lambertian sources. They may be visualized as a collection of random, isotropically orientated point dipole sources of amplitude $\mathbf{A}$, emitting polychromatic radiation. We assume that the fluctuations of the sources may be described by a statistical ensemble which is stationnary. The cartesian components of $\mathbf{A}$ are uncorrelated random variables, with equal variance and zero mean : $\langle A_j A_k^* \rangle = 0$ and $\langle |A_j|^2 \rangle = \langle |A_k|^2 \rangle$, for $j, k = x, y, z$, the brackets denoting an ensemble average.

Let us first assume that we have isolated one point of the source (alternatively $\mathbf{r}_1$ or $\mathbf{r}_0$), and one point of the detector (alternatively $\mathbf{r}_0$ or $\mathbf{r}_1$). The point of the source emits polychromatic radiation, and may be visualized as a superposition of uncorrelated monochromatic point sources, each of them emitting at a single frequency ω. For one of them, by taking the square modulus of (6.1), and averaging over the ensemble of point sources, we obtain

$$\mathcal{I}_0(\omega) I(\mathbf{r}_0, \mathbf{r}_1, \omega) = \mathcal{I}_1(\omega) I(\mathbf{r}_1, \mathbf{r}_0, \omega) \ , \tag{6.2}$$

where $I(\mathbf{r}_0, \mathbf{r}_1, \omega) = \langle |\mathbf{E}(\mathbf{r}_0, \mathbf{r}_1, \omega)|^2 \rangle$ is the intensity reaching the point detector at $\mathbf{r}_0$, due to a monochromatic point source of frequency ω and intensity $\mathcal{I}_1(\omega) = \langle |\mathbf{A}_1(\omega)|^2 \rangle$ at $\mathbf{r}_1$. In order to obtain a relationship involving the polychromatic point source, we simply add the intensities. This leads to :

$$\mathcal{I}_0 \, I(\mathbf{r}_0, \mathbf{r}_1) = \mathcal{I}_1 \, I(\mathbf{r}_1, \mathbf{r}_0) \ , \tag{6.3}$$

where $\mathcal{I}_0 = \int \mathcal{I}_0(\omega)\, d\omega$ and $I(\mathbf{r}_0, \mathbf{r}_1) = \int I(\mathbf{r}_0, \mathbf{r}_1, \omega)\, d\omega$, the integration being performed over the spectrum of the source.

In the first case [Fig. 10(a)], the point source at $\mathbf{r}_0$ radiates with an intensity $\mathcal{I}_0$, and the point detector at $\mathbf{r}_1$ produces a signal $S_1 = \mathcal{D}_1 I(\mathbf{r}_1, \mathbf{r}_0)$. In the second case [Fig. 10(b)], the point source at $\mathbf{r}_1$ radiates with an intensity $\mathcal{I}_1$ and the detectors at $\mathbf{r}_0$ measures a signal $S_0 = \mathcal{D}_0 I(\mathbf{r}_0, \mathbf{r}_1)$. Using (6.3), one shows that the two signal S_0 and S_1 are equal if

$$\frac{\mathcal{D}_0}{\mathcal{I}_0} = \frac{\mathcal{D}_1}{\mathcal{I}_1} \,. \tag{6.4}$$

Starting with the reciprocity relations (6.3) and (6.4) from a point in the source to another point in the detector, one can derive relations for an extended *incoherent* source and an extended optical detector (see Fig. 11).

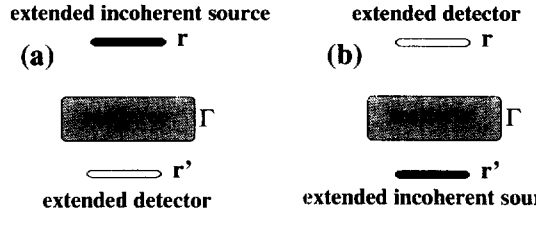

Fig. 11: *Geometry considered for the derivation of the reciprocity theorem with an incoherent source and an extended optical detector.*

The signal in each situation results from the incoherent superposition of the signal coming from each point of the source, and reaching each point of the detector :

$$S_1 = \int \mathcal{D}(\mathbf{r}) I(\mathbf{r}, \mathbf{r}')\, d\mathbf{r} d\mathbf{r}' \,, \tag{6.5}$$

$$S_2 = \int \mathcal{D}(\mathbf{r}') I(\mathbf{r}', \mathbf{r})\, d\mathbf{r}' d\mathbf{r} \,. \tag{6.6}$$

In the preceding equations, the response of the detector $\mathcal{D}(\mathbf{r})$ is assumed to be position dependent. The reciprocity relation (6.3) now gives

$$\mathcal{I}(\mathbf{r}') I(\mathbf{r}', \mathbf{r}) = \mathcal{I}(\mathbf{r}) I(\mathbf{r}, \mathbf{r}') \,, \tag{6.7}$$

where $\mathcal{I}(\mathbf{r})$ is the intensity across the source. The two signals S_1 and S_2 will be equal if

$$\frac{\mathcal{D}(\mathbf{r}')}{\mathcal{I}(\mathbf{r}')} = \frac{\mathcal{D}(\mathbf{r})}{\mathcal{I}(\mathbf{r})} \,. \tag{6.8}$$

Note that the simplest situation, in which (6.8) is satisfied, is when the source has a uniform intensity, and the efficiency of the detector is position independent.

The reciprocity relations (6.7) and (6.8) can be obtained with *unpolarized and incoherent* sources only [140]. Nevertheless, *this is not restrictive.* Indeed, reciprocity for any kind of sources and detectors can be obtained from these general relations [140, 142]. The reason is that it is always possible to add some optical filters, polarizers, pinholes... in order to modify the properties of the illumination and the detection. All these optical devices can be included in the scattering system, i.e., inside the boundary Γ. Thus, any scheme can be reduced to a scattering system enclosed by a boundary Γ, illuminated by an incoherent source, the output light being collected by an extended optical detector.

C. Application to SNOM set-ups

An important consequence of the reciprocity relations is that there is no fundamental difference between an illumination-mode and a collection-mode SNOM. More precisely, for any collection-mode set-up, it is possible to design an equivalent illumination-mode set-up, as we now proceed to show.

Let us consider the classical PSTM/STOM, which is a collection-mode instrument, represented in Fig. 12(a). The illuminating beam is, in this case, a highly-directional monochromatic linearly-polarized laser beam. To simulate such a beam, one can use an extended incoherent source placed behind a pinhole, followed by a narrow-band filter and a polarizer. The beam can then be collimated by a lens, followed by a filter with gaussian transparency (Fig. 12b). The resulting beam possesses the same features as the laser beam used in PSTM/STOM. It is now possible to apply our reciprocity principle to the set-up in Fig. 12(b). If the detector is replaced by an incoherent source of same shape and size, and the source by an extended detector, also with same shape and size, one obtains the *equivalent* illumination-mode set-up [Fig. 12(c)], which is an inverted PSTM/STOM [143]. According to the reciprocity principle, it produces the same images as the classical collection-mode PSTM/STOM. Note that, in order to obtain an equivalent set-up, the optical devices used to construct the laser beam from the incoherent source have to be kept in front of the detector.

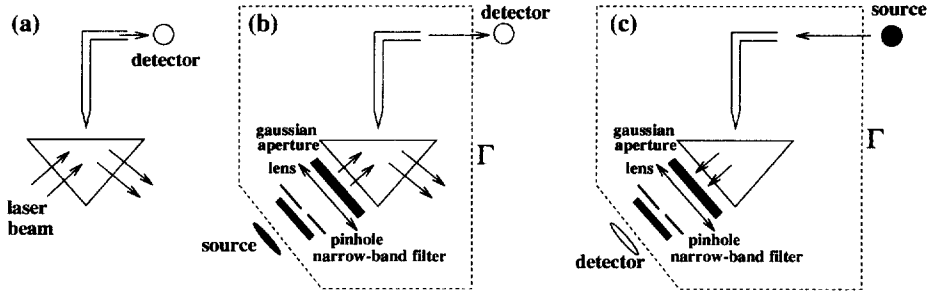

Fig. 12: *(a) representation of a PSTM/STOM. (b) schematic PSTM/STOM where the illuminating laser beam is simulated by an incoherent source followed by suitable optical devices. (c) illumination mode equivalent to the PSTM/STOM, resulting from the application of the reciprocity theorem to the configuration in (b).*

D. Role of coherence

The procedure we have used in order to find the equivalent illumination-mode instrument to the classical PSTM/STOM is general and may be used, for example, to design an equivalent set-up to the classical illumination-mode SNOM. We do not develop the entire procedure here, since the details can be found in ref.[140]. We insist on a rather subtle consequence of the reciprocity principle concerning the coherence of the illumination/detection scheme.

In the illumination-mode SNOM, the light is collected by the objective of a classical far-field microscope [1, 4]. Each pixel of the detector generates *independently* a signal proportional to its illumination. The signal delivered by the detector is the incoherent sum of all these contributions, which can be understood as *partially coherent detection* [140]. Indeed, in the reciprocal set-up, each pixel becomes an *independent* source. In other words, different pixels of the source are uncorrelated. Obviously, a laser does not fulfill this condition. Thus, if one illuminates with a laser beam through the objective of the far-field microscope, the collection-mode set-up that is obtained [63] *is not* the reciprocal set-up of the classical illumination-mode SNOM.

We are now able to discuss the main differences between the images produced by illumination-mode SNOM and those obtained by PSTM/STOM. Starting from the illumination-mode SNOM,

we have seen that the reciprocal set-up must use an incoherent source. In addition, the symmetry of the reciprocal source must be the same as the symmetry of the detector. In contrast, the typical PSTM/STOM uses a higly directional and coherent illumination, which explains why the images in illumination-mode SNOM ressemble more closely the actual structure of the sample than those obtained in PSTM/STOM. It has been shown both experimentally [144] and theoretically [145] that, under partially coherent and symmetric illumination, the PSTM/STOM also produces images that closely resemble the sample structure [144, 145].

E. Conclusion

Using the reciprocity theorem of electromagnetism, we have demonstrated that the illumination and collection modes of the SNOM are equivalent, which means that they have the same potential imaging capabilities and resolution. We have given a general rule to construct the equivalent collection-mode set-up to any illumination-mode set-up.

This result demonstrates that the differences that have been observed between the two modes are only due to differences in instrumentation and operating conditions. For example, the images in one case may present a better resolution, because of the use of coated fibers instead of uncoated ones, because the operating modes allows a smaller tip-sample distance, or because the degree of coherence of the illumination/detection scheme is not the same. But these differences are not fundamental. The same operating mode can be used in illumination and in detection mode. The same probes can also be used in both modes. Moreover, the coherence of the PSTM/STOM, for example, can be reduced [144, 146, 112, 147].

7. Image Formation

We now address the problem of the image formation with near-field optical microscopes. How the image is related to the optical properties and the topography of the sample has remained (to a large extent) an open question for the last decade. In this section, we derive a closed-form expression of the link between the detected signal and the structure of the sample.

Starting with the reciprocity theorem of electromagnetism, we obtain an *exact* expression for the signal, where all multiple interactions between the tip and the sample are taken into

account. We show that the signal is related to the induced current density in the sample, or to the induced polarization for non-magnetic materials. We give an explicit expression of the signal for the three main families of near-field optical set-ups, namely, collection-mode, illumination-mode and apertureless SNOM.

Finally, we address the problem of the existence of an impulse response (in direct space) or a transfer function (in Fourier space) in the image formation process in near-field optics. A number of authors have introduced different impulse responses or transfer functions [148, 136, 98, 149, 150, 151, 152, 153]. One of the major interest of the concept of impulse response is that it enables the inverse problem to be solved [98, 99, 145, 152, 153]. The existence of transfer functions in near-field optics is not trivial. One may define a transfer function between the signal and the near field (either a component or the square modulus of the electric field, for instance). It may also be possible to define a transfer function between the signal and the sample profile. The problem is more involved, if there is both topography and dielectric contrast. In that case, the concept of equivalent surface profile may be useful [99]. The necessary requirements for the existence of such a transfer function are linearity and translational invariance. The first requirement is satisfied in linear optics. The second implies that the effect of the tip is the same at any point of the sample. In other words, multiple scattering between the tip and the sample should be negligible. We will show that the concept of transfer function arises naturally, when using the reciprocity theorem and a single-scattering approximation (first Born approximation).

A. Collection mode

In collection-mode SNOM, the sample is usually illuminated by a laser focussed onto a part of the sample. An optical fiber (waveguide) with a sharp tip is scanned along the sample, at subwavelength distance, and the signal delivered by an optical detector at the end of the fiber is recorded. The scan can be performed at a constant distance from the mean surface (constant-height mode), or following the topography (constant-distance mode). An alternative procedure is to keep the optical signal constant by varying the distance between the tip and the sample, and to record the tip position (constant-intensity mode). It will be shown in §9 that the constant-height and constant-intensity modes are equivalent, provided that the maximum height of the sample surface is much smaller than the wavelength [154]. Moreover, we will see

that the constant-distance mode may induce topographical artifacts in the signal, due to the z-motion of the probe [155, 156, 157]. For these reasons, we will assume here that the scan is performed in the *constant-height mode*.

The basic question addressed here is : "what is measured by a near-field microscope in the collection mode ?" This involves two problems, namely, how the electromagnetic field is coupled to the guide through the tip and how the near-field is related to the structure of the sample. It has been often assumed that the signal is related to the square modulus of the electric field above the sample (see e.g. ref. [50]). Yet, there is no reason to believe that the different components of the field will have the same coupling efficiency. Van Labeke and Barchiesi [135] have proposed a heuristic model that predicts different coupling factors for different polarizations. The present work confirms the need for accounting for polarization and explicitly shows how to do it. Moreover, it gives an expression of the signal, showing its dependence on the sample structure.

(i) Basic expression of the signal

We establish the relationship between the signal delivered by a collection-mode SNOM and the current distribution inside the sample. To this aim, we use the reciprocity theorem of electromagnetism, and some standard relationships on guided waves, following the notation of ref.[80]. All the fields considered here are monochromatic, with frequency ω. A temporal dependence $\exp(-i\omega t)$ is assumed and omitted everywhere.

Inside an infinitely long waveguide with axis Oz, the field can be decomposed into modes $\mathbf{E}_\lambda^{(+)}$ propagating towards $z > 0$, and modes $\mathbf{E}_\lambda^{(-)}$ propagating towards $z < 0$. Thus, in a general situation, the field is :

$$\mathbf{E} = \sum_\lambda A_\lambda^{(+)} \mathbf{E}_\lambda^{(+)} + \sum_\lambda A_\lambda^{(-)} \mathbf{E}_\lambda^{(-)} \ , \tag{7.1}$$

where $A_\lambda^{(\pm)}$ is the amplitude of the mode $\mathbf{E}_\lambda^{(\pm)}$. If a current distribution, described by the current density $\mathbf{j}(\mathbf{r})$, is present in the guide and excites several modes, the mode $\mathbf{E}_\lambda^{(+)}$ has an amplitude $A_\lambda^{(+)}$ given by [80]

$$A_\lambda^{(+)} = \frac{-2\pi Z_\lambda}{c} \int \mathbf{j}(\mathbf{r}) \cdot \mathbf{E}_\lambda^{(-)}(\mathbf{r})\, d\mathbf{r} \ , \tag{7.2}$$

where the integral is performed over the volume of the current distribution. Equation (7.2) is

valid for a waveguide of arbitrary section, with perfectly conducting walls, and filled with a homogeneous material with dielectric constant ϵ_p and relative permeability μ_p. The impedance Z_λ is defined by $Z_\lambda = ck_\lambda/\epsilon_p\omega$ for TM polarization and $Z_\lambda = \mu_p\omega/ck_\lambda$ for TE polarization, k_λ being the wave vector of the mode λ along the z-axis. In what follows, we will consider a SNOM probe, i.e., a finite waveguide with a sharp tip at one extremity. For simplicity, we will assume that *far from the tip*, the guiding part of the probe is a waveguide with perfectly conducting walls. Thus, far from the tip, the field expansion (7.1), together with (7.2), is valid to describe the field reaching the detector.

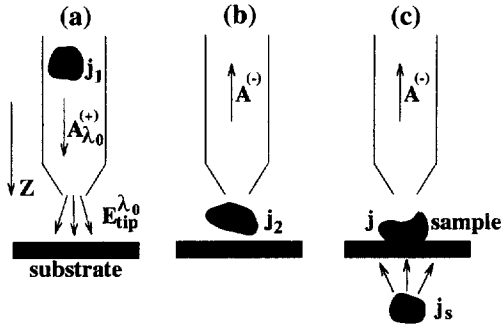

Fig. 13: *Sketch of the system. (a) The current* $\mathbf{j}_1$ *emits a field* $\mathbf{E}^{\lambda_0}_{tip}$ *in the presence of the flat substrate, (b) A current distribution* $\mathbf{j}_2$ *excites a field propagating into the waveguide, (c) A general situation where* $\mathbf{j}_2$ *is splitted into a source with an externally driven current density* $\mathbf{j}_s$ *and a sample with an induced current density* $\mathbf{j}$.

Let us assume that the probe is placed in front of a flat substrate, with the tip apex, or the probe aperture, located at a point $\mathbf{r}_{tip}$ (Fig. 13). This system (probe + flat substrate) is our reference system. In order to derive a general relationship between the amplitudes of the modes and the induced sources inside the sample, we apply the reciprocity theorem. To this aim, we consider two different configurations 1 and 2, involving the reference system with different sources [Figs. 13 (a) and (b)]. First, we assume that a *suitable current distribution* $\mathbf{j}_1$ placed inside the guide and *far from the tip* excites *a single mode* propagating towards $z > 0$, with label λ_0, and amplitude $A^{(+)}_{\lambda_0} = 1$ [Fig. 13(a)]. Using (7.2), this implies that :

$$\int_{V_1} \mathbf{j}_1(\mathbf{r}) \cdot \mathbf{E}^{(+)}_\lambda(\mathbf{r})\, d\mathbf{r} \;=\; 0$$

$$\int_{V_1} \mathbf{j}_1(\mathbf{r}) \cdot \mathbf{E}_\lambda^{(-)}(\mathbf{r})\, d\mathbf{r} = \frac{-c}{2\pi Z_\lambda} \delta_{\lambda\lambda_0} , \tag{7.3}$$

where V_1 is the volume of the current distribution $\mathbf{j}_1$, and $\delta_{\lambda\lambda_0} = 1$ if $\lambda = \lambda_0$, $\delta_{\lambda\lambda_0} = 0$ if $\lambda \neq \lambda_0$. Part of the mode is transmitted at the tip and produces a field denoted by $\mathbf{E}_{tip}^{\lambda_0}(\mathbf{r} - \mathbf{r}_{tip})$.

In a second configuration, we consider a current distribution $\mathbf{j}_2(\mathbf{r}, \mathbf{r}_{tip})$, of finite volume V_2, placed outside of the probe in the reference system, e.g., between the tip and the substrate [Fig. 13(b)]. This current distribution radiates a field $\mathbf{E}_2(\mathbf{r}, \mathbf{r}_{tip})$. Both $\mathbf{j}_2$ and $\mathbf{E}_2$ depend on $\mathbf{r}_{tip}$ because the tip is a scatterer that modifies the field and reacts on the source (multiple scattering). The reciprocity theorem states that (see Appendix D) :

$$\int_{V_1} \mathbf{j}_1(\mathbf{r}) \cdot \mathbf{E}_2(\mathbf{r}, \mathbf{r}_{tip})\, d\mathbf{r} = \int_{V_2} \mathbf{j}_2(\mathbf{r}, \mathbf{r}_{tip}) \cdot \mathbf{E}_{tip}^{\lambda_0}(\mathbf{r} - \mathbf{r}_{tip})\, d\mathbf{r} . \tag{7.4}$$

The field $\mathbf{E}_2$ in the left-hand side in (7.4) is evaluated inside the volume V_1, i.e., inside the guide and far from the tip, so that expansion (7.1) can be used. Moreover, the dot product of $\mathbf{j}_1$ with each mode of the expansion can be evaluated using (7.3). This leads to :

$$A_{\lambda_0}^{(-)}(\mathbf{r}_{tip}) = \frac{-2\pi Z_{\lambda_0}}{c} \int_{V_2} \mathbf{j}_2(\mathbf{r}, \mathbf{r}_{tip}) \cdot \mathbf{E}_{tip}^{\lambda_0}(\mathbf{r} - \mathbf{r}_{tip})\, d\mathbf{r} . \tag{7.5}$$

Equation (7.5) is a rigorous expression of the amplitude of the mode $\mathbf{E}_{\lambda_0}^{(-)}$ excited by an arbitrary current distribution $\mathbf{j}_2$. The only restriction to the validity of this result is that the media should satisfy reciprocity [141] (see also Appendix D). Let us recall that $\mathbf{E}_{tip}^{\lambda_0}$ is the field *emitted* by the tip used in illumination mode, when only the mode $\mathbf{E}_{\lambda_0}^{(+)}$ is excited. Note that, in general, $\mathbf{j}_2$ excites several modes and that (7.5) gives the amplitude of each mode. In (7.5), $\mathbf{j}_2$ is the current density *in the presence of the tip and the substrate*. At this step, we have not neglected multiple interactions between the tip, the substrate, and the current distribution $\mathbf{j}_2$. Let us remark that the general expression of the amplitude (7.5) does not involve the near field at the tip position. It establishes a direct relationship between the amplitude of the mode and the *current distribution* $\mathbf{j}_2$.

(ii) Case of a point dipole

Let us examine the coupling of a radiating point dipole $\mathbf{p}_2$ with the probe/substrate system. This corresponds to a current density $\mathbf{j}_2(\mathbf{r}) = -i\omega\mathbf{p}_2\, \delta(\mathbf{r} - \mathbf{r}_2)$, where $\mathbf{r}_2$ is the position of the

dipole. Equation (7.5) reduces to :

$$A_{\lambda_0}^{(-)} = 2\pi Z_{\lambda_0} \frac{i\omega}{c} \mathbf{p}_2 \cdot \mathbf{E}_{tip}^{\lambda_0}(\mathbf{r}_2 - \mathbf{r}_{tip}) \ . \tag{7.6}$$

Equation (7.6) is an exact expression of the amplitude of the mode excited by a point dipole, in near-field coupling with the probe. Note that all multiple interactions are taken into account in this expression, namely, $\mathbf{p}_2$ is the dipole moment in the presence of the probe and the substrate. This result could be useful, for instance, in the analysis of the fluorescence signal emitted by a single molecule, where the radiating dipole model is a good approximation. In particular, (7.6) involves a dot product, thus showing the sensitivity of the signal to the orientation of the molecule, as observed experimentally [38]. The dependence of the signal on the dipole orientation is also crucial when one uses the molecule lifetime as a source of contrast, this lifetime being strongly affected by the dipole orientation with respect to the surrounding objects [158].

(iii) Case of a mesoscopic sample

The current density $\mathbf{j}_2$ in (7.5) can be splitted into the current density $\mathbf{j}_s$ of the sources illuminating the system and the current density $\mathbf{j}$ of the sample under investigation, assumed to be deposited on the flat substrate [Fig. 13(c)]. In an experiment, $\mathbf{j}_s$ is the *externally driven* current density of the sources, placed in the far field of the sample. In particular, $\mathbf{j}_s$ is assumed to be unaffected by the presence of the sample. For the most general sample, the current density $\mathbf{j}$, which depends on the tip position, is :

$$\mathbf{j}(\mathbf{r}, \mathbf{r}_{tip}) = -i\omega \mathbf{P}(\mathbf{r}, \mathbf{r}_{tip}) + \nabla \times \mathbf{M}(\mathbf{r}, \mathbf{r}_{tip}) \ , \tag{7.7}$$

where $\mathbf{P}$ and $\mathbf{M}$ are the (position- and frequency-dependent) polarization and magnetization densities in the sample, respectively. For a local and linear material, they can be related to the electric and magnetic fields, $\mathbf{E}$ and $\mathbf{H}$, by the local constitutive relations $\mathbf{P} = \epsilon_0(\overleftrightarrow{\epsilon} - 1)\mathbf{E}$ and $\mathbf{M} = (\overleftrightarrow{\mu} - 1)\mathbf{H}$, where $\overleftrightarrow{\epsilon}$ and $\overleftrightarrow{\mu}$ are the dielectric and magnetic tensors of the sample, respectively. As usual in optics, $\overleftrightarrow{\epsilon}$ is a generalized dielectric tensor, that accounts for both the free and bound charges of the material. Thus, (7.5) can describe the dielectric, metallic and magnetic response of the sample, which may be anisotropic.

For the sake of simplicity, we limit the following discussion to a non-magnetic, isotropic sample. In this case, the polarization density is $\mathbf{P} = \epsilon_0(\epsilon - 1)\mathbf{E}$, where ϵ is the position-

and frequency-dependent dielectric constant of the sample. Dropping the upperscript $(-)$ for brevity, (7.5) can be cast in the form :

$$A_{\lambda_0}(\mathbf{r}_{tip}) = A^{bg}_{\lambda_0} + A^{sa}_{\lambda_0}(\mathbf{r}_{tip}) \ , \tag{7.8}$$

where :

$$A^{bg}_{\lambda_0} = \frac{-2\pi Z_{\lambda_0}}{c} \int_{V_s} \mathbf{j}_s(\mathbf{r}) \cdot \mathbf{E}^{\lambda_0}_{tip}(\mathbf{r} - \mathbf{r}_{tip})\, d\mathbf{r} \tag{7.9}$$

is the background component of the signal (V_s denotes the volume of the external sources) and

$$A^{sa}_{\lambda_0}(\mathbf{r}_{tip}) = \frac{2\pi Z_{\lambda_0}}{c}\, i\omega\epsilon_0 \int_V [\epsilon(\mathbf{r}) - 1]\, \mathbf{E}(\mathbf{r}, \mathbf{r}_{tip}) \cdot \mathbf{E}^{\lambda_0}_{tip}(\mathbf{r} - \mathbf{r}_{tip})\, d\mathbf{r} \ . \tag{7.10}$$

is the contribution of the sample (V is the volume of the sample).

The first term (7.9) does not depend on the sample properties. When the tip scans the sample *at a constant height*, this term contributes to the background of the mode amplitude only. Nevertheless, if the tip is forced to move along the z direction, for instance by the tip-sample distance regulation mechanism, this term may vary and induce a variation in the mode amplitude, this variation being *independent on the sample.* This z-motion artifact will be studied in details in §9 (see also refs.[155, 156, 157]).

The second term (7.10) describes the purely optical contribution. Assuming that the scan is performed at a constant height, this term is the only dynamic part of the mode amplitude. It relates the mode amplitude to the dielectric contrast in the sample. Let us stress that the dielectric contrast, as written in (7.10), shows the dependence on both the optical properties (dielectric constant) and the topography of the sample. Indeed, the function $[\epsilon(\mathbf{r}) - 1]$ vanishes out of the sample domain.

(iv) What is measured ?

In optics, one detects the time-averaged intensity rather than the amplitude of the field. Thus, the detected signal is the energy flux in the guide. For simplicity, we consider a monomode guide, so that the signal is proportional to $|A_{\lambda_0}|^2$. Accounting for the difference of magnitude between the two terms $A^{(back)}_{\lambda_0}$ and $A^{(dyn)}_{\lambda_0}$, the signal can be written to first order :

$$|A_{\lambda_0}|^2 = |A^{bg}_{\lambda_0}|^2 + 2\mathrm{Re}\{A^{bg*}_{\lambda_0} A^{sa}_{\lambda_0}\} \ , \tag{7.11}$$

where Re denotes the real part of a complex number. Only the second term $S = 2\mathrm{Re}\{A^{bg*}_{\lambda_0} A^{sa}_{\lambda_0}\}$ varies when the tip is scanned at a constant height. Inserting (7.10) into (7.11), we obtain :

$$S = -4\pi\epsilon_0 \frac{\omega}{c} \mathrm{Im}\left\{ Z_\lambda A^{bg*}_{\lambda_0} \int_V [\epsilon(\mathbf{r}) - 1]\, \mathbf{E}(\mathbf{r}, \mathbf{r}_{tip}) \cdot \mathbf{E}^{\lambda_0}_{tip}(\mathbf{r} - \mathbf{r}_{tip})\, d\mathbf{r} \right\} . \tag{7.12}$$

Note that the signal S has the structure of an interference term between the field $\mathbf{E}^{\lambda_0}_{tip}$ and the field $\mathbf{E}$ in the sample. The major consequence of this remark is that both the *phase* and the *amplitude* of the near field are recorded, as was pointed out earlier [97, 100] (see also §5), and it was recognized that the output is essentially a hologram of the sample. Yet, note that this conclusion was derived from the heuristic assumption that the signal is proportional to the *near-field intensity*, an assumption that turns out to coincide with the exact result (7.12) only when the vectorial nature of the field disappears. In fact, (7.12) accounts for the polarization effect in the detection process. The signal produced by the mode $\lambda_0^{(-)}$ propagating towards the detector is related to the projection of the near field on the direction of the field $\mathbf{E}^{\lambda_0}_{tip}$, emitted through the tip by the mode $\lambda_0^{(+)}$, in the absence of the sample. It is clear that $\mathbf{E}^{\lambda_0}_{tip}$ is the fundamental quantity to measure in order to interpret the image. Note that several methods have been proposed to characterize the field emitted by a SNOM probe [159, 38].

Finally, (7.12) shows explicitely the dependence of the signal on the sample structure (dielectric constant and topography). The result could be easily extended to a magnetic material, as mentionned previously. This expression is exact in the sense that all mutliple interactions are taken into account. Indeed, the field $\mathbf{E}(\mathbf{r}, \mathbf{r}_{tip})$ in (7.12) is the self-consistent field inside the sample. In particular, it depends on the tip position.

(v) Existence of an impulse response

Equation (7.12) provides a linear relationship between the signal and the dielectric contrast in the sample. Nevertheless, the dependence of the field $\mathbf{E}$ on the tip position breaks the translational invariance, so that it is *a priori* not possible to introduce an impulse response (or a transfer function). Nevertheless, we *demonstrate* that a single scattering approximation, which is accurate in NFO for a broad class of samples [112, 115, 153, 160, 151], allows to introduce naturally an impulse response describing the imaging system. This result gives a rigorous theroretical basis to the concepts of impulse response and transfer functions introduced

previously in NFO, either heuristically or numerically [148, 136, 98, 149, 150, 152, 153, 95].

Assuming that only single scattering takes place inside the sample, the first Born approximation can be used in (7.12). It amounts to approximating the field $\mathbf{E}(\mathbf{r}, \mathbf{r}_{tip})$ by the field $\mathbf{E}^{(0)}$ illuminating the sample, i.e., the field radiated by the sources $\mathbf{j}_s$ in the reference system (tip + substrate), in the absence of the sample. Due to the extended illumination used in collection-mode SNOM (the sources are in the far field, and the incident field is uniform accross the scanning area), the illuminating field has the form $\mathbf{E}^{(0)}(\mathbf{r} - \mathbf{r}_{tip})$. Note that $\mathbf{E}^{(0)}$ is in general non uniform, because of the interaction of the tip with the *substrate*. For example, a local enhancement may be observed. In other words, the tip may interact strongly with the *substrate*, but we assume a weak interaction with the *sample*. An example of such a situation would be a dielectric agregate deposited on a metallic substrate. The remarkable consequence of this approximation is that an impulse response $H(\mathbf{r} - \mathbf{r}_{tip})$ exists between the dielectric contrast and the signal :

$$S = \int_V H(\mathbf{r} - \mathbf{r}_{tip}) [\epsilon(\mathbf{r}) - 1] \, d\mathbf{r} \ , \tag{7.13}$$

where the impulse response is given by :

$$H(\mathbf{r} - \mathbf{r}_{tip}) = -4\pi\epsilon_0 \frac{\omega}{c} \operatorname{Im} \left\{ Z_\lambda A_{\lambda_0}^{bg*} \mathbf{E}^{(0)}(\mathbf{r} - \mathbf{r}_{tip}) \cdot \mathbf{E}_{tip}^{\lambda_0}(\mathbf{r} - \mathbf{r}_{tip}) \right\} \ . \tag{7.14}$$

Let us use this result to draw some conclusions on the imaging capabilities of the technique :

1. Obviously, the contribution to the signal comes from regions where the field $\mathbf{E}_{tip}^{\lambda_0}$ takes significant values. In this respect, the field $\mathbf{E}_{tip}^{\lambda_0}$ is a weighting factor. Yet, this condition is not sufficient, since the illuminating field $\mathbf{E}^{(0)}$ may be perpendicular to $\mathbf{E}_{tip}^{\lambda_0}$, so that the contribution would be zero.

2. The nice feature that appears is that one may select, as a weighting factor, any cartesian component of $\mathbf{E}_{tip}^{\lambda_0}$ by illuminating with a linearly polarized field $\mathbf{E}^{(0)}$. This can be used to enhance the resolution.

3. An important consequence of the existence of this impulse response is that the inverse problem can be solved, in principle, by a *deconvolution* [98, 99, 145].

4. Finally, (7.14) shows that *multiple scattering (i.e., strong interactions) between the tip and the sample is not necessary to obtain good resolution.* The achievable resolution is given

by the width of the impulse response. A sufficient condition, to achieve high resolution, is to have a high confinement of at least one component of $\mathbf{E}_{tip}^{\lambda_0}$. Moreover, the illuminating field $\mathbf{E}^{(0)}$ may also be confined, due to the interaction of the tip and the flat substrate.

Several experiments have been done by both illuminating and detecting with the same tip, which is known as the *reflection mode* [161, 162, 163]. In that case, the fields $\mathbf{E}_{tip}^{\lambda_0}$ and $\mathbf{E}^{(0)}$ coincide and the overlapping integral is maximal. An interesting property is that this configuration yields an improvement of the resolution very similar to what happens in confocal microscopy. Let us suppose, for the sake of illustration, that the field $\mathbf{E}_{tip}^{\lambda_0}$ is Gaussian with a width a. Then the product $\mathbf{E}_{tip}^{\lambda_0} \cdot \mathbf{E}^{(0)}$ is a Gaussian with width a for a plane-wave illumination and a Gaussian with width $a/\sqrt{2}$ for an illumination by $\mathbf{E}_{tip}^{\lambda_0}$. Note that, in this particular case, the transfer function does not depend on the polarization. A slightly different set-up has been used by Bozhevolnyi et al.[163]. They detected the cross-polarized light, which amounts to using, as an impulse response, the product $\mathbf{E}_{tip}^{\lambda_0} \cdot \mathbf{E}'^{\lambda_0}_{tip}$, where $\mathbf{E}'^{\lambda_0}_{tip}$ is the field emitted by the tip, when the polarization has changed by 90°. It is clear that the depolarizing properties of the tip are involved in this particular case.

Equation (7.14) is the explicit expression of the impulse response relating the signal to the dielectric contrast in the sample. Nevertheless, some attempts to define an impulse response (or a transfer function) between the signal and the *near-field intensity* illuminating the tip were reported in the literature [136, 151, 152, 95]. In the present framework, it is possible to derive an explicit expression of this transfer function. The derivation is reported in Appendix F. In particular, it shows that 1) such a transfer function exists under the same conditions as the impulse response defined in (7.14), and 2) the transfer function does not, in general, relates the signal to the square modulus of the electric field, due to polarization effects.

B. Illumination mode

The illumination-mode SNOM closely follows the original idea of Synge [51]. The nanosource, achieved for instance by illuminating through a metal-coated optical fiber tip [33], is scanned along the surface and the scattered field is recorded by a detector in the far field, either in transmission or reflection. The link between the signal and the properties of the sample can be established along similar lines : we use the reciprocity theorem to derive an exact relationship

between the induced current in the sample and the field at the detector position. Then, we show that a single scattering approximation leads to the introduction of an impulse response to describe the imaging properties of the microscope.

(i) Basic equation

To proceed, we consider two different situations, as depicted in Figs. 14 (a) and (b). The reference geometry involves a flat substrate and the probe at subwavelength distance. We use the same notations as in the previous paragraph.

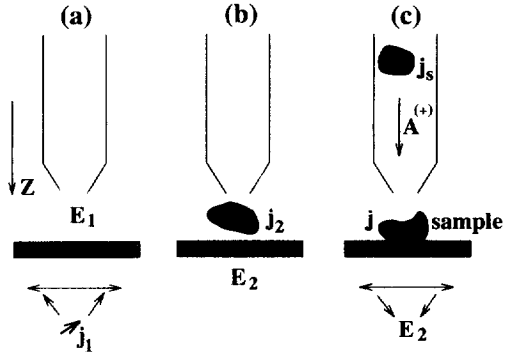

Fig. 14: *Sketch of the system. (a) The current* $\mathbf{j}_1$ *(assumed ponctual) produces a field* $\mathbf{E}_1$ *in the reference system, (b) A current distribution* $\mathbf{j}_2$ *produces a field* $\mathbf{E}_2$ *in the same system, (c) A general situation where* $\mathbf{j}_2$ *is splitted into a source with an externally driven current density* $\mathbf{j}_s$ *and a sample with an induced current density* $\mathbf{j}$.

In the first configuration [Fig. 14(a)], a current density $\mathbf{j}_1$, placed in the far field (i.e. at the location of the detector), produces a field $\mathbf{E}_1$. This field is a function of $\mathbf{r} - \mathbf{r}_{tip}$ because the source current density is placed in the far field, so that the *incident* field on the system is assumed to be uniform. Thus, the system is translationaly invariant. In practice, this condition is fulfilled on the scanning area. For simplicity, we assume that the detector in the experiment is a *point detector* placed in the far field (i.e., a small detector in the focal plane of a lens), and we can use a current density $\mathbf{j}_1(\mathbf{r}) = \mathbf{j}_1\delta(\mathbf{r} - \mathbf{r}_1)$. Note that we could also introduce the dipole moment $\mathbf{p}_1$, such that $\mathbf{j}_1 = -i\omega\mathbf{p}_1$. But in order to emphasize the similiraty with the

study of the collection-mode set-up, it is preferable to work with the current $\mathbf{j}_1$. In the second configuration [Fig. 14(b)], a current distribution $\mathbf{j}_2(\mathbf{r}, \mathbf{r}_{tip})$, placed for instance between the tip and the substrate, produces a field $\mathbf{E}_2(\mathbf{r}, \mathbf{r}_{tip})$. The reciprocity theorem yields (see Appendix D) :

$$\mathbf{j}_1 \cdot \mathbf{E}_2(\mathbf{r}_1, \mathbf{r}_{tip}) = \int_{V_2} \mathbf{j}_2(\mathbf{r}, \mathbf{r}_{tip}) \cdot \mathbf{E}_1^{\mathbf{j}_1}(\mathbf{r} - \mathbf{r}_{tip})\, d\mathbf{r} \ . \tag{7.15}$$

We use the notation $\mathbf{E}_1^{\mathbf{j}_1}$ for the field $\mathbf{E}_1$ to keep in mind that this field depends on the choice of the current $\mathbf{j}_1$ (especially on its orientation which controls the polarizxation of $\mathbf{E}_1^{\mathbf{j}_1}$). This notation is similar to that used for the collection-mode set-up, $\mathbf{E}_{tip}^{\lambda_0}$, where the mode λ_0 was chosen. Splitting the current density $\mathbf{j}_2$ into the contribution of the sources $\mathbf{j}_s$ located inside the guide and far from the tip (illumination mode) and the contribution of the sample $\mathbf{j} = -i\omega\epsilon_0[\epsilon(\mathbf{r}) - 1]\,\mathbf{E}(\mathbf{r}, \mathbf{r}_{tip})$ [see Fig. 14(c)], one obtains :

$$A_{\mathbf{j}_1}(\mathbf{r}_{tip}) = A_{\mathbf{j}_1}^{bg} + A_{\mathbf{j}_1}^{sa}(\mathbf{r}_{tip}) \ , \tag{7.16}$$

where $A_{\mathbf{j}_1} = \mathbf{p}_1 \cdot \mathbf{E}_2(\mathbf{r}_1, \mathbf{r}_1)$ is proportional to the component of the *field at the detector position* along the direction of $\mathbf{j}_1$. It contains a *background* contribution $A_{\mathbf{j}_1}^{bg}$, independent on the sample, and which does not vary with $\mathbf{r}_{tip}$ when the scan is performed at a constant height, and a *dynamic* contribution $A_{\mathbf{j}_1}^{sa}$, which contains the information on the sample :

$$A_{\mathbf{j}_1}^{bg} = \int_{V_s} \mathbf{j}_s(\mathbf{r}) \cdot \mathbf{E}_1^{\mathbf{j}_1}(\mathbf{r} - \mathbf{r}_{tip})\, d\mathbf{r} \tag{7.17}$$

$$A_{\mathbf{j}_1}^{sa}(\mathbf{r}_{tip}) = -i\omega\epsilon_0 \int_V [\epsilon(\mathbf{r}) - 1]\,\mathbf{E}(\mathbf{r}, \mathbf{r}_{tip}) \cdot \mathbf{E}_1^{\mathbf{j}_1}(\mathbf{r} - \mathbf{r}_{tip})\, d\mathbf{r} \ . \tag{7.18}$$

Equations (7.17) and (7.18) give the field at the detector position, produced in the illumination-mode configuration in Fig. 14(c). The second term (7.18) gives the optical information on the sample, and depends on both its dielectric constant and topography. Note that (7.17) and (7.18) do not give the expression of the field $\mathbf{E}_2$, but of its *projection on the direction of* $\mathbf{p}_1$. This describes the *polarization effect* in the imaging process in illumination-mode SNOM. For example, to determine the x-component of $\mathbf{E}_2$, one needs to know the field $\mathbf{E}_1^{\mathbf{j}_1}$ produced when the reference system (probe + substrate) is illuminated by a point current $\mathbf{j}_1$ of unit amplitude, polarized along Ox, and located at the detector position.

(ii) What is measured ?

The dynamic part S of the signal produced by the component of $\mathbf{E}_2$ along the direction of $\mathbf{j}_1$ is $S = 2\mathrm{Re}\{A^{bg*}_{\mathbf{j}_1} A^{sa}_{\mathbf{j}_1}\}$, assuming that $|A^{sa}_{\mathbf{j}_1}| \ll |A^{bg}_{\mathbf{j}_1}|$. Using (7.17) and (7.18), we obtain :

$$S = 2\omega\epsilon_0 \,\mathrm{Im}\left\{ A^{bg*}_{\mathbf{j}_1} \int_V [\epsilon(\mathbf{r}) - 1]\, \mathbf{E}(\mathbf{r}, \mathbf{r}_{tip}) \cdot \mathbf{E}^{\mathbf{j}_1}_1(\mathbf{r} - \mathbf{r}_{tip})\, d\mathbf{r} \right\} . \tag{7.19}$$

Equation (7.19) gives an exact expression of the signal in illumination-mode SNOM, when a point detector is used, and when only one component of the field is measured (polarized detection). Let us stress that all multiple interactions between the sample and the reference system are taken into account in this expression. In particular, $\mathbf{E}(\mathbf{r}, \mathbf{r}_{tip})$ is the field inside the sample, which depends on the tip position $\mathbf{r}_{tip}$.

(iii) Impulse response

As for the collection-mode configuration, an impulse response appears naturally when the first Born approximation is used to describe the field in the sample. In this particular case, this approximation amounts to replacing $\mathbf{E}(\mathbf{r}, \mathbf{r}_{tip})$ in (7.19) by the field produced by the tip in the reference system. Assuming for simplicity that only one mode, denoted by λ_0, is excited, this field is $\mathbf{E}^{\lambda_0}_{tip}(\mathbf{r} - \mathbf{r}_{tip})$ [see Fig. 13(a)]. One obtains :

$$S = \int_V H(\mathbf{r} - \mathbf{r}_{tip})\, [\epsilon(\mathbf{r}) - 1]\, d\mathbf{r} , \tag{7.20}$$

where

$$H(\mathbf{r} - \mathbf{r}_{tip}) = 2\omega\epsilon_0 \,\mathrm{Im}\left\{ A^{bg*}_{\mathbf{j}_1}\, \mathbf{E}^{\lambda_0}_{tip}(\mathbf{r} - \mathbf{r}_{tip}) \cdot \mathbf{E}^{\mathbf{j}_1}_1(\mathbf{r} - \mathbf{r}_{tip})\, d\mathbf{r} \right\} . \tag{7.21}$$

(iv) Similarity with the impulse response in collection-mode SNOM

Except for a constant factor, (7.21) is identical to (7.14), obtained for collection-mode SNOM. In fact, $A^{bg}_{\mathbf{j}_1}$, and $\mathbf{E}^{\mathbf{j}_1}_1$ in (7.21) are exactly $A^{bg}_{\lambda_0}$ and $\mathbf{E}^{(0)}$ in (7.14). This result is not surprising because in §6, we have demonstrated that the collection-mode and illumination-mode instruments are equivalent. Since the two modes are equivalent, their impulse response (when they exist) are identical, as expected. Note that some precise equivalence rules (involving the coherence and polarization properties of the illumination/detection scheme) are given in §6, in order to obtain the same signals in both modes. As a consequence of this equivalence, all the conclusions

we obtained from the structure of the impulse reponse in collection-mode SNOM (localization of the field emitted by the probe and resolution, polarization effect, inverse problem and role of multiple scattering) are valid in illumination-mode SNOM.

C. Apertureless set-up

A number of groups have built near-field optical microscopes by using the apertureless technique, where detection is performed in the far field. In order to send light to the detector, a tip is brought into the vicinity of the sample. It is mechanically excited and vibrates in the z-direction at a well-controlled frequency. The amplitude of the scattered light is related to the amplitude of the near-field that locally illuminates the tip. Differences in the set-ups are due to the nature of the tip (AFM or STM tips [76, 110, 74, 77, 90], levitating particles[73, 164]), operating mode (constant-height or constant-distance mode), illumination, coherence of the illumination, polarization control, etc.

Since the first papers were published, the understanding of the mechanism has considerably changed. Basically, the tip may play two roles. On one hand, it acts as a nanodetector, as depicted in Fig. 15(a). On the other hand, it acts as a nanosource, as shown in Fig. 15(b). Further interactions (multiple scattering) between the tip and the substrate may appear, if there is a strong coupling (e.g., a metallic particle above a metallic surface supporting surface plasmons polaritons). It has been argued very often that such multiple scattering is necessary to obtain good resolution [90]. Indeed, excellent resolutions have been demonstrated for metallic surfaces[71, 72]. Yet, as for the collection and illumination modes, we will see that strong interaction is not necessary. The present goal is to analyse the imaging capabilities of this type of set-up. We have shown that, for collection and illumination-mode SNOM, the signal can be related to the dielectric contrast of the sample. In this section, we outline a similar analysis, which is also based on the reciprocity theorem. We first establish an exact expression for the signal, and relate it to the sample properties. Then, we show that an impulse response can be introduced, when multiple scattering between the sample and the reference system (tip + substrate) can be neglected.

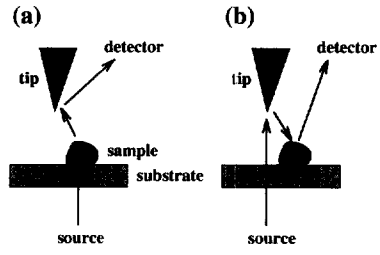

Fig. 15: *Sketch of the apertureless set-up. (a) Nanodetector mechanism. (b) Nanosource mechanism.*

(i) Exact expression for the signal

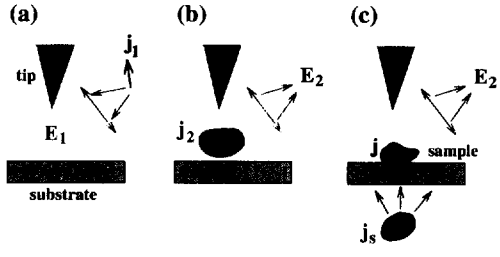

Fig. 16: *Sketch of the system. (a) A current* $\mathbf{j}_1$ *at the location of the detector produces a field* $\mathbf{E}_1$ *in the presence of the tip and the flat substrate. (b) A current density* $\mathbf{j}_2$ *produces a field* $\mathbf{E}_2$*. (c) A general experimental situation with a source current density* $\mathbf{j}_s$ *and a sample.*

Consider a tip placed at subwavelength distance from a flat substrate, at a position $\mathbf{r}_{tip}$. The tip may be either an AFM tip [76] or an STM tip [72, 77, 74, 110]. In a first situation [Fig. 16 (a)], a point dipole $\mathbf{p}_1$ is placed in the far field, at the location of the detector (assumed to be point-like for simplicity). As for the illumination-mode SNOM, we work with the point current density $\mathbf{j}_1 = -i\omega\mathbf{p}_1$, rather than with the dipole moment $\mathbf{p}_1$ itself. We denote by $\mathbf{E}_1^{\mathbf{j}_1}(\mathbf{r} - \mathbf{r}_{tip})$ the field produced in this situation. The field $\mathbf{E}_1^{\mathbf{j}_1}$ can be highly confined, because of the tip effect. This type of tip effect can be described in the electrostatic approximation, and some analytical expressions of the field behavior close to the tip can be found, e.g., in ref. [165]. In a second situation [Fig. 16 (b)], a current density $\mathbf{j}_2$, placed in the near field, produces a field $\mathbf{E}_2(\mathbf{r}, \mathbf{r}_{tip})$. Using the reciprocity theorem, and splitting the current distribution $\mathbf{j}_2$ into a source

distribution $\mathbf{j}_s$ and the current density in the sample, we obtain an expression for the signal which is formally the same as (7.19), derived for the illumination-mode SNOM :

$$S = 2\omega\epsilon_0 \,\mathrm{Im}\left\{ A_{\mathbf{j}_1}^{bg*} \int_V [\epsilon(\mathbf{r}) - 1]\, \mathbf{E}(\mathbf{r}, \mathbf{r}_{tip}) \cdot \mathbf{E}_1^{\mathbf{j}_1}(\mathbf{r} - \mathbf{r}_{tip})\, d\mathbf{r} \right\} . \tag{7.22}$$

The only differences between (7.22) and (7.19) are (1) in apertureless SNOM, the source distribution does not excite a mode in an optical fiber, used to produce a localized illumination of the sample, but the source directly illuminates the sample from the far field, as in conventional microscopy, and (2) the fields $\mathbf{E}_1^{\mathbf{j}_1}$ are very different in both situations. Let us recall that in (7.22), S is the dynamic part of the signal produced at the detector position by the component of the electric field polarized along the direction of $\mathbf{j}_1$.

(ii) Impulse response

As for the illumination-mode configuration, an impulse response appears naturally when the first Born approximation is used to describe the field in the sample. Within this approximation, we replace the field $\mathbf{E}(\mathbf{r}, \mathbf{r}_{tip})$ in (7.22) by the field $\mathbf{E}_{tip}^{(0)}(\mathbf{r} - \mathbf{r}_{tip})$ illuminating the sample. This field is the field radiated by the sources $\mathbf{j}_s$ in the reference system (tip + flat substrate). Note that, due to the near field interaction between the subwavelength tip and the *substrate* and the tip effect, this field may exhibit subwavelength lateral variations, as well as a local enhancement [166]. We finally obtain :

$$S = \int_V H(\mathbf{r}, \mathbf{r}_{tip})\, [\epsilon(\mathbf{r}) - 1]\, d\mathbf{r} , \tag{7.23}$$

where

$$H(\mathbf{r}, \mathbf{r}_{tip}) = 2\omega\epsilon_0 \,\mathrm{Im}\left\{ A_{\mathbf{j}_1}^{bg*}\, \mathbf{E}_{tip}^{(0)}(\mathbf{r} - \mathbf{r}_{tip}) \cdot \mathbf{E}_1^{\mathbf{j}_1}(\mathbf{r} - \mathbf{r}_{tip}) \right\} . \tag{7.24}$$

Let us emphasize that (7.24) accounts for the two roles of the tip, namely, nanosource and nanodetector (see Fig. 15). Note that in Fig. 15, only single scattering takes place inside the sample, thus the existence of an impulse response is not in contradiction with the presence of the two mechanisms. We see that the resolution depends on both the confinement of the illuminating field $\mathbf{E}_{tip}^{(0)}$ and of the field $\mathbf{E}_1^{\mathbf{j}_1}$. As for the other modes, *multiple scattering in the sample is not necessary to achieve high resolution.* Note that a confocal type set-up can be used to run such an experiment, where the illuminating field $\mathbf{E}^{(0)}$ *coincides with the field* $\mathbf{E}_1$, so that

the impulse response is even more localized, as discussed for the collection-mode set-up. In this particular arrangement, *the enhancement, due to the singularity close to the tip, is squared.* Such an experiment has been reported in the literature [110, 74].

Finally, we note that (7.24) accounts for the polarization properties of the imaging process in apertureless SNOM. The signal S is produced by a measurement of the field at the detector position, *projected on a direction* $\mathbf{u}_1$. This signal is proportional to the projection of the illuminating field $\mathbf{E}^{(0)}$ on the field $\mathbf{E}_1$, produced in the reference system by a dipole $\mathbf{p}_1$ oriented along $\mathbf{u}_1$, and placed at a detector position. The importance of polarization effects in apertureless SNOM was pointed out by several authors [84, 83]. The role of the polarization of the incident field and of the detector position was particularly emphasized. We hope that this model will be helpful in understanding the basic mechanisms leading to these observations.

(iii) Nanosource or nanodetector ?

Recently, numerical calculations have been performed to investigate the importance of multiple scattering between the tip and the sample, and also to compare the relative importance of the two mechanisms : nanosource and nanodetector (see Fig. 15) [83]. The system studied was a small gold cylinder levitating above a one-dimensional dielectric rough surface, illuminated by a p-polarized plane wave (see Fig. 17).

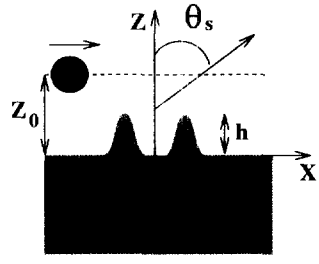

Fig. 17: *Sketch of the system, where a small cylinder is placed above the surface.*

The detector detects the flux of the Poynting vector in the far field for different observation angles. In a numerical simulation, it is possible to evaluate separately the field radiated by the tip (nanodetector mechanism) and the field radiated by the currents induced on the sample

(nanosource mechanism). If the field radiated by the sample shows high-resolution structures, it means that the cylinder has locally illuminated the sample, acting as a nanosource. The results displayed in Fig. 18(a) show that for a tip sample distance of 0.017λ, the intensity scattered by the surface does not display any resolution. Thus, the small cylinder is not a nanosource. By contrast, for the same tip-sample distance, the signal scattered by the tip reproduces the surface profile (see Fig. 18), as shown in ref.[83]. Thus, in this case, the tip mainly acts as a nanodetector.

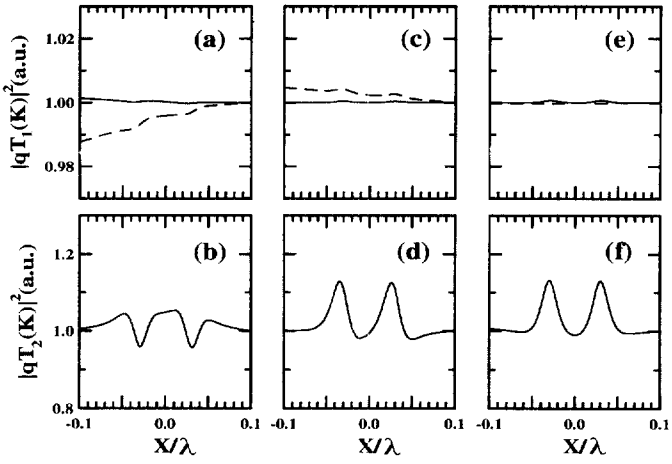

Fig. 18: *Intensity scattered in the far field by the sample (nanosource mechanism) (a), (c), (e) and by the tip (nanodetector mechanism)(b), (d), (f) for different scattering angles when the tip is scanned at constant height* $z_0 = 0.017\lambda$ *along the surface. The index of the substrate is 1.5. The dielectric constant of the tip is* $-9.89 + i1.05$*, its radius is* $\lambda/100$*. The topographical bumps are Gaussian, with height* $\lambda/200$ *and width* $\lambda/100$*. The incident field is p-polarized and* $\theta_0 = 50°$*. Left column,* $\theta_s = 0°$*, central column* $\theta_s = 50°$*, right column* $\theta_s = 90°$*. From A. Madrazo et al., ref.[83]*

The results in ref.[83] shows that, even with a metallic tip, the nanosource mechanism appears only for very small tip-sample distance, with a highly reflective substrate. Nevertheless, the calculations were carried out with a cylinder (i.e., the 2D equivalent of a sphere). With a sharp tip, a stronger local enhancement of the field could be obtained, due to the singularity of the field at the tip [165]. In this case, the nanosource mechanism may become more important.

D. Conclusion

We have shown that the image formation in the different families of near-field optical microscopes can be described in the same framework. In each case, the signal can be related to the polarization density in the sample. The explicit form of the relationship has the structure of an overlapping integral of the dot product of the polarization density with a reference field $\mathbf{E}_1$. *This reference field is always the field produced in the reference system* (tip +substrate, without sample), *when the detector of the experiment is replaced by a source*. We emphasize that, in the derivation of this result, the only assumption is that the reciprocity theorem holds (see Appendix D). We also emphasize that the reference field $\mathbf{E}_1$ depends on the polarization of the detection. By changing the polarization, the reference field is changed. The field $\mathbf{E}_1$ plays a central role in the analysis. It is therefore important to be able to measure it experimentally. A possible solution is to have a point dipole acting as a local sensor of the field. The experiments described by Betzig *et al.* [38] are essentially this type of measurement. Another method was proposed recently to measured the near field emitted by a probe [159]. Note that producing an interference with a linearly polarized plane wave might allow to record separately the diferent components of the field $\mathbf{E}_1$.

At the cost of an approximation, the concept of impulse response can be introduced naturally. The only assumption is that multiple scattering between the sample and the rest of the system is negligible. For instance, consider a molecule placed between a coated optical fiber and a metallic substrate. There is obviously a strong interaction between the tip and the substrate, but multiple scattering between the *molecule* and the rest of the system can often be neglected, because the scattering cross-section is small. This condition is generally satisfied if the samples are small compared to the wavelength and if their dielectric contrast is low. The general form of the impulse response is then the product of the illuminating field $\mathbf{E}^0$ (i.e., the

field produced in the reference system by the real sources of the experiment set-up) with the reference field $\mathbf{E}_1$:

$$H(\mathbf{r}-\mathbf{r}_{tip}) = K\,\mathrm{Im}\{K'\,\mathbf{E}_1\cdot\mathbf{E}^0(\mathbf{r}-\mathbf{r}_{tip})\}\ , \tag{7.25}$$

where K and K' are constants that depend on the set-up. This formalism should be useful in designing and evaluating the performances of near-field experiments. A number of features can be studied with this approach, such as the role of polarization, illumination or detection scheme, etc. Moreover, the concept of impulse response provides a quantitative tool to compare the potential capabilities of different SNOM set-ups.

8. Influence of Coherence

Coherence theory allows to discuss the statistical properties of random fields [167]. For our purpose, we distinguish two main classes of random fields. On one hand, thermally produced fields are random in nature. Indeed, they are produced by the uncorrelated radiative desexcitation of a large number of excited states. On the other hand, a fully coherent beam can acquire a random nature after being scattered by a random system (e.g., a particle suspension or a rough surface). The coherence of a field is measured by the degree of correlation of the field with itself at different points and/or different times. We can distinguish between spatial coherence and temporal coherence. Spatial coherence is fully characterized by the correlation function $\left\langle E_\alpha(\mathbf{r},t)E^*_\beta(\mathbf{r}',t)\right\rangle$ whereas temporal coherence is described by the correlation function at the same point and different times $\left\langle E_\alpha(\mathbf{r},t)E^*_\beta(\mathbf{r},t')\right\rangle$, where the brackets denote an ensemble average. In what follows, we are mostly concerned by the spatial coherence of the light. In the first subsection, we study the structure of the near field above a slightly rough surface, illuminated by a plane wave. We show that the near-field intensity acquires a *wavy structure* that has been observed experimentally [168, 169, 146, 112], and which is in fact closely related to the statistical properties of the surface itself. In the second subsection, we analyse the role of spatial coherence in the image formation process, when using a microscope operating in the collection mode. We show that the best illumination conditions are obtained when using a partially *spatially incoherent symmetric beam*, as has been first observed experimentally by Chabrier *et al.* [144]. This discussion is intimately connected to the discussion in §6.D.

A. Near field and coherence

The goal of this section is to analyse the link between a random rough surface and the near field. Thus, we consider a surface illuminated by a beam. We do not account for any imaging system in this study. Our aim is to describe statistically the structure of the near-field intensity. We introduce the main concepts used for the statistical description of a random rough surface and a random field. Then, we derive a relationship between these quantities. A consequence is that the knowledge of the statistical properties of the near field yields some information on the statistical properties of the surface. One could imagine reconstructing the statistical parameters of a surface from a measurement of the near-field intensity. The advantage of such a near-field technique is that it does not suffer from the frequency cut-off at ω/c that exists in far-field techniques [170].

(i) Statistical description of a rough surface

The statistical description of a random rough surface is based on the assumption that the surface is a random process, with zero mean-value. Let us consider a particular realization of the surface, described by the function $z = S(\mathbf{r}_{||})$. The probability of having $h < S(\mathbf{r}_{||}) < h + dh$ is given by $P(\mathbf{r}_{||}, h)dh$, where $P(\mathbf{r}_{||}, h)$ is the height probability density function. In what follows, we assume that the process is homogeneous, so that the probability density function is isotropic and translationally invariant. The rms height δ can be derived from the knowledge of $P(h)$ as

$$\delta^2 = \langle S(\mathbf{r}_{||})S(\mathbf{r}_{||})\rangle = \int_{-\infty}^{\infty} h^2 P(h)\, dh \ , \tag{8.1}$$

where the brackets denote an ensemble average. For convenience, we assume that the probability density function is Gaussian, as is commonly observed for natural surfaces. In order to compute higher moments of the random process, such as the correlation function $< S(\mathbf{r}_{||})S(\mathbf{r}'_{||}) >$, we need the joint probability density function $P(|\mathbf{r}_{||,1} - \mathbf{r}_{||,2}|, h_1, h_2)$, which only depends on the distance $|\mathbf{r}_{||,1} - \mathbf{r}_{||,2}|$, as a consequence of the isotropy and translational invariance of the random process. With this new quantity, the second moment can be evaluated :

$$\langle S(\mathbf{r}_{||,1})S(\mathbf{r}_{||,2})\rangle = \int_{-\infty}^{\infty} dh_1 \int_{-\infty}^{\infty} dh_2\, h_1 h_2 P(|\mathbf{r}_{||,1} - \mathbf{r}_{||,2}|, h_1, h_2) \ . \tag{8.2}$$

The correlation function $C(|\mathbf{r}_{\|,1} - \mathbf{r}_{\|,1}|)$ is then defined by

$$\langle S(\mathbf{r}_{\|,1}) S(\mathbf{r}_{\|,2}) \rangle = \delta^2 C(|\mathbf{r}_{\|,1} - \mathbf{r}_{\|,2}|) \ . \tag{8.3}$$

The characteristic length a of the correlation function [e.g., $C(r) = \exp(-r^2/a^2)$] is a measure of the distance between consecutive peaks and valleys of the surface. An example of a rough surface numerically generated, with a correlation length $a = \lambda/2$ and a roughness $\delta = \lambda/200$, is shown in Fig. 19 (a).

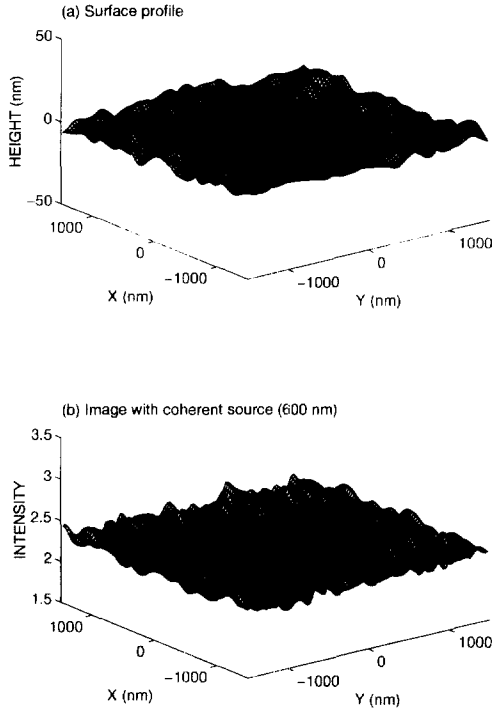

Fig. 19: *Near-field speckle pattern. (a) Random rough surface, $\delta = 0.005\lambda$, $a = 0.5\lambda$. (b) Near-field intensity along a plane separated by a distance $z_0 = 0.1\lambda$ from the mean surface. Illumination in total internal reflection. $\lambda = 600$ nm. Incident field polarized along the x-axis. The calculated speckle contrast is $C_I = 2.7\%$. Reprinted from ref.[171] with kind permission from Elsevier Science-NL, Sara Burgerhartstraat 25, 1055 KV Amsterdam, The Netherlands.*

When dealing with random processes, spectral analysis is a powerful tool. Since the surface is assumed to be translationally invariant, its Fourier transform cannot be defined. Thus, it is convenient to introduce a new function $S_A(\mathbf{r}_{||})$, which is equal to $S(\mathbf{r}_{||})$ over a finite area A and zero outside. The Fourier transform of this function is defined by

$$S_A(\mathbf{r}_{||}) = \frac{1}{4\pi^2} \int S_A(\mathbf{k}_{||}) \exp[-i\mathbf{k}_{||} \cdot \mathbf{r}_{||}] \, d\mathbf{k}_{||} \, . \tag{8.4}$$

We also introduce the Fourier transform of the correlation function

$$C(\mathbf{r}_{||}) = \int g(\mathbf{k}_{||}) \exp[i\mathbf{k}_{||} \cdot \mathbf{r}_{||}] \, d\mathbf{k}_{||} \, . \tag{8.5}$$

The Wiener-Khinchin theorem [167] relates the power spectral density of the surface and the fourier transform of the correlation function :

$$\lim_{A \to \infty} \frac{\langle S_A(\mathbf{k}_{||}) S_A(-\mathbf{k}_{||}) \rangle}{A} = 4\pi^2 \delta^2 g(\mathbf{k}_{||}). \tag{8.6}$$

(ii) Statistical description of the intensity

Upon reflection or transmission at a rough surface, a plane wave acquires a random complex amplitude. Thus, the intensity pattern displays bright and dark spots. In far-field imaging, this type of scattering pattern is usually called a *speckle pattern* [172, 173]. It is characterized by the probability density function of the intensity and by its correlation function. Since it is not a zero mean value process, a useful parameter is the speckle contrast C_I, defined as the fluctuation of the intensity, normalized by the mean intensity :

$$C_I = \frac{\sqrt{\langle I^2 \rangle - \langle I \rangle^2}}{\langle I \rangle} \tag{8.7}$$

To characterize the granular structure of the speckle pattern, it is also convenient to introduce the intensity correlation function :

$$\langle I(\mathbf{r}_{||}) I(\mathbf{r}'_{||}) \rangle \tag{8.8}$$

A remarkable property of the so-called *fully developped* speckle pattern, obtained in the far field for a roughness larger than a wavelength, is that the intensity has Gaussian statistics, which

means that the speckle is an universal phenomenon, independent of the specific properties of the medium that scatters the light. In addition, the correlation function (the size of the speckle grain) only depends on the width of the illuminating beam.

In this section, we are concerned with speckle patterns in the near field, produced by surfaces with a rms roughness smaller than $\lambda/10$. More specifically, we study the near-field intensity along a plane parallel to the surface, at a subwavelength distance z_0. When z_0 increases, one passes continuously from the near-field regime to the far-field regime, as the evanescent waves disapears. We show that, in the near field, the structure of the speckle pattern is closely related to the structure of the surface.

(iii) Link between the statistical properties of the surface and the field

Using first-order perturbation theory, we have established in §5 a linear relation between the field and an equivalent surface profile, connecting the topography and the dielectric contrast of the sample. In what follows, we limit the discussion to rough surfaces separating two homogeneous media. We focus our attention on the random structure of the near-field intensity due to to the random roughness of the sample. However, the concept of equivalent surface profile allows us to extend all the results to inhomogeneous rough surfaces.

For the sake of brevity, we only present the main results. A detailed derivation is reported in ref. [171]. To first order in the parameter δ/λ, the intensity in a plane $z = z_0$ can be cast in the form :

$$I(\mathbf{r}_{\|}) = I^{(0)}(\mathbf{r}_{\|}) + I^{(1)}(\mathbf{r}_{\|}) \ , \tag{8.9}$$

where $I^{(0)} = |\mathbf{E}^{(0)}|^2$ and $I^{(1)} = 2\mathrm{Re}[\mathbf{E}^{(0)*} \cdot \mathbf{E}^{(1)}]$. $\mathbf{E}^{(0)}$ and $\mathbf{E}^{(1)}$ are the illuminating and scattered fields, respectively (see §5). In the case of an illumination by a plane wave, $I^{(0)}$ is a function of z_0 only. At this point, we can obtain a fundamental result. Since the field is linearly related to the surface, the intensity is linearly related to the surface. Thus, if the height probability density function is Gaussian, then the intensity probability distribution function is also Gaussian. We now analyse the intensity correlation function. Inserting (8.9) in (8.8) yields :

$$\langle I(\mathbf{r}_{\|}) I(\mathbf{r}'_{\|}) \rangle = \langle I^{(0)}(z_0) I^{(0)}(z_0) \rangle + \langle I^{(1)}(\mathbf{r}_{\|}) I^{(1)}(\mathbf{r}'_{\|}) \rangle \ . \tag{8.10}$$

The first term is a constant, whereas the second term is a second-order dynamic contribution.

Note that it is unnecessary to account for the second-order terms of the form $\langle I^{(0)}(z_0)I^{(2)}(\mathbf{r}')$. $I^{(0)}$ being deterministic, these terms leads can be factorised into $\langle I^{(0)}(z_0)\rangle\langle I^{(2)}(\mathbf{r}')$, and give a second-order contribution to the constant term in (8.10). We seek a relationship between the surface profile and the intensity in the speckle pattern. By inserting the expression of the scattered field in Eq. (8.9), the correlation function can be cast in the form [171] :

$$\langle I^{(1)}(\mathbf{r}_{||})I^{(1)}(\mathbf{r}'_{||})\rangle = \mathrm{Re}\left\{\int \delta^2 g(\mathbf{k}_{||})\tilde{F}(\mathbf{k}_{||},\mathbf{k}_{||}^{inc},\omega,z_0)\exp[i\mathbf{k}_{||}\cdot(\mathbf{r}_{||}-\mathbf{r}'_{||})]\,d\mathbf{k}_{||}\right\} . \qquad (8.11)$$

In other words, *the power spectral density of the speckle pattern is the power spectral density of the surface multiplied by a filter* $\tilde{F}(\mathbf{k}_{||},\mathbf{k}_{||}^{inc},\omega,z_0)$. The detailed expression of the filter is given in ref.[171]. For the following discussion, we only need to know that the filter is explicitly given in terms of the frequency, the wave vector of the incident beam and the distance z_0 between the observation plane and the surface.

A simple result that appears at once by inspection of (8.11) is that the intensity correlation function varies as δ^2. Thus, *the speckle contrast varies like the rms roughness* δ. A numerically generated example of near-field speckle intensity is shown in Fig. 19(b). The corresponding surface is shown in Fig. 19(a). It is seen that the speckle pattern is not isotropic, which is due to the fact that the illuminating beam is oblique, breaking the symmetry of the physical system. When a symmetric illumination is used, this feature disappears.

It is well-known that the speckle contrast decreases when the coherence of the illuminating beam decreases. This is fully described by the filter function $\tilde{F}$ through its dependence on the frequency and the incident wave vectors. For a partially coherent beam, one should merely add the contributions of the different frequencies and wave vectors. We show in Fig. 20 a set of numerically simulated images of a small block on which a random roughness has been superimposed. The surface is displayed on Fig. 20(a), and images with different coherence states are shown in Fig. 20 (b,c,d). It is clearly seen that the speckle contrast is considerably reduced when the coherence of the illumination decreases.

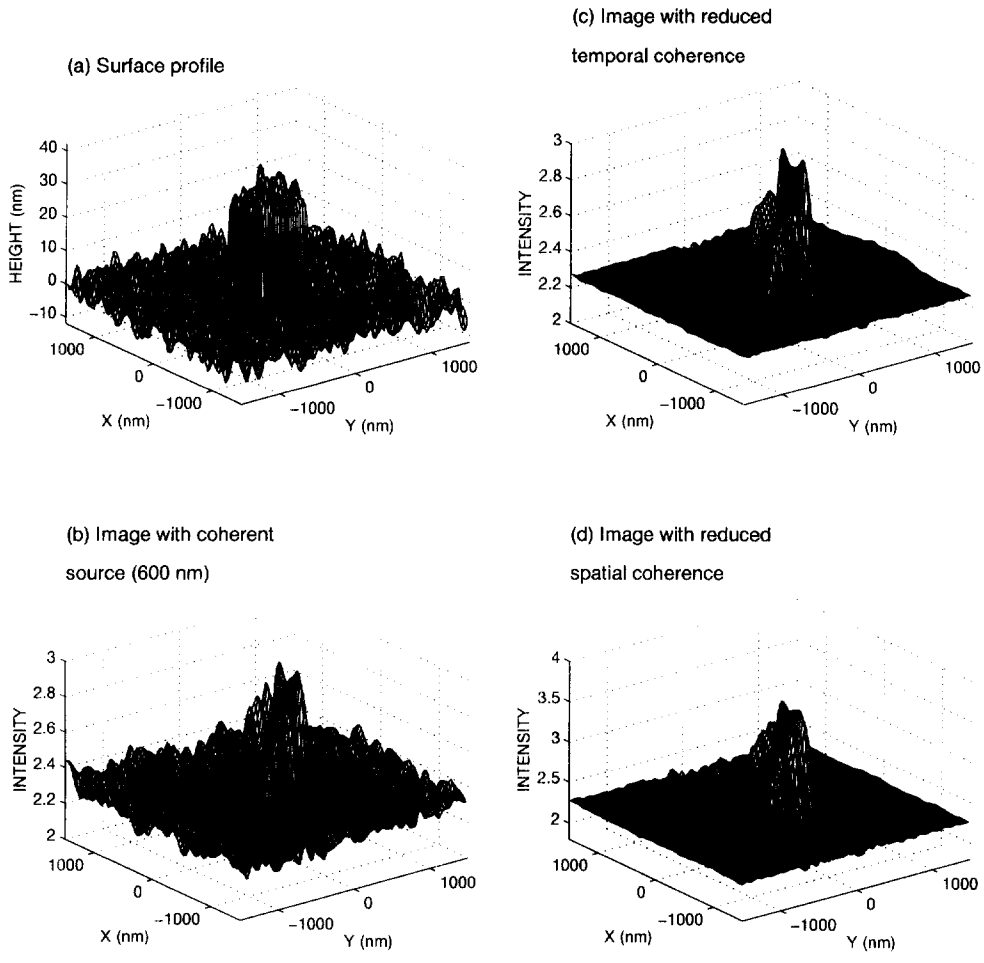

Fig. 20: *Near-field intensity above a surface composed of a deterministic profile (parallelepiped with side=1200 nm, height=15 nm and index n=1.5) and a randomly rough component (rms height $\delta = 1$ nm, correlation length a : 120 nm). (a) Surface profile. (b) Near-field intensity in a plane at a distance $z_0 = 30$ nm for a coherent illumination (plane wave in total internal reflection, $\lambda = 600$ nm). (c) Near-field intensity in a plane at a distance $z_0 = 30$ nm for a non-monochromatic wave. The spectrum is flat in the range 400-800 nm. (d) Near-field intensity in a plane at a distance $z_0 = 30$ nm for an illumination by a set of monochromatic uncorrelated plane waves polarized along the x-axis, with angles of incidence in the range (45°, 65°). The plane of incidence is $y - z$.Reprinted from ref.[171] with kind permission from Elsevier Science-NL, Sara Burgerhartstraat 25, 1055 KV Amsterdam, The Netherlands.*

B. Transfer function and coherence

In the preceding section, we analysed the influence of the coherence on the near-field speckle. We now address the influence of coherence on the image formation process, and emphasize the importance of the coherence of the illumination beam on the near-field intensity. Thus, we summarize the work presented in ref.[145].

To address this problem, we use the concept of impulse response introduced in §5. For plane-wave illumination, the near-field intensity can be cast in the form :

$$I^{(1)}(\mathbf{r}_{||}, z_0) = \int H_c(\mathbf{r}_{||} - \mathbf{r}'_{||}, \mathbf{k}_{||}^{inc}, \omega, z_0) S(\mathbf{r}'_{||})\, d\mathbf{r}'_{||} \ , \tag{8.12}$$

where the coherent impulse response H_c was introduced in §5. Its analytic expression in Fourier space is given in Appendix E. In order to obtain the near-field intensity for an incoherent illumination, we need to add the intensities for each plane wave and each frequency. Here, we assume that the angular spectrum of the incident field is delta correlated [145], so that two different plane waves do not interfere. As usual, at optical frequencies, different frequencies do not interfere. Taking into account the linearity between the intensity and the impulse response H_c, it is clear that we can define an incoherent impulse response :

$$H_{in}(\mathbf{r}_{||} - \mathbf{r}'_{||}, z_0) = \int d\omega \int d\mathbf{k}_{||}^{inc}\, H_c(\mathbf{r}_{||} - \mathbf{r}'_{||}, \mathbf{k}_{||}^{inc}, \omega, z_0) \ , \tag{8.13}$$

which describes the link between the near-field intensity in a plane $z = z_0$ and the surface profile.

In Fig. 21, we illustrate the importance of the coherence of the illumination. In fact, we consider only spatial coherence, since we use a single frequency. In other words, we represent the impulse response for a partially coherent monochromatic and symmetric illumination, and compare it with the coherent impulse response. It turns out that, with this kind of illumination, the impulse response is no longer disymmetric and its width is considerably reduced. Thus, the incoherent illumination yields better images than the illumination by a single homogeneous or inhomogeneous plane wave [174, 145]. Note also in Fig. 21 the better resolution obtained by using p-polarization.

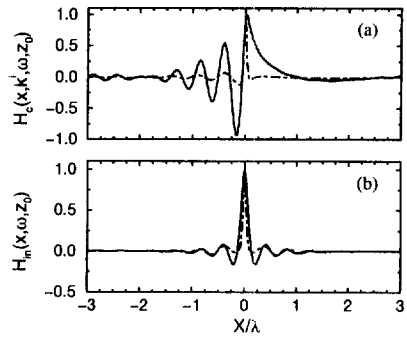

Fig. 21: *Impulse response normalized by its peak value.* $z_0 = \lambda/20$. *(a) Coherent impulse response for an illumination in total internal reflection, with an angle of incidence* $\theta_i = 45^o$. *(b) Incoherent impulse response obtained by integrating over all directions of incidence. The frequency is fixed. Solid curves : s-polarization, dashed curves : p-polarization. From ref.[145].*

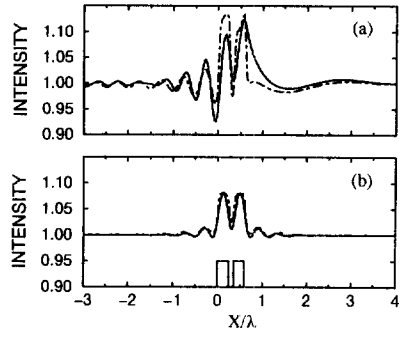

Fig. 22: *Near-field intensity above a 2D structure at a height* $z_0 = 0.0625\lambda$. *The structure contains two ridges with height* $h = \lambda/40$, *width* $w = \lambda/4$, *separated by a distance* $l = \lambda/10$ ($\lambda = 633$ *nm*). *(a) Near-field intensity for a plane wave illumination in total internal reflection. (b) Near-field intensity for a spatially partially incoherent illumination obtained by adding 43 different plane waves equally spaced between* -84^o *and* 84^o. *Solid curves : s polarization, dashed curves : p-polarization. The structure is represented at the bottom of figure (b). From ref.[145].*

To illustrate the importance of coherence for imaging, we show the intensities above the structure depicted in Fig. 22(a), for coherent illumination [Fig. 22(b)], and for incoherent illumination [Fig. 22(c)]. The calculations are done by solving exactly the Lippman-Schwinger equation (5.1), as described in refs.[78, 79]. It is obvious that the images are better resolved and easier to interpret when an incoherent illumination is used. It is also important to note that imaging in detection mode with *spatially partially incoherent illumination* is equivalent to imaging in the illumination mode *with a laser source* (see §6), which explains why the images produced with an illumination-mode SNOM are easier to interpret than the images obtained by a PSTM/STOM set-up. Finally, we mention that a partially coherent and symmetric illumination was achieved experimentally in a modified PSTM/STOM set-up. It was observed that the distorsion between images and objects was reduced, in comparison to a coherent illumination [144], as the theory presented here explains.

9. Operating Modes and Artifacts

A crucial point in SNOM is the regulation of the tip-sample distance. In order to obtain subwavelength resolution, this distance must be much less than the wavelength (typically of the order of a few nanometers). Three operating modes have been used so far. We compare these modes, and pay particular attention to the coupling between the optical signal and the motion of the tip induced by the regulation mechanism.

A. Three operating modes

Let $\mathcal{S}(x, y, z)$ be the optical signal detected when the tip is located at the point (x, y, z), the z-axis being normal to the sample mean surface. We do not need to describe, at this stage, the type of microscope that is used. It may be an illumination-mode, collection-mode or apertureless SNOM. In order to regulate the subwavelength tip-sample distance, three operating modes have been used : 1) In the *constant-height mode*, the tip is moved in the plane $z = z_0$, and one records the optical signal $\mathcal{S}(\mathbf{r}_{||}, z_0)$. 2) In the *constant-intensity mode*, the optical signal $\mathcal{S}$ is kept constant with a feedback system, forcing the tip to follow a surface $z = h(\mathbf{r}_{||})$, which does not, in general, reproduce the topography of the sample. Recording the motion of the tip [i.e., the surface $z = h(\mathbf{r}_{||})$] produces the image. This mode has been used extensively

in PSTM/STOM [66, 67, 68]. 3) In the *constant-distance mode*, the tip is forced to follow a surface $z = f(\mathbf{r}_{||})$ by an auxiliary non-optical distance-control mechanism. The optical signal that is recorded is $\mathcal{S}[\mathbf{r}_{||}, f(\mathbf{r}_{||})]$. The distance-control mechanism can use the tunneling current of an STM [3, 61], an AFM [109] or the lateral friction forces between the tip and the sample (shear forces) [57, 56]. In all these cases, we assume that the tip follows the topography of the sample, and $f(\mathbf{r}_{||})$ is the convolution of the sample profile by a function which describes the probe geometry [155]. Nevertheless, recent experimental studies of the shear-force mechanism have shown that the shear-force images do not always reproduce the sample topography [58, 59].

Our purpose is to compare the images obtained in these different modes. We first prove that the constant-height and constant-intensity modes are equivalent, in the sense that they lead to proportional images. Then, we concentrate on the constant-distance mode to study the presence of artifacts induced by the z-motion of the probe. The presence of such artifacts was put forward a few years ago [33], and received particular attention recently [58, 155, 175, 156, 88, 157]. We study how the presence of this artifact depends on the sample features (topography and dielectric contrast) and on the illumination conditions, and also show how it may misguide the evaluation of the optical resolution.

B. Equivalence of constant-height and constant-intensity images

(i) Analytical proof

Let us assume that the SNOM tip scans a sample with both topography and dielectric contrast (see Fig. 23). We split the optical signal into two contributions :

$$\mathcal{S}(\mathbf{r}_{||}, z) = \mathcal{S}^{(0)}(z) + \mathcal{S}^{(1)}(\mathbf{r}_{||}, z) \ . \tag{9.1}$$

$\mathcal{S}^{(0)}$ is the signal detected when the tip scans the flat substrate, without any sample. It is a function of the tip-substrate distance z only. $\mathcal{S}^{(1)}$ is the signal resulting from the presence of the sample, i.e., from a scattering process. It is given by (7.12) for collection-mode SNOM, by (7.19) for illumination-mode SNOM and apertureless SNOM. If the sample is a weak scatterer (as usually in SNOM), we can assume that $|\mathcal{S}^{(1)}| = \eta |\mathcal{S}^{(0)}|$, with $\eta \ll 1$.

When the tip scans the sample in the constant-intensity mode, it follows a surface $z = h(\mathbf{r}_{||})$ along which the optical signal is constant. We write $h(\mathbf{r}_{||}) = z_0 + \delta h(\mathbf{r}_{||})$ where $z_0 =< h(\mathbf{r}_{||}) >$,

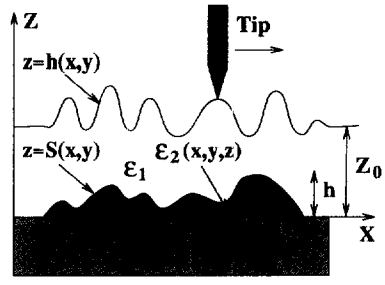

Fig. 23: *Geometry of the problem.* $z = S(x, y)$ *is the surface profile of the inhomogeneous sample, and* $z = h(x, y)$ *is a constant-signal surface.*

the brackets denoting the background value of a function of $\mathbf{r}_{\|}$. For a flat portion of substrate without sample, the scattered signal $\mathcal{S}^{(1)}$ is zero, so the total signal is given by $\mathcal{S}^{(0)}(z_0)$ (see Fig. 23). Along $z = h(\mathbf{r}_{\|})$, the total optical signal remains constant and one can write for each $\mathbf{r}_{\|}$

$$\mathcal{S}^{(0)}[h(\mathbf{r}_{\|})] + \mathcal{S}^{(1)}[\mathbf{r}_{\|}, h(\mathbf{r}_{\|})] = \mathcal{S}^{(0)}(z_0) \ . \tag{9.2}$$

We now perform a first-order expansion around $z = z_0$ in (9.2) :

$$\frac{d\mathcal{S}^{(0)}}{dz}(z_0)\,\delta h(\mathbf{r}_{\|}) + \mathcal{S}^{(1)}(\mathbf{r}_{\|}, z_0) + \frac{\partial \mathcal{S}^{(1)}}{\partial z}(\mathbf{r}_{\|}, z_0)\,\delta h(\mathbf{r}_{\|}) = 0 \ . \tag{9.3}$$

In order to compare the three terms in (9.3), let us introduce the length scales L_0 and L_1 of $\mathcal{S}^{(0)}$ and $\mathcal{S}^{(1)}$, respectively. The three terms in (9.3) are of the order of $\delta h/L_0$, η and $\eta\delta h/L_1$, where $\delta h = \sup|\delta h(\mathbf{r}_{\|})|$. With the condition $\delta h \ll L_1$, the third term can be neglected, so that $\delta h \sim \eta L_0$, which leads to a simple equation for the constant-intensity surface :

$$h(\mathbf{r}_{\|}) = \frac{-\mathcal{S}^{(1)}(\mathbf{r}_{\|}, z_0)}{\dfrac{d\mathcal{S}^{(0)}}{dz}(z_0)} + z_0 \ . \tag{9.4}$$

Equation (9.4) shows that the image obtained by scanning in the constant-intensity mode and recording the vertical position $h(\mathbf{r}_{\|})$ of the tip is proportional to the image obtained by scanning along a line at a constant height $z_0 =< h(\mathbf{r}_{\|}) >$ and recording the optical signal [whose dynamic part is $\mathcal{S}^{(1)}(\mathbf{r}_{\|}, z_0)$].

Equation (9.4) is valid if the conditions $\delta h \sim \eta L_0$ and $\delta h \ll L_1$ are fulfilled. L_0 is the z-variation of the signal resulting of the interaction between the tip and the flat substrate. It is

roughly given by the decay length of the evanescent waves involved in this interaction. These evanescent waves are created by the scattering at the tip apex or aperture. If a is the radius of curvature of the apex, or the aperture diameter, we have $L_0 \sim a$. The condition $\delta h \sim \eta L_0$ implies that the amplitude of the z-motion of the probe must be smaller than the decay length of the field illuminating the sample, which seems reasonable from a practical point of view, because if $\delta h > L_0$, one may loose the optical signal. The second condition $\delta h \ll L_1$ is more severe. L_1 is the decay length of the evanescent waves involved in the scattering process between the tip and the sample, and, in general, its evaluation is very difficult. Nevertheless, if the first Born approximation can be used to describe the field inside the sample (see §7), $\mathcal{S}^{(1)}$ results from the convolution product of the dielectric contrast by the illuminating field, whose lateral extension is of the order of a. Thus, the smallest decay length of the evanescent waves involved in the tip-sample scattering process is of the order of a. In this case, we can take $L_1 \sim a \sim L_0$, and the two conditions $\delta h \sim \eta L_0$ and $\delta h \ll L_1$ are automatically fulfilled.

(ii) Numerical illustration

To illustrate the result in (9.4), we present two-dimensional numerical calculations obtained by an exact method, described in refs.[78, 79]. It consists in solving the volume integral equation (5.1) by a moment method [118], without any approximation.

We consider a collection-mode geometry, where the sample is illuminated in transmission by an extended field, and the detection is performed by a local probe in the upper medium. In order to simplify the calculation, we do not take into account the coupling with the probe in the calculation. We calculate the square modulus of the electric field at the tip position, assuming that the true signal would be linearly related to this quantity. We know, from §7, that it is true for s-polarization only. For p-polarization, some polarization effects in the coupling with the probe modes may enhance the contribution of one component of the electric field. Nevertheless, we present the results for both polarizations, assuming that the conclusion remains valid with the polarization-dependent probe coupling taken into account.

We consider two different samples. The first one [Fig. 24(a)] consists of two gold particles of width $w = \lambda/10$, height $h = \lambda/40$ and dielectric constant $\epsilon_2 = -8.7 + 1.3i$ embedded in a glass substrate ($\epsilon_3 = 2.25$). The particles are separated by a distance $l = \lambda/8$. This sample

is illuminated in transmission by a monochromatic plane wave ($\lambda = 633$ nm) with an angle of incidence $\theta_i = 45^o$, corresponding to an illumination in total internal reflection. The second sample [Fig. 24(b)] is a glass substrate presenting two ridges of rectangular cross-section. The ridges are of height $h = \lambda/40$ and width $w = \lambda/4$, and are separated by a distance $l = \lambda/10$. This sample is illuminated in transmission by a monochromatic partially coherent light ($\lambda = 633$ nm) composed of 43 uncorrelated plane waves with different angles of incidence equally spaced between -84^o and 84^o.

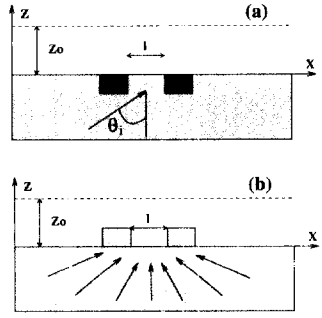

Fig. 24: *Samples used in the numerical simulations. (a) Two gold particles width $w = \lambda/10$, height $h = \lambda/40$ embedded in a glass substrate. $l = \lambda/8$. Illumination in transmission with one plane wave ($\lambda = 633$ nm, $\theta_i = 45^o$). (b) Glass substrate with two ridges of rectangular cross-section. The ridges are of height $h = \lambda/40$ and width $w = \lambda/4$. $l = \lambda/10$. Illumination in transmission with monochromatic spatially incoherent light ($\lambda = 633$ nm). From ref.[154].*

We show in Fig.25(a) two constant-intensity curves (dashed lines, right-hand side ordinate axis) and two constant-height curves (solid lines, left-hand side ordinate axis) calculated above the sample in Fig. 24(a). The light is p-polarized. In the constant-height calculations (solid lines), the height is $z_0 = 0.03\lambda$ for the curve numbered (1) and $z_0 = 0.09\lambda$ for the curve numbered (2). These heights correspond to the mean value of the constant-intensity curves numbered (1) and (2), respectively. According to (9.4), the two curves (1) (and the two curves 2) are proportional, which confirms that constant-intensity and constant-height images are similar, even with inhomogeneous samples without topography. Note also that the curves at different heights are not the same. In curves (1), the separation between the particles is better

defined than in curves (2). In fact, curves (2) correspond to a detecting distance greater than that in curves (1) and high spatial frequencies of the scattered field have been attenuated.

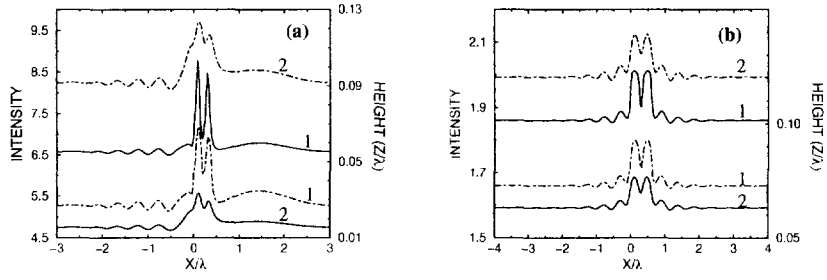

Fig. 25: *(a) Constant-intensity curves (dashed lines, right-hand side ordinate axis) and constant-height curves (solid lines, left-hand side ordinate axis) with $z_0 = 0.03\lambda$ (curve 1) and $z_0 = 0.09\lambda$ (curve 2). p-polarization. The sample and the illuminating field are described in Fig. 24(a). (b) Same as (a) with $z_0 = 0.07\lambda$ (curve 1) and $z_0 = 0.12\lambda$ (curve 2). The sample and the illuminating field are described in Fig. 24(b). From ref.[154].*

We show in Fig. 25(b) the same curves calculated above the sample in Fig.24(b). The sample is a purely topographic surface, under partially spatially coherent monochromatic illumination. In the constant-height calculations (solid lines), the height is $z_0 = 0.07\lambda$ for curve (1) and $z_0 = 0.12\lambda$ for curve (2), corresponding to the mean value of the constant-intensity curves (1) and (2), respectively. The light is p-polarized. As in Fig. 25(a), the two curves (1) (and the two curves 2) are proportional. The spatially incoherent illumination provides an image that closely follows the surface profile, as was recently demonstrated in the constant-height mode [145]. This is clearly visible in Fig. 25(b), for both constant-height and constant-intensity curves.

We present in Fig. 26 the same results for s-polarization. The conclusions still hold. Note that in the case of the gold particles under coherent illumination [Fig. 26(a)], the image is less resolved than for p-polarization [Fig. 25(a)]. Even the presence of two particles is not clear in the image in s-polarization.

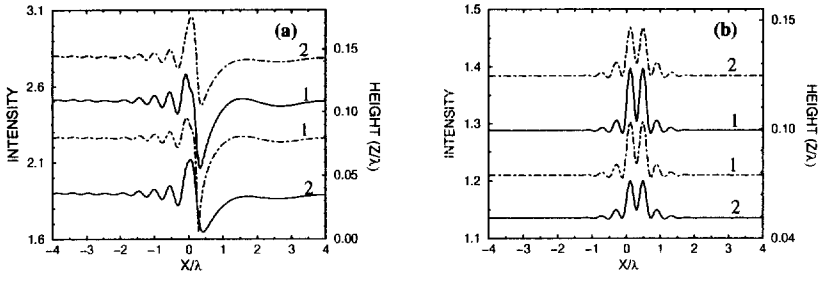

Fig. 26: *Same as Fig. 25 for s-polarization.*

(iii) Conclusion

Using a perturbation theory, we have proved analytically that for collection-mode SNOM, the constant-intensity image $z = f(\mathbf{r}_{||})$ is proportional to the constant-height image at the height $z_0 =< f(\mathbf{r}_{||}) >$. The demonstration applies to both inhomogeneous and purely topographic samples, and the conclusion was checked with rigorous numerical calculation. There is no reason to prefer the constant-height or the constant-intensity mode, except for practical reasons. For example, the constant-intensity mode provides a way to regulate the tip-sample distance, and is widely used in STOM/PSTM.

C. Constant-distance mode and z-motion artifacts

We now turn to the study of the constant-distance mode. In this mode, the z-motion of the probe, induced by a non-optical distance regulation mechanism, can couple to the purely optical information of the image, which creates an artifact that may lead to a wrong interpretation of the images. It was demonstrated experimentally in ref.[33] that the SNOM image may contain two different contributions : a purely optical one and one reproducing the motion of the tip. Recently, some systematic experimental studies of this artifact were reported [155, 175, 88], and showed that *many experimental images previously reported might be dominated by non-optical contrast mechanisms.* It was also demonstrated that the artifact depends on the sample under study, and the set-up that is used. We try to understand the physical origin and the importance of the artifact as a function of the illumination conditions and the sample features. To proceed,

we use first-order pertubation theory. Then, we illustrate the conclusions by exact numerical simulations.

(i) Perturbative approach

In the constant-distance mode, let $z = f(\mathbf{r}_{\|}) = z_0 + \delta f(\mathbf{r}_{\|})$ be the path followed by the tip, with $z_0 = \langle f(\mathbf{r}_{\|}) \rangle$. The optical signal $\mathcal{S}$ along this path is

$$\mathcal{S}[\mathbf{r}_{\|}, z = f(\mathbf{r}_{\|})] = \mathcal{S}^{(0)}[f(\mathbf{r}_{\|})] + \mathcal{S}^{(1)}[\mathbf{r}_{\|}, z = f(\mathbf{r}_{\|})] \,, \tag{9.5}$$

where $\mathcal{S}^{(0)}$ is the signal that would exist without the sample (for the tip in front of a flat substrate) and $\mathcal{S}^{(1)}$ is the signal resulting from the interaction with the sample. As in §9.B, we perform an expansion around z_0 to first order in $\delta f = \sup|\delta f(\mathbf{r}_{\|})|$ (see §9.B for a discussion of the validity of this expansion). This leads to :

$$\mathcal{S}[\mathbf{r}_{\|}, z = f(\mathbf{r}_{\|})] = \mathcal{S}^{(0)}(z_0) + \frac{d\mathcal{S}^{(0)}}{dz}(z_0)\,\delta f(\mathbf{r}_{\|}) + \mathcal{S}^{(1)}(\mathbf{r}_{\|}, z_0) \tag{9.6}$$

The first term in (9.6) is the background value of the image, the third one gives the purely optical contrast, and the second one reflects the coupling between the optical information and the z-motion of the tip. Due to this term, the optical image will depend on the path followed by the tip $\delta f(\mathbf{r}_{\|})$. This problem is avoided in the constant-height mode, because $\delta f(\mathbf{r}_{\|}) = 0$, so the second term in (9.6) always vanishes.

Let us consider a worst case scenario. Suppose that the second term dominates the right-hand side in (9.6). Then, the detected signal $\mathcal{S}$, given by the left-hand side of Eq. (9.6), is proportional to the motion of the tip $\delta f(\mathbf{r}_{\|})$. The resulting image is mainly *an optical read-out of the motion of the tip*, as that obtained with a conventional AFM, and does not contain any information on the optical properties of the sample. Moreover, the resolution of such an image does not result from optical mechanisms, but only from the interaction used to control the tip-sample distance.

This dramatic artifact was pointed out by several authors [33, 58, 155, 175, 88]. It is a consequence of the z-dependence of the optical signal $\mathcal{S}$, which has two (generally coupled) origins : 1) the z-dependence of the coupling efficiency with the probe, and 2) the z-dependence of the near field itself. Thus, as pointed out in ref.[155], the problem is very complicated

because the presence of the artifact will depend on the probe feature, on the sample properties (topography and dielectric contrast) and on the illumination/detection conditions.

To proceed further and obtain some physical insight, we simplify the problem. We deal with a collection-mode configuration. We know that it is not a limitation, because reciprocity can be used to extend all the results to an illumination-mode system (see §6). We also assume that the field inside the sample is correctly described by the first Born approximation (see §5, §7). In this case, the problem of the scattering of the illuminating field by the sample (leading to the induced polarization field inside the sample) and the problem of the detection of this polarization field by the probe can be decoupled. In this study, we choose to concentrate on the artifact induced by the scattering process by the sample. Moreover, the precise description of the coupling with the probe would require the knowledge of the field emitted by the probe when working in illumination mode (see §7). Thus, we will not consider the coupling with the probe, and calculate the square modulus of the electric field without the presence of the detecting tip. But we should keep in mind that the z-dependence of the coupling with the probe may induce an additional artifact, which is not studied here.

We consider a three-dimensional sample with variations in both topography and dielectric constant (see Fig. 27). The system is illuminated in transmission or reflection by a monochromatic field of wavelength λ, which is either a plane wave (coherent illumination) or a set of uncorrelated plane waves (spatially partially coherent illumination). In both cases, the illuminating intensity $I^{(0)}$ depends only on z.

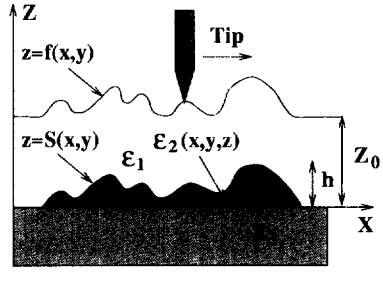

Fig. 27: *Scattering geometry.* $z = S(x, y)$ *is the surface profile of the inhomogeneous sample and* $z = f(x, y)$ *is the path followed by the tip.*

A first-order expansion of the intensity along the path followed by the tip leads to :

$$I[\mathbf{r}_{\|}, z = f(\mathbf{r}_{\|})] = I^{(0)}(z_0) + \frac{dI^{(0)}}{dz}(z_0)\,\delta f(\mathbf{r}_{\|}) + I^{(1)}(\mathbf{r}_{\|}, z_0)\ , \tag{9.7}$$

by a procedure analogous to that used to obtain (9.6). The first term on the right-hand side contributes to the background of the image. The second one is proportional to the path followed by the tip and leads to the artifact. The third one contains the purely optical information. Its relationship with the sample properties (topography and dielectric contrast) can be described with the concept of equivalent surface profile (see §5).

(ii) Optical content of the image

The right-hand side in (9.7) shows that the most general image is a superposition of a purely optical signal (third term) and a term proportional to the path followed by the tip (second term). The relative weight of these two terms determines the optical content of the near-field optical image. If the second term dominates, the use of NFO does not add any information to the AFM, STM or shear force images. NFO is of interest only in the situations in which the contrast is dominated by the third term in (9.7).

Proceeding as in ref.[155], we introduce the following parameter :

$$\Gamma = \frac{1}{I^{(1)}(\mathbf{r}_{\|}, z_0)}\ \frac{dI^{(0)}}{dz}(z_0)\ , \tag{9.8}$$

which is a measurement of the optical content of the image. $\Gamma = 0$ corresponds to a purely optical image. A large value of $|\Gamma|$ corresponds to a low optical content of the image, and a domination of the contrast by the artifact. In any relevant NFO image, $|\Gamma|$ should be minimized.

The value of Γ depends on both the operating conditions and the properties of the sample under study, which makes the prediction of the presence of this artifact very difficult. A systematic study has been reported in ref.[156], but is beyond the scope of the present review. Note that this study [156] is in agreement with experimental observations reported in ref.[155]. Nevertheless, (9.8) gives an idea of the expected behaviour of the images. *For a given sample*, the artifact will increase with the z-dependence of the illuminating intensity $I^{(0)}$. This z-dependence may be important for an illumination in total internal reflection, or in reflection on a very reflective substrate. We illustrate this point by numerical simulation in the following section. The complexity of the problem also lies in the dependence of the artifact on the

sample properties [155]. *For given illumination conditions*, Γ depends on the sample through the scattered intensity $I^{(1)}$ which is related to the equivalent surface profile, defined by (5.4). We keep in mind that, when the scattering potential of the sample increases (this scattering potential being well represented by the equivalent surface profile), the purely optical contribution increases. Thus, the artifact becomes less visible. We also illustrate this behaviour in the following subsection.

(iii) Numerical illustration of the presence of the artifact

We now study the near field scattered by one particle deposited on a flat semi-infinite substrate. The scattering geometry is depicted in Fig. 28. It is two-dimensional, namely, invariant in the y-direction. In order to investigate the dependence of the artifact on the illumination conditions and the scattering potential of the sample, we consider only one particle on the flat substrate. The case of two particles will be studied in the next section devoted to the resolution issue.

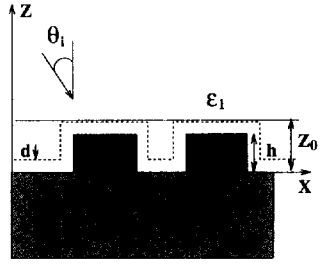

Fig. 28: *Geometry of the two-dimensional sample used in the exact numerical calculations. Solid line : path followed in the constant-height calculations. Dashed line : path followed in the constant-distance calculations.*

The system is illuminated in transmission or in reflection by a monochromatic plane wave of wavelength $\lambda = 633$ nm, with an angle of incidence θ_i. We provide exact numerical calculations [78, 79] of the total near-field intensity either at a constant height or at a constant distance from the sample. For the sake of brevity, we present the results for s-polarization. All the conclusions remains valid for p-polarization, the results being reported elsewhere [156].

We display in Fig. 29 the case of an illumination in transmission at $\theta_i = 0^o$. We have

used two different particles having the same dimensions ($w = 0.1\lambda$ and $h = 0.015\lambda$), but a different dielectric constant : (a) $\epsilon_2 = 2.25$ (glass), (b) $\epsilon_2 = -9 + i$ (gold). Varying ϵ_2 is a way to vary the scattering potential of the sample, and thus the level of the scattered intensity $I^{(1)}$ [see (5.4)]. The substrate is glass ($\epsilon_3 = 2.25$). The solid curves correspond to constant-height calculations along the solid line in Fig. 28 with $z_0 = 0.0225\lambda$ (14 nm). The dashed curves correspond to constant-distance calculations along the dashed line in Fig. 28 with $d = 0.0075\lambda$ (5 nm). The location of the particle is indicated at the bottom of the figure. For the two particles, the constant-height and constant-distance curves are similar. At normal incidence in transmission, $dI^{(0)}/dz$ vanishes. Thus, $\Gamma = 0$ and the constant-distance image is purely optical. No z-motion artifact induced by the scattering process is to be expected and the constant-height and constant-distance images are similar. Note that another kind of artifact, induced by the coupling with the probe, could exist in such a configuration.

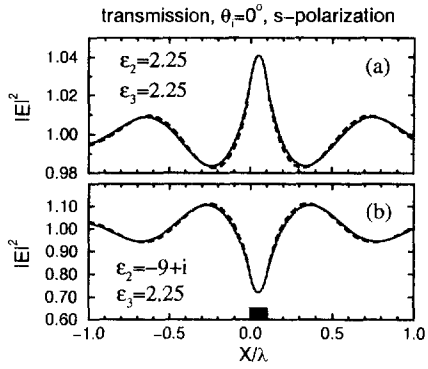

Fig. 29: *Near-field intensity above the sample in Fig. 28 with only one surface particle. s-polarization. Illumination in transmission, $\theta_i = 0^o$. Particle size : $w = 0.1\lambda$, $h = 0.015\lambda$. Solid line : constant-height calculations with $z_0 = 0.0225\lambda$. Dashed line : constant-distance calculations with $d = 0.0075\lambda$. Two values of the particle dielectric constant ϵ_2 are used. The dielectric constant of the substrate $\epsilon_3 = 2.25$ is fixed. From ref.[156].*

We show in Fig. 30 the same calculation, but for an incidence in total internal reflection ($\theta_i = 60^0$). $dI^{(0)}/dz$ and Γ do not vanish in this case. Moreover, $dI^{(0)}/dz$ is negative, and one

expects a contribution in the intensity of a term proportional to the path followed by the tip in inverse contrast [second term in (9.7)], as is clearly seen in Fig. 30. The constant-height (solid line) and constant-distance (dashed line) curves look different. A superposition of a purely optical signal (looking like a constant-height signal) and a signal proportional to the path followed by the tip in contrast reversal (artifact) can be easily identified in the constant-distance curves. The scattering potential of the particle being small ($\epsilon_2 = 2.25$), the purely optical contribution $I^{(1)}$ does not dominate the contrast of these curves. They are strongly dominated by the z-motion artifact. Experimental observations of such artifacts in contrast reversal with an illumination in total internal reflection have been reported [176, 177].

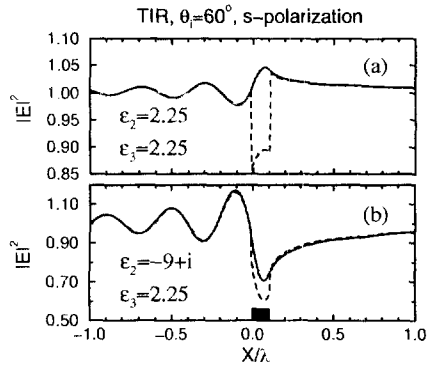

Fig. 30: *Same as Fig. 29 with an illumination in total internal reflection ($\theta_i = 60^o$). From ref.[156].*

When the scattering potential increases [Fig. 30(b), corresponding to a gold particle], the contribution of the optical term $I^{(1)}$ increases. For the gold particle, the purely optical contribution is so important that the presence of the artifact is strongly attenuated, which means that the contrast is now dominated by the third term in (9.7). The constant-height and constant-distance curves are practically identical. The presence of the artifact is no longer visible.

In summary, the presence of the artifact depends on the competition between a term sensitive to the illumination conditions, and a term sensitive to the scattering potential of the

sample. For a given sample, varying the illumination conditions, we found a situation where the contrast is dominated by the z-motion artifact, and another situation where the contrast is mainly optical. In the former case, we increased the scattering potential of the sample, to show that the presence of the artifact could disappear. Note that the same study can be performed with a reflection geometry, leading to the same conclusion [156]. The presence of the artifact is critical, in this case, when the substrate has a high reflectivity. The illuminating intensity $I^{(0)}$ is an interference pattern between the incident and the reflected field, thus providing a z-dependence in $I^{(0)}$ leading to an artifact. An experimental study of such artifacts in reflection can be found in ref. [175].

(iv) Resolution issue

In the previous section, we showed how the purely optical information of the image could be hidden by an artifact induced by the z-motion of the probe. We now discuss the resolution issue.

It was pointed out in Refs.[155, 175, 88] that the artifact may lead to a wrong interpretation of the purely optical resolution. We illustrate this important point with exact numerical calculations of the field scattered by two particles ($\epsilon_2 = 2.25$) deposited on a flat glass substrate ($\epsilon_3 = 2.25$) and separated by a distance $l = 0.075\lambda$ (47 nm). The geometry is displayed in Fig. 28. The illumination is done in transmission, with an s-polarized monochromatic plane wave ($\lambda = 633$ nm).

Figures 31(a) and (b) show the constant-height (solid line) and constant-distance (dashed line) curves for $\theta_i = 0^0$ and $\theta_i = 60^0$, respectively. In Fig. 31(a), the presence of the artifact is not visible. As shown previously, at normal incidence, the constant-height and constant-distance curves are almost identical. In contrast, in total internal reflection (Fig. 31b), the two images are clearly different. For the constant-height curve, the intensity distribution does not reproduce the surface profile. Even the presence of the two particles is not clear in this (purely optical) image. In contrast, the constant-distance curve exhibits strong variations at the precise location of the particles, with an extremely high resolution. But, these strong variations of the signal have their origin in the second term in (9.7), which is responsible for the artifact. In fact, a signal proportional to the path followed by the tip in inverse contrast can be easily

recognized in the intensity at constant distance. This is the signature of the z-motion artifact with an illumination in total internal reflection, for which $dI^{(0)}/dz$ is negative. In conclusion, the resolution in the constant-distance image does not have its origin in an optical interaction with the sample, but in an optical read-out of the z-motion of the tip.

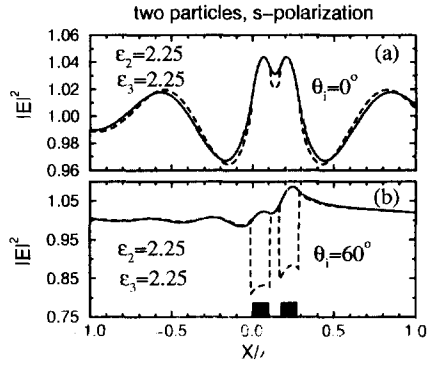

Fig. 31: *Near-field intensity above the sample in Fig. 28 with two particles separated by a distance* $l = 0.075\lambda$. *The dielectric constants* ϵ_3 *and* ϵ_2 *are fixed to a value of 2.25. (a)* $\theta_i = 0°$. *(b)* $\theta_i = 60°$. *Other parameters as in Fig. 29. From ref.[156].*

Note that another theoretical analysis of the optical resolution was presented in ref.[157]. The conclusions are similar to that presented in this section. In particular, three contributions to the signal were indentified : a z-motion artifact, a topographically induced optical contrast, and an optical contrast induced by the variations in optical properties. Concerning the resolution issue, the concept of *sensitivity window* was introduced, which determines the scale on which the contrast (and thus the resolution) is purely optical.

(v) Conclusion

We have studied the z-motion artifact that may appear in the images obtained in the constant-distance mode. We have illustrated the complexity of the problem, and shown that no general rule can be found in order to predict the presence or not of an artifact. A given set-up may

produce images with artifacts for a given sample, and almost purely optical images for another sample. We believe that the only safe way to produce images with purely optical contrast and resolution is to perform constant-height (or constant-intensity) scans [155, 88].

10. Concluding Remarks

Near-field instruments have proved to be unique tools to study a number of phenomena, such as surface waves and optical nanostructures. A large number of interesting properties has been demonstrated, such as magneto-optics effects, fluorescence detection, local spectroscopy, etc. So far, the major progresses have been the consequence of novel experimental techniques. Numerous techniques have been developed, and are available. Despite these impressive successes, the topic is not mature. There is a need for the development of instruments providing routinely images with resolution better than 50 nm. There is also a need for a comparative evaluation of the different techniques, which must be achieved for near-field optics to become a widely spread technique.

Comparatively, the understanding of the images and the understanding of the contrast mechanisms have been relatively slow. Therefore, there has been a lack of conceptual tools to analyse and compare different instruments. We hope that this review might be useful in summarizing the progress in the understanding of the images in the last five years. We also hope that it will serve to establish a common core of conceptual tools to describe and evaluate near-field optical instruments.

Acknowledgements

It is a pleasure to thank the following people for many helpful and stimulating discussions : P.M. Adam, R. Bachelot, D. Barchiesi, J.L. Bijon, A.C. Boccara, J.C. Bolomey, S.I. Bozhevolnyi, Y. Chen, H. Cory, U.Ch. Fischer, F. de Fornel, N. García, C. Girard, O. Keller, O. Marti, A. Madrazo, E.R. Méndez, M. Nieto-Vesperinas, L. Novotny, D.W. Pohl, P. Royer, J.J. Saenz, A. Sentenac, D. van Labeke, M. Vaez-Iravani and P.J. Valle.

Appendix A : Angular spectrum

We establish in this appendix the angular spectrum representation of a field propagating in a vacuum. Let $\mathbf{E}(\mathbf{r})$ be a monochromatic field of frequency ω propagating into the half-space $z > 0$, and created by sources that are assumed to be localized into the half-space $z < 0$. For $z > 0$, $\mathbf{E}$ satisfies the propagation equation (a temporal dependence $\exp(-i\omega t)$ is assumed and omitted)

$$\nabla^2 \mathbf{E}(\mathbf{r}) + k_0^2 \, \mathbf{E}(\mathbf{r}) = 0 \tag{A1}$$

where $k_0 = \omega/c$.

The field must have a finite energy for any $z \geq 0$:

$$\int |\mathbf{E}(\mathbf{r}_{||}, z)|^2 \, d\mathbf{r}_{||} < \infty \ , \tag{A2}$$

where $\mathbf{r}_{||} = (x, y)$. Moreover, $\mathbf{E}$ must satisfy the vector form of Sommerfeld's radiation condition in the half-space $z > 0$ [178] :

$$\lim_{r \to \infty} r(\nabla \times \mathbf{E} - ik_0 \frac{\mathbf{r}}{|\mathbf{r}|} \times \mathbf{E}) = 0 \ . \tag{A3}$$

With these two conditions, we assume that the field $\mathbf{E}(\mathbf{r})$ has a two-dimensional Fourier transform in any plane $z = constant$:

$$\mathbf{E}(\mathbf{r}_{||}, z) = \int \tilde{\mathbf{E}}(\mathbf{k}_{||}, z) \exp(i\mathbf{k}_{||} \cdot \mathbf{r}_{||}) \, d\mathbf{k}_{||} \ , \tag{A4}$$

where $\mathbf{k}_{||} = (k_x, k_y)$. For a more precise discussion of the validity of (A4), see, e.g., [179].

Introducing (A4) into (A1) leads to :

$$\frac{\partial^2}{\partial z^2} \tilde{\mathbf{E}}(\mathbf{k}_{||}, z) + \gamma^2 \, \tilde{\mathbf{E}}(\mathbf{k}_{||}, z) = 0 \ , \tag{A5}$$

with $\gamma^2 = k_0^2 - \mathbf{k}_{||}^2$.

The solution to (A5), which is compatible with the radiation condition (A3), is [179]:

$$\tilde{\mathbf{E}}(\mathbf{k}_{||}, z) = \mathbf{e}(\mathbf{k}_{||}) \exp(i\gamma z) \ , \tag{A6}$$

with $\gamma = (k_0^2 - \mathbf{k}_{||}^2)^{1/2}$, and the determination $\mathrm{Re}(\gamma) > 0$ and $\mathrm{Im}(\gamma) > 0$.

Introducing (A6) into (A4) leads to the expression of the angular spectrum (or plane-wave expansion) of the field, valid for $z \geq 0$:

$$\mathbf{E}(\mathbf{r}) = \int \mathbf{e}(\mathbf{k}_{||}) \exp(i\mathbf{k}_{||} \cdot \mathbf{r}_{||} + i\gamma z) \, d\mathbf{k}_{||} \ . \tag{A7}$$

Appendix B : Green's dyadic

When describing the radiation of an electric dipole $\mathbf{p}$ in §4, we introduced a *partial* Green's dyadic $\overleftrightarrow{\mathbf{G}}_0(\mathbf{r})$, in the sense that it was not defined at the origin $\mathbf{r} = 0$. This dyadic describes the radiation of a dipole source at exterior points [165], namely, at points outside the dipole itself.

We now give the expressions of the *full* Green dyadic, also defined at interior points [165], for free space, and for two homogeneous half-spaces separated by a flat interface.

Free space Green's dyadic

The full Green dyadic $\overleftrightarrow{\mathbf{G}}(\mathbf{r}, \mathbf{r}')$ in free space is the sum of two contributions [165, 180] :

$$\overleftrightarrow{\mathbf{G}}(\mathbf{r}, \mathbf{r}') = PV_\delta \; \overleftrightarrow{\mathbf{G}}_0(\mathbf{r}, \mathbf{r}') - \frac{1}{k_0^2}\delta(\mathbf{r} - \mathbf{r}') \; \overleftrightarrow{\mathbf{L}}_{V_\delta} \tag{B1}$$

where PV means that any integration of $\overleftrightarrow{\mathbf{G}}_0$ should be defined in principal value, V_δ being the exclusion volume. The expression of $\overleftrightarrow{\mathbf{G}}_0$ is :

$$\overleftrightarrow{\mathbf{G}}_0(\mathbf{r}, \mathbf{r}') = (\overleftrightarrow{\mathbf{I}} + \frac{1}{k_0^2}\nabla\nabla)\frac{\exp(ik_0|\mathbf{r} - \mathbf{r}'|)}{4\pi|\mathbf{r} - \mathbf{r}'|} \;, \quad \mathbf{r} \neq \mathbf{r}' \tag{B2}$$

where $\overleftrightarrow{\mathbf{I}}$ is the unit dyadic. It is also given by (4.2), when $\mathbf{r}'$ coincides with the origin.

The full Green dyadic $\overleftrightarrow{\mathbf{G}}$ satisfies the equation [165] :

$$\nabla \times \nabla \times \overleftrightarrow{\mathbf{G}}(\mathbf{r}, \mathbf{r}') - k_0^2 \; \overleftrightarrow{\mathbf{G}}(\mathbf{r}, \mathbf{r}') = \delta(\mathbf{r} - \mathbf{r}') \; \overleftrightarrow{\mathbf{I}} \tag{B3}$$

together with the vector form of Sommerfeld's radiation condition :

$$\lim_{|\mathbf{r}-\mathbf{r}'|\to\infty} |\mathbf{r} - \mathbf{r}'| \left(\nabla \times \overleftrightarrow{\mathbf{G}} - ik_0 \frac{\mathbf{r} - \mathbf{r}'}{|\mathbf{r} - \mathbf{r}'|} \times \overleftrightarrow{\mathbf{G}} \right) = 0 \; . \tag{B4}$$

The dyadic $\overleftrightarrow{\mathbf{L}}_{V_\delta}$ in (B1) describes the singularity of the full Green dyadic at $\mathbf{r} = \mathbf{r}'$. Its expression depends on the shape of the exclusion volume V_δ. Some examples are given in ref. [165, 180].

From a physical point of view, $\overleftrightarrow{\mathbf{L}}_{V_\delta}$ describes the radiation of the exclusion volume V_δ on itself. as the linear dimension δ of the volume tends to zero (self field). For a spherical exclusion volume, $\overleftrightarrow{\mathbf{L}}_{V_\delta} = \overleftrightarrow{\mathbf{I}}/3$. Thus, the singularity becomes $-1/3k_0^2 \, \delta(\mathbf{r} - \mathbf{r}') \; \overleftrightarrow{\mathbf{I}}$. This expression can be

understood as a macroscopic self-field correction to the radiation of an electric point dipole [181]. In a microscopic theory, where the longitudinal and transverse internal electrodynamics of the dipole are treated separately, the self-field correction is different. It contains both the transverse and longitudinal delta functions, with different coefficients [181, 182]. Nevertheless, many "microscopic" descriptions of near-field problems using the coupled-dipoles method *without microscopic self-field corrections* have been presented [47, 183, 184, 138, 185, 186]. These approaches are strictly equivalent to a macroscopic theory, when the singularity of the full Green dyadic described above is taken into account. A nice discussion can be found in ref.[187].

Green's dyadic of two half-spaces with a flat interface

The upper medium ($z > 0$) is homogeneous and isotropic, with a dielectric constant ϵ_1. The lower medium ($z < 0$) is also homogeneous and isotropic, with a dielectric constant ϵ_3. The notations are consistent with those used in §5.

The Green dyadic does not have any analytical expression in direct space in this case. Let us introduce its Fourier transform in the xy plane :

$$\overleftrightarrow{\mathbf{G}}\,(\mathbf{r}_{\|}-\mathbf{r}'_{\|},z,z') = \frac{1}{4\pi^2}\int \overleftrightarrow{\mathbf{S}}^{-1}\,(\mathbf{k}_{\|})\cdot\,\overleftrightarrow{\mathbf{g}}\,(\mathbf{k}_{\|},z,z')\cdot\,\overleftrightarrow{\mathbf{S}}\,(\mathbf{k}_{\|})\,\exp[i\mathbf{k}_{\|}\cdot(\mathbf{r}_{\|}-\mathbf{r}'_{\|})]\,d\mathbf{k}_{\|}\;. \tag{B5}$$

The matrix $\overleftrightarrow{\mathbf{S}}$ has the following expression [121, 186] :

$$\overleftrightarrow{\mathbf{S}}\,(\mathbf{k}_{\|}) = \frac{1}{k_{\|}}\begin{bmatrix} k_x & k_y & 0 \\ -k_y & k_x & 0 \\ 0 & 0 & k_{\|} \end{bmatrix}\;, \tag{B6}$$

where $k_{\|} = |\mathbf{k}_{\|}|$. $\overleftrightarrow{\mathbf{S}}$ is the rotation matrix from the basis $(k_x, k_y, 0)$ to the basis $(k_{\|}, 0, 0)$. In this basis, the expressions of the non-vanishing components of $\overleftrightarrow{\mathbf{g}}$ for $z > 0$ and $z' > 0$ are the following :

$$\begin{aligned}
g_{yy} &= \frac{-i}{2\gamma_1}\left[\frac{\gamma_1-\gamma_3}{\gamma_1+\gamma_3}\exp[i\gamma_1(z+z')] + \exp(i\gamma_1|z-z'|)\right]\;,\\
g_{xx} &= \frac{-i\gamma_1}{2\epsilon_1 k_0^2}\left[\frac{\epsilon_1\gamma_3-\epsilon_3\gamma_1}{\epsilon_1\gamma_3+\epsilon_3\gamma_1}\exp[i\gamma_1(z+z')] + \exp(i\gamma_1|z-z'|)\right]\;,\\
g_{zz} &= \frac{ik_{\|}^2}{2\epsilon_1\gamma_1 k_0^2}\left[\frac{\epsilon_1\gamma_3-\epsilon_3\gamma_1}{\epsilon_1\gamma_3+\epsilon_3\gamma_1}\exp[i\gamma_1(z+z')] - \exp(i\gamma_1|z-z'|)\right] + \frac{1}{\epsilon_1 k_0^2}\,\delta(z-z')\;,
\end{aligned}$$

$$g_{xz} = \frac{-ik_{||}}{2\epsilon_1 k_0^2}\left[\frac{\epsilon_1\gamma_3 - \epsilon_3\gamma_1}{\epsilon_1\gamma_3 + \epsilon_3\gamma_1}\exp[i\gamma_1(z+z')] - \exp(i\gamma_1|z-z'|)\mathrm{sign}(z-z')\right] ,$$
$$g_{zx} = \frac{ik_{||}}{2\epsilon_1 k_0^2}\left[\frac{\epsilon_1\gamma_3 - \epsilon_3\gamma_1}{\epsilon_1\gamma_3 + \epsilon_3\gamma_1}\exp[i\gamma_1(z+z')] + \exp(i\gamma_1|z-z'|)\mathrm{sign}(z-z')\right] , \tag{B7}$$

where $k_0 = \omega/c$. Equations (B5) and (B7) represent the full Green dyadic for the system with a flat interface, which has a singularity at $\mathbf{r} = \mathbf{r}'$. In Fourier space, this singularity is visible in the slow decay of the components of $\overleftrightarrow{\mathbf{g}}$ as $k_{||} \to \infty$ [79]. In numerical calculations, the singularity has to be handled with care [79, 188].

The Green dyadic $\overleftrightarrow{\mathbf{G}}$ is a solution of the equation

$$\nabla \times \nabla \times \overleftrightarrow{\mathbf{G}}(\mathbf{r}_{||} - \mathbf{r}'_{||}, z, z') - \epsilon_f(z)\, k_0^2\ \overleftrightarrow{\mathbf{G}}(\mathbf{r}_{||} - \mathbf{r}'_{||}, z, z') = \overleftrightarrow{\mathbf{I}}\ \delta(\mathbf{r}_{||} - \mathbf{r}'_{||})\,\delta(z - z') , \tag{B8}$$

satisfying the boundary conditions at the interface $z = 0$ (i.e., the conditions satisfied by the electric field $\mathbf{E}$) and Sommerfelds's radiation condition at infinity (B4). In (B8), $\overleftrightarrow{\mathbf{I}}$ is the unit dyadic, and $\epsilon_f(z)$ is ϵ_1 for $z > 0$ and ϵ_3 for $z < 0$.

Appendix C : Lippmann-Schwinger equation

We establish here the integral equation (5.1) for a monochromatic field of frequency ω. A temporal dependence $\exp(-i\omega t)$ is assumed and omitted everywhere. The geometry and the notations are defined in §5.

In an inhomogeneous, linear, isotropic and local medium, the total electric field satisfies the wave equation

$$\nabla \times \nabla \times \mathbf{E}(\mathbf{r}) - \epsilon(\mathbf{r})\, k_0^2\, \mathbf{E}(\mathbf{r}) = 0 , \tag{C1}$$

where $\epsilon(\mathbf{r})$ is the position-dependent dielectric constant of the medium and $k_0 = \omega/c$.

The field $\mathbf{E}^{(0)}$ that would exist for the flat interface at $z = 0$ obeys a similar equation :

$$\nabla \times \nabla \times \mathbf{E}^{(0)}(\mathbf{r}) - \epsilon_f(z)\, k_0^2\, \mathbf{E}^{(0)}(\mathbf{r}) = 0 , \tag{C2}$$

where $\epsilon_f(z)$ is ϵ_1 for $z > 0$ and ϵ_3 for $z < 0$. If we define the scattered field as $\mathbf{E}_s = \mathbf{E} - \mathbf{E}^{(0)}$, substracting (C2) from (C1) leads to the equation satisfied by $\mathbf{E}_s$:

$$\nabla \times \nabla \times \mathbf{E}_s(\mathbf{r}) - \epsilon_f(z)\, k_0^2\, \mathbf{E}_s(\mathbf{r}) = k_0^2\,[\epsilon(\mathbf{r}) - \epsilon_f(z)]\,\mathbf{E}(\mathbf{r}) . \tag{C3}$$

The right-hand side in (C3) is the source term of the scattered field. It vanishes everywhere, except when $\epsilon \neq \epsilon_f$, i.e., inside the inhomogeneous layer deposited on the interface $z = 0$. Inside this layer, one has $\epsilon(\mathbf{r}) - \epsilon_f(z) = \epsilon_2(\mathbf{r}) - \epsilon_1$.

We now make use of the Green dyadic for the system of two half-spaces separated by a flat interface at $z = 0$ (see Appendix B). Equation (B8) provides an integral solution to (C3) [122] :

$$\mathbf{E}_s(\mathbf{r}) = k_0^2 \int [\epsilon_2(\mathbf{r}') - \epsilon_1] \overleftrightarrow{\mathbf{G}} (\mathbf{r}_{||} - \mathbf{r}'_{||}, z, z') \, \mathbf{E}(\mathbf{r}') \, d\mathbf{r}' \ . \tag{C4}$$

Adding the illuminating field to (C4) leads to the Lippmann-Schwinger equation for the total electric field :

$$\mathbf{E}(\mathbf{r}) = \mathbf{E}^{(0)}(\mathbf{r}) + k_0^2 \int [\epsilon_2(\mathbf{r}') - \epsilon_1] \overleftrightarrow{\mathbf{G}} (\mathbf{r}_{||} - \mathbf{r}'_{||}, z, z') \, \mathbf{E}(\mathbf{r}') \, d\mathbf{r}' \ . \tag{C5}$$

Appendix D : Derivation of the reciprocity theorem

We recall here the derivation of the reciprocity theorem [141], which is the basis of many developments presented in this review.

In a situation 1, let V_1 be a source volume with a current density $\mathbf{J}_1(\mathbf{r})$ radiating at a frequency ω. Let us call $\mathbf{E}_1(\mathbf{r})$ and $\mathbf{H}_1(\mathbf{r})$ the fields created by this source in the presence of a scatterer described by its constitutive tensors $\overleftrightarrow{\epsilon}(\mathbf{r}, \omega)$ and $\overleftrightarrow{\mu}(\mathbf{r}, \omega)$. In a situation 2, let V_2 be a source volume with a current density $\mathbf{J}_2(\mathbf{r})$ radiating at the same frequency ω. Let us call $\mathbf{E}_2(\mathbf{r})$ and $\mathbf{H}_2(\mathbf{r})$ the fields created by this source in the presence of the same scatterer.

The fields in each situation satisfy Maxwell's equations, that is, with $k = 1, 2$:

$$\begin{aligned} \nabla \times \mathbf{E}_k &= i\omega \mathbf{B}_k \\ \nabla \times \mathbf{H}_k &= \mathbf{J}_k - i\omega \mathbf{D}_k \ , \end{aligned} \tag{D1}$$

together with the constitutive relations

$$\begin{aligned} \mathbf{D}_k(\mathbf{r}) &= \epsilon_0 \overleftrightarrow{\epsilon}(\mathbf{r}, \omega) \, \mathbf{E}_k(\mathbf{r}) \\ \mathbf{B}_k(\mathbf{r}) &= \mu_0 \overleftrightarrow{\mu}(\mathbf{r}, \omega) \, \mathbf{H}_k(\mathbf{r}) \ . \end{aligned} \tag{D2}$$

From (D1), one easily obtains for each point $\mathbf{r}$ the following equality :

$$\begin{aligned}(\mathbf{H}_2 \cdot \nabla \times \mathbf{E}_1 - \mathbf{E}_1 \cdot \nabla \times \mathbf{H}_2) + (\mathbf{E}_2 \cdot \nabla \times \mathbf{H}_1 - \mathbf{H}_1 \cdot \nabla \times \mathbf{E}_2) &= i\omega(\mathbf{B}_1 \cdot \mathbf{H}_2 - \mathbf{H}_1 \cdot \mathbf{B}_2) \\ &\quad -i\omega(\mathbf{D}_1 \cdot \mathbf{E}_2 - \mathbf{E}_1 \cdot \mathbf{D}_2) \\ &\quad +\mathbf{J}_1 \cdot \mathbf{E}_2 - \mathbf{J}_2 \cdot \mathbf{E}_1 \ . \quad \text{(D3)}\end{aligned}$$

The lef-hand side in (D3) can be rewritten in the form $\nabla \cdot (\mathbf{E}_1 \times \mathbf{H}_2 - \mathbf{E}_2 \times \mathbf{H}_1)$. With the use of (D2), one shows that the two first terms in the right-hand side vanish, provided that $\overleftrightarrow{\epsilon}$ and $\overleftrightarrow{\mu}$ are *symmetric* tensors.

Finally, one obtains for each point $\mathbf{r}$:

$$\nabla \cdot (\mathbf{E}_1 \times \mathbf{H}_2 - \mathbf{E}_2 \times \mathbf{H}_1) = \mathbf{J}_1 \cdot \mathbf{E}_2 - \mathbf{J}_2 \cdot \mathbf{E}_1 \ . \quad \text{(D4)}$$

By integrating (D4) over all space, the left-hand side is transformed into a surface integral over a sphere whose radius tends to infinity. The asymptotic expressions of the fiels for $kr \to \infty$ in a direction defined by the vector $\mathbf{k} = (\mathbf{k}_{||}, \gamma)$ are [179] .

$$\mathbf{E}_k(kr \to \infty) = 2i\pi\,\gamma(\mathbf{k}_{||})\,\mathbf{e}_k(\mathbf{k}_{||})\,\frac{\exp(ikr)}{r} \ , \quad \text{(D5)}$$

$$\mathbf{H}_k(kr \to \infty) = 2i\pi\,\gamma(\mathbf{k}_{||})\,\mathbf{h}_k(\mathbf{k}_{||})\,\frac{\exp(ikr)}{r} \ , \quad \text{(D6)}$$

where $\mathbf{e}_k(\mathbf{k}_{||})$ and $\mathbf{h}_k(\mathbf{k}_{||})$ are the angular spectra of the fields, as defined in Appendix A. As a consequence of Maxwell's equations,

$$\omega\mu_0\,\mathbf{h}_k(\mathbf{k}_{||}) = \mathbf{k} \times \mathbf{e}_k(\mathbf{k}_{||}) \ . \quad \text{(D7)}$$

By making use of (D5), (D6) and (D7), one shows that $(\mathbf{E}_1 \times \mathbf{H}_2 - \mathbf{E}_2 \times \mathbf{H}_1)$ vanishes identically in the far field, so that its integral over a sphere of infinite radius disappears. Finally, one is left with

$$\int_{V_1} \mathbf{J}_1(\mathbf{r}) \cdot \mathbf{E}_2(\mathbf{r})\,d\mathbf{r} = \int_{V_2} \mathbf{J}_2(\mathbf{r}) \cdot \mathbf{E}_1(\mathbf{r})\,d\mathbf{r} \ . \quad \text{(D8)}$$

In the case of dipole sources, the current densities are given by

$$\mathbf{J}_k(\mathbf{r}) = -i\omega\mathbf{p}_k\,\delta(\mathbf{r} - \mathbf{r}_k) \ , \quad \text{(D9)}$$

which gives, after introduction into (D8),

$$\mathbf{p}_1 \cdot \mathbf{E}_2(\mathbf{r}_1) = \mathbf{p}_2 \cdot \mathbf{E}_1(\mathbf{r}_2) \ . \quad \text{(D10)}$$

Appendix E : Expression of the impulse response H_c

Expression of the dyadic $\overleftrightarrow{\mathbf{L}}$

The first-order expression of the total field is [see (5.7)] :

$$\mathbf{E}(\mathbf{r}) = \mathbf{E}^{(0)}(\mathbf{r}) + \int \mathbf{e}^{(1)}(\mathbf{k}_{||}) \exp(i\mathbf{k}_{||} \cdot \mathbf{r}_{||} + i\gamma z)\, d\mathbf{k}_{||} \tag{E1}$$

with

$$\mathbf{e}^{(1)}(\mathbf{k}_{||}) = \hat{S}_{eq}(\mathbf{k}_{||}^{inc} - \mathbf{k}_{||})\ \overleftrightarrow{\mathbf{L}}\,(\mathbf{k}_{||}, \mathbf{k}_{||}^{inc})\, \mathbf{e}_{inc}\ . \tag{E2}$$

The choice of an adequate basis allows us to write the dyadic $\overleftrightarrow{\mathbf{L}}$ in a simple manner. The electric field, evaluated in a homogeneous medium, verifies $\nabla \cdot \mathbf{E} = 0$. Thus, its angular spectrum $\mathbf{e}^{(1)}(\mathbf{k}_{||})$ has only two independant components. Let us write :

$$\mathbf{e}^{(1)}(\mathbf{k}_{||}) = e_s^{(1)}(\mathbf{k}_{||})\, \mathbf{a}_s(\mathbf{k}_{||}) + e_p^{(1)}(\mathbf{k}_{||})\, \mathbf{a}_p(\mathbf{k}_{||}) \tag{E3}$$

where the basis vectors are defined by

$$\begin{aligned} \mathbf{a}_s(\mathbf{k}_{||}) &= \mathbf{a}_z \times \hat{\mathbf{k}}_{||} \\ \mathbf{a}_p(\mathbf{k}_{||}) &= \mathbf{a}_s(\mathbf{k}_{||}) \times \frac{\mathbf{k}_{||} + \gamma \mathbf{a}_z}{\sqrt{\epsilon_1} k_0}\ , \end{aligned} \tag{E4}$$

$\mathbf{a}_z$ being the unit vector along the z-axis, $\hat{\mathbf{k}}_{||}$ the unit vector $\mathbf{k}_{||}/|\mathbf{k}_{||}|$ and $k_0 = \omega/c$.

In the basis $(\mathbf{a}_s, \mathbf{a}_p)$, the expression for $\overleftrightarrow{\mathbf{L}}$ is :

$$\overleftrightarrow{\mathbf{L}}\,(\mathbf{k}_{||}, \mathbf{k}_{||}^{inc}) = \frac{i(\gamma_3 - \gamma_1)}{4\pi^2} \begin{bmatrix} t_s^{inc}\, \hat{\mathbf{k}}_{||}.\hat{\mathbf{k}}_{||}^{inc} & t_p^{inc}\, \dfrac{\gamma_1^{inc}}{n_1 k_0}\, \hat{\mathbf{k}}_{||}^{inc}.(\mathbf{a}_z \times \hat{\mathbf{k}}_{||}) \\ \dfrac{t_s^{inc}\, n_1 k_0\, \gamma_3\, \hat{\mathbf{k}}_{||}.(\mathbf{a}_z \times \hat{\mathbf{k}}_{||}^{inc})}{\gamma_1\gamma_3 + |\mathbf{k}_{||}|^2} & \dfrac{t_p^{inc}\, \{\gamma_1^{inc}\, \gamma_3\, \hat{\mathbf{k}}_{||}^{inc}.\hat{\mathbf{k}}_{||} + |\mathbf{k}_{||}|\, |\mathbf{k}_{||}^{inc}|\}}{\gamma_1\gamma_3 + |\mathbf{k}_{||}|^2} \end{bmatrix} \tag{E5}$$

We have introduced the notations $n_j = \sqrt{\epsilon_j}$, $\gamma_j = \gamma_j(\mathbf{k}_{||}) = (\epsilon_j k_0^2 - \mathbf{k}_{||}^2)^{1/2}$, $\gamma_j^{inc} = \gamma_j(\mathbf{k}_{||}^{inc})$. t_s^{inc} and t_p^{inc} are the Fresnel transmission factors accross the interface $z = 0$, for an s- and a p-polarized incident plane wave, respectively. In the basis $(\mathbf{a}_s, \mathbf{a}_p)$, the dyadic $\overleftrightarrow{\mathbf{T}}\,(\mathbf{k}_{||}^{inc})$ introduced in (5.6) is diagonal, its two diagonal elements being t_s^{inc} and t_p^{inc}.

Impulse response H_c

The impulse response, defined in (5.10), has an analytical expression in Fourier space. Its Fourier transform, defined by (5.11), is given by

$$\tilde{H}_c(\mathbf{k}_{||}, \mathbf{k}_{||}^{inc}, \omega, z_0) = f(\mathbf{k}_{||}^{inc} - \mathbf{k}_{||}, \mathbf{k}_{||}^{inc}, \omega, z_0) + f^*(\mathbf{k}_{||}^{inc} + \mathbf{k}_{||}, \mathbf{k}_{||}^{inc}, \omega, z_0) \ , \tag{E6}$$

with

$$f(\mathbf{k}_{||}, \mathbf{k}_{||}^{inc}, \omega, z_0) = 4\pi^2 \overleftrightarrow{\mathbf{T}}(\mathbf{k}_{||}^{inc})\, \mathbf{e}_{inc}^* \cdot \overleftrightarrow{\mathbf{L}}(\mathbf{k}_{||}, \mathbf{k}_{||}^{inc})\, \mathbf{e}_{inc} \exp[i\gamma_1(\mathbf{k}_{||})z_0 - i\gamma_1^*(\mathbf{k}_{||}^{inc})z_0] \ . \tag{E7}$$

The superscript $*$ denotes the complex conjugate.

Appendix F : Transfer function in collection-mode SNOM

We examine here the existence of a transfer function in collection-mode SNOM, relating the *signal* (i.e., the flux of the Poynting vector in the guiding part of the probe) to the near field illuminating the tip. The first step is to derive the form of the coupling coefficient between an incident plane wave and the mode amplitude in the guide.

Coupling between a plane wave and a mode in the probe

We first study the coupling between an illuminating plane wave and the mode in the SNOM probe, working in collection mode. The plane wave can be either homogeneous (propagating) or inhomogeneous (evanescent). We start with (7.6), which gives an exact expression of the amplitude $A_{\lambda_0}^{(-)}$ of the mode λ_0, when the probe is in near-field coupling with a point dipole $\mathbf{p}$ placed at a position $\mathbf{r}_d$ in vacuum (without any loss of generality, we can assume here that no substrate is present) :

$$A_{\lambda_0}^{(-)} = 2\pi Z_{\lambda_0} \frac{i\omega}{c} \mathbf{p} \cdot \mathbf{E}_{tip}^{\lambda_0}(\mathbf{r}_d - \mathbf{r}_{tip}) \ , \tag{F1}$$

where $\mathbf{r}_{tip}$ is the tip position. Let us recall that $\mathbf{E}_{tip}^{\lambda_0}$ is the field emitted by the tip in vacuum (no substrate here), when the mode λ_0 excites the tip in illumination mode.

We now assume that the tip does not react on the dipole, so that $\mathbf{p}$ is *independent on the tip position.* Since we look for the coupling coefficient (or transfer function) of a particular

wave vector, let us introduce the angular spectrum of the field $\mathbf{E}_{tip}^{\lambda_0}$ emitted by the tip :

$$\mathbf{E}_{tip}^{\lambda_0}(\mathbf{r}-\mathbf{r}_{tip}) = \int \mathbf{t}_{\lambda_0}(\mathbf{k}_{\|}) \exp\left[i\mathbf{k}_{\|}\cdot(\mathbf{r}_{\|}-\mathbf{r}_{tip\|}) + i\gamma(z-z_{tip})\right] d\mathbf{k}_{\|} \ , \ \text{for } z > z_{tip} \ , \qquad \text{(F2)}$$

where we have used the notation $\mathbf{r} = (\mathbf{r}_{\|}, z)$, $\gamma = (k_0^2 - \mathbf{k}_{\|}^2)^{1/2}$, with $\mathrm{Re}(\gamma) > 0$ and $\mathrm{Im}(\gamma) > 0$. The orientation of the z-axis is indicated in Fig. 13(a). Inserting (F2) into (F1), one obtains :

$$A_{\lambda_0}^{(-)} = 2\pi Z_{\lambda_0} \frac{i\omega}{c} \int \mathbf{t}_{\lambda_0}(\mathbf{k}_{\|})\cdot\mathbf{p} \exp\left[i\mathbf{k}_{\|}\cdot(\mathbf{r}_{d\|}-\mathbf{r}_{tip\|}) + i\gamma(z_d - z_{tip})\right] d\mathbf{k}_{\|} \ . \qquad \text{(F3)}$$

In order to obtain the amplitude of the mode due to a single illuminating plane wave, we introduce the angular representation of the field radiated by the dipole $\mathbf{p}$:

$$\mathbf{E}_{ill}(\mathbf{r}) = \int \frac{|\mathbf{k}_{\|}|}{\gamma} \overleftrightarrow{h}(\mathbf{k}_{\|})\cdot\mathbf{p} \exp\left[i\mathbf{k}_{\|}\cdot(\mathbf{r}_{\|}-\mathbf{r}_{d\|}) - i\gamma(z-z_d)\right] d\mathbf{k}_{\|} \ , \ \text{for } z < z_d \ , \qquad \text{(F4)}$$

where $\overleftrightarrow{h}(\mathbf{k}_{\|})$ is an operator that extracts the transverse component of $\mathbf{p}$. Let us define the spectral component $\mathbf{e}_{ill}$ of the illuminating field $\mathbf{E}_{ill}$ as follows :

$$\mathbf{E}_{ill}(\mathbf{r}) = \int \mathbf{e}_{ill}(\mathbf{k}_{\|}) \exp(i\mathbf{k}_{\|}\cdot\mathbf{r}_{\|} - i\gamma z)\, d\mathbf{k}_{\|} \ . \qquad \text{(F5)}$$

An identification with (F4) shows that ;

$$\mathbf{e}_{ill}(\mathbf{k}_{\|}) = \frac{|\mathbf{k}_{\|}|}{\gamma} \overleftrightarrow{h}(\mathbf{k}_{\|})\cdot\mathbf{p} \exp(-i\mathbf{k}_{\|}\cdot\mathbf{r}_{d\|} + i\gamma z_d) \ . \qquad \text{(F6)}$$

From (F6), one immediately obtains :

$$\mathbf{t}_{\lambda_0}(\mathbf{k}_{\|})\cdot\mathbf{e}_{ill}(-\mathbf{k}_{\|}) = \frac{|\mathbf{k}_{\|}|}{\gamma} \mathbf{t}_{\lambda_0}(\mathbf{k}_{\|})\cdot\overleftrightarrow{h}(-\mathbf{k}_{\|})\cdot\mathbf{p} \exp(i\mathbf{k}_{\|}\cdot\mathbf{r}_{d\|} + i\gamma z_d) \ . \qquad \text{(F7)}$$

Noting that $\mathbf{t}_{\lambda_0}(\mathbf{k}_{\|})$ is transverse, we have the identity $\mathbf{t}_{\lambda_0}(\mathbf{k}_{\|})\cdot\overleftrightarrow{h}(-\mathbf{k}_{\|})\cdot\mathbf{p} = \mathbf{t}_{\lambda_0}(\mathbf{k}_{\|})\cdot\mathbf{p}$, so that (F7) gives :

$$\mathbf{t}_{\lambda_0}(\mathbf{k}_{\|})\cdot\mathbf{p} \exp(i\mathbf{k}_{\|}\cdot\mathbf{r}_{d\|} + i\gamma z_d) = \frac{\gamma}{|\mathbf{k}_{\|}|} \mathbf{t}_{\lambda_0}(\mathbf{k}_{\|})\cdot\mathbf{e}_{ill}(-\mathbf{k}_{\|}) \ . \qquad \text{(F8)}$$

Using (F8), (F3) can be rewritten in the form :

$$A_{\lambda_0}^{(-)}(\mathbf{r}_{tip}) = 2\pi Z_{\lambda_0} \frac{i\omega}{c} \int \frac{\gamma}{|\mathbf{k}_{\|}|} \mathbf{t}_{\lambda_0}(\mathbf{k}_{\|})\cdot\mathbf{e}_{ill}(-\mathbf{k}_{\|}) \exp(-i\mathbf{k}_{\|}\cdot\mathbf{r}_{tip\|} - i\gamma z_{tip})\, d\mathbf{k}_{\|} \ . \qquad \text{(F9)}$$

We see that the coupling between a plane wave with wave vector $\mathbf{k}_{\|}$, either homogeneous ($|\mathbf{k}_{\|}| \leq \omega/c$) or inhomogeneous ($|\mathbf{k}_{\|}| > \omega/c$), is maximum if its electric field is parallel to $\mathbf{t}_{\lambda_0}(\mathbf{k}_{\|})$.

Transfer function

It is interesting to discuss the existence of a transfer function between the optical signal $S = |A_{\lambda_0}^{(-)}|^2$ and the near field intensity, defined as the square modulus of the illuminating electric field. Such a transfer function has been found numerically for s-polarization for two-dimensional systems [136]. Nevertheless, its existence was questionned recently [95].

Let us assume that the illuminating field $\mathbf{e}_{ill}$ is the sum of a background contribution $\mathbf{e}^{(0)}$ and a dynamic part $\mathbf{e}^{(1)}$, with the condition $|\mathbf{e}^{(1)}| \ll |\mathbf{e}^{(0)}|$. For example, $\mathbf{e}^{(0)}$ is an incident field and $\mathbf{e}^{(1)}$ is the field scattered by a sample with a topography much smaller than the wavelength and a low dielectric contrast (see §5). For simplicity, we assume that the incident field is a plane wave with wavevector $\mathbf{k}_{||}^{inc}$, so that $\mathbf{e}^{(0)}(\mathbf{k}_{||}) = \mathbf{e}_0\,\delta(\mathbf{k}_{||} - \mathbf{k}_{||}^{inc})$. In this case, dropping the superscript $(-)$ for brevity, the signal can be written to first order :

$$S(\mathbf{r}_{tip}) = |A_{\lambda_0}^{bg}(\mathbf{r}_{tip})|^2 + 2\mathrm{Re}\left\{A_{\lambda_0}^{bg*}(\mathbf{r}_{tip})\,A_{\lambda_0}^{sa}(\mathbf{r}_{tip})\right\}\ , \tag{F10}$$

where the background amplitude, whici does not contain any information on the sample, is :

$$A_{\lambda_0}^{bg}(\mathbf{r}_{tip}) = 2\pi Z_{\lambda_0}\frac{i\omega}{c}\frac{\gamma_{inc}}{|\mathbf{k}_{||}^{inc}|}\,\mathbf{t}_{\lambda_0}(\mathbf{k}_{||}^{inc})\cdot\mathbf{e}_0\,\exp(i\mathbf{k}_{||}^{inc}\cdot\mathbf{r}_{tip||} - i\gamma_{inc}z_{tip})\,d\mathbf{k}_{||}\ , \tag{F11}$$

and the sample contribution is :

$$A_{\lambda_0}^{sa}(\mathbf{r}_{tip}) = 2\pi Z_{\lambda_0}\frac{i\omega}{c}\int\frac{\gamma}{|\mathbf{k}_{||}|}\,\mathbf{t}_{\lambda_0}(\mathbf{k}_{||})\cdot\mathbf{e}^{(1)}(-\mathbf{k}_{||})\,\exp(-i\mathbf{k}_{||}\cdot\mathbf{r}_{tip||} - i\gamma z_{tip})\,d\mathbf{k}_{||}\ . \tag{F12}$$

We have used the notation $\gamma_{inc} = \gamma(\mathbf{k}_{||}^{inc})$,

The first term in (F10) is independent on $\mathbf{r}_{tip}$ and contributes to the background of the image only. The dynamic part of the signal, containing the information on the sample, is :

$$\begin{aligned} S^{sa}(\mathbf{r}_{tip}) &= 8\pi^2|Z_{\lambda_0}|^2k_0^2\,\mathrm{Re}\int\frac{\gamma_{inc}^*\gamma}{|\mathbf{k}_{||}^{inc}||\mathbf{k}_{||}|}\exp[-i\mathbf{k}_{||}^{inc}\cdot\mathbf{r}_{tip} + i(\gamma_{inc}^* - \gamma)z_{tip}] \\ &\quad\times\mathbf{t}_{\lambda_0}^*(\mathbf{k}_{||}^{inc})\cdot\mathbf{e}_0^*\,\mathbf{t}_{\lambda_0}(\mathbf{k}_{||})\cdot\mathbf{e}^{(1)}(-\mathbf{k}_{||})\,\exp(-i\mathbf{k}_{||}\cdot\mathbf{r}_{tip||})\,d\mathbf{k}_{||}\ . \end{aligned} \tag{F13}$$

Similarly, the dynamic part of the near-field intensity at the location of the tip is :

$$\begin{aligned} I^{sa}(\mathbf{r}_{tip}) &= 2\mathrm{Re}\left\{\mathbf{E}^{(0)*}(\mathbf{r}_{tip})\cdot\mathbf{E}^{(1)}(\mathbf{r}_{tip})\right\} \\ &= 2\mathrm{Re}\int\exp[-i\mathbf{k}_{||}^{inc}\cdot\mathbf{r}_{tip} + i(\gamma_{inc}^* - \gamma)z_{tip}]\,\mathbf{e}_0^*\cdot\mathbf{e}^{(1)}(-\mathbf{k}_{||}) \\ &\quad\times\exp(-i\mathbf{k}_{||}\cdot\mathbf{r}_{tip||})\,d\mathbf{k}_{||}\ . \end{aligned} \tag{F14}$$

Following ref.[136], we now try to define a transfer function as the ratio between the Fourier transform of the signal and the Fourier transform of the *near-field intensity*. Both Fourier transforms are evaluated in the plane $z = z_{tip}$, and performed with respect to $\mathbf{r}_{tip\|}$. From (F13) and (F14), one immediately sees that such a transfer function exists when the vectorial nature of the field disappears, namely, for a two-dimensional sample under s-polarized illumination. In this case, our approach justifies the existence of the transfer function defined in ref.[136], under the following conditions : (1) no multiple scattering between the sample and the tip, and (2) the dynamic component of the signal and the near field intensity is much smaller than the background.

In the case of a full vectorial problem, the transfer function introduced in [136] between the signal and the *intensity* does not exist, because of the presence of the projection operator in (F13), which projects the field on $\mathbf{t}_{\lambda_0}$. Thus, each component $e_{ill,k}$ of $\mathbf{e}_{ill}$ has a specific coupling factor $t_{\lambda_0,k}$. A similar conclusion was obtained in §7 : we have shown that an impulse response connecting the *vector* polarization density to the signal to the signal can be defined.

References

[1] D.W. Pohl, W. Denk, and M. Lanz, Appl. Phys. Lett. **44**, 651 (1984).

[2] A. Lewis, M. Isaacson, A. Harootunian, and A. Muray, Ultramicroscopy **13**, 227 (1984).

[3] U. Dürig, D. Pohl, and F. Rohner, J. Appl. Phys. **59**, 3318 (1986).

[4] E. Betzig, A. Lewis, A. Harootunian, M. Isaacson and E. Kratschmer, Biophys. J. **49**, 269 (1986).

[5] *Near-Field Optics*, edited by D. Pohl and D. Courjon (Kluwer, Dordrecht, 1993).

[6] Ultramicroscopy (special issue) **57 (2-3)**, (1995).

[7] Ultramicroscopy (special issue) **61 (1-4)**, (1995).

[8] *Photons and Local Probes*, edited by O. Marti and R. Möller (Kluwer, Dordrecht, 1995).

[9] *Optics at the Nanometer Scale : Imaging and Storing with Photonic Near Fields*, edited by M. Nieto-Vesperinas and N. García (Kluwer, Dordrecht, 1996).

[10] D. Pohl, in *Advances in Optical and Electron Microscopy*, edited by C. Sheppard and T. Mulvey (Academic Press, London, 1990).

[11] D. Pohl, in *Scanning Tunneling Microscopy II*, edited by R. Wiesendanger and H.-J. Güntherodt (Springer-Verlag, Berlin, 1992).

[12] H. Heinzelmann and D. Pohl, Appl. Phys. A **59**, 89 (1994).

[13] D. Courjon and C. Bainier, Rep. Prog. Phys. **57**, 989 (1994).

[14] J. Trautman, E. Betzig, J. Weiner, D. DiGiovanni, T. Harris, F. Hellman and E. Gyorgy, J. Appl. Phys. **71**, 4659 (1992).

[15] H. Bielefeldt, B. Hecht, S. Herminghaus, J. Mlynek, and O. Marti, in *Near Field Optics, NATO ASI series*, edited by D. Pohl and D. Courjon (Kluwer, Dordrecht, 1993), pp. 281–286.

[16] O. Marti, H. Bielefeldt, B. Hecht, S. Herminghaus, P. Leiderer and J. Mlynek, Opt. Commun. **96**, 225 (1993).

[17] P. Adam, L. Salomon, F. de Fornel, and J.-P. Goudonnet, Phys. Rev. B **48**, 2680 (1993).

[18] P. Dawson, K. Smith, F. de Fornel, and J.-P. Goudonnet, Ultramicroscopy **57**, 287 (1995).

[19] P. Dawson, F. de Fornel, and J.-P. Goudonnet, Phys. Rev. Lett. **72**, 2927 (1994).

[20] B. Hecht, H. Bielefeldt, L. Novotny, Y. Inouye and D. Pohl, Phys. Rev. Lett. **77**, 1889 (1996).

[21] J. Krenn, W. Gotschy, D. Somitsch, A. Leitner and F. Aussenegg, Appl. Phys. A **61**, 541 (1995).

[22] J. Krenn, R. Wolf, A. Leitner, and F. Aussenegg, Opt. Commun. **137**, 46 (1997).

[23] D. Tsai, J. Kovacs, Z. Wang, M. Moskovits, V. Shalaev, J. Suh and R. Botet, Phys. Rev. Lett. **72**, 4149 (1994).

[24] S. Bozhevolnyi, I. Smolyaninov, and A. Zayats, Phys. Rev. B **51**, 17916 (1995).

[25] S. Bozhevolnyi, Phys. Rev. B **54**, 8177 (1996).

[26] S. Bozhevolnyi and F. Pudonin, Phys. Rev. Lett. **78**, 2823 (1997).

[27] N. van Hulst, M. Moers, and E. Borgonjen, in *Photons and Local Probes, NATO ASI series*, edited by O. Marti and R. Möller (Kluwer, Dordrecht, 1995), pp. 165-180.

[28] I. Hörsch, R. Kusche, O. Marti, B. Weigl and K.J. Ebeling, J. Appl. Phys. **79**, 3831 (1996).

[29] J. Knight, N. Dubreuil, V. Sandoghdar, J. Hare, V. Lefèvre-Seguin, J. Raimond and S. Haroche, Opt. Lett. **20**, 1515 (1995).

[30] E. Betzig and J. Trautman, Science **257**, 189 (1992).

[31] A. Piednoir, F. Creuzet, C. Licoppe, and J. Ortega, Ultramicroscopy **57**, 282 (1995).

[32] M. Fujihira, in *Optics at the Nanometer Scale : Imaging and Storing with Photonic Near Fields, NATO ASI series*, edited by M. Nieto-Vesperinas and N. García (Kluwer, Dordrecht, 1996), pp. 205–221.

[33] E. Betzig, in *Near Field Optics, NATO ASI series*, edited by D. Pohl and D. Courjon (Kluwer, Dordrecht, 1993), pp. 7–15.

[34] M. Fujihira, H. Monobe, H. Muramatsu, and T. Ataka, Ultramicroscopy **57**, 118 (1995).

[35] M. Fujihira, H. Monobe, N. Yamamoto, H. Muramatsu, N. Chiba, K. Nakajima and T. Ataka, Ultramicroscopy **61**, 271 (1995).

[36] A. Jalocha and N. Van Hulst, J. Opt. Soc. Am. B **12**, 1577 (1995).

[37] A. Jalocha and N. Van Hulst, Opt. Commun. **119**, 17 (1995).

[38] E. Betzig and R. Chichester, Science **262**, 1422 (1993).

[39] R. Kopelman and W. Tan, Science **262**, 1382 (1993).

[40] J. Massanell *et al.*, in *Optics at the Nanometer Scale : Imaging and Storing with Photonic Near Fields, NATO ASI series*, edited by M. Nieto-Vesperinas and N. García (Kluwer, Dordrecht, 1996), pp. 181–190.

[41] S. Jiang, J. Ichibashi, H. Monobe, M. Fujihira and M. Ohtsu, Opt. Commun. **106**, 173 (1994).

[42] E. Betzig, J. Trautman, R. Wolfe, E. Gyorgy, P. Finn, M. Kryder and C.-H. Chang, Appl. Phys. Lett. **61**, 142 (1992).

[43] V. Safarov, V. Kosokubin, C. Hermann, G. Lampel, C. Marlière and J. Peretti, Ultramicroscopy **57**, 270 (1995).

[44] J.-M. Vigoureux, C. Girard, and D. Courjon, Opt. Lett. **14**, 1039 (1989).

[45] J.-M. Vigoureux, F. Depasse, and C. Girard, Appl. Opt. **31**, 3036 (1992).

[46] J.-M. Vigoureux and D. Courjon, Appl. Opt. **31**, 3170 (1992).

[47] C. Girard and D. Courjon, Phys. Rev. B **42**, 9340 (1990).

[48] D. van Labeke and D. Barchiesi, J. Opt. Soc. Am. A **9**, 732 (1992).

[49] D. van Labeke and D. Barchiesi, in *Near Field Optics, NATO ASI series*, edited by D. Pohl and D. Courjon (Kluwer, Dordrecht, 1993), pp. 157-178.

[50] C. Girard and A. Dereux, Rep. Prog. Phys. **59**, 657 (1996).

[51] E. Synge, Philos. Mag. **6**, 356 (1928).

[52] C. McCutchen, J. Opt. Soc. Am. **57**, 1190 (1967).

[53] E. Ash and G. Nicholls, Nature **237**, 510 (1972).

[54] G. Massey, Appl. Opt. **23**, 658 (1984).

[55] U. Fischer, J. Opt. Soc. Am. B **3**, 1239 (1986).

[56] R. Toledo-Crow, P. Yang, Y. Chen, and M. Vaez-Iravani, Appl. Phys. Lett. **60**, 2957 (1992).

[57] E. Betzig, P. Finn, and J. Weiner, Appl. Phys. Lett. **60**, 2484 (1992).

[58] C. Durkan and I. Shvets, J. Appl. Phys. **79**, 1219 (1996).

[59] C. Durkan and I. Shvets, J. Appl. Phys. **80**, 5659 (1996).

[60] U. Fischer, J. Koglin, and H. Fuchs, J. Microsc. **176**, 231 (1994).

[61] J. Koglin, U. Fischer, and H. Fuchs, Phys. Rev. B **55**, 7977 (1997).

[62] B. Hecht, H. Heinzelmann, and D. Pohl, Ultramicroscopy **57**, 228 (1995).

[63] E. Betzig, M. Isaacson, and A. Lewis, Appl. Phys. Lett. **51**, 2088 (1987).

[64] M. Vaez-Iravani, R. Toledo-Crow, and Y. Chen, J. Vac. Sci. Technol. A **11**, 742 (1993).

[65] E. Buckland, P. Moyer, and M. Paesler, J. Appl. Phys. **73**, 1018 (1993).

[66] R. Reddick, R. Warmack, and T. Ferrel, Phys. Rev. B **39**, 767 (1989).

[67] F. de Fornel, J.-P. Goudonnet, L. Salomon, and E. Lesniewska, Proc. SPIE **1139**, 77 (1989).

[68] D. Courjon, K. Sarayeddine, and M. Spajer, Opt. Commun. **71**, 23 (1989).

[69] H. Danzebrink, J. Microscopy **176**, 276 (1994).

[70] H. Danzebrink, O. Ohlsson, and G. Wilkening, Ultramicroscopy **61**, 131 (1995).

[71] M. Specht, J. Pedarnig, W. Heckl, and T. Hänsch, Phys. Rev. Lett. **68**, 476 (1992).

[72] J. Pedarnig, M. Specht, W. Heckl, and T. Hänsch, in *Near Field Optics, NATO ASI series*, edited by D. Pohl and D. Courjon (Kluwer, Dordrecht, 1993), pp. 273–280.

[73] S. Kawata, Y. Inouye, and T. Sugiura, Jpn. J. Appl. Phys. **33**, L1725 (1994).

[74] Y. Inouye and S. Kawata, Opt. Lett. **19**, 159 (1994).

[75] P. Gleyzes, A. Boccara, and R. Bachelot, Ultramicroscopy **57**, 318 (1995).

[76] N. van Hulst, M. Moers, O. Noordman, R. Tack, F. segerink and B. Bölger, Appl. Phys. Lett. **62**, 461 (1993).

[77] F. Zenhausern, M. O'Boyle, and H. Wickramasinghe, Appl. Phys. Lett. **65**, 1623 (1994).

[78] J.-J. Greffet, Opt. Commun. **72**, 274 (1989).

[79] F. Pincemin, A. Sentenac, and J.-J. Greffet, J. Opt. Soc. Am. A **11**, 1117 (1994).

[80] J. Jackson, *Classical Electrodynamics* (Wiley, New york, 1975).

[81] O. Martin, C. Girard, and A. Dereux, J. Opt. Soc. Am. A **13**, 1801 (1996).

[82] J.-C. Weeber, E. Bourillot, A. Dereux, J.-P. Goudonnet, Y. Chen and C. Girard, Phys. Rev. Lett. **77**, 5332 (1996).

[83] A. Madrazo, R. Carminati, M. Nieto-Vesperinas, and J.-J. Greffet, J. Opt. Soc. Am. A **15**, 109 (1998).

[84] M. Xiao, J. Opt. Soc. Am. A **14**, 2977 (1997).

[85] V. Tychinsky, Opt. Commun. **74**, 41 (1989).

[86] V. Tychinsky, Opt. Commun. **81**, 131 (1991).

[87] M. Totzeck and M. Krumbugel, Opt. Commun. **112**, 189 (1994).

[88] B. Knoll, F. Keilmann, A. Kramer, and R. Guckenberger, Appl. Phys. Lett. **70**, 2667 (1997).

[89] C. Bainier, D. Courjon, F. Baida, and C. Girard, J. Opt. Soc. Am. A **13**, 267 (1996).

[90] F. Zenhausern, Y. Martin, and H. Wickramasinghe, Science **269**, 1083 (1995).

[91] N. Garcia and M. Nieto-Vesperinas, Appl. Phys. Lett. **66**, 3399 (1995).

[92] R. Carminati, Phys. Rev. E **55**, R4901 (1997).

[93] S. Bozhevolnyi, O. Keller, and I. Smolyaninov, Opt. Lett. **19**, 1601 (1994)

[94] S. Bozhevolnyi, E. Bozhevolnaya, and S. Berntsen, J. Opt. Soc. Am. A **12**, 2645 (1995).

[95] S. Bozhevolnyi, B. Vohnsen, E. Bozhevolnaya, and S. Berntsen, J. Opt. Soc. Am. A **13**, 2381 (1996).

[96] A. Meixner, M. Bopp, and G. Tarrach, Appl. Opt. **33**, 7995 (1994).

[97] A. Sentenac and J.-J. Greffet, Ultramicroscopy **57**, 246 (1995).

[98] J.-J. Greffet, A. Sentenac, and R. Carminati, Opt. Commun. **116**, 20 (1995).

[99] R. Carminati and J.-J. Greffet, Ultramicroscopy **61**, 11 (1995).

[100] S. Bozhevolnyi and B. Vohnsen, Phys. Rev. Lett. **77**, 3351 (1996).

[101] R. Kaiser, Y. Lévy, N. Vansteenkiste, A. Aspect, W. Seifert, D. Leipold and J. Mlynek, Opt. Commun. **104**, 234 (1994).

[102] A. Roberts and J. Murphy, Opt. Commun. **128**, 41 (1996).

[103] A. Landragin, G. Labeyrie, C. Henkel, R. Kaiser, N. Vansteenkiste, C. Westbrook and A. Aspect, Opt. Lett. **21**, 1591 (1996).

[104] C. Henkel, K. Molmer, R. Kaiser, N. Vansteenkiste, C. Westbrook and A. Aspect, Phys. Rev. A **55**, 1160 (1997).

[105] E. Bourillot, F. de Fornel, J.-P. Goudonnet, D. Persegol, A. kevorkian and D. Delacourt, J. Opt. Soc. Am. A **12**, 95 (1995).

[106] D. Courjon, C. Bainier, and M. Spajer, J. Vac. Sci. Technol. B **10**, 2436 (1992).

[107] M. Moers *et al.*, in *Near Field Optics, NATO ASI series*, edited by D. Pohl and D. Courjon (Kluwer, Dordrecht, 1993), pp. 79–86.

[108] S. Bozhevolnyi, I. Smolyaninov, and O. Keller, Appl. Opt. **34**, 3793 (1995).

[109] N. van Hulst, M. Moers, and B. Bölger, J. Microsc. **171**, 95 (1993).

[110] R. Bachelot, P. Gleyzes, and A. Boccara, Opt. Lett. **20**, 1924 (1995).

[111] F. de Fornel, E. Bourillot, P. Adam, L. Salomon, G. Chabrier and J.-P. Goudonnet, in *Near Field Optics, NATO ASI series*, edited by D. Pohl and D. Courjon (Kluwer, Dordrecht, 1993), pp. 59–70.

[112] F. de Fornel, P. Adam, L. Salomon, J.-P. Goudonnet, A. Sentenac, R. Carminati and J.-J. Greffet, J. Opt. Soc. Am. A **13**, 35 (1996).

[113] L. Novotny, D. Pohl, and P. Regli, J. Opt. Soc. Am. A **11**, 1768 (1994).

[114] D. Barchiesi and D. van Labeke, Microsc. Microanal. Microstruct. **5**, 435 (1994).

[115] R. Carminati and J.-J. Greffet, J. Opt. Soc. Am. A **12**, 2716 (1995).

[116] C. Girard and A. Dereux, Phys. Rev. B **49**, 11334 (1994).

[117] J. Elson, Phys. Rev. B **30**, 5460 (1984).

[118] R. Harrington, *Field Computation by Moment Methods* (IEEE Press, Piscataway, 1993).

[119] O. Martin, A. Dereux, and C. Girard, J. Opt. Soc. Am. A **11**, 1073 (1994).

[120] O. Martin, C. Girard, and A. Dereux, Phys. Rev. Lett. **74**, 526 (1995).

[121] A. Maradudin and D. Mills, Phys. Rev. B **11**, 1392 (1975).

[122] G. Agarwal, Phys. Rev. B **14**, 846 (1976).

[123] J. Cites, M. Sanghadasa, C. Sung, R. Reddick, R. Warmack and T. Ferrel, J. Appl. Phys. **71**, 7 (1992).

[124] U. Schollwöck and H. Wagner, in *Near Field Optics, NATO ASI series*, edited by D. Pohl and D. Courjon (Kluwer, Dordrecht, 1993), pp. 247–254.

[125] A. Madrazo and M. Nieto-Vesperinas, J. Opt. Soc. Am. A **14**, 618 (1997).

[126] E. Kröger and E. Kretschmann, Z. Physik **237**, 1 (1970).

[127] G. Agarwal, Phys. Rev. B **15**, 2371 (1977).

[128] U. Fano, J. Opt. Soc. Am. **31**, 213 (1941).

[129] S. Rice, Commun. Pure Appl. Math. **4**, 351 (1951).

[130] F. Toigo, A. Marvin, V. Celli, and N. Hill, Phys. Rev. B **15**, 5618 (1977).

[131] J.-J. Greffet, Phys. Rev. B **37**, 6436 (1988).

[132] D. Jackson and D. Winebrenner, J. Acoust. Soc. Am **83**, 961 (1988).

[133] L. Kazandjian, J. Acoust. Soc. Am. **92**, 1679 (1992).

[134] V. Tatarskii, J. Opt. Soc. Am. A **12**, 1254 (1995).

[135] D. van Labeke and D. Barchiesi, J. Opt. Soc. Am. A **10**, 2193 (1993).

[136] R. Carminati and J.-J. Greffet, Opt. Commun. **116**, 316 (1995).

[137] A. Madrazo and M. Nieto-Vesperinas, J. Opt. Soc. Am. A **13**, 785 (1996).

[138] C. Girard and X. Bouju, J. Opt. Soc. Am. B **9**, 298 (1992).

[139] C. Girard, A. Dereux, O. Martin, and M. Devel, Phys. Rev. B **52**, 2889 (1995).

[140] E. Mendez, J.-J. Greffet, and R. Carminati, Opt. Commun. **142**, 7 (1997).

[141] L. Landau, E. Lifchitz, and L. Pitaevskii, *Electrodynamics of continuous media* (Pergamon Press, Oxford, 1984).

[142] D. Kermisch, J. Opt. Soc. Am. **67**, 1357 (1977).

[143] D. van Labeke, F. Baida, D. Barchiesi, and D. Courjon, Opt. Commun. **114**, 470 (1995).

[144] G. Chabrier, F. de Fornel, E. Bourillot, L. Salomon and J.-P. Goudonnet, Opt. Commun. **107**, 347 (1994).

[145] R. Carminati, J.-J. Greffet, N. Garcia, and M. Nieto-Vesperinas, Opt. Lett. **21**, 501 (1996).

[146] F. de Fornel, P. Adam, L. Salomon, and J.-P. Goudonnet, Opt. Lett. **19**, 1082 (1994).

[147] P. Adam, L. Salomon, F. de Fornel, and J.-P. Goudonnet, Opt. Commun. **105**, 7 (1994).

[148] C. Pieralli, Opt. Commun. **108**, 203 (1994).

[149] D. Barchiesi and C. Pieralli, SPIE **2782**, 535 (1996).

[150] J. Kann, T. Milster, F. Froehlich, R. Ziolkowski and J. Judkins, J. Opt. Soc. Am. A **12**, 1677 (1995).

[151] J.-C. Weeber, F. de Fornel, and J.-P. Goudonnet, Opt. Commun. **126**, 285 (1996).

[152] J.-C. Weeber, Ph.D. thesis, Université de Bourgogne, France, 1996.

[153] J.-J. Greffet and R. Carminati, in *Optics at the Nanometer Scale : Imaging and Storing with Photonic Near Fields*, *NATO ASI series*, edited by M. Nieto-Vesperinas and N. García (Kluwer, Dordrecht, 1996), pp. 1–26.

[154] R. Carminati and J.-J. Greffet, Opt. Lett. **21**, 1208 (1996).

[155] B. Hecht, H. Bielefeldt, Y. Inouye, D. Pohl and L. Novotny, J. Appl. Phys. **81**, 2492 (1997).

[156] R. Carminati, A. Madrazo, M. Nieto-Vesperinas, and J.-J. Greffet, J. Appl. Phys. **82**, 501 (1997).

[157] S. Bozhevolnyi, J. Opt. Soc. Am. B **14**, 2254 (1997).

[158] L. Novotny, Appl. Phys. Lett. **69**, 3806 (1996).

[159] D. van Labeke, D. Barchiesi, and F. Baida, J. Opt. Soc. Am. A **12**, 695 (1995).

[160] D. Barchiesi, C. Girard, O. Martin, D. van Labeke and D. Courjon, Phys. Rev. E **54**, 4285 (1996).

[161] D. Courjon, J.-M. Vigoureux, M. Spajer, K. Sarayeddine and S. Leblanc, Appl. Opt. **29**, 3734 (1990).

[162] U. Fischer, U. Durig, and D. Pohl, Appl. Phys. Lett. **52**, 249 (1988).

[163] S. Bozhevolnyi, M. Xiao, and O. Keller, Appl. Opt. **33**, 876 (1994).

[164] T. Sugiura, T. Okada, Y. Inouye, O. Nakamura and S. Kawata, Opt. Lett. **22**, 1663 (1997).

[165] J. van Bladel, *Singular Electromagnetic Fields and Sources* (Clarendon, Oxford, 1991).

[166] C. Girard, O. Martin, and A. Dereux, Phys. Rev. Lett. **75**, 3098 (1995).

[167] L. Mandel and E. Wolf, *Optical Coherence and Quantum Optics* (Cambridge University Press, Cambridge, 1995).

[168] T. Ferrell, S. Sharp, and R. Warmack, Ultramicroscopy **42–44**, 408 (1992).

[169] N. van Hulst, F. Segrink, F. Achten, and B. Bölger, Ultramicroscopy **42–44**, 416 (1992).

[170] N. García and M. Nieto-Vesperinas, Phys. Rev. Lett. **71**, 3645 (1993).

[171] J.-J. Greffet and R. Carminati, Ultramicroscopy **61**, 43 (1995).

[172] J. Goodman, *Introduction to Fourier Optics* (McGraw-Hill, New York, 1968).

[173] J. Goodman, *Statistical Optics* (Wiley, New York, 1985).

[174] N. García and M. Nieto-Vesperinas, Opt. Lett. **20**, 949 (1995).

[175] V. Sandoghdar, S. Wegscheider, G. Krausch, and J. Mlynek, J. Appl. Phys. **81**, 2499 (1997).

[176] H. Yamada, H. Tokumoto, S. Akamine, K. Fukuzawa and H. Kuwano, J. Vac. Sci. Technol. B **14**, 812 (1996).

[177] Y. Martin, F. Zenhausern, and H. Wickramasinghe, Appl. Phys. Lett. **68**, 2475 (1996).

[178] Chen-To Tai, *Dyadic Green Functions in Electromagnetic Theory*, 2nd ed. (IEEE Press, Piscataway, 1994).

[179] M. Nieto-Vesperinas, *Scattering and Diffraction in Physical Optics* (Wiley, New York, 1991).

[180] A. Yaghjian, Proc. IEEE. **68**, 248 (1980).

[181] O. Keller, in *Optics at the Nanometer Scale : Imaging and Storing with Photonic Near Fields*, *NATO ASI series*, edited by M. Nieto-Vesperinas and N. García (Kluwer, Dordrecht, 1996), pp. 63-93.

[182] O. Keller, Phys. Rep. **268**, 85 (1996).

[183] B. Labani, C. Girard, D. Courjon, and D. Van Labeke, J. Opt. Soc. Am. B **7**, 936 (1990).

[184] C. Girard and X. Bouju, J. Chem. Phys. **95**, 2056 (1991).

[185] M. Xiao and S. Bozhevolnyi, Opt. Commun. **130**, 337 (1996).

[186] O. Keller, M. Xiao, and S. Bozhevolnyi, Surface Science **280**, 217 (1993).

[187] A. Lakhtakia, Opt. Commun. **79**, 1 (1990).

[188] A. Castiaux, C. Girard, A. Dereux, O. Martin and J.-P. Vigneron, Phys. Rev. E **54**, 5752 (1996).