Apertureless near-field optical microscopy: Differences between heterodyne interferometric and non-interferometric images

Ruben Esteban a,b,c,*, Ralf Vogelgesang a, Klaus Kern a,d

a Max Planck Institut für Festkörperforschung, 70569 Stuttgart, Germany
b Centro de Física de Materiales, Centro Mixto CSIC-UPV/EHU and Donostia International Physics Center (DIPC), 20018 Donostia-San Sebastián, Spain
c Laboratoire Charles Fabry, Institut d’Optique, CNRS, Université Paris-Sud, Campus Polytechnique, RD 128, 91127 Palaiseau Cedex, France
d Institut de Physique des Nanostructures, Ecole Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland

1. Introduction

The possibly best known limitation to the wealth of information achievable with traditional optical microscopy is the classical diffraction limit. However, different techniques under current development can overcome this limitation. We focus in this paper on apertureless Scanning Near-Field Optical Microscopy (aSNOM) [1,2].

aSNOM typically uses a sharp, oscillating tip to gain information about the sample with lateral resolution in the order of the tip apex radius. In an active configuration [3,4], the tip apex interacts strongly with a closely situated substrate and gives rise to intense and very localized near fields. The near fields are, however, not directly accessible, and image construction relies on the detection of scattered far fields. A first crucial aspect is thus the relationship between the near-field interaction and the scattered fields [5]. The method used to detect and analyze such far fields is also relevant. For example, demodulation at the higher harmonics of an oscillating tip has been shown useful to improve lateral resolution and suppress background [6]. The background, for example from the bulk or the tip, complicates the discrimination of the high resolution information and its suppression is convenient for experimental interpretation.

In recent studies the elastically scattered radiation was detected using both interferometric [7,8] and non-interferometric [9–11] schemes. These two schemes are not equivalent, though, and the differences go beyond the possibility to obtain a stronger signal and phase information with the former. The differences have been qualitatively discussed and experimentally shown [12,13]. To better understand the differences, we start by considering the recorded signal $S_n$ obtained from demodulation with a dual-phase lock-in amplifier at the $n$-th harmonic of the oscillation angular frequency $\omega$

$$S_n \propto \frac{1}{T} \int_0^T I_{\text{det}} e^{i(n \omega + \Theta)} \, dt,$$  

where $T = 2\pi / \omega$ is the oscillation period and $\Theta$, zero in the following if not mentioned otherwise, allows to introduce the internal phase (relative to the tip oscillation) of the lock-in amplifier. $I_{\text{det}}$ is the current generated at the photodetector, and it differs between interferometric and non-interferometric schemes [14]. For the latter case and under convenient conditions it takes the form

$$I_{\text{det}} \propto \int_{A_{\text{det}}} |E_{\text{sca}}|^2 \, dA_{\text{det}}$$

where $A_{\text{det}}$ is the area of the detector, oriented perpendicular to the propagation direction of the detected radiation. The complex-valued $E_{\text{sca}} = |E_{\text{sca}}| e^{i\Theta_{\text{sca}}}$ refers to the electrical field of the...
signal scattered by the tip–substrate system. To illustrate one
significant difficulty that appears in non-interferometric
detection, we consider a sample consisting of an object with deep-
subwavelength structure supported by a substrate, which itself
hardly varies spatially and only on a scale of a wavelength
or more.

Intuitively, it is appealing to introduce a conceptual decom-
position of the scattered fields [13]. With $E_{\text{obj}} = |E_{\text{obj}}| \exp(i \theta_{\text{obj}})$ we refer to components that stem from the interaction between the
tip apex and the subwavelength object. The remainder, $E_{\text{bkg}} =
|E_{\text{bkg}}| \exp(i \theta_{\text{bkg}})$, may be considered a slowly varying “back-
ground”. $E_{\text{obj}}$ and $E_{\text{bkg}}$ are related to those spatial frequencies
in the recorded images that are higher and lower than roughly half
a wavelength. Only $E_{\text{obj}}$ carries high spatial resolution information.
But it is difficult, if not impossible, to give an exact, fully
satisfactory algebraic definition and we use the terms $E_{\text{obj}}$ and
$E_{\text{bkg}}$, here merely to facilitate the discussion, without attempting
a rigorous distinction. Nevertheless, one important conclusion can
be drawn immediately even from this conceptual distinction. A
self-homodyne term $|E_{\text{bkg}}| E_{\text{bkg}} \cos(\theta_{\text{bkg}} - \theta_{\text{obj}})$ occurs in the
total scattered field, $E_{\text{det}}^2 = |E_{\text{bkg}} + E_{\text{obj}}|^2$, which mixes both
contributions. The strength and phase of $E_{\text{bkg}}$ thus affect the value of
Eq. (1) in a non-trivial way for non-interferometric detection. In
general, this will complicate image interpretation even if $E_{\text{bkg}}$
would be perfectly constant [15].

We now consider an interferometric detection scheme. That is,
before impinging onto the photodetector, the scattered signal
interferes with a collinear, planar wave $E_{\text{ref}}^p = E_{\text{ref}}^p e^{i \phi_{\text{ref}}^p}$. Conside-
ing $E_{\text{ref}}^p$ constant and sufficiently large, the relevant $I_{\text{det}}$ expression
for harmonic $n > 0$ is

$$ I_{\text{det}} \propto \int_{\text{area}} |E_{\text{ref}}|^2 |E_{\text{obj}}| e^{i \phi_{\text{ref}} + \phi_{\text{obj}}} \, dA_{\text{det}}. $$

(3)

Only the interferometric term appears because the constant
contribution $|E_{\text{ref}}|^2$ will be filtered out by the harmonic demodu-
lation and we assume the term $|E_{\text{ref}}|^2$ negligibly small. The super-
index $p$ indicates the polarization selected with the reference
beam. The complex exponential phase factor in the integrand
Corresponds to homodyne in quadrature, heterodyne, balance
homodyning, phase-shifting or related interferometric techniques
[14,16,17]. The expressions already indicate that interferometry
gives access to optical phase information, and that the presence of
$E_{\text{ref}}^p$ allows to boost the detected signal.

A superficial look at the integrands of Eqs. (2), (3) might
suggest an otherwise simple relation between the module of the
non-interferometric and heterodyne interferometric signals, with
the differences related to the presence of $|E_{\text{ref}}|^2$ or $|E_{\text{ref}}|$, respec-
tively. However, the presence of the phase $\theta_{\text{obj}}$ renders the
differences more intricate. For example, the phase modulation due
to the tip oscillation affects also the demodulated signal amplitude
$S_n$ [14,18,19,6]. If we explicitly decompose $E_{\text{det}}^m = E_{\text{obj}} + E_{\text{bkg}}$ for the
interferometric case [12] as we did for non-interferometric detec-
tion, the relevant integrand in Eq. (3) turns into

$$ \left( |E_{\text{obj}}| e^{-i \phi_{\text{obj}}} + |E_{\text{bkg}}| e^{-i \phi_{\text{bkg}}} \right) |E_{\text{ref}}|^2 e^{i \phi_{\text{ref}}} $$

(4)
in the case, the simple additive form seems promising for filtering out
the slowly varying $E_{\text{bkg}}^m$, for example by the high harmonic
demodulation in Eq. (1), and thus better discriminating the high
resolution information contained in $E_{\text{obj}}^m$. Notice that taking the
module $|S_n|$ of the recorded signal does reintroduce a certain
cross-talk between both contributions. Thus, it may be better to
consider the full complex signal, for example, if the recorded image
is spatially filtered in a post-processing step.

These considerations and previous work [20,12] suggest that the
different dependence of $|S_n|$ on tip position results in clearer
images for heterodyne interferometric than for non-interfero-
metric measurements. To our knowledge, a detailed quantitative
account, however, has yet to be given. Numerical simulation
[10,36] of scenarios closely following the experimental conditions
is a challenging prerequisite. In particular, it appears necessary to
model large probe tips [21,22] which may be interacting strongly
with the sample. It is also not sufficient to compute only the near
fields or the total scattering cross section. The amplitude and
phase distribution of the electric fields over the detector area
must be computed, and the simulations must capture accurately
the behavior of at least the first few derivatives of the signal $I_{\text{det}}$
as they play a key role in the value of the different harmonics $S_n$
[6,14].

In our recent work, we have developed simulations that
consider these aspects [21], but we did not analyze the influence
of interferometry on the resulting quantities. In the following we
simulate approach curves and one-dimensional scans and infer
different qualitative features of non-interferometric and hetero-
dyne interferometric detection schemes relevant to experimental
and theoretical work.

2. Geometry and simulations

For our simulations we use a multiple multipole method, as
implemented in the numerical platform MaX-1 [24]. It employs
expansions that are vectorial solutions (or excellent numerical
approximations) of the Maxwell equations. MaX-1 minimizes the
error at the boundaries, in our case to less than 0.5 percent
average value. Moderately increasing the number of free param-
eters used to minimize the error did not significantly alter the
results of this paper, which supports a good numerical conver-
gence of the obtained results.

A silicon tip illuminated by an external excitation oscillates
along the vertical $z$ direction parallel to its axis, with $z$ decreasing
as the tip approaches the top of a sample. Fig. 1(a) sketches the
geometry with one sample consisting of a glass sphere of 2 µm
diameter containing fully a 10 nm radius gold spherical inclusion,
the center of both aligned along $z$ and their upper point at 1 nm
distance. Fig. 1(b) represents the fields for this geometry and a
particular tip position and will be, together with Fig. 1(c),
explained more in detail later on. We will be also interested in
a large homogeneous gold sphere of the same 2 µm diameter and
in two smaller gold spheres in vacuum (Fig. 4(a)), of 10 and 30 nm
radius, their top surface at the same $z$ position and their centers
separated 60 nm in the $x$ direction; the $x$ direction is perpendicu-
lar to $z$ and tangent to the different spheres at their upper point,
with $xz$ containing the tip axis. The origin of the coordinate axis is
at the upper point of the gold substrate or 1 nm over the 10 nm
gold inclusions. To minimize the effect in non-interferometric
detection of the self-homodyne term $|E_{\text{bkg}}| E_{\text{obj}} \cos(\theta_{\text{bkg}} - \theta_{\text{obj}})$
from a constant contribution from an extended substrate, we
subtract the fields scattered by the bare 2 µm large spheres, either
glass or gold, when illuminated in absence of any probing tip or
inclusion.

The modeled tip is a $C^{(1)}$ continuous cone with cylindrical
symmetry, capped by two hemispheres. The smaller hemisphere
probes the sample and has a radius $R = 10$ nm. The 1400 nm long
tips are used for the homogeneous gold sample, while the tip is
500 nm long for the other substrates. The cone half angle, i.e the
angle between the generatrix and the axis, is 10°. The excitation
makes a 70° angle with the axis of the tip, it is $p$ (TM) polarized
with $xz$ being the plane of incidence and its wavelength $\lambda$ is
$\sim 514.5$ nm. At this frequency, $\varepsilon = 2.1$, 17.76 + 0.508i and
$-3.95 + 2.58i$ for the glass, silicon [25] and gold [26], respectively.
Simple planar waves excite the 500 nm long tips, but we use a
The signal is detected by a dual-phase lock-in amplifier at a higher harmonics of the frequency of oscillation \( \omega \) of the probing tip, according to Eq. (1). During the oscillation the \( z \) position of the lower apex point follows \( z^{\text{tip}} = z_{\text{tip}0} + A(1 + \cos(\omega t)) \), where \( A \) is the oscillation amplitude. The value of \( \omega \) does not influence the final result. \( z_{\text{tip}0} \) is the corresponding \( z \) position when the tip is closer to the samples. For scans of the samples containing a 2 \( \mu \)m diameter glass sphere (Figs. 1 and 3) \( z_{\text{tip}0} \) is chosen so that the distance between the lower point of the oscillating tip and the glass substrate is kept constant for all \( x \). The tip scanning movement is thus not perfectly parallel along the \( x \) direction, but bends slightly. For the sample consisting of a 10 and a 30 nm sphere (Fig. 4), the scan trajectory follows \( x \) exactly and \( z_{\text{tip}0} \) is always the same.

Occasionally, the near fields will be of interest, in particular the near-field enhancement, i.e. the maximum of the electric field module in the proximity of the tip apex normalized to the module of the incident electric field. More relevant to experimental work, we also consider the scattered far fields. A rotationally symmetric lens of numerical aperture 0.342 whose axis is placed in the backscattering direction collects elastically scattered radiation. We model the lens as ideal, in the sense that it converts without power losses a spherical wave at the coordinate origin into a beam of constant phase at the output plane of the lens. We neglect any effects of further propagation or additional optical devices on the resulting beam. A planar photodetector perpendicular to the incoming radiation serves to convert the optical into an electrical signal. Precisely the different relationships between scattered fields and photodetector response distinguish heterodyne interferometric from non-interferometric detection schemes, as described by Eqs. (3) and (2), respectively. For the interferometric case, the scattered signal \( E_{\text{scat}} \) is mixed with a reference planar wave of identical propagation direction but significantly stronger amplitude and a slightly shifted frequency. The polarization of the reference planar wave selects the \( p \)-polarized, in plane (\( xz \) plane) component of the scattered electric fields incident onto the detector. Notice that the considered scenarios are symmetric with respect to the \( xz \) plane and an integral analogous to Eq. (3) cancels for the out-of-plane field component. As Eq. (3) depends on the strength of the reference beam, no attempt is made to compare the absolute strengths of the interferometric and non-interferometric signals. While the discussion in this paper mostly focuses on the absolute signal amplitude \( |S_n| \), it is instructive to also consider the signed value \( S_n \) for the non-interferometric case. In this case, each sign corresponds to one of the two possible phases, which are separated by 180°.

We first study the evolution of the signal as the tip–substrate distance is varied. These traces, also referred to as approach curves [28], are frequently helpful in understanding aSNOM behavior. If the signal is dominated by the tip–substrate interaction in a small sample volume in the close proximity of the tip apex, as desired, the strength of the signal should significantly increase for distances smaller than the apex radius. Nonetheless, while approach curves are useful, a more complete aSNOM analysis requires to implicitly include the imaging process [10]. We study the evolution of the signal for line scans of the patterned substrates. Notice that this requires a two-dimensional movement of the tip as it oscillates along \( z \) and scans along \( x \).

3. Results

To illustrate the differences between heterodyne interferometric and non-interferometric schemes, we plot in Fig. 2 approach curves over the upper point of the 2 \( \mu \)m diameter homogeneous gold sphere, for different oscillations amplitudes and second and third harmonic demodulation. \( |S_n| \) markedly increases for small tip–substrate distance and both harmonics when the oscillation amplitude is small, a manifestation of the expected discrimination of the information from the tip–substrate interaction. Small oscillation amplitudes also result in a lower signal level, however, which often requires a compromise. For a given oscillation amplitude, the change of behavior for short distances is more clearly appreciated for the third than for the second harmonic, both for interferometric and for non-interferometric signals.

The former similarities notwithstanding, the heterodyne interferometric (Fig. 2(e), (f)) and non-interferometric (Fig. 2(a)–(d)) approach curves are different. How changing the tip position affects the recorded signal clearly depends on the detection scheme used. The differences are perhaps more apparent at the second harmonic and large oscillation amplitudes. A more pronounced increase in \( |S_n| \) is observed for short tip–substrate distances in the interferometric case. Notice also that the non-interferometric signal can equal zero, which suggests a change of sign as confirmed by plotting \( S_n \) instead of \( |S_n| \) (Fig. 2(e), (d)).
We next consider the gold inclusion in the glass substrate. Fig. 1(b) shows the electric field modulus in the proximity of the apex for a given tip position [21]. The small gap between the tip and the inclusion distinctly exhibits the strongest fields. We include in Fig. 1(c) additional information on the behavior of the near and far fields. For this particular geometry, but not in general [29], the near-field enhancement broadly resembles the field distribution in the absence of the tip. Both are characterized by a clear maximum for $x$ close to 0, corresponding to the position of the inclusion, with the most significant difference being the stronger fields in the presence of the tip. When the tip is at $x \approx 0$, the near-field enhancement is $\sim 35.5$. In comparison, the scattered electric near-field module at 1 nm above the substrate in the absence of the tip reaches a maximum of only about twice the excitation. The maximum field in the absence of the tip is not exactly at $x = 0$ due to the oblique illumination.

The near-field magnitude is relatively weak when the tip apex is situated above the bare substrate but increases strongly above a gold inclusion, particularly at $x = 0$, which indicates an enhanced interaction [30–32]. In contrast, the non-interferometric, undemodulated far field signal $S_{0}$ is only weakly dependent on the tip position and shows no clear signature from the inclusion. Similarly, the modulus of the undemodulated heterodyne interferometric signal (not shown here) reveals a differently sloped but otherwise nearly featureless scan. However, a close correlation between the strength of this interaction and the detected far field signal would be convenient to interpret measurements in an active configuration.

Higher harmonic demodulation allows regaining high resolution information. We consider a fixed tip-sample distance of 1 nm at the lower oscillation point and an increasing oscillation amplitude. Fig. 3 shows second and third harmonic demodulated signal scans along $x$ according to Eq. (1) for both heterodyne interferometric and non-interferometric detection. Again, some common aspects emerge. A local extremum at $x \sim 0$ reveals the presence of the inclusion. It is especially distinct for small oscillation amplitudes and illustrates the aSNOM capabilities for nanometer resolution [8] of subsurface objects [23,33]. The traces for small oscillation amplitude presented here resemble the behavior of the interaction strength, as observed from the near-field enhancement near the apex in Fig. 1(c).

Looking in closer detail, a narrow global maximum appears for all considered heterodyne interferometric detection traces (Fig. 3(e), (f)) when the tip is situated directly above the gold inclusion. This maximum rides on a slowly varying background contribution whose average value and small slope depend on the oscillation amplitude. Thus, there is a consistent and relatively straightforward connection with the near-field signal.

For non-interferometric detection, a clear global maximum of the $S_{n}$ signal—which should be indicative of the tip being close to the inclusion—appears only for certain experimental conditions (Fig. 3(a), (b)). In other cases the extremum at $x \sim 0$ becomes even a local minimum. Notice also the often significant slope for large $x$. This slope could be related with the self-homodyne term $|E_{0}|^{2}/|E_{bg}|^{2}$ due to nearly constant background fields. The fields scattered by the isolated glass spherical substrate were subtracted and thus do not contribute, but the large tip makes other significant contributions to be expected. Including strong scattering from a real substrate may lead to a more marked effect and more difficult to interpret non-interferometric images.

Plotting the signed non-interferometric signal $S_{n}$ (Fig. 3(c), (d)) results in a more regular evolution with oscillation amplitude. In this case, the scans for the second harmonic and all considered oscillation amplitudes show a local minimum for $x \sim 0$ and, outside the central region, a negative slope for increasing $x$. They become a maximum and a positive slope for the third harmonic traces. Notice that sign of the slope and the nature of the extremum depends on the reference angle $\Theta_{bg}$ of the lock-in amplifier—any 180° shift corresponding to a sign change.

Last we consider scans over a sample consisting on two gold spheres of 10 and 30 nm radius. We have considered the spheres placed in vacuum, which is not realistic but simplifies the numerical burden and should nonetheless give important insights. Fig. 4(a) represents a snapshot of the electric fields module for a given tip position, while Fig. 4(b) shows the near-field enhancement as the tip moves along $x$ without oscillating. The lower point of the tip apex is kept constant at 2 nm over the upper point of the spheres. The obtained trace presents marked maxima when the tip is close to any of the gold spheres, with a clearly stronger response in the case of the larger sphere. Stronger fields for larger radius is unsurprising from previous work [34], although a more detailed study for the present case would require to consider the spectral response.
interferometric schemes. We have discussed in previous works [35,5,21] the detrimental effects of simplifications such as using small tips. To increase the reliability of the obtained conclusions, we are interested in models that approach typical experimental conditions as closely as possible. An important step towards this objective is the inclusion of large tips and, in some of the treated cases, extended substrates: The 1400 nm tip used for the approach curves is several wavelengths long and numerically close to the even longer tips of typical experiments. The somewhat shorter tips (500 nm) considered in the one-dimensional scans help to reduce the numerical complexity of the simulation, while being likely long enough to capture most significant phenomena. The homogeneous spheres of 2 μm diameter serve as an approximately flat substrate in the proximity of the apex.

Some differences remain between a large sphere and the more typical semi-infinite flat substrate, such as the existence of whispering gallery modes in the former. Real samples are significantly more complicated than either, and we chose the large sphere for numerical simplicity. In this respect, we discussed in the introduction how non-interferometric detection is affected by any constant (independent of tip position) contributions to the scattered fields, whereas heterodyne interferometric detection typically is not. This contribution can depend strongly on sample details, and may significantly complicate image interpretation in realistic experiments. To reduce the effect in the present study, we subtracted the constant backscattering component from the 2 μm diameter spheres in isolation.

For the non-interferometric scheme, we can simply consider just the amplitude $|S_n|$ at the dual-phase lock-in amplifier according to Eq. (1) or we can in addition associate 180° phase jumps with a sign change. The latter implies working with negative values, conceptually not a problem because our demodulated signal corresponds to the coefficient of a Fourier series, which can be non-positive. The simulated heterodyne interferometric results focus on the amplitude of the magnitude measured by the lock-in amplifier. Still, for accurate results we must consider the phase of the scattered optical fields. Notice that it is necessary to include the effect of the collecting lens geometry to correctly predict the phase distribution on the photodetector.

The simulated approach curves (Fig. 2) offer a direct connection to much work done in apertureless near-field optical microscopy. They serve to illustrate how the dependence of the signal amplitude $|S_n|$ on tip position is different for heterodyne interferometric and non-interferometric schemes.

The simulation of non-spatially homogeneous samples allows to study the differences in more detail. We consider as experimentally favorable a monotonous dependence between the final signal and the strength of the tip apex–substrate interaction, which we have associated to the near-field enhancement near the tip apex. Under this criterion, and remembering the results displayed in Figs. 1(c) and 4(b), a strong signal should be obtained when the tip is directly over a gold sphere and a weaker and nearly constant value when far away.

The obtained non-interferometric signal often departs from expectations. It frequently seems convenient to consider the signal and treat $S_n$ instead of $|S_n|$; still, even then the results are often difficult to interpret. This difficulty is already apparent in the strong slope of some of the traces for the inclusion in glass, but it is particularly evident for the sample consisting of two gold spheres.

The heterodyne interferometric results are significantly more encouraging. The correspondence between near-field enhancement and far fields is not exact, with notably the maxima from the small inclusion in Fig. 4 more marked for the near-field enhancement, but it seems nonetheless quite satisfactory and a straightforward interpretation of the demodulated results was
possible. For other geometries more complex images will likely result—for example, if the background contribution to the demodulated signal becomes sufficiently large and adds destructively to the signal of interest. In this case, it may be convenient to retain the complex-valued demodulated S0 signal, and not only its amplitude as in the present discussion. This should facilitate spacial filtering of the recorded image. Preliminary results appear to support these assertions.

In summary, we have contrasted non-interferometric and heterodyne interferometric detection schemes for apertureless SNOM. Both are capable of extracting local information about the chosen substrate. The correspondence between the tip–substrate interaction strength and the modulus of the measured signal is more robust to changes on the scanning conditions and more directly interpretable for heterodyne interferometric detection. For non-interferometrically detected signals, considering the sign of the signal, as obtained from the phase of the lock-in output signal, seems beneficial.

References