Technische Universität München

Quantitative Near-field Phonon-polariton Spectroscopy

Nenad Ocelić

Vollständiger Abdruck der von der Fakultät für Physik
der Technischen Universität München
zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzender:	Univ.-Prof. Dr. H. Friedrich
Prüfer der Dissertation:	1. Univ.-Prof. Dr. M. Stutzmann
	2. Hon.-Prof. Dr. W. Baumeister

Die Dissertation wurde am 19.07.2007 bei der Technischen Universität München
eingericht und durch die Fakultät für Physik am 24. 08. 2007 angenommen.

Contents

Abstract

Scattering-type scanning near-field optical microscopy (s-SNOM) is a versatile optical imaging technique which circumvents the diffraction barrier by detecting the light scattered from a sharp probing tip oscillating above the specimen surface. Significant improvements in the theoretical modeling, signal detection and applicability of the s-SNOM are presented in this thesis.

For the first time it is demonstrated that the s-SNOM operating at infrared frequencies can be employed for the nanoscale mapping of crystallinity in polar materials. To this end, the crystal lattice degradation caused by ion beam implantation in a SiC crystal is investigated by near-field infrared spectroscopy in the 9-11 µm wavelength range. It is found that the strength of the phonon-polariton resonant near-field interaction between the s-SNOM probe and the sample rapidly diminishes with the accumulation of defects in the crystal lattice. The near-field optical contrast between the crystalline and amorphous SiC exceeds 30 dB, making the near-field resonance strength a very sensitive measure of the crystal lattice quality. S-SNOM images of SiC crystal surfaces patterned by focused ion beam (FIB) implantation reveal a spatial optical resolution better than 50 nm ($\lambda/200$).

It is additionally shown that IR s-SNOM is sensitive to the stacking sequence of atomic layers in a crystal (polytypism). Longitudinal optical phonon frequencies in SiC polytypes are separated by only 2-3 cm^{-1}, resulting in very fine near-field optical contrasts. A proof is presented here that the commonly employed non-interferometric and homodyne interferometric s-SNOM signal detection methods do not yield reproducible results with weak contrasts like those observed with SiC polytypes. The reason for such behavior is found in the interference between the near-field and background scattering. To overcome this problem and obtain accurate and reproducible near-field spectra even with weak structural contrasts, a new ”pseudo-heterodyne” interferometric method for s-SNOM signal detection has been developed. The pseudo-heterodyne technique avoids the background interference by applying a sinusoidal phase modulation to the interferometer reference wave. It also of-

fers the advantage of a simple experimental implementation and applicability in the entire wavelength range from near-UV to far-IR.

Unambiguous material identification from the s-SNOM spectra has been up to now hindered by quantitative discrepancies between the calculated and the measured near-field contrasts. To resolve this problem, a ”monopole” model of the probe-sample near-field interaction is introduced in this work. It is derived by representing the s-SNOM probe as a prolate spheroid and subsequently reducing it to point charges which give the dominant contribution to the near-field interaction. The new model provides an unprecedented quantitative agreement with the experimentally observed near-field contrasts. Moreover, the final closed-form analytical solution is simple enough to be used in the inverse way to determine the complex dielectric function of the sample from the demodulated near-field signal at each pixel in an image.

The monopole model combined with the pseudo-heterodyne near-field spectroscopy has the potential to transform the infrared s-SNOM into a powerful analytical tool for noninvasive optical probing of the local chemical composition and structural properties on the nanometer scale.

Zusammenfassung

Die optische Streulicht-Nahfeldmikroskopie (scattering-type near-field optical microscopy, s-SNOM) ist eine vielseitige optische Mikroskopiemethode, bei der das Beugungslimit durch die Detektion des Streulichts einer über der Probenoberfläche gerasterten Sonde überwunden wird. Im Rahmen dieser Arbeit wird eine deutliche Verbesserung der theoretischen Beschreibung, der Signaldetektion, sowie der Anwendbarkeit dieses Mikroskopieverfahrens vorgestellt.

Es wird erstmals demonstriert, dass Streulicht-Nahfeldmikroskopie mit Infrarotbeleuchtung (IR s-SNOM) dazu benutzt werden kann, die Kristallinität von polaren Materialien auf der Nanometerskala abzubilden. Dazu wird die Schädigung der Kristallstruktur eines ionenimplantierten Siliziumkarbid Kristalls (SiC) durch Nahfeldspektroskopie im infraroten Wellenlängenbereich ($\lambda = 9\text{-}11\,\mu\text{m}$) untersucht. Die Stärke der phonon-resonanten Nahfeldwechselwirkung zwischen der Nahfeldsonde und der Probe nimmt dabei mit zunehmender Defektdichte im Kristallgitter rasch ab. Der nahfeldoptische Kontrast zwischen kristallinem und amorphen SiC beträgt mehr als 30 dB, weshalb die Stärke der Nahfeldresonanz ein extrem empfindliches Mass für die Kristallqualität darstellt. Anhand der nahfeldoptischen Bilder der mit fokussierten Ionenstrahlen (FIB) strukturierten SiC Oberflächen kann eine laterale optische Auflösung von unter 50 nm ($\lambda/200$) nachgewiesen werden.

Weiterhin wird gezeigt, dass das IR s-SNOM auch auf die Stapelfolge von Atomlagen in einem polaren Kristall (Polytypismus) sensitiv ist. Die longitudinalen optischen Phononfrequenzen von SiC Polytypen unterscheiden sich um lediglich 2-3 cm^{-1}, was zu einem sehr geringen Nahfeldkontrast führt. Anhand dieses Kontrasts wird gezeigt, dass die weithin verbreitete nicht-interferometrische oder homodyn-interferometrische Strreulichtdetektion keine reproduzierbaren Resultate von schwachen Kontrasten ergibt. Der Grund hierfür ist die Interferenz von Nahfeld- und Hintergrund-Streulicht. Dieses Problem wird durch eine „pseudo-heterodyne“ interferometrische Detektionsmethode überwunden. Das in dieser Arbeit neu entwickelte Detektionsverfahren liefert erstmals präzise und reproduzierbare

Nahfeldspektren von SiC Polytypen. Die pseudo-heterodyne Technik vermeidet störende Interferenzen durch eine sinusförmige Phasenmodulation des Referenzstrahls. Weitere Vorteile sind die einfache experimentelle Implementierung, sowie die Anwendbarkeit im kompletten Spektralbereich zwischen UV und Ferninfrarot.

Bisher war eine eindeutige Materialidentifizierung mit dem s-SNOM durch quantitative Unterschiede zwischen den theoretischen und experimentellen Nahfeldspektren oftmals nicht möglich. Um dieses Problem zu lösen, wird in der vorliegenden Arbeit ein neues Modell für die Nahfeldwechselwirkung zwischen Nahfeldsonde und Probe eingeführt. Bei der Herleitung des sogenannten „Monopol“-Modells wird die Nahfeldsonde zunächst durch ein Spheroid ersetzt. Die elektrischen Felder am Apex lassen sich näherungsweise durch eine Punktladung beschreiben, mit deren Hilfe eine analytische Näherung der nahfeldoptischen Wechselwirkung mit der Probenoberfläche gefunden wird. Die Ergebnisse des Modells zeigen eine hervorragende quantitative Übereinstimmung mit experimentellen Nahfeldkontrasten. Aufgrund der geschlossenen analytischen Lösung könnte das Modell in Zukunft angewendet werden, um aus den experimentellen Nahfeldsignalen an jedem Bildpunkt die komplexe dielektrische Funktion der Probe zu berechnen.

Die Streulicht-Nahfeldmikroskopie stellt bereit jetzt ein vielversprechendes Verfahren zur optischen Nanoanalytik dar. Sie bietet einmalige Möglichkeiten zur zerstörungsfreien Charakterisierung sowohl der lokalen chemischen Zusammensetzung als auch von lokalen strukturellen Materialeigenschaften. Die experimentellen und theoretischen Entwicklungen der vorliegenden Arbeit sollten in Zukunft zu weiteren faszinierenden Anwendungen führen, etwa zur zerstörungsfreien Charakterisierung nanoskaliger Spannungsfelder oder zur Analyse der Nanokompositen mit einer Auflösung in nm-Bereich.

1 Introduction

We are surrounded by a large variety of natural and artificial objects displaying interesting structural properties on the nanometer scale, i.e. the length scale extending from about one to hundred nanometers. Examples range from structured materials such as nanocomposites and quantum dots, over electronic components like transistors and memory cells, to biological objects including single molecules, protein complexes, viruses, etc. The structure of these objects cannot be directly observed by conventional optical microscopes due to the inherent resolution barrier imposed by the wave-like nature of light. In particular, the diffraction of light waves limits the smallest distance between two resolvable points to about half the wavelength. Consequently, even the best optical microscopes cannot resolve features smaller than about 200 nm.

The demand for the spatial resolution significantly better than 200 nm has led to the development of numerous alternative imaging techniques exploiting a wide variety of contrast mechanisms and in some cases surpassing even the atomic resolution. An overview of the most widespread high-resolution imaging methods is presented at the beginning of Chapter 2. The apertureless or scattering-type scanning near-field optical microscopy (s-SNOM) employed in this work is a nondestructive optical imaging technique which circumvents the diffraction barrier with the aid of a sharp probing tip placed in the immediate proximity of the specimen surface. The tip is illuminating by a focused laser beam and the light scattered by the tip is collected and recorded. The amplitude and phase of the scattered light depend on the near-field interaction between the probe and the sample, thereby enabling an optical map of the sample to be constructed by scanning the s-SNOM probe along the specimen surface. The highest attainable resolution is determined only by the sharpness of the probing tip and is thus independent of the light wavelength. A more detailed description of this technique is provided in Chapter 2, including the typical experimental setup and a simple theoretical model which qualitatively explains the signal and contrast generation in s-SNOM.

In addition to high-resolution optical microscopy, s-SNOM can also be applied for spectroscopy if the sample is repeatedly scanned while sequentially changing the illumination wavelength. This is of particular advantage in the infrared (IR) wavelength range where the vibrational spectral fingerprints enable chemical identification of compounds present in the sample. The conventional far-field IR spectroscopy thereby suffers from the poor resolution, on the order of several micrometers. The combination of s-SNOM and IR spectroscopy therefore provides the means to obtain both high chemical sensitivity and high spatial resolution. Furthermore, particularly sharp and strong resonances were discovered in the s-SNOM spectra of samples supporting quasi-particles known as surface polaritons. The physics behind surface polaritons and an overview of their properties and applications is given in Chapter 3.

By comparing the measured near-field spectra of polariton-resonant materials with predictions of the simple point dipole model presented in Chapter 2, a quantitative disagreement between the experiment and the model is found regarding both the resonance position and its strength. This discrepancy has been identified as the primary factor limiting the material identification capabilities of the IR s-SNOM. Chapter 5 is for this reason devoted to the derivation of a new and improved model capable of s-SNOM signal prediction on a quantitative level. The new ”monopole” model, first derived as a part of this thesis, departs from the commonly employed reduction of the s-SNOM probing tip to a point dipole and instead represents it by a prolate spheroid in uniform electric field. It is further demonstrated that only a small part of this spheroid interacts with the sample, and that the interacting part can be approximated by a point charge (monopole). An analytical expression for the charge induced by the near-field interaction between the s-SNOM probe and the sample is finally derived. Comparisons to the experiment demonstrate that the monopole model represents a major improvement over the dipole model and succeeds in quantitative prediction of near-field material contrasts measured by s-SNOM.

Technically, the measurement of pure near-field contrasts in s-SNOM is known to be complicated by the large background scattering generated by the vibrating s-SNOM probe. The monopole model derived in Chapter 5 is extended in Chapter 6 to account for the background scattering. On this basis, the twofold background influence on the s-SNOM signal is revealed and quantified. One part of the background influence is additive with respect to the near-field signal, and is routinely suppressed by the standard higher-harmonic demodulation method. The other part of the background contribution is multiplicative with

respect to the near-field signal and a proof that it needs to be eliminated for a reliable near-field spectroscopy of surface-polariton resonant samples is presented here. For a complete multiplicative background elimination, a new interferometric detection method has been proposed and implemented in this thesis. Comparison of the s-SNOM signal detection techniques demonstrates the advantages of the new ”pseudo-heterodyne” over the other methods either in terms of background suppression power or in the ease of implementation and available spectral range.

Building on the improvements in s-SNOM signal modeling and detection presented in Chapters 5 and 6, the applicability of s-SNOM is extended to the measurement of structural contrasts in materials of the same chemical composition in Chap 7. In particular, the way to measure crystal quality degradation due to ion implantation or radiation damage is presented. It is shown that even minute variations in crystal structure such as those arising from the different stacking order of layers in a crystal (polytypism) can be detected in s-SNOM and reproduced by the monopole model. It is finally suggested that with the aid of the monopole model the nanometer-scale chemical and structural composition of a sample on the might be recovered even with no prior information on the possible sample composition.

2 Scattering-type Near-field Optical Microscopy

2.1 Introduction

Light is our primary source of information about the environment which surrounds us. This is true not only for accomplishing everyday tasks, but also in the domain of scientific image acquisition because optical imaging methods are simple and reliable, fast and inexpensive and also contactless and noninvasive. Despite all these advantages, alternative imaging techniques had to be developed because of the fundamental limit on how small details can be resolved using optical methods. The optical resolution barrier is a consequence of the wave-like nature of electromagnetic (EM) radiation which is for this reason susceptible to diffraction. Due to diffraction, the image of a point source of light produced by an optical microscope is not a point but rather a set of concentric bright and dark rings known as Airy pattern or Airy disk (Fig. 2.1).

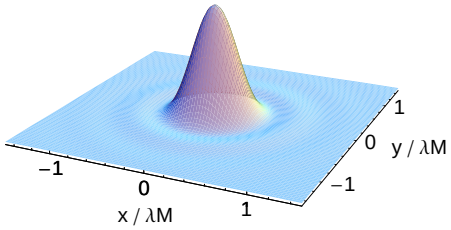


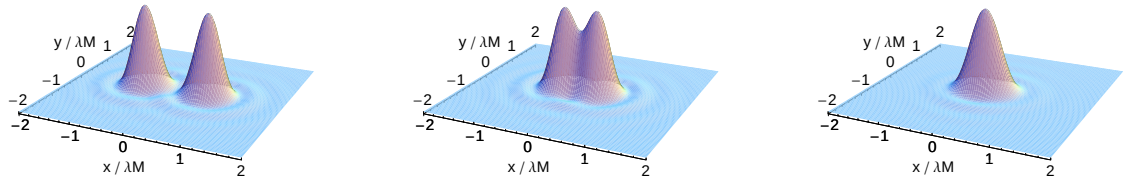
Figure 2.1: Airy pattern: Intensity distribution in the image plane of a point source viewed by an optical microscope with a numerical aperture $N_A = 1$ and magnification M.

Two closely separated point sources are usually considered just resolved when the central maximum of the first source's Airy pattern coincides with the first zero of the second source's Airy pattern, as shown in Fig. 2.2(b). This criterion, known as the Rayleigh criterion [1], was used by Ernst Abbe[2] to established the following quantitative relation

between the smallest distance Δx_{min} between two point source resolvable by an optical microscope and the operating light wavelength λ:

$$\Delta x_{min} = 0.61 \frac{\lambda}{n \sin\theta}, \tag{2.1}$$

where θ represents the half angle subtended by the cone of light captured by the microscope objective, and n is the refractive index of the medium between the sample and the objective. The product $n \sin\theta$ thereby defines the numerical aperture of the objective, denoted as N_A.



(a) (b) (c)

Figure 2.2: Image of two point-like sources of light obtained by an optical microscope with $N_A = 1$ and magnification M: (a) separated by $\Delta x = 1.2\lambda$ and easily resolved, (b) separated by $\Delta x = 0.6\lambda$ and just resolved according to the Rayleigh criterion, and (c) separated by $\Delta x = 0.3\lambda$ and not resolved.

It should be noted that the factor of 0.61 in Eq. 2.1 is a direct consequence of the Rayleigh criterion and other criteria might yield different results. For example, according to the Sparrow criterion [3], two point sources are resolved if a saddle exists along the line connecting the central maxima in their respective images, and the smallest resolvable distance equals $\Delta x = 0.47\,\lambda/n$ for $\theta = 90$. Even with the Rayleigh criterion, the resolution can be pushed to about the same value using oil immersion objectives for which $n \approx 1.3$. Hence it is customary to say that the resolution of optical microscopes is limited to about half the wavelength. Since the point spread function of the microscope is known, it might seem natural to attempt the deconvolution of a recorded image by post-processing it. However, this turns out to be an ill-posed inverse problem because significantly different distributions of sources can produce very similar intensity distributions in the images[4, 5].

Since there are numerous natural and artificial objects with overall sizes or substructures on the order of 100 nm and less, in many cases the resolution obtainable with visible light ($\lambda = 400..800$ nm) is not sufficient. Although ultraviolet (UV) light or even X-rays could be theoretically used to obtain much better resolution, there are several practical obstacles preventing this. Depending on the actual wavelength, it is difficult or even impossible

to fabricate lenses which are transparent to such light and also focus it with sufficient precision. Furthermore, the investigated samples may suffer from rapid degradation due to the ionization damage.

To circumvent these problems, alternative techniques have been developed. The most useful technique for obtaining high-resolution structure of crystalline material is based on the diffraction of X-rays. It provides atomically resolved maps of the electron density in crystals with a wide range of unit cell sizes. The unit cells may thereby contain between several atoms in simple inorganic crystals to over 10^3 atoms in large biological macromolecules. Even larger structures of up to 10^6 atoms can be studied by X-ray crystallography albeit the resolution is limited to about $0.4\,\mathrm{nm}$ in that case[6]. Limitations of the crystallography approach follow from the fact that many substances do not readily crystallize in their natural state and also that dynamic studies cannot be performed with matter in its crystallized state.

Whereas the crystallography only works with a huge number of identical units, it is possible to resolve single nanometer-sized objects by using electrons instead of photons for probing the sample[7, 8]. Unlike photons, electrons are charged particles and can be focused by magnetic lenses which are hollow and do not pose a problem for the transmission of particles. The de Broglie wavelength $\lambda = h/p$ of electrons is smaller than $0.1\,\mathrm{nm}$ already for electrons with kinetic energies on the order of $1\,\mathrm{keV}$, so that the attainable resolution is mostly limited by imperfections of magnetic lenses. High Resolution Transmission Electron Microscopy (HRTEM) can even achieve a true sub-atomic resolution in crystalline material by exploiting phase contrasts of electron waves exiting the sample. Although nowadays an indispensable tool for structural analysis in the materials and life sciences, transmission electron microscopy has to be used in high vacuum conditions and at low temperatures, making it unsuitable for monitoring naturally occurring dynamic processes. TEM is also notorious for the highly demanding sample preparation. Although the latter obstacle is avoided in Scanning Electron Microscopy (SEM, [9–11]), this comes at the expense of an order of magnitude lower resolution. In the case of both TEM and SEM, the exposure time is limited to relatively short periods because of the deteriorating influence of the electron beam on the sample.

Owing to the recent progress in fluorescence imaging, a completely non-invasive and non-destructive high-resolution imaging can be performed by purely optical means. There are two techniques that need to be mentioned in this context. One of them is known as the

STimulated Emission Depletion (STED, [12, 13]) microscopy. It uses two diffraction-limited pulses for illuminating the sample, the first of them being a Gaussian-shaped fluorescence excitation pulse. The excitation pulse is immediately followed by a second, ring-shaped (toroidal) depletion pulse tuned to an emission line of the fluorescent dye. After the second pulse depletes the excited energy levels of dye molecules by means of stimulated emission, the fluorophores remain in an excited states only in the center of the depletion pulse. The exact non-depleted spot size depends on the wavelength of light and the intensity of the depletion pulse compared to the intensity needed for saturating the depletion[14], and may be as small as 20-30 nm in diameter. By scanning this spot through the sample, a fluorescence map of the sample is obtained with the respective subwavelength resolution. The same effect can also be achieved by another method, known as the PhotoActivated Localization Microscopy (PALM, [15]). It relies on the photoactivation of only a small fraction of total fluorophore population at a time, such that each activated dye molecule is separated from all other activated dye molecules by at least one far-field resolvable distance Δx_{min}. The location of each active (fluorescing) dye molecule is then precisely determined by fitting the center position of its image to the theoretical point spread function (Airy pattern). After all activated fluorophores have been irreversibly bleached, the next generation of dye molecules is activated and the procedure is repeated until the complete fluorescence map of the sample has been obtained. The resolution which can be achieved this way is generally better than 25 nm, and theoretical considerations indicate it can be improved down to a few nanometers.

If the density of fluorophores is so low that they can all be directly resolved, their localization with extremely high precision may be performed even without resorting to photoactivable dyes [16, 17]. For highest localization and longest tracking times of fluorescent molecules, it is necessary to increase the number of photons they emit before bleaching to the largest values possible [18]. By depleting oxygen molecules from the sample, it is possible to register up to 10^6 photons instead of the usual $10^3 - 10^4$ photons per dye molecule and consequently enable the Fluorescence Imaging with One Nanometer Accuracy (FIONA, [19]). Even without special manipulation of the sample, it is possible to extract more information form the optical signal if the number of photons hitting each pixel in an image and their arrival times are both recorded [16, 20]. Alternatively, it is also possible to exploit other kinds of luminescence to gain spatial and structural information about the sample. Especially the electroluminescence and radioluminescence are

frequently employed in the semiconductor analysis and materials research, respectively.

Even more options exist if the distances between two or more sources are to be measured instead of their absolute positions. The suitable methods include Förster Resonance Energy Transfer (FRET, [21]) between fluorescent molecules and shift of the plasmon resonance due to interaction of closely spaced metallic nanoparticles [22]. The former works best over short distances of up to 10 nm, whereas the latter is intended for longer distances up to 70 nm. Finally, if the motion of an object needs to be tracked without knowing its absolute position, interferometric methods can provide a resolution below 1 nm even with non-labeled samples. In that sense, the Differential Interference Contrast (DIC) microscopy is sensitive to displacements on the order of $1\text{pm}\,\text{Hz}^{-1/2}$[23].

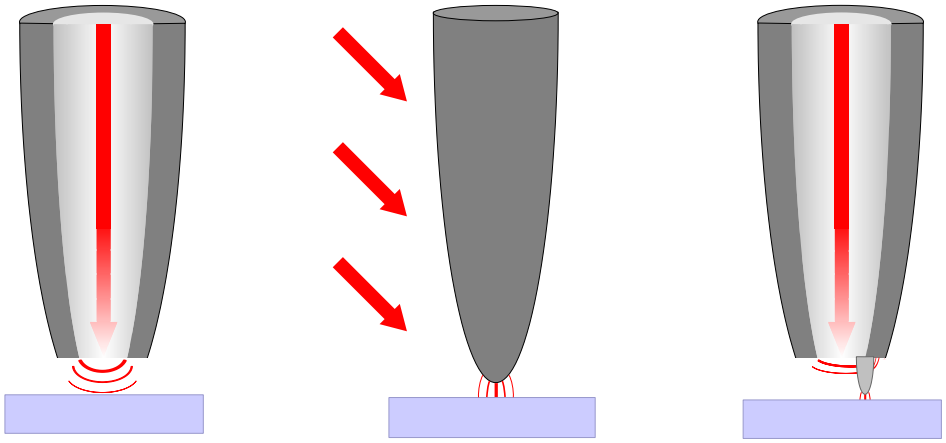
2.2 Scanning Near-field Optical Microscopy

There exists an entire family of high resolution imaging techniques known as Scanning Probe Microscopy (SPM) which has not yet been mentioned in this introduction. Different SPM variants cover a wide spectrum of contrast mechanisms, including mechanical (AFM [24]), electric (STM [25], KPFM [26], EFM [27], SCM [28]), magnetic (MFM [29, 30], MRFM [31, 32]) and optical (SNOM [33, 34], PSTM [35]) phenomena, with or without labeled specimens. SPM techniques routinely reach resolutions in the nanometer range and some are even capable of resolving details on the atomic scale. The common characteristic of all SPM variants is the use of sharp probing tips to scan the surface of the specimen. Either the probe or the specimen are mechanically moved from pixel to pixel, line by line, to obtain a raster image of the specimen by recording the probe-sample interaction as a function of their relative position. Some general shortcomings of the SPM techniques follow from this kind of image acquisition. In particular, the probing is limited to the surface of the sample, and the relatively long image acquisition times are necessary to complete the scanning process.

Historically, the SPM was founded with the invention of the Scanning Tunneling Microscope (STM [25, 36, 37]) in 1981 by G. Binnig and H. Rohrer, who solved the problem of achieving and maintaining very precise control over the distance between the probing tip and the sample surface despite all external disturbances like noise and vibration. The STM measures the current of electrons tunneling between the probe and the sample surface, which makes it applicable only to conducting and semiconducting materials. The

Atomic Force Microscope (AFM, [24]) was developed shortly afterward to enable imaging of the various samples independent of their conductivity. In the AFM, the probing tip is attached to a small cantilever which deflects in response to the variations in the interaction force between the probing tip and the surface of the specimen. With very sharp probing tips, even sub-atomic resolution can be achieved[38] and different kinds of forces can be probed depending on the way the probing tip is coated or functionalized.

The SPM principles can be extended to optical probing as well. The basic idea actually predates the development of STM and is attributed to Synge who already in 1928 proposed a method for high resolution optical imaging by opening a small aperture in an otherwise opaque screen. The aperture would be held in the immediate proximity of the specimen, thereby preventing the effects of diffraction to escalate before the transmitted light reaches the specimen. This idea was first implemented in practice by Ash and Nicholls in 1972 [39] using microwaves of about 3 cm wavelength. It took another 12 years before the first near-field optical microscope was constructed on that principle independently by D. Pohl [34] and by A. Lewis [33] and their coworkers. Today, the most commonly used aperture probes are metal-coated single-mode optical fibers with a small opening at their end [40], as illustrated in Fig. 2.3(a). The resolution improves as the aperture is made smaller, but this simultaneously leads to a rapid decrease in the transmission efficiency. The resolution of aperture SNOM is thus in practice limited to about $\lambda/10$.



(a) (b) (c)

Figure 2.3: Comparison of SNOM variants: (a) aperture, (b) apertureless, and (c) tip-on-aperture.

An alternative method that provides completely wavelength-independent resolution was first published in 1985 by Wessel who suggested to exploit the enhanced field in the vicin-

ity of a small particle for illuminating the sample [41]. The first near-field optical microscopes based on this principle were built in 1994 by Zenhausern et al. [42] and by Inouye and Kawata [43]. They used metalized AFM probes instead of the originally proposed small particles to create an optical "nano-focus", illustrated in Fig. 2.3(b). The size of the nano-focus depends only on the tip apex radius, thus providing the same resolution with the visible and infrared light, in both cases on the order of 10 nm [44–46] and below [47]. The imaging mechanism in the apertureless SNOM is based on the optical interaction between the probe and the specimen, mediated by the evanescent fields. The interaction modifies the amplitude and/or phase of light waves scattered by the probe and subsequently detected by far-field methods. Because the scattered light represents the measured quantity in apertureless SNOM, this SNOM variant is also called the scattering-type SNOM (s-SNOM).

Among possible scattering channels, elastic scattering is largely predominant. However, Raman scattering can also appear [48] and can be extracted from the elastic scattering background by filtering. The same can be done with light of doubled frequency which is generated by virtue of second harmonic generation if the field intensity at the tip apex is strong enough [49, 50]. Single- and two-photon fluorescence excitation by s-SNOM probes have also been reported [51, 52]. It should be noted that unlike most other scanning probe techniques, the near-field optical microscopy can also detect objects located at some distance *below* the sample surface [53, 54]. Finally, being built upon the AFM, the s-SNOM can also provide the topography of the sample surface simultaneously with the optical image.

Besides all the advantages of the s-SNOM, there exists an obstacle for its application in optical imaging due to the already mentioned large background scattering. The background scattering is not related to the probe-sample near-field interaction but can nevertheless produce contrasts in s-SNOM images, thereby falsifying them. For this reason, a large portion of development efforts has been devoted to the suppression of the background [47, 55–58] ever since the first experimental realizations of the s-SNOM. This thesis is no exception to the said trend and includes an entire chapter (Chapter 6) devoted to the theoretical modeling and experimental suppression of the background scattering.

It is also worth noting that a hybrid between the aperture and apertureless SNOM has been recently introduced by H. Frey et al.[59, 60]. The probes used in this method are known as tip-on-aperture (TOA) probes, illustrated in Fig. 2.3(c). With the TOA, the evanescent field generated by the aperture is additionally enhanced and confined by the

small tip. The presence of the small tip thus addresses both main problems of aperture SNOM, namely the poor topographic resolution and the transmission-resolution tradeoff. Nevertheless, a widespread usage of TOA probes is still hindered by the very demanding fabrication procedure. Furthermore, the aperture limits the range of wavelengths the TOA probes transmit sufficiently well and renders them unsuitable for operation with IR and longer wavelengths.

2.3 Scattering-type Near-field Optical Microscope

Based on the presented overview of imaging techniques, the apertureless or scattering-type scanning near-field microscopy (s-SNOM) can be identified as the only method offering nanometer-scale optical resolution in the mid-infrared spectral range. The investigation of optical properties in the mid-IR spectral range is in turn motivated by the fact that mid-IR photon energies correspond to the energies of vibrational modes of chemical bonds between atoms in molecules and crystals. The vibrational energy spectra are highly specific to each kind of molecule or crystal and are thus often regarded as ”fingerprints” of a material’s chemical composition and structure. The sensitivity of IR s-SNOM (also named the scattering-type Scanning Near-field Infrared Microscope, s-SNIM [61]) to chemical composition has already been demonstrated several times [44, 46, 62–65]. One of the topics elaborated in this thesis is the extension of the s-SNOM application range from analyzing the chemical composition to the investigation of the crystal structure [66, 67], presented in Chapter 7.

As a reference for the further analysis of the s-SNOM principles and applications, the anatomy of the scattering type near-field optical microscope used in this work is presented in Fig. 2.4. The description of its functional units is provided in the following sections. More details about this setup can also be found in [68].

2.3.1 Atomic Force Microscope

The atomic force microscope (AFM) shown on the left-hand side of Fig. 2.4 represents the core upon which the s-SNOM is built. It consists of a piezoelectric scanner capable of positioning the sample with nanometer precision in the x- and y-directions and with angstrom precision in the z-direction.

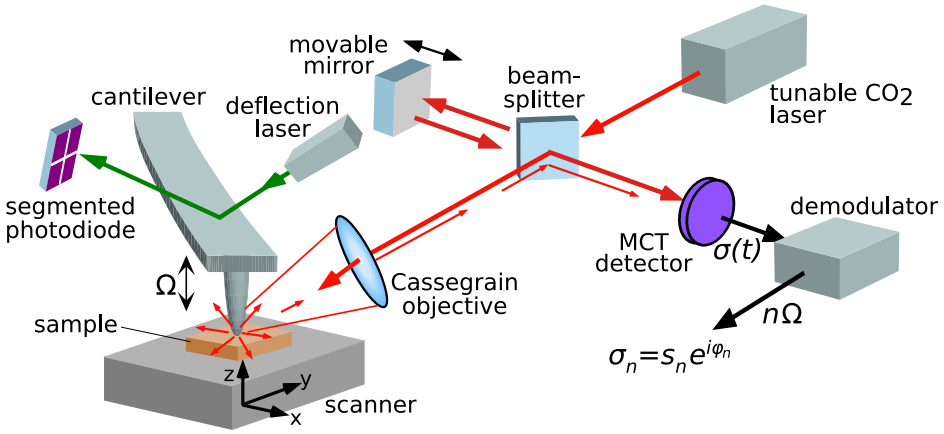


Figure 2.4: Schematic representation of the s-SNOM setup. Adapted from [45].

The probing tip is affixed to the bottom side of the cantilever. The cantilever is driven to oscillation in z-direction at its resonant frequency Ω, with an amplitude of about 20 to 50 nm. The probe and the sample are maintained at such a distance that the probe touches the sample surface once per oscillation period. This mode of AFM operation is known as the ”tapping” or intermittent contact mode.

The amplitude of the cantilever vibration is monitored by reflecting a focused laser beam from the top side of the cantilever and detecting it by a four-segment photodiode which is sensitive to the beam deflection from its central position. The measured vibration amplitude is used to regulate the tip-sample distance in a closed loop that moves the sample away from the tip if the vibration amplitude decreases below a preselected value (set point), and the other way around. By recording the vertical position of the sample as a function of its lateral position, a three-dimensional relief or topography of the sample is acquired.

2.3.2 Probe Illumination

What distinguishes a s-SNOM from an ordinary AFM is the presence of an additional high-power laser which is focused onto the probing tip. The light from this laser polarizes the probing tip which responds by emitting scattered radiation in all directions. The light scattered back along the illumination path is collected and recorded as the optical signal. It is important to note that the optical near-field interaction between the probe modifies the amplitude and phase of the scattered light and thus enables an optical map of the sample to be constructed.

For the experiments described in this thesis, a CO_2 laser tunable in the $9-11$ µm wave-

length range was employed as the light source. Reflective optical elements were used to steer, focus and collect the light in order for the same path to be usable both with invisible IR beam and with auxiliary visible beam which served as the aid for adjusting the optical components.

2.3.3 Detection of Scattered Light

There are several known methods to detect the light scattered by a s-SNOM probe. A detailed analysis and comparison of these methods is presented in Chapter 6. Measurements presented in this work were carried out using an interferometric setup shown in Fig. 2.4. It is based on a Michelson interferometer whose reference mirror is either moved between two positions $\lambda/8$ apart in a stepwise fashion (Sect. 6.3.3), or continuously vibrated with an amplitude equal to 0.21λ (Sect. 6.4.2). In both cases the use of an interferometer serves to coherently amplify the signal and, more importantly, to obtain the information about the scattered signal amplitude and phase.

However, the useful information about the sample is not directly obtained form the total light intensity measured by the detector. Rather, the demodulation of the detector output voltage at an integer multiple of the probe vibration frequency (higher harmonic) is necessary for obtaining reproducible high-resolution optical map of the sample. The reason for this is explained shortly.

2.3.4 Demodulation of the Detector Output

It was already mentioned in the introduction that there are two contributions to the light scattered by the probing tip. One of them depends on the near-field interaction between probe and the sample and represents the useful signal, also known as the ”near-field signal”. The other part consists of the reflections from optical elements and the direct scattering by the probing tip, both completely independent of the near-field interaction. This part is not useful in optical imaging and represents the unwanted s-SNOM background signal. It has been observed that the light scattered by a probing tip contains much larger proportion of the background than the near-field signal [69].

Fortunately, the background varies only slightly with the change of the tip-sample distance as long as it remains on the order of the tip radius. Over the same distances, the evanescent fields which mediate the near-field interaction vanish almost completely, as described in

Sect. 4. It is therefore clear that for slight displacements of the probe just above the sample surface, the change in the scattered field will be mostly caused by the change in the near-field interaction strength, rather than the background scattering. The small movements of the probing tip required to separate the near-field from the background signal are usually obtained by vibrating the probing tip in the intermittent contact with the sample surface. If we denote the tip vibration frequency by Ω, the periodic detector output signal $u(t)$ can be expressed in the form of the Fourier series

$$u(t) = \sum_{n=\infty}^{\infty} u_n e^{in\Omega t} \tag{2.2}$$

with $u_{-n} = u_n^*$ since the function $u(t)$ is always real. Only the first few terms in this series are usually sufficient to account for the slow change in the background signal. It was experimentally determined that already the first two terms ($|n| = 0, 1$) at infrared frequencies and three terms ($|n| = 0..2$) at visible frequencies suffice[69]. Further terms ($|n| \geq 2$, i.e. $|n| \geq 3$) are needed to represent the rapidly changing near-field signal. It is therefore exactly those higher harmonics that should be recorded in order to extract the pure near-field contribution to the scattering signal.

Mathematically, the coefficients u_n in the Fourier series from Eq. 2.2 are obtained from the measured detector output voltage $u(t)$ as

$$u_n = \frac{1}{T} \int_{-T/2}^{T/2} u(t)\, e^{-in\Omega t}\, dt. \tag{2.3}$$

In practice, a lock-in amplifier can be used to extract single coefficients u_n. Alternatively, a data acquisition system capable of performing Fast Fourier Transform (FFT) can be employed to calculate several coefficients u_n in parallel.

The exact detector output voltage in each particular experiment depends on several parameters beyond the near-field interaction. These parameters include the detector responsivity, illumination intensity and incidence angle, numerical aperture of the objective used for light focusing and collection, the shape of the probing tip, and so on. To be able to compare the results of the measurements made in different experimental conditions, it is thus customary to consider only the *relative contrasts* between different regions in the optical map of the sample.

2.4 Dipole Model of Probe-Sample Interaction

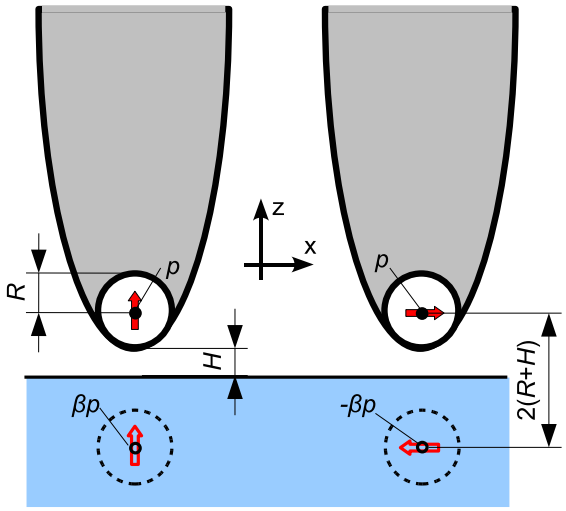


Figure 2.5: Point dipole model of the probe-sample interaction for perpendicular (left) and parallel (right) orientations of the probe dipole with respect to the sample surface.

There exists a simple model for predicting relative s-SNOM contrasts between different materials. It is known as the (point) dipole model, first developed in the context of Raman scattering [70, 71] and later reused to describe the probe-sample interaction in the apertureless near-field optical microscopy [57, 72, 73].

Within the point dipole framework, the probing tip is represented by a point dipole assuming the position and polarizability of a sphere of radius R inscribed into the probing tip apex as depicted in Fig. 2.5. The dipole polarizability consequently equals $\alpha = 4\pi R^3(\varepsilon - 1)/(\varepsilon + 2)$ and the dipole moment induced by the external illumination is $p = \alpha E_0$, with E_0 being the strength of the illumination electric field.

The second constituent of the point dipole model is the ”mirror image” of the probe dipole, formed in the sample. For the tip dipole p normal to the sample surface (Fig. 2.5, left), the mirror dipole has the strength $p' = \beta p$, with the ”reflection coefficient” $\beta = (\varepsilon_s - 1)/(\varepsilon_s + 1)$ and ε_s the dielectric constant of the sample. The mirror dipole acts back on the probe, thereby polarizing it even stronger. This additional probing dipole moment increases the strength of the mirror dipole, which in turn induces further polarization of the tip, and so on.

Mathematically, such an interaction can be expressed in the form of an infinite geometric

series

$$p = p_0 \sum_{n=0}^{\infty} g^n, \tag{2.4}$$

with p_0 being the initial dipole moment. The factor $g = \Delta p/p$ represents the relative increase in the probe dipole moment p after a single reflection from the sample. When the infinite sum in Eq. 2.4 is evaluated, the dipole moment p represents the total dipole moment of the tip as a result of its interaction with the sample.

An alternative way to obtain the dipole moment p relies on the self-consistency condition. We start by noting that the mirror dipole strength p' includes the initial probe dipole moment p_0 and the interaction-induced part p_i. Since p_i is induced by p', we get the following system of equations:

$$\begin{aligned} p' &= \beta\,(p_0 + p_i), \\ p_i &= f\,p', \end{aligned} \tag{2.5}$$

where f is a so far undetermined function of the distance between the tip and the sample. Inserting p' from the upper equation 2.5 into the lower one, we obtain $p_i = \beta\, f\,(p_0 + p_i)$. Substituting $p_i = p - p_0$ and $\beta f = g$, this can be expressed as

$$p = p_0 + g\,p. \tag{2.6}$$

Equations 2.4 and 2.6 describe the same physical situation so their solutions must be equal. And indeed, in both cases we obtain

$$p = \frac{p_0}{1-g}. \tag{2.7}$$

The function f contained in $g = \beta f$ should be determined at this point for a complete solution. Its value depends on the orientation of the dipole, as will be shown next.

Let us first consider the dipole oriented perpendicularly to the sample surface (i.e. parallel to the z-axis), shown in Fig. 2.5 left. Following the coordinate system choice of Fig. 2.5, this orientation can also be named the z-orientation. If the distance between the sphere representing the tip and the sample is H, the separation between the point dipole and its mirror image is $D = 2(R+H)$. The field produced by the mirror dipole p' at the position

of the tip dipole p is equal to

$$E_z = \frac{p'}{2\pi D^3} \tag{2.8}$$

The field E_z induces a dipole moment $\Delta p = \alpha E_z$ in the probe. The function f is thus $f_z = \Delta p/p' = \alpha/2\pi D^3$. Expressing the distance D in terms of the tip-sample separation H, the factor g in Eq. 2.7 becomes $g_z = \beta\, f_z = \beta\alpha/16\pi(R+H)^3$. We can thus assemble the final solution for the vertical tip dipole moment

$$p_z = E_0 \frac{\alpha}{1 - \frac{\alpha\beta}{16\pi(R+H)^3}}. \tag{2.9}$$

Instead of the dipole moments, it is customary to use the effective polarizability $\alpha_{\text{eff}} = p/E_0$. The value of $\alpha_{\text{eff,z}}$ is obvious from Eq. 2.9:

$$\alpha_{\text{eff,z}} = \frac{\alpha}{1 - \frac{\alpha\beta}{16\pi(R+H)^3}}. \tag{2.10}$$

For a dipole oriented parallel to the sample surface (x-axis in Fig. 2.5), its mirror image exhibits an anti-parallel orientation, i.e. $p' = -\beta p$. The mirror dipole field at the tip dipole position has also got a negative sign and equals

$$E_x = -\frac{p'}{4\pi D^3}. \tag{2.11}$$

This results in $q_x = \beta\, f_x = \beta\alpha/32\pi(R+H)^3$ and the final expression analogous to Eq. 2.10 is then easily obtained:

$$\alpha_{\text{eff,x}} = \frac{\alpha}{1 - \frac{\alpha\beta}{32\pi(R+H)^3}}. \tag{2.12}$$

Due to the linearity of the system there is no cross-talk between the parallel and the perpendicular dipoles, so that any dipole orientation can be decomposed into the *z*- and *x*-orientations. However, this is rarely needed if the true tip response is to be mimicked because an elongated probing tip has a high preference for the field along the z-axis, making it the only component worth considering. An exception may be the illumination containing only the x-component of the field (e.g. s-polarized light), which only probes $\alpha_{\text{eff,x}}$. But in this case the signal level would be very low due to the missing field enhancement by the probe.

If a dipole is driven by an oscillating field, it will respond by an oscillating dipole moment.

Far away from the dipole, its radiated field approaches the form of a spherical wave. The field oscillation amplitude is thereby directly proportional to the dipole moment p and thus to α_{eff} as well. The knowledge of the effective polarizability of the tip thus is sufficient to predict the *relative* signal strength detected in the far field. It is thereby customary to define the scattering coefficient $\sigma = s e^{i\varphi}$ as the ratio of the incident field E_0 and the scattered field E at some point in space:

$$E = \sigma E_0. \tag{2.13}$$

Since σ defined this way is proportional to the effective polarizability α_{eff}, the following identity holds for any two materials 1 and 2 whose response is measured in the same setup and under the same conditions:

$$\alpha_{\text{eff},1} : \alpha_{\text{eff},2} = \sigma_1 : \sigma_2. \tag{2.14}$$

In other words, the quantities σ and α_{eff} can be used interchangeably as long as only relative contrasts are measured.

Finally, it should be noted that the point dipole model was originally introduced in a slightly different form than used in some later publications [46, 74, 75] and presented here. The original version included the radiation of the mirror dipole to give [57]

$$\alpha_{\text{eff}} = \frac{\alpha(1+\beta)}{1 - \frac{\alpha\beta}{16\pi(R+H)^3}}. \tag{2.15}$$

This approach was followed in numerous later publications. A modification of the original expression (Eq. 2.15) was also suggested, with an additional factor $(1+r_p)$ accounting for the indirect illumination of the tip via the reflection from the surface[76]. Thereby r_p stands for the Fresnel reflection coefficient of the p-polarized light. In Sect. 5.2 it will be shown that, strictly speaking, none of the above variants is completely correct, and the proper way to handle the mirror dipole radiation will be derived.

Finally, it should be noted that due to its simplicity, the dipole model lends itself well to extensions. In particular, an analytical solution has been obtained for systems consisting of up to four mutually interacting dipoles [75].

2.5 Scattering-type Near-field Optical Spectroscopy

The full power of infrared near-field microscopy is revealed only in conjunction with spectroscopy. Infrared nano-spectroscopy resulting from this combination can exploit vibrational fingerprints of molecules and crystals to provide chemical and structural analysis of specimens with the resolution on the order of 10 nm [44, 64, 67]. However, the current s-SNOM state of the art permits imaging only at one fixed wavelength at a time. Although indications exist that true parallel near-field spectroscopy may be possible [77], the development has not yet matured to the point where it could be used for routine measurements. For the time being, the near-field spectroscopy has to be performed sequentially, following the procedure outlined below.

2.5.1 Construction of Near-field Spectra

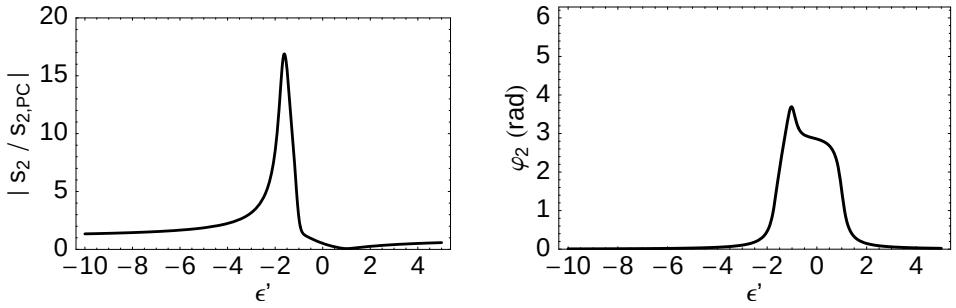
To construct a complete near-field spectrum of a specimen, the scattering signal resulting from the near-field interaction between the tip and the sample has to be recorded at a number of wavelengths. For a successful material identification it is also necessary that the said wavelengths cover the spectral region where the vibrational modes of the investigated molecule or crystal can be excited. Once a sufficiently broad spectral region has been covered and the desired density of wavelengths has been reached, the near-field spectrum can in principle be assembled from those measurements.

However, the scattering signals recorded at different wavelengths cannot be directly compared to each other because of the varying laser power, beam alignment and phase, detector sensitivity and probing tip radiation pattern, which can all change with the wavelength. To overcome this problem, the measured material response at each wavelength must be normalized to the response of a reference material which is known to exhibit no spectral variation over the wavelengths of interest. In the mid-IR, the most commonly used reference materials are gold and undoped silicon. It is thereby important to measure the signal from the reference material under exactly the same conditions as the signal from the investigated material. In practice, this usually means that both the signal from the investigated material and the reference material should be acquired in a single scan. To facilitate this, a thin (30-60 nm) Au film is often evaporated on a part of the sample surface and the scan region is chosen such that it includes a portion of the surface covered by the Au film.

The normalized signal may finally be assembled into a near-field spectrum which can be

compared with other experiments or theoretical predictions.

2.5.2 Phonon-enhanced Near-field Interaction



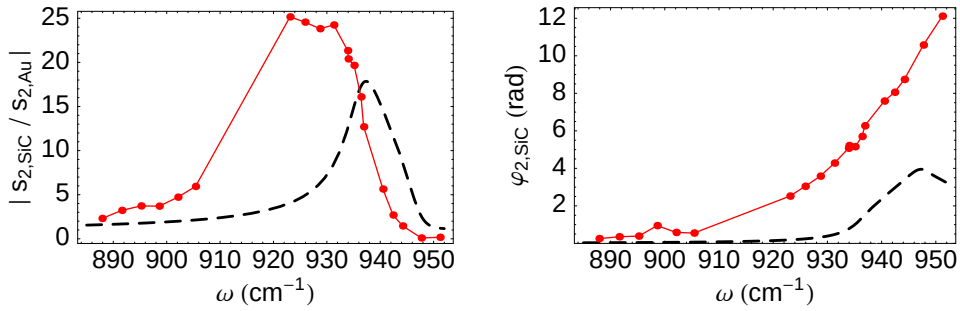
(a) (b)

Figure 2.6: Dipole model prediction for the amplitude (a) and phase (b) of the s-SNOM signal vs. real part ε' of the sample dielectric function. The imaginary part of ε is set to $\varepsilon'' = 0.2$ and the tip radius and vibration amplitude are set to $R = 35\,\text{nm}$ and $A = 25\,\text{nm}$, respectively. The signal is normalized to the response of a perfect electric conductor.

One intriguing theoretical prediction of the dipole model is the existence of a near-field resonance where the real part ε' of the sample dielectric function $\varepsilon = \varepsilon' + i\varepsilon''$ assumes values close to $\varepsilon' = -2$, accompanied by low values of the imaginary part ε''. This resonance is clearly seen in Fig. 2.6(a) displaying the predicted effective polarizability $\alpha_{\text{eff,z}}$ of the coupled probe-sample system as a function of ε' in the overall dielectric function $\varepsilon = \varepsilon' + 0.2i$.

Such near-field resonances indeed exist and have been first observed using crystalline silicon carbide (SiC) as the sample and infrared light with wavelengths between 9.2 µm and 11.2 µm [63]. The near-field spectrum of SiC typically obtained by s-SNOM within the aforementioned wavelength range is shown in Fig. 2.7(b). The observed phenomenon is attributed in [63] to the phonon-enhanced near-field interaction since it relies on resonant excitation of the crystal lattice vibrations in the sample. The physics behind this effect is elaborated in more detail in Ref. [68] and also in Chapter 3 of this thesis.

Since the position and the height of the spectral peak related to the near-field resonance depend on the energies of vibrational modes in the examined crystal, they are highly material-specific, just like the far-field IR spectra. This provides the opportunity for infrared spectral fingerprinting with ultra high resolution, better than $\lambda/500$ [66, 67]. Unfortunately, by



(a) (b)

Figure 2.7: Second-harmonic (a) amplitude and (b) phase spectra of a crystalline SiC sample normalized to Au. Shown are experimentally obtained data (points, full line) and the prediction by the dipole model (dashed line), obtained with the SiC optical constants from [78]. The probing tip radius was $R = 35\,\text{nm}$ and the vibration amplitude $A = 25\,\text{nm}$.

comparing the experimental spectra and theoretical predictions in Fig. 2.7 (a) and (b), we see that although there is a *qualitative* agreement between them, the agreement is *quantitatively* not very good. Unambiguous material identification from the near-field spectra thus requires an improvement in the theoretical description of the probe-sample interaction and motivates the development of a new model in Chapter 5.

3 Polaritons

Sect. 2.4 introduced the dipole model of the probe-sample interaction in the scattering-type near-field optical microscopy. This model represents the s-SNOM probe by a point dipole to which the sample responds by a redistribution of charge close to the surface. Above the surface, the electric field due to the polarization of the sample is equivalent to the field that would be produced by a "mirror image" of the probe dipole below the sample surface (Fig. 2.5). In the quasi-electrostatic limit, the ratio of the image dipole moment p' to the probe dipole moment p is equal to

$$\beta = \frac{p'}{p} = \frac{\varepsilon_s - 1}{\varepsilon_s + 1}. \tag{3.1}$$

An interesting property of the electrostatic "reflection coefficient" β is its ability to assume absolute values significantly larger than 1 under special circumstances. The denominator in Eq. 3.1 indicates that this happens when the dielectric function $\varepsilon_s = \varepsilon_s' + i\,\varepsilon_s''$ is close to $\varepsilon_s' = -1$ simultaneously with $\varepsilon'' < 1$. However, such values of the dielectric function are not encountered in electrostatics at all. They are not very common in electrodynamics either, since dielectrics are characterized by $\varepsilon_s' \geq 1$, and metals have largely negative ε_s' and high values of ε_s''. Nevertheless, many materials, including all metals and semiconductors as well as some insulators are characterized by at least one narrow material-specific range of wavelengths with the dielectric function close to $\varepsilon_s \approx -1$. As we have seen in Sect. 2.5, such situations are of special importance in the near-field microscopy because they give rise to near-field resonances. The attention will be now briefly devoted to the analysis of physical mechanisms responsible for the negative values of the dielectric function, first from the classical and then from the quantum-mechanical point of view.

3.1 Lorentz Oscillator Model

In a very simple classical approximation, the sample is treated as a collection of uncoupled (mechanical) harmonic oscillators each with mass m and force constant k. The oscillators couple to the electric field through their dipole moment p_x, related to the oscillator elongation x as $p_x = ex$. This approximation is known as the Lorentz oscillator model and its short outline will be presented here. For a more detailed treatment, standard textbooks such as [79] or [80] can be consulted.

If the Lorentz oscillators are driven by the external electric field of the form $E_0 e^{-i\omega t}$, their equation of motion reads

$$m\ddot{x} + \gamma m\dot{x} + kx = eE_0 e^{-i\omega t}, \tag{3.2}$$

where a phenomenological damping term $\gamma m\dot{x}$ has been additionally included. The amplitude of the steady-state solution of Eq. 3.2 is given by

$$x_0 = \frac{eE_0/m}{\omega_0^2 - \omega^2 - i\omega\gamma}, \tag{3.3}$$

where $\omega_0 = \sqrt{k/m}$.

From the dipole moment amplitude $p_0 = ex_0$ we obtain the polarizability

$$\alpha = \frac{p_0}{E_0} = \frac{e^2/m}{\omega_0^2 - \omega^2 - i\omega\gamma}. \tag{3.4}$$

A regular three-dimensional array of Lorentz oscillators with a volume density N has the polarization density equal to

$$\mathbf{P} = N\alpha\mathbf{E_0} = \frac{Ne^2/m}{\omega_0^2 - \omega^2 - i\omega\gamma}\mathbf{E_0}. \tag{3.5}$$

From Eq. 3.5, the susceptibility $\chi(\omega)$=$\mathbf{P}/\mathbf{E_0}$ can be directly extracted:

$$\chi(\omega) = \frac{Ne^2/m}{\omega_0^2 - \omega^2 - i\omega\gamma}, \tag{3.6}$$

and therewith the dielectric function $\varepsilon(\omega) = \chi(\omega) + 1$, which can also be written in the form

$$\varepsilon(\omega) = \frac{f}{\omega_0^2 - \omega^2 - i\omega\gamma} + 1 \tag{3.7}$$

with $f = Ne^2/m$ being the oscillator strength. However, a single oscillator experiences not only the external driving field, but also the field due to its neighbors. This effects leads to the following relation between the dielectric function $\varepsilon(\omega)$ and the polarizability $\alpha(\omega)$:

$$\frac{\varepsilon(\omega)-1}{\varepsilon(\omega)+2} = \frac{N\alpha(\omega)}{3\varepsilon_0}, \tag{3.8}$$

also known as the Clausius-Mossotti equation.

In the case of weak damping, the dielectric function obtained from 3.8 can still be put in the form of Eq. 3.7 by shifting the eigenfrequency ω_0 to $\omega' = \omega_0 - f/3$.

Finally, it should be noted that more than one type of oscillators can exist in a given material. The dielectric function is then obtained by summing the susceptibility over all resonances:

$$\varepsilon(\omega) = 1 + \sum_j \frac{f_j}{\omega_j^2 - \omega^2 - i\,\omega\,\gamma_j}, \tag{3.9}$$

where ω_j, f_j and γ_j are the natural frequency, oscillator strength and damping coefficient pertaining to j-th resonance, respectively.

If some resonance frequency ω_0 is well separated from all other ω_j, $j \neq 0$, the dielectric function in the vicinity of ω_0 can be described in a very simple way. To this end, the influence of all resonances below ω_0 ($\omega_j \ll \omega_0$) can be neglected, and the contribution of resonances above ω_0 ($\omega_j \gg \omega_0$) can be collected into the single term

$$\varepsilon_\infty = 1 + \sum_{\omega_j > \omega_0} \frac{f_j}{\omega_j^2}. \tag{3.10}$$

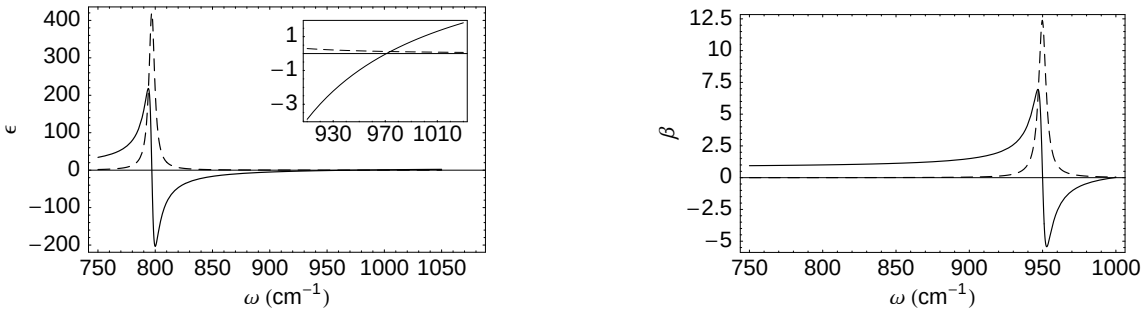
The dielectric function around ω_0 is then given by

$$\varepsilon(\omega) = \varepsilon_\infty + \frac{f_0}{\omega_0^2 - \omega^2 - i\,\omega\,\gamma_0}. \tag{3.11}$$

Eq. 3.11 will be used in the following to determine when the reflection factor β in Eq. 3.1 is expected to peak.

A typical behavior of the dielectric function ε and the electrostatic reflection factor β is shown in Fig. 3.1. Both quantities were calculated according to Eq. 3.11 with the following parameters: $\varepsilon_\infty = 6.56$, $f_0 = 2\times10^6\,\mathrm{cm}^{-2}$, $\omega_0 = 797\,\mathrm{cm}^{-1}$ and $\gamma_0 = 6\,\mathrm{cm}^{-1}$, which provide an approximate fit to the SiC crystal dielectric function.

From Fig. 3.1 we see that there are two frequencies where the real part of the dielectric



(a) (b)

Figure 3.1: Real (full line) and imaginary part (dashed line) of (a) the SiC dielectric function ε_s, and (b) the quasi-electrostatic reflection coefficient $\beta = (\varepsilon_s - 1)/(\varepsilon_s + 1)$ as a function of the frequency ω.

function assumes the value $\varepsilon_s' \approx -1$. The lower of them lies just above the resonance frequency ω_0. However, at this frequency the phase between the driving field $\mathbf{E_0}$ and the material polarization $\mathbf{P}$ is around 90, implying that the velocity (or current) is in phase with the driving force (i.e. electric field). In such a situation the energy is efficiently dissipated by the oscillator, resulting in a large imaginary part ε_s'' of the dielectric function. Due to the large ε_s'', the reflection coefficient β given by Eq. 3.1 does not exhibit a maximum at this point. Instead, the maximum value of β is reached at the second frequency for which $\varepsilon_s' = -1$, located just below the end of the negative-ε_s' range. The polarization lags behind the driving field by almost 180, so the absorption is low at that point. The low value of ε_s'' permits the reflection factor β to assume large values, limited only be the damping coefficient γ_0.

3.2 Bulk Polaritons

The Lorentz oscillator model used so far represents a purely classical description of the material response to the EM field. A more accurate quantum-mechanical treatment of the light-matter interaction yields the result in the same form as the Eq. 3.7[81]. However, the oscillator strength f has to be calculated in a different way because the quantum-mechanical description of the system departs substantially from the Lorentz oscillator model. In particular, the classical oscillator is replaced by a quantum system capable of making transition between two states j and l whose respective energy levels are separated by $\Delta E = \hbar\omega_0$. The overall oscillator strength f resulting from N such uncoupled oscillators

per unit volume is then given by

$$f = \frac{2N\omega_0}{\varepsilon_0 \hbar} |H_{jl}^D|^2, \tag{3.12}$$

where $H_{jl}^D = \langle l|e\mathbf{r}|j\rangle$ is the matrix element of the dipole transition between the states j and l. For dipole-forbidden transitions, quadrupole matrix elements should be taken instead of H_{jl}^D. However, they are much weaker than the dipole transitions and generally do not suffice to make the dielectric function negative. In any case, the calculation of the oscillator strength f from first principles is a demanding task. Rather than from Eq. 3.12, the oscillator strength is thus in practice usually determined by measuring the reflectivity of the sample and fitting the measured values to Eq. 3.11.

Apart from the aforementioned quantitative difference between the classical and the quantum-mechanical treatment of the light-matter interaction, there is also a qualitative one. In particular, the classical picture allows for an arbitrary oscillation amplitude and consequently an arbitrary oscillator energy. On the other hand, a quantum-mechanical harmonic oscillator is quantized, i.e. possesses only discrete energy levels. Furthermore, the quantization applies equally well to the electromagnetic field oscillations and the oscillations in matter involving either single particles or collective oscillations of many particles. The quanta of electromagnetic field are thereby known as photons, and the quanta of oscillations in matter are named according to the nature of the oscillation. For example, collective oscillations of electrons against the ions in a metal ("plasma oscillations") have quanta know as **plasmons** [82]. Analogously, the crystal lattice vibrations have **phonons** [83] as their quanta, and the excitation of electron-hole pairs in semiconductors appears in quanta known as **excitons** [84–86].

Of particular interest here is the coupling between the electromagnetic waves and the oscillations in matter. For this coupling to be efficient, it is necessary that the oscillations in matter be associated with the polarization of the material. For plasmons and excitons this is always the case, whereas the phonons can exhibit this property only in polar crystals, i.e. crystals with at least partially ionic character. In addition, the interaction of light with phonons is only possible for optical phonons, characterized by atoms in the unit cell oscillating against each other. This is in contrast to acoustic phonons, where atoms in a unit cell oscillate in phase with each other, so their oscillations are not accompanied by a polarization unless the material is piezoelectric.

In all cases where the electromagnetic waves couple to the polarization waves in matter,

mixed modes are formed. Quantum mechanics requires those mixed waves to be quantized, and the resulting quanta are quasi-particles known as **polaritons** [87, 88]. In particular, when photons couple to plasmons, the resulting quasi-particles are called plasmon polaritons. In case of photon-phonon coupling we get phonon polaritons, and so on. Since the polaritons are formed by coupling of two bosons, they also possess integer angular momenta.

Properties of all kinds of polaritons largely depend on the energy difference between photons and the quanta of matter polarization. When the energy mismatch is large, it is customary to describe the resulting polaritons as photons propagating in a dielectric medium with $\varepsilon \geq 1$. Such approach is also supported by the polaritons' dispersion relation, which is in the region of weak coupling just a straight line with the constant slope $\omega/k = c/\sqrt{\varepsilon}$. This can be seen in the lower part of Fig. 3.2 which contains a plot of the dispersion relation derived from the dielectric function in Fig. 3.1(a). It should be noted that although this dispersion is based on the classical model (Eq. 3.11), it is consistent with the quantum-mechanical results [81].

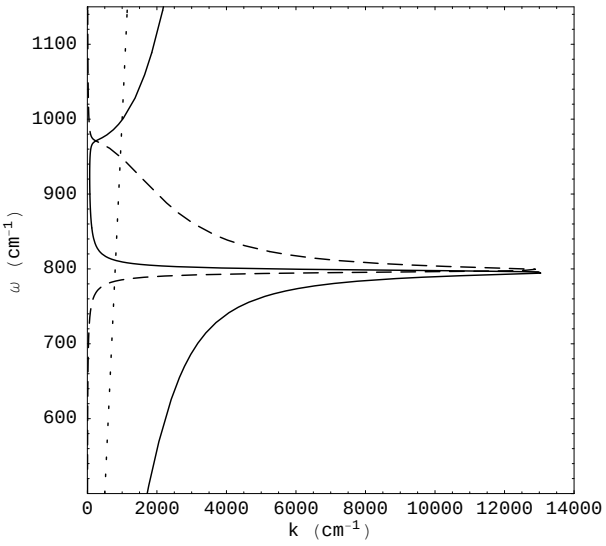


Figure 3.2: The dispersion relation $\omega = ck/\sqrt{\varepsilon}$ of phonon polaritons in a SiC crystal, derived form the dielectric function shown in Fig. 3.1(a). The full line represents the real part of the wave vector, and the dashed line the imaginary part. The dotted line depicts the wave vector of the photons in vacuum (i.e. the light line).

As the photon energy $\hbar\omega$ approaches $\hbar\omega_0$, the coupling between electromagnetic and polarization waves becomes stronger. Photons gain additional momentum due to this coupling

and their dispersion relation deviates from the straight line. This is also accompanied by a fast rise in the imaginary part k'' of the wave vector k, represented by the dotted line in Fig. 3.2. Large values of k'' signify a rapid extinction of the polaritons as they propagate through the material. By comparing Fig. 3.2 and Fig. 3.1(a), we can conclude that the extinction around the frequency $\omega \approx \omega_0$ is mainly due to absorption because of large values of ε'' around $\omega \approx \omega_0$.

3.3 Surface Polaritons

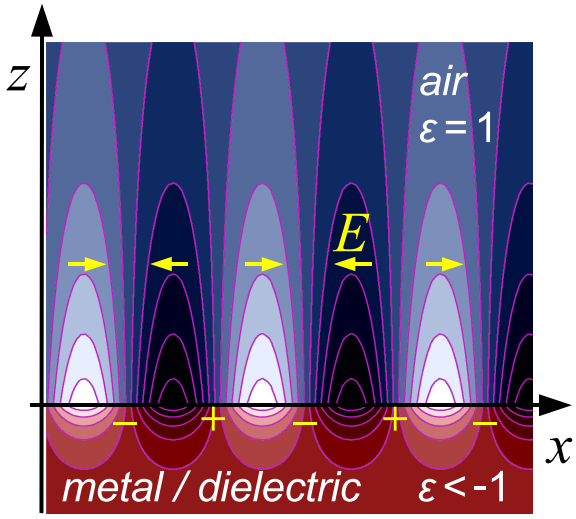


Figure 3.3: Surface polaritons propagating along the interface between two materials with opposite signs of their dielectric function. In the direction perpendicular to the surface, the electric field of surface polaritons decays exponentially.

Going to frequencies ω above ω_0, the imaginary part k'' of the wave vector remains large over the entire range of frequencies with negative ε'. In the range of negative ε', the polarization in the material opposes the external electric field and causes the photons to be reflected from the sample surface, thus again contributing to the extinction of polaritons. Obviously, polaritons cannot propagate through the material in this frequency range. Nevertheless, modes exist that propagate *along the surface* of the sample and are evanescent perpendicular to the surface, as illustrated in Fig. 3.3. Such modes are known as **surface polaritons** and they obey a dispersion relation different from the bulk (volume) polaritons [89–91]. On the interface between a medium with a dielectric constant ε_m and a material

with a dielectric function $\varepsilon(\omega)$ the dispersion relation of surface polaritons is given by:

$$k_{\mathrm{SP}} = \frac{\omega}{c}\sqrt{\frac{\varepsilon(\omega)\,\varepsilon_m}{\varepsilon(\omega)+\varepsilon_m}}. \tag{3.13}$$

A plot of this relation is shown in Fig. 3.4 using the same parameters as for the volume

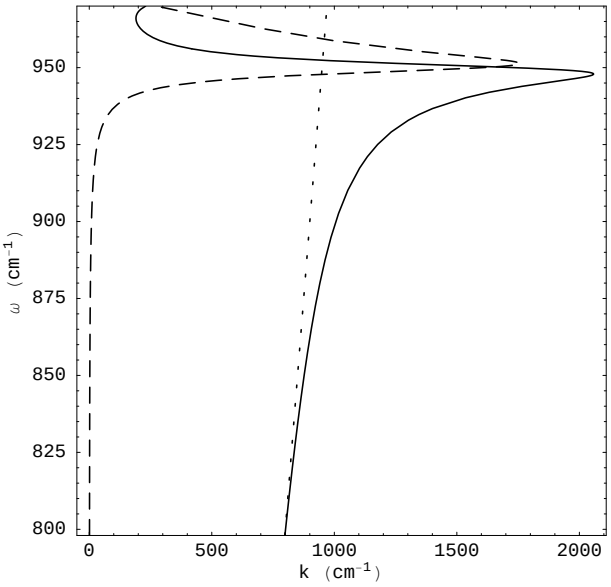


Figure 3.4: The dispersion relation $\omega = ck/\sqrt{\varepsilon/(\varepsilon+1)}$ of the surface phonon polaritons on an interface between a SiC crystal and vacuum calculated using the dielectric function from Fig. 3.1(a). Shown is the negative-ε frequency region between $\omega_{\mathrm{TO}} = 800\,\mathrm{cm}^{-1}$ and $\omega_{\mathrm{LO}} = 970\,\mathrm{cm}^{-1}$. The full line represents the real part of the wave vector, and the dashed line the imaginary part. The dotted line depicts the wave vector of the photons in vacuum (i.e. the light line).

polaritons in Fig. 3.2. We see that the momentum $\hbar k_{\mathrm{SP}}$ of surface polaritons is in principle larger than the momentum $\hbar\omega/c$ of photons outside the sample. Where this is true, the surface polaritons can neither be directly excited nor can they radiate into free space if the surface of the sample is flat. Alternatively, when the momentum of surface polaritons is smaller than the momentum of photons, the surface polaritons are no more bound to the surface and thus quickly cease to exist.

The boundaries of the spectral region with negative ε' bear special meaning and names depending on the kind of polaritons they are associated with. In connection with the lattice oscillations, the lower boundary ω_0 is known as the transverse optical phonon frequency, denoted by ω_{TO}. The upper boundary $\varepsilon'(\omega) = 0$ is the longitudinal optical phonon fre-

quency, ω_{LO}. Its name is derived from the fact that in the absence of damping, the polarization field in the material is equal and opposite to the driving field and thus cancels it, making the existence of *longitudinal* optical waves possible at this frequency. These longitudinal waves are pure polarization and not really electromagnetic waves since they are characterized by vanishing magnetic field H and induction B. Being longitudinal, they also do not couple directly to the free-propagating light which consists of transverse waves only.

In metals, conduction electrons are usually depicted forming a free electron gas. Since the conduction electrons are free to move through the metal, there is no restoring force associated with their displacement and the natural frequency ω_0 of their oscillations is equal to zero. The restoring force appears only when electrons attempt to move out of the metal. This is the case for collective oscillations of electrons against the positive ions in the metal. Such longitudinal oscillations are known as electron plasma oscillations and their eigenfrequency is the plasma frequency ω_{P}. From the perspective of polaritons, the plasma frequency ω_{P} plays the same role for plasmon polaritons as the longitudinal optical phonon frequency ω_{LO} does for phonon polaritons.

Finally, it should be added that the frequencies ω_{TO} and ω_{LO} (or their analogs) are not independent of each other. In general, the following relation between the two frequencies and the oscillator strength f holds:

$$\omega_{\mathrm{LO}}^2 = \omega_{\mathrm{TO}}^2 + \frac{f}{\varepsilon_\infty}. \tag{3.14}$$

Eq. 3.14 is a variant of the Lyddane-Sachs-Teller equation [92] and enables the oscillator strength to be determined if the frequencies ω_{TO} and ω_{LO} are known together with the high-frequency dielectric constant ε_∞. All these quantities can experimentally obtained from e.g. reflectivity measurements or with the aid of Raman spectroscopy.

The remaining oscillator parameter is the damping coefficient γ which is much harder to determine. Nevertheless, a rough comparison of the damping coefficients pertaining to different kinds of polaritons can be made. Instead of the damping coefficient γ, the oscillator relaxation time $\tau = 1/\gamma$ is often specified. Phonons in defect-free crystals have in general relatively long relaxation times, on the order of $\tau = 10^{-12}$ s. With the notable exception of silver, the relaxation time of plasmons in metals is one to two orders of magnitude shorter than that of phonons. On the other hand, the relaxation time of excitons in semiconductors may vary over several orders of magnitude depending on the preparation and purity of the

sample.

Since the strength of the near-field resonance observed in s-SNOM is ultimately limited by the damping of the polaritons, we can conclude that the phonon polaritons should yield the strongest near-field resonances. In Chapter 7 it will be shown how this resonances can be utilized as a very sensitive method for analyzing the crystal structure on the nanometer scale. Another, very intriguing application of surface phonon polaritons is the coherent emission of thermal radiation [93], enabled by the diffraction of surface phonon polaritons by a grating etched in the surface of a SiC crystal [94].

For the near-field microscopy it is important to note that the density of surface polariton states, dk_{SP}/dE peaks around the frequency ω_{SP} where $\varepsilon'(\omega_{\mathrm{SP}}) = -\varepsilon_m$, which is the same frequency where the quasi-electrostatic reflection coefficient β has a maximum. It might therefore seem logical to ascribe the near-field resonance described in Sect. 2.5 to the excitation of surface polaritons. However, in Sect. 5.2 it will be shown that the waves participating in the near-field interaction between a sample and a probing tip have spatial frequencies far above the spatial frequency of both propagating surface polaritons and volume polaritons. For this reason, the distinction between the type of polaritons does not seem to be of importance in the context of s-SNOM near-field resonances.

Nevertheless, surface polaritons lend themselves naturally to the investigation by scanning near-field optical microscopy which is a surface scanning method by design. SNOM and the closely related scanning photon tunneling microscopy (PSTM) have been often employed for mapping the surface polariton propagation through various structures such as fabricated waveguides [95] and corrugated surfaces [96]. By making a series of such measurements at different illumination frequencies, the dispersion curve of the surface polaritons can be traced out [97].

It is interesting to note that the surface polaritons are also responsible for the existence of the near-field "perfect lens" [98]. The near-field perfect lens comprises only a slab of material with $\varepsilon = -\varepsilon_{\mathrm{m}}$ and produces an almost perfect image of the source at a distance $x_{\mathrm{image}} = d - x_{\mathrm{source}}$, where d represents the lens thickness. Image formation is thereby mediated by surface polaritons whose dispersion relation, Eq. 3.13, yields infinitely large wave vectors k_{SP} for $\varepsilon = -\varepsilon_{\mathrm{m}}$, hence posing no limit to the spatial frequencies which can be transferred by surface polaritons. However, the condition $\varepsilon = -\varepsilon_m$ can be exactly satisfied only in the absence of damping. In practice, the finite damping $\varepsilon'' \neq 0$ limits the magnitude of k_{SP} as shown e.g. in Fig. 3.4 and degrades the perfect lens properties

exponentially with the lens thickness d. The resolution of a "perfect lens" is hence always limited and such a lens is in general only useful in the near-field regime $x_{\text{source}} \leq d \ll \lambda$. Subwavelength resolution can nevertheless be achieved, as was demonstrated using both surface plasmons [99, 100] and surface phonon polaritons [101, 102].

One further peculiarity of surface polaritons is the existence of guided modes on cylindrical or conical waveguides exhibiting no cutoff regardless of the waveguide diameter [103]. The surface polariton wave vector thereby rapidly increases with shrinking waveguide diameter, resulting in the surface polariton super-focusing effect [104–106] with potential applications in near-field optical microscopy [107]. On the other side, thin rectangular waveguides support surface polariton modes with largely sub-wavelength field confinement and propagation length in the centimeter range [108–111]. They could provide the basis for optical interconnects in integrated circuits and for addressing individual elements in such circuits [112–115].

3.4 Particle Polaritons

A particularly interesting situation arises when a material supporting surface polaritons forms a small particle with the size much smaller than the light wavelength. Surface polaritons in such particles are localized in all three spatial dimensions and often referred to as the particle polaritons [116, 117]. They are highly interesting because of the large field enhancement they provide in the vicinity of resonantly excited particles [118]. The spectral position and shape of such resonances, also known as Fröhlich resonances, depend on the material the particle is composed of, as well as the particle's shape. Analytical solution for the field enhancement can be obtained in case of ellipsoidal particles with the semi-axes $c \leq b \leq a \ll \lambda$. It turns out that a dipole mode is excited in such small particles, with the maximum field enhancement at the surface of the particle given by

$$\gamma = \frac{\varepsilon - \varepsilon_m}{l_a(\varepsilon - \varepsilon_m) + \varepsilon_m}, \tag{3.15}$$

where l_a is the depolarization or shape factor defined by

$$l_a = \int_0^\infty \frac{abc}{2\sqrt{(q+a)^3(q+b^2)(q+c^2)}} \mathrm{d}q. \tag{3.16}$$

A maximum in the field enhancement γ occurs for $\varepsilon = \varepsilon_m(1-1/l_a)$ and is independent of the particle size as long as the approximation $a \ll \lambda$ is valid. In case of spherical particles we have $l_a = 1/3$, and $\varepsilon = -2\varepsilon_m$ is the condition for the maximum field enhancement.

It has been proposed that a chain of closely spaced nano-particles similar in size should enable one-dimensional energy transport with lateral confinement on the subwavelength scale by dipole-dipole coupling between the particles[119]. Experimental confirmation of the energy transfer in such an arrangement of particles and was provided in [120]. An alternative arrangement of n small particles of decreasing size in a self-similar chain was predicted to have focusing properties [121]. In the self-similar chain, larger particles drive smaller ones without being disturbed and an overall field enhancement of γ^n can be achieved. This results in an efficient nanolens with potential applications in the near-field microscopy and especially in particular single-molecule Raman spectroscopy which requires field enhancements on the order of 10^3 for the weak Raman scattering to be detectable.

Nanoparticles can also act as a very efficient sensors when functionalized such that they preferentially bind a certain target molecule. Binding of the target molecules modifies the dielectric constant ε_m of the medium around the nanoparticle and causes a slight spectral shift of the particle plasmon resonance [122]. This shift can be observed in the scattering spectrum of the functionalized nanoparticles even for a very small amount of target molecules.

Finally, metallic nanoparticles can be enclosed in a dielectric shell of a given ε_m and a chosen thickness, or, alternatively, a dielectric core can be coated by a metal shell to deliberately shift plasmon resonances to wavelengths between about $0.5\,\mu$m and $5\,\mu$m [123]. Tuning the absorption of nanoshells to about $\lambda = 1.1\,\mu$m, targeted and localized heating can be achieved in living organisms since the physiological absorption is very low in this spectral region. This opens the possibility for targeted drug delivery and photothermal therapy of tumors [124, 125].

The application mentioned above do not represent an exhaustive list. A vast effort has been invested in the research of nanoparticles in the last decade, and with no sign of a decline. Consequently, a number of novel applications of particle polartitons as well as methods for fabrication and detection of polariton-supporting nanoparticles can be expected in the future.

4 Near-field Optical Probes

In Sect. 2.1 it was mentioned that the diffraction of light limits the resolving power of conventional optical elements such as lenses or concave mirrors. For the same reason, these elements cannot be used to focus the light to dimensions below roughly one wavelength (λ). Near-field optical microscopy (Sect. 2.2) overcomes this limitation by utilizing a largely subwavelength light source located in the immediate vicinity of the sample.

Denoting the radius of a SNOM light source by $\Delta\rho$, and the spatial frequencies in the plane containing the source by ν_ρ, we can make use of the classical "uncertainty relation" $\Delta\rho\,\Delta\nu_\rho \geq 1$. This relation is applicable to any pair of variables related by the Fourier transform, and in this particular case it yields $\Delta\nu_\rho \gg \lambda^{-1}$ since $\Delta\rho \ll \lambda$. From the Fourier optics it is well known that spatial frequencies $\nu_\rho > \lambda^{-1}$ represent exponentially decaying (evanescent) waves. This means that the field near a SNOM light source is predominantly evanescent.

Being related to high spatial frequencies (large ν_ρ), the evanescent waves are the key to the high resolution. However, the results of optical probing are recorded by a detector located far away from the source in terms of the light wavelength. Obviously, the evanescent waves do not extend to the detector and therefore do not contribute to the detected signal. In other words, an attempt in high-resolution optical imaging based on the strong field confinement alone is not guaranteed to yield the desired information. On the contrary, this is only possible if the evanescent fields can be converted into the propagating waves which then convey the information to the detection point.

The two major SNOM probe types (Sect. 2.2) can now be compared from this point of view. Aperture SNOM utilizes a small opening in an otherwise opaque screen to confine the light. Apertures smaller than about half the wavelength exhibit transmission efficiency decreasing with the fourth power of the aperture size [126–128]. This results in a poor coupling of evanescent into propagating waves for small apertures and effectively limits the resolution of an a-SNOM to roughly $\lambda/10$. In contrast, a long and sharp metallic tip

of a s-SNOM is much more efficient in conversion between evanescent and propagating waves because the movement of free charges is essentially unrestricted along the tip, i.e. in the direction perpendicular to the sample surface. The confinement of charge by a s-SNOM probe takes place only in the plane parallel to the surface. This plane is orthogonal to the electric field oscillations, so the charge confinement does not impede the process of absorption or radiation of electromagnetic waves. Owing to this property, the s-SNOM exhibits a wavelength-independent resolution limited only by the radius of the probing tip apex [45].

From the preceding description it is obvious that the shape and function of a s-SNOM probe very much resemble those of an antenna. And indeed, the simplest radio antennas have the form of elongated metallic rods, with their length optimized for efficiency. In the region of infrared and larger wavelengths their analogues can be easily imagined and fabricated [129–131] since most metals are good conductors at those wavelengths. In the visible light range there are no such good conductors, but it is to some extent possible to exploit small-particle plasmon resonances (Sect. 3.4 and 4.1) to achieve efficient coupling between the near and far field [132, 133]. To compensate for the lack of good conductors at visible light wavelengths, alternative antenna geometries have also been proposed, yielding significantly higher field enhancements than elongated rods or small spheroidal particles. They include bow-tie antennas [134–136], triangular [137] and C-shaped [138] apertures, and even self-similar chains of small particles with progressively decreasing size [121].

This analogy between s-SNOM probes and antennas can be formalized by actually defining the optical antenna as an *object or device which efficiently couples the free-space radiation to highly localized electromagnetic field and vice versa* [139]. From this definition, one important difference between classical radio antennas and optical antennas follows: whereas radio antennas have to transport as much charge as possible to one of their terminals, for optical antennas it is additionally the concentration of this charge that matters. The reason for this additional requirement on optical antennas can be understood by noting that an optical antenna is more efficient if it produces stronger EM fields because this leads to larger useful signals, and it is also better if it produces tighter field confinement because this results in higher resolution. Fortunately, the charge confinement favorably affects the field strength, so that both performance criteria can be simultaneously improved. The proportionality between the charge concentration and the field strength is thereby often attributed to the ”lightning rod effect” [140]. It relies on the fact that the field strength near

an equipotential surface is inversely proportional to the local radius of curvature. For the near-field optical microscopy this means that in the wavelength range where good electric conductors exist, sharper probes will produce both better field confinement and higher field strength. In the remainder of this chapter, these two quantities will be studied in a more quantitative manner.

4.1 Field Enhancement

A detailed numerical analysis of the field enhancement by s-SNOM probing tips can be found in many papers, e.g. [141], [142], [143], [144], [145], [146], and [147]. However, an *analytical* expression for the field enhancement would be more interesting in the context of this work since it would permit the probe-sample interaction to be modeled in a relatively simple way (Chapter 5). The best known analytically solvable approximation to the s-SNOM probing tip is a prolate spheroid in a uniform electric field, treated in Refs. [71], [70], [148], and [145]. The adoption of this model is justified by noting that the apex of a typical probing tip does indeed look much like the apex of a spheroid. Furthermore, the tip apex is also expected to be the decisive part for the near-field interaction with the sample since it resides in the closest proximity of the sample. The largely different shape far away from the apex certainly influences the radiation characteristics of the probe, but not the nature of the probe-sample interaction which gives rise to contrasts in s-SNOM.

On the other hand, treating the illumination as a homogeneous field can only be a good approximation over short lengths. This sets an upper bound to the length of a probe which can be accurately modeled in the electrostatic approximation. Based on electrodynamic calculations of the field enhancement by realistically modeled probes [143], one can conclude that the spheroid length should be kept below about $\lambda/4$ in an electrostatic calculation. Otherwise, the field enhancement can be significantly overestimated due to the neglected dephasing effects.

Using an electrostatic approximation, the electric potential around the probing tip can be approximated by the potential Φ outside a prolate spheroid in a uniform field E_0. In spheroidal coordinates, Φ can be expressed as [145]

$$\Phi = E_0 F \eta \left[\xi + A\left(-1 + \frac{\xi}{2} \ln \frac{\xi+1}{\xi-1}\right)\right], \tag{4.1}$$

where A is a constant equal to

$$A = \frac{(\varepsilon - 1)\,\xi_o}{\xi_0 \frac{dQ_1}{d\xi}(\xi_0) - \varepsilon\,\mathscr{L}_1(\xi_o)}.$$

The first term in Eq. 4.1, $\Phi_0 = E_0 F\,\eta\,\xi$, is the potential of the external field and the second part is the potential of the spheroid, i.e. the induced charge. The notation used in the above expressions and the rest of this section is explained in Table 4.1.

Variable	Description
η, ξ, φ	spheroidal coordinates
$\mathscr{L}_n$	Legendre function of the second kind, n-th order
E_0	external electric field strength, oriented along the long spheroid axis
ξ_0	spheroid boundary
F	half the distance between the spheroid foci
L	length of the semi-major spheroid axis
R	radius of the spheroid curvature at its apex

Table 4.1: Notation used in Eq. 4.1.

For convenience, equations transforming the spheroidal coordinates to the more familiar cylindrical coordinates are provided:

$$\rho = F\sqrt{(1-\eta^2)(\xi^2-1)}, \tag{4.2}$$

$$z = F\,\eta\,\xi, \tag{4.3}$$

while the azimuthal angle φ has the same meaning in both coordinate systems.

The electric field E outside the spheroid can be easily obtained as the negative gradient of the potential Φ. The final expression is rather complicated and the result is thus best presented graphically. This has been done in Fig. 4.1 for a spheroid with semi-major axis L and a curvature radius at its apex $R = 0.1\,L$.

Looking at the vector field plot in Fig. 4.1, it can be easily seen that the field close to the spheroid ends looks much like a field of an extended dipole with its constituent charges residing near the opposite end of the spheroid. This observation is one of the cornerstones of the tip-sample interaction model derived later in Chapter 5.

Fig. 4.1 provides a qualitative insight into the field confinement and enhancement of the prolate spheroid. For a quantitative analysis, the enhancement factor $\gamma = E/E_0$ at the

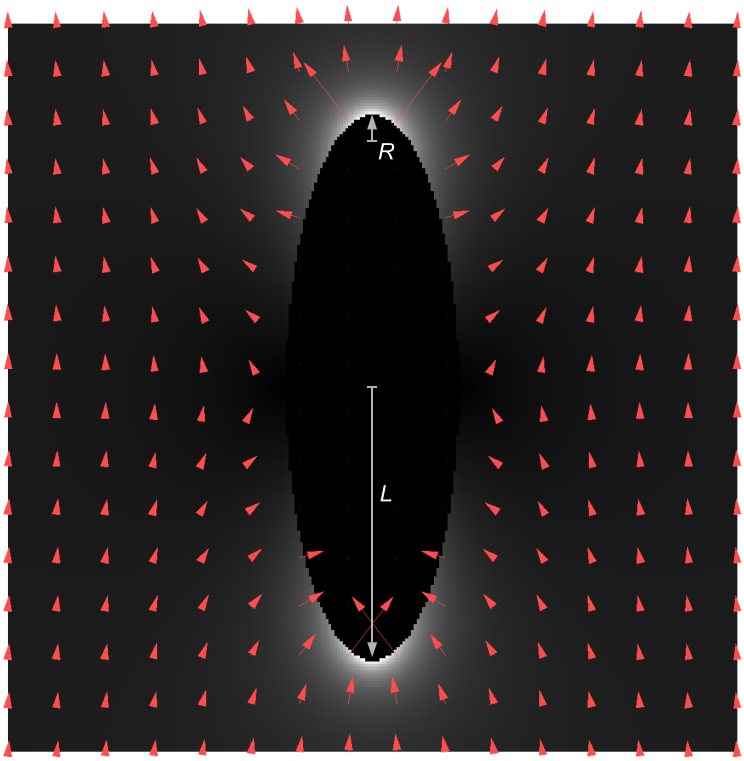


Figure 4.1: Electric field in the vicinity of a spheroid with $R = 0.1L$ located in a uniform external electric field oriented along the spheroid axis.

bottom of the spheroid can be derived from Eq. 4.1. For a point at the intersection of the surface and the axis of the spheroid one obtains

$$\gamma = \frac{L(\varepsilon-1)(2F + R\ln\frac{L-F}{L+F})}{2F(L-\varepsilon R) - LR(\varepsilon-1)\ln\frac{L-F}{L+F}} + 1 \tag{4.4}$$

where F is the focal distance $F = \sqrt{L(L-R)}$. For convenience, Eq. 4.4 can also be rewritten in terms of relative dimensions $r = R/L$ and $f = \sqrt{1-r}$:

$$\gamma = \frac{(\varepsilon-1)(2f + r\ln\frac{1-f}{1+f})}{2f(1-\varepsilon r) - r(\varepsilon-1)\ln\frac{1-f}{1+f}} + 1, \tag{4.5}$$

The spheroid shape enters the right-hand side of Eq. 4.5 only through the ratio r of the spheroid curvature radius R to its half-length L. The field enhancement γ is thus completely determined by the ratio r and the dielectric constant ε of the spheroid. The field enhancement γ as a function of r and ε is shown in Fig. 4.2. The dielectric constant ε was thereby taken in the Drude form $\varepsilon(\omega) = \varepsilon_\infty - \omega_p/\omega(\omega + i\gamma)$ with $\omega_p = 9\,\mathrm{eV}$, $\gamma = 0.1\,\mathrm{eV}$

and $\varepsilon_\infty = 2$ in order to obtain values similar to a real metal.

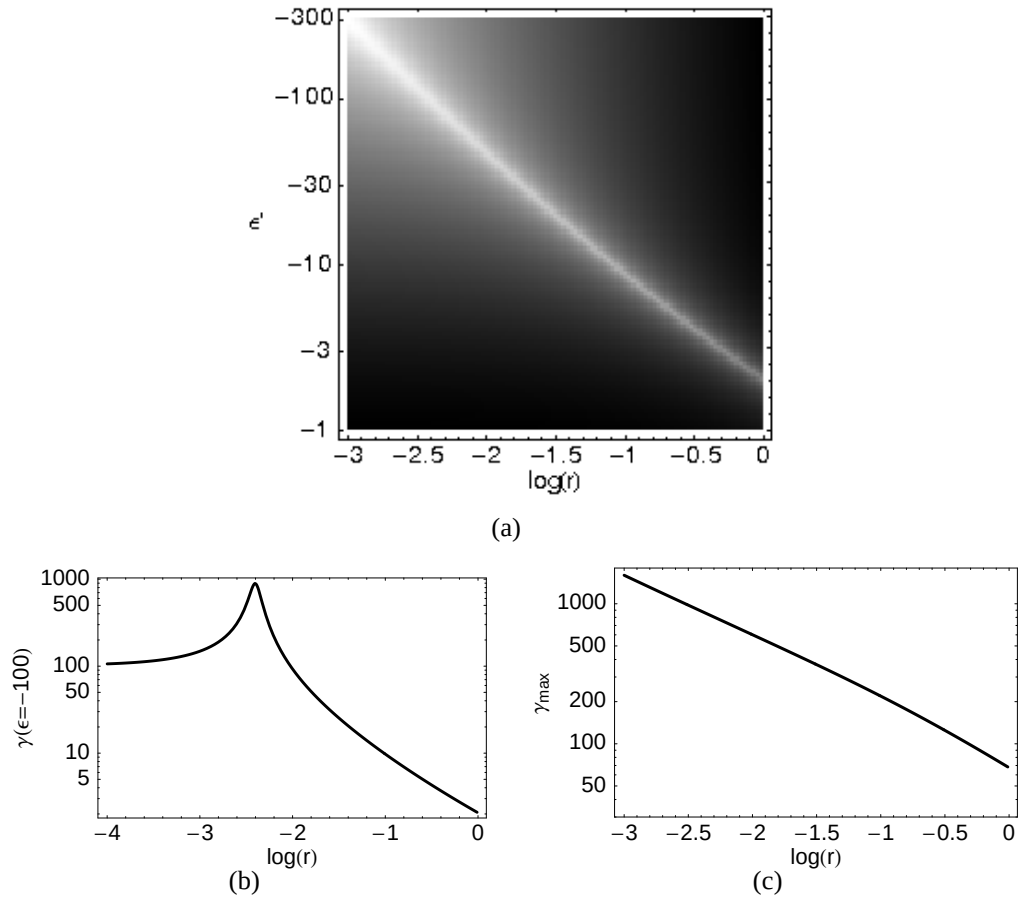


Figure 4.2: (a) Field enhancement γ at the spheroid apex as a function of the spheroid dielectric constant $\varepsilon(\omega) = 2 - 9\,\mathrm{eV}/\omega(\omega + 0.1i\,\mathrm{eV})$ and the ratio $r = R/L$ of the spheroid curvature radius to its half-length, (b) γ as a function of r for $\varepsilon' = -100$, and (c) maximum field enhancement γ as a function of r.

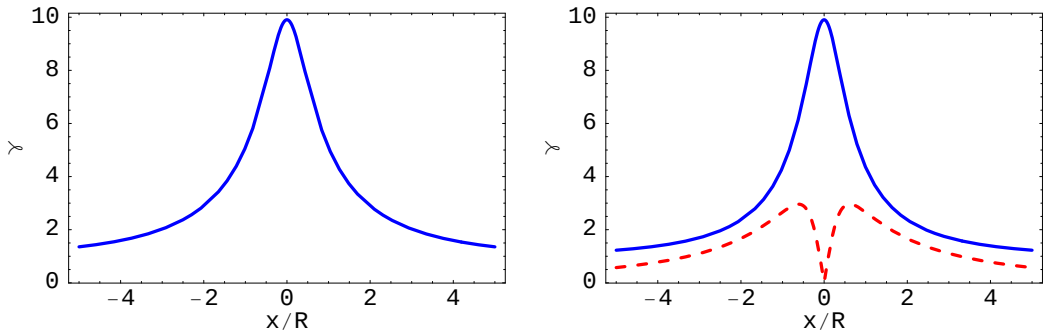
Two conclusions can be immediately drawn from Fig. 4.2(a) and the line profile along $\varepsilon' = -100$ shown in Fig. 4.2(b). First, for each r, the field enhancement at the surface of the spheroid exhibits a sharp peak which shifts towards more negative ε' (lower frequencies ω) as the spheroid shape factor r decreases. This is a manifestation of the small-particle plasmon resonance, explored in more detail in Sect. 3.4. Second, higher field enhancement γ can be achieved with sharper tips (lower r). This is quantified in Fig. 4.2(c) where the maximum field enhancement γ_{max} is shown as a function of r. Based on an extrapolation, it might seem that an arbitrary large enhancement can be achieved by decreasing $r = R/L$ sufficiently enough by making the spheroid very long (large L) and thin (small R). In practice, however, the length of the tip must be kept significantly below the light wavelength for

the electrostatic approximation to be valid. Furthermore, the tip radius can only be reduced down to a few nanometers before the bulk dielectric constant becomes inappropriate for describing the optical properties of the material [149]. For this reason the field enhancement factors in the mid-IR range are limited to about three orders of magnitude in the best case.

Refs. [145] and [143] contain a more detailed electrodynamic treatment of prolate spheroids and a comparison with the electrostatic calculation of the type presented above. A good agreement between the electrostatic and electrodynamic calculations can be inferred from them for spheroids shorter than roughly $\lambda/5$. Field enhancement factors up to several hundreds are indeed obtained for thin spheroids even with the retardation effects included. In the visible spectral range, such high field enhancement is attributed to the small-particle plasmon resonances. In the infrared, the lightning rod effect on its own is sufficient to achieve the same effect.

4.2 Field Confinement

Here we consider the second important aspect of the optical antenna function: the field confinement. Based on Fig. 4.1, the confinement can already be estimated to be about one spheroid apex radius R. For a more quantitative analysis, Fig. 4.3 shows the horizontal line profiles extracted at the bottom of the spheroid in Fig. 4.1.



(a) (b)

Figure 4.3: (a) Total field enhancement below the spheroid from Fig. 4.1 and (b) its two orthogonal components: γ_z (full line) and γ_ρ (dashed line).

As expected, the field enhancement in Fig. 4.3 drops to about half its maximum value already one tip radius R far form the axis. However, the curves in Fig. 4.3 can be used

to estimate the resolution obtainable using such a tip for near-field microscopy purposes only after the method of detection and the structure of the sample have been taken into account. If the probing tip serves both to collect and to radiate light, a signal enhancement proportional to the square of the field enhancement follows from the optical reciprocity principle [150, 151]. Additionally, in the case of a non-interferometric signal detection, photodetectors measure intensity of the emitted light, proportional to the square of the field strength. This makes the cumulative enhancement of the detected signal proportional to γ^4 when both aforementioned effects are combined. The enhancement factor γ^4 applies to the signal generated by linear processes like fluorescence or elastic and Raman scattering. In the case of non-linear response associated with e.g. two-photon fluorescence excitation [51] or second harmonic generation [49], the detected signal enhancement is proportional to an even higher power of the field enhancement γ. The curve from Fig. 4.3(a) raised to the appropriate power finally determines the point spread function (PSF) of the near-field optical microscope using a probing tip whose field enhancement is shown in Fig. 4.3. The width of the PSF equals the optical resolution when imaging samples consisting of point-like objects, i.e. objects much smaller than the tip radius. For extended objects whose dimensions are comparable to or larger than the tip radius, the emitted signal can be calculated as a convolution of the PSF and the object response. This issue will not be pursued further since the standard FWHM resolution criterion is not applicable in such cases and the results are dependent on the actual criterion used.

5 Monopole Model of the Near-field Interaction

5.1 Monopole Model

5.1.1 Motivation

The efficiency of s-SNOM probes has been examined in Chapter 4. Based on an electrostatic model of the probing tip as a prolate spheroid, field enhancement and confinement have been calculated. In the context of the near-field optical microscopy, these two properties determine the detected signal level and attainable resolution, respectively. At the same time, they provide no information about one of the most fundamental issues in s-SNOM, the optical contrast. In order to explain and predict optical contrasts, it is necessary to understand the interaction between the probing tip and the investigated sample. An accurate model of the tip-sample interaction is of particular importance to the apertureless near-field spectroscopy since it represents the key to the identification of materials based on their spectral fingerprints.

The first attempt at describing the tip-sample interaction was the point dipole model [57, 70–73], described in Sect. 2.4. This model succeeds in a *qualitative* explanation of almost all phenomena experimentally observed in s-SNOM imaging, including material contrasts [44, 46, 64], near-field resonance [63] and its blue-shift with increasing probe-sample distance [61]. It is also simple enough to give a direct insight into the major determinants of tip-sample interaction. However, the point dipole model does not provide a good *quantitative* prediction of the measured signal. In particular, it fails to correctly reproduce both the spectral position and the magnitude of near-field resonances, as shown in Sect. 2.5.

Apart from the analytical dipole model, the tip-sample interaction has also been the object of several numerical studies. Different computational techniques have been employed,

including finite element [152] and boundary element [75] methods, multiple multipole method [76], multipole expansion [153] and a time-domain approach [154]. Numerical modeling of the tip sample interaction, especially at the infrared wavelengths, is a highly complex task. The complexity stems from the widely different length scales which need to be covered simultaneously. On one side there are the tip length and the illumination wavelength, both possibly larger than 10 µm. On the other side, the system must be modeled with precision better than the tip radius, usually in the nanometer range. The resulting dynamic range of four orders of magnitude usually requires some compromises to be made before a numerical solution can be obtained. A common simplification is the reduction of the problem dimensionality from three to two dimensions by imposing the axial symmetry of the system. Additionally, the probes are usually artificially truncated instead of considering their full length and an idealized tip shape is assumed. Finally, the probing tip vibration is rarely taken into account due to the large computational effort involved, although it represents an essential aspect of the s-SNOM signal detection [155].

For one or more of the above reasons, no model has so far accurately reproduced the experimentally obtained spectra [63, 66] of near-field resonant samples. Such outcome also precludes an unambiguous material identification based on near-field vibrational spectra. With constant improvement in the computer power and modeling techniques, it is likely that some numerical method will eventually succeed at this task. However, unless it turns out to be much faster than methods currently in use, it will most likely be too slow to be used in the inverse way and determine the optical constants of a material based on the measured near-field optical signal at each pixel in an image. An improved analytical model would therefore be the most important step towards the promised local optical constants on the nanometer scale [58]. An attempt at achieving this goal is presented in this chapter.

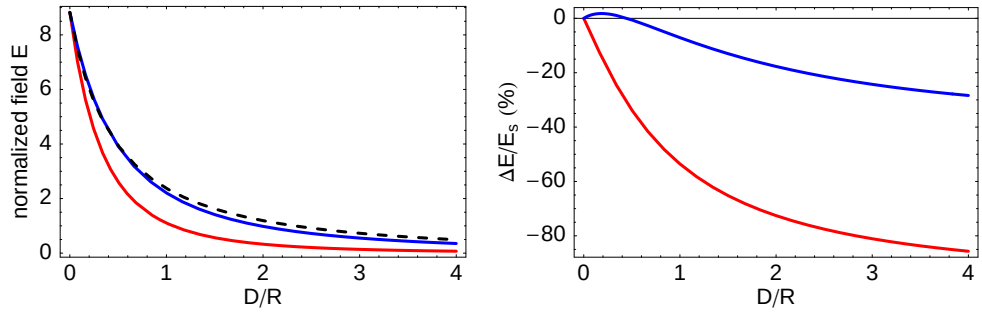
5.1.2 Overview

To derive an improved model of the tip-sample near-field interaction, the reason for the failure of the dipole model will be identified and subsequently corrected. As described in Sect. 2.4 and depicted in Fig. 2.5, the dipole model reduces the probing tip to a point dipole located near its end. We have already seen in Fig. 4.1 that the field around a spheroid looks much more like a field of an extended (i.e. finite), and not a point-like dipole. At a large distance from the spheroid, the point-like and extended dipoles are hardly distinguishable. Nevertheless, significant differences exist in the immediate vicinity of the spheroid. This

claim is quantitatively supported by Eq. 4.1 which yields the following expression for the electric field $E_s(D) = E(D) - E_0$ generated by the spheroid along its axis ($\rho = 0$) as a function of the distance D from the spheroid boundary:

$$E_s(D) = \frac{\frac{2F\,(L+D)}{D^2+L\,(2D+R)} + \ln\frac{L-F+D}{L+F+D}}{2\frac{F\,(L-\varepsilon R)}{LR(\varepsilon-1)} - \ln\frac{L-F}{L+F}} E_0. \tag{5.1}$$

The exact solution from Eq. 5.1 will now be compared to the field of a point dipole and an extended dipole. The field of point dipole along the spheroid axis is given by $E_{pd}(D) = A_{pd}/(R+D)^3$. The dipole is thereby positioned in the center of curvature of the spheroid apex, as prescribed by the dipole model (Sect. 2.4). In contrast, the field of an extended dipole consists of two monopole contributions: $E_{ed}(D) = A_m(\,(R+D)^{-2} + (2L - R + D)^{-2}\,)$. For short distances D from the spheroid ($D \ll L$) the point charge far from the observation point can be neglected. Based on Fig. 4.1, we expect the remaining single monopole field $E_m = A_m/(R+D)^2$ to match the exact result E_s better than the dipole field E_{pd}. For a meaningful comparison, the constants A_{pd} and A_m are chosen so that fields of different models assume the same value at the tip apex, i.e. so that $E_{pd}(0) = E_m(0) = E_s(0)$. The results for the spheroid from Fig. 4.1 with $R/L = 0.1$ are displayed in Fig. 5.1.



(a) (b)

Figure 5.1: (a) Electric field E_s generated by a spheroid in homogeneous external field as a function of the distance D from the spheroid end. Shown are the exact result from Eq. 5.1 (dashed line), monopole field $E_m = A_m(R+D)^{-2}$ (upper full line) and dipole field $E_{pd} = A_{pd}(R+D)^{-3}$ (lower full line), scaled to the value E_s at the spheroid surface, $D = 0$. (b) Relative deviation of the monopole field E_m (upper line) and dipole field E_{pd} (lower line) from the true spheroid field E_s.

Evidently, the monopole field is in a much better agreement with the exact result. The conformity is particularly good close to the spheroid surface, giving less than 10% error at $D = R$ where the spheroid field has already decayed to about 20% of the value at the

surface. On the other hand, the point dipole field deviates from the exact result by 50% at $D = R$ and more than 80% at a distance of $D = 4R$. Such large error is certainly one of the main reasons for the quantitative discrepancy between the dipole model and the experimentally observed near-field spectra shown in Sect. 2.5.

At the same time, Fig. 5.1 also suggests that the probing tip might be replaced by a point charge (monopole) Q_0 for the purpose of modeling the near-field interaction. The opposite charge -Q_0 constituting the extended dipole p_0 resides on the other end of the tip (i.e. the model spheroid), so it can be neglected when calculating the interaction with the sample.

The charges Q_0 and -Q_0 and the resulting dipole moment $p_0 \approx 2Q_0L$ are directly induced by the external illumination field E_0, which by definition makes them independent of the near-field interaction with the investigated sample. Material contrasts useful for near-field optical microscopy stem from an additional polarization induced due to the interaction of the probe with the sample. An attempt will be made here to describe this additional polarization of the probe by another extended dipole p_i consisting of monopoles Q_i and -Q_i, as illustrated in Fig. 5.2(b). Of those two charges, only Q_i is assumed to participate in the near-field interaction, while the existence of -Q_i is required by the electric neutrality of the probing tip. The exact strength and location of the interaction-induced charges $\pm Q_i$ shown in Fig. 5.2(b) will be determined in Sects. 5.1.4 and 5.1.5, respectively. Although the charge -Q_i is considered negligible in the near-field interaction, it is nevertheless needed to calculate the dipole moment p_i which contributes to the radiated EM field.

In summary, the part of the probe participating in near-field interaction is represented by two point charges, Q_0 and Q_i. For this reason the entire model is perhaps best described as the ”monopole model”. Such name also stands in a clear contrast to the dipole model which it is supposed to replace. For easier comparison of the two models, they are illustrated side-by-side in Fig. 5.2.

In the following, a more careful justification of the assumptions behind the monopole model is presented, together with the derivation of the expression for the interaction-induced charge Q_i and the corresponding dipole moment p_i.

5.1.3 Spheroid in Homogeneous External Field

A nice fit of the monopole field E_m to the exact spheroid field E_s shown in Fig. 5.1 is not obtained for every spheroid’s shape. As a matter of fact, the shape factor $r = 0.1$

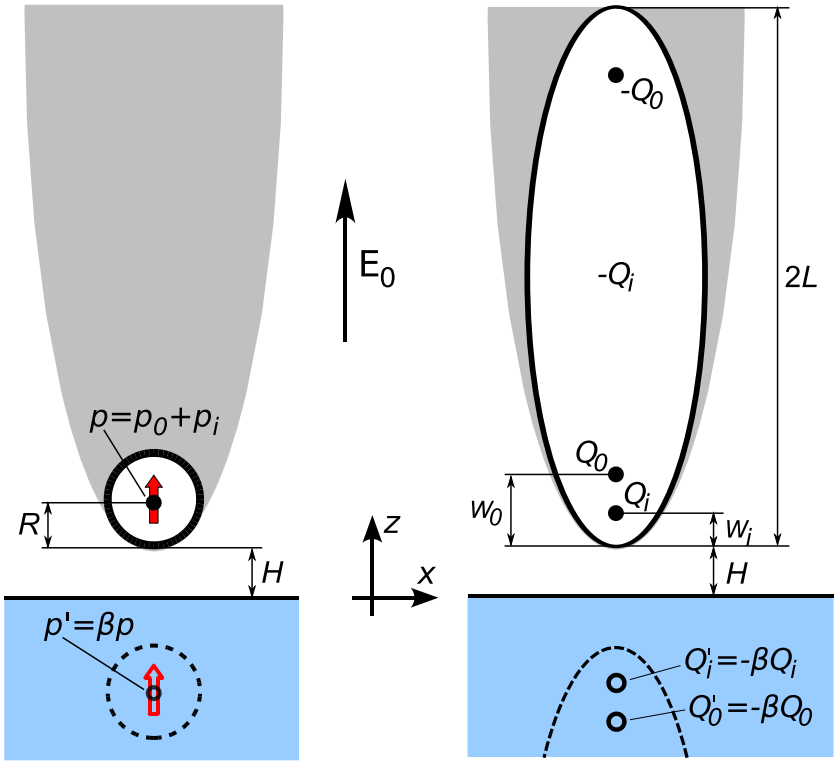


Figure 5.2: Comparison of the point dipole and the monopole model. (a) In the point dipole model, the tip is first reduced to a small sphere in a uniform electric field E_0 to determine its polarizability. The sphere is then further reduced to a point dipole located in its center for calculating the interaction with the sample. (b) In the monopole model, the tip is approximated by a spheroid in a uniform electric field $\mathbf{E}_0$ to obtain the field enhancement close to the tip apex. The spheroid is then reduced to a finite dipole p_0 which produces the same field enhancement at the tip apex. The dipole p_0 consists of the monopoles Q_0 and -Q_0, of which only Q_0 (positioned closer to the sample surface) participates in the near-field interaction. As the consequence of the interaction, an additional point charge Q_i is induced close to the spheroid focus, whereas the opposite charge -Q_i is distributed along the spheroid.

chosen in Fig. 5.1 turns out to be very close to the optimum value. For $r = 0.01$, the error $|E_s - E_m|/E_s$ at $D = R$ reaches almost 25% and further increases with the distance D. If a better fit is required over a broader range of shape factors r, the slope of the monopole field outside the spheroid can be adjusted by slightly shifting the charge location. Taking W_0 to be the required monopole distance from the spheroid boundary, this is accomplished by the following function:

$$E_m(D) = \frac{W_0^2 E_s(0)}{(W_0 + D)^2}, \tag{5.2}$$

where the factor $W_0^2 E_s(0)$ in the numerator ensures the correct field enhancement at the spheroid surface. The position W_0 can be calculated by imposing the condition $E_m(D_1) = E_s(D_1)$ for a particular distance D_1. The choice $D_1 = R$ seems to be a reasonable trade-off between the field enhancement overestimate at shorter distances and the corresponding underestimate at larger distances.

The exact solution for the charge position W_0 is rather complicated, but a simple approximate result can be derived with an accuracy better than 5% for all $r < 0.3$:

$$W_0 \approx \frac{1.31\,RL}{L+2R}. \tag{5.3}$$

The average error of about 3% made by Eq. 5.3 is completely negligible, especially because there is no need for the $E_m(D)$ to intersect the true function $E_s(D)$ exactly at $D_1 = R$. To demonstrate the effect of the charge shift, the errors with and without the shift are compared in Fig. 5.3. Two values of r were selected, bounding roughly the range of practicable shape factors conforming with the assumption $R \ll L \ll \lambda$ for λ in the mid-IR range. Owing to the shift, the error $\Delta E = |E_s - E_m|$ has been constrained to below 4% of the maximum field strength $E_s(D = 0)$ in all cases.

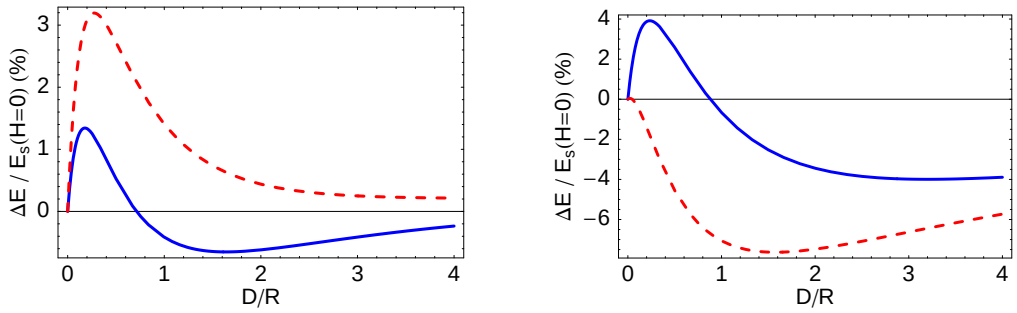


Figure 5.3: Approximation of the spheroid field E_s by a monopole field E_m, with (full line) and without (dashed line) a corrective charge shift. The deviation of the approximate result $E_m(D)$ from the exact result $E_s(D)$ is plotted, normalized to $E_s(0)$. Two spheroid shape factors $r = R/L$ are shown: (a) $r = 0.2$ and (b) $r = 0.01$.

As seen in Fig. 5.3, even without shifting the charge Q_0, the error $|E_s(D) - E_m(D)|/E_s(0)$ remains well within 10% of $E_s(0)$. This error turns out to be relatively small compared to the approximations which will be made in subsequent sections. The shift of the Q_0 position may thus be regarded as an optional correction.

In conclusion, the field of a point charge successfully reproduces the field of a spheroid over short distances $D \ll L$ and for spheroids with shape factors $0.01 \leq r \leq 0.2$, relevant for modeling the s-SNOM experiments. In the absence of the sample, this permits the entire tip under an external illumination with $\lambda \gg L$ to be reduced to a single point charge Q_0. This approximation works well as long as the distance D to the sample remains within several tip radii R, which is sufficient to cover the typical tip-sample distances in the apertureless near-field optical microscopy.

Finally, the value of Q_0 can be obtained by comparing Eq. 5.2 to the general expression for a monopole field in cgs units, $E_m = Q/R^2$, from which $Q_0 = W_0^2 E_s(0)$ follows. With the relation

$$E_s(0) = E(0) - E_0 = (\gamma_0 - 1)\, E_0, \tag{5.4}$$

Q_0 can also be expressed as

$$Q_0 = W_0^2\, (\gamma_0 - 1)\, E_0, \tag{5.5}$$

where $\gamma_0 = E(0)/E_0$ is the field enhancement at the point $(z, \rho) = (L, 0)$, i.e. on the intersection of the spheroid axis with the spheroid surface. The value of γ_0 can be derived by setting $D = 0$ in Eq. 5.1:

$$\gamma_0 = \frac{\frac{2F}{R} + \ln\frac{L-F}{L+F}}{2\frac{F\,(L-\varepsilon R)}{LR(\varepsilon-1)} - \ln\frac{L-F}{L+F}} + 1. \tag{5.6}$$

For $R \ll L$, Eqs. 5.6 and 5.3 simplify to

$$\gamma_0 = \frac{\frac{2L}{R} + \ln\frac{R}{4L}}{2\frac{F\,(L-\varepsilon R)}{LR(\varepsilon-1)} - \ln\frac{L-F}{L+F}} + 1, \text{ and} \tag{5.7}$$

$$W_0 = 1.31\, R, \tag{5.8}$$

respectively. Further simplifications of the denominator in Eq. 5.7 are not possible without producing large errors in Q_0 for some possible values of ε.

5.1.4 Spheroid in External Monopole Field

The next topic to be considered is the influence of the tip on the sample. As with the dipole model (Sect. 2.4), the method of images can be employed. If the tip is represented by a point charge Q_0 at a distance H *above* the surface, the sample response is equivalent to a point charge $Q_0' = -\beta Q_0$ at the same distance H *below* the surface (Fig. 5.2(b)). The

"reflection factor" β is known from electrostatics to be related to the dielectric constant of the sample as $\beta = (\varepsilon_s - 1)/(\varepsilon_s + 1)$. A proof that the electrostatic concept of the mirror image can be successfully applied over very short distances even with fields oscillating at optical frequencies is presented in Sect. 5.2 .

The response of the probing tip to the mirror charge $Q_0' = -\beta Q_0$ induced in the sample should be determined next. It is thereby particularly important to find out whether this response looks like the induction of another point charge Q_i close to the tip apex. This would justify one of the basic assumptions of the monopole model, sketched in Fig. 5.2.

Before attacking the more complex problem of a spheroid in the monopole field, a simpler case of a grounded conducting sphere will be presented for introduction. To this end, let us recall that the entire surface of any grounded perfectly conducting object always resides on the zero potential. The same obviously applies to our sphere in the presence of an external point charge Q_e. From standard textbooks it is known that the charge distribution on the surface of the sphere in this case produces an electric field equivalent to that of a certain point charge Q_i within the sphere [156]. The exact relation between the two charges is $Q_i = Q_e z_i/R$, with R being the sphere radius and $z_i = R^2/z_e$ the distance of the fictive "internal" point charge Q_i from the center of the sphere. The external charge Q_e is located at a distance z_e form the sphere center (Fig. 5.4(a)).

If the sphere is deformed along one axis to become a prolate spheroid, its response to the external charge Q_e is no more equivalent to a *point* charge. It has been shown that its response is instead equivalent to a certain *line* charge distribution along the line connecting the two spheroid foci[157]. This situation is illustrated in Fig. 5.4(b).

The hypothesis that the spheroid in the field of a point charge can be replaced by another point charge will now be examined utilizing results from [158] and [159]. These two references provide semi-analytical solutions to the problem of a grounded perfectly conducting spheroid in the presence of a point charge. To make use of these results, from this point on the derivation of the monopole model has to be restricted to perfectly conducting probe materials, too. On the other hand, the transition from grounded to isolated spheroids will be made before the final result is derived.

The distribution of line charge $q_i(z)$ required to maintain the surface of a grounded spheroid at zero potential in the presence on an external point charge Q_e is given by the following

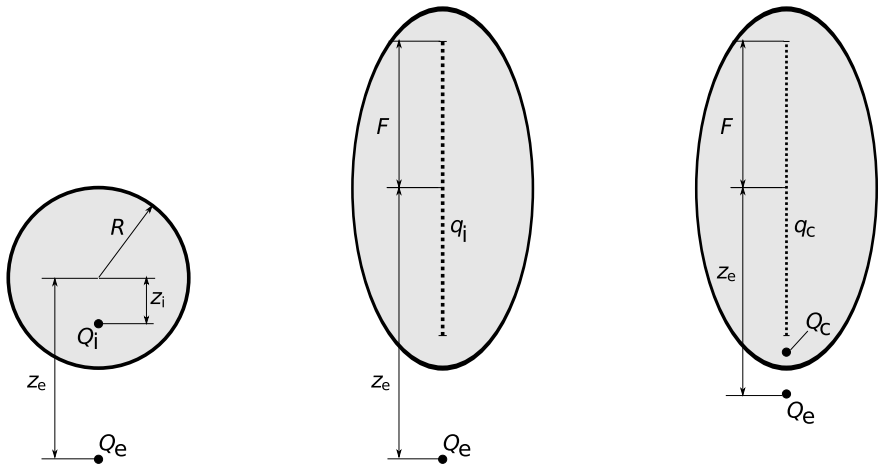


Figure 5.4: (a) Electric potential outside a grounded perfectly conducting *sphere* in the presence of an external charge Q_e is equivalent to the potential obtained by replacing the sphere by a point charge Q_i. (b) Electric potential outside a grounded perfectly conducting *spheroid* in the presence of an external charge Q_e is equivalent to the potential of a certain non-uniform line charge distribution q_i (cf. Fig. 5.5). (c) For short distances between the spheroid and the external charge, convergence problems are encountered in Eq. 5.9 while calculating the line charge q_i. The problems can be avoided by separating the equivalent charge into the point charge contribution Q_c and a reduced amount of line charge q_c.

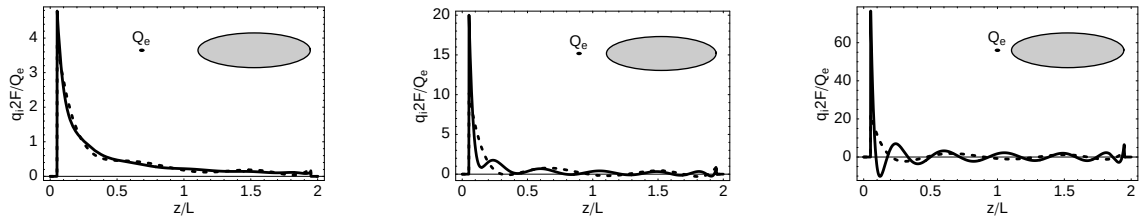
series[158]:

$$q_i(z) = -\frac{Q_e\,\Theta(F^2-z^2)}{2F}\sum_{n=0}^{\infty}(2n+1)\frac{\mathscr{L}_n((L+D)/F)\,P_n(L/F)}{\mathscr{L}_n(L/F)}P_n(z/F). \qquad (5.9)$$

The notation from Table 4.1 has been employed here with the addition of two new symbols: the unit step function Θ and the Legendre polynomials P_n. As indicated in Fig. 5.4, the origin of the z-coordinate in Eq. 5.9 is set to the spheroid center.

Three examples are given in Fig. 5.5 showing the (fictive) line charge distribution along the spheroid axis calculated using Eq. 5.9.

The convergence speed of the series in Eq. 5.9 can be estimated from Fig. 5.5 by observing the difference between the calculation with $n = 6$ and with $n = 12$ terms in the sum from Eq. 5.9. For large distances D, Fig. 5.5(a) reveals a rapid convergence, which becomes slower as the distance D is reduced in Fig. 5.5(b) and (c). For thin spheroids ($R \ll L$), Eq. 5.9 exhibits serious convergence problems with external charge distances below about



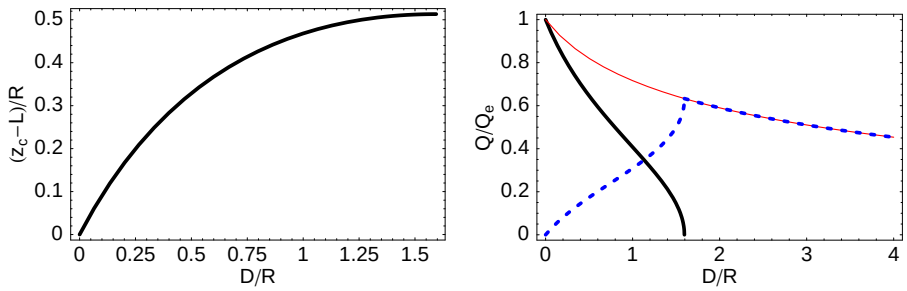
(a) (b) (c)

Figure 5.5: Line charge distribution q_i induced along the axis of an $r = 0.1$ spheroid as calculated from the first 6 terms (dashed line) and 12 terms (full line) in Eq. 5.9. The external point charge is located at: (a) $D = 4r$, (b) $D = 2r$, and (c) $D = r$.

$D = 3R/2$. For such distances, an alternative approach has been suggested in [158]. It involves the separation of the induced charge q_i into a point charge Q_c and a reduced line charge distribution q_c. This model is illustrated in Fig. 5.4(c). The point charge component Q_c in the alternative approach is located at

$$z_c = \frac{(L+D)(1+r) - 2\sqrt{r((1+D/L)^2 - f^2)}}{(1-r)}, \tag{5.10}$$

measured from the spheroid center. It is easy to show that z_c lies between the focus $z = F$ and the spheroid boundary $z = L$. This can also be seen in Fig. 5.6, where the position z_c is plotted together with the amount of charge at this position. The amount of remaining linear charge is also shown in Fig. 5.6(b) and demonstrates how the induced charge gradually changes it character from a point-like to line charge distribution. Its mean position also simultaneously shifts away from the spheroid end towards the spheroid center.



(a) (b)

Figure 5.6: (a) Position z_c of the induced point charge component according to Eq. 5.10. (b) Partition of the total induced charge Q_t (thin full line) into the point charge component Q_c (thick full line) and line charge component q_c (dashed line). All quantities are calculated for a spheroid with $r = 0.1$.

In conclusion, the charge induced in the spheroid only roughly resembles a point charge. Moreover, it changes both its position and its distribution depending on distance of the external charge. However, for short separations D between the spheroid and the external charge Q_e, the induced charge does indeed remain concentrated within roughly one curvature radius R from the spheroid end. Consulting the figures 5.5 and 5.6 again, we see that, on average, the highest charge concentration seems to be around the spheroid focus ($z = F$). Since for simplicity only one fixed position of the induced charge Q_i is allowed in the monopole model (Fig. 5.2), $z = F$ seems to be the best choice for this position. For $R \ll L$, the position of the focus is given by $z = F \approx L - R/2$, measured from the spheroid center, or $W_i \approx R/2$ from the spheroid end.

It should be emphasized once more that the reduction of charge induced in the spheroid by an external point charge Q_e to another point charge Q_i at fixed location inside the spheroid is a crude approximation. Its use is motivated exclusively by the huge simplification of the final solution it enables.

5.1.5 Amount of Induced Charge

To complete the approximation of the spheroid response to an external point charge Q_e by another monopole Q_i, the value of Q_i remains to be determined. This can be done by noting that Legendre polynomials $P_n(z/F)$ encountered in Eq. 5.9 are oscillatory and integrate to zero on the interval $[-1, 1]$ for all n except $n = 0$. This means that only the first term ($n = 0$) in the sum in Eq. 5.9 determines the total charge on the spheroid[159] and all subsequent terms just redistribute it along the spheroid axis. Since $P_0(x) = 1$, the net charge Q_t induced on a grounded perfectly conducting spheroid in the field of an external point charge Q_e is determined by Legendre functions $\mathscr{L}_0$ in Eq. 5.9:

$$Q_t = -Q_e \frac{\mathscr{L}_0(\,(L+D)/F\,)}{\mathscr{L}_0(\,L/F\,)} = -Q_e \frac{\ln \frac{L+F+D}{L-F+D}}{\ln \frac{L+F}{L-F}}. \tag{5.11}$$

As seen in Fig. 5.5, not all of the charge Q_t is concentrated close to the spheroid end. Since only the charge found in the proximity of the sample is considered relevant for the near-field interaction in the monopole model, the induced monopole Q_i can be assigned only a fraction of the total induced charge Q_t. An estimate of its value should be made bearing in mind that for describing the probe-sample interaction, the mirror image $Q_i' = -\beta Q_i$ of the monopole Q_i assumes the role of the external charge Q_e. If the probing tip is in

contact with the sample surface, only the charge density in the immediate neighborhood of the contact area is important for the interaction. Moving the tip away from the surface, a larger portion of its mirror image has to be considered for an accurate description of the electric field at the position of Q_i. Following this rationale, Fig. 5.7 shows the fraction of Q_t located within the distance $\Delta z = R + D$ from the spheroid end, calculated according to the model in [158]. Curves in Fig. 5.7 exhibit a similar behavior for various spheroid shapes. Moreover, they can be considered constant to a first approximation for distances D larger than $D_{min} = W_i = R/2$, the smallest distance permitted by the model. A reasonable estimate for this constant based on Fig. 5.7 seems to be around $g = 0.7 \pm 0.1$. This value will be used as the starting point for further refinement.

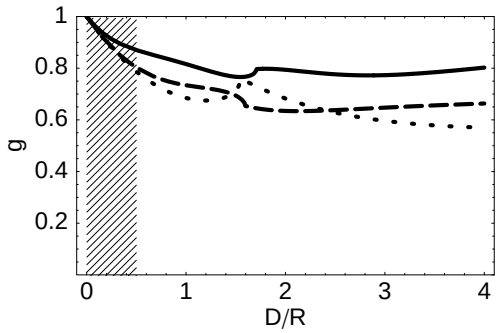


Figure 5.7: Fraction of the total induced charge Q_t found within the range $R + D$ from the tip end. Three different shape factor are shown: $r = 0.2$ (full line), $r = 0.1$ (dashed) and $r = 0.05$ (dotted). The shaded part of the plot from $D = 0$ to $D = 0.5R$ denotes external charge distances which do not appear in the monopole model.

One important refinement is the transition from a grounded spheroid considered so far to an isolated spheroid of the monopole model shown in Fig. 5.2. When the charge Q_t is induced in a grounded spheroid, the surplus charge -Q_t is assumed to flow out of the spheroid. Since the surplus charge cannot leave an *isolated* spheroid, it must distribute in such a way that the entire surface of the spheroid still remains at the same potential. This requirement is satisfied by an ellipsoid if every slice of thickness Δz contains an equal amount of charge $\Delta Q = -Q_t \Delta z / 2L$ [160]. This means the charge -Q_t is uniformly spread over the length of an isolated spheroid, and some of it resides within the range $\Delta z = R + D$ considered relevant for the near-field interaction in Fig. 5.7(b). To account for this charge, the factor g derived above has to be decreased to $g' = g - (R + D)/2L$. With this correction, the strength of the

induced monopole Q_i is equal to

$$Q_i = g' Q_t = -(g - \frac{R+D}{2L}) Q_e \frac{\ln \frac{L+F+D}{L-F+D}}{\ln \frac{L+F}{L-F}}. \tag{5.12}$$

Since the approximation $R \ll L$ has been introduced already in Sect. 5.1.3, we can expand $F = \sqrt{L-R}$ as $F = L - R/2$. This way Eq. 5.12 becomes

$$Q_i \approx -(g - \frac{R+D}{2L}) Q_e \frac{\ln \frac{2L-R/2+D}{R/2+D}}{\ln \frac{2L+R/2}{R/2}}. \tag{5.13}$$

Neglecting $R/2$ and D in comparison to $2L$, Eq. 5.14 is finally obtained:

$$Q_i \approx -(g - \frac{R+D}{2L}) Q_e \frac{\ln \frac{4L}{R+2D}}{\ln \frac{4L}{R}}. \tag{5.14}$$

5.1.6 Near-field Interaction

Eq. 5.14 represents the final form of the expression for Q_i. To evaluate it, the external charge Q_e in Eq. 5.14 and its distance D have to be specified now. In this regard, it should be noted that there actually exist *two* parts of the induced charge Q_i shown in Fig. 5.2 and again in Fig. 5.8 with more details. For the first part, $Q_{i,0}$, the external charge Q_e is the mirror image Q'_0 of the charge Q_0 induced by the illumination field. For the second part, $Q_{i,1}$, the external charge Q_e is the mirror image Q'_i of the monopole Q_i itself. The corresponding distances D_0 and D_1 are easy to derive from Fig. 5.8:

$$\begin{aligned} D_0 &= 2H + W_0, \quad \text{and} \\ D_1 &= 2H + W_i, \end{aligned} \tag{5.15}$$

where H denotes the distance between the probing tip and the sample surface and the position W_0 of the monopole Q_0 is given by Eq. 5.3.

Inserting Eqs. 5.15 into Eq. 5.14, the following relations between the image charges Q'_0 and Q'_i and the monopoles $Q_{i,0}$ and $Q_{i,1}$ they induce in the probe are obtained:

$$Q_{i,0} = - f_0 Q'_0 = -(g - \frac{2H + W_0 + R}{2L}) \frac{\ln \frac{4L}{4H+2W_0+R}}{\ln \frac{4L}{R}} Q'_0, \text{ and}$$

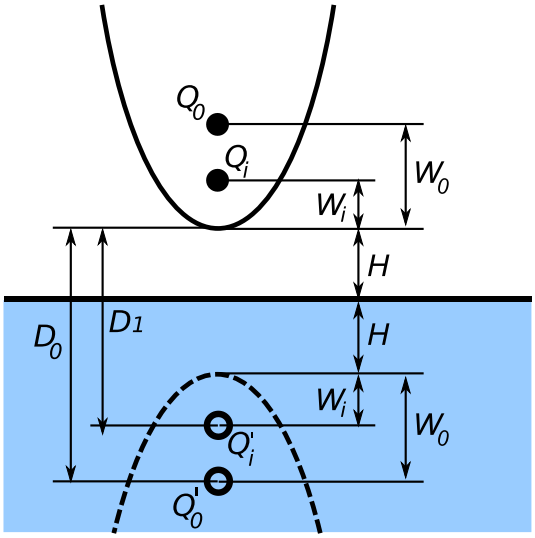


Figure 5.8: Charges participating in the near-field interaction between the probe and the sample, together with their positions according to the monopole model.

$$Q_{i,1} \;=\; -\;\; f_1 Q_{i,1} = \; -(g - \frac{2H + W_i + R}{2L}) \; \frac{\ln \frac{4L}{4H+2R}}{\ln \frac{4L}{R}} \; Q'_i. \tag{5.16}$$

The induced charges $Q_{i,0}$ and $Q_{i,1}$ should finally be added together to obtain the monopole Q_i,

$$Q_i = Q_{i,0} + Q_{i,1}, \tag{5.17}$$

which measures the strength of the near-field interaction between the probe and the sample.

Since the image charges Q'_0 and Q'_i can be expressed as

$$\begin{aligned} Q'_0 &= -\beta\, Q_0, \text{ and} \\ Q'_i &= -\beta\, Q_i, \end{aligned} \tag{5.18}$$

and the monopole Q_i comprises $Q_{i,1}$ which is proportional to the mirror image of Q_i itself, we have obtained a recursive definition for Q_i. Such a situation has been already encountered with the dipole model in Sect. 2.4, so the same approach of searching for a self-consistent solution can be utilized here. That way the value of Q_i is obtained by solving the equation

$$Q_i = \beta\,(f_0\, Q_0 + f_1\, Q_i). \tag{5.19}$$

The solution of Eq. 5.19 is the near-field-interaction-induced charge Q_i referred to in Fig. 5.2.

$$Q_i = \frac{\beta\, f_0}{1-\beta\, f_1}\, Q_0. \tag{5.20}$$

Beside the charge Q_i, the electric neutrality requires the existence of the opposite charge -Q_i somewhere on the spheroid. The dipole moment p_i of the charge distribution generated by Q_i and -Q_i oscillates with the driving field frequency and thus radiates light. To determine the dipole moment p_i, we recall the fact that the only distribution of charge -Q_i which does not break the equipotential property of the spheroid surface is the uniform distribution over the spheroid length. The symmetry of this distribution allows us to effectively replace it a single charge -Q_i located in the spheroid middle since the dipole moment is linear in the charge position. The average distance between Q_i and -Q_i is then $L - R/2 \approx L$, and the resulting dipole moment p_i equals

$$p_i = Q_i L = \frac{\beta\, f_0}{1-\beta\, f_1}\, Q_0 L. \tag{5.21}$$

The dipole moment p_i emits radiation which can be taken as the measure of the near-field interaction between the probe and the sample. In addition to p_i, there is also the dipole moment $p_0 = 2Q_0L$ which depends on the size and shape of the probe and on the illumination field strength, but is completely independent of the sample and its properties. This allows a dimensionless "near-field contrast factor" $\eta = p_i/p_0$ to be defined as an illumination-independent measure of the s-SNOM signal, comparable between different samples and different measurements:

$$\eta = \frac{p_i}{p_0} = \frac{1}{2}\,\frac{\beta\, f_0}{1-\beta\, f_1} \tag{5.22}$$

Inserting f_0 and f_1 from Eqs. 5.16 into Eq. 5.22 and setting $W_i = R/2$, we obtain

$$\eta = \frac{\beta\,(2Lg - 2H - W_0 - R)\,\ln\frac{4L}{4H+2W_0+R}}{4L\ln\frac{4L}{R} - \beta\,(4Lg - 4H - 3R)\,\ln\frac{4L}{4H+2R}} \tag{5.23}$$

Eq. 5.23 can be written in a simpler form by noting that all lengths scale with the spheroid half-length L. Substituting h, r, and w respectively for H/L, R/L, and W_0/L, Eq. 5.23 becomes

$$\eta = \frac{\beta\,(2g - 2h - w - r)\,\ln(h + w/2 + r/4)}{4\ln(r/4) - \beta\,(4g - 4h - 3r)\,\ln(h + r/2)}. \tag{5.24}$$

The parameter w in Eq. 5.24 is the position of the initial monopole Q_0. It can be taken to be equal to $w = r$ for simplicity. Alternatively, a slightly more accurate value can be determined with the aid of Eq. 5.8, yielding

$$w = \frac{W_0}{L} = \frac{1.31\,r}{1+2\,r}. \tag{5.25}$$

5.1.7 Effective Polarizability

The overall dipole moment p_{eff} of the charge distribution in the probe can be constructed by adding the contributions from the dipole p_0 induced by the external field and the dipole p_i induced by the near-field interaction:

$$p_{\text{eff}} = p_0 + p_i = 2\,Q_0\,L\,(1+\eta). \tag{5.26}$$

The value of Q_0 required for calculating p_{eff} is provided by Eq. 5.5: $Q_0 = W_0^2\,(\gamma_0 - 1)\,E_0$. In the limit of a perfectly conducting spheroid ($\varepsilon_{\text{probe}} \to -\infty$), the field enhancement factor γ_0 from Eq. 5.7 simplifies to

$$\gamma_0 = \frac{2\,L/R + \ln\frac{R}{4eL}}{\ln\frac{4L}{e^2R}} + 1, \tag{5.27}$$

or, equivalently:

$$\gamma_0 - 1 = \frac{2 + r\ln\frac{r}{4e}}{r\ln\frac{4}{e^2 r}}, \tag{5.28}$$

Finally, inserting Q_0 from Eq. 5.5 into Eq. 5.26, the effective polarizability α_{eff} of the spheroid interacting with the sample is easily obtained:

$$\alpha_{\text{eff}} = \frac{p_{\text{eff}}}{E_0} = 2\,(\gamma_0 - 1)\,w^2\,L^3\,(1+\eta) \tag{5.29}$$

Together with Eqs. 5.28, 5.25 and 5.24 which determine the quantities γ_0, w and η, Eq. 5.29 represents the complete expression for calculating the relative strength of the s-SNOM signal based on the monopole model. It can be directly compared to the effective polarizability obtained from the point dipole model, Eq. 2.10. Since the probe in the monopole model is assigned the polarizability of a prolate spheroid of length $2L$, it obviously provides a much better approximation to the real probes than the dipole model in which the probe assumes the polarizability of a small sphere of radius R, $R \ll L$.

Still, the model spheroid length is constrained to $2L \ll \lambda$ due to the electrostatic treatment of the problem. On the other hand, standard s-SNOM (AFM) probing tips usually have lengths of about 10 µm or more, which is comparable to or greater than λ for mid-IR and shorter illumination wavelengths. For this reason, a significant discrepancy between the polarizability of real probes and a small spheroid to which Eq. 5.29 applies should be expected. This problem will be addressed in Chapter 6.

Although the entire probing tip can radiate light, the largest part of a real probe is too far from the sample to give a significant contribution to the near-field interaction. The length $2L$ considered in the electrostatic model should therefore provide a satisfactory description of the near-field interaction. This permits the contrast factor η in the form of Eq. 5.24 to be used for prediction of true near-field optical contrasts in s-SNOM experiments.

Technically, the near-field part of the polarizability α_{eff} related to the dipole moment p_i and the contrast factor η is extracted from the overall emitted signal by the higher-harmonic demodulation method described in Sect. 2.3. This method suppresses the detection of light scattered due to the interaction-independent dipole moment p_0, also known as the s-SNOM background signal. A detailed analysis of this issue is presented in Chapter 6.

5.1.8 Parameters of Monopole Model

Eq. 5.23 contains all parameters which determine the near-field optical contrast between two materials within the monopole model framework. Each parameter was described at the point where it was introduced, but an overview of their meaning and permissible values is presented here for convenience.

The first quantity found in Eq. 5.23 is the electrostatic "reflection coefficient" β, defined as $\beta = (\varepsilon_s - 1)/(\varepsilon_s + 1)$, where ε_s is the complex dielectric constant of the sample. The quantity H stands for the probe-sample separation, i.e. the height of the probing tip above the sample. R is the radius of the spheroid apex curvature which can be obtained from an electron micrograph of the probing tip or from the topographic resolution it provides. For commercial probes, the radius R can also be found in data sheets provided by the manufacturers. According to the assumptions of the model, both H and R have to be much smaller than L, the length of the major semi-axis of the spheroid representing the probe.

The total length of the spheroid ($2L$) is equal to the probe length if the latter is much shorter than the wavelength λ. This condition is usually not satisfied in practice and the length L

has to be determined by other means. Thereby the constraint $2L \ll \lambda$ has to be observed, since otherwise the field enhancement, and consequently the interaction strength, would be overestimated. In Sect. 5.4, L will be determined as a free parameter determined such that it provides the best fit to the available experimental results. Unlike an ordinary fit parameter, it is not readjusted to each experimental data set separately. Rather, its value is determined only once and used unchanged in all subsequent calculations.

Finally, the least precisely defined parameter is the factor g, related to the proportion of the total induced charge Q_t that is relevant for the near-field interaction. In a simple approach used here, it was replaced by a constant estimated to about 0.7 ± 0.1 based on Fig. 5.7(b). In practice, some corrections can also be incorporated into the factor g on a phenomenological basis, such as the effect of using a probe with finite conductivity instead of a perfectly conducting spheroid. Further corrections may also include the radiation resistance which has been neglected so far. Given that the near-field coupling is essentially capacitive in its nature, the electric resistance due to finite conductivity and the radiation resistance cause a slight phase difference between the driving field and the response of the probe. This phase difference may be accounted for by using a complex factor g with a small imaginary component, as will be done in Sect. 5.4.

At this point it is still an open question how sensitive are the results obtained by the monopole model to the variation of its parameters. Due to the complex interplay between the parameters, the relative impact of each parameter on the final result depends to a large degree on the values of all other parameters and on the dielectric function of the sample. For this reason no simple answer can be provided here, but this issue remains an important topic to study in future publications.

5.1.9 Possible Improvements

To conclude, a new model for describing the interaction of a s-SNOM probing tip with the investigated sample has been derived in this section. The final expression, Eq. 5.29, is somewhat more complicated than the corresponding expression in the dipole model (Eq. 2.10), but it still represents a completely analytical closed form solution. A detailed derivation of Eq. 5.29 was presented here to give the justification for assumptions of the model and to indicate directions for possible future improvements. Among them, the provision for varying the position of the monopole Q_i with the probe-sample distance would probably make the biggest difference. Another possible improvement would be to repre-

sent the interacting part of the probe by several monopoles $Q_1..Q_n$ which would provide a better match to the line charge distribution of Eq. 5.9. This extension would also enable the proper treatment of the uniformly distributed surplus charge -Q_i, which is expected to give the interaction a slightly dipolar character. From electrodynamic calculations, the effective probe length L could be derived without resorting to a fit to experimental data. Finally, a significant advancement would be made by extending the model to dielectric probes based on Refs. [161] and [162], thereby enabling the prediction of signal obtained using polariton-resonant probes.

5.2 Mirror Image Radiation

The expressions 2.10 and 5.29 describing the near-field interaction account for the presence of the sample only through its influence on the probe polarization. However, the probe induces an oscillating charge distribution in the sample which obviously emits some radiation as well. It was thus proposed in the original form of the dipole model (Eq. 2.15) to treat the mirror dipole βp the same way as the tip dipole p. Such an approach was consistently avoided in this work based on the following argument against the equivalence of the two dipoles for emitting radiation: while the tip dipole p is assumed to be a real dipole, its mirror image $p' = \beta p$ is just a fictive construct to describe the field distribution above the sample surface in the *electrostatic limit*. Consequently, the rapidly evanescent fields are properly accounted for by such a construct, but the propagating waves are not since they reflect from the sample with a different reflection coefficient which also varies with the angle of incidence. Therefore, the electrostatic mirror image p' cannot be used to calculate the intensity of waves radiated from the sample surface.

The same conclusion can be reached on the energy conservation basis. Namely, $|\beta| > 1$ for all materials with negative dielectric constant ε (or more precisely, the real part of it, $\varepsilon' = \mathrm{re}(\varepsilon)$). This implies that upon reflection of the wave from the sample surface, more energy would be re-emitted than received in the first place, constituting a violation of the energy conservation principle. On the other hand, the reflection coefficient $|\beta| > 1$ does not contradict the energy conservation when applied to the evanescent waves because they do not transport energy away from the source.

The correct way to treat the radiation reflected by the sample will now be deduced building on the Fourier optics approach from Sect. 4. Just like any other field, the field of a (point-

like or extended) dipole can be expanded into plane waves. The electrostatic "reflection coefficient" $\beta = (\varepsilon_s - 1)/(\varepsilon_s + 1)$, Eq. 3.1, is expected to apply only to waves not significantly affected by the retardation. These can only be the highly confined, i.e. evanescent components, not detectable in the far field. From this reasoning it follows that the reflection coefficient of propagating waves must differ from that of evanescent waves, and it may also never exceed unity.

5.2.1 Propagating vs. Evanescent Waves

The difference in the reflection coefficient of waves propagating from the probe towards the detector and those mediating the near-field interaction will be now determined quantitatively. For this purpose, let us first consider a plane wave with the free-space wave vector $\mathbf{k_0}$. Let the wave intersect the sample surface, and let $\mathbf{k}_\rho$ be the projection of $\mathbf{k_0}$ onto the surface. The component of the wave vector $\mathbf{k_0}$ perpendicular to the surface is then given by $k_z = \sqrt{k_0^2 - k_\rho^2}$. It is important to note that the same relation holds even if $k_\rho > k_0$[163]. In that case k_z is imaginary, and the wave is evanescent. Provided the sample is non-magnetic ($\mu = 1$), the wave has a wave vector $k_0' = \sqrt{\varepsilon_s}\, k_0$ in the half-space below the sample surface. For the transverse magnetic (TM, or "p") polarization of the wave the reflection from the surface is determined by the following relation, valid both for real and complex k_z[163]:

$$r_p = \frac{\varepsilon_s k_z - k_z'}{\varepsilon_s k_z + k_z'}. \tag{5.30}$$

In analogy to k_z, the z-component of $\mathbf{k_0'}$ is equal to $k_z' = \sqrt{k_0'^2 - k_\rho^2} = \sqrt{\varepsilon_s k_0^2 - k_\rho^2}$. The continuity of the EM field component parallel to the surface assures that k_ρ is equal above and below the surface.

For the transverse electric (TE, or "s") polarization, a different reflection coefficient is obtained [163]. However, this result will not be needed for the present analysis because a dipole oscillating perpendicular to the surface emits only TM-polarized waves. The dominant polarization of an elongated optical antenna perpendicular to the surface is also TM.

By expanding k_z and k_z' in Eq. 5.30 and substituting $k_\rho = \kappa k_0$, we finally obtain

$$r_p(\kappa) = \frac{\varepsilon_s \sqrt{1 - \kappa^2} - \sqrt{\varepsilon_s - \kappa^2}}{\varepsilon_s \sqrt{1 - \kappa^2} + \sqrt{\varepsilon_s - \kappa^2}}. \tag{5.31}$$

There are two limiting cases in Eq. 5.31 worth attention. For $\kappa = 0$ we obtain $r_p(0) = (\sqrt{\varepsilon_s} - 1)/(\sqrt{\varepsilon_s} + 1)$, the well known Fresnel reflection coefficient for normally incident waves. Even for $k \neq 0$ we can be certain that $|r_p| \leq 1$ as long as $|\kappa| \leq 1$ since $\sqrt{\varepsilon_s} = n$ is nothing but the index of refraction whose real part is never negative (unless $\mu < 0$, but materials with $\mu \neq 1$ were explicitly excluded here). On the other side, letting $\kappa \to \infty$ we get $r_p \to (\varepsilon_s - 1)/(\varepsilon_s + 1) = \beta$, the electrostatic reflection coefficient. It was already noted that $|\beta| \geq 1$ for $\varepsilon' < 0$.

To summarize, we have seen that the reflection coefficient r_p assumes different values for waves with different spatial frequencies κ. Only waves characterized by $\kappa \leq 1$, i.e. $k_\rho \leq k_0$ propagate and can be detected in the far field. To those waves, the electrostatic reflection coefficient β does not apply because r_p approaches β only in the limit $\kappa \to \infty$. As a consequence, β can be employed for calculating the influence of the sample on the near-field interaction, but not for radiation of waves into free space.

5.2.2 Reflection of Evanescent Waves

Although r_p only tends to β for $k_\rho \to \infty$, in both the dipole and the monopole model it is assumed that all waves reflect with the same coefficient β, regardless of their spatial frequency. Since all waves have finite spatial frequency, this is obviously just an approximation. Fig. 5.9 can be used as a starting point to estimate the error made in Eqs. 2.5 and 5.18 by applying the same reflection factor β to all plane waves (evanescent or not). It shows the magnitude of the Fresnel coefficient r_p over a large range of spatial frequencies covering both propagating and evanescent waves. The limits $r_p(\kappa \to 0) = (n-1)/(n+1)$ and $r_p(\kappa \to \infty) = (\varepsilon_s - 1)/(\varepsilon_s + 1) = \beta$ are clearly observable in Fig. 5.9.

Of particular interest here is the range of evanescent wave vectors $k > k_0$ for which $r_p(\kappa) \approx \beta$. From Fig. 5.9 we can conclude this approximation is in fact excellent for $\kappa > 3$. This value should be compared to the actual spectrum of plane waves constituting the field close to the probe. If it is found that the major part of the spectral density is contained in the components with $\kappa > 3$, the approximation that all waves reflect with the factor β is justified.

Waves with larger κ decay exponentially faster, so the distance between the probe and the plane of observation must be carefully chosen to obtain meaningful results. In typical experimental conditions, the probing tip vibrates above the sample with an amplitude approximately equal to the tip radius R. This makes the average probe-sample distance equal

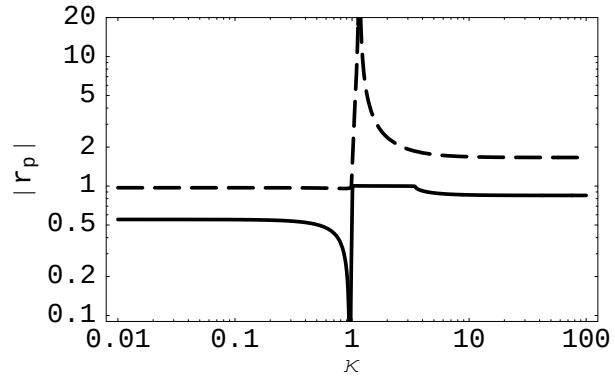


Figure 5.9: Magnitude of the Fresnel reflection coefficient r_p as a function of wave vector $\kappa = k_\rho / k_0$ calculated for a Si surface ($\varepsilon = 12 + 0.1i$, full line) and a SiC surface at about 11μm ($\varepsilon = -4 + 0.3i$, dashed line).

to R, and the average distance between the probe and its mirror image is $2R$. This is a very short distance compared to the probe length $2L$, and the assumption $R \ll L$ of the monopole model is satisfied. In Sect. 5.1.3 it was shown that in this case the field of the spheroid can be approximated by a field of point charge located on the spheroid axis at the distance R from the spheroid boundary. In the plane $z = 3R$ measured from the charge position, the full width $\Delta\rho$ of the resulting monopole field distribution is equal to $\Delta\rho = 6R$ at its half maximum. Invoking the relation $\Delta k_\rho \, \Delta\rho \geq 2\pi$, we get $\Delta k_\rho \geq 2\pi/6R$ or $\Delta\kappa \geq \lambda/6R$.

With the probe radius $R = 25\,$nm, and the wavelength $\lambda = 3\,\mu$m on the lower end of the mid-IR range we obtain $\Delta\kappa = 20$. This is significantly larger than the threshold $\kappa = 3$ derived from Fig. 5.9. The error made in Eqs. 5.18 where the image charges Q_0' and Q' were equated to -βQ_0 and -βQ is therefore negligible throughout the mid-IR wavelength range this thesis is focused on.

On the other side, $\lambda = 0.6\,\mu$m from the visible wavelength range yields a much lower value of only $\Delta\kappa = 4$ with the same probe radius $R = 25\,$nm. Obviously, Eqs. 5.18 become rather crude approximations with the visible light. A more detailed investigation of this issue is beyond the scope of this thesis, but it would generally have to be taken into account for a quantitative prediction of the s-SNOM signal at visible or near-IR wavelengths.

5.2.3 Reflection of Propagating Waves

The attention will now be transferred from the evanescent field which mediates the near-field interaction to the propagating field which enables the near-field interaction to be de-

tected and measured in the far field. The reflection of waves propagating towards the detector should be accounted for using the reflection coefficient r_p from Eq. 5.31 with an appropriate choice of κ ($\kappa \leq 1$) depending on the angle of observation. For a plane wave with declination θ from the surface normal (z-axis), κ is given by

$$\kappa = k_\rho / k_0 = \sin\theta. \tag{5.32}$$

If the sample is illuminated along the same path the scattered light is detected, the reflection coefficient of the incoming and emitted waves is the same. Owing to two reflections, the field scattered by the probe is then effectively increased by a factor $(1 + c\,r_p(\kappa))^2$ where the complex coefficient c accounts for the possible path difference between the direct and reflected radiation.

The situation may become more complex if an asymmetry between the illumination and collection paths exists. Two different reflection coefficients r_p might then be required. Additionally, if the illumination and collection are performed via an objective with a large numerical aperture (N_A), the plane-wave approximation might become inadequate. In particular, the contribution of the direct and reflected wave should be weighted differently and a proper integration over all angles covered by the objective might be required as well.

An important consequence of the factor $(1 + c\,r_p(\kappa))^2$ can be demonstrated by setting $\kappa = 1$ ($\theta = 90$, grazing incidence) in Eq. 5.31. We obtain $r_p = -1$, implying that the signal level is significantly reduced for illumination and observation under large angles θ. However, it would not be correct to conclude that opposite extreme ($\theta \approx 0$, normal incidence) is preferable. The radiation pattern of a dipole oriented along the z-axis has a minimum (zero) along the same axis ($\theta = 0$). Invoking the reciprocity principle, the same factor should be applied to the absorption and to the emission of radiation, adding a factor $\kappa^2 = \sin^2\theta$ to the scattered field. The combined effect of $(1 + c\,r_p)^2\kappa^2$ factor with $c = 1$ is plotted in Fig. 5.10, clearly showing that angles close to the surface normal ($\theta \lesssim 10$) should be avoided because of poor detection efficiency.

According to Fig. 5.10, incidence angles θ between about 30° and 70° should be chosen for optimal measurement conditions. The (soft) upper limit of about 70° is suggested not only because of the reduced detection efficiency on some materials, but also because of large differences in signal strength due to far-field reflection coefficient r_p which might obscure the near-field effects.

It should be noted that due to the way they were derived, the results form Fig. 5.10 apply

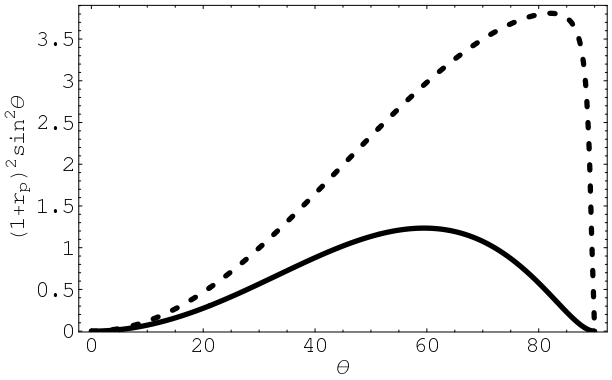


Figure 5.10: Factor $(1+r_p(\theta))^2 \sin\theta^2$ as a function of the normal angle of incidence $\theta = \arcsin(\kappa)$, calculated for a silicon surface (full line) and a gold surface (dashed line).

only to a small dipole-like probe with length $L \ll \lambda$ perpendicular to the sample surface. If the probe is inclined from the surface normal, the range of preferred incidence angles θ extends more towards the normal incidence ($\theta = 0$). Probes with lengths larger than the wavelength ($L > \lambda$) may exhibit several radiation lobes with maxima occurring closer to the surface normal [164], thereby also making smaller incidence angles θ acceptable.

5.2.4 Scattering coefficient

The results derived in Sect. 5.2.3 call for an extension to the effective polarizability α_{eff} from Eq. 5.29 in order to include the contribution of the reflection from the sample to the radiated field. Two corrections are required: the illumination field E_0 is increased by a factor $1 + c\,r_p$, and the mirror dipole moment $p'_{\text{eff}} = c\,r_p\,p_{\text{eff}}$ adds to the field directly scattered by the probe. These two effects do not directly modify the effective polarizability α_{eff}, but they do influence the scattered field. For this reason they will be assigned to the scattering contrast $\sigma = s e^{i\varphi}$, defined in Sect. 2.4 as

$$\sigma = E_s / E_0, \tag{5.33}$$

i.e. as the ratio of the scattered to the incident field strength, both measured at some fixed point in space. According to its definition, σ must be proportional to the total illumination field $E_0(1 + c\,r_p)$ and to the probe dipole moment p_{eff} to which its mirror image p'_{eff} should

be added:

$$\begin{aligned}\sigma &= f_\sigma (1+c r_p)^2 \alpha_{\text{eff}} \\ &= f_\sigma (1+c r_p)^2 (\gamma_0 - 1) w^2 L^3 (1+\eta). \end{aligned} \tag{5.34}$$

The reflection factor $r_p = r_p(\kappa)$ is defined by Eq. 5.31, and its argument κ is equal to $\kappa = \sin\theta$ where θ is the normal angle of light incidence on the sample. The coefficient c accounts for the difference between the waves directly incident on (emitted from) the probe and those reflected from the sample. Assuming plane wave illumination, the coefficient c should contain only the phase retardation by $\Delta\varphi = L k_0 \cos\theta$, i.e. $c = \exp(-i\Delta\varphi)$. Finally, f_σ is a proportionality constant dependent on the exact experimental parameters, including the observation point, illumination and detection angle with respect to the probe axis, and numerical aperture of the focusing objective. Its exact value is not important since it cancels when the relative contrasts σ_A/σ_B of two materials A and B are calculated.

It is important to note that there are two completely different length scales involved in Eq. 5.34 over which the optical constants of the specimen should be considered. The first of them is related to the reflection factor β contained in the near-field contrast η (Eq. 5.24). The area of the sample surface relevant for β is determined by field confinement below the probing tip. From Chapter 4 and Sect. 5.2.2 it follows that the confinement is always on the scale of the tip radius R, although its precise extent depends on the sample distance H from the probing tip.

On the other hand, for the reflection factor of propagating waves (r_p), the entire area of the sample in focus should in principle be considered. Even in the best case, the smallest possible diameter of this area is still on the wavelength (λ) scale. For this reason, the diffraction effects due to large structures present on the sample surface can be observed in s-SNOM images. This issue complicates the interpretation of s-SNOM images and should be considered in quantitative analyses of experimental results.

As an illustrative example, the scattering signal measurement at the boundary of two large surfaces with different dielectric constants ε_A and ε_B is presented in Fig. 5.11. Crossing the boundary A-B by a distance small compared to the probing tip length, the material directly below the tip changes, but, to a first approximation, the illumination of the probe remains constant. Consequently, β should be calculated using different values of ε_s in the regions A and B, whereas r_p should be calculated with roughly the same value of ε_s in points 1

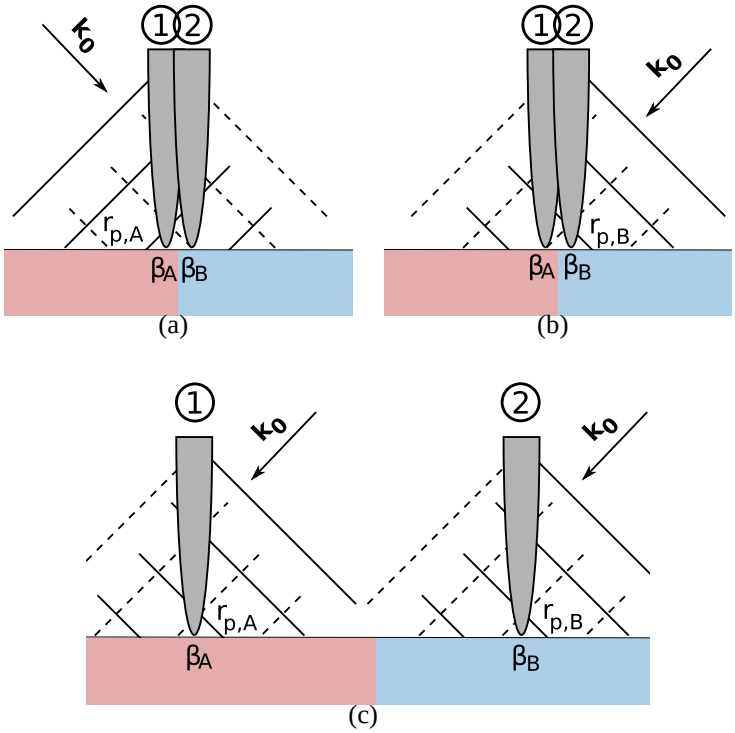


Figure 5.11: Choosing the far-field reflection coefficient r_p: (a) Probe close to the material boundary, illumination from side 1: $r_{p,1} \approx r_{p,2} \approx r_{p,\mathrm{A}}$, (b) probe close to the material boundary, illumination from side 2: $r_{p,1} \approx r_{p,2} \approx r_{p,\mathrm{B}}$, and (c) probe far from the material boundary: $r_{p,1} = r_{p,\mathrm{A}}$, $r_{p,2} = r_{p,\mathrm{B}}$. For simplicity, the diffraction of light by the probe and the material boundary has been ignored.

and 2 shown in Fig. 5.11 (a). The correct value of r_p thereby depends on the illumination direction, as illustrated in Fig. 5.11 (a) and (b). On the other hand, at distances large from the material boundary compared to the probing tip length, both β and r_p assume different values at points 1 and 2 in Fig. 5.11(c) since they should be calculated using ε_{A} in case 1 and ε_{B} in case 2. Finally, for the intermediate distances comparable to the tip length, oscillations in the recorded signal are expected as a consequence of the light diffraction on the boundary between the two surfaces (cf. e.g. Fig. 6.2 in [165]).

5.3 Anisotropy

So far it has been assumed that all materials are isotropic, i.e. that light propagates through each material with a speed independent of polarization and propagation direction. In practice, many materials interesting from the fundamental or technical point of view are

anisotropic. This includes SiC, the material known to exhibit a near-field resonance in the mid-IR, and also used as a model material throughout this thesis. The effects of anisotropy thus require a brief consideration here.

In general, anisotropy makes theoretical predictions of the near-field interaction of a s-SNOM probe with the material considerably more difficult. To reduce the complexity, the analysis will be confined to uniaxial crystals, i.e. crystals with the property that electric field oscillations along two of the three orthogonal axes are characterized by the same dielectric function, whereas the dielectric function along the third axis (the crystallographic *c-axis*) has a different value.

The Fresnel reflection coefficient of p-polarized waves given by Eq. 5.31 does not correctly describe the reflection from anisotropic crystals. For a uniaxial crystal cut perpendicular to the c-axis, the reflection coefficient should be modified as follows[166]:

$$r_{p,\perp} = \frac{\sqrt{\varepsilon_\perp \varepsilon_\parallel (1-\kappa^2)} - \sqrt{\varepsilon_\parallel - \kappa^2}}{\sqrt{\varepsilon_\perp \varepsilon_\parallel (1-\kappa^2)} + \sqrt{\varepsilon_\parallel - \kappa^2}}, \tag{5.35}$$

where $\varepsilon_\perp$ and $\varepsilon_\parallel$ are the dielectric constants perpendicular an parallel to the c-axis, respectively, and $\kappa = k_\rho / k_0$ is the ratio of the spatial frequency k_ρ in the sample surface plane to the wave vector $k_0 = \omega/c$ of a plane wave. If the crystal is cut parallel to its c-axis and the c-axis lies in the plane of incidence, the reflection factor is given by

$$r_{p,\parallel} = \frac{\sqrt{\varepsilon_\perp \varepsilon_\parallel (1-\kappa^2)} - \sqrt{\varepsilon_\perp - \kappa^2}}{\sqrt{\varepsilon_\perp \varepsilon_\parallel (1-\kappa^2)} + \sqrt{\varepsilon_\perp - \kappa^2}}. \tag{5.36}$$

In the quasi-electrostatic limit $\kappa \to \infty$, both $r_{p,\perp}$ and $r_{p,\parallel}$ tend to the same value

$$\beta_\parallel = \frac{\sqrt{\varepsilon_\perp \varepsilon_\parallel} - 1}{\sqrt{\varepsilon_\perp \varepsilon_\parallel} + 1}. \tag{5.37}$$

If a crystal is cut perpendicular to c-axis, the reflection factor β from Eq. 5.37 applies to all electric field polarization directions in the sample surface plane. For a crystal cut parallel to the c-axis, such as the sample whose spectra are shown in Sect. 5.4, the reflection factor depends on the direction of electric field oscillations relative to the c-axis. In particular, the

waves with the electric field perpendicular to the c-axis reflect with the factor

$$\beta_{\perp} = \frac{\varepsilon_{\perp} - 1}{\varepsilon_{\perp} + 1}, \tag{5.38}$$

and the waves with the horizontal component of the electric field parallel to the c-axis reflect with the factor $\beta_{\parallel}$. Due to the symmetry of the monopole field, all angles in the surface plane contribute equally to the near-field interaction. The coefficient of evanescent waves reflection from the surface of a crystal cut parallel to the c-axis can be thus approximated by assuming equal importance of both orientations,

$$\beta'_{\perp} = (\beta_{\parallel} + \beta_{\perp})/2. \tag{5.39}$$

However, this approach seems to provide only an approximate description of how the field of a point charge reflects off an anisotropic crystal according to a more complete treatment of the same problem in Refs. [167] and [168]. There it was shown that the response of an anisotropic crystal to the presence of a point charge above its surface can be described by the usual mirror image of the point charge and a certain surface charge distribution within the crystal. Furthermore, the additional surface charge effectively reduces to a line charge in the case of weak anisotropy. On this basis, the point dipole model of the near-field interaction (Sect. 2.4) was extended to anisotropic samples in [169]. However, the dipole model has already been shown in Sect. 2.5.1 to be incapable of a quantitative prediction of the near-field interaction, so the formalism from [169] cannot be used here. On the other hand, the monopole model has not been yet extended to anisotropic media in a way analogous to the approach in [169]. For this reason, the anisotropy will be accounted for only partially through the near-field reflection factor β given by either Eq. 5.37 or 5.39, depending on the cut direction.

5.4 Monopole vs. Dipole Model

The obvious question at this point is whether the monopole model in the form of Eq. 5.29 offers an advantage over the dipole model, Eq. 2.10. Only the comparison to the experimental results can provide the answer. For this comparison to be possible, the vibration of the s-SNOM probe and the subsequent scattering signal demodulation at a higher harmonic of the vibration frequency have to be taken into account.

5.4.1 Higher-harmonic Demodulation

As explained in Sect. 2.3, the light scattered by a vibrating s-SNOM probe contains a lot of background scattering, not related to the near-field interaction. In the monopole model, the background scattering is attributed to the directly induced dipole p_0, and the near-field signal to the interaction-induced dipole p_i in Eq. 5.26. Please note that a more complete model of the background scattering and its suppression is derived in Chapter 6, whereas the present section provides only the basics necessary to understand the experimental technique employed for the near-field signal detection.

It was already mentioned in Sect. 2.3.4 that by demodulating the signal at higher harmonics of the probe vibration frequency Ω, the pure near-field signal can be extracted from the total scattered field. To see how the higher harmonic demodulation technique [56, 57, 170] succeeds in extracting the near-field signal, the n-th harmonic σ_n of the scattering coefficient σ from Eq. 5.34 will be calculated next. Mathematically, σ_n is equal to the n-th Fourier series coefficient of σ with respect to the probe vibration frequency Ω, i.e. calculated over the fundamental period $T = 2\pi/\Omega$:

$$\sigma_n = \frac{1}{T}\int_{-T/2}^{T/2} \sigma(t)\, e^{-in\Omega t}\, dt \tag{5.40}$$

The dependence of σ on the time variable t comes from the varying height $h = H/L$ as a consequence of the s-SNOM probe vibration. Denoting the amplitude of the vibration by $A = aL$, we get

$$h(t) = \frac{A(1+\cos\Omega t)}{L} = a(1+\cos\Omega t). \tag{5.41}$$

Since only the near-field contrast factor η depends on the distance in Eq. 5.34, Eq. 5.40 can be written as

$$\sigma_n = k_\sigma (1+cr_p)^2 (\gamma_0 - 1) w^2 L^3 \frac{1}{T}\int_{-T/2}^{T/2} (1+\eta(t))\, e^{-in\Omega t}\, dt. \tag{5.42}$$

For $n > 0$ the constant background part vanishes and the above expression reduces to

$$\sigma_{n>0} = k_\sigma (1+cr_p)^2 (\gamma_0 - 1) w^2 L^3 \frac{1}{T}\int_{-T/2}^{T/2} \eta(t)\, e^{-in\Omega t}\, dt. \tag{5.43}$$

In Chapter 6 it will be shown that due to slight variations in the coefficient c, the harmonic index n larger than $n = 1$ has to be chosen in order for the background term to be really

negligible. For this reason Eq. 5.43 will be compared to the experiment only for $n \geq 2$.

The calculations can be further simplified by noting that only relative contrasts are measured in s-SNOM, expressed through the ratio of scattering coefficients σ_n resulting from the interaction of the probe with different materials. This enables all constant factors in Eq. 5.43 to be neglected since they cancel when the ratio $\sigma_{n,A}/\sigma_{n,B}$ is evaluated for two materials A and B. This leaves us with the ratio $\sigma_{n,A}/\sigma_{n,B} = \eta'_{n,A}/\eta'_{n,B}$, where η'_n is defined as

$$\eta'_n = (1 + c\, r_p)^2 \frac{1}{T} \int_{-T/2}^{T/2} \eta(t)\, e^{-in\Omega t}\, dt. \tag{5.44}$$

Due to the complicated form of the function η given by Eq. 5.24, the integral in 5.44 cannot be evaluated analytically. For further analysis we therefore have to rely on numerical integration. This is not a demanding task because the integrand $\eta(t)\, e^{-in\Omega t}$ is well behaved for all physically sensible values of parameters consistent with the assumptions $r = R/L \ll 1$ and $h = H/L \ll 1$ with which the monopole model was derived.

There is nevertheless one possible issue that has to be discussed. The denominator of η, $4\ln(r/4) - \beta\,(4g - 4h - 3r)\ln(h + r/2)$ contains a difference between two terms that can in principle become equal to each other. However, they could only cancel exactly if the reflection coefficient β would be a purely real number. This is never the case in practice because all materials exhibit at least some amount of damping which makes their dielectric function ε, and therewith the factor $\beta = (\varepsilon - 1)/(\varepsilon + 1)$, a complex number. It has also been argued in Sect. 5.1.8 that by including the effects of finite conductivity of the probe and the radiation resistance into the factor g, it becomes a complex number and guarantees the finite values of the integrand even for purely real β.

By inserting the appropriate dielectric function ε for different materials into Eq. 5.44, we can try to predict the near-field optical contrasts between them. The dielectric function thereby enters Eq. 5.44 through two different reflection coefficients: β, the reflection coefficient of evanescent waves, and r_p, the reflection coefficient of propagating waves. It was already argued in Sect. 5.2.4 that ε for calculating β may be different from ε used in r_p. Dealing with e.g. nano-composite materials whose structure is homogeneous on the scale of the probe length, but exhibits structural variations on a smaller scale, we have a situation where the dielectric function entering β varies, but the value of r_p remains unchanged.

Finally, if the value of r_p does not change over the area of interest, it is sufficient to calculate

the near-field contrast factor

$$\eta_n = \frac{1}{T}\int_{-T/2}^{T/2} \eta(t)\, e^{-in\Omega t}\, dt \tag{5.45}$$

to obtain the correct scattering signal ratios since $\sigma_{n,A}/\sigma_{n,B} = \eta_{n,A}/\eta_{n,B}$ in this case.

5.4.2 Material Contrast

The near-field contrast factor η as defined by Eq. 5.24 requires two parameters, g and L, to be specified before it can be evaluated. These two parameters are not directly measurable in an experiment and the theoretical considerations gave only some estimates of their values. In particular, the real part of the factor g was found in Sect. 5.1.5 to be $\mathrm{re}(g) \approx 0.7$ and for the imaginary part of the constant g the constraint $\mathrm{im}(g) \ll 1$ has been established. It is also known that the effective probe length $2L$ must be much smaller than the wavelength, i.e. $2L \ll \lambda$. In the absence of other means to obtain g and L, they were determined by searching for values which are in a good agreement with the available experimental data presented in the remainder of this chapter, including the material contrast, approach curves and near-field spectra of phonon-polariton resonant samples. The best values found are $g = 0.7e^{0.06i}$ and $L = 300\,\mathrm{nm}$. It is thereby important to note that the same values of these parameters will be subsequently used for all comparisons to the measured data and not adjusted to each experiment separately.

As the first simple example, the near-field contrast between gold and silicon will be calculated from Eq. 5.45. The required dielectric constants in the mid-infrared frequency range around $\lambda = 10\mu\mathrm{m}$ are $\varepsilon_{Au} \approx (-5+i)\,10^3$ [171] and $\varepsilon_{Si} \approx 12 + 0.1i$[172]. This gives $\beta_{\mathrm{Au}} \approx 1$ and $\beta_{\mathrm{Si}} \approx 0.85$. The known parameters of the experiment are the measured vibration amplitude $A = 16\,\mathrm{nm}$ and the tip radius $R \approx 25\,\mathrm{nm}$, obtained from the manufacturer data sheet (Nanosensors, model PPP-NCHPt). The sample was prepared by evaporating a 50 nm thick Au layer on top of a Si crystal and a sharp transition between Au and Si was produced by scratching away the Au film using sharp tweezers. The s-SNOM signal obtained by scanning the probe perpendicular to the Au-Si boundary is shown in Fig. 5.12. It can be immediately noticed that the optical signal-to-noise ratio is significantly lower on the Au-side than on the Si-side in Fig. 5.12. This can be explained by a much higher roughness of the evaporated Au film compared to the polished Si surface. In particular, the roughness of the Au film was about 5 nm, whereas that of the Si surface was about 0.7 nm

according to the simultaneously recorded AFM topography signal.

Apart from the above effect, the behavior of the optical signal in Fig. 5.12 can be described by two main features: a slow but steady increase from the Si-side towards the Au-side, and a larger, abrupt increase at the Si-Au boundary. The steady increase in the optical contrast with the distance from the boundary corresponds to the continuous transition form the arrangement illustrated in Fig. 5.11(b) to the one in Fig. 5.11(c). In the former case the probe illumination is almost the same on both Si and Au, whereas in the latter case the illumination of the probe is stronger on the Au side than on Si because of the larger far-field reflection coefficient r_p. In Sect. 5.2.4, this part of the scattering coefficient was explicitly introduced as a separate factor, independent of the near-field contrast. Consequently, the near-field contrast between Si and Au should in principle be determined from the abrupt change in the s-SNOM signal at the boundary between the two materials. However, there exist two problems which prevent the direct readout of the signal levels at the boundary. The first of them is a topography artifact which appears when the probing tip starts climbing onto the Au film and partially loses contact with the surface. This causes a ≈ 50 nm wide notch in the optical signal that must be avoided in the analysis. The second effect is related to the interaction of the probe with the Au film edge which produces its own near fields sensed by the probe and also creates a larger interaction area than a flat surface alone. In Fig. 5.12 this effect manifests as an offset to the near-field signal reaching its peak value of $\approx 8\%$ of the signal level at the boundary and decaying to negligible values over a distance of about 150 nm.

To investigate how much do the above issues influence the experimental determination of the pure near-field contrast, three different methods are employed to measure the contrast between Si and Au in Fig. 5.12. The results obtained this way are compared to each other and to the theoretical predictions In Table 5.1. The first method (1) uses the average near-field signal level between 150 and 200 nm from the edge on both sides (full lines in Fig. 5.12) to determine the near-field contrast between Si and Au. The said distance was chosen as the best compromise between the long-range illumination effect favoring shorter distances from the edge and the short-range artifacts favoring larger distances. Still, the variation in the illumination over the average 350 nm separation between the measurement points on the Si and Au side may lead to an underestimated ratio $|\eta_{n,\mathrm{Si}}/\eta_{n,\mathrm{Au}}|$ by up to 4%. The second method (2) aims at correcting this problem by extrapolating the linear part of the near-field signal trace on each side to the point $x = 0$ where the contrast between

Si and Au is calculated. This procedure, depicted by the dashed lines in Fig. 5.12, is expected to yield the most accurate results for the near-field contrast because it is immune to both the long-range illumination effects and the short-range artifacts described in preceding paragraph. Finally, the third method (3) avoiding the illumination effects, but not the edge artifacts is also presented in Table 5.1 for completeness. It uses the values closest to the edge on both sides, marked by the dotted lines in Fig. 5.12. Although the values can be read very precisely with this method, its accuracy is the worst of all three methods because of the edge effects and the noise which can significantly alter the results since no averaging is applied.

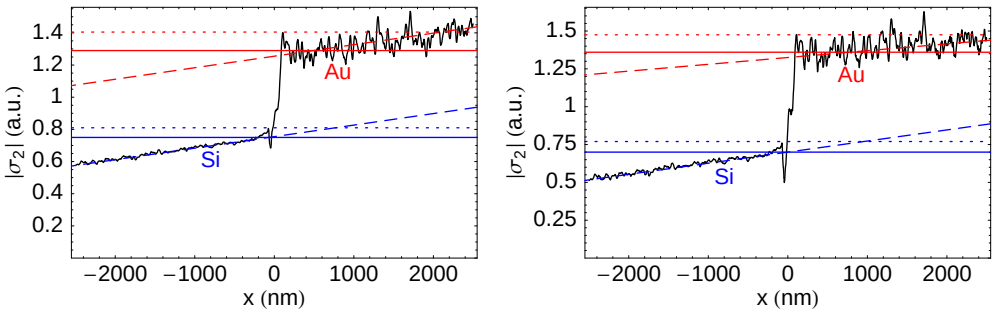


Figure 5.12: Line scan of the transition area between large flat Au and Si surfaces. Shown is the amplitude of the detected scattering signal demodulated at the (a) second, and (b) third harmonic of the probe vibration frequency.

The comparison of the experimental results in Table 5.1 indicates that all three methods yield almost the same results, each differing from the others by less than 4%. This value is comparable to the estimated measurement errors of $2-6\%$, which means that any of the three methods may be used in practice if an error of up to $\approx 5\%$ can be tolerated. Further comparison of the experimental results to the theoretical predictions for the Si/Au near-field optical contrast demonstrates that the monopole model is very successful in predicting the measured near-field contrast for both the second and the third harmonic of the detected scattering signal. Furthermore, the predicted values lie within the bounds of the experimental error of the most accurate measurement (2). The dipole model, on the other hand, overestimates the Si/Au near-field contrast by more than 25%.

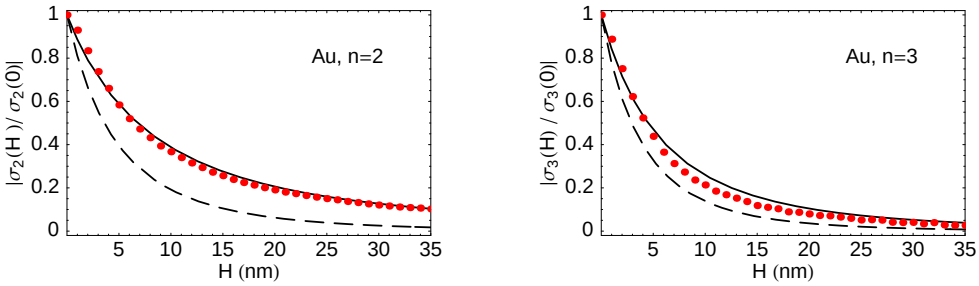
	$\lvert\sigma_{2,\mathrm{Si}}/\sigma_{2,\mathrm{Au}}\rvert$	$\lvert\sigma_{3,\mathrm{Si}}/\sigma_{3,\mathrm{Au}}\rvert$
Experiment (1)	0.59 ± 0.02	0.52 ± 0.02
Experiment (2)	0.61 ± 0.01	0.54 ± 0.01
Experiment (3)	0.59 ± 0.03	0.52 ± 0.03
Monopole model	0.61	0.55
Dipole model	0.79	0.77

Table 5.1: Near-field optical contrast between Si and Au measured close to the material boundary. The experimental values denoted (1), (2) and (3) were obtained from the full, dashed and dotted line pairs at $x = 0$ in Fig. 5.12, respectively.

The superiority of the monopole model over the dipole model could have been anticipated in this example because the monopole model contains two parameters which were adjusted to provide the best fit to the experimental data. A hypothetical introduction of free parameters into the dipole model would surely provide just as good agreement to the measured Si/Au contrast as the monopole model does. As shown in [75], even the distance-dependence of the near-field signal can be fitted with an extended version of the dipole model in which the dipole assumes the polarizability and the center position of a spheroid instead of a sphere. However, such a model results in an even worse fit to the near-field spectra of resonant samples due to the reduced near-field coupling strength caused by the shift of the dipole away from the surface. Even if two parameters are introduced into the dipole model through the adjustable position and polarizability of the dipole, it still seems to be capable only of improving one aspect of its performance at the expense of another one. This may be the reason why no extension to the point dipole model has been reported so far capable of simultaneously predicting data from all kinds of experiments with the same set of parameters. On the other hand, the monopole model already succeeds at this task, as will be shown next with two more challenging examples.

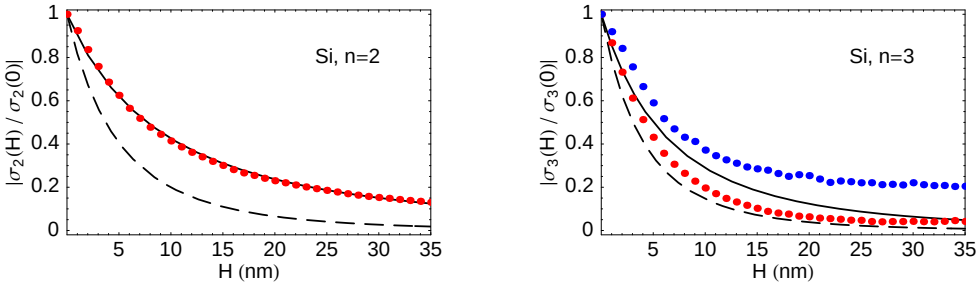
A more rigorous and thus more interesting test for the theoretical models is the near-field signal behavior as a function of the tip-sample distance. This kind of plot is usually called an approach curve, although a more proper name would be a retract curve. This is because such curves are experimentally obtained by switching off the closed-loop distance control of the AFM and moving the sample away from the probe. The absence of precise distance regulation makes the approach curves more noisy than the signal in the normal operation mode. Additionally, the near-field signal level at significant probe-sample distances is very weak, so that averaging over a larger number of approach curves is often performed to obtain more reliable results.

The second and third harmonic approach curves obtained as the average value of 100 single approach curves above a gold surface are presented in Fig. 5.13. The measurements were carried out with a probing tip (Nanosensors, PPP-NCHPt) which radius $R = 20\,\text{nm}$ was estimated from the topographical resolution, and the measured vibration amplitude was $A = 18\,\text{nm}$. The theoretical values were obtained from the monopole model using the same values of parameters g and L as above.



(a) (b)

Figure 5.13: Approach curves on a gold surface demodulated at (a) second harmonic ($n = 2$) and (b) third harmonic ($n = 3$). Shown are the experimentally obtained values (dots), and predictions by the monopole model (full line) and dipole model (dashed line). The values are normalized to the signal at $H = 0$



(a) (b)

Figure 5.14: Approach curves on a silicon surface demodulated at (a) second harmonic ($n = 2$) and (b) third harmonic ($n = 3$). Shown are the experimentally obtained values (dots), and predictions by the monopole model (full line) and dipole model (dashed line). The values are normalized to the signal at $H = 0$. The lower experimental approach curve (red dots) in part (b) was obtained by complex averaging of raw data, whereas the absolute vales were averaged for the upper curve (blue dots).

The agreement between the observed second harmonic ($n = 2$) signal in Fig. 5.13(a) and the prediction by the monopole model is almost exact. This is an indication that the distance

dependence in the monopole model is indeed a good approximation of the true tip-sample interaction. On the other hand, the dipole model predicts a much faster decay of the signal than actually observed. This is not surprising since the dipole field decay $1/z^3$ is much faster than monopole one, $1/z^2$. Similar comments can be made about the third harmonic in Fig. 5.13(b) as well.

The same kind of measurement repeated on a silicon surface is shown Fig. 5.14. The second harmonic signal on Si again shows excellent agreement with the monopole model prediction, and the dipole model exhibits an even larger discrepancy than above the Au surface. Unfortunately, a meaningful comparison between the experimental and theoretical approach curves at $n = 3$ on Si is not possible because of a large amount of complex noise contained in the recorded optical signal. Due to this noise, different results are obtained depending on whether the absolute value is taken before or after averaging the measurement results, as shown in Fig. 5.14(b). Assuming that the true third-harmonic approach curve would lie between the two experimental curves shown, it would be again very close to the monopole model prediction. However, further investigations are necessary to decide whether this is indeed the case.

Altogether, it seems reasonable to conclude that the monopole model represents a significant improvement over the dipole model concerning the distance dependence of the near-field interaction.

5.4.3 Near-field Spectra of Resonant Samples

The denominator in the Eq. 5.24 will be now examined again. It was already mentioned that the expression $4\ln(r/4) - \beta\,(4g - 4h - 3r)\ln(h + r/2)$ cannot be equal to zero because of the imaginary part of the constants β and g. However, depending on the actual values of these factors, the denominator of η can in some case approach zero. This gives rise to the near-field resonance discussed in Sect. 2.5, first demonstrated on a silicon carbide surface around $\lambda = 11\mu$m[63]. The effect was found to occur for the real part of the dielectric constant around $\varepsilon' \approx -2$, i.e. the real part of the reflection coefficient $\beta' \approx +3$.

The strength of the near-field interaction between a probing tip and SiC surface in contact with it is plotted according to the monopole and dipole models in Fig. 5.15(a). We see that both models predict a resonant response, but the predicted position and the strength of the resonance are different. The probe polarizability, and therefore also the near-field interaction strength is greater in the monopole model, so that the resonance condition is reached

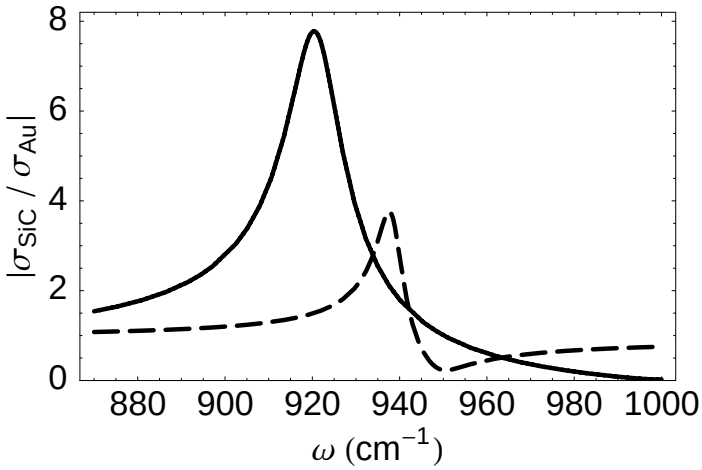


Figure 5.15: Scattering signal generated by the near-field interaction between a probing tip and SiC surface, relative to Au surface. Predictions by the monopole model (full line) and dipole model (dashed line) are displayed for the case of a stationary (non-oscillating) probe in contact with the sample.

at lower values of re(β). From Fig. 3.1(b) we see that lower values of re(β) correspond to lower frequencies, where the damping is also weaker, thus explaining why the predicted magnitude of the near-field resonance is larger in the monopole model. However, neither of these predictions can be directly compared to the measured values since the near-field signal is recorded by a s-SNOM only at a higher harmonic of the probe vibration frequency to suppress the background signal.

When the tip vibration is taken into account, the situation becomes more complicated since both the interaction strength and its distance dependence must be correctly modeled at the same time. The prediction for the harmonics of the scattered field in the region of near-field resonances therefore represents the ultimate test for the theoretical models.

Fig. 5.16 displays the comparison of the theoretically predicted and experimentally observed near-field spectra of a 4H-SiC crystal cut parallel to the c-axis (cf. Sect. 5.3). For theoretical predictions, the polarizabilities given by Eq. 5.29 (monopole model) and Eq. 2.10 (point dipole model) were used and the signal demodulation was performed according to Eq. 5.40. The experimental parameters used were the probing tip radius $R \approx 35\,\text{nm}$ (MikroMasch, CSC37/Ti-Pt) and the vibration amplitude $A = 25\,\text{nm}$, and the values of g and L were not changed ($g = 0.7e^{0.06i}$, $L = 300\,\text{nm}$). To minimize the influence of the far-field reflection factor r_p, the measurement was performed about 200 nm far from the boundary of SiC crystal surface and 30 nm thick Au film evaporated onto the SiC crystal.

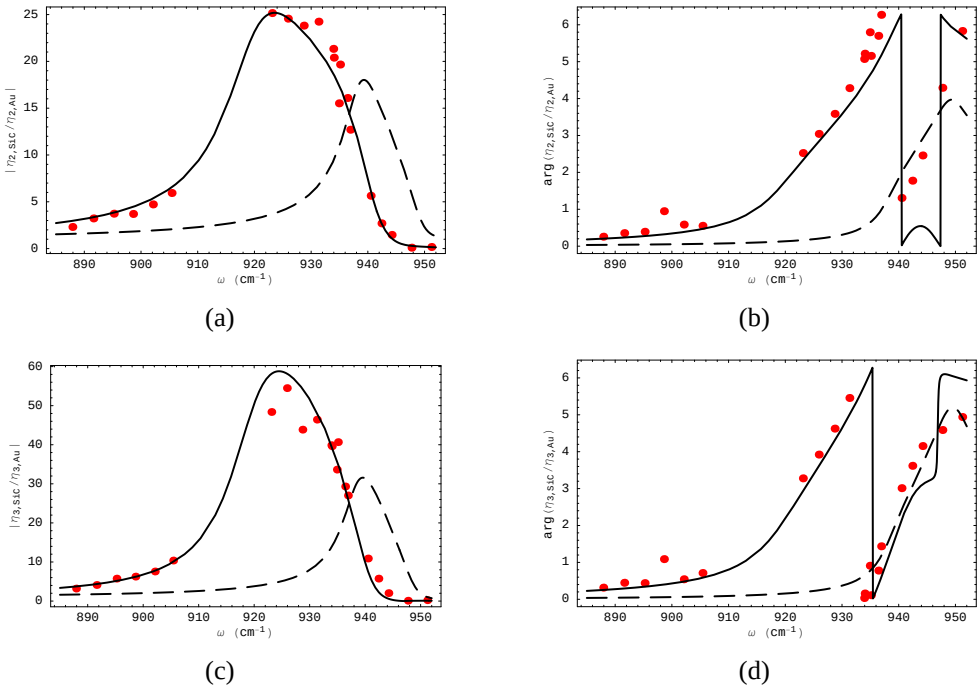


Figure 5.16: Near-field spectra of a 4H-SiC crystal cut parallel to c-axis. Shown are the values obtained experimentally (dots) and predictions by the monopole (full line) and dipole model (dashed line) for the second and third harmonic amplitude and phase.

From Fig. 5.16 it is obvious that the monopole model reproduces the experimental spectra much better than the dipole model. Apart from the signal phase in the region above the resonance, the monopole model exhibits a quantitative agreement with the experiment. As a result, the ultimate goal of recovering the optical constants on the nanometer might finally become reachable.

6 Background-free Detection of Near-field Signals

In Chapter 5, a new ”monopole” model of the near-field interaction between s-SNOM probes and investigated samples was presented. Even though the near-field interaction was shown to be mediated primarily by the evanescent fields in the gap between the probe and sample, the interaction can still be detected and measured far from the probe. This is possible because of an additional dipole moment p induced in the probe as a consequence of the near-field interaction. The dipole moment p oscillates with the illumination field frequency ω and therefore radiates light which can be utilized to measure the near-field interaction strength. This light is usually referred to as the ”near-field signal” or even ”near-field scattering”. Strictly speaking, the latter term should be understood as a shorthand for ”near-field-interaction-induced scattering”.

The near-field interaction described by the monopole model only takes place after an initial dipole p_0 has been induced in the probe by the illumination field E_0 (Fig. 5.2). By definition, p_0 is the constant part of the total tip dipole moment $p = p_0 + p$, not affected by the near-field interaction. For this reason, the light radiated by p_0 conveys no information useful for near-field microscopy and represents the s-SNOM ”background signal” or ”background scattering”.

It is long known that the background signal may cause various artifacts in s-SNOM images and should be suppressed as much as possible[43, 56, 57, 69, 173]. The mechanism of background signal generation and methods for its separation from the near-field signal are examined in this chapter. A new ”pseudo-heterodyne” s-SNOM signal detection technique is thereby introduced and its advantages for near-field spectroscopy are demonstrated in comparison to alternative techniques.

6.1 Background artifacts

We start the investigation of the background signal by deriving the total scattering contrast σ of the coupled tip-sample system, including both the near-field and the background contributions. For the model spheroid from Chapter 5, the result was already derived in Eq. 5.34, repeated here for convenience:

$$\sigma = k_\sigma (1 + c r_p)^2 (\gamma_0 - 1) w^2 L^3 (1 + \eta). \tag{6.1}$$

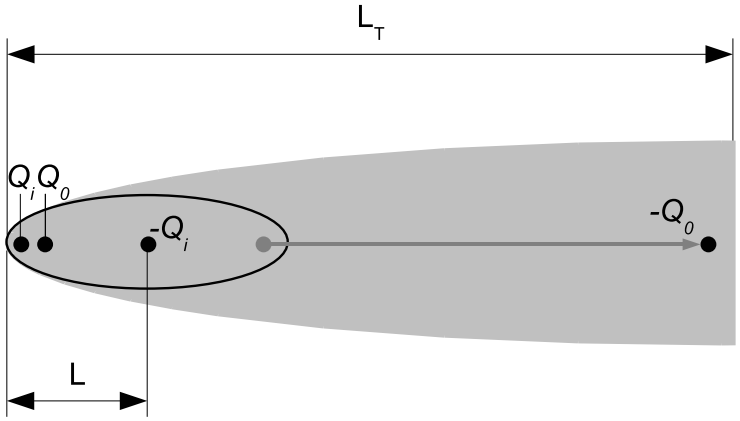


Figure 6.1: Approximate calculation of the background dipole moment $p_{0,T}$ of a real tip: the model spheroid background $p_0 = 2Q_0L$ is increased to $p_{0,T}$ in proportion to the ratio of the true tip length L_T to the model spheroid length $2L$. This is equivalent to shifting the charge $-Q_0$ to the real tip end without changing its value, yielding $p_{0,T} = Q_0 L_T$.

For a real tip, the situation is more complicated because the model probing tip length $2L$ differs from the true tip length L_T. While the model tip length $2L$ measures the portion of the tip responsible for the near-field interaction, the background scattering is produced by the entire length L_T of the tip shaft in focus. Unless $L_T \ll \lambda$, the scattering by the entire tip requires a full electrodynamic treatment. Since this would make the analysis considerably more difficult, a simple estimate will be made here instead. In particular, the difference between the model spheroid length $2L$ and the probing tip length L_T will be accounted for by increasing the contribution of the background scattering in proportion to the probe length. This is equivalent to moving the charge $-Q_0$ from the point close to the spheroid end to a point near the opposite probe end without changing its value, as illustrated in Fig. 6.1. Although it represents just a rough estimate, the result obtained is much more

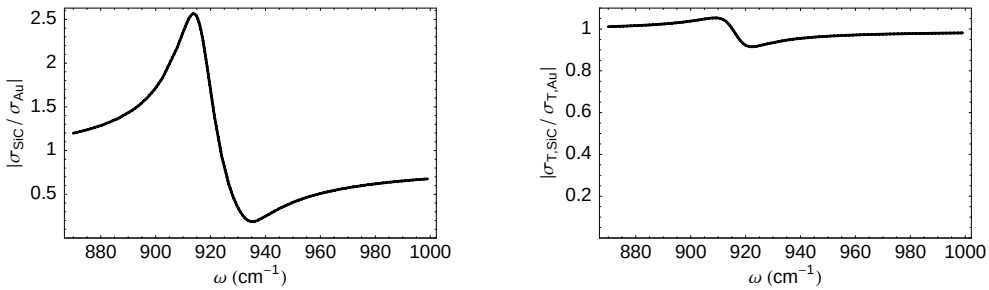
satisfying than the alternative approach of increasing the tip length to L_T before the charge Q_0 is calculated, which would inevitably lead to a large overestimate in both the near-field and background signal strength because of the neglected retardation effects. With this approximation, the background dipole moment of the tip becomes $p_{0,T} \approx Q_0 L_T$, resulting in the total dipole moment

$$p_T = p_{0,T} + p = L\,Q_0\,(\chi_T + \eta), \tag{6.2}$$

with the factor $\chi_T = L_T/2L$. In analogy to Eq. 5.34, the approximate scattering coefficient σ_T of the probing tip may be written as

$$\sigma_T = k_\sigma\,(1 + c\,r_p)^2\,(\gamma_0 - 1)\,w^2 L^3\,(\chi_T + \eta). \tag{6.3}$$

To determine the numerical value of χ_T, we can turn to Sect. 5.4, where a good fit to experimental values was obtained by using the value $2L \approx 0.6\mu$m. Given the tip length $L_T \approx 15\mu m$, the background signal is expected to be between one and two orders of magnitude stronger than the near-field contribution. The effect of adding the background scattering to the near-field signal is illustrated in Fig. 6.2 using the SiC near-field resonance as an example. For $\chi_T = 24$, the maximum relative DC contrast between SiC and Au measured by a long tip (Fig. 6.2(b)) is reduced by a factor of about 20 compared to the contrast obtained using a small spheroid (Fig. 6.2(b)) and about 50 compared to the pure near-field contrast (Fig. 5.15). Since the SiC near-field resonance represents an example



(a) (b)

Figure 6.2: DC scattering contrast σ for a $r = 20$nm probe in contact with a SiC crystal surface, relative to Au surface. Shown are the scattering contrasts calculated for (a) a small spheroid with $L_T = 2L$, and (b) a long tip with $L_T = 24L$. The parts (a) and (b) can also be compared to Fig. 5.15, where the same calculation with $L_T = 0$ is shown.

of very strong contrasts, it is likely that under such circumstances the near-field contrast between two non-resonant materials would be hardly noticeable at all. Yet despite such large *relative* contrast loss, the *absolute* complex contrast is not affected at all because the background term is additive with respect to the near-field signal. Therefore, it should still be possible to recover the pure near-field contrast, provided the background contribution can be eliminated.

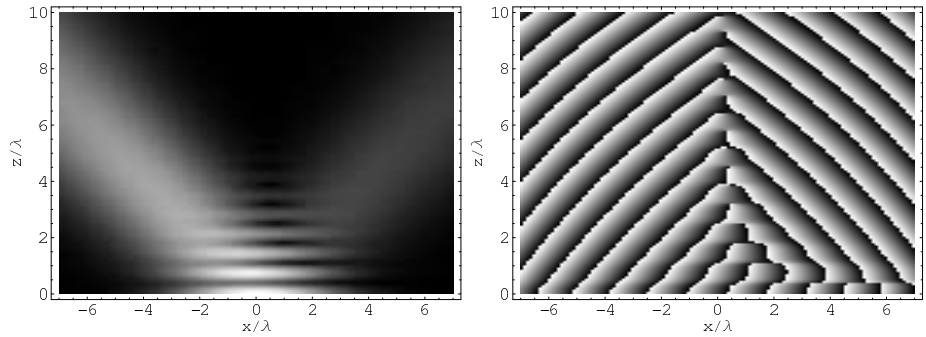
6.1.1 Diffraction Artifacts

Besides the loss of the relative contrast, there exists one further reason why the background has to be suppressed. In particular, the background scattering is generated by the entire probing tip body and therefore depends on the exact tip shape and the illumination profile along the tip shaft. The illumination profile can change over short and long distances in very complicated ways due to perturbations of the illumination field. The perturbations can be caused, for example, by diffraction of the waves reflected from an inhomogeneous surface. Since the perturbations modify the illumination profile along the tip shaft, they necessarily influence the background scattering. The key point here is that changes in the background scattering have exactly the same impact on the detected signal as the changes in the near-field interaction. Yet the background signal is much stronger than the near-field part, so even a small change in the background signal may be sufficient to overwhelm the near-field contrast. An example of this effect observed experimentally in the visible spectral range can be found in [174], Fig. 4.6. It clearly demonstrates that diffraction can produce optical contrasts which do not resemble the original structure of the surface and also do not coincide with it spatially. Both of these properties are disadvantageous for the near-field microscopy and should be avoided for reliable near-field imaging.

6.1.2 Topography Artifacts

The diffraction artifacts can be often recognized when they appear in the near-field images because they exhibit variations on the wavelength scale and bear no resemblance to the sample topography. Although in this case the topography helps to spot the diffraction artifacts, it can cause another kind of artifacts on its own. Namely, it has been observed that an optical contrast can be obtained solely due to the small variations in the sample surface height, with no relation to the optical properties of the material under the probe

[175]. Unlike the diffraction artifact, the topography artifact induces the optical signal variations as fast as the change in the sample height. For this reason it is easily mistaken for the true optical contrast. The origin of this artifact can be understood by examining the interference of the directly incident and reflected waves above the sample, as shown in Fig. 6.3. A region of "standing waves" is formed where the two waves strongly interfere. This interference makes the scattering amplitude dependent on the height above the surface. Since the probing tip always follows the topography of the sample, its vertical position in



(a) (b)

Figure 6.3: Gaussian beam with waist size 2λ reflecting off a silicon crystal surface. Shown are (a) amplitude and (b) phase of the resulting wave.

the standing wave field can change slightly as the sample is scanned below the tip. It is thereby important to note that the standing wave field is susceptible to variations on the wavelength scale and does not follow the rapid local height variations. As a consequence, the background scattering changes in response to the sample topography. Although the background signal change due to a few tens of nanometers topography is quite small, it is important to remember that only a few percent change in the background scattering can represent a significant variation in comparison to the pure near-field contrasts since $p_i \ll p_{0,T}$, i.e. $\eta \ll \chi_T$. A detailed analysis of the topography artifacts can be found in [56] and [176].

The examples presented above serve to motivate the efforts to suppress the background scattering as much as possible. One obvious way to reduce the background would be to tailor the probing tip to the shortest possible length without reducing the near-field signal strength. In the best case, the total tip length should be equal to the effective tip length participating in the near-field interaction, $L_T \approx 2L$. However, no experimental results in this direction have been reported yet. Even if such an approach eventually turns out to be effective, it still would not be able to completely eliminate the background, despite the

presumably large effort required. Rather, it would only bring the near-field and background signal strength to the approximately same order of magnitude.

Because the background signal can never be completely eliminated, a method is needed to separate the background from the near-field signal after it has been generated. As will be shown next, two different techniques are actually simultaneously required for a complete background suppression because there are two different contributions of the background to the measured scattering signal: one of them is additive, and the other one multiplicative with respect to the near-field signal.

6.2 Additive s-SNOM Background

6.2.1 Origin of the Additive s-SNOM Background

The scattering coefficient σ_T from Eq. 6.3 can be naturally divided in two terms; the near-field scattering coefficient σ_N and the background scattering coefficient σ_B defined as

$$\sigma_N = k_\sigma (1+cr_p)^2 (\gamma_0 - 1) w^2 L^3 \eta, \quad \text{and} \tag{6.4}$$

$$\sigma_B = k_\sigma (1+cr_p)^2 (\gamma_0 - 1) w^2 L^3 \chi_T. \tag{6.5}$$

Added together, they give the total scattering coefficient of the probing tip, $\sigma_T = \sigma_N + \sigma_B$. According to its definition, χ_T is a constant number and does not depend on the probe-sample distance or interaction. On the other side, the near-field contrast factor η (Eq. 5.24) is a measure of the near-field interaction strength and vanishes for large distances between the probe and the sample. Based on these facts, the extraction of the pure near-field signal might look like a straightforward task. In particular, if the scattering coefficient $\sigma_T(H \to \infty) = \sigma_B$ measured with the probe far away from the sample is subtracted from the scattering coefficient $\sigma_T(H = 0)$ obtained with the probe in contact with the sample, the difference $\sigma_T(H = 0) - \sigma_T(H \to \infty) = \sigma_N(H = 0)$ corresponds to the near-field scattering coefficient at $H = 0$.

Unfortunately, the approach outlined above does not work for the same reason the topography artifacts (Sect. 6.1.2) appear. Namely, background signal changes with the probe distance from the sample due to the interference between the waves directly incident on the probe and those reflected from the sample. This observation demands the complex coefficient c in Eqs. 6.4 and 6.5 to be treated as a function of the distance H between the

probe and the sample. As a consequence, $\sigma_B(H \to \infty) \neq \sigma_B(H = 0)$, and no definite conclusion about the near-field scattering coefficient σ_N can be made based on the difference $\sigma_{T,\infty} - \sigma_{T,0}$.

An alternative method of background signal suppression[43] relies on the measurement of the change in the scattering coefficient σ_T over short distances H, on the order of the probing tip radius $R \ll \lambda$. Over such distances, the illumination field due to the interference between the direct and reflected waves remains almost constant, so that the background term σ_B only changes very slightly and a large part of the scattering coefficient change can be attributed to the near-field signal. In principle, the required distance variation may be achieved by moving either the probe or the sample. Since it has to be done repeatedly one or more times for each pixel in an image, it is in practice more convenient to vary the tip-sample distance by letting the probing tip vibrate. This enables the average distance to be regulated based on the tip vibration amplitude and the distance variation can also be performed significantly faster since the tip mass is much lower than the mass of the entire sample and its carrier.

Due to the tip vibration, the scattering contrast σ_T is periodically modulated with the fundamental frequency equal to the tip vibration frequency Ω. Being a periodic function of time, σ_T can be decomposed into harmonic components with frequencies $\Omega_n = n\Omega$ and amplitudes $\sigma_{T,n}$. The amplitudes σ_n are thereby equal to the Fourier series coefficients of σ_T, as explained in Sect. 5.4.1.

Since the Fourier transform is a linear operation, we can separate the harmonic amplitudes $\sigma_{T,n}$ into a ear-field and a background part the same way as the original function σ_T to obtain $\sigma_{T,n} = \sigma_{N,n} + \sigma_{B,n}$. To estimate the ratio $\sigma_{N,n}/\sigma_{B,n}$, let us recall that the background signal oscillates on the wavelength (λ) scale (Fig. 6.3), whereas the near-field signal diminishes over a few tip radii $R \ll \lambda$ (Fig. 5.1). Since higher harmonics have higher frequencies, they measure faster changes in the scattering signal and favor the increase in the ratio $\sigma_{N,n}/\sigma_{B,n}$ with increasing index n. This means that by choosing a sufficiently large n, the pure near-field scattering contrast can be recorded. This approach to background suppression is known as the higher-harmonic demodulation technique [56, 57, 170].

In the following, a quantitative analysis of the relation between the near-field and background scattering coefficients $\sigma_{N,n}$ and $\sigma_{B,n}$ will be presented in order to find the experimental parameters which maximize the ratio $\sigma_{N,n}/\sigma_{B,n}$.

6.2.2 Background Scattering in Absence of Sample

To reduce the complexity of the analysis, several simplifications will be made. First of all, only plane wave illumination will be considered. Second, the probing tip vibration amplitude A will be assumed to be small with respect to the illumination wavelength, $\lambda \gg A$. With the illumination in the form of a focused Gaussian beam and the typical experimental parameters $A \approx 20\text{nm}$, $\lambda \approx 10\mu\text{m}$, the said approximations should not have a significant influence on the end result. Third, the probing tip will be reduced to a point scatterer at its apex for a first qualitative analysis, and fourth, the reflection from the sample will be ignored for the beginning. The last two simplifications have considerable influence on the end result and will be reconsidered later.

In practice, having no reflections from the sample means that the probe is actually located far away from the sample, in which case no interference pattern akin to Fig. 6.3 is formed. Under this condition, we can assume the incident field magnitude is constant over the entire tip trajectory as long as the tip vibration amplitude is small compared to the wavelength and therefore also to the focal spot size. However, the incident (and thus also the scattered) field phase does change as the tip oscillates. If we denote the mean tip-sample distance by H_0, the optical path difference to an arbitrary point $H_0 + \Delta H$ on the tip trajectory is equal to $\Delta S = \Delta H \cos\theta$, as depicted in Fig. 6.4. Substituting $\lambda = 2\pi/k_0$, we obtain the phase difference between the points H_0 and $H_0 + \Delta H$:

$$\Delta\varphi = k_0 \Delta S = k_0 \Delta H \cos\theta. \tag{6.6}$$

Since the probing tip oscillates according to the function $\Delta H = A\cos\Omega t$, the illumination field phase φ continuously changes according to the function $\varphi = \Phi_0 \cos\Omega t$. The phase oscillation amplitude Φ_0 is equal to $\Delta\varphi$ from Eq. 6.6 at the point of the maximum elongation $\Delta H = A$:

$$\Phi_0 = A k_0 \cos\theta. \tag{6.7}$$

The same path difference is encountered again with the radiated waves, thereby doubling the total optical phase variation in the scattering coefficient. This requires Eq. 6.5 to be rewritten as

$$\sigma_B = k_\sigma \, e^{i 2\Phi_0 \cos\Omega t} \, (\gamma_0 - 1) \, w^2 L^3 \, \chi_T, \tag{6.8}$$

where the reflection form the sample has been neglected because the sample is assumed to

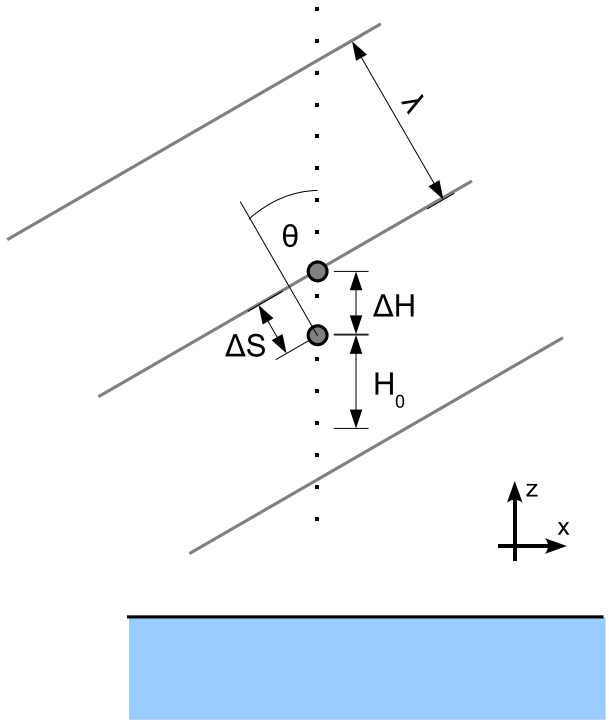


Figure 6.4: Optical path length change due to the motion of the scattering source.

be out of focus.

The additional factor $e^{i2\Phi_0\cos\Omega t}$ in Eq. 6.8 can be interpreted as a sinusoidal phase modulation of the scattering signal. Since the $\cos\Omega t$ factor stands in the exponential, the phase modulation is obviously not a linear process and it necessarily produces higher harmonics of the modulation frequency Ω. The amplitudes of the higher harmonics can be obtained form the Fourier series decomposition of the function $\sigma_B(t)$. All factors in Eq. 6.8 except $f = e^{i\varphi} = e^{i2\Phi_0\cos\Omega t}$ are constant and are not affected by the Fourier expansion. The Fourier coefficients f_n of the function f can be expressed in terms of the Bessel functions of the first kind, J_n [155, 176, 177], yielding

$$f_n = J_n(2\Phi_0)\, i^n. \tag{6.9}$$

Since we are considering small vibration amplitudes $A \ll \lambda$, we have $Ak_0 \ll 1$ and therewith $\Phi_0 \ll 1$. This permits the Taylor expansion of coefficients f_n around $\Phi_0 = 0$:

$$f_n = \frac{(i\Phi_0)^{|n|}}{|n|!} + O(\Phi_0)^2. \tag{6.10}$$

With the aid of Eq. 6.10, the background scattering coefficients can be calculated:

$$\sigma_{B,n} \approx k_\sigma \frac{(i\Phi_0)^{|n|}}{|n|!}(\gamma_0 - 1)\, w^2 L^3 \chi_T. \tag{6.11}$$

Given that $\Phi_0 \ll 1$, the above expression reveals a rapid decrease in the amplitude of the background scattering coefficient with the harmonic index n, just as expected from the qualitative arguments presented in the introduction to Sect. 6.2.

Note that the absolute value $|n|$ has been used in Eq. 6.10 to explicitly allow for all integer values of n, both positive and negative. The provision for such values of n is necessary for the following reason: when the modulation described Eq. 6.8 is applied to the optical carrier wave oscillating at frequency ω, the carrier gets split into sidebands with lines at frequencies $\omega_n = \omega + n\Omega$, i.e. both higher and lower than the carrier wave frequency ω. Since the expansion coefficients $\sigma_{B,n}$ are symmetric around $n = 0$ and they overlap upon the measurement of the carrier wave intensity by a photodetector, for practical purposes it is often irrelevant whether n is taken to include negative values or not.

6.2.3 Estimating Probe Vibration Amplitude

A short digression will be made here before proceeding to the analysis of the background scattering in the presence of the sample. Returning to Eq. 6.11, it can be easily seen that it provides an interesting possibility to obtain the probing tip vibration amplitude from the ratio of two subsequent background scattering coefficients $\sigma_{B,n-1}$ and $\sigma_{B,n}$, $n \geq 1$. In particular,

$$\frac{\sigma_{B,n}}{\sigma_{B,n-1}} = \frac{i\Phi_0}{n} = \frac{iAk_0 \cos\theta}{n}. \tag{6.12}$$

With the probe far from the sample, there are no reflection from the sample surface ($c = 0$) and also no near-field interaction ($\eta = 0$). The measured scattering coefficients $\sigma_{T,n}$ are thus equal to the background coefficients $\sigma_{B,n}$. As a corollary, we get a purely optical and contactless means to measure the absolute tip vibration amplitude:

$$A = \frac{n\,|\sigma_{B,n}|\,\lambda}{2\pi|\sigma_{B,n-1}|\cos\theta}. \tag{6.13}$$

The only prerequisites for this are the known illumination wavelength λ and the illumination angle θ. Since in most cases the DC signal ($\sigma_{B,0}$) is not available either because

it is not measured at all or because it is contaminated by reflections and scatterings from sources other than the probing tip, the lowest n acceptable is usually $n = 2$. The background signal at $n = 2$ can be rather weak, so the detection sensitivity must be high enough to measure this signal.

As a conclusion, let us note that the vibration amplitude A obtained this way should be used only as an estimate because Eq. 6.13 was derived for a single point scatterer and not for an elongated tip and also the illumination angle is not very precisely defined for a focused light beam. Nonetheless, with the wavelength $\lambda \approx 10\,\mu\text{m}$ and vibration amplitudes $A \approx 20\,\text{nm}$, the results obtained from the background signal measurements with the aid of Eq. 6.13 were experimentally found to differ by only about 10% to 20% from the vibration amplitudes determined by the standard procedure of monitoring the cantilever deflection signal decrease as the average distance between the probe and the sample in the intermittent contact mode is continuously reduced.

6.2.4 Background Scattering in the Proximity of the Sample

Eqs. 6.11 and 6.13 are only valid when the probe oscillates far from the sample so the reflections from the sample surface can be ignored. Clearly, near-field signals must be measured in the immediate vicinity of the sample surface. Close to the surface, the probe is illuminated by both the directly incident and the reflected waves which together form a "standing wave" pattern, shown in Fig. 6.3. In the region of strong interference between incident and reflected waves, it is mostly the *amplitude* of the illumination field that changes with the probe-sample distance H, rather than its phase. However, each interfering wave itself still has a constant amplitude and a changing phase. The phase change of the directly incident wave in the form of Eq. 6.6 has been derived from Fig. 6.4. By symmetry, only the sign of the phase change $\Delta\varphi$ in Eq. 6.6 has to be inverted for the reflected wave, yielding:

$$\Delta\varphi_r = -\Delta\varphi_d = -k_0\,\Delta S = -k_0\,\Delta H\cos\theta. \tag{6.14}$$

With this information, Eq. 6.5 can be improved to account for the changing phase depending on the probing tip position:

$$\sigma_B = k_\sigma \left(e^{i\Delta\varphi} + c\,r_p\,e^{-i\Delta\varphi}\right)^2 (\gamma_0 - 1)\,w^2 L^3\,\chi_T, \tag{6.15}$$

with $\Delta\varphi = A k_0 \cos\theta \cos\Omega t = \Phi_0 \cos\Omega t$, obtained by expanding ΔH in Eq. 6.14 as $\Delta H = A\cos\Omega t$. To determine the Fourier coefficients of σ_B, it suffices to consider only the function $f = (e^{i\Phi_0\cos\Omega t} + c\,r_p\,e^{-i\Phi_0\cos\Omega t})^2$ since all other values are constant by definition. Expanding the square in f', we get:

$$f = e^{i2\Phi_0\cos(\Omega t)} + c^2 r_p^2 e^{-i2\Phi_0\cos(\Omega t)} + 2c\,r_p. \tag{6.16}$$

The Fourier expansion of Eq. 6.16 can now be performed term-by-term. The Fourier coefficient of the last term in Eq. 6.16 is just $2c\,r_p\delta_n$, where δ_n is the Kronecker delta symbol. The remaining two terms can be immediately noted down by analogy to coefficients f_n in Eq. 6.10. This way we obtain

$$f_n \approx \frac{(i\Phi_0)^{|n|}}{|n|!} + c^2 r_p^2 \frac{(-i\Phi_0)^{|n|}}{|n|!} + 2c\,r_p\delta_n, \tag{6.17}$$

or, equivalently,

$$f_n \approx \frac{(i\Phi_0)^{|n|}}{|n|!}(1 + c^2 r_p^2 (-1)^n) + 2c\,r_p\delta_n. \tag{6.18}$$

With f_n from Eq. 6.18, the Fourier series coefficients $\sigma_{B,n}$ assume the following form:

$$\begin{aligned} \sigma_{B,0} &\approx (1+c\,r_p)^2\,(\gamma_0 - 1)\,w^2 L^3\,\chi_T, \quad \text{and} \\ \sigma_{B,n\neq 0} &\approx k_\sigma \frac{(i\Phi_0)^{|n|}}{|n|!}(1 + c^2 r_p^2 (-1)^n)\,(\gamma_0 - 1)\,w^2 L^3\,\chi_T. \end{aligned} \tag{6.19}$$

6.2.5 Overall Scattering Coefficient

To obtain the total scattering coefficient including both the near-field and background contributions, the effect of the scattered field phase modulation by the vibrating tip on the near field signal has to be determined. The initial expression for σ_N is easy to derive by analogy to the background signal σ_B from Eq. 6.15:

$$\sigma_N = k_\sigma \left(e^{i\Delta\varphi} + c\,r_p\,e^{-i\Delta\varphi}\right)^2 (\gamma_0 - 1)\,w^2 L^3\,\eta. \tag{6.20}$$

However, the analogy between σ_B and σ_N cannot be extended to the Fourier coefficients $\sigma_{B,n}$ and $\sigma_{N,n}$ because unlike χ_T which is constant, the near-field contrast factor η depends on the probe-sample distance. For this reason there are two quantities which depend on the time variable t in Eq. 6.20: $\Delta\varphi$ and η. The integral which needs to be evaluated to obtain

the expansion coefficients $\sigma_{N,n}$ thus reads (cf. Eq. 5.43):

$$\sigma_{N,n\neq 0} = k_\sigma\,(\gamma_0 - 1)\,w^2 L^3\,\frac{1}{T}\int_{-T/2}^{T/2} f(t)\,\eta(t)\,e^{-in\Omega t}\,dt, \tag{6.21}$$

with $f(t) = \left(e^{i\Phi_0\cos\Omega t} + c\,r_p\,e^{-i\Phi_0\cos\Omega t}\right)^2$ as before. The factors f and η in the integrand of Eq. 6.21 cannot be separated before integration, but convolution theorem can be applied to the resulting Fourier coefficients f_n and η_n. This way we obtain

$$\sigma_{N,n} \propto \sum_{m\in\mathbf{Z}} \eta_{n-m}\,f_m \tag{6.22}$$

As already mentioned, the coefficients $f_m \propto (i\Phi_0)^{|m|}/|m|!$ given by Eq. 6.18 drop rapidly with the increasing absolute value of m due to $\Phi_0 \ll 1$, so only the term $\eta_n f_0$ should be kept in sum 6.22. From Eq. 6.18, we get $f_0 = (1 + c\,r_p)^2$, and the near-field scattering coefficient becomes

$$\sigma_{N,n} \approx k_\sigma\,(1 + c\,r_p)^2\,(\gamma_0 - 1)\,w^2 L^3\,\eta_n. \tag{6.23}$$

It is important to keep in mind that although the terms $\eta_{n-m}\,f_m$, $m \neq 0$ in Eq. 6.22 can be neglected in comparison to $\eta_n f_0$ because $f_m \ll f_0$, the background signal $\sigma_{B,n} \propto \chi_T\,f_n$ detectable at the same frequency as $\sigma_{N,n}$ cannot be ignored *a priori* due to the large value of χ_T which might compensate for the small values of f_n.

Combining $\sigma_{B,n}$ from Eq. 6.19 and $\sigma_{N,n}$ from Eq. 6.23, the overall scattering coefficient $\sigma_{T,n\neq 0}$ can be constructed:

$$\begin{aligned}\sigma_{T,n} \;&=\; \sigma_{B,n} + \sigma_{N,n} \approx\; k_\sigma\,(\gamma_0 - 1)\,w^2 L^3\,\cdot \\ &\quad\cdot\left(\chi_T\,\frac{(i\Phi_0)^{|n|}}{|n|!}\left(1 + c_B^2 r_p^2\,(-1)^n\right) + \eta\,(1 + c_N\,r_p)^2\right).\end{aligned} \tag{6.24}$$

Before proceeding to the quantitative comparison of $\sigma_{N,n}$ and $\sigma_{B,n}$, the introduction of two different coefficients c_B and c_N has to be explained. It is clear that if the probing tip is approximated by a point scatterer, there is only one value of c necessary:

$$c_B = c_N = c = e^{-i\,k_0 H_0\cos\theta}. \tag{6.25}$$

The meaning of the quantities H_0 and θ is illustrated in Fig. 6.4. As soon as the true tip size is taken into account, two different coefficients c_B and c_N are needed because, at least

according to the monopole model, the total tip length L_T is not the same as the length $2L$ participating in the near-field interaction so the background and near-field scattering originate from different positions. Further progress in determining the coefficients c_B and c_N is hard to make because in addition to the probe distance from the sample, they are dependent on several other factors, including the exact probe shape and length together with the illumination profile along it.

Assuming the background part will be suppressed by choosing a sufficiently large n, we need to determine only the factor c_N. If the probe is approximated by a spheroid of length $2L$ as in the monopole model, and the point scatterer (dipole) which Eq. 6.25 refers to is placed in the center of the spheroid, we have $H_0 = H + L$ and therewith

$$c_N = e^{-ik_0(H+L)\cos\theta}. \tag{6.26}$$

One further correction to c_N is needed to account for using a focused illumination instead of a plane wave illumination assumed in 6.26. If the focal spot has e.g. a Gaussian profile centered on the tip, the sample might only reflect the tail of this profile onto the tip, thereby making the absolute value of the coefficient c_N smaller than unity. Because of the large uncertainty in the actual beam profile and position relative to the probing tip, later in Chapter 7 the absolute value of c_N will be adjusted based on the fit between the theory and the experiment.

Besides the factor c_N, there exists one more problem with Eq. 6.24, namely the reflection coefficient r_p. While it is easy to determine on flat homogeneous surfaces, structured surfaces may give rise to diffraction, thereby changing the effective value of r_p. Beside the perturbations in probe illumination field caused by the reflected waves, a part of the incident light intensity can also be diffracted into surface waves if the sample material supports them. Except when this effect is exploited to investigate the properties of these surface modes such as surface phonon or plasmon polaritons [97], it represents an unwanted disturbance which can interfere with the measurement of material-specific near-field contrasts. As discussed in Sect. 3.3, surface polaritons propagate on the interface between two materials with opposite signs of ε' (real part of the dielectric function). This condition is always satisfied by near-field resonant samples ($\varepsilon' \approx -2$) in air or vacuum, so that special care must be taken to minimize the excitation of surface polaritons in order to obtain reliable near-field spectra of resonant samples. Furthermore, a strong local perturbation of the electric field may be caused by small polariton resonant particles or gaps. This effect

can be exploited to map the eigenfields of resonant structures[178], but it should also be avoided for the analysis of the chemical or structural composition of the sample.

It is important to note that perturbations of the illumination field influence not only the background but also the near-field signal, in the form of a multiplicative factor. For this reason, their manifestation in s-SNOM images can be eliminated neither by the higher-harmonic demodulation technique nor by any other known background suppression technique. However, two possible solutions to this problem will be proposed here as an aid for future attempts to factor the variations in the illumination field out of s-SNOM images.

The first proposed solution consists of recording the near-field scattering signal $\sigma_{N,n}$ demodulated at two different harmonics n and m and finding the ratio $\sigma_{N,n}/\sigma_{N,m}$. In the next section it will be shown that this ratio is dependent on the optical properties of the investigated material, but any multiplicative factor constant over the oscillation cycle of the probe cancels when the ratio $\sigma_{N,n}/\sigma_{N,m}$ is calculated. Furthermore, this procedure might also obviate the normalization step in the process of the near-field spectrum construction (sect. 2.5.1). However, a big disadvantage of such approach is the reduction in the near-field contrast it causes. Preliminary calculations indicate that this way only about 10% contrast between Au and Si surfaces should be expected instead of about 35% obtained by direct comparison of scattering coefficients (Sect. 5.4.1).

Alternatively, all Fourier coefficients $\sigma_{T,n}$ of the scattering coefficient σ_T having non-negligible amplitudes (with the possible exception of the DC term $\sigma_{T,0}$) can be simultaneously recorded, and the function $\sigma_T(H)$ reconstructed from them. In the next step, Eq. 6.24 with η from Eq. 5.24 is used to fit the experimentally obtained function $\sigma_T(H)$, with the dielectric function ε of the sample and the reflection coefficient r_p as the adjustable parameters. To make this possible, the remaining parameters χ_T, c_B, c_N as k_σ should be pre-calibrated on a sample with known optical properties. It might be an interesting topic for further research to find out whether a reasonable fit between the theory and the experiment could be achieved this way.

6.2.6 Near-field vs. Background Scattering

It has already been mentioned several times that the background contribution $\sigma_{B,n}$ to the total scattering coefficient $\sigma_{T,n} = \sigma_{B,n} + \sigma_{N,n}$ becomes negligible compared to the near-field contribution $\sigma_{N,n}$ if the scattering signal is measured at a sufficiently high harmonic n. It has been experimentally observed that ”sufficiently high” in practice means second

or third harmonic ($n = 2$ or $n = 3$) in the infrared, and third or fourth harmonic ($n = 3$ or $n = 4$) in the visible spectral range[69]. It was also found that actual choice depends on the probe vibration amplitude A and the measurement precision required. However, this topic was never investigated theoretically. Based on Eqs. 6.19 and 6.23, the first quantitative prediction of the near-field−to−background signal ratio (NFBR) for different harmonics n and vibration amplitudes A is now possible.

For that purpose, the near-field contrast factor η appearing in Eq. 6.23 will be calculated utilizing the monopole model of the near-field interaction developed in Chapter 5. Further, since no reliable way to determine the factor c in Eq. 6.19 has been derived, the reflections from the surface will be completely ignored by setting $c = 0$ in both Eq. 6.19 and 6.23. Because both the factor c and the reflection coefficient r_p are bound by $|c| \leq 1$ and $|r_p| \leq 1$, their combined contribution must also lie within the unit circle $|c\,r_p| = 1$. Given no other prior information, the choice $c\,r_p = 0$ could thus be interpreted as an average case, in which the reflections from the surface neither increase not decrease the near-field and background scattering.

To calculate the NFBR, the parameters pertaining to the experimental conditions have to be specified. Since all measurements shown in this work were made using a CO_2 laser operating between roughly $\lambda = 9\,\mu\text{m}$ and $\lambda = 11\,\mu\text{m}$, the wavelength will be set to $\lambda \approx 10\,\mu\text{m}$ in the calculations. The probing tip will be described by its true length $L_T = 15\,\mu\text{m}$ and the effective length for the near-field interaction $2L = 600\,\text{nm}$. Finally, a typical tip radius of about $R \approx 25\,\text{nm}$ will be used.

Fig. 6.5 displays the NFBR obtained with the parameter values specified above for a sample made of crystalline silicon. We see that the DC background term $\sigma_{B,0}$ dominates the overall signal, being almost two orders of magnitude larger than any other component, including $\sigma_{N,0}$. At the fundamental probe vibration frequency Ω, the near-field and background signals $\sigma_{N,1}$ and $\sigma_{B,1}$ are almost equal when the amplitude A is small with respect to the tip radius R. The background scattering coefficient $\sigma_{B,1}$ increases faster than its near-field counterpart $\sigma_{N,1}$ as A increases, so that the measured scattering coefficient $\sigma_{T,1}$ consists mostly of the background term $\sigma_{B,1}$ for amplitudes $A \approx R$ typically encountered in experiments. For $A \leq R$ and $n = 2$, the near-field signal $\sigma_{N,2}$ finally exceeds the background signal $\sigma_{B,2}$ by more than one order of magnitude. The ratio of $\sigma_{N,2}$ to $\sigma_{B,2}$ reduces significantly for larger vibration amplitudes due to the quadratic increase of the background, as suggested by Eq. 6.10 ($\sigma_{B,n} \propto (A/\lambda)^n$) [57]. Going to even higher harmonics, the background

contribution becomes negligible compared to the near-field signal under the experimental conditions investigated here ($A/\lambda < 0.01$).

Finally, it is interesting to compare the predicted NFBR to the experiment. Although the plots in Fig. 6.5 are intended to provide only a rough estimate, the measured NFBR on Si surface $\sigma_{N,2}/\sigma_{B,2} \approx 15$ is very close to the predicted NFBR≈20 in Fig. 6.5(b) for the same vibration amplitude $A \approx 25\,\text{nm}$ as in the experiment. The background signal $\sigma_{B,3}$ was below the detection threshold in the same experiment. Such an outcome is also consistent with the predicted NFBR of about $\sigma_{N,3}/\sigma_{B,3} \approx 2000$.

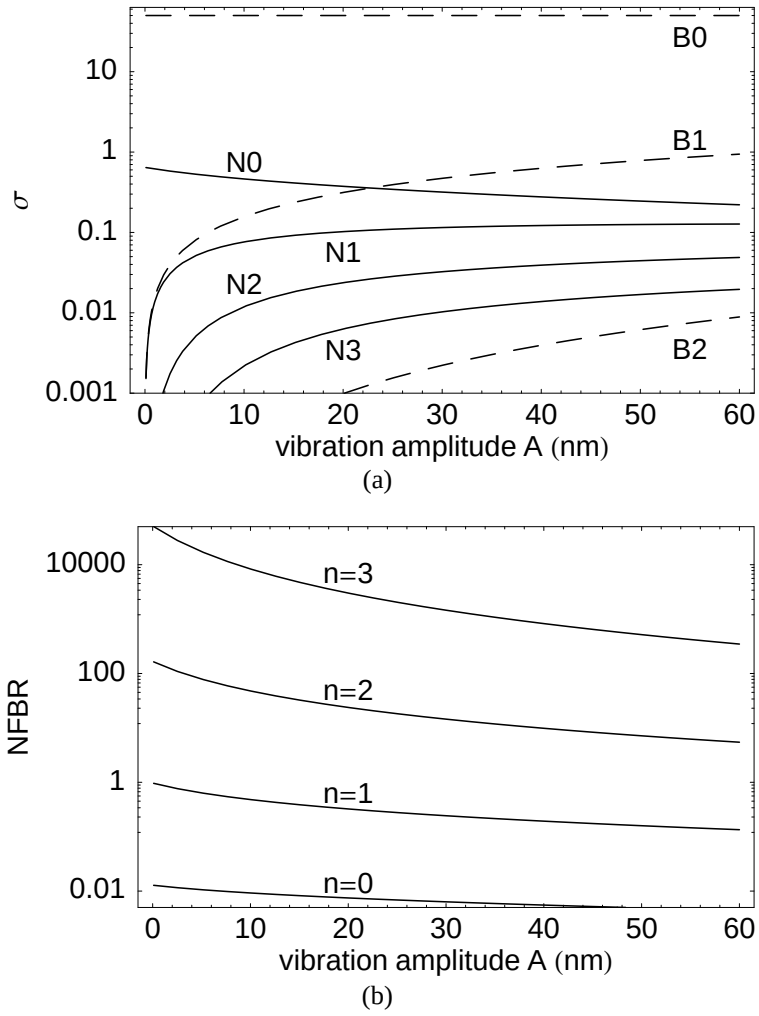


Figure 6.5: (a) Amplitude of the background signal harmonics $\sigma_{B,n}$ (dashed lines, B0..B2) and near-field signal harmonic $\sigma_{N,n}$ (full lines, NF0..NF3) for $n = 0..3$ as a function of the probing tip vibration amplitude A . (b) Ratio $\sigma_{N,n}/\sigma_{B,n}$ of the curves form part (a) . The near-field interaction was calculated using Eq. 5.34 for a Si sample ($\varepsilon \approx 12$) and the probing tip radius $R = 25\,\text{nm}$.

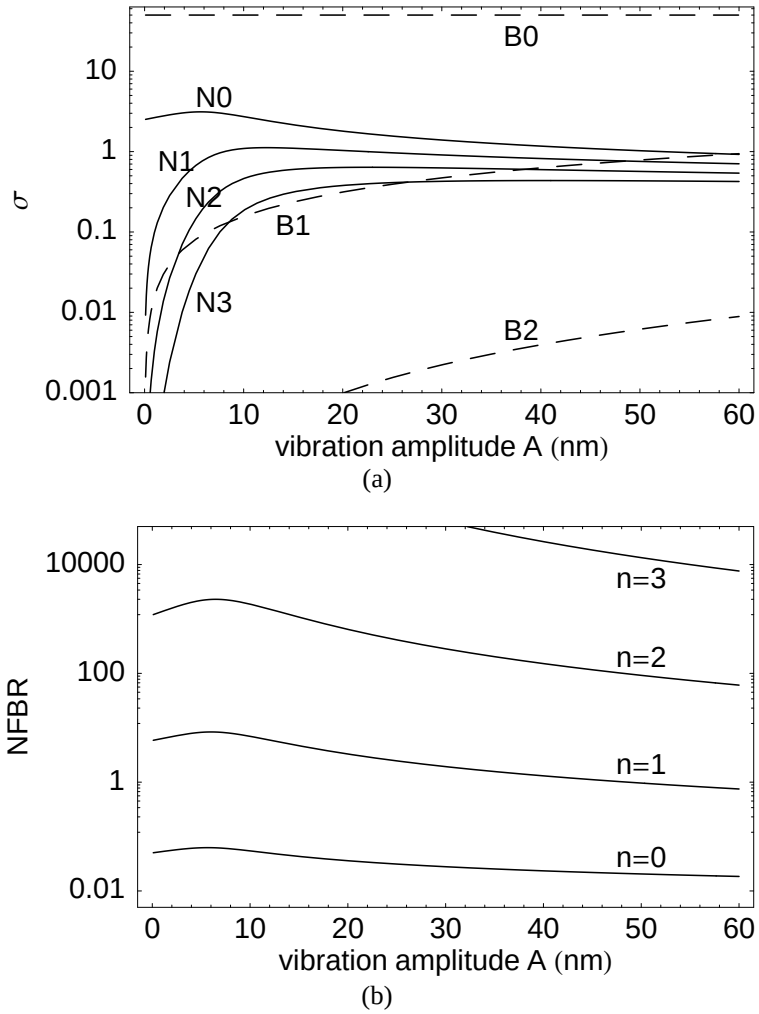


Figure 6.6: (a) Amplitude of the background signal harmonics $\sigma_{B,n}$ (dashed lines, B0..B2) and near-field signal harmonic $\sigma_{N,n}$ (full lines, NF0..NF3) for $n = 0..3$ as a function of the probing tip vibration amplitude A. (b) Ratio $\sigma_{N,n}/\sigma_{B,n}$ of the curves form part (a) . The near-field interaction was calculated using Eq. 5.34 for a resonant SiC sample ($\omega = 930 cm^{-1}$) and the probing tip radius $R = 25\,\text{nm}$.

After an ordinary dielectric material, the analysis of NFBR will be performed for a near-field resonant sample. To this end, Fig. 6.6 shows the same results as in Fig. 6.5, but for a resonant SiC sample instead of Si. The background signal is not affected by the change of the sample, so $\sigma_{B,n}$ still loses a factor λ/A or about two orders of magnitude per harmonic n, as expected from Eq. 6.10. However, the near-field scattering coefficient $\sigma_{N,n}$ is about one order of magnitude higher on SiC than on Si and the background scattering $\sigma_{B,n}$ is proportionally less important. For this reason, a satisfying near-field-to-background ratio of about 10 can be achieved already using the first harmonic ($n = 1$) and very small vi-

bration amplitudes ($A \approx 6$nm). With the second harmonic, the background is negligible compared to the near-field signal for every vibration amplitude A shown in Fig. 6.6, as was also confirmed experimentally.

Another important conclusion which can be drawn by comparing Fig. 6.6 and 6.5 is that the overall signal level remains almost the same, but higher harmonics are generated with higher efficiency on a resonant material such as SiC than on a dielectric like Si. For a typical tip vibration amplitude $A \approx R$, the ratio $\sigma_{N,n+1} : \sigma_{N,n}$ on SiC is as high as 0.7, in contrast to only 0.25 on Si. This observation indicates that the unmodulated (DC) approach curves for different samples do not simply scale while keeping the same slope. Rather, their steepness changes so that the higher harmonics of the demodulated signal have larger amplitudes as the coupling between the probe and the sample increases. At the end of Sect. 6.2.5, it was suggested that this property of the near-field signal might be used to suppress the variations in the s-SNOM contrast due to non-uniform reflection of light from the sample and also to eliminate the need for normalization of all near-field spectra to a standard reference material.

6.2.7 Near-field vs. Background Scattering and Disturbances

From Fig. 6.5 and 6.6, it might seem that using a higher harmonic always represents a better choice since the NFBR is increased this way. This would be true in the absence of noise and other disturbances in the signal. In practice they set a limit to the signal level that can still be measured. The quality of the s-SNOM images obtained is therefore ultimately determined by the ratio of the near-field signal to the background and disturbances (NFBDR). The disturbances are thereby understood to cover all kinds of unwanted interferences with the measured signal, including the thermal and shot noise as well as the signal distortion due to the non-linearity of detectors and amplifiers. It also includes other kinds of disturbances, like capacitive charge pick-ups or inductive cross-talk between currents in different leads or cables.

To estimate the influence of the noise on the measured s-SNOM signal it will be assumed here that a sufficiently high illumination power can be supplied to bring the detector just below the point of saturation. The CO_2 lasers used in this work were found capable of providing enough power to allow this, thus justifying the assumption. A very good detection system may provide a dynamic range of about 100dB, implying that the noise and distortion floor should be set to 10^{-5} of the maximum signal level. The maximum signal

level is determined almost exclusively by $\sigma_{B,0}$, making the total background plus noise coefficient ζ_n equal to $\zeta_n = \sigma_{B,n} + 10^{-5}\sigma_{B,0}$. The ratio $\sigma_{N,n}/\zeta_n$ will be used in the following as an estimate of the NFBDR. At $\lambda \approx 10\mu$m, the s for Si and SiC samples is shown in Fig. 6.7(a) and (b), respectively. For lower vibration amplitudes, best results are obtained with the second harmonic, whereas the third harmonic should be used if the vibration amplitude is large. The border between the two cases shifts to lower amplitudes A for samples yielding stronger near-field interaction. It is interesting to note that with the experimental parameters considered here, the fourth harmonic never represents the optimal choice in the mid-IR because its signal level is closer to the noise floor than the second or third near-field harmonics are to their background counterparts.

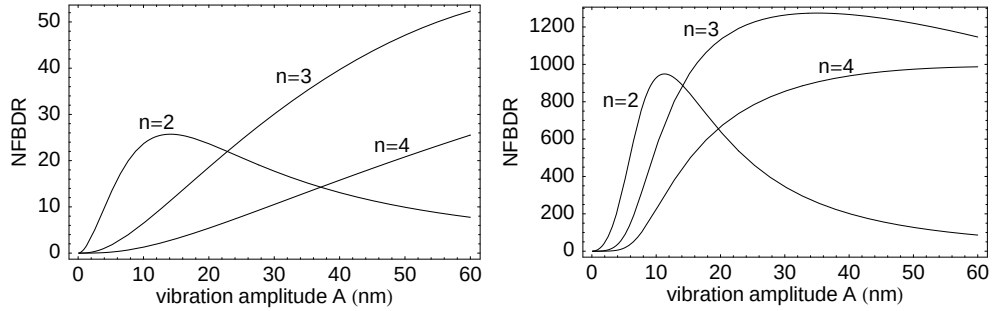


Figure 6.7: Ratio $\sigma_{N,n}/\zeta_n$ of the near-field signal $\sigma_{N,n}$ to the background and disturbances ζ_n (NFBDR) for $n = 2..4$, assuming a detector dynamic range of 100 dB. (a) Si sample, (b) resonant SiC sample ($\omega = 930cm^{-1}$).

It should be noted that the NFBDR shown in Fig. 6.7 is a theoretical maximum obtainable only under ideal conditions and long data acquisition intervals. If the illumination power is not sufficient to take advantage of the entire dynamic range of the detector, or if the acquisition time is limited due to fast scanning of the sample, the noise level might be significantly higher. If the dynamic range is reduced to 80 dB, the second harmonic might provide an advantage over the other choices of the demodulation order n in a much larger range of vibration amplitudes A, as shown in Fig 6.8(a).

Another interesting comparison can be made between the infrared and visible illumination. If the illumination wavelength $\lambda = 600$ nm is taken instead of $\lambda = 10\mu$m, the background scattering becomes a much larger problem. For this reason, the NFBDR similar to Fig. 6.8(a) might be obtained, with $n = 4$ representing the best choice for most tip vibration amplitudes. However, the accuracy of this prediction is rather limited because two assumptions of the monopole model are violated in the visible range. In particular, there are no

perfectly conducting materials which can be used to make the probing tips. For $R = 25\,\text{nm}$ and $\lambda = 600\,\text{nm}$, there is also no model spheroid length $2L$ for which $R \ll L \ll \lambda$. For this reason, the results in Fig. 6.8(a) should be taken with caution.

Nevertheless, a qualitative comparison between the s-SNOM imaging in the visible and infrared spectral ranges can still be made on this basis. The background scattering is much larger and the probing tips are less efficient as optical antennas in the visible range. Furthermore, strong and sharp near-field resonances such as those of SiC are expected to occur only at infrared frequencies because the plasmon resonances in metals at visible frequencies are more strongly damped than the phonon-polariton resonances in the infrared. Compared to the conventional far-field microscopy, the improvement in the resolution achievable by a s-SNOM is also much larger for the infrared than for the visible light. For all these reasons, it is clear that the scattering-type near-field microscopy offers much more advantages when used in the infrared than in the visible wavelength range.

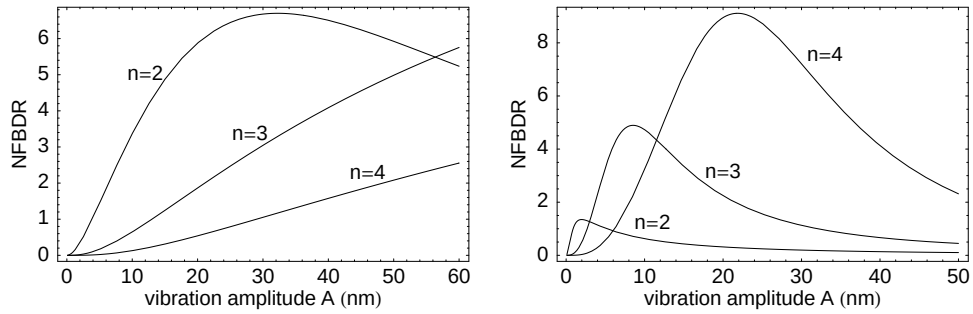


Figure 6.8: NFBDR for the same probing tip and the sample as in Fig. 6.7(a), but with (a) dynamic range reduced from 100 dB to 80 dB, or (b) illumination wavelength $\lambda = 0.6\,\mu\text{m}$ and effective tip length $L = 80\,\text{nm}$ instead of $\lambda = 10\,\mu\text{m}$ and $L = 250\,\text{nm}$.

In conclusion, we have seen that the higher harmonic demodulation technique can easily improve the signal-to-background ratio by several orders of magnitude. However, this improvement is only possible at the cost of the reduced signal-to-noise ratio. As a consequence, an optimal choice of the probing tip vibration amplitude A and the demodulation order n exists, depending on the illumination wavelength and the sample response. With the optimal set of parameters and a near-field resonant sample, the ratio of the near-field signal to the background and noise may in theory exceed 60 dB. In practice, somewhat lower values of up to 40 dB are observed due to limited measurement time and other sources of noise such as mechanical instabilities, which have not been included in this analysis.

6.3 Multiplicative s-SNOM background

6.3.1 Origin of Multiplicative s-SNOM Background

In Sect. 6.2 the light scattered by s-SNOM probing tip was explicitly divided in two components: near-field scattering and background scattering. Those two components are coherent with each other, so the total scattered field E_T measured at some point in space is obtained by summing the field E_N due to the near-field scattering and field E_B arising from the background scattering to obtain $E_T = E_N + E_B$.

Since the probing tip periodically oscillates with a frequency Ω, both E_N and E_B are periodic functions of time and can be expanded in Fourier series,

$$\begin{aligned} E_N(t) &= \sum_{n=-\infty}^{\infty} e^{in\Omega t} E_{N,n} = E_0 \sum_{n=-\infty}^{\infty} e^{in\Omega t} \sigma_{N,n}, \text{ and} \\ E_B(t) &= \sum_{n=-\infty}^{\infty} e^{in\Omega t} E_{B,n} = E_0 \sum_{n=-\infty}^{\infty} e^{in\Omega t} \sigma_{B,n}, \end{aligned} \tag{6.27}$$

where E_0 is the strength of the probe illumination field and the approximate values of the scattering coefficients $\sigma_{N,n} = E_{N,n}/E_0$ and $\sigma_{B,n} = E_{B,n}/E_0$ are given by Eqs. 6.23 and 6.19, respectively. In Sect. 6.2.6, the background scattering coefficients $\sigma_{B,n}$ have been shown to rapidly decrease in comparison to the near-field scattering coefficients $\sigma_{N,n}$ as the index n is increased. By choosing a sufficiently high harmonic n, in theory, the background field can be made negligibly low.

However, this technique alone is not sufficient to completely eliminate the background contribution from the *detected* signal [69] because the detectors of light cannot directly measure the electric field strength. Instead, they produce output currents of voltages proportional to the number of photons they absorb. Yet the number of photons striking the detector is proportional to the light intensity I, and not to its field strength E_T. Given that the intensity I is a quadratic function of the electric field strength ($I \propto |E_T^2|$), the harmonic components of the background field E_B inevitably appear mixed with the harmonics of the near-field scattering E_N in the detector output. Because mixing yields pairwise products of the near-field and background scattering coefficients, this occurrence of the background signal will be referred to as the *multiplicative* background interference, also known as the interferometric background effect [179].

Although the interference of the background and near-field signal harmonics cannot be

prevented, its influence on the outcome of the near-field signal measurement depends on the actual signal detection technique used. In that regard, two simple detection techniques will be analyzed first and demonstrated to be completely or partially susceptible to the near-field−background interference. Pure near-field signals can be measured utilizing more sophisticated detection techniques presented in Sect. 6.4.

6.3.2 Non-interferometric detection

The simplest near-field signal detection scheme uses just the light scattered by the probing tip, which is collected and sent to the detector. Such a setup [43] is illustrated in Fig. 6.9.

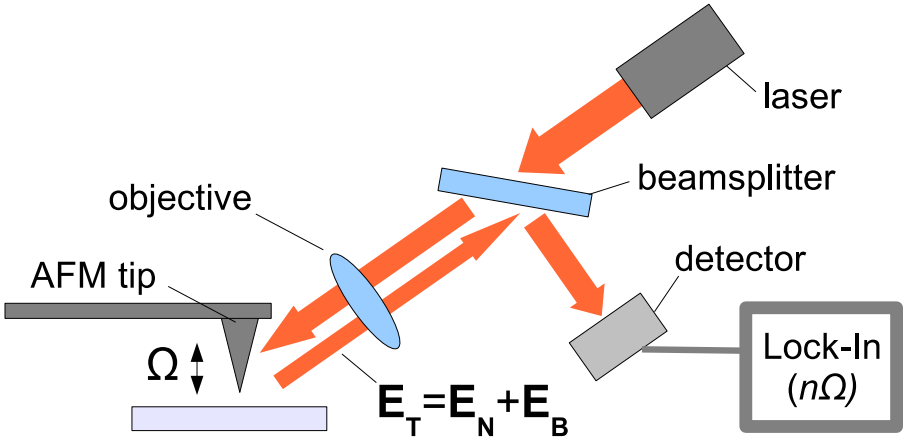


Figure 6.9: Non-interferometric detection setup

The field E_D at the detector position in Fig. 6.9 is equal to the total scattered field E_T which comprises the near-field and background components, E_N and E_B. Substituting E_N and E_B from Eq. 6.27, we have

$$E_D(t) = E_N(t) + E_B(t) = E_0 \sum_{n=-\infty}^{\infty} e^{in\Omega t}(\sigma_{N,n} + \sigma_{B,n}). \tag{6.28}$$

The detector output voltage u is proportional to the light intensity, $u \propto |E_D^2|$. The Fourier series of the output voltage u is then

$$u = \sum_{n=-\infty}^{\infty} e^{in\Omega t} u_n = k_u \left| \sum_{m=-\infty}^{\infty} e^{im\Omega t}(\sigma_{N,m} + \sigma_{B,m}) \right|^2, \tag{6.29}$$

with k_u being a proportionality constant whose exact values thus depends on the applied

laser power, detector responsivity, and the properties of the probing tip and the objective used for the sample illumination and the collection of the scattered light.

To determine the expansion coefficients u_n, the identity $|Z|^2 = ZZ^*$ valid for any $Z \in \mathbb{C}$ will be applied to the sum on the left of Eq. 6.29. After expansion of the left side of Eq. 6.29, the coefficients of all terms containing the complex exponential $e^{in\Omega t}$ should be added together to get the Fourier coefficient u_n. The terms proportional to $e^{in\Omega t}$ which corresponds to the detector output at frequency $n\Omega$ can be obtained only from the pairwise products of the form $X_m e^{im\Omega t} Y_l^* e^{-il\Omega t}$, where $m-l=n$ and X,Y can be any of σ_N,σ_B. There are four such products for each given n and m:

$$u_n = k_u \sum_{m=-\infty}^{\infty} \sigma_{N,m}\sigma_{N,l}^* + \sigma_{N,m}\sigma_{B,l}^* + \sigma_m\sigma_{B,l}^* + \sigma_{B,m}\sigma_{N,l}^*, \tag{6.30}$$

with $l = m-n$. Where more convenient, $\sigma_{N,-l}$ can be substituted for $\sigma_{N,l}$ and $\sigma_{B,-l}$ for $\sigma_{B,l}$ since both E_N and E_B are even functions of time because the probe height variation is given by the $\cos\Omega t$ term in Eq. 5.41.

From Fig. 6.5 one can infer that the terms containing $\sigma_{B,0}$ must be much larger than any other term not containing $\sigma_{B,0}$ in Eq. 6.30. Neglecting all but the four terms involving $\sigma_{B,0}$, we get:

$$u_n \approx k_u \left(\sigma_{B,0}\sigma_{B,n}^* + \sigma_{B,0}\sigma_{N,n}^* + \sigma_{N,n}\sigma_{B,0}^* + \sigma_{B,n}\sigma_{B,0}^* \right). \tag{6.31}$$

The complex scattering coefficients $\sigma_{N,n}$ and $\sigma_{B,n}$ can be expanded into real-valued amplitude and phase components. If we in this sense substitute $s_n e^{i\varphi_n}$ for $\sigma_{N,n}$ and $b_n e^{i\psi_n}$ for $\sigma_{B,n}$, the complex conjugate operator just changes the sign of the phases φ_n and ψ_n. This way we obtain

$$u_n = k_u b_0 \left[b_n e^{i(\psi_0-\psi_n)} + s_n e^{i(\psi_0-\varphi_n)} + s_n e^{i(\varphi_n-\psi_0)} + b_n e^{i(\psi_n-\psi_0)} \right], \tag{6.32}$$

which simplifies to [69]:

$$u_n = 2k_u b_0 \left[b_n \cos(\psi_n-\psi_0) + s_n \cos(\varphi_n-\psi_0) \right]. \tag{6.33}$$

Taking advantage of the higher-harmonic demodulation, the background harmonic amplitude b_n can be made much smaller than the near-field signal amplitude s_n. So for a

sufficiently high demodulation order n, we get

$$u_n \approx 2k_u b_0 s_n \cos(\varphi_n - \psi_0). \tag{6.34}$$

From Eq. 6.34 we see that there is a multiplicative background contribution $b_0 = |\sigma_{B,0}|$ to the detected signal, regardless of the demodulation order n. For this reason, all background artifacts described at the beginning of this chapter can still affect the measured signal even if the detector output is demodulated at a higher harmonic of the probe vibration frequency Ω.

However, there exists an important difference between the additive and the multiplicative background interference which will be explained by example. To this end, let us assume that the amplitude of the background scattering coefficient $\sigma_{B,0}$ has changed by 10% for some reason. In the case of the *multiplicative* background $b_0 = |\sigma_{B,0}|$ in Eq. 6.34, the detected signal u_n would change by the same amount (10%). Since the background b_0 is completely equivalent to the near-field signal s_n in Eq. 6.34, the variation in the background can be misinterpreted as the change in the near-field signal by the same amount, independent of the harmonic n. In contrast, the same 10% change in the *additive* background would in general not cause an error in the measured near-field signal equal to 10%. Rather, it would alter the square bracket in Eq. 6.33 by $0.1b_n$, meaning that the actual error depends on the ratio of s_n to b_n, i.e. the NFBR (Sect. 6.2.6). If we take the values from Fig. 6.5 to estimate the ratio b_n/s_n for $A = 20\,\text{nm}$, we get $b_0/s_0 \approx 100$ and $b_2/s_2 \approx 0.05$. If the former ($n = 0$) case, a 10% change in the background would be 10 times larger than the entire near-field signal s_0, whereas in the latter ($n = 2$) case the change would amount to only about 0.5% s_2. We can thus conclude that when unsuppressed, the additive background can be much more detrimental to the near-field microscopy than the multiplicative background.

However, there is another source of unreliability contained in Eq. 6.34. The cosine factor $\cos(\varphi_n - \psi_0)$ might change due to both the background phase ψ_0 and the near-field phase φ_n [69]. The background phase is an inherently unreliable quantity because it depends, amongst other things, on the adjustment of the optical components which influence the illumination profile. Moreover, all reflections from the sample surface and from the optical elements such as lenses have the same effect on the detected signal as the unmodulated background term $\sigma_{B,0}$. They can thus be simply added to it, making the exact relation between the background phase ψ_0 and the near-field phase φ_0 even harder to predict and control.

Even if the background phase could be controlled by careful adjustment of the optical components, the near-field phase φ_n would still be determined by the tip-sample interaction and could therefore change, e.g., due to a near-field resonance. As a consequence of the phase change, the amplitude of the detected signal u_n would be modified. Such coupling between the near-field signal phase and the measured signal amplitude is undesirable since it can lead to false conclusions about the strength of the near-field interaction. This topic will be elaborated further in Sect. 6.5.

6.3.3 Homodyne detection

The shortcomings of the non-interferometric scattered light detection technique described in Sect. 6.3.2 can be avoided to a large degree by performing an interferometric measurement employing a well defined and controllable reference beam. As illustrated in Fig. 6.10, the reference wave R interferes with the wave T scattered by the probing tip. Formally,

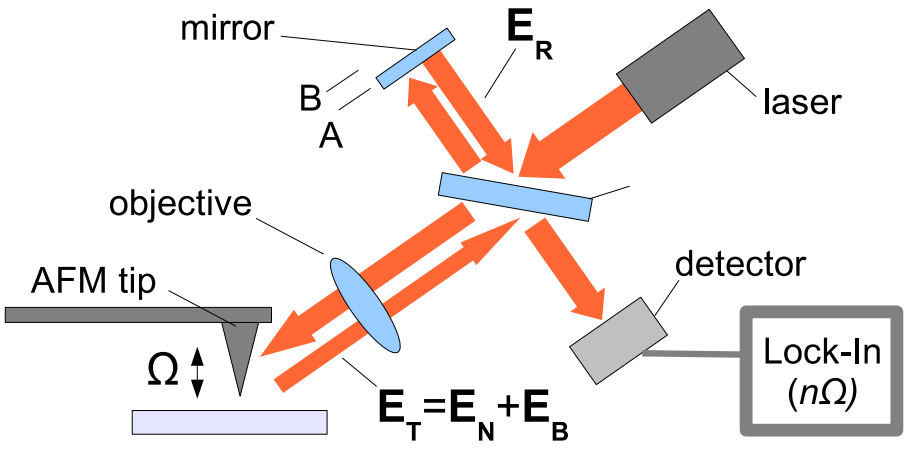


Figure 6.10: Homodyne signal detection

there is no difference between the reference and the background wave concerning the interference with the near-field signal. We can thus easily extend Eq. 6.34 by adding the interference term between the near-field scattering E_N and reference field $E_R = E_0 r_R e^{i\psi_R}$:

$$u_n = 2k_u s_n [b_0 \cos(\varphi_n - \psi_0) + r_R \cos(\varphi_n - \psi_R)]. \tag{6.35}$$

For efficient suppression of the background term $b_0 \cos(\varphi_n - \psi_0)$, it is necessary to have a strong reference wave, $r_R \gg b_0$. If this is the case, the background term in Eq. 6.35 can be

omitted, reducing u_n to

$$u_n = 2k_u r_R s_n \cos(\varphi_n - \psi_R). \tag{6.36}$$

After obtaining u_n, it was suggested in Ref. [45] to repeat the same measurement with the reference phase shifted by $\Delta\psi_R = 90$, marked as position B in Fig. 6.10. From this measurement a second value is obtained,

$$w_n = 2k_u r_R s_n \sin(\varphi_n - \psi_R). \tag{6.37}$$

Having both the sine and the cosine component of a given scattered field harmonic, both its amplitude and phase can be reconstructed. This is most easily done by constructing the complex quantity

$$U_n = u_n + i w_n = 2k_u r_R s_n e^{i(\varphi_n - \psi_R)}. \tag{6.38}$$

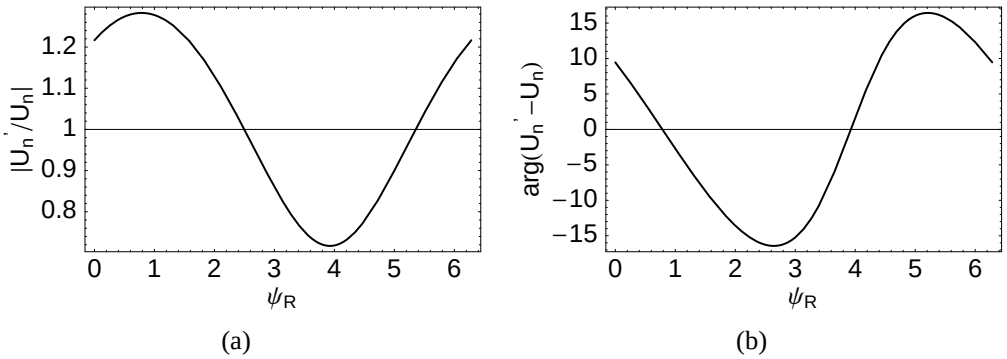
U_n is thus proportional to the near-field scattering coefficient $\sigma_n = s_n e^{i\varphi_n}$, with the proportionality equal to $k_h = 2k_u r e^{-i\psi_R}$. As already explained in Sect. 2.5, this constant needs not be determined if relative measurements are performed.

The most important advantage of the homodyne detection technique over the non-interferometric method is the decoupling of the signal amplitude s_n from its phase φ_n. Owing to the independent recovery of the scattered field amplitude and phase, the homodyne method is suitable for near-field spectroscopy of both strong and weak oscillators such as SiC[63] and polystyrene[62, 64], respectively.

However, the amplitude and phase measurements by the homodyne method are only as accurate as is the approximation made between Eq. 6.35 and Eq. 6.36. There it was assumed that the background scattering is negligible with respect to the reference wave. In a typical experiment, the reference wave gets reflected by a flat mirror (Fig. 6.10), with more than 95% efficiency. On the other hand, it was found that the probing tip returns only about 3-5% of the incident light intensity. The ratio of the reference and background wave intensities is therefore close to $I_R : I_B = 25$. Although such ratio may seem large enough to neglect the background, the background influence is actually much larger because it is the field strength that enters Eq. 6.35, not the intensity. Since $E_R : E_B$ is only about 5, the influence of the background scattering can have a noticeable impact on the measured near-field scattering coefficient.

The influence of the multiplicative background interference on the results obtained using the homodyne detection method can be estimated by comparing the signal amplitude U_n

obtained from Eq. 6.38 with the background omitted to the signal amplitude U_n' obtained with the background contribution included. The worst-case appears when the background and near-field signal phases are equal, i.e. $\psi_0 = \varphi_n$. For this case the behavior of U_n'/U_n as a function of the reference phase ψ_R is shown in Fig. 6.11. The error made by reconstructing the complex scattering signal amplitude via the Eq. 6.38 can be read out directly from Fig. 6.11. The maximum error in the measured amplitude of the near-field contrast s_n amounts to over $\pm 28\%$ and the error in its phase φ_n to about ± 16. Such a large error significantly reduces the accuracy of the near-field contrast measurements.



(a) (b)

Figure 6.11: The ratio of complex signal amplitudes U_n' and U_n calculated with the background contribution included (Eq. 6.35) and omitted (Eq. 6.36), respectively. Shown are (a) the amplitude and (b) the phase of U_n'/U_n as functions of the reference wave phase ψ_R for $\psi_0 = \varphi_n$ and $E_R : E_B = 5$.

Although there exist reference phases ψ_R in Fig. 6.11 for which the amplitude error is zero, and other phases ψ_R for which the phase error is zero, both errors can never be eliminated simultaneously. It is also not possible to attenuate the background without affecting the near-field signal because they both originate from the same spot (the probing tip). Although attenuating them together does indeed increase the ratio $E_R : E_B$, this comes at the cost of the proportionally reduced near-field signal, which is highly unfavorable for achieving a good signal-to-noise ratio. The complete elimination of the multiplicative background therefore requires more sophisticated detection methods, presented in the next section.

6.4 Background-free Detection of Scattered Light

6.4.1 Heterodyne detection

The first detection method capable of a complete multiplicative background suppression was the heterodyne method, introduced in [180] and [58]. It uses a reference wave of a slightly different wavelength than the light used for probe illumination. The reference must nevertheless be coherent with the illumination, and the difference in their wavelengths must be very small compared to the wavelengths. Only under such conditions can an interference between the scattered wave and the reference wave be observed. As illustrated in Fig. 6.12(a), the required wavelength difference may be produced by an acousto-optic modulator (AOM), exploiting the Doppler shift of light upon its diffraction by traveling acoustic waves. The angular frequency shift $\Delta\omega$ of the light wave is thereby equal to the angular frequency Δ of the acoustic waves, times the diffraction order N ($\Delta\omega = N\Delta$). This is equivalent to a wavelength change of $\Delta\lambda = \lambda\Delta\omega/\omega$, which for a typical AOM frequency $\Delta \approx 100$MHz and a wavelength of e.g. $\lambda = 3\mu$m, amounts to only one part in about 10^6.

In the heterodyne detection setup shown in Fig. 6.12(a), the signal and the reference waves are brought to interference by using a variant of the Mach-Zehnder interferometer. Just like in the homodyne method, the near-field signal interferes with the background and the reference wave. However, since the reference and the background now have different frequencies, their interference with the near-field signal can be distinguished from one another. In particular, the frequencies $\Delta \pm n\Omega$ found in the detector output can only come form the interference between the frequency-shifted reference wave and the scattered signal, as indicated in Fig. 6.12(b). This way the multiplicative background is avoided and the remaining additive background found at frequencies $\Delta \pm n\Omega$ can be suppressed by choosing n large enough [69, 181], i.e. by the higher-harmonic demodulation method described in Sect. 6.2.

Formally, the detector output signal demodulated at a frequency $\Delta + n\Omega$ can be expressed as [58, 174]

$$U_n = k_u r_R s_n e^{i(n\psi_T - \varphi_n)}, \tag{6.39}$$

where ψ_T is the phase difference between the mechanical oscillation of the probe and the sinusoidal voltage driving the probe.

The expression for the demodulated detector output voltage in Eq. 6.39 is approximate only

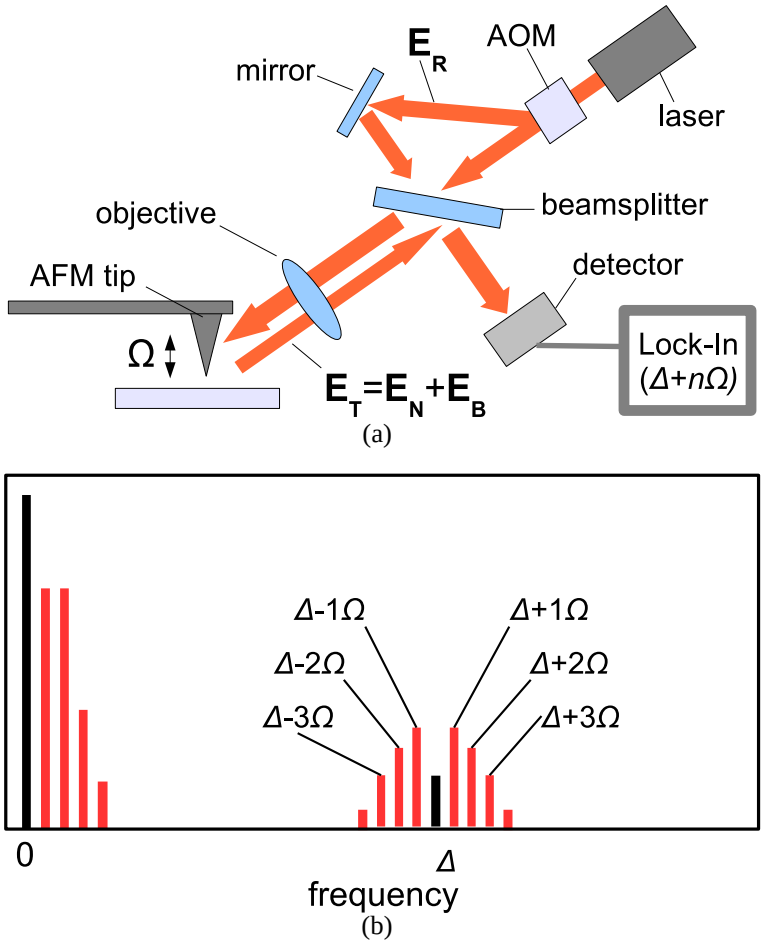


Figure 6.12: (a) Heterodyne detection setup from [58], and (b) the spectrum of the detector output signal.

to the extent that the additive background has been neglected. In particular, the multiplicative background is indeed completely eliminated by the heterodyne detection technique because the signal is extracted at the frequency $\Delta \pm n\Omega$ which cannot be produced by the interference of the near-field and background signals. The heterodyne technique can thus achieve the highest degree of background suppression among all three detection methods presented so far. Consequently, the heterodyne method yields the most reliable near-field images.

However, there are some practical disadvantages of the heterodyne method as shown in Fig. 6.12. First of all, the phase of U_n depends on the phase of the probe vibration, which can change depending on the mechanical interaction of the probe and the sample. Second, AOMs typically operate in the VHF frequency range ($30 - 300\,\mathrm{MHz}$), where the light detectors typically provide a lower signal-to-noise ratio than in the sub-MHz range occupied

by the near-field signal alone. Further, the frequency shifted beam at the output of the AOM is separated from the undiffracted beam by only about ten milliradians, requiring an additional beam expander and making the setup less compact. Additionally, the output angle of the frequency-shifted beam changes with the wavelength, so the setup must be re-adjusted for each selected spectral line. The AOMs operating in the infrared are opaque to visible light so the optical alignment is further hampered by the unavailability of a pilot beam. Finally, the commercially available AOMs cover only a limited portion of the IR spectral range.

Although some of the aforementioned problems can be avoided [182], this can only be done at the expense of increased complexity and cost of the setup. In any case, the implementation of the heterodyne method is a challenging task, especially for the infrared spectroscopy applications. As a consequence, no successful spectroscopic near-field measurement using the heterodyne technique has been reported so far.

In the next section, a simpler alternative to the heterodyne method will be presented. It provides the solution to all problems mentioned above while retaining the same degree of the background suppression.

6.4.2 Pseudo-heterodyne detection

A new method for background-free s-SNOM signal detection is introduced in this section. The method is based on the sinusoidal phase modulation of the reference wave and provides a complete multiplicative background elimination and permits the simultaneous measurement of the near-field signal amplitude and phase. Since the reference wave is modulated and not shifted in frequency, this method is known as the ”pseudo-heterodyne” method [183]. Fig. 6.13(a) contains a schematic representation of the pseudo-heterodyne setup as implemented in [173].

It is in essence a Michelson interferometer configuration and uses only one beam splitter and a vibrating mirror. It is therefore much easier to implement than a heterodyne setup. Also, the required components are readily available over the entire near-UV to far-IR spectral range. Furthermore, the vibrating mirror can be driven by piezoelectric actuators with kHz frequencies and thus leave the useful signal in the sub-MHz range where light detectors offer a higher signal-to-noise ratio compared to those required by the heterodyne technique.

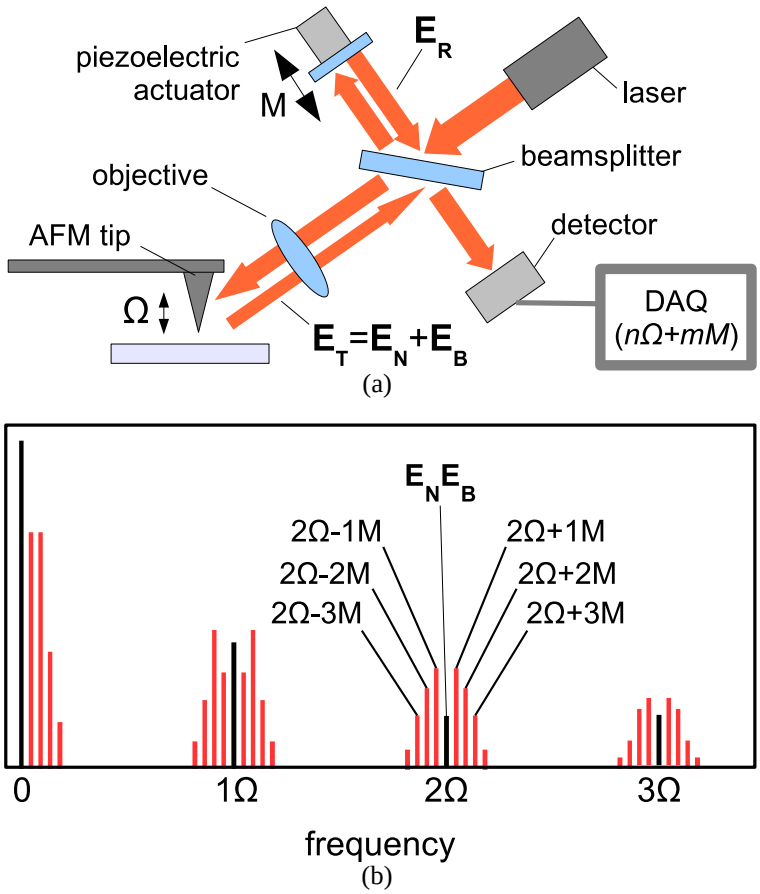


Figure 6.13: (a) Pseudo-heterodyne detection setup, and (b) the schematic representation of the corresponding detector output spectrum.

It now remains to be explained how the near-field signal is recovered by the pseudo-heterodyne technique and how the background interference is thereby avoided. This can be best done by analyzing the detector output spectrum, schematically shown in Fig. 6.13(b). Such a spectrum is a result of the interference between the phase-modulated reference wave field E_R described by

$$E_R = k_e \rho \, e^{i\gamma \cos(Mt)}, \tag{6.40}$$

and the total scattered wave field $E_T = E_N + E_B$, described by Eq. 6.28,

$$E_T(t) = E_0 \sum_{n=-\infty}^{\infty} \sigma_{T,n} e^{in\Omega t}. \tag{6.41}$$

The angular frequency M in Eq. 6.40 denotes the reference wave phase modulation frequency, and the angle γ stands for the phase modulation amplitude (also know as the phase

modulation depth). The complex reference wave reflection coefficient $\rho = r_R e^{i\psi_R}$ has a magnitude r_R and also comprises the phase offset ψ_R that accounts for the average optical path difference between reference and signal waves.

In order to calculate the amplitudes of the spectral components in the detector output, the Fourier decomposition of E_R must be known. The Fourier expansion coefficients of a phase-modulated field were already encountered in Sect. 6.2, Eq. 6.9. In the same manner, the coefficients ρ_m in the Fourier series of interest here,

$$E_R = \sum_{m=-\infty}^{\infty} \rho_n e^{imMt}, \tag{6.42}$$

are equal to

$$\rho_m = r_R J_m(\gamma)\, e^{i\psi_R + im\pi/2}. \tag{6.43}$$

It should be noted that the fundamental angular frequency M involved in the Fourier expansion of E_R in Eq. 6.42 must be different from the frequency Ω in Eq. 6.41. The interference of the two fields E_T and E_R then yields signal at frequencies $\nu_{n,m} = n\Omega + mM$. If the reference wave modulation frequency M is thereby lower than the tip vibration frequency Ω, each of the scattered signal harmonics with frequencies $n\Omega, n > 0$ splits into sidebands containing frequencies $\nu_{n,m} = n\Omega + mM$, depicted in Fig. 6.13(b).

In contrast, the detector output spectrum with no reference wave contains only exact multiples $n\Omega$ of the probing tip vibration frequency. Those are the frequencies where the multiplicative interference between the near-field and background signal is found. With the phase-modulated reference, the multiplicative background still appears at the same frequencies $\nu_{n,0} = n\Omega$, but not at the sideband frequencies $\nu_{n,m} = n\Omega + mM$, $n, m \neq 0$. By extracting the signal from the sidebands which contain the doubly-modulated signal, the multiplicative background interference is thus completely avoided. This is in a full analogy with the heterodyne method, where the same effect is achieved by demodulating the detector output signal at frequencies $\nu_n = \Delta + n\Omega$.

In addition to the elimination of the multiplicative background interference, the pseudo-heterodyne approach also enables the simultaneous measurement of both amplitude and phase of the scattered field's harmonics. This can be done by combining the measured signal at two frequencies $\nu_{n,m}$ and $\nu_{n,l}$ where m is an even and l an odd integer. To demonstrate this, let us first derive the amplitude $u_{n,m}$ of the detected signal at a frequency $\nu_{n\neq 0, m\neq 0}$. Quite generally, the detector output u is given by the intensity of the light resulting from

the interference between the scattered field E_T and the reference field E_R:

$$u = k_u |E_T + E_R|^2 = k_u \left| \sum_{n=-\infty}^{\infty} \sigma_{T,n} e^{in\Omega t} + \sum_{m=-\infty}^{\infty} \rho_m e^{imMt} \right|^2 . \tag{6.44}$$

The detector output amplitude $u_{n,m}$ demodulated at a frequency $\nu_{n,m} = n\Omega + mM$ is comprised of the complex coefficients $c_{n,m}$ found in terms of the form $c_{n,m}e^{i(n\Omega+mM)t}$ obtained after expanding the left side of Eq. 6.44. With the aid of the identity $|Z|^2 = ZZ^*$ we see that the coefficients $c_{n,m}$ can only be the products of the form $\sigma_{T,n}\rho^*_{-m}$ or $\sigma^*_{T,-n}\rho_m$. Since both E_T and E_R are even functions of time, we can write $\sigma^*_{T,-n} = \sigma^*_{T,n}$ and $\rho^*_{-n} = \rho^*_n$. The detected signal amplitude $u_{n,m}$ at the frequency $\nu_{n\neq 0,m\neq 0}$ is thus equal to

$$u_{n,m} = k_u \, (\sigma_{T,n}\rho^*_m + \sigma^*_{T,n}\rho_m), \tag{6.45}$$

Substituting ρ_m from Eq. 6.43 and expressing the complex coefficient $\sigma_{T,n}$ as $\sigma_{T,n} = s_{t,n}e^{i\varphi_{t,n}}$ we obtain

$$u_{n,m} = 2k_u r_R \, J_m(\gamma) \, s_{t,n} \cos(\varphi_{t,n} - \psi_R - m\pi/2). \tag{6.46}$$

From Eq. 6.46 it is clear that the successive lines in a sideband are in alternation proportional to the real part (cos) and imaginary part (sin) of the complex Fourier coefficient τ_n. In particular, the real part is obtained for even m, and the imaginary part for odd m. Consequently, the complex scattering coefficient $\sigma_{T,n}$ can be recovered from two signal amplitudes $u_{n,m}$ and $u_{n,l}$ as

$$\sigma_{T,n} = k_p \, (\frac{u_{n,m}}{J_m(\gamma)} + i \frac{u_{n,l}}{J_l(\gamma)}), \tag{6.47}$$

where $k_p = \frac{e^{i\psi_R}}{2k_u r_R}$ and $m \neq 0$ is an even and l an odd integer.

Eq. 6.47 simplifies even further if $J_m(\gamma)$ is made equal to $J_l(\gamma)$ by a suitable choice of the modulation depth γ. For $l = 1$ and $m = 2$, the required modulation depth obtained from $J_1(\gamma) = J_2(\gamma)$ is γ_{12}=2.63. In the experiment, such modulation can be obtained by vibrating the reference mirror with an amplitude $\Delta l_R = \frac{\gamma_{12}}{2}\frac{\lambda}{2\pi} \approx 0.21\lambda$. The scattering coefficient $\sigma_{T,n}$ is in this case just

$$\sigma_{T,n} = \frac{k_p}{J_1(\gamma_{12})} \, (u_{n,2} + i u_{n,1}). \tag{6.48}$$

As already explained is Sect. 2.5, the value of the constant k_u contained in k_p needs not

be determined for relative contrast measurements usually performed in the near-field microscopy. It is sufficient to consider just the complex amplitude $U_n \propto \sigma_{T,n}$ of the detector output signal, defined as

$$U_n = u_{n,2} + i\,u_{n,1}. \tag{6.49}$$

As long as the experimental conditions are kept constant, the proportionality constant between U_n and $\sigma_{T,n}$ remains unchanged. Finally, taking advantage of the higher-harmonic demodulation, the background contribution $\sigma_{B,n}$ to the overall scattering coefficient $\sigma_{T,n}$ can be made negligibly small. In that case we have $\sigma_{T,n} = \sigma_{N,n} + \sigma_{B,n} \approx \sigma_{N,n}$, so that the measured complex amplitude U_n is proportional to the pure near-field scattering coefficient $\sigma_{N,n}$. The desired near-field scattering coefficient amplitude s_n is then proportional to the modulus of U_n, and the phase φ_n is equal to the argument of the complex value U_n.

6.5 Experimental Comparison of Detection Techniques

After the theoretical presentation of the known s-SNOM detection methods, in the following it will be experimentally demonstrated that the pseudo-heterodyne method succeeds in a complete multiplicative background elimination, whereas the non-interferometric and homodyne interferometric detection schemes do not. For convenience, the schematic illustration of the three detection techniques is provided in the top row of Fig. 6.14. The bottom row of Fig. 6.14 contains the corresponding detector output spectra in the vicinity of the probe vibration frequency Ω, recorded in the absence of the sample. The frequency Ω was equal to about 30 kHz and the reference wave modulation frequency M was set to 400 Hz. The doubly modulated signal resulting from the interference between the background and the reference wave was measured by a MCT photodetector (Kolmar Technologies model KMPV-0.2-J1/AC) which had its maximum responsivity around the wavelength $\lambda \approx 10\,\mu$m, emitted by the tunable CO_2 laser used in this experiment. The experimental spectra shown in the bottom row of Fig. 6.14 were calculated via the Fourier transform of the recorded photodetector output voltage.

The experimental spectra for each of the three detection methods in Fig. 6.14 nicely confirm the theoretical predictions from the preceding two sections. In particular, we see that the non-interferometric method and the homodyne method yield signal in the detector out-

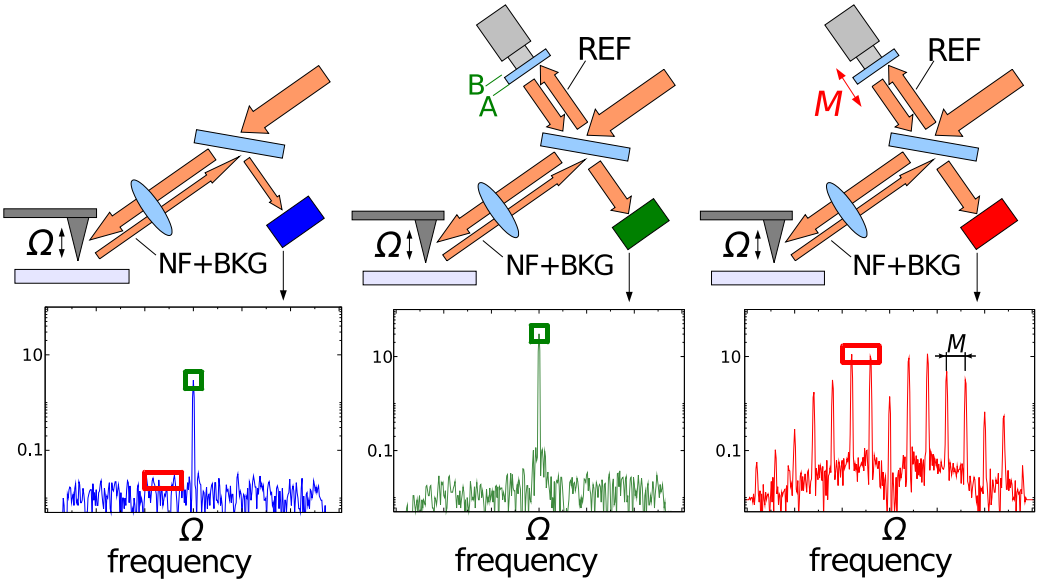


Figure 6.14: Example signal spectrum obtained using different detection methods: (a) non-interferometric intensity detection, (b) homodyne interferometric detection, and (c) pseudo-heterodyne interferometric detection.

put only at the exact multiples of the frequency Ω, marked by squares in Fig. 6.14(a) and (b). In the case of non-interferometric detection, this signal is the interference product of the scattered field component $E_{T,n}$ and the background field component $E_{B,0}$. With homodyne detection, the detector output signal at the frequency $n\Omega$ is a superposition of the interference between the scattered field component $E_{T,n}$ and both the reference field E_R and background scattering field $E_{B,0}$ (cf. Eq. 6.35). Unless $E_R \gg E_{B,0}$, the homodyne measurement of the scattering signal thus incorporates a systematic error, investigated in Sect. 6.3. The experimental manifestation of this effect is shown in Fig. 6.15(c) and described later in this section.

With the pseudo-heterodyne method, the near-field signal is reconstructed from two different frequencies, $n\Omega + 1M$ and $n\Omega + 2M$. These two frequencies are marked by rectangles in Fig. 6.14(c) and (a). From the part (a) it is evident that no signal is present at those two frequencies when a non-interferometric measurement is performed, apart from the noise. The multiplicative background can therefore not affect the pseudo-heterodyne measurements since it appears only at frequencies $n\Omega$, not used for signal reconstruction with the pseudo-heterodyne technique.

To demonstrate the impact of the multiplicative background on the s-SNOM images, a SiC

sample exhibiting a sharp structural transition from 4H to 6H crystal structure is imaged utilizing all three detection methods described above. This sample was chosen because it simultaneously provides a moderate contrast both in the amplitude and the phase of the near-field signal. This contrast is caused by slightly different dielectric properties of the 4H and 6H SiC polytypes. Measured at $\lambda = 10.63\mu$m, it amounts to about 20% in the amplitude s_2 and 30° in the phase φ_2[67]. A more detailed examination of the near-field spectra obtained with the same sample is presented in Sect. 7.5.

Fig. 6.15(d) shows an image of the 4H-6H SiC polytype transition obtained by the pseudo-heterodyne method with three different reference phases ψ_R in sequence, designated as A, B, and C. The total change in the reference phase ψ_R from A to C was about 180° and was achieved by applying a voltage offset to the piezoelectric actuator driving the reference mirror. Apart from the reference phase ψ_R, all other settings were left unchanged. Since neither the amplitude nor the phase contrast between the SiC polytypes changes as ψ_R is varied in Fig. 6.15(d), this experiment provides evidence that the multiplicative background interference has indeed been successfully eliminated.

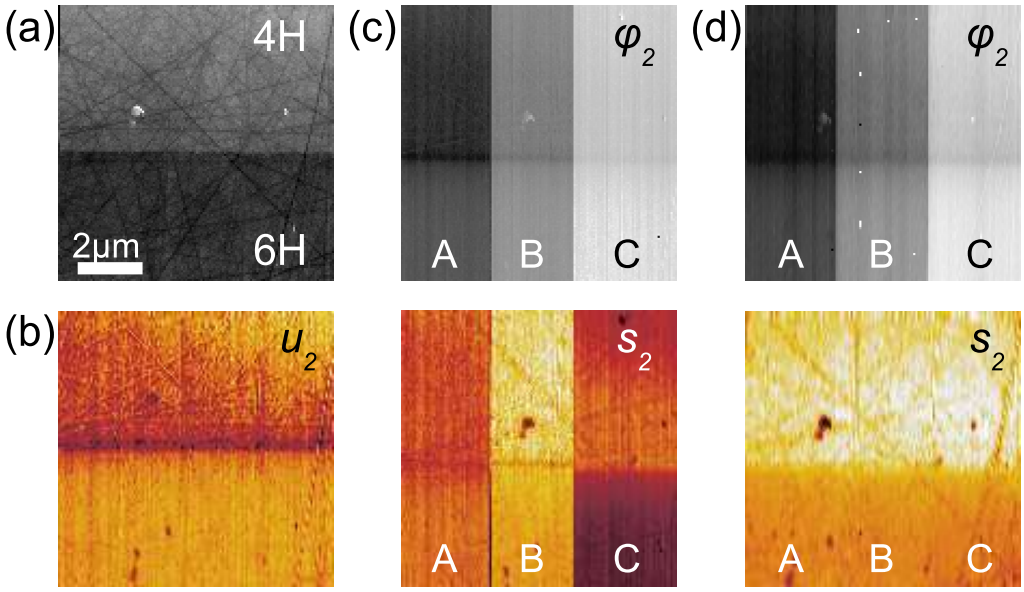


Figure 6.15: (a) Topography of a 6H-4H SiC polytype transition, exhibiting a≈ 2 nm height step generated in the process of sample surface polishing due to different hardness of 4H and 6H SiC polytypes. (b-d) Second-harmonic optical signal from the same sample obtained by (b) the non-interferometric, (c) homodyne and (d) pseudo-heterodyne signal detection techniques.

When this is not the case, an image like Fig. 6.15(c) may be obtained. It contains the same

SiC polytype transition, but imaged using the homodyne detection. The experiment was performed under the same conditions as before, except for the vibrating mirror which was replaced by a stationary one. To obtain the cosine and the sine components of the measured signal harmonic, each line was scanned twice, with $\Delta l = \lambda/8$ shift of the reference mirror between the two passes. With three different initial reference phase offsets ψ_R in sequence (Fig. 6.15(c), parts A-C), three different amplitude and phase contrast between the 4H and 6H SiC polytypes are obtained. Furthermore, the contrast variation turns out to be strong enough to cause even a slight contrast inversion in part A.

Finally, the non-interferometric detection was performed in the same setup. To this end, the reference beam was simply blocked. Although the measurement was performed under the same conditions as before, the resulting image of the same sample area, Fig. 6.15(b) displays an inverse amplitude contrast between the two SiC polytypes compared to Fig. 6.15(d). The origin of such contrast inversion should be sought in the Eq. 6.34. It contains the cosine factor $\cos(\varphi_n - \psi_0)$ which translated the phase difference into the amplitude contrast so that no definite conclusion about the near-field signal can be reached based on such measurements.

In conclusion, we have seen that the non-interferometric detection method is not suitable for near-field microscopy of samples exhibiting variations in the near-field signal phase. The homodyne method is significantly better in that regard and is in fact able to measure distinct optical resonances as shown in [63] and [64]. Still, the homodyne method is not sufficient for near-field imaging of weak optical contrasts. Such contrast may arise from e.g. structural differences between isomers or polytypes of the same material, or from different materials with closely separated resonances. For such application, the pseudo-heterodyne method should be used since it succeeds in a complete multiplicative background suppression. Furthermore, the pseudo-heterodyne method enables simultaneous measurement of the near-field signal amplitude and phase, thus reducing the image acquisition time by a factor two compared to the homodyne method. In comparison with the heterodyne s-SNOM, the phase-modulation technique provides the advantage of simpler and more compact experimental realization together with the applicability in a much broader spectral range.

7 Crystal Structure Mapping by Local Phonon-Polariton Spectroscopy

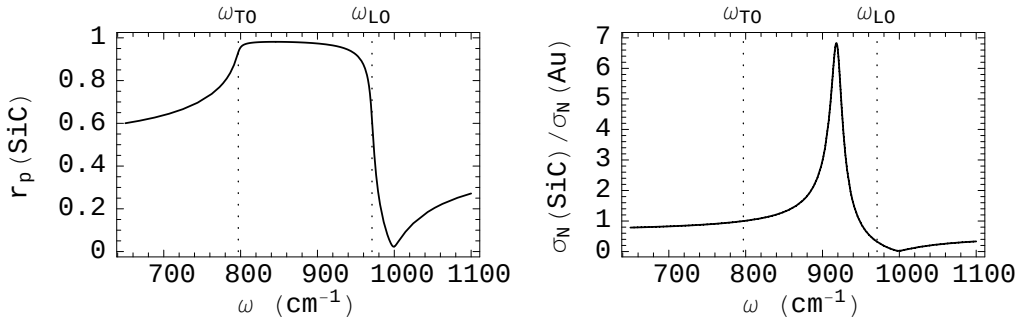
7.1 Introduction

Infrared spectroscopy is a powerful technique for determination of the chemical and structural composition of various organic and inorganic materials based on their spectral "fingerprints". Infrared spectral fingerprints reflect the energies of vibrational modes in molecules and crystals which are highly specific for the given material. The energies of vibrational modes mostly lie in the 30meV - 400meV range, coinciding roughly with the mid-infrared spectral region (about $2.5\,\mu$m to $30\,\mu$m).

In order for a given vibrational mode to be observable in infrared spectroscopy, the corresponding chemical bond must possess at least partially polar character. If this condition is satisfied, photons with energy close to the eigenfrequency of such vibrational mode are efficiently absorbed, and a minimum in transmission through the sample occurs. Formally, this effect can be described by an increase in the extinction coefficient $\kappa = \mathrm{im}(\sqrt{\varepsilon})$, where ε is the dielectric function of the material. The Lorentz oscillator model in the form of Eq. 3.11 can be successfully applied to predict the behavior of the dielectric function ε around the vibration mode eigenfrequency provided the oscillator strength f and damping γ are known.

Particularly interesting is the case of polar crystals whose lattice vibrations couple to the light so strongly that the dielectric function ε turns negative in a limited frequency range above the eigenfrequency ω_0, as shown in Fig. 3.1(a). This region, known as the residual or *Reststrahlen* band, is bounded by the transverse and the longitudinal optical phonon frequencies defined in Sect. 3.3. Within the Reststrahlen band, the reflection coefficient

$r_p = (\sqrt{\varepsilon}-1)/(\sqrt{\varepsilon}+1)$ is close to unity and almost all light is reflected from the sample as shown in Fig. 7.1(a).



(a) (b)

Figure 7.1: (a) Far-field reflection coefficient $r_p = (\sqrt{\varepsilon}-1)/(\sqrt{\varepsilon}+1)$ of p-polarized EM waves normally incident on a SiC crystal. (b) Near-field signal σ_{N} emitted by an s-SNOM probing tip in contact with SiC crystal surface, normalized to the near-field signal σ_{N} on an Au surface.

Note that Fig. 7.1(a) applies only to the reflection of propagating waves under normal incidence. In Sect. 5.2 it was shown that the near-field interaction between a s-SNOM probing tip and the sample is mediated by evanescent waves which reflect with a different reflection coefficient $\beta = (\varepsilon-1)/(\varepsilon+1)$. The mutual interaction between the tip and the sample finally results in a scattered light spectrum like the one shown in Fig. 7.1(b), calculated according to Eq. 5.29.

A prominent feature of Fig. 7.1(b) is the sharp peak within the Reststrahlen band. In Sect. 3.2, it was argued that such behavior should be traced back to the resonant excitation of surface phonon polaritons. From Fig. 7.1(b), we can conclude that the near-field phonon-polariton resonance is much sharper than the far-field Reststrahlen band which comprises it. Considering in addition the huge improvement in the resolution compared to the far-field IR spectroscopy [45, 62], near-field infrared spectroscopy clearly represents a very sensitive technique for material identification. In that regard, it has already been shown that the infrared s-SNOM can identify materials exhibiting near-field phonon-polariton resonances [63]. Furthermore, even in the absence of a polariton resonance, s-SNOM can distinguish and identify materials based on their IR absorption lines, as demonstrated for two polymers in a blend [46] and single viruses [65].

In each of the above cases, the material identification was performed by recording the near-field spectrum of the sample and comparing it to the theoretical prediction. In the

following, it will be shown that the same procedure can be applied to differentiate between crystals of the same chemical composition but different crystal structure. Although near-field optical contrasts due to differences in crystal structure are rather weak, the required degree of sensitivity can be achieved by exploiting the phonon-polariton near-field resonance. This enables the crystal quality degradation due to radiation damage to be measured and the polytypes of the same material to be distinguished and identified by s-SNOM. The former can be done in a quantitative way by using the monopole model introduced in Chapter 5, and the latter additionally requires the use of the pseudo-heterodyne detection method from Sect. 6.4.2 in order for the measured optical contrasts to be reproducible.

The application of s-SNOM for crystal quality mapping and polytype identification will be demonstrated for SiC crystals. The primary reason for performing the experiments with SiC is its phonon-polariton resonance around $\lambda = 10.75\,\mu m$, lying within the wavelength range covered by tunable CO_2 lasers. This greatly facilitates the experiments since CO_2 lasers represent the most powerful and reliable sources of coherent mid-IR radiation. Beside the experimental convenience, SiC is on its own a prospective material for high-power and high-temperature electronic applications[184]. A brief survey of its most important properties will thus be presented first.

7.2 Properties of SiC

7.2.1 Electric Properties

SiC is an exceptional semiconducting material able to retain its semiconducting behavior even at temperatures above 500°C [185]. This characteristic is a consequence of the wide bandgap of SiC ($\approx 3\,$eV) and the resulting low intrinsic carrier concentration ($\sim 10^{-6}\text{cm}^{-1}$). Combined with the high breakdown field ($> 3\,$MV/cm) and excellent thermal conductivity ($\approx 5\,$W/cmK), this property permits SiC to sustain large power densities [186]. The high breakdown field and the wide energy bandgap of SiC enable much faster power switching than in silicon power-switching devices with an equivalent power rating [187]. Finally, all aforementioned characteristics together give SiC an additional advantage over Si when applied in high-power radio-frequency (RF) signal generation and amplification.

However, the huge application potential of SiC still remains unexplored because of the

several difficulties encountered with SiC electronic devices fabrication. One of the main obstacles, the poor quality and high price of SiC wafers, might have been recently resolved by a new "repeated α-face" (RAF) SiC crystal growth method [188]. Still, the insulating oxide layer on SiC needed to fabricate MOSFET devices is known to be thermally unstable and of unreliable insulation properties. This obstacle can be avoided e.g. by using the Metal-Semiconductor Field Effect Transistor (MESFET) design [186, 189] which does not require an insulating oxide layer and also allows for a faster operation than MOSFET. Finally, the contacting and other supporting electronic elements which can withstand high operating temperatures of SiC are being steadily improved. The ongoing efforts to develop and optimize SiC electronic devices can therefore benefit from an analytical tool that enables non-destructive high-resolution imaging of crystal structure and doping profiles.

7.2.2 Polytypism

SiC properties such as energy gap or intrinsic carrier concentration mentioned above may significantly differ from crystal to crystal depending on their exact structure. Apart from the possible defects in the crystal lattice, there exist variations in the structure of perfect SiC crystals due to different stacking sequence of layers along one direction in space. The freedom in the stacking order is a consequence of the fact that for each 2-D layer of densely packed spheres designated by A in Fig. 7.2, there are two possible arrangements of densely packed spheres on top of layer A. These two arrangements are labeled B and C in Fig. 7.2.

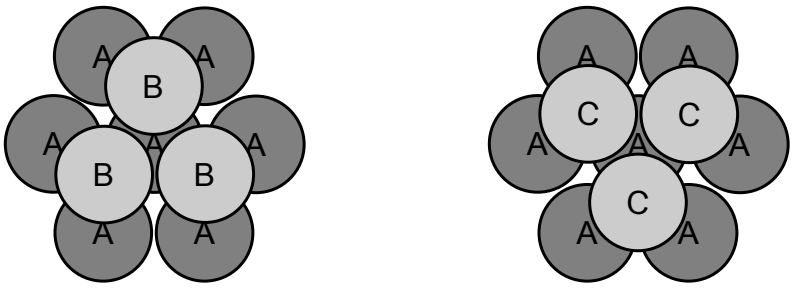
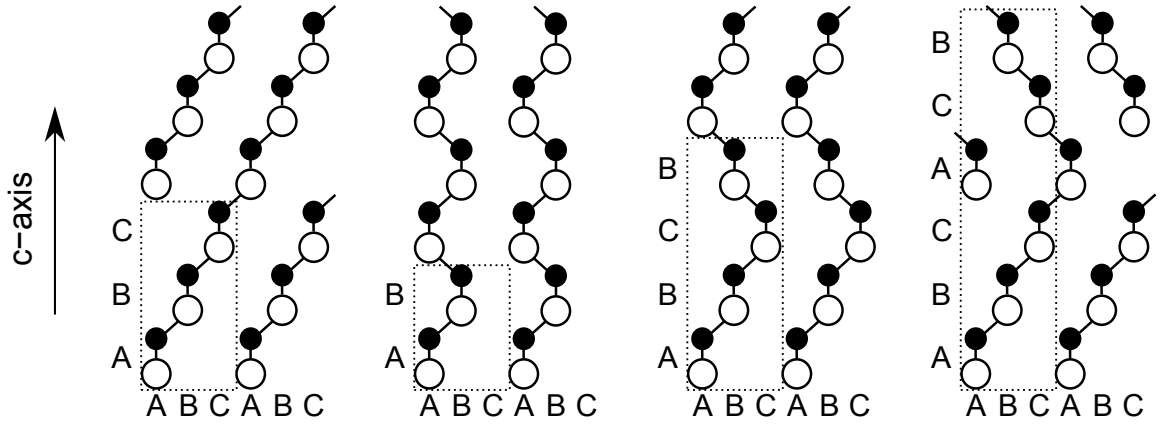


Figure 7.2: Two possible arrangements B and C of closely packed sphere layers on top of the layer A.

All crystals with the same stacking sequence, i.e. with the same ordering of layers A, B, and C, belong to the same **polytype**. More than 200 SiC polytypes are known, of which the most common ones are the cubic polytype 3C, hexagonal polytypes 2H, 4H, 6H and

rhombohedral polytype 15R. Stacking sequences of the first four of them are illustrated in Fig. 7.3 which displays the cross-sections through the crystallographic $(11\bar{2}0)$ plane. This plane is defined by the c-axis of the crystal and the base line delineated in Fig. 7.4.



(a) 3C-SiC (b) 2H-SiC (c) 4H-SiC (d) 6H-SiC

Figure 7.3: Stacking sequence of Si-C layers in the most frequently encountered SiC polytypes. Shown are cross-section along the crystallographic $(11\bar{2}0)$ plane with the base line designated in Fig. 7.4 below. Larger circles represent Si atoms, smaller disks stand for C atoms.

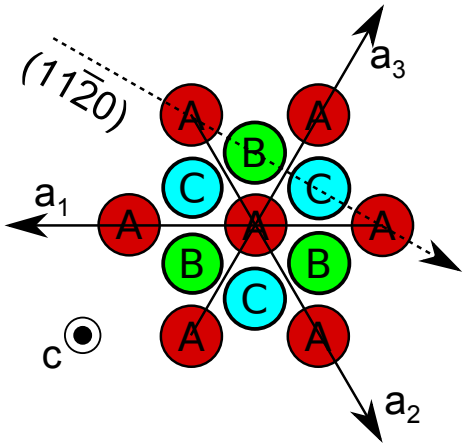


Figure 7.4: Projection of atom positions onto the (0001) plane in layers denoted by A, B, and C in Figs. 7.2 and 7.3. Full lines are the crystallographic axes, and the dashed line represents the base line of cross-sections shown in 7.3.

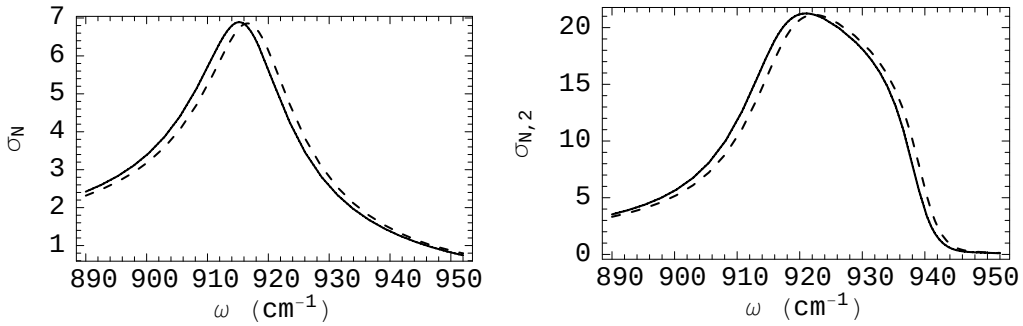
Properties relevant for electronic applications are given in Table 7.1 below for the three technically most interesting SiC polytypes: 3C, 4H and 6H. For comparison, corresponding properties of Si are also given in the same table.

	3C-SiC	4H-SiC	6H-SiC	Si
Bandgap (eV)	2.4	3.26	3.03	1.12
Breakdown field (MV/cm)	>1.5	3.0	3.2	0.3
Thermal conductivity (W/cmK)	3.2	4.9	4.9	1.3
Intrinsic carrier concentration (cm^{-3})	5×10^{-9}	5×10^{-9}	10^{-6}	10^{10}
Electron mobility (cm^2/Vs)	800	$\parallel$c: 900 $\perp$c: 800	$\parallel$c: 60 $\perp$c: 400	1430
Hole mobility (cm^2/Vs)	40	115	90	480
Saturated electron velocity (10^7 cm/s)	2.5	2	2	1

Table 7.1: Comparison of electrical properties of SiC and Si at $T = 300\,\mathrm{K}$.

7.2.3 Anisotropy

Regarding the optical properties of SiC, it should be noted that all SiC polytypes with the exception of the cubic 3C polytype are anisotropic. The 4H and 6H polytypes of SiC presented here are both uniaxial like many other technically relevant crystals including e.g. silicon. This enables the formalism from Sect. 5.3 to be applied here. Fig. 7.5 shows the difference between the cuts perpendicular and parallel to the c-axis, calculated for a 4H-SiC crystal using the monopole model (Eq. 5.29) with the respective near-field reflection coefficients $\beta_\perp$ and $\beta_\parallel$ given by Eqs. 5.37 and 5.39. A slight resonance shift of about $2\,\mathrm{cm}^{-1}$ is noticeable both for the DC signal and the signal demodulated at the second harmonic (2Ω).



(a) (b)

Figure 7.5: Near-field signal from a 4H-SiC crystal cut perpendicular to the c-axis (full line) and parallel to c-axis (dashed line), calculated according to the monopole model. Shown are (a) the signal emitted by a stationary s-SNOM probing tip in contact with SiC crystal, and (b) the second-harmonic demodulated signal from a vibrating tip (b). The values are normalized to the equivalent signal above an Au surface.

According to the theoretical prediction in Fig. 7.5, the signal level in the resonance maximum does not depend on the cut direction. Given that the anisotropy was treated only approximately here, it is possible that a more complete investigation will yield different results. No experimental data exists in this regard, so the effects of anisotropy on the near-field signal obviously require a more systematic investigation in the future. In the remainder of this chapter only crystals cut perpendicular to the c-axis will be compared, thus avoiding the possible scattering signal variations due to different orientations.

7.3 Structural Contrasts in Ion-Implanted SiC

7.3.1 Ion Implantation in SiC

The study of the near-field contrasts arising from differences in the crystal structure requires a modification of the structure to be performed without influencing the chemical composition of the sample. One possible way to achieve such structural modifications is to damage the crystal by highly energetic ions impinging on the surface of the crystal. The ions cannot be stopped immediately at the point of impact with the surface of the crystal, but they gradually lose their energy in a series of collisions with the lattice. Thereby the atoms in the crystal lattice can be displaced from their initial positions, developing point defects in the lattice. Since the incoming ions can be scattered in random directions after each impact, the distribution of ion end positions is quite complex. In general, random deviations of ion trajectories from a straight line in the crystal are known as straggling. The average final position of implanted ions strongly depends on the ion mass and energy, with lighter and faster ions exhibiting larger straggling [190]. If the ions are light and fast enough, they are eventually stopped far below the surface. However, the damage to the crystal lattice close to the surface still remains as the consequence of the implantation, offering the possibility to investigate the effects of degrading crystal quality on the near-field signal without a change in the chemical composition of the material.

In the experiment presented here, Be^{2+} ions with an energy of 60 keV were used to achieve the desired effect. The implantation was performed in a focused ion beam facility by R. Wernhardt at the Ruhr-Universität Bochum, Lehrstuhl für Angewandte Festkörperphysik. The expected distribution of Be ions upon implantation in SiC was calculated using SRIM software (Stopping and Range of Ions in Matter, [191]), giving an average depth of stopped

Be ions $\bar{Z} = 300\,\text{nm}$ and the average straggle $\Delta Z = 100\,\text{nm}$. Monte-Carlo-type simulations of the same experiment using the TRIM code [191] yield similar results, with $\bar{Z} = 275\,\text{nm}$ and $\Delta Z = 75\,\text{nm}$. The calculated concentration of Be ions as a function of the depth below the SiC crystal s

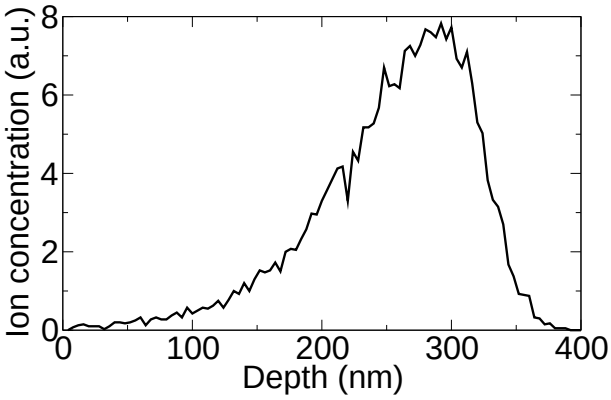


Figure 7.6: TRIM-calculated concentration of implanted 60keV Be^{2+} ions as a function of depth below the SiC crystal surface.

Given the total Be^{2+} ion dose of $10^{15}\,\text{cm}^{-2}$, the concentration of Be atoms within the first 100 nm below the surface amounts to only about 60 Be atoms per million Si and C atoms according to the TRIM simulation, and is thus negligible. As a consequence, any near-field optical contrast observed between the implanted and unimplanted areas of the crystal can be attributed with high certainty to the differences in the crystal structure, and not to the chemical composition of the material.

7.3.2 Near-field Infrared Images of FIB-Patterned SiC

The focused ion beam (FIB) implantation allows arbitrary implantation patterns to be realized on the surface of the crystal. This feature was exploited to facilitate the comparison between the near-field response if the ion-beam damaged and intact SiC crystal. For this purpose, the pattern shown in Fig. 7.7 was ”drawn” on the surface of a 6H-SiC crystal cut perpendicular to its *c*-axis and polished prior to the FIB implantation. The FIB-patterned crystal was then imaged in a s-SNOM, using the homodyne setup described in Sect. 6.3.3. A Ti-Pt-coated probing tip (MikroMasch, model CSC37/Ti-Pt) with the resonance frequency of about $\Omega = 30\,\text{kHz}$ was used in the experiment. The tip radius was specified by manufacturer to be about $R = 35\,\text{nm}$, and the tapping amplitude was set to $A = 25\,\text{nm}$. Since s-SNOM is based on the atomic force microscope (AFM), it enables the sample topography

to be recorded simultaneously with the optical scattering signal. The information provided by s-SNOM optical and mechanical channels will be now compared before proceeding to the near-field infrared spectra.

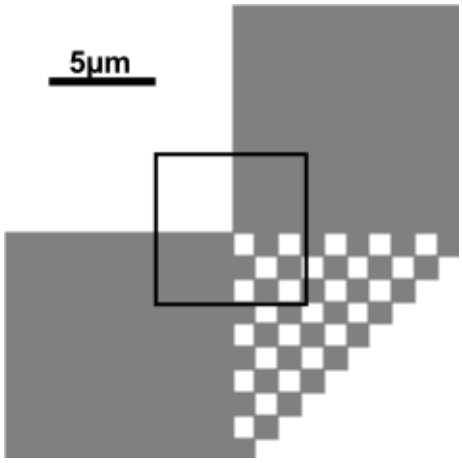
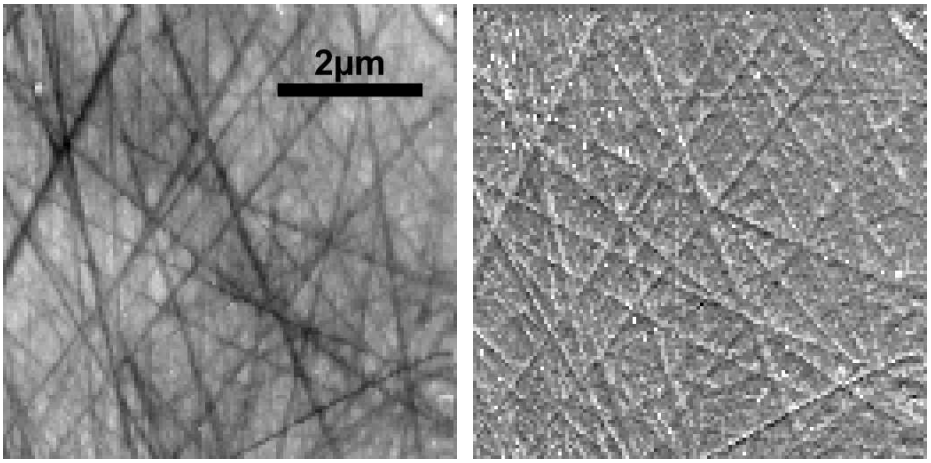


Figure 7.7: Pattern created in 6H-SiC crystal by focused ion beam implantation of 60keV Be^{2+} ions. The area imaged in the experiment is delineated by the black rectangle.

The experimentally obtained topography image of the FIB-patterned SiC surface is shown in Fig. 7.8(a). Obviously, the topography image reveals little information apart from the 10-20 nm deep scratches produced in the process of crystal surface polishing. A slight correlation between Fig. 7.7 and Fig. 7.8(a) is a consequence of material expansion (swelling, [192]) due to increased disorder in the crystal. However, the height variations due to swelling would be hard to notice without looking at the implantation pattern (Fig. 7.7), and the transition between the implanted and unimplanted areas is so blurry that the small implanted squares are not recognizable at all. The mechanical phase, shown in Fig. 7.8(b) reveals just the polishing scratches, thus providing even less information than the topography.

Fig. 7.9 contains the simultaneously recorded optical image obtained by demodulating the detector output signal at the second harmonic of the tapping frequency ($2\Omega \approx 60\,\mathrm{kHz}$) obtained with the CO_2 laser tuned to the wavelength $\lambda = 925\,\mathrm{cm}^{-1}$. From Fig. 7.9 it is clear that the optical signal reveals much more information than the topography. Especially the amplitude image in Fig. 7.9(a) reveals every detail of the implanted structure from Fig. 7.7. The optical phase image in Fig. 7.9(b) also resembles the implanted pattern, but the contrast is much weaker and less clear than in the case of the optical signal amplitude. Interestingly,



(a) (b)

Figure 7.8: Height (a) and tip vibration phase (b) image of the FIB-patterned SiC. Image size: 6 µm×6 µm

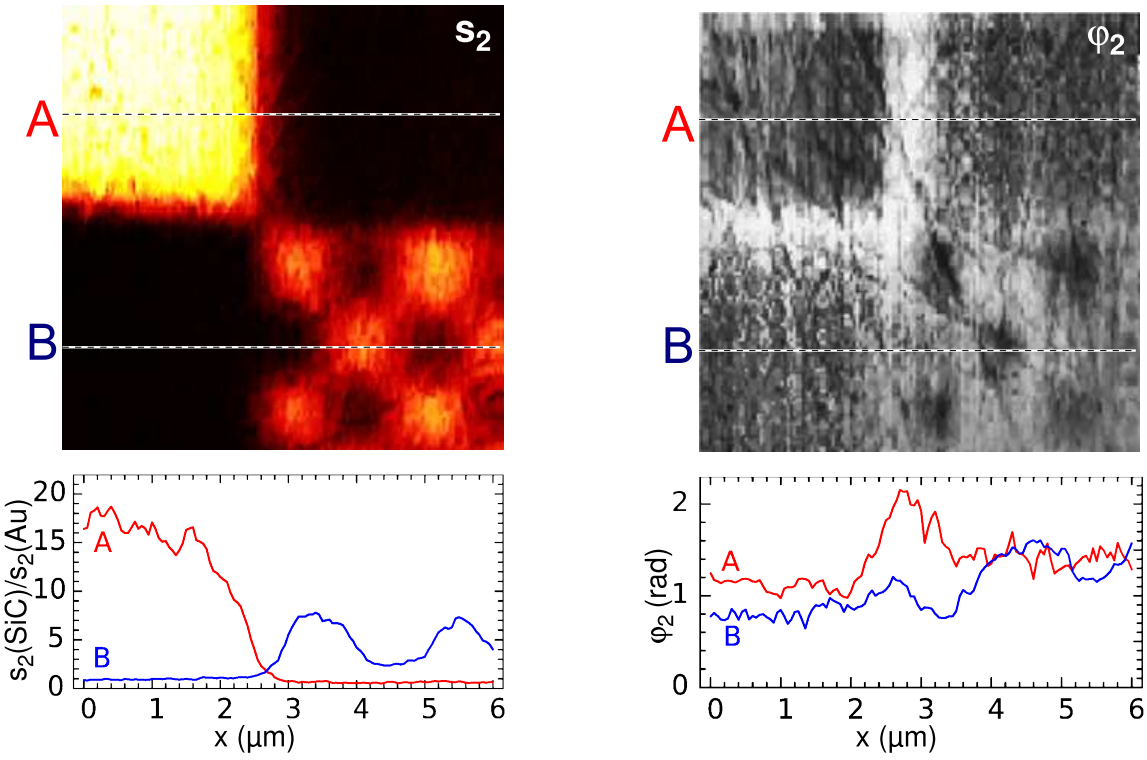
exactly the opposite situation is found in the optical amplitude and phase images at another wavelength, $\lambda = 899\,\mathrm{cm}^{-1}$, as shown in Fig. 7.10.

The reason for this behavior of the scattering signal as well as the appearance of bright edges in the phase image in Fig. 7.10(b) will become clear when the complete near-field spectra are examined in the next section.

7.3.3 Near-field Spectra of Ion-implanted SiC

The complete near-field signal spectra will be extracted at four different positions marked as A,B,C and D in Fig. 7.11. These four positions comprise two unexposed areas (A and B) and two areas exposed to the ion beam (C and D). However, each of the four areas has received a different effective implantation dose, as can be concluded from the different amplitudes of the signal in Fig. 7.9(a) at each of the positions A-D. The origin of this difference is explained below.

Starting with the area D, we can assume it has received a full dose of $10^{15}\,\mathrm{cm}^{-2}$. However, this dose consists of the ions which hit the SiC surface within the region D as well as those which hit the surrounding area but recoil into D. According to the TRIM simulation, the average lateral straggle ΔX of 60 keV Be^{2+} ions normally incident on a SiC surface is about 90 nm, with significant number of ions stopped at a lateral distances as large as $X = 250\,\mathrm{nm}$ away from the entry point. In addition, a focused ion beam necessarily contains ions incident under a range of angles around the normal which contributes to an even larger



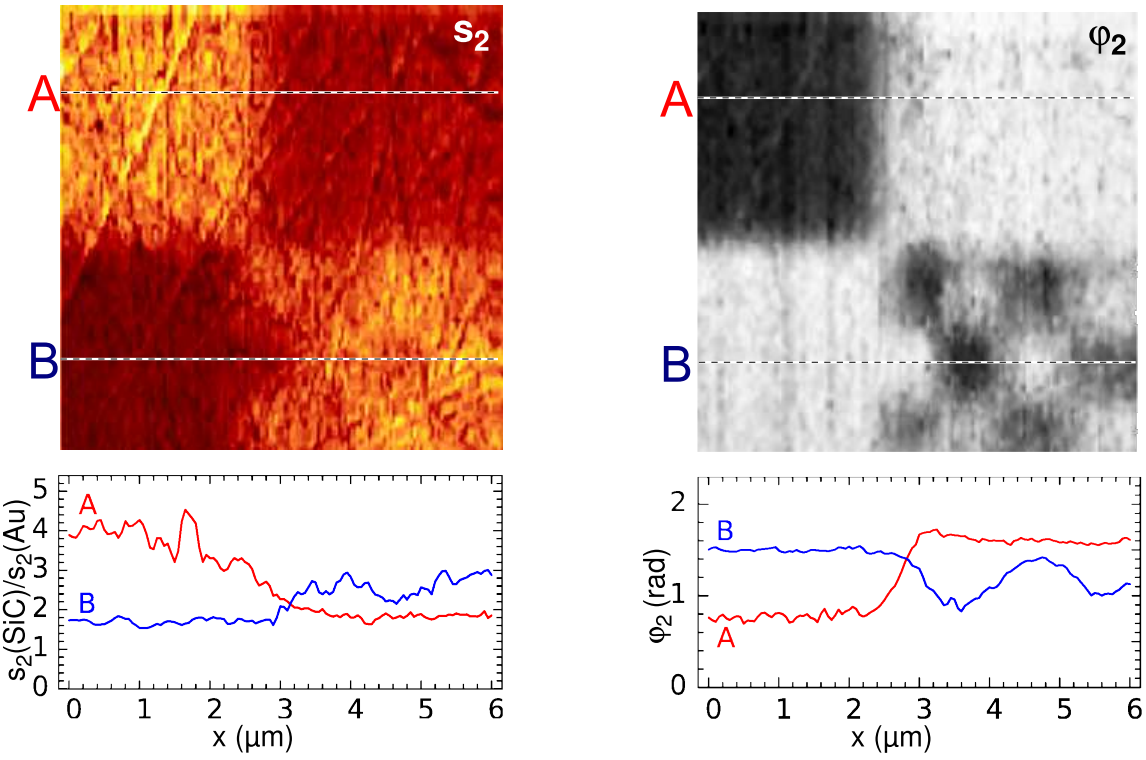
(a) (b)

Figure 7.9: Near-field optical images of focused-ion-beam patterned SiC. Shown are the demodulated second harmonic signal amplitude s_2 (a) and phase φ_2 (b) recorded at $\omega = 925\,\mathrm{cm}^{-1}$.

lateral straggle. Finally, the ion beam focus has a Gaussian profile with a finite width, not guaranteed to be better than 200 nm FWHM in this experiment [193]. All this effects average out on a large implanted surface like the area D, but not in an area like C which is surrounded by unimplanted regions. The damage induced by ion implantation in the region C is thus lower than in the region D.

Since there is an amount of ions ”missing” in the region C, it obviously ended up in the surrounding regions similar to the one marked by B. For this reason, the region B accumulated some damage from the ions targeted at the region C and its response will differ from the completely unimplanted SiC. Of the four marked regions, only the region A is far away from any implanted regions and can be assumed to be free of the damage to the crystal structure.

One important consequence of the above effects are blurred edges of both large and small squares in Figs. 7.9 and 7.10. In Sect. 7.6 it will be proved this is indeed a consequence of the implantation process and not the s-SNOM resolution limit since under different



(a) (b)

Figure 7.10: Near-field optical images recorded at $\omega = 899\,\mathrm{cm}^{-1}$. Shown are the demodulated second harmonic signal amplitude s_2 (a) and phase φ_2 (b) .

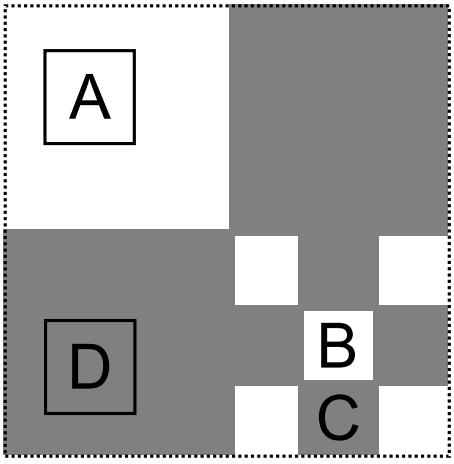


Figure 7.11: Scanned area of the implanted pattern from Fig. 7.7 with marked areas A-D where the near-field spectra are extracted.

implantation conditions much sharper edges can be achieved.

Before the spectra can be assembled from single scans, they have to be normalized to a signal on a reference material (see Sect. 2.5.1). Since no reference material was available in

the scanned region, the normalization had to be performed in two steps. First, the spectrum of unimplanted SiC was obtained from a separate measurement on the same crystal but at a different location, close to a 50 nm thick Au film. The so-obtained spectrum was then taken as the normalized spectrum of the unimplanted SiC in the region A. The signal in other areas (marked B-D) was then scaled at each wavelength, so that their ratios to the signal in the area A remain the same as measured in the experiment. The final result is shown in Fig. 7.12.

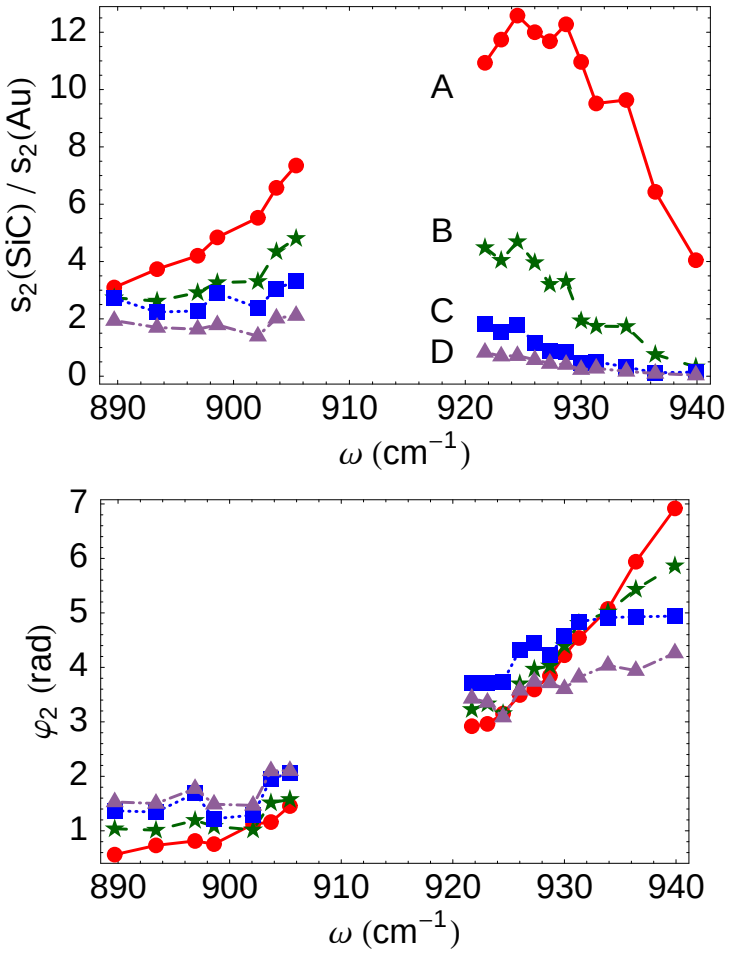


Figure 7.12: Second-harmonic near-field spectra of SiC sample with different Be^{2+} implantation doses A-D.

From Fig. 7.12 we see that the magnitude of the near-field resonance diminishes rapidly with the damage to the crystal lattice. A shift of the resonance maximum to lower frequencies with the increasing ion dose seems to take place as well, but it cannot be directly observed due to the missing CO_2 laser lines in the range between $\omega = 905\,\mathrm{cm}^{-1}$ and $\omega = 920\,\mathrm{cm}^{-1}$.

The phase spectra in Fig. 7.12(b) exhibit a decreasing slope as the implantation dose increases. This is an indication of broader and less pronounced resonances, consistent with the conclusions drawn from the amplitude spectra. The most heavily implanted region (D) has the highest phase $\varphi_{2,D}$ around $\omega = 900\,\mathrm{cm}^{-1}$. Due to its broad resonance, the total phase rotation in the region D is the lowest of all investigated regions so the signal in the region D has the lowest phase at $\omega = 940\,\mathrm{cm}^{-1}$. The crossing with the phase $\varphi_{2,A}$ of the unimplanted region A occurs around the unimplanted SiC resonance maximum at $\omega = 925\,\mathrm{cm}^{-1}$, thus explaining the large amplitude, but small phase contrast in Fig. 7.9. The phase signal $\varphi_{2,C}$ of the small implanted squared (region C) crosses $\varphi_{2,A}$ at a different wavelength ($\omega \approx 930\,\mathrm{cm}^{-1}$), so it does not lie between the phases $\varphi_{2,D}$ and $\varphi_{2,A}$ in the frequency region between $\omega \approx 920\,\mathrm{cm}^{-1}$ and $\omega \approx 930\,\mathrm{cm}^{-1}$. This explains the appearance of bright edges in the phase image in Fig. 7.10(b).

7.4 Quantitative Determination of SiC Crystal Quality

The spectra in Fig. 7.12 display a clear trend towards a broader and weaker resonance as the damage to the crystal lattice increases. It is therefore obvious that there is a correlation between the near-field spectra and the structure of the crystal. To express this relationship in a quantitative way, a theoretical model capable of reproducing the measured values sufficiently well is required. This was the principal motivation for deriving the monopole model in Chapter 5. The monopole model in the form of Eq. 5.34 will be employed here to determine the structural composition of SiC most consistent with the observed near-field spectra in the regions A-D.

7.4.1 Effective Medium Approximation

As the first step towards a quantitative statement about the ion-implantation damage in SiC, the amount of damage needs to be related to the dielectric function of the material. In general, a composite medium made of two different materials can be considered as a homogeneous material for the purposes of far-field spectroscopy if its components are mixed on the scale much smaller than the wavelength. In the context of s-SNOM, the field is confined to a volume V roughly equal to a half-sphere of radius equal to the probing

tip radius R, $V \approx 2/3\,\pi R^3$. The volume V represents a rough upper bound to the volume over which the material components have to be homogeneously mixed. The damage due to the ion implantation is limited to dimension on the order of 1nm and is smaller than the probing volume radius (20-30 nm). This enables a single dielectric constant ε to be assigned to the entire volume of the material within the probing radius. Such an approach is generally known as the effective medium approximation (EMA). The dielectric function of such effective medium must lie between the Wiener bounds given by

$$\varepsilon_{NS} = x_1\,\varepsilon_1 + x_2\,\varepsilon_2 \tag{7.1}$$

and

$$\varepsilon_{FS} = \left(\frac{x_1}{\varepsilon_1} + \frac{x_2}{\varepsilon_2}\right)^{-1}, \tag{7.2}$$

where x_1 and $x_2 = (1 - x_1)$ are the volume fractions of the two compounds and ε_1 and ε_2 their respective dielectric functions. The first bound, ε_{NS}, applies to the case of no screening (NS), e.g. lamellae of material 1 and material 2 lying parallel to the direction of light propagation. The second bound, ε_{FS}, applies to the case of full screening (FS), realized e.g. by lamellae of the two materials lying perpendicular to the direction of light propagation.

More sophisticated models like the Bruggeman EMA [194] result in values for effective ε closer to the limit of no screening, ε_{NS}, than to the full screening, ε_{FS}. In addition, measurements of the far-field transmission and reflection spectra of ion-implanted SiC [195] seem to be consistent with the dielectric function given by Eq. 7.1. For this reason, the use of an expression more complicated than Eq. 7.1 would not be justified here.

The two components whose dielectric functions enter Eq. 7.1 still need to be determined. It might seem obvious that they comprise the initial and final stages of the implantation process, i.e. the purely crystalline and the purely amorphous SiC. However, an ion transferring its kinetic energy to the SiC crystal lattice does not cause its immediate amorphization. Instead, point defects such as vacancies and interstitials are first produced. These defects increase the imperfection of the lattice and therewith the phonon-polariton damping, represented by the damping coefficient γ. This suggests the dielectric function of the form

$$\varepsilon(\omega) = x\,\varepsilon_{\mathrm{aSiC}}(\omega) + (1-x)\,\varepsilon_{\mathrm{cSiC}}(\gamma,\omega), \tag{7.3}$$

where the parameters γ and x should be adjusted until the best fit to the experiment is obtained. The parameter γ is thereby the damping coefficient of the damaged but still crystalline SiC (c-SiC), while the parameter x signifies the fraction of the amorphous SiC (a-SiC) material in the total volume.

7.4.2 Damping and Amorphization

With the aid of Eq. 7.3, ion-implantation damage to SiC can be expressed in terms of amorphization degree x and damping coefficient γ. Both of these parameters are known for the unimplanted region A. In particular, $x = 0$, and the damping coefficient of high-quality crystalline SiC is $\gamma = 6 \pm 1\,\mathrm{cm}^{-1}$.

To be able to reproduce the unimplanted SiC spectrum (A) using Eqs. 5.44 and 5.24, optical constants determining the dielectric function of SiC are required. According to [78], the 6H-SiC crystal used in this experiment can be described by the transverse and longitudinal optical phonon frequencies $\omega_{TO,\parallel} = 788\,\mathrm{cm}^{-1}$ and $\omega_{LO,\parallel} = 964\,\mathrm{cm}^{-1}$ parallel to the c-axis, and $\omega_{TO,\perp} = 797\,\mathrm{cm}^{-1}$ and $\omega_{LO,\perp} = 970\,\mathrm{cm}^{-1}$ perpendicular to the c-axis. Finally, the dielectric functions are obtained from Eq. 3.11 with $\varepsilon_{\infty,\parallel} = 6.72$ and $\varepsilon_{\infty,\perp} = 6.56$ and the anisotropy is accounted for using the quasi-electrostatic reflection coefficient β given by Eq. 5.39.

The parameters of Eq. 5.24 pertaining to the probing tip are its measured vibration amplitude $A = 25\,\mathrm{nm}$ and the radius $R \approx 35\,\mathrm{nm}$, specified by the manufacturer (MikroMasch, CSC37/Ti-Pt). A good fit to the experimental spectrum A is obtained using the parameter $g = 0.71e^{0.11i}$. This is slightly different form $g = 0.7e^{0.06i}$ which was found to reproduce the near-field spectra of 4H-SiC shown in Fig. 5.16. Since the 4H-SiC crystal was cut parallel to the c-axis, whereas the 6H-SiC examined here was cut perpendicular to the c-axis, the small difference in the constant g might actually be needed to compensate for the approximate treatment of anisotropy effects, discussed in Sect. 5.3. The effective tip length of $2L = 0.6\,\mu\mathrm{m}$ was used, as in Chapter 5.

Finally, the most problematic part of Eq. 5.44, the weighting factor c of the reflected waves, was set to $c_0 = 0.2e^{2\pi L/\lambda \cos 45}$ to account for the illumination under 45° as well as for the fact that the reference spectrum for the area A was obtained close to, but not exactly on the boundary between Au film and SiC surface. The resulting spectrum of the region A is shown in Fig. 7.13 and provides a reasonably good fit to the experiment.

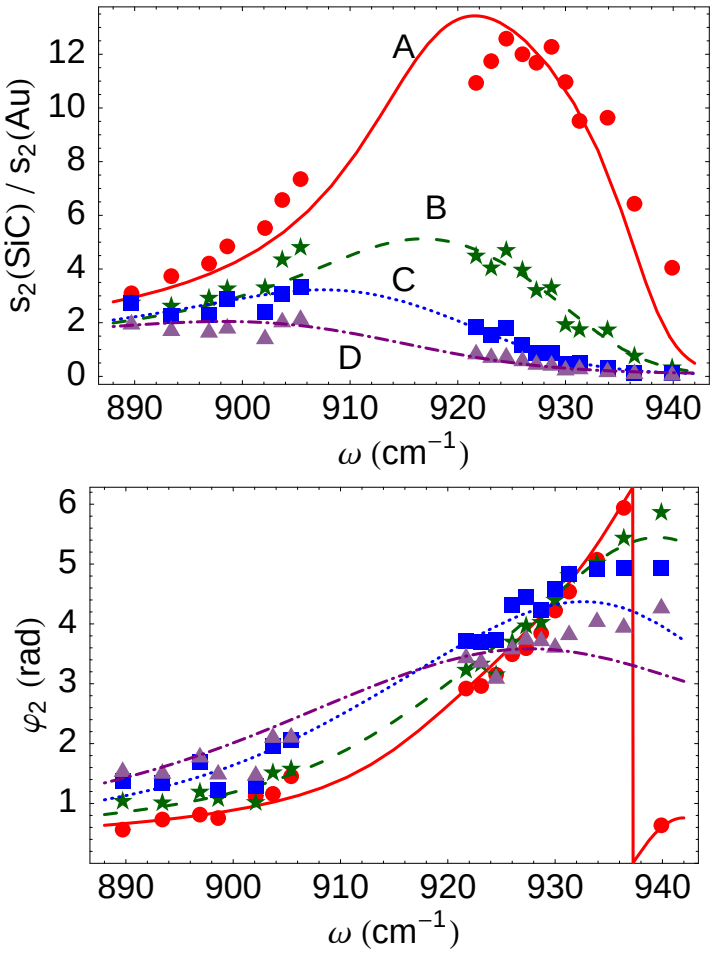


Figure 7.13: Measured second-harmonic near-field spectra of SiC sample with different Be^{2+} implantation doses A-D (points), together with the theoretical signal spectrum prediction by the monopole model with parameters from table 7.2 (lines).

Fig. 7.13 also contains the the spectra of the regions B-D. They were calculated the same way as the spectrum A, with two differences. First, the entire pre-factor $(1+cr_p)$ in Eq. 5.44 was replaced by $(1+cr_p)(1+cr_{p,A})/(1+c_0 r_{p,A})$ to properly account for the normalization procedure described in Sect. 7.3.3 above. The weighting factor c was thereby set to $0.4e^{2\pi L/\lambda \cos 45}$ since all data except the reference spectrum assigned to A were obtained far from the Au film. However, due to the small dimensions of areas B and C, the reflections from adjacent squares probably caused some cross-talk between the spectra B and C. This effect was neglected here because it could not be precisely quantified.

Second, the parameter x in Eq. 7.3 was different from zero in areas B-D, so that the dielectric function $\varepsilon_{\mathrm{aSiC}}$ of amorphous SiC from [196] had to be included. Additionally, a

variable damping constant γ was used in the dielectric function $\varepsilon_{\mathrm{cSiC}}$ of crystalline SiC, as required by Eq. 7.3. The actual values of x and γ have been determined by numerically searching for the best fits to the experimental spectra. It should be noted that a simple linear regression was not applicable in this case because both parameters enter the Eq. 5.24 in a highly non-linear way. Attempts to use automated non-linear regression algorithms returned values for x and γ which obviously produced worse fit to the data than the values obtained manually by a trial-and-error procedure. To obtain the best values of x and γ and their respective uncertainties by an objective method, a systematic search for the global minimum of an error function was eventually performed in the two-dimensional x-γ space. In order to correctly reproduce both the minima and the maxima of the experimental spectra, it was thereby necessary to construct an error function f_e which measures mostly the relative, rather than the absolute deviation of the calculated from the measured data at each spectral point. In particular, the following function was used:

$$f_e = \left(\frac{\left|\rho_{2,\mathrm{exp}} - \rho_{2,\mathrm{theo}}\right|}{\left|\rho_{2,\mathrm{exp}}\right| + 0.1\left|\rho_{2,max}\right|} \right)^2, \tag{7.4}$$

where $\rho_{2,\mathrm{exp}}$ and $\rho_{2,\mathrm{theo}}$ are, respectively, the experimental and theoretical complex second-harmonic optical contrast between SiC and Au. A constant equal to 10% of the maximal amplitude contrast $\left|\rho_{2,max}\right|$ in each spectrum was added to the denominator of Eq. 7.4 to reduce the influence of the large relative error contained in the experimental points with low amplitudes. Finally, the results obtained by this procedure are summarized in Table 7.2, with the error bounds estimated from the points in x-γ space where the error function $f_e(x,\gamma)$ exceeded its lowest value by more than 50%.

Region	γ/cm^{-1}	x
A	6 ± 1	0
B	11 ± 2	0.04 ± 0.01
C	11 ± 3	0.13 ± 0.01
D	15 ± 5	0.20 ± 0.03

Table 7.2: Parameters γ and x in Eq. 7.3 obtained by fitting the prediction of the monopole model to the near-field spectra of ion-implanted SiC shown in Fig. 7.13.

The plausibility of the dielectric function in the from of Eq. 7.3 and the values obtained for the fit parameters γ and x will be now discussed in more detail. To this end, it has already been mentioned in Sects. 7.3.1 and 7.4.1 that ion implantation tends to cause point defects,

and the material becomes amorphous only after enough point defects have accumulated close to each other. Furthermore, the far-field reflectivity and transmittivity measurements in [195] indicate that upon implantation of B^+ ions (very similar to Be^{2+}ions used in this work), the pure crystalline SiC is first converted to the defective, but still crystalline material, with negligible presence of the amorphous phase. The fraction of the amorphous SiC in the total volume starts to increase only after the fraction of the defective crystalline SiC has saturated.

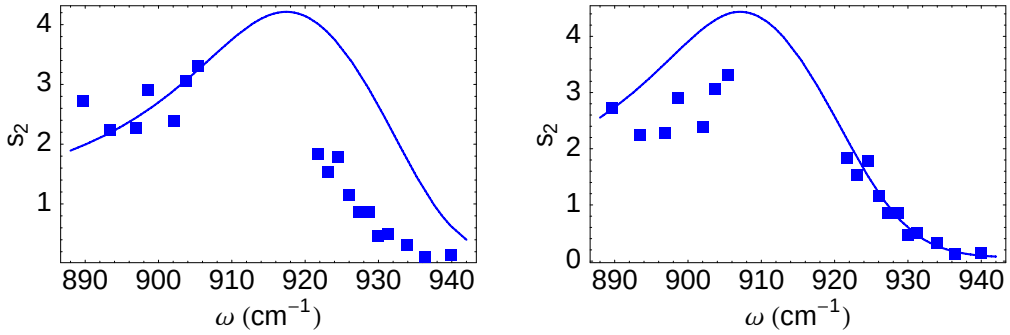


Figure 7.14: Second-harmonic near-field spectra of the implanted area C modeled by adjusting (a) only the damping coefficient γ, or (b) only the amorphization degree x.

Such observations are similar to the conclusion that can be made based on the fit parameters in Table 7.2. In particular, the main difference between the intact area A and the low implantation dose in area B is the almost doubled damping γ, accompanied by only 4% increase in the amorphization degree x. In contrast, the damping increases by only about 15% from B to C and from C to D, whereas the amorphization degree rises by $\approx 8\%$ in each step. This indicates that enough damage has been accumulated between the regions B and C to bring the damping in the crystalline SiC close to saturation value from which point on the amorphization degree starts rising more rapidly. However, a sharp separation between these two phases found in [195] is not observed here.

The necessity of treating the implanted SiC as the mixture of the described two components is further supported by Fig. 7.14. It displays an attempt to reproduce the measured spectrum C by adjusting only one parameter. In particular, changing the parameter x while keeping $\gamma = 6$ yields the result shown in Fig. 7.14(a). This model, equivalent to representing the medium by a mixture of the defect-free crystalline SiC and amorphous SiC, fails to correctly predict the near-field signal at frequencies below the resonance. Similarly, allowing only the damping of the crystalline SiC to be changed and setting $x = 0$, the

spectra in Fig. 7.14(b) are obtained, falling into disagreement with the experiment at frequencies above the resonance maximum. The correct maximum position, signal amplitude and phase above and below the resonance are reproduced simultaneously and correctly only when values close to those listed in Table 7.2 are used. This permits a fairly quantitative determination of the crystal quality, limited mostly by the accuracy of the model dielectric function, Eq. 7.3, and the description of the tip-sample near-field interaction given by Eq. 5.34.

7.5 Identification of SiC Polytypes

Near-field optical contrasts due to the different implanted ion doses encountered in Sect. 7.4 were large enough for the homodyne detection method to yield reproducible results. In this section a more difficult task of identification of two SiC polytypes is presented. The polytypes to be distinguished were known to be nitrogen-doped 4H-SiC and 6H-SiC, with the concentration of N atoms on the order of 10^{18} per cm^3. Due to the relatively high level of doping, the dielectric function of the two SiC polytypes has to be constructed by adding the contributions to the susceptibility by both phonon- and plasmon-polaritons. The resulting expression reads[196]:

$$\varepsilon(\omega) = \varepsilon_\infty(1 + \frac{\omega_{LO}^2 - \omega_{\mathrm{TO}}^2}{\omega_{TO}^2 - \omega^2 - i\,\omega\,\Gamma} + \frac{\omega_p^2}{-\omega^2 - i\,\omega\,\gamma}). \tag{7.5}$$

A different set of parameters applies to each polytype, and considering that the anisotropy has to be accounted for, it yields a total of 24 parameters. The parameters, together with the values found in the literature are summarized in Table 7.3.

Parameter	6H-SiC$\parallel$c	6H-SiC$\perp$c	4H-SiC$\parallel$c	4H-SiC$\perp$c
ε_∞ [196]	6.72	6.56	6.78	6.56
ω_{TO} (cm^{-1})[78]	788	797	782	797
ω_{LO} (cm^{-1})[78]	964	970	967	971
Γ (cm^{-1})[197]	5.5	5.9	6.6	6.6
ω_p (cm^{-1})[197]	120	230	220	275
γ (cm^{-1})[197]	250	500	450	450

Table 7.3: Values of the parameters in Eq. 7.5 describing the investigated 6H and 4H SiC polytypes, parallel and perpendicular to the crystallographic c-axis.

The main difference between the 6H- and 4H-SiC parameters pertaining to phonon polari-

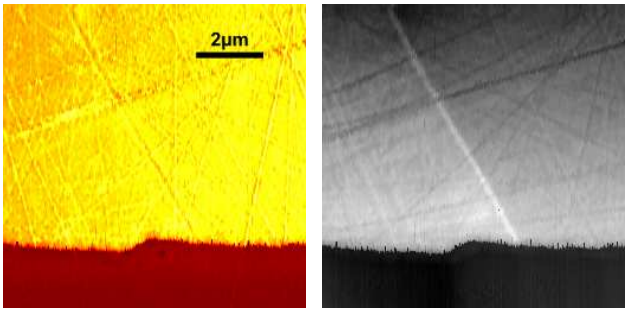
tons is the ca. $4\,\mathrm{cm}^{-1}$ shift of the ω_{TO} and ω_{LO} frequencies for electric field oscillations parallel to the c-axis. Perpendicular to the c-axis, the ω_{TO} and ω_{LO} frequencies are almost identical, which is not surprising since the structure of the two polytypes differs only along the c-axis. The damping coefficients Γ were taken from [197], where they were obtained as fit parameters to Raman scattering measurements on doped 4H- and 6H-SiC crystals. On the plasmon side, the most important difference between the polytypes is the higher plasma frequency ω_{p} for the 4H polytype. This difference can be traced back to the lower effective electron mass parallel to c-axis in 4H-SiC, found to be $m_{\parallel,4\mathrm{H}} = 0.48 m_0$, compared to $m_{\parallel,6\mathrm{H}} = 1.4 m_0$ in 6H-SiC. The estimate for the plasma frequency $\omega_p = \sqrt{4\pi n_N e^2/m_e}$ was obtained from the the dopant concentration $n_N = 1.7 \times 10^{18}\,\mathrm{cm}^{-3}$ which was found to be in the best agreement with the experiment, as will be explained later. The plasmon damping γ was then interpolated between the values given for different doping levels in [197] to the same value $n_N = 1.7 \times 10^{18}\,\mathrm{cm}^{-3}$.

Although the plasmon properties differ significantly between the polytypes, the plasma frequency ($\approx 250\,\mathrm{cm}^{-1}$) lies far below the investigated frequency range between 900 and $950\,\mathrm{cm}^{-1}$. The plasmon contribution to the dielectric function is thus small compared to the contribution by phonon polaritons. As a consequence, only weak contrasts between the polytypes are expected in the region of the near-field phonon-polariton resonance, requiring the use of the pseudo-heterodyne detection method (Sect. 6.4.2) to obtain reproducible results. Otherwise, spurious contrasts can be observed, as shown in Sect. 6.5.

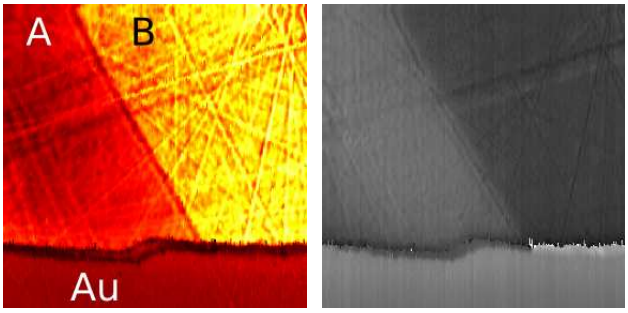
7.5.1 Near-field Optical Images of 6H/4H Polytype Transition

Near-field images of the transition between 4H-SiC and 6H-SiC polytypes obtained at $\lambda = 10.70\mu\mathrm{m}$ ($\omega = 935\,\mathrm{cm}^{-1}$) and $\lambda = 10.55\mu\mathrm{m}$ ($\omega = 948\,\mathrm{cm}^{-1}$) using the pseudo-heterodyne detection method are shown in Fig. 7.15. No contrast between polytypes in the optical amplitude image at $\omega = 935\,\mathrm{cm}^{-1}$ is observable, but a small phase contrast nevertheless exists at this wavelength. The image taken at $\omega = 948\,\mathrm{cm}^{-1}$ reveals both a weak but unambiguous contrast in the near-field signal amplitude and a larger contrast in the signal phase. This indicates that the polytypes can indeed be distinguished by purely optical means. For the purpose of later identification, the polytype to the left of the boundary will be denoted by A, and the polytype to the right of the boundary by B, as shown in Fig. 7.15(b).

For comparison, the simultaneously obtained topography of the sample is shown in Fig. 7.16,

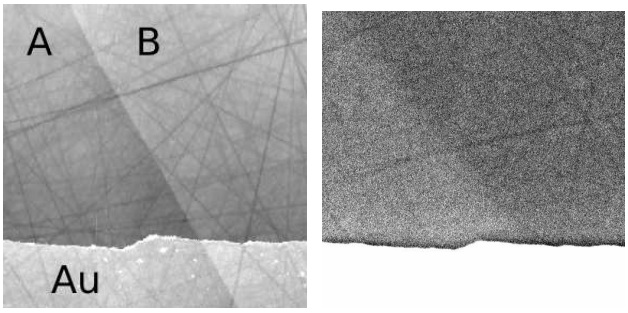


(a)



(b)

Figure 7.15: Near-field optical images of 4H-6H SiC polytype transition. Shown are the near-field signal amplitude s_2 (left) and phase φ_2 (right) at (a) 935 cm^{-1} and (b) 948 cm^{-1}. Image size: 9 µm×9 µm.



(a) (b)

Figure 7.16: (a) Topography and (b) SEM image of a 4H-6H SiC polytype transition.

together with a separately recorded scanning electron micrograph of the same area. The topography image displays a clearly visible height step across the boundary between the two polytypes, running from upper left towards lower right corner of the image. This step is most likely generated in the polishing process due to slightly different hardness of the polytypes. The elevated area at the bottom of the images in Fig. 7.16 is the Au film evaporated onto the SiC crystal surface for the purpose of signal normalization (see

Sect. 2.5.1). The same highly conductive Au surface is easily distinguished in the SEM image in Fig. 7.16(b). A slight contrast between the two polytypes is also present in the SEM image. However, the identification of the polytypes A and B based on either the topography or the SEM image is not possible without some further information on the properties of each of the two polytypes. Furthermore, without knowing that the topography and SEM images actually contain the 4H-SiC and 6H-SiC polytypes, the materials imaged in Fig. 7.16 (a) and (b) could not be identified at all.

7.5.2 Spectral Identification of SiC Polytypes

It will be shown now that the infrared near-field spectroscopy enables an unambiguous identification of the polytypes A and B based on their IR dielectric function given by Eq. 7.5 and parameters from Table 7.3. To this end the near-field spectra of the polytypes A (disks) and B (squares) from the polytype boundary in Fig. 7.16 are shown in Fig. 7.17, together with the theoretical near-field spectrum prediction for 6H-SiC (full line) and 4H-SiC (dashed line). Since the full line coincides with disks and dashed line with squares in Fig. 7.17, the polytype marked as A in Fig. 7.16 can clearly be identified as 6H-SiC and the polytype B as 4H-SiC.

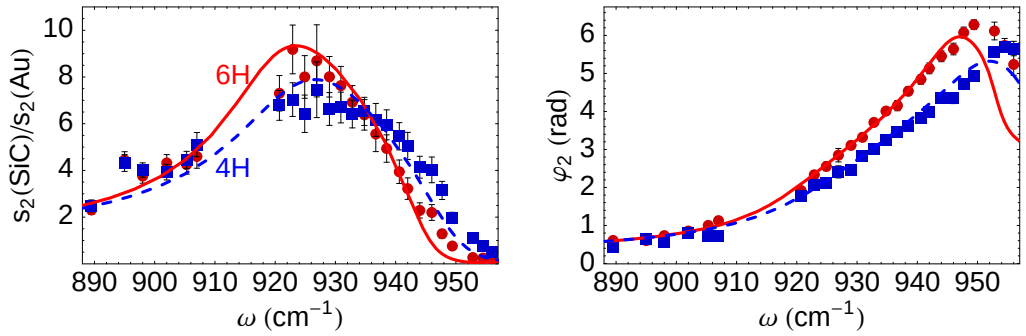


Figure 7.17: Second-harmonic near-field spectra of two different doped SiC polytypes. Points represent the measured values, whereas the full and dashed line represent the monopole model prediction for near-field signal spectra of doped ($n \approx 10^{18}$cm^{-3}) 6H-SiC and 4H-SiC respectively.

It should be noted here that the excellent agreement between the experimentally obtained spectra and the theoretical prediction based on the monopole model (Chapter 5) was obtained by adjusting the dopant concentration n_N to provide the best fit to the experiment. The adjustment of n_N to fit the experimental spectra was not only justified but also necessary because the exact dopant concentration was only very approximately known to be on

the order of $10^{18}\,\mathrm{cm}^{-3}$. Furthermore, the concentration of free carriers is not exactly equal to the dopant concentration, anyway.

Even if n_N was not know precisely, the distinction between 6H and 4H SiC polytypes would be possible since the 4H-SiC spectrum appears shifted to higher frequencies and has a lower maximum than the 6H-SiC spectrum regardless of the plasma frequency ω_p determined by n_N. Furthermore, the same conclusion is reached if the dipole model is used to predict the near-field signal [67]. However, the monopole model allows a more precise characterization of the sample, including the determination of an otherwise very loosely specified parameter n_N. Provided that the results of the monopole model are confirmed by future studies, this model could enable the identification of components contained in a sample with nanometer-scale resolution, even without the prior information on what the components might be.

This and some other possible applications of the findings presented in this chapter will be discussed in the next section.

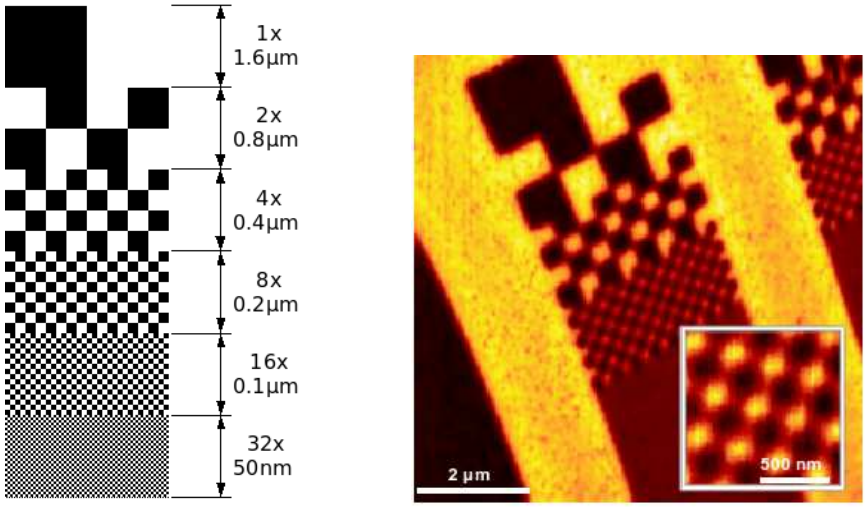
7.6 Applications

The results presented in this chapter demonstrate the potential of near-field infrared spectroscopy as a quantitative tool for investigation of crystal structure with a nanoscale resolution, limited only by the probing tip radius. The best attainable resolution thus surpasses the standard (far-field) infrared spectroscopy by roughly three orders of magnitude without a loss in the sensitivity to chemical and structural properties of the sample. As stated in Sect. 7.4, the near-field spectroscopy also bears the potential to exceed the sensitivity of the far-field IR spectroscopy since the near-field interaction between the probing tip and a crystalline sample exhibits sharper resonances then the Reststrahlen bands observable in far-field spectroscopy. This effect can be used e.g. to map the implantation profiles of donor or acceptor atoms in semiconductors. This is particularly important for SiC, where ion implantation is the only viable method of patterned doping required for fabrication of electronic devices. The diffusion-based doping process, standard in Si electronics fabrication, cannot be applied to SiC because of the practically non-existent diffusion at temperatures below 1800 C [198].

The ability to determine the crystal structure on the nanometer scale also enables the mapping of the crystal quality, i.e. defects in the crystal structure [67]. Again, this possibility is

especially interesting in conjunction with SiC whose widespread usage in the high-power electronic circuits is still held back by the insufficient crystal quality [188].

As a more exotic example, it has been suggested that ion-implanted SiC can serve as an extraordinary data storage medium offering essentially unlimited lifetimes and the ability to withstand extreme environmental conditions [199]. In that case the data storage density corresponding to the 1µm × 1µm squares in Fig. 7.9 would equal 100 Mbit/cm^2, which is roughly equal to the density of bits on a compact disc (CD). The resolution observed in Fig. 7.9 can be significantly improved by implanting heavier ions instead of Be^{2+}. Heavier ions exhibit less straggling below the surface of SiC crystal and thus permit much sharper structures to be drawn. As an example, Fig. 7.18(b) contains a near-field optical image of SiC crystal patterned by implantation of 50 keV Ga^{2+} ions, with squares down to 200 nm in size are clearly distinguishable. Squares in Fig. 7.18 smaller than 200 nm cannot be clearly resolved primarily because of the unwanted contamination of the SiC crystal surface by stray ions produced by blanking of the ion beam as it is scanned over an unimplanted area between two implanted squares.

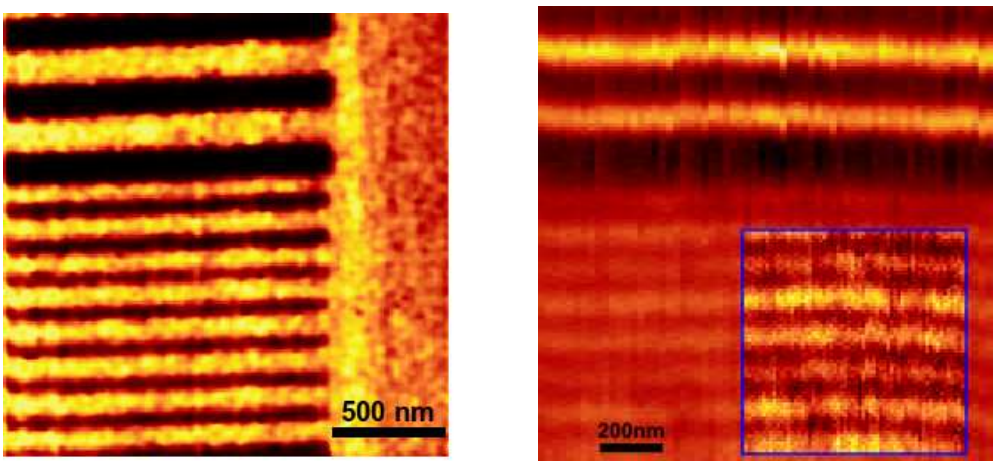


(a) (b)

Figure 7.18: (a) Implantation pattern and (b) s-SNOM images of 50 keV Ga^+ FIB-implanted SiC crystal surface. The implantation dose was about 5×10^{14} ions/cm^2.

To prove that the resolution limit in Fig. 7.18(b) is a technical, rather than a fundamental one, a line pattern implanted by FIB in SiC and imaged in s-SNOM is shown in Fig. 7.19. No beam blanking is performed within one line in Fig. 7.19, thereby significantly reducing the contamination of the unimplanted areas and making the lines 100 nm wide clearly

visible. 50 nm lines can also be resolved after a slight contrast enhancement, as shown in Fig. 7.19(b). The lateral straggle of 50 keV Ga^+ ions amounts to about 20 nm, thus preventing even smaller structures to be created.



(a) (b)

Figure 7.19: s-SNOM images of a line pattern created in a N-doped SiC crystal by FIB implantation of 50 keV Ga^+ ions. Stripes 200 nm and 100 nm wide are easily resolved in part (a). With the aid of the contrast enhancement in the lower right corner of the part (b), 50 nm wide stripes are also revealed. The blurring of the implanted stripes is caused by the ≈ 20 nm lateral straggle of Ga^+ ions.

Considering that the s-SNOM resolution has been repeatedly proven to be limited only by the probing tip radius and can routinely reach below 20 nm [44–46], it can be inferred that even the 50-nm structures do not push the s-SNOM resolution to its limits. Furthermore, it was shown in [75] that objects with sizes as small as $1/3$ of the probing tip radius can be detected and imaged by s-SNOM. Consequently, it should be possible to investigate down to 5-nm defects or irregularities in engineered structures of sizes as small as 20 nm using commercially available metal-coated probes. This is by no means an end because it has already been shown that electrochemically etched metal probes with $R \leq 5$ nm can be produced and successfully used in s-SNOM [74], providing a fourfold resolution improvement over the metal-coated tips.

It is important to note that the same resolution and sensitivity can be achieved in the entire spectrum from the visible to the terahertz frequencies. The visible and near-IR wavelengths are thereby useful for identification of metals, the mid-IR range for identification of polar semiconductors or insulators, and the THz frequencies for measuring the carrier density in doped semiconductors. Considering that this entire range of wavelengths can be combined

in a single instrument, the scattering-type near-field optical microscopy may soon become an indispensable tool for the analysis of materials and devices on the nanometer scale.

List of Publications

Research Papers

1. N. Ocelic, R. Hillenbrand, *Subwavelength-scale tailoring of surface phonon polaritons by focused ion-beam implantation*, Nat. Mater. **9**, 606-609 (2004),
2. A. Huber, N. Ocelic, D. Kazantsev, R. Hillenbrand, *Near-field imaging of mid-infrared surface phonon polariton propagation*, Appl. Phys Lett. **87**, 081103 (2005),
3. A. Huber, N. Ocelic, T. Taubner, R. Hillenbrand, *Nanoscale Resolved Infrared Probing of Crystal Structure and of Plasmon-Phonon Coupling*, Nano Lett. **6**, 774-778 (2006),
4. A. Cvitkovic, N. Ocelic, J. Aizpurua, R. Guckenberger, R. Hillenbrand, *Infrared imaging of single nanoparticles via strong field enhancement in a scanning nanogap*, Phys. Rev. Lett. (2006),
5. N. Ocelic, A. Huber, R. Hillenbrand, *Pseudo-heterodyne detection for background-free s-SNOM imaging*, Appl. Phys. Lett. **89**, 101124 (2006),
6. A. Cvitkovic, N. Ocelic, and R. Hillenbrand, *Analytical model for quantitative prediction of material contrasts in scattering-type near-field optical microscopy*, Opt. Express **15**, 8550-8565 (2007).

Conference Proceedings

1. M. Brehm, H. G. Frey, R. Guckenberger, R. Hillenbrand, D. Kazantsev, F. Keilmann, N. Ocelic, T. Taubner, *Consolidating apertureless SNOM*, J. Korean Phys. Soc. **47**, S80-S85 (2005),
2. A. Huber, N. Ocelic, R. Hillenbrad, *Local excitation and interference of surface phonon polaritons studied by near-field infrared microscopy*, J. Microsc. (to be published).

Patent Applications

1. N. Ocelic, R. Hillenbrand, *Optical Device For Measuring Modulated Signal Light*, WO2007039210,
2. F. Keilmann, A. Schliesser, M. Brehm, N. Ocelic, *Method and Device for Sampling Periodically Repeating Events*, WO2007045461.

Bibliography

[1] L. Rayleigh, “Investigations in optics, with special reference to the spectroscope,” Philos. Mag. **8**, 261 (1879).

[2] E. Abbe, “Beitrage zur Theorie des Mikroskops und der mikroskopischen Wahrnehmung,” Schultze Arch. Mikrosc. Anat. **9**, 413 (1873).

[3] C. M. Sparrow, “On spectroscopic resolving power,” Astrophys. J. **44**(2), 76 (1916).

[4] W. Carrington, R. Lynch, E. Moore, G. Isenberg, K. Fogarty, and F. Fredric, “Superresolution 3-dimensional images of fluorescence in cells with minimal light exposure,” Science **268**(5216), 1483–1487 (1995).

[5] J. Butler, J. Reeds, and S. Dawson, “Estimating solutions of 1st kind integral-equations with nonnegative constraints and optimal smoothing,” SIAM J. Numer. Anal. **18**(3), 381–397 (1981).

[6] J. R. Helliwell and M. Helliwell, “X-ray crystallography in structural chemistry and molecular biology,” Chem. Commun. **14**, 1595–1602 (1996).

[7] M. Knoll and E. Ruska, “Das Elektronenmikroskop,” Z. Physik **78**, 318–339 (1932).

[8] F. B. and J. M. Howe, *Transmission Electron Microscopy and Diffractometry of Materials* (Springer, 2005).

[9] L. Reimer, *Scanning Electron Microscopy: Physics of Image Formation and Microanalysis* (Springer, 1998).

[10] M. von Ardenne, “Das Elektronen-Rastermikroskop. Theoretische Grundlagen.” Z. Phys. **109**, 553–572 (1938).

[11] M. Knoll, “Aufladepotentiel und Sekundäremission elektronenbestrahlter Körper,” Z. tech. Phys. **16**, 467–475 (1935).

[12] T. A. Klar and S. W. Hell, “Subdiffraction resolution in far-field fluorescence microscopy,” Opt. Lett. **24**(14), 954–956 (1999).

[13] T. A. Klar, S. Jakobs, M. Dyba, A. Egner, and S. W. Hell, "Fluorescence microscopy with diffraction resolution barrier broken by stimulated emission," Proc. Natl. Acad. Sci. U. S. A. **97**(15), 8206–8210 (2000).

[14] S. W. Hell, "Toward fluorescence nanoscopy," Nat. Biotechnol. **21**(11), 1347–55 (2003).

[15] E. Betzig, G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz, and H. F. Hess, "Imaging intracellular fluorescent proteins at nanometer resolution," Science **313**(5793), 1642–5 (2006).

[16] R. J. Ober, S. Ram, and E. S. Ward, "Localization accuracy in single-molecule microscopy," Biophys. J. **86**(2), 1185 – 1200 (2004).

[17] S. Ram, E. S. Ward, and R. J. Ober, "Beyond Rayleigh's criterion: A resolution measure with application to single-molecule microscopy," Proc. Natl. Acad. Sci. U. S. A. **103**(12), 4457 – 4462 (2006).

[18] R. E. Thompson, D. R. Larson, and W. W. Webb, "Precise nanometer localization analysis for individual fluorescent probes," Biophys. J. **82**(5), 2775 – 2783 (2002).

[19] A. Yildiz and P. R. Selvin, "Fluorescence imaging with one nanometer accuracy: application to molecular motors." Acc. Chem. Res. **38**(7), 574–82 (2005).

[20] M. P. Gordon, T. Ha, and P. R. Selvin, "Single-molecule high-resolution imaging with photobleaching," Proc. Natl. Acad. Sci. U. S. A. **101**(17), 6462 – 6465 (2004).

[21] P. R. Selvin, "The renaissance of fluorescence resonance energy transfer," Nat. Struct. Biol. **7**(9), 730 – 734 (2000).

[22] C. Sonnichsen, B. M. Reinhard, J. Liphardt, and A. P. Alivisatos, "A molecular ruler based on plasmon coupling of single gold and silver nanoparticles," Nat. Biotechnol. **23**(6), 741–745 (2005).

[23] W. Denk and W. W. Webb, "Optical measurement of picometer displacements of transparent microscopic objects," Appl. Optics **29**(16), 2382 – 2391 (1990).

[24] G. Binnig, C. F. Quate, and C. Gerber, "Atomic force microscope." Phys. Rev. Lett. **56**(9), 930–933 (1986).

[25] G. Binnig, H. Rohrer, C. Gerber, and E. Weibel, "Tunneling through a controllable vacuum gap," Appl. Phys. Lett. **40**(2), 178 – 180 (1982).

[26] M. Nonnenmacher, M. P. Oboyle, and H. K. Wickramasinghe, “Kelvin probe force microscopy,” Appl. Phys. Lett. **58**(25), 2921 – 2923 (1991).

[27] H. J. Butt, “Electrostatic interaction in atomic force microscopy,” Biophys. J. **60**(4), 777 – 785 (1991).

[28] J. R. Matey and J. Blanc, “Scanning capacitance microscopy,” J. Appl. Phys. **57**(5), 1437 – 1444 (1985).

[29] J. J. Saenz, N. Garcia, P. Grutter, E. Meyer, H. Heinzelmann, R. Wiesendanger, L. Rosenthaler, H. R. Hidber, and H. J. Gunterodt, “Observation of magnetic forces by the atomic force microscope,” J. Appl. Phys. **62**(10), 4293 – 4295 (1987).

[30] Y. Martin and H. K. Wickramasinghe, “Magnetic imaging by force microscopy with 1000-a resolution,” Appl. Phys. Lett. **50**(20), 1455 – 1457 (1987).

[31] J. A. Sidles, “Noninductive detection of single-proton magnetic-resonance,” Appl. Phys. Lett. **58**(24), 2854 – 2856 (1991).

[32] D. Rugar, C. S. Yannoni, and J. A. Sidles, “Mechanical detection of magnetic-resonance,” Nature **360**(6404), 563 – 566 (1992).

[33] A. Lewis, M. Isaacson, A. Harootunian, and A. Muray, “Development of a 500-a spatial-resolution light-microscope,” Ultramicroscopy **13**(3), 227–231 (1984).

[34] D. Pohl, W. Denk, and M. Lanz, “Optical stethoscopy - image recording with resolution lambda/20,” Appl. Phys. Lett. **44**(7), 651–653 (1984).

[35] R. C. Reddick, R. J. Warmack, and T. L. Ferrell, “New form of scanning optical microscopy,” Phys. Rev. B **39**(1), 767 – 770 (1989).

[36] G. Binning, H. Rohrer, C. Gerber, and E. Weibel, “Surface studies by scanning tunneling microscopy,” Phys. Rev. Lett. **49**(1), 57 – 61 (1982).

[37] G. Binnig and H. Rohrer, “Scanning tunneling microscopy,” Surf. Sci. **126**(1-3), 236 – 244 (1983).

[38] F. J. Giessibl, S. Hembacher, H. Bielefeldt, and J. Mannhart, “Subatomic features on the silicon (111)-(7x7) surface observed by atomic force microscopy,” Science **289**(5478), 422 – 425 (2000).

[39] E. Ash and G. Nicholls, “Super-resolution aperture scanning microscope,” Nature **237**(5357), 510–& (1972).

[40] E. Betzig and J. Trautman, "Near-field optics - microscopy, spectroscopy, and surface modification beyond the diffraction limit," Science **257**(5067), 189–195 (1992).

[41] J. Wessel, "Surface-enhanced optical microscopy," J. Opt. Soc. Am. B-Opt. Phys. **2**(9), 1538–1541 (1985).

[42] F. Zenhausern, M. Oboyle, and H. Wickramasinghe, "Apertureless near-field optical microscope," Appl. Phys. Lett. **65**(13), 1623–1625 (1994).

[43] Y. Inouye and S. Kawata, "Near-field scanning optical microscope with a metallic probe tip," Opt. Lett. **19**(3), 159–161 (1994).

[44] R. Hillenbrand and F. Keilmann, "Material-specific mapping of metal/semiconductor/dielectric nanosystems at 10 nm resolution by backscattering near-field optical microscopy," Appl. Phys. Lett. **80**(1), 25–27 (2002).

[45] T. Taubner, R. Hillenbrand, and F. Keilmann, "Performance of visible and mid-infrared scattering-type near-field optical microscopes," J. Microsc.-Oxf. **210**(14), 311–314 (2003).

[46] M. B. Raschke, L. Molina, T. Elsaesser, D. H. Kim, W. Knoll, and K. Hinrichs, "Apertureless near-field vibrational imaging of block-copolymer nanostructures with ultrahigh spatial resolution," ChemPhysChem **6**(10), 2197–2203 (2005).

[47] F. Zenhausern, Y. Martin, and H. Wickramasinghe, "Scanning interferometric apertureless microscopy - optical imaging at 10 angstrom resolution," Science **269**(5227), 1083–1085 (1995).

[48] R. M. Stockle, Y. D. Suh, V. Deckert, and R. Zenobi, "Nanoscale chemical analysis by tip-enhanced Raman spectroscopy," Chem. Phys. Lett. **318**(1-3), 131–136 (2000).

[49] A. V. Zayats and V. Sandoghdar, "Apertureless scanning near-field second-harmonic microscopy," Opt. Commun. **178**(1-3), 245–249 (2000).

[50] I. I. Smolyaninov, A. V. Zayats, and C. C. Davis, "Near-field second-harmonic imaging of ferromagnetic and ferroelectric materials," Opt. Lett. **22**(21), 1592 – 1594 (1997).

[51] X. S. Xie, E. Sanchez, and L. Novotny, "Near-field and two-photon fluorescence imaging of single molecules and biomolecules," Abstr. Pap. Am. Chem. Soc. **216**(0), U84–U84 (1998).

[52] E. J. Sanchez, L. Novotny, and X. S. Xie, "Near-field fluorescence microscopy based on two-photon excitation with metal tips," Phys. Rev. Lett. **82**(20), 4014–4017 (1999).

[53] N. Anderson, P. Anger, A. Hartschuh, and L. Novotny, "Subsurface Raman imaging with nanoscale resolution," Nano Lett. **6**(4), 744–749 (2006).

[54] T. Taubner, F. Keilmann, and R. Hillenbrand, "Nanoscale-resolved subsurface imaging by scattering-type near-field optical microscopy," Opt. Express **13**(22), 8893–8899 (2005).

[55] S. Kawata and Y. Inouye, "Scanning probe optical microscopy using a metallic probe tip," Ultramicroscopy **57**(2-3), 313–317 (1995).

[56] M. Labardi, S. Patane, and M. Allegrini, "Artifact-free near-field optical imaging by apertureless microscopy," Appl. Phys. Lett. **77**(5), 621–623 (2000).

[57] B. Knoll and F. Keilmann, "Enhanced dielectric contrast in scattering-type scanning near-field optical microscopy," Opt. Commun. **182**(4-6), 321–328 (2000).

[58] R. Hillenbrand and F. Keilmann, "Complex optical constants on a subwavelength scale," Phys. Rev. Lett. **85**(14), 3029–3032 (2000).

[59] H. G. Frey, S. Witt, K. Felderer, and R. Guckenberger, "High-resolution imaging of single fluorescent molecules with the optical near-field of a metal tip," Phys. Rev. Lett. **93**(20) (2004).

[60] H. G. Frey, F. Keilmann, A. Kriele, and R. Guckenberger, "Enhancing the resolution of scanning near-field optical microscopy by a metal tip grown on an aperture probe," Appl. Phys. Lett. **81**(26), 5030–5032 (2002).

[61] T. Taubner, F. Keilmann, and R. Hillenbrand, "Nanomechanical resonance tuning and phase effects in optical near-field interaction," Nano Lett. **4**(9), 1669–1672 (2004).

[62] B. Knoll and F. Keilmann, "Near-field probing of vibrational absorption for chemical microscopy," Nature **399**(6732), 134–137 (1999).

[63] R. Hillenbrand, T. Taubner, and F. Keilmann, "Phonon-enhanced light-matter interaction at the nanometre scale," Nature **418**(6894), 159–162 (2002).

[64] T. Taubner, R. Hillenbrand, and F. Keilmann, "Nanoscale polymer recognition by

spectral signature in scattering infrared near-field microscopy," Appl. Phys. Lett. **85**(21), 5064–5066 (2004).

[65] M. Brehm, T. Taubner, R. Hillenbrand, and F. Keilmann, "Infrared spectroscopic mapping of single nanoparticles and viruses at nanoscale resolution," Nano Lett. **6**(7), 1307–1310 (2006).

[66] N. Ocelic and R. Hillenbrand, "Subwavelength-scale tailoring of surface phonon polaritons by focused ion-beam implantation," Nat. Mater. **3**(9), 606–609 (2004).

[67] A. Huber, N. Ocelic, T. Taubner, and R. Hillenbrand, "Nanoscale resolved infrared probing of crystal structure and of plasmon-phonon coupling," Nano Lett. **6**(4), 774–778 (2006).

[68] T. Taubner, "Inrarotspektroskopie im Nahfeld einer Tastspitze," Ph.D. thesis, Technische Universität München (2004).

[69] R. Hillenbrand, B. Knoll, and F. Keilmann, "Pure optical contrast in scattering-type scanning near-field microscopy," J. Microsc.-Oxf. **202**(29), 77–83 (2001).

[70] A. Wokaun, J. P. Gordon, and P. F. Liao, "Radiation damping in surface-enhanced Raman-scattering," Phys. Rev. Lett. **48**(14), 957 – 960 (1982).

[71] J. Gersten and A. Nitzan, "Electromagnetic theory of enhanced raman-scattering by molecules adsorbed on rough surfaces," J. Chem. Phys. **73**(7), 3023–3037 (1980).

[72] P. Aravind and H. Metiu, "The effects of the interaction between resonances in the electromagnetic response of a sphere-plane structure - applications to surface enhanced spectroscopy," Surf. Sci. **124**(2-3), 506–528 (1983).

[73] I. S. Averbukh, B. M. Chernobrod, O. A. Sedletsky, and Y. Prior, "Coherent near field optical microscopy," Opt. Commun. **174**(1-4), 33–41 (2000).

[74] M. B. Raschke and C. Lienau, "Apertureless near-field optical microscopy: Tip-sample coupling in elastic light scattering," Appl. Phys. Lett. **83**(24), 5089–5091 (2003).

[75] A. Cvitkovic, N. Ocelic, J. Aizpurua, R. Guckenberger, and R. Hillenbrand, "Infrared imaging of single nanoparticles via strong field enhancement in a scanning nanogap," Phys. Rev. Lett. **97**(6) (2006).

[76] J. Renger, S. Grafstrom, L. M. Eng, and R. Hillenbrand, "Resonant light scattering by near-field-induced phonon polaritons," Phys. Rev. B **71**(7) (2005).

[77] M. Brehm, A. Schliesser, and F. Keilman, "Spectroscopic near-field microscopy using frequency combs in the mid-infrared," Opt. Express **14**(23), 11,222 (2006).

[78] M. Hofmann, A. Zywietz, K. Karch, and F. Bechstedt, "Lattice-dynamics of SiC polytypes within the bond-charge model," Phys. Rev. B **50**(18), 13,401–13,411 (1994).

[79] C. F. Bohren and D. R. Huffman, *Absorption and Scattering of Light by Small Particles* (John Wiley & Sons Inc, 1998).

[80] P. W. Milonni and J. H. Eberly, *Lasers* (Wiley-Interscience, 1988).

[81] C. Klingshirn, *Semiconductor Optics* (Springer, 2005).

[82] D. Bohm and D. Pines, "A collective description of electron interactions:.3. coulomb interactions in a degenerate electron gas," Phys. Rev. **92**(3), 609–625 (1953).

[83] J. Frenkel, *Wave mechanics: elementary theory* (The Clarendon press, 1932).

[84] J. Frenkel, "On the transformation of light into heat in solids. I," Phys. Rev. **37**(1), 17–44 (1931).

[85] G. H. Wannier, "The structure of electronic excitation levels in insulating crystals," Phys. Rev. **52**(3), 0191 – 0197 (1937).

[86] N. F. Mott, "The conduction band and ultra-violet absorption of alkali-halide crystals." Trans. Faraday Soc. **34**(1), 0500–0505 (1938).

[87] U. Fano, "Atomic theory of electromagnetic interactions in dense materials," Phys. Rev. **103**(5), 1202 – 1218 (1956).

[88] J. Hopfield, "Theory of the contribution of excitons to the complex dielectic constant of crystals," Phys. Rev. **112**(5), 1555–1567 (1958).

[89] R. W. Alexander, G. S. Kovener, and R. J. Bell, "Dispersion curves for surface electromagnetic-waves with damping," Phys. Rev. Lett. **32**(4), 154–157 (1974).

[90] E. Schuller, H. Falge, and G. Borstel, "Dispersion curves of surface phonon-polaritons with backbending," Phys. Lett. A **54**(4), 317–318 (1975).

[91] H. Falge, A. Otto, and W. Sohler, "Dispersion of surface and bulk phonon-polaritons on alpha-quartz measured by attenuated total reflection," Phys. Status Solidi B-Basic Res. **63**(1), 259–269 (1974).

[92] R. H. Lyddane, R. G. Sachs, and E. Teller, "On the polar vibrations of alkali halides," Phys. Rev. **59**(8), 673–676 (1941).

[93] F. Marquier, K. Joulain, J. P. Mulet, R. Carminati, J. J. Greffet, and Y. Chen, "Coherent spontaneous emission of light by thermal sources," Phys. Rev. B **69**(15) (2004).

[94] J. J. Greffet, R. Carminati, K. Joulain, J. P. Mulet, S. P. Mainguy, and Y. Chen, "Coherent emission of light by thermal sources," Nature **416**(6876), 61–64 (2002).

[95] J. C. Weeber, J. R. Krenn, A. Dereux, B. Lamprecht, Y. Lacroute, and J. P. Goudonnet, "Near-field observation of surface plasmon polariton propagation on thin metal stripes," Phys. Rev. B **6404**(4) (2001).

[96] S. I. Bozhevolnyi and V. S. Volkov, "Observation of propagation of surface plasmon polaritons along line defects in a periodically corrugated metal surface," Opt. Lett. **26**(10), 734 – 736 (2001).

[97] A. Huber, N. Ocelic, D. Kazantsev, and R. Hillenbrand, "Near-field imaging of mid-infrared surface phonon polariton propagation," Appl. Phys. Lett. **87**(8) (2005).

[98] J. B. Pendry, "Negative refraction makes a perfect lens," Phys. Rev. Lett. **85**(18), 3966–3969 (2000).

[99] N. Fang, H. Lee, C. Sun, and X. Zhang, "Sub-diffraction-limited optical imaging with a silver superlens," Science **308**(5721), 534 – 537 (2005).

[100] D. O. S. Melville and R. J. Blaikie, "Super-resolution imaging through a planar silver layer," Opt. Express **13**(6), 2127 – 2134 (2005).

[101] D. Korobkin, Y. Urzhumov, and G. Shvets, "Enhanced near-field resolution in mid-infrared using metamaterials," J. Opt. Soc. Am. B-Opt. Phys. **23**(3), 468 – 478 (2006).

[102] T. Taubner, D. Korobkin, Y. Urzhumov, G. Shvets, and R. Hillenbrand, "Near-field microscopy through a SiC superlens," Science **313**(5793), 1595–1595 (2006).

[103] J. Takahara, S. Yamagishi, H. Taki, A. Morimoto, and T. Kobayashi, "Guiding of a one-dimensional optical beam with nanometer diameter," Opt. Lett. **22**(7), 475–477 (1997).

[104] A. J. Babadjanyan, N. L. Margaryan, and K. V. Nerkararyan, "Superfocusing of surface polaritons in the conical structure," J. Appl. Phys. **87**(8), 3785 – 3788 (2000).

[105] N. A. Janunts, K. S. Baghdasaryan, K. V. Nerkararyan, and B. Hecht, "Excitation and superfocusing of surface plasmon polaritons on a silver-coated optical fiber tip," Opt. Commun. **253**(1-3), 118 – 124 (2005).

[106] V. S. Gurevich and M. N. Libenson, "Surface-polaritons propagation along micropipettes," Ultramicroscopy **57**(2-3), 277 – 281 (1995).

[107] A. Bouhelier, J. Renger, M. R. Beversluis, and L. Novotny, "Plasmon-coupled tip-enhanced near-field optical microscopy," J. Microsc.-Oxf. **210**(16), 220–224 (2003).

[108] D. Sarid, "Long-range surface-plasma waves on very thin metal-films," Phys. Rev. Lett. **47**(26), 1927 – 1930 (1981).

[109] P. Berini, "Plasmon-polariton modes guided by a metal film of finite width," Opt. Lett. **24**(15), 1011 – 1013 (1999).

[110] P. Berini, "Plasmon-polariton waves guided by thin lossy metal films of finite width: Bound modes of symmetric structures," Phys. Rev. B **61**(15), 10,484 – 10,503 (2000).

[111] R. Zia, M. D. Selker, P. B. Catrysse, and M. L. Brongersma, "Geometries and materials for subwavelength surface plasmon modes," J. Opt. Soc. Am. A-Opt. Image Sci. Vis. **21**(12), 2442–2446 (2004).

[112] J. A. Dionne, L. A. Sweatlock, H. A. Atwater, and A. Polman, "Plasmon slot waveguides: Towards chip-scale propagation with subwavelength-scale localization," Phys. Rev. B **73**(3) (2006).

[113] S. A. Maier, M. L. Brongersma, P. G. Kik, S. Meltzer, A. A. G. Requicha, and H. A. Atwater, "Plasmonics - A route to nanoscale optical devices," Adv. Mater. **13**(19), 1501 (2001).

[114] A. Boltasseva, T. Nikolajsen, K. Leosson, K. Kjaer, M. S. Larsen, and S. I. Bozhevolnyi, "Integrated optical components utilizing long-range surface plasmon polaritons," J. Lightwave Technol. **23**(1), 413–422 (2005).

[115] J. C. Weeber, A. Dereux, C. Girard, G. C. des Francs, J. R. Krenn, and J. P. Goudonnet, "Optical addressing at the subwavelength scale," Phys. Rev. E **62**(5), 7381–7388 (2000).

[116] U. Fischer and D. Pohl, "Observation of single-particle plasmons by near-field optical microscopy," Phys. Rev. Lett. **62**(4), 458–461 (1989).

[117] B. Lamprecht, J. R. Krenn, A. Leitner, and F. R. Aussenegg, "Particle-plasmon decay-time determination by measuring the optical near-field's autocorrelation: influence of inhomogeneous line broadening," Appl. Phys. B-Lasers Opt. **69**(3), 223–227 (1999).

[118] T. Klar, M. Perner, S. Grosse, G. von Plessen, W. Spirkl, and J. Feldmann, "Surface-plasmon resonances in single metallic nanoparticles," Phys. Rev. Lett. **80**(19), 4249–4252 (1998).

[119] M. Quinten, A. Leitner, J. R. Krenn, and F. R. Aussenegg, "Electromagnetic energy transport via linear chains of silver nanoparticles," Opt. Lett. **23**(17), 1331–1333 (1998).

[120] S. A. Maier, P. G. Kik, H. A. Atwater, S. Meltzer, E. Harel, B. E. Koel, and A. A. G. Requicha, "Local detection of electromagnetic energy transport below the diffraction limit in metal nanoparticle plasmon waveguides," Nat. Mater. **2**(4), 229–232 (2003).

[121] K. R. Li, M. I. Stockman, and D. J. Bergman, "Self-similar chain of metal nanospheres as an efficient nanolens," Phys. Rev. Lett. **91**(22) (2003).

[122] G. Raschke, S. Kowarik, T. Franzl, C. Sonnichsen, T. A. Klar, J. Feldmann, A. Nichtl, and K. Kurzinger, "Biomolecular recognition based on single gold nanoparticle light scattering," Nano Lett. **3**(7), 935–938 (2003).

[123] S. J. Oldenburg, R. D. Averitt, S. L. Westcott, and N. J. Halas, "Nanoengineering of optical resonances," Chem. Phys. Lett. **288**(2-4), 243–247 (1998).

[124] S. R. Sershen, S. L. Westcott, N. J. Halas, and J. L. West, "Temperature-sensitive polymer-nanoshell composites for photothermally modulated drug delivery," J. Biomed. Mater. Res. **51**(3), 293–298 (2000).

[125] L. R. Hirsch, R. J. Stafford, J. A. Bankson, S. R. Sershen, B. Rivera, R. E. Price, J. D. Hazle, N. J. Halas, and J. L. West, "Nanoshell-mediated near-infrared thermal therapy of tumors under magnetic resonance guidance," Proc. Natl. Acad. Sci. U. S. A. **100**(23), 13,549–13,554 (2003).

[126] H. A. Bethe, "Theory of diffraction by small holes," Phys. Rev. **66**(7/8), 163 – 182 (1944).

[127] C. J. Bouwkamp, "On bethes theory of diffraction by small holes," Philips Research Reports **5**(5), 321 – 332 (1950).

[128] F. J. G. I. de Abajo, "Light transmission through a single cylindrical hole in a metallic film," Opt. Express **10**(25), 1475 – 1484 (2002).

[129] D. B. Rutledge, S. E. Schwarz, and A. T. Adams, "Infrared and submillimeter antennas," Infrared Phys. **18**(5-6), 713 – 729 (1978).

[130] I. Wilke, W. Herrmann, and F. K. Kneubuhl, "Integrated nanostrip dipole antennas for coherent 30 THz infrared radiation," Appl. Phys. B-Lasers Opt. **58**(2), 87 – 95 (1994).

[131] F. J. Gonzalez and G. D. Boreman, "Comparison of dipole, bowtie, spiral and log-periodic IR antennas," Infrared Phys. Technol. **46**(5), 418 – 428 (2005).

[132] T. Kalkbrenner, U. Hakanson, A. Schadle, S. Burger, C. Henkel, and V. Sandoghdar, "Optical microscopy via spectral modifications of a nanoantenna," Phys. Rev. Lett. **95**(20) (2005).

[133] T. Kalkbrenner, M. Ramstein, J. Mlynek, and V. Sandoghdar, "A single gold particle as a probe for apertureless scanning near-field optical microscopy," J. Microsc.-Oxf. **202**(24), 72–76 (2001).

[134] P. Muhlschlegel, H. J. Eisler, O. J. F. Martin, B. Hecht, and D. W. Pohl, "Resonant optical antennas," Science **308**(5728), 1607–1609 (2005).

[135] P. J. Schuck, D. P. Fromm, A. Sundaramurthy, G. S. Kino, and W. E. Moerner, "Improving the mismatch between light and nanoscale objects with gold bowtie nanoantennas," Phys. Rev. Lett. **94**(1) (2005).

[136] K. Sendur and W. Challener, "Near-field radiation of bow-tie antennas and apertures at optical frequencies," J. Microsc.-Oxf. **210**(15), 279–283 (2003).

[137] A. Naber, D. Molenda, U. C. Fischer, H. J. Maas, C. Hoppener, N. Lu, and H. Fuchs, "Enhanced light confinement in a near-field optical probe with a triangular aperture," Phys. Rev. Lett. **89**(21) (2002).

[138] X. L. Shi, L. Hesselink, and R. L. Thornton, "Ultrahigh light transmission through a C-shaped nanoaperture," Opt. Lett. **28**(15), 1320–1322 (2003).

[139] L. Novotny and S. J. Stranick, "Near-field optical microscopy and spectroscopy with pointed probes," Annu. Rev. Phys. Chem. **57**(105), 303–331 (2006).

[140] P. F. Liao and A. Wokaun, "Lightning rod effect in surface enhanced Raman scattering," J. Chem. Phys. **76**(1), 751 – 752 (1982).

[141] O. J. F. Martin and C. Girard, "Controlling and tuning strong optical field gradients at a local probe microscope tip apex," Appl. Phys. Lett. **70**(6), 705–707 (1997).

[142] F. Demming, J. Jersch, K. Dickmann, and P. I. Geshev, "Calculation of the field enhancement on laser-illuminated scanning probe tips by the boundary element method," Appl. Phys. B-Lasers Opt. **66**(5), 593 – 598 (1998).

[143] Y. C. Martin, H. F. Hamann, and H. K. Wickramasinghe, "Strength of the electric field in apertureless near-field optical microscopy," J. Appl. Phys. **89**(10), 5774–5778 (2001).

[144] J. L. Bohn, D. J. Nesbitt, and A. Gallagher, "Field enhancement in apertureless near-field scanning optical microscopy," J. Opt. Soc. Am. A-Opt. Image Sci. Vis. **18**(12), 2998–3006 (2001).

[145] N. Calander and M. Willander, "Theory of surface-plasmon resonance optical-field enhancement at prolate spheroids," J. Appl. Phys. **92**(9), 4878–4884 (2002).

[146] J. Renger, S. Grafstrom, L. M. Eng, and V. Deckert, "Evanescent wave scattering and local electric field enhancement at ellipsoidal silver particles in the vicinity of a glass surface," J. Opt. Soc. Am. A-Opt. Image Sci. Vis. **21**(7), 1362–1367 (2004).

[147] R. Esteban, R. Vogelgesang, and K. Kern, "Simulation of optical near and far fields of dielectric apertureless scanning probes," Nanotechnology **17**(2), 475–482 (2006).

[148] W. Denk and D. Pohl, "Near-field optics - microscopy with nanometer-size fields," J. Vac. Sci. Technol. B **9**(2), 510–513 (1991).

[149] U. Kreibig and M. Vollmer, *Optical Properties of Metal Clusters* (Springer, 1995).

[150] R. J. Potton, "Reciprocity in optics," Rep. Prog. Phys. **67**(5), 717–754 (2004).

[151] R. Carminati, M. Neito-Vesperinas, and J. J. Greffet, "Reciprocity of evanescent electromagnetic waves," J. Opt. Soc. Am. A-Opt. Image Sci. Vis. **15**(3), 706–712 (1998).

[152] R. Fikri, D. Barchiesi, F. H'Dhili, R. Bachelot, A. Vial, and P. Royer, "Modeling recent experiments of apertureless near-field optical microscopy using 2D finite element method," Opt. Commun. **221**(1-3), 13–22 (2003).

[153] S. V. Sukhov, "Role of multipole moment of the probe in apertureless near-field optical microscopy," Ultramicroscopy **101**(2-4), 111–122 (2004).

[154] H. Hatano and S. Kawata, "Applicability of deconvolution and nonlinear optimization for reconstructing optical images from near-field optical microscope images," J. Microsc.-Oxf. **194**, 230 – 234 (1999).

[155] J. N. Walford, J. A. Porto, R. Carminati, J. J. Greffet, P. M. Adam, S. Hudlet, J. L. Bijeon, A. Stashkevich, and P. Royer, "Influence of tip modulation on image formation in scanning near-field optical microscopy," J. Appl. Phys. **89**(9), 5159–5169 (2001).

[156] J. D. Jackson, *Classical Electrodynamics* (Wiley & Sons, 1998).

[157] J. C. E. Sten and I. V. Lindell, "An electrostatic image solution for the conducting prolate spheroid," J. Electromagn. Waves Appl. **9**(4), 599 – 609 (1995).

[158] I. V. Lindell, G. Dassios, and K. I. Nikoskinen, "Electrostatic image theory for the conducting prolate spheroid," J. Phys. D-Appl. Phys. **34**(15), 2302–2307 (2001).

[159] D. V. Redzic, "An electrostatic problem - a point-charge outside a prolate dielectric spheroid," Am. J. Phys. **62**(12), 1118 – 1121 (1994).

[160] D. V. Redzic, "Image of a moving spheroidal conductor," Am. J. Phys **60**, 506–508 (1992).

[161] I. V. Lindell and K. I. Nikoskinen, "Electrostatic image theory for the dielectric prolate spheroid," J. Electromagn. Waves Appl. **15**(8), 1075–1096 (2001).

[162] D. V. Redzic and S. S. Redzic, "Image charge inclusions in the prolate dielectric spheroid," J. Phys. D-Appl. Phys. **38**(21), 3991 – 3994 (2005).

[163] L. Novotny and B. Hecht, *Principles of Nano-Optics* (Cambridge University Press, 2006).

[164] L. M. Matarrese and K. M. Evenson, "Improved coupling to infrared whisker diodes by use of antenna theory," Appl. Phys. Lett. **17**, 8–10 (1970).

[165] M. Brehm, "Inrarot-Mikrospektroskopie mit einem Nahfledmikroskop," Ph.D. thesis, Technische Universität München (2006).

[166] F. Engelbrecht and R. Helbig, "Effect of crystal anisotropy on the infrared reflectivity of 6H-SiC," Phys. Rev. B **48**(21), 15,698 – 15,707 (1993).

[167] I. V. Lindell, K. I. Nikoskinen, and M. J. Flykt, "Electrostatic image theory for an anisotropic half-space slightly deviating from transverse isotropy," Radio Sci. **31**(6), 1361 – 1368 (1996).

[168] I. V. Lindell, K. I. Nikoskinen, and A. Viljanen, "Electrostatic image method for the anisotropic half space," IEE Proc.-Sci. Meas. Technol. **144**(4), 156 – 162 (1997).

[169] S. C. Schneider, S. Grafstrom, and L. M. Eng, "Scattering near-field optical microscopy of optically anisotropic systems," Phys. Rev. B **71**(11) (2005).

[170] G. Wurtz, R. Bachelot, and P. Royer, "Imaging a GaAlAs laser diode in operation using apertureless scanning near-field optical microscopy," Eur. Phys. J.-Appl. Phys **5**(3), 269–275 (1999).

[171] M. A. Ordal, L. L. Long, R. J. Bell, S. E. Bell, R. R. Bell, R. W. Alexander, and C. A. Ward, "Optical-properties of the metals Al, Co, Cu, Au, Fe, Pb, Ni, Pd, Pt, Ag, Ti, and W in the infrared and far infrared," Appl. Optics **22**(7), 1099 – 1119 (1983).

[172] K. J. Marsh and J. A. Savage, "Infrared optical materials for 8-13 mum - current developments and future prospects," Infrared Phys. **14**(2), 85 – 97 (1974).

[173] N. Ocelic, A. Huber, and R. Hillenbrand, "Pseudoheterodyne detection for background-free near-field spectroscopy," Appl. Phys. Lett. **89**(10) (2006).

[174] R. Hillenbrand, "Nahfeldoptische Amplituden- und Phasenkontrastmikroskopie zur nanoskopischen optischen Abbildung von Materialkontrast und optisch resonanten Partikeln," Ph.D. thesis, Technische Universität München (2001).

[175] B. Hecht, H. Bielefeldt, Y. Inouye, D. W. Pohl, and L. Novotny, "Facts and artifacts in near-field optical microscopy," J. Appl. Phys. **81**(6), 2492–2498 (1997).

[176] P. G. Gucciardi, G. Bachelier, and M. Allegrini, "Far-field background suppression in tip-modulated apertureless near-field optical microscopy," J. Appl. Phys. **99**(12) (2006).

[177] S. Hudlet, S. Aubert, A. Bruyant, R. Bachelot, P. M. Adam, J. L. Bijeon, G. Lerondel, P. Royer, and A. A. Stashkevich, "Apertureless near field optical microscopy: a contribution to the understanding of the signal detected in the presence of a background field," Opt. Commun. **230**(4-6), 245–251 (2004).

[178] R. Hillenbrand, F. Keilmann, P. Hanarp, D. S. Sutherland, and J. Aizpurua, "Coherent imaging of nanoscale plasmon patterns with a carbon nanotube optical probe," Appl. Phys. Lett. **83**(2), 368–370 (2003).

[179] I. Stefanon, S. Blaize, A. Bruyant, S. Aubert, G. Lerondel, R. Bachelot, and P. Royer, "Heterodyne detection of guided waves using a scattering-type Scanning Near-Field Optical Microscope," Opt. Express **13**(14), 5553–5564 (2005).

[180] Y. Sasaki and H. Sasaki, "Heterodyne detection for the extraction of the probe-scattering signal in scattering-type scanning near-field optical microscope," Jpn. J. Appl. Phys. Part 2 - Lett. **39**(4A), L321–L323 (2000).

[181] F. Keilmann and R. Hillenbrand, "Near-field microscopy by elastic light scattering from a tip," Philos. Trans. R. Soc. Lond. Ser. A-Math. Phys. Eng. Sci. **362**(1817), 787–805 (2004).

[182] L. Gomez, R. Bachelot, A. Bouhelier, G. P. Wiederrecht, S. H. Chang, S. K. Gray, F. Hua, S. Jeon, J. A. Rogers, M. E. Castro, S. Blaize, I. Stefanon, G. Lerondel, and P. Royer, "Apertureless scanning near-field optical microscopy: a comparison between homodyne and heterodyne approaches," J. Opt. Soc. Am. B-Opt. Phys. **23**(5), 823–833 (2006).

[183] M. Vaeziravani and R. Toledocrow, "Phase-contrast and amplitude pseudoheterodyne interference near-field scanning optical microscopy," Appl. Phys. Lett. **62**(10), 1044–1046 (1993).

[184] H. Morkoc, S. Strite, G. B. Gao, M. E. Lin, B. Sverdlov, and M. Burns, "Large-band-gap SiC, III-V nitride, and II-VI ZnSe-based semiconductor-device technologies," J. Appl. Phys. **76**(3), 1363 – 1398 (1994).

[185] M. E. Levinshtein, P. A. Ivanov, M. S. Boltovets, V. A. Krivutsa, J. W. Palmour, M. K. Das, and B. A. Hull, "High-temperature (up to 773 K) operation of 6-kV 4H-SiC junction diodes," Solid-State Electron. **49**(7), 1228 – 1232 (2005).

[186] C. E. Weitzel, "Comparison of SiC, GaAs, and Si-R-F-MESFET power densities," IEEE Electron Device Lett. **16**(10), 451 – 453 (1995).

[187] M. Bhatnagar and B. J. Baliga, "Comparison of 6H-SiC, 3C-SiC, and Si for power devices," IEEE Trans. Electron Devices **40**(3), 645 – 655 (1993).

[188] D. Nakamura, I. Gunjishima, S. Yamaguchi, T. Ito, A. Okamoto, H. Kondo, S. Onda, and K. Takatori, "Ultrahigh-quality silicon carbide single crystals," Nature **430**(7003), 1009 – 1012 (2004).

[189] R. J. Trew, "Experimental and simulated results of SiC microwave power MESFETs," Phys. Status Solidi A-Appl. Res. **162**(1), 409 – 419 (1997).

[190] M. S. Janson, M. K. Linnarsson, A. A. Hallen, and B. G. Svensson, "Ion implantation range distributions in silicon carbide," J. Appl. Phys. **93**(11), 8903 – 8909 (2003).

[191] J. F. Ziegler, J. P. Biersack, and U. Littmark, *The Stopping and Range of Ions in Solids* (Pergamon Press, 1985).

[192] R. Nipoti, E. Albertazzi, M. Bianconi, R. Lotti, G. Lulli, M. Cervera, and A. Carnera, "Ion implantation induced swelling in 6H-SiC," Appl. Phys. Lett. **70**(25), 3425 – 3427 (1997).

[193] R. Wernhardt, Personal communication.

[194] D. A. G. Bruggeman, "Calculation of various physics constants in heterogenous substances I Dielectricity constants and conductivity of mixed bodies from isotropic substances," Ann. Phys.-Berlin **24**(7), 636 – 664 (1935).

[195] E. Wendler and G. Peiter, "Optical properties of partially damaged ion implanted SiC layers by use of effective medium models," J. Appl. Phys. **87**(11), 7679 – 7684 (2000).

[196] H. Mutschke, A. C. Andersen, D. Clement, T. Henning, and G. Peiter, "Infrared properties of SiC particles," Astron. Astrophys. **345**(1), 187–202 (1999).

[197] H. Harima, S. Nakashima, and T. Uemura, "Raman-scattering from anisotropic LO-phonon-plasmon-coupled mode in n-type 4H-SiC and 6H-SiC," J. Appl. Phys. **78**(3), 1996–2005 (1995).

[198] J. B. Casady and R. W. Johnson, "Status of silicon carbide (SiC) as a wide-bandgap semiconductor for high-temperature applications: a review," Solid-State Electron. **39**(10), 1409 – 1422 (1996).

[199] S. Kalbitzer, "Semiconductors for optical memories," Curr. Opin. Solid State Mat. Sci. **6**(4), 271–279 (2002).

Acknowledgments

I would like to thank Prof. Martin Stutzmann for supervising my thesis. His valuable recommendations and careful reviewing of the manuscript have greatly improved the quality of the work presented here.

I am also greatly indebted to Prof. Wolfgang Baumeister for the opportunity to carry out my scientific work in the outstanding environment of his department at the Max-Planck Institute of Biochemistry.

Special thanks go to my mentor, Dr. Rainer Hillenbrand. In his lab I got the chance to work in the exciting field of nano-optics and a unique opportunity to try all my ideas in practice. The optimal research atmosphere he created and the freedom he gave me were crucial for the successful development of this thesis. I appreciate very much our extensive arguments and his numerous comments on the manuscript which helped me to significantly improve the level and clarity of presentation.

Dr. Fritz Keilmann has many times provided new insights into the physics of the problems at hand and his vast experience often proved decisive for solving experimental puzzles. I am also thankful for his optimism which helped me to relax and recharge with enthusiasm.

Dr. Thomas Taubner introduced me to the science and art of the near-field optical microscopy at the beginning of my doctoral studies. I am also grateful to him for useful practical hints and for constructive discussions.

I wish to express my special gratitude to Andreas Huber for making available the experimental data which I used to support for some of the theoretical results derived in this thesis. It was a great pleasure to work with him on several projects during the past three years.

Dr. Reinhard Guckenberger was always ready to share his technical expertise in the AFM mechanics and electronics with us, thereby helping us to keep our instruments, and experiments, running.

Insightful discussions with Scott Carney, Javier Aizpurua and Rainer Hegerl clarified important topics in the theory of the electric field calculations and image processing.

I am thankful to all past and present members of Martinsrieder SPM group, including Mirjam Beuttler, Dr. Markus Brehm, Dr. Heinrich Frey, Thomas Ganz, Dr. Daniel Häfliger, Florian Huth, Dr. Nader Issa, Dr. Dmitri Kazantsev, Hans-Georg von Ribbeck, Dr. Jens Schiener and Albert Schliesser for the enjoyable atmosphere in the lab and for stimulating discussions at our Friday meetings.

I wish to express my deepest gratitude to my parents, grandparents and my brother who supported me in many ways from the beginning of my doctoral studies to their end.

The last paragraph of the acknowledgments is reserved for the first one in my heart, my girlfriend and coworker Antonija Cvitković. Our scientific cooperation provided the clues to some long-standing questions, and both the motivation and experimental support for the model of the probe-sample interaction derived in this thesis. Our discussions during the long walks around the Starnberger lake helped me to develop a deeper understanding of not only the near-field optics, but also the life, universe and (almost) everything. Privately, she made all my years spent in a foreign country feel just like always being at home. For what you are, and for all you have done - a big, big thanks!