Institut für Angewandte Physik
Fakultät Physik
Bereich Mathematik und Naturwissenschaften
Technische Universität Dresden

Investigation of nanometer scale charge carrier density variations with scattering-type scanning near-field microscopy in the THz regime

Dissertation zur Erlangung des akademischen Grades
Doktor rerum naturalium
(Dr. rer. nat.)

vorgelegt von

Frederik Kuschewski

geboren am 07.09.1988 in Münster (Westf.)

eingereicht am 23. April 2019
verteidigt am 16. Juli 2019

Erstgutachter: Prof. Dr. Lukas M. Eng
Zweitgutachter: Prof. Dr. Hartmut Roskos

Für die Freiheit

Abstract

Near-field microscopy is a versatile technique for non-destructive detection of optical properties on the nanometer scale. Contrary to conventional microscopy techniques, the resolution in near-field microscopy is not restricted by the diffraction limit, but by the size of the probe only. Typically, wavelength-independent resolution in the range of few ten nanometers can be achieved. Many fundamental phenomena in solid states occur at such small length scales and can be probed by infrared and THz radiation. In the present work, nanoscale charge carrier distributions were investigated with near-field microscopy in classic semiconductors and state-of-the-art graphene field-effect transistors. A CO_2 laser, the free-electron laser FELBE at the Helmholtz-Zentrum Dresden Rossendorf and a photoconductive antenna were applied as radiation sources for illumination of the samples.

In the theoretical part of the work, the band model for charge carriers in semiconductors is briefly explained to derive typical charge carrier densities of such materials. The influence of the charge carriers to the light-matter interaction is introduced via the Drude model and evaluated for both infrared and THz radiation. In field-effect transistors, charge carrier density waves can occur when strong AC fields are coupled into the device. The phenomena in such transistors are introduced as a more complex material system. To describe the near-field coupling of the samples to the nanoscopic probe, the dipole model is introduced and extended for periodic charge carrier density, as elicited by low repetition-rate excitation lasers. Consequently, sidebands occur as new frequencies in the signal spectrum, allowing for a more sensitive probing of such transient processes.

Experimental investigations of these sidebands were performed with a CO_2 laser setup on a bulk germanium sample which was excited with femtosecond laser pulses. New frequencies up to the 8th sideband could be observed. The results show a characteristic near-field decay for all sidebands when the probe-sample distance is increased. A nanoscale material contrast in the sidebands signatures has been demonstrated via near-field scans on a gold / germanium heterostructure.

Near-field signatures of graphene-field effect transistors have been examined utilizing FELBE. The results match the predicted behavior of charge carriers in such a device and in particular represent the first direct observations of the plasma waves. In collaboration with the group of Prof. Dr. Hartmut G. Roskos (Goethe-Universität Frankfurt), the plasma wave velocity in the graphene field-effect transistor has been derived via fitting to the model for two datasets on different devices

from independent fabrications. The obtained velocity is in good agreement with literature values. The results promise the application of field-effect transistors as THz detectors and emitters and may lead to faster communication technology.

Kurzfassung

Nahfeldmikroskopie ist eine vielseite Technik für das zerstörungsfreie Auslesen von optischen Eigenschaften auf der Nanoskala. Im Gegensatz zur konventionellen Mikroskopie ist die Auflösung nicht durch Beugungseffekte, sondern durch die Größe der genutzten Sonde begrenzt. Überlicherweise werden wellenlängenunabhängig Auflösungen von einigen zehn Nanometern erreicht. Viele fundamentale Prozesse in der Festkörperphysik treten auf Längenskalen dieser Größenordnung auf und können mit Infrarot- und THz-Strahlung untersucht werden. In dieser Arbeit wurden nanoskalige Ladungsträgerverteilungen mit Rasternahfeldmikroskopie untersucht, einerseits in klassischen Halbleitern, anderseits in state-of-the-art Graphen Feldeffekttransistoren. Zur Beleuchtung der Proben wurden ein CO_2 Laser, der freie-Elektronen Laser FELBE am Helmholtz-Zentrum Dresden-Rossendorf und eine photoleitende Antenne verwendet.

Im theoretischen Teil der Arbeit wird das Bändermodell für Ladungsträger in Halbleitern erklärt, um daraus typische Ladungsträgerdichten in diesen Materialien abzuleiten. Der Einfluss der Ladungsträger auf die Interaktion mit Strahlung wird durch das Drude-Modell eingeführt und für Infrarot- und THz-Strahlung abgeschätzt. In Graphen Feldeffekttransistoren können Ladungsträgerdichtewellen auftreten, wenn starke Wechselfelder in das Bauelement eingekoppelt werden. Die Prozesse in solchen Transistoren werden als komplexeres Materialsystem eingeführt. Um die Nahfeldkopplung der Proben an die Sonde zu beschreiben, wird das Dipol-Modell eingeführt und für periodische Ladungsträgerdichten erweitert, wie sie bspw. durch Pumplaser mit niedrigen Repetitionsraten erzeugt werden können. In der Folge entstehen Seitenbänder als neue Frequenzen im Signalspektrum, welche eine sensitivere Messung solcher transienten Prozesse ermöglichen.

Experimentelle Untersuchungen des erweiterten Dipol-Modells wurden mit einem CO_2 Laser Aufbau an einem Germaniumkristall durchgeführt, welcher mit Femtosekunden Laserpulsen angeregt wird. Neue Frequenzen im Spektrum konnten bis zu dem achten Seitenband beobachtet werden. Die Resultate zeigen den typischen Abfall des Nahfeldes, wenn der Abstand zwischen Sonde und Probe vergrößert wird. Ein Materialkontrast auf der Nanoskale im Seitenband-Signal konnte durch laterale Rasternahfeld-Scans auf einer Gold/Germanium Heterostruktur gezeigt werden.

Die Nahfeldsignaturen der Graphen Feldeffekttransistoren wurden mit FELBE untersucht. Die Resultate stimmen mit dem vorausgesagtem Verhalten der Ladungsträger in einem solchen Bauteil überein und sind die erste direkte Beobachtung solcher Plasmawellen. In Kooperation mit der Gruppe um Prof. Dr. Hartmut G. Roskos (Goethe-Universität Frankfurt) wurde die Geschwindigkeit der Plasmawelle durch Regression der Daten berechnet. Dabei wurden zwei Datensätzen

an Bauteilen von unabhängigen Fabrikationsprozessen genutzt. Die berechnete Geschwindigkeit ist in guter Übereinstimmung mit Literaturwerten. Die Resultate verheißen die Anwendung von Feldeffekttransistoren als THz Sender und Detektoren und könnten zu schnellerer Kommunikationstechnologie führen.

Contents

1 Introduction 7

2 Fundamentals 11
2.1 Semiconductors . . . 12
2.2 Plasma Waves in Graphene Field-Effect Transistors . . . 25
2.3 Near-Field Microscopy . . . 30
2.3.1 Aperture-SNOM . . . 33
2.3.2 Scattering-SNOM . . . 35
2.4 THz Optics . . . 37

3 SNOM-Theory 49
3.1 Dipole Model . . . 49
3.2 Detection and Demodulation . . . 54
3.3 Pump-induced Sidebands in SNOM . . . 60
3.4 Field Enhancement by Resonant Probes . . . 66

4 Near-Field Microscope Setups 73
4.1 FELBE THz SNOM . . . 73
4.2 Pump-modulated s-SNOM . . . 81
4.3 THz Time-Domain-Spectroscopy SNOM . . . 84

5 Sideband Results 91
5.1 Pump-induced Sidebands in Germanium . . . 91
5.2 Fluence Dependence . . . 94
5.3 Higher-order sidebands . . . 95
5.4 Oscillation Amplitude . . . 97
5.5 Technical Aspects of the Sideband Demodulation . . . 99
5.6 Nano-optic Material Contrast . . . 102

6 Field-Effect Transistors **107**
6.1 Device Design 108
6.2 Data Analysis 110
6.3 Near-Field Overview Scans 111
6.4 Plasma Wave Examination 115
6.5 Conclusion 124

7 Discussion and Outlook **127**

A Appendix **131**
A.1 Scanning Probe Microscopy 131
A.2 Atomic Force Microscope 144

List of Figures **175**

Bibliography **179**

List of Abbreviations

a-SNOM aperture-type scanning near-field optical microscopy

AC alternating current

AFM atomic force microscopy / atomic force microscope

c-AFM conductive atomic force microscopy

CB conduction band

CNP charge neutrality point

DC direct current

ELBE Elektronen Linearbeschleuniger für Strahlen hoher Brillanz und niedriger Emittanz / Electron Linac for beams with high Brilliance and low Emittance

EOS electro-optic sampling

GaAs gallium arsenide

FEL free-electron laser

FELBE free-electron laser beamline at ELBE

FET field-effect transistor

FFT fast Fourier transform

FTIR Fourier-transform infrared spectroscopy

FWHM full width at half maximum

Ge germanium

GFET graphene field-effect transistor

HOPG highly oriented pyrolytic graphite

HZDR Helmholtz-Zentrum Dresden-Rossendorf

IR Infrared

KPFM	Kelvin probe force microscopy
lHe	liquid helium
LIA	Lock-In amplifier
MFM	magnetic force microscopy
NA	numerical aperture
NF	near field
PCA	photo-conductive antenna
PFM	piezoresponse force microscopy
PML	perfectly matched layer
PPLN	periodically poled lithium niobate
PPP	PointProbe Plus, AFM tip standard by Nanoworld AG
QCL	quantum cascade laser
RMN	Rocky Mountain Nanotechnology, Inc.
RMS	root mean square
s-SNOM	scattering-type scanning near-field optical microscopy
SB	sideband
SEM	scanning electron microscopy
SNIM	scanning near-field infrared microscopy
SNOM	scanning near-field optical microscopy
SNR	signal-to-noise ratio
SPM	scanning-probe microscopy
TDS	time-domain spectroscopy
TELBE	THz facility at ELBE

THz Terahertz

VB valence band

WR weighted near-field response

1. Introduction

There's Plenty of Room at the Bottom.

(Richard P. Feynman)

Obtaining information about our surroundings by means of optical analysis is one of the most frequently applied methods in science, but also in our daily lives. The eye-sight is one of the most developed senses of the human body. It has an enormous dynamic range, angular-resolved detection of the incoming light and a processing rate which is fast enough to correctly observe most changes in our environment, like movements of objects.

Although the eyes' functionality is very impressive, it has a number of limitations. It cannot detect the polarisation of the light and is sensitive to the intensity only. The biggest limitation is certainly the wavelength regime the eye is sensitive to. Covering the range from $\lambda \approx$ 400 - 700 nm, only a tiny portion of the whole electromagnetic spectrum can be observed by the human eye. While physicists and electrical engineers have developed sources and detectors for almost the whole electromagnetic spectrum and can deduce huge amounts of information with their techniques, one part of the spectrum has just been opened for exploration in the recent years.

The THz range is roughly defined as the part of the spectrum where the frequency of the light f is in the range of $f =$ 0.1 - 10 THz (or $\lambda =$ 30 - 3000 μm). It lies in between the infrared and microwave area of the electromagnetic spectrum. This range has not been covered by sources or detectors until a few decades ago, when laser scientists started to find novel sources and detectors to access at least certain laser lines in this regime. Nowadays, with the invention of short pulsed laser sources, a larger portion of the spectrum up to a few THz can be covered by optical rectification in electro-optic crystals, and just recently Schottky-diode based sources developed by electrical engineers became available for scientists. Although the first quantum cascade lasers now operate in the THz range, the range of 3 - 10 THz is still mostly uncovered by radiation sources, especially when large powers are needed.

Many interesting phenomena can be investigated with light-matter interaction experiments in the infrared and THz regime [1], with most of the origins for these phenomena lying at the nanometer scale. Unfortunately, conventional microscopy techniques are limited by diffraction [2] and deliver a spatial resolution proportional to the wavelength employed in the microscope. This limits the achievable resolution to roughly ten μm for the infrared regime and one millimeter for low THz radiation. This resolution is far too low for the investigation of many processes, and a resolution of a few hundred nanometers is needed to gain deep insight into fundamentals of solid state physics as shown in this work.

In order to overcome the diffraction limit, another approach of optical exploration must be taken. Instead of probing the far field of the radiation, the near field may be accessed, that carries field distributions that are not limited by the applied wavelength. This can be performed via a small aperture, as proposed by Synge [3], or - employing Babinet's principle - by a small scatterer located close to the sample surface [4]. In such a setup, the resolution is limited by the size of the aperture/scatterer (few ten nm) and not by the employed wavelength. In experimental setups, this concept is realized by either an optical fiber [5] as an aperture, or by a metallic tip [4] as a small scatterer. When the fiber or tip is scanned over the sample, we refer to this as a scanning near-field optical microscope (SNOM). As visible light is applied for illumination in the two references above, the enhancement in resolution compared to a conventional microscope is not very large. The enhancement becomes much larger when infrared or THz radiation is needed for explorations. However, the fiber based approach is not fruitful due to large losses in the fiber and only the tip-based scattering-type scanning near-field optical microscopy (s-SNOM) is applied for such long wavelengths. The gain in resolution easily reaches a factor of hundreds in the infrared [6] and of several thousands [7] in the THz regime.

This work features the application of s-SNOM to investigate nanoscale charge carrier density distribution with both infrared and THz laser systems. In the first chapter of the work, the investigated material systems namely semiconductors and graphene field-effect transistors are introduced. While the theory of classic semiconductors is well understood and may be familiar to many readers, the exciting phenomena occuring in a graphene field-effect transistor is a topic under investigation at the moment. In both material systems, the charge carriers play the dominant role and are investigated with infrared and THz radiation. Using the Drude theory it is described that the charge carriers in Germanium need to purposely be excited by a short pulse laser to probe them with an infrared laser. However, the theory shows that THz radiation is much more sensitive to lower charge carrier concentrations. It can be applied to investigate charge carriers

without an additional excitation. The second chapter of the work features the dipole model to describe an s-SNOM system. This analytical theory is extended to a system where a low-repetition rate excitation laser is used to excite high charge carrier densities. The main conclusion is that new frequencies appear as sidebands in the signal spectrum of the near fiel. For the theoretical investigation of more realistic tip systems, a numerical approach is taken to estimate the field enhancement of different probes.

The experimental part of the work features two sections. The first section describes the s-SNOM setups in detail. All three setups utilize a home-built atomic force microscope which was specially designed for the application of THz radiation and multiple laser sources at once. The first setup is located at the free-electron laser FELBE at the Helmholtz-Zentrum Dresden-Rossendorf. The narrow-band laser is tunable from 5 - 230 μm wavelength and the setup is capable to work with almost any wavelength in this range. In this work, the laser was tuned to a wavelength of 150 μm (2 THz). At this wavelength, the performance of different commercial s-SNOM probes was investigated and compared. The results match the predictions made in the theoretical part of the work. The second setup is used to investigate transient processes in Germanium with a CO_2 probe laser a femtosecond laser for excitation. The third setup features a pulsed and broadband low THz exciation and detection via time-domain spectroscopy. Here, the whole spectrum of the near-field response can be deduced in the low THz range.

The second part of the experimental work gives a detailed analysis of the results obtained with these setups. The transient states in Germanium are investigated with a demodulation technique feeaturing cascaded Lock-In amplifiers. The existence of the sidebands is proven and parameter studies are performed for both physical properties of the system and technical settings of the Lock-In amplifiers. A nano-optical material contrast is shown in the sidebands via lateral scans. Using the free-electron laser at 150 μm wavelength, the near-field response of graphene field-effect transistors is studied in detailed. Here, charge carrier density waves could be observed and via comparison to the theoretical model the wave velocity is derived. The results match the expectations from theoretical calculations.

2. Fundamentals

I was like: 'Oh my God, that sounds cool!'

(Donna Strickland,
about a lecture on electro-optics and laser physics.)

This chapter features the fundamentals of both the sample systems and the experimental techniques used in this work.

Charge carriers are strongly influencing the light-matter interaction of infrared and THZ radiation. In order to estimate charge carrier densities and their impact on the permittivity in semiconductors, calculations of densities via the Fermi-Dirac statistic and the Drude model are performed for a Germanium crystal. Besides the intrinsic charge carriers, pulsed lasers can be applied to deliberately excite charge carriers in the conduction band and valence band in semiconductors. Their transient influence on the permittivity is investigated in this chapter as well. As a second sample system, Graphene field-effect transistors are introduced. It will be presented that a two-dimensional electron gas under the influence of an applied bias voltage can actually be described by hydrodynamic equations, leading to fascinating properties in the THz range.

Conventional optical microscopes detect far-field radiation and hence are diffraction-limited. Their resolution is limited by the applied wavelength, as shown in this chapter in two different ways. To circumvent this fundamental limit of optics, the near field can be probed directly. Several near-field microscopy techniques featuring this capability are illustrated here. A detailed theoretical description of s-SNOM, the technique chosen for this work, can be found in chapter 3. As some readers may be unfamiliar with THz optics, typical emitter and detector concepts are presented as the last part of this chapter.

2.1. Semiconductors

Semiconductors are defined as materials with an electric conductivity G located in between the conductivity of metals ($G \geq 10^4$ S/cm $= 10^4$ A/V·cm) and insulators ($G \leq 10^{-8}$ S/cm $= 10^{-8}$ A/V·cm) covering a huge range of 12 orders of magnitude. These borders are not strictly defined and the material classes can in general be overlapping in their conductivity. The second property classifying a material as a semiconductor is a rising conductivity when the temperature is increased. Semiconductors are of enormous interest in the scientific community (the search engine Google Scholar lists almost four million results for the keyword semiconductor) due to their electronic properties and how they can be taylored via doping for different applications as well as their application in many micro- and nanoelectronic devices.

Semiconductors can be crystalline or amorphous. In this work, only crystalline semiconductors are investigated and the theoretical description is limited to this subgroup in the following. Semiconducting materials are typically made of elements of the fourth group in the periodic table (e. g. silicon, germanium) or compound semiconductors with a mix of group III and V or group II and VI elements. Silicon and Germanium crystallize in a diamond structure, while the compound semiconductors mentioned in this work mostly crystallize in the zyncblende structure.

Band Structure

The scope of this work is on the electro-optic properties of semiconductors. In general, semiconductors have energy bands in their electronic structure which arise from the atomic orbitals of the element. When a crystal is formed, the orbitals of different atoms overlap and the electrons are delocalized, i.e. they are not bound to a certain atom anymore and form a large quantum-mechanical system. Due to the Pauli principle, two electrons cannot have the same set of quantum numbers. As a result the electrons gradually fill up states with different energy eigenvalues. This behavior is illustrated in figure 2.1.1 for the s- and p-orbital. For large interatomic distances (e. g., in liquids/gases) the orbitals have distinct energies. When the distance is descreased, energy bands form. At the distance of the lattice constant a, the distance of the bands determines the so-called bandgap.

The full band structure can be calculated with so-called $\vec{k} \cdot \vec{P}$ models, also called $\vec{k} \cdot \vec{P}$ pertubation theory. It is an often used, semi-empirical approach to calculate band structures via Bloch waves and pertubation theory in quantum mechanics. Details will not be discussed in this work and can be studied e. g. in [9]. The band structures for a germanium (Ge) and a gallium arsenide

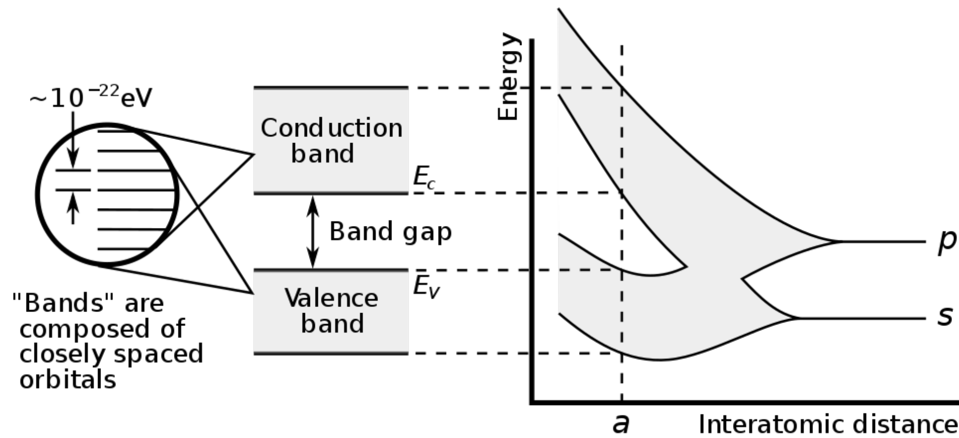

Figure 2.1.1.: *The formation of energy bands from atomic orbitals in a semiconductor arises when the distance in between the atoms becomes small. In a crystal with the lattice constant a, the distance of the valence and conduction band is $E_g = E_c - E_v$. Electron states can have energies within the bands (grey) but not in between them. Illustration taken from [8].*

(GaAs) crystal are shown in 2.1.2. Here, the electron energy is plotted as a function of the electron wavevector $\vec{k}$. In the GaAs crystal, the highest point of the valence band (VB) and the lowest point of the conduction band (CB) are both located at the Γ-point. When an electron is lifted from the valence band to the conduction band no change of $\vec{k}$ is needed and hence no momentum needs to be be transferred. For Ge, the highest point of the valence band is also at the Γ-point, but the lowest point of the conduction band is at the L-point. Here, a momentum transfer of $|\vec{k}| = |L-\Gamma|$ is needed to overcome the bandgap. Therefore, GaAs is a so-called direct semiconductor and Ge is an indirect semiconductor.

In the ground state at $T = 0$ K, the valence band is completely filled and the conduction band is completely empty. At higher temperatures, electrons are lifted from the valence band to the conduction band and leave an unoccupied state, a so-called hole. The occupation density of the energy states is determined by the Fermi-Dirac statistic. The propability f that a state with the energy E is populated is given by

$$f(E) = \frac{1}{e^{(E-E_F)/k_B T} + 1} \tag{2.1.1}$$

where E_F is the Fermi level, k_B the Boltzmann constant and T the temperature. In intrinsic semiconductors, the Fermi level is close to the middle of the bandgap. Figure 2.1.3 shows the

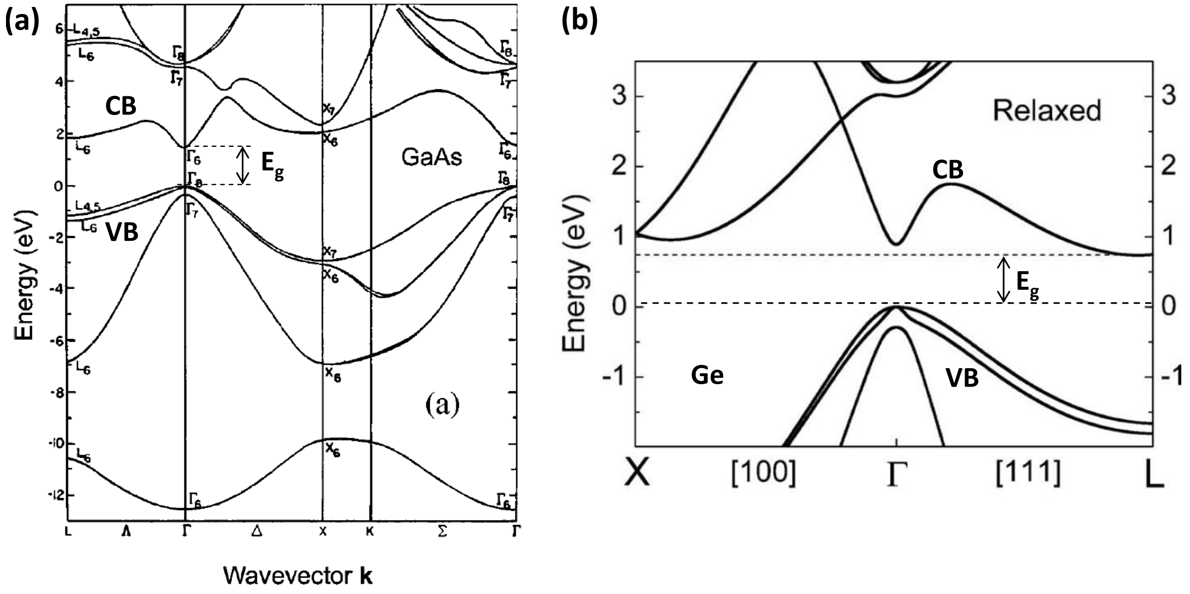

Figure 2.1.2.: *Bandstructures of Semiconductors. (a) GaAs has a direct bandgap, with minimum in the CB and maximum in the VB at the Γ-point. (b) Ge has a indirect bandgap with the CB-mininum and VB-maximum are different wavevectors. Illustration taken from [10] and [11] and adapted.*

population in a germanium crystal for different temperatures. The valence band has energy states up to 0 eV and the conduction band from 0.67 eV upwards. The bandgap in between cannot be populated. For $T = 0$ K, f is one up to the Fermi level and jumps to zero for states above the Fermi level. When the temperature is increased, distribution is getting broader and the drop in f is smoothened. However, the population above the bandgap is only significantly above zero for high temperatures. At room temperature $T = 293$ K, the probability that the lowest energy state of the conduction band is populated is only $f = 1.72 \cdot 10^{-6}$.

The Fermi-Dirac statistic determines the probability that a state with a certain energy E is populated, but does neither state anything about the number of the states nor about the overall number of electrons in the conduction band. This information is given by the density of states g_C for a three-dimensional material. For electrons in the conduction band in a semiconductor it is given by

$$g_C(E) = \frac{8\pi\sqrt{2}}{h^3}(m_e^*)^{3/2}\sqrt{E - E_C} \quad \text{for} \quad E \geq E_C. \tag{2.1.2}$$

Here, E_C is the lower energy limit of the conduction band, h is the Planck constant and m_e^* is the

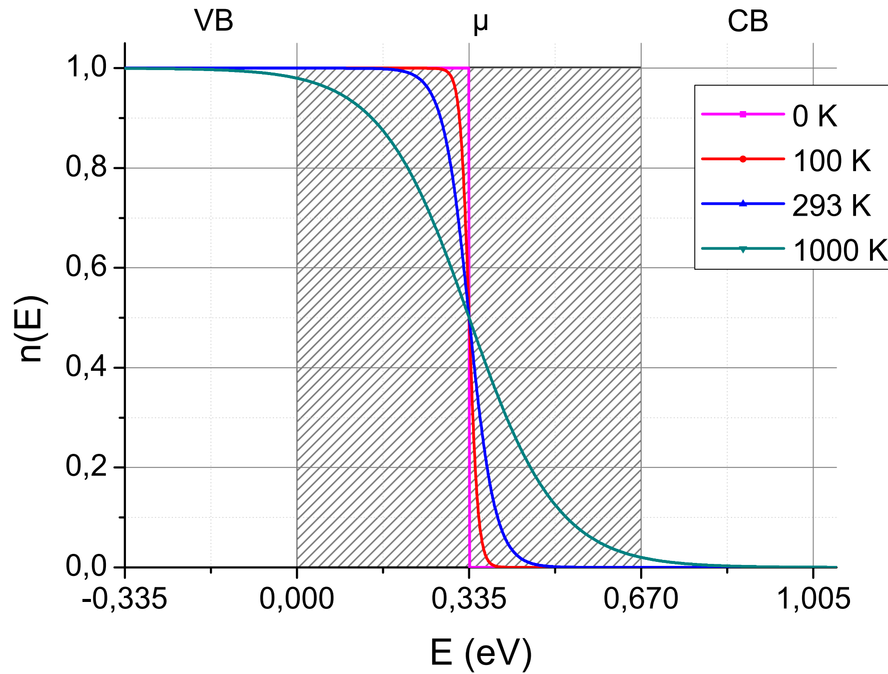

Figure 2.1.3.: *Population of valence and conduction band in germanium: For T = 0 K, the value drops from one to zero at the Fermi level. For higher temperatures, the distribution becomes broader. The grey area is the bandgap which cannot be filled with electrons, hence only at large temperatures, a significant amount of electrons is found in the conduction band.*

effective mass[1] of the electrons in the conduction band. For the valence band, it is given by

$$g_V(E) = \frac{8\pi\sqrt{2}}{h^3}(m_h^*)^{3/2}\sqrt{E_V - E} \quad \text{for} \quad E \leq E_V, \tag{2.1.3}$$

with E_V as the highest energy level of the valence band and m_h^* the effective mass of the holes.

In order to calculate the overall number of electrons and holes (or rather their densities n and p, respectively) in the semiconductor, the density of states must be multiplied with the Fermi-Dirac

[1] In a crystal, the electrons and holes react stronger or weaker to electric fields than in free space. This effect is mathematically compensated for with an effective mass m_e^* and m_h^* being lower or higher than the electron mass m_e.

statistic and needs to be integrated over the energy range of interest. Hence we obtain

$$n = \int_{E_C}^{\infty} g_C(E) f(E) dE \qquad \text{and} \qquad p = \int_{-\infty}^{E_V} g_V(E) f(E) dE. \tag{2.1.4}$$

This equation cannot be solved analytically. However, for a non-degenerate[2] semiconductor, the Fermi-Dirac distribution can be approximated via $f(E) = e^{(E_F - E)/kT}$. For this case the electron and hole densities are given with

$$n = 2\left(\frac{2\pi m_e^* kT}{h^2}\right)^{3/2} e^{(\mu - E_C)/kT} \qquad \text{and} \tag{2.1.5}$$

$$p = 2\left(\frac{2\pi m_h^* kT}{h^2}\right)^{3/2} e^{(E_V - \mu)/kT} \tag{2.1.6}$$

For undoped germanium at room temperature the effective masses are $m_e^* = 0.08 \cdot m_e$ and $m_h^* = 0.28 \cdot m_e$ [12]. The derived values are

$$n = 9.38 \cdot 10^{11}/\text{cm}^3 \qquad \text{and} \tag{2.1.7}$$

$$p = 6.20 \cdot 10^{12}/\text{cm}^3. \tag{2.1.8}$$

The mismatch of the results violates the charge conservation of the system and is a result of the approximation used for non-degenerate semiconductors. To calculate the intrinsic charge carrier density n_i some sources simply argue that charge carrier conservation must apply and calculate

$$n_i = n = p = \sqrt{np} = 2.47 \cdot 10^{12}/\text{cm}^3, \tag{2.1.9}$$

giving a fair approximation for the intrinsic charge carrier density. The experimental value usually stated for pure, high quality germanium is $n = 2.4 \cdot 10^{13}/\text{cm}^3$ [12] and thus is one order of magnitude larger than the derived value. This mismatch is a consequence of the approximations performed during the calculation. However, the next section will show that both the theoretical and experimental values are too small to significantly influence the s-SNOM response in the infrared or THz range.

[2] A degenerate semiconductor is a heavily doped semiconductor in which the Fermi level is close to the conduction band. In this work, only non-degenerate semiconductors are used.

Drude Model

In this work, s-SNOM is utilized to probe several material systems with laser sources operating at wavelengths from mid-infrared to millimeter waves. To estimate the impact of free charge carriers on the optical properties in such a large wavelength range, the Drude model [13] is the most commonly used theory. It will briefly be introduced below.

Let us assume the carrier density in a crystal is quasi-homogenic with no interaction in between the electrons. When an AC electric field $E \propto e^{-i\omega t}$ is applied to the crystal, all electrons exhibit the same displacement from their ground state. The equation of motion for a single electron with the charge e in the 1D-case is then

$$m_e^* \ddot{x} = -eE \tag{2.1.10}$$

with the solution

$$x = \frac{eE}{m_e^* \omega^2}. \tag{2.1.11}$$

Each displaced electron contributes with a dipole moment of $p = -ex$. If we know the carrier concentration n (for semiconductors, we calculated it in chapter 2.1 above), we can derive the polarization of the crystal:

$$P = -nex = -\frac{ne^2}{m_e^* \omega^2} E. \tag{2.1.12}$$

Via Maxwell's equations, we can then calculate the permittivity induced by the free electrons. The real part is given by

$$\varepsilon(\omega) = \frac{D(\omega)}{\varepsilon_0 E(\omega)} = 1 + \frac{P(\omega)}{\varepsilon_0 E(\omega)} = 1 - \frac{ne^2}{\varepsilon_0 m_e^* \omega^2} = 1 - \frac{{\omega_p}^2}{\omega^2}, \tag{2.1.13}$$

where $D(\omega)$ is the electric displacement field and $\omega_p = ne^2/\varepsilon_0 m_e^*$ is the so-called plasma frequency. Here, the most important parameter is the carrier density n. In many cases, we can influence this property via doping or the application of an electric potential to the material. A third option is the induction of free charge carriers via the absorption of light described in section 2.1. If the frequency ω of the incident light is smaller than the plasma frequency ω_p, the permittivity ε is negative and the material is reflecting the light. For frequencies above the plasma frequency, ε is larger than zero and most of the light will be transmitted into the material. A more

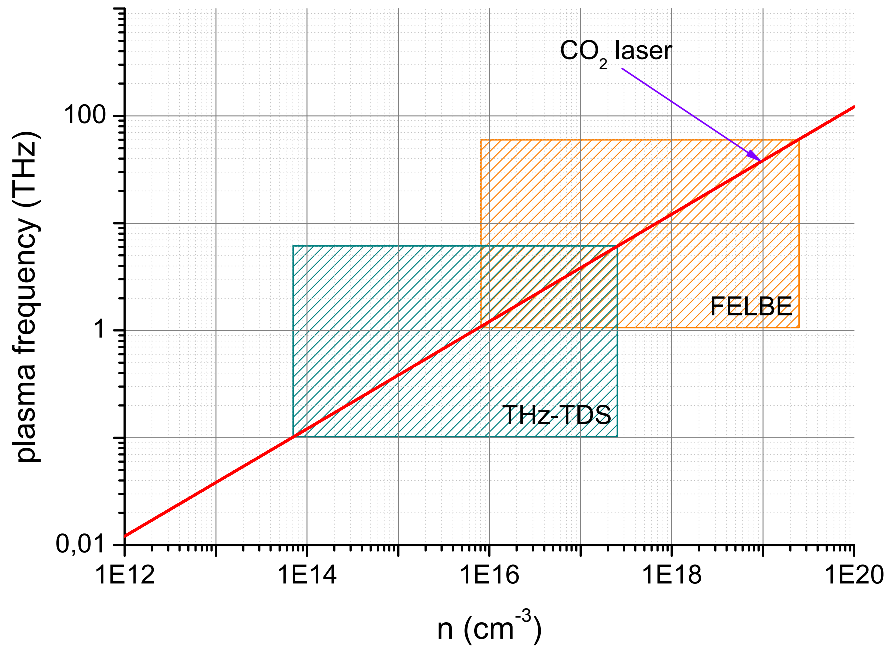

Figure 2.1.4.: *Plasma frequency in germanium as a function of the charge carrier density. The frequencies reachable with the applied laser sources in this work are included in the plot.*

figurative explanation is that the electrons in the material have a maximum oscillation frequency (the plasma frequency). For frequencies up to this level, the electrons are displaced, but can completely compensate the external electric field at the surface by their movement. As they oscillate, they emit electromagnetic radiation, which is the reflection. When the frequency of the incident light is too high, the electrons cannot compensate the external field completely and the crystal becomes transparent.

As shown in chapter 3.1, the strongest signal in s-SNOM is expected for slightly negative values of ε. As ε changes its sign at ω_p, the plasma frequency is a good starting point to estimate a suitable laser source for the investigation of a certain electron density. The plasma frequency of germanium as a function of the electron density is plotted in figure 2.1.4. It covers many orders of magnitude, as germanium can be doped to the level of $n \approx 10^{18}/\text{cm}^3$. Assuming realistic carrier densities in undoped, doped, and excited Germanium, we can conclude that an extremely broad range of laser sources is needed to investigate effects linked to plasma frequencies.

To this point, the Drude model was presented in the simplest case with just one type of charge

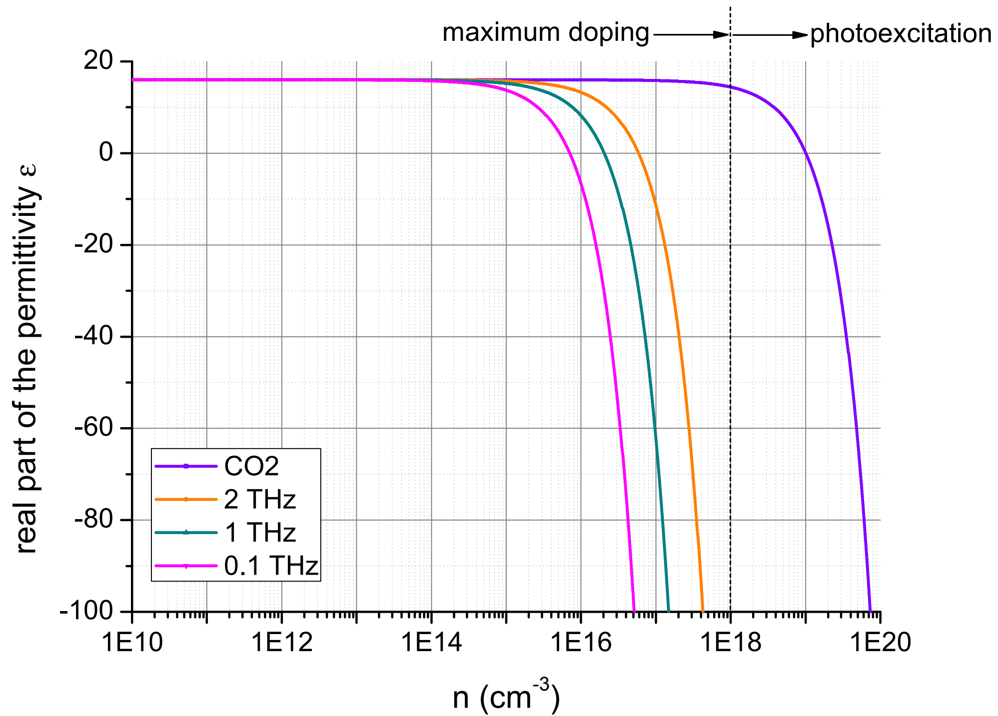

Figure 2.1.5.: *Permittivity as a function of carrier concentration for different laser sources utilized in this work. A frequency of 2 THz was used with the free-electron laser for the investigation of graphene field-effect transistors. The emission spectrum from a photoconductive antenna used in the TDS setup typically has a main peak at 1 THz and a lowest frequency of 0.1 THz.*

carriers contributing to the formula. For semiconductors, the holes in the valence band must be added as well as the electron scattering times in the material in order to include damping via scattering with phonons in the model. The real part of the permittivity is then given by [14]

$$\text{Re}[\varepsilon(\omega, n)] = \varepsilon_b - \frac{ne^2}{\varepsilon_0}\left(\frac{{\tau_e}^2}{{m_e}^*(1+\omega^2{\tau_e}^2)} + \frac{{\tau_h}^2}{{m_h}^*(1+\omega^2{\tau_h}^2)}\right). \tag{2.1.14}$$

Here, $\varepsilon_b = 16$ is the bulk permittivity and $\tau_{e,h}$ are the scattering times for the electrons and holes. It can be calculated by

$$\tau_{e,h} = \frac{m^*_{e,h} \cdot \mu_{e,h}}{e}, \tag{2.1.15}$$

where $\mu_e = 3900$ cm^2/Vs and $\mu_h = 1900$ cm^2/Vs are the electron and hole mobilities for germanium [12]. The result of equation 2.1.14 is depicted in figure 2.1.5 for several laser wavelengths

utilized in this work. For low carrier concentration, only the bulk permittivity ε_b contributes to the overall permittivity. For higher carrier concentrations, the permittivity becomes negative for all laser wavelengths. When the charge carrier density is increased, the Drude term in the equation leads to negative permittivity values. For a CO_2 laser, the carrier concentration must be very high ($\geq 10^{19}$/cm^3) in order to reach a negative ε, hence strong photoexcitation must be applied (see the next section). For lower frequencies, the needed carrier density needed is much lower and can be induced by doping as well. Hence, THz radiation enables much more sensitive examinations of small carrier densities as reported in [15]. In that work, the mapping of charge carrier densities is compared using a 2.54 THz alcohol laser and a CO_2 laser at the same Si-based transistor. The alcohol laser is clearly superior for sensing small carrier densities and their profiles.

Absorption of Light

As mentioned above, electrons can be moved from the valence band when enough energy (and momentum for indirect semiconductors) is transferred to the electron. As discussed before, this occurs due to thermal excitation, but can also be realized via photoexcitation. For the latter, a photon is absorbed in the crystal and excites an electron into the conduction band. The energies required to overcome the bandgap usually lie in the range of visible to near-infrared light (few hundred meV to some eV). In this section we regard single-photon absortion only, processes with more than one absorped photon, so-called multi-photon processes are neglected.

For indirect bandgap semiconductors such as Ge, a momentum of $|\vec{k}| = |L - \Gamma|$ needs to be transferred as well. As photons have a small momentum, the absorption needs to be supported by a third particle to fulfill the law of momentum conservation. For example, phonons have enough momentum and support the absorption of photons in indirect semiconductors. At room temperature, a high phonon density is present in the crystal, hence the absorption of light in indirect and direct semiconductors is comparable. Absorption spectra for different semiconductors are shown in figure 2.1.6. For photon energies larger than the bandgap, the absorption coefficient α is in the range of $\sim 10^5$ cm^{-1} and light does not penetrate deeply into the crystal. The steep fall of the absorption for larger wavelengths is a result of the lack of energy of the photons to excite electrons into the conduction band.

Physicists of all fields use the absorption of laser light to excite electrons into the conduction band and in order to investigate the charge carrier dynamics in time-resolved pump-probe experiments. In this work, absorption is used to excite charge carrier densities high enough to enhance the response of an s-SNOM system at the CO_2 laser wavelength of 10.6 μm. As shown in figure 2.1.5,

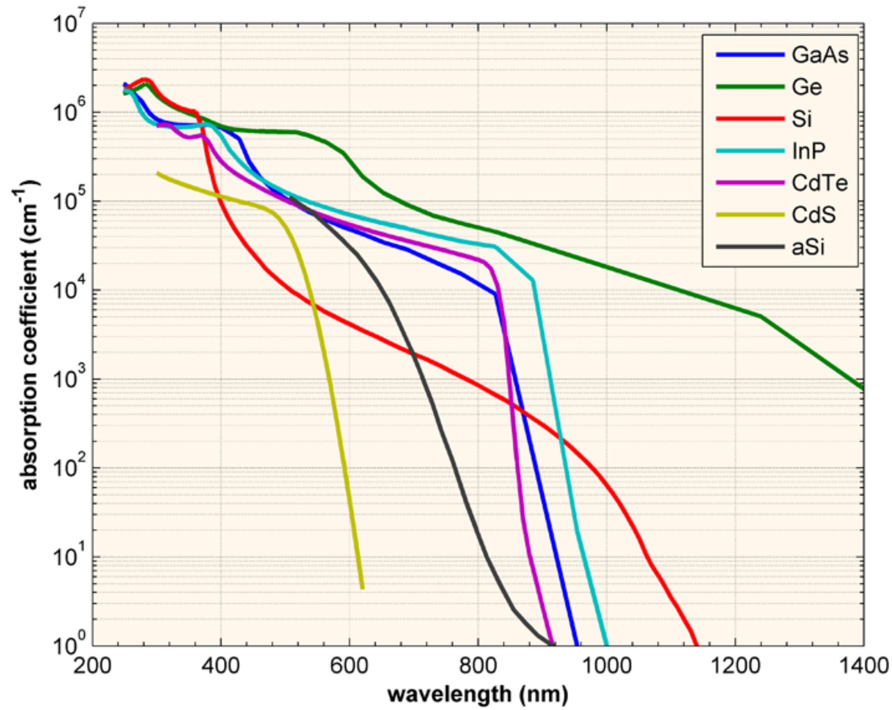

Figure 2.1.6.: *Absorption spectra of different semiconductors in the visible and near-infrared regime at room temperature (taken from [16]). Germanium has a high absorption coefficient and absorbs light even more than direct-bandgap semiconductors and a few orders of magnitude more than the most prominent indirect semiconductor silicon.*

carrier densities of up to 10^{19}/cm^{-3} are needed to to reach a slightly negative permittivity and to strongly enlarge the s-SNOM signal if it is operated with a CO_2 laser. To reach such high values, a pulsed laser system is needed as discussed in the following.

In general, the electron density in a semiconductor after a laser pulse is determined by the energy density ζ of the laser pulse, the absorption coefficient α, the refractive index η and the laser frequency ω [17]. The electron density is given by

$$n = \frac{8\pi\eta\alpha\zeta}{\hbar\omega(\eta+1)^2}. \tag{2.1.16}$$

In this work, a Ti:Sa amplified laser system[3] was used to excite electrons in an undoped germanium crystal. The laser covers a spectral range of $\lambda =$ (800 $\pm$ 10) nm due to its short pulse-length

[3]Legend, Coherent Inc.

of about 100 fs. For simplicity the center wavelength of 800 nm (corresponding to f = 374.75 THz ⇔ ω =2354.49 THz) is used for further consideration. At this wavelength, the refractive index of Ge is given by η = 4.71 and the absorption coefficient is α = 5.06·10^4/cm [18]. To excite a charge carrier density of 10^{19} /cm^3 a energy density of

$$\zeta = \frac{nh\omega(\eta+1)^2}{8\pi\alpha\eta} \approx 0.09\frac{\text{mJ}}{\text{cm}^2} \tag{2.1.17}$$

is needed. The laser system used in this work can easily supply mJ pulses at a repetition rate of 1 kHz. It is focused below one mm at the tip position to reach the needed power density with small overall laser powers to minimize thermal impact on the system. The pulse energy density itself is still below the damage threshold of $zeta$ = 50 mJ/cm^2 of the Ge crystal [17, 19].

When the electrons are excited into the conduction band, they stay there with an average lifetime τ_{CB} until they recombine back into the valence band. Here, several processes of recombination play a role. The dominant process and hence the carreir lifetime is determined by the charge carrier concentration. The processes are discussed in the following.

In band-band recombination, an excited electron directly recombines with a hole and emits a photon with the energy of the bandgap. It is the dominant process for semiconductors with few impurities and low doping concentration for low carrier concentrations. Lifetimes can reach several hundred μs for undoped, indirect semiconductors [20]. The second process is the recombination via impurities (also called Shockley-Read-Hall or SRH recombination). Here, the electron falls into an energy state which lies in the bandgap due to an impurity atom or an unsatisfied atomic bond as a result of imperfect crystal growth. The energy is transferred to a phonon. From that state, it recombines with a hole and moves into the valence band. When the crystal has many impurities, this process has a much higher probability than band-band recombination and dominates the lifetime of the charge carriers. The third process is Auger recombination. Here, two electrons in the conduction band scatter where one gives its energy to the other one and recombines into the valence band. No photon is emitted during the process. It only occurs at very high charge carrier densities when electron-electron scattering becomes more likely. It is the dominant process for charge carrier densities $n \geq 10^{18}$ cm^{-3} and lowers the lifetimes to 100 ns [20]. When such a high charge carrier densitiy is reached, it is considered an electron-hole plasma.

An overview of carrier lifetimes in germanium is shown in figure 2.1.7. The lifetime depends on many parameters as the crystal growth, the type of doping and if the carriers density is determined by the doping concentration ($k = 0.01$) or induced via photoexcitation ($k = 100$). k is a dimensionless parameter used in the simulation. In this work, undoped germanium was pumped

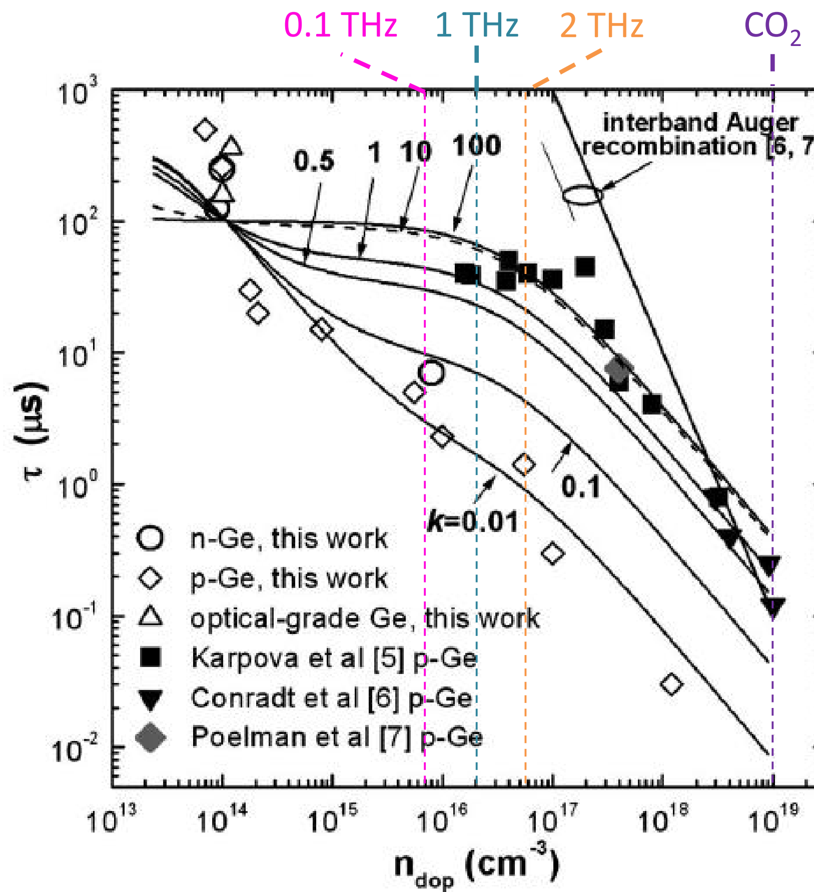

Figure 2.1.7.: *Lifetime of charge carriers in germanium as a function of the carrier density. The solid lines correspond to simulated values, the symbols to the experimental results. With increasing concentration the lifetime decreased down by several orders of magnitude. The colored lines marks the positions where the permittivity changes sign for different laser sources as predicted by the Drude model. Illustration taken from [20] and adapted.*

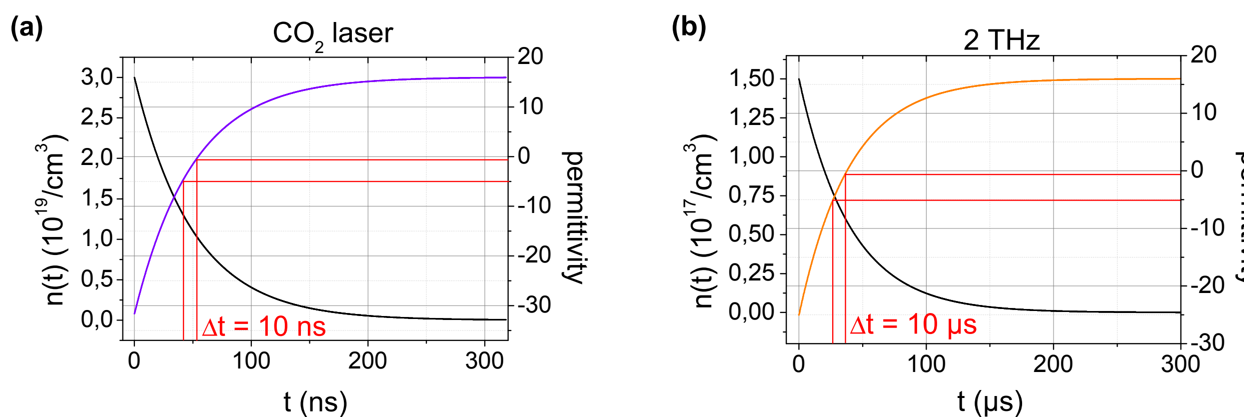

Figure 2.1.8.: *Temporal behavior carrier densities and the corresponding permittivity in germanium after photoexcitation for probing with a CO_2 laser (a) and a 2 THz source (b). The time window of enlarged s-SNOM is depicted by the red lines. Calculations were performed with the same material parameters as in section 2.1 for both cases due to lack of experimental reports on the material parameters during such a dynamic process.*

with a pulsed laser, hence the values for $k = 100$ and the experimental values indicated with black squares and triangles apply. The colored stripes mark the positions where the carrier density is high enough to change the sign of the permittivity for the different laser sources used in this work and, hence, strongly increased scattering cross-sections in s-SNOM are expected. For the CO_2 laser, carrier lifetimes are definitely shorter than one μs and rather in the range of 100 ns. This value is in agreement with other surface-sensitive techniques as reported in [14], who find 50 ns in their experiment. For the THz regime, the lifetimes are in the range of ten to hundreds of μs.

Altogether it is not easily possible to state an exact lifetime for charge carriers when they are induced by a focused laser. On the one hand, the carrier lifetime increases as soon as the first charge carriers recombine. Secondly, the carrier density decreases due to diffusion and hence the lifetime increases even without the recombination of charge carriers. The diffusion coefficient also depends on the carrier density. Overall, there are many phenomena interacting with each other in such a dynamic process. The full determination of the interaction of these phenomena is complex and is not part of this work. Interested readers may refer to [17], where these effects are discussed further.

In order to estimate the period of time with an increased s-SNOM signal ($\varepsilon = -1$ to $\varepsilon = -5$, see chapter 3), the temporal behavior of the permittivity is calculated for two cases in figure 2.1.8.

For the first case (CO_2 laser), a photoexcited charge carrier density of $n = 3 \cdot 10^{19}$ cm^{-3} excited at $t = 0$ with a constant carrier lifetime of 50 ns is assumed. For the second case (2 THz) an photoexcited charge carrier density of $n = 1.5 \cdot 10^{17}$ cm^{-3} and a lifetime of 40 μs are chosen as parameters. For both cases, the carrier densityis assumed to decay exponentially and the corresponding permittivity is calculated with the Drude model. The results are depicted in figure 2.1.8. The exponential decay of the carrier density is shown on a timescale of 300 ns (a). In the same time, the permittivity changes from $\varepsilon =$ -30 to its bulk value of $\varepsilon =$ 16. The period of time when a strong s-SNOM signal is expected is only 10 ns long. In (b) the timescale is three orders of magnitude longer and the time of enlarged s-SNOM signal is 10 μs.

The CO_2 experiments performed in section 5.1 use a pulsed excitation laser with a repetition rate of 1 kHz ($\widehat{=} 1$ ms). This leads to a small duty cycle of 10^{-5}, and the expected signals are consequently low. Still, the modulation of the charge carrier density could clearly be observed. However, the effects would be much larger for the same experiment performed with a THz laser source, as the duty cycle would be 10^{-2} for this case due to much longer charge carrier lifetimes.

2.2. Plasma Waves in Graphene Field-Effect Transistors

In the 90s, the two famous theoreticians M. Dyakonov and M. Shur published a series of seminal papers [21–24] on the behaviour of electronic transport in a 2-dimensional conducting channel in a field effect transistor (FET). They showed that the charge carriers motion can be described by hydrodynamic equations, which is the reason they are often named electron fluid. This fluid has some fascinating properties and develops a so-called plasma wave, which can show resonant behavior in certain cases. The scientific impact of their theoretical work is large, however until today the evidence for the existence of the plasma waves is indirect and phenomenological [25–27]. Particularly, they have not been observed directly, as the conducting channel in conventional semiconductor devices is buried under encapsulation material and cannot be accessed directly. There is strong interest in a direct observation because of an ongoing debate whether such FETs can also be used as emitters of THz radiation arising from the non-thermal excitation of plasma waves by a strong source-drain current. This was predicted by the work of Dyakonov and Shur [21, 24] and some experimental indications for this type of emission exist [28], however the results could be explained by thermal emission as well.

While plasma waves have been studied and been taken to practical applications [29–32] in many classical semiconductor devices, the examination of such waves in graphene field-effect transis-

tors (GFETs) only recently became part of fundamental research in the recent years. The impact of electric fields in graphene properties is the main topic in the first work on graphene by Novoselov and Geim [33] and has been exploited in GFETs since then [34–36]. For this work, GFETs with the channel of the transistor located on top of the device and without an encapsulation layer were produced, hence the plasma waves could be investigated directly. The devices have sizes in the low micrometer range, thus an investigation with a sub-diffraction technique is mandatory and has been performed via s-SNOM in this work. The results are presented in section 6. The theoretical background and the resulting properties of a gated, two-dimensional electron fluid are discussed further in this section.

The main results of Dyakonov and Shur from their theoretical reports [21, 24] are presented in the following. They discuss the properties of a classic field effect transistor as shown in 2.2.1. The source and drain electrodes are separated by a layer which is non-conducting when the gate voltage is zero[4]. When a voltage is applied, electrons form a thin conducting channel next to the gate insulator, which can be regarded as a two-dimensional electron gas. The electron transport behavior is dominated by scattering on phonons or impurities if the channel length L is long. If the channel length L is small enough and the material is of good quality, the average momentum relaxation time τ is much longer than the transit time of the electrons to move from drain to source. Hence, an average electron passes the device without a single scattering event with a phonon or an impurity, a motion often called ballistic transport. However, the electron density in the two-dimensional layer can easily reach n=10^{12} cm^{-2} leading to an average electron distance of $\approx$ 100 Å. As the average distance in between electron-electron scattering is in the same range, the number of electron-electron collision during transit is high even for a small device with a length of $L = 1$ μm. The transport behavior is thus dominated by electron-electron scattering and consequently the electron gas behaves like a fluid which can be described by hydrodynamic equations.

For thin insulating layers in between the gate and the channel, the electron density n is given by

$$n = \frac{CU_g}{e}, \tag{2.2.1}$$

[4]This is a result of different doping (n and p) of source and drain and the layer in between them. Without a gate voltage, current cannot flow from source to drain as one of the pn-junctions cannot be passed by electrons. However, this effect does not play a role in the theory.

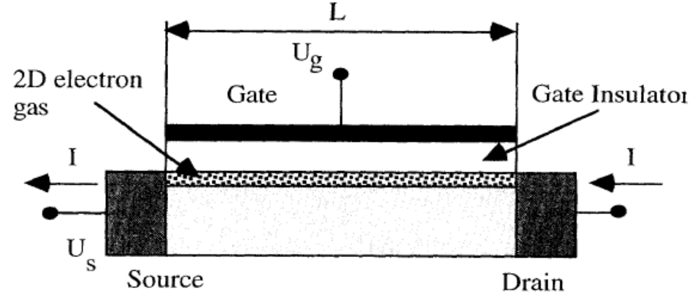

Figure 2.2.1.: *Typical field-effect transistor. Source and drain are separated by a non-conducting layer. When a gate voltage is applied to the gate, a conducting, thin channel of a 2D electron gas is formed. Illustration taken from [21].*

where, C is the capacitance per unit area and U_g is the effective gate voltage in the channel[5]. The Euler equation of motion for the one-dimensional case is given by

$$\frac{\partial v}{\partial t} + v\frac{\partial v}{\partial x} = -\frac{e}{m_e^*}\frac{\partial U_g}{\partial x} - \frac{v}{\tau}. \tag{2.2.2}$$

Here, $v(x,t)$ is the local electron velocity, $\partial U_g/\partial x$ is the longitudinal electric field in the channel and the last term accounts for damping from electron scattering at phonons and impurities. When the equation of motion is solved, the continuity equation

$$\frac{\partial U_g}{\partial t} + \frac{\partial (U_g v)}{\partial x} = 0 \tag{2.2.3}$$

must be taken into account. The equations 2.2.2 and 2.2.3 coincidence with the hydrodynamic equations of shallow water given in many textbooks on hydrodynamics (e. g. [37, 38]). In shallow water, the water level is the analogon to the charge carrier density and is determined by gravity. In the case of a field-effect transistor, the electron density is controlled by the gate voltage and not by the Poisson equation as in the standard two-dimensional case. This fact has drastic consequences on the behavior of the electron fluid, which will be discussed subsequently.

As a first case, let us assume the following boundary conditions of the system with a constant current in driven from source to drain and a constant voltage U_g applied to the gate. For the

[5]In classic field effect transistors, a threshold voltage must be reached before the channel is formed. As the graphene channel is present at all times, no threshold must be reached and the offset can be neglected here.

boundary conditions this gives us

$$U_g(x=0) = \text{const} \quad \text{and} \tag{2.2.4}$$

$$j(x=L) = \text{const}, \tag{2.2.5}$$

where the current density $j = n \cdot v$ is given by the electron density and the electrons' propagation speed. Dyakonov and Shur define the boundary conditions as a constant voltage in between source and gate and a constant current density at the drain ($x = L$), however this is arbitrarily chosen and can be interchanged.

As a next step, they regard the influence of a small, random fluctuation of the propagation speed and the gate voltage. Let the voltage and velocity at $x =$0 be modulated

$$U_g(x=0) = U_0 + U_{ext}e^{i\omega t} \quad \text{and} \tag{2.2.6}$$

$$v(x=0) = v_0 + v_{ext}e^{i\omega t}. \tag{2.2.7}$$

Taking the boundary conditions and into account, the real and imaginary part of $\omega = \omega' + i\omega''$ can be derived. They read as follows:

$$\omega' = \frac{|s^2 + v_0^2|}{2Ls}\pi l \tag{2.2.8}$$

$$\omega'' = \frac{s^2 + v_0^2}{2Ls} \ln \left| \frac{s + v_0}{s - v_0} \right| \tag{2.2.9}$$

with $s = \sqrt{eU_0/m_e^*}$, the speed of the plasma wave. Here, l is an odd integer for $|v_0| \leq s$ and an even integer for $|v_0| \geq s$. As a realistic case is given when $|v_0| \leq s$, this will be considered further.

The important result at this point is the logarithmic dependence of ω''. The term $(s+v_0)/(s-v_0)$ is the reflection ratio of the wave at $x = L$. It is larger than one for $v_0 \leq s$, hence the reflection leads to an amplification in the system! This instability to an external fluctuation is an interesting feature of such a device. By running a constant current through the device, strong plasma oscillations are induced in the device and hence electromagnetic waves should be emitted. Conclusively, the device could in general be a radiation source. Mechanisms opposing a growth to an infinitely strong wave are external and internal friction by electron scattering at phonons or impurities and internal friction due to the viscosity of the electron fluid.

However, with other boundary conditions, the device can also be used as a detector. This is the

second case which will be discussed in this work. Let us assume we apply a voltage $U_g = U_0$ at the gate and run no current through the device. We now induce an external AC field (e.g. by illumination) in between gate and source. The boundary conditions for this case are

$$U_g(x=0) = U_0 + U_{ext}\cos(\omega t) \quad \text{and} \tag{2.2.10}$$
$$j(x=L) = 0. \tag{2.2.11}$$

We will in general search for solutions of 2.2.2 and 2.2.3 in the form of $U_g(x,t) = U_0 + U_1(x,t) + U_2(x,t) + ...$ and $v(x,t) = v_0 + v_1(x,t) + v_2(x,t) + ...$ Here, U_1 is proportional to U_{ext}, U_2 to U_{ext}^2, etc. The same applies for the terms of v. For the first order, the differential equations reform to

$$\frac{\partial v_1}{\partial t} + \frac{\partial u_1}{\partial x} + \frac{v_1}{\tau} = 0 \tag{2.2.12}$$
$$\frac{\partial u_1}{\partial t} + s^2 \frac{\partial v_1}{\partial x} = 0, \tag{2.2.13}$$

where $u_1 = eU_1/m$. The solutions of 2.2.2 and 2.2.3 for the first-order case are given by

$$U_g = \text{Re}[(C_1 e^{ik_0 x} + C_2 e^{-ik_0 x})e^{-i\omega t}] \quad \text{and} \tag{2.2.14}$$

$$v = \text{Re}[\omega k_0^{-1} s^{-2}(C_1 e^{ik_0 x} + C_2 e^{-ik_0 x})e^{-i\omega t}] \tag{2.2.15}$$

where C_1 and C_2 are given in [24], and depend linearly on the field U_{ext}, a parameter which cannot be controlled with absolute measures in our experiment. $k_0 = k_0' + ik_0''$ has a real and an imaginary part. The dispersion of the wave is given with

$$k_0' = \frac{\omega}{s}\sqrt{\left(\frac{(1+\omega^{-2}\tau^{-2})^{1/2}+1}{2}\right)} \quad \text{and} \tag{2.2.16}$$
$$k_0'' = \frac{\omega}{s}\sqrt{\left(\frac{(1+\omega^{-2}\tau^{-2})^{1/2}-1}{2}\right)} \tag{2.2.17}$$

We have now derived the full wave equation induced by an AC field at the gate for the first order of the differential equation. Higher orders are much less efficiently excited and are neglected at this point. When the solutions of the first order at $x = 0$ and $x = L$ are calculated, we find

$$\Delta U = U_1(x=0) - U_1(x=L) \neq 0! \tag{2.2.18}$$

The plasma wave induces constant a potential in between the source and drain and hence a DC current is flowing through the device. This effect is quite commonly named rectification and can be detected with simple electronics, although the external AC field may have a frequency way higher than the maximum operating frequency of the transistor (in this work ω is in the THz range). While Dyakonov and Shur evaluate this behavior further in their publications [22–24] and interested readers can refer to their reports for more details, we discussed their findings in sufficiently to analyze the experimental findings of section 6.

As mentioned before, the DC current induced by electromagnetic radiation has been observed in many cases, however the scientific community disagrees on the origin of the findings. Some scientists believe the source are plasma waves as predicted by Dyakonov and Shur [25–27, 39, 40], while others believe the origin is a photothermal effect as shown in [41]. A shown above, the plasma wave show a decaying electric field in the channel, while a photothermal effect is not associated to an electric field in the channel. The detected signals in s-SNOM depend on the local electromagnetic field as shown in section 3.2, hence an investigation with s-SNOM is a suitable tool to perform an exploration of the plasma wave. As shown in section 6, plasma waves could be found using s-SNOM on GFET devices. They show a characteristic response to the gate voltage, finally proving the existence of the waves with a direct approach.

2.3. Near-Field Microscopy

Near-field microscopy is the most important scanning probe technique for this work. It circumvents the diffraction limit of conventional microscopes and enables the access to many fundamental phenomena in solid state physics at the nanometer scale. Unfortunately, there are many abbreviations for this technique. Some variations are due to different order and highlighting of the words, like SNOM (scanning near-field optical microscopy), NSOM (near-field scanning optical microscopy), NFOM (near-field optical microscopy) and near-field optical-scanning (NFOS) microscopy. As the focus of the research went to longer wavelengths in the last decade, there is also the term SNIM (scanning near-field infrared microscopy). Some other names are obsolete, like optical stethoscopy. In the THz range, the most common term is THz near-field microscopy, which does not have an abbreviation, since SNTHzM cannot be enunciated. In this work, both infrared and THz near-field microscopes are utilized. Due to a lack of an abbreviation for those two techniques, SNOM is chosen in the text, as it is the oldest and most common term. However, the fundamentals of these techniques are the same. Evanescent waves are probed to break the

diffraction limit of conventional microscopy.

In conventional microscopy, the achievable resolution Δ is defined by the distance two objects must have to be seperated in the image. It is fundamentally limited by the wavelength used in the microscope. In 1873, Ernst Abbe, a pioneer in modern optics, determined the lower limit of the resolution [2] as

$$\Delta = \frac{\lambda}{2n\,\mathsf{sin}(\theta)}. \tag{2.3.1}$$

Here, λ is the utilized wavelength, n is the refractive index of the medium in between the object and the microscope and θ is half of the opening angle of the light collected with the lense. Increasing the refractive index (e. g. with immersion oil) and the opening angle may improve the resolution to a certain degree, however for a fixed wavelength, there will always be a limit for the resolution.

There are several ways to derive equation 2.3.1 or a similar formula with different approaches. Two of them are presented below.

Rayleigh Criterion

The simplest way to derive this formula is the heuristic approach with the so-called Rayleigh criteria. Let us assume a slit as a light source. The intensity distribution behind the slit on a screen is given by

$$I(\alpha) \propto \left(\frac{\mathsf{sin}\left(\frac{\pi b\,\mathsf{sin}(\alpha)}{\lambda}\right)}{\frac{\pi b\,\mathsf{sin}(\alpha)}{\lambda}} \right)^2. \tag{2.3.2}$$

Here, b is the width of the slit and α is the propagation angle of the light after the slit. The intensity has a maximum value at $\alpha = 0$ and further maxima at higher angles with the criterion of

$$b \cdot \mathsf{sin}(\alpha) = \left(k + \frac{1}{2}\right) \cdot \lambda \quad \text{with} \quad k = 1, 2, 3, ... \tag{2.3.3}$$

and minima at

$$b \cdot \mathsf{sin}(\alpha) = k \cdot \lambda \quad \text{with} \quad k = 1, 2, 3, ... \tag{2.3.4}$$

For small angles we can substitute sin$(\alpha) \approx \alpha$. Hence the main maximum has a half-width Λ of

$$\Delta = l\,\tan(\alpha) \approx \frac{l}{b} \cdot \lambda \tag{2.3.5}$$

at a screen at the distance l. Let us now assume two slits as light sources. Rayleigh argues that the maxima of the two sources can be distinguished only if they are separated by at least the length of the half-width. As Δ depends on the wavelength, the achievable resolution is clearly limited by it. Higher resolution is only possible when the distance of the screen is the same range as the width of the slit. This behavior becomes more obvious when taking a look at paragraph 2.3.1.

Pendry's Superlense Approach

In the highly recognized paper introducing superlensing as a concept for unlimited resolution in microscopy, John B. Pendry gives a very conclusive explanation for the diffraction limit [42]. Considering a point dipole in a sample (x-y-plane) with the frequency ω in front of a lense, w.l.o.g. the emitted electric field can be written as

$$\vec{E}(\vec{r}, t) = \sum_{k_x, k_y} \vec{E}(k_x, k_y) \cdot e^{i(k_x x + k_y y + k_z z - \omega t)}. \tag{2.3.6}$$

The wavevector $\vec{k}$ can be written as

$$\vec{k}^2 = k_x^2 + k_y^2 + k_z^2 = \frac{\omega^2}{c^2}. \tag{2.3.7}$$

Let us now regard waves traveling into z-direction. The k_z component can be expressed for two cases

$$k_z^I = \sqrt{-k_x^2 - k_y^2 + \frac{\omega^2}{c^2}} \qquad \text{for} \qquad \frac{\omega^2}{c^2} \geq k_x^2 + k_y^2, \tag{2.3.8}$$

$$k_z^{II} = i\sqrt{k_x^2 + k_y^2 - \frac{\omega^2}{c^2}} \qquad \text{for} \qquad \frac{\omega^2}{c^2} \leq k_x^2 + k_y^2. \tag{2.3.9}$$

Inserting this back into equation 2.3.6, we derive two cases for the field:

$$\vec{E}(\vec{r}, t) = \sum_{k_x, k_y} \vec{E}(k_x, k_y) \cdot e^{i(k_x x + k_y y - \omega t)} \cdot \underbrace{e^{ik_z^I z}}_{\text{prop.}} \tag{2.3.10}$$

$$\vec{E}(\vec{r}, t) = \sum_{k_x, k_y} \vec{E}(k_x, k_y) \cdot e^{i(k_x x + k_y y - \omega t)} \cdot \underbrace{e^{ik_z^{II} z}}_{\text{evan.}}. \tag{2.3.11}$$

Equation 2.3.10 will lead to propagating waves which will reach the lense and are detected in a macroscopic experiment. The in-plane components are restricted by $k_x^2 + k_y^2 \leq \omega^2 c^{-2}$, hence there is a maximum wavevector k_{max}, thus limiting the smallest detectable lateral change of optical information in the real space. Pendry approximates the limit of the resolution to

$$\Delta \approx \frac{2\pi}{k_{max}} = \frac{2\pi}{\sqrt{k_x^2 + k_y^2}} \geq \frac{2\pi c}{\omega} = \lambda. \tag{2.3.12}$$

For the evanescent case, the wavevectors k_x and k_y do not have an upper limit and thus can carry information of the optical properties with infinite resolution. The key to circumvent the diffraction limit is the detection of these waves. The following two chapters feature the ideas for successfully performing such an experiment.

2.3.1. Aperture-SNOM

Roughly 50 years after Abbe determined the limit of conventional microscopy, E. Synge proposed a method to achieve resolutions beyond the diffraction limit [3]. His idea was simple: If a small aperture is positioned close to the sample surface, the illuminated area is determined by the size of the aperture and hence the resolution is not limited by the wavelength anymore. At the time he proposed the idea, he was not able to build such a small aperture, neither could he place it precisely enough above the sample surface. Also, light detectors at that time were too insensitive to detect the small amount of light passing through such a small aperture. The technical developments allowing for such an experiment needed decades to be developed. He died way before his fundamental idea for aperture-SNOM (a-SNOM) could ultimately pay off.

The first report on such super-resolution technique came by Ash and Nicholls in 1972 [5], where they used microwaves with $\lambda = 3$ cm and achieved a resolution of $\lambda/20$. Here, the long wavelength facilitated the experiment drastically, as the aperture had a size of 1.5 mm and no nanopositioning system was needed to place the aperture on the sample. For shorter wavelengths, scanning probe microscopy needed to be developed first. The first report in the visible regime was published in 1984 by Pohl et al. [43], where a resolution of $\lambda/20$ was achieved at $\lambda = 488$nm.

Today's aperture-SNOM still use the same principle. An edged glass fiber is coated with a metal film to achieve a small hole at the end. To approach and scan the sample surface, the so-called tuning-fork method is applied, which works in a very similar way as an atomic force microscope (AFM) in non-contact mode (see section A.1 for a detailed description of the method). The fiber oscillates with a small amplitude in the x-y-direction. When it reaches the sample surface, the

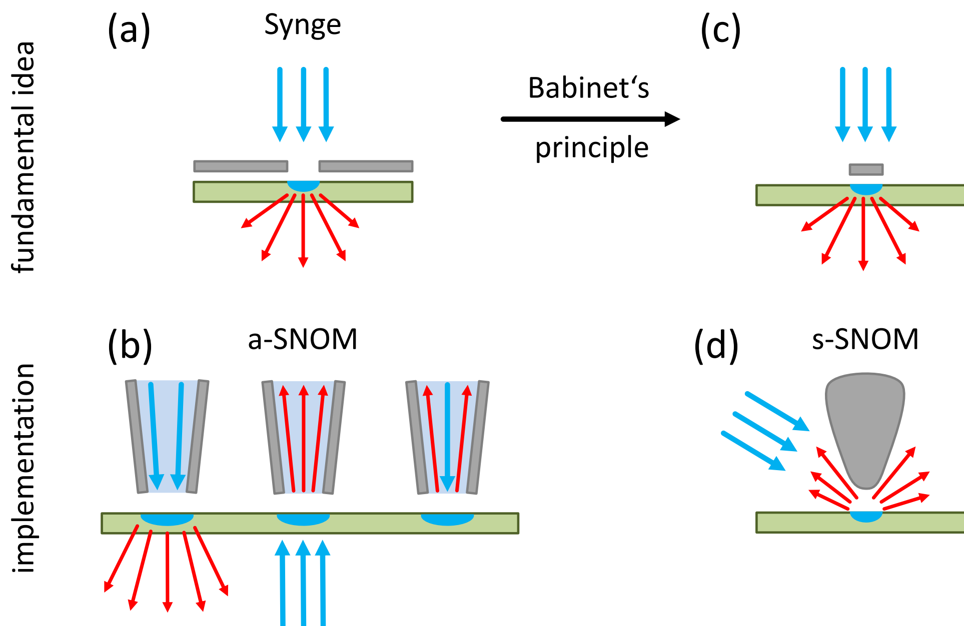

Figure 2.3.1.: *Concepts of SNOM: (a) shows the idea of Synge to place a small aperture close to a sample. The illumination comes from the top (blue) and the sample is illuminated only below the hole (blue semi-ellipsoid). The scattered light (red) is passing the aperture and travels in all directions. The experimental implementation of three different a-SNOM is shown in (b), where metal-coated glass fibers are applied. The illumination is always illustrated with blue, the scattered light with red arrows. The relevant part of the sample, contributing to the scattered signal, is shown with the blue semi-ellipsoids. When Babinet's principle is applied to Synge's idea, the aperture is replaced with a small scatterer (c). The experimental implementation is shown in (d), where a metallized AFM tip is used as such a small scatterer close to the sample surface.*

oscillation is damped and the eigenfrequency shifts. This shift is used as the control parameter in the a-SNOM. The optical resolution achieved with this technique is limited by two factors. Firstly, the amount of light transmitted through a nanoscale hole is very small, hence the detector must be sensitive enough for the application. Secondly, the light penetrates into the metal coating via the skin effect [44], leading to a lower resolution limit of roughly 30-50 nm [45].

Today, a-SNOM is not longer a topic of fundamental research in physics but a rather common and commercially available technique applied in many other scientific areas like biology. A recent and detailed review article is found in [46]. One major point why a-SNOM has less impact in fundamental solid state physics in the recent years can be explained with the trend towards shorter and longer wavelengths. Here, compared to the visible regime, many more phenomena are far less explored (compare section 2.4). At smaller wavelengths, the diffraction-limit is less problematic, as an UV or X-ray microcope achieves good resolution intrinsically. For larger wavelengths, the interest for tools for the investigations on the nanometer scale is unevenly higher due to the reduced resolution with conventional techniques. However, the power transmittance of this fiber-based technique is too tiny to be utilized in this field of research.

Figure 2.3.1 (a) illustrates the idea of Synge. An aperture is placed above a sample and illuminated. The light reaches the sample only in the area where the aperture's hole is placed. The scattered light from the sample can be detected. The experimental implementation of Synge's idea is depicted in (b). Three types of a-SNOM are shown. All utilize a metal coated glass fiber with a small opening at their ends. On the left, the illumination is performed through the fiber and the light is collected after transmission through the sample. In the second case (center), the illumination comes from the bottom and the scattered light of the small area is collected with the fiber. Both these two methods have the disadvantage that only transparent samples can be investigated. In the third case (right), both illumination and collection of the light is performed through the fiber. This can be performed on absorbing samples, however the light must go through the small hole of the fiber two times, strongly reducing the amount of light which can finally be detected.

2.3.2. Scattering-SNOM

According to Babinet's principle, the diffraction pattern of an aperture and its complementary body is the same. For this reason, the aperture proposed by Synge can be exchanged by a small scatterer placed close to the sample surface, depicted in figure 2.3.1 (c). When this idea is transferred to a-SNOM, the fiber can be replaced by a metal AFM tip (d). Here, the illumination is coming from the side and the tip scatters the light arising from the area below the tip towards

the detector. Although this vivid explanation of the s-SNOM working principle is plausible, the theory of the fundamental mechanisms and the detection principle is much more complex and described in detail in section 3.1. In the following paragraph the state-of-the-art of s-SNOM is briefly described.

The tip-sample system is usually illuminated from the top/side direction via a non-dispersive parabolic mirror (or less often a lense) and most groups collect the backscattered light with the same mirror [7, 15, 47–52] which has the advantage that only one parabolic mirror (with 6 degrees of freedom) has to be properly aligned. However, some groups also use a forward scattering detection scheme [53] for measurements less sensitive to beam fluctuactions. It is also applied in time-domain-spectroscopy-SNOM (TDS-SNOM) for easier alignment of the setup and due to a larger dynamic range of the detectors [54–56] allowing for the correct separation of the large background and small near-field signals. Although the setups are most commonly illuminated from the top under an angle, some groups use an attenuated total reflection (ATR) illumination scheme from the bottom [57]. This strongly reduces influences of the background, but has the disadvantage that a prism is needed as a light-transmissive optical element, taking away the advantage of a completely wavelength-independent setup. Today, this illumination is rarely applied in s-SNOM as the data interpretation is complex when the light has passed the sample and the prism two times. Recently, ATR illumination was applied in tip-enhanced Raman spectroscopy (TERS) [58].

All s-SNOM techniques have two aspects in common: Firstly, the background is sufficiently suppressed via the modulation of the tip sample distance, as the near-field is only present close to the sample surface (see section 3.2.3). While most group apply higher-harmonic demodulation via lock-in amplifiers, there is also a report on phase-domain sampling of the cantilevers' phase correlating it to the detectors' signal with data post-processing [59]. Secondly, the resolution only depends on the apex of the utilized tip and not on the wavelength. The typical resolution lies in the range of a few ten nanometers [7, 15, 60], even values of 3 nm were reported [45].

Since the first s-SNOMs reported by Inouye et al. and Zenhauser et al. in 1994 [4, 45], the technique has come a long way. Only two years later, the first mid-infrared s-SNOM showed a resolution of 17 nm at a wavenlength of $\lambda = 10.6$ μm [6], the clear proof that the wavelength does not play a role in the resolution of s-SNOM. After that, homodyne [49], heterodyne [50], pseudo-heterodyne [51] and synthetic optical holography [61] have been developed to seperate amplitude and phase of the scattered near field signal. Broadband radiation allows for the deduction of nano-optical spectra via nano-FTIR [62]. In the recent years, low-temperature s-SNOM is in

the scope of several researchers [63–67]. Another field of research was opened through the integration of pump-probe setups with s-SNOM for time-resolved measurements with fs resolution [68–70].

In general s-SNOM maps a material contrast at the nanoscale (see section 3.1). With correct referencing, it can even deliver quantitative values for the dielectric constants [52]. Almost every effect influencing the optical properties that can be measured at the macroscopic scale can be measured with s-SNOM as well. Below, some examples are given. Areas of different electron densities can be identified with the technique [15] including also transient phenomena like plasmon resonances [71, 72]. Phonon resonances can be identified [63, 73–77] as well as vibrational bonds of chemical reaction products [62, 78, 79]. It has been successfully combined with *Raman* spectroscopy [58, 80] and can also be used to directly map stress and strain [81, 82].
The integration of s-SNOM to infrared and THz beamlines of large scale facilities is becoming more and more common to utilize the unique characteristics of either ultra-broadband radiation from synchrotrons [53, 77, 83–85] or narrow-band radiation from free-electron lasers [7, 63, 86, 87] realized at the Helmholtz-Zentrum Dresden-Rossendorf by the group of Prof. Eng.

Low THz s-SNOM has seen a huge interest in the last few years. Many scientific groups working with s-SNOM reported on their progress in this field. There are few light sources available offering narrow-band radiation, either tunable and from free-electron lasers with low availability [7] or only at distinct laser lines [15, 48, 88]. These sources offer the investigation of resonant behaviour. Other groups use short and broadband THz pulses with time-domain spectroscopy to deduce nano-spectroscopic information in the low THz range [54–56, 89]. One advantage of this approach is the direct probing of the near-field compared to detection schemes where the near-field needs to interfere with a background before it can be detected.

In this work THz s-SNOM utilizing both broadband radiation from a photoconductive antenna as well as narrow-band radiation from the free-electron laser FELBE has been applied. The results are presented in section 4.3 and 6.

2.4. THz Optics

THz radiation is part of the electromagnetic spectrum and is quite roughly determined. Definitions cover frequencies of 0.1 THz up to the so-called multi-THz range of 40 THz[6]. The main

[6] As the term THz has become a buzzword in the last decade, it is quite often used for mid-infrared radiation as well.

reason part of the electromagnetic spectrum is commonly named THz gap is the lack of available THz sources, espacially from 1 to 10 THz . Only in the last three decades, THz sources have become available to more than just a few pioneer scientists and almost none of them are turn-key devices. The main reason for the lack of sources is that concepts of optics and electronics cannot simply be transferred to this range. To achieve an inversion population the lasing material must be cooled down [7] with liquid helium. Also, electronic devices reaching such high frequencies become highly inefficient and thus deliver very low radiation power. They same is valid for THz detectors. Although there has been quite remarkable progress in the field, to the best of the authors' knowledge, the area of 3-10 THz still is not covered by table-top laser sources suitable for s-SNOM. In this section, an overview of physical phenomena in the THz range, of THz sources and THz detectors will be given.

Physical Phenomena in the THz regime

Many interesting physical phenomena in THz range exist, and there has been a significant rise in the number of publications in the last years. Raising no claim to completeness, some are listed below.

- Phonon energies can lie in the low THz range. e. g., for GaAs, acoustic and optical phonons have frequencies of 0.5 and 0.8 THz [90].
- As shown in section 2.1, charge carrier densities of 10^{16} to 10^{18} /cm^3 have plasma frequencies in the THz range. Thus, THz radiation is very sensitive to small electron concentrations.
- Zeeman splitting effects can be investigated with THz radiation [91].
- Another effect controlled by applied magnetic fields is the cyclotron resonance, which can lie in the THz region and is of relevance for THz detection. [92].
- Dopants in semiconductors can have ionization energies of few meV [93] (e. g. an Sb donor in Ge will induce a resonance of the system at 2 THz).
- In so called highly correlated material systems, energies of lattice vibrations, spin waves and electron-hole pairs can have energies in the low THz range. As these phenomena have the same energy range, they are strongly entangled and show fascinating properties. [94].
- Surface Plasmons exist also at THz frequency [95].
- Magnons usually have resonances in the low THz regime [96].

[7] 1 THz $\widehat{=}$ 300 μm $\widehat{=}$ 4.1 meV $\widehat{=}$ 47.6 K

- Quite recently, non-linear effects due to ultra-strong THz fields could be observed in graphene [97].

Interested readers may also refer to the review paper by Lewis [1].

THz sources

The lack of high power and reliable THz sources has been the main reason for the slow development in THz science and technology in the past. However, as interest in the exploration of the THz regime has risen in the last decades, now more and more THz sources are available using different concepts. As THz science as a field of research may be relatively unfamiliar to some readers, the most common concepts will be briefly introduced in this section. For more details, readers may consult the review articles and book given in [98–100]. The first THz sources were incoherent and broadband radiation sources, like the globar source and the mercury lamp (the latter is still used in many FTIR applications today), however state-of-the-art THz sources are all coherent. Hence, we will focus on these sources in this section.

The first high power cw sources in the THz regime were alcohol lasers, running at distinct laser lines. The most common lasermedium is CH_3OH delivering radiation with a frequency of 2.54 THz. They can have output powers of $\geq$ 100 mW [101] and have even been utilized for THz near-field microscopy [15]. However, the cavity of such a laser must be several meters large, the pump power is usually supplied by a very strong CO_2 laser and hence a heavy duty and precise cooling system must be installed to guarantee stable laser performance. Overall, this makes the use of such a laser very impractical.

Other narrow-band cw sources are quantum cascade laser (QCLs). They can in general be designed to a desired wavelength and newest designs are quite tuneable [102]. There are more and more companies offering turn-key solutions, however the devices are expensive. The output power ranges around 1 mW. Also, cryogenic temperatures are needed to operate the systems. With state-of-the-art closed-cycle cooling systems, this has become less challenging in the last years. In s-SNOM QCL with a frequency of 3 THz have successfully been applied [103] as well as mid-infrared QCLs [104]. However, the regime from 3 to 10 THz is still not covered.

A very new cw source for THz scientists are high-power Schottky diodes. With frequency upmixing they now reach almost one THz and have even been applied for s-SNOM [48]. The are commercially available with powers of up to a few ten mW [105]. However, the upconversion efficiency of the diodes will decrease further for higher frequencies and probably never reach the range of few THz.

While the sources mentioned above are narrow-band emitters, there are also pulsed and broadband emitters available. A quite common concept is the utilization of short and broadband visible or infrared laser pulses for optical rectification. The fs pulse is shot on a crystal which shows a significant non-linearity in intense electric fields. Most commonly, ZnTe is used for this application. Here, the non-linear dependence of the polarisation to the electric field leads to difference frequency generation of all spectral components included in the laser pulse. This creates a ps THz pulse with a spectrum ranging from few ten GHz up to roughly 5 THz. The disadvantage is that even for small THz powers of few ten μW, the crystal is already at the breakdown input power due to very low conversion efficiency. Recently, lithium niobate has been utilized for pulsed THz emission. Here, the breakdown of the crystal can be circumvented through strong dispersion at visible wavelengths. As many scientists are interested in large THz fields, this concept is mostly used with low repetition-rate laser sources to create intense THz pulses. Due to the low repetition rate, it has not been applied with s-SNOM yet. Both types of sources are explained in more detail in [106].

In this work both broadband and narrow-band THz sources were applied. The working principles are presented in detail in the follwing.

The broadband source is a **photoconductive antenna** (PCA). The concept is depicted in figure 2.4.1. Two electrodes are seperated by a semiconductor[8]. Between the electrodes, a DC voltage is applied. In the ground state, almost no current is running in between the electrodes. Electrons are now excited into the conduction band by a fs laser pulse (red) and leave holes in the valence band. The charge carriers are separated by the applied voltage and induce a current flow rising and then falling again in between the electrodes. This process lasts for a short period of time and is limited by the charge carrier lifetime[9].

Moving charges will always induce an electromagnetic field following $E_{ind} \propto dI(t)/dt$. As a result, broadband THz radiation (blue) is emitted from the device. For this work, a large-area device combining many of such antennas was used[10]. It has a spectrum from roughly 200 GHz to 5 THz and a peak emission frequency of 1-1.5 THz. The output power reaches up to 1 mW [107]. As fs lasers are used to excite the antenna, it is a suitable source to be combined with other time-resolved techniques and can be detected via time-domain spectroscopy (TDS, see following section). The emission spectrum of a PCA is quite similar to the spectrum of a rectification emitter. However, the efficiency of rectified emission is very low and cannot be improved further due to

[8] most commonly low-temperature grown GaAs
[9] for GaAs typically in the range of few ns
[10] Tera-SED 10, LASER QUANTUM LTD

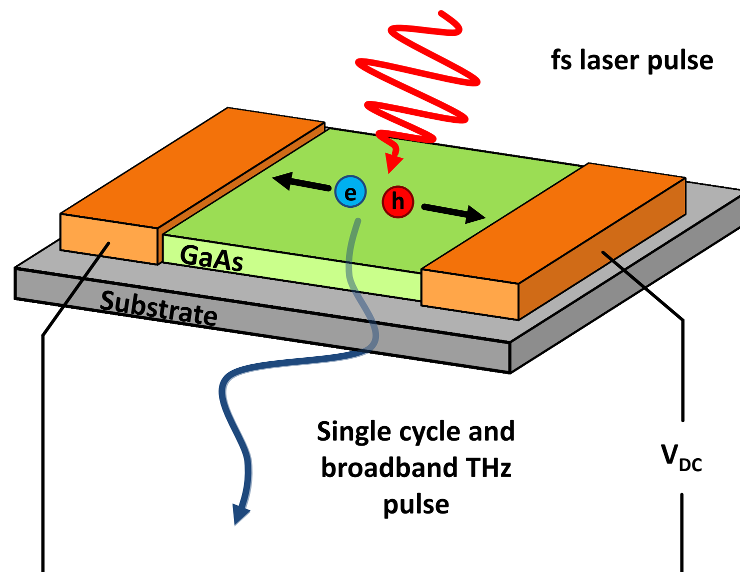

Figure 2.4.1.: *Schematic drawing of a PCA. On a substrate material, two electrodes (orange) and a suitable semiconductor (green) in between them are grown. When a fs laser pulse hits the semiconductor, charge carriers are separated and transferred between the electrodes, inducing an electromagnetic field propagating out of the device.*

the breakdown limit of the crystals when the excitation laser is focused in the crystal. PCAs are already at least as strong as rectified emitters, but still under development and become more efficient every year. Although PCAs are active devices as they need to be driven with a power supply, they will very likely be the broadband emitter of choice for many applications in the future.

The narrow-band THz source applied in this work is the **free-electron laser FELBE** located at the Helmholtz-Center Dresden-Rossendorf. Free-electron lasers (FEL) deliver intense and narrow-band radiation and are usually tuneable over a wide spectral range. FEL theory is complex, but can be managed. However, the technical implementation is costly and challenging, the operating costs are also high due to large power consumption (e. g. by LHe cooling). Hence only few of this large facility lasers are available. To the best of our knowledge, the s-SNOM utilized in this work is the only near-field microscope at such a beamline and can be used in the wavelength range of $\lambda =$5 - 230 μm. In the following, the FEL working principle will be explained.

A schematic picture of an FEL is shown in figure 2.4.2. A relativistic electron beam is guided into an array of alternating magnets. Here, the electrons experience an undulating movement around their propagation direction. On each curve, they emit electromagnetic radiation as they are accelerated radially. This kind of radiation is called synchrotron radiation or also magneto-bremsstrahlung. The wavelength of the emitted radiation is on the one hand determined by the distance of the undulator's magnet and their magnetic field strength. On the other hand, the electron energy is an important and tunable parameter. Additionally, one needs to consider relativistic effects, namely the Lorentz contraction of the undulator as viewed from the electrons coordinate system and secondly the change of the wavelength for an observer outside the FEL due to the relativistic Doppler effect. The overall wavelength is given by [109]

$$\lambda = \frac{\lambda_u}{2\gamma^2}\left(1+\frac{K^2}{2}\right) \quad \text{with} \quad \gamma = \frac{1}{\sqrt{1-v^2/c^2}} \quad \text{and} \quad K = \frac{eB_u\lambda_u}{2\pi m_e c}. \tag{2.4.1}$$

B_u is the magnetic field of the undulator and λ_u the distance of two equally poled magnets.

At FELBE, the accelerator delivers electron bunches with a repetition rate of $f_{rep} =$13 MHz with electron energies of up to 40 MeV. The cavity length is matched to a value that a round trip of the light matches this frequency and for every cycle new electrons are injected into the undulator. The electron bunches are longer than the wavelength λ_u of the undulator. All electrons in general emit at a random phase and the radiation is incoherent. Coherent emission is achieved by so-called micro-bunching. Here, the electromagnetic field in the cavity interacts with the electrons, decreasing or increasing their energy depending on their phase compared to the electromagnetic

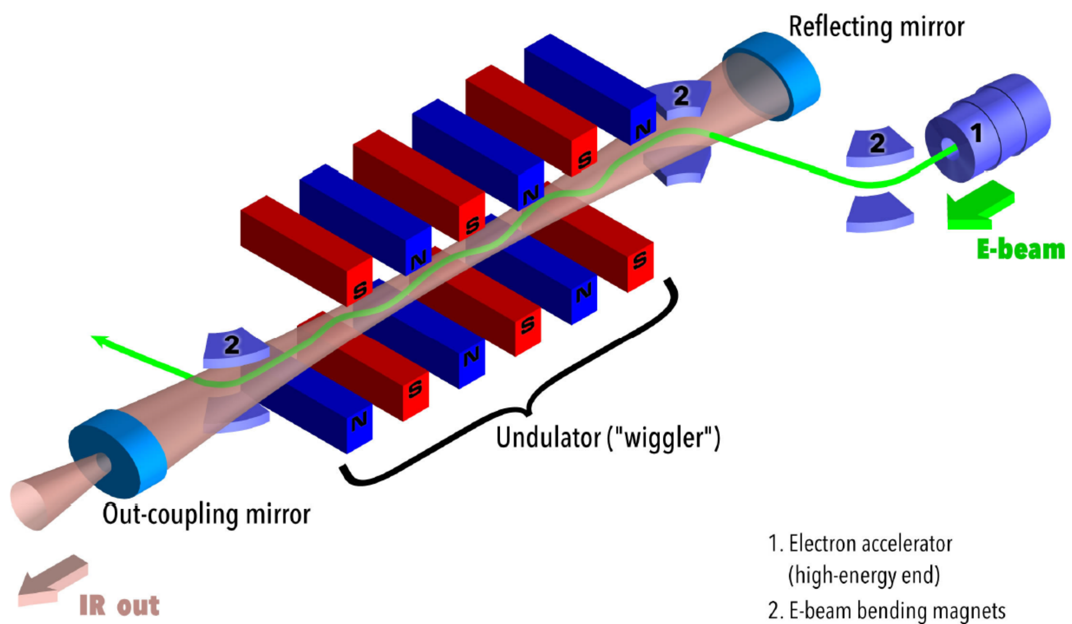

Figure 2.4.2.: *Schematic drawing of a free-electron laser taken from [108]. An electron accelerator (1) delivers electron bunches which are guided by dipolar magnets (2) into the cavitity. The electrons' movement in the cavity is influence by alternating static magnetic fields and they emit radiation at each wiggle until they reach the end of the undulator and are directed towards a beam dump. The emitted radiation is stored in the laser cavity. In contrast to a usual laser systems, the outcoupling is established via a hole instead of a partially transmitting laser mirror. The reason is the tuneability of the laser and the lack of materials which are partially transmitting over the large spectral range of the FEL.*

field. The randomly distributed electrons become concentrated in microbunches. Their distance matches the wavelength of the emitted electromagnetic field and coherence is achieved. Interested readers may refer to [110] for a short theoretical essay of free-electron lasers, [109] for a much more detailed discussion and [108] for a review of state-of-the-art technologies and their future applications.

Free-electron lasers in general can be used to create strong coherent radiation from hard X-rays to millimeter waves. A list with many parameters of the roughly 50 facilities worldwide can be found in [111]. The free-electron laser FELBE delivers radiation from 5-230 μm wavelength with two different undulators and delivers powers of several watts at any wavelength. Depending of the settings, more than 100 watts of power can be reached [112]. In contrast to most other facilities, FELBE delivers pulses with a repetition rate high enough for the usage in s-SNOM. In this work, the U100 undulator of FELBE was used with 2 THz ($\widehat{=}$ 150 μm) radiation.

THz detection

In general, THz detection concepts can be classified in two categories. On the one hand, detection systems which detect intensities $I \propto E^2$ of the THz field and can detect incoherent radiation are e. g. applied in FTIR spectroscopy. They can directly detect any THz radiation and thus also pick up any random background noise, like blackbody radiation from the surroundings. On the other hand, coherent detection systems are more complex in their application and can only be applied for pulsed and coherent detection. However, as they directly detect E and usually integrate the signals over a certain amount of time, random background fluctuations are averaged to zero and efficiently suppressed. In this work, both coherent and incoherent detection systems were applied. Their working principles are explained below.

The most prominent **incoherent detection** systems for THz radiation are photodetectors with shallow impurities such as gallium-doped germanium and pyroelectric detectors like bolometers [113]. Both achieve best sensitivity and signal-to-noise ratio when they are cooled to cryogenic temperatures. Assuming the electrons' distribution in the detector to be given by the Maxwell-Boltzmann statistic, 2 THz radiation corresponds to a temperature of $\sim$ 95 K[11].

In this work a LHe-cooled hot-electron bolometer[12] was used to detect 2 THz radiation from the free-electron laser FELBE. Here, the InSb sensor is cooled to a temperature of 4.2 K. The electrons are only weakly coupled to the crystal lattice and interaction with phonons is minimized.

[11] 2 THz $\widehat{=}$ 8.2 meV, k_B = 8,62$\cdot 10^{-2}$ meV/K $\Rightarrow$ T = 95 K.
[12] QMC Instruments QFI/XBI

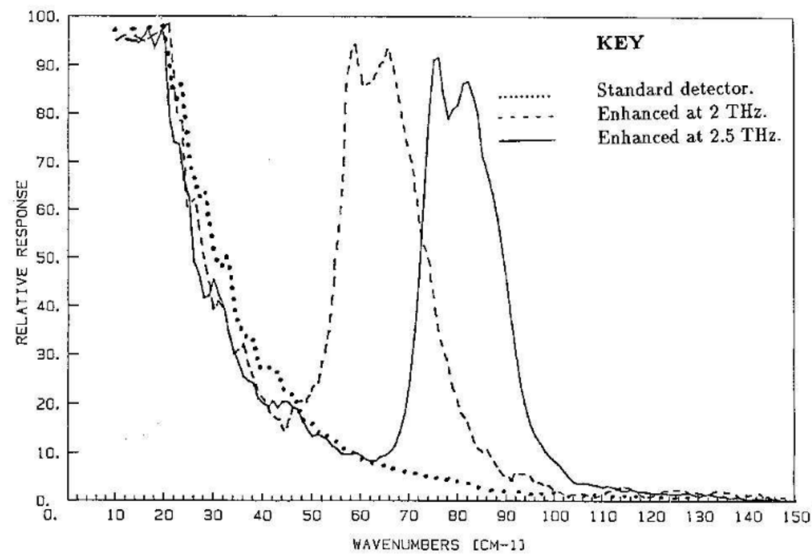

Figure 2.4.3.: *Relative response of standard and magnetically tuned InSb hot-electron bolometers. Figure taken from [114]. The InSb applied in this work is tuned to roughly 2 THz (dahed line).*

Whenever photons are absorbed by the material, the temperature rises and the electron mobility $\mu_e \propto T^{3/2}$ changes [114]. The low effective mass m_e^* of the charge carriers allow for fast detection of the changes. In InSb bolometers, the effective relaxation time of the system is roughly 0.3 μs and the bandwidth (-3 dB) is stated to be 1 MHz. For a typical s-SNOM, this should be the lower bandwidth limit of the detector. During this work, signals with even higher frequencies of 1.6 MHz were detected.

Standard InSb bolometers can be used to detect radiation with frequencies below 1.5 THz. The key to use such a bolometer for higher frequencies is to exploit the cyclotron resonance of the electrons in the crystal. As already mentioned above, this resonance often lies in the THz range. Here, a permanent magnetic field B is used to tune the cyclotron resonance to the desired frequency $\omega_c = eB/m_e^*c$. As this resonance is accompanied by strong absorption, the highest sensitivity of the detector can be deliberately tuned to the desired frequency. Typical sensitivity spectra of such a detector are shown in figure 2.4.3. Here, the response peak is only a few hundred GHz wide. By using an inhomogeneous magnetic field, the resonance peak can be broadened as the effective cyclotron resonance is spatially varying within the InSb crystal. However, a broader peak corresponds to lower sensitivity at a fixed frequency and the tuning should be chosen wisely for each application.

For **coherent detection**, either photoconductive antennas or electro-optic sampling can be ap-

plied. In photoconductive antennas the detection principle is as follows: No DC voltage is applied when the fs laser pulse creates free charge carriers in the semiconductor. If a THz pulse hits the antenna at the same time, the carriers are transported to the electrodes due to the presence of the THz field and create a net current which can be measured. Although the alignment is relatively simple, the time-resolution of this technique is limited by the carrier lifetime and are in this sense inferior to electro-optic sampling, where the time-resolution is only limited by the laser pulse length.

Electro-optic sampling is a highly efficient and low-noise detection method for pulsed and synchronized THz radiation. As it measures the THz field in the time-domain it is often called time-domain spectroscopy or TDS. The term spectroscopy applies as the pulse information can be fourier-transformed into the frequency-domain directly allowing for spectral investigation. A typical setup is shown in figure 2.4.4. A focused THz pulse and a linearly polarized fs laser pulse are focused onto an electro-optic crystal[13]. If there is no temporal overlap of the beams (a), the fs laser pulse first passes the electro-optic crystal and then a $\lambda/4$ waveplate, where its polarization is changed to circular. The two polarization components are now equally split up by a Wollaston prism (WP) and detected by two highly sensitive photodiodes. If the laser pulse and the THz pulse have a temporal overlap at the electro-optic crystal (b), the linear polarization of the laser pulse is rotated via the Pockels effect induced by the THz pulse. Please note that the Pockels effect is sensitive to the THz electric field and not to its intensity. After passing the $\lambda/4$ waveplate, the beam has elliptical polarization and is unequally split up by the Wollaston prism. The intensity change at the photodiodes can be measured with high sensitivity and precision by differential amplifiers and hence leads to the most sensitive detector for THz radiation. As the measurement is usually averaged over thousands or millions of pulses, random background fluctuations of the electric field are simply averaged out. In comparison, incoherent detectors will show random background field fluctuations as a noise floor. If the temporal delay of the fs laser and the THz is changed via a delay line, the amplitude and phase of the THz pulse can be measured with a temporal resolution limited only by the pulse length of the fs laser. For this reason, this is often called sub-cycle resolution, as the resolution is shorter than the THz pulse. Full amplitude and phase information makes the coherent detection inherently superior to incoherent detection, as phase information is lost unless it is combined with a phase-sensitive detection method like interferometry. Whenever possible, coherent detection should be applied.

[13]There are other devices such as PCAs which can be applied as well, however electro-optic crystals were used in this work and the theoretical description is regarding this method only.

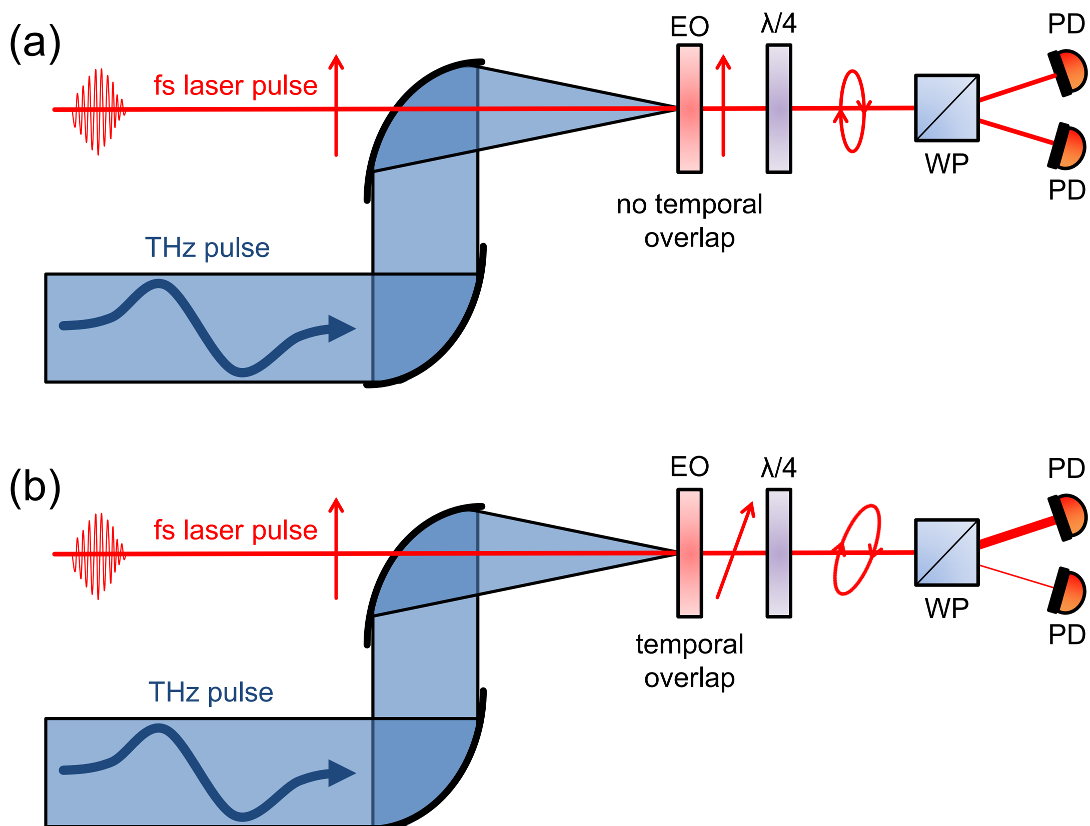

Figure 2.4.4.: *Electro-optic sampling setup. (a) shows the principle with no temporal overlap of the THz (blue) and fs laser (red) pulse. The fs laser is equally split at the photodiodes (PDs). With temporal overlap (b), the polarisation of the laser pulsed is turned at the electro-optic crystal (EO) and the beam is unequally split up to the PDs.*

Electro-optic sampling can in general be applied for any combination of synchronized THz pulse and fs laser. The only requirement is a detection crystal which is transparent and show the Pockels effect for both the THz radiation and fs laser. It can be applied for mid-infrared [68] and for low THz radiation [56]. To achieve a larger signal-to-noise ration, a chopper is usually placed into the beam path of the THz pulse and the signal at the photodiodes is modulated with the chopping frequency. Using a Lock-In amplifier demodulating the signal using the chopping frequency as the reference frequency, it reaches a dynamic range of more than 90 dB [115] and is hence the most sensitive THz method available.

THz s-SNOM is a field of research which is still under exploration. It has been applied with narrow band radiation from cw sources [15, 116] and with the tuneable free-electron laser FELBE [7] for mapping charge carrier densities. In a recent report, a fully electronic THz source has been applied for the very low THz range [48]. Another track is the utilisation of broadband THz radiation and detection in a TDS s-SNOM system. Here, first reports are from 2008 [54], while quality has been improved drastically in the last years [55, 56]. Recently, a TDS s-SNOM has been applied to study the properties of graphene [89]. Now, even commercial systems are available [117]. However, there is still much open room for exploration.

3. SNOM-Theory

Alles, was noch nicht gewesen ist, ist Zukunft, wenn es nicht gerade jetzt ist.

(Dr. Angela Merkel)

This chapter provides a discussion on the theoretical background of scattering-type scanning near-field microscopy. As the focus of this thesis is on experimental work, the fundamental theory introduced by Knoll and Keilmann is explained briefly. Although this analytic description of a model scattering-type near-field microscope is rather simple, it still predicts the experimental behaviour quite well. The experimental detection of the predicted signatures with Lock-In amplifiers at higher harmonics is explained as well. Moreover, several experimental methods for phase-resolved detection are presented. Furthermore, the model is expanded for systems with non-constant sample properties in section 3.3. Here, amplitude-modulation leads to the creation of sidebands in the s-SNOM signal which can be detected with enhanced signal-to-noise ratio. As a last point, more realistic tip geometries are investigated with numerical methods to explain the divergence of the dipole model to experimental observations.

3.1. Dipole Model

The most common analytic description of a scattering-type near-field microscope was introduced by Berhard Knoll and Fritz Keilmann in their seminal paper in 2000 [118]. Although the model is analytic and relatively simple, in many cases the predictions of this approach match quite well to the experimental results and can even be used for quantitative analysis of complex permittivity at the local scale [52]. This section is summarizing Knoll's and Keilmann's model as it is needed for further discussions in this chapter and for the interpretation of the experimental results.

Firstly, they propose a sphere much smaller than λ as an approximation for the scattering tip,

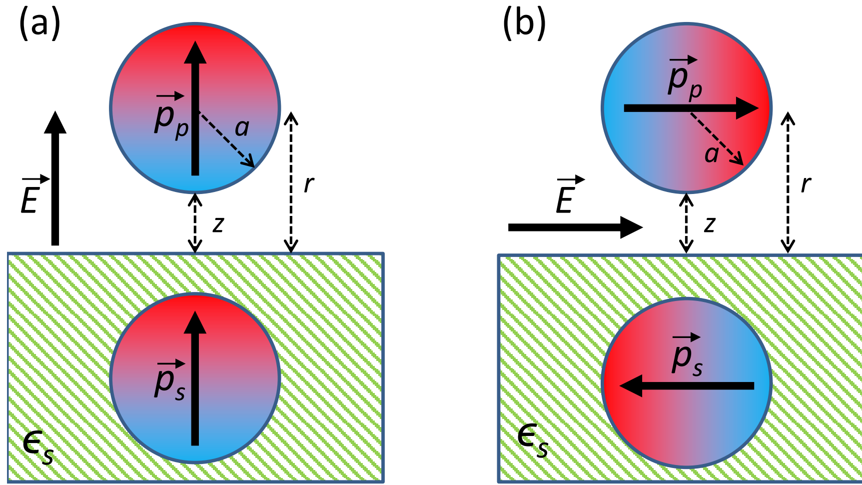

Figure 3.1.1.: *Schematic sketch of the dipole model. (a) shows the case for the electric field E direction perpendicular to the surface. The sphere with the radius a and the distance z to the sample surface has a dipole moment $\vec{p}_p$ (red and blue color) in the same direction. The induced mirror dipole has the same direction for $|\varepsilon_s| \leq 1$, but not necessarily the same size. In the case of field direction parallel to the surface (b), the dipole moments show in opposite direction.*

which is made of an isotropic material with a radius $a \ll \lambda$ and the polarizability α. In textbooks on classical electrodynamics (e. g. [119]), the polarizability of such a sphere in vacuum is given by

$$\alpha = 4\pi a^3 \cdot \left(\frac{\epsilon_p - 1}{\varepsilon_p + 2} \right). \tag{3.1.1}$$

Here, ε_p is the complex dielectric constant of the sphere, labeled with $_p$ as it is the probe material in the s-SNOM. Whenever there is an electric field present at the sphere, a dipole moment $\vec{p}_p$ will be induced

$$\vec{p}_p = \alpha \vec{E}. \tag{3.1.2}$$

Secondly, they assume the center of the sphere is located at a distance r from the sample filling the lower halfspace and, firstly, consider the case of $\vec{E}$ perpendicular to the surface (see figure 3.1.1 (a)). The dipole of the sphere has a field at the sample surface underneath the tip given by $\vec{E}_{\text{surf}} = \vec{p}/(2\pi r^3)$, leading to surface charges which will again induce an electric field in the upper halfspace. This effect is well known and is best described by an image dipole $\vec{p}_s$ which is located

in the sample at a distance of r below the surface. This dipole moment is labeled with $_s$ for the sample and is given by

$$\vec{p}_s = \beta \cdot \vec{p}_p = \left(\frac{\varepsilon_s - 1}{\varepsilon_s + 1}\right) \cdot \vec{p}_p. \tag{3.1.3}$$

The image dipole's an electric field alters the total field at the position of sphere. Hence, the dipole moment p_p will be larger as well and is now given by

$$\vec{p}_p = \alpha \left(\vec{E} + \frac{\vec{p}_s}{16\pi r^3}\right) = \left(\frac{\alpha}{1 - \frac{\alpha\beta}{16\pi r^3}}\right) \cdot E. \tag{3.1.4}$$

The whole system now exhibits a total field induced by both the sphere's and the image dipole. At large distances, it can be describe as a single dipole with the effective polarizability

$$\alpha_{\text{eff}} = \frac{\alpha(1+\beta)}{1 - \frac{\alpha\beta}{16\pi r^3}}. \tag{3.1.5}$$

for the perpendicular case.

For the case of the incident field being polarized parallel to the sample surface, two changes in the calculation must be considered. Firstly, the tip dipole's field at the sample surface is given by $\vec{E}_{\text{surf}} = -\vec{p}/(4\pi r^3)$ and the image dipole moment is then given by $\vec{p}_s = -\beta\vec{p}_p$ as depicted in figure 3.1.1 (b). Following the same path as before, the effective polarizability reads

$$\alpha_{\text{eff}} = \frac{\alpha(1-\beta)}{1 - \frac{\alpha\beta}{32\pi r^3}} \tag{3.1.6}$$

for the parallel case.

Please note that all thoughts to this point neglect phase retardation in between the two dipoles and is thus often called the quasi-electrostatic solution to the problem. To find the scattering cross-section of the system, the Mie-theory for scattering of electromagnetic waves on small particles is applied. In the Rayleigh-limit ($a \ll \lambda$) the cross-sections for scattering and absorption are given by

$$C_{\text{sca}} = \frac{k^4}{6\pi}|\alpha_{\text{eff}}|^2 \qquad C_{\text{abs}} = k\text{Im}\{\alpha_{\text{eff}}\}, \tag{3.1.7}$$

where $k = 2\pi/\lambda$ is the wavevector. The scattering cross-section as a function of r is depicted in figure 3.1.2 for a laser wavelength of 10.6 μm. The sphere and sample are made of platinum

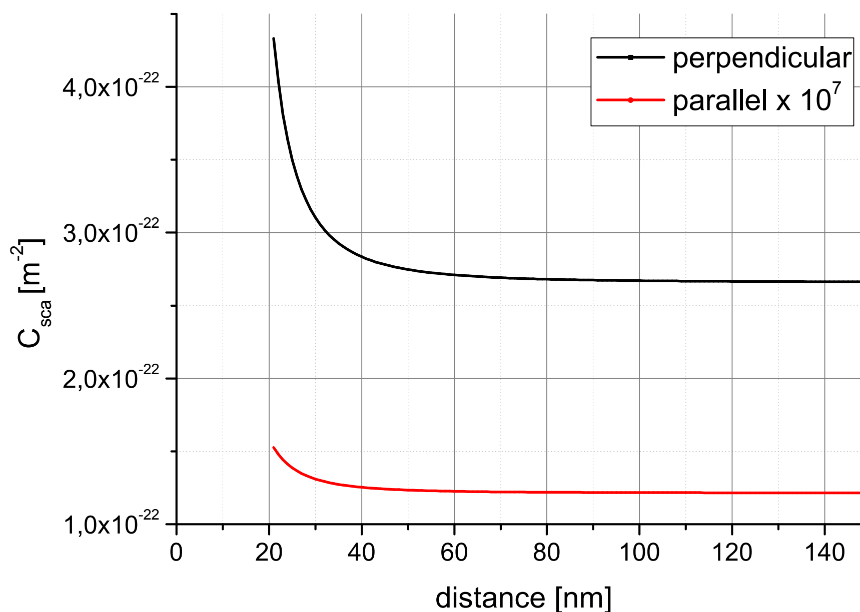

Figure 3.1.2.: *Scattering cross-section of a small sphere ($a = 20$ nm) with $\varepsilon_p = -1450$ and $\varepsilon_s = -4679$ (both taken from [120]). The wavelength is 10.6 μm. Please note that values with a distance below a are neglected as the sphere touches the surface and the cross-section has unphysically large values.*

and gold, respectively. The typical non-linear distance-dependence for small distances is visible for both the perpendicular and parallel case of the calculation. At a distance of a few ten nanometers, the near-field signal completely vanishes. The cross-section in the perpendicular case is more than a factor of 10^7 larger than in the parallel case. The reason for this is that the sample response function $\beta = 1.0004$ is close to 1 and the effective polarizability becomes tiny. Figuratively speaking, the image dipole $\vec{p}_s$ in figure 3.1.1 (b) has the exact same dipole moment $\vec{p}_p$ and thus they cancel each other out. The overall scattering cross-section is very small and also the contrast of the signal close to the surface and far away from it is small. The next section will illustrate the experimental way to find such small signals.

To this point, the shape of the tip has been approximated by a sphere. To incorporate the tip's elongated shape in the theory, the dipole model has been expanded to the so-called finite dipole model introduced by Cvitkovic et al. [121], where the tip is approximated by a spheroid with the long axis perpendicular to the sample surface. However, the model is restricted to lengths of the spheroid of $L \leq \lambda/4$ and is not suitable for large tip sizes as supplied by Rocky Mountain Nanotechnology (see section A.2). To investigate these tips' behavior in more detail on a theoretical

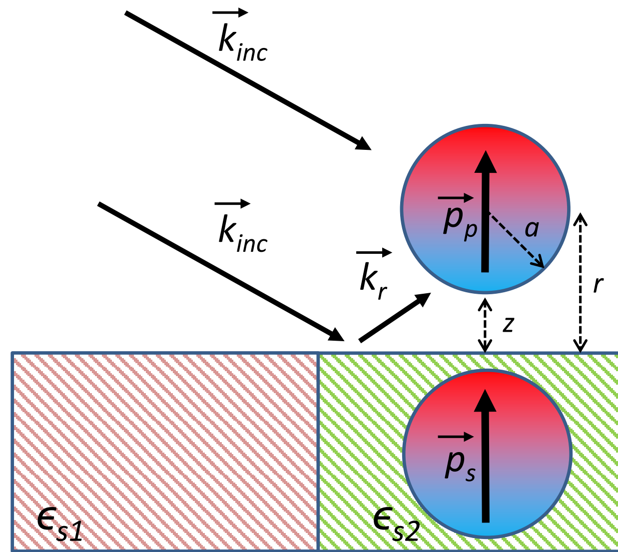

Figure 3.1.3.: *Modification of the dipole model in [121] for the perpendicular case. The incident light does not only illuminate the sphere, but also the reflected light is also taken into account. When the sphere is located on different materials the amount of reflected light can change strongly. As the sphere is located close to the sample surface, the reflections illuminating it are located very close by and hence do not annihilate the sub-diffraction resolution.*

level, a numerical simulation is a suitable choice and performed in section 3.4.

In her model, Cvitkovic also takes the sample reflectivity into account, as it has been discussed by others, e. g. [122]. Experimentally it is impossible to illuminate the probe without illuminating the sample beyond the tip. The reflected beam will contribute to the effective field at the probe (see figure 3.1.3). When scanning different sample areas, the amount of reflected light needs to be taken into account via the Fresnel coefficients of the investigated sample. Here, the effective probe polarizability is given by

$$\alpha_{\text{eff},r_n} = \frac{\alpha(1 + r_n)^2}{1 - \frac{\alpha\beta}{16\pi r^3}} \tag{3.1.8}$$

for the perpendicular case, where r_n is the Fresnel coefficient of the sample. The same arguments apply for the parallel case. Note that the sub-diffraction resolution is not compromised by this reflection as the tip is only a few ten nanometers above the sample surface in an s-SNOM

experiment. Hence, the reflection contributing to the effective field at the sphere is laterally only few ten nanometers apart from the sphere.

3.2. Detection and Demodulation

As discussed above, scattering cross-sections in near-field microscopy are small, compared to far-field experiments they are tiny. This fact is an obstacle on its own. Additionally, most of the illuminating radiation is not even contributing to the near-field interaction at all. Whenever a light, infrared or THz source is utilized to illuminate a tip for a near-field experiment, its focus will be much larger than the tip apex and hence create a large area of intense radiation around the tip. The area at the tip end which actually contributes to the near-field interaction is small. This problem worsens for larger wavelengths, as the focus size increases due to diffraction and the radiation density at the tip end decreases. Ultimately, a much larger amount of far-field radiation is scattered directly from larger structures (e. g. the tip shank and large structures on the sample) and eventually reach the detector than the scattered radiation from the near-field interaction.

In order to deduce the tiny near-field signal from these far-field contributions, the most commonly used technique is called higher-harmonic demodulation [118, 123]. The idea is to use the short interaction lengths of the near-field as an advantage for probing these small signals. By periodically changing the tip-sample distance $z = r - a$, the tip exhibits stronger and weaker near-field coupling during each oscillation, hence the scattered signal amplitude is modulated as well. These small changes with a distinct frequency can be detected with Lock-In amplifiers with enormous sensitivity, gain and noise reduction. Signals in the nV range can be detected with a dynamic reserve of 100 dB, hence the near-field signal can be extracted from the much larger background noise. The principles and state-of-the art of Lock-In amplifiers applied in this work are well explained in [124] and will not be discussed in detail here.

Let us now assume the tip oscillates with an amplitude of z_{osc} and a frequency Ω close above the sample surface. The distance can then be expressed by

$$z(t) = z_0 + z_{\text{osc}} \cdot \Big(1 + \cos(\Omega t)\Big). \tag{3.2.1}$$

Consequently, $r(t) = z(t) + a$ and thus $\alpha_{\text{eff}}(t)$ are now time-dependend.

The overall scattered field reaching the detector is now consisting of near-field E_{nf}, field scattered at the tip shank E_{ts} and background E_{bg} far-field radiation, while all field scale linearly with the

size of the incoming field E_{inc}. The background field $E_b g$ does not depend on the oscillation frequency of the tip, as it is scattered only at parts in the setup which are located at fixed positions. The field can be it can now be expressed by a Fourier series

$$E_{sca}(z) = E_{nf}(z(t)) + E_{ts}(z(t)) + E_{bg} = E_0 + E_1 \cos(\Omega t) + E_2 \cos(2\Omega t) + ... \quad (3.2.2)$$

where E_0 is by far the largest field and does not depend on the time. As the detector measures only intensities, the interference of these fields leads to

$$\begin{aligned} I_{\mathrm{det}} \propto\ & const. \\ & + \Big(E_1 E_2 + 2E_0 E_1\Big) \cos(\Omega t) + \Big(0.5E_1^2 + 2E_0 E_2\Big) \cos(2\Omega t) \\ & + \text{higher harmonics.} \end{aligned} \quad (3.2.3)$$

It is feasible to assume that only mixed terms with E_0 are detected, as all other terms are much smaller and can be neglected. As E_0 is assumed to be constant, the change in signal strength on the detector is originating from the change in electric field E_{sca}. For the field E_{ts} scattered from the tip shank, the change is linear with z and thus only contributing to E_1. For the near-field contribution E_{nf}, the distance dependence is strongly non-linear (see equation 3.1.5) and hence contributes to E_1, E_2 and all higher harmonics. Accordingly, the detector signal demodulated at higher harmonics of Ω leads to the near-field information.

This technique is schematically shown in figure 3.2.1. The near-field microscope is illustrated in (a). The tip end, shank and the sample are illuminated by a diffraction-limited focused light source. The tip is oscillating close to the sample surface with a frequency Ω. Light which is scattered from the tip shank does not exhibit near-field interactions and therefore has a linear dependence on the tip-sample distance z (b). As discussed in chapter 3.1, the near-field signal (c) is strongly non-linear with z and mostly contributes on the first few nanometers above the sample surface. The lower row of the figure schematically shows the Fourier coefficients in the frequency domain. The far-field (d) only has a contribution at the tip oscillation frequency Ω. The near-field however has components at multiples of Ω as well (e). When these higher harmonics are detected with a Lock-In amplifier, the deduction of the near-field information is possible.

The detection method presented above is often called direct detection or self-homodyne detection mode as it utilizes the background field E_{bg} to amplify the small near-field contributions. It has the obvious disadvantage that the size of the background field cannot be determined easily

and the amplification is unknown. Additionally, the phase relation in between background field and near field are unknown. In order to have a well-determined amplification and phase relation, s-SNOM can be combined with interferometric techniques. Similar to macroscopic experiments, the disentanglement of these two optical properties can deliver important information about the characteristics of a material, i.e. the real and imaginary part of the dielectric constant. In the following, various techniques to deduce the amplitude and phase-information separately on the nanoscale will be explained, namely interferometric methods and time-domain spectroscopy as applied in this work.

In far-field optics, many of such techniques and concepts have been developed and applied for more than a century. The most famous and widely applied one is surely the Michelson interferometer, which takes the scattered signal from an object and correlates it to a defined reference beam. When the length of the reference beam is changed, the phase can be deduced from the data. Recently, the technique gained public interest again when it was applied with outstanding engineering expertise in the LIGO collaboration [125]. With their advanced Michelson interferometer, the existence of gravitational waves was proven in 2015. Their effort was awarded with a Nobel Prize in physics just two years later.

The combination of a Michelson interferometer with an s-SNOM is a well-known concept [49, 51, 126] and schematically shown in figure 3.2.2. It allows for the disentaglement of the amplitude and phase information of the scattered near-field signal. The laser source is divided in two arms by a beamsplitter. One arm illuminates the near-field microscope, the other arm is reflected at a mirror. The scattered signal from the s-SNOM and the reference beam are aligned to a detector and interfere. To change the phase ϕ_{ref} of the reference beam and near-field, the reference mirror can be moved along its optical axis. This setup is usually called **homodyne interferometric detection** in s-SNOM.

In a homodyne interferometer, the detected intensity at the is given by

$$I_{\text{det}}(z) = |E_{sca}(z) + E_{ref}|^2 = |E_{sca}(z)|^2 + |E_{ref}|^2 + 2|E_{sca}(z)||E_{ref}| \cos(\phi_{hom}) \tag{3.2.4}$$

where $E_{sca}(z)$ is the overall field scattered from the microscope (equation 3.2.2), E_{ref} is the field from the reference arm of the interferometer and $\phi_{hom} = \phi_{ref} - \phi_{sca}$ is the optical phase difference at the detector. When higher-harmonic demodulation is applied, similar assumptions as in equation 3.2.3 can be made. $|E_{ref}|^2$ does not depend on the tip-sample distance and will not be detected via Lock-In demodulation. $|E_{sca}(z)|$ is small compared to $|E_{ref}|$, thus only the mixed term $2|E_{sca}(z)||E_{ref}| \cos(\phi)$ will be detected. Here, similar to the self-homodyne case discussed

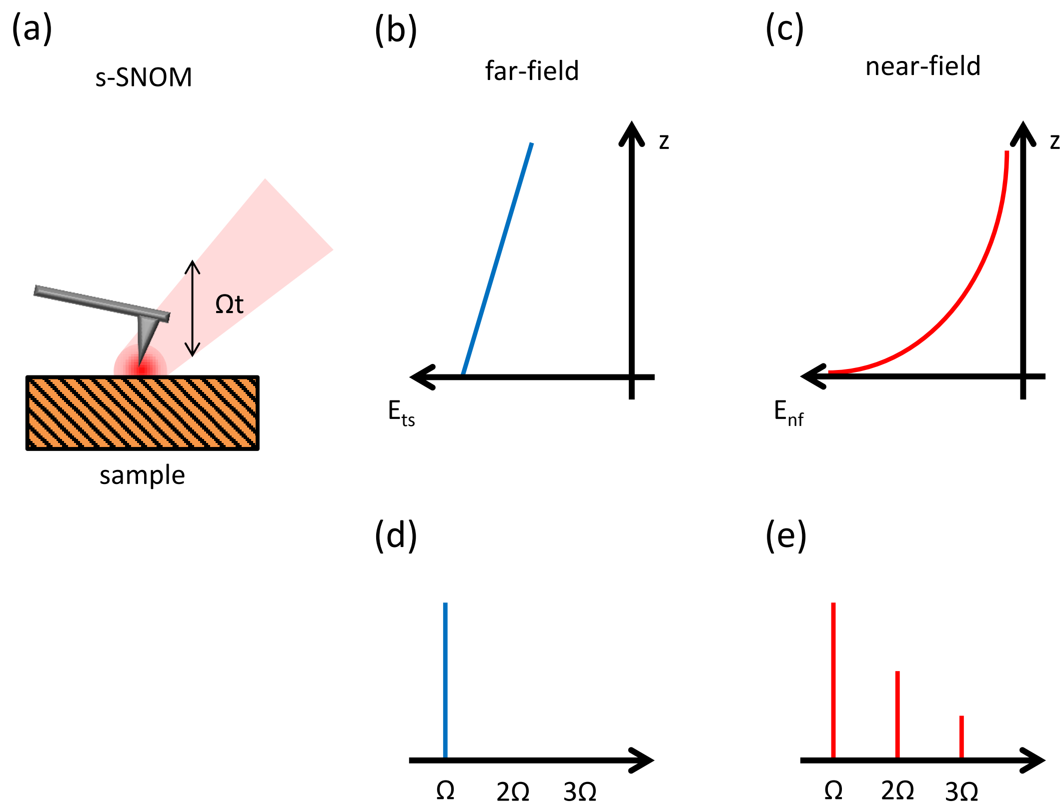

Figure 3.2.1.: *Schematic illustration of higher-harmonic-demodulation: (a) the near-field microscope is working in tapping or non-contact mode with an oscillation frequency* Ω*. (b) Far-field reflections, e. g. from the tip shank have a linear dependence with the distance* z*. Hence, the spectrum of the detected signal (d) has only one Fourier coefficient at* Ω*. The near-field (c), however, is strongly non-linear with the distance* z *and consequently also carries higher harmonics at multiples of* Ω*.*

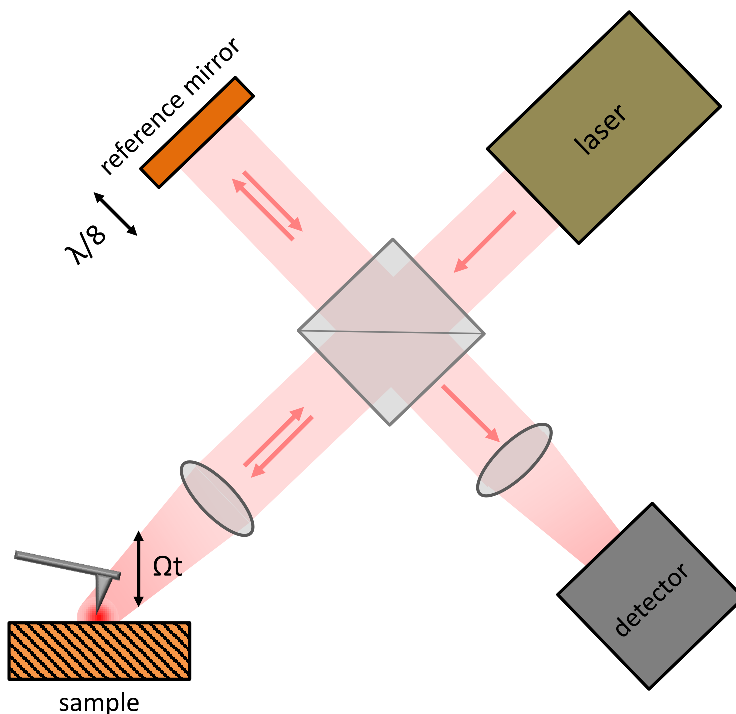

Figure 3.2.2.: *Homodyne detection setup. A Michelson interferometer is combined with s-SNOM. A laser beam is divided by a beamsplitter. One half reflected by the reference mirror, the other half is used to illuminate the s-SNOM. The backscattered near-field signal and the reference beam are interfering at the detector.*

above, higher-harmonic demodulation will extract the near-field information for 2Ω and higher.

The advantages of this method are the are firstly the multiplication of the near-field with a well defined reference. This leads to a known, constant, controlloable and ultimately much larger signal amplification. Secondly, the phase information is encoded in the signal. By changing the reference arm length, the near-field amplitude A and phase ϕ can then be obtained. In general, two measurements with different reference phase settings are sufficient to correctly extract this information. Let us assume that two measurements of the scattered intensity I_1 and I_2 with a

reference phase shift of $\pi/2$ are performed. The measured intensities are then

$$\begin{aligned} I_1 &= 2|E_{sca}(z)||E_{ref}|\cos(\phi_{ref} - \phi_{sca}), \\ I_2 &= 2|E_{sca}(z)||E_{ref}|\sin(\phi_{ref} - \phi_{sca}), \\ \rightarrow A^2 &= I_1^2 + I_2^2 = 4|E_{sca}(z)|^2|E_{ref}|^2, \\ A &= \sqrt{I_1^2 + I_2^2} = 2|E_{sca}(z)| \cdot |E_{ref}|. \end{aligned} \tag{3.2.5}$$

Please note that higher-harmonic demodulation will again suppress far-field contributions to the signal and A is directly proportional to the near-field amplitude. To determine the phase, simple calculations have to be performed:

$$\begin{aligned} I_2/I_1 &= \sin(\phi_{ref} - \phi_{sca})/\cos(\phi_{ref} - \phi_{sca}) \\ &= \tan(\phi_{ref} - \phi_{sca}) = \tan(\phi_{hom}), \\ \rightarrow \phi_{hom} &= \arctan(I_2/I_1). \end{aligned} \tag{3.2.6}$$

Higher-harmonic demodulation will suppress the influence of any far-field phase from E_{sca} in the experiment and deduce optical phase in between the scattered near-field signal and the reference beam.

Although successfully performed by e. g. Taubner et al. [49], the experimental implementation of homodyne interferometry is challenging as the shift by $\pi/2$, which corresponds to a displacement of the reference mirror by $\lambda/8$, must be exact and constant. Otherwise the equations 3.2.5 and 3.2.6 do not apply anymore. In a typical scanning time of several minutes, thermal drifts will lead to unstable optical path lengths and consequently unstable phases - the reason for the rare application of this technique.

To tackle these problems, two methods can be applied, both using oscillating reference mirrors. Firstly, a p-i control loop can be added to the setup, keeping the measured amplitude at the maximum at one selected harmonic by adding a constant phase to ϕ_{ref} [126]. Here, much smaller and arbitrary oscillation amplitudes can be applied, but the method still affected by drifts. This also halves the measurement time compared to Taubner's method. The second method is the so-called **pseudo-heterodyne detection** [51] which is patented by the company Neaspec. The oscillation amplitude is 0.21 λ, inducing sidebands to the higher harmonics. It is phase-stable and not effected by drifts, can in principle be adapted for any wavelength, but need large oscillations amplitudes of the reference mirror. For large wavelengths, this can be very challenging.

It is mathematically much more complex but nowadays the most commonly applied method in s-SNOM simply due to the fact that the NeaSNOM is the most widely distributed commercial instrument for s-SNOM.

Another approach is the **heterodyne detection** method. Here, the optical frequency of the reference arm is changed typically with a acousto-optical modulators (AOM) [50], leading to a beating signal at the detector which can be used to separate the near-field amplitude and phase. Although the method is very stable against thermal drifts, the lack of acousto-optical modulators for many wavelengths and especially in the THz regime makes this approach unsuitable for many applications and is nowadays rarely used in s-SNOM. The detection scheme has recently been used for THz s-SNOM with a fully electronic radiation source [48].

In the THz region, homodyne and heterodyne detection have successfully been applied [15, 48] with cw sources. Mostly, THz radiation is generated with electro-optic crystals or photo-conductive antennas and hence very broadband. Here, **time-domain spectroscopy (TDS)** is applied to detect the scattered radiation and delivers in situ electric field and phase information of the scattered signals, as described in section 2.4. Typically, the time-domain spectroscopy setup is aligned using the direct reflection of the THz beam from the sample, before the tip is placed at the focus position. When the tip is oscillating above the sample surface, the near-field is modulated and hence the overall field at the electro-optic crystal is modulated as well. To demodulate the near-field signal, the demodulation frequency at the Lock-In amplifier has to be changed from the chopper's freuquency to Ω, 2Ω, $\ldots$ in order to easily extract the full near-field near-field information. This was first demonstrated by von Ribbeck [54] in 2008 and significantly improved in sensitivity by Moon [55, 56] in 2012 and 2015. This setup is described in more detail in section 4.3, where THz TDS s-SNOM is applied.

3.3. Pump-induced Sidebands in SNOM

In section 3.1 the permittivity ε_s is assumed to be constant during all times. This is true for many near-field experiments, but does not apply to pump-probe experiments, where the sample properties are changed periodically with a short pulse pump laser. This section will discuss the impact of a non-constant sample permittivity. Experimentally, this is most commonly performed with two different types of pulsed laser sources:

1. Either with a laser oscillator where the repetition rate is defined through the laser cavity length and the time of a round-trip of the laser pulse inside it. A typical repetition rate is $\sim$

80 MHz.

2. Or with a regenerative laser amplifier storing and amplifying the laser energy in the cavity before releasing the whole strong pulse at once via a pockels cell. Here, the repetition rate is much smaller but can in general be chosen arbitrarily. Usually this lies in the range of several Hz to hundred kHz.

The former case does not effect the thoughts of the previous section. The repetition rate of 80 MHz is much higher than the modulation frequencies of an s-SNOM. Also, the detector speed is usually too slow to directly detect such high frequencies. In the latter case, the repetition rate of the pulsed laser can be much lower than the modulation frequencies in s-SNOM, which has serious consequences on the s-SNOM theory and especially the demodulation technique in particular. In this section, not only the induced changes are discussed, but we also present how it is possible to take advantage of this modification.

Assuming the s-SNOM system now exhibits two modulations (of the tip-sample distance $z(t)$ and the change of the sample permittivity $\varepsilon_s(t)$), so-called amplitude modulation will be discussed in general, before the conclusions will be adapted to s-SNOM theory.

Let us assume a high frequency carrier signal $u_1 = U_1 \cos(f_1 t)$, where U_1 is changed via an addition of a low frequency and small modulation signal $u_2 = U_2 \cos(f_2 t)$, with $f_1 \gg f_2$. The amplitude modulated signal is then given by

$$\begin{aligned} u_{\text{am}} &= \Big(U_1 + U_2 \cos(f_2 t)\Big) \cdot \cos(f_1 t) \\ &= U_1 \cos(f_1 t) + U_2 \cos(f_2 t) \cdot \cos(f_1 t) \\ &= U_1 \cos(f_1 t) + \frac{U_2}{2}\Big(\cos((f_1 - f_2)t) + \cos((f_1 + f_2)t)\Big) \end{aligned} \tag{3.3.1}$$

with $f_1 \gg f_2$. Equation 3.3.1 clearly shows that the signal now consists of components at three different frequencies namely f_1, $f_1 + f_2$ and $f_1 - f_2$, where the component at f_1 is the largest.

This theory will now be applied to the basic s-SNOM theory of section 3.1. As shown in equation 3.2.2, the scattered signal can be expressed as

$$E_{\text{sca}}(z) = E_0 + E_1 \cos(\Omega t) + E_2 \cos(\Omega t) + \ldots \tag{3.3.2}$$

Let us for a moment assume the tip-sample distance is not modulated, but instead the permettivity of the sample ε_s is periodically changed with a frequency f_p. Hence, $\beta = (\varepsilon_s - 1)/(\varepsilon_s + 1)$

is also modulated with f_p. Depending on the value of f_p in the ground state and the strength of the modulation, $beta$ can have very different shapes. In general, it can be written by the Fourier series

$$\beta(t) = \beta_0 + \beta_1 \cos(f_p t) + \beta_2 \cos(2 f_p t) + ... \tag{3.3.3}$$

That means $\beta(t)$ will change the effective polarizability of the system, the scattering cross-section and thus also the scattered electric field (equation 3.2.2) with the time t. Due to the non-linearity and asymmetry in β of the polarizability of the the system (see equation 3.1.5), the scattering cross-section will contain all frequency components of β (e. g. f_p, $2f_p$, $3f_p$, $\ldots$). Hence, the overall scattered field (including the modulations arising from Ω and f_p) will now be modulated

$$E_{\text{sca}}(z) = E_0 + E_0^\beta + \left(E_1 + E_1^\beta(t)\right)\cos(\Omega t) + \underbrace{\left(E_2 + E_2^\beta(t)\right)\cos(2\Omega t)}_{=\xi_2} + \ldots, \tag{3.3.4}$$

where each time-dependend $E_n^\beta(t)$ component will carry frequency components of f_p, $2f_p$ and so on. The corresponding fields will be labelled by $E_{n,m}^\beta$. If we now look at ξ_2 and apply the amplitude modulation from equation 3.3.1 this leads to

$$\begin{aligned}
\xi_2 =& \left(E_2 + \sum_m E_{2,m}^\beta(t)\right)\cos(2\Omega t)\\
=& \left(E_2 + E_{2,1}^\beta \cos(f_p) + E_{2,2}^\beta \cos(2f_p) + E_{2,3}^\beta \cos(3f_p) + \ldots\right)\cos(2\Omega t)\\
=& E_2 \cos(2\Omega t)\\
&+ \frac{E_{2,1}^\beta}{2}\left(\cos((2\Omega - f_p)t) + \cos((2\Omega + f_p)t)\right)\\
&+ \frac{E_{2,2}^\beta}{2}\left(\cos((2\Omega - 2f_p)t) + \cos((2\Omega + 2f_p)t)\right)\\
&+ \frac{E_{2,3}^\beta}{2}\left(\cos((2\Omega - 3f_p)t) + \cos((2\Omega + 3f_p)t)\right)\\
&+ \ldots
\end{aligned} \tag{3.3.5}$$

We can now show that the scattered field of a system with periodic change of ε_s will in general lead to sidebands with distances mf_p around the frequencies of the higher harmonics $n\Omega$. Please note, the shape and origin of the periodic change were not defined in this discussion and will drastically change the prefactors β_m and consequently $E_{n,m}^\beta$. For example, it can be a short pulse

laser source as in an optical pump-probe experiment or a sinusoidal change of an applied bias voltage at the sample. Calculating these prefactors is in general straight forward. However, for the comparison to a to a real experiment such as the pump-probe experiment performed in section 5.1, one would need a number of parameters with much higher accuracy (e. g., pump fluence, carrier lifetime, field enhancement of the pump by the tip, ...). The fourier transformation for a realistic case was calculated twice, once with a numeric approach starting from the dipole model and a second time from a COMSOL simulation comparable to the simulations performed in section 3.4.3. For both cases, the numerical fourier transformation suffered from numerical artifacts making a quantitative analysis impossible.

The system becomes even more complex when the sample reflectivity is taken into account (see equation 3.1.8). In this case, not only β is a function of f_p and its higher harmonics, but also r_n. The scattering cross-section now depends on several functions with the parameters Ω and f_p

$$\alpha_{\text{eff}} = \frac{\alpha(1+\boldsymbol{r_n})^2}{1-\frac{\alpha\boldsymbol{\beta}}{16\pi\boldsymbol{r}^3}}, \tag{3.3.6}$$

where the bold symbols indicate the time dependent quantities. $r_n(f_p,t)$ is the time-dependent change of the Fresnel coefficient, $\beta(f_p,t)$ represents the time-dependent near-field coupling of tip and sample, and $r(\Omega,t)=z(\Omega,t)-a$ is the time-dependent tip-sample distance. When this is implemented in equation 3.3.5, additional fields $E^{r_n}_{2,m}$ will contribute to ξ_2:

$$\begin{aligned}\xi_2 =& \Big(E_2 + \sum_m E^{\beta}_{2,m}(t) + E^{r_n}_{2,m}(t)\Big)\cos(2\Omega t)\\ =& \Big(E_2 + E^{\beta}_{2,1}\cos(f_p) + E^{\beta}_{2,2}\cos(2f_p) + E^{\beta}_{2,3}\cos(3f_p) + \dots\\ &+ E^{r_n}_{2,1}\cos(f_p) + E^{r_n}_{2,2}\cos(2f_p) + E^{r_n}_{2,3}\cos(3f_p) + \dots\Big)\cos(2\Omega t)\end{aligned} \tag{3.3.7}$$

These additional fields are changing the overall amplitudes of the harmonics and sidebands. However, as these additional prefactors have their origin in the reflection directly beyond the tip, they will not compromise the sub-diffraction resolution of this method.

A schematic illustration of the frequency spectrum of an amplitude-modulated s-SNOM experiment is shown in figure 3.3.1. A pulsed pump laser (green) is introduced in (a), leading to sidebands in the far-field (b) and near-field signatures (b) with a distance of $m \cdot f_p$ to the higher harmonics. Here, only sidebands up to $m=3$ are shown due to simplicity.

Experimentally, the direct detection of the sidebands can be exploited for a vastly increased sensi-

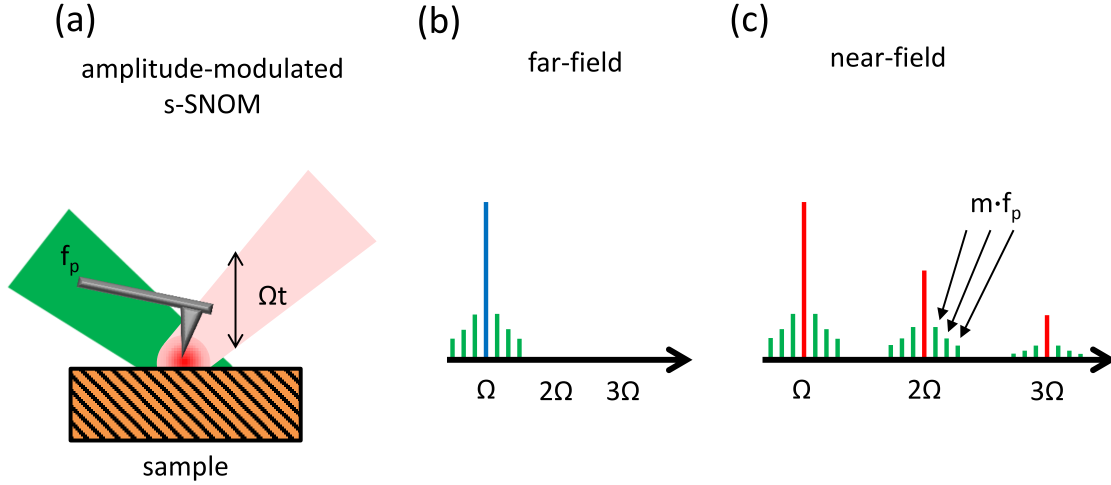

Figure 3.3.1.: *Schematic illustration of amplitude-modulated s-SNOM: (a) the near-field microscope is complemented with a pulsed pump laser with the repetition rate r_p. (b) The far-field signal now has sidebands with the distances of $m \cdot f_p$ to the Ω-peak. The same behavior is depicted for the near-field in (c) for the higher harmonics of Ω.*

tivity to changes of the sample properties. When the scattering cross-section is calculated for typical sample permettivities ε_s (e. g. semiconductor, gold) in the mid-infrared regime, the changes are relatively small. A few values are given in table 3.1. For a germanium sample in the ground state, the cross-section is $2.31 \cdot 10^{-22}$ m^{-2}, for a gold sample it is $3.49 \cdot 10^{-22}$ m^{-2}. When the germanium sample is excited with a strong pump laser as described in section 2.1, the permittivity can easily reach $\varepsilon_s = -100$ for a probing laser wavelength of 10.6 μm. However, due to carrier lifetimes on the range of few ten μs [20], the germanium will relax to the ground state in $\approx$ 100 μs. Directly after the pump pulse, the cross-section is only 12% larger than the cross-section of the ground state. However, when the permittivity passes the resonant regime of $\varepsilon_s = -1 \ldots -5$, the scattering cross section rises strongly (e. g. $1.15 \cdot 10^{-21}$ m^{-2} for $\varepsilon_s = -3$). For a strongly resonant system ($\varepsilon = -1.5 + 1.2i$), the cross-section can become an order of magnitude larger than the ground state's value. The time period where the cross section is actually enlarged can be estimated by roughly the double of the decay time [127] and is in this case few ten μs. Due to this fact and a typical repetition rate of only 1 kHz, the duty cycle of this pump-probe system is only a few percent and the time-averaged change of the signal is too low to be detected with the sensitivity of a typical s-SNOM.

ε_s	$C_{perp}[m^{-2}]$	**sample type**
$-4679.9 + 1674.5i$	$3.49 \cdot 10^{-22}$	gold [128]
16	$2.31 \cdot 10^{-22}$	germanium [128]
$-1.5 + 1.2i$	$1.15 \cdot 10^{-21}$	strongly resonant system
$-100 + 3.5i$	$3.74 \cdot 10^{-22}$	germanium after pump

Table 3.1.: *Scattering cross-section for the dipole model at a distance of 25 nm for different sample permettivities. The wavelength is 10.6 μm and the tip is made of platinum ($\varepsilon_p = -1450$) with a radius of 20 nm.*

In contrast, if the sidebands are detected directly, the pump-induced changes can easily be detected [47, 127]. Without a periodic excitation, sidebands are non-existent. Theoretically, the contrast in between the pumped and unpumped system is infinitely high. In a real experiment, the noise floor of the detection method limits the contrast to $\sim$ 95 % [47] which is still an extremely large value for a s-SNOM experiment. In section 5.1, a two-Lock-In detection method is described and applied to directly deduce the sideband signal in the experiment.

In a pump-probe experiment, the change of sample properties can also induce macroscopic artifacts in the measurement. When the sample properties are changed on a macroscopic scale due to the diffraction limited pump laser pulse, the background field E_0 can also exhibit an unknown periodic change due to changed reflection on the sample surface. The strength of the reflection is determined by the Fresnel coefficients, which are non-linear functions of the refractive index. Additionally, the refractive index exhibits a non-linear change when the Drude model for semiconductors is applied (section 2.1). Following the previous concept E_0 is then a function of

$$E_0\Big(r_n(t)\Big) = E_0^0 + E_0^1(f_p) + E_0^2(2f_p) + \dots \tag{3.3.8}$$

which will consequently induce sidebands. As E_0 is the largest field contributing to the detected signal, one can assume that this will cause sidebands in equation 3.2.3 with extremely large amplitudes. As E_0 is a macroscopic quantitiy, this artifact will undermine the sub-diffraction resolution and must be handled carefully. To minimize it, the laser focus of the pump laser should be small compared to the focus of the probe laser to make sure the overall change of the backgroud field reaching the detector E_0 is small. In experiments with other stimuli (such as AC bias voltages, temperature jumps, ...), the excited area should be much smaller than the probe laser focus as well. In addition, it should always be checked if the sideband signal vanishes on a material which

cannot be excited by the pump laser (e. g., a dirt particle or a small gold reference patch on the investigated sample) and if the sideband signal shows contrast with sub-diffraction resolution.

Despite the challenges arising from artifacts as described above, this technique has successfully been applied in this work. The results are presented in section 5.1.

3.4. Field Enhancement by Resonant Probes

The analytic description of s-SNOM offers many insights into the fundamentals of such a model system, however the predicted scattering cross-sections are much smaller than the findings in an actual experiment. In the seminal paper by Knoll and Keilmann [118] they find the detected power in the experiment to be four to five magnitudes larger than the predictions by their model. Already at this point, they correctly attributed this to the antenna effect of the tip (often referred to by ”lightning rod effect”). Without this effect, the detection of the small near-field signatures in s-SNOM would be impossible.

The field enhancement has been investigated in numerous studies so far [15, 80, 88, 129–131] and depending on the geometry, wavelength and material the calculated enhancement is ranging from below 10 [129] up to 25000 [131]. One study investigating the enhancement in the THz range used the geometry of a self-assembled tip system leading to field enhancements of ~ 3500 [88]. Unfortunately, the assembly of this tip costs a lot of effort and is not commercially available. Another theoretical study used an excessively long and thin wire [131] leading to unrealistic high values of the field enhancement with a factor of more than 25000. In a real experiment, this large field would instantly damage the tips' apex.

For the THz near-field experiments performed in section 4.3 a commercially available Rocky Mountain Nanotechnology tip [132] was chosen. The tip's properties for the application in s-SNOM are discussed in section A.2 in general . Due to its geometry with a tip shank length of 80 μm a strong field enhancement in the THz regime is expected, an antenna effect which is investigated with a numerical simulation. In this section, numerical simulations of the field enhancement are performed for different tip geometries with COMSOL Multiphysics [1] with a finite-element-method in the frequency domain. The model was developed by Lisa Ortmann in her master thesis [133] for the same geometry used in the dipole model and adapted by the author for other tip geometries.

The main features of the model are explained in this section. For a more detailed discussion see [133], the COMSOL general introduction [134] and the wave optics package documentation [135].

[1] version 5.3, `www.comsol.com`

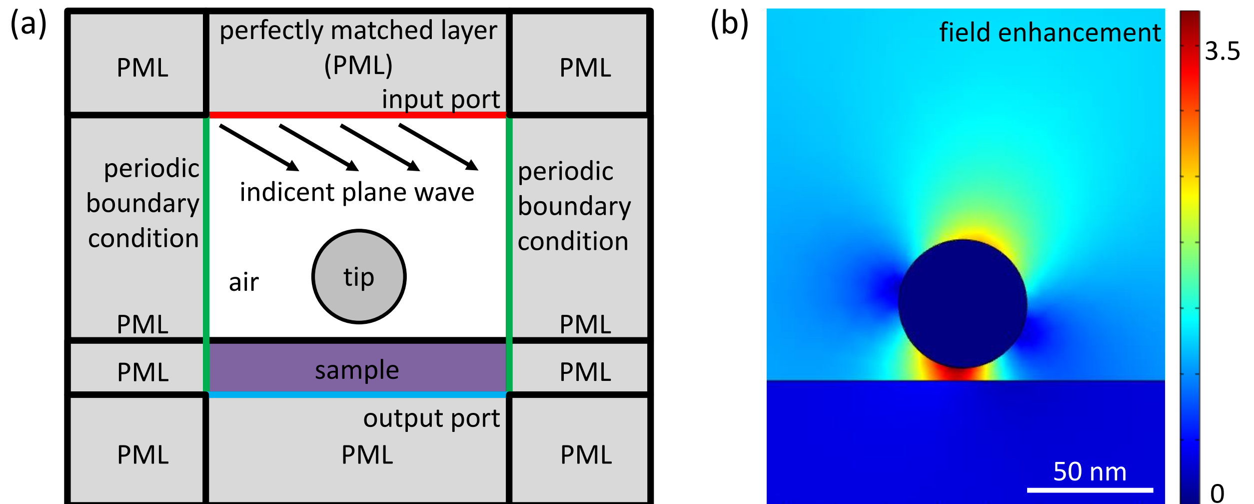

Figure 3.4.1.: *Field enhancement in s-SNOM: (a) Sketch of the 2D COMSOL simulation geometry (not to scale). The tip is located 5 nm above the sample surface in air. The tip-sample system is surrounded by perfectly matched layers (orange) to avoid interference. The incident field is injected via an input port (red) and can exit the simulation via the output port (blue). The green lines mark two planes with a periodic boundary condition. The calculated field enhancement is shown in (b). Simulation with 300 μm wavelength. The strongest field is located in between sphere and sample and is a factor of 3.5 larger than the incident field. The sphere also shows a field enhancement at the top. The slightly asymmetric behaviour is a result of the 60 degree incident angle of the light.*

The scatterer is located five nm above the sample surface and is either a sphere or a hemiellipsoid with a height of 10 or 80 μm. All three geometries have an apex radius of ≈ 10 nm. A rectangular geometry around the scatterer is used to implement periodic boundary conditions and to avoid unwanted interference effects due the finite size of the simulation. A sketch of the simulation geometry is shown in figure 3.4.1 (a). All materials are isotropic justifying a 2D simulation instead of a time-consuming 3D calculation. The sphere has a radius of 50 nm and is made of platinum with a strongly negative permittivity. As the sample material, germanium with $\varepsilon = 16$ was chosen. Sphere and sample are located inside a rectangle assembled out of an input and output port and two layers with a periodic boundary condition. The input port injects a plane wave with an angle of 60 degree with p-polarization. The output port ensures that no reflection is going back into the system which would cause an infinite amount of energy stored in the model. The periodic boundary condition allows for the virtual calculation of an infinite space. The perfectly matched layers (PML) suppress any reflection of the waves at the outer edges of the model and hence eliminate artifacts arising from the finite model size. Please note that the PML size is changing with the wavelength. For the calculation, the mesh size is chosen to an equally spaced quadratic mesh in the PML. In the sample, the mesh is automatically conceived by the COMSOL software with a minimum element size of 1 nm. For the region of the sphere and the air, the mininum element size is 0.1 nm.

The result of a simulation with a wavelength of 300 μm $\widehat{=}$ 1 THz is shown in figure 3.4.1 (b). The strongest field enhancement of 3.5 is reached directly in between sphere and sample. The sphere also shows an increased electric field at its top ≈ 2.5 and a reduced field at its side. The slightly asymmetric behavior is a result of the 60 degree input angle of the incident field. The areas of low field are mostly determined by this angle. Nonetheless, the area of strongest field enhancement is only slightly influenced by the angle and depends much stronger on the chosen geometry.

More realistic geometries were chosen for the simulation presented in figure 3.4.2. Here, the sphere was replaced with a hemi-ellipsoid with a length L of either 10 or 80 μm in simulation (a) and (b) respectively to take the height of different tip types into account. They represent either a Nanosensors PPP-NCLPt (10 μm, named PPP in the follwing) or a Rocky Mountain Nanotechnology 25PT300B (80 μm, named RMN) tip. The shorter semi-axis is chosen to a value that both geometries have the same apex radius of $\approx$ 10 nm, a realistic value for a new tip. At the upper end of the tip, the edges of the hemi-ellipsoid were rounded to avoid artifacts in the simulation. The wavelength is 300 μm $\widehat{=}$ 1 THz. The field enhancement of the PPP tip reaches a value of $\approx$ 250, whereas the RMN tip produces an enhancement of 530 due to its larger long axis and

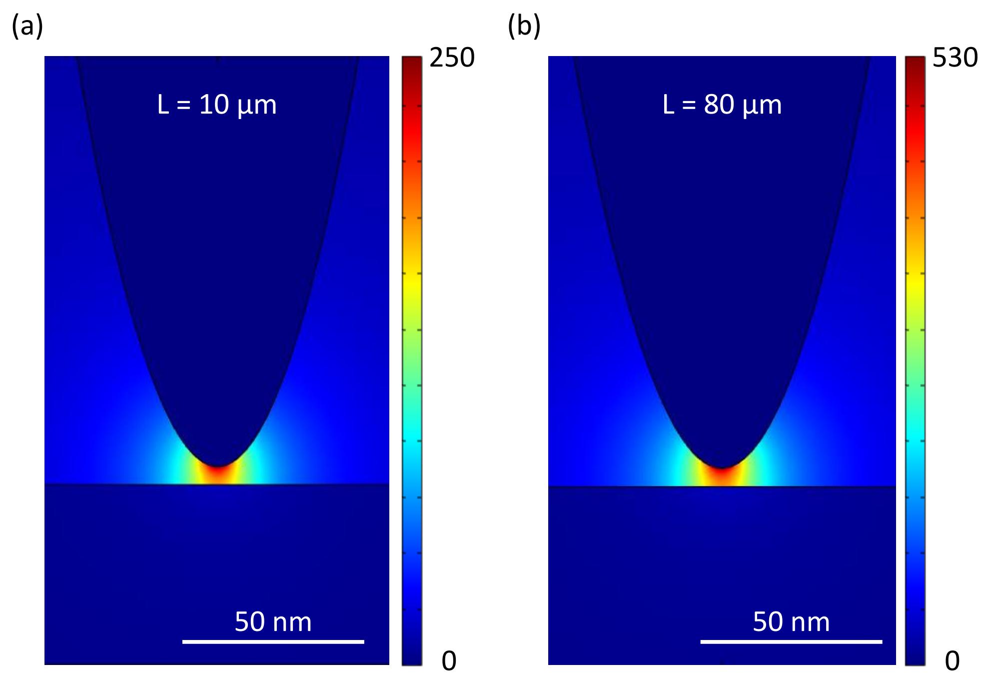

Figure 3.4.2.: *Field enhancement with enlarged near-field probes: For a hemiellipsoid with 10 μm height (a), the field enhancement reaches up top 250. For a height of 80 μm the enhancement is 530.*

better coupling to longer wavelengths. Also the lateral size of the field enhancement is slightly larger with the latter tip. The influence of the 60 degree incident angle is not visible in the field distribution, which is determined completely by the long axis of the tip.

To determine the enhancement efficiency for a large wavelength range the simulations shown in 3.4.1 and 3.4.2 were repeated for a number of wavelengths covering the whole range from the far-infrared down to low THz radiation. In each solution the largest value was determined with COMSOL. The results are depicted in figure 3.4.3.

The field enhancement for the sphere is $\approx$ 3.81 for all wavelengths from 5 to 300 μm. Only at the highest value, the enhancement is slightly larger with $\approx$ 3.84. Here, the simulation is much larger due to the increased PML thickness and the change of enhancement is probably a numerical artifact.

The behavior is completely different for the PPP and the RMN tip. For this geometry, the enhancement is already more than 20 for a wavelength of 5 μm. The PPP tip shows a overall rising trend for larger wavelengths and a resonance at $\approx$ 15 - 16 μm. The position of the peak is in the same range as reported in [131] and is attributed to the first eigenmode of the long axis of the tip. For an ideal antenna illuminated under an angle of 60 degree without a sample, the first eigenmode is found at $\lambda = 2 \cdot L \sin(60°) \Leftrightarrow L/\lambda \approx 0.6$. In the simulation it is found at $L/\lambda \approx 0.63$ and in good agreement with the theoretical value. The field enhancement reaches values of up to 290 for $\lambda =$750μm. The enhancement seems to reach a plateau of $\approx$ 250 for values up to 300 μm and is rising again for larger wavelengths.

For the RMN tip, the simulation is much larger due to the eight times larger tip size which strongly increases the mesh size and the calculation time. Due to the scalability of the problem, the same behaviour, is expected from the simulation, but simply shifted to longer wavelengths. However, the results seem to be less stable with more variation in the field enhancement compared to the smaller PPP tip. The field enhancement is slightly lower for small wavelengths. For wavelengths larger than 17.5 μm, the enhancement with the RMN tip is in general larger than for the PPP tip and reaches values of more than 1200 for $\lambda =$ 750 μm. Despite lower stability, the resonance peak at $\lambda = L/0.61 \approx$ 130 μm is still visible for this tip as well.

The simulation shows that the dipole model is not sufficient to quantitatively describe a s-SNOM system, in particular not at long wavelengths. The enhanced model by Cvitkovic et al. for elongated tip shapes is limited to $L \leq \lambda/4$ and cannot be used at this point. The field enhancement is very strong for geometries which are approximating the tips shape in a much more realistic way. At the longest wavelengths, the difference in between the sphere and the RMN tip is more

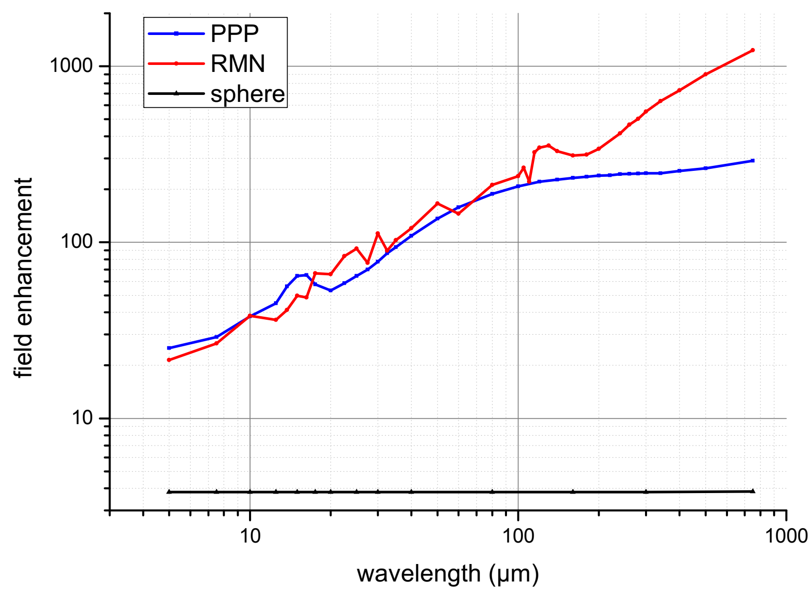

Figure 3.4.3.: *Field enhancement at different wavelengths. The sphere (black) shows no wavelength dependence. The PPP (blue) and RMN (red) tip show rising field enhancement for larger wavelengths. The long axis geometric resonance is visible at ≈16 μm and ≈130 μm, respectively.*

than a factor of 300. Hence, the detected near-field signal of a sphere in a real experiment would simply perish in the noise level. The simulation also justifies the application of the RMN tip for a THz near-field experiment despite its weaker performance as a non-contact AFM tip (compare section A.2). Although the commercially available RMN tips have been successfully utilized for scanning microwave microscopy (SMM) for several years now [136–138], the tip is barely applied for s-SNOM. To the best of the author's knowledge, the first application was performed during this work in 2017. Recently, a first report by Keilmann and Hillenbrand [48] was published where the tip was applied for s-SNOM in the range of $\lambda =$400-600 μm.

4. Near-Field Microscope Setups

Any sufficiently advanced technology is indistinguishable from magic.

(Arthur C. Clarke)

This chapter features the near-field microscopy setups utilized in this work. The fundament of the s-SNOM are the self-built non-contact AFMs described in the appendix (A.2). The narrow-band THz setup is located at the free-electron laser FELBE at Helmholtz-Zentrum Dresden-Rossendorf (HZDR) and is described first. It has been used for the investigation of graphene field-effect transistors (see section 6). The second setup is stationed at the TELBE beamline at the same accelerator facility. It has been employed for sideband measurements with a CO_2 laser and reassembled for the combination with a time-domain spectroscopy investigations in the low THz regime.

4.1. FELBE THz SNOM

The setup at the free-electron laser FELBE features a self-built AFM as described in the appendix A.2, however some details of the cantilever defelction readout are unique compared to this setup. The AFM is connected to the commercial controller[1] with a self-built four-channel second stage preamplifier[2] before it is processed via a four-channel interface[3] and fed into the controller. The setup features reliable performance for many AFM measurement types (as described in the appendix A.2).

All optical elements in the setup must be non-dispersive to untilize the setup in the enormous wavelength range FELBE can deliver (λ =5 - 230 μm). Hence, only metal-coated reflecting optical

[1]RHK Technology Inc., R9
[2]Specification are listed in A.2.
[3]RHK Technology Inc., PSD interface

elements are employed. The size of the elements is 2 inches as the FEL can easily reach such diameters. The setup is shown schematically in figure 4.1.1 and includes a visible alignment laser and a CO_2 laser for pre-examination to save precious time during FEL shifts. They can be coupled in via magnetically mounted mirrors and are directed through two apertures to make sure the beams are collinear. The alignment laser can be aligned by eye, for the invisible lasers, a 1 cm^2 512x512 pixel pyroelectric sensor[4] is applied. The CO_2 laser and FEL have a polarisation parallel to the optical table when entering the setup (s-pol.). It can be changed via a set of flip mirrors by 90 degree. The whole setup can be flooded with nitrogen gas to avoid absorption of the free-electron laser due to water vapor at ambient conditions.

Afterwards, the beam is split up into two equal parts with a geometrical beam-splitter, i.e. a gold-coated silicon wafer with a sharp edge produced with a water jet cutter. This unconventional beamsplitter is chosen due to the lack of an affordable non-dispersive cubic or thin film beam-splitter. The only material known to cover this extremely large regime with a spectrally flat response in transmission is diamond and simply too expensive in the large size of 2 inch needed for the setup. Moreover, unwanted interferences of the FEL can occur in a thin film beamsplitter. As a consequence of the geometrical beamsplitter, the focus of the parabolic mirror at the s-SNOM is less well defined due to a smaller numerical aperture. Due to a focusing angle of roughly 35 degrees to the sample surface, the focus will contain both p- and s-polarized components when the incident polarization is solely p- or s-polarized. This can be compensated by changing the incident polarization to an angle in between zero and 90 degree. This procedure is time-consuming and has not been performed in this work. The usual splitting ratio is 50:50, but can be changed with a micrometer screw by shifting the position of the beam-splitter in the direction of the plane where no realignment of the setup is necessary. In some cases, the deliberate change of the ratio can be desirable, i.e., when especially low powers at the sample are required and most of the scattered near-field signal shall be collected. In this work, the signal-to-noise ratio needed to be as good as possible, which was found for a ratio change towards higher values in the direction of the tip. A possible explanation is a better focusing and hence less scattering of the FEL at larger structures of the sample and/or AFM, which will eventually also reach the detector and create noise in the signal.

The beam is focused on the metallized tip with a parabolic mirror[5]. For best focussation and collection of the scattered light, it can be moved in three dimensions via a nanopositioning system[6].

[4]Ophir Optronics Solutions, Inc., Spricon Pyrocam III
[5]90 degree off-axis parabolic mirror, effective focal length 7.5 cm.
[6]mechOnics ag, MS30, 50 nm precision, closed loop.

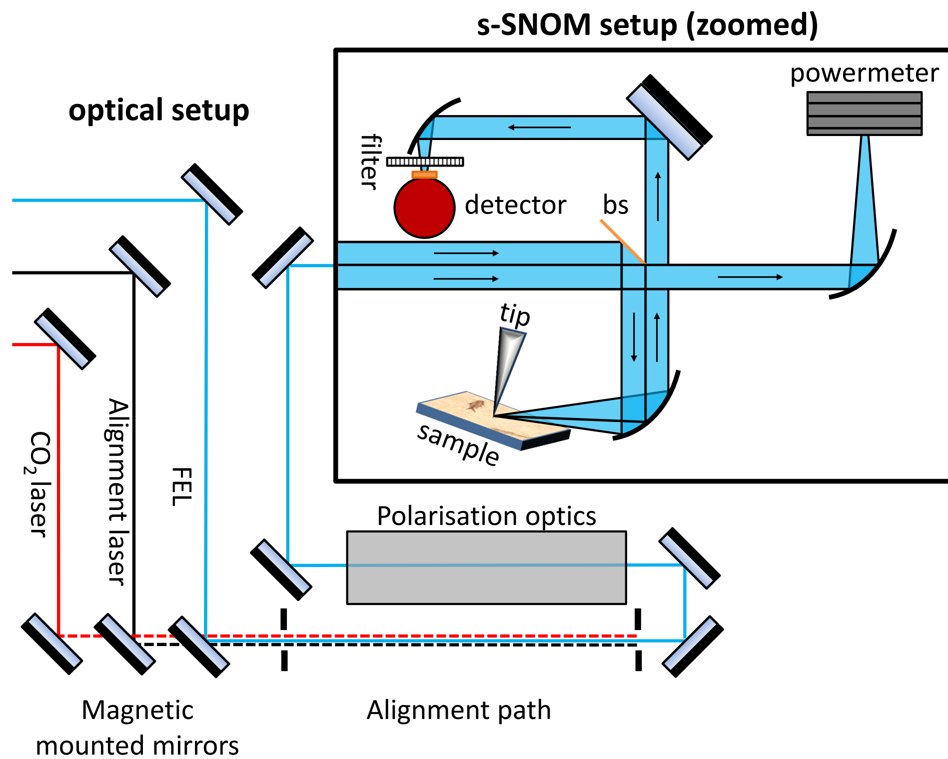

Figure 4.1.1.: *Setup at the free-electron laser FELBE: the different laser sources can be coupled in via magnetic mounted mirrors to allow for fast switching with good repositioning of the mirrors during experiments. After the alignment through two apertures, only the FEL (blue) is shown in the drawing. The near-field microscope is shown in more detail. A beamsplitter (bs) cuts the beam into two parts to illuminate the AFM tip and to measure the power at a powermeter. The back-scattered light is detected in the third arm of the setup.*

With a home-made LabView software[7], six positions can be saved and addressed e. g. for different laser sources and polarisations. The second part of the beam is directed into a powermeter to record the incident power at all times. A MCT detector[8] is applied to detect the CO_2 laser, other models are available for the wavelength range from mid-IR up to $\lambda = 28$ μm. For the detection of longer wavelengths from ~ 32 μm , a LHe cooled detection system is applied[9]. For wavelengths up to 100 μm a Ge:Ga photodetector is utilized, for longer wavelenghts up to the mm-range, an InSb superconducting bolometer is used. Both entraces have polyethylene windows with metal-meshed filters to cut off undetectable wavelength which would only enlarge the heat impact in the cryostat from the outside. A free-standing wire grid polarizer can be put in front of the detector to analyze the polarisation of the scattered light or to detect a certain polarization of it. All detectors systems in the setup measure AC signals only and have a bandwidth up to 1 MHz which is required for the detection of s-SNOM signals. Their signals are fed into the AFM controller to record the scattered near-field signal.

As described in appendix A.2, a gold sample is typically chosen to align the setup before the sample of interest is investigated. In figure 4.1.2, the near-field decay on a gold sample measured with the FEL at λ =150 μm is shown for a Nanosensors PPP-NCLPt tip both in linear (a) and logarithmic (b) scale for best visibility. Here, s-polarized illumination was chosen and a p-polarized filter was in front of the InSb detector to collect only the p-polarized scattered field and to suppress background noise [139]. The pulse-to-pulse stability of the FEL is not very high, thus the noise spectrum of the FEL contains any frequency below 13 MHz. If randomly scattered light (i.e., from parts of the AFM or oxygen molecules in the air) reaches the detector, it induces noise in the detector and consequently in the SNOM signal. Although the quality of the signal is quite remarkable already, this measurement has been performed before the setup was purged with nitrogen, hence a further improvement of the signal is possible. Interestingly, the wavelength is already long enough that the near-field decay is visible in the first harmonic as the demodulation technique is more efficient for longer wavelengths [50]. However, the decay length is twice as large as in the second harmonic and hence the resolution of a scan will be much better in the second harmonic.

As discussed in section 3.4.3, full-metal Rocky Mountain Nanotechnology tips should show a much better near-field coupling than silicon-based Nanosensors PPP-NCLPt tips. In order to compare

[7] Software written and implemented by Hans-Georg von Ribbeck

[8] Teledyne Technologies Inc., J15D12-M204-S100U-60 with PA-101 amplifier

[9] QMC Instruments Model QFI/3BI(2+Ge:Ga) in a TK1840 cryostat, a detailed manual can be found in [114]. The sensitivity spectra can be found in fig. 2.4.3. In this work the detector was enhanced at 2 THz.

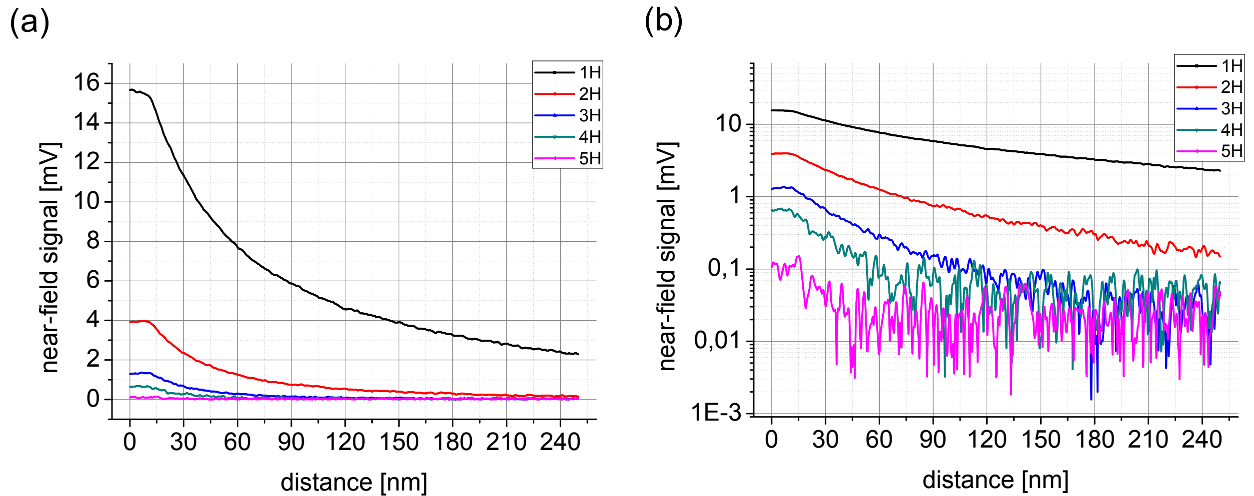

Figure 4.1.2.: *Near-field signal as a function of the distance on a gold sample measured at $\lambda = 150$ μm up to the fifth harmonic of the cantilever oscillation measured with a Nanosensors PPP-NCLPt tip. The input polarization is s-polarized, the filter passes only the p-polarized component of the scattered light. The near field can be detected with sufficient signal-to-noise ratio up to the fourth harmonic. The noise level is below 100 μV at a signal level in the range of several mV for an incident power of 18 mW, a oscillation amplitude of 80 nm and a lock-in integration time of 100 ms. Please note the linear (a) and logarithmic scale (b) for best visibility. The constant signal up to 15 nm is a consequence of a low tapping amplitude setpoint for this measurement. The values below 15 nm can be ignored.*

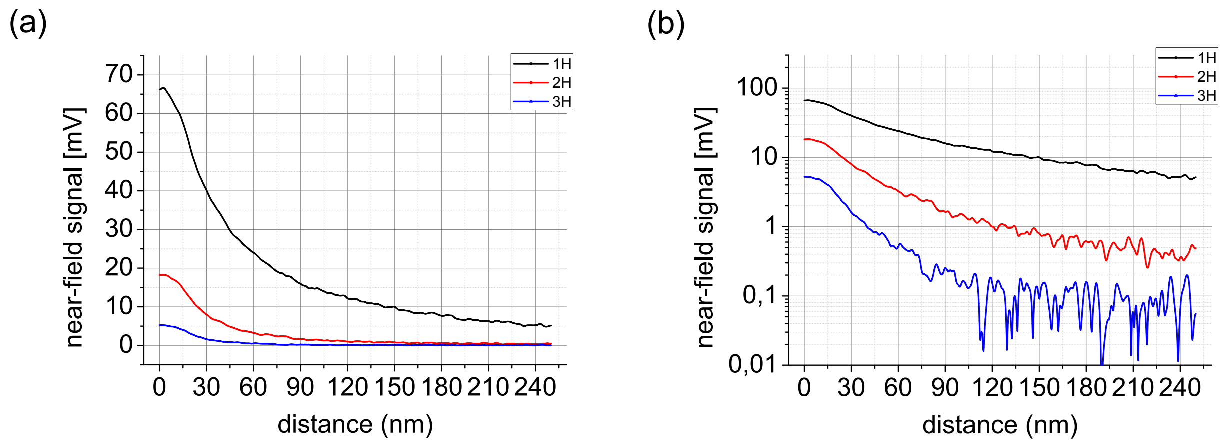

Figure 4.1.3.: *Near-field signal as a function of the distance on a gold sample measured at $\lambda = 150\ \mu$m measured with a RMN tip. The input polarization is s-polarized, the filter detects the p-polarized component of the scattered light. The near-field can clearly be detected up to the third harmonic. The noise level is in the range of 100 μV at a signal level in the range of several ten mV for an incident power of 11 mW, an oscillation amplitude of 80 nm and a lock-in integration time of 100 ms. Please note the linear (a) and logarithmic scale (b) for best visibility.*

the performance, the same experiment as discussed above was repeated with a RMN 25PtIr200B-H tip with a tip shaft length of 80 μm. One major difference in this experiment is the shorter focal length of the parabolic mirror focusing the beam onto the tip, which was included into the setup for better performance. With an effective focal length of 50 mm, the numerical aperture increases by a factor of $\approx$ 1.5 compared to the 75 mm effective focal length mirror used with the silicon-based tip. The result is depicted in figure 4.1.3. In general, the RMN tip shows a similar decay behavior when the tip-sample distance is increased. Notably, even the decay length matches quite well for both fresh tip models. The most significant difference is the much larger signal with the RMN tip compared to the PPP tips. The detected signal is much larger although a smaller illumination power of 11 mW was applied to the setup. A more detailed comparison is performed in the following.

Tip Performance Comparison

For a more quantitative analysis of the tips' characteristics and their performance at 2 THz on a gold sample, the distance dependence of the near-field is analyzed and compared for all tips

used with the free-electron laser at λ =150 μm in the table 4.1. The beam-splitting ratio. The data chosen for the comparison were the overall near-field signals demodulated at the second harmonic of the cantilever oscillation, the corresponding noise levels far away from the sample surface, the input laser power and the numerical aperture (NA) used in the experiment. From this values, the weighted near-field response (WR) per input power at the tip and signal-to-noise ratio was deduced for an overall number of eleven measured curves with fresh and used tips. To correct for the higher power density at the tip apex due to the different parabolic mirrors, we assume the NA has to be considered with its second power as the focus diameter will scale with the NA in its first power. As the amount of near-field collected by the parabolic mirror depends on the solid angle covered by it, this has to be accounted for as well. Here we choose the NA, although this simple assumption does not consider many details of near-field microscopy, such as unequal near-field emission for different propagation angles. However, as all measurements were performed with similar alignment and settings of the setup, these assumptions might influence calculated values but do not have an effect on their ratios. Overall, the WR is given by

$$\text{WR} = \frac{\text{signal}[\text{mV}]}{\text{power}[\text{mW}] \cdot \text{NA}^3}. \tag{4.1.1}$$

The signal-to-noise ratio is calculated with the basic definition of SNR=signal/noise.

The PPP tip measurements were mostly obtained with powers in the range of several ten mW, whereas the RMN tips give much stronger signals already with powers of 12 and 16 mW. The signal itself ranges below 5 mV for the PPP tips and in the range of 20 mV for RMN tips. The noise level in both measurement is determined by the FEL noise itself and can drastically change in between two beamtimes due to different alignment and machine parameters. On the one hand, the FEL noise will directly couple into the scattered near-field signal and into the background field E_0, as any fluctuation of the power density at the tip will change the near-field coupling efficiency. However, this should scale linearly with the near-field signal. On the other hand, scattering of the FEL on sample structures, tip shaft, cantilever, cantilever chip, water molecules in air and other scatterers ultimately reaching the detector will cause noise in the detected signal, hence alignment and the sample structure will influence the noise level. Here, the scaling in between near-field signal and noise can scale in a different way. These two factors can be observed when the PPP tips are compared. The tendency is a larger noise level for larger signals, however the signal-to-noise ratio ranges from 2.75 up to 16. In general, fresh tips should have a lower coupling efficiency, as the effective probe radius is smaller. This tendency can be observed when fresh and worn PPP tips are compared. Here, the WR is higher for used tips and can be almost a factor of

type	power [mW]	signal [mV]	noise [mV]	NA	comment	WR	SNR
PPP	17	0.55	0.2	0.33	fresh	0.9	2.75
PPP	40	0.8	0.07	0.33	fresh	0.6	11.42
PPP	48	1.1	0.15	0.33	fresh	0.6	7.33
PPP	55	1.3	0.1	0.33	fresh	0.7	13
PPP	44	1.1	0.15	0.33	fresh	0.7	7.33
PPP	29	4	0.25	0.33	used	3.8	16
PPP	24	5	0.4	0.33	used	5.8	12.5
PPP	50	3.5	0.3	0.33	used	1.9	11.67
PPP	50	1.8	0.3	0.33	used	1	6
RMN	12	17.5	0.5	0.5	fresh	11.7	**35**
RMN	16	25	1.5	0.5	fresh	**12.5**	16

Table 4.1.: *Comparison of tip performances with the free-electron laser at $\lambda = 150\ \mu m$. The best weighted response and signal-to-noise ratio (marked in bold) were obtained with RMN tips in the setup shown in 4.1.1.*

ten larger. The signal-to-noise ratio does not show a clear tendency for old and used tips and lies around ten for most tips.

The two RMN tips used for this analysis were fresh and show a much better performance with the FEL: The weighted response is a factor of 10 larger than for any new PPP tip. This allows for measurements at much lower incident power. For sample systems sensitive to heat impact, this can be a significant argument to use RMN tips in the THz range. Also, the signal-to-noise ratio is larger for the RMN tips. One tip reaches a SNR of 35 and is more than a factor of two better than any PPP tip. Overall, RMN tip are clearly superior to PPP tips for the application in the THz range.

It is not trivial to compare this data to the simulation performed in section 3.4.3, as the measured data depends in many ways on alignment and other experimental parameters. Also, the simulation accounts for the field enhancement and not for the overall near-field signal. However, the tendency of enhanced coupling and hence better performance is covered by both simulation and experiment.

4.2. Pump-modulated s-SNOM

In order to investigate pump-induced sidebands as predicted in section 3.3, a copy of the setup described in 4.1 was utilized with few modifications. The instrumentation with a CO_2 probe laser is located in the TELBE lab at HZDR, where many different laser sources are available, allowing for a systematic investigation sidebands induced by a NIR pump laser. The results on the sidebands are analysed in section 5.1.

The modified s-SNOM setup is depicted in figure 4.2.1. In most parts, it is similar to the setup at the free-electron laser so the description will focus on the parts which are different in this experiment. To avoid incoupling of the strong pump laser into the four-quadrant diode of the AFM, a high-pass optical filter[10] is mounted in front of the diode, sufficiently blocking the pump laser. The AFM signal second stage preamplifier is a commercially available four channel device supplied by RHK technology[11]. Instead of using mirror optics to change the polarisation of the CO_2 laser, a polarizer/attenuator[12] utilizing two free-standing wire grids placed consecutively in a mechanical holder is applied. Here the two wire grids can be rotated individually, with the second grid determining the outcoupling polarization and the first determining the outcoupling power. With this setup the polarisation and power can be changed steplessly and without any beam-steering as the wires are free-standing. Here, p-polarized light was used for pump and probe and no polarisation filter was used in front of the detector.

Excitation of the sample under the tip is experimentally challenging as the tip's shade is preventing the pump laser to reach the sample directly at the this position, and it is hard to confirm that the pump laser[13] focus is actually at the tip end. After many experimental trials using shadow images and pyroelectric detectors power measurements of the beams reflected at the sample, a collinear scheme of the pump and probe was confirmed to be the most prosperous approach to achieve an overlap at the focus position. To achieve collinear beams, an additional mirror is included into the setup. It is made of a gold-coated silicon wafer with a hole drilled by the author to transmit the pump laser. The CO_2 laser has a much larger diameter and is reflected at the parts around the hole. As the detection path is aligned to collect radiation from the tip, parasitic incoupling of the pump laser radiation scattered at the tip into the detector must be avoided. To ensure this, a germanium filter is placed directly in front of the detector effectively blocking

[10] Edmund Optics 84-714, OD4
[11] RHK Technology Inc, Beetle AFM Preamp.
[12] LASNIX, continuously variable attenuator model 401
[13] Legend, Coherent Inc.

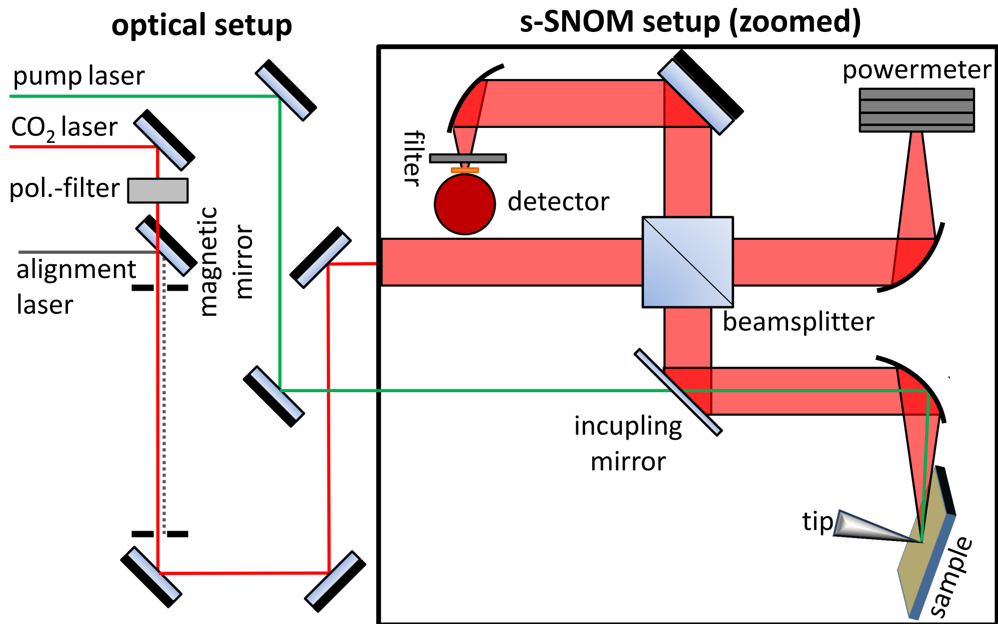

Figure 4.2.1.: *Setup for sideband investigations: For alignment, a visible laser can be coupled into the setup, using two apertures to achieve collinearity with the CO_2 (red) laser. The near-field microscope is shown in more detail. A beamsplitter cuts the beam into two parts to illuminate the AFM tip and to measure the power at a powermeter. The back-scattered light is wavelength-filtered with a germanium crystal detected in the third arm of the setup. To couple in the pump laser (green), a mirror with a drilled hole is employed in the arm of the AFM.*

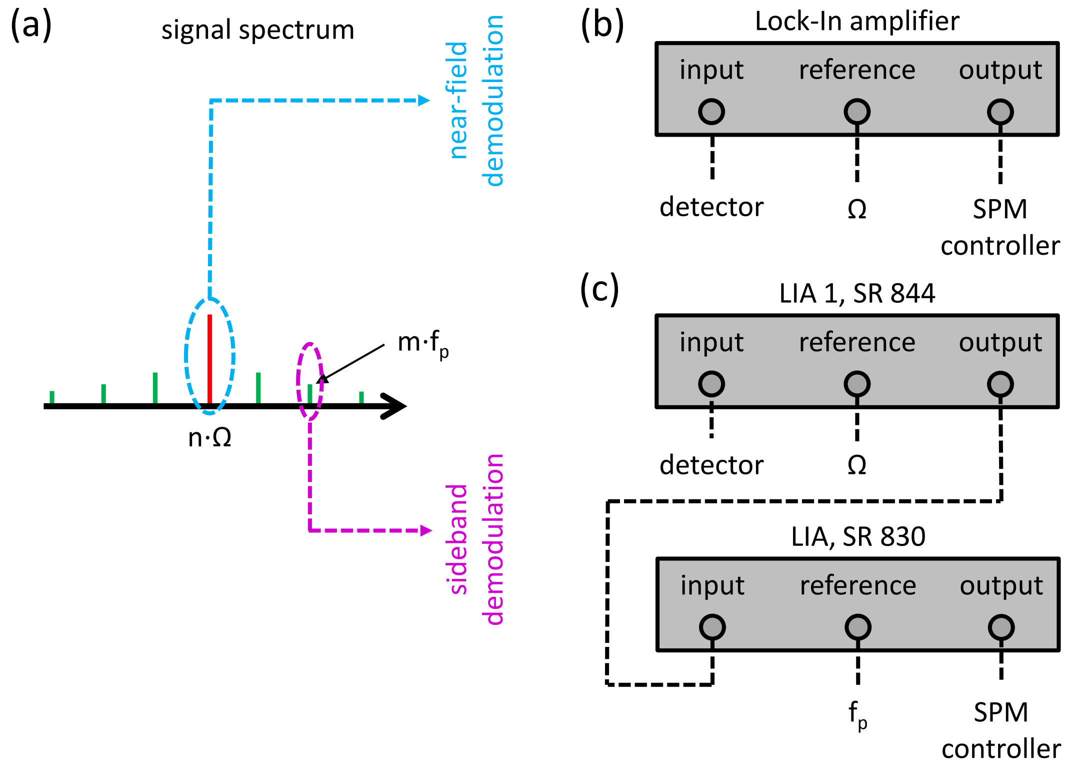

Figure 4.2.2.: *Sideband demodulation setup: (a) Expected signal spectrum with higher-harmonic peak (red) and sidebands (green). The classic near-field demodulation is depicted in (b) where the detector signal and the cantilever oscillation frequency is used for higher-harmonic demodulation. For the demodulation of the sidebands (c), two cascaded Lock-In amplifiers are used, where the second Lock-In demodulates the desired sideband.*

any radiation from the pump laser as its wavelength $\lambda_p = (800 \pm 10)$ nm is strongly absorbed by the semiconductor (see section 2.1). The s-SNOM performance of this setup before incuding the pump laser is described and evaluated in section A.2.

For demodulation of the sidebands predicted in section 3.3, a cascaded Lock-In amplifier scheme was applied. The basics are depicted in figure 4.2.2. The expected signal spectrum is sketched in (a). The red bar indicates the nth harmonic of the cantilever oscillation Ω and the green bars show the sidebands with a spectral distance of $m \cdot f_p$ to the main peak. When the conventional near-field shall be detected, the detector signal is fed into a Lock-In amplifier with the reference frequency Ω (b). Choosing the desired harmonic and applying a steep filter, the Lock-In amplifier extracts only the main peak of the nth harmonic and hence the unmodulated near-field signal which is sent to the SPM controller for data acquisition.

For the demodulation of the sidebands, two Lock-In amplifiers are utilized. The first Lock-In[14] needs to have the option of short integration times τ_i and a fast output channel, specifically τ_i must be below the time of two consecutive pump pulses and the bandwidth must be larger than the repetition rate of the pump laser. If the Lock-In amplifier does not support this speed, the output signal will be averaged over more than a period of the pump laser and any temporal change of the near-field signal will be averaged. Hence no sidebands could be detected. Moreover, the Lock-In amplifier also needs a broad filter which still passes the sidebands in the frequency domain. The demodulated and at this point amplitude-modulated near-field signal amplitude is then fed into a second Lock-In amplifier[15] and uses the pump laser repetition rate as a reference to demodulate the signal at the mth harmonic of the sideband frequency. The output signal is passed to the SPM controller for data acquisition.

A parameter study for different Lock-In amplifier settings has been performed in [127] in detail for an amplitude-modulated signal from a waveform generator and will not be discussed at this point. The experimental results of this work are discussed in chapter 5.1, where higher-order sidebands could be observed on a germanium sample.

4.3. THz Time-Domain-Spectroscopy SNOM

Additionally, a time-domain spectroscopy (TDS) s-SNOM setup has been installed to be combine this setup with the accelerator-based TELBE beamline. It enables phase-resolved detection of broadband THz near-field pulses. As beamtime is rare and valuable, the setup was aligned first with a table-top laser source to simulate the experimental behavior as realistic as possible before the utilization of TELBE can be performed. The major obstacles of TELBE are a low repetition rate of 100-200 kHz and very unstable pulses, including amplitude and timing jitters, making a post-mortem data sorting mandatory [140]. One the one hand, the setup described in this section allows for gaining experience with a forward-scattering TDS-SNOM setup. On the other hand it allows for pre-alignment of the optics before the TELBE beam is guided into it during beamtimes.

The setup consists of a pump and a probe beam. Please note that the nomenclature of pump and probe in this section is used as common in time-domain spectroscopy. While the pump beam excites the THz pulse, the probe beam is used to probe the scattered near-field pulse. Both beams are supplied by a pulsed laser from a regenerative amplifier[16] with a center wavelength of 800 nm

[14] Standford Research Systems SR 844 in this setup
[15] Standford Research Systems SR 830 in this setup
[16] Coherent Inc., RegA

and a pulse length of 100 fs. A repetition rate of 200 kHz is applied. This unusual low frequency is a major obstacle for a decent signal-to-noise ratio in s-SNOM as explained in the following: Assuming a typical tapping frequency of 150 kHz of the AFM tip, the Nyquist-Shannon sampling theorem [141] yields a minimum sampling frequency of 300 kHz for the first harmonic, 600 kHz for the second harmonic and so on. To avoid undersampling, a low tapping frequency tip (ideally below 20 kHz) is mandatory. Moreover, a tip with a strong field enhancement at THz frequencies is beneficial to achieve the strongest near-field coupling in between tip and sample. Both needs are fulfilled by the Rocky Mountain Nanotechnology 25Pt300B tip (80 μm long, see section A.2), which has been successfully applied in this section.

The setup utilizes the same AFM as in section 4.2. However, the optical setup follows a forward scattering geometry, where the light reflected from the sample and the light scattered from the tip is guided towards the detector. This is quite unusual in s-SNOM as much more background radiation reaches the detector and makes the setup more susceptible to artifacts. In the infrared range, there has only been one publication [53] using this scheme in the last years, as the applied MCT detectors are quite easily saturated from the strong far-field radiation. In contrast, for time-domain spectroscopy s-SNOM, all reported experiments have been performed in forward scattering direction [54–56, 89]. As the detection scheme for TDS setups has an enormous dynamic range and operate far from saturation, the near-field can be detected although they do not interfere with a large background field. The alignment is more complex as spatial and temporal overlap must be achieved in the setup. Typically, the signal is first aligned via the direct far-field reflection from the sample and the near-field is then optimized in a second step. For low THz radiation up to frequencies of few THz as applied here, a ZnTe crystal is typically used as an electro-optic detector of the radiation.

The TDS-SNOM setup is depicted in figure 4.3.1. The incoming beam of the above described pulsed laser is divided by a beamsplitter (BS) into a strong pump beam ($P \approx 800$mW) and a much weaker probe beam ($P \approx 1$ mW). The pump is guided via a delay stage[17] and hits a photoconductive antenna (PCA)[18], which emits p-polarized radiation. As the antenna is a large-area emitter and emits with an opening angle of several ten degrees, a pair of lenses is applied to achieve a decent collimation of the THz beam before it is focussed with a parabolic mirror[19], on the AFM tip. The forward-scattered part of the near-field is collected with a second parabolic mirror using the same focal length and then focused with a third parabolic mirror onto a ZnTe

[17] Newport Corp., M-IMS series
[18] Laser Quantum GmbH, TeraSED 10, specification can be found in [107]
[19] 90 degree off-axis parabolic mirror, focal length 75 mm

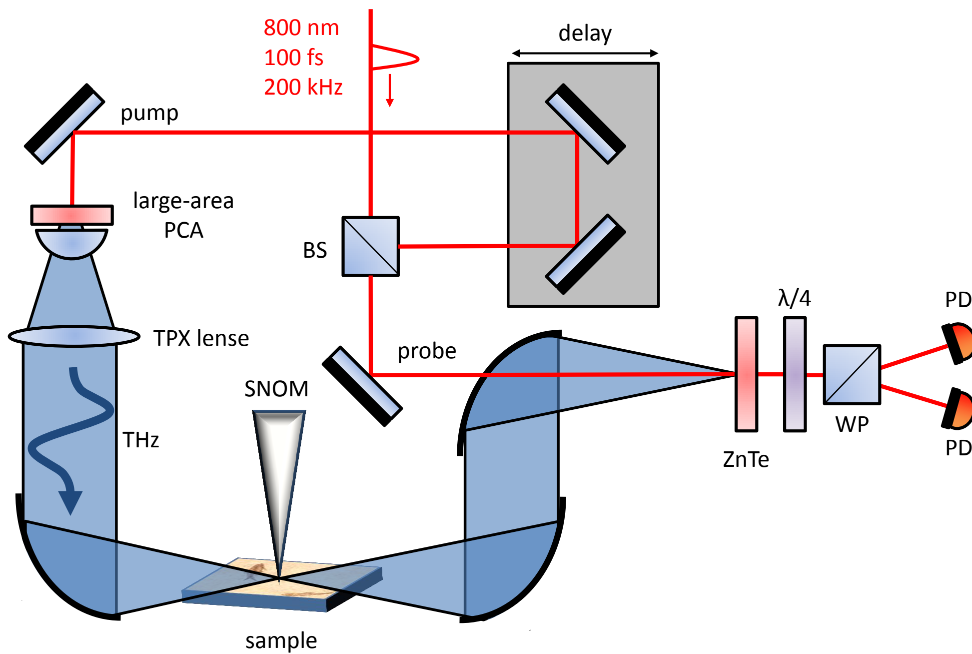

Figure 4.3.1.: *Time-domain spectroscopy setup: The incoming laser pulse is divided into two parts. One excites the photoconductive antenna which emits a THz pulse. The second part is used for probing of the THz pulse via electro-optic sampling at a ZnTe crystal.*

crystal. The linearly polarized probe beam is guided through a hole in the parabolic mirror to the crystal as well, then passes a $\lambda/4$ waveplate leading to a circular polarized beam. When it passes the Wollaston prism (WP), it is equally split up on two balanced detectors (see section 2.4).

Similar to concepts discussed in section 2.4, this setup allows for direct probing of the THz field's amplitude and phase. As the oscillating tip modulates the scattered near-field amplitude, the near field can be detected with higher-harmonic demodulation when the signal of the balanced detector is fed into a Lock-In amplifier[20]. The setup can be used for both local spectroscopic investigations of the material and for broadband scanning of nanostructures. In the latter, the detected signal is spectrally integrated. Hence, the setup is in general a powerful device to detect nanoscopic material compositions (similar to FTIR spectroscopy) and small variations in it, but the dectection of resonant sample areas is not as straight forward as compared to a narrow-band laser source. The broadband excitation of the sample will always excite all resonant features in the spectrum at the same time, hence a spectral separation of different material properties via the amplitude information may in some cases not be possible. For the latter examination, a narrow-band source is the better choice.

The field emitted from the photoconductive antenna is shown in the time- and the frequency-domain in figure 4.3.2. Please note that a chopper in the pump arm of the laser is used to detect the THz signal via a lock-in amplifier. The tip is not part of the setup at this point. The emitted electric field as a function of the time delay is shown in (a) and set via the delay stage. Before the photoconductive antenna is excited, no THz field is emitted. When the pulse hits the antenna, a large single-cycle THz pulse with a length of roughly 2 ps is emitted. It is followed by fluctuations which are most likely the result of water vapor absorption in the air. The spectrum of this pulse is shown in (b). The strongest emission occurs at frequencies in the range of few hundred GHz. The power loses several orders of magnitude at frequencies above 1 THz and reaches the noise floor at roughly 3 THz. The difference of the spectrum compared to the specifications of the commercial photoconductive antenna, where higher frequencies are reached, can be explained by many factors. Firstly, the THz radiation travels an unusual long distance before it reaches the detector. Absorption and scattering at water molecules in air can strongly change the spectrum. Secondly, as mentioned before, the emission of the THz radiation has a large opening angle and not all frequencies are emitted isotropically. When the emission is not fully collected by the lenses, certain frequencies can be suppressed. Thirdly, as the beam cannot be perfectly collimated due to the

[20] In this setup, an external Stanford Research Systems SR 830 Lock-In amplifier is applied which was connected to a LabView software to control the delay lines.

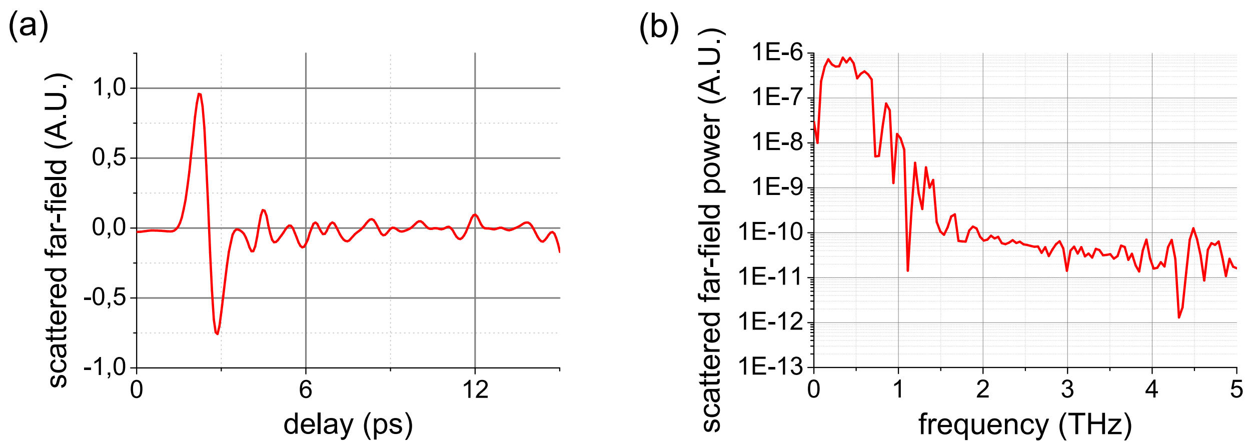

Figure 4.3.2.: *Emitted single-cycle THz pulse from the photoconductive antenna in the time-domain (a) and the frequency-domain (b). The sample used for this measurement is a gold sample with a tickness of 100 nm gold on a silicon wafer which should not influence the spectrum of the THz pulse. Please note the logarithmic scale in (b).*

large area emission, the focused THz beam at the ZnTe crystal will be far from a diffraction-limited focus and some frequencies will be more pronounced in certain areas of the focus than others. Please note that the far-field alignment of the setup has been carried out by maximizing the signal, the spectrum was only regarded after the alignment was finished. With other alignment, it might likely be possible to achieve a broader spectral response.

After the tip is included into the setup and approached to the gold sample, the chopper is removed from the setup and the scattered near-field signal is detected via higher-harmonic demodulation of the scattered signal. The setup is again aligned by maximizing the signal through the adjustment of the optical elements. The result is shown in figure 4.3.3. The scattered signal in the time-domain (a) does not show a single cycle anymore, but rather a multi-cycle pulse with a length of more than 5 ps and at least three oscillations before the signal vanishes in the noise floor. The signal-to-noise ratio is in general much lower since the near-field measurement signal is much smaller but the noise floor of the setup stays the same. The reduce of signal-to-noise ratio has been observed in other TDS near-field experiments as well [54–56]. The scattered near-field in the frequency-domain shows a strong change compared to the far-field emission of the antenna. Here, the peak frequency lies in the range of 2 to 2.5 THz and decays slower to smaller and larger frequencies. This is a result of the dipole-antenna enhancement of the tip discussed in section 3.4.3 for this 80 μm long AFM tip. A qualitative comparison of the experiment to the sim-

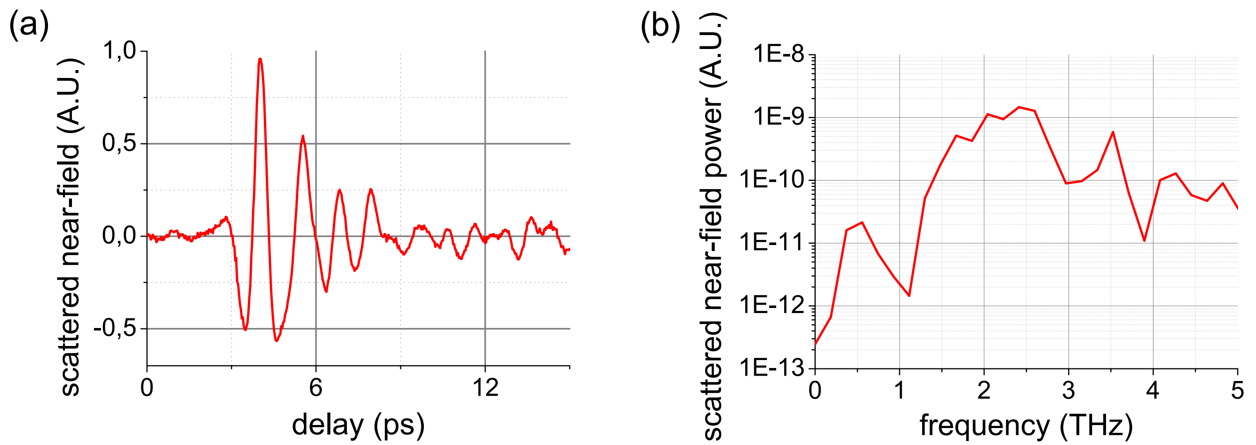

Figure 4.3.3.: *Emitted THz near-field pulse from the RMN tip in the time-domain (a) and the frequency-domain (b). The pulse is now a multicycle pulse with a broad resonance around 2 THz. Please note the logarithmic scale in (b).*

ulation cannot be performed due to too many uncertainties considering tip shape, experimental settings and other difficulties, still it is impressive that the most pronounced frequencies of the experiment and the simulation are in good agreement. This may change from tip to tip as they are etched individually and do not have the high reproducibility as offered by silicon-based ones. As the setup is aligned on a gold sample for each tip anyway, a far-field and near-field spectrum should be collected on this spectrally flat sample to allow for comparison of other materials for reference.

In order to demonstrate the capability of mapping material contrast on the nanometer scale with this setup, a Si/SiO_2 nanostructure has been chosen as a prove-of-principle sample. Here, the main peak of the near-field pulse in the time domain has been chosen for best signal-to-noise ratio in the spectrally integrated lateral 500x500 nm scan of the sample. The topography is shown in figure 4.3.4 (a), where a diagonal stripe is visible in between two areas of elevated material. The corresponding near-field response is shown in (b). Here, the diagonal stripe shows a much stronger response and can be identified as silicon with this measurement. The free charge carriers (p-doped with $n \approx 10^{15}$ cm^{-3} contribute to the THz near-field response as predicted in section 2.1 on the Drude model, whereas the SiO_2 is an insulator in the THz range and does not have a strong near-field response. This finding has been confirmed in other near-field experiments as well [48, 117]. The resolution cannot be determined with high accuracy due to low signal-to-noise ratio, however can be confirmed to be below 100 nm from the data. When compared to the better signal-to-noise ratio presented in similar experiments [55, 56, 117], the difference can

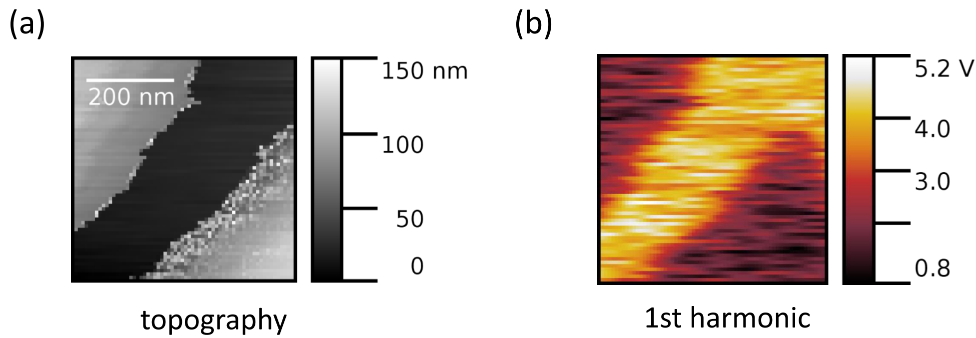

Figure 4.3.4.: *500 nm, 64x64 pixel broadband THz near-field scan of a Si/SiO$_2$ structure. (a) shows a stripe of Si surrounded by two areas of SiO$_2$. The near-field response is shown in (b), where Si has a much stronger response due to a doping concentration of $\approx 10^{15}$ cm^{-3}.*

be explained by the much lower sampling rate of 200 kHz in this setup.

Ultimately, the goal of this experimental work is to use TELBE instead of the PCA for probing with a high-power and narrow-band THz source. Unfortunately. the combination of this setup with the TELBE beamline has been unsuccessful when first tried. After a 24-hour shift with the TELBE beam, the post-mortem pulse analysis and data correction [140] of the signal performed by Thales de Oliveira did not show evidence of near-field interaction in between tip and sample. Most likely, the pulse-to-pulse stability of TELBE and the amplitude jitter is insufficient for an s-SNOM experiment and the correction accuracy is not high enough to push the noise level to a point where near-field signals can be deduced from the background. A second approach is planned for the end of March 2019, where the pulse-to-pulse stability of TELBE will be optimized for this experiment.

5. Sideband Results

Physics is like sex: sure, it may give some practical results, but that's not why we do it.

(Richard P. Feynman)

This chapter covers the systematic investigation of sidebands predicted in section 3.3 was performed with the setup described in section 4.2. To demonstrate the capabilities of the setup, a bulk germanium sample was chosen as a sample for proof-of-principle measurements. The properties of such a semiconductor are discussed in section 2.1. Below, experimental evidence of higher-order sidebands in a pump-probe s-SNOM setup is presented. To gain insight into fundamental aspects of such a system, parameter studies of physical and technical properties have been performed. The sub-diffraction resolution and the nanoscale material contrast compared to a gold island are presented as well. As a last step, the results will be compared to the theory.

5.1. Pump-induced Sidebands in Germanium

The main aim of this experiment is the confirmation of the predictions made in section 3.3, namely the observation of higher-order sidebands. For this, the setup as described in 4.2 has been utilized with a germanium sample[1], which electronic properties have been discussed in section 2.1. A CO_2 laser has been applied for probe and a Ti:Sa pulsed laser ($f_{rep} = 1$ kHz) to excite the sample. For probing the near-field, a *Nanosensors PPP-NCLPt,* $f \approx$ *160 kHz.* tip has been used.

For each parameter studied in this section at least two retract curves were obtained to ensure the reproducibility of each measurement. Overall, the influence of nine parameters such as the pump fluence, the harmonic of the near-field and sidebands, as well as the oscillation amplitude of the

[1] 1cm^2, single crystal, n-type, $\rho = 30\ \Omega$ cm

tip was investigated. Additionally, technical parameter settings of the employed Lock-In amplifier have been in the scope of this work. To present the characteristics of the system to the reader while keeping the length of this chapter reasonable, only some retract curves are presented in its entirety. Later on, the data is presented in the following way: For each curve, the sideband amplitude data of the first 10 nm above the sample surface is averaged. As the tip will be in this regime during lateral scans, this method gives a reasonable measure to compare the influence of the parameters. When a signal-to-noise ratio is calculated, this value is compared to the RMS noise of the same retract curve for the distance of 75-150 nm to the sample surface. In plots, the notation is chosen as follows for an unequivocal presentation: Whenever the near-field is demodulated it is noted with NF with additional information on the harmonic order, e.g. the first harmonic is labeled with NF1H, second harmonic with NF2H, and so on. When a certain sideband is demodulated, the harmonics of both the near-field and the sideband are given. The second sideband of the first harmonic is labeled NF1H SB2H and the fourth sideband of the third harmonic with NF3H SB4H.

Figure 5.1.1 (a) shows a set of typical near-field retract curves on germanium with a probe and pump laser power of 25 mW and 5 mW, respectively, for the first three harmonics of the cantilever oscillation. The second and third harmonic show the typical near-field decay with a decay length to half of its contact value of $\sim$ 15-20 nm. Compared to the measurements in the THz region, the first harmonic inhibits stronger far-field contribution for this measurement with the CO_2 laser ($\lambda = 10.6\ \mu$m). The signal decays and reaches the point of zero at around 35 nm distance of the sample before it rises again towards larger distances. This behaviour is a result of either a field minimum or maximum of the background roughly at distance. At this point the tip modulation does not change the overall scattered field, hence the modulation is not visible at the detector and the Lock-In amplifier detects a value of zero. The change in direction is associated with a 180 degree phasejump in the Lock-In phase (not shown in the figure). This is a common effect in the first harmonic with such low wavelengths and will induce artifacts during lateral scans as the background field is not constant. For this reason, the first harmonic is in many cases not recorded. With regard to this novel technique of sideband demodulation it is however fruitful to demodulate and analyse this signal to better understand the physical background of the sidebands' behavior as far-field effects may influence the higher harmonics as well if the effect is too large.

The corresponding first-order sidebands[2] are shown in (b). NF2H SB1H and NF3H SB1H show a

[2] The Lock-In amplifier settings for this measurement are τ_i=100 μs, 12 dB/octave filter for the SR 844 and the output was set to the amplitude R. For the SR830 an integration time of τ_i= 30 ms and a 12 dB/octave filter were used

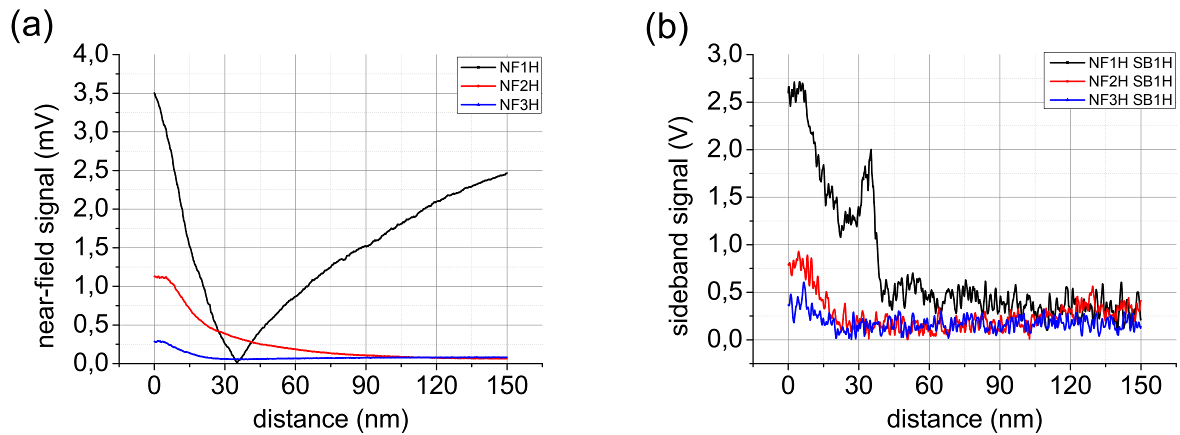

Figure 5.1.1.: *Retract curves with a pump power of 5 mW. The near-field signals of the first three harmonics are shown in (a) and were measured with an internal Lock-In amplifier of the R9. The corresponding first sidebands are shown in (b). Please note that the sideband signal is three orders of magnitude larger due to amplification of the external Lock-In amplifiers before it is fed into the data acquisition.*

typical near-field decay with a slightly short decay length as in (a). The signal-to-noise ratio of the sidebands is smaller than in the overall near-field signal, but is still large enough to perform a scan on the sample. The NF1H SB1H signal shows an artifacts in form of an unphysical peak at the distance of $\sim$ 35 nm. It is associated with the first harmonic reaching zero at this position. At this point, the cascaded Lock-In demodulation inherently produces this error. When the first harmonic reaches values close to zero, the signal stability is lost as the SR844 Lock-In amplifier is not delivering a stable near-field signal but rather a random noise floor, containing all frequencies - including 1 kHz. This directly couples into the second Lock-In demodulation and produces this peak without a physical meaning. For distances higher than 40 nm, the signal slowly decays and reaches the noise level for values above 120 nm. Assuming the decay length is associated with the lateral resolution of the s-SNOM, the resolution of NF2H SB1H and NF3H SB1H will be superior to the resolution of NF1H SB1H, however it is quite remarkable that NF1H SB1H shows a clear near-field characteristic although the corresponding NF1H signal suffers from far-field effects.

5.2. Fluence Dependence

The influence of the pump laser's power is investigated in steps of 1 mW average power up to the point when a critical heat impact to the AFM is reached ($P \approx 8$ mW)[3]. In the near-field signals of the second and the third harmonic (see figure 5.2.1 (a)), no significant influence is found for pump powers up to 7 mW. Although the induced charge carriers strongly increase the near-field response, the duty cycle of this effect is too low to induce a visible change in the overall near-field signal. At higher powers, the signals drops by $\sim$ 25 % as a result of the heat impact. At this point, the control loop becomes instable which leads to an effectively larger distance from the sample surface when the retract curves are obtained. As the near-field signal drops strongly with the distance, an unstable control loop couples quite strongly into the near-field signal. For the first harmonic, this effect is visible for high power and is even more pronounced. The data of the first harmonic below the power of 7 mW is not very stable either. Here, the data points follow a non-steady pattern which cannot be explained directly by theory and may be influenced by various artifacts. When the raw data is analysed, a shift of the point of $z = 0$ near-field by up to 10 nm can be identified, hence the maximum value at a distance of $z = 0$ nm exhibits a quite strong change for each measurement. As the data does not show a random pattern but rather follow a quite slow change (each point corresponds to the retract curves and needs more than a minute to collect and they were obtained in chronological order), a slow drift (e. g. change of the background field due to a thermal drift of the optical setup) is most likely the origin of the change in signal. The exact source of the drift is unclear, however, most drifts will simply change the amplitude and phase of the background field. As the supression of the background effect is much less efficient in the first harmonic than at higher harmonics, this effect is not visible for NF2H and NF3H.

The data for the corresponding first-order sidebands is presented in (b). For NF2H SB1H and NF3H SB1H a rise in amplitude is observed with the strongest gradient from 3 to 5 mW. For powers from 5 to 7 mW, a plateau is reached before the heat impact induces artifacts to the signal. The NF1H SB1H amplitude shows a linear behaviour with rising pump powers up to 7 mW of pump power. The rise in signal with increasing pump powers fits in general to the theory: As predicted by the Drude theory and the expected rise in the scattering cross-section in s-SNOM (see section 2.1), the strongest contribution to the near-field signal is expected when the permittivity of the sample becomes slightly negative. When the pump power is strong enough to excite the needed charge

[3] Due to low absorption of the probe laser, its heat impact can be neglected. Furthermore a cw heat impact does not influece the feedback control loop as much as a pulsed laser.

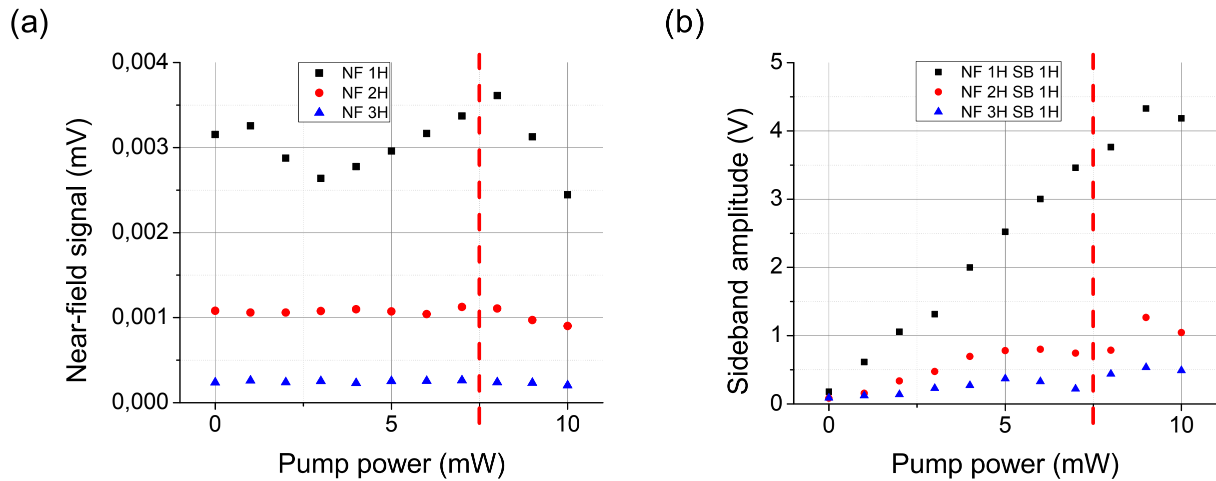

Figure 5.2.1.: *Pump power dependence of the near-field signal (a) and the corresponding first sidebands (b). The dashed red line indicates the power where the heat impact has a significant impact to the s-SNOM and the data becomes unreliable.*

carrier density, a further increase of power has a smaller contribution to the sideband signal. From the data, the pump power where the permittivity becomes negative is most probably in the range of 3-5 mW. A pump laser focus size of roughly 400 μm (FWHM) was determined using a beam profile camera[4], the probe laser focus was a bit larger with a size of roughly 600 μm. From this value and a indident angle of $\sim$ 55 degree, the fluence for a power of 4 mW is 0.55 mJ/cm^2. This value has the same order as observed in previous studies [47, 127] and is also in the same range as reports on macroscopic experiments [19, 142, 143] stating 1-15 mJ/cm^2.

5.3. Higher-order sidebands

The existence of the first sideband in s-SNOM and the application for nanoscale imaging has been proven before [47, 127]. Moreover, it is a fundamental question how many orders of the sideband can be detected as the theory discussing the sidebands does not consider the damping in the amplitude modulation. Here, the first sideband may still suffer from artifacts of the heat transfer into the tip-sample system e. g. via the mechanical oscillation of the tip. The heat which is induced by the laser in the sample changes the cantilever's temperature and its properties. This could lead to a periodic change in oscillation amplitude of the cantilever with a repetition rate of 1 kHz and

[4]Spricon Pyrocam III

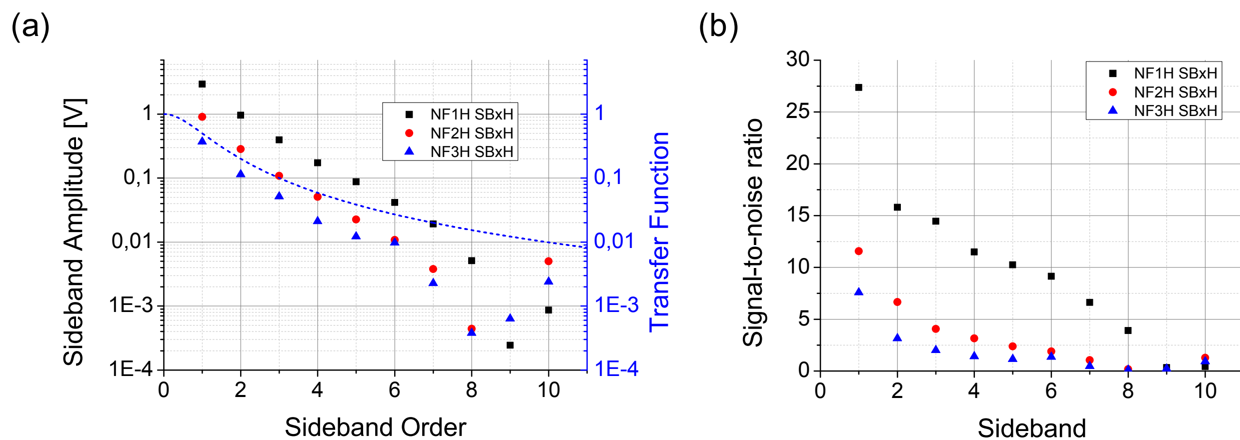

Figure 5.3.1.: *Sideband amplitude for higher sidebands of all three harmonics (a). Please note the logarithmic scale. The transfer function of the SR844 Lock-In amplifier is shown in the dashed line (blue). The calculated signal-to-noise ratio is depicted in (b).*

hence contributes to the signal although it is not related to the change of the sample properties. As the oscillation amplitude and the corresponding near-field signal have a linear relationship to amplitudes up to 200 nm [144], such a heat artifact cannot occur in higher-order sidebands. Hence, the higher-order sidebands are expected to be less affected by such artifacts.

A systematic investigation of higher sidebands is depicted in figure 5.3.1. (a) shows the sideband amplitude with logarithmic scale for the different orders of the sideband[5]. The sidebands amplitude of all three harmonics follow an exponential decay quite well up to the seventh order. Above that, the noise level is reached and the data becomes untrustworthy. For the first-order sideband, the value slightly deviates from the exponential behaviour and is too large in all three harmonics. The origin of this deviation cannot be verified with this dataset, but might be a result of the heat impact discussed above. The blue dashed line shows the transfer function of the SR844 Lock-In amplifier filter. The sideband amplitude shows a stronger decrease than the filter, hence the sidebands decay is of a physical origin and not an artifact of the demodulation technique. When the data is compared to the transfer function, the drop below the noise level for the 7th-order harmonic and higher is clearly visible. The signal-to-noise ratio for the measurements is calculated in (b). It is becoming smaller for higher-order sidebands and shows in general a rather linear behavior for the first harmonic and a faster decay for the second and the third

[5] The Lock-In amplifier settings for this measurement are τ_i=100 μs, 12 dB/octave filter for the SR 844 and τ_i=30 ms, 24 dB/octave filter for the SR 830

harmonic. The ratio is by far the largest for the first harmonic and sufficient up to the seventh sideband. For the second harmonic, the ratio is good enough for a scan up to the fourth sideband and for the third harmonic only the first and second sidebands are suitable. For the first sidebands, the signal-to-noise ratio is significantly larger than for higher-order sidebands. This is a result of the deviation in the first-order sideband, where an increased amplitude is observed. The noise level does not show such an increase and hence the calculated signal-to-noise ratio is larger. However, as the increased signal might be the result of the heat impact, it might carry unwanted contributions. To avoid the detection of these contributions, $m \geq 2$ should be preferred during measurements. In general, the detection of such high-order sidebands is quite remarkable and indicates a very small damping of the amplitude modulation.

5.4. Oscillation Amplitude

The oscillation amplitude of the cantilever is a parameter in s-SNOM which is commonly chosen to values which roughly matches distance where the near-field has decayed to almost zero. In most reports, the amplitude ranges around 60 nm, a quite large value for tapping mode AFM already. Some scientist increase this to values above 100 nm as larger near-field signals are found (e. g. [89]). In this range, the near-field amplitude has a linear dependence to the oscillation amplitude of the tip [144]. However, increasing the amplitude too much will cause severe artifacts in the s-SNOM measurement as the mechanical oscillation becomes anharmonic and inherently carries higher harmonics of the driving frequency [145]. Hence, higher harmonics not associated with near-field interactions will be found in s-SNOM signals [146]. To investigate the influence of the oscillation amplitude on the s-SNOM signals and the sidebands, the oscillation amplitude was increased while measuring the first three harmonics of the near-field and the corresponding first sidebands. Due to limited amount of time in the laboratory, this measurement could not be performed for higher sidebands. The dependency of the near-field and sideband amplitudes on

the oscillation amplitude is depicted in figure 5.4.1. The second and third near-field harmonics follow a linear behavior when the oscillation amplitude is increased from 40 to 120 nm. When being increased to 160 nm, the rise in amplitude becomes smaller as the additional distance covered by the tip is in the range where the near-field in close to zero. This behavior is quite typical for an s-SNOM experiment [144]. For the first harmonic, the signal rises by $\sim$20 % when the oscillation amplitude is increased from 40 to 60 nm, for larger values the signal is relatively stable. Depending on the alignment of the setup and the position of the tip at the sample, this might

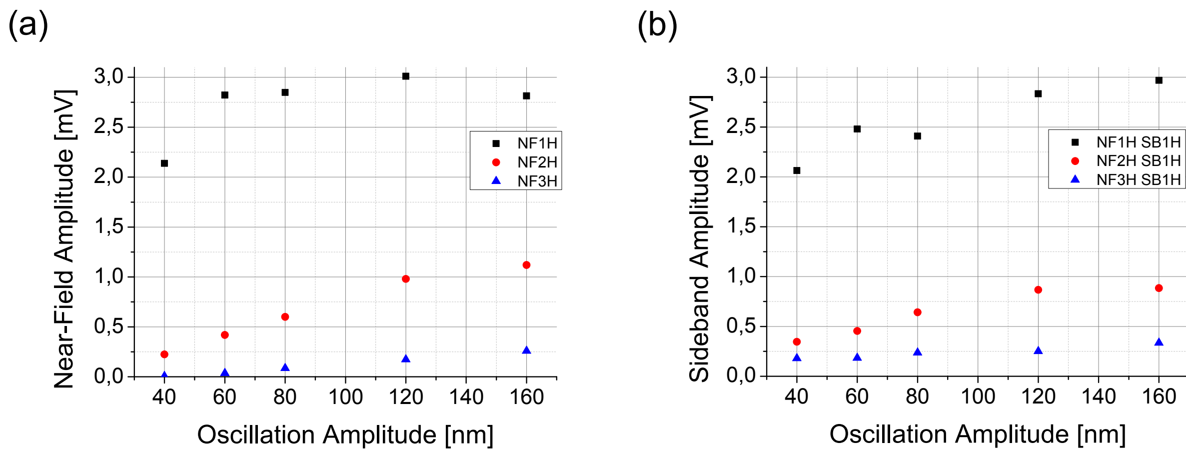

Figure 5.4.1.: *Near-field (a) and corresponding sideband (b) amplitudes for different oscillation amplitudes of the tip. The Settings for the SR 844 were chosen to an integration time of 100 μs, a 12 dB filter and a sensitivity of 30 mV/V. For the SR 830, the integration is 30 ms and the filter is 24 dB.*

change quite drastically, as it strongly depends on the background field. Please note that this measurement was performed several hours after the measurements of the higher-order sidebands summarized in figure 5.3.1. During this time, the tip was worn off a bit, leading to a larger effective tip radius and a slightly enhanced near-field signal. The sideband data is shown in figure 5.4.1 (b). The NF2H SB1H and NF3H SB1H amplitudes in general follow the same behavior as the corresponding near-field signals. They rise with the oscillation amplitude up to 120 nm, although the slope is a bit smaller. For amplitudes above 120 nm, the rise decreases. For NF1H SB1H there is a rising trend towards larger amplitudes, but the fluctuations of the NF1H data is visible in the sideband as well.

In order to find a suitable oscillation amplitude for such measurements, the signal-to-noise ratio is determined and depicted in figure 5.4.2. (a) shows the data for the near-field signals. The values for the first harmonic decrease constantly from $\sim$25 to $\sim$15 when the oscillation amplitude is increased following a rather linear behaviour. For the second harmonic, the value decreases as well. However, a strongly enlarged signal-to-noise ratio is calculated. The reason is the enhanced near-field amplitude while the noise floor does not change from 40 to 60 nm amplitude. Most likely the noise floor is limited by the Lock-In amplifier itself for small amplitudes, while it is limited by the above discussed artifacts from the higher harmonic demodulation and/or anharmonities from the cantilever oscillation. To investigate this further, more data points in the range from 40 to 80 nm would be needed. For the third harmonic, the signal-to-noise ratio rises rather linearly

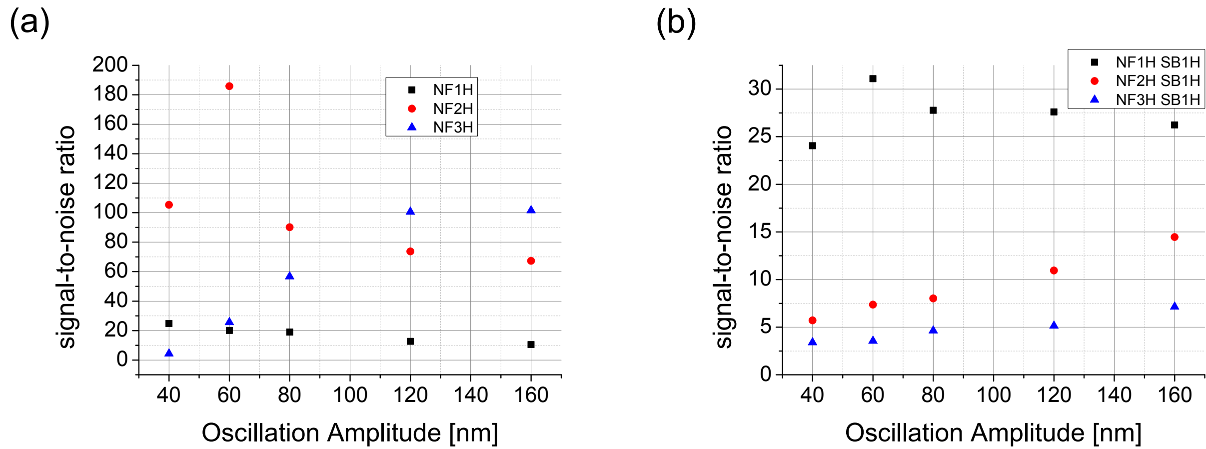

Figure 5.4.2.: *Calculated near-field (a) and corresponding sideband (b) signal-to-noise ratios for different oscillation amplitudes of the tip.*

up to 120 nm and does not rise further afterwards. The corresponding data for the first order sidebands is plotted in (b). Here, both NF2H SB1H and NF3H SB1H show a strong rise towards higher amplitudes and do not saturate for amplitudes above 120 nm. The NF1H SB1H has by far the largest amplitude and does not show a clear trend.

In conclusion, the sweet spot of oscillation amplitude lies in the range of 80 to 120 nm. Here, both the absolute signal strengths and the signal-to-noise ratios of all signals is in a suitable regime. Only NF1H has quite a comparably small SNR, but this signal carries too many artifacts for a proper near-field scan and is usually neglected for this purpose. The achievable resolution in a scan is a bit better for smaller amplitudes but will still be fine for an amplitude around 100 nm.

5.5. Technical Aspects of the Sideband Demodulation

In this section the influence of the Lock-In amplifier settings on the sideband signals is examined. As the cascaded method for the sideband demodulation is new in this field, this examination enables the optimization of setting and significantly improves the quality of the data.

The influence of the **SR 844 sensitivity** is evaluated in table 5.1. Here, largest sensitivity was chosen in a way that the cascaded SR830 Lock-In amplifier could still receive the signal without any processing error, meaning that both the input and output ranges were not fully exploited. For the evaluation, the NF2H SB1H signal was taken into account. The signal-to-noise ratio is best for a sensitivity of 30 mV/V (10 % of the full range exploited), however the small change when using 10 mV/V could arise from statistical noise fluctuations during the measurement. When the

	10 mV/V	**30 mV/V**	**100 mV/V**
Signal [V]	2.64	0.95	0.222
Noise [V]	0.22	0.07	0.026
SNR	11.56	12.15	8.65

Table 5.1.: *Calculated signal-to-noise ratio for different sensitivities of the SR844 Lock-In amplifier. For this measurement an integration time of 100 μs and a filter of 12 dB/octave has been used for the SR844 Lock-In amplifier. The SR830 was set to an integration time of 30 ms and a filter of 24 dB/octave. The pump power of this measurement is 5 mW and the tip had an oscillation amplitude of 160 nm.*

output range of the SR 844 is exploited only by 2%, the signal-to-noise ratio drops down by almost 30 %. The conclusion of this measurement can be transferred to other harmonics and sidebands without hesitation.

The influence of the **SR844 filter** is analysed in table 5.2. Here, all possible filter settings are tested. The trend shows larger signal-to-noise ratios for steeper filters when the NF2H SB1H is demodulated. This measurement cannot be transferred to the measurement of higher sidebands, as the filter will determine how strongly signals around the main peak of the harmonic are cut off. I.e., a 24 dB/octave filter will only pass 4% of the second sideband and 1% of the third sideband. For the measurements of the higher-order sidebands described above, the filter was set to 12 dB/octave, which passes signals up to the 7th sideband. A smaller filter may pass even more sidebands. To gain a full picture, the measurement of higher sidebands should be repeated with different filter settings.

The last technical setting of the **SR 844** investigated is the **integration time** τ_i of the Lock-In amplifier. Again, NF2HSB1H is measured while τ_i is changed. The obtained signal-to-noise ratios are given in table 5.3. A massive drop by a factor of more than 400 in the signal amplitude passed by the SR844 is observed when the integration time becomes larger than 1 ms and hence larger than the period of the excitation laser. The optimal integration time is 300 μs, were the integration is long enough to suppress the noise, but still easily passes the 1 kHz modulation. For the investigation of higher-order sidebands, the modulation $n\cdot$ 1 kHz becomes faster and hance the optimal integration time should tend towards smaller values as the modulation.

As the SR 830 Lock-In amplifier is the last device in the cascade, its settings should influence the

	0 dB/oct	**6 dB/oct**	**12 dB/oct**	**18 dB/oct**	**24 dB/oct**
Signal [V]	7.65	6.19	5.61	4.60	4.53
Noise [V]	0.70	0.66	0.51	0.42	0.35
SNR	10.97	9.38	10.99	11.02	12.89

Table 5.2.: *Calculated signal-to-noise ratio for different filters of the SR844 Lock-In amplifier. For this measurement an integration time of 100 μs and a sensitivity of 10 mV/V has been used for the SR844 Lock-In amplifier. The SR 830 has an integration time of 30 ms, a filter of 24 dB/octave and a sensitivity of 10 mV/V. The pump power of this measurement is 5 mW.*

	100 μs	**300 μs**	**1 ms**	**3 ms**
Signal [V]	2.761	0.966	0.076	0.006
Noise [V]	0.246	0.061	0.005	0.001
SNR	11.22	15.72	14.31	7.029

Table 5.3.: *Calculated signal-to-noise ratio for different integration times τ_i of the SR844 Lock-In amplifier. For this measurement a filter of 12 dB/octave and and a sensitivity of 10 mV/V is used for the SR844 Lock-In amplifier. The SR 830 had an integration time of 30 ms, a filter of 24 dB/octave and a sensitivity of 20 mV/V. The pump power of this measurement was 5 mW.*

	1 ms	3 ms	10 ms	30 ms	100 ms	300 ms
Signal [V]	1.111	0.969	0.947	0.895	1.001	0.818
Noise [V]	0.176	0.156	0.105	0.076	0.049	0.025
SNR	6.284	6.223	9.034	11.74	20.71	32.74

Table 5.4.: *Calculated signal-to-noise ratio for different integration times τ_i of the SR830 Lock-In amplifier. For this measurement a filter of 12 dB/octave, an integration time of 100 μs and a sensitivity of 10 mV/V is used for the SR844 Lock-In amplifier. The SR 830 has a filter of 24 dB/octave. The pump power of this measurement is 5 mW.*

signal-to-noise ratio as expected for any Lock-In amplifier (as decribed in [124]). Firstly, steep filters will enlarge the signal-to-noise ratio at the cost of quick onset of the signal when the modulation is started. Secondly, long integration times will enlarge the signal-to-noise ratio at the cost of measurement speed. The obtained values for different **SR 830 integration times** are given in table 5.4. Here, the SNR clearly rises for larger integration times as the random noise is suppressed more efficiently. The best trade-off in between SNR and integration time needs to be found for each experiment. For many measurements, an integration time of 10 ms will be sufficient for a sideband scan. This value is in the typical regime used for imaging in near-field microscopy and will not decrease the scanning speed. If a larger SNR is needed, an integration time above 30 ms needs to be chosen and will increase the scanning time to more than an hour for a 128x128 pixel scan. For such long scans, the AFM setup must be very stable against drifts and should be in thermal equilibrium before the scan is started.

As a last technical parameter, the **SR 830 filter** is regarded. When it is changed, a slight change towards smaller signals and lower noise, is observed. When the signal-to-noise ratio is calculated, no clear trend can be observed. The results are given in table 5.5.

5.6. Nano-optic Material Contrast

The ultimate proof that the measured sidebands are related to the optical information of the near-field is a near-field scan with sub-diffraction resolution. Here, a germanium crystal with a gold nanostructure on top is chosen as a demonstration system. In the ideal case, the sidebands should not appear on the gold areas and only on the germanium crystal, as the charge carrier density is increased by the excitation laser only in the latter case. Figure 5.6.1 (a) shows the

	6 dB	12 dB	18 dB	24 dB
Signal [V]	3.78	2.31	2.58	2.58
Noise [V]	0.32	0.26	0.26	0.22
SNR	11.84	8.78	9.99	11.54

Table 5.5.: *Calculated signal-to-noise ratio for different filter settings of the SR830 Lock-In amplifier. For this measurement a filter of 12 dB/octave, an integration time of 100 μs and a sensitivity of 10 mV/V is used for the SR844 Lock-In amplifier. The SR 830 has an integration time of 30 ms. The pump power of this measurement is 5 mW.*

topography of the sample. A undefined gold patch with a size of roughly 10x5 μm and a height of 50 nm was fabricated via lithography on a single crystal germanium substrate[6]. At the edges, a thin area of elevated topography is observed. The top row of the figure moreover shows the near-field amplitude (b) of the second harmonic and the corresponding first sideband (c) without the excitation laser. In the near-field, a clear material contrast in between the gold patch and the substrate can be observed. The dark rim around the gold area corresponds to the elevated areas in the topography. From this data, it can be concluded that the elevated area is most likely residual resin from the fabrication process. The scan is not detailed enough to determine the lower limit of the near-field resolution, however an upper limit of $\sim$ 150 nm can be stated. In the sideband data, there are artifacts visible which need to be addressed before evaluating the scan with the excitation laser, such as a slightly more noisy and increased signal on the gold and clearly elevated signals at the edges. All data shown before were spectra collected at a single point of the sample, hence the p-i control loop of the AFM did not need to compensate for the fast changes in the topography which occur during a scan. While scanning, the feedback loop will compensate for any fluctuation in the topography. As the sample is not completely flat and the control loop does not compensate perfectly, this will in general induce changes in the actual tip-sample distance containing all frequencies, including 1 kHz. As the near-field is very sensitive to the tip-sample distance, this 1 kHz signal will directly couple into the near-field and sideband demodulation. The gold film is much rougher than the germanium and much more vulnerable to this artifact. By adjusting the control loop, the artifact could be minimized, however a slightly larger signal on gold still still remains. The largest artifacts can be observed at the edges themselves, where the control loop needs to compensate strongly for the topography. The scan direction is left to right

[6]The characteristics of this semiconductor are discussed in section 2.1

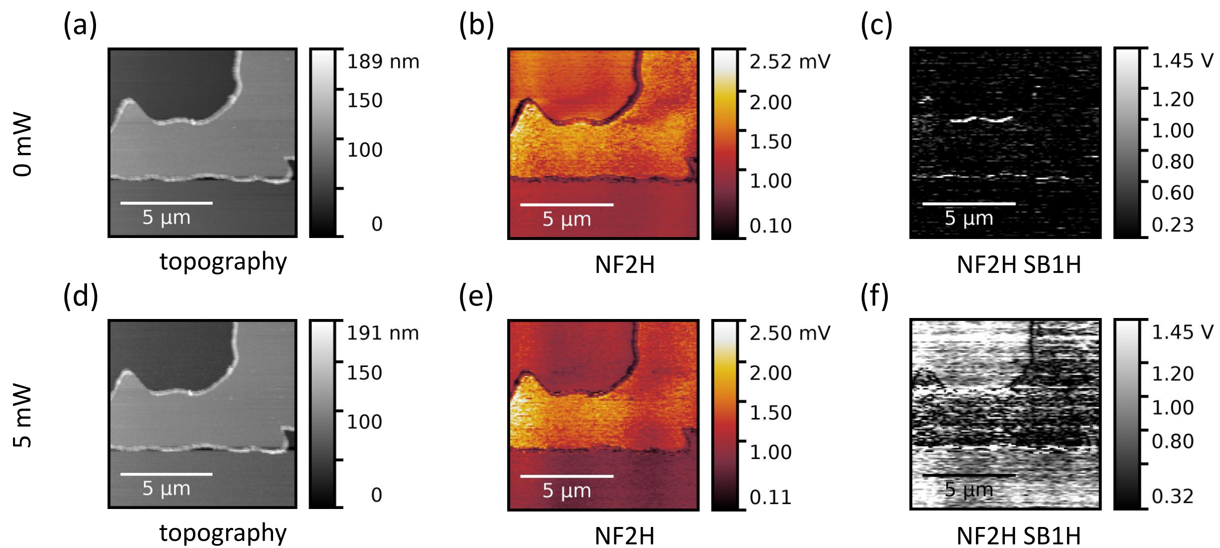

Figure 5.6.1.: *Nano-optical investigation of a gold patch on a germanium substrate. The upper row shows the measurement without the excitation laser, the lower row with an excitation power of 5 mW. To suppress background interference from the CO_2 probe laser, the 1d-FFT filter of Gwyddion [147] was applied in (b), (c), (e) and (f). For the measurement an integration time of 100 μs, a filter of 12 dB and a sensitivity of 30 mV/V was used at the SR 844. At the SR 830 the integration time was chosen to 30 ms, the filter was chosen to 12 dB and the oscillation amplitude was 100 nm.*

in this picture explaining why the edges along the scan direction are more sensitive to this artifact than the edges perpendicular to it.

The lower row shows the scan with a pump exciation power of 5 mW. Both the topography (d) and the second harmonic in the near-field (e) show no significant change. In the sideband (f), a clear material contrast in between the gold and the germanium can be observed. The germanium appears bright, while the gold area is darker. Additionally, the sideband phase is only stable on the germanium (not shown). The resolution of the sideband can be estimated to a value below 150 nm and is smaller than the diffraction limit of both the CO_2 probe and the Ti:Sa excitation laser. The signal on the gold rises to a value of few hundred mV, which might be related to heat induced artifacts in the first sideband as discussed above. Unfortunately, the limited lab time did not allow for a lateral scan of this structure with the demodulation of a higher-order sideband. A second explanation for the enlarged signal on gold can be a unwanted change of the background field due to the mascroscopic change of the reflectivity of the germanium substrate which will directly

induce sidebands in the full data of the scan. To exclude this artifact an inverted sample (almost completely coated with gold) would be a good candidate. For such a sample, the background field would not change under the influence of the background laser.

Although the scan nicely demonstrates the capabilities on nano-optic investigations of transient states, the contrast in this measurement is much lower than expected (an artifact- and noise-free measurement should show a contrast of infinity, as discussed in section 3.3) and is not in the same range as shown in previous experiments [47, 127]. One possibility could be the increased impact of heat in this experiment. While a Nd:YAG laser ($\lambda = 1064$ nm) has been used in [47, 127], this experiment was performed with a Ti:Sa short pulse laser ($\lambda \approx 800$ nm). The Ti:Sa laser has a stronger absorption coefficient (see figure 2.1.6) and more photons are absorbed in a smaller volume. Additionally, the laser in this experiment has more energy per photon which needs to be transferred during absorption. As this energy is mostly transferred to phonons, the heat impact into the system is larger than compared to the Nd:YAG experiment.

6. Field-Effect Transistors

> We choose to go to the Moon! We choose to go to the Moon, not because it is easy, but because it is hard; because that goal will serve to organize and measure the best of our energies and skills, because that challenge is one that we are willing to accept, one we are unwilling to postpone, and one we intend to win.
>
> *(John F. Kennedy)*

The second experiment featured in this section is the systematic investigation of plasma waves in graphene field-effect transistors. Many of such devices have been investigated with THz near-field microscopy and other complementary scanning probe microscopy techniques during the author's PhD time. Two device are analysed in full detail in this chapter. For this, the theory of s-SNOM and higher-harmonic demodulation of chapter 3.2.1 is combined with the theory of plasma waves in field-effect transistors presented in section 2.2. The experimental results are compared to the predictions of this theory and compared to macroscopic characterizations of the devices.

In section 2.2.1 the fascinating physical phenomena occurring in a two-dimensional electron fluid of a field-effect transitor (FET) are discussed. As the electron fluid in a conventional FET is buried under the gate electrode, the possibilities for investigation were very limited until the recent years. With the implementation of graphene in FETs, this limitation has been overcome. The graphene can be processed without a capping layer and now allows for direct access to the electron fluid.

The investigation of graphene field-effect transistors (GFETs) was proposed by Prof. Dr. Hartmut Roskos and Dr. Amin Soltani (Goethe-Universität Frankfurt / Physikalisches Institut) during the author's PhD thesis and carried out in close collaboration. Additionally to this thesis, the results are discussed in a publication which is close to submission. The GFETs were supplied by the group of Prof. Dr. Jan Stake and fabricated by Marlene Bonmann (Chalmers University of Technology, Sweden / Microtechnology and Nanoscience, Terahertz and Millimeter Wave Laboratory).

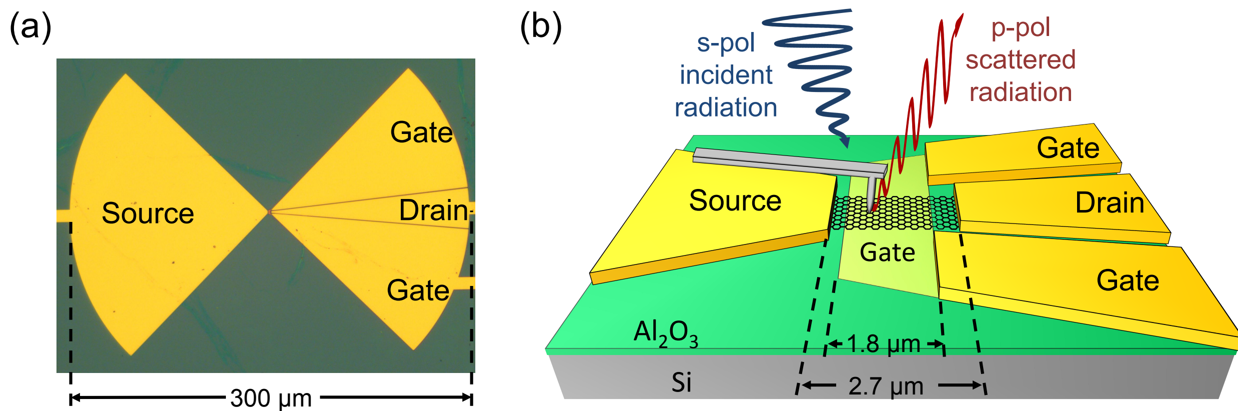

Figure 6.1.1.: *A top view optical microscope image of the GFet device is shown in (a). A bow-tie antenna (Au) with a size of 300 μm is designed to also include the transistors' connections of source, drain and gate. A schematic illustation of the device and the near-field measurement is shown in (b).*

6.1. Device Design

The devices in general consist of two large bowtie antennas with the FET located in the middle of them. The connections for source, drain and gate electrodes are established through the bow-tie antennas which are connected to wires enabling the connection of source measurement units (see microscope picture in (a)). The antenna has a radius of 300 μm for efficient incoupling of radiation from a few hundred GHz up to a few THz. A schematic close-up drawing of the transistor and the near-field measurement is depicted in (b). The device consists of several layers. The top layers are the gold source and drain electrodes with a height of 150 nm and the gate electrode with a height of 300 nm. The channel in between drain and source is connected through a graphene sheet with a length of 2.7 μm . To control the electric potential of the graphene, the gate electrode is partially convering the area beyond the channel (called bottom gate electrode during the further reading). The bottom gate has been implemented with sloped edges into the structure to achieve a smooth transition from the area with the electrode below the graphene (gated region) to the area without the gate electrode (ungated region). Without this feature, the graphene layer would presumably exhibit strong mechanical stress on a sharp edge. As some of the devices had breakdowns during the first FEL shifts, this potential origin for the breakdown was removed together with others. TO control the charge carrier density of the channel, the gate electrode is separated from the channel with a a layer of 25 nm thick Al_2O_3, which also prevents a current flow in between the gate electrode and the graphene. Although not visible in the illus-

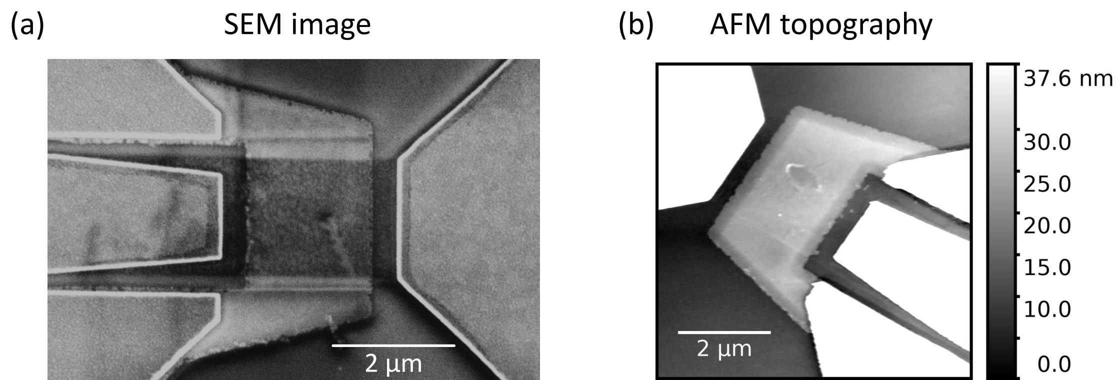

Figure 6.1.2.: *SEM (a) and AFM (b) images of typical samples investigated. Both show the high quality of the device, but also some residuals from the sample fabrication. The adjustment of the greyscale in (b) was chosen to enhance the visibility of the bottom gate. Please note the SEM image and AFM scan were performed on two different devices.*

tration, the gate electrode on top and beyond the Al_2O_3 are connected. The substrate is a 280 µm thick silicon wafer.

An overview of the structure is shown in figure 6.1.2. (a) features a typical SEM image of the device. The top electrodes have a very good quality with sharp edges and a uniform surface. The edges of the bottom gate electrode are less well defined and have some holes and cracks. The graphene is easily observed as the thick dark stripe connecting the source and drain. It is more grainy than the other areas and also shows a 1 μm long structure in the bottom right corner of the graphene. An AFM measurement of another device is depicted in (b). Although the resolution is clearly worse than in the SEM image, the topography is nicely mapped. Here, the scale is adjusted to a the point where the sloped edges of the bottom gate electrode and the slightly grainy graphene layer become visible. In this scan a non-perfect sample processing is visible in the area of the graphene, where an elliptic feature is visible in the topography. This feature is a result of an imperfect sample fabrication process and could be a hole in the graphene layer or a result of a non-uniform Al_2O_3 layer. Such fluctuations from the fabrication process can strongly disturbe the s-SNOM response in the channel and must be taken into account carefully. Some samples were contaminated to a degree where quantitative analysis of the near-field measurements becomes untrustworthy.

6.2. Data Analysis

In order to analyze the collected data and deduce quantitative values, the theory of near-field detection and the prediction made by Dyakonov and Shur need to be combined. During the collaboration on this topic, several methods were proposed by Dr. Amin Soltani until the collaborators decided to choose the method presented below for the analysis and fitting of the data.

The near field can in general be split up into two parts $E_{nf} = E_m + E_{pl}$, where E_m represents the field attributed to the material and E_{pl} the plasma wave. The total field present at the sample is then

$$E_{total} = E_b e^{i\phi_b} + E_m + \underbrace{E_1 e^{-i(k'_{pl}x+\phi_{pl})} e^{-k''_{pl}x}}_{\text{plasma wave}}. \tag{6.2.1}$$

Here, E_b is the background field which accounts for scattering at the GFET structure,the sample, tip and the AFM. The plasma wave equation is taken from section 2.2. As we consider only the first order of the plasma wave, k_{pl} is used as the plasma wave vector instead of k_0 (see equation 2.2.17). As the coefficients C of the plasma wave equation cannot determined in absolute measures in our experiment anyway, they are represented by E_1 in this calculation. The detector in our experiment detects intensities only, thus any contribution with $|e^{i\omega t}| = 1$ and can be neglected. The intensity at the detector reads

$$\begin{aligned} I_{total} =& E_{total} \cdot E^*_{total} \\ =& \Big(E_b e^{i\phi_b} + E_m + E_1 e^{-i(k'_{pl}x+\phi_{pl})} e^{-k''_{pl}x} \Big) \\ & \cdot \Big(E_b e^{-i\phi_b} + E_m + E_1 e^{i(k'_{pl}x+\phi_{pl})} e^{-k''_{pl}x} \Big) \\ =& E_b^2 + E_b E_m e^{i\phi_b} + E_b E_1 e^{i\phi_b} e^{i(k'_{pl}x+\phi_{pl})} e^{-k''_{pl}x} \\ & + E_m E_b e^{-i\phi_b} + E_m E_1 e^{i(k'_{pl}x+\phi_{pl})} e^{-k''_{pl}x} \\ & + E_m^2 + E_1 E_b e^{-i\phi_b} e^{-i(k'_{pl}x+\phi_{pl})} e^{-k''_{pl}x} \\ & + E_m E_1 e^{-i(k'_{pl}x+\phi_{pl})} e^{-k''_{pl}x} + E_1^2 e^{-2k''_{pl}x}. \end{aligned} \tag{6.2.2}$$

Due to higher-harmonic demodulation (see section 3.2) the term containing only E_b can be neglected. As these background fields are much larger than E_m or E_1, only mixed terms of back-

ground and near-field remain [118]. Hence the detected intensity reads

$$I_{total} \propto 2E_m E_b \cos(\phi_b) + E_1 e^{-k''_{pl}x} 2E_b \cos(k'_{pl}x + \phi_{pl} + \phi_b). \tag{6.2.3}$$

For the sake of simplicitiy, ϕ_b can assumed to be constant in the gated region, i.e. $\phi_b = 0$ leading to the final equation of

$$I_{total} \propto 2E_b \cdot \Big(E_m + E_1 e^{-k''_{pl}x} \cos(k'_{pl}x + \phi_{pl})\Big). \tag{6.2.4}$$

Here, k'_{pl} and k''_{pl} depend on the plasma wave velocity s which depends on the applied gate voltage. They are given by

$$k'_{pl} = \frac{\omega}{s}\sqrt{\left(\frac{(1+\omega^{-2}\tau^{-2})^{1/2}+1}{2}\right)}, \tag{6.2.5}$$

$$k''_{pl} = \frac{\omega}{s}\sqrt{\left(\frac{(1+\omega^{-2}\tau^{-2})^{1/2}-1}{2}\right)} \quad \text{and} \tag{6.2.6}$$

$$s = \sqrt{\frac{eU_0}{m_e^*}}. \tag{6.2.7}$$

Hence, a change of the gate voltage should result in a change of the deduced plasma wave velocity. For this reason, linescans along the channel were repeated with different gate voltages.

6.3. Near-Field Overview Scans

The GFETs were investigated with the s-SNOM setup as described in section 4.1 with a FEL wavelength of $\lambda = 150\ \mu$m[1]. For characterization of the samples, both overview near-field scans and linescans along the channel were performed. To achieve the best incoupling efficiency of the THz field to the antenna, the incoupling field is mostly polarized along the antenna (s-polarized, blue). The s-SNOM tip scatters the near-field into the InSb detector. With a p-polarized filter is mounted in front of the detector, only the p-polarized part of the scattered light is shown.

Many of such devices were investigated with the FEL in between 2017 and 2019 with both silicon-based (PPP) and full-platinum (RMN) tips[2]. Below, overview near-field scans are shown for two

[1]The experiment was also tried unsuccessfully at 230 μm and at 75 μm. For the latter case, the device antenna has an insufficient incoupling efficiency to induce a plasma wave. For the former case the reason is unknown.

[2]Nanosensors PPP-NCLPt and Rocky Mountain Nanotechnology 25PtIr200B-H

devices of the same device, GT1 and JT1 at a frequency of 2 THz are presented and discussed.

An overview near-field scan of the GT1 device with a PPP tip is shown in figure 6.3.1. The topography is shown in (a). The tip radius is obviously much larger than in the measurement shown in figure 6.1.2 due to tip wear off. The corresponding mechanical phase of the cantilever is shown in (b). The phase is changed especially at edges of the structure as the tip exhibits stronger interaction with the sample in this area and consequently amplitude and phase of the cantilever are shifted (see appendix A.1). In the first harmonic (c), the different areas are clearly visible. The bottom gate region is clearly the brightest and is relatively uniform. Although the gold is covered by a 20 nm Al_2O_3 layer, the near-field is probed through this transparent layer [148]. This commonly known effect has been utilized in several near-field experiments [56, 149]. Although previous studies show a slightly stronger response of graphene with broadband THz radiation [89], the graphene is not directly visible at this wavelength. While the drain and the top gate electrodes show a response in the same range, the source side has a much lower near-field amplitude. As the cantilever is located over the source side for all measurements shown here, we attribute this to either shadowing effects of the cantilever or to the asymmetry of the tip shaft leading to a change in illumination of the antennas. Two facts support this thesis: Firstly, for some measurements, the sample was rotated by 180 degree. In this case, the source side showed a much stronger response than drain and gate. Such a measurement is shown in figure 6.3.3. Secondly, when this measurement was repeated with the RMN tip, the difference in amplitude was much smaller (see figure 6.3.2). Compared to the PPP tip, the RMN is more symmetric and has a much longer tip shaft and thus the cantilever is located further above the sample and will create less shadowing effect on the source antenna. At all the edges of the device, the signal is reduced due to scattering. This artifact is well known in s-SNOM but must be taken into account carefully when the data is analysed.

The second and third harmonic near-field signals are shown in (d). Here, two features in the bottom gate region become visible: Thin stripes of enhanced signal at the upper and lower end of this region, as well as enhanced signal at the source side. The former could be related to scattering at the edge, but should in this case be most pronounced in the first harmonic signal and not in the second harmonic. In the third harmonic, the effect is not visible anymore, but the noise level might simply be too high there to observe such a small artifact. More likely, the origin for this stripe could be an artifact from the higher-harmonic demodulation. When the mechanical phase of the cantilever is analysed, instable conditions are found around the same positions where the stripes occur. When the mechanical phase becomes unstable, the Lock-In amplifiers for higher-

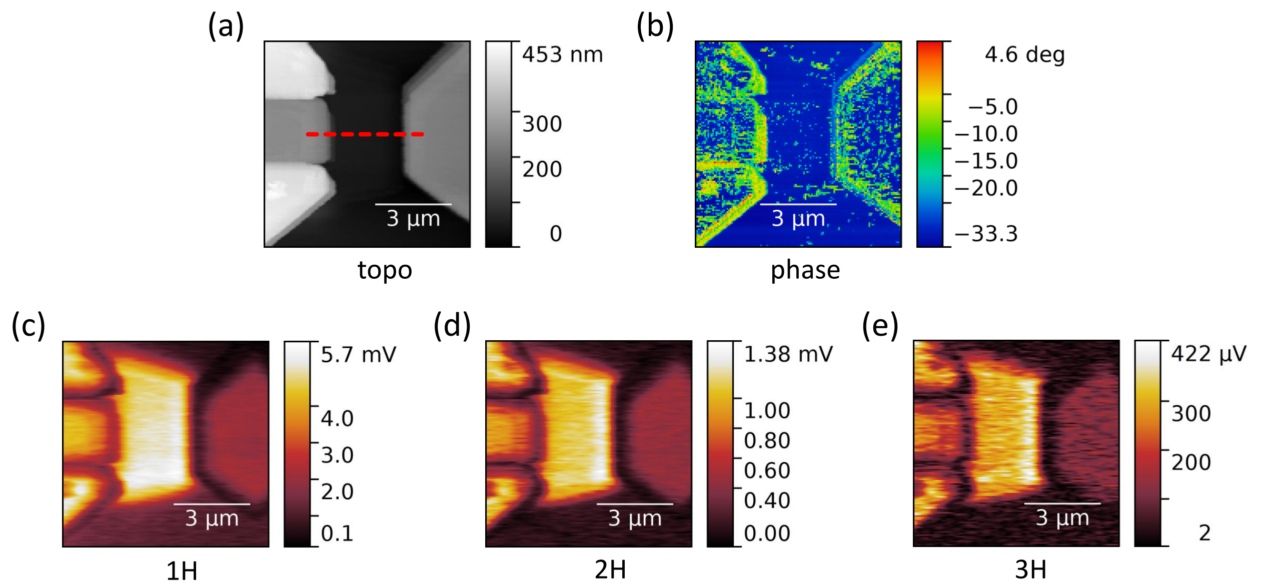

Figure 6.3.1.: *8x8 μm near-field scan of the GT1 device with a PPP tip. (a) shows the topography, (b) the meachnical phase of the cantilever. The integration time of the internal R9 Lock-In amplifiers is 100 ms with a filter of 48 dB. First, second and third harmonic are presented in (c), (d) and (e), respectively. The dashed red line marks the position where linescans are performed for quantitative analysis of the data.*

harmonic demodulation can produce artifacts as they compare the detector signals to the AC drive voltage of the AFM controller. When the driving phase does not have a stable relationship to the actual phase of the cantilever and it changes faster than the Lock-In amplifier's integration time, this unfortunately causes artifacts in the near-field measurement. These artifacts might be shifted a bit to the left side, as the Lock-In integration of 100 ms will cause a delay. In this scan, the scanning direction is from right to left. The artifact is clearly visible in the color scale, but is only 5 % larger than the amplitude at the bottom gate region and hence is not very large compared to the strongest feature of the measurement, which is described in the following. The brightest feature in this measurement is a clearly enhanced signal at the source side of the bottom gate region. This signal might be attributed to the plasma wave predicted in section 2.2 and will be analysed further below in section 6.4. The length of this feature is extended slightly further up and down than expected by the area of graphene, however the resolution of the tip is clearly not very good.

An overview scan of the JT1 device with identical geometry obtained with the RMN tip is shown in

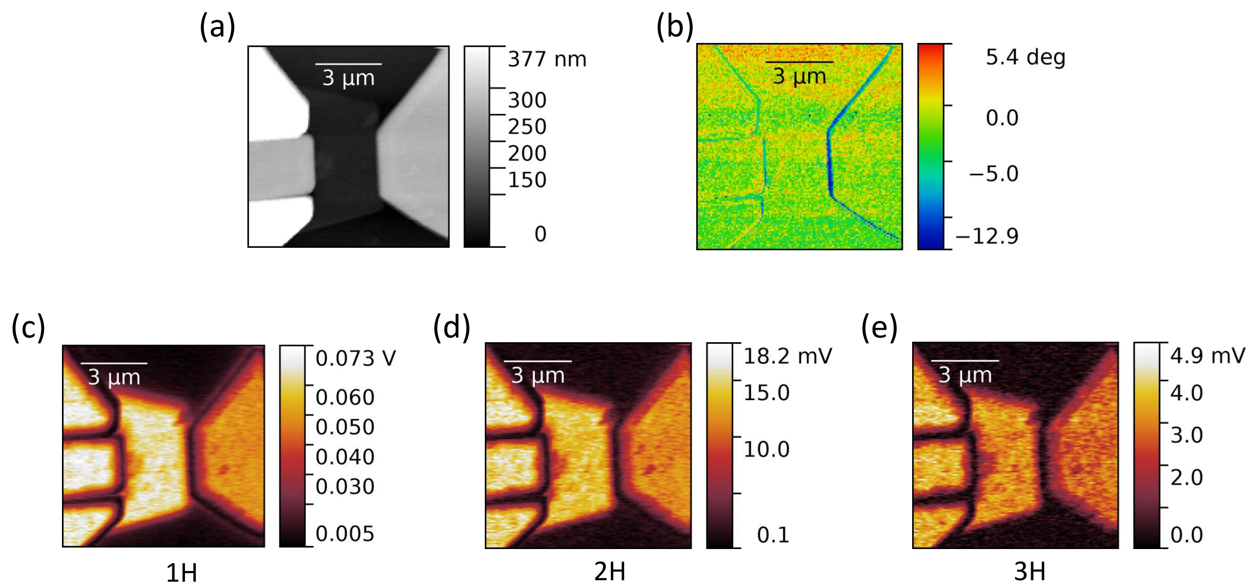

Figure 6.3.2.: *9x9 μm near-field scan of the JT1 with a RMN tip. The integration time of the internal R9 Lock-In amplifiers is 50 ms with a filter of 48 dB. First, second and third harmonic are presented in (a), (b) and (c).*

figure 6.3.2 for comparison. The topography is similar to the devices described before and shown in (a). In the channel region some dirt particles are visible. The mechanical phase is shown in (b). Compared to the measurement with the PPP tip, the phase is much more stable and only changes at the edges of the electrodes. (c) shows the first harmonic data of the near-field. Here, the difference in between the signals of the source and drain electrodes is much smaller. The signal-to-noise ratio is much larger although the integration time of the Lock-In amplifier was reduced to 50 ms. This matches the conclusion made in section 4.1, where a better near-field response of the RMN tip in the THz region is expected. The signal also carries less artifacts compared to the measurement with the PPP tip, e. g. the edge-effects are observed at much smaller length scales. The background suppression works better in this measurement, which can be observed at the low signal level of the Al_2O_3 in the first harmonic. In the second harmonic (d), artifacts in the channel region due to dust particles are clearly appearing as dark dots. In the third harmonic (e), the noise level increases, but still beats the results of the PPP tip by far. In this measurement, no feature attributed to a plasma wave can be observed in the overview scan. From macroscopic rectification measurements of the device, a ten times smaller response to THz radiation of the JT1 device compared to the GT1 device is estimated. This much smaller sensitivity might quite likely be the reason why the plasma wave is not observed in the near-field scans. Unfortunately, interpretation of this data via slow line scans with better sensitivity is not possible due to the dirt

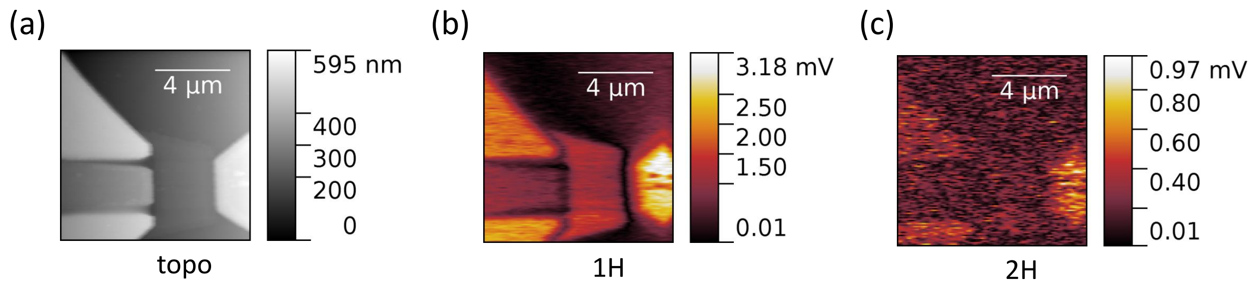

Figure 6.3.3.: *10x10 μm near-field scan of the TF1 with a PPP tip. Here, the device was turned by 180 degrees and in order to investigate the asymmetries of the setup. The device has been assembled for reference measurements and does not contain a graphene layer. (a) shows the topography. The first harmonic is depicted in (b). Here, the source electrode is much brighter than the gate and drain electrodes. The second harmonic contains a high level of noise, but shows the strongest signal at the source electrode as well. This measurement shows clearly the asymmetry of the system as described above.*

in the channel, which influences the near-field data too heavily.

6.4. Plasma Wave Examination

As observed in figure 6.3.1, a feature which might be attributed to a plasma wave is visible on the source side of the channel. To gain quantitative statements about this feature, linescans were performed with different voltages applied at the gate. To achieve a high signal-to-noise ratio, the integration time of the Lock-In amplifier was increased to 1 s and the linetime was chosen to to 100 s, the maximum value possible with the R9 AFM controller. Each line consists of 512 pixels. To assure the quality of the device and to confirm its functionality, the following steps were performed during the experiment: After each near-field measurement, an IV-curve of the transistor was collected to verify the device did not degrade in between the measurements. At the beginning and at the end of the experiment, the DC response on the THz field (see section 2.2, often called rectification in literature) of the device was verified using a chopper in the THz beam and a Lock-In amplifier while measuring the current induced in between source and drain due to irradiation of the device. This indirect proof of the plasma wave assures that the device in principle is worth to be examined.

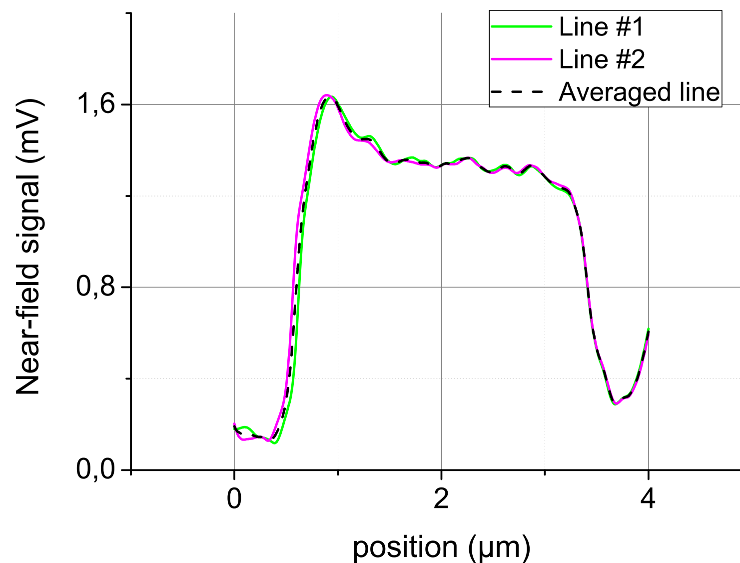

Figure 6.4.1.: *Near-field data for two individual lines after drift correction. The averaged signal is plotted with the dashed line in black. The lines are in very good agreement and show the reproducibility of the method.*

The near-field experiment was performed as describe below. Each line was collected at least twice per scanning direction. They were checked for artifacts occurring from, e. g., large dirt particles. The correct position perpendicular to the line (up and down in figure 6.3.1) was confirmed after each change of the gate voltage. Due to drifts during this long linetime, a slight shift along the scanning direction of each line is possible. To compensate this, each lineshift was corrected via the topography data before multiple lines were averaged. To exclude other artifacts, the forward and backward lines were compared for each measurement, which showed a good reproducibility of the method. As an example, the outcome of this procedure for the second harmonic is plotted in figure 6.4.1 for the measurement of the GT1 device with a gate voltage of 0 V. Here, the two lines (green, magenta) are in good agreement. The averaged line (black, dotted) deviates only slightly from the two single lines. Due to the long integration time of the Lock-In amplifier, the peaks in the near-field are not exactly at the same position in forward and backward direction when compared to the topography data. When scanning forward (or left), the peaks are shifted towards the left and when scanning backwards (or right), the peaks are shifted towards the right. Since the absolute position of the peaks compared to the gate electrode is of high importance, the middle positions of the peaks in forward and backward direction were chosen as the correct position of the peaks.

GT1 device analysis

The near-field signal of such a corrected linescan obtained on the GT1 device with a PPP tip is presented in figure 6.4.2. For the correct determination of the gate region, the sample was investigated with a commercial closed-loop AFM[3]. The positions are marked with blue lines in the third harmonic near-field information. On the source side, a peak with an amplitude of roughly 0.5 mV is clearly visible exactly at the start of the gate region. As resolution of s-SNOM is limited and the gates' DC field is not completely bound to the end of the bottom gate, the signal is enlarged also outside the gate region. The peak decays after several hundred nanometers towards the drain to a value of less than 0.4 mV, which is the near-field amplitude arising from the gold bottom gate electrode itself (E_m in the fitting procedure). The wave is almost completely overdamped, a second lobe is barely visible in this data due to strong damping of the electrons in the graphene. The origin of small modulation (few percent) visible in the range of 1 to 2.3 μm is not clear. However, smallest grains and residuals from the sample fabrication influence the distant-dependent near-field quite strongly. This could be an explanation of the modulation. Due to the very long linetime, the topography data is reliable enough to properly correlate the near-field data to such small features in the topography. A second possible explanation could be a plasmon mode in the graphene which could be exited via the tip and reflect at the graphene borders simililar to [71, 72]. However, such plasmon modes are not expected in graphene at this THz frequency [150]. Outside the channel, the signal decays as E_m changes when the tip is not above the gate anymore and due to edge effects when the tip reaches the source and drain electrode, also visible in the overview scan. To prove the origin of the peak is not related to the topography, devices with the same geometry but without the graphene layer were investigated as well. The general behavior of this device is similar, however no plasma wave could be observed (see figure 6.4.3). The signal-to-noise ratio is not as good as compared to the GT1 device, however it is still good enough to observe a feature as large as observed in figure 6.4.2.

Equation 6.2.4 was fitted to the experimental data as shown in figure 6.4.4 (a). A slight change of the slope is visible when the gate voltage is changed and will be analysed further below. However, the change is expected to be fairly small, as the plasma wave velocity only weakly reacts to the charge carrier density with $s \propto n^{1/4}$. For ideal graphene $n \propto |U_g - U_{CNP}|$, where U_{CNP} is the charge neutrality point, which will be discussed below. In a realistic system suffering from impurities, defects and inhomogenities, this dependence becomes much smaller (especially around U_{CNP}).

[3]Asylum Research Cypher AFM

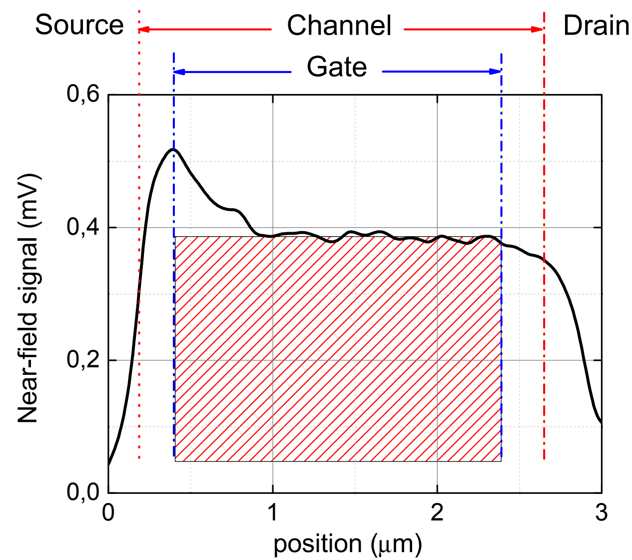

Figure 6.4.2.: *Third harmonic near-field data of a linescan performed on the GT1 device with a gate voltage of 0 V and a power of 45 mW at the sample. The dotted lines mark the positions of the channel and gate region. The red rectangle shows the contribution to the near-field signal arising from the gold gate, which is represented by E_m in equation 6.2.3.*

.

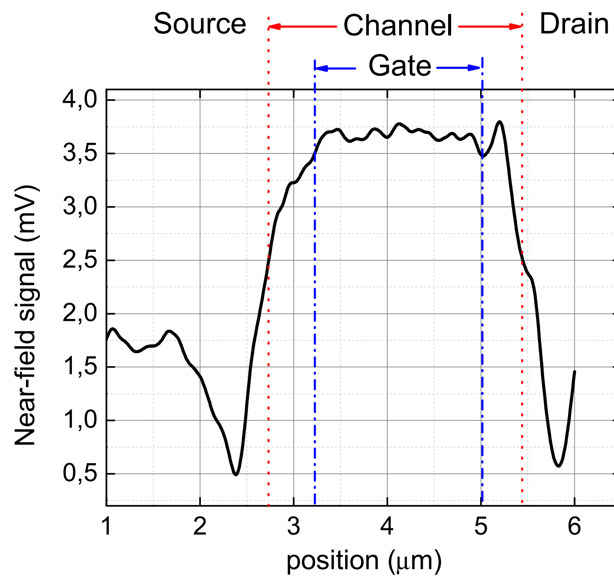

Figure 6.4.3.: *First harmonic near-field data of a linescan performed on the BT1 reference device with standard illumination. Here, no feature associated with a plasma wave could be observed. The noise level is larger than compared to the GT1 device, but still low enough to detect a large feature as visible in 6.4.2.*

.

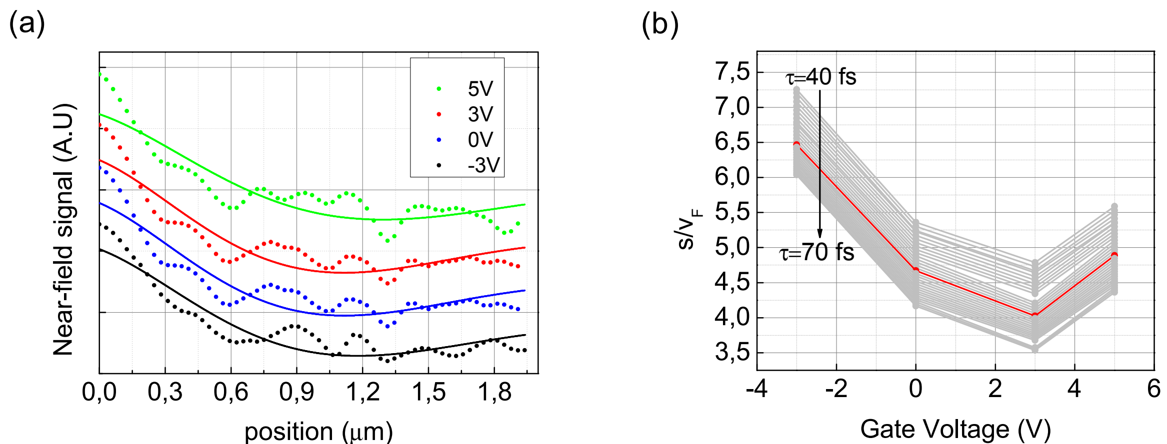

Figure 6.4.4.: *Linescan data: (a) Corrected third harmonic near-field data (dots) for different gate voltages on the GT1 device. The corresponding fits of the data are plotted with solid lines. Please note that the data is plotted with offsets for better visibility. (b) shows the deduced values for the plasma wave velocity with the red line for* $\tau = 55$ *fs.*

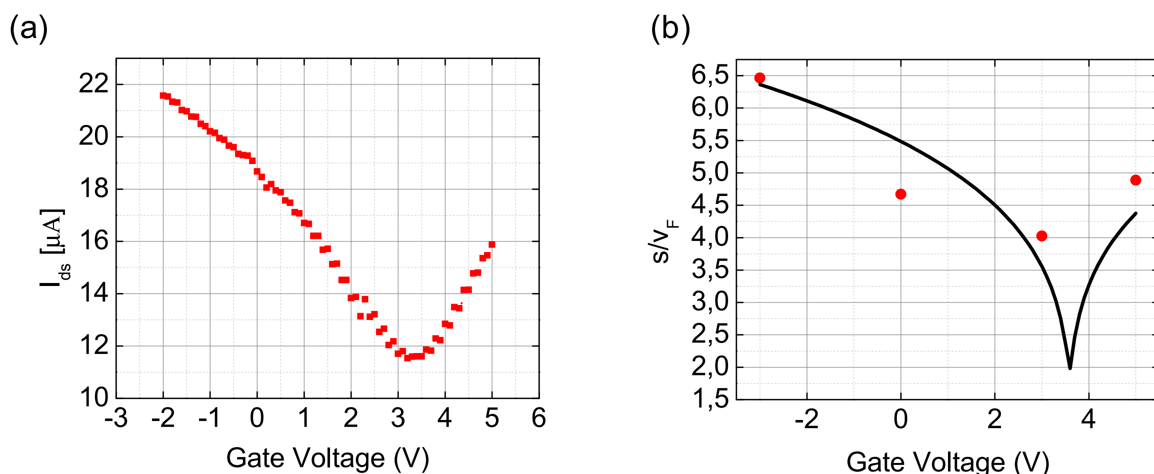

Figure 6.4.5.: *(a) IV-Curve of the GT1 device under THz radiation. The CNP is at 3.2 V. The calculation of the plasma-wave velocity s and the measured values are shown in (b). Both the macroscopic and the nanoscopic measurements are in decent agreement.*

The values for s deduced assuming different electron scattering times ranging from 40-70 fs, values which have been reported in several experiments for exfoliated graphene [151–153]. The peak height of the measured data is a bit larger than expected from the theory. From the dataset the origin of this mismatch cannot be explained, however one of the small oscillations visible in the data could also be present at a position of 0 μm and enlarge the signal. The results are shown in figure 6.4.4 (b). The red dots show the calculation for a scattering time of $\tau = 55$ fs compared to the Fermi velocity v_F[4]. The obtained plasma wave velocity is 6.5 v_F for a gate voltage $U_g =$ -3 V and decreases down to 4 v_F at a voltage of 3 V. For higher voltages, it rises again up to almost 5 v_F. The grey data shows the fitting results for τ = 40 fs up to τ = 70 fs. The behavior is in general the same, however smaller scattering times lead to larger s and larger scattering times to smaller s.

GFETs have been investigated in their application as transistors both in theory [154] and application [35, 36, 155, 156]. As mentioned before, the charge neutrality point is an important quantity in such a device. It is the point where an equal amount of electrons and holes is reached in the graphene through the application of a gate voltage. On the one hand, the CNP determines the point at which the conductivity of the GFET is the lowest. On the other hand, from plasma wave theories extended to GFETs [157] the expected lowest plasma wave velocity is at the CNP.

[4]The Fermi velocity corresponds to the Fermi energy E_F and hence is the velocity of the fastest electron in the semiconductor when cooled down close to zero temperature. Here, we assume $v_F = 1\mu$m/ps.

Consequently, the CNP is very well suited to connect macroscopic and nanoscopic results of the experiment. For the GT1 device, the conductivity of the channel was measured in between each linescan performed on the device to evaluate the performance, possible drifts and the influence of the strong THz radiation on the device. A typical IV-curve is depicted in 6.4.5 (a). Here, the current I_{ds} has the lowest value of $\sim$ 12 μA at a gate voltage of 3.2 V and rises up to $\sim$ 21.5 μA for a gate voltage of -2 V. The ratio in between these two values is the so-called ON/OFF ratio and is tiny compared to a any commercially available transistor based on classic semiconductors, where a ratio of 10000 can be found. Graphene-based transistors can have a ratio 500 times larger than the devices of this work [158], when they are optimized for this particular issue. In this case, the design was optimized for near-field measurements, explaining this huge difference. However, the ON/OFF ratio is not important for the plasma waves' behavior. (b) shows the comparison of the plasma wave velocity to the IV-curve. Here, the measured CNP was used to calculate s with the theory of [157]. The deduced values for s are shown with the red dots. They are in decent agreement with the calculation shown in black. To check the consistency further, more experimental values are needed. Unfortunately, the steep decrease of s at the CNP can neither be confirmed nor be ruled out with this measurement as voltage step sizes of roughly 0.2 V would be needed to further examine the plasma wave's behavior at the CNP.

HT3 device analysis

A second experiment on another device was performed few month later to confirm the experimental results. The device was fabricated in an independent process run, another tip was used and the setup was realigned completely. The data analysis after the experiment was performed in the same way as described above. The results are summarized in figure 6.4.6.

Overall, similar results could be obtained. For this device, the plasma wave occured at the drain side of the channel. This could be a reason of changes of the asymmetries in the setup, e. g. from different alignment of the laser. However, as explained in section 2.2, the boundary conditions on source and drain are arbitrarily chosen and can be interchanged without loss of generality and the plasma wave can be induced from either side of the channel. Hence, the data analysis has been performed in the same way as before. The linescans and their fits to equation 6.2.4 are presented in figure 6.4.6 (a). For this device, the noise level is lower and the fits are in better agreement with the experimental values. From the fits, the plasma wave velocity is deduced assuming electron scattering times in between $\tau =$ 25 - 60 fs. The red line marks the fit with $\tau =$ 40 fs. The plasma wave velocity is $s \approx 5.75 v_F$ for a gate voltage of -2V and decreases below $4 v_F$ before it increases

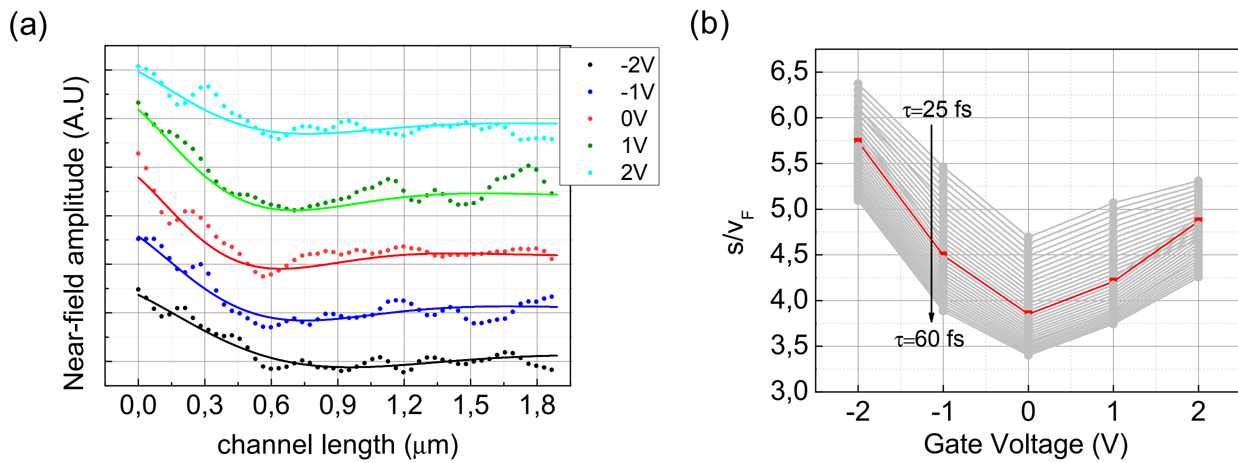

Figure 6.4.6.: *Linescan data: (a) Corrected second harmonic near-field data (dots) for different gate voltages on the HT3 device. The corresponding fits of the data are plotted with the lines. (b) shows the deduced values for the plasma wave velocity.*

to $s \approx 5v_F$ for a positive gate voltage of 2V. The plasma wave velocity is a bit lower than observed on the GT1 device, but is in the same range and can be very like varying from device to device due to uncertainties during the fabrication.

The corresponding IV-curve for device HT3 is shown in figure 6.4.7 (a). The absolute current and the ON/OFF ratio are similar to the GT1 device, however the CNP is at 0.3 V here. This difference emphasizes the sensitivity of the graphene's electrical potential to the fabrication details and/or possibly the air-exposure history as chemical reactions cannot be excluded under ambient conditions. The deduced plasma wave velocities for $\tau = 40$ fs are compared to the theorectical values of [157]. Here, the data fits quite well to the calculation (see figure 6.4.7 (b)), however the density of measurement points is again not sufficient to properly map the area of very low s around the CNP.

In this measurement run, the rectification of the device was studied in more detail. Figure 6.4.7 (c) features a measurement of the rectification current as a function of the gate voltage. The functional dependence is as expected [34]. The data shows a strong hysteresis where the minimum is shifted by $\approx$ 0.8 V, depending on the measurement direction. This behavior can be explained by significant influence of impurities in the device. At lower and larger voltages, the rectification is rising. Here, the device is more sensitive towards the THz radiation. The rectification data roughly fits to the IV-curve and to the near-field measurement, although the hysteresis is much more pronounced for this measurement. During the measurement of several hours a drift of the IV-curve

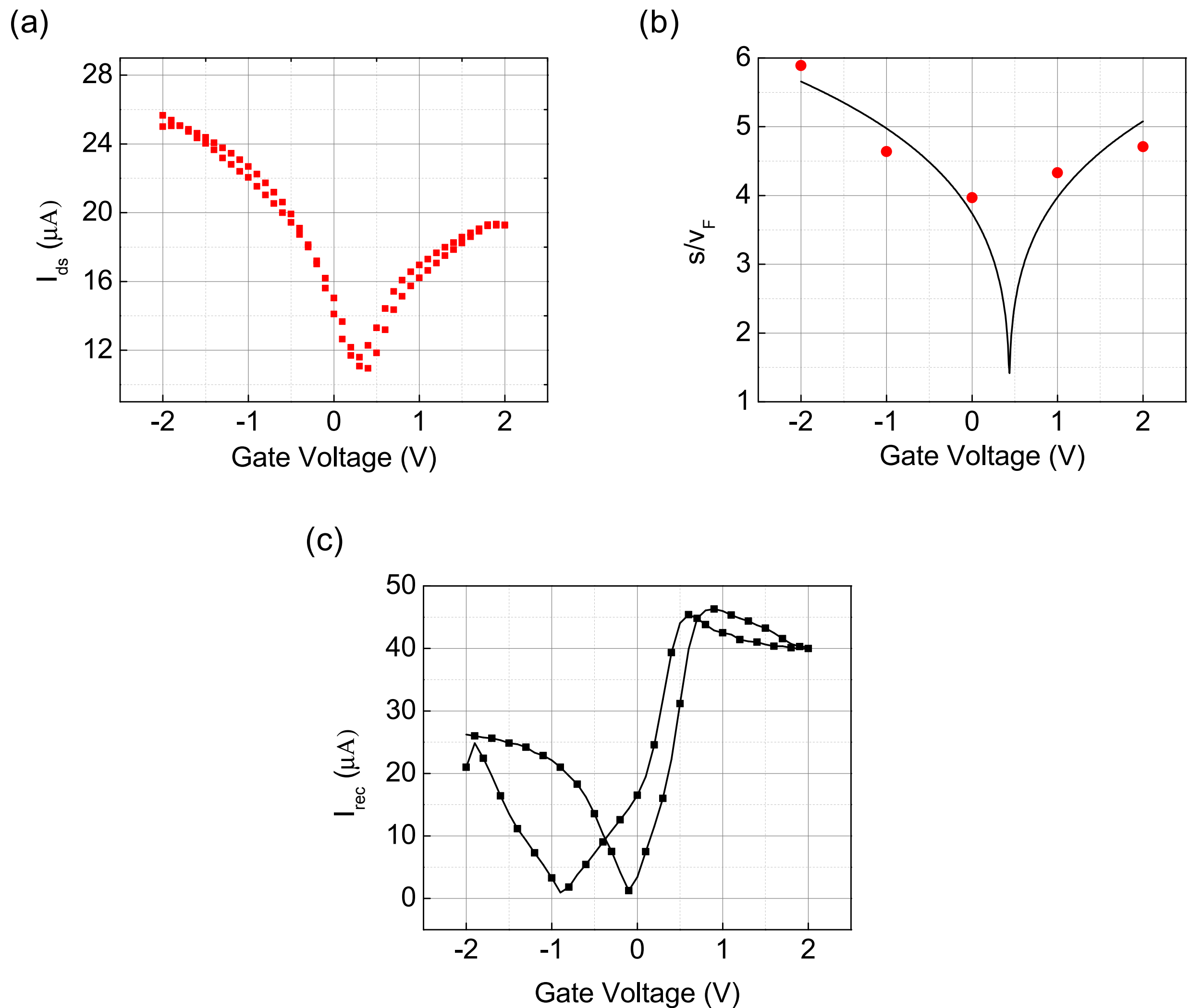

Figure 6.4.7.: *(a) IV-Curve of the HT3 device under THz radiation. The CNP is at 0.3 V. The calculation of the plasma-wave velocity s and the measured values are shown in (b). The rectification as a function of the applied gate voltage is shown in (c), here, the largest values could be achieved far from the CNP. When direction of the measurement is changed, a hysteresis of $\approx$ 0.8 V can be observed. Both the macroscopic and the nanoscopic measurements are in decent agreement.*

could be observed. As the rectification measurement was performed in the first few minutes of exposure to the THz field, this could explain a drifts at later stages of the experiment and thus the mismatch to the IV-curve.

6.5. Conclusion

With the results obtained above, we are certain that the plasma wave predicted by Dyakonov and Shur has successfully been mapped directly for the first time. Two devices from different fabrication runs showed near-field responses which can be attributed to plasma waves. Both datasets were collected months apart from each other obtained with different AFM tips and alignment. The reaction to an applied bias delivered the expected response of the charge carrier dynamics in the GFETs. The deduced plasma wave velocities are in good agreement to literature expectations [159, 160]. Although the velocities depend on many parameters, the calculations lead to plasma wave velocities of $s \approx= 4 - 8v_f$ at room temperature, completely matching our findings.

The origin of the small oscillations in the GT1 device remains unverified, however some possible origins are discussed at this point. For the following thoughts, please note the signature of the small oscillations remains unchanged with different gate voltages. The change of signal is small compared to the main peak of the plasma wave, but can influence the deduced plasma wave velocities, as the fitting of an overdamped plasma wave will be quite sensitive to such changes.

Firstly, small scatterers could be the reason for these unexpected signal changes. They could be located in between the graphene and the Al_2O_3 or the between Al_2O_3 and the gold electrode. They can lead to small areas of enhanced local field and lead to fluctuations in the signal. A second possibility are dirt particles which can be located in between the gold gate electrode and the Al_2O_3 or the Al_2O_3 and the graphene respectively. In this case the effective distance of the tip to the gold gate could be changed, leading to a change to the very distance-sensitive near-field signal. This behavior was observed in devices with lower processing quality and is visible in figure 6.3.2. A third point to consider is the reflection of the plasma wave at the tip. Here, a wave could possibly travel to another edge and reflect back to the tip. In this case we expect the wavelength of the standing wave pattern to be half of the plasma wave, however we observe rather a quarter. In the recent years, tip-launched plasmons have been in the scope of many scientist [71, 72], however such plasmons are not expected on graphene at the wavelength of 2 THz [150]. From the plasma waves' theory when higher orders of the differential equations are regarded, however such higher harmonics are excited very inefficiently and would occur at multiples of the incident

THz frequency [157] and would not be detected with the InSb bolometer. In the HT3 devices, the small oscillations are not as stable compared to the GT1 device, however the same though apply and none of the proposed origins explains the observes fluctuations.

As the plasma wave's signal is fairly small and the free-electron laser FELBE is noisy compared to any table-top laser, long Lock-In integration were needed for the measurement. Collecting a complete dataset with rectification measurement, IV-curves and near-field linescans takes several hours, which implies the issue of drifts. While drifts in the AFM could be taken care of by checking the tip's position after each linescan, the drifts of the CNP cannot be compensated for. However, the shift of the CNP was strongest shortly after it was placed in the setup and irradiated for the first time. When the quantitative near-field investigations were started, drifts of the CNP were comparably small.

For higher certainty and better quality during the investigation, a device with lower damping is needed. In general the graphene could be processed on or encapsulated with hexagonal boron nitride to strongly increase the electron scattering time [161]. This would certainly allow for a wave appearing with several maxima or even a standing wave in the channel of the device. A second possibility is an experiment at low temperatures, which is possible with a cryogenic near-field microscope located in the same laboratory [63, 64]. Here, standing wave patterns could be observed when the device is cooled down to liquid helium temperatures [25, 26]. However, such an experiment at the free-electron laser is very challenging, as cryogenic s-SNOMs are far more complex and tend to suffer from dirt-condensation of the devices due to condensation of particles at the sample. However, with rapid development of electronic radiation sources in the few-hundred GHz range, a successful combination of a low temperature s-SNOM with such a table-top device seems to be an approach with prospect of success. A table-top radiation source would grant much more measurement time and would certainly help to collect more data points with smaller spacing in between the gate voltages.

The applied gate voltage can in general lead to an electrostatic force in between tip and sample, changing the effective distance of the tip and hence the near-field response. In our measurement, one could argue that the peaks arising from the plasma wave could be broadened due to the effectively larger tip distance, as the resolution of the near-field microscope is worse in this case. However, we observed the plasma waves on two devices with totally different CNP. An electrostatic artifact would occur at the same voltage (most likely at 0 V) for both devices and is for this reason quite unlikely. To fully exclude this artifact, a Kelvin-Probe Force Microscopy (see appendix A.1) control loop running at the same time can be applied. However, graphene screens

DC potential quite well [162], minimizing the impact of electrostatic forces acting in between tip and sample. During this measurement we refrained from the use of this technique, as a voltage at the tip could either break or influence the device strongly.

This measurements shown in this chapter show that the plasma wave is involved in the process of THz detection in FETs. Nonetheless, it does not exclude that photothermal effects are occuring at the same time. As the plasma wave is the origin for THz emission in such a device, the knowledge gained with this measurement can be utilized to optimize FET-based THz sources. If the emission efficiency rises to a value suitable for most applications, FET-based emitters would by far be the cheapest THz sources on the market as semiconductor processing is a scaleable process. Another field of application would be THz communication via FETs, which could enable unforeseen data rates. At this moment it is unclear whether the optimization of these devices to such a point will suceed or not.

7. Discussion and Outlook

Der Mensch ist nie zufrieden.

(Rosa Luxemburg)

Goal of this work was the application of infrared and THz s-SNOM to investigate nanoscale charge carrier distributions. While a new method has been introduced to employ table-top infrared radiation for the probing of the pump-induced carriers, THz radiation from the free-electron laser has been applied to investigate plasma waves in graphene field-effect transistors.

The **theoretical part** of the work contains two chapters. The first chapter gives an overview of the sample systems, as well as the applied experimental techniques in this work. The influence of charge carriers to the permittivity of semiconductors is described with the Drude model. As the charge carriers can deliberately be excited by short and intense laser pulses, transient states are described in this section as well. The main conclusion is an enhanced sensitivity of THz radiation to smaller charge carrier densities, both in ground and transient states. However, with intense pump laser systems, infrared radiation can still be a good choice for probing these transient states. Graphene field-effect transistors and their fascinating properties are theoretically described as a second sample system. Afterwards, near-field microscopy and THz optics are described conceptual. The second chapter gives a detailed description of s-SNOM theory with the analytical dipole model that describes several detection techniques in order to suppress background scattering. The model is extended to include the occurance of sidebands appearing under the influence of low repetition-rate pulsed laser. At last, a numerical simulation is chosen to investigate the influence of the tip shape in infrared and THz s-SNOM.

In this work, three different near-field **setups** have been set up and are described in a dedicated chapter. While the setup at the free-electron laser FELBE for the investigation of GFETs and the table-top infrared setup for the investigation of sidebands are similar, TDS-SNOM employs another detection technique, allowing for the deduction of the full amplitude and phase informa-

tion of the near-field interaction.

With the FELBE setup, the superiority of full platinum Rocky Mountain Nanotechnology tips compared to standard silicon-based tips for THz-SNOM has been shown clearly. The possibility to extract the full amplitude and phase information in the TDS-SNOM setup demonstrates that these tips have a geometrical resonance in the low THz range, as expected by numerical simulations performed in the theory chapter. A nanoscale material contrast on a Si/SiO_2 structure could be observed with the TDS-SNOM as well. Although the THz radiation had a pulse repetition rate of only 200 kHz and the setup has thus operated close to the Shannon-Niquist sampling limit, the measurement performed proof that the setup can be combined with the TELBE beamline.

The **sidebands** induced with an intense near-infrared pulse laser were analyzed in detail in the application chapter with both distant-dependent measurements and near-field scans. Here, physical and technical aspects were analyzed in detail. The existence of higher-order sidebands has been proven and a nanoscale material contrast via the demodulation of sidebands has been shown on a germanium/gold nanostructure.
The demodulation technique applied in this work is rather slow leading to high demands on the stability of the whole setup. Here, the application of state-of-the-art Lock-In amplifiers [124] would certainly pay off due to their ability to demodulate more than one sideband at once. Another way to investigate the sidebands could also be a sample where the sample properties are modulated electrically instead of optically. This would simplify the measurement and - depending on the sample design - could strongly reduce artifacts if the excited areas are small.

Both macroscopic and nanoscopic measurements of **graphene field-effect transistor** devices are described and analyzed in the second result chapter. Signatures associated with plasma waves could be found and a parameter study for different gate voltages show an agreement of the data with theoretical predictions. Unfortunately, the damping of plasmons in the graphene is too large at this frequency to excite a propagating wave which would show several maxima. The attempt to use a higher THz frequency was unsuccessful as the applied bow-tie antenna is not designed to work at this frequency. The origin of smaller oscillation visible in the dataset remains unclear and can be part of future investigation.
As the investigation of the plasma waves is of high interest for the scientific community and also for the implementation of FET transistors both for THz detection and emission, some improvements for the experiment are suggested at this point: The most promising approach is the enhancement of the mobility/scattering time of the charge carriers in the graphene to exhibit less damping in the channel. This could be performed with low-temperature s-SNOM at the free-

electron laser FELBE, where standing plasma-waves could be observed in the channel. However, this approach would not change the influences of local scatterers, is technically very challenging and falls into the category of a high-risk-high-gain experiment. A second approach is the implementation of hexagonal boron nitride (h-BN) into the device. With the deposition of graphene onto h-BN, the mobility of the charge carriers can be enhanced by a factor of ten [163], which will easily result in a standing plasma wave in the channel. However, the deposition of h-BN is not yet as well established in the collaboration as classic device fabrication methods for other materials and might be demanding.

In conclusion, the plasma waves predicted by Diakonov and Shur could directly be observed for the first time. The results open the pathway to further investigations and the application of such devices in micro- and nano-electronics.

A. Appendix

Nothing compares to the simple pleasure of a bike ride.

(John F. Kennedy)

A.1. Scanning Probe Microscopy

Scanning Probe Microscopy (SPM) is an umbrella term for a vast number of techniques which are generally located in the field of microscopy. The main idea of SPM is to utilize a nano-probe to measure one or several physical quantities with a point-by-point approach to obtain a picture with 2D information. In this chapter, the first scanning-probe technique is briefly explained followed by the fundamentals of atomic force microscopy, the basis for all techniques applied in chapter A.2. Lastly, the concepts of near-field microscopy are presented.

As the AFM used in the experiments is home-build and was specially designed for this work, the well-established atomic force microscopy and related techniques are briefly explained in the appendix. The performance of the device is crucial for all near-field measurements shown in this work and is evaluated in section A.2.

The term Scanning Probe Microscopy in general pools various techniques which are applied in order to investigate a sample with very high resolution using a small probe and a point-by-point approach. The development of this technique had an enourmous impact on the scientific community. Before its invention, all microscopy techniques were either strongly limited in resolution (e.g. optical microscopes) or invasive (e.g. scanning electron microscopes) and could only extract few of the physical properties. The first SPM was the scanning tunneling microscope invented by

Binning and Rohrer in 1982 [164]. Their developement was awarded the *Nobel Prize* only four years later. Today, a scanning tunneling microscope can easily reach (sub-)atomic resolution.

Binning and Rohrer used piezo actuactors and an active control loop for their first STM. They measured the tunneling currect flowing from tip to the sample to successfully image single atoms. Keeping the current constant, the sample was scanned line by line to obtain a 2D image of the topography of the sample. To avoid the tip from crashing into the surface or leaving the tunneling regime, a piezo actuator moved the tip up and down via a control loop. This general concept of actively controlling the size of a measured quantitity and line-by-line data aquisition is applied in almost all scanning probe techniques. It is the key feature for well defined measurement conditions and minimal impact on tip and sample due to wear-off.

Countless types of SPM have been developed in the recents decades. This chapter features the theory of several modes of operation applied in this work using the home-built T-Bone atomic force microscope. Typical measurements with these SPM/AFM modes are shown in the setup chapter A.2.

Atomic Force Microscopy

To overcome the limitation to conductive samples in STM, Binning invented the *Atomic Force Microscope* (AFM) only few years later [165]. Instead of measuring and actively controlling the tunneling current, the force in between tip and sample was chosen as a physical quantity. The working principle of an AFM is schematically shown in A.1.1. The tip is mounted to a cantilever and brought into close vicinity of the sample. The backside of the cantilever illuminated by a focused laser and reflects the beam onto a four-quadrant diode. The laser beam is usually aligned to the center of the detector in order to level the signals of each quadrant. The signals processed by an AFM controller, which usually calculates the so-called top-bottom signal $TB = A + B - (C + B)$ and the left-right signal $LR = A + C - (B + D)$. Whenever the tip reaches a topographic feature of the sample, the cantilever is bent and the point of the reflection is shifted upwards on the four quadrant diode, leading to change in the TB signal. If the tip exhibits a torque, the reflection on the four-quadrant diode is shifted in left or right direction resulting in a change of LR. The readout by the four-quadrant photodiode is particularly sensitive if the distance in between cantilever and the photodiode is large. For best performance, the spot size of the laser focus should be smaller than the cantilever width to avoid interference of light scattered on the sample with light from the reflection. A second approach for the cantilever read-out is an interferometric detection via

an optical fiber placed on top of the cantilever. Gaining enormous sensitivity on the bending of the cantilever [166], the sensitivity on torsion is lost completely. Such read-out is usually applied to measure extremely small forces present for example in magnetic force microscopy [167].

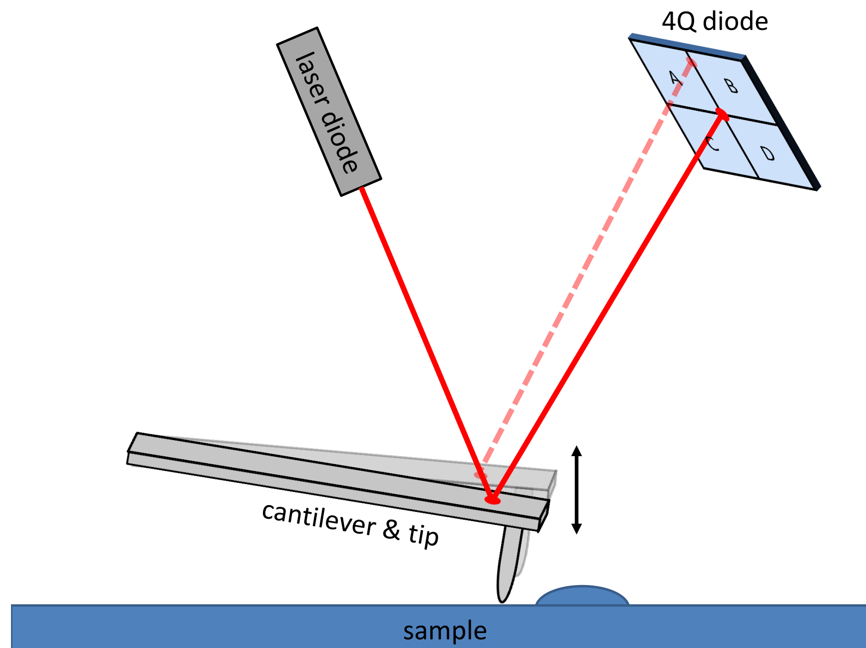

Figure A.1.1.: *Schematic drawing of an AFM. The tip is mounted on a cantilever and touch the sample surface on the other side. A laser diode is focused on the cantilever back, the reflection hit a four-quadrant photodiode. Whenever the tip moves over an object on the sample, the cantilever is bent (shown with the transparent sketch) and the reflected beam is hitting the four-quandrant photodiode on a different position (dashed line).*

Most commercially available tips are made out of monocrystalline n-doped silicon and have very high quality standarts and reproduceability. The stiffness k of the cantilever is usually in the range of 0.2 - 50 N/m [168] and the effective tip apex are below 10 nm. If the tip is coated with a conductive coating, the radius is in the range of 20 nm. Tips can be coated with conducting diamond, some companies also offer tips made of silicon nitride [169].

The forces acting on the tip when it is approached to the sample are on the one hand attractive, long-ranging van-der-Waals and electrostatic forces, on the other hand short ranging repulsive forces due to the Pauli principle. The combination of these forces is commonly approximated by the Lennard-Jones-(12,6)-potential given with

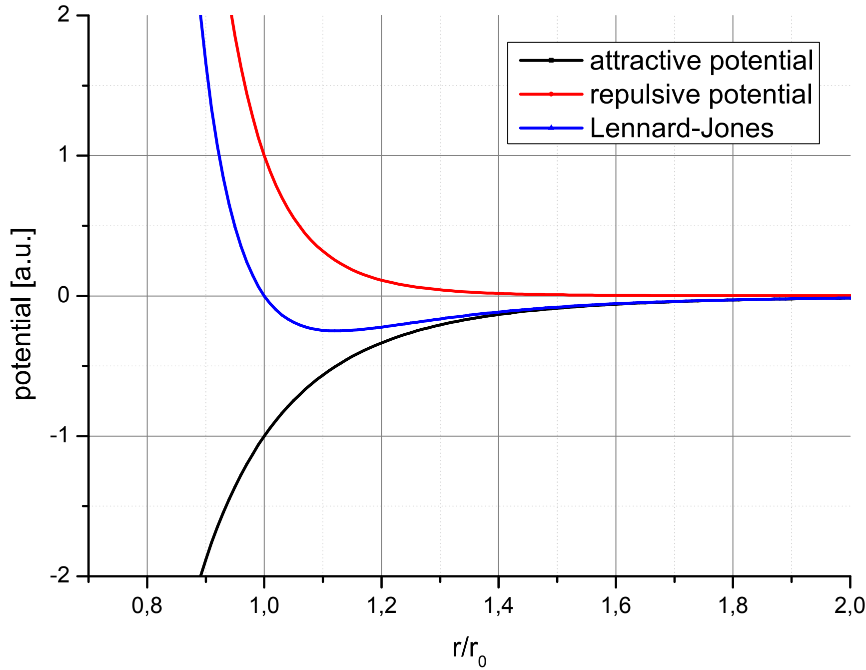

Figure A.1.2.: *Lennard-Jones potential for the 12th potency on the repulsive term. Here, two point atoms are assumed to interact. The long-range interaction is dominated by the attractive potential (black), the short range is dominated by the repulsive potential (red). Together, they add up to the blue Lennard-Jones potential. The potential crosses zero at a value of r_0, the lowest point is reached at $\approx 1.12r_0$. The depth of the combined potential depends on the factors C_r and C_a. In the shown case, both were chosen to one for simplicity.*

$$C_{LJ} = \underbrace{\frac{C_r}{r^{12}}}_{repulsive} - \underbrace{\frac{C_a}{r^6}}_{attractive} \tag{A.1.1}$$

where C_r and C_a are material constants. The 12th potency is chosen for simplicity during calculation and can also be chosen to other values larger than 6. The attractive term is negative and dominating the bahavior for large distances leading to an attractive force in between tip and sample. For smaller distances the repulsive term with positive values becomes much larger laeding to a repulsive force. The point of zero force acting on the tip is at $\approx 1.12r_0$, whereas r_0 is the point where the potential crosses zero.

Contact Mode

On the long list of AFM modes, contact mode was the first one to be successfully implemented, due to its the simplicity. The cantilever is approached to the sample surface until it is bend to a point where the TB signal reaches a desired value. The TB signal is constantly compared to the setpoint in the AFM controller. Whenever it is lower than the setpoint, sample and tip are approached further. In the opposite case, the tip is pulled back. With optimized p-i-gain settings, the tip's displacement can be controlled in the sub-Å range. During a lateral scan, topographic features change the displacement which is then compensated by the control loop. From this data, the topographic information can be recovered with the AFM controller software. The mode is simple to handle and easily reaches atomic resolution but has the disadvantage of solely operating in the repulsive regime, where the tip is worn off quickly and might even damage the sample during scanning. Due to the forces acting on the tip and sample, the mode is also not suitable for soft materials (e.g. biologic samples.) For these reasons, contact mode is usually only chosen when very high topopgraphic sensitivity is needed or when an ohmic contact in between tip and sample is needed, like conductive AFM (c-AFM) or piezoresponse force microscopy (AFM).

Piezoresponse Force Microscopy

In piezoresponse force microscopy (PFM) the AFM is operated in *contact mode* while an AC voltage is applied in between tip and sample. This mode is used to map the spontaneous formation of ferroelectric domains via the converse piezoelectric effect.

Piezoelectric materials develop mechanical stress when they exhibit an electric field. The strain is given with

$$e_j = d_{ij} E_i \quad \text{with} \quad i = 1,2,3 \quad \text{and} \quad j = 1,2,3,4,5,6 \tag{A.1.2}$$

when *Voigt notation* is applied [170]. e_j with $j = 1,2,3$ represents normal stress along the three axis, whereas $j = 4,5,6$ denotes shear stress. The symmetry of the investigated crystal determines the non-zero elements of the 6x3 piezoelectric tensor d_{ij}. An overview for different lattice structures is given in [170].

Of all the piezoelectric material groups, some are pyroelectric, meaning they develope a spontaneous polarization below the Curie temperature T_C. Here, the low symmetry leads to a uneven distribution of charges in the unit cell. To reduce electrostatic energy, the crystal forms domains of different polarization directions. Depending on the polarization of a domain, the crystal will

react differently on an applied field. Material systems, where the polarization can be deliberately changed in between states by a strong electric fields, are called ferroelectric.

The strain and hence the deformation of a crystal can be measured with an AFM with nanoscale resolution. A schematic drawing of the setup is shown in A.1.3. Compared to a normal AFM setup, the sample should be grown or mounted on a conductive material to allow for the application of a voltage in between tip and sample. For simplicity, the polarization of the crystal is either facing up or down (so-called c_+- and c_-- domains). The area below the tip exhibits an electric field leading to an expansion or shrinking of the crystal along the z-axis, causing the tip to get slightly displaced. To separate the expansion from the topography, an AC voltage is applied. Influences on the displacement are constant as the setpoint is chosen constant, influences of the AC voltage are modulating the four-quadrant diode signal at the AC frequency. With a lock-in amplifier, this signal can be detection with enourmous sensitivity, signal-to-noise ratio and amplification [124].

When performing PFM measurements, the AC frequency must be chosen specifically for each application. For materials with large piezoelectric tensors, a non-resonant frequency of few kHz is the best choice, as it offers very stable operation even when scanning parameters change. For materials with small piezoelectric tensors, the signal at low frequencies is too small to be detected. Here, the PFM frequency is sweeped to find the *PFM* contact resonance frequency. It is usually location at 3-5 times the mechanical eigenfrequency of the cantilever. A typical sweep is shown in the setup chapter in figure A.2.4. Around the resonance peak, the PFM amplitude is very large and can be detected easily. The drawback of this method is clearly the smaller stability of the mode. Whenever the parameters and thus the eigenfrequency of the tip-sample system change (e.g. due to dirt pick-up), the amplitude and phase strongly change as well and the reliability of the data becomes questionable. To tackle this problem, the resonance frequency can be tracked during the measurement [171], a mode only recommended for more experienced users.

Tapping Mode

The next sections all fall into the category of dynamic modes in AFM. The motivation to develope these application is degradation of tips during contact mode AFM [172, 173]. In all dynamic modes, the mechanical contact of tip and sample is kept minimal to avoid the tips' wear. Thereby, a higher reproducibility during measurements is achieved and resources such as time and money are saved.

In all dynamic modes, the cantilever is mechanically excited via a piezo actuator, causing the tip mechanically oscillate above the sample surface. To find the mechanical resonance of the can-

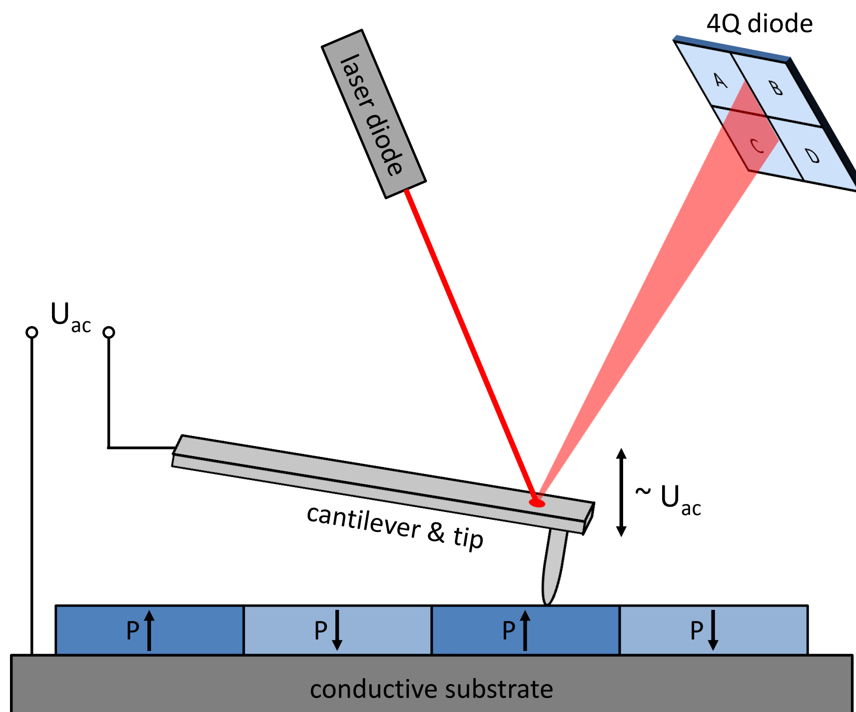

Figure A.1.3.: *Schematic drawing of a PFM setup. An AC voltage is applied in between tip and sample. On a piezoelectric substrate, the field will lead to an expansion/shrinking of the sample (blue). The displacement of the cantilever is mapped with the beam deflection on the four-quandrant diode (depicted with the transparent beam). The signal is demodulated with a lock-in amplifier. On different piezoelectric domains (blue colors), the amplitude and phase of the four-quadrant diode signal compared to the AC voltage changes.*

tilever, the excitation frequency is sweeped while the oscillation signal is obtained at the four-quadrant diode. A typical frequency sweep is shown in the setup chapter in figure A.2.6. In most dynamic modes, the tip is driven at its fundamental resonance frequency

$$f_0 = \frac{1}{2\pi}\sqrt{\frac{k}{m^*}}, \tag{A.1.3}$$

whrere m^* is the effective mass and k the stiffness of the cantilever [174]. Depending on the shape and the material of the cantilever, the resonance frequency can range from several kHz up to the MHz range. Most probes for dynamic modes have frequencies above 100 kHz. Some techniques also use higher eigenmodes of the cantilever [175], but are not in the scope of this work.

Tapping mode ensures the tips' wear is kept small while keeping the experimental effort low. The mode is actually so simple and fail-safe that an unexperienced scientist can be trained to use this mode within a matter of hours.

When driven with small excitation amplitudes, the cantilever acts as a harmonic oscillator. The cantilevers response to the excitation is shown as a function of the driving frequency in figure A.1.4. The height A_0 and the full width half maximum (FHWM) of the peak is determined by the so-called quality factor Q of the resonance, which is directly connected to damping of the oscillator. When the tip is far away from the sample, it is determined mainly by the drag of the air surrounding the cantilever. When the tip is approached to the sample, the forces discussed in figure A.1.2 are getting stronger, leading to increased an damping and thus a lower Q factor. The resonance peak is shifted towards frequencies, the height becomes smaller and the width larger.

In tapping mode, the excitation frequency is kept constant and the oscillation amplitude is chosen as the main physical property for the control loop. When the sample is approached, the cantilever exhibits stronger damping, resulting in a decreased oscillation amplitude at f_0 (see figure A.1.4). The tip is usually approached to a setpoint of $0.5 \ldots 0.8 \cdot A_0$. The control loop works in a similar way as in contact mode allowing for the mapping of topographic feature. In many cases (e.g s-SNOM), the tip oscillates with amplitudes of 50 nm or more. In this case, the tip exhibits both the attractive and the repulsive regime of the Lennard-Jones potential. The tips wear is significantly reduced compared to contact mode due to less time spent in the repulsive regime.

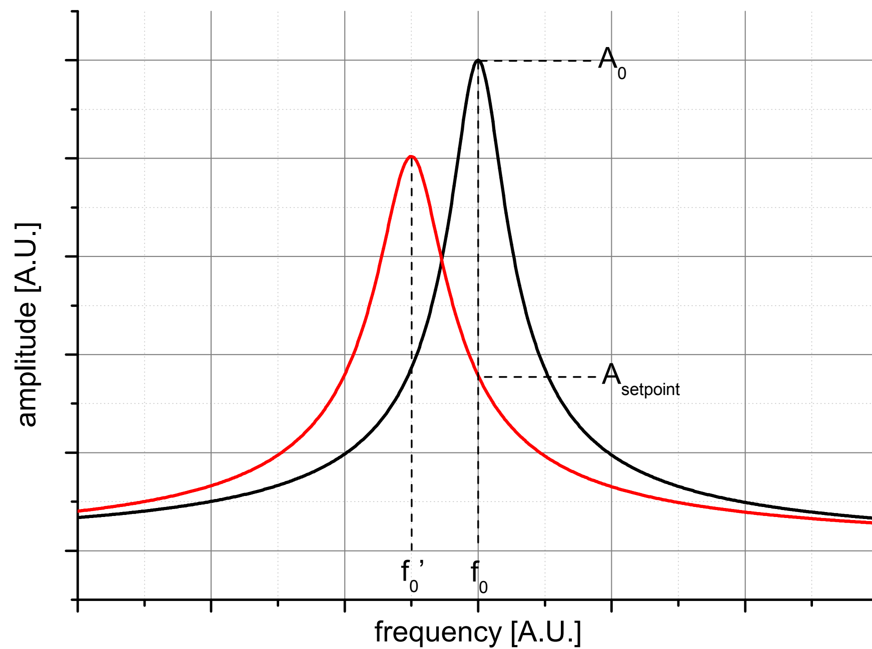

Figure A.1.4.: *Frequency spectra in tapping mode: The black curve shows the response of an harmonic oscillator in air, having a peak with an amplitude A_0 at f_0. When the the tip is approached to the sample and the the setpoint $A_{setpoint}$ is reached, the damping of the oscillator increases, the peak is getting smaller and broader. Also, the resonance frequency is shifted to a smaller value f_0'. Please note that the shift of the peak is not to scale and increased manually for the sake of clarity.*

Non-Contact Mode

In non-contact mode, the repulsive regime is never reached. For this reason it is sometimes called True Non-Contact. Here, the physical property measured and fed into the control loop is the resonance frequency. In order to track the frequency, a phase-locked loop is applied comparing the actual resonance frequency to the excitation frequency via a phase measurement. Whenever the resonance frequency is higher than the setpoint, the tip will approached further to the sample and vice versa. In ambient conditions, the setpoint is usually chosen to a few ten to hundred Hz below the resonance frequency far away from the sample in air. The resonance frequency can be measured with mHz precision allowing for very high sensitivity even when the tip is in the range of weak interaction in the attractive regime. Therefore, the wear of the tip is minimal unless the control loop fails to correctly control the excitation frequency.

At the same time, a second control loop can be applied to keep the amplitude of the oscillation constant via a change of the excitation amplitude. This allows for artifact free measurements in techniques sensitive to the tip ocillation amplitude like s-SNOM, or magnetic force microscopy (MFM). A schematic drawing of the frequency spectrum in this mode is shown in A.1.5.

The main disadvantage of this mode is its complexity. Only experienced scientists will be capable to correctly apply this mode and almost any mistake will lead to the tip crashing into the surface. For this reason, non-contact mode is mostly applied when smallest forces need to be detected e.g. in MFM.

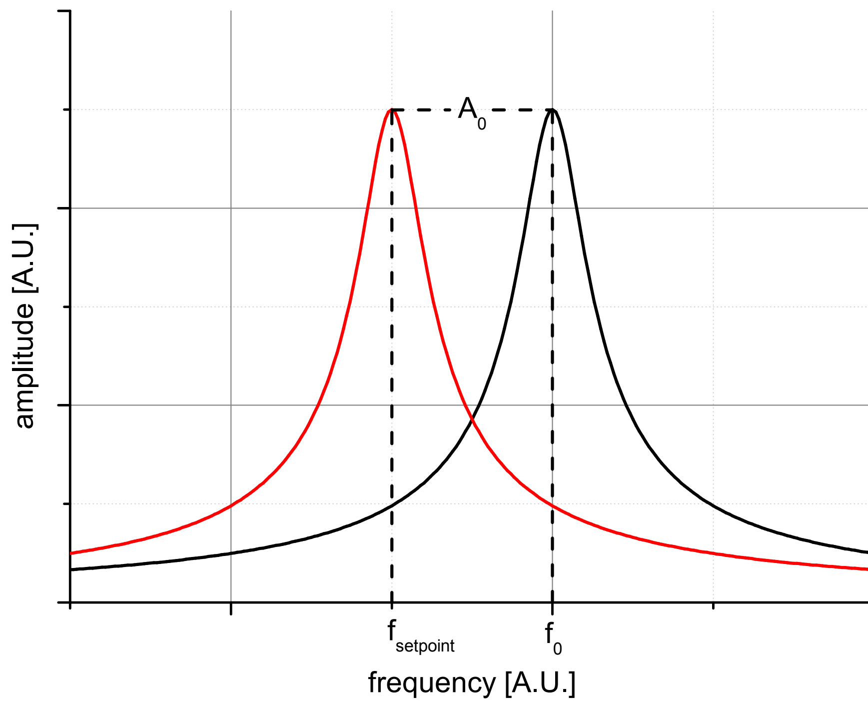

Figure A.1.5.: *Frequency spectra in non-contact mode: The black curve shows the response of the harmonic oscillator far away from the sample. It has a peak with an amplitude A_0 at f_0. When the the tip is approached to the sample the resonance frequency shifts to lower values. The tip is approached until a chosen frequency $f_{setpoint}$ is reached. At the same time, a second control loop keeps the amplitude constant at A_0 to avoid influences of changing oscillation amplitude to the system.*

Kelvin Probe Force Microscopy

Kelvin Probe Force Microscopy (KPFM) is a technique to measure the electrostatic potential of a sample on the nanoscale. This can be used to determine the work function of a material and also to avoid influences of electrostatic interaction in between tip and sample.

The idea goes back to the capacitor method of Lord Kelvin, which he used to determine the contact potential difference (CPD) of parallel capacitor plates. He connected the two plates via a wire and altered their distance. Hence, their capacitance C changed and a current flew through the wire

$$\frac{dQ}{dt} = I = \frac{dC}{dt} \cdot U_{CPD}. \tag{A.1.4}$$

To nullify the running current, a voltage U_K can be applied to the capacitor:

$$I = \frac{dC}{dt} \cdot \big(U_{CPD} - U_K\big). \tag{A.1.5}$$

In AFM, the capacitor is the tip sample system. The tips used for KPFM have a metallic coating, hence the work function is well defined and one can determine the work function of the sample underneath the tip. On insulating samples, U_K corresponds to the electrostatic potential due to surface charges.

The implementation of KPFM is done via an additional AC voltage applied in between tip and sample. The overall voltage is now

$$U = \big(U_{CPD} - U_K + U_{AC} \cdot \sin(\omega t)\big). \tag{A.1.6}$$

The energy stored in the capacitor is

$$E = \frac{1}{2}CU^2 \quad \Leftrightarrow \quad F = \frac{1}{2}\frac{dC}{dz}U^2, \tag{A.1.7}$$

where F is the force acting in between the tip and sample. The force acting on the cantilever is measured via the beam deflection. When U is plugged into equation A.1.7, the force has three components $F = F_{DC} + F_{omega} + F_{2\omega}$. The DC part

$$F_{DC} = \frac{1}{2}\frac{dC}{dz}\left(\big(U_C PD - U_K\big)^2 + \frac{1}{2}U_{AC}\right) \tag{A.1.8}$$

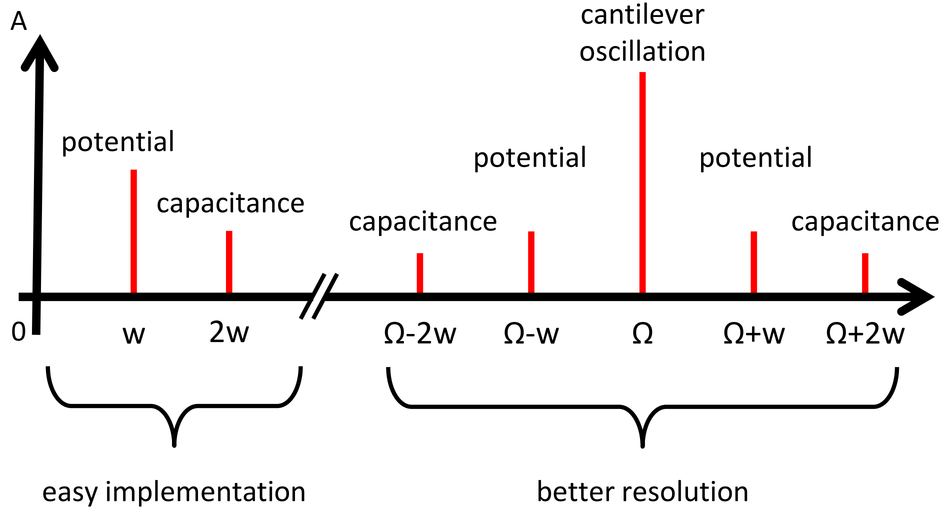

Figure A.1.6.: *Beam deflection (TB) frequency spectrum in KPFM: ω and 2ω terms (left) can easily be measured directly. Amplitude modulated terms (right) offer better resolution, but need higher experimental effort to be extracted. Here, the reference frequency (e.g. $\Omega + \omega$) for the lock-in amplifier must be generated via mixing.*

does not depend on the frequency ω_{AC}. One part contains a component at ω and is given by

$$F_\omega = \frac{dC}{dz}\left((U_{CPD} - U_K)^2 + U_{AC}\sin(\omega t)\right) \tag{A.1.9}$$

and can be used to extract the contact potential difference via nullifying F_ω or the component in four quadrant diode signal, respectively. Another component is found at 2ω given with

$$F_{2\omega} = -\frac{1}{4}\frac{dC}{dz}U_{AC}^2\cos(2\omega t). \tag{A.1.10}$$

It can be used to determine the nanoscale capacitance.

The impementation of KPFM can be performed with either tapping or non-contact mode with the use of lock-in amplifiers. To avoid influences of large-scale electrostatic forces acting in between cantilever and sample the signal should not demodulated directly on the frequencies ω or 2ω. Instead, the mixed frequency of the cantilever oscillation Ω and the Kelvin modulation frequency ω can be used to extract only the influences of the forces acting in between tip apex and sample. To extract the local surface potential, the signal is demodulated at $\Omega + \omega_{AC}$ and the output is inserted into a p-i control loop to keep the output at zero at via the application of U_K. A sketch

of the frequency spectrum is shown in A.1.6.

A.2. Atomic Force Microscope

The atomic force microscope utilized in this work was specially designed for the application of THz-SNOM and for pump-probe-SNOM experiments requiring more than one laser source at once. Hence, illumination and light collection should be possible from both sides. To fulfill this demand, a slim AFM setup offering a large field of view to the tip is necessary. Although this light design comes at the cost of mechanical stability, the AFM should still allow for topography measurements with sub-monolayer resolution in the Ångström range.

The AFM should be applicable to most common modes like contact, tapping and non-contact mode. As Scientists of all fields are more and more interested in electronic properties on the nanoscale, the tip must be properly wired to apply an electric potential or to measure the current flowing from tip to the sample. This allows for PFM, KPFM and various other measurement types.

The design was finished by the author during his time of his diploma thesis and was afterwards constructed with CNC-milling in the mechanical workshop of the Faculty of Physics at the Technische Universität Dresden. With additional home-build and commercial parts the AFM was first assembled by Jonathan Döring and afterwards characterized in detail by the author. In this chapter, the AFM is described schematically and fundamental tests of several operation modes are presented. Although not all measurement types have actually been used to gain the main results of this work, this was an essential and also time-consuming part of the authors' PhD time and is consequently described in detail in this chapter.

Setup

The AFM consists of two main parts, namely the foot and the head, which can be separated from each other in a matter of seconds. This allows for quick tip changes and easy maintenance. An experienced scientists needs less than ten minutes for a tip change and a complete realignment of the device. During the measurements the head is placed on the foot via three micrometer adjustment screws [1] fixing the position of the head. The screws have ball-rollers at their ends which perfectly fit into a hole and a elongated groove in the foot and the re-positioning of the AFM head after replacing it on its foot reaches an accuracy of few ten µm. When scanning, the tip always stays at its position and the sample is moved with a xyz scanning stage [2]. This is mandatory for near-field microscopy as the whole optical setup is fixed and aligned onto the tip.

The **foot** of the AFM acts as a heavy duty basement of the whole device. Two differently sized pillars are connected by the base plate and the larger one offers the hole and the groove for two micrometer adjustment screws. One hole is simply circular to fix the position of the head at one point, the other elongated groove fixes two angles and thus a second defined point. The third point is defined by the third micrometer screw at the pillar, completing the three fixed points needed to completely define three-dimensional objects' position. With this three-point-system the repositioning of the head is highly reproducible, a very important requirement of the system: The tip should be almost at the same position after a tip change to avoid unnecessary and time-consuming realignment of the optical setup for the near-field measurement - especially when beamtime is limited at large facilities such as ELBE at the Helmholtz-Center Dresden-Rossendorf.

The foot is made of steel, which has enough stiffness to achieve very high eigenfrequencies and allows for a measurement with low mechanical noise. The bottleneck of the mechanical stability is the thinner pillar, but its strong eigenmodes oscillate in x and y direction, whereas the the adjustment screws' position is not fixed in that direction on this pillar, hence the impact of this bottleneck is minimal. The sheer weight of the foot is roughly three *kg*, which reflects the mechanical stability.

The **head** of the AFM is shaped like the large letter T, hence it is referred to as the T-Bone AFM. On each of its three ends, a clearance hole is drilled to attach the micrometer screws. The screws allow for choosing the height (and also the angle if necessary) of the head with respect to the foot. In this configuration, different sample scanners (e.g. piezo tubes or compact commercial systems) can be used with the system. The head is completely made of Invar-steel, a iron-nickel allow with

[1] Newport Corporation, Model 9302

[2] attocube systems AG, ANSxyz50 or ANSxyz100

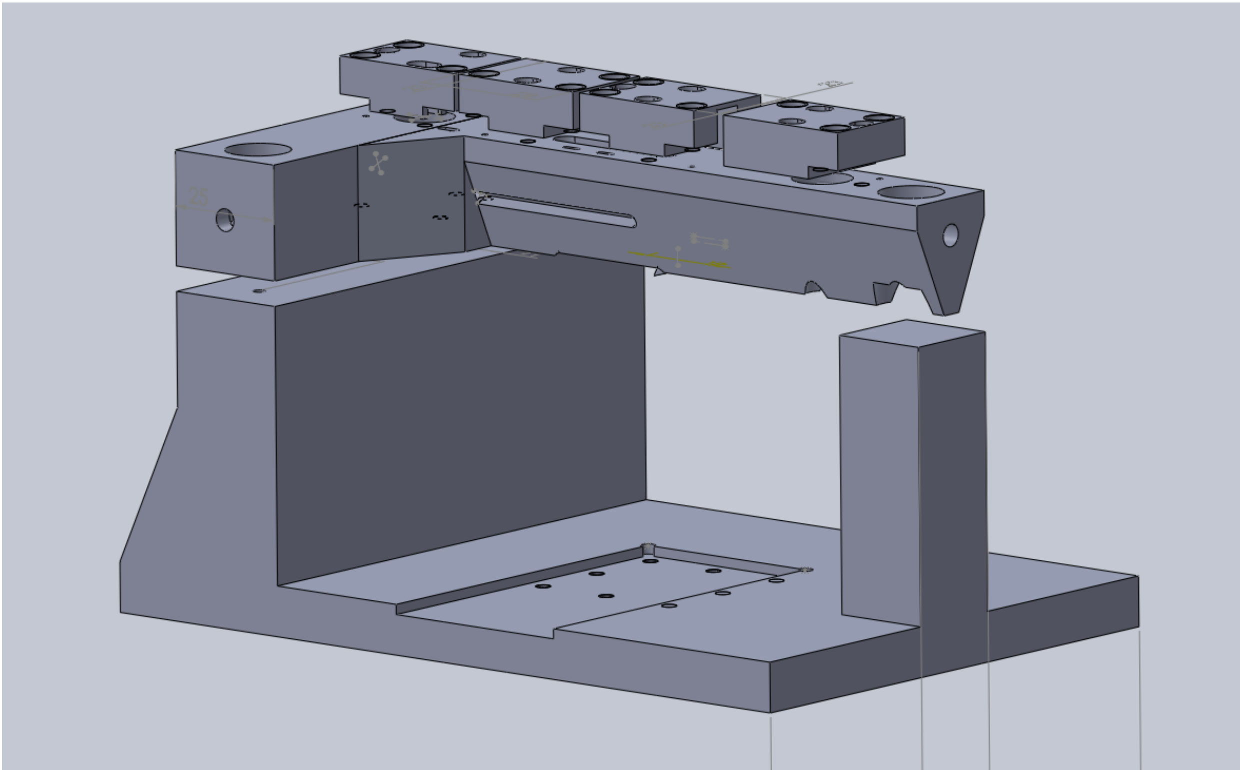

Figure A.2.1.: *Screenshot of the T-Bone AFM in the Solidworks software. The foot, head and mirror elements are shown. The commercially available parts (micrometer screws, positioners and scanners are not included in this model).*

an extremely low thermal expansion coefficient of $\alpha_t = 1.7 \cdot 10^{-6}$K^{-1} at room temperature. This is very useful to avoid any influence of temperature changes on the position and the alignment of the head. The discovery of this material was awarded with a Nobel Prize in physics to Charles Édouard Guillaume in 1920, as it is very well suited to produce length standards.

The design of the instrument was carried out in the commercially available software SolidWorks, a 3D-CAD program with various tools to engineer devices for a vast number of applications. A screeshot of the AFM in the software is shown in figure A.2.1.

To have the possibility to detect bending of the cantilever as well as torque acting on it, a beam deflection optic was chosen for the detection of the cantilever's movement. The tip and beam deflection optics are installed on the long-axis of the AFM head (see figureA.2.2). The whole design is conceived for maximum stability and minimal incoupling of stray light into the detection path. To illuminate the tip, a commercially available laser-diode [3] and the corresponding driver

[3] *IMM Photonics GmbH, IMK-0714-E-K-U-LD-650571A*

[4] are used. For highest power stability, the power supply is established via batteries. The red laser beam is then focused with a lense [5] on the cantilever. The beam direction can be adjusted with two mirror elements which are mounted with the same concept as the head is placed on the foot: Three micrometer adjustment screws are used to tilt the angle of the mirror elements in the desired way to steer the beam with mrad precision. When correctly aligned, the position of each element is fixed with two allen screws. With the height of the elements the best focussation of the beam on the cantilever can be reached. The elements are also designed in a heavy duty way and made of Invar steel to achieve the best mechanical stability. The actual reflecting elements are gold-coated pieces of a 500 µm thick silicon wafer, which are glued into each element. The reflected beam is then directed with two additional mirror elements through a lense [6] onto a four-quadrant silicon diode, which is mounted at the lower end of the head. The mirror elements have a compact design to avoid any stray light from incoupling into the beam path.

The four-quadrant diode[7] is a silicon based photodiode with an active area of 5x5 mm^2. And detects light with wavelengths of $\lambda =$ 320 - 1100 nm, with a peak sensitivity of 0.72 A/W at $\lambda_p =$ 960 nm. The diode can be operated with a reverse voltage of up to 30 Volts, however, in order to avoid heat loss, the voltage is set to 5V here. The diode has an intrinsic cutoff frequency of $f_c =$ 25 MHz at a reverse voltage of 10 V and a load resistance of 50 Ω. At λ_p, the noise equivalent power is 1.4$\cdot$10^{-14} W/Hz$^{1/2}$. For the AFM used at the free-electron laser FELBE, the signal is transferred through two individial preamplifiers, where the second stage amplification can be adjusted via different resistor settings. For the standard settings, the bandwidth is still in the range of 10 MHz. The AFM used for both the sideband and the TDS measurement, the signal is amplified with a commercially available preampfilifier.[8]

The signal of the 4Q-preamplifier is then coupled into an interface[9] where is is processed and passed to the AFM controller.
The cantilever is mounted with a spring to a commercially available alignment chip[10] for µm-accuracy realignment after a tip change. For non-contact modes, the tip has to be mechanically driven at its resonance frequency. To achieve this, a piezo-ceramic material is connected with two electrodes and placed in between the head and the alignment chip. It is electrically insulated from

[4] IMS-WKL-O1
[5] Edmund Optics model 45-489, 84 mm focal length
[6] Edmund Optics Model 33-355, 36 mm focal length
[7] Hamamatsu Photonics K.K., model S5980
[8] RHK Technology, Inc. Beetle AFM preamp
[9] RHK Technology, Inc.,R9 PSD Interface
[10] *NanoWorld AG, Nanosensors PointProbe® Plus XY-Alignment Series*

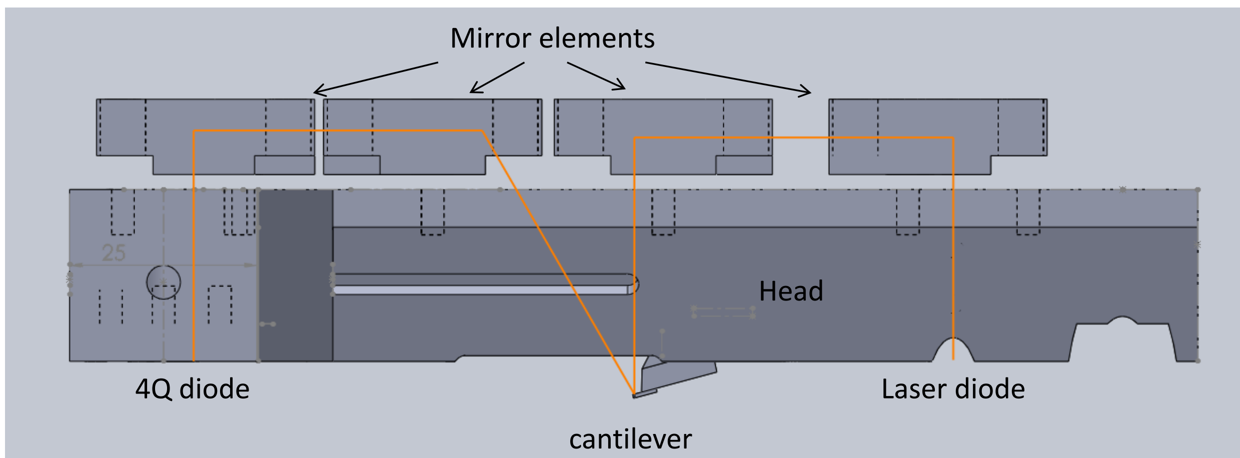

Figure A.2.2.: *Screenshot of the T-Bone head in the Solidworks software. The orange line displays the beam deflection beampath starting at the laser diode and ending at the 4-quadrant diode. The commercially available parts (e.g. the micrometer screws) are not included.*

the AFM by two sheets of ceramic[11]. At resonance frequency, an AC driving voltage of less than 10 mV_{pp} is applied to achieve a cantilever oscillation of 50 to 100 nm.

Positioning and scanning is performed with a compact stack [176] consisting of two lateral [12] and one vertical positioner [13] and one xyz scanner [14]. The full positioning range is 5x5x5 mm^3 and the scan range is 40x40x4.4 (or 40x40x24) μm^3. The whole stack allows for atomic resolution in a stable environment.

All measurements shown in this chapter were performed with a commercial, and fully digital AFM controller[15]. The feature of this state-of-the-art device are well explained in [177]. Most important are six internal, fully digital Lock-In amplifiers for the application in s-SNOM allowing for the demodulation of several harmonics at once. The scanning procedures in the software[16] were provided by Steffen Porthun (Schaefer Technologie GmbH / RHK Technology, Inc.) and Dr. Peter Milde (Institute of Applied Physics, Technische Universität Dresden) and then modified for this special application.

The head also offers several cable passages to establish more measurement modes like PFM,

[11] MACOR, Corning Inc.
[12] *attocube systems AG, Px101*
[13] *attocube systems AG, Pz101*
[14] *attocube systems AG, ANSxyz50 or ANSxyz100*
[15] *RHK Technology, Inc., R9*
[16] *RHK Technology, Inc., Rev9*

Kelvin Force Microscopy and others. The following sections describe the established techniques and give specifications for each mode to display the performance of the device. For a general explanation of these modes, please refer to section A.1.

Contact Mode

When first testing the AFM, dozens of parameters of the device are unknown and vary with the utilized AFM tip type. Hence, a simple AFM mode is a good choice to start with test and specify the device and the control parameters before moving to more complex modes. Conclusively, the contact mode was chosen. The most important unknown parameters are the setpoint and gain values of the distant control as well as the scanner calibration. To start with, the following guidelines fro setting up such a device can be used.

When the beam-deflection is aligned to the center of the 4Q-diode, the setpoint[17] can be chosen to roughly 10% of the overall signal on the diode. However, the correct direction of the setpoint (negative or positive values) is unknown due to the many stages where the AFM signal is processed. The proper algebraic sign must be found by trial-and-error. The p-gain and i-gain should be set to a value that full z-piezo distance will be traveled in several seconds. The algebraic signs again play a role, as they depend on the definition of the gain formula used by the AFM controller. RHK and most of the companies supplying AFM controllers use formula in a way that both p- and i-gain should have the same sign.

After using these guideline, test measurements in the contact mode were performed on a calibration grid[18]. Figure A.2.3 shows a scan of an calibration grid obtained with a *BudgetSensors, ContE-G, Innovative Solutions Bulgaria Ltd.* tip. The Si/SiO_2 stripe structure has a pitch of (4 ± 0.1) μm and a height of (20 ± 2) nm. As the stripes are parallel and equally spaced, the non-linearity of the scanner can be observed. In this scan the non-linearity is neglectable for values below 60% of its maximum travel range. Due to the properties of piezo scanner materials (hysteresis, creep, ...), this will drastically change with the scan speed and linetime. The non-linearity can easily be noticed in the top right quadrant of the scan. Compared to the lower left quadrant, the stripes are enlarged by roughly 55% in both dimensions. Various surface roughness values can extracted with the open source software Gwyddion [147] and the RMS (root mean square) roughness is $\approx$2.4 Å along the stripes for the raw data. When corrected with a line slope filter (b), it decreases

[17] In this case the top-bottom signal.
[18] Anfatec Instruments AG, UMG01 [178]

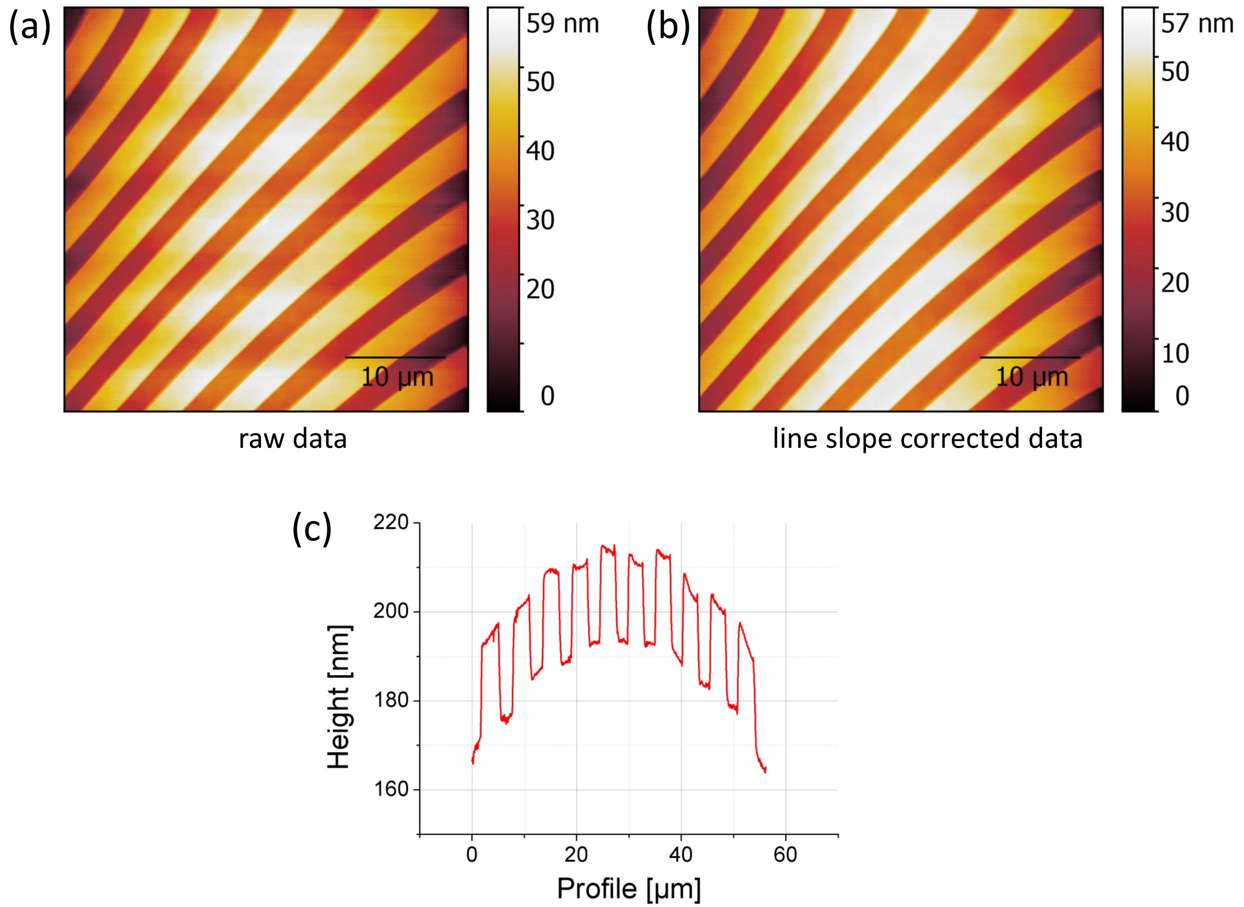

Figure A.2.3.: *1024x1024 pixel, $\sim$ 40 μm scan of UMG01 calibration grid. Raw data is shown in (a). The topography stripes reveal the 20 nm steps of the Si/SiO$_2$ structure. (b) shows the line slope corrected data with slightly lower roughness. An extracted profile is shown in (c). Here the quadratic scanner artifact in z-direction can be observed. Please note that this first measurement did not yet have a correct scanner calibration.*

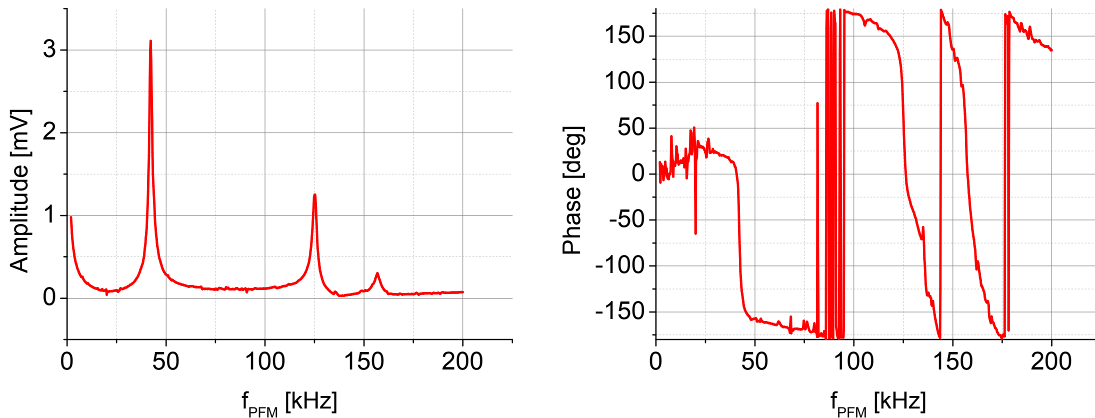

Figure A.2.4.: *Out-of-plane PFM frequency sweep on a PPLN sample. In the range of 2 to 200 kHz, three peaks are visible and connected with strong changes in the phase diagram by 180 degree.*

to ≈1.9 Å. The z-piezo non-linearity can be observed in (c). The extracted line profile (top left to bottom right of the scan) shows a parabolic behaviour with a 20 nm change on the full scan range, a small value - especially when compared to piezo tubes. As this scanner artifact is quite common in AFM, *Gwyddion* offers data postprocessing prozedures to correct this parabolic behaviour (not shown here).

Piezoresponse Force Microscopy

Piezoresponse Force Microscopy (PFM) is a commonly used AFM mode to characterize the piezoelectric response of solids on the nanoscale [REFs]. The mode can be run in parallel during contact mode measurements. It simply needs an AC voltage supply and one or two Lock-In amplifiers (LIA) in addition to the standard contact mode AFM with a conductive tip.

In order to map the ferroelectric domains, a suitable AC voltage and frequency needs to be selected. When the tip is in contact this is best done with a frequency sweep, measuring the amplitude and phase response with constant AC voltage. A typical frequency response is shown in fig. A.2.4. The peaks correspond to mechanical resonances of the tip-sample system. For different applications, suitable voltages can lie in totally different regimes. For materials with strong piezoelectric response an off-resonant frequency below 10 kHz can be selected. This setting is most stable against small changes in the system (e.g. drifts, dirt pick-up, tip wear-off, ...). However, the

PFM signal for this case is usually quite small. For samples with smaller piezoelectric response, the frequency of a resonance peak is often a better choice, as the PFM amplitude is much larger and the phase is very sensitive to slight changes of piezoelectric responses (e.g. by different types of domains). The measurement is also much more sensitive to changes of the environment, as small changes of the resonance frequency can change amplitude and phase dramatically.

To evaluate the performance of this mode, lithium niobate (LNO) has been examined. The crystal has a polarisation along the z-axis which can be either pointing up or down. When heating above the Curie temperature of 1150 °C [179] and cooling down again, domains form in a random pattern with the two different polarizations (so called c_+- and c_--domains) [180]. Due to the large piezoelectric response of these domains can be easily observed with PFM and thus make LNO a very good material for PFM test measurements. Furthermore, the polarisation of LNO can be deliberately changed by large electric fields. With nano-structured electrodes, domains can be written into the crystal with desired forms and sizes. One design is a pattern of large stripes with the width of few μm, so called periodically poled lithium niobate (PPLN). Typically used for its application in quasi-phase matching in non-linear optics, it is commercially available at low costs. The capability of the T-Bone to map nanoscale domains is demonstrated with such a sample.

For a perfectly aligned AFM and a homogenius electric field[19] the PPLN measurement no contrast in the out-of-plane amplitude channel is expected as c_+- and c_--domains should have the same response. The phase should show a 180 degree contrast in between the two domains, as the polarisation shows in the opposite direction. The out-of-plane channel should theoretically show no response and thus no contrast at all.

The PPLN measurement is depicted in fig. A.2.5. The topography (a) is relatively flat and has a RMS roughness of few nm. It shows a small amount of dirt particles on the sample of which one is picked on the last third of the measurement. Ferroelectric domains cannot be observed with the topography data. The PFM AC voltage was selected off-resonant (< 10 kHz) and with 1 V_{pp} amplitude. The out-of-plane PFM amplitude (b) shows a contrast in between c_+- and c_--domains. 7 μm thick stripes are visible in the whole scan range. Depending on the position in the scan the response of the brighter stripe can be more than twice as large as the darker lines. The phase difference (c) of the stripes is up to 140 degree and does not reach the expected value due to an offset in the Lock-In measurement. Notwithstanding, the domains can clearly be mapped in with this measurement.

[19] Due to the finite tip size, the field will never be homogenious.

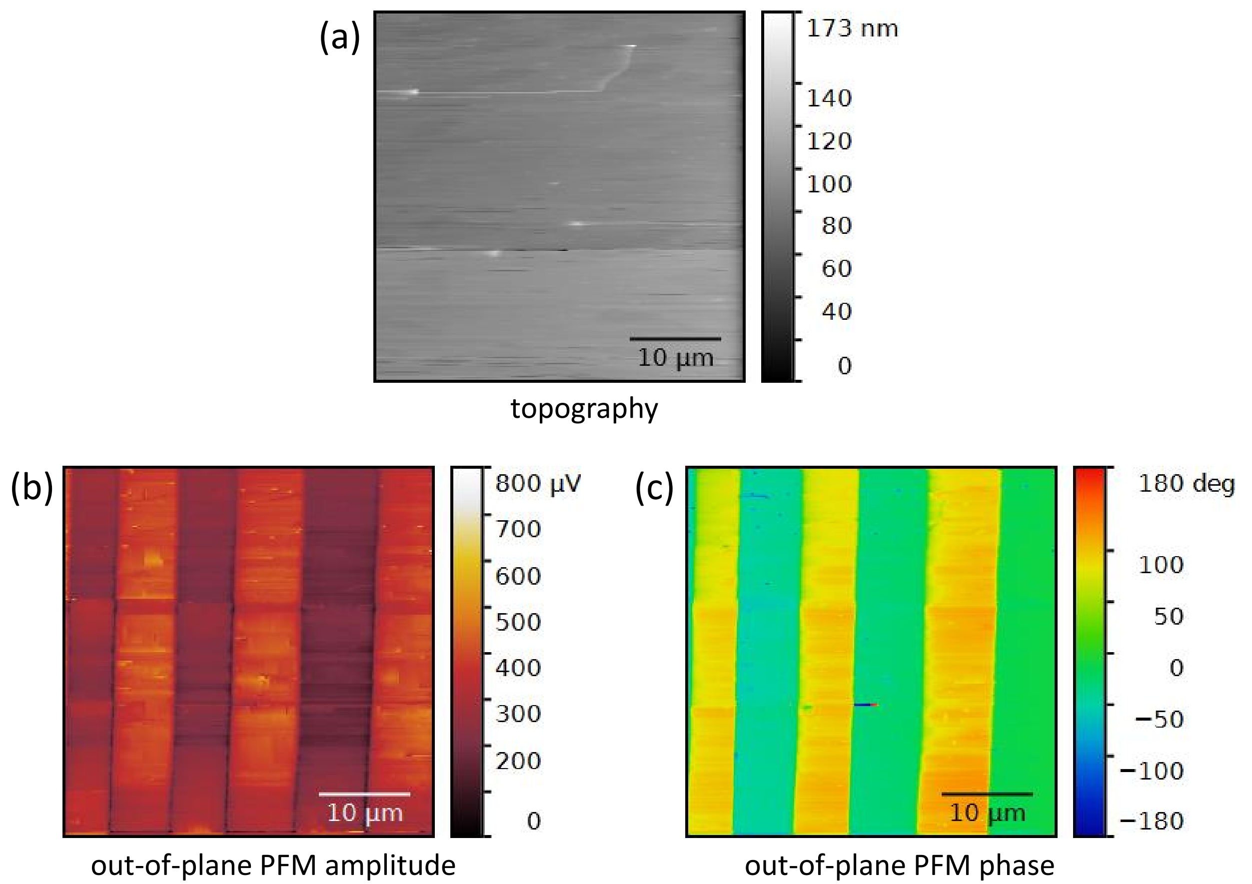

Figure A.2.5.: *512x512 pixel PFM measurement of PPLN: (a) shows the topography. (b) and (c) show the out-of-plane amplitude and phase response.*

Tapping Mode

A very commonly used AFM mode is the tapping mode, where the tip oscillates above the sample surface and the oscillation amplitude is controlled with a p-i control loop A.1. It is also often called intermittent contact mode and is mostly applied at room temperature and atmospheric pressure. It features high stability and is user-friendly and very fail-safe due to the simplicity of the distance control. In almost any case of a complication during a measurement (e.g. dirt-pick up), the tip will be pulled away of the surface instead of crashing into it. Using modern, commercial AFMs, a simple topographic characterization of a sample can be done in few 10 minutes, including the mounting of tip and sample - another reason for the wide distribution of the mode.

Firstly, the tip resonance frequency of the free oscillation must be found when the tip is mechanically excited. A typical sweep with an AC excitation voltage of few mV is depicted in figure A.2.6. An overview sweep is shown in (a). A clear and sharp peak in the amplitude is visible at 250 kHz, associated with a phasejump of π. A zoomed sweep in depicted in (b). The deviation from a damped harmonic oscillator (which would show a Lorentzian distribution) is a consequence of a crosstalk with different mechanical resonances close to the main peak. The quality factor Q of a resonance peak at f_{peak} with an amplitude A_{peak} is determined by the width Δf of the peak. Δf is give by the frequencies where the mechanical amplitude has fallen to $A_{peak}/\sqrt{2}$. For this example, the Q-factor is $Q = \Delta f / f \approx 450$.

To approach tip and sample a setpoint of $\approx 80\,\%$ of the free oscillation amplitude is a good value to start with. The p-gain and i-gain should be set to a value that full z-piezo distance will be traveled in several seconds. For tapping mode, the p-gain can be three orders of magnitude smaller than the i-gain. When the setpoint is reached and the tip is close to the sample surface, it is necessary to verify that the tip reached the regime of repulsive forces. This can best be checked with a so-called retract curve, where the feedback loop is frozen and the tip is pulled away from the surface in a linear motion whilst the amplitude and phase of the cantilever oscillation is measured. A typical example for this shown in figure A.2.7. Here, the feedback is stopped and the tip-sample distance is increased up to 500 nm. (a) shows the mechanical amplitude of the cantilever. On the first 20 nm, the amplitude rises quickly as the cantilever shortly touches the surface during each oscillation. For larger distances, the amplitude rises much slower due to long-range electrostatic forces interacting in between tip and sample. The small oscillation is an artifact resulting from interference of the light reflected from the cantilever and stray light from the sample surface. The strength of the interference can change with the alignment of the beam deflection and the sample. For this measurement its size is neglectible. If the interference is larger, the beam-

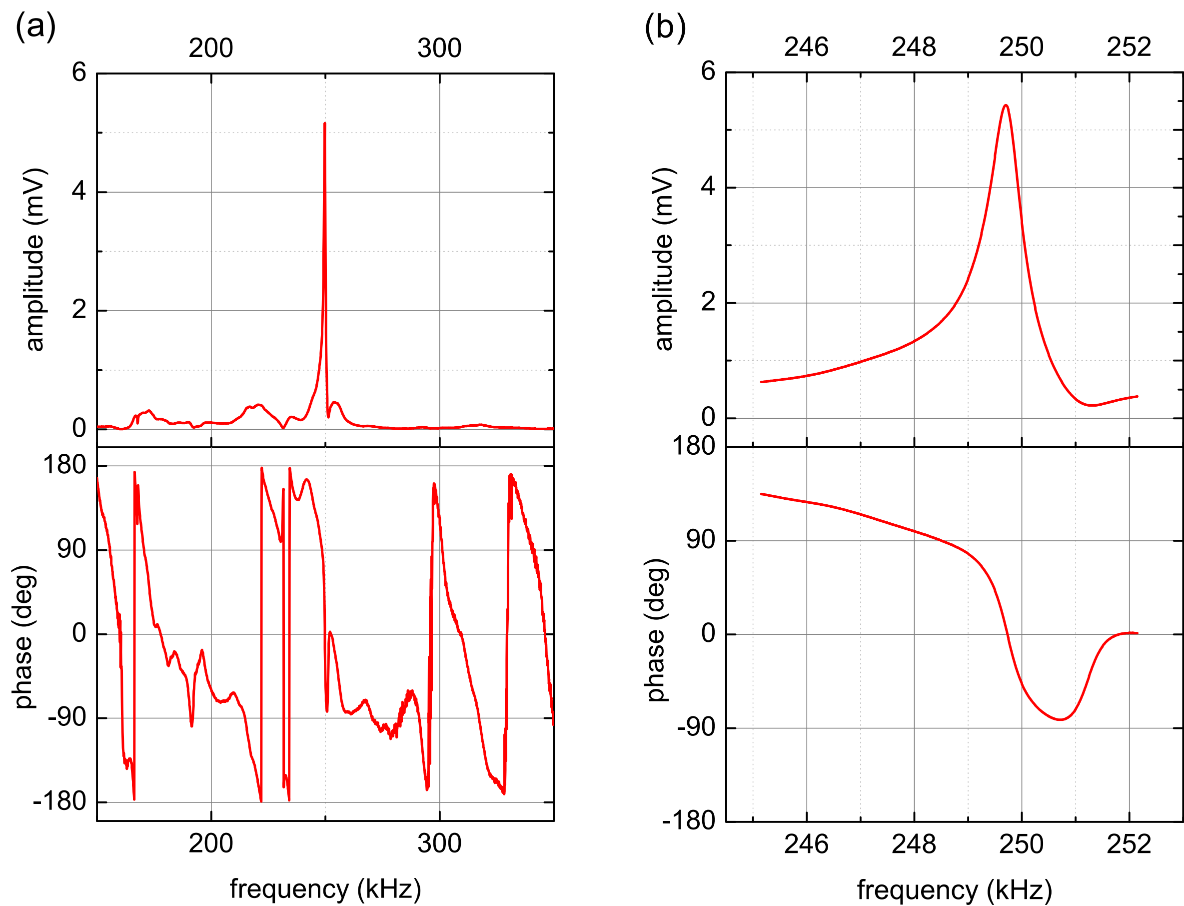

Figure A.2.6.: *Frequency sweep of a BudgetSensors Tap300E-G tip. An overview sweep is shown in (a). The first-order mechanical resonance of the cantilever is at* $\sim$*250 kHz and connected with a phasejump of* π*. A zoomed sweep is shown in (b). The* Q*-factor of the resonance is* $\sim$ *450.*

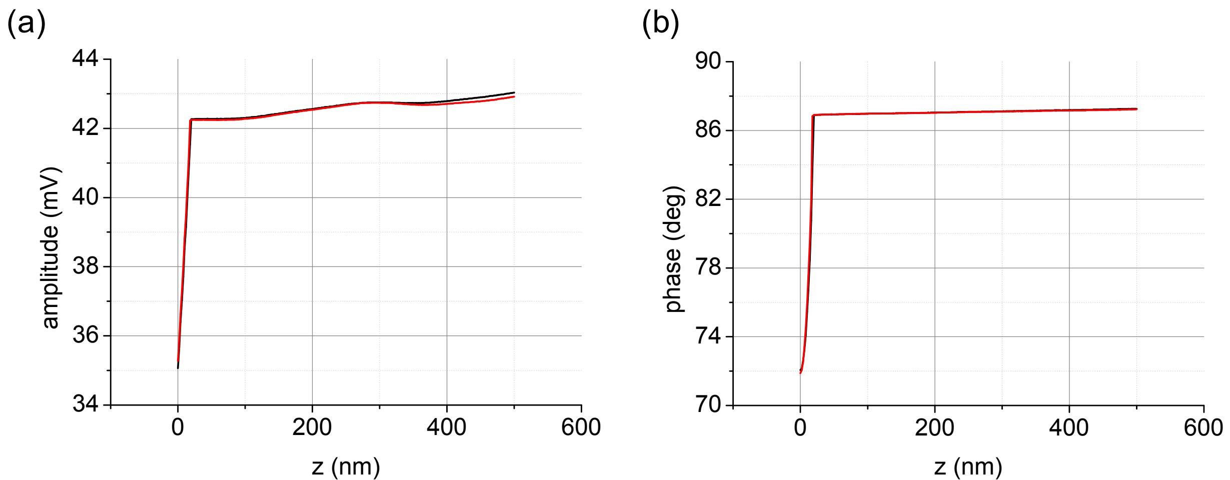

Figure A.2.7.: *Retract curve on the UMG01 grating. (a) shows the amplitude and (b) the phase data as a function the tip-sample distance. The measurement as been performed twice (black and red curve).*

deflection must be realigned. In most cases the focus at the cantilever is too large and must be improved with the first two mirror elements. For near-field microscopy (and many other measurement types) this measurement is of high importance and must be very reproducible which is the reason always to collect at least two curves for each measurement. The corresponding phase data is shown in (b). It also has a strong change on the first 20 nm and is relatively flat for higher distances.

To investigate the performance of the tapping mode, the *Anfatec UMG01* test grating was measured again. Figure A.2.8 shows the topography of the grating. Here, the forward (piezo voltage rising from zero) and backward (piezo voltage decreasing to zero) scanning directions are compared. While the forward direction (a) suffers strongly from non-linearity in the top and right areas of the scan, the backward direction is less affected by this problem and shows the strongest non-linearity when the voltage becomes small again. The origin of this behaviour is the creep and the hysteresis of the piezo scanner. As a consequence both directions should always be checked whenever the dimensions of a certain structure is unclear. Alternatively, a closed-loop scanner can be implemented into the device which costs are three to four times larger than the costs of an open loop scanner. The RMS roughness along the stripes lies in the range of 500 pm to 1 nm - too large for the investigation of monolayers. As huge scientific interest lies in the investigation of 2d-materials such as graphene or sheets of transition metal dichalcogenides, the capability to map single atomic steps needed to verified.

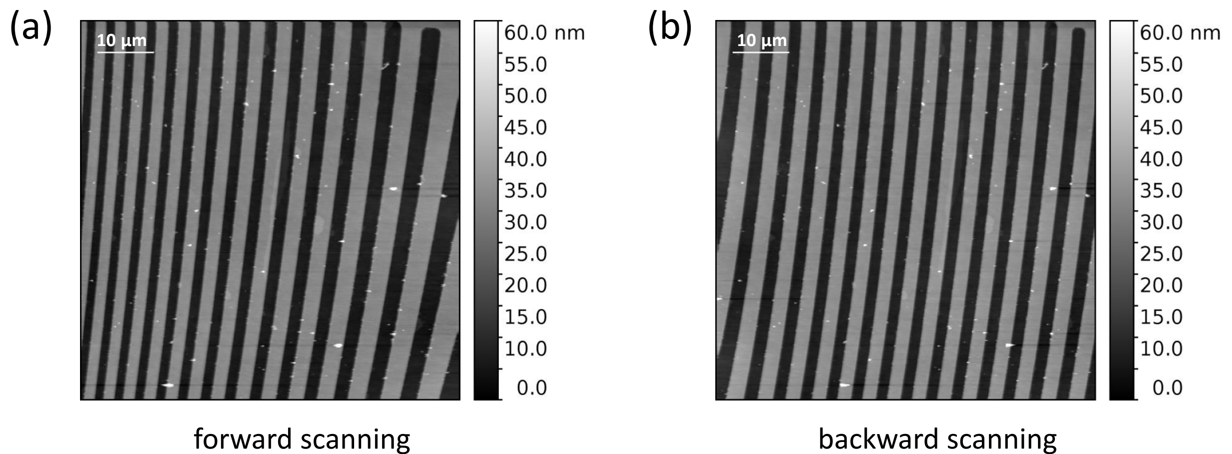

Figure A.2.8.: *Forward (a) and backward (b) scan in tapping mode of the UMG01 test grating after the substraction of a second-order polynomial background. Non-linearities occur stronger in the forward direction.*

A common sample with such single atomic steps is highly oriented pyrolytic graphite (HOPG). The creation of such steps can simply be achived via cleaving of the sample. A topography scan of a HOPG surface is depicted in figure A.2.9. To reduce noise in (a) a 2d FFT filter is applied. Secondly, a second degree polynomial background is substracted to remove scanner artifacts. Now the small steps in the topography are clearly visible. The red line shows the position of the extracted line profile in (b). A small step in the topography occurs at $x \approx 2.4\,\mu$m. The step height of 3.4 Å matches well to the literature value of 3.35 Å [181]. This measurement clearly proofs the AFMs' capability to map single atomic steps. In the raw data, the RMS roughness at the leayers is $\approx$ 1 Å, which marks the lower limit of the T-Bone AFM in tapping mode.

Non-contact Mode

The non-contact mode (also referred to as nc-AFM, *True Non-Contact* mode, dynamic force microscopy) also uses an oscillating tip to map the topography of a sample. However, the p-i control loop measures and controls the the oscillation amplitude and the phase simultaniously. This guarantees a constant oscillation amplitude of the tip at all times and can reduce artifacts from steps in the topography. Also, this mode has an extreme sensitivy to map smallest forces with very high resolution [167, 182] as the oscillation frequencies in the range of 100 kHz can be measured with few mHz precision. This mode is rather suitable for flat samples and not very fail-safe if large features apprear the topography. In case of an unwanted incident like a dirt-pick up or a too

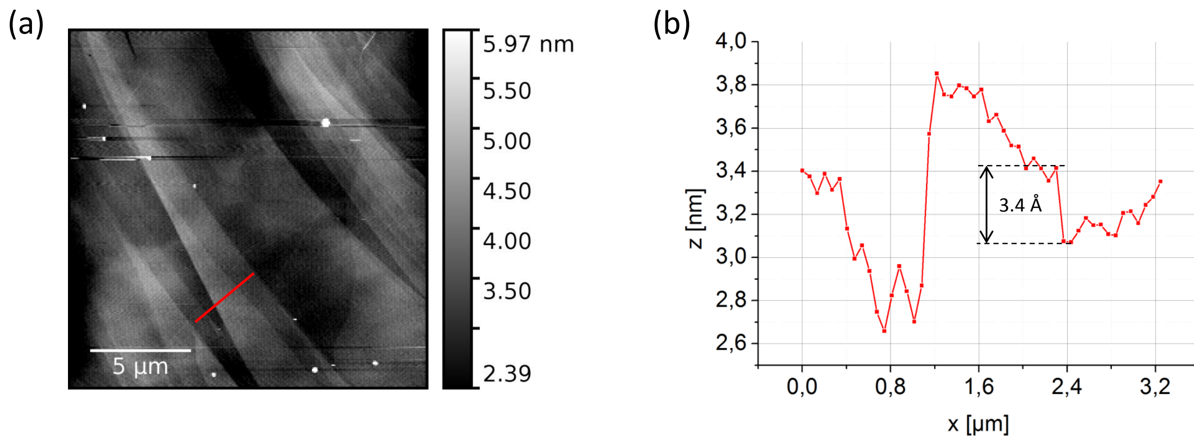

Figure A.2.9.: *512x512 pixel, 17.5 μm scan of a HOPG sample. (a) shows steps of different height in the topography. The red line marks the position of the extracted profile depicted in (b). Two steps are clearly visible, a multilayer step at $x \approx 1.2$ μm and a single step at $x \approx 2.4$ μm.*

large feature in the topography the tip will crash into the surface in almost any case. This mode also requires a much more experienced user compared to the tapping mode, as more parameters need to be set. Besides the control loop for the oscillation amplitude, a second p-i control loop needs be configured for the setpoint Δf of the desired frequency shift. The parameters strongly influence each other and must also be varied from sample to sample. To find good starting values before approaching the sample, a procedure is installed in the Rev9 software to test if the control loop can follow changes in amplitude and phase precisely enough for a real measurement. It changes the setpoints in between two values with a certain periodicity and the gain values can are adjusted in a way that the control loop can reach the changed setpoints in a reasonable time without overshooting.

The following two test measurements show the capability of the T-Bone to utilize this mode and also demonstrate an example of how strongly the settings can change on different samples.

A measurement of the UMG01 test grating is depicted in figure A.2.10. The topography data (second degree polynomial corrected) for the right-to-left direction is presented in (a). The stripes are clearly visible, however the tip used for this measurement is blunt, which explanes the addional structure on the lefthand side of each stripe. The surface roughness is in the range of 1 Å on the stripes. The frequency shift is shown in (b). The setpoint is -20 Hz and on the stripes it is controlled with a precision of 100 mHz (RMS). This is an excellent value considering this mea-

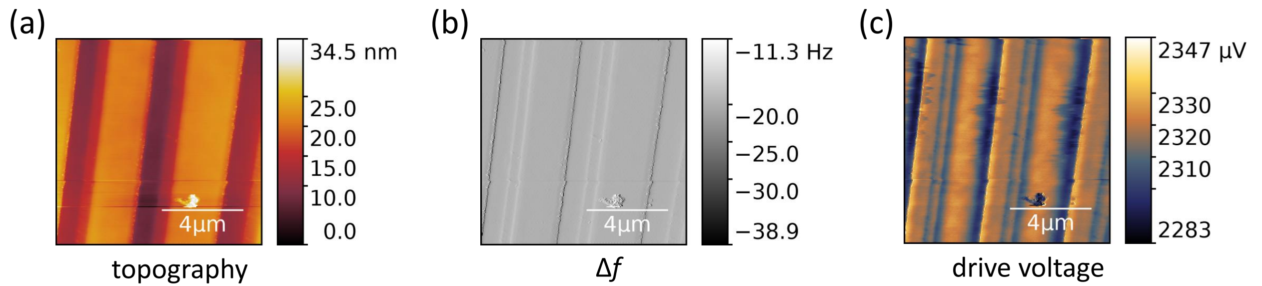

Figure A.2.10.: *256x256 pixel, 10 μm scan of the Anfatec UMG01 test grating. (a) shows the topography. The 4 μm pitch stripes with 20 nm height are clearly visible. The frequency shift is shown in (b). When moving onto a stripe, there is a short shift towards lower frequencies. When moving off, the shift goes towards higher frequencies. The drive voltage is depicted in (c) and clearly depends on the scanning direction*

surement has been performed in ambient condition. Whenever the tip moves from the lower to the higher level of the grating, the cantilever gets closer to the surface and the eigenfrequency shortly drops to a lower value due to the increased force interaction in between tip and sample. As the control loops need a limited time to compensate for this shift, thin black lines appear on the righthand side of each stripe. On the lefthand side it is vice versa. Although the edges are not very sharp in the drive voltage (c) due to the worn-off tip, the control loop keeps the drive voltage constant with a precision of less than 1 μV on the stripes.

As a second example and to verify the non-contact mode is capable of mapping single atomics steps the HOPG sample has been investigated another time. Figure A.2.11 (a) shows the topography of the measurement after a second degree polynomial correction and a 2d FFT filter to reduce 50 Hz noise. The small steps of the HOPG are again visible. A profile is extracted at the red line and plotted in (b). Here, the capability of mapping signal atomic steps is proven for this mode. The slight mismatch of a 4 Åcompared to the literature value of 3.35 Åcan be a result of the nonlinear behaviour of the open-loop scanner. The RMS noise of the frequency shift is $\sim$ 200 mHz and in the same range as on the test grating. The RMS noise in the drive channel is one order of magnitude larger though, with 10 μV (not shown).

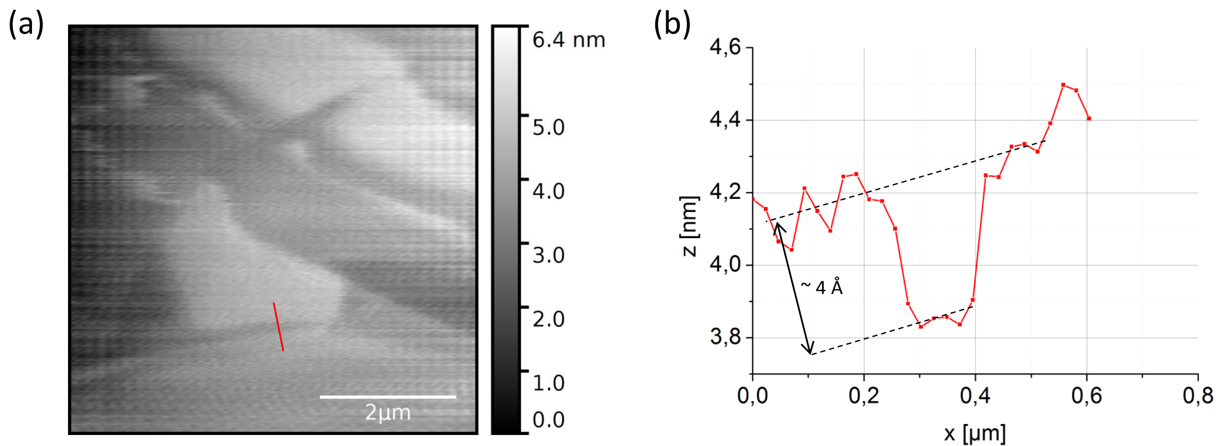

Figure A.2.11.: *256x256 pixel, 6 μm scan of a HOPG sample. (a) shows steps of different height in the topography. The data is post-processed with a 2d FFT filter and a second degree polynomial background correction for the sake of clarity. The red line marks the position of the extracted profile depicted in (b). A single layer valley is in the data. The step height of $\sim$ 4 Å is in fair agreement with the literature vale of 3.35 Å.*

Kelvin-Force Microscopy

Electrostatic interactions in between tip and sample can have huge influences on certain samples and sometimes need to be compensated both in tapping and non-contact mode. Without compensation it is often not possible to reach the surface with the tip which causes artifacts the topography measurement and consequently also in other modes like s-SNOM. For conducting materials and structures it is sufficient to connect tip and sample to the same ground potential to solve this problem. In certain cases (isolators, pyroelectric materials) the electrostatic potential must be actively compensated with a voltage applied to a conductive tip (see chapter A.1 for different implementations of this technique). For example, localized surface charges cannot simply drain to the ground connection through a bulk crystal [64] and are trapped at their position. Their stray fields must be compensated by KPFM.

Here non-contact AFM is combined with KPFM to measure the nanoscale surface potential of a $Si_{0.23}Ge_{0.77}$ thin film sample (figure A.2.12). The Kelvin signal was demodulated at the first sideband of the cantilever oscillation. Whenever there is a mismatch of the surface potential of tip and sample the sideband is non-zero and a control loop adds a DC voltage to the conducting tip to nullify the mismatch. Please note that the demodulation frequency is not a fixed value, but is

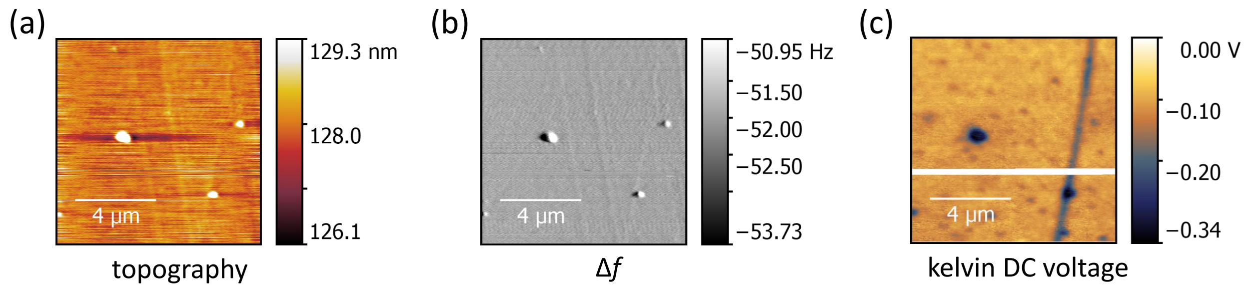

Figure A.2.12.: *512x512 pixel, 10 μm scan of a SiGe thin film sample sample. (a) show the topography after a second degree polynomial filter was applied. Besides a few particles, the topography is very flat. A slightly lower stripe in the data is visible from the top middle to the bottom middle. The Δf signal (b) shows very stable measurement conditions. The Kelvin signal is shown in (c). Here, the particles and one inclinated stripe are clearly visible. After two thirds of the scan, the control loop jumps to a positive value before moving back. The reason for this behaviour is unclear.*

steadily updated as the cantilever oscillation and driving frequency changes. Although expected changes of the surface potential due to the crosshatch-pattern [183, 184] could not be observed, the measurement still shows the capabilities of the *T-Bone* to measure surface potentials.

The topography is depicted in figure A.2.12 (a). The vertical stripe of slightly lower topography is a typical feature of such samples [183, 184]. The lattice mismatch of SiGe and Si is $\sim$ 1% [127] which corresponds to strains in the range of GPa. When the top layer is grown, the impurities and dislocations partially relieve the stress, and create a crosshatch structure often seen as stripes in the topography. Please note the height of these steps is usually even smaller than a single atomic step again proving the very low noise of the topography measurements with the T-Bone. The corresponding Δf measurement (b) shows a stable operation and the frequency only shifts significantly when the tip reaches one of the particles on the sample surface. The Kelvin DC voltage applied to nullify the sideband is depicted in (c). Here, one can clearly observe small changes in the topography even by smallest particles and also one long, slightly inclinated stripe of lower DC voltage going through the whole scan range. When a profile is extracted across this stripe, a jump of 800 mV in the Kelvin voltage can be observed on a length scale of roughly 350 nm. The resolution is not a good value compared to other reports [185], and it is quite likely limited by the structure itself and not by the measurement method. The RMS roughness on areas without features is in the range of 6 mV, which is a good value considering the fact potential differences

of different materials often lie in the range of eV and values of several volts need to be applied on pyroelectric materials [186].

s-SNOM

The main purpose to design the *T-Bone* AFM was the need for a new near-field microscope with large field of view to apply it in several setups covering optical ranges from mid-infrared to THz radiation. The first tests were carried out with a CO_2 laser [20] setup and MCT detector [21] in a self-homodyne detection scheme (see chapter 4.1). This combination of stable laser and sensitive detector with large dynamic range is used on a day-to-day basis in many research groups worldwide. In the follwing chapters the very first s-SNIM measurements obtained with the *T-Bone* AFM are presented.

Infrared Near-field Examination on a Gold Sample

In order evaluated the quality of near-field signal, tip quality and signal-to-noise ratio the measurement of the near-field signal as a function of the tip-sample distance (retract-curve) is commonly used after alignment of the signal to its maximum value. A large layer $\approx$ 100 nm of evaporated gold on a polished silicon waver leads to a relatively large near-field signal and has a flat spectral response and thus is applicable for alignment and as a reference sample for all wavelengths from infrared to THz radiation.

Such a measurement from the developing stage of the T-Bone AFM is shown in figure A.2.13. A cantilever tapping amplitude of 20 nm and a laser power of 15 mW were chosen. The typical near-field decay is clearly visible in (a) for all harmonics (represented with different colors) when the distance in between tip and sample is increased. The second harmonic is drawn in black. Here the near-field signal is the largest. Also, the signal-to-noise ratio has a good quality. However, the demodulated near-field signal does not go down to zero for larger distances although the near-field is only present on the first few 10 nm above the sample surface. This is a far-field artifact as higher-harmonic-demodulation does not yet work perfectly for the second harmonic [118]. The decay length to half of the initial signal strength is an indicator for the achieveable resolution in a lateral scan and is reached at a distance of 24 nm. The third harmonic (red) has

[20] *Access Laser model Lasy 3S*

[21] *Teledyne Judson Technologies model J15D16-M204-S01M-60 with PA-101 preamp.*

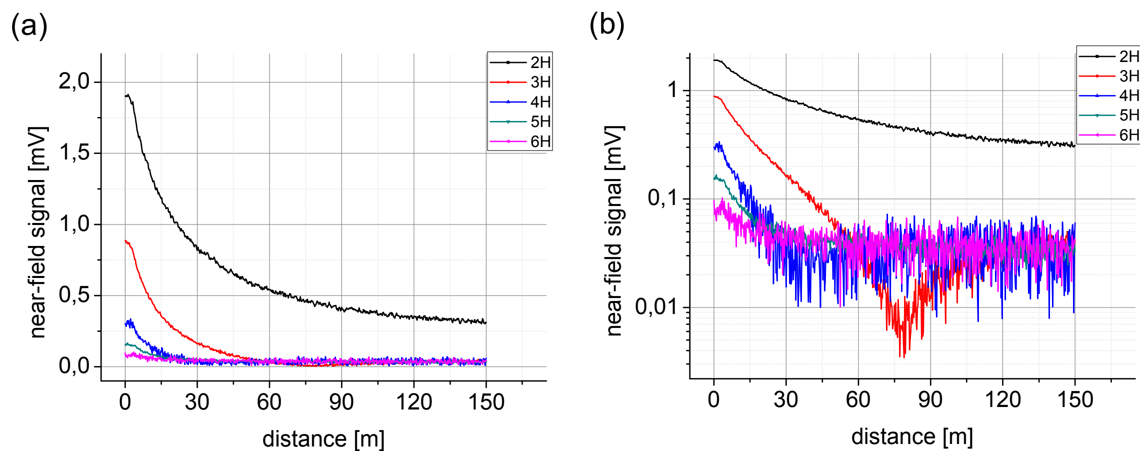

Figure A.2.13.: *Near-field signal as a function of tip sample distance on a gold sample. (a) shows the typical near-field decay for different harmonics (color scale). The same plot is depicted with a logarithmic scale in (b) for better visibility. The dip in the third harmonic signal at 80 nm is a result of a phaseshift in the Lock-In measurement and a residual of far field contributions.*

a much shorter decay length of 11.5 nm already. As the signal goes down to zero for larger distances, far-field influences are neglectable at this point. For the fourth harmonic this value goes down to 7.5 nm. However, the signal-to-noise ratio significantly rises, which is more clearly visible in the logarithmic plot in (b). The noise level is in the range of (40 $\pm$ 10) μV for the third and higher harmonics, whereas the signal level gets lower with each harmonic. The last detectable signal was measured at the sixth harmonic, where the preamp is already confronted with signals above the stated bandwidth limit. At the sixth harmonic, the signal-to-noise ratio is too low for a lateral scan. For qualitative measurements the second harmonic is usually a good trade-off in between signal strength and resolution, for qualitative and high-resolution measurements higher harmonics should be chosen. Please note that the combination of a CO_2 laser with a *J15D16* MCT detector is not ideal. With a *J15D12* detector, the sensitivity is twice as good and increases the signal-to-noise ratio significantly. This upgrade was implemented later to the setup.

Infrared near-field imaging

To dermine the resolution in a lateral scan and confirm that the obtained signals in the retract-curves are no artifacts, a scan on a sample with a defined material contrast is usually chosen.

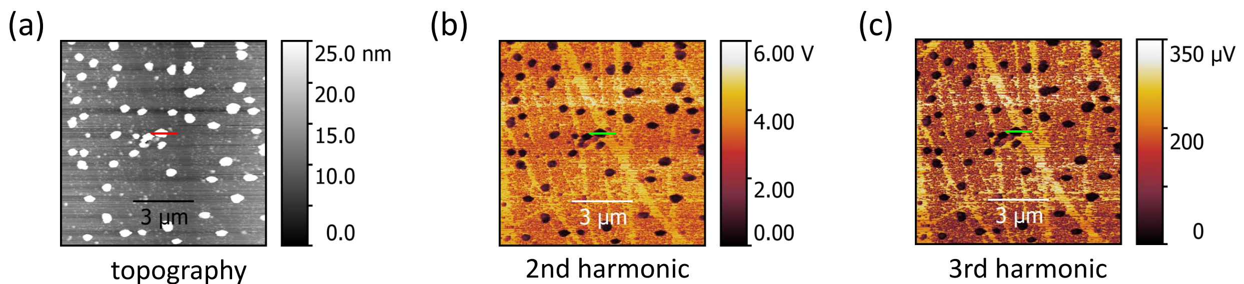

Figure A.2.14.: *512x512 pixel, 10 μm near-field scan of a gold sample. The topography is shown in (a), the correspondig near-field signals of the second and third harmonic are shown in (b) and (c) respectively.*

As the gold samples used to align the near-field usually are contaminated with dirt particles, a material contrast is expected during scanning. As the sample does not need to be changed for this measurement, this saves time and is thus often the material of choice.

The first near-field scan ever obtained with the T-Bone is shown in figure A.2.14. A platinum-iridium-coated tip [22] with a oscillation frequency of 160 kHz was used. The topography (a) shows several larger dirt particles with a lateral dimension of hundreds and a height of several ten nanometers. On a much smaller scale, there are also thousands of much smaller particles visible almost forming a closed film on the sample. The origin is unclear, as well as the way of production, storage and age of the sample. The corresponding near-field amplitudes of the second and third harmonic are shown in (b) and (c). Here, a material contrast is clearly visible in both harmonics in between the large dirt particles and the gold. The largest signals are found on some stripe which seem to be almost completely free of contaminations. The second and third harmonic show he same information, yet the contrast in the third harmonic is larger as far field effects are better surpressed.

From this data, a profile was extracted at the marked position (red and green lines) to determine the resolution of the scan. The data is depicted in figure A.2.15. The topography shows a particle of $\sim$ 600 nm lateral size and a height of $\sim$ 70 nm. The near-field signals of the second and third harmonic were recorded simultaneously using one internal Lock-In amplifier of the *R9* controller for the third harmonic and one external Lock-In[23] for the second harmonic. Please note the second harmonic signal is amplified by a factor of 1000 due to this fact. On the dirt particle, the signals are low, and on the gold they are high. Although there is no clear convention to deter-

[22] *Nanoworld AG, Nanosensors PPP-NCLPt*
[23] *Stanford Research Systems SR 844*

mine the resolution for such a measurement, one way is to measure the distance to overcome 80% of the signal change in between the two levels [7]. Here, a lateral resolution of 60 and 62 nm is achieved for the third harmonic and second harmonic respectively. Higher values can be achieved in s-SNIM [6], however it fits to the value determined in [7] with the same kind of tip.

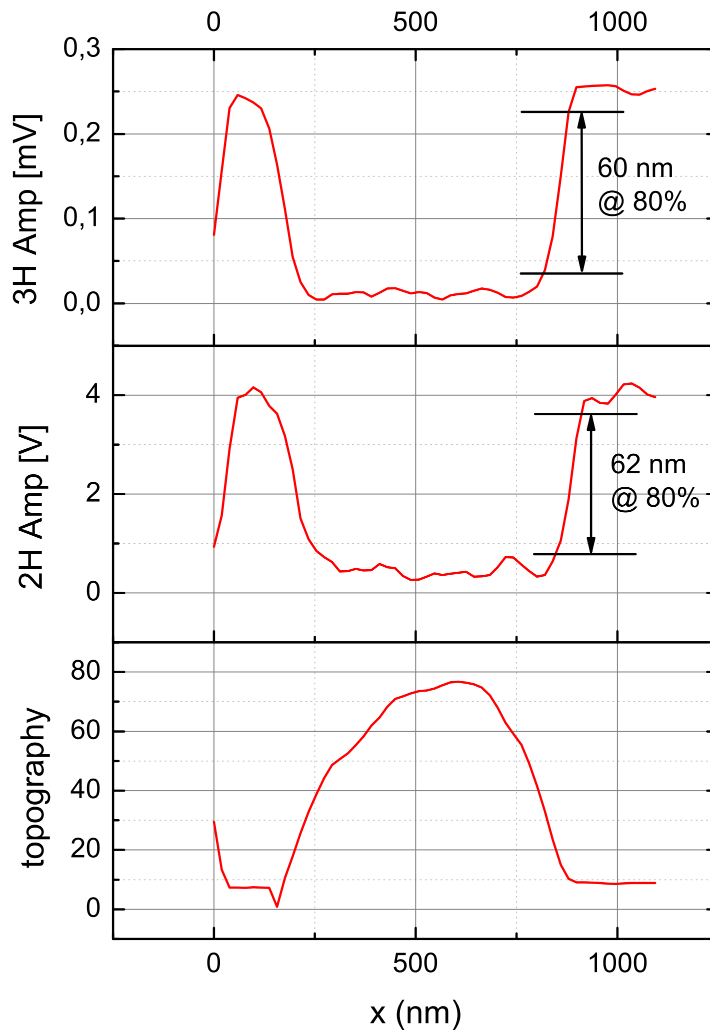

Figure A.2.15.: *Line profile extracted from figure A.2.14 (red and green markers): the topography (bottom) shows a dirt particle with ∼ 600 nm width and ∼ 70 nm height. The corresponding near-field signals (middle, top) show jumps in the signal strength at a length scale of 60 nm.*

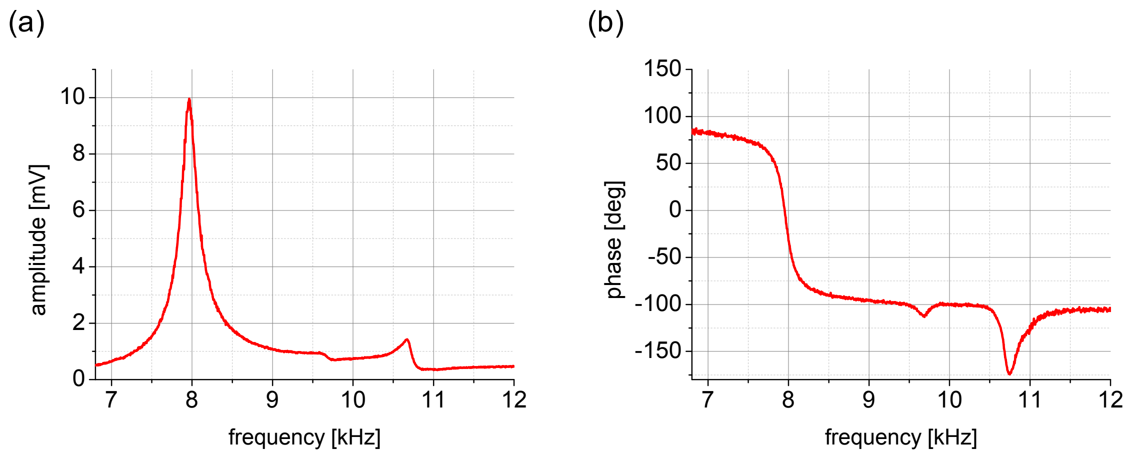

Figure A.2.16.: *Frequency sweep of a BudgetSensors ContE-G tip. The amplitude is depicted (a), the corresponding phase in (b).*

Low-Repetition-Rate Tapping Mode

AFM tips applied in tapping mode usually have high stiffness and mechanical eigenfrequencies in the range of > 100 kHz. This allows for fast measurements and easy tuning of the p-i control loop. It also has the advantage that electronic noise at such high frequencies is lower when higher harmonic demodulation is applied in s-SNIM.

One scientific case of this PhD work was the implementation of the TELBE THz light source [140] to the T-Bone for near-field microscopy. The repetition rate of this accelerator-based light source is planned to reach 13 MHz repetition rate, but at the time of the thesis has been strictly limited to 100 kHz, which opens a demand to run the s-SNOM with low cantilever eigenfrequencies. Considering the Niquist-Shannon sampling theorem [141], the highest detectable frequency when sampling a signal with 100 kHz sampling rate is 50 kHz. When s-SNIM signals up to the third harmonic shall be detected, the maximum eigenfrequency must not exceed 50 kHz / 3 $\approx$ 16.7 kHz. This is a very low value for tapping mode AFM and s-SNIM which makes unconventional methods necessary. One way is simply to use tips with such low eigenfrequencies and stiffnesses, which are actually used for contact-mode and related techniques. As the THz near-field response of these conventional probes is unclear, two kinds of probes were tested during this work. Firstly, a standard silicon-based and PtIr-coated probe was used. Secondly, full platinum probes by Rocky Mountain Nanotechnology were used. Made of a platinum wire with a tip shank length of 80 μm these tips have an antenna resonance in the low THz regime (see section 3.4.3). This chapter presents results with these different tips for this unconventional application.

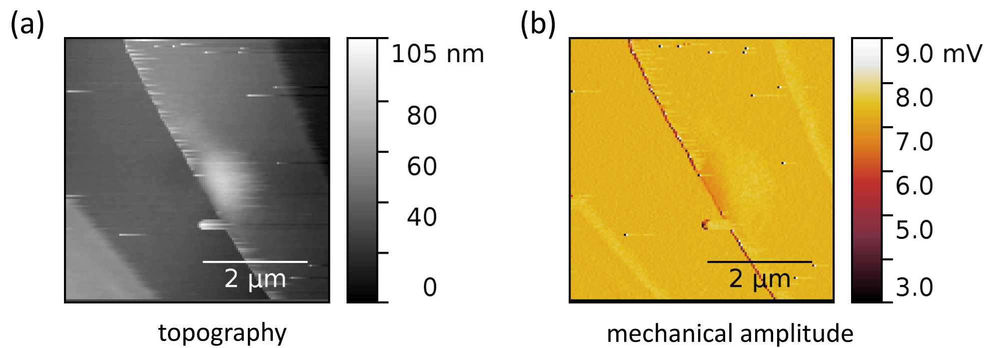

Figure A.2.17.: *128x128 pixel, 5 μm scan of the UMG01 test grating. The topography is depicted in (a), revealing artifacts espacially at the edges. The corressponding mechanical amplitude is shown in (b).*

Figure A.2.16 shows a low frequency sweep of a ***BudgetSensors ContE-G*** tip. The amplitude data (a) has a broad Lorentzian peak at $\sim$ 8 kHz connected with a 180 degree phasejump in the phase (b). The Q-factor of the resonance is only 46, one order of magnitude smaller than the value for a conventional non-contact tip (see figure A.2.6). Thus the control loop of the AFM controller must be set to much be set to much higher p & i values in order to reach stable feedback conditions, which is a result of the smaller change in amplitude of the cantilever oscillation when it reaches the sample surface and the resonance frequency slightly changes.

To find a suitable set of gain parameters, the test grating [24] was applied once again to confirm the AFM's capability to correctly present nanoscale topography values. Here, aim of the measurement was not to achieve an Ångström-level topography sensitivity, but to find the best values without increasing the line times to a value which is impractical for a daily useage of the AFM. The measurement is shown in A.2.17. With a scanning speed of 1 μm/s the RMS noise in the topography channel (a) is $\leq$ 1 nm along the stripes, but does not reach the Å-level. Scanning artifacts become obvious at the edges and especially at dirt particles. Additionally, the amplitude (b) breaks down to one almost one third of the setpoint when it reaches the edge in the topography. For comparison, the breakdown in classic tapping mode is rather in the range of 10%. Consequently measurements with this tip must be carefully evaluated, as features of the topography can easily couple into the demodulated near-field signal.

PICTURES!!!The second candidate for low-frequency tapping mode is the **Rocky Mountain Nanotechnology 25PT300B** tip [132]. Made of solid platinum wire, it can reach very small tip radii

[24] Anfatec Instruments AG, UMG01 [178]

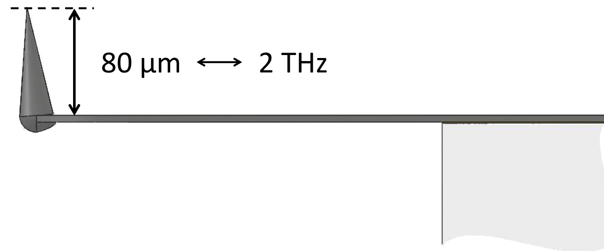

Figure A.2.18.: *Schematic side view of the Rocky Mountain Nanotechnology 25PT300B tip. Illustration from [132], label included by the author.*

below 10 nm. The cantilever is mounted on a plastic substrate and thus does not fit into the alignment chip used in the T-Bone and needs to be mounted by eye to the correct position. Furthermore the back of the cantilever is not flat and consequently the beam deflection is harder to align. For a trained scientist it is still possible to correctly mount the tip in few ten minutes. The performance of this tip is analysed at this point in more detail in this section, as it is applied in a THz Time-Domain-Spectroscopy SNOM setup (see chapter 4.3). The stiffness of 22 N/m is one order of magnitude larger than the *BudgetSensors* tip and half the size of conventional high-frequency tips, which allows for easier handling in tapping mode AFM. The tip shank is 80 μm long and thus much larger than a conventional silicon-based tip. The side view of the tip is shown in figure A.2.18. For low THz radiation, the antenna acts like a dipole antenna with a center frequency of $\sim$ 2 THz and strongly enhances the THz field at the tip apex (see section 3.4 for a detailed discussion).

A mechanical frequency sweep is depicted in A.2.19. The overview sweep shown in (a) shows a clear peak at 20 kHz. Far from the resonance, the cantilever barely oscillates and the signal is at the noise level. A zoomed scan is shown in (b). The sharp resonance is assiciated with a clear phasejump of π. Although the number of pixels in this sweep is low, the Q-factor of the resonance can be determined to $\sim$ 450 with the procedure applied in chapter A.2.

The quality of topography control is crucial for near-field measurements. Consequently retract-curves and a scan on a Si/SiO_2 nanostructure were performed with this kind of tip to evaluate the performance. The measurement of the retract-curve is depicted in A.2.20. The mechanical amplitude as a function of the tip-sample distance is shown in (a). Here, a very steep breakdown

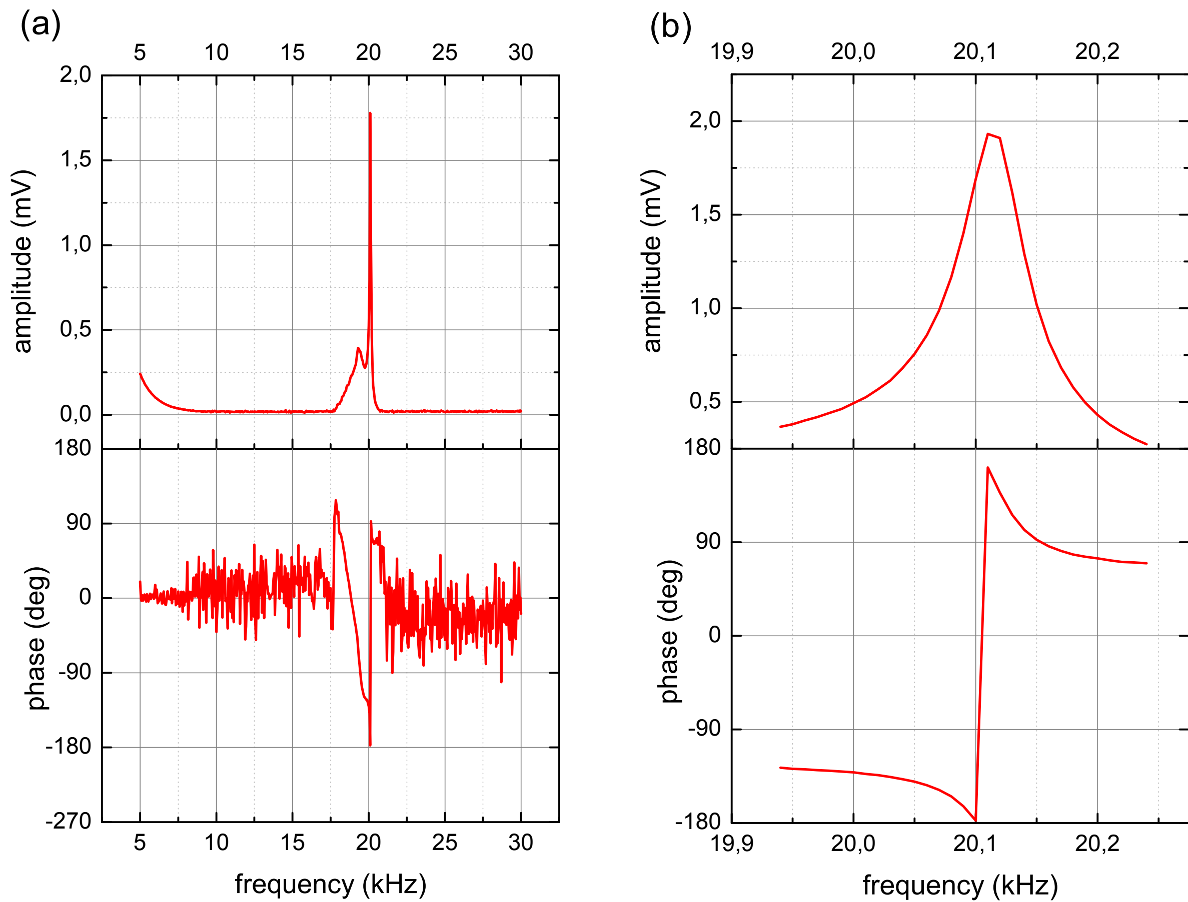

Figure A.2.19.: *Frequency sweep of a Rocky Mountain Nanotechnology 25PT300B tip. (a) shows an overview sweep, (b) a zoomed sweep of mechanical amplitude and phase, respectively.*

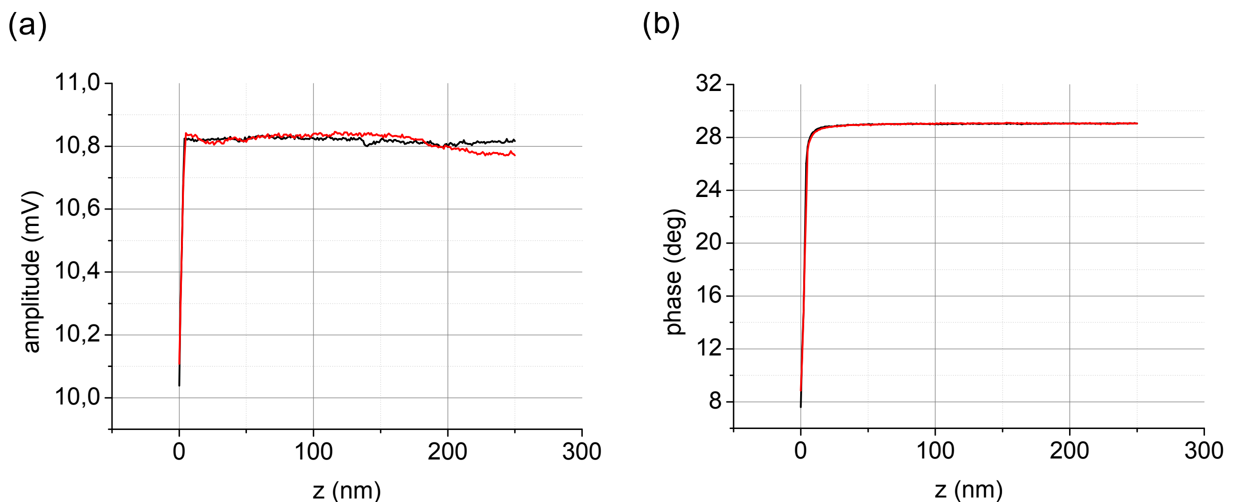

Figure A.2.20.: *Retract curve on a Si/SiO$_2$ nano-structure. (a) shows the amplitude and (b) the phase data as a function the tip-sample distance. The measurement has been performed twice (black and red curve).*

of the cantilever's oscillation is observed on the first ten nanometers. When the tip is further away from the sample surface, the amplitude is not as stable as observed in A.2.7, however it is still in the range of 1 %. The phase (b) is much more stable. Here, a difference in between the two measurements (black and red) is barely visible.

To confirm the capability to map the topography with a *25PT300B* tip on the nanoscale a Si/SiO2 nanostructure was measured with the T-Bone (see figure A.2.21). A complex checkered pattern is visible in the topography with heights up to 100 nm. The scan direction shown is from right to left. Here, the gain values have been chosen a bit too low, as the edges are sharp on the right-hand side of the topography features connected with a strong breakdown of the mechanical amplitude down to one third of the setpoint and not very sharp on the left-hand side of the features connected with an increased amplitude. However, larger gain values will significantly increase the topographic noise or may even lead to unstable conditions. Considering the low oscillation frequency and a scan speed of 2 μm/s, this is not surprising. The RMS roughness of the topography is in the range of 1.5 nm along the stripes. This value is not comparable to values achieved with the conventional method in this chapter, however it is good enough for many applications in near-field microscopy. Imperfect p-i control of the scan is not as problematic with this tip as compared to metal-coated silicon tips. When the coating is worn off, silicon tips are usually useless for near-field microscopy. Here, the tip is completely made of platinum and cannot lose any coating. When working with these tips, the near-field signal usually rises strongly

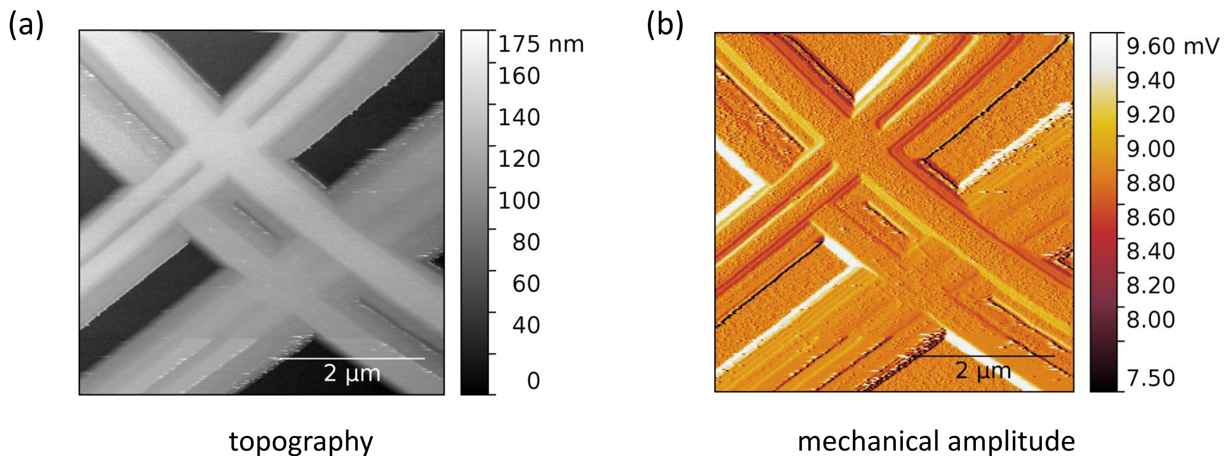

Figure A.2.21.: *256x256 pixel, 5 μm topography measurement of a Si/SiO_2 nano-structure. (a) shows the topography after a line-substract filter and (b) the mechanical amplitude for the same measurement. Artifacts can be observed in both channels at the edges.*

at the cost of resolution when the tip becomes more and more blunt.

For comparison with silicon-based s-SNOM tips, the tip was tested in the well-known and rather simple CO_2 laser setup to confirm its general applicability in near-field microscopy. A MCT detector[25] was utilized for excitation and detection of the near-field signal. A detailed description of the setup is presented in section 4.1. Here, the tip acts as a small scatterer close to the sample surface.

A complete characterisation of the *25PT300B* tip with a CO_2 s-SNOM measurement is depicted in figure A.2.22. The near-field amplitude of the second and the third harmonic on a gold sample as a function of the tip-sample distance is shown in (a) and (b), respectively. The reproducibility is good. When in contact, the absolute values of the amplitude are in the same range as with conventional tips (see figure A.2.13). The signal decay to half of the initial values is reached at distances of 11 and 7 nm which indicates small tip radius and consequently a very good resolution for lateral scans in near-field microscopy. The value of the third harmonic goes down to zero for large distances confirming a good far-field suppression. For this measurement, the signal-to-noise ratio is even better than with the conventional tip in section A.2. However, the noise level in these retract curves can be strongly influenced by laser noise and changes on a daily basis.

A well suited sample to verify the near-field microscope is capable of mapping a material con-

[25] Judson Technologies Inc., model J15D12-M204-S01M-60 with PA-101 preamp

trast and determining the achievable lateral resolution, is a gold nanostructure on a semiconductor/insulator substrate [7]. The topography of such a structure is shown in (c). The gold is covering the top left area of the scan and has a height of $\sim$ 30 nm. Although the lithography process is of doubtful quality, as can be seen by the form of the structure and residuals at the edges of the gold from the fabrication process, the sample shows three distinct signal levels in the near-field amplitude (d) and (e). The residuals are probably made of resin from the fabrication process and do not contribute to the near-field signal. The substrate is an undoped germanium single crystal. The signal level is higher than the level on the residuals. The gold clearly has the largest near-field signal as it has a negative permittivity for a wavelength of 10.6 μm. The third harmonic shows a better contrast compared to the second harmonic in between the different materials due to better far-field suppression. To determine the resolution, a zoomed scan (f) was performed in the area highlighted in (e) with the green rectangle. The three signal levels can be observed again and a profile was extracted at the position of the green line and is plotted in (g). With the method applied in section A.2, the resolution is determined to 15 nm - a comparably good value for s-SNOM. Overall this tip shows a very good performance for s-SNOM and due to its resonance in the low THz regime it is now also applied by other groups [48] and the company *Neaspec* for THz near-field microscopy.

At this point it must also be mentioned that there is also an exciting report of phase-domain sampling [187], where s-SNIM is applied with low sampling rate and standard high-frequency tips. The cantilever amplitude and phase is recorded and correlated to the sampled s-SNIM detector signal for each laser pulse. The data analysis performed after the experiment is technically challenging. The implementation of this technique for the combination of the *T-Bone* and *TELBE* this is part of Thales de Oliveira's work and will not be explained in detail in this thesis as it is a complex process, experimentally challenging and has not been applied for any measurements presented in this work. Just to name a few obstacles of this technique: The signal cannot be observed on the fly when the experiment is performed, also the amount of data sampled with such a techniques is very large and needs to be post-processed correctly.

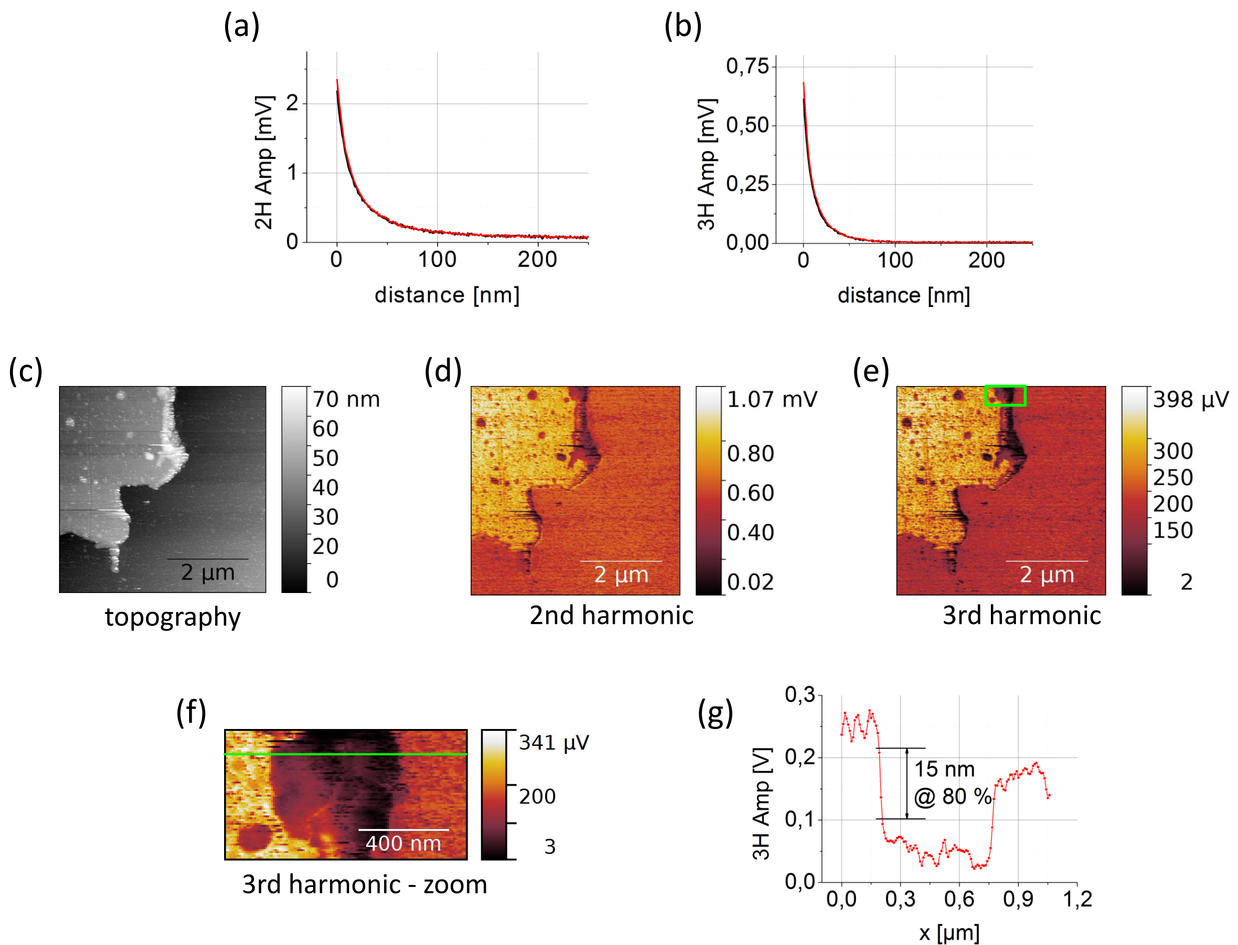

Figure A.2.22.: *Near-field examination with a Rocky Mountain Nanotechnology 25PT300B tip. Retract curves on gold are shown in (a) and (b). Topography and corresponding near-field signatures of a gold nanostructure on a germanium substrate are depicted in (c), (d) and (e). The size of the scan is 5 μm with 256x256 pixel resolution. A zoomed near-field scan (127x67 pixel) with a size of 1100 nm x 550 nm is shown in (f), where a profile (green) is extracted and plotted in (g).*

List of Figures

2.1.1 Formation of energy bands in semiconductors. 13
2.1.2 Band structures of GaAs and Ge. 14
2.1.3 Fermi-Dirac statistic in Germanium. 15
2.1.4 Plasma frequency in germanium as a function of the charge carrier density. . 18
2.1.5 Permittivity as a function of charge carrier density. 19
2.1.6 Absorption of light in semiconductors. 21
2.1.7 Lifetime of charge carriers in germanium. 23
2.1.8 Temporal behavior of the permittivity in excited germanium. 24
2.2.1 Schematic drawing of a field-effect transistor. 27
2.3.1 Concepts of SNOM. 34
2.4.1 Schematic drawing of a photo-concutive antenna. 41
2.4.2 Schematic drawing of a free-electron laser. 43
2.4.3 Response of InSb detectors. 45
2.4.4 Electro-optic sampling. 47

3.1.1 Dipole-Dipole model for s-SNOM. 50
3.1.2 Scattering cross-section in s-SNOM. 52
3.1.3 Modified dipole-dipole model in s-SNOM. 53
3.2.1 Illustration of higher-harmonic-demodulation. 57
3.2.2 Homodyne interferometric detection in s-SNOM. 58
3.3.1 Schematic illustration of amplitude-modulated s-SNOM. 64
3.4.1 Field enhancement simulation. 67
3.4.2 Field enhancement with resonant probes. 69
3.4.3 Resonance of different probe types. 71

4.1.1 Near-field setup at FELBE. 75
4.1.2 Retract curves with 2 THz radiation and PPP tips at FELBE. 77

4.1.3 Retract curves with 2 THz radiation and RMN tips at FELBE. 78
4.2.1 Optical setup for sideband investigations. 82
4.2.2 Sideband demodulation setup. 83
4.3.1 Time-domain spectroscopy near-field setup. 86
4.3.2 THz pulse of a photoconductive antenna. 88
4.3.3 THz near-field pulse on a gold sample. 89
4.3.4 THz near-field material contrast. 90

5.1.1 Near-field and sideband retract curves on germanium. 93
5.2.1 Near-field and sideband fluence dependence. 95
5.3.1 Amplitude and signal-to-noise ratio of higher sidebands. 96
5.4.1 Near-field and sideband signals as a function of the oscillation amplitude. . . 98
5.4.2 Near-field and sideband signal-to-noise ratio as a function of the oscillation amplitude. 99
5.6.1 Near-field and sideband nano-optical contrast. 104

6.1.1 Schematic illustration of a graphene-field effect transistor. 108
6.1.2 SEM and AFM images of graphene field-effect transistors. 109
6.3.1 Near-field scan of the GT1 device. 113
6.3.2 Near-field scan of the JT1 device. 114
6.3.3 Near-field scan of the TF1 device. 115
6.4.1 Averaging of multiple lines 116
6.4.2 Near-field linescan of the GT1 device. 118
6.4.3 Near-field linescan of a reference device device. 119
6.4.4 Fitted data and extracted plasma wave velocity in the GT1 device. 119
6.4.5 IV curve and calculated plasma wave velocity in the GT1 device. 120
6.4.6 Fitted data and extracted plasma wave velocity in the HT3 device. 122
6.4.7 IV curve calculated plasma wave velocity and rectification in the HT3 device. 123

A.1.1 Schematic drawing of an atomic force microscope. 133
A.1.2 Lennard-Jones potential. 134
A.1.3 Schematic drawing of a piezoresponse force microscopy setup. 137
A.1.4 Frequency spectra in tapping mode AFM. 139
A.1.5 Frequency spectra in non-contact mode AFM. 141
A.1.6 Signal spectrum in Kelvin Probe Force Micrcoscopy. 143

A.2.1 Screenshot of the T-Bone AFM model. 146
A.2.2 Beam deflection setup of the T-Bone AFM 148
A.2.3 Contact mode topography image of the UMG01 test grating. 150
A.2.4 Out-of-plane PFM frequency sweep. 151
A.2.5 PFM scan of a PPLN sample. 153
A.2.6 Tapping mode frequency sweep. 155
A.2.7 Mechanical retract curve in tapping mode AFM. 156
A.2.8 Tapping mode topography image of the UMG01 test grating. 157
A.2.9 Contact mode topography image of a HOPG sample. 158
A.2.10 Non-contact mode topography image of the UMG01 test grating. 159
A.2.11 Non-contact mode topography image of a HOPG sample. 160
A.2.12 KPFM scan of a SiGe thin film sample. 161
A.2.13 Retract curves on a gold sample with a CO_2 laser. 163
A.2.14 Near-field scan of a contaminated gold sample. 164
A.2.15 Extracted near-field line profile. 166
A.2.16 Frequency sweep of low frequency ContE-G tip. 167
A.2.17 Tapping mode topography scan with ContE-G tip. 168
A.2.18 Schematic side view of the Rocky Mountain Nanotechnology 25PT300B tip. . 169
A.2.19 Tapping mode frequency sweep of an RMN 25PT300B tip. 170
A.2.20 Mechanical retract curve of an RMN tip. 171
A.2.21 Si/SiO_2 tapping mode topography scan with an RMN tip. 172
A.2.22 Near-field scan with an RMN tip and a CO_2 probe laser. 174

Bibliography

1. Lewis, R. A. Physical phenomena in electronic materials in the terahertz region. *Proc. IEEE* **95,** 1641–1645 (2007).
2. Abbe, E. Beiträge zur Theorie des Mikroskops und der mikroskopischen Wahrnehmung. *Archiv für mikroskopische Anatomie* **9,** 413–418 (1873).
3. Synge, E. XXXVIII. A suggested method for extending microscopic resolution into the ultra-microscopic region. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **6,** 356–362 (1928).
4. Inouye, Y. & Kawata, S. Near-field scanning optical microscope with a metallic probe tip. *Opt. Lett.* **19,** 159–161 (1994).
5. Ash, E. & Nicholls, G. Super-resolution Aperture Scanning Microscope. *Nature* **237,** 510–512 (1972).
6. Lahrech, A., Bachelot, R., Gleyzes, P. & Boccara, A. C. Infrared-reflection-mode near-field microscopy using an apertureless probe with a resolution of λ/600. *Opt. Lett.* **21,** 1315–1317 (1996).
7. Kuschewski, F. *et al.* Narrow-band near-field nanoscopy in the spectral range from 1.3 to 8.5 THz. *Appl. Phys. Lett.* **108,** 113102 (2016).
8. Wikipedia. `https://en.wikipedia.org/wiki/Band_gap#/media/File:Solid_state_electronic_band_structure.svg`. Last visited November 18th, 2018.
9. Voon, L. & Willatzen, M. *The kp method: electronic properties of semiconductors.* doi:`10.1007/978-3-540-92872-0` (Springer Science & Business Media, 2009).
10. Vurgaftman, I., Meyer, J. R. & Ram-Mohan, L. R. Band parameters for III–V compound semiconductors and their alloys. *J. Appl. Phys.* **89,** 5815–5875 (2001).
11. El Kurdi, M., Fishman, G., Sauvage, S. & Boucaud, P. Band structure and optical gain of tensile-strained germanium based on a 30 band k·p formalism. *J. Appl. Phys.* **107,** 013710 (2010).

12. Sze, S. M. & Ng, K. K. *Physics of semiconductor devices.* doi:10.1002/0470068329 (John Wiley & Sons, 2006).
13. Drude, P. Zur Elektronentheorie der Metalle. *Annalen der Physik* **306,** 566–613 (1900).
14. Hein, P., Gallant, M. & van Driel, H. Influence of excitation wavelength on the reflectivity of photogenerated plasmas in germanium. *Solid State Commun.* **39,** 601–604 (1981).
15. Huber, A. J., Keilmann, F., Wittborn, J., Aizpurua, J. & Hillenbrand, R. Terahertz Near-Field Nanoscopy of Mobile Carriers in Single Semiconductor Nanodevices. *Nano Lett.* **8,** 3766–3770 (2008).
16. PVEducation. https://www.pveducation.org/pvcdrom/absorption-coefficient. Last visited November 21th, 2018.
17. Auston, D. H. & Shank, C. V. Picosecond Ellipsometry of Transient Electron-Hole Plasmas in Germanium. *Phys. Rev. Lett.* **32,** 1120–1123 (1974).
18. Aspnes, D. E. & Studna, A. A. Dielectric functions and optical parameters of Si, Ge, GaP, GaAs, GaSb, InP, InAs, and InSb from 1.5 to 6.0 eV. *Phys. Rev. B* **27,** 985–1009 (1983).
19. Alcock, A. J., Corkum, P. B. & James, D. J. A fast scalable switching technique for high-power CO2 laser radiation. *Appl. Phys. Lett.* **27,** 680–682 (1975).
20. Gaubas, E. & Vanhellemont, J. Dependence of carrier lifetime in germanium on resisitivity and carrier injection level. *Appl. Phys. Lett.* **89,** 142106 (2006).
21. Dyakonov, M. & Shur, M. Shallow water analogy for a ballistic field effect transistor: New mechanism of plasma wave generation by dc current. *Phys. Rev. Lett.* **71,** 2465–2468 (1993).
22. Dyakonov, M. & Shur, M. Two dimensional electronic flute. *Appl. Phys. Lett.* **67,** 1137–1139 (1995).
23. Dyakonov, M. & Shur, M. Choking of electron flow: A mechanism of current saturation in field-effect transistors. *Phys. Rev. B* **51,** 14341–14345 (1995).
24. Dyakonov, M. & Shur, M. Detection, mixing, and frequency multiplication of terahertz radiation by two-dimensional electronic fluid. *IEEE Trans. Electron Devices* **43,** 380–387 (1996).
25. Dyer, G. C. *et al.* Enhanced performance of resonant sub-terahertz detection in a plasmonic cavity. *Appl. Phys. Lett.* **100,** 083506 (2012).
26. Kang, S., Burke, P. J., Pfeiffer, L. N. & West, K. W. Resonant frequency response of plasma wave detectors. *Appl. Phys. Lett.* **89,** 213512 (2006).

27. Boppel, S. *et al.* 0.25-μmGaN TeraFETs Optimized as THz Power Detectors and Intensity-Gradient Sensors. *IEEE Transactions on Terahertz Science and Technology* **6,** 348–350 (2016).
28. Knap, W. *et al.* Terahertz emission by plasma waves in 60 nm gate high electron mobility transistors. *Appl. Phys. Lett.* **84,** 2331–2333 (2004).
29. Knap, W., Deng, Y., Rumyantsev, S. & Shur, M. S. Resonant detection of subterahertz and terahertz radiation by plasma waves in submicron field-effect transistors. *Appl. Phys. Lett.* **81,** 4637–4639 (2002).
30. Lisauskas, A. *et al.* Rational design of high-responsivity detectors of terahertz radiation based on distributed self-mixing in silicon field-effect transistors. *J. Appl. Phys.* **105,** 114511 (2009).
31. Bauer, M. *et al.* Antenna-coupled field-effect transistors for multi-spectral terahertz imaging up to 4.25 THz. *Opt. Express* **22,** 19235–19241 (2014).
32. Grzyb, J. & Pfeiffer, U. THz Direct Detector and Heterodyne Receiver Arrays in Silicon Nanoscale Technologies. *J. Infrared Millim. Terahertz Waves* **36,** 998–1032 (2015).
33. Novoselov, K. S. *et al.* Electric Field Effect in Atomically Thin Carbon Films. *Science* **306,** 666–669 (2004).
34. Zak, A. *et al.* Antenna-Integrated 0.6 THz FET Direct Detectors Based on CVD Graphene. *Nano Lett.* **14,** 5834–5838 (2014).
35. Meric, I. *et al.* Current saturation in zero-bandgap, top-gated graphene field-effect transistors. *Nat. Nanotechnol.* **3,** 654 (2008).
36. Lemme, M. C., Echtermeyer, T. J., Baus, M. & Kurz, H. A Graphene Field-Effect Device. *IEEE Electron Device Letters* **28,** 282–284 (2007).
37. Landau, L. & Lifshitz, E. *Fluid Mechanics* (Pergamon Press, 1959).
38. Weiyan, T. *Shallow water hydrodynamics. Mathematical theory and numerical solution for a two-dimensional system of shallow water equations* (Elsevier, 1991).
39. Drexler, C. *et al.* Helicity sensitive terahertz radiation detection by field effect transistors. *Journal of Applied Physics* **111,** 124504 (2012).
40. Shchepetov, A. *et al.* Oblique modes effect on terahertz plasma wave resonant detection in InGaAs/InAlAs multichannel transistors. *Applied Physics Letters* **92,** 242105 (2008).
41. Cai, X. *et al.* Sensitive Room-Temperature Terahertz Detection via Photothermoelectric Effect in Graphene. *Nature Nanotechnology* **9.** doi:10.1038/nnano.2014.182 (May 2013).

42. Pendry, J. Negative Refraction Makes a Perfect Lens. *Phys. Rev. Lett* **85,** 3966–3969 (2000).

43. Pohl, D. W., Denk, W. & Lanz, M. Optical stethoscopy: Image recording with resolution λ/20. *Appl. Phys. Lett.* **44,** 651–653 (1984).

44. Chambers, R. The anomalous skin effect. *Proc. R. Soc. Lond. A* **215,** 481–497 (1952).

45. Zenhausern, F., O'Boyle, M. P. & Wickramasinghe, H. K. Apertureless near-field optical microscope. *Appl. Phys. Lett.* **65,** 1623–1625 (1994).

46. Paiva, J. S., Jorge, P. A., Rosa, C. C. & Cunha, J. P. Optical fiber tips for biological applications: From light confinement, biosensing to bioparticles manipulation. *Biochim. Biophys. Acta* **1862,** 1209 –1246 (2018).

47. Kuschewski, F. *et al.* Optical nanoscopy of transient states in condensed matter. *Sci. Rep.* **5** (2015).

48. Liewald, C. *et al.* All-electronic terahertz nanoscopy. *Optica* **5,** 159–163 (2018).

49. Taubner, T., Keilmann, F. & Hillenbrand, R. Nanomechanical Resonance Tuning and Phase Effects in Optical Near-Field Interaction. *Nano Lett.* **4,** 1669–1672 (2004).

50. Hillenbrand, R. & Keilmann, F. Complex Optical Constants on a Subwavelength Scale. *Phys. Rev. Lett.* **85,** 3029–3032 (2000).

51. Ocelic, N., Huber, A. & Hillenbrand, R. Pseudoheterodyne detection for background-free near-field spectroscopy. *Appl. Phys. Lett.* **89,** 101124 (2006).

52. Tranca, D. *et al.* High-resolution quantitative determination of dielectric function by using scattering scanning near-field optical microscopy. *Sci. Rep.* **5,** 11876 (2015).

53. Pollard, B., Maia, F. C. B., Raschke, M. B. & Freitas, R. O. Infrared Vibrational Nanospectroscopy by Self-Referenced Interferometry. *Nano Lett.* **16,** 55–61 (2016).

54. Von Ribbeck, H.-G. *et al.* Spectroscopic THz near-field microscope. *Opt. Express* **16,** 3430–3438 (2008).

55. Moon, K. *et al.* Quantitative coherent scattering spectra in apertureless terahertz pulse near-field microscopes. *Appl. Phys. Lett.* **101,** 011109 (2012).

56. Moon, K. *et al.* Subsurface Nanoimaging by Broadband Terahertz Pulse Near-Field Microscopy. *Nano Lett.* **15,** 549–552 (2015).

57. Hamann, H. F., Gallagher, A. & Nesbitt, D. J. Enhanced sensitivity near-field scanning optical microscopy at high spatial resolution. *Appl. Phys. Lett.* **73,** 1469–1471 (1998).

58. L. Heilman, A & J. Gordon, M. Tip-enhanced near-field optical microscope with side-on and ATR-mode sample excitation for super-resolution Raman imaging of surfaces. *J. Appl. Phys.* **119,** 223103 (2016).

59. Wang, H., Wang, L. & Xu, X. Scattering-type Scanning Near-Field Optical Microscopy with Low-Repetition-Rate Pulsed Light Source through Phase-Domain Sampling. *Nat. Commun.* **7,** 13212 (2016).

60. Berweger, S., Atkin, J. M., Olmon, R. L. & Raschke, M. B. Light on the Tip of a Needle: Plasmonic Nanofocusing for Spectroscopy on the Nanoscale. *J. Phys. Chem. Lett.* **3,** 945–952 (2012).

61. Schnell, M., Carney, P. & Hillenbrand, R. Synthetic Optical Holography for Rapid Nanoimaging. *Nat. Commun.* **5,** 3499 (2014).

62. Huth, F. *et al.* Nano-FTIR Absorption Spectroscopy of Molecular Fingerprints at 20 nm Spatial Resolution. *Nano Lett.* **12,** 3973–3978 (2012).

63. Döring, J., von Ribbeck, H.-G., Fehrenbacher, M., Kehr, S. C. & Eng, L. M. Near-field resonance shifts of ferroelectric barium titanate domains upon low-temperature phase transition. *Appl. Phys. Lett.* **105,** 053109 (2014).

64. Lang, D. *et al.* Infrared nanoscopy down to liquid helium temperatures. *Rev. Sci. Instrum.* **89,** 033702 (2018).

65. McLeod, A. *et al.* Nanotextured phase coexistence in the correlated insulator V2O3. *Nature Physics* **13** (2016).

66. Yang, H., Hebestreit, E., E Josberger, E. & Raschke, M. A cryogenic scattering-type scanning near-field optical microscope. *Rev. Sci. Instrum.* **84,** 023701 (2013).

67. Ni, G. *et al.* Fundamental limits to graphene plasmonics. *Nature* **557,** 530 (2018).

68. Eisele, M. *et al.* Ultrafast multi-terahertz nano-spectroscopy with sub-cycle temporal resolution. *Nat. Photonics* **8,** 841–845 (2014).

69. Wagner, M. *et al.* Ultrafast and Nanoscale Plasmonic Phenomena in Exfoliated Graphene Revealed by Infrared Pump–Probe Nanoscopy. *Nano Lett.* **14,** 894–900 (2014).

70. Huber, M. A. *et al.* Femtosecond photo-switching of interface polaritons in black phosphorus heterostructures. *Nat. Nanotechnol.* **12,** 207 (2017).

71. Fei, Z. *et al.* Gate-tuning of graphene plasmons revealed by infrared nano-imaging. *Nature* **487,** 82–85 (2012).

72. Chen, J. *et al.* Optical nano-imaging of gate-tunable graphene plasmons. *Nature* **487,** 77–81 (2012).

73. Li, P. *et al.* Hyperbolic phonon-polaritons in boron nitride for near-field optical imaging and focusing. *Nat. Commun.* **6,** 7507 (2015).

74. Kehr, S. C. *et al.* Anisotropy Contrast in Phonon-Enhanced Apertureless Near-Field Microscopy Using a Free-Electron Laser. *Phys. Rev. Lett.* **100,** 256403 (2008).

75. Dai, S. *et al.* Tunable Phonon Polaritons in Atomically Thin van der Waals Crystals of Boron Nitride. *Science* **343,** 1125–1129 (2014).

76. Li, P. *et al.* Optical Nanoimaging of Hyperbolic Surface Polaritons at the Edges of van der Waals Materials. *Nano Lett.* **17,** 228–235 (2017).

77. Hermann, P. *et al.* Characterization of semiconductor materials using synchrotron radiation-based near-field infrared microscopy and nano-FTIR spectroscopy. *Opt. Express* **22,** 17948–17958 (2014).

78. Firkala, T. *et al.* Near-Field Optical Examination of Potassium n-Butyl Xanthate/Chalcopyrite Flotation Products. *Minerals* **8,** 118 (2018).

79. Pollard, B., Muller, E. A., Hinrichs, K. & Raschke, M. B. Vibrational nano-spectroscopic imaging correlating structure with intermolecular coupling and dynamics. *Nat. Commun.* **5,** 3587 (2014).

80. Kusch, P. *et al.* Combined Tip-Enhanced Raman Spectroscopy and Scattering-Type Scanning Near-Field Optical Microscopy. *The Journal of Physical Chemistry C* **122,** 16274–16280 (2018).

81. Huber, A., Ziegler, A., Köck, T & Hillenbrand, R. Infrared nanoscopy of strained semiconductors. *Nat. Nanotechnol.* **4,** 153–157 (2009).

82. Gigler, A. M. *et al.* Nanoscale residual stress-field mapping around nanoindents in SiC by IR s-SNOM and confocal Raman microscopy. *Opt. Express* **17,** 22351–22357 (2009).

83. Muller, E. A. *et al.* Nanoimaging and Control of Molecular Vibrations through Electromagnetically Induced Scattering Reaching the Strong Coupling Regime. *ACS Photonics* **5,** 3594–3600 (2018).

84. Hermann, P. *et al.* Near-field imaging and nano-Fourier-transform infrared spectroscopy using broadband synchrotron radiation. *Opt. Express* **21,** 2913–2919 (2013).

85. Bechtel, H. A., Muller, E. A., Olmon, R. L., Martin, M. C. & Raschke, M. B. Ultrabroadband infrared nanospectroscopic imaging. *Proceedings of the National Academy of Sciences* **111,** 7191–7196 (2014).
86. Fehrenbacher, M. *et al.* Plasmonic Superlensing in Doped GaAs. *Nano Lett.* **15,** 1057–1061 (2015).
87. Kehr, S. *et al.* Near-field examination of perovskite-based superlenses and superlense-enhanced probe-object coupling. *Nat. Commun.* **2,** 249 (2011).
88. Mastel, S. *et al.* Terahertz Nanofocusing with Cantilevered Terahertz-Resonant Antenna Tips. *Nano Lett.* **17,** 6526–6533 (2017).
89. Zhang, J. *et al.* Terahertz Nanoimaging of Graphene. *ACS Photonics* **5,** 2645–2651 (2018).
90. Xu, W. Generation of terahertz acoustic-phonon signals by heated electrons in $Al_xGa_{1-x}As$ / GaAs-based quantum wires. *Appl. Phys. Lett.* **68,** 1353–1355 (1996).
91. Crooker, S. A., Baumberg, J. J., Flack, F., Samarth, N. & Awschalom, D. D. Terahertz Spin Precession and Coherent Transfer of Angular Momenta in Magnetic Quantum Wells. *Phys. Rev. Lett.* **77,** 2814–2817 (1996).
92. Spencer, B. F. *et al.* Terahertz cyclotron resonance spectroscopy of an AlGaN/GaN heterostructure using a high-field pulsed magnet and an asynchronous optical sampling technique. *Appl. Phys. Lett.* **108,** 212101 (2016).
93. Jaros, M. Ground states of shallow donors in silicon and germanium. *J. Phys. C.: Solid State Phys.* **4,** 1162–1167 (1971).
94. Kampfrath, T., Tanaka, K. & Nelson, K. A. Resonant and nonresonant control over matter and light by intense terahertz transients. *Nat. Photonics* **7,** 680 (2013).
95. Jeon, T.-I. & Grischkowsky, D. THz Zenneck surface wave (THz surface plasmon) propagation on a metal sheet. *Appl. Phys. Lett.* **88,** 061113 (2006).
96. Wang, Z. *et al.* Magnetic field dependence of antiferromagnetic resonance in NiO. *Appl. Phys. Lett.* **112,** 252404 (2018).
97. Hafez, H. A. *et al.* Extremely efficient terahertz high-harmonic generation in graphene by hot Dirac fermions. *Nature* **561,** 507 (2018).
98. Ferguson, B. & Zhang, X.-C. Materials for terahertz science and technology. *Nature materials* **1,** 26 (2002).
99. Tonouchi, M. Cutting-edge terahertz technology. *Nat. Photonics* **1,** 97 (2007).

100. Lee, Y.-S. *Principles of terahertz science and technology* (Springer Science & Business Media, 2009).

101. Mueller, E. R., Fontanella, J. & Henschke, R. *Stabilized, integrated, far-infrared laser system for NASA/Goddard Space Flight Center* in *Proc. 10th Int. Symp. Space Terahertz Technol* (2000).

102. Rösch, M. *et al.* Heterogeneous terahertz quantum cascade lasers exceeding 1.9 THz spectral bandwidth and featuring dual comb operation. *Nanophotonics* **7,** 237–242 (2018).

103. Degl'Innocenti, R. *et al.* Terahertz Nanoscopy of Plasmonic Resonances with a Quantum Cascade Laser. *ACS Photonics* **4,** 2150–2157 (2017).

104. Muller, E. A., Pollard, B., Bechtel, H. A., van Blerkom, P. & Raschke, M. B. Infrared vibrational nanocrystallography and nanoimaging. *Science Advances* **2.** doi:10.1126/sciadv.1601006 (2016).

105. Lytid SAS. http://lytid.com/wp-content/uploads/2018/08/TeraSchottky-final-lab-serie-pg.pdf. Last visited November 28th, 2018.

106. Ekström, C. *Generation of Short Intense Terahertz Pulses Through Optical Rectification* (Diploma Thesis, Lund University, 2011).

107. Laserquantum Ltd. https://www.laserquantum.com/download-ds.cfm?id=707. Last visited November 29th, 2018.

108. Socol, Y. High-power free-electron lasers–technology and future applications. *Opt. Laser Technol.* **46,** 111 –126 (2013).

109. Brau, C. Free-electron lasers. *Adv. Elec. El. Phys.* **22** (1990).

110. Hopf, F. A., Meystre, P., Scully, M. O. & Louisell, W. H. Classical Theory of a Free-Electron Laser. *Phys. Rev. Lett.* **37,** 1215–1218 (18 1976).

111. Cohn, K., Colson, W., Price, M., Blau, J. & Ng, J. Free electron lasers in 2015. *Proceedings of FEL2015* (2015).

112. Klopf, J. M. Personal Correspondence. 2018.

113. Richards, P. L. Bolometers for infrared and millimeter waves. *J. Appl. Phys.* **76,** 1–24 (1994).

114. Thomas Keating Ltd. http://www.terahertz.co.uk/images/qmci/woodcraft/pdf/QFI3BI%202+GeGa%20Operating%20Manual.pdf. Last visited December 1st, 2018.

115. Vieweg, N. *et al.* Terahertz-time domain spectrometer with 90 dB peak dynamic range. *Journal of Infrared, Millimeter, and Terahertz Waves* **35,** 823–832 (2014).

116. Giordano, M. C. *et al. Phase-resolved terahertz near-field nanoscopy of a topological insulator phonon-polariton mode* in *43rd International Conference on Infrared, Millimeter, and Terahertz Waves (IRMMW-THz)* (2018).

117. Neaspec GmbH. `https://www.neaspec.com/application/terahertz-near-field-microscopy-below-30nm-spatial-resolution/`. Last visited January 2nd, 2019.

118. Knoll, B. & Keilmann, F. Enhanced dielectric contrast in scattering-type scanning near-field optical microscopy. *Opt. Commun.* **182,** 321 –328 (2000).

119. Jackson, J. *Klassische Elektrodynamik* (Walter de Gruyter, 1983).

120. Rakic, A. D., Djurivsic, A. B., Elazar, J. M. & Majewski, M. L. Optical properties of metallic films for vertical-cavity optoelectronic devices. *Appl. Opt.* **37,** 5271–5283 (1998).

121. Cvitkovic, A., Ocelic, N. & Hillenbrand, R. Analytical model for quantitative prediction of material contrasts in scattering-type near-field optical microscopy. *Opt. Express* **15,** 8550–8565 (2007).

122. Schneider, S. C. *Scattering Scanning Near-Field Optical Microscopy on Anisotropic Dielectrics* (Doktorarbeit, Technische Universität Dresden, 2007).

123. Wurtz, G., Bachelot, R. & Royer, P. Imaging a GaAlAs laser diode in operation using apertureless scanning near-field optical microscopy. *Eur. Phys. J. AP* **5,** 269–275 (Mar. 1999).

124. Zurich Instruments AG. `https://www.zhinst.com/sites/default/files/li_primer/zi_whitepaper_principles_of_lock-in_detection.pdf`. Last visited October 9th, 2018.

125. The LIGO Scientific Collaboration. Optical properties of metallic films for vertical-cavity optoelectronic devices. *Class. Quantum Grav.* **32,** 074001 (2015).

126. Cebula, M. *Phasengeregelte homodyne Interferometrie für s-SNOM* (Diplomarbeit, Technische Universität Dresden, 2016).

127. Kuschewski, F. *Zeitaufgelöste Nahfeldmikroskopie* (Diplomarbeit, Technische Universität Dresden, 2014).

128. RefractiveIndex.INFO. `refractiveindex.info`. Last visited March 11th, 2019.

129. Thomas, S., Wachter, G., Lemell, C., Burgdörfer, J. & Hommelhoff, P. Large optical field enhancement for nanotips with large opening angles. *New J. Phys.* **17,** 063010 (2015).

130. Wang, Y, Cai, W, Yang, M, Liu, Z & Shang, G. A simple model of the scanning near-field optical microscopy probe tip for electric field enhancement. *Optica Applicata* **47,** 119–130 (Jan. 2017).

131. Haidary, A., Miyahara, Y. & Grütter, P. *Antenna and Plasmonic Properties of Scanning Probe Tips at Optical and Terahertz Regimes* in *Proceedings of the 2014 COMSOL Conference* (2014).

132. Rocky Mountain Nanotechnology LLC. `http://rmnano.com/tech.html`. Last visited September 24th, 2018.

133. Ortmann, L. *Theoretische Betrachtungen zur Nahfeldmikroskopie mit anisotropen Superlinsen* (Masterarbeit, Technische Universität Dresden, 2015).

134. Comsol Multiphysics GmbH. `https://cdn.comsol.com/documentation/5.4.0.225/IntroductionToCOMSOLMultiphysics.de_DE.pdf`. Last visited November 11th, 2018.

135. Comsol Multiphysics GmbH. `http://people.ee.ethz.ch/~fieldcom/pps-comsol/documents/User%20Guide/WaveOpticsModuleUsersGuide.pdf`. Last visited November 26th, 2018.

136. Berweger, S. *et al.* Microwave Near-Field Imaging of Two-Dimensional Semiconductors. *Nano Lett.* **15,** 1122–1127 (2015).

137. Tselev, A. *et al.* Near-field microwave scanning probe imaging of conductivity inhomogeneities in CVD graphene. *Nanotechnology* **23,** 385706 (2012).

138. Brinciotti, E. *et al.* Probing resistivity and doping concentration of semiconductors at the nanoscale using scanning microwave microscopy. *Nanoscale* **7,** 14715–14722 (35 2015).

139. Esslinger, M., Dorfmüller, J., Khunsin, W., Vogelgesang, R. & Kern, K. Background-free imaging of plasmonic structures with cross-polarized apertureless scanning near-field optical microscopy. *Rev. Sci. Instr.* **83,** 033704 (2012).

140. Green, B. *et al.* High-Field High-Repetition-Rate Sources for the Coherent THz Control of Matter. *Sci. Rep.* **6,** 22256 (2016).

141. Jerri, A. J. The Shannon sampling theorem—Its various extensions and applications: A tutorial review. *Proc. IEEE* **65,** 1565–1596 (1977).

142. Corkum, P. B. & Keith, D. Controlled switching of 10-micrometer radiation using semiconductor étalons. *J. Opt. Soc. Am. B.* **2,** 1873–1879 (1985).

143. Jamison, S. & Nurmikko, A. Generation of picosecond pulses of variable duration at 10.6 μm. *Appl. Phys. Lett.* **33,** 598–600 (1978).

144. Von Ribbeck, H.-G. *THz Near-Field Microscopy and Spectroscopy* (Doktorarbeit, Technische Universität Dresden, 2014).

145. Hillenbrand, R., Stark, M. & Guckenberger, R. Higher-harmonics generation in tapping-mode atomic-force microscopy: Insights into the tip–sample interaction. *Appl. Phys. Lett.* **76,** 3478–3480 (2000).

146. Gucciardi, P. G. *et al.* Artifacts identification in apertureless near-field optical microscopy. *J. Appl. Phys.* **101,** 064303 (2007).

147. Gwyddion. `http://gwyddion.net/`. Last visited September 12th, 2018.

148. Rajab, K. Z. *et al.* Broadband dielectric characterization of aluminum oxide (Al_2O_3). *J. Micro. and Elect. Pack.* **5,** 2–7 (2008).

149. Taubner, T., Keilmann, F. & Hillenbrand, R. Nanoscale-resolved subsurface imaging by scattering-type near-field optical microscopy. *Opt. Express* **13,** 8893–8899 (2005).

150. Tomadin, A. Personal Correspondence. 2018.

151. Monteverde, M. *et al.* Transport and Elastic Scattering Times as Probes of the Nature of Impurity Scattering in Single-Layer and Bilayer Graphene. *Phys. Rev. Lett.* **104,** 126801 (2010).

152. Hong, X., Zou, K. & Zhu, J. Quantum scattering time and its implications on scattering sources in graphene. *Phys. Rev. B* **80,** 241415 (2009).

153. George, P. A. *et al.* Ultrafast Optical-Pump Terahertz-Probe Spectroscopy of the Carrier Relaxation and Recombination Dynamics in Epitaxial Graphene. *Nano Lett.* **8,** 4248–4251 (2008).

154. Lu, N., Wang, L., Li, L. & Liu, M. A review for compact model of graphene field-effect transistors. *Chin. Phys. B* **26,** 036804 (2017).

155. Kim, S. *et al.* Realization of a high mobility dual-gated graphene field-effect transistor with Al2O3 dielectric. *Appl. Phys. Lett.* **94,** 062107 (2009).

156. Xia, F., Farmer, D. B., Lin, Y.-m. & Avouris, P. Graphene Field-Effect Transistors with High On/Off Current Ratio and Large Transport Band Gap at Room Temperature. *Nano Lett.* **10,** 715–718 (2010).

157. Tomadin, A. & Polini, M. Theory of the plasma-wave photoresponse of a gated graphene sheet. *Phys. Rev. B* **88,** 205426 (2013).

158. Lu, Y. *et al.* High-On/Off-Ratio Graphene Nanoconstriction Field-Effect Transistor. *Small* **6,** 2748–2754 (2010).

159. Ryzhii, V. Terahertz Plasma Waves in Gated Graphene Heterostructures. *Jpn. J. Appl. Phys.* **45,** L923–L925 (2006).

160. Ryzhii, V., Satou, A. & Otsuji, T. Plasma waves in two-dimensional electron-hole system in gated graphene heterostructures. *J. Appl. Phys.* **101,** 024509 (2007).

161. Geim, A. K. & Grigorieva, I. V. Van der Waals heterostructures. *Nature* **499,** 419 (2013).

162. Wagner, T., Köhler, D., Milde, P. & Eng, L. M. Probing the local surface potential and quantum capacitance in single and multi-layer graphene. *Appl. Phys. Lett.* **103,** 023102 (2013).

163. Dean, C. R. *et al.* Boron nitride substrates for high-quality graphene electronics. *Nature Nanotechnol.* **5,** 722 (2010).

164. Binnig, G., Rohrer, H., Gerber, C. & Weibel, E. Surface Studies by Scanning Tunneling Microscopy. *Phys. Rev. Lett.* **49,** 57–61 (1982).

165. Binnig, G., Quate, C. F. & Gerber, C. Atomic Force Microscope. *Phys. Rev. Lett.* **56,** 930–933 (1986).

166. Martin, Y., Williams, C. C. & Wickramasinghe, H. K. Atomic force microscope–force mapping and profiling on a sub 100-Å scale. *J. Appl. Phys.* **61,** 4723–4729 (1987).

167. Milde, P. *et al.* Unwinding of a Skyrmion Lattice by Magnetic Monopoles. *Science* **340,** 1076–1080 (2013).

168. NanoWorld AG. `https://www.nanosensors.com/`. Last visited November 30th, 2018.

169. Bruker Corp. `https://www.brukerafmprobes.com/c-111-silicon-nitride-probes.aspx`. Last visited November 30th, 2018.

170. De Jong, M., Chen, W., Geerlings, H., Asta, M. & Persson, K. A database to enable discovery and design of piezoelectric materials. *Sci. Data* **2,** 150053 (2015).

171. Oxford Instruments PLC. Cypher AFM User Guide.

172. Liu, J., Notbohm, J. K., Carpick, R. W. & Turner, K. T. Method for Characterizing Nanoscale Wear of Atomic Force Microscope Tips. *ACS Nano* **4,** 3763–3772 (2010).

173. Chung, K.-H. Wear characteristics of atomic force microscopy tips: A review. *Int. J. Pr. Eng. Man.* **15,** 2219–2230 (2014).

174. Giessibl, F. J. Atomic Resolution of the Silicon (111)-(7x7) Surface by Atomic Force Microscopy. *Science* **267,** 68–71 (1995).

175. Garcia, R. & Herruzo, E. The emergence of multifrequency force microscopy. *Nat. Nanotechnol.* **7,** 217–226 (2012).

176. Attocube Systems GmbH. `http://www.attocube.com/attomotion/premium-line/`. Last visited September 12th, 2018.

177. RHK Technologies, Inc. `http://www.rhk-tech.com/wp-content/uploads/2015/12/RHK-R9-BROCHURE-display.pdf`. Last visited January 23nd, 2019.

178. Anfatec Instruments GmbH. `http://www.anfatec.de/cantilevers/umg01.html`. Last visited September 11th, 2018.

179. Boston Piezo Optics Inc. `https://www.bostonpiezooptics.com/lithium-niobate`. Last visited September 12th, 2018.

180. Schröder, M. *et al.* Conducting Domain Walls in Lithium Niobate Single Crystals. *Adv. Funct. Mater.* **22,** 3936–3944 (2012).

181. AIST-NT Inc. `http://nanoprobes.aist-nt.com/apps/HOPG%20info.htm`. Last visited September 14th, 2018.

182. Sweetman, A. *et al.* Mapping the force field of a hydrogen-bonded assembly. *Nat. Commun.* **5** (2014).

183. Gray, M., Hsu, J., Giovane, L & Bulsara, M. Effect of anisotropic strain on the crosshatch electrical activity in relaxed GeSi films. *Phys. Rev. Lett* **86,** 3598 (2001).

184. Fitzgerald, E. A. *et al.* Relaxed Ge x Si1- x structures for III–V integration with Si and high mobility two-dimensional electron gases in Si. *J. Vac. Sci. Technol. B: Microelectron. and Nanometer Struct. Process. Meas. Phenom.* **10,** 1807–1819 (1992).

185. Zerweck, U., Loppacher, C., Otto, T., Grafström, S. & Eng, L. Accuracy and resolution limits of Kelvin probe force microscopy. *Phys. Rev. B* **71,** 125424 (2005).

186. Döring, J. *et al.* Low-temperature nanospectroscopy of the structural ferroelectric phases in single-crystalline barium titanate. *Nanoscale* **10,** 18074–18079 (2018).

187. Wang, H., Wang, L. & Xu, X. Scattering-type scanning near-field optical microscopy with low-repetition-rate pulsed light source through phase-domain sampling. *Nat. Commun.* **7,** 13212 (2016).

Acknowledgements

Without the support of many people, this work would never have been possible. At this point, I would like to thank everyone, who supported me during my studies.

I met Prof. Dr. Lukas Eng for the first time when i was trying to switch universities during my studies. When the bureaucracy was not allowing me to continue my studies at TU Dresden and i was talking to brick wall, you solved the issue with a single call. Since then, you have been supervising my work. In these years, I profited of your advice in countless occasions. Always pushing me further, I ended up performing experiments I was not convinced would ever work. My group leader Dr. Susanne Kehr has been my second supervisor. The fruitful discussions in numerous 1-to-1 meetings were always a good guidance for my work. I will never have the impressive precision you have when it comes to details of a physical phenomena.

Furthermore, I would like to thank Prof. Dr. Roskos for the interesting and challenging collaboration and of course for the review of this thesis. I hope it is a good read.

Thanks goes to the s-SNOM group, starting with Hans-Georg von Ribbeck, who first performed THz near-field microscopy passed a lot of knowledge needed for this work. I really enjoyed working with the other PhD students: Markus Fehrenbacher, Denny Lang, Lukas Wehmeyer, Tobias Nörenberg and Hamed Pour. Special thanks goes to Thales de Oliveira: You were always a good mate, in relaxed and hard times. I will miss working with you!

Thanks goes to the smp2-group in general, i want to name the veterans Alexander Haussmann and Peter Milde and thank for their knowledge and interpretation of any event happening both in the institute and the university. During the design of the AFM i worked a lot together with the technicians Ralf Raupach and Daniel Dietrich. Without them, the T-Bone would never have reached the working stage.

Most of my work was performed at the HZDR. The list of people who directly or indirectly supported it is long. I want to thank the FELBE people, espacially Stephan Winnerl and Mike Klopf for their scientific input and organizational support during my work. The TELBE team with their

former leader Michael Gensch is a bunch of tough people. Thank you for your input, the possibility to work in your lab, and for the support during the TDS measurements. Special thanks go to Sergey Kovalev and Nilesh Awari for the alignment of the lasers. The support by the ELBE team running the accelerator-based experiments at HZDR is kindly acknowledged. It is amazing that they can get such a large machine running.

I want to thank the Rosa Luxemburg Foundation for both the financial support and their guidance during my time as a scholarship holder. I enjoyed the time ”off-topic” during my thesis a lot. It is good to get distracted every once i a while.

I was accompanied by many friends during my time at the university. As almost no one of them would even dare to take a look at a thesis about near-field optics, i will refrain from writing a list here. Thank you for all the goods times we had together!

Last but not least i want to thank my family. You were always there, in good and bad times. Three decades of unconditional support brought me to this very point.

Erklärung entsprechend § 5.5 der Promotionsordnung

Hiermit versichere ich, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht. Die Arbeit wurde bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Die Dissertation wurde im Institut für Angewandte Physik unter wissenschaftlicher Betreuung von Prof. Dr. phil. II habil. Lukas M. Eng angefertigt. Frühere Promotionsverfahren haben nicht stattgefunden.

Ich erkenne die Promotionsordnung des Fachbereiches Mathematik und Naturwissenschaften der Technischen Universität Dresden an.

Datum: Unterschrift: