Optical Sensing through Metasurface Dispersion Spectroscopy

Bachelor Thesis

Marius Graml

ARC Centre of Excellence for Transformative Meta-Optical Systems
Centre for Gravitational Astrophysics, Applied Metrology Labs

Regensburg University of Applied Sciences
Australian National University

Germany, Australia
22/08/2023

Statutory Declaration

1. I am aware that this copy of the bachelor thesis becomes the property of the Regensburg University of Applied Sciences as an examination submission.

2. I hereby declare that I have independently written this Bachelor's thesis, have not submitted it elsewhere for examination purposes, and have used only the specified sources and materials. Additionally, any verbatim and paraphrased quotations have been appropriately cited.

_____

Place, Date, Signature

Presented by: Marius Graml
Student ID: 3233111
Study Program: Electrical Engineering and Information Technology (B.Eng.)
Time Frame: 01/07/2023 - 01/10/2023
Supervisor: Prof. Dr. Armin Sehr
Secondary Supervisor: Prof. Dr. Roland Schiek

Abstract

This thesis presents a collaborative bachelor's project between the ARC Centre of Excellence for Transformative Meta-Optical Systems (TMOS) and the Applied Metrology Labs (AML), affiliated with the Centre of Gravitational Astrophysics, at the Australian National University. It serves as the final project to obtain the degree of Bachelor of Engineering from the Regensburg University of Applied Sciences.

The primary research objective is the development of a dispersion based metasurface sensor with the capability of detecting extremely low analyte concentrations. Towards this goal, this project aims to develop the dispersion spectroscopic technique which will be employed by the sensor to detect changes in the metasurface frequency response. Therefore, the research focuses on establishing and analyzing a state-of-the-art dispersion spectroscopy setup to detect the phase signal encoding the analyte information. This primarily involves enhancing the system's Signal-to-Noise Ratio (SNR) and verifying its capacity to identify actual signals, ultimately constructing a suitable sensing system capable of detecting resonance shifts of the metasurface's frequency response below 1nm.

For that, the measurement system is thoroughly analyzed, encompassing its noise behavior, the measurement signal amplitude, and the validation of a real signal measurement using a hydrogen cyanide (HCN) vapor cell. The implementation of a serial Acousto-optic Modulator (AOM) structure results in a noise limit of $1.9 \frac{\mu \text{rad}}{\sqrt{\text{HZ}}}$, which can be further reduced to an impressive noise limit of $0.6 \frac{\mu \text{rad}}{\sqrt{\text{HZ}}}$ by transitioning to a parallel structure, surpassing the performance of most state-of-the-art spectroscopy setups. Theoretical considerations and simulations suggest that the serial AOM structure provides the highest measurement signal amplitude. The real signal validation, achieved through measurements within the HCN resonance area, exhibits favorable alignment between the theoretical assumption derived from simulation and the real measurement, demonstrating a deviation of 4.2%.

Leveraging the noise measurements, the minimal detectable resonance shifts of the system for specific metasurfaces are predicted. This is facilitated by simulating the phase responses of two distinct metasurfaces with slightly different bandwidths and digitally integrating them into the system. Calculations reveal that the system is projected to detect a minimal resonance shift of $\Delta \lambda_{min} = 0.8707$nm for the slightly narrower metasurface. This prediction suggests that the system has the potential to achieve its goal of detecting a minimal resonance shift below 1nm, which is notably impressive considering the utilized metasurface's relatively broad bandwidth.

Moreover, the optimal AOM-induced frequency shifts for both metasurfaces are calculated as 0.7958THz and 0.8539THz, significantly exceeding the current capabilities of standard AOMs. Nevertheless, both the minimal detectable resonance shift and the optimal AOM-induced frequency shift can be notably reduced by narrowing the bandwidth of the metasurface, potentially favoring the parallel AOM structure over the serial one in terms of SNR at a certain bandwidth.

Dedication

To the obstacles that became stepping stones.

Acknowledgements

Contents

Chapter 1

Introduction

This chapter provides a foundational understanding of the subjects relevant to the ongoing project. It begins with a concise overview of spectroscopy and optical sensing, highlighting the fundamental process of concentration measurement using spectroscopy and the advantages of dispersion spectroscopy over absorption spectroscopy, along with its limitations. The subsequent segment examines the primary criteria for optical sensors within the optical sensing domain. The focus then shifts to metamaterials and metasurfaces, exploring their functionality in optical sensing and concentration measurement. It is explained how metasurfaces can effectively overcome the limitations commonly associated with dispersion spectroscopy while still retaining its core benefits. In addition, potential applications of the metasurface-based dispersion spectroscopy sensor are discussed before presenting the final research goal.

For a comprehensive grasp of the project's considerations, it is strongly advised to acquaint oneself with the theoretical background, which incorporates essential formulas and intricate physical considerations. These details can be found in the appendix A located at the end of this thesis. By delving into the theoretical background provided in the appendix, readers can gain a deeper understanding of the project's fundamental principles. This comprehensive resource empowers the audience to further comprehend the project's methodology, results, and implications, fostering a more enriched and informed appreciation of the research undertaken.

1.1   Spectroscopy and Optical Sensing

Spectroscopy is a scientific discipline used to study the interaction between matter and electromagnetic radiation. At the core of spectroscopy lies the fundamental theory that light is composed of various wavelengths, and each element in the periodic table responds uniquely to electromagnetic waves by altering their properties, such as amplitude or phase. This response encompasses a range of scattering phenomena, including emission, absorption, transmission, and reflection, which depend on the frequency of the wave as exhibited by the Lorentz oscillator model in section A.2.2.

A variety of spectroscopy techniques are available, each distinguished by the manner in which they investigate the interaction between electromagnetic waves and the material under study. Prominent examples include ultrafast laser spectroscopy, nuclear magnetic resonance spectroscopy, absorption spectroscopy, and dispersion spectroscopy. In the context of this thesis, particular emphasis is placed on absorption spectroscopy and dispersion spectroscopy due to their relevance and significance in examining the concentration of gases which is the goal of this project.

1.1.1   Absorption Spectroscopy

Absorption spectroscopy measures how much light is absorbed by a sample as a function of wavelength. This technique delivers the absorption spectrum of the analyte which is characteristic for its type and properties, such as concentration and pressure. The amount of light absorbed by the sample is defined by the Beer-Lambert law [1] (Eq. 1.1).

$$A(\omega) = \epsilon(\omega, c)cz \tag{1.1}$$

The parameter A describes the absorbance of the electromagnetic wave, $\epsilon(\omega)$ is called the molar attenuation coefficient (not permittivity) and takes the wavelength of the wave into account, c is the molar concentration

of the medium and z the optical path length. Due to the molar attenuation coefficient being a nonlinear function of the analyte concentration c [1], the absorption of light is not strictly linearly proportional to the concentration, as one might initially perceive based on the Beer-Lambert law. This nonlinear correlation becomes increasingly prominent at higher concentrations, as demonstrated in Eq. A.47. This is because the charge density per unit volume, denoted by the parameter N, is linearly connected to the analyte concentration c through the equation $N = N_A * c$, where N_A represents the Avogadro constant [1]. Consequently, the electric susceptibility is also linearly proportional to the analyte concentration. However, the absorption coefficient exhibits a nonlinear relationship with the electric susceptibility (see Eq. A.54), resulting in a nonlinear correlation between the analyte concentration and the absorption coefficient.

Instead of determining the absorbance, the transmittance T of the intensity can also be calculated [1] (Eq. 1.2).

$$T(\omega) = 10^{-A(\omega)} = 10^{-\epsilon(\omega)cz} \tag{1.2}$$

Hence, the relationship between the Beer-Lambert law (Eq. 1.2) and the absorption coefficient α from section A.2.3.1 can be shown, as the transmittance T is defined as $T(z) = \frac{I(z)}{I_0}$ where $I(z)$ is the intensity of the electric field after the distance z and I_0 is the intensity of the initial field, i.e., the field at $z = 0$. Since the base 10 can be transformed to base e by $10^{-\epsilon cz} = e^{-2.3\epsilon cz}$, it is obtained (Eq. 1.3):

$$I(z) = (E_0 e^{-\frac{1}{2}\alpha z})^2 = I_0 e^{-2.3\epsilon cz} \tag{1.3}$$

Therefore, a direct relationship between the absorption coefficient α, electric susceptibility, molar attenuation coefficient ϵ, and molar concentration c of the medium can be observed which confirms the nonlinear relationship between the molar attenuation coefficient and the molar concentration.

To summarize, Eq. 1.1 and 1.2 demonstrate that absorption is frequency dependent, primarily due to the frequency dependence of the molar attenuation coefficient $\epsilon(\omega)$. Moreover, since in absorption spectroscopy the transmitted light is commonly detected, this technique introduces disadvantages compared to dispersion spectroscopy, which is primarily due to the nonlinear relationship between the molar attenuation coefficient (or absorption coefficient) and the molar concentration as further explained in section 2.1.

1.1.1.1 Methods

Multiple absorption spectroscopy techniques exist, varying in complexity and sensitivity. One of the simplest methods is Tunable Laser Direct Absorption Spectroscopy (TLDAS), which only involves measuring the power transmitted through a sample. When sweeping the wavelength of a diode laser across a specific absorption line of the analyte, TLDAS utilizes the optical absorption to estimate the concentration of the gas by employing the Beer-Lambert law [2]. However, TLDAS is notably limited in that it relies on detecting a small signal change against a significant background. This means that any noise introduced by the light source or the optical system can compromise the detectability of the technique, resulting in limited sensitivity being far from the ideal shot noise level. To overcome this, there are two main approaches. Firstly, reducing noise in the signal can be achieved through the use of modulation techniques [3]. Secondly, increasing the absorption can be accomplished by placing the gas within a cavity, allowing the light to pass through the sample multiple times, thereby increasing the interaction length.
A common modulation technique used in absorption spectroscopy is Wavelength Modulation Spectroscopy (WMS). This technique leverages the characteristic that technical noise tends to decrease as the frequency increases. By encoding and detecting the absorption signal at a high frequency, where the noise level is low, WMS improves the SNR of the measurement [4]. As a consequence of frequency modulation, the beam traversing the medium manifests a beat note comprising a carrier frequency and a sinusoidal modulation signal. Subsequently, when this signal is demodulated by the photodetector, the modulation signal is converted into an amplitude modulation and the received signal represents not the actual transmission spectrum but rather its first derivative [5,6].
Another approach to enhance the detectability of TDLAS is by increasing the interaction length of the light with the analyte. This can be achieved by placing the analyte within a cavity where the light undergoes multiple reflections. Techniques employing this method are referred to as Cavity-Enhanced Absorption Spectroscopy (CEAS).

1.1.2 Dispersion Spectroscopy

Dispersion spectroscopy exploits the different phase speeds of electromagnetic waves in a medium depending on their frequency. The approach of this technique is to measure the phase delays of electromagnetic waves as a function of frequency, allowing for the reconstruction of the refractive index profile, which is characteristic for the analyte and its properties, such as concentration. So, the core principle of dispersion spectroscopy revolves around the linear relationship between the analyte's refractive index and the phase shift caused by the analyte, as demonstrated in Eq. A.59.

A widely used dispersion spectroscopy technique is Heterodyne Phase-Sensitive Dispersion Spectroscopy (HPSDS) [2, 7]. The main concept involves modulating the intensity of a tunable laser at a high frequency to generate an optical spectrum consisting of three tones. This spectrum is then swept across the desired spectral range. When the optical signals interact with a sample, changes in the refractive index near molecular transitions result in a relative phase shift among the three optical waves. This phase shift is reflected in the phase of the beat note when the optical signals are detected by a photodetector. By measuring the phase shift of the beat note, the dispersion spectrum of the sample can be determined. From this information, the gas species present in the sample and their characteristics can be inferred.

To delve deeper into the principle of dispersion spectroscopy, section 2.1 provides an in-depth exploration of a previously investigated dispersion spectroscopy approach, similar to WMS and HPSDS but specifically tailored for dispersion spectroscopy. This approach employs a highly noise stable interferometer technique, making it not only suitable for explaining dispersion spectroscopy in greater detail but also serving as the baseline model for the project. The main distinction of HPSDS from this baseline model is that the laser beam passing through the sample consists of three tones instead of two, while still operating on the same fundamental principle.

1.1.2.1 Principle of Concentration Measurement

One primary objective of dispersion spectroscopy sensing applications is to determine the concentration of the analyte, which can be achieved through the linear relationship between the refractive index n, the phase shift ϕ, and the concentration c for small concentrations, as shown in Eq. A.59. Eq. 1.4 portrays the relationship, wherein the proportional constants a and b facilitate the direct determination of the analyte concentration from the phase shift signal.

In the context of improved dispersion spectroscopy techniques like HPSDS or the baseline model, the measurement of analytes relies on a differential phase shift $\Delta\phi$ instead of a single phase shift ϕ. The reason behind this choice is that the differential phase shift effectively eliminates phase noise, which will be further discussed in section 1.1.3.

$$c = a * n = b * \phi \tag{1.4}$$

The entire procedure of measuring the analyte concentration involves first obtaining the curve of the phase signal $\phi_{known}(\omega)$ for a known analyte concentration c_{known}. This curve, depicted as the black curve in Fig. 1.1, represents a fundamental phase response of a medium as illustrated in Fig. A.7. Next, the analyte is illuminated with the wavelength λ_0, which corresponds to the maximum absolute value of the phase signal, and the new phase shift ϕ_{new} is measured. This value is represented by the red dot on the green curve in Fig. 1.1. By utilizing Eq. 1.4, the ratio of the measured differential phase shifts provides a constant that indicates the concentration of the analyte as shown in Eq. 1.5.

$$\frac{e * \phi_{new}}{e * \phi_{known}} = \frac{c_{new}}{c_{known}} = k$$
$$c_{new} = k * c_{known} \tag{1.5}$$

The wavelength λ_0 is chosen in the measurement process because at this specific wavelength, the difference between the two phase shifts is maximized, resulting in a higher SNR. The identical procedure is employed for measuring the analyte concentration in absorption spectroscopy. The only distinction lies in the fact that, in this case, the analysis is performed on the transmission spectrum rather than the dispersion spectrum.

It is important to emphasize that this approach is only suitable for low concentrations, as the relationship between the propagation constant and the analyte concentration is typically nonlinear [8], which aligns with the reasons discussed for the absorption coefficient in section 1.1.1. The nonlinear association between the propagation constant and the electric susceptibility (as observed in Eq. A.55) further contributes to the nonlinear correlation between the analyte concentration and the propagation constant. Consequently, both
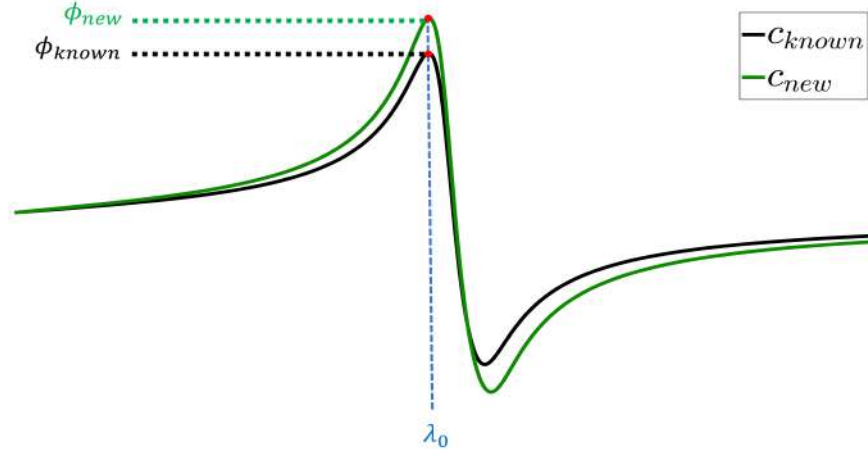
Figure 1.1: Determining analyte concentration from the phase signal $\phi(\omega)$.

the refractive index and the phase shift, being linearly linked to the propagation constant (see Eq. A.55 and Eq. A.59), also exhibit nonlinear dependence on the analyte concentration. The nonlinear behavior of the phase shift can be visualized by plotting these parameters as functions of the parameter χ_0 from Eq. A.47, which is linearly proportional to N (see Eq. A.48) and consequently to the analyte concentration (see Fig. 1.2a).
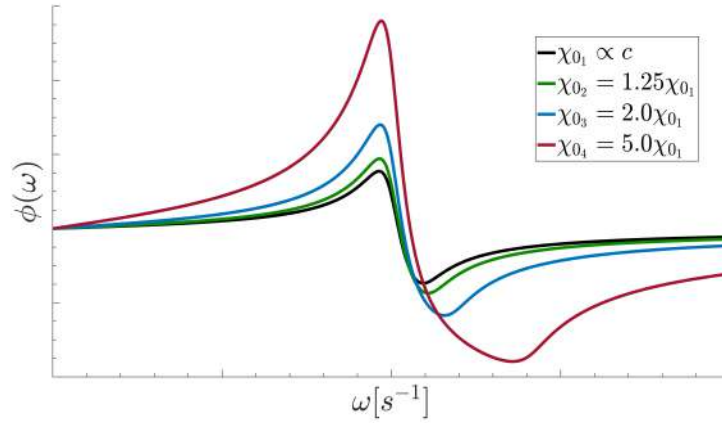
(a) $\phi(\omega)$ for different concentrations represented by different values of χ_0.
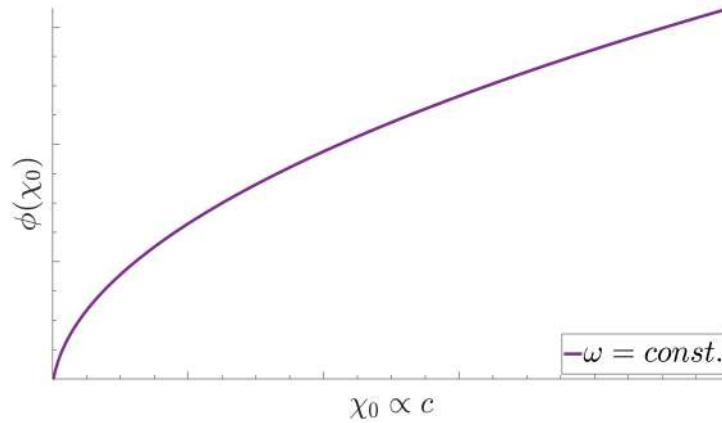
(b) Phase signal ϕ as a function of χ_0 and thus concentration c.

Figure 1.2: Nonlinearity between phase response of medium and its concentration.

Fig. 1.2a clearly illustrates that the phase curves representing $\phi(\omega)$ exhibit increasingly noticeable distortions as the χ_0-values increase. This nonlinearity is further evident in Fig. 1.2b, where the phase shift is plotted as a function of χ_0 and, consequently, the analyte concentration. This means, to accurately measure higher concentrations, a calibration curve must be established, mapping each analyte concentration c_{new} to its corresponding scaling factor k through a function $c_{new} = f(k)$. Consequently, for every value of k the corresponding concentration c_{new} is known. Nevertheless, since in most cases very low concentrations want to be measured, the linear proportionality usually applies.

1.1.3 Evaluation

The main advantage dispersion spectroscopy offers over absorption spectroscopy is that the nonlinearity of the refractive index is less significant in light detection than the nonlinearity of the absorption coefficient. This is because dispersion spectroscopy detects the phase shift of the transmitted light, while absorption spectroscopy detects its amplitude. The relationship between the refractive index and the phase shift is linear, as shown in Eq. A.59, whereas the relationship between the transmission and the absorption coefficient α (or molar attenuation coefficient ϵ) is nonlinear and given by $T(\omega) \propto e^{-\epsilon(\omega)} \propto e^{-\alpha(\omega)}$, resulting in greater significance of the nonlinearity of the refractive index. Consequently, the relationship between concentration and the refractive index approximates linearity over a wide range of concentrations [8], while a linear relationship between absorbed light and concentration is only valid up to 10% of sample absorption [9]. Since linearity is critical for accurately detecting analyte properties, especially concentration as explained above, many absorption spectroscopy systems are optimized to work only within the linear regime and thus provide only a small dynamic range. This limitation makes absorption spectroscopy less practical, less flexible and less sensitive compared to its dispersion counterpart.

Another advantage of phase measurement is its inherent immunity to fluctuations in optical power [7, 10]. This immunity eliminates the need for signal calibration, which is often required in various absorption spectroscopy techniques such as TLDAS.

In addition to that, dispersion spectroscopy offers specific techniques that can enhance the SNR. One such technique, not only used in HPSDS but also in various other dispersion spectroscopy techniques, is its noise stability. As the electromagnetic waves contributing to the differential phase shift propagate through the same medium, they encounter the same noise sources. Since the phase signal is obtained from the phase difference between the waves, any phase noise present is automatically canceled out which enhances the overall noise immunity of the system. This feature is also advantageous for mitigating phase errors caused by laser frequency noise, specifically Rayleigh-Backscattering effects [11–13], which can adversely affect the accuracy of the phase measurement. Laser frequency noise typically exhibits a characteristic 1/f-profile in the Amplitude Spectral Density (ASD), increasing at low frequencies [14]. However, in the case of the differential phase signal, which serves as the spectroscopic readout, its ASD appears flattened compared to the ASDs of the individual phase signals, allowing for a noise reduction of at least five orders of magnitude for 1/f phase noise [15].

Moreover, one specific advantage applicable to HPSDS and the baseline model in section 2.1 is their simpler implementation compared to other dispersion spectroscopy techniques like Frequency Modulation Spectroscopy (FMS) or Chirped Laser Dispersion Spectroscopy (CLaDS). Despite their simpler implementation, the HPSDS approach maintains a similar dynamic range and performance level as CLaDS [2]. The baseline model even exhibits superior performance compared to the majority of other spectroscopy setups in terms of its noise limit. Therefore, using this architecture as a basis, the metasurface setup is expected to maintain a relatively straightforward and uncomplicated design.

Nevertheless, dispersion spectroscopy also faces disadvantages. As illustrated in Fig. A.7, the phase shift encountered by a propagating wave in a medium is typically exceedingly small. As a result, dispersion spectroscopy necessitates highly sensitive and precise measurement techniques, often entailing complex hardware and methodologies. This complexity is even further exacerbated when dealing with very low concentrations. In contrast, the implementation of absorption spectroscopy is typically less complex, but it also provides lower precision compared to dispersion spectroscopy. Moreover, dispersion spectroscopy lacks selectivity, as it can be difficult to ensure that only the desired analyte influences the detected phase shift, leading to potential interferences from other species. Additionally, this technique has limited flexibility, for example in terms of laser wavelength, as the resonance frequency of the analyte being analyzed determines the required wavelength of the laser as well as in terms of analyte detection since common dispersion spectroscopy techniques are insensitive to molecules such as anti-bodies or cancer cells. All these challenges can be addressed by incorporating metasurfaces into the concept of dispersion spectroscopy, as discussed in section 1.2.3.2 and

1.2.3.3 by still maintaining the mentioned advantages dispersion spectroscopy offers.

1.1.4 Optical Sensing

The different spectroscopy techniques, such as absorption or dispersion spectroscopy, are applied in the field of optical sensing. Optical sensing identifies chemical or biological analytes by illuminating a sample and detecting the encoded information in the transmitted or reflected light. While passing through the sample, the light interacts with the analyte and its properties, such as intensity or phase, may be changed. These changes can then be detected and measured to retrieve the encoded information about the analyte, such as type and concentration. The performance and evaluation of optical sensors and especially optical biosensors depends on different factors.

The most important parameter for evaluating a sensor's performance, regardless of its design mechanism, is sensitivity. Sensitivity is typically defined as the change in the sensor's response signal per unit change in the quantity or property of the analyte being measured. For example, if the sensor's response is measured in terms of the phase shift of an electromagnetic wave, as discussed in section A.2.3.2, with changes in analyte concentration, sensitivity can be calculated by taking the first derivative $\frac{\partial \phi}{\partial c}$ of the phase shift with respect to the analyte concentration. Thus, sensitivity is determined by differentiating the output signal with respect to the input signal in the response curve. Typically, sensors that exhibit a steep and linear response curve are considered to have the best performance because they offer very high and stable sensitivity across the entire range of operation. Especially in the case of biosensors, the sensor performance is commonly characterized by a Limit of Detection (LOD), which refers to the lowest extractable signal that can be distinguished from noise with a defined degree of confidence and therefore depends on the SNR. According to International Union of Pure and Applied Chemistry (IUPAC), the LOD is defined as [16]

$$LOD = \mu_B + 3\sigma_B \tag{1.6}$$

with μ_B and σ_B being the mean and the standard deviation of the sensor response to a blank sample. By incorporating metasurfaces in optical sensing, a substantial improvement in sensitivity is anticipated compared to conventional dispersion spectroscopy methods. This enhancement stems from the metasurface's ability to generate enhanced electromagnetic near fields, significantly increasing the interaction between light and analytes as explained in detail in section 1.2.3.2.

Selectivity is another essential sensor parameter, which refers to the ability of a sensor to distinguish a specific analyte from other interfering analytes. Poor selectivity can result in the sensor response being distorted by the presence of other analytes, leading to inaccurate measurements. Especially for metasurface-based sensors, special techniques exist in order to improve the sensors sensitivity and selectivity as explained in section 1.2.3.2 and 1.2.3.3.

One additional vital requirement is label-free detection. Conventional methods often rely on labeling the analyte molecules for detection purposes, which means employing an immobilized ligand to capture the target analyte and utilizing specialized detection reagents to analyze the interaction between the two molecules. However, labeling can cause structural or functional changes in biomolecules and distort experimental outcomes. In some instances, it can even render the biomolecule being studied completely inactive. Label-free quantification of molecular interactions offers an appealing alternative to traditional label-based approaches, making the overall detection process more reliable and shorter. Label-free sensing employs optical biosensors to measure changes that occur when the analyte binds to a biosensor surface. In the case of metasurface sensing, the alteration in the refractive index of the environment surrounding the metasurface is detected, as explained in greater detail in section 1.2.3. This approach allows for real-time monitoring of the interaction, enabling to perform analyses that were previously unattainable using conventional methods, which only provide endpoint results. Real-time detection, particularly, plays a pivotal role in pathogen detection, as further elaborated in section 1.3.

Multiplexed detection holds significant importance in biosensors as it greatly impacts detection times and analysis accuracy. The primary advantage of multiplexed analysis lies in its ability to detect multiple biomarkers qualitatively or quantitatively within a single sample. This is particularly valuable when maximizing the information obtained from a given sample is essential for reaching a conclusive result. Moreover,

multiplexed analysis offers a substantial reduction in overall detection time and resources by allowing for the simultaneous detection of multiple analytes in a single measurement step. This proves advantageous when sample volumes are limited, as it minimizes the amount of analyte samples required, allowing for a more efficient analysis process. The successful development of a multiplexed LSPR biosensor by Kim et al., capable of identifying up to four different bacterial species in a single test [17], suggests that metasurface sensors could also possess the capability for such multiplexed detection.

Finally, the compactness of a biosensor while maintaining high sensitivity is another crucial requirement to consider, especially when the sensor is intended for use as a point-of-care (POC) device. The compact design of a biosensor ensures portability and ease of use in various settings, including clinics, remote locations, or even home environments. POC devices are of paramount importance in monitoring health status, detecting diseases, and identifying pathogens at an early stage, as elaborated in section 1.3. However, despite the reduced size, it remains essential to maintain a high level of sensitivity to enable accurate and reliable detection of analytes.

In conclusion, the metasurface-based sensor proves to be an ideal solution, meeting all the required criteria. Its inherent label-free and real-time detection capabilities, along with the potential for multiplexed detection, make it highly advantageous. The compact size of the metasurface contributes to the sensor's overall compactness. Furthermore, the enhanced sensitivity and selectivity, as detailed in the subsequent sections, demonstrate the sensor's potential to surpass conventional dispersion spectroscopy techniques in terms of performance. Moreover, the metasurface-based approach also offers higher design flexibility compared to common dispersion spectroscopy techniques. Despite offering these numerous advantages, the metasurface-based sensor retains the key benefits of dispersion spectroscopy, including a linear relationship between the measured signal and analyte concentration, immunity to fluctuations in optical power, inherent laser noise stability, and a straightforward design. These combined advantages position the metasurface-based sensor as a promising and versatile tool for various optical sensing applications, paving the way for new possibilities in the field of spectroscopy and beyond.

1.2   Metamaterials and Metasurfaces

All natural materials are composed of atoms. According to the different arrangements of atoms, natural materials are basically classified into three types - crystals with periodic arrangement of atoms, noncrystals with random arrangement of atoms and quasicrystals, which consists of arrays of atoms that are ordered but not strictly periodic [18]. Similar to natural materials, metamaterials are macroscopic composites of periodic or nonperiodic subwavelength-scale artificial particles with certain shapes, materials and geometries. Referring to natural materials, these nanoparticles are called meta-atoms. The main idea of metamaterials is to mimic and extend the behavior of real materials to electromagnetic waves [18]. This means, metamaterials are able to interact stronger with electromagnetic waves and manipulate their properties, such as amplitude, phase, wavelength, and polarization, in a more advanced way than natural materials can do, which opens the door for manipulating electromagnetic waves in a more pliable way. The key for this process are the metamaterial's meta-atoms.
The meta-atoms, the fundamental modules of metamaterials, are electric and/or magnetic resonators, also called scatterers. These structures resonantly couple to the electric and magnetic fields of the incident electromagnetic wave, showcasing responses that are characterized by their polarizability [19]. This desired response is based on the fundamental principle of scattering, as explained in section A.2.2, where the re-radiation of each meta-atom is a result of the presence of electric and magnetic modes, often referred to as Mie-type or Fano resonances. Since the phenomenon of scattering is the foundation for all other types of light-matter interaction as mentioned in A.2.2, being able to control the scattering effect of the meta-atoms allows manipulating all properties of the incident light. So, the key is to artificially engineer the meta-atoms in a way so that they respond to an incident electromagnetic wave in a desired manner which can be done by changing the properties of the meta-atoms, such as size, shape or arrangement. As examined in appendix A, natural atoms also exhibit resonance behavior, which also corresponds to their polarizability. Hence, meta-atoms in metamaterials take the scattering role of natural atoms in natural materials but can, as a result of their flexible design, manipulate the incident light in a more customizable way, resulting in extremer values of effective medium parameters (e.g., negative permittivity, negative permeability and zero index of refraction), which could not be achieved by natural materials [18]. Thus, compared to natural effective media,

the main advantage of metamaterials is the flexibility of design. Meta-atoms can be engineered and tailored as required and hence may provide countless choices in terms of scattering behavior whereas natural atoms are very limited. Furthermore, meta-atoms can be arbitrarily arranged but it is difficult to arrange natural atoms to specific orders. As a result, homogeneous metamaterials with periodically arranged particles, called super crystals, inhomogeneous metamaterials with quasiperiodically structures, called quasicrystals, and random metamaterials with randomly arranged meta-atoms, called super noncrystals, can be realized. Since the accumulative response of all meta-atoms, which are defined by their polarizability, determines the behavior of the metamaterial, each metamaterial is defined by its electric susceptibility, according to Eq. A.17, and thus permittivity and refractive index.

In the realm of optical sensing, the unique ability of meta-atoms acting as optical resonators and confining light within nanoscale electromagnetic hotspots can be leveraged to enhance the sensitivity of various detection techniques. The strong resonances observed in metamaterials generate near-field hotspots, resulting in significant light-matter interactions with analytes in close proximity to the resonators, thus increasing the change of amplitude or phase of the detected wave. Binding of analytes to the resonators induces local changes in the refractive index, which can then be detected by monitoring the amplitude or phase response shift of the metamaterial, as elaborated further in section 1.2.3.

Metasurfaces, as a subgroup of metamaterials, represent ultra-thin 2D planar arrays of sub-wavelength resonant nanostructures. Equal to metamaterials, metasurfaces are capable of arbitrarily modulating the electromagnetic response over a broad spectral range by changing the wave's properties, including spectra, phase, and polarization [20]. However, due to their different structure, metasurfaces exhibit lower radiation losses, smaller thickness and more compactness than bulky metamaterials [21] and are easier to fabricate and more cost effective [22], positioning them as highly promising components in integrated optical platforms including sensor technologies.

Fig. 1.3a - 1.3c show the same metasurfaces from different perspectives. Each single 3D-object represents a meta-atom, i.e., electromagnetic resonator. As depicted in Fig. 1.3a, the meta-atoms have a size of about $500nm$ which is about the wavelength of blue light. The images deliver a good comparison of different metasurfaces and illustrate the different shapes and arrangements of the meta-atoms pretty well.
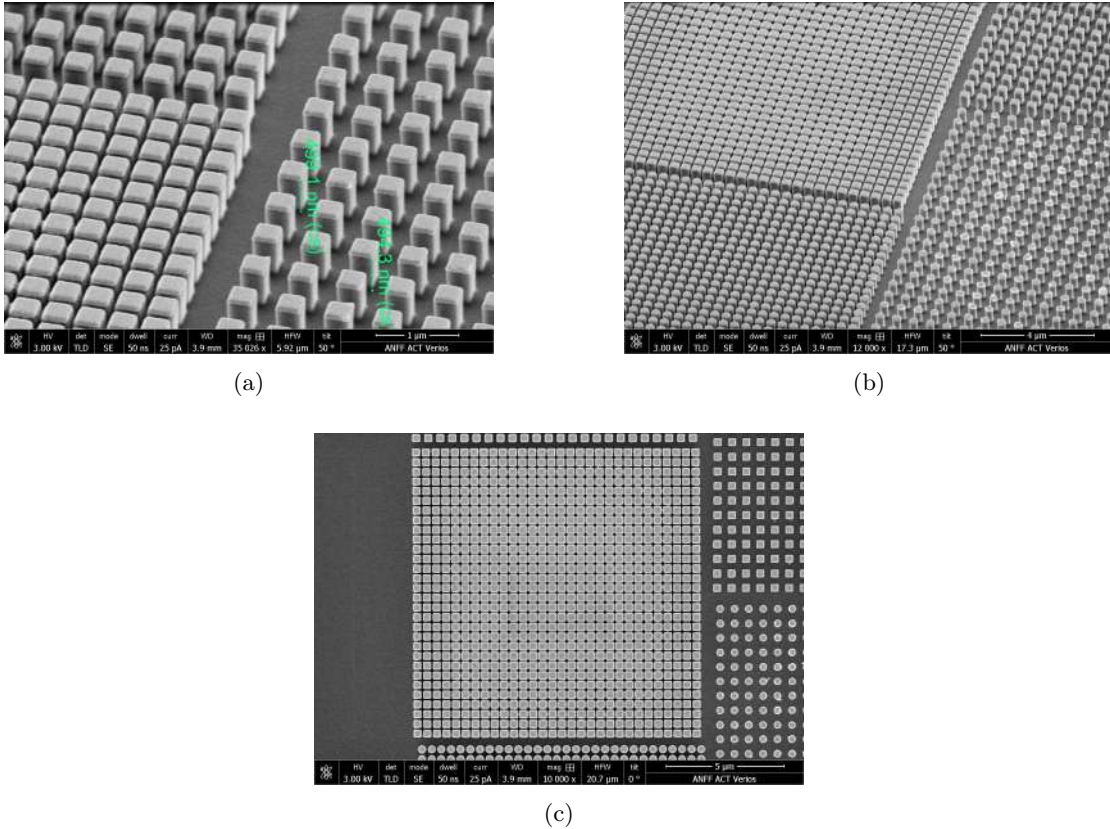

(a)

(b)

(c)

Figure 1.3: Metasurfaces from different perspectives.

1.2.1 Types of Metamaterials

Basically, two classes of metamaterials can be defined, plasmonic metamaterials and dielectric metamaterials. Whereas plasmonic metamaterials have been investigated for many years, dielectric metamaterials have been discovered recently and show huge potential in overcoming plasmonic metamaterials' disadvantages by enabling strong Mie-type resonances of both electric and magnetic character.

Plasmonic metamaterials are metallic nanostructures, usually made out of gold, silver and aluminium, and have extensively been studied over the last ten years. In these structures, the resonances arise from collective oscillations of the electrons in the metal which is called Localized Surface Plasmon Resonance (LSPR) [20]. By varying size, shape or interparticle distance of the plasmonic particles, the resonance frequency can be turned across a wide spectral range [20]. The drawback of plasmonic nanostructures are the strong absorption losses of metals at optical frequencies which hampers the performance of most of their potential applications in nanophotonics [22, 23].

Dielectric metamaterials have emerged as a more recent approach in nanostructure design, effectively addressing the issue of high losses in plasmonic metamaterials. Unlike metallic nanoparticles, dielectric nanoparticles exhibit minimal losses at optical frequencies while still exhibiting resonant behavior analogous to plasmonic resonances [23]. The resonances in dielectric particles arise from the oscillations of circular displacement currents in the dielectric medium [23]. Basically, for spherical dielectric nanoparticles, these resonances can be described by the so called Mie theory and are characterized by electric and magnetic modes. Nevertheless, due to the advantages offered by a large number of geometrical degrees of freedom, non-spherical nanoparticles are receiving more and more research attention [23]. In such complex systems, the original Mie solution does not directly apply, but numerical techniques can be employed to solve Maxwell's equations in three dimensions, enabling the determination of the electric and magnetic modes [23].

Due to their lower losses, dielectric nanostructures exhibit higher Q-factors, which is crucial for enhancing the detection sensitivity of metasurfaces in optical sensing applications as further examined in section 1.2.3.2. However, it is important to note that the modes in dielectric nanostructures are generally less localized and extend over larger volumes compared to their plasmonic counterparts [20]. Depending on specific circumstances, this property can either be advantageous or disadvantageous when utilizing dielectric metasurfaces in optical sensing applications (see 1.2.3.2). Moreover, while plasmonic metasurfaces offer a broader dynamic range, dielectric metasurfaces provide a superior LOD due to their higher Q-factor [24]. In the context of determining very low analyte concentrations, a higher LOD holds greater advantages than a wide dynamic range.

1.2.2 Optical Response of Metamaterials

As previously mentioned, much like natural materials that interact with light based on their electric susceptibility (see section A.2.2), which stems from the polarizability of their atoms or molecules, the optical response of metamaterials and metasurfaces is given by the cumulative resonant behavior of their meta-atoms which follows from their polarizability (see Eq. A.17). As shown by Decker et al. [19], for dielectric metamaterials the effective electric and magnetic polarizabilities of the meta-atoms can be described as (Eq. 1.7)

$$\alpha^e_{eff} = \frac{\alpha^e_0}{\omega^2_{e,0} - \omega^2 - j\gamma_e\omega}$$
$$\alpha^m_{eff} = \frac{\alpha^m_0}{\omega^2_{m,0} - \omega^2 - j\gamma_m\omega}$$

(1.7)

with $\alpha^{e,m}_0 = \frac{4Ac_0}{n_d}\gamma_{e,m}$ taking the area A and the refractive index n_d of a meta-atom into account and $\omega_{e,m}$ and $\gamma_{e,m}$ being the resonance frequency and damping coefficient of the meta-atom. Upon comparison of Eq. 1.7 and Eq. A.47 with the consideration of Eq. A.17, it can be observed that both natural and artificial dielectric materials exhibit a similar frequency response in their optical behavior since both equations possess a Lorentzian line shape. The different algebraic signs result from different definitions of the incident wave. Whereas for Eq. A.47 the electromagnetic wave was defined as $\underline{E} = Ee^{j(\omega t - kz)}$, in Eq. 1.7 the wave was defined as $\underline{E} = Ee^{-j(\omega t + kz)}$ which corresponds to a flip of the frequency response at the horizontal axis. This means, the amplitude and phase response curve of dielectric metamaterials is predominantly determined by the resonance frequency and damping coefficient of the corresponding meta-atoms whereby the resonance frequency defines the position and the damping coefficient the Full Width at Half Maximum (FWHM) of the

curve. Consequently, the frequency response is similar to the one of the electric susceptibility (see Fig. A.5), but mirrored at the horizontal axis. Hence, the resulting phase and amplitude response curve of a dielectric metasurface with respect to wavelength are illustrated in Fig. 1.4.
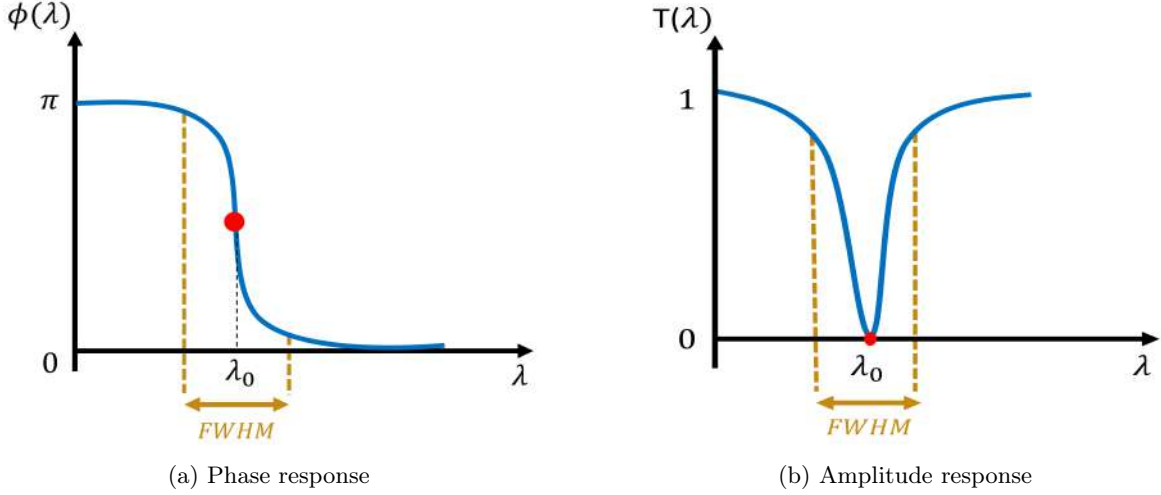

(a) Phase response (b) Amplitude response

Figure 1.4: Metasurface frequency response and corresponding FWHM, λ_0 represents the resonance wavelength.

The optical response of dielectric metamaterials can be applied to plasmonic metamaterials as well. As plasmonic meta-atoms are modeled as LC circuits where the resonance frequency is determined by $f_0 = \frac{1}{2\pi\sqrt{LC}}$ [22], they basically represent a damped harmonic oscillator, equally to natural atoms which are described by the Lorentz oscillator model from section A.2.2. As a result, both individual plasmonic meta-atoms and entire plasmonic metasurfaces exhibit similar system responses as a damped harmonic oscillator or a second-order system, akin to natural atoms or mediums characterized by their susceptibility which has also been demonstrated by Cui et al. [18].

It is important to mention that the frequency response described above applies primarily to straightforward metasurfaces. More advanced metasurfaces also exist, featuring phase responses that range from 2π to 0 or even exhibit overlapping resonances, as demonstrated in section 5.3. Nevertheless, the discussed response serves as a fundamental foundation for the subsequent considerations presented in this thesis.

1.2.3 Functionality in Optical Sensing

Metamaterials and especially metasurfaces can be utilized in many different fields of application. One area, in which metasurfaces show huge potential, is optical sensing for the detection of analytes. The key of this method lies in a change of the refractive index of the medium surrounding the metasurface which alters the resonance behavior of the metasurface. As this change in resonance behavior affects both the amplitude and phase responses of the metasurface, two main techniques emerge for detecting analytes and their properties: measuring the intensity or the phase of the light wave.
Both approaches enable label-free optical sensing and can be further enhanced through surface functionalization and texturing (see section 1.2.3.3). While utilizing light intensity to extract analyte information has been extensively studied, this project focuses on exploiting the light's phase as the measurement signal. By exploring phase-based measurements, the research aims to unlock new possibilities and advancements in optical sensing, contributing to the development of more sensitive and advanced sensing technologies.

1.2.3.1 Analyte Induced Resonance Shift

The important effect, which enables metasurfaces being used for analyte detection, is their interaction with the analytes. The meta-atoms of the metasurface generally interact with their surroundings through their electromagnetic modes. Any changes in the refractive index of the environment, caused - for instance - by the presence of analyte molecules, can influence the modes and thus alter the resonance behavior of the meta-atoms. By detecting these changes in resonance, valuable information about the analyte surrounding the

metasurface can be obtained. In order to comprehend this interaction, it must be considered as a quantum mechanical process at a microscopic level as demonstrated by Arnold et al [25].

One homogeneous spherical meta-atom with radius R is considered. A general oscillating EM mode in each meta-atom can be perceived as multiple single photon resonant states. If one analyte molecule is attached to a meta-atom, a single photon resonant state induces a dipole moment in the molecule through its electromagnetic near field. Both the near field and the dipole moment contain energy, as per the Planck-Einstein relation $\hbar\omega$ and the energy pE of a dipole in a electric field E. However, the direction of the dipole's field is opposite to that of the photon's resonant field, similarly to Fig. A.1. This implies that the energy of the photon is negatively shifted relative to the energy of the dipole. For one molecule positioned at $\vec{r}_i$, this energy shift can be expressed as $\Delta\hbar\omega$, and dividing the energy of the dipole by the energy of the photon's mode, which can be approximated by the volume integral of its energy density, yields Eq. 1.8

$$\left(\frac{\Delta\hbar\omega}{\hbar\omega}\right)_i = \frac{-\alpha|\underline{\vec{E}}(\vec{r}_i)^2|}{2\int\epsilon_s|\underline{\vec{E}}(\vec{r})|^2 dV} \tag{1.8}$$

where α is the polarizability of the molecule. Thus, the resonance shift for one photon induced by one molecule can be expressed as shown in Eq. 1.9.

$$(\Delta\omega)_i = \frac{-\alpha|\underline{\vec{E}}(\vec{r}_i)^2|}{2\int\epsilon_s|\underline{\vec{E}}(\vec{r})|^2 dV}\omega \tag{1.9}$$

Expanding this approach for more molecules attached to the resonator and for generalizing one single photon resonant state to a general TE mode, the resonance shift is given by Eq. 1.10 [25].

$$\Delta\omega = \frac{-\alpha N}{4\pi\epsilon_0 R^3(n_s^2 - n_m^2)}\omega \tag{1.10}$$

Here, N represents the number of molecules attached to the meta-atom, R the radius of the spherical meta-atom, n_s the refractive index of a single meta-atom and n_m the refractive index of the environment around the metasurface. This expression can also be written in terms of the wavelength (Eq. 1.11) [26].

$$\Delta\lambda = \frac{\alpha N}{4\pi\epsilon_0 R^3(n_s^2 - n_m^2)}\lambda \tag{1.11}$$

Due to the relationship between the polarizability α of a molecule and its refractive index n given by the Lorentz-Lorenz equation (Eq. 1.12) [27]

$$\alpha = \frac{3M(n^2-1)}{4\pi\rho(n^2+2)N_A} \tag{1.12}$$

where M and ρ are the molecular weight and molar volume of the molecule and N_A the Avogadro constant, the resonance shift $\Delta\lambda$ (Eq. 1.11) can be expressed regarding the refractive index n of the molecule (Eq. 1.13).

$$\Delta\lambda = \frac{3M(n^2-1)N}{16\pi^2\epsilon_0\rho R^3 N_A(n^2+2)(n_s^2-n_m^2)}\lambda \tag{1.13}$$

It is important to note that the linear relationship between the number of molecules N (i.e., the concentration) and the resonance shift indicated by this expression is only applicable within a certain range. If the meta-atom becomes saturated with analyte molecules, no further resonance shift will be induced, regardless of the external concentration. Additionally, due to limitations in measurement sensitivity, it is impossible to detect every single molecule-induced resonance shift. Therefore, the relationship between concentration and resonance shift is assumed to be nonlinear at very high and very low analyte concentrations, but remains linear in between. As a result, the response curve of the metasurface-based concentration sensor may resemble the graph depicted in Figure 1.5, aligning with existing literature [24].

Since the metasurface sensor is typically used in situations where reliable measurement sensitivity and the detection of low analyte concentrations are desired, it is supposed to operate within the linear regime. As a result, the metasurface sensor retains the main benefit of dispersion spectroscopy, namely the linear relationship between the analyte concentration and the signal that encodes this information over a wide dynamic range.

It is important to acknowledge that Eq. 1.13 is a simplified representation of the impact of a single molecule on
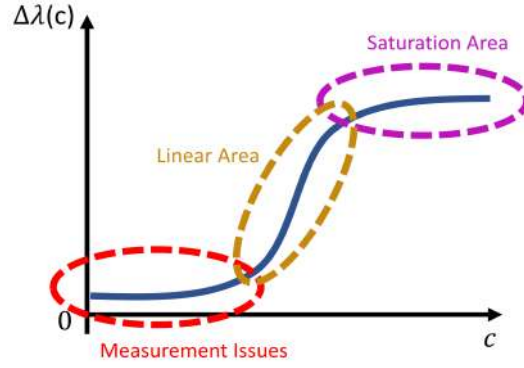
Figure 1.5: Assumed response curve of metasurface-based concentration sensor. The sensor exhibits nonlinear behavior for very low and very high concentrations. In between, the sensor shows linearity.

a meta-atom. In reality, meta-atoms are usually not spherical in shape and may exhibit optical heterogeneity, making the description of resonance shifts significantly more complex. Furthermore, the modes normally do not only occur inside the meta-atoms - as assumed in the above derivation - but especially between the meta-atoms outside the metasurface in the surrounding environment. Nevertheless, the previously discussed concept of interaction between the modes of the meta-atoms and the analyte molecules is also valid for these outer modes, thus providing valuable insights into the process causing the resonance shift. In order to accurately depict the resonance behavior of meta-atoms, numerical methods are commonly employed in practice due to the intricacy of the various processes taking place inside and around the metasurface concurrently.

In conclusion, the resonance shift in the metasurface's frequency response depends on the change of refractive index of the surrounding environment. When the refractive index is increased, the resonance wavelength undergoes a red shift of $\Delta\lambda$, i.e., the resonance wavelength of the metasurface increases by $\Delta\lambda$ or the resonance frequency decreases by $\Delta\omega$ (Fig 1.6). A blue shift represents the opposite behavior.

(a) Phase response

(b) Amplitude response
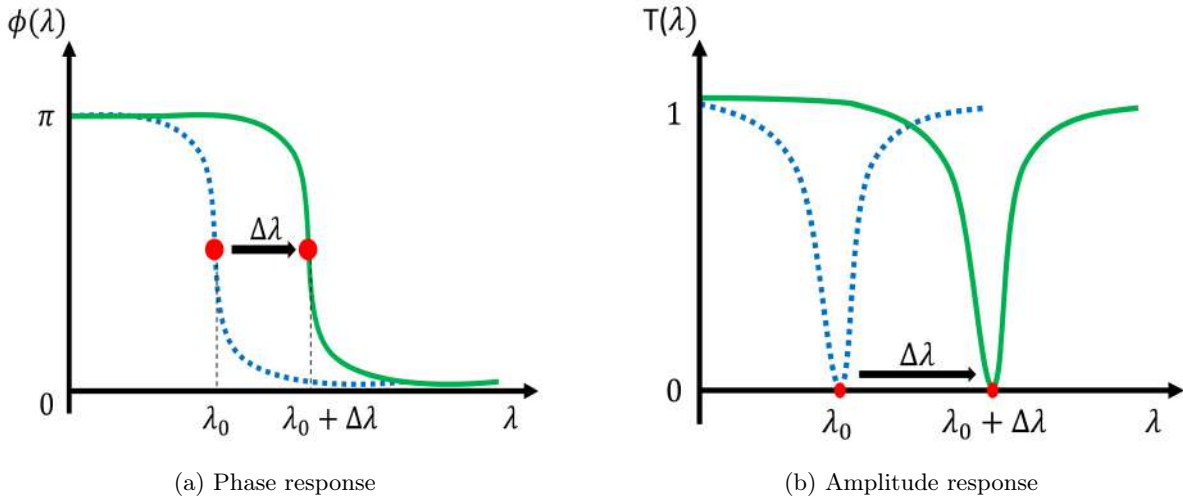
Figure 1.6: Influence of resonance shift to phase (a) and amplitude (b) response.

The primary objective of metasurface sensors is to accurately determine the concentration of the analyte. This is achieved by measuring the resonance shift, which is directly proportional to the concentration. The underlying principle of this measurement is elaborated in section 2.2.1, where the resonance shift in the phase response is utilized as a key indicator.

1.2.3.2   The Q-Factor as Key Parameter

As explained before, the interaction of the metasurface with the analyte results from the modes of the meta-atoms which are resonantly enhanced near fields induced by the incident electromagnetic wave. This

enhancement arises from the gradual accumulation of energy from multiple cycles of incident electromagnetic waves [28] and is basically described by the Q-factor. The Q-factor is a crucial parameter for all types of resonators and is often used to assess their performance. It is the measure of the energy stored in the resonator relative to the energy lost in radiation, i.e., scattering, or thermal energy [20]. Thus, a high Q-factor implies lower energy loss, resulting in slower damping of oscillations and longer ringing times. So, while the Q-factor of an element relates to the losses, which in turn correspond to the damping coefficient in second order systems, this links directly in to the bandwidth of a resonator with respect to its centre frequency. Consequently, the Q-factor is mathematically defined by Eq. 1.14

$$Q = 2\pi \frac{E_{stored}}{E_{lost\ per\ cycle}} = \frac{\omega_0}{\Delta\omega} \tag{1.14}$$

where ω_0 represents the resonance frequency of the resonator and $\Delta\omega$ the resonance width, also called FWHM. This width indicates the bandwidth over which the resonator's response is at or above half its maximal absolute value. Consequently, the higher the Q-factor of the metasurface the narrower its amplitude and phase response as depicted by Fig. 1.7.

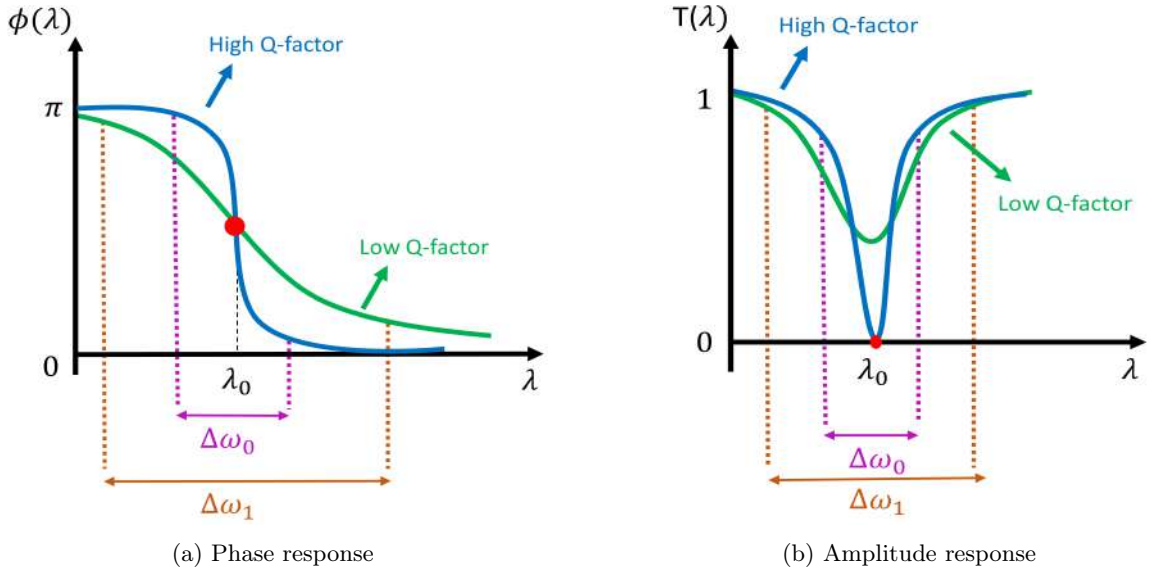(a) Phase response (b) Amplitude response

Figure 1.7: Phase (a) and amplitude (b) response for different Q-factors and corresponding FWHM.

In the context of using metasurfaces for sensing applications, the Q-factor is a critical parameter. When dealing with low concentrations of analyte molecules, conventional spectroscopy techniques may not produce a detectable output signal since the light properties would not be changed sufficiently by the molecules (Fig. 1.8a). The system is not sensitive enough to the analyte molecules. The same applies when analyte molecules exhibit sluggish resonance behavior, indicated by a small refractive index, due to a strong attraction between positive and negative charges in the molecule corresponding to a high damping coefficient (see Eq. A.47 and Eq. A.49) as depicted in Fig. 1.8a. By implementing metasurfaces, this issue can be effectively compensated by enhancing their Q-factor.

A higher Q-factor corresponds to less absorption losses and stronger electric fields of the resonators, which increases the interaction between the meta-atoms' electric field and the analyte. The stronger the interaction of the fields with the analyte, the more significant the changes of the transmitted light's properties, such as amplitude or phase, compared to the incident field and the easier and more precise the detection of these changes which comprehend the encoded information about the analyte. In other words, a stronger interaction of the electric fields of the meta-atoms, resulting from a higher Q-factor, leads to a higher sensitivity of the metasurface system to the analyte. This sensitivity is indicated by a higher absolute slope value of the amplitude and phase response of the metasurface, as well as a smaller FWHM, as shown in Fig. 1.7. This means, in the context of analytes with low refractive indices or small concentrations, enhancing the metasurface's Q-factor can effectively counteract the minor resonance shift, which is proportional to the analyte concentration or refractive index change as showcased by Eq. 1.13. The higher Q-factor leads to

more substantial phase changes and ultimately yields heightened sensor sensitivity as depicted in Fig. 1.8b. Thereby it is shown that a connection between the resonance shift itself, which directly encodes the analyte information as explained in section 1.2.3.1, and the enhanced near fields resulting from a higher Q-factor, actually exists although it is not immediately apparent in Eq. 1.13. This connection is established because the resonance shift is encoded in the amplitude or phase of the output wave which are influenced by the interaction of the enhanced near fields with the analyte molecules or the Q-factor, respectively.
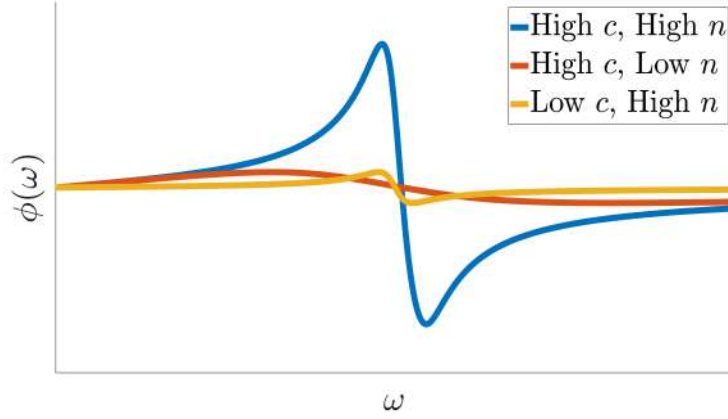
(a) Comparing the phase response curve of media with high concentration c and high refractive index n, high concentration c and low refractive index n, and low concentration c and high refractive index n. It is evident that a low refractive index or a low concentration only induce small phase changes in light passing through the medium, limiting common dispersion spectroscopy for low concentrations and unresponsive media.
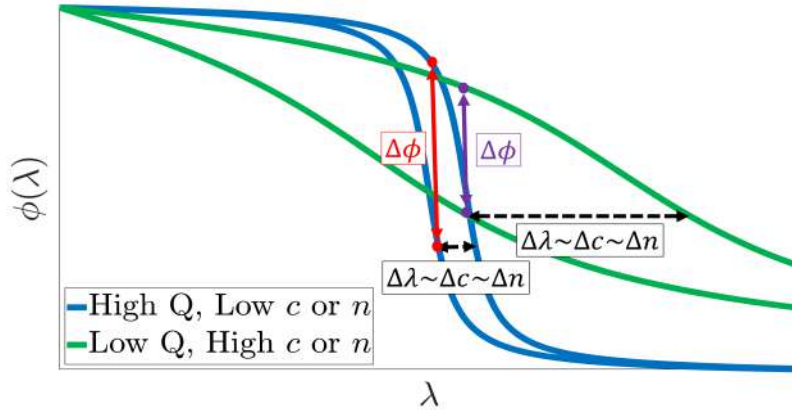
(b) Analyzing impact of Q-factor on metasurface sensitivity. It is demonstrated how a higher Q-factor effectively counteracts resonance shifts caused by slight concentration or refractive index changes. The central idea revolves around encoding the resonance shift within the detected phase signal which can be influence by the Q-factor. Although the sensor with a low Q-factor (green) shows a larger resonance shift due to higher concentration or refractive index changes, the sensor with a higher Q-factor (blue) yields a more substantial detectable signal, $\Delta\phi$. This is because the higher Q-factor compensates for the slight resonance shift, leading to an amplified detectable signal $\Delta\phi$.

Figure 1.8: Revealing the capability of metasurfaces to overcome sensitivity limitations in conventional dispersion spectroscopy.

Looking at it from another angle, the initial dispersion spectroscopy system, where an analyte is illuminated and the transmitted light is detected and analyzed, is extended by incorporating a metasurface into the transduction mechanism. The metasurface descriptively represents an amplifier increasing the Q-factor of the whole system by "accumulating" the incident electromagnetic field through its meta-atoms and thus enhancing the interaction of the light with the analyte, resulting in an intensified modulation of the incident light and thus a better detection of the encoded analyte information. Referring to spectroscopy techniques, this entire concept can be regarded as a variation of CEAS.

In conclusion, the capability to confine light into nanoscale electromagnetic hotspots and significantly en-

hance the field strength is the key process of metasurface-based sensors making them superior to the initial concept of dispersion spectroscopy and enabling them to excel in applications such as single-molecule spectroscopy [21].

The interaction of the incident electromagnetic field and the analyte molecules also explains the advantages and disadvantages of plasmonic and dielectric metasurfaces in terms of optical sensing applications. Due to the higher Q-factor of dielectric metasurfaces and therefore more enhanced electric fields of the resonators, dielectric metasurfaces are basically more sensitive to analytes than plasmonic ones. Furthermore, since the modes of dielectric nanoparticles span larger volumes, dielectric metasurfaces are more sensitive to analytes, which are further away from the surface, than plasmonic metasurfaces. This is often the case for refractive index sensing [20] which is applied in this project. However, if the analyte molecules are very close to the surface, plasmonic metasurfaces provide higher sensitivity due to their higher electric field density (Fig. 1.9).

As a result, since dielectric metasurfaces provide lower losses and consequently a higher Q-factor, which enhances the overall sensitivity of the optical sensing system and leads to a higher LOD, and higher sensitivity for analyte molecules that are located further away from the surface, a dielectric metasurface is employed in this project.
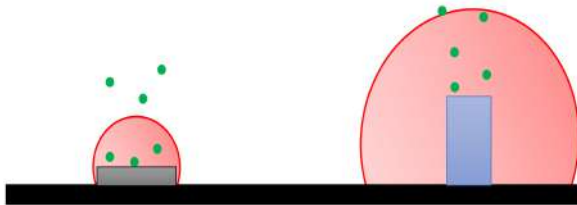

Figure 1.9: Enhanced electric field for a plasmonic (left) and dielectric (right) meta-atom. The green dots represent the analyte molecules. Dielectric meta-atoms span larger volumes, leading to lower electric field density.

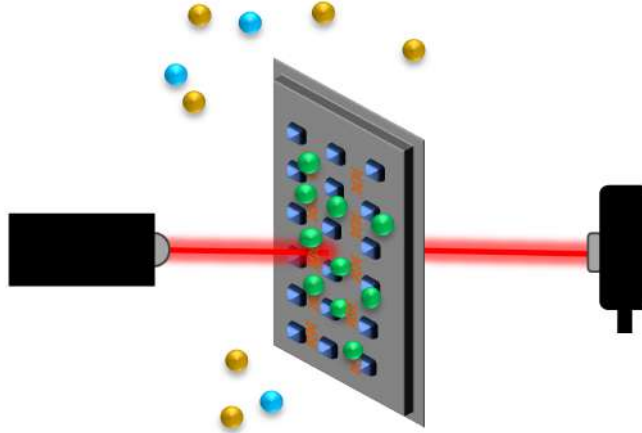### 1.2.3.3   Surface Functionalization and Design Flexibility

Primary hurdles faced by optical sensors utilizing dispersion spectroscopy techniques are their sensitivity, selectivity, and design flexibility as evaluated in section 1.1.3. This means, the induced phase shift caused by the medium is extremely small and diminishes with lower concentrations, making it challenging to measure very low concentrations accurately. Additionally, since the analyte molecules move randomly in free space, their interaction with the laser beam is unpredictable, and other molecules that do not represent the analyte can also induce phase shifts, leading to distorted analyte analysis. This is illustrated in Fig. 1.10a, where the green molecules symbolize the analyte, while the blue and golden molecules represent unrelated species. Furthermore, the resonance wavelength of the analyte is fixed, determining the laser wavelength used, which limits flexibility in sensor design.

Regardless of the metasurface type and measurement technique, all metasurface-based sensors basically suffer from almost the same key limitations. First, they are sensitive only to analytes that are in sufficiently close proximity to the surface to interact with the enhanced near fields of the resonators. Second, the transduction mechanism is intrinsically nonspecific, i.e. any analyte that changes the local environment - and therefore its refractive index - induces the resonance wavelength of the metasurface and thus the response of the sensor. However, both limitations of optical metasurface sensors, namely sensitivity and selectivity, can be overcome through functionalization and texturing of the metasurface to suit the desired application. This technique usually relies on the addition of target-binding molecules (bioreceptors), such as proteins, nucleic acids, antibodies or viruses, to the sensing surface, in order to bond the analyte of interest by establishing specific forces (e.g., van der Waals or electrostatic forces) [20]. Fig. 1.10b illustrates this process, showcasing nucleic acids as the bioreceptors in this particular case.

This analyte capturing enhances the sensitivity of the sensor in two ways. One the one hand, the analyte molecules (green) are basically closer attached to the surface. The nearer the molecules to the surface the stronger the interaction of the electric fields with the analyte and the more significant the change of the incident light's properties. On the other hand, the analyte concentration automatically increases which also increases the resonance shift, according to Eq. 1.13.
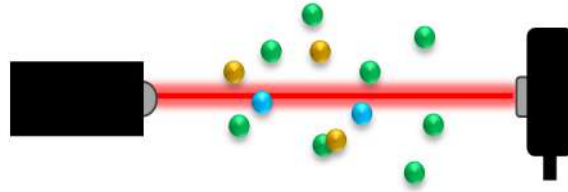
(a) Common Dispersion Spectroscopy. Many techniques suffer from sensitivity, selectivity and design flexibility.

(b) Surface functionalization of metasurface. Nucleid acids are utilized as bioreceptors to bind the analyte molecules (depicted in green) onto the metasurface. This strategic approach significantly enhances the sensitivity and selectivity of the metasurface-based sensor as the specific bonding between the nucleic acids and the analyte ensures that the measurement signal remains unaffected by other molecules (blue and golden) that are not bonded by these nucleic acids.

Figure 1.10: Concept and advantages of surface functionalization.

The selectivity of the metasurface sensor is improved since the bioreceptors only bond the analyte of interest. This means, the resonance shift of the metasurface is mainly induced by the desired analyte (green) and rarely distorted by other analyte molecules (blue and golden).

Furthermore, as mentioned earlier in this chapter, the behavior of metasurfaces, including their resonance frequency, is primarily determined by the geometry and material of the meta-atoms. This implies that each metasurface can be individually fabricated to cater to specific application requirements, offering a high degree of flexibility and freedom in sensor design. For instance, the metasurface can be adapted to the given light source, as the frequency response can be chosen from fabrication whereas conventional dispersion spectroscopy techniques depend on the resonance frequency of the analyte molecules. These significant advantages highlight the potential value in incorporating metasurfaces into existing and well-established dispersion spectroscopy techniques. The enhanced flexibility, sensitivity, and selectivity offered by metasurfaces make them a promising avenue for improving and advancing current sensor technologies, especially for detecting analyte concentrations as subsequently explained.

1.3 Applications

The use of metasurface-based sensors in analyzing biological parameters can upgrade the quality and performance of cost-effective, portable sensor devices which play salient roles in improving the standards of living and healthcare services.

As previously examined, unlike many common sensor techniques that are limited to specific analytes or concentrations, metasurface-based sensors have the remarkable ability to detect almost any desired analyte with high accuracy, even at low concentrations, which is attributed to the enhanced near fields generated by the resonance behavior of the meta-atoms, the enhanced selectivity due to surface functionalization and their label-free detection [20] capability. Furthermore, the rapid performance of metasurface-based sensors, facilitated by their label-free and multiplexed detection capabilities [29], contributes to their immense po-

tential in detecting life-threatening diseases, such as the Coronavirus [24,30], as well as pathogens in general since current detection methods often involve lengthy detection times, which can lead to severe consequences such as pathogen outbreaks, diseases, and even loss of life [31]. Additionally, due to their compact size, these sensors have significant promise as POC devices, offering invaluable benefits to the medical and environmental sensing sector as well as the food industry.

In conclusion, metasurface-based sensors, with their real-time, label-free, and multiplexed detection capabilities, combined with high sensitivity, selectivity, and compact design, enabling them being used as POC devices, can fulfill the key requirements mentioned in section 1.1.4 and thus hold significant potential as high-performing biosensors in various sectors. They offer particular promise in the detection of pathogens and cancer cells, making them invaluable tools for improving healthcare and environmental sensing as further discussed below.

1.3.1 Pathogen Detection

Across all segments, the market is generating a huge need especially for pathogen detecting biosensors which are sensors being capable of detecting bacteria, viruses or microorganisms that can cause disease or illness from minor to severe and life-threatening conditions. These organisms can be spread through various means, such as through the air, contaminated food or water, or contact with infected surfaces or animals. Current pathogen-diagnostic methods often suffer from inefficiency and slowness, particularly in resource-limited regions where suitable solutions are in high demand. Therefore, there is an urgent need to develop faster and more accurate diagnostic methods that do not rely on complicated and expensive assay steps, as it can be observed in the fields of food and environmental monitoring as well as medical diagnostics.

Ensuring the detection of foodborne and waterborne pathogens plays a crucial role in safeguarding public health. These pathogens can stem from various sources and matrices and commonly infect individuals through the consumption of contaminated food and water.

According to the World Health Organization (WHO), approximately 420,000 people lose their lives each year due to tainted food, with children under the age of 5 accounting for around 30% of the fatalities [32]. While there have been notable advancements in the detection techniques for foodborne pathogens in recent years, certain challenges, particularly regarding detection speed, persist. Culture-based detection, a commonly employed method for identifying foodborne bacteria, fungi, and viruses, still requires several days to yield results. Specifically, the detection of fungi can take up to 7 days, while the integration of Polymerase Chain Reaction (PCR) can reduce the detection time for bacteria to 3 days [33], which is still very long compared to detection times of metasurface sensors, which typically amount to just a few seconds or minutes.

Also the contamination of water by waterborne pathogens and the associated diseases pose significant concerns worldwide. Unsafe drinking water, primarily linked to diarrheal diseases, contributes significantly to the overall disease burden. In 2019, diarrheal diseases caused approximately 1.5 million deaths worldwide, with children under 5 years of age being particularly vulnerable, as it is the fourth-leading cause of death among this age group worldwide [34]. The importance of improving water quality, as recognized by the United Nations in the Millennium Development Goals [35], emphasizes the need for more efficient water quality control methods. Traditional approaches involve collecting water samples from the network's endpoint and conducting time-consuming laboratory analysis, often relying on cell cultures or other slow methods. These methods can take days to weeks to produce results, depending on the pathogen's ability to grow visible colonies [36]. Furthermore, the process involves labor-intensive steps such as culture medium preparation, inoculation, and colony counting, limiting its accessibility to trained personnel.

The field of medical diagnostics faces similar challenges. Recent outbreaks of diseases like Ebola and Covid-19, along with the rise of bacterial resistance to antibiotics, emphasize the need for new and effective diagnostic tools that enable early and rapid pathogen detection. Classical microbiological and immunoserological methods, as well as modern diagnostic platforms like ELISA, chemiluminescence analysis, PCR, flow cytometry, and mass spectrometry (MALDI), are widely adopted in centralized laboratories of medical hospitals and centers for accurate verification of infectious agents [37]. However, these diagnostic tools often necessitate costly equipment and highly skilled personnel, making them not always accessible to small hospitals, particularly in resource-constrained environments with decentralized medical infrastructures [37]. Moreover, the lengthy testing times associated with these methods hinder the timely delivery of evidence-based care, thereby contributing to the inappropriate use of antibiotics. This misuse of antibiotics is a significant factor driving antimicrobial resistance, which further complicates the effective management of illnesses [37]. The limitations of standard diagnostics are particularly pronounced in non-traditional and resource-poor health-

care settings, emphasizing the need for alternative approaches.

This means, in comparison to the rapid response offered by metasurface sensors, which provide real-time detection capabilities, the current detection times for pathogens in the fields of food/environmental monitoring and medical diagnostics remain notably lengthy and inefficient. Moreover, the existing methods require trained personnel for detection and involve sending analyte samples to specific laboratories, resulting in measurement results that can take hours or even days to obtain. This process is not only costly but also carries the risk of harm to humans if pathogens are not promptly detected. Consequently, the application of metasurface biosensors presents a viable solution to these challenges. By offering real-time detection capabilities, metasurface biosensors significantly reduce the time required for pathogen detection. Additionally, these sensors can be utilized by individuals without extensive technical backgrounds, making them accessible for use in any location. As a result, the adoption of metasurface biosensors addresses the drawbacks of the current detection methods, contributing to a decrease in diseases and illnesses. This is particularly advantageous in less developed countries where limited infrastructure and financial resources make it challenging to conduct extensive culture-based measurements and train highly educated personnel.

1.3.2 Cancer Cell Detection

Metasurface biosensors also have the potential to make significant contributions to medical diagnostics in the field of cancer cell detection.

In 2020, the WHO reported a global cancer burden estimate of 19.3 million new cases and 10 million deaths [38]. Different types of cancers exhibit distinct molecular characteristics, and their optimal treatments can vary significantly. Therefore, accurate detection and differentiation of these cancers at an early stage are crucial for timely clinical decision-making and medical interventions since early identification of cancer, as highlighted by the WHO, enhances the likelihood of treatment response, resulting in improved survival rates with reduced morbidity and lower treatment costs [39].

Metasurface-based sensors offer significant advantages in the detection of cancer cells as they operate in the terahertz region with high sensitivity. Studies have indicated that tumors possess higher water content compared to healthy cells [40, 41]. As water exhibits strong absorption throughout the terahertz range, the refractive index of cancer cells is expected to be higher than that of normal tissue, as demonstrated by Pickwell et al. in the case of basal cell carcinoma (i.e., skin cancer) [42]. Consequently, the response of terahertz radiation to cancer cells differs from that of unaffected cells, making it possible to detect these differences using a highly sensitive metasurface biosensor. Additionally, cancer cells can be indirectly detected by targeting their associated tumor markers, which typically represent protein molecules. As the resonance region of protein molecules is also located in the THz region [43], optical metasurface sensors offer not only high sensitivity towards cancer cells but also towards their marker molecules. This dual capability enables the detection of both cancer cells and their corresponding biomarkers, providing a comprehensive approach for cancer diagnostics.

1.4 Research Goal

The primary objective of the entire project is to develop a highly sensitive metasurface-based sensor, capable of detecting analytes at extremely low concentrations. To achieve this, the metasurface-based sensor leverages the power of dispersion spectroscopy and harnesses the metasurface's capability to shift its frequency response in response to changes in the environment's refractive index. This innovative technique is anticipated to outperform traditional dispersion spectroscopy methods, particularly due to the enhanced light-matter interaction resulting from the metasurface's augmented near-fields.

This bachelor project, in particular, marks the initial phase of the larger project, primarily focused on establishing the experimental setup, conducting its characterization, and demonstrating its performance through both experimental and simulation-based approaches. Due to the unavailability of the desired metasurface initially, the experimental setup employs an HCN vapor cell as a substitute, enabling meaningful noise and real signal measurements. The real signal measurements are then compared to simulations to ensure alignment with theoretical expectations. The project places particular emphasis on maximizing the SNR by reducing noise sources and enhancing the measurement signal's amplitude.

Furthermore, to gain insights into the behavior of the system when combined with an actual metasurface, an older metasurface with a broader bandwidth, although not meeting the exact specifications, is digitally integrated into the experimental setup. This investigation allows for assessing the combined performance of

the system and the metasurface, aiming to predict the minimal detectable resonance shift of the metasurface and identifying the optimal AOM frequency shift. The overarching objective is to achieve a significant increase in the SNR to attain a minimal detectable resonance shift below 1nm.

These preliminary investigations lay a strong foundation for the subsequent phases of the project, where the desired metasurface will be incorporated, and further experimentation and analysis will be conducted to achieve the ultimate goal.

Chapter 2

Dispersion Interferometry

2.1 Baseline Model

To delve into the principle of dispersion spectroscopy mentioned in section 1.1.2 in greater detail, a previously investigated interferometer setup considered as a modified Sagnac interferometer [44] is employed. As the objective of this project is to expand this concept and apply it to metasurfaces, it requires a comprehensive examination of its underlying principles and mechanisms. Furthermore, this model offers significant immunity to frequency and environmental noise, making it a suitable reference for the metasurface setup in this project. The fundamental concept of the modified Sagnac interferometer is illustrated in Figure 2.1. This technique is referred to as homodyne spectroscopy, as the demodulated signal from the photodetector is a DC signal.
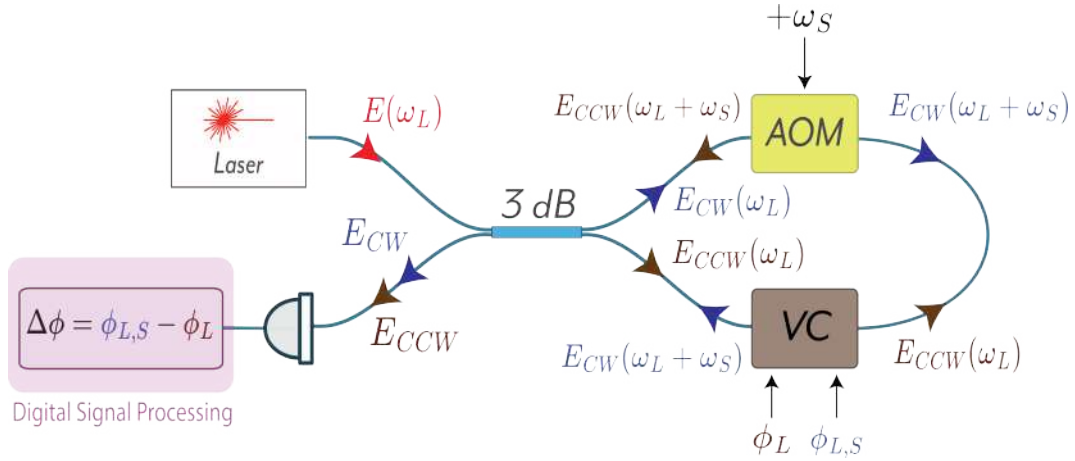

Figure 2.1: Concept of a modified Sagnac interferometer, serving as baseline model.

A tunable laser source emits a single optical tone of frequency ω_L. The tone is split into clockwise (CW) and counterclockwise (CCW) paths. The electromagnetic wave E_{CW} with frequency ω_L in the CW path is frequency shifted by the AOM with a shift frequency ω_S. Consequently, the wave enters the vapor cell (VC), which contains a HCN gas as analyte, with a frequency $\omega_L + \omega_S$. The electromagnetic wave E_{CCW} with frequency ω_L in the CCW path enters the vapor cell unshifted and is only frequency shifted by the AOM after it exits the vapor cell. The vapor cell is thus transited by two optical frequencies ω_L (E_{CCW}) and $\omega_L + \omega_S$ (E_{CW}), each experiencing a different phase shift (see section A.2.3.2) corresponding to the dispersion profile of the HCN gas equally shaped as in Fig. A.6. The two fields exit the loop, however, at the same frequency. Hence, when the two counterpropagating fields are recombined at the coupler and observed with a photodetector, their interference results in signal whose phase $\Delta\phi$ represents the difference of both phase shifts ϕ_L and $\phi_{L,S}$ of the signals E_{CW} and E_{CCW} and is therefore equal to the differential dispersion experienced by the two fields. By sweeping the laser frequency over a specific spectral range, the differential phase shift of the two fields can be extracted for multiple frequencies, allowing the dispersion profile of the analyte to be reconstructed. Mathematically, it is described as follows.

According to Eq. A.59, the waves in the CW and CCW path experience different phase shifts due to their

different frequencies given by Eq. 2.1 and Eq. 2.2

$$\phi_{CCW} = \phi_L = \frac{\omega_L l_{cell}}{c_0}(n(\omega_L) - 1) \tag{2.1}$$

$$\phi_{CW} = \phi_{L,S} = \frac{(\omega_L + \omega_S)l_{cell}}{c_0}(n(\omega_L + \omega_S) - 1) \tag{2.2}$$

where ω_L is the frequency of the laser, ω_S the frequency shift induced by the AOM, n the frequency dependent refractive index of the analyte sample in the vapor cell and l_{cell} the length of the vapor cell. According to Eq. A.52, the CCW and CW fields, after transiting the vapor cell and the AOM, can be expressed by Eq. 2.3 and Eq. 2.4. Since both waves propagate along the same distance l_{cell}, the z-dependency of the waves can be neglected.

$$\underline{E}_{CCW}(t) = E_{CCW}e^{\frac{1}{2}\alpha(\omega_L)}e^{j((\omega_L+\omega_S)t+\phi_{CCW}(z))} = A_1 e^{j\phi_1(t)} \tag{2.3}$$

$$\underline{E}_{CW}(t) = E_{CW}e^{\frac{1}{2}\alpha(\omega_L+\omega_S)}e^{j((\omega_L+\omega_S)t+\phi_{CW}(z))} = A_2 e^{j\phi_2(t)} \tag{2.4}$$

where $\alpha(\omega)$ represents the frequency dependent absorption coefficients experienced by the fields when transiting the VC. After being recombined by the coupler, both fields are detected as superposition by the photodetector. Since both signals have the same frequency, the detected signal is the intensity of their interference. The general description of the intensity for the superposition of two electromagnetic waves is expressed by Eq. 2.5.

$$
\begin{aligned}
I(t) &= |\underline{E}_{CCW}(t) + \underline{E}_{CW}(t)|^2 = |A_1 e^{j\phi_1(t)} + A_2 e^{j\phi_2(t)}|^2 \\
&= (A_1 e^{j\phi_1(t)} + A_2 e^{j\phi_2(z,t)})(A_1 e^{-j\phi_1(t)} + A_2 e^{-j\phi_2(t)}) \\
&= A_1^2 + A_2^2 + A_1 A_2 e^{j\phi_1(t)}e^{-j\phi_2(t)} + A_1 A_2 e^{j\phi_2(t)}e^{-j\phi_1(t)} \\
&= A_1^2 + A_2^2 + A_1 A_2[e^{j(\phi_1(t)-\phi_2(t))} + e^{j(\phi_2(t)-\phi_1(t))}] \\
&= A_1^2 + A_2^2 + A_1 A_2[cos(\phi_1(t) - \phi_2(t)) + jsin(\phi_1(t) - \phi_2(t)) \\
&+ cos(\phi_1(t) - \phi_2(t)) - jsin(\phi_1(t) - \phi_2(t))] \\
&= A_1^2 + A_2^2 + 2A_1 A_2 cos(\phi_1(t) - \phi_2(t)) = A_1^2 + A_2^2 + 2A_1 A_2 cos(\Delta\phi(t))
\end{aligned} \tag{2.5}
$$

Since the electromagnetic waves in the Sagnac interferometer are emitted by the same laser and are assumed to be split equally by the coupler, it applies $A_1 = \frac{E}{2}e^{\frac{1}{2}\alpha(\omega_L)}$ and $A_2 = \frac{E}{2}e^{\frac{1}{2}\alpha(\omega_L+\omega_S)}$. Considering the general equation for the intensity of the superposition of two electromagnetic waves (Eq. 2.5), the measured intensity by the photodetector in the Sagnac interferometer is given by Eq. 2.6.

$$
\begin{aligned}
I(t) &= A_1^2 + A_2^2 + 2A_1 A_2 cos(\Delta\phi(t)) \\
&= \frac{E^2}{4}e^{\alpha(\omega_L)} + \frac{E^2}{4}e^{\alpha(\omega_L+\omega_S)} + 2\frac{E}{2}e^{\frac{1}{2}\alpha(\omega_L)}\frac{E}{2}e^{\frac{1}{2}\alpha(\omega_L+\omega_S)}cos((\omega_L + \omega_S)t \\
&+ \phi_{CCW} - (\omega_L + \omega_S)t - \phi_{CW}) \\
&= \frac{E^2}{4}e^{\alpha(\omega_L)} + \frac{E^2}{4}e^{\alpha(\omega_L+\omega_S)} + \frac{E^2}{2}e^{\frac{1}{2}(\alpha(\omega_L)+\alpha(\omega_L+\omega_S))}cos(\phi_{CCW} - \phi_{CW}) \\
&= \frac{E^2}{4}e^{\alpha(\omega_L)} + \frac{E^2}{4}e^{\alpha(\omega_L+\omega_S)} + \frac{E^2}{2}e^{\frac{1}{2}(\alpha(\omega_L)+\alpha(\omega_L+\omega_S))}cos(\phi_{CW} - \phi_{CCW}) \\
&= \frac{E^2}{4}e^{\alpha(\omega_L)} + \frac{E^2}{4}e^{\alpha(\omega_L+\omega_S)} + \frac{E^2}{2}e^{\frac{1}{2}(\alpha(\omega_L)+\alpha(\omega_L+\omega_S))}cos(\Delta\phi)
\end{aligned} \tag{2.6}
$$

where $\Delta\phi$ is the differential phase between the frequency shifted and unshifted field. An expression for this phase shift can be derived by combining Eq. 2.1 and Eq. 2.2.

$$\Delta\phi = \phi_{CW} - \phi_{CCW} = \phi_{L,S} - \phi_L = \frac{l_{cell}}{c_0}((\omega_L + \omega_S)(n(\omega_L + \omega_S) - 1) - \omega_L(n(\omega_L) - 1)) = \frac{l_{cell}}{c_0}\Delta n \tag{2.7}$$

This phase shift is received from the detected signal using homodyne digital interferometry, which extracts the phase of the cosine signal [44]. By sweeping the laser frequency ω_L within a particular spectral range, the resulting phase signal is expressed as Eq. 2.7 but becomes a function of ω, as indicated by Eq. 2.8.

$$\Delta\phi(\omega) = \frac{l_{cell}}{c_0}\Delta n(\omega) \tag{2.8}$$

Due to the linear relationship between the refractive index $n(\omega)$ and the phase shift $\phi(\omega)$ as shown by Eq. A.59 as well as by Eq. 2.8, the differential phase shift $\Delta\phi(\omega)$ encodes the refractive profile of the analyte present in the vapor cell. Since the obtained information is the differential phase shift $\Delta\phi(\omega)$, the received profile is actually similar to the first derivative of the refractive index, with the similarity dependent on the magnitude of the frequency shift ω_S in comparison to the refractive index profile's bandwidth. The precise choice of the frequency shift is of particular significance as it has a direct influence on the magnitude of the differential phase signal and, consequently, the SNR. Notably, there even exists an optimal frequency shift. In the subsequent section, the utilization of the differential phase signal to determine the analyte concentration is briefly elucidated.

2.1.1 Principle of Concentration Measurement

The concentration measurement process utilizing the differential phase signal closely resembles the one described in section 1.1.2.1. Since the shift frequency is a fixed parameter, each measured signal is expected to exhibit the same shape. By comparing the differential signal of an analyte with unknown parameters to a reference signal with known parameters, the parameters of the unknown analyte, such as concentration, can be determined since the linear relationship between the measured signal and the concentration still holds true for the differential phase signal (see Fig. 2.2).
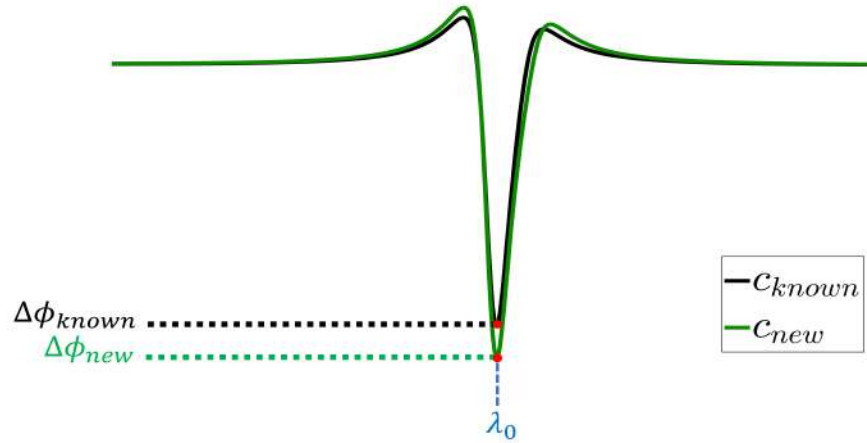

Figure 2.2: Concept of analyte concentration measurement for differential phase signal. The process is very similar to the one in section 1.1.2.1.

2.1.2 Evaluation

The primary advantage of various dispersion spectroscopy techniques, including the Sagnac interferometer approach, lies in the detection of a differential phase shift rather than a simple phase shift. As this differential phase shift arises from the phase differences between two electromagnetic waves propagating along the same path and originating from the same light source, both phase signals share the same noise characteristics, allowing for effective noise cancellation through their subtraction. Consequently, this noise cancellation mechanism results in a high SNR for dispersion spectroscopy.

Furthermore, the Sagnac approach exhibits similarities with the WMS technique discussed in section 1.1.1.1. Similar to WMS, the received signal in this approach does not directly represent the dispersion spectrum, but rather its first derivative. However, the key distinction lies in the method of operation: while WMS modulates the frequency of the signal, in this approach, the frequency is shifted. This implies that, in addition to its relatively simple setup architecture, another advantage of the Sagnac technique is the reduction of technical noise due to the frequency shift, as already explained in section 1.1.1.1.

Moreover, the utilization of the same propagation path for both electromagnetic waves in the Sagnac interferometer concept provides another significant advantage as this approach eliminates any distortions in

the phase signal that could result from differential path lengths. Consequently, the system achieves highly accurate and reliable measurement of the phase information, reaching an impressive noise limit of 0.8 $\frac{\mu \text{rad}}{\sqrt{\text{Hz}}}$. This noise limit surpasses what most HPSDS or CLaDS methods can achieve, which usually yield phase sensitivities on the order of a milliradian [44].

The combination of these advantages creates a strong basis for extending the Sagnac approach by incorporating a metasurface into the setup. The objective is to preserve all these benefits while synergizing them with the metasurface's heightened sensitivity, selectivity, and design flexibility. The result is a dispersion interferometry setup characterized by remarkably low noise and exceptionally high sensitivity, ultimately leading to a significant increase in the SNR.

2.2 Metasurface Dispersion Interferometry

The dispersion spectroscopy technique discussed in section 2.1 has demonstrated impressive capabilities in measuring analyte concentrations with high precision [2, 44]. However, as mentioned before, there is still room for improvement, particularly in terms of sensitivity, selectivity and flexibility, to detect even lower analyte concentrations. This project proposes a novel approach that leverages metasurfaces to enhance the performance of well-investigated dispersion spectroscopy techniques. The expected outcome of this project is a cutting-edge dispersion spectroscopy system with unprecedented sensitivity.

The main idea of this novel approach is to replace the conventional vapor cell used in section 2.1 by a metasurface. As explained in chapter 1.2, metasurfaces exhibit an extremely high sensitivity to analytes and therefore induce a much bigger change of refractive index than an analyte in a vapor cell does compared to vacuum. However, integrating the metasurface into the existing setup from section 2.1 requires transitioning to a unidirectional configuration, which necessitates additional modifications to avoid distortions of the differential phase that were previously automatically compensated by the Sagnac interferometer concept. Additionally, due to the unidirectional configuration, the beat note at the photodetector produces an AC signal, leading to the characterization of this approach as heterodyne spectroscopy. Fig. 2.3 depicts the basic concept of the project.

A tunable laser source emits a single optical tone of frequency ω_L. The tone is split equally into two paths. The electromagnetic wave E_2 with frequency ω_L in the upper path is frequency shifted by the AOM with a shift-frequency ω_S. The electromagnetic wave E_1, also with frequency ω_L, reaches the second coupler unshifted. The waves may experience distinct path lengths, leading to a phase difference $\Delta\phi_1$ as a potential outcome. Both electromagnetic waves are recombined at the second coupler, i.e., in the following two paths the superposition (or beat note) of the two initial waves E_1 and E_2, and thus of the two frequencies ω_L and $\omega_L + \omega_S$, propagates. In the lower path, the beat note enters the metasurface (MS) containing the analyte and both frequencies ω_L and $\omega_L + \omega_S$ are shifted in phase according to the phase response of the metasurface (Fig. 1.4a) and thus produce a differential phase shift $\Delta\phi_2$ which corresponds to the differential phase shift $\Delta\phi$ in Fig. 2.1. Consequently, the total phase difference between the waves E_1 and E_2 now consists of the phase difference $\Delta\phi_2$, which was induced by the metasurface, and the phase difference $\Delta\phi_1$ which results from potential different path lengths in the setup before the second coupler. This phase difference, representing the phase of the beat note formed by the two frequencies, is then detected by a photodetector and extracted using a phasemeter. In the upper path, the beat note remains unaffected by any phase shift and is directed towards the second photodetector to measure its phase, solely representing $\Delta\phi_1$, which is subsequently subtracted from the signal in the upper path resulting in the final measurement signal $\Delta\phi_2$. The incorporation of a second interferometer path and the subsequent subtraction of $\Delta\phi_1$ are of significant importance as they significantly reduce the SNR, as discussed in detail in section 5.1.

The distinction between the superposition observed in Fig. 2.1 and the one in the lower path of the current setup is that the latter results in a beat note due to the disparate frequencies of the interfering waves, whereas the former involved waves with the same frequency, resulting in an interference signal. Nevertheless, the detected intensity can still be described by Eq. 2.5 by replacing $\underline{E}_{CCW}$ and $\underline{E}_{CW}$ with $\underline{E}_1$ and $\underline{E}_2$. However, since both waves might propagate along different path distances, the z-dependency of the waves cannot be neglected. If reflected light is abandoned, both waves are described by Eq. 2.9 and Eq. 2.10.

$$\underline{E}_1(z,t) = E_1 e^{j(\omega_L t - k_1 z + \phi_L)} = E_1 e^{j\phi_1(z,t)} \tag{2.9}$$

$$\underline{E}_2(z,t) = E_2 e^{j((\omega_L + \omega_S)t - k_2 z + \phi_{L,S})} = E_2 e^{j\phi_2(z,t)} \tag{2.10}$$
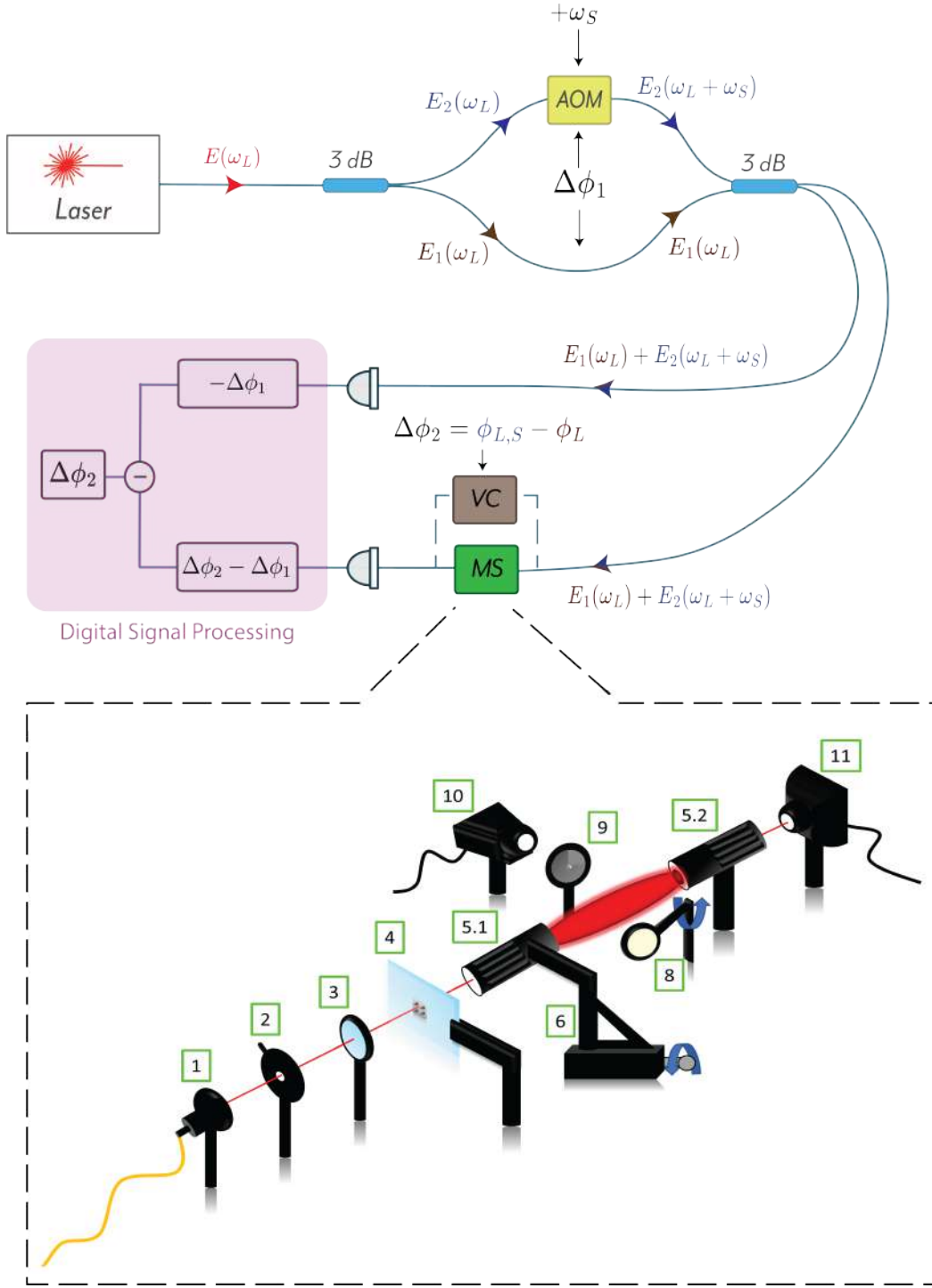
Figure 2.3: The experimental concept of the metasurface dispersion spectroscopy approach, representing a setup similar to the Sagnac interferometer discussed in section 2.1. It, however, adopts a unidirectional configuration, incorporating an additional path to mitigate signal distortions caused by variations in path lengths. The dashed box depicts the metasurface breadboard. This experimental setup is further explained in section 4.1.

The z-dependency can be removed if the phase difference $\Delta\phi_1$ is added to the absolute phase of one of the waves (Eq. 2.11).

$$\underline{E}_1(t) = E_1 e^{j(\omega_L t + \phi_L + \Delta\phi_1)} = E_1 e^{j\phi_1(t)} \tag{2.11}$$

According to Eq. 2.5, the measured intensity of the superposition of both waves, i.e., the beat note, in the lower path follows then as (Eq. 2.12):

$$
\begin{aligned}
I(t) =& |E_1 e^{j\phi_1(t)} + E_2 e^{j\phi_2(t)}|^2 = E_1^2 + E_2^2 + 2E_1 E_2 cos(\phi_1(t) - \phi_2(t)) = \\
& E_1^2 + E_2^2 + 2E_1 E_2 cos(\omega_L t + \phi_L + \Delta\phi_1 - (\omega_L + \omega_S)t - \phi_{L,S}) = \\
& E_1^2 + E_2^2 + 2E_1 E_2 cos(\omega_S t + \phi_{L,S} - \phi_L - \Delta\phi_1) = \\
& E_1^2 + E_2^2 + 2E_1 E_2 cos(\omega_S t + \Delta\phi_2 - \Delta\phi_1)
\end{aligned}
\tag{2.12}
$$

Since the initial laser beam E is equally split two times, it applies $E_1 = E_2 = \frac{E}{4}$ and thus:

$$
I(t) = \frac{E^2}{8} + \frac{E^2}{8} cos(\omega_S t + \Delta\phi_2 - \Delta\phi_1)
\tag{2.13}
$$

The detected intensity in the upper path results in the same way but without the differential phase shift $\Delta\phi_2$ which is induced by the metasurface in the upper path (Eq. 2.14).

$$
\begin{aligned}
I(t) =& E_1^2 + E_2^2 + 2E_1 E_2 cos(\phi_1(t) - \phi_2(t)) = \\
& E_1^2 + E_2^2 + 2E_1 E_2 cos(\omega_L t + \Delta\phi_1 - (\omega_L + \omega_S)t) = \\
& E_1^2 + E_2^2 + 2E_1 E_2 cos(\omega_S t - \Delta\phi_1) = \\
& \frac{E^2}{8} + \frac{E^2}{8} cos(\omega_S t - \Delta\phi_1)
\end{aligned}
\tag{2.14}
$$

If the phase signals of both beat notes are extracted by a phasemeter and finally subtracted, the desired differential phase shift $\Delta\phi_2$, which corresponds to the differential phase shift $\Delta\phi$ in Fig. 2.1, is received. (Eq. 2.15).

$$
\Delta\phi_2 - \Delta\phi_1 + \Delta\phi_1 = \Delta\phi_2
\tag{2.15}
$$

As the laser frequency ω_L scans across a specific spectral range, similar to the original configuration discussed in section 2.1, the differential phase shift $\Delta\phi_2$ becomes a function of frequency ω, approximating the first derivative of the metasurface's phase response. Compared to the narrower bandwidth of the analyte discussed in section 2.1, the metasurface's phase response has a much broader bandwidth, approximately 20 times wider, which results in the differential phase shift bearing a closer resemblance to the first derivative of the metasurface's phase response. This concept is visually illustrated in Fig. 2.4. The span of the frequency shift induced by the AOM is a critical factor in improving the SNR, as elaborated upon in section 4.3.5 and 5.3.2.
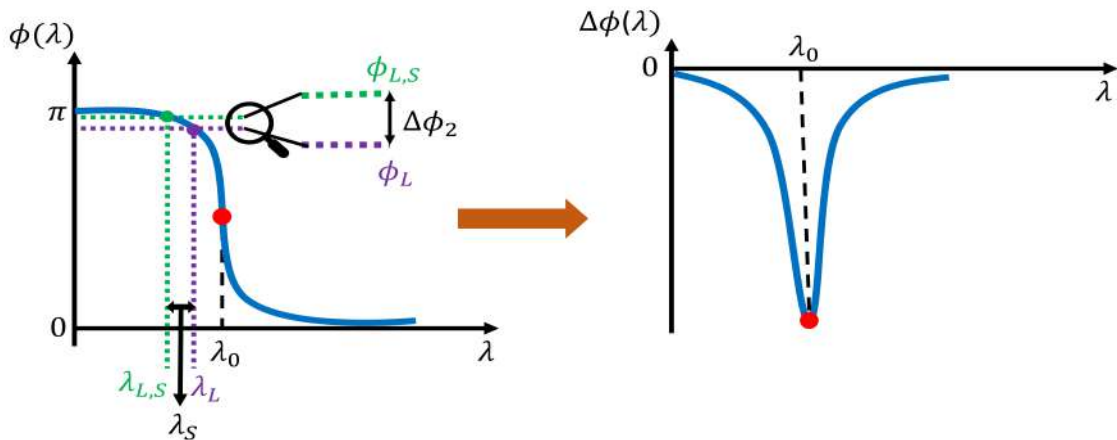

Figure 2.4: Obtaining the first derivative of the metasurface's phase response. The concept is the same as in section 2.1.

Although the project's objective is to investigate metasurface dispersion spectroscopy, the initial characterization of the setup is conducted using an HCN vapor cell in place of the metasurface as indicated in Fig. 2.3. This is done for two reasons. Firstly, the metasurface was not available at the project's outset. Secondly, the

behavior of the HCN vapor cell is well-documented and even calibrated by the National Institute of Standards and Technology (NIST) [45], which offers advantages for analyzing real signal measurement results. Once the setup characterization is finalized, the HCN vapor cell will be replaced by the metasurface.

2.2.1 Principle of Concentration Measurement

The process of determining the concentration of the analyte from the differential phase signal $\Delta\phi_2(\omega)$ is similar to the approach described in section 1.1.2.1 and 2.1. However, whereas in 1.1.2.1 and 2.1 the concentration was directly derived from the differential phase signal due to its linear relationship to the refractive index, in the metasurface setup an additional interim stage must be done. As oppose to the dispersion spectroscopy approach, the analyte concentration is not linearly proportional to the differential phase but to the resonance shift $\Delta\lambda$ (see Eq. 1.13). This means, the resonance shift $\Delta\lambda$ must first be determined from the differential phase signal, which is illustrated in Fig. 2.5. From the resonance shift $\Delta\lambda$, the corresponding analyte concentration can be determined.
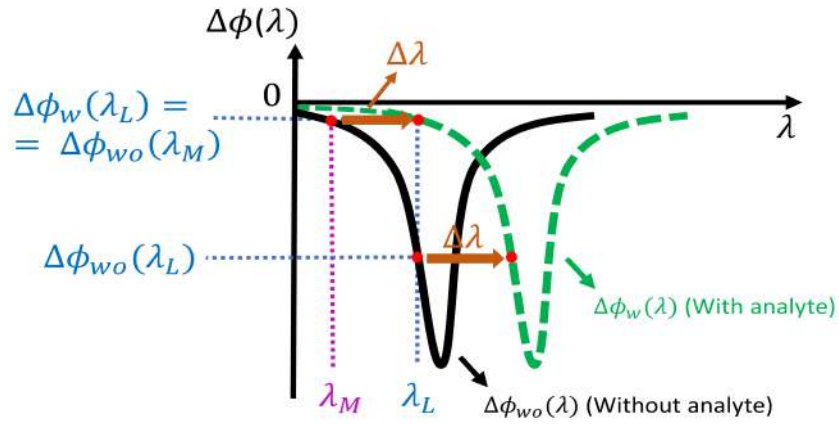

Figure 2.5: Determination of the analyte concentration. Once the reference signal, represented by the black curve, is obtained by measuring the differential phase without the analyte, the metasurface sample containing the analyte is illuminated with a specific wavelength, denoted as λ_L. This illumination leads to a corresponding value of the differential phase signal with the analyte, represented by the green curve. The value of the green curve aligns with a value of the black curve which is related to a wavelength denoted as λ_M. The disparity between these two wavelengths, λ_L and λ_M, represents the resonance shift.

For obtaining the resonance shift $\Delta\lambda$, the initial step is to measure the differential phase signal, denoted as $\Delta\phi_{wo}(\lambda)$, of the metasurface without the presence of the analyte. This is achieved by sweeping the laser wavelength across the bandwidth of the metasurface, resulting in the black curve as shown in Fig. 2.5. A wavelength λ_L is then determined as the point where the differential phase signal exhibits the highest absolute slope, indicated by the red dot in the middle of the black curve. Once the wavelength λ_L is determined, the metasurface is then illuminated with this wavelength in the presence of the analyte resulting in a measurement of the differential phase value denoted as $\Delta\phi_w(\lambda_L)$. As this value must also be part of the differential phase signal $\Delta\phi_{wo}(\lambda)$, it corresponds to a specific wavelength denoted as λ_M. Since the differential phase signal $\Delta\phi_{wo}(\lambda)$ is known, the corresponding wavelength value λ_M associated with the differential phase value $\Delta\phi_{wo}(\lambda_M) = \Delta\phi_w(\lambda_L)$ is also known. Thus, the resonance shift can be determined as $\Delta\lambda = \lambda_L - \lambda_M$.

The rationale behind selecting λ_L as the wavelength with the highest absolute slope is to maximize the disparity between the differential phase shifts $\Delta\phi_w(\lambda_L)$ and $\Delta\phi_{wo}(\lambda_L)$, leading to a more accurate determination of the resonance shift. In Fig. 2.5, the separation between the curves is exaggerated compared to actuality which underscores the necessity for a significant difference between the two differential phase signals to achieve precise detection of the resonance shift. Additionally, within the region of the steepest absolute slope, the differential phase signals exhibit approximately linear characteristics. As illustrated in section 2.1, a linear behavior proves to be advantageous in dispersion spectroscopy, offering significant benefits in some cases.

An alternative approach could involve determining the phase shift at a single frequency. However, this would result in the loss of the inherent advantage of dispersion spectroscopy mentioned in section 2.1.2, where the phase noise of both waves is subtracted, leading to a decrease in the SNR. The resonance shift detection would still follow the same procedure as for the differential phase signal, but without the benefit of noise cancellation, which would impact the accuracy and sensitivity of the measurement.

2.2.2 Evaluation

The setup involving the metasurface differs from the original configuration outlined in section 2.1 in four significant ways.

Firstly, the substitution of the vapor cell with the metasurface necessitates the opening of the Sagnac interferometer loop, as the metasurface does not permit bidirectional propagation. However, this alteration introduces a drawback as it can result in the waves following different path lengths, leading to an additional differential phase shift $\Delta\phi_1$. Since this phase shift is not induced by the analyte, it distorts the measurement of the analyte concentration. To address this issue, a second interferometer path is implemented which preserves the differential phase shift $\Delta\phi_1$ until the conclusion of the measurement process in order to nullify it by subtracting it from the detected phase of the beat note in the metasurface path. Therefore, this additional path serves as a substitute for the bidirectional configuration of the Sagnac interferometer, since the bidirectional architecture naturally cancels out any variations in path lengths experienced by the two electromagnetic waves, as they traverse the exact same propagation paths. The significant importance of this second path is showcased in section 5.1 by analyzing the noise behavior of the system. The reason for neglecting the potential additional phase difference due to different path lengths after the second coupler is the low frequency of the envelope of the beat note ω_S (as shown in Eq. 2.12) compared to the high frequency of the waves propagating before the coupler. According to Eq. A.59, the phase shift of a wave when propagating through a medium is linearly proportional to the wave's frequency ω. Since the shift frequency ω_S, representing the frequency of the envelope of the beat note, is very small (RF frequency) compared to the laser frequency ω_L (optical frequency), any phase difference resulting from different path lengths after the coupler would be much smaller than the differential phase $\Delta\phi_1$ or the phase shifts induced by the metasurface and can therefore be neglected. This theoretical assumption finds experimental validation in section 4.3.

Secondly, in the metasurface setup, the interfered waves have different frequencies, resulting in the generation of a beat note at the photodetector. This beat note exhibits a time-dependent absolute phase, in contrast to the Sagnac interferometer setup where the interfered waves have the same frequency and produce a time-independent phase signal. Thus, the dispersion spectroscopy approach from 2.1 is denoted as homodyne detection whereas the metasurface approach as heterodyne detection.

Thirdly, in the Sagnac interferometer setup, the interference occurs after the waves pass through the analyte, whereas in the metasurface setup, the interference takes place before the waves transit through the metasurface. However, this distinction is not problematic as both the analyte in the vapor cell and the metasurface only alter the phase and amplitude of the signals, making them linear and time-invariant (LTI) systems. Due to the linearity property of these systems, it does not matter whether the superposition of waves occurs before or after entering the system, as the overall effect on the signals remains the same.

Fourthly, while the Sagnac technique employed a spectroscopy method similar to WMS, the metasurface configuration can be regarded as a combination of WMS and CEAS (see section 1.2.3.2), but utilized specifically for dispersion spectroscopy. The AOM-induced frequency shift bears resemblance to the WMS technique, which decreases technical noise, while the implementation of meta-atoms as optical resonators acts as an optical cavity, amplifying the optical near fields and thereby increasing the interaction distance between the light and the analyte. As a consequence, the metasurface approach takes advantage of both the WMS as well as the CEAS technique, effectively enhancing the overall SNR by reducing the system noise and amplifying the measurement signal's amplitude.

Chapter 3

Setup Simulation

In order to evaluate the real signal measurement obtained from the HCN setup in section 5.2, it is essential to conduct a simulation of the differential phase signal acquired from the setup. By comparing the measurement with its theoretical representation, it can be ensured that the experimental results align with the expected outcomes.

The simulation entails four essential steps. Firstly, the transmission spectrum of the HCN gas is simulated, taking into consideration three broadening phenomena known as natural, collision and Doppler broadening. Secondly, the imaginary part of the susceptibility of the HCN gas is determined based on the obtained transmission spectrum. Thirdly, employing the Kramers-Kronig relations, the real part of the susceptibility is derived from its imaginary counterpart, allowing the calculation of the real part of the refractive index. Lastly, the differential phase shift is determined by analyzing the phase shifts of the two frequency components passing through the HCN gas shown in Fig. 2.3, which are derived from the refractive index.

Transmission Spectrum

The simulation of the transmission spectrum can be divided into two distinct stages. Initially, the width of the transmission spectrum is determined, which is equivalent to the width of the absorption spectrum [46]. Hence, this initial step provides both the incorrectly shaped transmission spectrum and absorption spectrum. In the subsequent stage, the transmission spectrum is appropriately scaled by accurately scaling the absorption spectrum and then deriving the transmission spectrum based on the correctly scaled absorption spectrum.

Stage 1

As explained in appendix A, the transmission spectrum of an atomic ensemble is derived from the scattered light emitted by the atoms and molecules in the medium. However, in practice, the spectrum of the re-radiation from a collection of atoms and molecules is not a perfectly sharp line but exhibits a broadened spectral lineshape due to three main types of interactions.

The first type of interaction arises from the finite natural lifetime, denoted as τ, of an excited atomic state. When atoms absorb photons during scattering, they get excited, but the excited state often lasts only a short time before the photon is re-emitted. According to the Heisenberg uncertainty principle, when applied to time and energy [47], a finite lifetime τ of an excited atomic state implies an associated energy uncertainty $\Delta E \geq \frac{\hbar}{\tau}$. Thus, a shorter excited state duration leads to higher energy uncertainty of the atomic state and greater uncertainty in the energy of the re-emitted photon. Consequently, this energy uncertainty leads to a natural spectral width $\Delta\omega$, which is determined by the relationship between energy and frequency of a photon as described by the Planck-Einstein relation $\omega = \frac{E}{\hbar}$ [48]. This phenomenon was already captured by the Lorentz oscillator model discussed in A.2.2 [48]. Due to the damping coefficient of the oscillating dipole, which represents the attracting force between the electron and the positive charged atom core, the emission of the dipole decays exponentially and can be described by a damped electromagnetic wave as $\underline{E}(z,t) = e^{-\gamma_n t} e^{-j(\omega_0 t + \phi(z))}$ [46]. Thus, in the frequency domain, the intensity of the observed wave does not exhibit an infinitely sharp spectral line - as it would for $\gamma_n = 0$ - but a broadened spectrum with a Lorentzian lineshape profile (see Eq. 3.1), representing the intensity of the re-radiation or transmission, respectively, similar to the imaginary part of the electric susceptibility discussed in A.2.2 [48, 49]. Consequently, the damping coefficient γ_n corresponds to the FWHM of the Lorentzian lineshape and indicates the frequency

uncertainty $\Delta\omega$ mentioned above.

$$I_N(f) = \frac{K_N}{(f - f_0)^2 + (\frac{\gamma_n}{2})^2} \tag{3.1}$$

The second type of interaction, known as collision broadening or pressure broadening, occurs when atoms collide. This phenomenon is particularly noticeable in gases and becomes more significant with increasing pressure. During these collisions, the re-radiated electromagnetic wave of an atom is randomly disrupted, resulting in a random phase shift of the radiated field at each collision. A sine wave that experiences random phase shifts at unpredictable times exhibits spectral broadening, and its spectrum follows a Lorentzian lineshape, similar to the natural broadening effect [46], which is fully described by the resonance frequency of the atom and its FWHM, given by the collision rate (i.e., mean number of collisions per second) [46]. Therefore, both lifetime and collision broadening can be represented by a Lorentzian lineshape function, where the overall linewidth is the sum of the individual linewidths [46]. Consequently, the intensity spectrum of the re-radiated electromagnetic field can be described by combining the phenomena of natural and collision broadening according to Eq. 3.2

$$I_{N,C}(f) = \frac{K_{N,C}}{(f - f_0)^2 + (\frac{\gamma_n + \gamma_c}{2})^2} \tag{3.2}$$

with γ_c being the linewidth resulting from collision broadening.

In this project, the HCN vapor cell utilized has a well-defined total linewidth for natural and collision broadening, as determined by the NIST [45]. The measurements conducted by NIST reveal that for the line center of 1549.730587nm, the broadening coefficient is $89\frac{\text{MHz}}{\text{Torr}}$. With the vapor cell operating at a pressure of 10Torr, the linewidth of the Lorentzian lineshape, which combines the contributions of γ_n and γ_c, is hence determined to be $0.89\frac{\text{GHz}}{\text{Torr}}$. Consequently, given the broadening coefficient $\eta = \gamma_n + \gamma_c = 0.89\frac{\text{GHz}}{\text{Torr}}$ and the line centre $\lambda_0 = 1549.730587$nm, the Lorentzian lineshape is fully characterized except for the proportional constant $K_{N,C}$ that will be incorporated later when considering the scaling of the absorption spectrum.

Thirdly, Doppler broadening in atomic physics refers to the broadening of spectral lines caused by the Doppler effect resulting from the distribution of velocities among atoms or molecules. When a particle moves towards the observer as a result of thermal motion, the emitted radiation undergoes a frequency shift towards higher frequencies. Similarly, when the emitter moves away, the frequency is shifted towards lower frequencies. The Doppler shift in frequency can be calculated using the formula (Eq. 3.3)

$$f = f_0(1 + (-)\frac{v}{c}) \tag{3.3}$$

where f is the observed frequency, f_0 is the actual frequency of the radiated wave, v is the velocity of the particle towards the observer and c is the speed of light.

As a result of the Doppler effect, the lineshape of an atom is shifted by the frequency $+(-)f_0\frac{v}{c}$. Variations in particle velocities lead to different Doppler shifts, and the combined effect of these shifts contributes to the broadening of absorption lines. The distribution of velocities is described by the Maxwell-Boltzmann distribution, resulting in an emitted radiation which follows a Gaussian lineshape given by Eq. 3.4 [48].

$$I_D(f) = K_D e^{-\frac{(x-\mu)^2}{2\sigma^2}} = K_D e^{-\frac{(f-f_0)^2}{2(\frac{\gamma_d}{2*\sqrt{2*ln(2)}})^2}} \tag{3.4}$$

The lineshape is centered around the resonance frequency f_0, and the standard deviation σ is related to the FWHM of the function, denoted as γ_d. To fully characterize the Gaussian lineshape, the FWHM must be determined as the resonance frequency is already known from [45]. According to Hill [50], the Half Width at Half Maximum (HWHM) of the Doppler-broadened component can be expressed by Eq. 3.5.

$$HWHM = \frac{\nu}{c}\sqrt{\frac{2N_A k T ln(2)}{M}} \tag{3.5}$$

The parameter ν represents the wavenumber of the spectral line transition in vacuum and is also provided by [45] as 6452.73448cm^{-1} for the HCN vapor cell. The parameter M corresponds to the molar mass of an HCN molecule, which is $27.0253\frac{\text{g}}{\text{mol}}$. N_A denotes the Avogadro constant, and c represents the speed of light. To obtain the FWHM in the correct units, the HWHM must be multiplied by 2 and the wavenumber must be converted to a frequency in Hz using the definition of the wavenumber $\nu = \frac{1}{\lambda} = \frac{f}{c}$, which leads to Eq. 3.6.

$$FWHM = \gamma_d = 2\nu\sqrt{\frac{2N_A kTln(2)}{M}} * 10^2 \tag{3.6}$$

The factor 10^2 accounts for the conversion between the wavenumber given in cm^{-1} and the speed of light given in $\frac{\text{m}}{\text{s}}$. Similar to the Lorentzian lineshape, the proportional constant K_D will be considered through the scaling of the absorption spectrum.

Due to the statistical independence of these three phenomena, the resulting transmission spectrum $\tilde{T}(f)$, which is not correctly scaled yet, is obtained by a convolution of the Lorentzian lineshape and the Gaussian lineshape, resulting in a so called Voigt profile (Eq. 3.7) [49, 51, 52]. It is worth noting that in many cases, the Doppler broadening is significantly larger than the broadening due to lifetime or collisions [48, 52], i.e., the Gaussian lineshape tends to dominate in solutions.

$$\tilde{T}(f) = I_{N,C}(f) \circledast I_D(f) \tag{3.7}$$

Stage 2

In order to correctly scale the transmission spectrum $\tilde{T}(f)$, the proportional constants $K_{N,C}$ and K_D need to be taken into account. This scaling is achieved by measuring the total absorption of the HCN vapor cell, which provides the overall proportional constant for the absorption spectrum, already incorporating $K_{N,C}$ and K_D. The total proportional constant is determined by the ratio of the intensity measured after and before passing through the vapor cell, given as $K_T = \frac{I(z=l)}{I(z=0)} = e^{-\alpha l}$, where l represents the length of the vapor cell. Thus, the relationship between this proportional constant and the actual absorption coefficient α defined in Eq. A.52 is given by $\alpha = -\frac{ln(K_T)}{l}$. Consequently, to obtain the correctly scaled absorption spectrum, the transmission spectrum $\tilde{T}(f)$ is multiplied by K_T and scaled by its peak value $\hat{\tilde{T}}(f)$ to obtain the absorption spectrum with a percentage axis, as shown in Eq. 3.8. Finally, the transmission spectrum can be obtained by subtracting the absorption spectrum from $100\% = 1$, as expressed in Eq. 3.8.

$$T(f) = 1 - K_T\frac{\tilde{T}(f)}{\hat{\tilde{T}}(f)} \tag{3.8}$$

Imaginary Electric Susceptibility

By considering the absorption coefficient, which is directly related to the imaginary part of the electric susceptibility (as shown in Eq. A.54), a relationship between the transmission spectrum and the imaginary part of the electric susceptibility can be derived (Eq. 3.9).

$$\frac{I(z)}{I_0} = T(f) = e^{-\alpha z} = e^{2k_0 \Im\{\sqrt{1+\underline{\chi}_e(f)}\}z} \tag{3.9}$$

For small values of $\underline{\chi}_e(f)$, the approximation $\alpha \approx -k_0\chi_I(f)$ is valid [51], leading to a simplified form of Eq. 3.9.

$$T(f) = e^{k_0\underline{\chi}_I(f)z} \tag{3.10}$$

By transforming Eq. 3.10, the imaginary part of the susceptibility can be obtained from the transmission spectrum (Eq. 3.11).

$$\chi_I(f) = \frac{ln[T(f)]}{k_0 z} \tag{3.11}$$

Real Refractive Index

The Kramers-Kronig relations, given by Eq. 3.12, establish a connection between the real part of the electric susceptibility and its imaginary part.

$$\chi_R(f) = -\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{\chi_I(f')}{f'-f}df' \tag{3.12}$$

Since this expression resembles a Hilbert transform, it is possible to determine the real part of the susceptibility by taking the Hilbert transform of its imaginary part (Eq. 3.13) [53].

$$\chi_R(f) = \mathcal{H}\{\chi_I(f)\} \tag{3.13}$$

As the Hilbert transform is readily available in standard signal processing toolboxes like Python or MATLAB, calculating the Hilbert transform is much simpler than directly solving the Kramers-Kronig integral.

From the real part of the susceptibility, the real part of the refractive index can be determined utilizing Eq. A.55.

$$n'(f) = \Re\{\sqrt{1 + \underline{\chi}_e(f)}\} = \sqrt{1 + \chi_R(f)} \tag{3.14}$$

Differential Phase Shift

Applying Eq. A.59 allows calculating the phase shifts induced by the HCN gas to the different frequency components. Subtracting both phase shifts delivers the differential phase signal (Eq. 3.15) which is the final measurement in the setup.

$$\Delta\phi(f, z) = \phi(f_{L,S}, z) - \phi(f_L, z) = -\frac{2\pi f_{L,S} z}{c_0}(n'(f_{L,S}) - 1) + \frac{2\pi f_L z}{c_0}(n'(f_L) - 1) \tag{3.15}$$

Figure 3.1 presents the simulation results for each of the four steps.

(a) Simulated transmission spectrum for frequency $f_L = 193.448$THz.

(b) Simulated imaginary susceptibility profile for frequency $f_L = 193.448$THz.

(c) Simulated refractive index profile for frequency $f_L = 193.448$THz.

(d) Simulated differential phase signal for the frequencies $f_L = 193.448$THz and $f_{L,S} = f_L + 240$MHz.
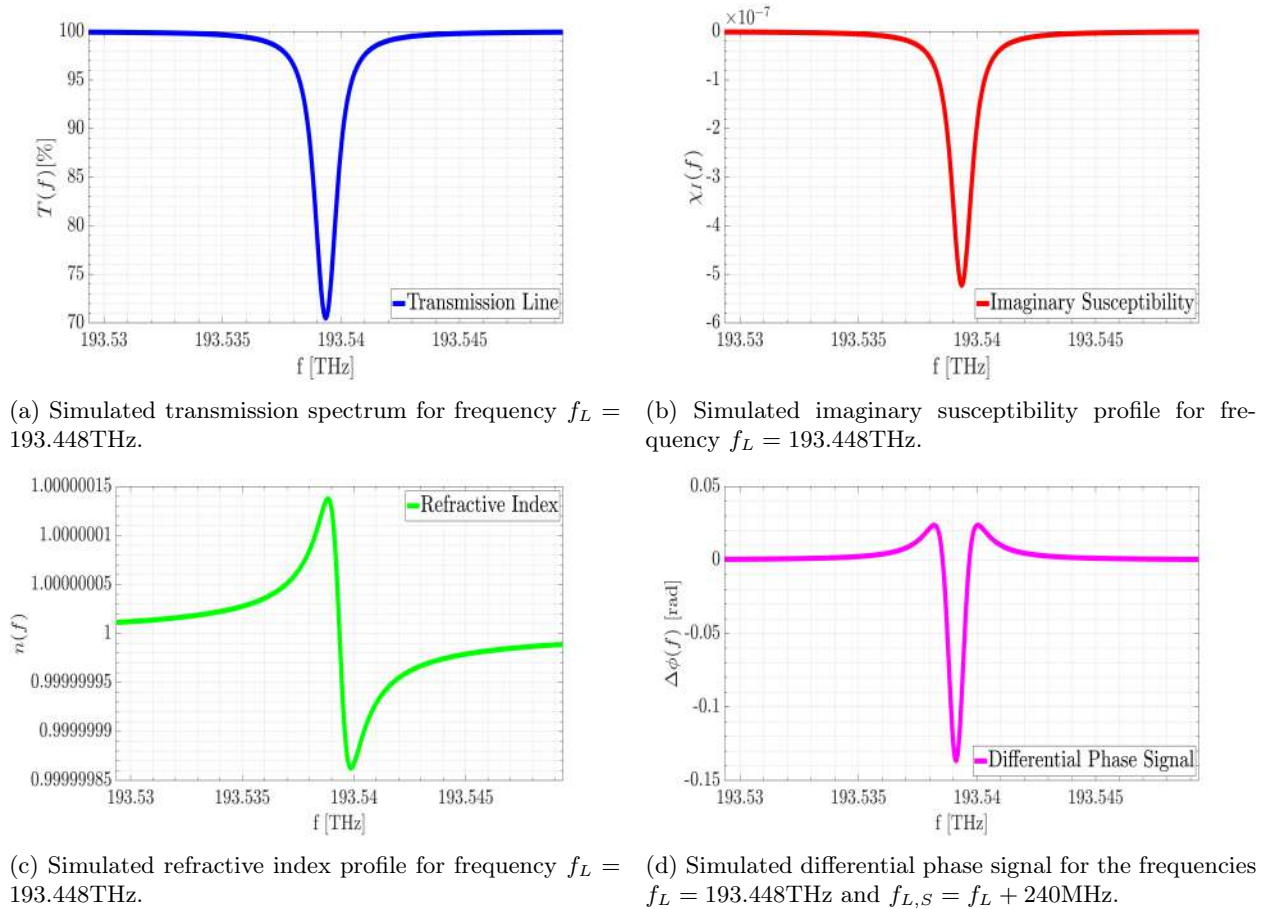
Figure 3.1: Setup simulation for a cell length of $z = 0.165$m , room temperature of $T = 296$K, and a scaling factor of $K_T = 0.295$.

Chapter 4

Experimental Implementation and SNR Enhancement

While metasurface-based dispersion spectroscopy is expected to offer greater sensitivity than conventional dispersion spectroscopy methods, the detectable phase shifts caused by the metasurface still remain extremely small. This necessitates that the entire sensing system must be as precise as feasible. Therefore, to ensure detectable phase shifts, it is particularly important to minimize noise sources as much as possible and maximize the differential phase signal amplitude in order to increase the SNR.

In this section, the physical implementation of the concept described in section 2.2 is explained, which consists of a free space optics part and a fiber optics part. Furthermore, the setup undergoes characterization to gain insights into its behavior. This involves calibrating the utilized laser, investigating the AOM as potential sources of noise and implementing measures to reduce its noise contribution before analyzing the final setup's performance in section 5.

Each ASD presented in the subsequent sections in chapter 4 and 5 is derived using an improved spectrum estimation technique developed by Troebs and Heinzel [54]. This method offers significant advantages in portraying density spectra for lengthy time series on a logarithmic frequency scale, allowing for the maximum amount of information to be displayed. It addresses the limitations of Welch's method, which provides lower frequency resolution for low frequencies and an inconvenient density of data-points at higher frequencies when plotted on a logarithmic axis as a result of the equidistant sample distance in the frequency domain [54]. The chosen approach, however, calculates an individual sample distance for each frequency, resulting in non-integer bin numbers specific for each sample which ensures an evenly distributed set of sample points along a logarithmically scaled frequency axis. In this project, the method involves employing 1000 spectral bins, applying a 10% segment overlap, and using an averaging factor of 1 for each frequency bin. For dividing the time series into multiple segments, a minimum 4-term Blackman-Harris window, also abbreviated as BH92, is applied. The main advantage of this window is the small height of the sidelope adjacent to the main peak of the frequency response. This height is 92dB below the main peak [55], leading to a reduced presence of error frequencies in the estimated spectrum. The measurements for noise analysis are performed off resonance of the HCN cell such that the characteristic phase signal is not obtained and only phase excursion due to noise is measured. Each measurement is carried out at a sampling rate of 2.4 kHz.

4.1 Experimental Setup

The physical setup, depicted in Fig. 2.3, comprises two main components. The first component is the metasurface breadboard, which represents the free space optics part of the experimental setup. The second component is the optical front end, derived from the Sagnac interferometer setup described in section 2.1, with slight modifications as explained in section 2.2. This component corresponds to the fiber optics part and includes all elements except the metasurface breadboard.

The entire equipment used for the free space optics component as well as the fiber optic component comprises the devices in Table 4.1.

Device	Functionality	Number
Fiber Collimator	Transforms the light output of the optical fiber into a free space collimated beam	1
Aperture	Changes the diameter of the light beam as necessary	2
Lenses	Adjust the light beam's focus as needed	3
Metasurface	Provides the information about the analyte concentration encoded in the resonance shift of its phase response	4
Microscope Objective	Spreads out the laser beam to investigate the alignment of the metasurface on the camera (5.1) or focuses the laser beam to concentrate maximal power on the photodetector (5.2)	5
Translation Stage	Facilitates the adjustment of the microscope objective	6
Mirrors	Reflects the light beam in order to control its direction in free space optics	7
Flip Mirror	Allows to flip the path of the light beam between the camera and the photodetector	8
Neutral Density Filter	Prevents the camera from reaching saturation	9
Camera	Assists in aligning the position of the metasurface, ensuring that the metasurface is accurately hit by the beam	10
Photodetector	Detects the intensity of the beat notes	11
Tunable Laser (1550nm)	Generates the initial laser beam	12
3dB-coupler	Superimposes and subsequently splits the two electromagnetic waves	13
AOM	Induces a frequency shift of $\omega_S = \omega_{S_1} = 200MHz$ and $\omega_{S_2} = 40MHz$ (for ω_{S_2} see section 4.3)	14
Fiber Polarization Controller	Brings the polarization of one pathway into alignment with the polarization of the other pathway	15
HCN Vapor Cell	Interim substitute for metasurface prior to its fabrication	16
Adjustable attenuator	Attenuates the optical power to prevent the photodetector from saturation	17
Moku:Pro	Measures the phase of the beat note, runs the AOM and sweeps the laser frequency	\
Optical Fiber	Guides the path of the light beam	\

Table 4.1: Setup equipment with corresponding functionalities and reference numbers.

4.1.1 Free Space Optics

Fig. 4.1 presents the metasurface breadboard, which is also illustrated in Fig. 2.3 and serves as the free space optics component of the setup. The devices utilized in this configuration are numbered according to Table 4.1.

The primary function of this breadboard is to ensure proper alignment of the laser beam to accurately target the metasurface. For that, a laser beam is launched from the fiber optics section to the free space optics part via a fiber collimator (1). Subsequently, the beam's diameter can be adjusted using an aperture (2) for alignment purposes. The following lens (3) enables adjustment of the beam's focus, directing the majority of its intensity onto the metasurface (4). A microscope objective (5.1) located behind the metasurface, which can be aligned using a translation stage (6), is employed to expand the laser beam which is necessary for obtaining a clear image on the camera (10). After being redirected by the mirror (7), the laser beam is either directed towards the camera (10), when the flip mirror (8) is employed, or directed towards another microscope objective (5.2). In the latter case, the expanded beam is refocused to maximize intensity on the photodetector (11). In the first case, the camera assists in aligning the metasurface with the laser beam since the metasurface itself is too small to be seen by the naked eye. During alignment, the mirror (7) and the translation stage (6) must be adjusted until the metasurface becomes visible on the camera, ensuring precise targeting of the laser beam to the metasurface. To prevent saturation of the camera due to the intense light
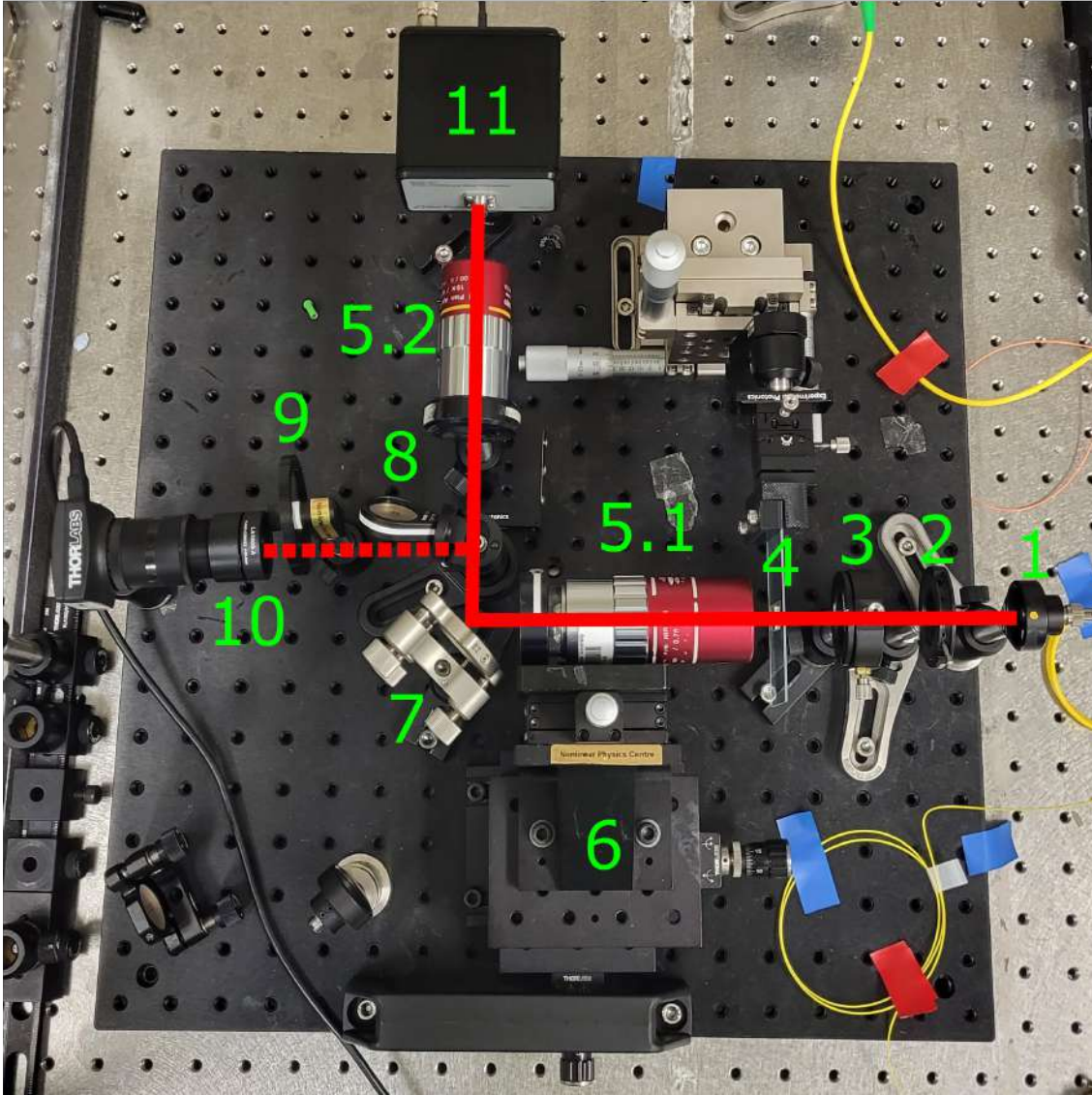
Figure 4.1: Metasurface breadboard with the numbered parts according to Table 4.1. The dashed red line depicts the path of the laser beam when the flip mirror is employed to direct it towards the camera. The solid red line represents the path taken by the laser beam after alignment when the setup is used for measurements.

beam, a neutral density filter (9) is used.

Therefore, the dashed red line in Fig. 4.1 illustrates the path of the laser beam when the flip mirror is employed to direct it towards the camera, aiding in the alignment of the metasurface with the laser beam. The solid red line represents the path taken by the laser beam after alignment when the setup is used for measurements.

4.1.2 Fiber Optics

The fiber optics component serves as the optical front end for the metasurface breadboard and connects to the phasemeter after the photodetectors. In the HCN setup used for system characterization, the fiber optics part also includes the HCN cell. The fiber optics part is depicted in Fig. 4.2 with the devices numbered according to Table 4.1.

Within this setup, a 1550nm laser (12) generates the laser beam, which is then guided through the optical fibers. The first 3dB-coupler (13.1) splits the laser beam into two separate beams with a 50:50 ratio. The first beam passes through both successive AOMs, where the first AOM (14.1), driven by a voltage amplitude of 400mV, shifts the laser beam by a frequency of 40MHz. Similarly, the second AOM (14.2), driven by a

voltage of 700mV, shifts the beam with a frequency of 200MHz before the beam reaches the second 3dB-coupler (13.2). The reason for using two serial AOMs is further explained in section 4.3. The second laser beam enters the second 3dB-coupler (13.2) without undergoing any modifications. Both beams combine at the second coupler, resulting in a beat note that consists of the initial laser frequency and a double shifted frequency of 240MHz. This beat note follows different paths afterward. The first path involves the metasurface (refer to section 4.1.1) or the HCN vapor cell (16), respectively, inducing varying phase shifts based on the wave frequencies. The other path simply directs the beat note to the end of the setup to subtract the phase difference $\Delta\phi_1$ from the phase signal received from the first path. The beat note in the second path is measured by the photodetector in Fig. 4.2 (11), while the photodetector measuring the first path is mounted on the metasurface breadboard (see Fig. 4.1). To prevent saturation of the photodetector, the power in the second path of the optical front end is attenuated using an adjustable attenuator (17). This is necessary as the photodetector saturates at 400μW. The signals from the photodetectors are then measured by the phasemeter of the Moku:Pro, which operates as a phase-locked loop and thus similar to a lock-in amplifier (see section 4.1.2.2). With the phasemeter set to a frequency of 240MHz, the phase extraction primarily focuses on the phase of the beat note in the signal, effectively canceling out much of the background noise, as further explained below.
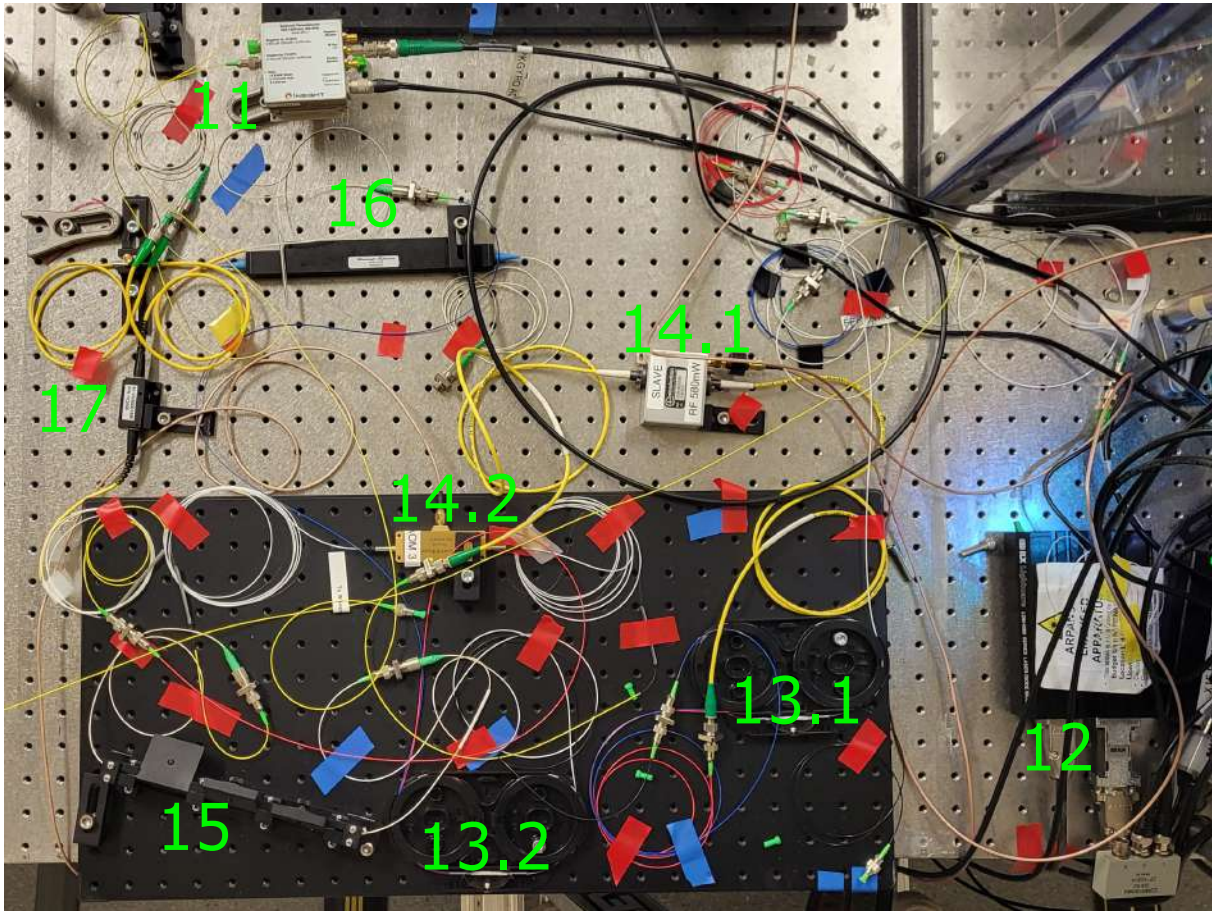
Figure 4.2: Front end breadboard with the numbered devices according to Table 4.1.

From the fiber optics setup, two devices, namely the AOM and the concept of the phasemeter of the Moku:Pro, are introduced a bit more since they play key roles in terms of the noise behavior of the system as discussed in section 4.3.

4.1.2.1 Acousto-Optic Modulator

An acousto-optic modulator is a device which allows the frequency of a laser beam to be modulated utilizing the acousto-optic effect described by Brillouin scattering.

In general, the propagation of an acoustic wave through a material leads to the deformation of the material, which in turn alters its refractive index. This means, sound waves produce a refractive index grating in the material which results in incoming light to be scattered and interfered similar to Bragg diffraction. From a quantum perspective, this process can be understood as the interaction between light photons and acoustic or vibrational quanta, known as phonons. Without going too much into detail, the big difference between the initial concept of Bragg diffraction and the acousto-optic effect is that for the acousto-optic effect the light is scattered from moving planes instead of fixed atoms in a crystal. So, because the refractive index grating is continuously moving, the output light is Doppler-shifted by an amount equal to the frequency f_s of the sound wave [56]. This shift is also known as Brillouin shift. From the perspective of solid state physics, this frequency shift can be considered as a result of creation and annihilation of phonons. In the so called Stokes Brillouin transition, a high-frequency photon is converted into a lower frequency Stokes photon, in the process creating a phonon. In the Anti-Stokes transition, a lower frequency photon is converted to a higher frequency photon while absorbing a phonon [57]. Such a shift in photon energy corresponds to a shift in frequency according to the Planck-Einstein relation $E = \hbar\omega$.

Technically, an acousto-optic modulator consists of a piezoelectric transducer attached to a material such as glass. An oscillating electric signal with frequency f_s drives the transducer to vibrate which creates sound waves in the material with frequency f_s representing the periodic modification of the refractive index in the crystal.

4.1.2.2 Phasemeter (Principle of a Lock-in Amplifier)

Dispersion spectroscopy relies on extracting encoded information from the phase of detected light, making the phasemeter a crucial component in the setup. A versatile tool called the Moku:Pro platform was used, which offers multiple instruments in one device, including a waveform generator and phasemeter. Both instruments were utilized for driving the AOMs and measuring the phases. To comprehend the concept of the Moku:Pro measuring the phase, it is essential to understand the technique it employs, known as phase-locked loop. Since the phase-locked loop technique is just an advanced version of a lock-in amplifier, discussing the basic concept of lock-in amplifiers is sufficient for understanding further considerations.

Lock-in amplifiers employ the knowledge about a signal's frequency to extract it from a noisy background. To achieve this, the input signal is split and separately multiplied by a reference signal that specifically represents the frequency of the signal component to be extracted and a 90° phase-shifted duplicate of the reference signal. The outputs of the mixers are then passed through low-pass filters, resulting in the two outputs I und Q, termed the in-phase and quadrature component. The amplitude A and the phase ϕ of the input signal are derived from I and Q by a transformation from Cartesian coordinates into polar coordinates. A schematic representation of this lock-in amplification process is depicted in Fig. 4.3.
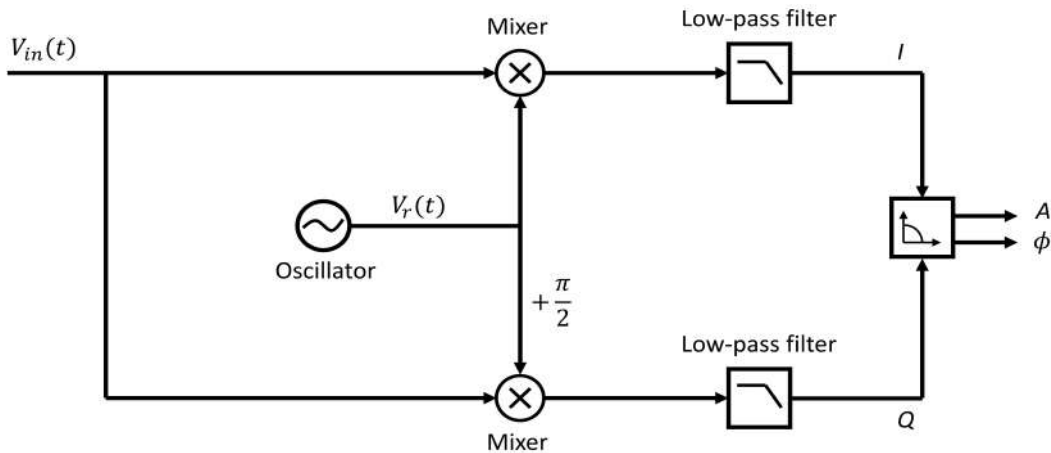

Figure 4.3: Basic concept of a lock-in amplifier. The input signal is split and separately multiplied with the reference signal and its 90° shifted version. Subsequently, each resulting signal is filtered by a low-pass filter. These filtered signals represent the I- and Q-components.

Mathematically, the process can either be described using real or complex numbers.
For that, the input signal is represented by a cosine signal (Eq. 4.1). In terms of using lock-in amplification as

phase measurement technique, the goal is to extract the phase from just this frequency component, assuming this component is superimposed by other frequencies which are considered as noise.

$$V_{in}(t) = A cos(\omega_{in} t + \phi) \tag{4.1}$$

The cosine signal can also be depicted using complex numbers (Eq. 4.2).

$$V_{in}(t) = \frac{1}{2} e^{j(\omega_{in} t + \phi)} + \frac{1}{2} e^{-j(\omega_{in} t + \phi)} \tag{4.2}$$

In the next step, the input signal is split up and multiplied by a reference cosine and sine signal given by $V_r(t)$ (Eq. 4.3), which is also called mixing. Using complex numbers, the split up can be replaced by multiplying the input signal by a complex reference signal (Eq. 4.4).

$$
\begin{aligned}
Z_1(t) =& V_{in}(t) * cos(\omega_r t) = A cos(\omega_{in} t + \phi) * cos(\omega_r t) = \\
& \frac{A}{2}(cos((\omega_{in} - \omega_r)t + \phi) + cos((\omega_{in} + \omega_r)t + \phi)) \\
Z_2(t) =& V_{in}(t) * (-sin(\omega_r t)) = A cos(\omega_{in} t + \phi) * (-sin(\omega_r t)) = \\
& \frac{A}{2}(sin((\omega_{in} - \omega_r)t + \phi) - sin((\omega_{in} + \omega_r)t + \phi))
\end{aligned}
\tag{4.3}
$$

$$
\begin{aligned}
\underline{Z}(t) =& I(t) + jQ(t) = V_{in}(t) * \underline{V}_r(t) = \\
& (\frac{A}{2} e^{j(\omega_{in} t + \phi)} + \frac{A}{2} e^{-j(\omega_{in} t + \phi)}) * e^{-j\omega_r t} = \\
& \frac{A}{2} e^{j((\omega_{in} - \omega_r)t + \phi)} + \frac{A}{2} e^{-j((\omega_{in} + \omega_r)t + \phi)}
\end{aligned}
\tag{4.4}
$$

Since applying a lock-in amplifier involves setting the oscillator frequency ω_r to match the frequency of the component that needs to be extracted, i.e., $\omega_r = \omega_{in}$, the process results in Eq. 4.5 and Eq. 4.6.

$$
\begin{aligned}
Z_1(t) =& \frac{A}{2}(cos(\phi) + cos(2\omega_{in} t + \phi)) \\
Z_2(t) =& \frac{A}{2}(sin(\phi) - sin(2\omega_{in} t + \phi))
\end{aligned}
\tag{4.5}
$$

$$\underline{Z}(t) = \frac{A}{2} e^{j\phi} + \frac{A}{2} e^{-j(2\omega_{in} t + \phi)} \tag{4.6}$$

The subsequent filtering removes all higher frequency components, resulting in Eq. 4.7 or Eq. 4.8.

$$
\begin{aligned}
Z_1 =& \frac{A}{2} cos(\phi) = I \\
Z_2 =& \frac{A}{2} sin(\phi) = Q
\end{aligned}
\tag{4.7}
$$

$$\underline{Z} = \frac{A}{2} e^{j\phi} = \frac{A}{2}(cos(\phi) + j sin(\phi)) = I + jQ \tag{4.8}$$

Finally, the amplitude and phase of the input signal can be determined using the relationship between Cartesian and polar coordinates for complex numbers as illustrated by Eq. 4.9.

$$
\begin{aligned}
A =& \sqrt{|I|^2 + |Q|^2} \\
\phi =& arctan(\frac{Q}{I})
\end{aligned}
\tag{4.9}
$$

To conclude, the essential elements of lock-in amplification are the mixer and low-pass filter. Together, they enable the extraction of signal components with a specific frequency by matching the frequency ω_r of the oscillator - also called lock-frequency ω_{lock} - to the desired components. The oscillator signal is multiplied with the input signal, resulting in a DC-signal for the desired component, as seen in Eq. 4.7 and 4.8, while other signal components with different frequencies are converted to higher frequencies. By applying the low-pass filter, the higher frequencies are eliminated, and the DC-signals can be used to determine the amplitude and phase of the original input signal which was superimposed by noise, i.e., other frequency components. This concept is crucial for reducing the noise and increasing the SNR in section 4.3.

4.2 Laser Calibration

The setup incorporates a distributed feedback diode laser that operates based on the principles of a Fabry-Pérot semiconductor laser. These lasers exhibit narrow bandwidth and rapid frequency modulation but are characterized as nonlinear since the gain is a function of energy flux [58]. Specifically for the real signal validation of the system within one of the resonance areas of the HCN vapor cell, as outlined in section 5.2, a precise calibration of the laser is necessary.

To achieve this, the laser wavelength is swept using a laser diode controller which allows for wavelength sweeping by adjusting either the temperature or current of the laser diode, with temperature changes producing more significant wavelength shifts compared to current adjustments. In this setup, a ramp signal with an amplitude of 400mVpp and a duty cycle of 50% was applied to the laser diode controller input to sweep the laser wavelength by modulating the current. The wavelength signal was measured using a wavelength meter. However, due to the wavelengthmeter's extremely low sample rate of approximately 4.67Hz, the frequency of the ramp signal had to be set very low at 1mHz to ensure an adequate number of samples. During a measurement period of 7 minutes, a single wavelength sweep lasting about 100 seconds was extracted. Fig. 4.4a illustrates a sweep line representing the relationship between the current, which is linear to the voltage of the applied ramp signal, and the wavelength.
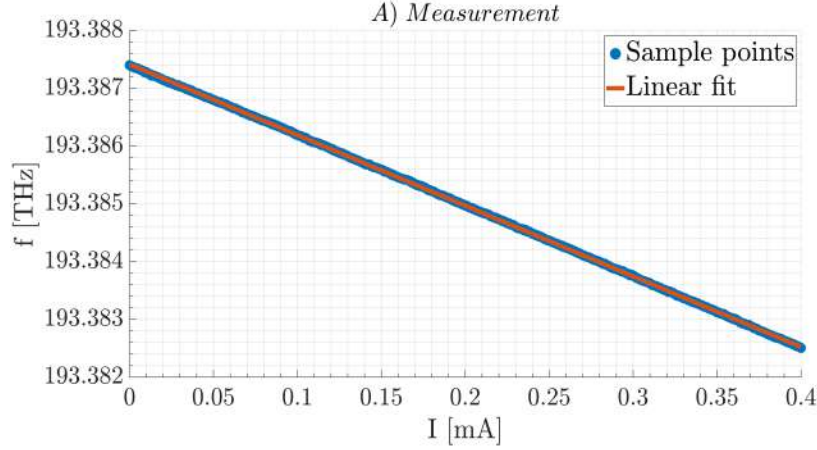
(a) Measured laser wavelength with respect to the laser diode current and corresponding linear fit curve. The nonlinearity of the laser is barely discernible.
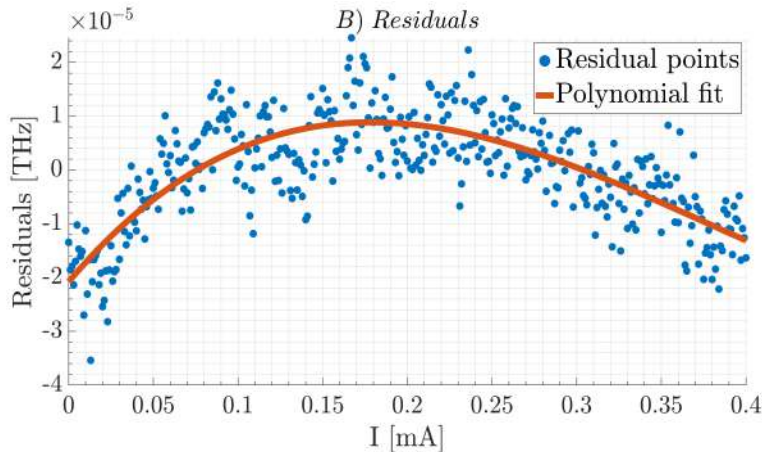
(b) Residuals of the measured signal and linear fit from plot 4.4a. The polynomial fit applied to the residual points reveals the nonlinear characteristics of the laser.

Figure 4.4: Laser characterization.

To highlight the laser's nonlinearity, which is only slightly discernible from the straightforward sweep measurement (plot 4.4a), the residuals between the sample points from the wavelength meter and their corre-

sponding linear fit curve are calculated. Plot 4.4b displays the residual samples along with their corresponding polynomial fit curve, effectively revealing the laser's slight concave nonlinear behavior.

The laser calibration performed will further be applied in section 5.2, as it will be utilized to mitigate any nonlinearities in the real signal measurement caused by laser tuning.

4.3   AOM Noise

As stated earlier in section 4.1.2, the final experimental setup employs two consecutive AOMs in the upper path instead of one single AOM. This decision was motivated by two primary factors: minimizing the AOM's contribution as a noise source and amplifying the amplitude of the differential phase signal. This section provides further clarification on these reasons.

4.3.1   Single AOM Structure

One possible source of noise was believed to be the AOM discussed in section 4.1.2.1. The AOM was thought to allow some of the input light to pass through without frequency shifting. This unshifted light leads to an additional term in the wave intensity measured by the photodetector that shares the same frequency as the differential phase signal. Thus, according to the theoretical considerations for the phasemeter in 4.1.2.2, the unshifted light is not canceled out through the demodulation and low pass filter of the phasemeter, causing an unwanted phase signal or noise, respectively.

To demonstrate this mathematically, another electromagnetic wave, $\underline{E}_3(z,t)$, representing the frequencies passing the AOM unshifted, is included in Eq. 2.12. As a result, Eq. 4.10 is obtained which describes the signal measured by the photodetector in the upper path by considering unshifted frequencies.

$$
\begin{aligned}
I(t) &= |E_1 e^{j\phi_1(t)} + E_2 e^{j\phi_2(t)} + E_3 e^{j\phi_3(t)}|^2 = \\
&E_1^2 + E_2^2 + E_3^2 + 2E_1 E_2 cos(\phi_1(t) - \phi_2(t)) + 2E_1 E_3 cos(\phi_1(t) - \phi_3(t)) + 2E_2 E_3 cos(\phi_2(t) - \phi_3(t)) = \quad (4.10) \\
&E_1^2 + E_2^2 + E_3^2 + 2E_1 E_2 cos(\omega_{S_1} t + \Delta\phi_2 - \Delta\phi_1) + 2E_1 E_3 cos(\Delta\phi_1) + 2E_2 E_3 cos(\omega_{S_1} t + \Delta\phi_2)
\end{aligned}
$$

with

$$
\begin{aligned}
\phi_1(t) &= \omega_L t + \phi_L + \Delta\phi_1 \\
\phi_2(t) &= (\omega_L + \omega_{S_1})t + \phi_{L,S_1} \\
\phi_3(t) &= \omega_L t + \phi_L \\
\Delta\phi_2 &= \phi_{L,S_1} - \phi_L
\end{aligned}
\qquad (4.11)
$$

where ω_L and $\omega_{S_1} = 200\text{MHz}$ are the frequency of the tunable laser and the shift frequency induced the AOM in the upper path. The phases ϕ_L and ϕ_{L,S_1} represent the phase shifts induced by the metasurface or vapor cell, respectively, for the corresponding frequencies, i.e., for ω_L and $\omega_L + \omega_{S_1}$.

Eq. 4.10 displays the extra terms that arise from the unshifted frequency components. Whereas the orange components are eliminated by the mixer and low pass filter of the phasemeter as background noise, the red component is problematic since it shares the same frequency as the blue component that generates the differential phase signal $\Delta\phi_2$. As a result, it corrupts the final measurement of the differential phase signal. To verify whether unshifted frequencies passing through the AOM are causing any problems, an approach was taken to increase the amplitude of the signal that is driving the AOM. This would result in higher amplitudes of sound waves propagating through the AOM crystal, reducing the number of electromagnetic waves that are not influenced by the sound waves.

Conducting noise measurements for different driving voltages and determining the ASDs for 2000 seconds suggested that unshifted frequencies passing through the AOM are likely a problem as increasing the signal amplitude from 400mV to 700mV generally resulted in less signal noise, as shown in Fig. 4.5a. While all other voltages surpass the 400mV measurement across the entire frequency range, the 500mV measurement is specifically surpassed by the 600mV and 700mV voltages at higher frequencies, leading to significantly lower noise levels for the 600mV and 700mV measurements, as evident in Fig. 4.5b. Upon comparing the noise limits of the 600mV and 700mV measurements in Fig. 4.5b, the 600mV measurement exhibited a slightly better noise limit, primarily due to an extremely small peak at the very end of its ASD (see Fig. 4.5a). However, considering the ASDs over the entire frequency range, the 700mV measurement surpassed the 600mV measurement across all frequencies, except for the small area at very high frequencies.

As the difference in noise limits between the 600mV and 700mV measurements is minimal, evaluating the performance over the entire frequency range provides a better indicator, clearly demonstrating the superior performance of the 700mV driving voltage over the 600mV one.
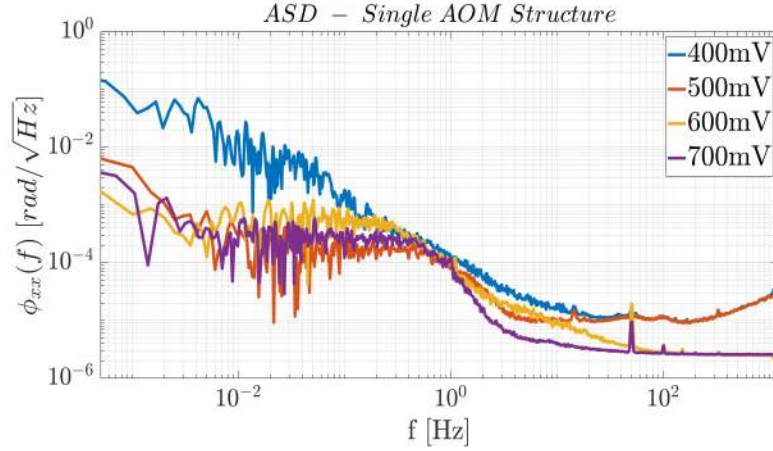
(a) ASD for different amplitudes driving the AOM. The results confirm the assumption that the AOM might be a noise source as the ASD decreases with higher driving amplitude. The flat noise shelves for low frequencies are an indicator for unshifted frequencies passing the AOM, similar to Rayleigh backscatter noise.
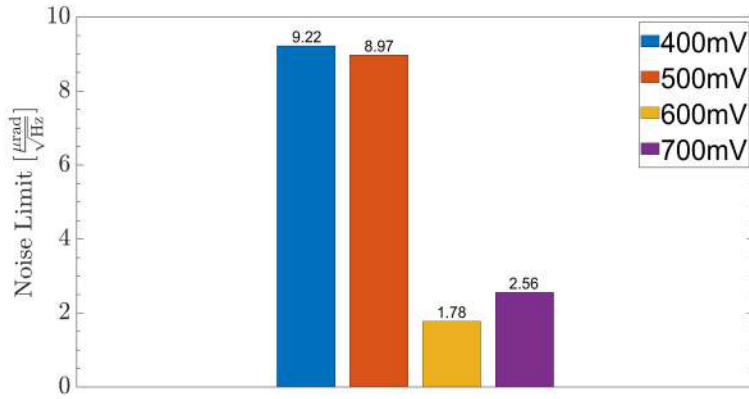
(b) Noise limits for the different driving voltages of the AOM. The plot shows that higher driving voltage generally reduces the noise limit. While the noise limit at 700mV is slightly higher than at 600mV, this should not cast doubt on the overall assumption, as the ASD plot demonstrates better performance of the 700mV voltage over nearly the entire frequency range.

Figure 4.5: Noise comparison for single AOM structure with respect to different driving voltages. The overall assumption that higher driving voltage reduces the AOM noise is generally validated.

In addition, especially the presence of a flat noise shelf in the ASDs (Fig. 4.5a) below 1Hz suggests the possibility of unshifted frequencies being re-coupled into the original laser beam, showing the same phase error as caused by Rayleigh backscattering within a fiber. Numerous studies have demonstrated that Rayleigh backscattering can introduce phase errors in the mHz range which arise from the interference between the actual metrology optical field and a distributed scattering field, manifesting as a flat noise shelf at low frequencies in the ASD [11–13]. Despite the red component not representing a scattered electromagnetic field, the theory of backscatter phase noise can be applied to it similarly, as it shares the same characteristics as a scattered field, i.e., it possesses the same frequency as the metrology field but differs in relative phase. As showcased by Fig. 4.5a, this shelf can be reduced by suppressing the amplitude of the unshifted frequency component, achieved by increasing the driving voltage of the AOM. However, regardless of what amplitude value is used for driving the AOM, it is challenging - if not impossible - to avoid frequencies passing the AOM unshifted. Consequently, the phase noise must be reduced differently by altering the setup in two ways. First, an additional AOM is placed parallel to the existing one in the lower path of the setup shown in Fig.

2.3. Second, another AOM is placed in series with the existing one in the same path. These modifications are expected to reduce the background noise as well as to increase the differential phase signal, as explained below, therefore improving the SNR.

4.3.2 Parallel AOM Structure

The utilization of a parallel AOM structure, in the context of Fig. 2.3, entails the addition of another AOM positioned parallel to the existing one in the lower path before the second coupler. From theory, the key benefit of this modification is that the term responsible for producing the differential phase signal operates at a frequency distinct from the other terms. Consequently, the lock-frequency of the phasemeter can be set to this unique frequency, removing all other intensity components, including those that arise from frequencies passing both AOMs unshifted.

Mathematically, the resulting intensity on the photodetector in the upper path is described by Eq. 4.12. Here, $\underline{E}_1(t) = E_1 e^{j\phi_1(t)}$ and $\underline{E}_2(t) = E_2 e^{j\phi_2(t)}$ represent the electromagnetic waves in the upper and lower path experiencing a phase shift by the AOMs. $\underline{E}_3(t) = E_3 e^{j\phi_3(t)}$ and $\underline{E}_4(t) = E_4 e^{j\phi_4(t)}$ describe the waves in the upper and lower path passing the AOMs unshifted.

$$
\begin{aligned}
I(t) =&|E_1 e^{j\phi_1(t)} + E_2 e^{j\phi_2(t)} + E_3 e^{j\phi_3(t)} + E_4 e^{j\phi_4(t)}|^2 = \\
& E_1^2 + E_2^2 + E_3^2 + E_4^2 \\
& + 2E_1 E_2 cos(\phi_1(t) - \phi_2(t)) + 2E_1 E_3 cos(\phi_1(t) - \phi_3(t)) \\
& + 2E_2 E_3 cos(\phi_2(t) - \phi_3(t)) + 2E_1 E_4 cos(\phi_1(t) - \phi_4(t)) \\
& + 2E_2 E_4 cos(\phi_2(t) - \phi_4(t)) + 2E_3 E_4 cos(\phi_3(t) - \phi_4(t)) = \\
& E_1^2 + E_2^2 + E_3^2 + E_4^2 \\
& + 2E_1 E_2 cos((\omega_{S_1} - \omega_{S_2})t + \Delta\phi_2 - \Delta\phi_1) + 2E_1 E_3 cos(\omega_{S_1}t + \phi_{L,S_1} - \phi_L) \\
& + 2E_2 E_3 cos(\omega_{S_2}t + \phi_{L,S_2} - \phi_L + \Delta\phi_1) + 2E_1 E_4 cos(\omega_{S_1}t + \phi_{L,S_1} - \phi_L - \Delta\phi_1) \\
& + 2E_2 E_4 cos(\omega_{S_2}t + \phi_{L,S_2} - \phi_L) + 2E_3 E_4 cos(\Delta\phi_1)
\end{aligned}
\tag{4.12}
$$

with

$$
\begin{aligned}
\phi_1(t) =&(\omega_L + \omega_{S_1})t + \phi_{L,S_1} \\
\phi_2(t) =&(\omega_L + \omega_{S_2})t + \phi_{L,S_2} + \Delta\phi_1 \\
\phi_3(t) =&\omega_L t + \phi_L \\
\phi_4(t) =&\omega_L t + \phi_L + \Delta\phi_1 \\
\Delta\phi_2 =&\phi_{L,S_1} - \phi_{L,S_2}
\end{aligned}
\tag{4.13}
$$

where $\omega_{S_1} = 200\text{MHz}$ and $\omega_{S_2} = 40\text{MHz}$ represent the shift frequency of the AOM in the upper and lower path and ϕ_{L,S_1} and ϕ_{L,S_2} describe the phase shifts induced by the metasurface or HCN vapor cell, respectively, for the frequencies $\omega_L + \omega_{S_1}$ and $\omega_L + \omega_{S_2}$.

In Eq. 4.12, the blue highlighted component represents the term which provides the actual differential phase signal $\Delta\phi_2 = \phi_{L,S_1} - \phi_{L,S_2}$. Since this term occupies the unique frequency $\omega_{S_1} - \omega_{S_2}$, the phasemeter's mixer and low pass filter remove all other frequency components when its lock-frequency is set to $\omega_{Lock} = \omega_{S_1} - \omega_{S_2}$. In this way, it can be ensured that the unshifted frequency components passing through the AOMs do not interfere with the differential phase signal as noise signals.

4.3.3 Serial AOM Structure

Next to the single and parallel structure, an alternative approach is to arrange the AOMs in a serial configuration. In terms of Fig. 2.3, this arrangement involves positioning both AOMs one after the other in the upper path, resulting in a different set of frequency components in the intensity measured by the photodetector, as described by Eq. 4.14. Here, $\underline{E}_1(t) = E_1 e^{j\phi_1(t)}$ represents the field which is shifted twice by both AOMs as desired, $\underline{E}_2(t) = E_2 e^{j\phi_2(t)}$ describes the field passing the second AOM unshifted, $\underline{E}_3(t) = E_3 e^{j\phi_3(t)}$ is the field passing the first AOM unshifted, $\underline{E}_3(t) = E_3 e^{j\phi_3(t)}$ represents the field passing both AOMs unshifted and $\underline{E}_5(t) = E_5 e^{j\phi_5(t)}$ describes the field in the lower path.

$$
\begin{aligned}
I(t) =& |E_1 e^{j\phi_1(t)} + E_2 e^{j\phi_2(t)} + E_3 e^{j\phi_3(t)} + E_4 e^{j\phi_4(t)} + E_5 e^{j\phi_5(t)}|^2 = \\
& E_1^2 + E_2^2 + E_3^2 + E_4^2 + E_5^2 \\
& + 2E_1 E_2 cos(\phi_1(t) - \phi_2(t)) + 2E_1 E_3 cos(\phi_1(t) - \phi_3(t)) + 2E_1 E_4 cos(\phi_1(t) - \phi_4(t)) \\
& + 2E_1 E_5 cos(\phi_1(t) - \phi_5(t)) + 2E_2 E_3 cos(\phi_2(t) - \phi_3(t)) + 2E_2 E_4 cos(\phi_2(t) - \phi_4(t)) \\
& + 2E_2 E_5 cos(\phi_2(t) - \phi_5(t)) + 2E_3 E_4 cos(\phi_3(t) - \phi_4(t)) + 2E_3 E_5 cos(\phi_3(t) - \phi_5(t)) \\
& + 2E_4 E_5 cos(\phi_4(t) - \phi_5(t)) = \\
& E_1^2 + E_2^2 + E_3^2 + E_4^2 + E_5^2 \\
& + 2E_1 E_2 cos(\omega_{S_2} t + \phi_{L,S_1,S_2} - \phi_{L,S_1}) + 2E_1 E_3 cos(\omega_{S_1} t + \phi_{L,S_1,S_2} - \phi_{L,S_2}) \\
& \color{red}{+ 2E_1 E_4 cos((\omega_{S_1} + \omega_{S_2})t + \phi_{L,S_1,S_2} - \phi_L) + 2E_1 E_5 cos((\omega_{S_1} + \omega_{S_2})t + \Delta\phi_2 - \Delta\phi_1)} \\
& + 2E_2 E_3 cos((\omega_{S_1} - \omega_{S_2})t + \phi_{L,S_1} - \phi_{L,S_2}) + 2E_2 E_4 cos(\omega_{S_1} t + \phi_{L,S_1} - \phi_L) \\
& + 2E_2 E_5 cos(\omega_{S_1} t + \phi_{L,S_1} - \phi_L - \Delta\phi_1) + 2E_3 E_4 cos(\omega_{S_2} t + \phi_{L,S_2} - \phi_L) \\
& + 2E_3 E_5 cos(\omega_{S_2} t + \phi_{L,S_2} - \phi_L - \Delta\phi_1) + 2E_4 E_5 cos(\Delta\phi_1)
\end{aligned}
\tag{4.14}
$$

with:

$$
\begin{aligned}
\phi_1(t) =& (\omega_L + \omega_{S_1} + \omega_{S_2})t + \phi_{L,S_1,S_2} \\
\phi_2(t) =& (\omega_L + \omega_{S_1})t + \phi_{L,S_1} \\
\phi_3(t) =& (\omega_L + \omega_{S_2})t + \phi_{L,S_2} \\
\phi_4(t) =& \omega_L t + \phi_L \\
\phi_5(t) =& \omega_L t + \phi_L + \Delta\phi_1 \\
\Delta\phi_2 =& \phi_{L,S_1,S_2} - \phi_L
\end{aligned}
\tag{4.15}
$$

As demonstrated by Eq. 4.14, the use of a serial AOM structure results in multiple frequency components for the measured intensity. In order to extract the frequency component that results from the double shifted field from the upper path and the field from the lower path (blue component), containing the differential phase signal $\Delta\phi_2$, the demodulation frequency of the phasemeter is set to $\omega_{lock} = \omega_{S_1} + \omega_{S_2}$, where ω_{S_1} and ω_{S_2} represent the shift frequencies of the first and second AOMs, respectively. However, an additional frequency component with the same frequency as the differential phase signal - represented by the red term in Eq. 4.14 - can result in noise since it interferes with the desired signal. So, at first glance, it may appear that the serial structure does not offer any advantage over the single AOM structure presented in Eq. 4.10. Nonetheless, since the unshifted frequency component in the serial structure arises from the fields that pass both AOMs unshifted, its amplitude is expected to be extremely small as the probability of an electromagnetic wave passing both AOMs unshifted is extremely small.

Moreover, the serial AOM structure offers advantages over the single or parallel AOM structures in terms of amplifying the amplitude of the differential phase signal. This, in turn, leads to an overall improvement in the SNR, as analyzed in section 4.3.5.
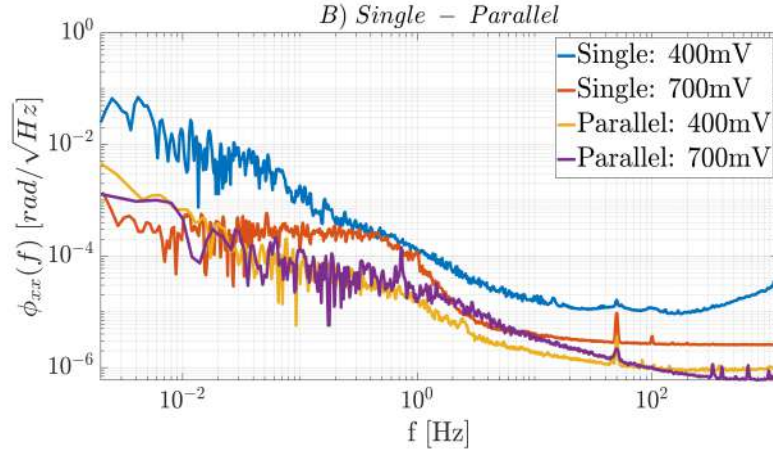
4.3.4 AOM Structure Comparison

In order to compare the practical noise characteristics of the three structures, noise measurements were carried out for each structure. Fig. 4.6 shows the ASDs and the corresponding noise limits for the single, parallel and serial AOM structures for a 400mV and 700mV amplitude of the signal driving the initial AOM in the upper path. The signals were measured for 1000 seconds.

The experimental results align with the previously established theoretical predictions, since both the parallel and serial AOM structures exhibit superior noise performance compared to the single structure, as evident by the noise limits in plot 4.6a. Additionally, irrespective of the driving voltage, the parallel structure surpasses its serial counterpart with an impressive minimal noise limit of $0.6 \frac{\mu rad}{\sqrt{Hz}}$, possibly due to residual unshifted frequency components passing through the serial AOM structure. This noise limit is noteworthy when compared with the state-of-the-art dispersion interferometer, the Sagnac interferometer, discussed in section 2.1, which exhibits a noise limit of $0.8 \frac{\mu rad}{\sqrt{Hz}}$.

Furthermore, the experimental observations corroborate the theoretical considerations by revealing a progressive increase in dependence on the driving voltage when transitioning from the parallel to the serial and ultimately to the single structure. This trend is evident when examining the ASDs in plot 4.6b and plot
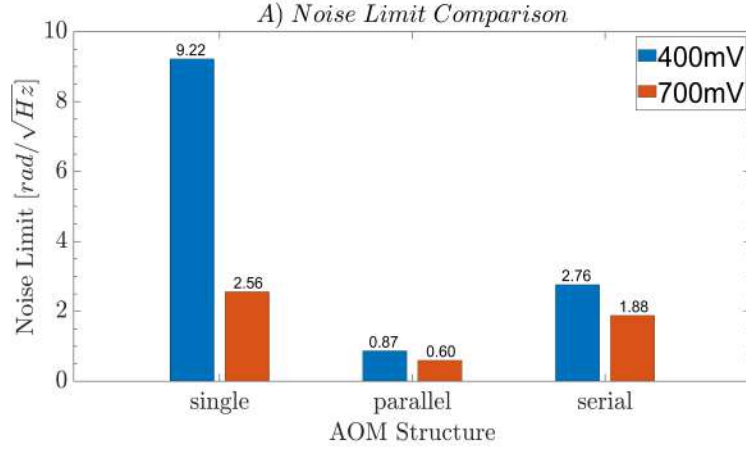
(a) Noise limit comparison of AOM structures. The parallel structure outperforms both the single and serial structure in terms of noise behavior. Additionally, the increasing dependence on the driving voltage as moving from the parallel to the serial and finally to the single structure confirms the theoretical considerations from previous sections.

(b) Comparison of single and parallel AOM structure. The parallel structure exhibits superior noise performance as it specifically demonstrates a reduction in the flat noise shelf for lower frequencies. Additionally, the parallel structure displays less dependence on the driving voltage compared to the single structure.
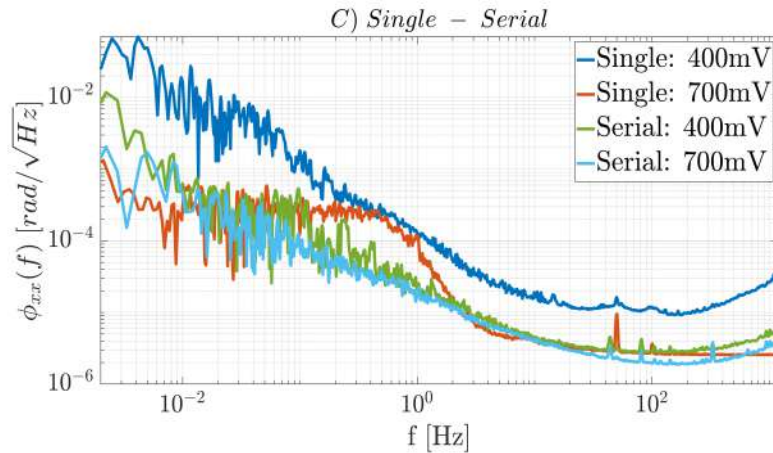
(c) Comparison of single and serial AOM structure. The serial structure exhibits superior noise performance since the serial AOM structure particularly showcases a reduction in the flat noise shelf for lower frequencies. Furthermore, the serial structure displays less dependence on the driving voltage compared to the single structure, similar to the parallel AOM structure.

Figure 4.6: Noise comparison for different AOM structures and driving voltages. The measurements confirm the theoretical presumptions.

4.6c and clearly illustrated in plot 4.6a, where the ratio between the noise limits of the 400mV and 700mV measurements increases as moving from the parallel to the serial and finally to the single structure. Specifically, for the parallel structure, the noise limit increases by $0.27 \frac{\mu \text{rad}}{\sqrt{\text{Hz}}}$ when the driving voltage is reduced from 700mV to 400mV. In contrast, the serial structure experiences a higher increase of $0.88 \frac{\mu \text{rad}}{\sqrt{\text{Hz}}}$, making it 3.3 times more sensitive to changes in driving voltage compared to the parallel one. This discrepancy was anticipated, as the design of the parallel and serial structures aims to cancel out unshifted frequency components, rendering the AOM driving amplitude irrelevant for the parallel structure. However, the serial structure retains a small amount of unshifted frequencies, leading to its greater dependence on the driving voltage. The contrast is even more evident when considering the single structure, where the signal noise exhibits a higher dependency on the driving voltage than for the parallel or serial structure. When comparing the noise limit from 700mV to 400mV, there is a significant increase of $6.66 \frac{\mu \text{rad}}{\sqrt{\text{Hz}}}$ corresponding to a factor of 3.60. In comparison, this increase is approximately 25 times higher than for the parallel structure and almost 8 times higher than for the serial structure, which reinforces the effectiveness of the parallel and serial AOM structures in suppressing signal noise and their independence from the driving voltage.

Additionally, the parallel and serial configurations successfully suppress the flat noise shelf at low frequencies, arising from the interference between unshifted frequencies and the metrology field, as discussed in section 4.3.1. This improvement is clearly evident in plots 4.6b and 4.6c below 1Hz, providing further confirmation of all the theoretical assumptions made previously including the consideration that unshifted frequencies pass through the AOM and can be treated as Rayleigh-Backscattered phase noise.

Moreover, the comparable noise levels observed for the serial and parallel AOM structures, as well as the Sagnac interferometer, validate that neglecting signal distortions due to path length differences after the second coupler in the metasurface architecture does not significantly affect the overall measured signal as the bidirectional architecture of the Sagnac interferometer cancels out these distortions automatically. As explained in section 2.2, these path differences only cause minor signal distortions since the beat note after the second coupler operates at an RF frequency, unlike the wave before the coupler, which represents an optical frequency.

In summary, when considering the noise behavior, the parallel AOM structure demonstrates the best performance, aligning with the theoretical expectations. However, since the primary objective of the setup characterization is to enhance the SNR, which relies not only on the amplitude of the noise signal but also on the amplitude of the measured signal, it is crucial to assess the impact of the AOM structures on the amplitude of the differential phase signal. This evaluation will be addressed in the subsequent section.

4.3.5   Influence on Differential Phase Signal

Altering the AOM structure in the experimental setup not only helps reduce setup noise but also has an impact on the magnitude of the differential phase signal. Thus, an alternative method to enhance the SNR is to maximize the value of the differential phase signal, which can be achieved by selecting the appropriate AOM structure based on the shape of the metasurface's phase response.

Assuming the phase response follows an S-shaped curve, as derived from theory, the optimal value of the differential phase signal can be obtained by employing a serial AOM structure. This is because the span between the phase shifts caused by the two frequencies passing through the metasurface in the serial structure always encompasses the span between the phase shifts resulting from the frequencies passing through the metasurface in the single or parallel structure as demonstrated by Fig. 4.7. Consequently, since the phase response curve of the metasurface is monotonically decreasing, the differential phase signal of the serial structure is always higher than the one of the single or parallel structure regardless of the shift frequencies. This shows that the serial structure enhances the value of the differential phase signal compared to the single AOM or parallel AOM structure. If the amplitude of the differential phase signal is enhanced significantly more than that of the noise signal resulting from the frequency component passing both AOMs unshifted, the SNR for the serial structure will be higher than that of the parallel one.

This principle also holds true when comparing the single and parallel structure. Given that the frequency range provided by the single structure inherently includes that of the parallel structure as depicted in Fig. 4.7, the differential phase signal value of the parallel structure consistently remains lower than that of the single structure. However, the investigation in the preceding section showcases the significantly greater reduction in signal noise achieved by the parallel structure as opposed to the single variant. Consequently, the notably reduced signal noise exhibited by the parallel structure is projected to offset the slightly higher differential phase signal value observed in the single structure, which, in turn, leads to the assumption that the parallel configuration will excel over the single arrangement in terms of SNR performance.
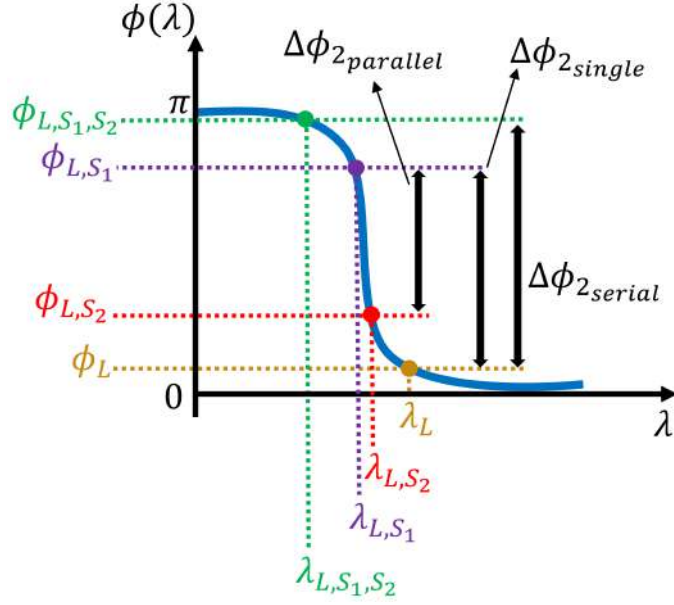
Figure 4.7: Influence of single, parallel and serial AOM structure on differential phase signal value. Due to the monotonically decreasing nature of the phase response, the span between the wavelengths that pass through the metasurface in the serial structure always encompasses the span between the wavelengths passing through the metasurface in the single or parallel structure. As a result, the amplitude of the differential phase signal is always greater in the serial structure compared to both the single and parallel structures.

So, the parallel structure reduces background noise more effectively than the serial structure, whereas the serial structure increases the differential phase signal more than the parallel structure. However, the unshifted frequency components in the serial structure are expected to be highly suppressed, as explained earlier, and thus the serial structure is assumed to overcompensate its additional background noise by increasing the differential phase signal. Consequently, the serial AOM structure is presumed yielding a higher SNR than the parallel structure and the single AOM structure. Therefore, the serial AOM structure will be selected for the final setup.

However, it is important to emphasize that in practical applications, the phase response of a metasurface may not always adhere to the expected S-shaped curve derived from theoretical analysis. In fact, there are instances where the metasurface is intentionally fabricated to exhibit a phase response that deviates from the typical curve. In such cases, the selection of the optimal AOM structure depends on the specific characteristics of the phase response, particularly its shape and bandwidth. Depending on the shape and bandwidth of the phase response, the parallel AOM structure can prove to be more advantageous than the serial structure if both structures are capable of providing the desired optimal frequency shift, but the parallel configuration offers improved noise performance. This distinction is explored in greater detail in section 5.3.2, where the influence of the AOM-induced frequency shift on the metasurface's phase response is further examined and discussed.

Chapter 5

System Performance Analysis

In this chapter, a comprehensive evaluation of the overall system performance is carried out based on the findings discussed in 4.3. This entails assessing the system's noise characteristics, stability, and its behavior in the resonance region of the HCN cell. The resonance behavior is consequently compared to the simulated setup described in section 3. Additionally, after the final characterization of the system, an incorporation of two metasurfaces into the setup is simulated. This simulation enables the prediction of the optimal AOM frequency shift and the system's minimal detectable resonance shift. Since the newly fabricated metasurface was not available at this stage, two older metasurfaces with an undesired broad bandwidth are employed for the simulation.

5.1 Noise Analysis

To evaluate the system's noise behavior, a final noise measurement was performed utilizing a serial AOM structure with AOMs driven at two different settings: 700mV and 200MHz, as well as 400mV and 40MHz. The measurement involves capturing the single phase signals for both channels outside the resonance area of the HCN gas for one hour and determining the differential phase signal, which represents the system noise, and the ASDs for all three time signals calculated using the same method as for previous ASDs. The obtained results are presented in Fig. 5.1. These plots provide important information about the signals measured in this setup in the time domain as well as in the frequency domain.

In Fig. 5.1a, it is shown that the individual phase signals exhibit significant fluctuations. As explained in section 2.2, the signal $\phi_{cell}(t)$, representing the phase signal from the metasurface path in Fig. 2.3, comprises $\Delta\phi_2(t)$ and $\Delta\phi_1(t)$, while $\phi_{free}(t)$, representing the phase signal in the upper path shown in Fig. 2.3, mainly consists of $\Delta\phi_1(t)$. Hence, since both signals demonstrate similar drift patterns, it can be concluded that these undesired signal distortions primarily arise from the phase signal $\Delta\phi_1(t)$, which originates from the variation in path lengths before the second coupler. The corresponding ASDs in Fig. 5.1c further support this conclusion. Since the ASDs for both single phase signals are nearly identical, it is evident that the main noise contribution comes from the signal $\Delta\phi_1$ representing the strong fluctuations in the time domain. However, in the metasurface setup, the differential phase signal, obtained by subtracting the single phase signals $\phi_{cell}(t)$ and $\phi_{free}(t)$, is used to extract the encoded information from the analyte. This effectively mitigates the extreme signal distortions resulting from varying path lengths before the second coupler, significantly reducing the impact of fluctuations in the individual phase signals and rendering them less problematic. Consequently, the differential phase signal $\Delta\phi_2$ appears as a straight line in comparison to the single phase signals depicted in Fig. 5.1a. Examining the differential phase signal in the frequency domain through its ASD, illustrated by the red curve in Fig. 5.1c, which results from the subtraction of the ASDs of the single phase signals represented by the blue and green curves, the ASD analysis reveals a noise limit of $1.9\frac{\mu\mathrm{rad}}{\sqrt{\mathrm{Hz}}}$ for the metasurface setup. Moreover, a noise suppression factor of 3×10^5 for low frequencies and 290 for high frequencies when utilizing the differential phase signal instead of a single phase signal as final measurement is evident. These results underscore the immense significance of noise suppression provided by the additional interferometer path in the metasurface setup. The minor residual fluctuations in the differential phase signal, as shown in Fig. 5.1b, or the residual noise, represented by the red curve in Fig. 5.1c, are probably mainly attributed to the modest transmission of unshifted frequency components through the serial AOM structure,
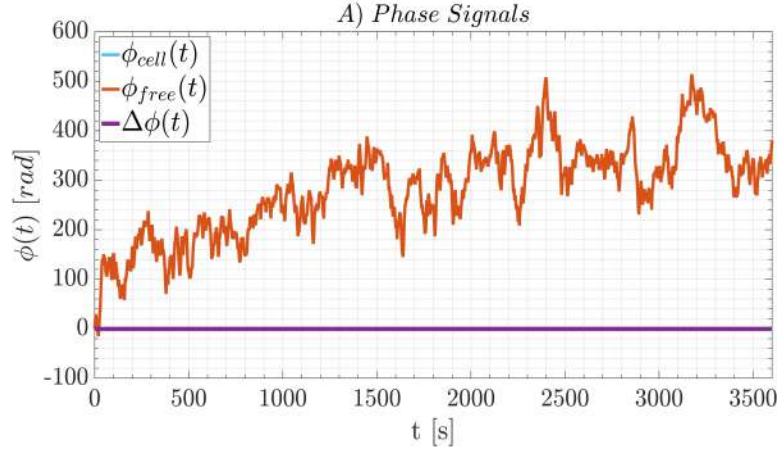
(a) Single phase signals for each path. They exhibit significant fluctuations, primarily attributed to the phase signal $\Delta\phi_1(t)$ which arises from the disparity in path lengths before the second coupler, as discussed in section 2.2. In comparison, the differential phase signal $\Delta\phi(t)$ exhibits significantly less noise, appearing as a straight line.
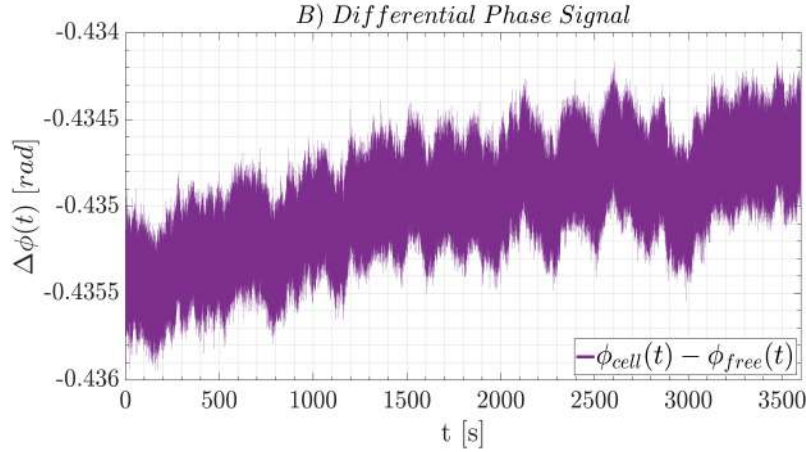
(b) Differential phase signal $\Delta\phi_2(t)$ (see section 2.2) off resonance, demonstrating notably smaller fluctuations compared to the single phase signals and thus less signal noise. Unshifted frequency components passing the serial AOM structure and different path lengths after the second coupler are probably the main reason for the residual fluctuations.
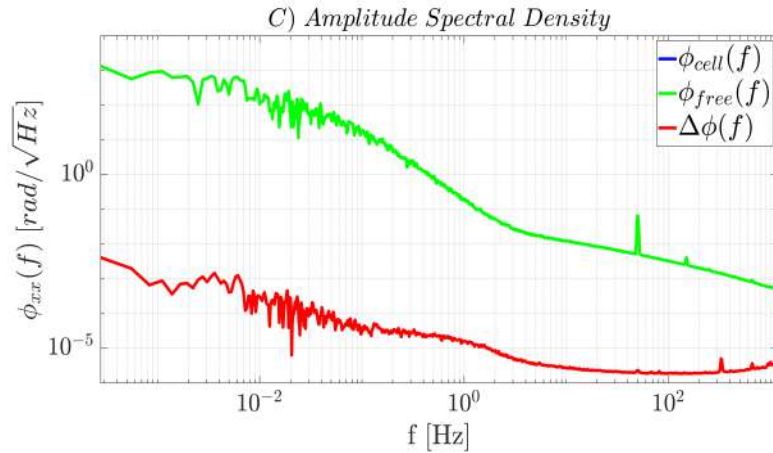
(c) ASDs for all three signals. The similarity of the ASDs of the single phase signals further supports the notion that the phase signal $\Delta\phi_1(t)$ primarily contributes to the signal noise. The noise caused by the fluctuations in the individual phase signals is effectively eliminated by subtracting the signals from both paths. This is evident from the red curve representing the noise level for the differential phase signal and a noise limit of $1.9\frac{\mu\mathrm{rad}}{\sqrt{\mathrm{Hz}}}$ for the system.

Figure 5.1: Noise analysis measurements. They conducted by measuring the single phase signals and the differential phase signal, along with calculating their corresponding ASD, in an off-resonance configuration.

coupled with minor remaining signal distortions after the second coupler.

When comparing the metasurface setup to the Sagnac interferometer setup, which has a noise limit of $0.8 \frac{\mu\text{rad}}{\sqrt{\text{Hz}}}$, several conclusions can be drawn. The transition from the Sagnac interferometer to the metasurface setup involves moving from a bidirectional configuration to a unidirectional one. Consequently, the unidirectional configuration of the metasurface setup introduces an additional noise source, which is the phase signal $\Delta\phi_1$. As previously demonstrated, this phase signal constitutes the primary noise source in the entire setup but is effectively canceled out by the second path of the metasurface interferometer. In essence, this second path takes on the role of substituting the bidirectional configuration of the Sagnac interferometer.

As mentioned in section 4.3, another significant noise source is the AOM structure. However, it is possible to fully avoid AOM noise by implementing a parallel structure, resulting in a noise limit of $0.6 \frac{\mu\text{rad}}{\sqrt{\text{Hz}}}$ (see section 4.3.4), or nearly avoid it with a serial structure, leading to a noise limit of $1.9 \frac{\mu\text{rad}}{\sqrt{\text{Hz}}}$. On the other hand, the Sagnac interferometer architecture inherently cancels out AOM noise through its bidirectional configuration. Therefore, when the metasurface setup utilizes a parallel AOM structure, it can be seen as an analogous approach to the Sagnac interferometer, where the bidirectional configuration is effectively replaced by a second interferometer path, thereby canceling out the noise source automatically canceled out by the Sagnac interferometer architecture. The similarity in noise limit between the Sagnac interferometer and the metasurface setup with a parallel AOM structure serves as further confirmation of their comparable performance. Any minor differences observed in the noise limit could be attributed to background noise in the laboratory or other negligible noise sources. Consequently, despite the higher complexity, the metasurface setup demonstrates comparable performance to the Sagnac interferometer and the slightly higher noise limit for the metasurface setup with a serial structure can mainly be attributed to the additional AOM noise.

In conclusion, the measurements exhibit promising results, showcasing a noise limit of approximately $1.9 \frac{\mu\text{rad}}{\sqrt{\text{Hz}}}$ as evident in the ASD of the differential phase signal. In comparison to the Sagnac interferometer approach documented in [44], which achieved a noise limit of $0.8 \frac{\mu\text{rad}}{\sqrt{\text{Hz}}}$ and serves as the baseline model, this noise limit is quite impressive considering that the setup using the serial AOM structure does not fully avoid all AOM noise as achieved inherently by the Sagnac approach. The metasurface configuration comes closest to the Sagnac approach's noise limit when using the parallel AOM structure. In this case, the metasurface setup demonstrates performance comparable to the baseline model, as the bidirectional configuration of the Sagnac interferometer is effectively substituted by the second interferometer path and the parallel AOM structure in terms of noise cancellation. Nevertheless, a noise limit of $1.9 \frac{\mu\text{rad}}{\sqrt{\text{Hz}}}$ still surpasses the majority of sophisticated dispersion spectroscopy architectures, such as HPSDS or CLaDS, which typically achieve noise limits in the milliradian range [44].

One approach to further decrease the noise limit of the system and thus increase its precision or limit of detection is to average multiple measurements, which effectively reduces the influence of background noise on the actual signal. However, if the measured signal contains drifts, such as the differential phase signal shown in plot 5.1b, averaging the signal over an excessively long period will also incorporate these drifts. Consequently, the final value becomes less precise, as the distortions caused by the incorporated drifts become more significant and overshadow the reduction in noise achieved by averaging, leading to a rise of the detection limit. The number of samples that should be used for averaging, in order to remove the background noise from the signal, is limited by the stability of the system. A more stable system exhibits smaller drifts in the measured signal and can therefore utilize a larger number of samples for averaging. By increasing the number of samples used for averaging, the background noise is further reduced in the actual signal, resulting in a lower limit of detection. To determine the optimal integration time, denoted as τ_{opt}, the Allan variance [14] is commonly employed. This quantity serves as a parameter to quantify the stability of a system. By analyzing the Allan variance, one can identify the integration time that maximizes the system's precision by minimizing the influence of noise through averaging. Since for white-noise dominated systems, the square root of the Allan variance is proportional to the detection limit of the system [14], the Allan variance can be utilized to illustrate the detection limit of a system as a function of the number of samples being used for averaging. Fig. 5.2 depicts the Allan plot representing the system utilized in this project. The minimum value of the Allan variance indicates the optimal integration time τ_{opt} which is for this system at about $\tau_{opt} = 128$s. Upon comparing this value with the optimal integration time of 60 seconds reported in other spectroscopy setups [14], it becomes evident that the provided system exhibits favorable stability.

In the context of the metasurface-based system from 2.2, the concept of averaging can be applied in the final step when the phase response of the metasurface is known, and the sample is illuminated with the wavelength that corresponds to the highest absolute slope value of the phase response. During this step,
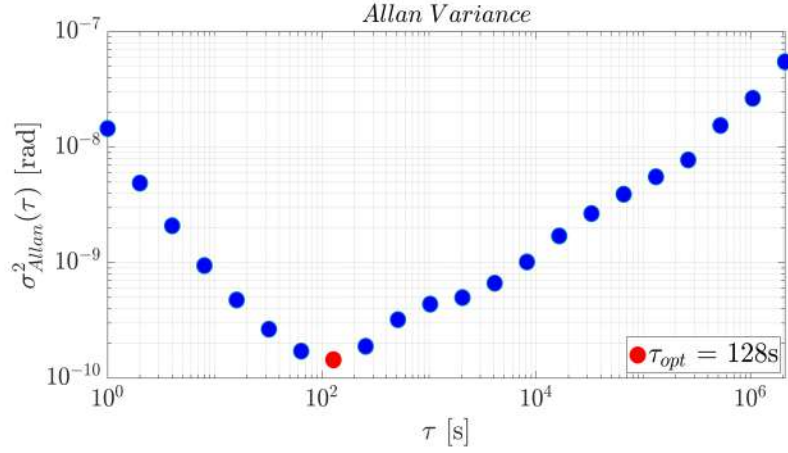
Figure 5.2: Allan plot for the metasurface setup. The red dot represents the optimal integration time, which is at $\tau_{opt} = 128$s.

measuring the signal with a maximal time of $\tau_{opt} = 128$s and subsequently averaging the measurements can effectively remove noise and improve the detection limit of the system.

5.2 Real Signal Validation

Validating the system's behavior requires not only assessing its noise performance but also determining its capability to detect a real signal. This means testing whether the system can effectively identify and register the presence of an actual signal amidst the noise and provide accurate measurements or responses accordingly. A reliable method for that is to measure the differential phase signal in one of the resonance regions of the HCN vapor cell and compare it to the corresponding theoretical prediction. This is accomplished by simulating the differential phase signal using the theoretical considerations presented in section 3 and utilizing the calibration data of the HCN vapor cell provided by NIST [45]. The gas cell used contains pure HCN gas at a calibrated pressure of 10Torr. The spectroscopic measurement is conducted specifically for the P10 line of the HCN gas, a common wavelength reference for 1550nm.

In order to obtain the dispersion profile, the laser wavelength is swept across the resonance area, which is accomplished using the same method employed for the laser calibration in section 4.2. This means, the resonance area is initially located manually by adjusting the temperature of the laser diode controller, before the ramp signal is applied to the controller input to sweep the laser wavelength across the resonance. The same ramp signal as for the laser calibration with an amplitude of 400mVpp, a duty cycle of 50% and a frequency of 10mHz is chosen since a low frequency provides more sample points for the wavelength axis, which is necessary because of the low sample rate of the wavelengthmeter.

The differential readout is obtained by simultaneously recording the wavelength signal and differential phase signal over a time axis for a measurement period of 3 minutes. The measurement results for the selected frequency are presented in Fig. 5.3. This figure showcases the recorded wavelength measurement, the differential phase signal plotted against time, and the dispersion profile, which illustrates the relationship between the differential phase signal and the corresponding wavelengths.

Comparing the wavelength signal from plot 5.3a with the differential phase signal over time from plot 5.3b highlights that the laser sweeps over the resonance area for each duty cycle, resulting in a corresponding dispersion profile. When the ramp signal decreases, it can be regarded as a "back sweep," causing the dispersion profile to be mirrored with respect to the middle of the ramp signal. In plot 5.3c, the differential readout is converted from time to optical frequency. This transformation entails reparameterizing the time axis using the calibration curve outlined in section 4.2, thereby eliminating the nonlinearity associated with laser tuning from the measurement. To accomplish this, a section of the wavelength signal (plot 5.3a) is extracted, which corresponds to a time interval covered by the differential phase signal (plot 5.3b). Since the sample rate of the wavelength meter is considerably lower than that of the phasemeter, the extracted wavelength signal is interpolated before being utilized as the x-axis for plot 5.3c which ensures a coherent mapping between the optical frequency shift and the differential phase signal.
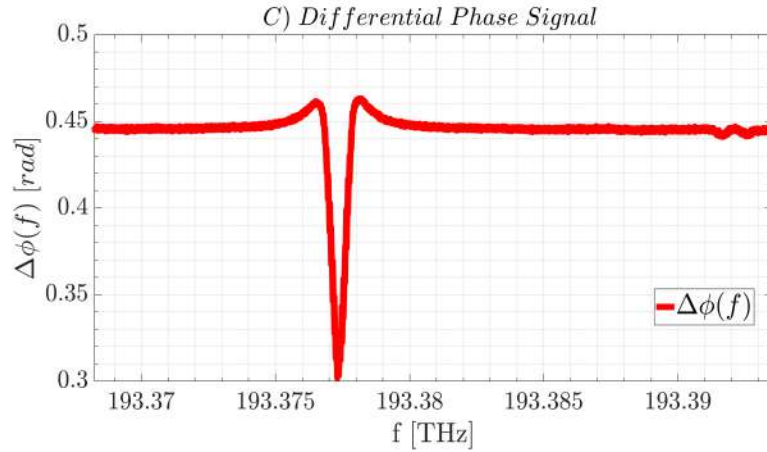
The subsequent step involves comparing the measured differential phase signal with the simulation obtained
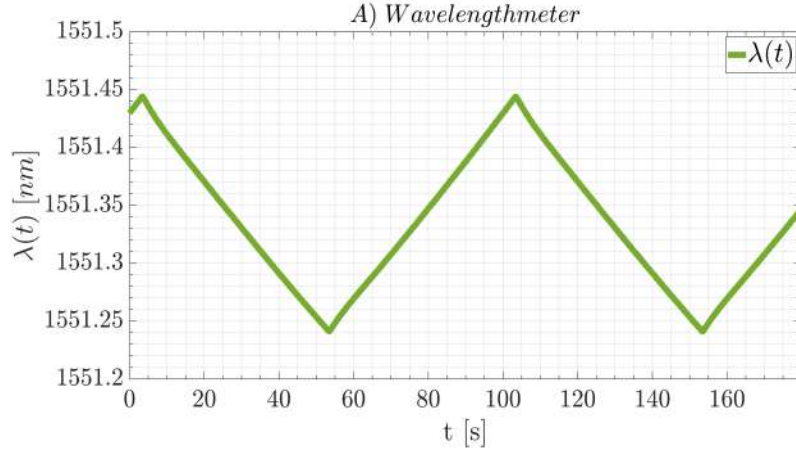
(a) Wavelength sweep achieved by applying a 10mHz ramp signal to the laser diode controller.
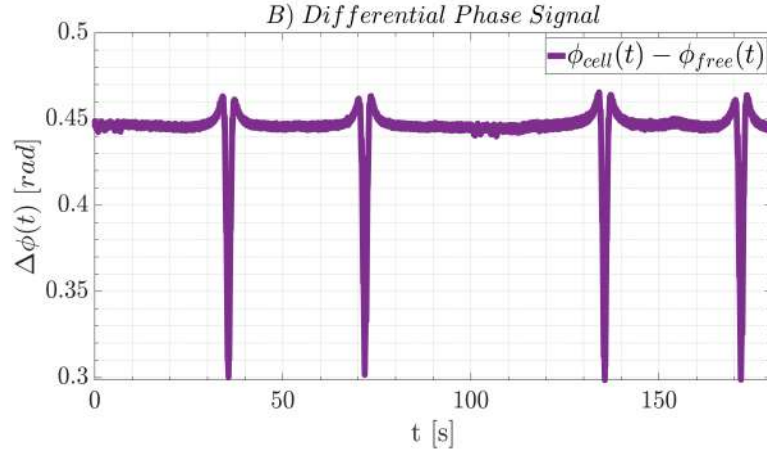

(b) Resonance signal evolving over time due to the wavelength sweep shown in Fig. 5.3a.

(c) Reparameterizing the time axis to the wavelength axis. The nonlinearity of the laser, which was investigated in section 4.2, is automatically eliminated.

Figure 5.3: Obtaining measurement results in resonance area by applying a 10mHz ramp signal.

based on the considerations outlined in section 3. To ensure an accurate comparison, both signals are adjusted on the x-axis to center the phase signals at the line center of the P10 transition. In order to visualize the differences between the measurement and the simulation, the residuals between these two signals are calculated. For that, the measured signal is subjected to a FIR low-pass filter with a cut-off frequency of 10Hz. This filtering process effectively removes high-frequency technical noise from the measurement, enabling a clearer representation of the deviations between the measured and simulated signals. As illustrated in Fig. 5.4a, the modeled signal exhibits good agreement with the experimentally obtained signals. A more distinct comparison can be observed by examining the residuals of both signals, depicted in Fig. 5.4b, which further highlights the concurrence between the two sets of data. By quantifying the accuracy as the standard deviation of the difference between the measured and the simulated dispersion spectra, a deviation of only 4.2% is achieved. This outcome is desirable when compared to other studies in the field, which reported a deviation of 4.5% [2].
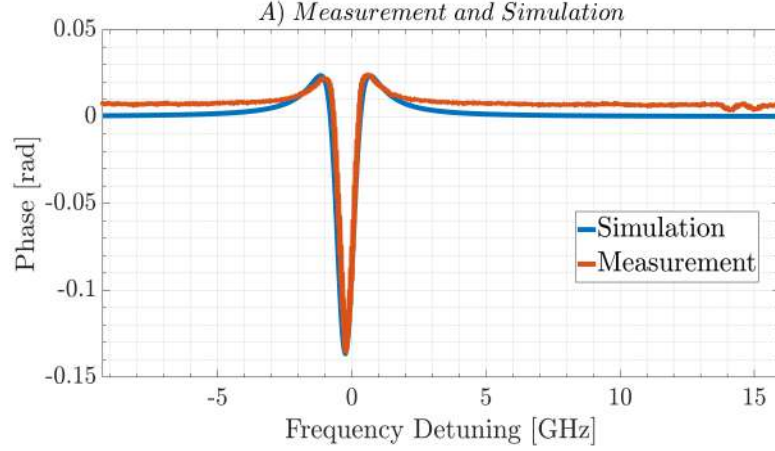
(a) HCN gas resonance measurement. Comparing the experimental differential phase signal received from measurement (orange) and the theoretical differential phase signal generated by the simulation from section 3 (blue).
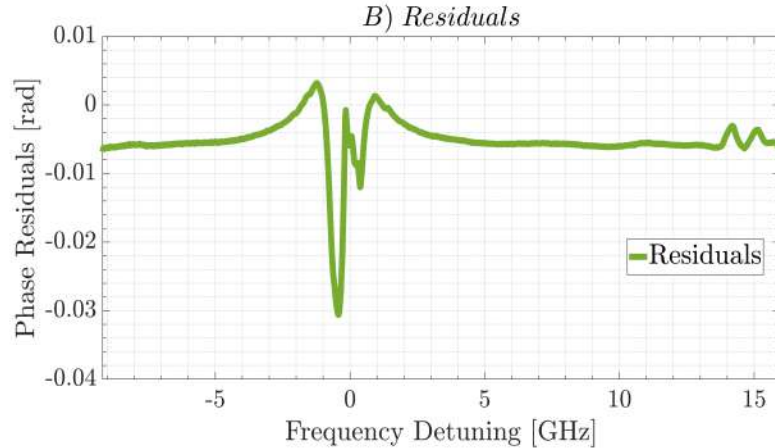
(b) The residuals between the simulation and the measurement in the HCN resonance area. An accuracy, given by the standard deviation of the residuals, of 4.2% is achieved.

Figure 5.4: Real signal validation. Comparison between measurement in HCN resonance area and corresponding simulation from section 3.

5.3   Simulated Metasurface Incorporation

As the setup is appropriately configured, the HCN vapor cell can be substituted with the metasurface. Initially, this replacement takes place digitally through simulation. For that, the simulated phase responses of two older versions of metasurfaces, obtained using the software Lumerical and depicted in plot 5.5a, were

employed due to the unavailability of the desired metasurface at that time. However, this limitation does not significantly affect the validity of the conclusions drawn from the simulations, as the results can be generally extrapolated to any metasurface. Following an initial analysis based on the simulation results, a physical metasurface can be introduced into the setup and subsequent measurements can be conducted.

The first metasurface used in the simulations exhibits a resonance frequency of approximately 1550nm, which is essential for compatibility with the setup utilizing a 1550nm laser, but a significantly wide undesired bandwidth. The second metasurface possesses a resonance frequency of about 1530nm and a slightly narrower bandwidth compared to the first one though it still retains an unusual and relatively broad bandwidth. While the wrong resonance frequency might seem problematic, it does not pose a major issue since the focus of the simulations lies in evaluating the influence of the metasurface's bandwidth. This means, to facilitate a better comparison, the frequency response of the second metasurface is assumed to be 1550nm, thereby aligning it with the same resonance frequency as the first metasurface. This adjustment allows for a more meaningful and insightful analysis between the two metasurfaces.

Moreover, due to the utilized laser's limited tuning range of 1nm, equivalent to 125GHz, the complete phase responses of both metasurfaces cannot be fully probed, as the laser tuning range covers only a fraction of the actual bandwidth of the metasurfaces.

It is important to note that the observed shapes of the phase responses differ from the theoretically anticipated shape discussed in section 1.2.2. This disparity arises due to the overlapping of two resonances in the metasurface, where one resonance is considerably narrower than the other. Without delving into intricate details, this occurrence is desired and attributable to the fact that the incident light couples more effectively with the broader resonance, while coupling is less efficient for a very narrow resonance. In practical terms, the incident light is initially coupled into the broader resonance of the metasurface, as it is relatively easier to achieve. Consequently, due to the overlap of the two phase responses, the mode coupled into the broader resonance is also partially coupled into the narrower resonance, leading to a sharper metasurface phase response.

Using MATLAB, the differential phase signals were derived from the phase responses, as shown in plot 5.5b. Additionally, the first derivatives of the differential phase signals were calculated and illustrated in plot 5.5c. The first derivative allows to identify the wavelength associated with the highest absolute slope value of the differential phase signal, indicated by the green or black square in both plot B and plot C. As discussed in section 2.2, this specific wavelength is important for detecting the concentration of the analyte in the final concentration measurement.

5.3.1 Minimal Detectable Resonance Shift

By considering the simulations and the system noise described in section 5.1, it is possible to predict the minimal detectable resonance shift of the metasurface when utilizing the system. For that, the noise limit NL of the system must be determined, which results from the squared mean $\mu^{(2)}$ of the given noise signal as this magnitude represents the power of a random signal. In order to align with previous considerations, where the ASD was used instead of the Power Spectral Density (PSD), the square root of the squared mean is utilized to represent the noise limit as depicted in Eq. 5.1.

$$NL = \sqrt{\mu^{(2)}} = \sqrt{\sigma^2 + \mu^2} \qquad (5.1)$$

Here, μ denotes the mean value of the statistical signal, while σ^2 represents its variance.

In determining the minimal detectable resonance shift, it is assumed that the mean of the noise signal is 0, despite the presence of an offset in the differential phase signal shown in Fig. 5.1b. This assumption is made based on the ergodicity of the system, which assumes that the statistical properties of the signal, involving its mean value, remain constant over time or different measurements, respectively. Consequently, when detecting the resonance shift, the mean value of the noise signal can be subtracted from the measurement and thus does not further influence it. As a result, the noise limit is simply defined by the standard deviation of the system's noise signal, leading to a noise limit of $NL = 0.2504$mrad for the final noise measurements from section 5.1.

The minimal detectable resonance shift Δf_{min} can then be calculated with the following expression (Eq. 5.2).
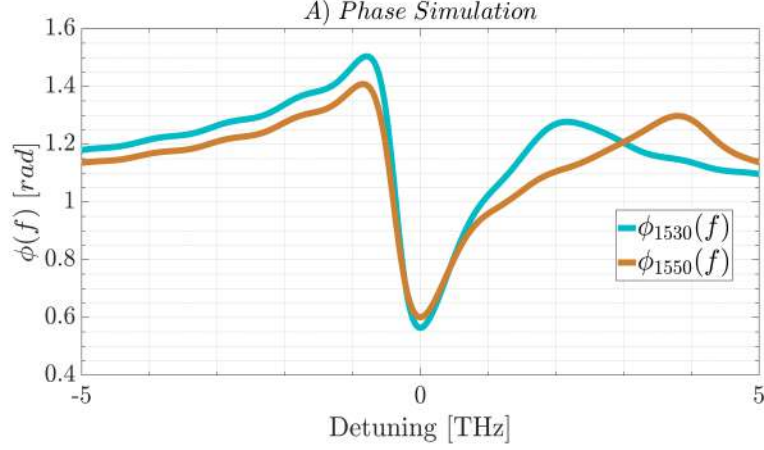
(a) Simulated phase response of the 1550nm and 1530nm metasurfaces. The deviation from the theoretical expectations described in section 1.2.2 can be attributed to the overlap between two resonances.
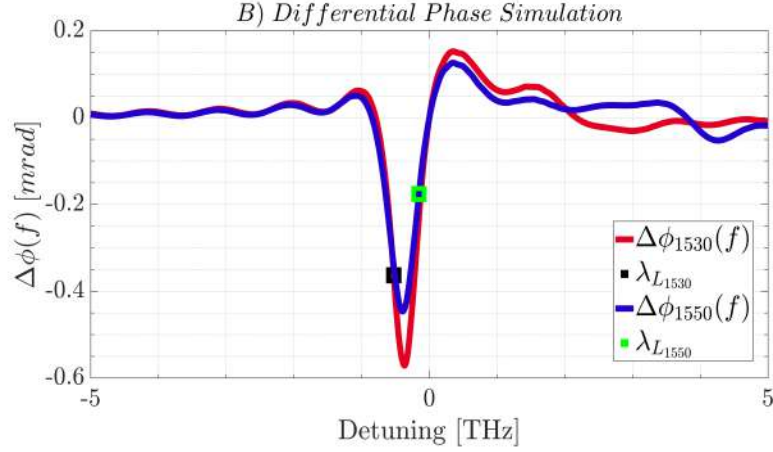
(b) Simulated differential phase signal of the 1550nm and 1530nm metasurfaces. The black or green square represent the highest absolute slope value of the signal and therefore the laser wavelength used for analyte detection (see section 2.2).
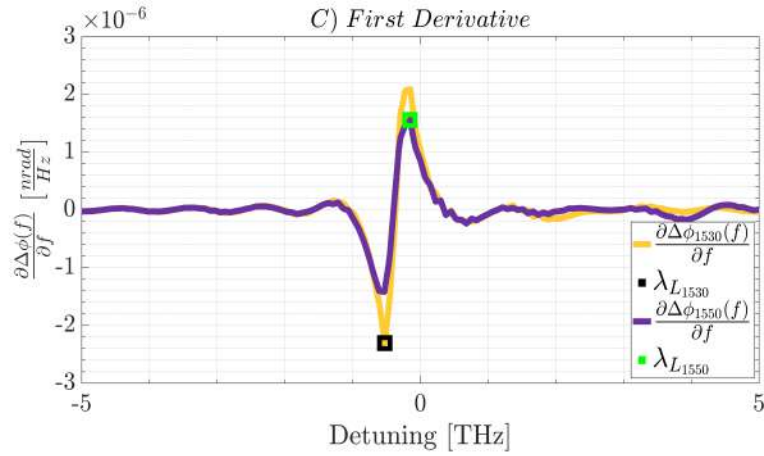
(c) First derivative of the differential phase signal of the 1550nm and 1530nm metasurfaces. The black and green square represent the maximal absolute value of the first derivative and thus the highest absolute slope value of the differential phase signal.

Figure 5.5: Metasurface simulations.

$$\Delta f_{min} = \frac{NL}{\Delta\phi_{dev}(f)_{max}} \tag{5.2}$$

where $\Delta\phi_{dev}(f)_{max}$ represents the maximal absolute value of the first derivative of the differential phase signal from Fig. 5.5c, depicted by the green or black square, respectively. As explained previously, this value corresponds to the laser wavelength used for the concentration measurement in the final stage.

Based on the maximal absolute slope values of the differential phase signals derived from the simulations, as shown in the second column of Table 5.1, the associated minimal detectable resonance shift for each metasurface is calculated as presented in the third and fourth column of the table, expressed in terms of frequency and wavelength.

Metasurface	$\Delta\phi_{dev}(f)_{max}$ $[\frac{nrad}{Hz}]$	Δf_{min} [GHz]	$\Delta\lambda_{min}$ [nm]
1550nm	1.5595×10^{-6}	160.5558	1.2867
1530nm	2.3046×10^{-6}	108.6436	0.8707

Table 5.1: Results of metasurface simulations.

The simulations demonstrate that the 1530nm metasurface outperforms the 1550nm metasurface in terms of sensitivity, which is attributed to the higher absolute slope. Given that the main goal of the project is to measure a resonance shift below 1nm, it is apparent that such a small shift would not produce a detectable signal for the 1550nm metasurface using the current system. Instead, it would be perceivable with a metasurface having the same bandwidth as the 1530nm one, but with a resonance wavelength at 1550nm. These results are particularly noteworthy given the 1530nm metasurface's unusual broad bandwidth, and the system's ability to predictably detect resonance shifts with such challenging characteristics. Hence, the simulation results showcase the system's sophisticated behavior, demonstrating its readiness for real metasurface measurements.

The theoretical results obtained from simulations can be experimentally proven by tuning the laser frequency across the metasurface's bandwidth and measuring the differential phase signal since the minimal detectable resonance shift directly corresponds to the smallest laser sweep essential for observing a differential phase signal on the phasemeter. This interconnection stems from the fact that any shift in the metasurface's frequency response is directly associated with a sweep in wavelength or frequency of the laser. Consequently, given the limitations of the 1550nm laser in this setup, which is restricted to a wavelength sweep of approximately 1nm, the theoretically predicted outcome is that the specified 1550nm metasurface would not generate a discernible signal on the phasemeter when sweeping the laser across its entire frequency range. A subsequent experimental confirmation, following the metasurface's integration into the system, yields an ambiguous differential phase signal that fails to resemble a small segment of the metasurface phase response illustrated in Fig. 5.5a. Therefore, to showcase the viability of metasurface-based dispersion spectroscopy, an alternative approach involves utilizing a laser capable of sweeping across a minimum range of 1.29nm, enabling the acquisition of a detectable differential phase signal through the use of the 1550nm metasurface, instead of employing a newly fabricated metasurface alongside the current laser configuration.

5.3.2   Optimal AOM Frequency Shift

Another parameter that can be obtained from the simulations is the optimal frequency shift induced by the AOMs. This shift is crucial for obtaining a maximal amplitude of the differential phase signal and, subsequently, maximizing the SNR.

According to the theoretical analysis, the phase response exhibits an S-shaped curve, as illustrated in various figures such as Fig. 4.7. In this scenario, the optimal frequency shift would be the largest possible, as the phase response is a monotonically decreasing function, and thus the serial AOM structure would always deliver the best result as already explained in section 4.3.5. However, the simulation results demonstrate that an S-shaped phase response is not always the case, as the phase response of the metasurface can vary depending on its design. Consequently, in some cases, a finite optimal frequency shift exists, which can be determined by simulating the differential phase signal for different frequency shifts and identifying the shift that yields the highest peak-to-peak value. This analysis generates a graph, shown in Fig. 5.6, which illustrates the peak-to-peak value of the differential phase signal as a function of the frequency shift induced by the AOMs.
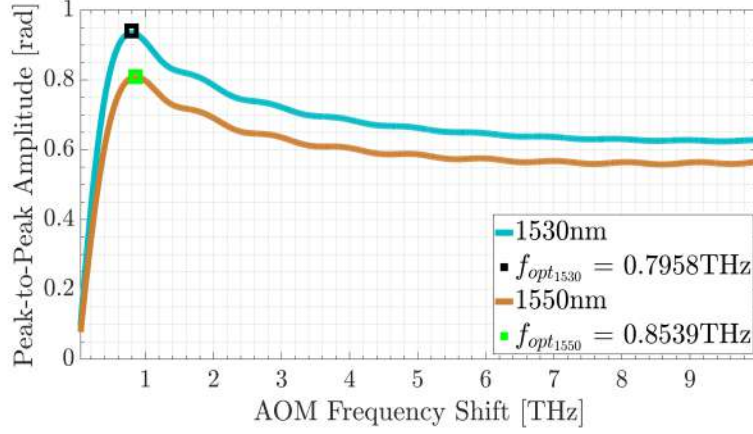
Figure 5.6: Optimal AOM induced frequency shift. The graphs show the peak-to-peak value of the differential phase signal as a function of the AOM frequency shift. The black and green square indicate the optimal shift which is 0.7958THz and 0.8539THz.

From the data presented in Fig. 5.6, it can be observed that the optimal frequency shift is 0.7958THz for the 1530nm metasurface and 0.8539THz for the 1550nm one. Considering that current AOMs are capable of inducing a maximum frequency shift of about 1.5GHz, it follows that the optimal frequency shift for this metasurface is still the maximum possible value. Consequently, for this specific metasurface, the serial AOM structure still proves to be the most suitable option since it consistently provides a larger frequency shift compared to the parallel or single structure, as demonstrated in section 4.3.5. However, if the metasurface exhibits a narrower bandwidth while maintaining a similar shape, the optimal frequency shift may not necessarily be achieved by employing the largest possible shift. In such cases, it is possible that the optimal frequency shift can also be attained using a parallel AOM structure instead of a serial one. Since the parallel structure exhibits superior noise performance compared to the serial AOM structure, as demonstrated both theoretically and experimentally in section 4.3, in scenarios where the metasurface possesses a narrower bandwidth, employing a parallel AOM structure may be more advantageous than using a serial structure.

Chapter 6

Conclusion and Future Work

In this project, initial steps were taken towards the development of a metasurface-based dispersion spectroscopy sensor which overcomes the limitations traditional dispersion spectroscopy techniques exhibit. These limitations include limited sensitivity, particularly towards specific analytes, as well as limited selectivity, design flexibility and compactness. However, the metasurface-based sensor retains the advantages of dispersion spectroscopy, such as linearity over a wide dynamic range, inherent immunity to optical power fluctuations, and improved SNR through the use of a differential phase shift instead of a single phase shift for retrieving the analyte information.

As a starting point, a metasurface-based dispersion spectroscopy architecture was set up and modified. The objective was to optimize the SNR of the system by boosting the amplitude of the differential phase signal and minimizing the system noise to the greatest extent possible. This led to the implementation of a unidirectional configuration for the metasurface architecture, complemented by an additional interferometer path, which was proven to be of utmost significance as it played a crucial role in reducing the overall system noise.

To further enhance the SNR, extensive efforts were made to identify and address potential noise sources within the new setup. Notably, the AOM was identified as a significant noise source, prompting the implementation of a serial AOM structure in place of a single AOM. The noise limit for the serial structure was measured to be $1.9 \frac{\mu\text{rad}}{\sqrt{\text{Hz}}}$, while for the parallel and single structure, it was measured to be $0.6 \frac{\mu\text{rad}}{\sqrt{\text{Hz}}}$ and $2.6 \frac{\mu\text{rad}}{\sqrt{\text{Hz}}}$. Therefore, despite its higher complexity, the metasurface setup with the parallel AOM structure achieved a noise limit that matches the inherent noise stability of the Sagnac interferometer approach, which served as the baseline model with a noise limit of $0.8 \frac{\mu\text{rad}}{\sqrt{\text{Hz}}}$, outperforming most dispersion spectroscopy methods. This achievement underscores the successful design and optimization of the metasurface setup, demonstrating its potential to compete with, and in some cases surpass, the noise performance of established and highly stable dispersion spectroscopy techniques. However, the decision to choose the serial AOM structure over the parallel one was based on the assumption that the considerably greater amplitude of the differential phase signal, achieved with the serial configuration, outweighs the slightly higher noise limit when considering the overall SNR. The low noise limit, combined with the metasurface's sensitivity, makes the given metasurface dispersion spectroscopy system likely to outperform many current state-of-the-art spectroscopy systems in terms of the LOD. The system sensitivity can be further improved by conducting measurements for a maximum time of 128s and averaging the results as shown by the Allan variance. Real signal validation involving measuring the resonance of HCN gas validated the system's reliability, as the measurements aligned well with simulations.

The integration of two metasurfaces into the system by simulation has demonstrated promising potential for detecting minimal resonance shifts below 1nm. This achievement is particularly remarkable given the fact that both metasurfaces retain undesired broad bandwidth characteristics. However, practically proving these results was not feasible, as the metasurface with the narrow enough bandwidth to achieve the desired performance had a resonance frequency that did not align with the frequency of the applied laser. Furthermore, the corresponding optimal AOM frequency shifts for these metasurfaces were predicted to be 0.8539THz and 0.7954THz. Current AOMs have a maximum frequency shift of about 1.5GHz, rendering this frequency shift too large. Nevertheless, fabricating a metasurface with a narrower bandwidth, easily achievable with current fabrication methods, is expected to significantly decrease the minimal detectable resonance shift as well as the optimal AOM frequency shift. If the bandwidth is narrow enough for the parallel AOM structure to achieve the optimal frequency shift, this structure might be a better choice due to its lower noise limit

compared to the serial structure.

In conclusion, the main goal of this project, which is to set up a system capable of detecting resonance shifts below 1nm, is potentially achieved.

The next crucial step in the project involves obtaining a properly designed metasurface that meets the specific requirements, including a sufficiently narrow bandwidth and the correct resonance frequency. Once this metasurface is available, the system can be tested on it to verify its performance. During testing, the goal is to observe a clear and distinguishable differential phase signal as the laser sweeps across its 1nm range. The simulations indicate that the bandwidth of the 1530nm metasurface should be adequate to capture a signal on the phasemeter.

After obtaining the reference curve for final analyte detection, depicted as the black curve in Fig. 2.5, the wavelength corresponding to the highest absolute slope value of the black curve needs to be identified. This wavelength will be used to achieve the highest SNR. However, since the newly fabricated metasurface is expected to have a much narrower bandwidth than the metasurfaces used for simulation, it should still be possible to detect a resonance shift even if the utilized laser wavelength deviates from the optimal value.

Finally, the first analyte concentration measurements can be conducted by attaching a known analyte concentration to the metasurface and illuminating it with the laser at the determined wavelength. Using the concept explained in 2.2, the resonance shift can be determined from the differential phase signal obtained during the measurement. To ensure accurate results, the mean value of the noise signal measured in section 5.1 should be subtracted from the received signal, as explained in section 5.3.1. This step helps avoid signal distortions and enhances the precision of the measurement.

In order to compare the system with other biosensing systems as a final step, the LOD serves as a proper parameter. This value will provide valuable insight into the sensitivity and efficacy of the metasurface-based dispersion spectroscopy sensor, potentially confirming its superiority over existing biosensing techniques.

Bibliography

[1] T. G. Mayerhöfer, S. Pahlow, and J. Popp, "The bouguer-beer-lambert law: Shining light on the obscure," *Chemphyschem : a European journal of chemical physics and physical chemistry*, vol. 21, no. 18, pp. 2029–2046, 2020.

[2] P. Martín-Mateos and P. Acedo, "Heterodyne phase-sensitive detection for calibration-free molecular dispersion spectroscopy," *Optics express*, vol. 22, no. 12, pp. 15 143–15 153, 2014.

[3] D. T. Cassidy and J. Reid, "Atmospheric pressure monitoring of trace gases using tunable diode lasers," *Applied optics*, vol. 21, no. 7, pp. 1185–1190, 1982.

[4] P. Kluczynski, J. Gustafsson, Å. M. Lindberg, and O. Axner, "Wavelength modulation absorption spectrometry — an extensive scrutiny of the generation of signals," *Spectrochimica Acta Part B: Atomic Spectroscopy*, vol. 56, no. 8, pp. 1277–1354, 2001.

[5] P. Werle, F. Slemr, K. Maurer, R. Kormann, R. Mücke, and B. Jänker, "Near- and mid-infrared laser-optical sensors for gas analysis," *Optics and Lasers in Engineering*, vol. 37, no. 2-3, pp. 101–114, 2002.

[6] G. C. Bjorklund, "Frequency-modulation spectroscopy: a new method for measuring weak absorptions and dispersions," *Optics letters*, vol. 5, no. 1, p. 15, 1980.

[7] P. Martín-Mateos, J. Hayden, P. Acedo, and B. Lendl, "Heterodyne phase-sensitive dispersion spectroscopy in the mid-infrared with a quantum cascade laser," *Analytical chemistry*, vol. 89, no. 11, pp. 5916–5922, 2017.

[8] T. G. Mayerhöfer, A. Dabrowska, A. Schwaighofer, B. Lendl, and J. Popp, "Beyond beer's law: Why the index of refraction depends (almost) linearly on concentration," *Chemphyschem : a European journal of chemical physics and physical chemistry*, vol. 21, no. 8, pp. 707–711, 2020.

[9] M. Nikodem and G. Wysocki, "Measuring optically thick molecular samples using chirped laser dispersion spectroscopy," *Optics letters*, vol. 38, no. 19, pp. 3834–3837, 2013.

[10] ——, "Molecular dispersion spectroscopy–new capabilities in laser chemical sensing," *Annals of the New York Academy of Sciences*, vol. 1260, no. 1, pp. 101–111, 2012.

[11] C. P. Bandutunga, Y. Zhang, T. G. McRae, M. B. Gray, and J. H. Chow, "Coherent rayleigh backscatter phase noise in digitally enhanced fiber interferometers," *Journal of Lightwave Technology*, vol. 39, no. 8, pp. 2625–2630, 2021.

[12] B. Canuel, E. Genin, G. Vajente, and J. Marque, "Displacement noise from back scattering and specular reflection of input optics in advanced gravitational wave detectors," *Optics express*, vol. 21, no. 9, pp. 10 546–10 562, 2013.

[13] D. J. Ottaway, P. Fritschel, and S. J. Waldman, "Impact of upconverted scattered light on advanced interferometric gravitational wave detectors," *Optics express*, vol. 20, no. 8, pp. 8329–8336, 2012.

[14] P. Werle, R. Mcke, and F. Slemr, "The limits of signal averaging in atmospheric trace-gas monitoring by tunable diode-laser absorption spectroscopy (tdlas)," *Applied Physics B Photophysics and Laser Chemistry*, vol. 57, no. 2, pp. 131–139, 1993.

[15] J. Wong, C. P. Bandutunga, P. Singh, M. B. Gray, and J. H. Chow, "Molecular dispersion spectroscopy using a code-division-multiplexed optical carrier," *Physical Review Applied*, vol. 19, no. 4, 2023.

[16] H. Altug, S.-H. Oh, S. A. Maier, and J. Homola, "Advances and applications of nanophotonic biosensors," *Nature nanotechnology*, vol. 17, no. 1, pp. 5–16, 2022.

[17] V. R. Samuel and K. Rao, "A review on label free biosensors," *Biosensors and Bioelectronics: X*, vol. 11, p. 100216, 2022.

[18] T. J. Cui, W. X. Tang, X. M. Yang, Z. L. Mei, and W. X. Jiang, *Metamaterials.* Boca Raton: Taylor & Francis, 2016. | "A CRC title.": CRC Press, 2017.

[19] M. Decker, I. Staude, M. Falkner, J. Dominguez, D. N. Neshev, I. Brener, T. Pertsch, and Y. S. Kivshar, "High-efficiency dielectric huygens' surfaces," *Advanced Optical Materials*, vol. 3, no. 6, pp. 813–820, 2015.

[20] B. I. Karawdeniya, A. M. Damry, K. Murugappan, S. Manjunath, Y. M. N. D. Y. Bandara, C. J. Jackson, A. Tricoli, and D. Neshev, "Surface functionalization and texturing of optical metasurfaces for sensing applications," *Chemical reviews*, vol. 122, no. 19, pp. 14 990–15 030, 2022.

[21] C. Liang, J. Lai, S. Lou, H. Duan, and Y. Hu, "Resonant metasurfaces for spectroscopic detection: Physics and biomedical applications," *Advanced Devices & Instrumentation*, vol. 2022, 2022.

[22] S. Tabassum, S. K. Nayemuzzaman, M. Kala, A. Kumar Mishra, and S. K. Mishra, "Metasurfaces for sensing applications: Gas, bio and chemical," *Sensors (Basel, Switzerland)*, vol. 22, no. 18, 2022.

[23] M. Decker and I. Staude, "Resonant dielectric nanostructures: a low-loss platform for functional nanophotonics," *Journal of Optics*, vol. 18, no. 10, p. 103001, 2016.

[24] M. L. Tseng, Y. Jahani, A. Leitis, and H. Altug, "Dielectric metasurfaces enabling advanced optical biosensors," *ACS Photonics*, vol. 8, no. 1, pp. 47–60, 2021.

[25] S. Arnold, M. Khoshsima, I. Teraoka, S. Holler, and F. Vollmer, "Shift of whispering-gallery modes in microspheres by protein adsorption," *Optics letters*, vol. 28, no. 4, pp. 272–274, 2003.

[26] F. Vollmer, D. Braun, A. Libchaber, M. Khoshsima, I. Teraoka, and S. Arnold, "Protein detection by optical shift of a resonant microcavity," *Applied Physics Letters*, vol. 80, no. 21, pp. 4057–4059, 2002.

[27] J. Wang, X.-Q. Xie, T. Hou, and X. Xu, "Fast approaches for molecular polarizability calculations," *The journal of physical chemistry. A*, vol. 111, no. 20, pp. 4443–4448, 2007.

[28] M. R. Hasan and O. G. Hellesø, "Dielectric optical nanoantennas," *Nanotechnology*, vol. 32, no. 20, p. 202001, 2021.

[29] B. Chocarro-Ruiz, A. Fernández-Gavela, S. Herranz, and L. M. Lechuga, "Nanophotonic label-free biosensors for environmental monitoring," *Current opinion in biotechnology*, vol. 45, pp. 175–183, 2017.

[30] H. Torun, B. Bilgin, M. Ilgu, N. Batur, M. Ozturk, T. Barlas, G. Guney-Esken, C. Yanik, S. Celik, O. Dogan, O. Ergonul, F. Can, I. Solaroglu, and M. C. Onbasli, "Rapid nanoplasmonic-enhanced detection of sars-cov-2 and variants on dna aptamer metasurfaces," *Advanced Devices & Instrumentation*, vol. 4, 2023.

[31] P.-E. Fournier, M. Drancourt, P. Colson, J.-M. Rolain, B. La Scola, and D. Raoult, "Modern clinical microbiology: new challenges and solutions," *Nature reviews. Microbiology*, vol. 11, no. 8, pp. 574–585, 2013.

[32] World Health Organization, "Food safety," 2022. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/food-safety

[33] M. Aladhadh, "A review of modern methods for the detection of foodborne pathogens," *Microorganisms*, vol. 11, no. 5, 2023.

[34] T. A. Ghebreyesus, J. Voegele, C. Russell, "State of the world's drinking water: An urgent call to action to accelerate progress on ensuring safe drinking water for all," *World Health Organization*, 2022. [Online]. Available: https://www.who.int/publications/i/item/9789240060807

[35] P. K. Pandey, P. H. Kass, M. L. Soupir, S. Biswas, and V. P. Singh, "Contamination of water resources by pathogenic bacteria," *AMB Express*, vol. 4, p. 51, 2014.

[36] J. Simões and T. Dong, "Continuous and real-time detection of drinking-water pathogens with a low-cost fluorescent optofluidic sensor," *Sensors (Basel, Switzerland)*, vol. 18, no. 7, 2018.

[37] M. L. Y. Sin, K. E. Mach, P. K. Wong, and J. C. Liao, "Advances and challenges in biosensor-based diagnosis of infectious diseases," *Expert review of molecular diagnostics*, vol. 14, no. 2, pp. 225–244, 2014.

[38] P. Wang, J. Lou, Y. Yu, L. Sun, L. Sun, G. Fang, and C. Chang, "An ultra-sensitive metasurface biosensor for instant cancer detection based on terahertz spectra," *Nano Research*, vol. 16, no. 5, pp. 7304–7311, 2023.

[39] World Health Organization, "Cancer," 2022. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cancer

[40] M. Bruehlmeier, U. Roelcke, P. Bläuenstein, J. Missimer, P. A. Schubiger, J. T. Locher, R. Pellikka, and S. M. Ametamey, "Measurement of the extracellular space in brain tumors using 76br-bromide and pet," *Journal of Nuclear Medicine*, vol. 44, no. 8, pp. 1210–1218, 2003. [Online]. Available: https://jnm.snmjournals.org/content/44/8/1210

[41] K. F. Ross and R. E. Gordon, "Water in malignant tissue, measured by cell refractometry and nuclear magnetic resonance," *Journal of microscopy*, vol. 128, no. Pt 1, pp. 7–21, 1982.

[42] E. Pickwell, A. J. Fitzgerald, B. E. Cole, P. F. Taday, R. J. Pye, T. Ha, M. Pepper, and V. P. Wallace, "Simulating the response of terahertz radiation to basal cell carcinoma using ex vivo spectroscopy measurements," *Journal of biomedical optics*, vol. 10, no. 6, p. 064021, 2005.

[43] B. Ferguson and X.-C. Zhang, "Materials for terahertz science and technology," *Nature materials*, vol. 1, no. 1, pp. 26–33, 2002.

[44] J. Wong, C. P. Bandutunga, Y. Zhang, M. B. Gray, and J. H. Chow, "Digitally enhanced molecular dispersion spectroscopy," *Optics letters*, vol. 45, no. 22, pp. 6290–6293, 2020.

[45] W. C. Swann and S. L. Gilbert, "Line centers, pressure shift, and pressure broadening of 1530-1560 nm hydrogen cyanide wavelength calibration lines," *Journal of the Optical Society of America B*, vol. 22, no. 8, p. 1749, 2005.

[46] B. E. A. Saleh and M. C. Teich, *Photonics*, third edition ed., ser. Wiley series in pure and applied optics. Hoboken, NJ: Wiley, 2019, vol. Part 2.

[47] Hilgevoord, Jan and Uffink, Jos, "The uncertainty principle (stanford encyclopedia of philosophy)," 2016. [Online]. Available: https://plato.stanford.edu/entries/qt-uncertainty/

[48] M. P. Silverman, *Probing the Atom: Interactions of Coupled States, Fast Beams, and Loose Electrons*. Princeton, NJ: Princeton University Press, 2000.

[49] G.-W. Truong, J. D. Anstie, E. F. May, T. M. Stace, and A. N. Luiten, "Accurate lineshape spectroscopy and the boltzmann constant," *Nature communications*, vol. 6, p. 8345, 2015.

[50] C. Hill, "Hitranonline - documentation: Units and definitions," 5/26/2023. [Online]. Available: https://hitran.org/docs/definitions-and-units/

[51] M. A. Zentile, J. Keaveney, L. Weller, D. J. Whiting, C. S. Adams, and I. G. Hughes, "Elecsus: A program to calculate the electric susceptibility of an atomic ensemble," *Computer Physics Communications*, vol. 189, pp. 162–174, 2015.

[52] L. Gianfrani, "Linking the thermodynamic temperature to an optical frequency: recent advances in doppler broadening thermometry," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 374, no. 2064, p. 20150047, 2016.

[53] K. A. Whittaker, J. Keaveney, I. G. Hughes, and C. S. Adams, "Hilbert transform: Applications to atomic spectra," *Physical Review A*, vol. 91, no. 3, 2015.

[54] M. Tröbs and G. Heinzel, "Improved spectrum estimation from digitized time series on a logarithmic frequency axis," *Measurement*, vol. 39, no. 2, pp. 120–129, 2006.

[55] G. Heinzel, A. Rüdiger, and R. Schilling, "Spectrum and spectral density estimation by the discrete fourier transform (dft), including a comprehensive list of window functions and some new flat-top windows," *Max Plank Inst*, vol. 12, 01 2002.

[56] A. Korpel, "Acousto-optics—a review of fundamentals," *Proceedings of the IEEE*, vol. 69, no. 1, pp. 48–53, 1981.

[57] C. Wolff, M. J. A. Smith, B. Stiller, and C. G. Poulton, "Brillouin scattering—theory and experiment: tutorial," *Journal of the Optical Society of America B*, vol. 38, no. 4, p. 1243, 2021.

[58] M. G. Noppe, "On the nonlinear theory of fabry–perot semiconductor lasers," *Laser Physics*, vol. 26, no. 5, p. 055004, 2016.

[59] N. L. Sharma, "Reflection, refraction, and multislit interference," *European Journal of Physics*, vol. 25, no. 1, pp. 1–5, 2004.

[60] K. Wilson and J. Walker, *Principles and techniques of biochemistry and molecular biology*, 7th ed. Cambridge: Cambridge University Press, 2010.

[61] S. Fishman, I. Guarneri, and L. Rebuzzini, "Stable quantum resonances in atom optics," *Physical review letters*, vol. 89, no. 8, p. 084101, 2002.

[62] M. Kerker, *Scattering of light and other electromagnetic radiation*, ser. Physical chemistry. New York: Elsevier Science, 2016.

List of Abbreviations

AOM Acousto-optic Modulator

ASD Amplitude Spectral Density

CEAS Cavity-Enhanced Absorption Spectroscopy

CLaDS Chirped Laser Dispersion Spectroscopy

FWHM Full Width at Half Maximum

HCN hydrogen cyanide

HPSDS Heterodyne Phase-Sensitive Dispersion Spectroscopy

HWHM Half Width at Half Maximum

IUPAC International Union of Pure and Applied Chemistry

LOD Limit of Detection

LSPR Localized Surface Plasmon Resonance

MS metasurface

NIST National Institute of Standards and Technology

POC point-of-care

SNR Signal-to-Noise Ratio

TLDAS Tunable Laser Direct Absorption Spectroscopy

VC vapor cell

WHO World Health Organization

WMS Wavelength Modulation Spectroscopy

List of Figures

List of Tables

Appendix A

Theoretical Background

This appendix imparts profound insights into the underlying physical processes within the project. It entails a thorough investigation of the interactions between electromagnetic waves and matter, with particular emphasis on the physics behind absorption and dispersion which are optical processes as a direct consequence of the fundamental physical phenomenon known as scattering.

A.1 Maxwell's Equations

Each electromagnetic phenomena can fundamentally be described by Maxwell's Equations. These equations were published in 1861 and 1862 by the Scottish scientist James Clerk Maxwell and form the foundation of classical electromagnetism. They provide a mathematical model for electronic, optical and radio technologies and basically describe how electric and magnetic fields are generated by charges, currents and changes of fields. These equations can be formulated in two different ways, as integral equations (Eq. A.1 - A.4) and differential equations (Eq. A.5 - A.8).

Integral Form

$$\oiint \vec{E} * d\vec{A} = \frac{1}{e_0} \iiint \rho dV \tag{A.1}$$

$$\oiint \vec{B} * d\vec{A} = 0 \tag{A.2}$$

$$\oint \vec{E} * d\vec{l} = -\frac{d}{dt} \iint \vec{B} * d\vec{A} \tag{A.3}$$

$$\oint \vec{B} * d\vec{l} = \mu_0 (\iint \vec{j} * d\vec{A} + e_0 \frac{d}{dt} \iint \vec{E} * d\vec{A})) \tag{A.4}$$

Differential Form

$$\nabla * \vec{E} = \frac{\rho}{\epsilon_0} \tag{A.5}$$

$$\nabla * \vec{B} = 0 \tag{A.6}$$

$$\nabla \times \vec{E} = -\frac{\partial \vec{B}}{\partial t} \tag{A.7}$$

$$\nabla \times \vec{B} = \mu_0 (\vec{j} + \epsilon_0 \frac{\partial \vec{E}}{\partial t}) \tag{A.8}$$

A.2 Maxwell's Equations in Matter

In the presence of matter, the direct applicability of Maxwell's Equations becomes limited, as it would require considering each charge in every atom. Therefore, an extension of Maxwell's Equations is necessary to acquire a suitable macroscopic depiction of electromagnetic fields in matter. To achieve this, two additional vector fields are introduced which are associated to the electric and magnetic field by Eq. A.9 and Eq. A.10, respectively.

$$\vec{D} = \epsilon_0 \epsilon_r \vec{E} = \epsilon \vec{E} = \epsilon_0 (1 + \chi_E) \vec{E} = \epsilon_0 \vec{E} + \vec{P} \tag{A.9}$$

$$\vec{H} = \frac{\vec{B}}{\mu_0 \mu_r} = \frac{\vec{B}}{\mu} = \frac{\vec{B}}{\mu_0 (1 + \chi_m)} = \frac{\vec{B}}{\mu_0 - \vec{M}} \tag{A.10}$$

The electric displacement, denoted as $\vec{D}$, incorporates the influence of polarization. Polarization arises as a response of the bound charges within the atoms of a medium to an external electric field. In essence, each atom in a dielectric material consists of a positively charged nucleus surrounded by a cloud of negatively charged electrons. Due to the attractive force between the electrons and the nucleus, the electrons cannot entirely depart from their positions, but they can be slightly shifted by the application of an electric field to the dielectric material. Consequently, the center of gravity of the negative charges does not align precisely with that of the positive charges. As a result, the charges establish dipoles and polarize other nearby dipoles, leading to an accumulation of dipoles within the dielectric. The resulting electric field generated by these dipoles opposes the external electric field, thereby diminishing the electric field strength within the dielectric material (Fig. A.1).
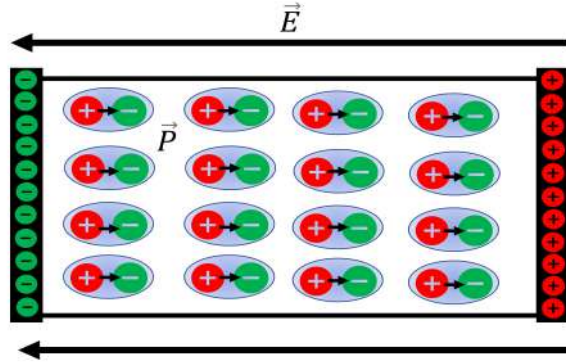
Figure A.1: Concept of polarization.

The dipole moment of one single atom or molecule induced by a local electric field $\vec{E}_{local}$ is defined according to Eq. A.11. It should be noted that the local electric field seen by a molecule is generally different from the macroscopic electric field $\vec{E}$ that would be measured externally.

$$\vec{p} = \epsilon_0 \alpha \vec{E}_{local} \tag{A.11}$$

The parameter α is called polarizability and describes the relative tendency of a charge distribution, like the electron cloud of an atom or molecule, to be distorted from its normal shape by a local electric field. Since always considering each dipole moment of each single atom or molecule in order to describe the material's behavior to an external electric field is quite unpractical, another quantity is introduced which is called polarization density and described by the vector field $\vec{P}$. The polarization density is defined for a certain volume element dV which carries a dipole moment dp (Eq. A.12) and can be seen as a parameter which takes the dipole moments of all electrons, induced by the external electric field, in the medium into account.

$$\vec{P} = \frac{d\vec{p}}{dV} \tag{A.12}$$

To enhance comprehension of the dielectric characteristics of a medium and the connection between polarization density and the applied electric field, it is helpful to conceptualize each medium as a fundamental system. In this system, the input signal is represented by the electric field, while the output signal is represented

Figure A.2: Basic relationship $\vec{E}(\vec{r}, t)$ and $\vec{P}(\vec{r}, t)$.

by the polarization density (or electric displacement) (Fig. A.2). This framework allows for a systematic examination of how the applied electric field influences the polarization behavior of the medium.

The behavior of the system, or more specifically, the medium, can be described by five key relationships between the electric field and the polarization density [46] which characterize the intricate connection between the applied electric field and the resulting polarization effects within the medium.

1) The dielectric medium is considered *linear* if the vector field $\vec{P}(\vec{r}, t)$ exhibits a linear relationship with the vector field $\vec{E}(\vec{r}, t)$. This linearity enables the principle of superposition to be applicable.

2) The medium is classified as *nondispersive* if its response to the applied electric field is instantaneous. However, nondispersiveness is an idealization as all physical systems have a finite response time.

3) The medium is deemed *homogeneous* if the relationship between the electric field and the polarization density is independent of the position $\vec{r}$ within the medium. In other words, the behavior remains consistent throughout the entire medium.

4) If the vectors $\vec{P}$ and $\vec{E}$ are parallel in the medium, it is considered *isotropic*. In this case, the relationship between $\vec{P}$ and $\vec{E}$ within the medium is direction-independent.

5) The medium is observed as *spatially nondispersive* if the relation between $\vec{P}$ and $\vec{E}$ is local, meaning that the polarization at a specific position $\vec{r}$ depends solely on the electric field at the same position.

For further considerations, to simplify the analysis, the system comprising the HCN vapor cell or metasurface is assumed to be homogeneous, isotropic, and spatially nondispersive. Consequently, for a more thorough examination of the relationship between the electric field and the polarization density, detailed considerations are only required for 1) and 2).

A.2.0.1 Linear, Nondispersive, Homogeneous, and Isotropic Medium

If the medium is *linear, nondispersive, homogeneous*, and *isotropic*, describing its properties becomes simplest. In such a case, the vectors $\vec{P}$ and $\vec{E}$ are parallel and have a consistent proportionality at every position and time [46]. The polarization density, as per Eq. A.9, is linked to the electric field through either the relative permittivity ϵ_r or the electric susceptibility χ_e (Eq. A.13), with both parameters being real scalar values (Fig. A.3).

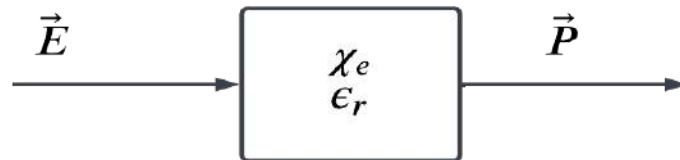
Figure A.3: Simplest relationship between $\vec{P}$ and $\vec{E}$.

$$\vec{P} = \epsilon_0 \chi_E \vec{E} = \epsilon_0 (\epsilon_r - 1) \vec{E} \tag{A.13}$$

The permittivity ϵ can be considered as the ability of a medium to resist to an electric field and avoid electric flux. The relative permittivity ϵ_r is a material's property and defined as the quotient of the electric permittivity ϵ of the dielectric and the permittivity ϵ_0 of the vacuum (Eq. A.14).

$$\epsilon_r = \frac{\epsilon}{\epsilon_0} \tag{A.14}$$

The electric susceptibility χ_e indicates the ability of a dielectric material to polarize in response to an applied electric field. The relationship to the permittivity results from the phenomena that if a dielectric material gets polarized, the material starts creating an internal field, oppose the external field, which will consequently reduce the electric flux within the material. Mathematically, the relationship is described by Eq. A.15.

$$\chi_e = \epsilon_r - 1 = \frac{\epsilon}{\epsilon_0} - 1 \tag{A.15}$$

In this simple case, the medium represents an LTI-system with the electric field $\vec{E}$ as input signal and the polarisation density $\vec{P}$ as output signal. Since the output signal occurs immediately after the input signal, the impulse response of the system is a dirac delta function $\delta(t)$ weighted by the parameter χ_e or ϵ_r (Eq. A.16), depending on $\vec{P}$ or $\vec{D}$ being the output signal.

$$h(t) = \chi_e \delta(t) \text{ or } h(t) = \epsilon_r \delta(t) \tag{A.16}$$

Comparing Eq. A.11 and Eq. A.13 indicates that there should also be a relationship between the electric susceptibility of a medium and the polarizability of the medium's atoms or molecules. If the local electric field $\vec{E}_{local}$ is parallel to the external electric field $\vec{E}$, which is the case for linear media, the comparison of both equations delivers Eq. A.17.

$$\chi_e = N\alpha \tag{A.17}$$

Since N is the number of molecules or atoms per unit volume, the electric susceptibility can be thought as the polarizability but for macroscopic fields instead of local fields. If the medium is not linear, the relationship between the electric susceptibility and the polarizability, is given, by the Clausius-Mossotti relation or the Lorentz-Lorenz equation.

A.2.0.2   Dispersive Medium

The focus of this project lies on dispersive media which are characterized by their inability to respond instantaneously to an input signal. The main property of a dispersive medium is that it does not response instantaneously to an input signal, which basically applies to all real world media. The time delay between the external electric field acting as the input signal and the resulting polarization density or electric displacement, which serves as the output signal, arises from the duration required for electrons to develop their dipole moment. As dispersive media ideally conform to the definition of Linear Time-Invariant (LTI) systems, the relationship between $\vec{P}$ and $\vec{E}$ can be effectively described by a linear differential equation of the form shown in Eq. A.18.

$$\vec{P}(t) = \frac{1}{a_0}\left(\sum_{m=0}^{M} b_m \frac{d^m \vec{E}(t)}{dt^m} - \sum_{n=1}^{N} a_n \frac{d^n \vec{P}(t)}{dt^n}\right) \tag{A.18}$$

As a result of the time-dependent nature of polarization and electric susceptibility, the description of the medium as a Linear Time-Invariant (LTI) system needs to be extended beyond the initial simple case depicted in Eq. A.16. Due to the delayed response of the system, the impulse response $h(t)$ is no longer a Dirac delta function. In theory, the electric field can be represented as a superposition of infinitely many weighted and shifted Dirac delta functions (Eq. A.19).

$$\vec{E}(t) = \int_{-\infty}^{\infty} \vec{E}(\tau)\delta(t - \tau)d\tau \tag{A.19}$$

Indeed, each Dirac delta function can be interpreted as an individual electric field impulse applied to the medium, leading to the induction of a corresponding polarization intensity as an output signal. As the medium is linear, the principle of superposition holds true. Therefore, the total polarization density $\vec{P}(t)$ is obtained by summing up all the individual output signals induced by the electric field impulses. Consequently,

the behavior of the system becomes more complex, involving a continuous distribution of impulse responses over time which can be mathematically represented as a convolution (Eq. A.20).

$$\vec{P}(t) = \epsilon_0 \int_{-\infty}^{\infty} \vec{E}(\tau)\chi_e(t-\tau)d\tau = \epsilon_0(\vec{E}(t) \circledast \chi_e(t)) \tag{A.20}$$

Therefore, it can be concluded that every linear, dispersive medium can be characterized as a Linear Time-Invariant (LTI) system, where the impulse response $h(t)$ is determined by the time-dependent electric susceptibility $\chi_e(t)$. In this system, the input signal is the electric field $\vec{E}$, and the output signal is the polarization density $\vec{P}$. The relationship between the input and output signals is represented by the convolution operation (Fig. A.4). As a result, the complete description of a linear, dispersive medium relies on the knowledge of its electric susceptibility $\chi_e(t)$.
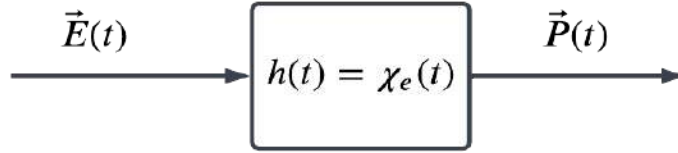

Figure A.4: Relationship between $\vec{P}$ and $\vec{E}$ for dispersive medium.

Alternatively, the medium can be characterized by its frequency response $H(\omega)$, which describes its behavior in the frequency domain. By applying the convolution theorem, Equation A.20 can be transformed into the frequency domain, resulting in Equation A.21.

$$\underline{\vec{P}}(\omega) = \epsilon_0\underline{\vec{E}}(\omega)\underline{\chi}_e(\omega) = \underline{H}(\omega)\underline{\vec{E}}(\omega) \tag{A.21}$$

The Fourier transform of $\chi_e(t)$, denoted as $\underline{\chi}_e(\omega)$, highlights its frequency-dependent nature. Hence, in the context of dispersive media, the electric susceptibility exhibits a frequency-dependent, complex value, which can either be described by its real amplitude and phase or by its real and imaginary part (Eq. A.22).

$$\underline{\chi}_e(\omega) = |\underline{\chi}_e(\omega)|e^{j\phi_{chi_e}(\omega)} = \Re\{\underline{\chi}_e(\omega))\} + j\Im\{\underline{\chi}_e(\omega)\} = \chi'_e(\omega) + j\chi''_e(\omega) \tag{A.22}$$

In accordance with Eq. A.21, the term $\epsilon_0\underline{\chi}_e(\omega)$ represents the frequency response of the medium when $\underline{\vec{E}}(\omega)$ is the input signal and $\underline{\vec{P}}(\omega)$ is the output signal (Eq. A.23).

$$\underline{H}(\omega) = \epsilon_0\underline{\chi}_e(\omega) = \frac{\underline{\vec{P}}(\omega)}{\underline{\vec{E}}(\omega)} = \epsilon_0|\underline{\chi}_e(\omega)|e^{j\phi_{\chi_e}(\omega)} \tag{A.23}$$

From Eq. A.23, it can be inferred that a phase difference $\phi_P(\omega) - \phi_E(\omega) = \phi_{\chi_e}(\omega)$ and an amplitude difference $|\underline{\vec{P}}(\omega)| - |\underline{\vec{E}}(\omega)| = \epsilon_0|\underline{\chi}_e(\omega)|$ generally exist between the polarization density and the electric field. The phase difference is determined by the phase of the electric susceptibility, while the amplitude difference is governed by the product of ϵ_0 and the real amplitude of the electric susceptibility.

The relationship between the electric field and the electric displacement can be obtained using a similar procedure as described in Eq. A.18 to Eq. A.21, resulting in Eq. A.24.

$$\underline{\vec{D}}(\omega) = \underline{\epsilon}(\omega)\underline{\vec{E}}(\omega) = \epsilon_0\underline{\epsilon}_r(\omega)\underline{\vec{E}} = \underline{H}(\omega)\underline{\vec{E}}(\omega) \tag{A.24}$$

The frequency-dependent permittivity $\underline{\epsilon}(\omega)$ is obtained as the Fourier transform of $\epsilon(t)$. Hence, similarly to the electric susceptibility, both the permittivity $\underline{\epsilon}(\omega)$ and the relative permittivity $\underline{\epsilon}_r(\omega)$ are complex values that vary with frequency.

It should be emphasized that both the phase difference and the amplitude difference between $\vec{P}(\omega)$ and $\vec{E}(\omega)$ or between $\vec{D}(\omega)$ and $\vec{E}(\omega)$, respectively, always depend on the frequency of the external electric field. Furthermore, from the Fourier transform it is known that each phase shift in the frequency domain corresponds to a time delay in the time domain (Eq. A.25).

$$x(t - a) \xrightarrow{\mathcal{F}} \underline{X}(\omega)e^{-j\omega a} \tag{A.25}$$

As elaborated in section A.2.3.2, this frequency-dependent time delay is a crucial factor for the phenomenon of dispersion. This phenomenon forms the foundation of dispersion spectroscopy, which can be employed to accurately detect analyte properties, including concentration.

The relationship between the magnetizing field $\vec{H}$ and the magnetic flux density $\vec{B}$ for magnetic materials can similarly be described as the relationship between the electric displacement $\vec{D}$ and the electric field $\vec{E}$. Moreover, the same applies for the magnetic susceptibility χ_m and the permeability μ which correspond to the electric susceptibility χ_e and the permittivity ϵ but for magnetic media. The magnetic behavior of a medium will not be discussed further.

A.2.1 Electromagnetic Waves

In this section, the description of electromagnetic waves will be explored, as optics primarily focuses on light. A convenient approach for understanding their fundamental behavior is by considering them as plane waves. For this type of waves it applies that the field vectors $\vec{E}$ of the electric field and $\vec{B}$ of the magnetic field are perpendicular to each other and have got the same phase. Additionally, they are also perpendicular to the direction of propagation whereby each plane wave only propagates in one direction z. Consequently, there is no field component in the direction of propagation.

It is assumed that the medium, in which the electromagnetic wave propagates, is linear, homogeneous and isotropic but may be nondispersive. This means, $\vec{E}$ and $\vec{H}$ are independent of their position $\vec{r}$ and direction but time-dependent. Moreover, the permittivity ϵ and the permeability μ are generally frequency-dependent and complex and the superposition principle applies. Furthermore, the waves are considered propagating in a source-free dielectric medium which does not develop electric currents and is far away from electric charges. In this case, it applies $\rho = 0$ and $\vec{j} = 0$.

With these assumptions, Maxwell's Equations in differential form (Eq. A.5 - Eq. A.8) become:

$$\nabla * \vec{E} = 0 \tag{A.26}$$

$$\nabla * \vec{H} = 0 \tag{A.27}$$

$$\nabla \times \vec{E} = -\mu \frac{\partial \vec{H}}{\partial t} \tag{A.28}$$

$$\nabla \times \vec{H} = \epsilon \frac{\partial \vec{E}}{\partial t} \tag{A.29}$$

Applying the ∇-operator to Eq. A.29 results in:

$$\nabla \times \nabla \times \vec{H} = \epsilon \nabla \times \frac{\partial \vec{E}}{\partial t} \tag{A.30}$$

Inserting Eq. A.28 into Eq. A.30 delivers:

$$\nabla \times \nabla \times \vec{H} = -\epsilon \mu \frac{\partial^2 \vec{H}}{\partial t^2} \tag{A.31}$$

Due to $\nabla \times \nabla \times \vec{H} = \nabla * \nabla * \vec{H} - \Delta \vec{H}$ and $\nabla * \vec{H} = 0$ (Eq. A.27) it follows:

$$\nabla \times \nabla \times \vec{H} = -\Delta \vec{H} = -\epsilon \mu \frac{\partial^2 \vec{H}}{\partial t^2} \tag{A.32}$$

Therefore, the following expression for the magnetic field is obtained (Eq. A.33).

$$\Delta \vec{H} - \epsilon \mu \frac{\partial^2 \vec{H}}{\partial t^2} = 0 \tag{A.33}$$

The same procedure from Eq. A.30 - Eq. A.33 can be done for the electric field (Eq. A.28) which results in the following expression (Eq. A.34).

$$\Delta \vec{E} - \epsilon \mu \frac{\partial^2 \vec{E}}{\partial t^2} = 0 \tag{A.34}$$

Eq. A.33 and Eq. A.34 are commonly referred to as the wave equations or d'Alembert equations. Considering the electric and magnetic field being harmonic waves, which is valid since any generic wave can be expressed as a sum of harmonic waves, the d'Alembert equations can be transformed into the frequency domain. This

transformation leads to the Helmholtz equations, where the fields are represented by the phasors $\underline{\vec{H}}$ and $\underline{\vec{E}}$, as given in Equation A.35 and Equation A.36.

$$\Delta \underline{\vec{H}} + \epsilon\mu\omega^2 \underline{\vec{H}} = 0 \tag{A.35}$$

$$\Delta \underline{\vec{E}} + \epsilon\mu\omega^2 \underline{\vec{E}} = 0 \tag{A.36}$$

In the case of plane waves, where only one propagation direction exists, the electric and magnetic fields are solely dependent on the propagation coordinate. Furthermore, due to the perpendicularity of the fields to each other and the direction of propagation, each field component represents a distinct vector component. In the case of a wave propagating along the z-direction, with the electric field oriented in the x-direction and the magnetic field in the y-direction, the Helmholtz equations can be simplified as follows (Equation A.37 and Equation A.38):

$$\frac{\partial^2 \underline{H}(z)}{\partial z^2} + \epsilon\mu\omega^2 \underline{H}(z) = \frac{\partial^2 \underline{H}(z)}{\partial z^2} + k^2 \underline{H}(z) = 0 \tag{A.37}$$

$$\frac{\partial^2 \underline{E}(z)}{\partial z^2} + \epsilon\mu\omega^2 \underline{E}(z) = \frac{\partial^2 \underline{E}(z)}{\partial z^2} + k^2 \underline{E}(z) = 0 \tag{A.38}$$

The simplified Helmholtz equations from Eq. A.37 and Eq. A.38 represent second-order differential equations. Solving these equations provides the solution for the electric field of the electromagnetic wave, given by A.39. Similarly, by replacing the electric field with the magnetic field, the solution for the magnetic field of the electromagnetic wave is obtained.

$$\underline{E}(z) = \underline{E}_1 e^{-jkz} + \underline{E}_2 e^{jkz} \tag{A.39}$$

With the wave number k:

$$k = \sqrt{\epsilon\mu\omega^2} = \omega\sqrt{\epsilon\mu} \tag{A.40}$$

Hence, the solution of Maxwell's Equations under specific restrictions is a superposition of two harmonic waves, propagating along one axis forward and backward. Since in most cases, the backward propagating wave is set to zero, i.e., $\underline{E}_2 = 0$, the final expression for a plane wave is (Eq. A.41):

$$\underline{E}(z,t) = \underline{E}(z)e^{j\omega t} = \underline{E}e^{-jkz}e^{j\omega t} = Ee^{j\phi}e^{jkz}e^{j\omega t} = Ee^{j(\omega t - kz + \phi)} \tag{A.41}$$

The parameter $\underline{E} = Ee^{j\phi}$ represents the complex amplitude, which consists of the real amplitude and the phase. The complex function $\underline{E}(z)$ is called the phasor of the wave and includes the complex amplitude and the dependence of propagation.

A.2.2 Light-Matter Interaction in Dielectric Media

As discussed in the previous section, electromagnetic waves are characterized by oscillating electric and magnetic fields propagating through space. In spectroscopy and optical sensing, the interaction between electromagnetic waves and matter plays a crucial role. This interaction primarily involves the interaction between the oscillating electromagnetic wave and the electrons within the atoms of the material.

The behavior of electrons in an isotropic material can be modeled as a damped harmonic oscillator, known as the Lorentz oscillator model, named after the Dutch physicist Hendrik Antoon Lorentz. This model is widely used to describe the dielectric behavior of materials, including the electric susceptibility χ_e and the electric permittivity ϵ, by establishing a relationship between the polarization density $\vec{P}$ and the electric field $\vec{E}$ from a microscopic perspective.

In classical mechanics, it is known that each harmonic oscillator or spring-mass-damping-system, respectively, can be described by a second-order differential equation (Eq. A.42).

$$\frac{d^2 x(t)}{dt^2} + \gamma \frac{dx(t)}{dt} + \omega_0^2 x(t) = \frac{F(t)}{m} \tag{A.42}$$

The parameter γ is called the damping coefficient and ω_0 the (angular) resonance frequency or the eigenfrequency of the system. The resonance frequency of an oscillator is defined as $\omega_0 = \sqrt{\frac{\kappa}{m}}$ where κ is the elastic constant and m the mass of the weight within the oscillating system. The input signal of this system is a

force $F(t)$ and the output signal is the position $x(t)$ of the mass.

In the Lorentz oscillator model, the motion of an electron is analogously described to a damped harmonic oscillator where m is the mass of an electron in the atom. The electron is modeled to be connected to the nucleus via a hypothetical spring, and its motion is damped by a hypothetical damper, which can be considered as the attracting force between the nucleus and the electron. The input signal or force $F(t)$ is the oscillating field of an electromagnetic wave, given by $E(t) = -\frac{F(t)}{e}$, where e represents the charge of the electron. The output signal is the position $x(t)$ of the electron, as described by Eq. A.43.

$$\frac{d^2\vec{x}(t)}{dt^2} + \gamma\frac{d\vec{x}(t)}{dt} + \omega_0^2\vec{x}(t) = \frac{\vec{F}(t)}{m} = -\frac{e}{m}\vec{E}(t) \tag{A.43}$$

Since the electric field is considered as an harmonic wave, Eq. A.43 can be transferred into the frequency domain which simplifies solving this second order differential equation (Eq. A.44).

$$(-\omega^2 + j\gamma\omega + \omega_0^2)\underline{\vec{X}}(\omega) = -\frac{e}{m}\underline{\vec{E}}(\omega) \tag{A.44}$$

The expression for the position $\underline{\vec{X}}(\omega)$ in the frequency domain therefore is (Eq. A.45):

$$\underline{\vec{X}}(\omega) = -\frac{e}{m(-\omega^2 + j\gamma\omega + \omega_0^2)}\underline{\vec{E}}(\omega) \tag{A.45}$$

In Eq. A.45, it is evident that when an electron interacts with an oscillating incident wave, it undergoes oscillatory motion at the same frequency as the incident wave, albeit with different amplitude and phase, which depend on the frequency of the wave. This interaction occurs when light passes through a medium and stimulates the electrons to oscillate, causing the charged particles to move and emit radiation in all directions at the same frequency as the incident wave, with each electron behaving like a point source. This phenomenon is called scattering and is fundamental to all optical processes such as reflection, refraction, diffraction, dispersion, polarization, fluorescence, and phosphorescence which are macroscopic manifestations of scattering effects at the atomic and subatomic levels [59]. The unique characteristics of each of these phenomena arise due to variations in the way electromagnetic waves are emitted from atoms as electrons drop to lower energy levels. Similarly to all second-order systems, the intensity of electron oscillations, and thus light scattering, is highest when the frequency of the incident wave is near the resonance frequency of the medium's electrons. Under these conditions, the intense electron oscillations result in interactions with adjacent atoms, leading to the conversion of vibrational energy into thermal energy which is also known as absorption.

From another point of view, when light is considered as photons, the mechanisms of scattering and absorption can be explained differently. When light interacts with matter, a photon is absorbed by an atom, leading to the temporary excitation of an electron to a higher energy level. Subsequently, the electron may be released, resulting in its relaxation to a lower energy level. During this de-excitation, the electron re-radiates the absorbed energy as an electromagnetic wave which is also accompanied with the emission of heat [60]. However, if the photons energy matches the difference between two atomic energy levels, the electron remains in the higher energy level without re-radiation and the absorbed energy is converted into thermal energy [61],i.e., the electromagnetic wave is absorbed. This occurs when the frequency of the light is near the resonance frequency, as given by the Planck-Einstein relation $E = \hbar\omega$, where E is the energy of the photon, ω is the angular frequency of the light, and $\hbar = \frac{h}{2\pi}$ is the reduced Planck's constant. So, even absorption could be viewed as a very unique form of scattering in which the atom does not emit electromagnetic radiation due to the missing relaxation of the electron.

Since this scattering effect fundamentally follows from the dipole moments of the electrons, it is directly related to the polarization density which is, according to Eq. A.12, the dipole moment per unit volume and thus can be determined by $\vec{\underline{P}}(\omega) = N\vec{\underline{p}}(\omega)$ with N being the density of charges per unit volume and $\vec{\underline{p}}(\omega)$ the dipole moment of one dipole. The dipole moment of two electronic charges with the charge $-e$ is defined as $\vec{p} = -e\vec{x(\omega)}$ where $x\vec{(\omega)}$ is the distance between the two charges. So, the polarization density $\vec{\underline{P}}(\omega)$ induced by the external electric field $\underline{\vec{E}}(\omega)$ is defined as (Eq. A.46):

$$\vec{\underline{P}}(\omega) = N\vec{\underline{p}}(\omega) = -Ne\underline{\vec{X}}(\omega) = \frac{Ne^2}{m(-\omega^2 + j\gamma\omega + \omega_0^2)}\underline{\vec{E}}(\omega) \tag{A.46}$$

Comparing Eq. A.46 with Eq. A.21, the susceptibility $\underline{\chi}_e(\omega)$ follows as (Eq. A.47)

$$\underline{\chi}_e(\omega) = \chi_0 \frac{\omega_0^2}{-\omega^2 + j\gamma\omega + \omega_0^2} = \chi_0 \frac{\omega_0^2(\omega_0^2 - \omega^2)}{(\omega_0^2 - \omega^2)^2 + (\gamma\omega)^2} - j\chi_0 \frac{\omega_0^2\omega\gamma}{(\omega_0^2 - \omega^2)^2 + (\gamma\omega)^2} \qquad (A.47)$$
$$= \chi'_e(\omega) + j\chi_e(\omega)''$$

with

$$\chi_0 = \frac{Ne^2}{\epsilon_0 m\omega_0^2} \qquad (A.48)$$

The electric susceptibility exhibits the form of a complex Lorentzian line shape which is unambiguously defined by its resonance frequency ω_0 and damping coefficient γ. The aforementioned relationships reveal the crucial role of the polarization density, particularly the electric susceptibility, in describing the fundamental process of light scattering and all other light-matter interactions in dielectric media and highlights the significance of understanding the electric susceptibility and its influence on light-matter interactions in dielectric media.

Fig. A.5 provides a graphical representation of $\underline{\chi}_e(\omega)$ with the parameter values $\omega_0 = 10\frac{1}{s}$, $\chi_0 = 1$, and $\gamma = 4\frac{Ns}{m}$ for the sake of simplicity. While these parameter values may not accurately reflect reality, this is not a concern as the main focus is on the shape of the curves. As expected, a second-order behavior of the electric susceptibility and thus the dispersive medium can be observed.
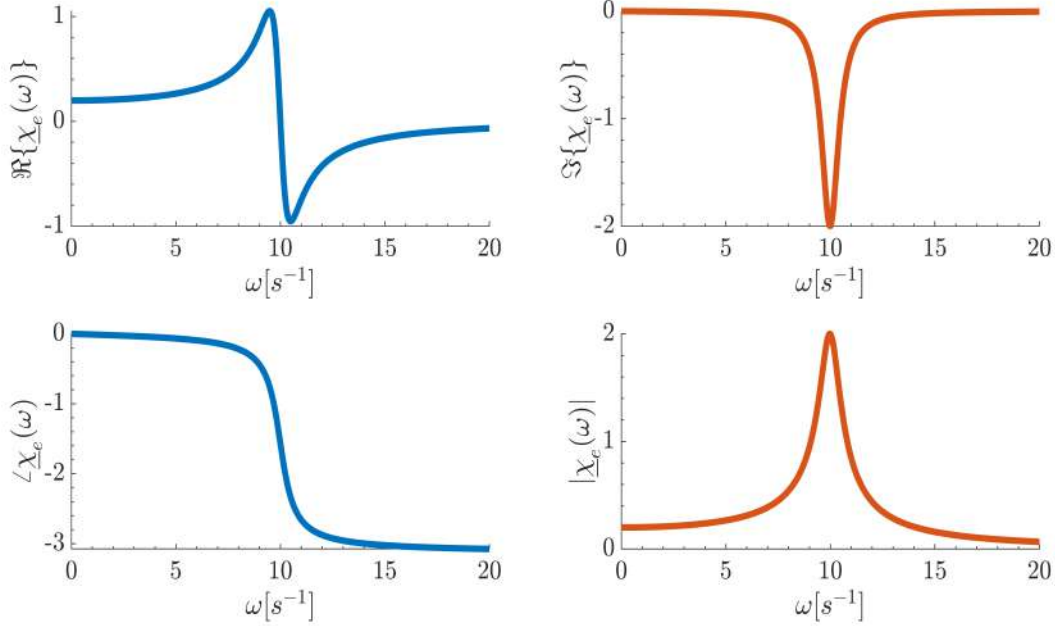

Figure A.5: Real and imaginary part and phase and absolute value of $\underline{\chi}_e(\omega)$ for the parameter values $\omega_0 = 10\frac{1}{s}$, $\chi_0 = 1$ and $\gamma = 4\frac{Ns}{m}$.

The graphics reveal several key insights about the susceptibility. First, at the resonance frequency ω_0, the real part of the susceptibility approaches zero while the imaginary part reaches its maximum value. Second, the real part of the susceptibility is significant across frequencies near and below resonance, while the imaginary part is limited to frequencies near resonance. Third, the susceptibility diminishes for frequencies well above the resonance frequency, and approaches χ_0 as ω approaches zero. The phase of the susceptibility ranges from zero to $-\pi$ and reflects a phase shift or a time delay between the electric field $\vec{E}$ and the polarization $\vec{P}$ or input and output signal of the medium. The magnitude of the susceptibility reaches its maximum at the resonance frequency, indicating that the polarization and energy storage in the material are most substantial at that frequency. As a result, absorption is highest at the resonance frequency.

A.2.2.1 The Refractive Index

In order to describe and characterize the optical properties of a medium, another parameter is defined which follows from the susceptibility and permittivity and is known as refractive index n. For non-magnetic materials (i.e., $\mu = \mu_0$), the refractive index is defined as a complex and frequency dependent value due to the reliance on the electric susceptibility (Eq. A.49).

$$\underline{n}(\omega) = \sqrt{\frac{\epsilon(\omega)}{\epsilon_0}} = \sqrt{1 + \underline{\chi}_e(\omega)} = n'(\omega) + jn''(\omega) \tag{A.49}$$

The real part of the refractive index indicates the speed of light in a medium as given by Eq. A.50 whereas the imaginary part, resulting from the imaginary part of the electric susceptibility, signifies the absorption of light by the medium (see A.2.3.1).

$$c(\omega) = \frac{c_0}{n'(\omega)} \tag{A.50}$$

In Eq. A.50, it is apparent that the speed of electromagnetic waves within a medium varies depending on their frequency. As the real part of the refractive index follows from the real part of the electric susceptibility, this variation in speed is a result of the polarization or alignment of dipole moments in the atoms of the medium induced by the external field. When electromagnetic radiation interacts with matter, the energy from the external electric field is temporarily stored in the medium as the electrons in the material align with the field. This stored energy is then re-radiated, but the beam's travel is slowed down due to the interaction with the material. From a different perspective, the electric field observed after passing through the medium is a superposition of the incident electric field and the electric field generated by the oscillating charges in the medium [62]. This superposition leads to a phase shift in the resulting electric field compared to the original electric field, causing the electromagnetic wave to appear to slow down in the medium. This phase shift or reduced speed is described by the real part n' of the refractive index.

In the frequency domain, the speed difference manifests as a phase difference between the polarization and the electric field (Eq. A.23), represented by the phase of the electric susceptibility $\underline{\chi}_e(\omega)$. This is because, as per Eq. A.25, each phase shift in the frequency domain corresponds to a time delay in the time domain. This time delay is responsible for the different propagation speed of the electromagnetic wave compared to the speed of light in vacuum.

Fig. A.6 shows the refractive index as a function of frequency with the same values as used for the electric susceptibility.
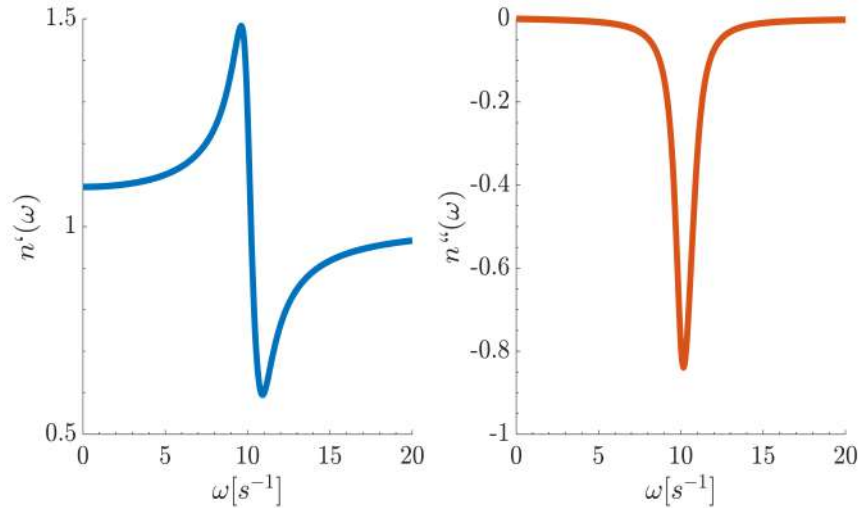

Figure A.6: Refractive index $\underline{n}(\omega)$ for the parameter values $\omega_0 = 10\frac{1}{s}$, $\chi_0 = 1$ and $\gamma = 4\frac{Ns}{m}$.

The real part of the refractive index, determined by the real part of the electric susceptibility, and the imaginary part of the refractive index, influenced by the imaginary part of the electric susceptibility, exhibit similar shapes. This similarity arises due to their inherent connection. The refractive index holds significant

importance in understanding the phenomena of absorption and dispersion, which will be discussed in detail in the subsequent section.

Note: In this thesis, the denomination "refractive index" always refers to its real part $n'(\omega)$.

A.2.3 Absorption and Dispersion

In section A.2.2, it has been elucidated that the interaction between light and matter can be explained by the interaction between waves and electrons in the medium, resulting in absorption and scattering phenomena. Optical sensing techniques utilize the reliance of these light-matter interactions on the characteristics and properties of the medium. By exploiting the principles of absorption and scattering the properties of the medium under study can be analyzed and investigated. Of particular importance for this project are absorption and especially dispersion, with dispersion being a resulting phenomenon of scattering (see A.2.3.2). The mathematical descriptions of these two phenomena will be discussed in the following sections.

The mathematical foundation for both phenomena is established by Eq. A.41, as well as Eq. A.21 and A.24, which indicate that the electric susceptibility and permittivity are characterized by frequency-dependent complex values. As a result, the wavenumber k, as defined in Equation A.40, is also generally complex and frequency-dependent due to its reliance on the permittivity and susceptibility. For nonmagnetic materials, where $\mu = \mu_0$, it applies

$$\underline{k}(\omega) = \omega\sqrt{\underline{\epsilon}(\omega)\mu_0} = \omega\sqrt{\epsilon_0(1+\underline{\chi}_e(\omega))\mu_0} = k_0\sqrt{1+\underline{\chi}_e(\omega)} = \beta(\omega) - j\frac{1}{2}\alpha(\omega) \tag{A.51}$$

with $k_0 = \frac{\omega}{c_0}$ as the wavenumber in vacuum.
Inserting Eq. A.51 into Eq. A.41 delivers Eq. A.52

$$\underline{E}(z,t) = Ee^{j(\omega t - \underline{k}z + \phi)} = Ee^{j(\omega t - (\beta - j\frac{1}{2}\alpha)z + \phi)} = Ee^{-\frac{1}{2}\alpha z}e^{j(\omega t - \beta z + \phi)} \tag{A.52}$$

where the value of $\underline{k}$ and thus the values α and β depend on the frequency ω of the given signal $\underline{E}(z,t)$. The real time signal is obtained by taking the real part of the complex signal (Eq. A.53).

$$E(z,t) = \Re\{\underline{E}(z,t)\} = Ee^{-\frac{1}{2}\alpha z}cos(\omega t - \beta z + \phi) \tag{A.53}$$

A.2.3.1 Absorption

As a result of the imaginary part of $\underline{k}(\omega)$, the plane wave undergoes a change in magnitude or rather real amplitude by traveling trough a medium. For $\alpha < 0$, the real amplitude of the plane wave is attenuated by the factor $e^{-\frac{1}{2}\alpha}$ so that the intensity of the plane wave, which is proportional to the squared real amplitude, is attenuated by $(e^{-\frac{1}{2}\alpha})^2 = e^{-\alpha}$. This describes the phenomenon called absorption. The parameter α is recognized as the absorption coefficient of the medium, is characteristic for each medium and related to the electric susceptibility and refractive index as following (Eq. A.54).

$$\alpha(\omega) = -2k_0\Im\{\sqrt{1+\underline{\chi}_e(\omega)}\} = -2k_0n''(\omega) \tag{A.54}$$

Since the absorption coefficient and imaginary part of the refractive index are linearly related, the curve of the absorption coefficient over the frequency aligns with that of the imaginary part of the refractive index (Fig. A.6) where the resonance frequency was set to $\omega_0 = 10s^{-1}$. This confirms the Lorentz oscillator model's prediction that electromagnetic wave absorption is at its maximum at the resonance frequency. This is because the photon energy at this frequency matches the difference between two atomic energy levels, resulting in a higher probability of photon absorption. For frequencies significantly above or below the resonance frequency, a medium becomes effectively non-absorptive.
Furthermore, this direct relationship between absorption and susceptibility or permittivity can also be observed in Eq. A.21 and Eq. A.24. In both equations, the absolute value of the susceptibility or permittivity represents the amplitude attenuation or amplification between the polarization $\vec{\underline{P}}(\omega)$ and the electric field $\vec{\underline{E}}(\omega)$ or between the displacement current $\vec{\underline{D}}(\omega)$ and the electric field $\vec{\underline{E}}(\omega)$. This indicates that these parameters must also be directly related to absorption.

A.2.3.2 Dispersion

The real part of $\underline{k}(\omega)$ is called the propagation constant $\beta(\omega)$ and is the rate at which the phase of the wave changes in the direction of propagation. The frequency-dependence of the propagation constant is the reason for the phenomenon called dispersion which is in optics the frequency-dependence of the phase velocity of an electromagnetic wave.

The relationship between the propagation constant and the susceptibility or the refractive index is indicated by Eq. A.55.

$$\beta(\omega) = k_0 \Re\{\sqrt{1 + \underline{\chi}_e(\omega)}\} = k_0 n'(\omega) \tag{A.55}$$

Thus, comparing Eq. A.50 and Eq. A.55, the propagation constant influences the speed of light in a medium as (Eq. A.56)

$$c(\omega) = \frac{c_0 k_0}{\beta(\omega)} \tag{A.56}$$

Similar to the absorption coefficient and imaginary part of the refractive index, the propagation constant is linearly related to the real part of the refractive index, as shown in Eq. A.55. Therefore, the curve of the propagation constant over frequency displays the same pattern as that of the real part of the refractive index in Fig. A.6. At frequencies slightly above or below the resonance frequency, the dispersion effect is most pronounced. At significantly higher or lower frequencies, the propagation constant remains relatively constant, resulting in an approximately non-dispersive medium. The propagation constant never becomes 1 for frequencies below the resonance frequency, but converges to 1 for frequencies above it. According to Eq. A.56, this indicates that the speed of electromagnetic waves in the medium is almost as fast as in a vacuum for frequencies much higher than the resonance frequency, while for lower frequencies, the speed is significantly slower than in vacuum. Therefore, there is always a phase shift for frequencies below resonance, but almost no phase shift for frequencies much higher than resonance.

This frequency dependent phase shift plays a critical role in dispersion spectroscopy techniques, as it is a unique characteristic of the medium and its properties, and causes the electromagnetic waves to have different phase speeds depending on the frequency and medium, resulting in a time delay between the incoming and outgoing waves.

The phase shift can be determined by considering two electromagnetic waves with the same frequency ω. The one wave is propagating through a medium ($\underline{E}_1(z,t)$, Eq. A.57), the other one is propagating through vacuum ($\underline{E}_2(z,t)$, Eq. A.58). Consequently, both waves have got different wavenumbers $\underline{k}(\omega)$. Absorption is neglected which means the wavenumber k is a real value $\beta(\omega)$. Since both waves have the same frequency, it is sufficient to only consider the phasors.

$$\underline{E}_1(z) = E_1 e^{-j\beta(\omega)z} = E_1 e^{-j\frac{\omega n'(\omega)}{c_0}z} \tag{A.57}$$

$$\underline{E}_2(z) = E_2 e^{-jk_0 z} = E_2 e^{-j\frac{\omega n_0}{c_0}z} = E_2 e^{-j\frac{\omega}{c_0}z} \tag{A.58}$$

The phase of the second wave, representing the wave propagating in vacuum, is considered as reference phase which means subtracting the phase of $\underline{E}_2(z)$ from the phase of $\underline{E}_1(z)$ delivers the phase change the wave receives when propagating through the medium described by the real part of the refractive index $n'(\omega)$ (Eq. A.59).

$$\phi(z,\omega) = -\frac{\omega n'(\omega)}{c_0}z - (-\frac{\omega}{c_0}z) = -\frac{\omega z}{c_0}(n'(\omega) - 1) \propto n'(\omega) \tag{A.59}$$

The linear relationship between the phase shift and the real part of the refractive index, $n'(\omega)$, is of great importance in dispersion spectroscopy, as will be discussed in section 1.1.2. The similarity in shape between the plot of the phase shift over frequency ω (Fig. A.7) and the plot of the real refractive index (Fig. A.6) is not surprising due to this proportionality.

Based on Eq. A.59, it can actually be demonstrated that dispersion is a consequence of scattering as it was claimed in section A.2.2. According to Eq. A.59, the electromagnetic wave after the medium can be described as (A.60):

$$\underline{E}_{after}(z) = E_0 e^{-j(k_0 z + \phi(z,\omega))} = E_0 e^{-j(\frac{\omega}{c_0}z - \frac{\omega}{c_0}z(n'(\omega)-1))} = E_0 e^{-j\frac{\omega}{c_0}z} e^{-j\frac{\omega}{c_0}z(n'(\omega)-1)} \tag{A.60}$$
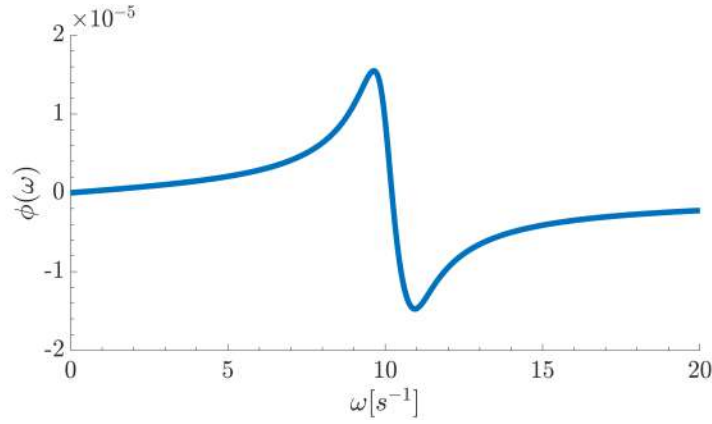
Figure A.7: Phase shift $\phi(\omega)$ for the parameter values $\omega_0 = 10\frac{1}{s}$, $\chi_0 = 1$, $\gamma = 4\frac{Ns}{m}$ and $z = 1m$.

Using the expression $e^x \approx 1 + x$, which means $e^{-j\omega\frac{z}{c_0}(n'(\omega)-1)} \approx 1 - \frac{j\omega(n'(\omega)-1)z}{c_0}$, Eq. A.60 can also be formulated as (Eq. A.61):

$$\underline{E}_{after}(z) = E_0 e^{-j\frac{\omega}{c_0}z} e^{-j\frac{\omega}{c_0}z(n'(\omega)-1)} = E_0 e^{-j\frac{\omega}{c_0}z} - \frac{j\omega(n'(\omega)-1)z}{c_0}E_0 e^{-j\frac{\omega}{c_0}z}$$

$$= \underline{E}_{before}(z) + \underline{E}_{medium}(z)$$

(A.61)

In Eq. A.61, the electric field after the medium is expressed as a superposition of the initial electric field incident on the medium $\underline{E}_{before}(z)$ and the electric field in the medium $\underline{E}_{medium}(z)$, which results from the scattering of the oscillating electrons induced by the external electric field and represents the polarization. The second summand of the equation, representing the scattering of the electrons, is characterized by the complex factor $-\frac{j\omega(n'(\omega)-1)z}{c_0}$, which causes the initial electromagnetic wave to undergo a negative phase shift. This indicates that the polarization, or rather the interaction of the electromagnetic wave with the electrons, leads to the phase shift or time delay of the incoming electric field. Additionally, since the complex factor of the second summand is frequency-dependent, the phase shift is also frequency-dependent, which confirms previous considerations.