UNIVERSITY OF SOUTHAMPTON

Faculty of Engineering and Physical Sciences
Zepler Institute

Integrating Terahertz Time-Domain Spectroscopy with Microfluidic Platforms and Machine Learning for Protein Hydration Studies

by

Nicholas Thomas Klokkou

MPhys

ORCiD: 0000-0002-0999-3745

A thesis for the degree of
Doctor of Philosophy

October 2022

University of Southampton

Abstract

Faculty of Engineering and Physical Sciences
Zepler Institute

Doctor of Philosophy

Integrating Terahertz Time-Domain Spectroscopy with Microfluidic Platforms and Machine Learning for Protein Hydration Studies

by Nicholas Thomas Klokkou

Terahertz Time-Domain Spectroscopy (THz-TDS) has been widely adopted as a technique of choice for sub-millimetre research over the last three decades, thanks to its high signal-to-noise ratio (SNR), broad bandwidth and straightforward phase retrieval. It allows the probing of picosecond dynamics of biomolecules, transient processes in semiconductors, imaging for security and medical applications, among many others. By measuring the amplitude of the electric field, the phase of the spectral information is preserved, and so complex parameters are easily extracted. However, manual preprocessing of the data is still required and iterative, numerical fitting of the material parameters is often needed, resulting in complexity, loss of accuracy and inconsistencies between measurements. The difficulties associated with analysing samples exhibiting strong terahertz absorption only compound with the aforementioned challenges. Aqueous solutions of proteins and their the surrounding water, or protein hydration shell, are uniquely probable with terahertz radiation, if this attenuation of the water can be addressed. This thesis aims to tackle both of these challenges. The use of machine learning techniques for interpreting spectroscopic THz-TDS data is described, achieved by training artificial neural networks (ANNs) with large data sets of simulated light-matter interactions, resulting in a computationally efficient method for material parameter extraction. The trained model improves on the accuracy of analytical methods that need approximations while being easier to implement and faster to run than iterative root-finding methods. Furthermore, a terahertz compatible, surface tension confined polytetrafluoroethylene (PTFE) based microfluidic flow cell is presented. THz-TDS measurements of a range of concentrations of aqueous bovine serum albumin (BSA) are shown to demonstrate the device's efficacy of probing protein hydration dynamics in a transmission configuration. The combination of the machine learning techniques and the microfluidic device comprised of such a chemically inert material presented here offer new analytical and practical techniques for the probing of notoriously difficult samples.

Contents

List of Figures

Declaration of Authorship

I declare that this thesis and the work presented in it is my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;

7. Parts of this work have been published as:
   N. T. Klokkou, D. J. Rowe, B. M. Bowden, N. P. Sessions, J. J. West, J. S. Wilkinson, and V. Apostolopoulos, "Structured surface wetting of a ptfe flow-cell for terahertz spectroscopy of proteins," *Sensors and Actuators B: Chemical*, vol. 352, p. 131003, 2022
   N. Klokkou, J. Gorecki, J. S. Wilkinson, and V. Apostolopoulos, "Artificial neural networks for material parameter extraction in terahertz time-domain spectroscopy," *Opt. Express*, vol. 30, pp. 15583–15595, Apr 2022
   J. Gorecki, N. Klokkou, L. Piper, S. Mailis, N. Papasimakis, and V. Apostolopoulos, "High-precision thz-tds via self-referenced transmission echo method," *Appl. Opt.*, vol. 59, pp. 6744–6750, Aug 2020
   N. T. Klokkou, D. J. Rowe, B. M. Bowden, N. P. Sessions, J. J. West, J. S. Wilkinson, and V. Apostolopoulos, "Thz spectroscopy of bsa in a surface-tension confined flow-cell," in *2021 46th International Conference on Infrared, Millimeter and Terahertz Waves (IRMMW-THz)*, pp. 1–2, 2021
   N. Klokkou, J. Gorecki, and V. Apostolopoulos, "Thz-tds parameter extraction via machine learning," in *2021 46th International Conference on Infrared, Millimeter and Terahertz Waves (IRMMW-THz)*, pp. 1–2, 2021

Signed:..... Date:.....

Acknowledgements

If I were to list all the people and the ways in which they have shown me the support and encouragement that made this PhD possible over the years, this section would prove longer than the thesis itself. I of course must thank my family, both immediate and extended for their love and support. My closest friends both from Kent and my time at university, who have always remained loyal even when I would all too frequently vanish into my studies. I am ever grateful for the friendship and guidance from those in the research groups of which I have been lucky to be a part of: the Terahertz and Lasers Laboratories, the Integrated Photonic Devices group and also the Polaritonics Group, who accepted me as an honorary member. From each of these I have friendships that will prove lifelong. I wish to thank my undergraduate tutor, Pavlos Lagoudakis, who continues to be a mentor and source of encouragement and motivation. I cannot overstate my appreciation for the unwavering support, patience, understanding and encouragement of Vasilis Apostolopoulos and James Wilkinson, who have been incredible supervisors throughout this PhD. I feel truly lucky to have been surrounded by so many wonderful people, thank you all.

'If you can't be a highway then just be a trail,
If you can't be the sun be a star;
It isn't by size that you win or you fail -
Be the best of whatever you are!'
- Uncle Malcolm

Chapter 1

Introduction

The terahertz band is generally defined as the frequency range of the electromagnetic spectrum between 100 GHz and 10 THz, a region that is often referred to as the *terahertz gap* due to the historic difficulty in generating and detecting radiation in this band. The challenges associated with targeting these frequencies are due to the limits of more widely established technologies for probing electromagnetic radiation on either side of the spectrum. The frequencies in the terahertz gap are too high for the electronics of radio-wave and microwave technologies, while the photon energies are too low (especially when background, thermal radiation is factored in) for band transitions in semiconductors to be useful. Terahertz Time-Domain Spectroscopy (THz-TDS) has proved itself as an excellent technique for unmasking this void in the electromagnetic spectrum for researchers over the last few decades and has become a staple for many laboratories around the world. However, challenges associated with this technique are still present, especially with regard to the analysis of data, and the nature of probing physical phenomena in highly absorbing media, such as water. One such phenomenon is the collective motions of proteins that exist on the picosecond timescale, and their interaction with the surrounding solvent. Obtaining the required measurements for these studies, however, requires multiple concentrations to be used, and micron scale path lengths. Further, the variation in terahertz measurements seen between research laboratories adds a further layer of variance to the measures and claims that are made. This thesis addresses these in two key ways. The first, is the development of material parameter extraction techniques with the use of machine learning, promising to increase reliability and ease of data analysis. The second is via the design and fabrication of a novel microfluidic flow cell, that has the advantages associated with its exploitation of surface tension forces, which is a unique solution to confining water to micron channels for a terahertz compatible device. Here, the background and concepts are introduced for the context required and appreciation for this work's motivation.

1.1 Terahertz Spectroscopy

In 1947, a paper entitled "A Pneumatic Infra-Red Detector" introduced what could be regarded as one of the first, room-temperature THz-compatible detectors, colloquially known today as a *Golay Cell* [6]. The device consists of a gas chamber with a membrane constituting one of the cavity walls that deflects visible light from its surface. Incident radiation heats the gas and causes it to expand, deforming the membrane. The deformation is then subsequently measured by the change in angle of reflected light from the membrane surface. With the early use case measuring terahertz emission from high power sources such as fusion reactor plasma [7], these detectors are in fact still used today, and achieve comparable resolution for imaging as with cryogenically cooled bolometers [8]. For terahertz emitters, simple glow lamps outputting black-body radiation were the among the most popular in early years, with even such a device being used with Fourier transform infrared (FTIR) spectroscopy to obtain a bandwidth between 90 GHz to 2.3 THz [9], albeit in this case requiring a cryogenically cooled detector. Many advancements have been made in coherent, terahertz emission, such as from frequency down-conversion in CO_2 gas lasers [10] and molecular lasers[11], with sources boasting the highest output powers of milliwatts and greater, later arriving in the form of the quantum cascade laser [12] and the p-Germanium laser [13].

1.1.1 Terahertz Time-Domain Spectroscopy

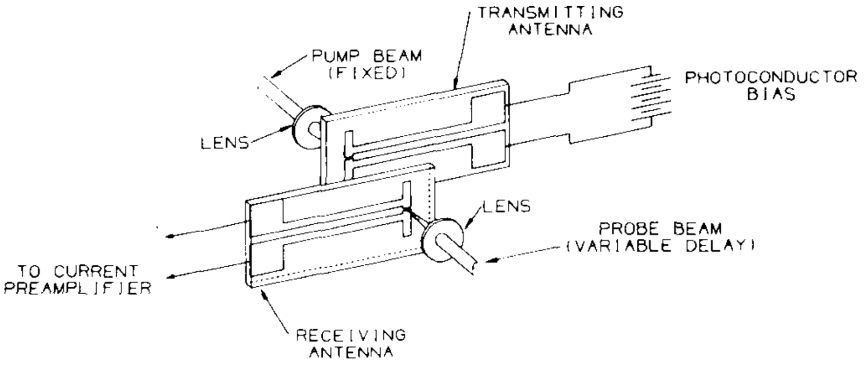

FIGURE 1.1: Figure from [14], where the first use of photoconductive antennas in a pump-probe configuration was reported. The optical beams excite the semiconductor between the antenna traces, with the a DC bias applied across the emitter and the detector outputting the measured current to a pre-amplifier. The terahertz pulse propagates through free space, with no lenses, so the antennas are separated by just two two mm.

In 1988, the first use of photoconductive antennas (PCAs) pumped by a pulse laser to both generate and detect terahertz radiation was reported [14]. This configuration was the beginning of Terahertz Time-Domain Spectroscopy (THz-TDS), allowing higher

signal-to-noise ratios (SNR) than that of FTIR spectroscopy [15], while having faster data acquisition times and straightforward phase retrieval. Figure 1.1 (from ref. [14]) shows the arrangement used in this early experiment, which is almost identical to the setup used today, with the main discrepancy being the lack of the modern terahertz optics in-between the emitter and detector.

The technique exploits the high repetition rate, coherence and short pulses from ultra-fast lasers to both generate and detect broadband terahertz pulses. The configuration is analogous to a pump-probe setup, whereby the laser pulses pass through a beam-splitter and one laser arm is retarded with a delay-line. A laser pulse from one arm is used to generate a terahertz pulse, the emission events of which are coherent with each laser pulse incident on the emitter. The other arm's pulse train is incident on a detector, providing an ultra-short detection window. The optical pulses, typically less than 100 fs, are much shorter than the > 1 ps terahertz pulses that are to be detected. The probing of the terahertz pulse, which is a convolution of the carrier envelope of the laser pulse with the terahertz waveform, is able to resolve high frequency components due to the optical pulse's envelope approaching that of a delta-function.

1.1.2 Terahertz wave generation and detection from photoconductive antennas

Both the detection and generation of terahertz radiation in the spectrometers used in this thesis are achieved with PCAs. They typically consist of a semiconductor, with metallic contacts deposited on the surface. The contacts create a conductive geometry, which can be of various designs, with the fundamental feature being a small gap, where the incident laser light is focused. A DC bias is applied across the contacts producing an electric field creating a dipole. Absorption of an ultra-fast pulse generates free carriers in the semiconductor, which due to the applied electrical bias, are accelerated creating a transient photocurrent. The photocurrent is a function of the carrier envelope of the optical pulse, which introduces new free carriers as the pulse propagates into the semiconductor, with the recombination of the carriers acting to reduce the photocurrent. The lifetime of the carriers is longer than that of the optical pulse, but short enough that the photocurrent depletes in less than a few picoseconds. Terahertz radiation is produced with an electric field strength that is proportional to the transient photocurrent. Therefore, for a short time domain pulse, and subsequently a broad pulse in the frequency domain, materials with shorter carrier lifetimes are necessary. Figure 1.2(a) (from ref. [16]) illustrates this time evolution. Figure 1.2(b-e) shows how the photocurrent is produced as a result of the carriers availability of mobile carriers, their generation proportional to the optical pulse intensity as it propagates into the photoconductive material.

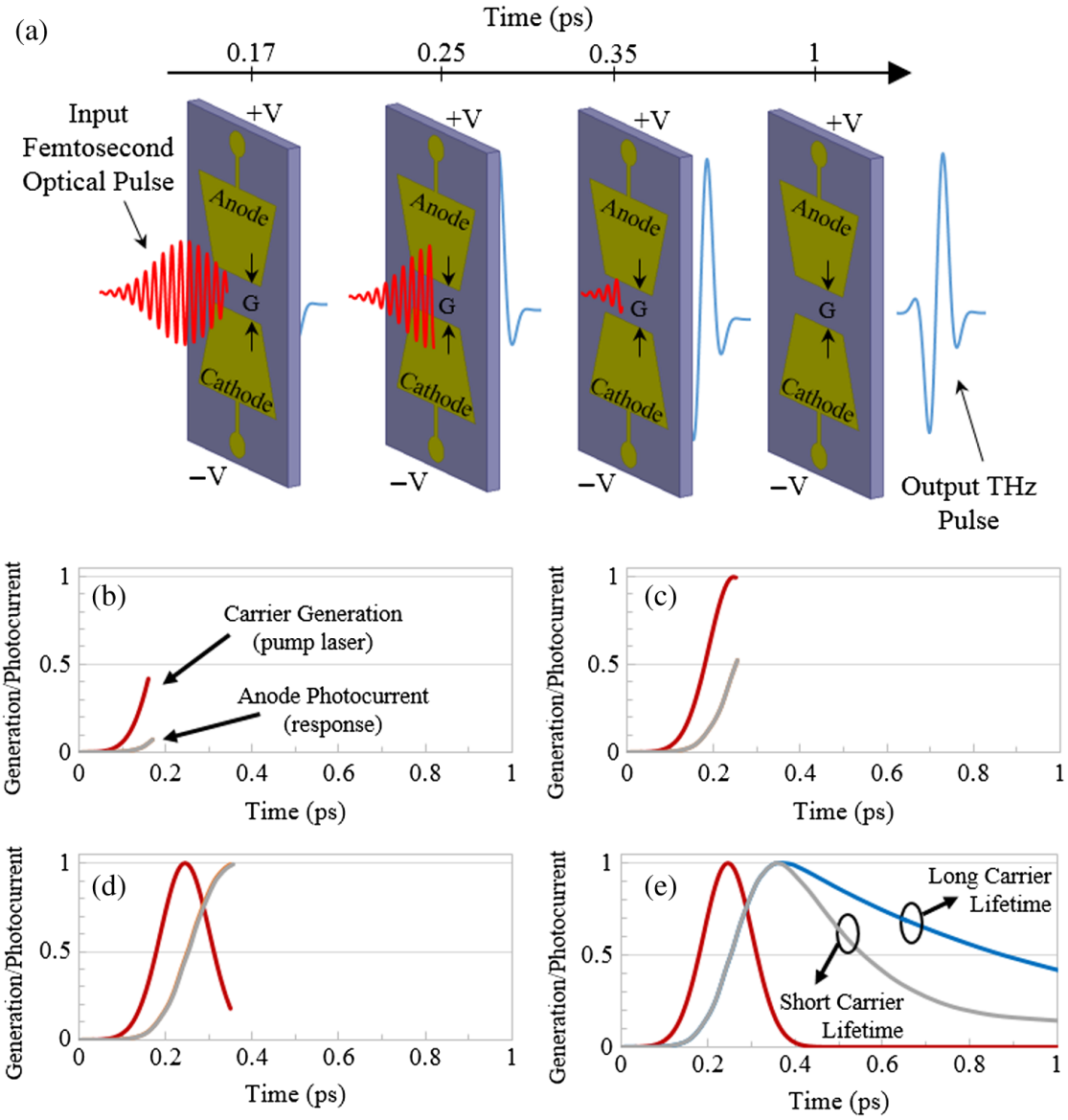

FIGURE 1.2: Figure from ref.[16], illustrating the time evolution of terahertz emission from a PCA. (a) A femtosecond optical pulse is incident on the semiconductor of the PCA, exciting carriers, which are accelerated by a DC bias. The transient current between the dipole emits terahertz radiation. (b-e) The carrier generation (red) from the interaction of the optical pulse with the semiconductor, the generated photocurrent in the antenna gap for long (grey) and short (blue) carrier lifetimes.

In the case of detection using PCAs, the mechanism is similar. Again, the incident pulse creates free carriers in the semiconductor, but no DC bias is applied across the contacts. Instead, a terahertz pulse is focused onto this active, now photoconductive region, and carriers are consequently moved by the incident electric field. This movement of carriers is realised as a measurable current. The optical pulse and the carrier lifetimes in the detector are short enough that this photocurrent is proportional to the amplitude of the electric field at a given point in time, depending on the delay between the two arms of the laser. Due to the coherence of the pump-probe system, this current measurement

will correspond to the same point in time of the evolution of the terahertz pulse. After the measurement is made, the delay line can be moved and in doing so, the entire waveform the of the terahertz pulse can be sampled.

1.1.3 Alternative emitters for THz-TDS

Of course, PCA emitters are just one subset of terahertz emission devices that are compatible with TDS. The closest sibling to the PCA emitter is the photodember emitter, where carrier generation and recombination are are generated by an ultrafast optical pulse to that of the PCA, but instead of an applied DC bias accelerating the carriers, the emitted terahertz radiation is generated by the diffusion of the generated carriers themselves. In the conventional case, the net THz radiation occurs orthogonal to the exciting optical pulse path as all other generated waves are destructively interfered. The introduction of a gold mask allows the lateral diffusion to reflect and constructively interfere to allow emission in the same direction as the optical pulse, known as the lateral photo dember effect [17]. Whilst benefiting from an electronically passive operation, these devices however have yet to become mainstream due to scaling challenges with regards to efficiency in pump pulse conversion and integration with fibre systems.

Optical rectification as a means to produce broadband terahertz pulses in an electrically passive way was shown as early as 1984 [18]. Unlike the photo dember effect, the $\chi^{(2)}$ non-linearity of a crystal can be both used to generate the terahertz radiation from an incident ultrafast optical pulse, but also sample the produced terahertz waveform. Typically the materials used are inorganic, examples including lithium niobate ($LiNbO_3$) and zinc telluride ($ZnTe$) but even broader bandwidths have been generated by transitioning to organic crystals, such as the reported 0.4 - 6.7 THz from 4-N,N-dimethylamino-4-N-methyl-stilbazolium tosylat (DAST) crystals [19]. The introduction of organic materials reduces the requirement for high powered, amplified laser systems necessary for conventional inorganic crystals.

Even broader bandwidths can be obtained through spintronic terahertz emitters where an ultrafast flow of electron spin in a magnetized ferromagnetic layer, generated by an incident optical pulse, is then coupled into a nonmagnetic layer whereby the spin-orbit coupling creates a transient charge current resulting in an emitted terahertz pulse. The more recent technique enables broad, gap-less bandwidths of up to 30 THz [20] and promises the advantage in cost effective production thanks to advances in nanomaterial design and fabrication.

1.1.4 Applications of THz-TDS

As introduced earlier, the *terahertz gap* refers to the region of the electromagnetic spectrum between the microwave and mid-infrared wavelengths. Many of the applications of terahertz spectroscopy are motivated in part by the resonances that occur at these frequencies, such as those of molecular bonds and phonons in crystal structures, examples shown in figure 1.3 (from ref. [21]). It follows that one such application of terahertz spectroscopy is in biomedical sensing, as large biological molecules such as proteins and nucleic acids have vibrational modes on this picosecond timescale [22].

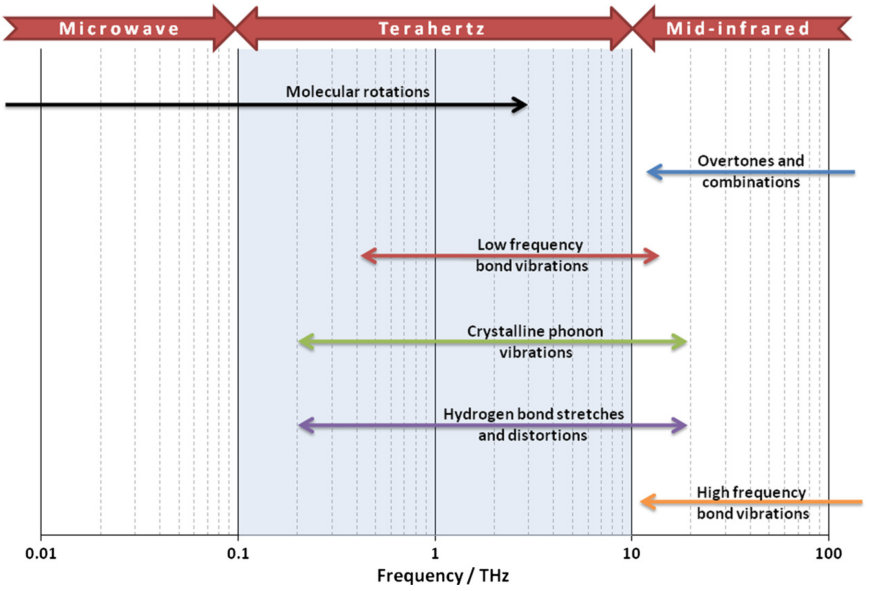

FIGURE 1.3: An illustration of the terahertz band in the electromagnetic spectrum (from ref. [21]), with examples of resonances that occur in this frequency regime.

Water, however, is extremely attenuating of radiation at terahertz frequencies, inhibiting penetration into biological material where it is abundant. Nevertheless, *in vivo* studies at the surface of biological matter have been carried out, such as imaging scar tissue with a THz-TDS reflection geometry [23]. Figure 1.4 shows the stark contrast between the scar tissue and the surrounding skin. This is due to the different water content of the skin versus the much less absorbing collagen in the scar tissue. This contrast was also visible through a plaster cast, indicating a promising application of the use of terahertz imaging for the monitoring of healing processes.

Similarly, exploiting water's strong attenuation has allowed imaging of the cornea of the eye, as well as tear film thinning with age [24]. This was able to achieved *in vivo* due to the non-ionising nature of terahertz radiation. Furthermore, the radiation has been used to study eye sclera, where the micro-structures present have a detectable polarisation dependence that indicates the health of the tissue [25]. Terahertz imaging has also been used to study the effects of Alzheimer's disease in the brain. Early results indicate a difference in absorption between healthy and afflicted parts of brain matter [26]. More

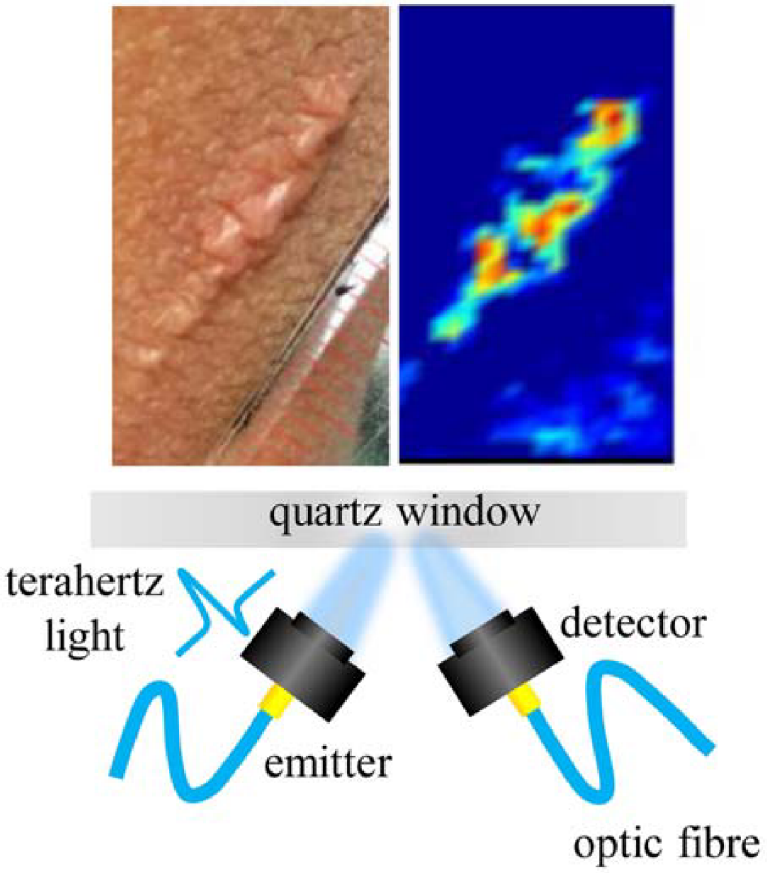

FIGURE 1.4: An example of imaging biological tissue with THz-TDS. Photograph of scar (left) and terahertz image of the scar (right) taken from a THz-TDS reflection set-up with translation stages for imaging (from Ref. [23]).

recently, by comparing frozen samples of both white and grey matter, it was found that healthy and diseased parts of the brain displayed different terahertz reflection spectra [27]. Further advancements in imaging techniques have allowed sub-wavelength resolution by employing an optically patterned silicon mask on the samples, which allowed comparably high resolution imaging of articular cartilage with sub-millimetre radiation [28].

1.2 Terahertz spectroscopy of proteins

Ideally, one would prefer to study biological material *in vivo*, where the environmentally sensitive matter of interest is sure to be in its native condition. However, the water that is abundant in biological tissue strongly absorbs terahertz radiation, this makes the study of biological tissue very challenging, and thus further trying to isolate particular biological macromolecules and their vibrational modes becomes quite difficult.

1.2.1 Protein structure

Proteins are biological macromolecules measuring a few nanometres across, each with a specific function for the survival of their host organism. A protein's function is dependent on its complete structure, which can be broken down into three main substructures.

The fundamental building blocks of proteins are the 20 amino acids that exist in genetic code. These are organic compounds that have a common central carbon atom and a hydrogen atom, a carboxyl group ($COOH$), and an amino group (NH_2). The amino acids are distinguished by the addition of a side chain attached to the central carbon atom. These can then be classed into three different groups, defined by the chemical properties that chain provides: those with hydrophobic side chains, charged residues or polar side chains. These are then joined together in a chain by a peptide bond, where the nitrogen of an acid's amino group bonding to the carbon of another carboxyl group, producing a free molecule of water. This chain, of many differing amino acids is called a peptide, and its sequence is known as the **primary structure** of a protein.

Following this, the peptide chain, or polypeptide, can fold into configurations that are energetically favourable, usually determined by the presence of certain amino acid side chains. Such **secondary structures** are mostly comprised of alpha helices or beta sheets, which are then arranged into *motifs*. The motifs can then form compact globular structures giving rise to a protein's **tertiary structure** [29].

For the study of biological macromolecules, specifically proteins, *ex vivo* measurements are required. However, in order for proteins to retain their proper function, it follows that they must maintain their complete structure. This is dependent on the conditions to which they are exposed. Salinity, pH and temperature of the surrounding solution, also known as buffer, are the major contributing factors. This is a key issue with *ex vivo* measurements, particularly when the probing radiation is highly attenuated by liquid water. The conditions are intrinsically linked to what can be probed by terahertz, as varying these conditions can activate/deactivate collective motions and can even completely denature the protein.

1.2.2 Protein studies in dry, frozen and crystallised states

It is logical to remove the largest source of difficulty from protein study: the water. The first use of THz-TDS to study a protein used lyophilised bovine serum albumin (BSA) and collagen compressed into pellets (mixed with polyethylene powder to control path-length) [30]. The pellets were then minimally hydrated, by increasing the relative humidity up to 77 %. However, other than the increase in absorption due to the extra water content, there was no change in spectral features, even compared to the

completely denatured sample. This was also the case for myoglobin [31] between 3.6 and 42 wt % of hydration. While this method allows for a good signal to noise ratio, the conditions are possibly too far from native to see any resonances from higher order structures. Rather than remove the water to reduce signal loss, it is possible to change its spectroscopic properties by changing its temperature. Liquid water's absorption has a strong temperature dependence [32]. Unlike in its liquid state, frozen water is mostly transparent below 1.5 THz. As such, studies have been conducted with the protein confined in ice. At temperatures above 220 K, the water in close proximity to the protein remains liquid and the higher order protein structures and their picosecond timescale motions are retained, despite the bulk water being solid. The point in which the temperature is high enough such that the water remains liquid in close enough proximity to the protein to allow these motions is referred to as the **dynamical transition**.

The dynamical transition of myoglobin was observed experimentally in 1989 [33] with inelastic neutron scattering, but was detected for the first time with terahertz radiation by Markelz in 2006 [34] in horse heart cytochrome c using a THz-TDS setup. By comparing native and denatured hen egg white lysozyme, THz-TDS was used to determine that this onset of collective motions was not actually dependent on the protein's secondary or tertiary structure [35], but was governed by the peptide structure and side chains themselves. This could mean that although ice is not representative of *in vivo* conditions, it can allow for studying directly the primary structure while 'freezing out' the higher order structures.

1.2.3 Aqueous solutions of proteins: hydration dynamics

When proteins are significantly hydrated, i.e. when they are in an aqueous solution, their interaction with the surrounding water extends beyond the most immediate layer of water molecules in proximity with the protein. A non-linear dependence of absorption as function of concentration of protein in solution was first observed in 2007 [36]. This led to the theory of the water extending out from the protein having different spectroscopic properties than the bulk water, now known as a **hydration shell**. A model to calculate the extent of this shell was presented in 2008, with the measurement of the hydration shell of ubiquitin [37].

The non-monotonic behaviour observed can be described by two contributing factors. Proteins are less absorbent than water in the terahertz band, and so initially there is a reduction in absorption of the solution as the protein is introduced, called the **terahertz defect**. In some cases, the protein can be modelled as hollow spheres due their minimal attenuation of the signal. However, the protein's hydration shell are more strongly absorbing than the bulk water, and so the absorption can actually increase when introducing more protein due to the addition of shell water, known as **terahertz excess**. There comes a point where all the bulk water has been displaced by protein and

shell water, which causes an inflection point in the concentration dependent absorption curve. At this point, equation 1.1 can be used to calculate the extent of this hydration shell, where α is the absorption coefficient and V is the volume.

$$\alpha = \alpha_{protein}\frac{V_{protein}}{V} + \alpha_{shell}\frac{V_{shell}}{V} + \alpha_{buffer}\frac{V - V_{protein} - V_{shell}}{V} \tag{1.1}$$

Stimulated unfolding of a protein by lowering the pH of the buffer can be detected by this disappearance of this hydration shell [38], evidenced by the lack of a non-linear behaviour in the concentration dependence. Furthermore, by modifying one of the protein's side-chains (from a polar to a less polar chain), a clear change in hydration shell was observed.

Correlation between a hydration shell's extent and protein function was observed [39], by looking at the number of water molecules in shells around peptides. The point in which the peptides are able to access enough water molecules shows a measurable onset of collective motions, which corresponds directly to the hydration required for the molecule to function biologically. An example of this relationship to protein function is the study of insect antifreeze proteins [40]. Here, the protein that's responsible for lowering the melting point of water in insects had a measured hydration shell that was more pronounced the closer the temperature of the solution was to the freezing point of water, strongly indicating that these long range dynamics are indeed key to its function.

The hydration shell for BSA was calculated to be 16 Åin diameter in 2014 using synchrotron radiation of 0.3 - 3.3 THz [41]. However, in 2015, a measurement of these shells indicated a value greater than this with 0.220 - 0.325 THz radiation [42]. Their findings revealed hydration shells with diameters of 16 Å, 19 Å, 25 Åfor lysozyme, myoglobin and BSA respectively. These diameters are also much larger than those previously calculated and measured with NMR, X-ray scattering and neutron scattering methods. The difference here suggests that the hydration shell may consist of motions that are on a range of timescales and distances from the protein.

A promising application of these measurements are in sensing a reaction or modification of a protein. A difference in hydration shell size of two specific labels/conjugations of immunoglobulin G antibody was detected [43]. However a more dramatic observation was made during the hydration shell measurement for the glycoprotein hemagglutinin from the H9 sub-type influenza A virus, where no shell could be detected when reacted with the antibody F10 [44].

While not a protein, its worth mentioning that carbohydrate macromolecules also have a long range hydration shell in solution. Measurements of shells around trehalose, lactose and glucose extend to 5.7 Å, 6.5 Å and 4.7 Å respectively [45]. The authors of this study suggest that the shell size is related to the number of oxygen atoms present in the molecule. It has also been proposed that the hydration shell is related to the

mechanism in which plants deploy sugars to protect themselves from extreme weather conditions [46].

The research in this field is ongoing, and the uniquely probable hydration shell is worthy of more exploration, as it relates to the complex domain of protein function and structure in a way that appears uniquely probably with terahertz radiation.

1.3 Machine learning for THz-TDS

Spectroscopic analysis was revolutionised by the increase in computational resources, particularly with the automation of the Fourier transform. The Fast-Fourier Transform (FFT) is an algorithm with a huge prevalence in of data processing, from image compression to astronomy. Today, it is argued that for general purpose computing there is another revolution: the access to artificial intelligence (A.I). Like with the FFT, the training of A.I to replace rudimentary algorithms with more sophisticated ones has become commonplace, with modern smartphones now containing optimised computer architecture to accelerate AI predictions for many of their functions and applications, from enhancing photos to predictive text. It is unknown whether this explosion in the prevalence of A.I. will continue on its exponential growth, or if like the hardware that runs it, will plateau in coming years in lieu of another software revolution. With the tools readily available and the benefits seen in so many fields of research and commercialisation it is of interest to see how this could be applied to terahertz science.

Artificial neural networks (ANNs), the subset of machine learning that has been the basis for modern applications and research, has been explored for the use in processing spectral data for medical imaging since 1989 [47], in part due to the increase in available computation resources but also due to the introduction of the efficient and effective back propagation training algorithm [48]. Since then, neural networks have been used to predict a myriad of parameters, from hydraulic head in ground water [49] to aircraft stability [50]. In the field of terahertz spectroscopy, ANNs have been used to classify organic compounds [51] and biological samples [52, 53], as well as detection of molecular concentrations in solutions [54, 55]. The aim of this work is to use ANNs to extract spectroscopic parameters from THz-TDS data as directly as possible, the ultimate goal being a completely hands-off approach.

1.3.1 Principles of artificial neural networks

An ANN is the general term given to the subset of machine learning whereby input data is processed in a way that is analogous to neurons in the brain. Figure 1.5 shows this analogy and how the an artificial neuron accepts input signals from other artificial

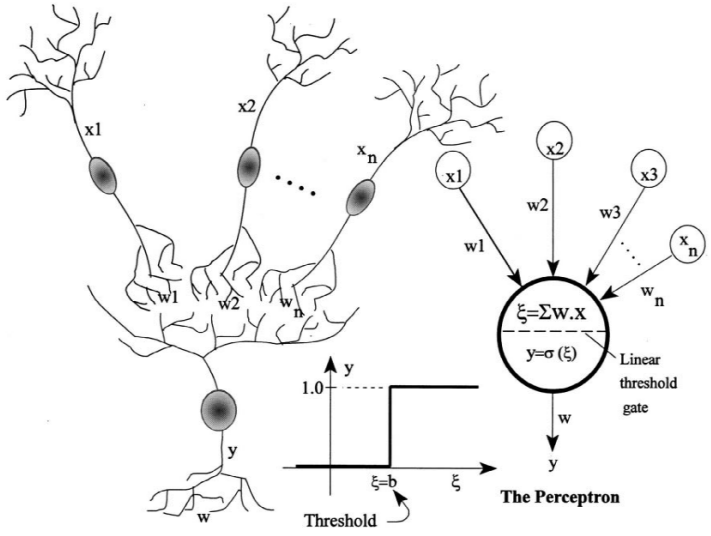

FIGURE 1.5: An example of an artificial neuron as an analogy to a living neuron, from ref. [56]. In a living brain, neurons activate depending on the electro-chemical signal strength received by neighbouring neurons. In the artificial case, this is simulated by using mathematical operations. The sum of the inputs is calculated and then inputted into a non-linear activation function to generate the output, y.

neurons. The output, or *activation*, of the artificial neuron is a function of the sum of each input multiplied by the weight of the connections. In this case, the termed 'activation function', is a step function. This provides normalisation to the output of each neuron but more importantly, non-linearity when multiple neurons are connected. For ANNs today, it is typical for each connected neuron to have an additional bias. Further, the step function is often replaced by a sigmoidal function or a similar hyperbolic tangent. Therefore the activation of a neuron becomes

$$y = \sigma\big(b + \sum_{i=1}^{n} w_i \cdot x_i\big), \tag{1.2}$$

where y is the activation (or output) of a neuron, b is a bias for that neuron and all the connections to it, w_i is the weight of each connection, and x_i is the input from each connected neuron. A network, or model, is typically constructed of layers of neurons, that are fully connected to those of the preceding and subsequent layers. The first layer, the input layer, is an array of values known as *features* with the final layer corresponding to an array of data known as *labels*. The addition of one or more layers in between, termed hidden layers, creates enough flexibility that with enough neurons, any function can be approximated by the network [57]. Figure 1.6 shows an ANN with one hidden layer. Each neuron 'activates' outputting a value that is a function of the values of the connecting neurons preceding it.

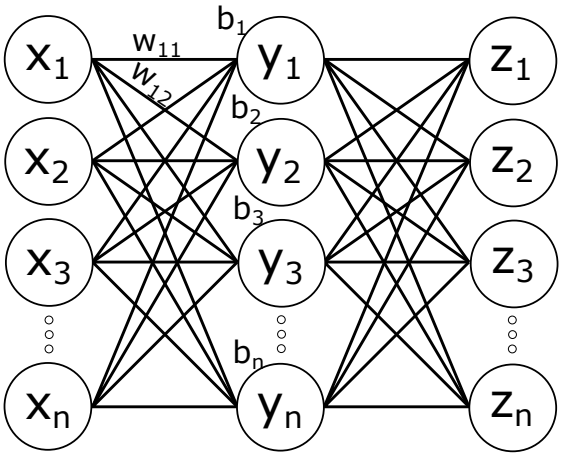

FIGURE 1.6: Fully connected artificial neural network (ANN), with one hidden layer. The combination of connections with individual weights, the non-linearity introduced by the activation function and the addition of a layer between the input and output provides the flexibility for this to approximate any function.

1.3.2 Training neural networks

With the highly flexible framework defined, the ANN requires specific values for its weights and biases for it to perform the desired function. This is achieved by *training* the network through means of back-propagation and a supply of training data. The training data is a collection of already known inputs and outputs that the network is encouraged to emulate. Back-propagation is the technique where the performance of the network is evaluated when predicting an output for a given input, and the weights and biases are adjusted based on this performance. The performance is characterised by the loss function, representing how close the prediction was to the known output. The two most prevalent algorithms for doing so are known as Levenberg-Marquardt back-propagation [58] and the stochastic gradient descent algorithms. They both function by analysing the loss of network, and adjusting the weights and biases in accordance with the the gradient of the loss function with respect to these weights and biases. The next iteration can trial new parameters than should be closer to a local minimum of the loss function. The former uses the entire network architecture to create the adjustments, which tends to be yield more accurate results and faster convergence at the cost of high memory usage and the inability to parallelise the workflow. SGD however, analyses random samples of the network to inform the next guess, reducing the memory overhead and even allowing optimisation for parallel processing on graphics processing units (GPUs) [59].

1.4 Microfluidics for terahertz spectroscopy

Typically, when water based samples are measured in a TDS system, separate refillable cells are used, usually consisting of two parallel windows and a spacer [60, 61, 62, 63]. This scheme allows less than 100 μm thick path-lengths for highly attenuating water to be achieved in a straightforward manner. However, this can become cumbersome when many different concentrations of a solution are required. Additionally, the constant adding and removing of samples from the spectrometer requires frequent referencing to account for any drift or miss-alignments that can happen throughout the lengthy experiment.

An alternative is to employ a microfluidic device. If constructed from terahertz compatible materials (low absorption and refractive index), they can provide the short path-lengths required while having additional benefits such as performing real-time spectroscopy, additional sample manipulation and expanded functionality, such as integrating electronics [64], photonics [65], heating elements [66] or mixing chambers [67]. Microfluidic platforms promise to house a wide array of functional components, able to target samples in unique ways and with extreme precision. For example, the ability to manipulate single biological cells [68, 69]

As well as spectroscopic considerations for integration with a THz-TDS system, mechanical rigidity is often a desirable characteristic, to avoid flexing of a device possibly changing the optical path-lengths or causing device failure due to leaking. Examples are silicon or quartz [70, 71], but they suffer from requiring specialised manufacturing processes. Polymers are often used for the rapid prototyping of microfluidic devices, such as polydimethylsiloxane (PDMS). This material is much faster to produce, but has wide variability in absorption in the terahertz band [72].

Polytetrafluoroethylene (PTFE) is a material that has excellent chemical resistance and bio-compatibility, along with a low, non-dispersive refractive index between 300 GHz and 3 THz, and negligible absorption over a broad terahertz spectrum. It is easily machinable making prototyping quick, but suffers from incompatibility with bonding agents and can be prone to flexing. However, its hydrophobicity can aid water confinement in compression sealed devices, where capillary action acts against the escaping of water [73]. PTFE is the material of choice for this work, where its hydrophobicity plays an essential role in the function in a device. Furthermore, it is possible to modify this hydrophobicity through plasma exposure reducing this effect [74, 75, 76]. This makes depositing materials such as gold more achievable due to the change in surface chemistry, but also presents an opportunity to create more sophisticated devices where water is controlled through surface tension effects [77, 78, 79, 80].

1.4.1 Flow in a microfluidic channel

It is important to understand the behaviour of fluid confined to a microfluidic channel. If the fluid exhibits laminar flow, the control of different concentrations of solutions is possible to predict and new solutions are able to replace existing ones by an advancing, uniform boundary. The Reynolds number [81], R_e, is a value that determines the state of flow. It is calculated as follows

$$R_e = \frac{QD_H}{\nu A}, \tag{1.3}$$

where Q is the flow rate, A is the cross-sectional area of the channel, ν is the fluid viscosity, and D_H is the the **hydraulic diameter**, a length value that depends on the shape of the cross section of the channel. It is important to avoid turbulent flow for the proper function of microfluidic devices in most cases (mixing chambers can benefit from turbulent flow to enhance diffusion), and the Reynolds number provides design restrictions when fabricating a channel of particular size and geometry with expected flow rates.

1.4.2 The effect of surface topology on wetting

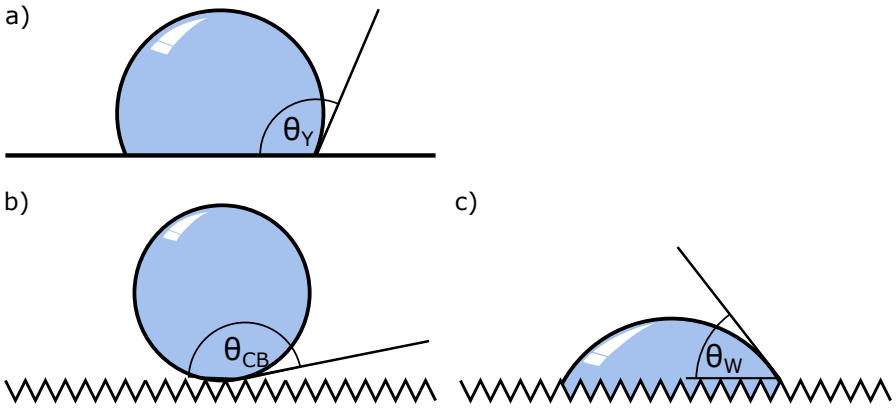

FIGURE 1.7: Contact angles for different surfaces and different regimes. The native, Young's contact angle a) is due to the interaction of a water droplet on a flat surface. This is then enhanced by a surface of significant roughness where the contact angle is a combination of the Young's contact angle and that of air (180 deg), known as the Cassie-Baxter state. If the liquid is able to overcome surface tension forces and wets the entire rough surface, the contact angle is significantly reduced entering the Wenzel regime.

PTFE is a hydrophobic material, meaning that it is 'water repelling', or 'non-wetting', due to its non-polar nature. To characterise this effect, a contact angle measurement is made. When a water droplet is placed on a material, it will bead or spread depending on how hydrophobic or hydrophilic the surface is. This is due to the balancing of the free surface energy of the surface, liquid and gas, at their respective interfaces.

The angle is measured between the liquid surface where it meets the solid surface. An angle below 90 degrees is said to be hydrophilic, greater than 90 degrees is hydrophobic and above 150 degrees is referred to as superhydrophobic. Figure 1.7(a) illustrates a droplet on a hydrophobic surface, with a corresponding Young's contact angle, θ_Y, corresponding to the interaction of the droplet due to the surface chemistry alone.

In addition to the substrate material's chemical composition, its surface topology also plays a significant role in its wetting characteristics. The material's roughness, r, or the presence of micro-structures with much smaller dimensions than the average droplet size can act to enhance the effective repulsion or attraction of water on a surface. If the liquid bridges the microstructures such that air pockets remain in the surface's 'valleys' illustrated by figure 1.7(b), the contact angle is always increased, and the system is said to be in the Cassie-Baxter state [82]:

$$\cos\theta_{CB} = f_1 \cos\theta_Y - f_2 \tag{1.4}$$

where f_1 and f_2 are the fractions of the surface area in contact with the liquid of the solid material and air respectively, and θ_{CB} is the effective, Cassie-Baxter contact angle. Alternatively, if the liquid is entirely in contact with the surface, occupying gaps between microstructures for example, it is said to be in the Wenzel state [83], and the measured contact angle, θ_W is given by equation 1.5

$$\cos\theta_W = r\cos\theta_Y \tag{1.5}$$

where r is the surface roughness. Figure 1.7(c) illustrates how a droplet behaves in such a regime, with the contact angle being reduced. There is also a transition that occurs between these two extremes, where the minimisation of the surface energy leads to partial penetration of the surface valleys, but in the case of very tall and narrow peaks (relative to micro-scale) of a material with a naturally high contact angle, and assuming that the wettability is increased only moderately, the two regimes will remain distinguished [84].

1.5 Thesis outline

Terahertz radiation can be used to probe various physical phenomena from charge carriers in semiconductors to vibrational modes in biological macromolecules. One such category of molecule is the protein. Consisting of a long chain of amino acids, folded into secondary and tertiary structures, proteins have molecular bonds which give the higher order structure collective vibrational motions which move on the picosecond

scale, therefore resonating with terahertz radiation. However, strong water absorption makes targeting these modes almost impossible, but presents the opportunity to investigate the hydration shell in proximity to the protein if the attenuation of the terahertz signal can be tamed. Furthermore, parameter extraction techniques for THz-TDS data, while having its advantages due to its straightforward phase retrieval, adds another layer of variability and complexity. Machine learning with artificial neural networks can possibility greatly assist, thanks to the huge development of software and hardware over recent years.

This thesis documents the development of new methods for analysing TDS data as well as a new experimental platform for measuring aqueous samples. Chapter 2 describes experimental methods and introduces some established parameter extraction techniques, while introducing a new method that promises to alleviate the issues of repeatability. Chapter 3 advances to a machine learning approach for complex refractive index extraction, where an entirely new approach is demonstrated. Then, the practical limitations of measuring aqueous proteins is discussed, with the development of a novel surface tension confined microfluidic device documented in chapter 4, with measurements of aqueous bovine serum albumin (BSA) to validate the platform. Finally, the results of this work are concluded and the next steps outlined in chapter 5.

Chapter 2

Terahertz time-domain Spectroscopy: Methods and Parameter Extraction

The foundation of this thesis is in the advancement of Terahertz time-domain spectroscopy (THz-TDS) experimental platforms and analysis techniques. This chapter aims to introduce the reader to the experimental and data processing techniques to provide the necessary context for the work in later chapters, where some of the nuances and limitations are discussed, as well as introducing a new parameter extraction method outlined in the final part of this chapter, of which was published [3]. The data for the publication was acquired by J. Gorecki, and with the thesis author contributing with the data analysis.

2.1 Introduction

The key characteristic of THz-TDS is the ability to measure the amplitude of the electric field in the time-domain instead of merely its intensity. An ultrafast laser (< 100 fs) is used to both generate and probe a broadband terahertz pulse, with each emission and detection event being coherent with the repetition rate of the pulsed laser. Measuring the amplitude as opposed to intensity enables a high signal-to-noise ratio (SNR) due to the background thermal radiation consisting of random phases and amplitudes which are not coherent with the emission and detection events. The frequency of the sampling of the terahertz waveform can occur at frequencies on the order of tens of megahertz, enabling a fast acquisition time as the background radiation is rapidly 'averaged out'.

The focus of this chapter is on the other major advantage that arises from this direct probing of the electric field: the preservation of phase information when extracting the

frequency components. By drastically simplifying the phase retrieval process, it is possible to extract the complex refractive index from a sample measurement without the need for Kramers-Kronig relations due to the possibility to numerically fit a relatively simple theoretical model containing just two unknown variables: the complex refractive index. Further, it is even possible to analytically extract the parameters when some approximations are made, at the cost of some accuracy. Figure 2.1 shows one setup used for the data collected in this work. It is a typical photoconductive antenna (PCA) emission and detection scheme consisting of a sub-100 femtosecond Ti:sapphire laser driving the apparatus, a delay-line comprised of a linear translation stage and a retroreflector, a PCA emitter for generation of the THz radiation, with lenses for columnating and focusing the beam to the PCA detector. These are just the optical components, and later in the chapter, the addition of electrical and mechanical components will be discussed.

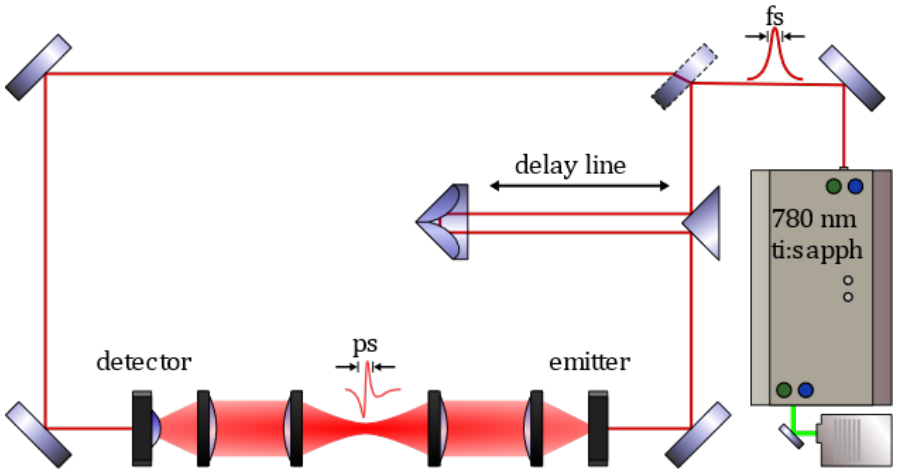

FIGURE 2.1: A schematic of the time-domain spectrometer. Ultrashort pulses (<100 fs) at 800 nm are produced by a mode-locked, ti:sapphire laser, which are then split. One arm is delayed by a retro-reflector on a linear stage (not shown) before continuing to the THz generation and detection apparatus.

To find a sample's refractive index in a transmission setup such as this, first a reference scan has to be taken. In the case of a single piece of solid material, this reference is air. However, in the case of a liquid sample, the reference is the windows of the cell that houses the fluid, in direct contact with each other. A scan of the THz pulse in either configuration is then taken. The sample is then placed in the focus of the terahertz beam visible in figure 2.1, and another scan of the pulse is taken. It is possible to calculate the average refractive index and absorption by simply looking at the delay of the peak (equation 2.1) and the difference in amplitude with respect to the reference signal,

$$n_{average} = 1 + \frac{c \cdot \Delta t_{pulsepeak}}{L} \tag{2.1}$$

where $\Delta t_{pulsepeak}$ is the time delay between the pulses propagating through the reference and sample, L is the sample thickness and c is the speed of light. However, there

is a broad spectrum of frequencies that constitute each pulse, and performing a Fourier transform gives enough information to extract more than just the average refractive index, but the frequency-dependent complex refractive index.

There are other schemes of TDS where an independent reference measurement is not necessary which have been demonstrated. Instead of a direct transmission configuration, an oblique reflection mode can be used where a sample is mounted against a quartz slide and the first terahertz pulse reflects off the face of the quartz, while the second pulse enters the quartz and reflects off the sample. By comparison of the two pulses, the complex refractive index of the sample can be determined [85, 86]. Alternatively, a sample can be placed on a polished mirror, in which internally reflected THz pluses are delayed in time as compared to the pulse reflected on the top surface [87]. By comparing the two reflected pulses, the refractive index can be obtained without an independent air reference. Complications arise due to the quality of the secondary pulses being highly dependent on the sample–mirror interface, requiring alignment of two flat surfaces without an air gap. Further, the use of terahertz reflection setups can impose additional alignment issues as compared to transmission modes, as one of the pulses will be displaced due to the refractive index of the sample.

A self-referencing method in a transmission mode, used to measure the strength of absorption peaks has been demonstrated in which a Fourier transform is applied to the time-domain signal with an ever-increasing window size [88]. In this manner, the frequency resolution of the data increases with the window size, allowing the strength of sharp absorption lines to be resolved. The method requires no independent reference scan, but is effective at measuring absorption peaks with a sharp spectral width, and is not applicable to obtaining the real refractive index. Internally reflected pulses transmitted through the sample can be utilised to increase the sensitivity of spectroscopy devices. Various groups have demonstrated techniques where the first transmitted pulse and the internally reflected "echo" pulses are both compared to the air reference, where the echo pulses contain extra information due to increased interaction with the sample [89, 90, 91]. A similar technique has been presented in transmission mode to determine refractive index, where the first pulse transmitted through the sample is compared to the echo pulse, which is internally reflected in the sample before being transmitted. However, this technique was only shown to produce the refractive index averaged over the terahertz pulse bandwidth instead of the frequency-dependent values [92], in a manner analogous to equation 2.1. Use of the echos produced by a substrate preceding the unknown sample has been demonstrated with a less conventional but effective algorithm where a decision tree is used to build an appropriate transfer function with materials of many layers [93, 94]. In this way, one does not need to explicitly write a cumbersome and long transfer function that contains all contributions from reflections and transmissions at each layers boundary and instead the complexity can be built up in discrete iterations. In this chapter, the clear pulse separation observed in sufficiently

high-refractive index samples will be exploited to allow the fitting of a transfer function with fewer terms that benefits from simpler numerical fitting at the expense of a more restricted sample parameters.

When gathering and analysing THz-TDS data it is crucial to understand the nuances of the experimental procedure of measuring the pulse. This chapter introduces the methods involved in taking a spectroscopic measurement with THz-TDS, with much of the focus on the parameter extraction techniques used in this work. A new method is introduced in the last section of this chapter, where in the case of thin samples of high refractive index, the need for an air sample measurement can be negated. The method relies on internal reflections within the sample to produce what will be referred to as 'echo pulses', which can be referenced to each other, requiring only one measurement scan to obtain all the necessary data. This method has a number of advantages over the traditional air referencing method, as both measurements can be taken in a single time-domain trace, which reduces the chances of drift in the spectrometer between measurements. In addition, this method allows for the sample to be held securely in the optical setup at all times, which reduces chances of misalignment between repeated sample mountings and can allow for samples to be held in difficult-to-access places such as sealed chambers and thermal stages. The method shows an improvement on the reputability as compared to conventional methods in some cases, while also being compatible with established root-finding algorithms and inspection techniques.

2.2 Considerations for the acquisition of THz-TDS data

In this section, the key components for the acquisition of data for THz-TDS measurements are described, along with certain features to give context for the parameter extraction techniques discussed. The introduction of each electrical and mechanical component necessary and how they are integrated with the terahertz optics is outlined, along with common considerations that should be taken into account when building and using a THz-TDS.

2.2.1 THz emission and detection characteristics from two types of PCA

In the terahertz setups featured in this work, multiple configurations of PCA emitters are used. The variety for measurements requiring the best signal-to-noise ratio (SNR) are known as large area emitters (Tera-SED 3x3 mm emitter from Laser Quantum). They consist of an array of periodic contacts, that cover the surface area of the device, with each, alternate pair of contacts forming a dipole, behaving much in the same way as described in chapter 1. Figure 2.2(a) shows a simplified schematic of what is contained in these types of emitters. They offer greater output powers than single dipole

emitters, while requiring lower DC biases, as well as the easier alignment of the incident laser pulses, requiring no focusing. Most of the generated THz pulse transmits out of the emitter at the semiconductor/air interface. However, some radiation is internally reflected producing *satellite pulses* as a consequence of the refractive index contrast of the PCA material and the surrounding air. In the case of the large area emitters, the photoconductor and its substrate are thin enough that the satellite pulses are visible in a 50 ps window, demonstrated in the accompanying time-domain trace in figure 2.2(a). These reflections are visible as oscillations in the frequency-domain which can make fitting problematic. In single dipole emitters requiring the lens, the silicon shares a similar refractive index to that of the gallium arsenide (GaAs) of the PCA. Therefore, there is minimal reflection at the boundary between the lens and the PCA, and the next measurable internal reflection occurs between the silicon lens and air, a much larger optical path. This results in a delay larger than that of the time delay window often limited by the delay-line (> 50 ps). The train of pulses is of note when analysing the data in all but the thinnest samples, where there is a small delay so windowing is always required. Both emitters are used in this work, with the 'echo method' at the end of this chapter requiring the silicon lens and single dipole configuration 2.2b), and the microfluidics integration requiring as high SNR as possible in chapter 4.

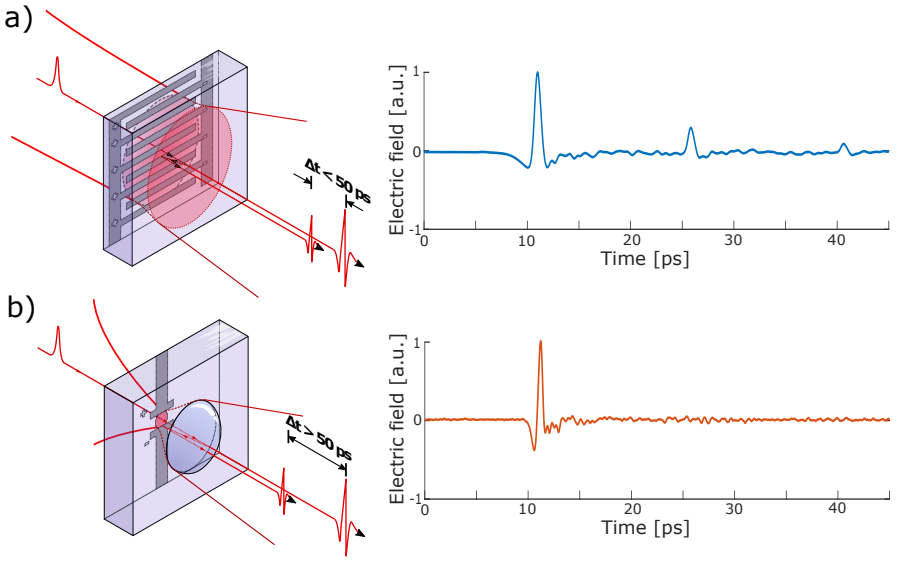

FIGURE 2.2: Terahertz emission from a large area PCA emitter (a) and a single dipole PCA (b) with significantly delayed satellite pulses. Satellite pulses are only captured in the time-domain of the large area emitter due to the silicon lens sufficiently increasing the path length of internal reflections.

The detectors used in the setup are of a similar configuration to the PCA emitter shown in figure 2.2, where the detection process is almost the reverse of that of the emission principle. The incident femtosecond pulse creates free carriers in the semiconductor of the detector between the contacts deposited on the surface. The incident terahertz pulse is focused using a silicon lens from the other side of the detector, to overlap with

this excited region, and the terahertz electric field moves the carriers generating a measurable current proportional to the terahertz electric field. The lifetime of the carriers are sufficiently short (sub picosecond), along with their generation by the femtosecond optical pulse, as compared to the period of the frequency components within the terahertz waveform. Therefore the current, proportional to the convolution of the two, effectively probes the amplitude of incident terahertz field at a given time delay between both pulses incident on the detector. The repetition rate of the laser is approximately 80 MHz, which nets a high SNR in a short period of time as this is the rate in which the emission and detection events occur. A delay-line is then used to offset the time in which the terahertz pulse is incident with the excited semiconductor from the detector, producing a time resolved electric field trace.

The PCA emitters used are driven by a function generator, which varies the applied DC bias continuously between $\pm$ 6 V at a frequency of 29 kHz. This is then referenced by a lock-in amplifier measuring the current from the detector to help reduced electrical noise in the system. The key variables here are the maximum voltage applied across the emitter along with the duty cycle. Each emitter has their own upper limits, which is a function of these along with the average power of the excitation pulses. It is possible to alter the characteristics (usable bandwidth and peak SNR) by choosing a shorter duty cycle but with a higher voltage, for example. The frequency of the function generator can be tuned to find a value that improves SNR by selectively suppressing background electrical noise, as long as it is sufficiently lower than that of the repetition rate of the laser. The lock-in amplifier will average the current in accordance with this signal, but the limits occur as a function of the movement of the delay-line.

2.2.2 Scanning a terahertz pulse with *fast* and *slow* operation

The time-delay offset between the arrival of the optical pulse and the terahertz pulse on the detector is controlled by introducing a delay between the emission and detection arms of the split Ti:sapphire laser beam. The path length of the 'emission line' is increased using a retro-reflector mounted on a translation stage. When a single data point is acquired, the current from the detector is measured by a lock-in amplifier, which is outputted to a computer, along with the position of the delay-line at that time. There are two modes of operation possible with the configuration used in the work, which here will be referred to as 'slow scan' and 'fast scan'. The slow scan mode of operation starts with the retro-reflector set at one position of on the motorised, linear stage, and sequentially moves in discreet amounts while the lock-in amplifier is averaging over relatively long periods of time ($\approx$ one second) at each delay-line position. Once this has repeated for the desired length of time (typically 50-100 ps), limited by the length of the linear translation stage, a scan is complete. For 1024 data points, it can take on the order of tens of minutes, which can make the scan susceptible to interruptions or laser

drift in the setup. There is the advantage, however, of having more granular control of the the averaging at each point, as the linear stage will not move until the averaging has been completed. Alternatively, if the linear stage can accelerate quickly enough, it can be set to continuously oscillate between two, predetermined endpoints, while the lock-in amplifier streams detected current to the computer, along with the corresponding positions of the delay-line. In the case for the setup in this work, the linear stage is too slow for this operation. Instead, an oscillating stage able to operate up to 20 Hz with a maximum delay of 50 ps ('APE scanDelay') is mounted on top of the linear translation stage. In this case, the lock-in amplifier is set to a much smaller time constant (< 1 ms), where each data point receives less averaging as it is acquired, but the entire scan can be averaged after the fact in software. This has the advantage of less susceptibility to fluctuations during acquisition, while also being able to display a waveform instantly, either on the computer or with an oscilloscope. However, care must be taken to select a short enough time-constant as not to suppress high frequency components present in the time-domain trace. Further, the 50 ps window can be too short to account for both the sample scan and the reference scan, if the delay of the pulse is too large. If so, it is possible to move the linear delay stage that the fast oscillator is mounted on, but this will cause the error in time delay (and subsequently phase and extracted refractive index) to be a function of both the error in the linear stage and the error in the oscillator itself. Such a scenario is therefore best avoided where possible.

2.2.3 Referencing and laser drift

When referencing a terahertz scan in order to obtain reliable material characterisation, one issue to highlight from the air referencing method is that a terahertz spectrometer may experience fluctuations in the laser power between taking a reference measurement and a sample measurement. This effect is especially compounded when it is required to take repeat measurements of samples over a long time period, such as several hours, to investigate a temporally changing system. Furthermore, when taking a reference measurement, the sample must be removed/added to the optical system, which may not be physically possible in some situations; for example, with a sample held in a sealed chamber or other enclosed space. When taking a measurement with an air reference, it is generally assumed the complex refractive index of the air is 1+0ik, i.e., the air has identical refractive index to that of a vacuum; however, in practical cases, there is often atmospheric moisture, which can cause absorption of the terahertz pulse. After the air reference is taken and the sample is placed in the spectrometer, the beam then passes through less water vapour, which can skew the values of imaginary refractive index and register in the system as an artificial gain in intensity. It is possible to counter the laser drift in a measurement by tracking the power output from the laser, and scaling the measured power accordingly. However, the assumption has to be made that this is indeed a linear correspondence with the measured spectra and function of

the PCAs. By taking repeated scans over a long period of time of a reference, along with measuring the power of the laser from an additional laser arm, it is suggested that this correlation holds (not shown). However, such modifications are not ideal, due to the assumptions about the linearity of both the detector and emitter that must be made. A more appropriate approach is to split the terahertz pulse itself and have a reference scan along side, however, this reduces signal and adds more complexity and noise to the system. Later in this chapter, a self referencing technique is introduced which avoids these issues entirely.

2.3 Measurement data preprocessing

As the name implies, THz-TDS is a spectroscopic tool. Resolving frequency components in a measurement requires Fourier analysis, typically achieved by the efficient and widely used algorithm: the Fast Fourier Transform (FFT). Before spectroscopic parameters can be extracted, however, some data processing must take place in both the time and frequency domains. Often, these steps come at some compromise, and care must be taken in each step.

2.3.1 Time-domain preprocessing: windowing function and zero padding

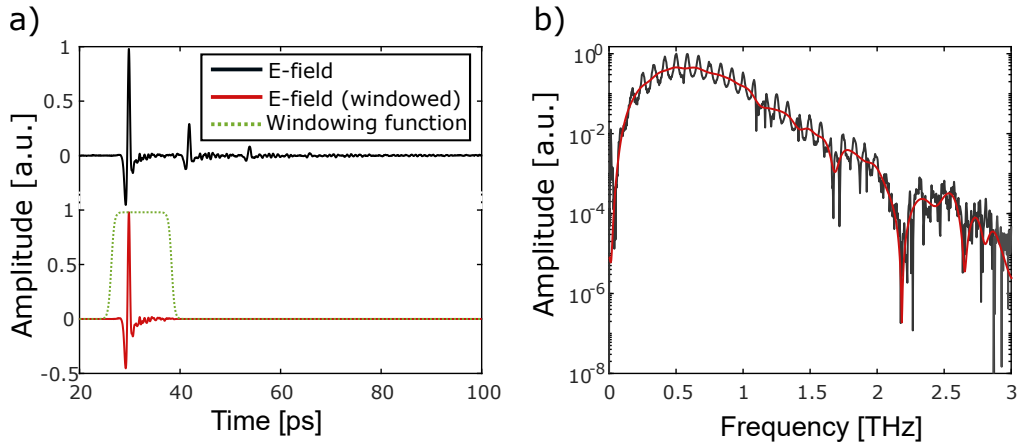

FIGURE 2.3: Time-domain trace of a pulse of a sample with satellite pulses due to internal reflections a) with the windowing function applied to remove trailing pulses. The frequency spectra is shown b) of both the raw pulse and the windowed pulse.

If the sample being measured has a sufficiently large enough refractive index and thickness, any additional reflections observed in the time-domain can be removed by using a windowing function, $W(t)$. As will be evident later, this simplifies the theoretical model and potentially reduces artefacts from the cutoff in time due to the finite travel of

the delay-line. This function when multiplied by the time-domain data retains the leading pulse and suppresses any subsequent reflections. An asymmetric, double-sigmoid function is used to smoothly zero the trailing electric field fluctuations while generating minimal additional frequency components when the Fourier transform is performed, given by the following

$$w(t) = \frac{1}{(1 + e^{-\alpha_1(t-t_0+\beta_1)}) \cdot (1 + e^{-\alpha_2(t_0-t-\beta_2)})}, \tag{2.2}$$

where β_1 and β_2 adjust the displacement of the sigmoid with respect to the centre at t_0. α_1 and α_2 give the maximum gradient of the rise and fall of the window respectively. Figure 2.3(a) shows the window function and its application to the pulse. This function has been chosen to minimise the spectral leakage caused by the sharp features of a discrete top-hat windowing function, while also providing more tunability than a Tukey window. The reflections cause severe oscillations to appear in the frequency domain, but applying the window function eliminates these as can be seen in figure 2.3(b). Depending on the thickness and refractive index of the sample, this window function must be altered as to maintain remove the pulses whilst still maintaining a high resolution in the frequency domain. Shortening the time window in this way reduces resolution due to the inverse relationship of the time and frequency domains,

$$f_s = \frac{1}{T_{window}}, \tag{2.3}$$

where f_s is the sampling frequency of the FFT of a time series of length T_{window}. Typically, windowing parameters will be chosen to accommodate as much of the pulse as possible while removing trailing reflections. A slow decent helps minimise the effects of cutting off any trailing signal information that may overlap with the subsequent reflected pulse.

2.3.2 Frequency-domain: phase unwrapping

If the frequency components of a measured pulse, have corresponding phases greater than 2π radians, the extracted phase information from a Fourier transform is wrapped, due the periodic nature of sinusoidal functions. It follows that so will the corresponding transfer function that maps one measurement onto the other, and the extracted real refractive index in particular will give erroneous results as the induced delay in a particular frequency component from the sample will be underestimated by an integer multiple of 2π radians. A phase unwrapping step is therefore required in order to obtain the correct refractive index. Figure 2.4 shows the phase of the Fourier transform of a measured pulse. The wrapping can be seen as saw-tooth function of frequency. The 'unwrap' function in Matlab searches for sudden changes of 2π radians and removes

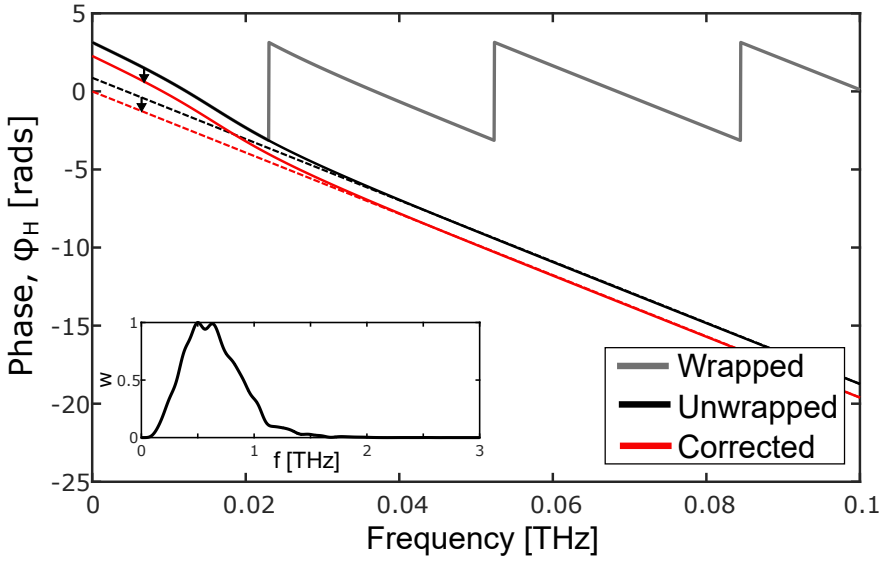

FIGURE 2.4: The phase of the Fourier transform of a terahertz pulse (windowed) both wrapped (grey) and unwrapped (black). The phase is then offset by fitting a weighted line (red) where the weights are defined by the square of the power spectrum (inset)

them to reconstruct a continuous trace. However, this can lead to a phase with an incorrect offset in the y-axis due to noise, and therefore uncertainty at the starting values. To provide a better prediction for this, a function is fitted to the phase, with weights determined by the magnitude of the power spectrum (see inset in figure 2.4). The intercept of the y-axis is determined from this fitted function and the offset is applied to the phase. An obvious problem presents itself with this approach, as the approximate shape of the real refractive index of the sample being measured must be known to determine what function needs to be fitted to correct the y-offset. Incorrect offsetting results in erroneous values of the real refractive index, predominantly in the low frequency range.

2.4 Extracting the frequency dependent, complex refractive index from an experimental transfer function

With the data ready to be analysed in the frequency-domain, the experimentally measured transfer function, $\widetilde{H}_{exp}(\omega)$, is obtained by straightforward division of the sample by the reference scans. The transfer function describes how the light-matter interaction with the sample is mapped from that of the known reference.

An example is shown in figure 2.5. A reference scan of air, followed by a sample scan of a 500 µm thick sample of semi-insulating GaAs (SI-GaAs) was taken in a THz-TDS, in 'fast scan' operation. The data was preprocessed in the methods previously described

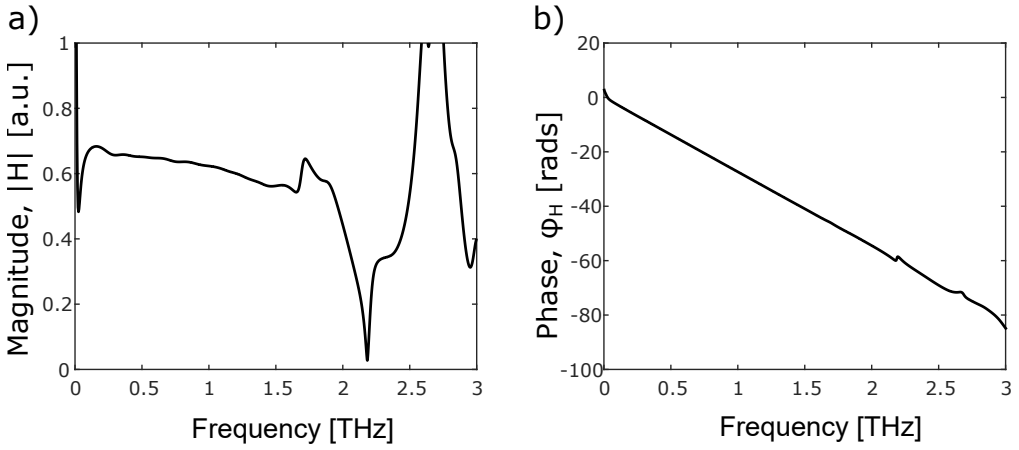

FIGURE 2.5: The magnitude a) and phase b) of an experimental transfer function obtained by dividing the preprocessed frequency spectra from a TDS measurement of 500 micron thick semi-insulating gallium arsenide (SI-GaAs) sample by a reference measurement of air.

before the division of amplitudes and phase subtracted shown in figure 2.5(a) and (b) respectively. The result is the experimental transfer function,

$$\widetilde{H}_{exp}(\omega) = \frac{|\widetilde{E}_{sam}(\omega)|}{|\widetilde{E}_{ref}(\omega)|} \cdot e^{i(\phi_{sam}(\omega) - \phi_{ref}(\omega))} \tag{2.4}$$

where $|\widetilde{E}_{sam}(\omega)|$, $\phi_{sam}(\omega)$, $|\widetilde{E}_{ref}(\omega)|$ and $\phi_{ref}(\omega)$ are the magnitudes and phases of the measured and Fourier transformed electric fields corresponding to the reference and sample scans respectively.

2.4.1 Modelling the theoretical transfer function with Fresnel coefficients

To understand how the transfer function relates to spectroscopic properties of that of the sample material, the interaction of the propagating pulse can be modelled using the Fresnel coefficients. They describe the transmission and reflection at a boundary between two materials (a and b), and the propagation through a single material, given by $r_{ab}(\omega)$, $t_{ab}(\omega)$ and p_a respectively. They are complex, frequency dependent and are related to the complex refractive index $\tilde{n}_a(\omega) = n(\omega) + ik(\omega)$ as follows,

$$r_{ab}(\omega) = \frac{\tilde{n}_a(\omega) - \tilde{n}_b(\omega)}{\tilde{n}_a(\omega) + \tilde{n}_b(\omega)} \tag{2.5}$$

$$t_{ab}(\omega) = \frac{2\tilde{n}_a(\omega)}{\tilde{n}_a(\omega) + n_b(\omega)} \tag{2.6}$$

$$P_a(\omega, L) = e^{-i\tilde{n}_a(\omega)\omega L/c} \tag{2.7}$$

where ω is the angular frequency, L is the propagation distance and c is the speed of light.

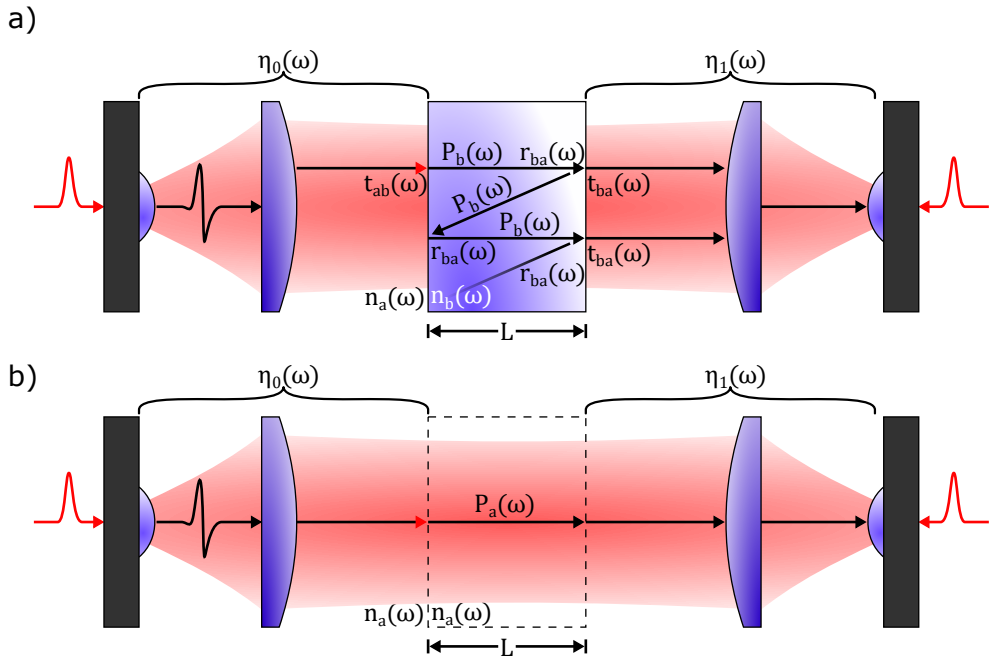

FIGURE 2.6: Theoretical model of the propagation of terahertz radiation through a sample a) and a corresponding reference b). The pulse is travels from the emitter, columnating lenses and the sample. The path of the radiation is indicated by arrows with the reflections, transmission and propagation labelled by the corresponding Fresnel coefficients. $\eta_0(\omega)$ and $\eta_1(\omega)$ are place holder variables to describe the path of the radiation between the emitter and sample, and sample and detector respectively.

Figure 2.6 shows how for a medium of refractive index $\tilde{n}_a$ surrounding a sample of thickness L and refractive index $\tilde{n}_b$, the light propagates and interacts with the sample. In this description, two additional terms are present, $\eta_0(\omega)$ and $\eta_1(\omega)$. These are placeholder terms describing the arbitrary power of the electric field from the emitter, the detection sensitivity, and any transformations that occur outside of the sample path due to lenses etc that are assumed identical between the two measurements. Therefore, it can be seen from figure 2.6(a) that the electric field corresponding to the sample measurement can be described as

$$E_{sam}(\omega) = \eta_0(\omega) \cdot t_{ab}(\omega) \cdot P_b(\omega) \cdot t_{ba}(\omega) \cdot \eta_1(\omega) \cdot \tilde{F}_P(\omega), \tag{2.8}$$

where the internal reflections form a Fabry-Perot etalon

$$\widetilde{F}_P(\omega) = \sum_{j=0}^{m} \left[r_{ba}^2(\omega) \cdot P_b^2(\omega) \right]^j, \tag{2.9}$$

where m is the total number of internal reflections that are detected. The division of the sample by the reference pulse results in many of the above terms cancelling out, giving the following general theoretical transfer function, It is important to note the lack of focusing lenses in figure 2.6. They are omitted for clarity, but also because the assumption will be made that even at a focus, the sample is thin enough, that a plane wave approximation can be made. Further, the focus is loose enough to negate effects of the Gouy shift.

$$\widetilde{H}_{theo}(\omega) = \frac{\widetilde{E}_{sam}(\omega)}{\widetilde{E}_{ref}(\omega)} = \frac{t_{ab}(\omega) \cdot P_b(\omega) \cdot t_{ba}(\omega)}{P_a(\omega)} \cdot \widetilde{F}_P(\omega). \tag{2.10}$$

For scans of samples where the combination of refractive index and thickness that satellite pulses can be effectively removed through applying a windowing function, the Fabry-Perot etalon term, $F_P(\omega)$, becomes unity and so the experimental transfer function simplifies to a *zero reflection transfer function*,

$$\widetilde{H}_{zero}(\omega) = \frac{4\tilde{n}_a(\omega)\tilde{n}_b(\omega)}{(\tilde{n}_a(\omega) + \tilde{n}_b(\omega))^2} \cdot e^{-i[\tilde{n}_b(\omega) - \tilde{n}_a(\omega)]\omega L/c}. \tag{2.11}$$

Alternatively, in the case where the sample's reflections are all measurable in the time-domain window (i.e. all the reflections that have peaks above the noise level are measurable in the limited time-domain window), the Fabry-Perot etalon term becomes a sum to infinity, which can be represented by the geometric series,

$$\widetilde{F}_P(\omega) = \sum_{j=0}^{\infty} \left[r_{ba}^2(\omega) \cdot P_b^2(\omega) \right]^j = \frac{1}{1 - r_{ba}^2(\omega) \cdot P_b^2(\omega)}, \tag{2.12}$$

and therefore one obtains the *infinite reflection transfer function*,

$$\widetilde{H}_{inf}(\omega) = \frac{\frac{4\tilde{n}_a(\omega)\tilde{n}_b(\omega)}{(\tilde{n}_a(\omega)+\tilde{n}_b(\omega))^2} \cdot e^{-i[\tilde{n}_b(\omega) - \tilde{n}_a(\omega)]\omega L/c}}{1 - \frac{(\tilde{n}_b(\omega)-\tilde{n}_a(\omega))^2}{(\tilde{n}_a(\omega)+\tilde{n}_b(\omega))^2} \cdot e^{-i2\tilde{n}_b(\omega)\omega L/c}}. \tag{2.13}$$

The infinite reflection transfer function has the advantage of retaining the spectral information not removed by windowing, but suffers from being much a more complicated expression. However, both the resulting theoretical transfer functions contain only two unknown variables: the real and imaginary components to the complex refractive index, but they are not analytically solvable, as they are complex and the exponential

terms are non-bijective. Therefore, numerical fitting is required to fit values of n and k to the transfer functions, or in some cases, approximations can be made.

2.4.2 Analytical inspection through small k approximation

In the case of a sample (and its reference) where the extinction coefficients are smaller than the real refractive index, ($k_b \ll n_b$ and $k_a \ll n_a$), it is possible to extract approximations for the parameters analytically. The contribution to the phase of the transfer function from the transmission Fresnel terms are insignificant as compared to that of the propagation coefficient. The real refractive index in the exponent is therefore taken as the only contributor to the phase of the transfer function, and so it follows that

$$\phi_H(\omega) \approx -(n_b(\omega) - n_a(\omega))\frac{\omega L}{c}. \tag{2.14}$$

The equation can then be rearranged for the sample's real refractive index, n_a,

$$n_b(\omega) = -\phi_H(\omega) \cdot \frac{c}{\omega L} + n_a(\omega). \tag{2.15}$$

The magnitude of the transfer function and the relation in equation 2.11 is taken, with values of the real refractive index calculated in 2.15 substituted to calculate the imaginary part of the refractive index, $k(\omega)$,

$$k(\omega) = \ln\left[\frac{1}{|H(\omega)|} \cdot \frac{4n_a(\omega)n_b(\omega)}{(n_b(\omega) + n_a(\omega))^2}\right] \cdot \frac{c}{\omega L}. \tag{2.16}$$

The method is particularly appropriate for slab samples of high refractive index and low absorption, measured in a reference of air. The limits of the technique are discussed in chapter 3.

2.4.3 Numerical fitting using the Newton-Raphson method

Once the theoretical and experimental transfer function has been obtained, the complex refractive index can be fitted to the model. As there are two unknown variables, the real and imaginary components of the complex refractive index, it is necessary to use an iterative numerical method to retrieve the desired parameters. A commonly used algorithm is the Newton-Raphson method. It is an efficient procedure whereby a root to a function is found by first making a guess value, and improving that guess by using the derivative of the function evaluated at that point. The next guess value

is taken to be where a straight line gradient intersects the x-axis and the process is repeated eventually converging on the root. The procedure terminates after the function is evaluated to be less than with a predetermined accuracy. A subsequent guess, x_{n+1}, is found by the expression

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}, \tag{2.17}$$

where f(x) = 0 is the equation to be solved, $f'(x_n)$ is the evaluation of the derivative with respect to x at the previous guess, x_n.

To implement this method for fitting experimental data to the transfer functions described previously, it is advantageous to use the natural logarithm of equations 2.11 and 2.11. This separates the phase and amplitude of a complex number, which, as shown previously with regard to the small k approximation method, they tend to be predominantly influenced by the real and the imaginary parts of the refractive index respectively. Therefore, the equation the method aims to solve is the following,

$$\ln\left[\widetilde{H}_{theo}(\tilde{n}(\omega))\right] - \ln\left[\widetilde{H}_{exp}(\omega)\right] = 0. \tag{2.18}$$

The equation 2.18 is independently solved at each frequency value, so for frequency ω_j, the evaluations of the natural logarithm of the zero and infinite transfer functions are as follows,

$$\widetilde{G}_{zero}(\tilde{n_b}) = \ln\left[\widetilde{H}_{zero}(\tilde{n})\right] = \ln\left[\frac{4\tilde{n}_a\tilde{n}_b}{(\tilde{n}_a+\tilde{n}_b)^2}\right] - \frac{i[\tilde{n}_b-\tilde{n}_a]\omega_j L}{c} \tag{2.19}$$

$$\widetilde{G}_{inf}(\tilde{n_b}) = \ln\left[\widetilde{H}_{inf}(\tilde{n})\right] = \widetilde{G}_{zero}(\tilde{n_b}) - \ln\left[1 - \frac{(\tilde{n}_b-\tilde{n}_a)^2}{(\tilde{n}_a+\tilde{n}_b)^2}\cdot e^{-i2\tilde{n}_b\omega_j L/c}\right] \tag{2.20}$$

and their respective derivatives

$$\frac{\partial}{\partial\tilde{n}_b}\widetilde{G}_{zero}(\tilde{n_b}) = \frac{1}{\tilde{n}_b} - \frac{2}{\tilde{n}_a+\tilde{n}_b} - \frac{i\tilde{n}_b\omega_j L}{c} \tag{2.21}$$

$$\frac{\partial}{\partial\tilde{n}_b}\widetilde{G}_{inf}(\tilde{n_b}) = \frac{\partial}{\partial\tilde{n}_b}\widetilde{G}_{zero}(\tilde{n_b}) + \left(\frac{2\frac{(\tilde{n}_b-\tilde{n}_a)^2}{(\tilde{n}_b+\tilde{n}_a)^2}\cdot e^{-2i\tilde{n}_b\omega_j L/c}}{1-\frac{(\tilde{n}_b-\tilde{n}_a)^2}{(\tilde{n}_a+\tilde{n}_b)^2}\cdot e^{-i2\tilde{n}_b\omega_j L/c}}\right)\cdot\left(\frac{1}{(\tilde{n}_b-\tilde{n}_a)^2} - \frac{1}{(\tilde{n}_b+\tilde{n}_a)^2} - \frac{i\omega_j L}{c}\right) \tag{2.22}$$

By taking the natural logarithm of the transfer function before performing any root-finding or minimisation method, the equation 2.18 will have fewer oscillations due

to the removal of the complex exponential term as discussed in the first reporting of these extraction algorithms [95]. It can be seen however, that this benefit is only the case in the zero reflection model, where equations 2.19 and 2.21 are significantly less complicated as compared to equations 2.20 and 2.22, where complex exponential terms are still present, producing oscillations in $\widetilde{G}$.

2.4.4 Note on the limits of Newton-Raphson fitting

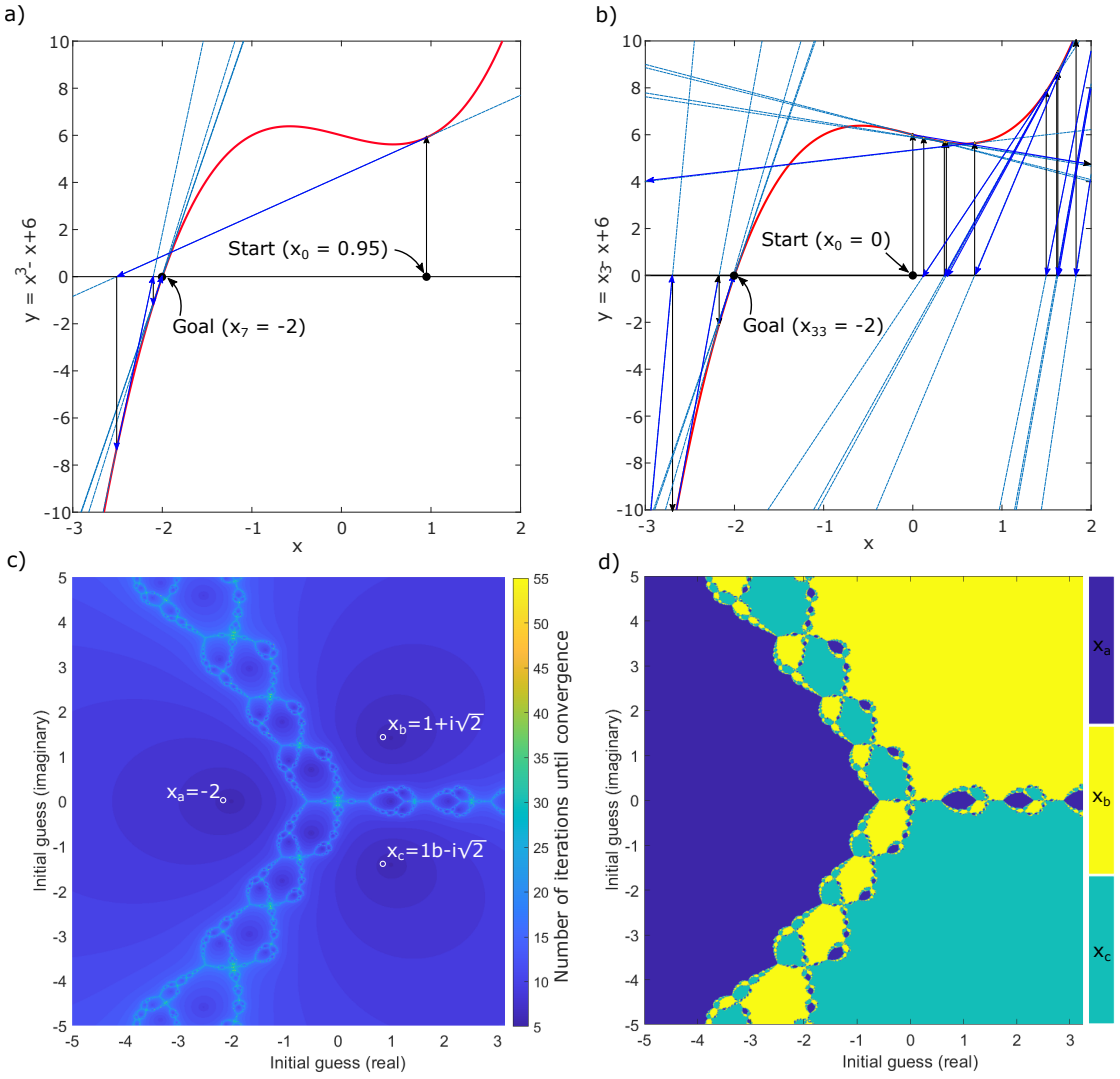

FIGURE 2.7: The Newton Raphson of an example function $y = x^3 - x + 6$ showing the difference in number of iterations needed for convergence between initial guesses (a) $X_0 = 0.95$ and (b) $X_0 = 0$. The number of iterations depending on the complex starting guess shows fractal features (c) with the algorithm converging on a solution that is not always near the starting guess (d).

Figure 2.7(a) shows an example of how the Newton Raphson method algorithm proceeds for solving the function $y = x^3 - x + 6$ with a starting guess, $x_0 = 0.95$. After just seven iterations, it converges onto the solution, x = -2, with a defined accuracy

of 0.1 %. However, even such a relatively simple function can cause problems if the guess is not chosen with care. For example, in figure 2.7(b), the initial guess is chosen as $x_0 = 0$, a not unlikely guess given the zeroth order term of the function. Starting from this guess takes 33 iterations to converge on the correct solution, due to the negative gradient and subsequent inflection point. Furthermore, in the complex domain, where the method is required for the complex parameter extraction of THz-TDS data, more complexity arises from this function having multiple roots and areas in the plane where initial guesses converge to unexpected roots. Figure 2.7(c) shows where the three complex roots of the example function are, and the colour corresponds to how many iterations it takes for an initial guess. Initial guess that are in close proximity to where the roots are located on the complex plane, lead to few iterations in order to converge, and this gradually increases. However, in the boundaries between other roots, fractal patterns emerge where there are sudden increases in iteration count. Figure 2.7(d) shows where the algorithm converges to if starting at a particular point. The concern here is that even for a relatively simple function, without the aid of such a diagram, it is not obvious which root the algorithm will converge to after it is initialised.

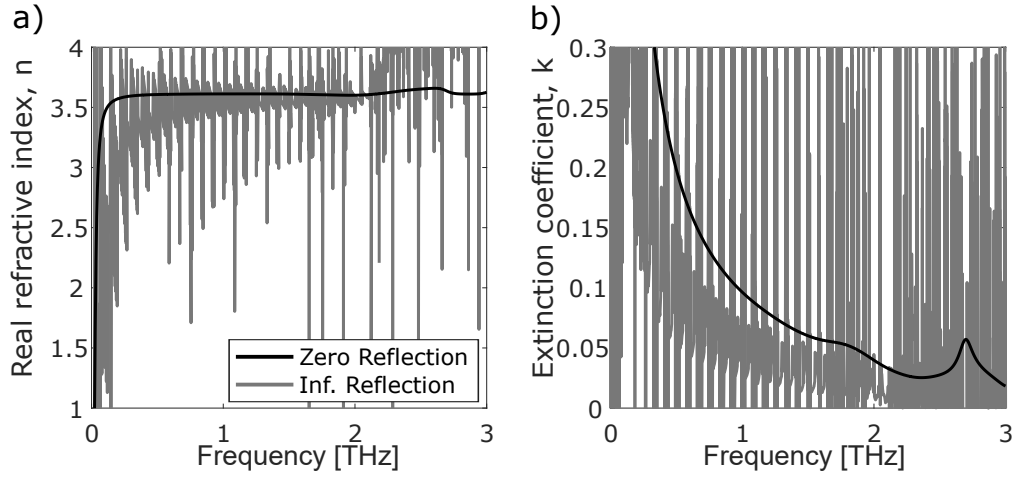

FIGURE 2.8: The extracted real refractive index (a) and imaginary refractive index (b) for a GaAs sample of 500 μm thickness using the Newton-Raphson Method, comparing the zero and infinite reflection transfer functions. The infinite reflection model shows poor convergence in both cases, where as the zero reflection model has a stable solution.

This volatility is displayed when the root finding method is applied to the case of the infinite reflection model. A scan of GaAs was taken where all reflections were visible in the time-domain window, as subsequent reflections were below the noise floor. Both the infinite reflection model 2.13 and the much simpler zero reflection model 2.11 (with the reflections removed in the time-domain) were fitted to the data. Figure 2.8 shows the comparison of the two fitting attempts. The zero reflection method results in a stable fitting with the expected refractive index of 3.5856 [96], but the infinite reflection model fails to converge in most cases. For this measurement at least, if one wants to

exploit the extra information offered by the echo pulses, an alternative fitting method or a different technique should be used.

2.5 The echo method

It is possible to exploit the extra information the multiple reflections from a sufficiently thin and high refractive index sample offers, while adding benefits to the reduced acquisition time, variability in measurements and more stable fitting. Here, the so called *echo method* is presented, whereby the first pulses transmitted through the sample is used as reference measurement to the following, 'echo pulse' resulting from the internal reflection. The major advantage of this technique is it removes the need for an air reference, and the same fitting methods can be used as described earlier in this chapter from one scan of the sample, reducing errors associated with long term fluctuations in the setup between reference and sample scans (such as laser beam drift) as well as removing the effect of water vapour absorption lines erroneously showing as gain in the extracted absorption spectra. The work was published in [3], and here the same technique is additionally applied to the 500 μm thick GaAs sample used as an example throughout the chapter.

2.5.1 The echo method transfer function for Newton-Raphson fitting

The transfer function for the echo method is the division of the pulse after one internal reflection (the echo pulse), $E_{2nd}(\omega)$ divided by the first pulse transmitting through the sample, $E_{1st}(\omega)$. The resulting transfer function includes only the reflection coefficients and propagation coefficients inside the material, as the transmission coefficients in and out of the sample material are now divided out. The transfer function follows as

$$\widetilde{H}_{echo}(\omega) = \frac{E_{2nd}(\omega)}{E_{1st}(\omega)} = \widetilde{r}_{ba}(\omega)^2 \cdot \widetilde{P}_a(\omega)^2, \tag{2.23}$$

where $\widetilde{r}_{ba}(\omega)$ and $\widetilde{P}_a(\omega)$ are the reflection and propagation coefficients respectively. The natural logarithm creates a function that has no exponential terms as follows,

$$\widetilde{G}_{echo}(\tilde{n}_b) = \ln\left[\frac{(\tilde{n}_b - \tilde{n}_a)^2}{(\tilde{n}_a + \tilde{n}_b)^2}\right] - \frac{-i2\tilde{n}_b\omega_j L}{c}, \tag{2.24}$$

where $\tilde{n}_a$ and $\tilde{n}_b$ are the complex refractive indices of the surrounding medium and the sample respectively at the frequency ω_j, L is the thickness of the sample and c the speed of light. This function is easily differentiated with respect to $\tilde{n}_b$ as,

$$\frac{\partial}{\partial \tilde{n}_b}\widetilde{G}_{echo}(\tilde{n_b}) = -\frac{4\tilde{n}_a}{\tilde{n}_a^2 - \tilde{n}_b^2} - \frac{i2\omega_j L}{c} \tag{2.25}$$

which is a function that is much more likely to converge with the Newton-Raphson method than previous transfer functions with reflections.

2.5.2 Results with Newton-Raphson fitting: Gallium Arsenide

A sample of semi-insulating gallium arsenide (SI-GaAs) with a thickness of 500 μm was cleaved from a 2 inch diameter wafer, and placed in the THz-TDS beam path as described earlier in the chapter. Multiple 30 second long scans of this sample were taken (36), at a scanning rate of 1 Hz. Additionally, a 2 minute long air reference scan was taken with the sample removed. A windowing function was applied to both the time-domain traces of the air scan, and at two positions of each sample scan to isolate the first transmission pulse and the first echo pulse. Figure 2.9(a) shows the time-domain traces of the air measurements and the GaAs sample, with the windowing applied. The inset shows the amplitude of the spectrum of the air and sample measurement raw, with that of the windowed pulses on top. It can be seen that the second pulse has a reduced bandwidth as the amplitude is reduced across the entire spectrum resulting in more frequency components being buried beneath the noise floor.

The Newton-Raphson algorithm was then used to extract the complex refractive index using the echo method (equations 2.24 and 2.25) and the conventional, zero reflection method (equations 2.19 and 2.21). The mean of the real and imaginary refractive index are shown in figures 2.9(b) and (c) respectively, with the standard deviation shown as a shaded region. Sub-figures (d) and (e) compare the variation in both showing the the standard deviation of the echo method divided by that of the zero reflection method. It can be seen that the consistency between measurements is higher for the self referencing method when extracting the real refractive index by an order of magnitude, but is much closer to unity when comparing the imaginary term. The real refractive index is consistent with the literature value of around n = 3.6 [97], although it is slightly larger here. This could be due to an error in thickness combined with the limits of time-delay accuracy with thin samples. The standard deviation in general for this parameter is very small for both extraction techniques. It must be noted here that there is a significant difference between the extracted imaginary values. This offset could be contributed to the error in phase at low frequencies, which can also be seen in the real refractive index as the values diverge with reduced frequency. It is possible that better phase correction is required, or that the length parameter needs to be adjusted.

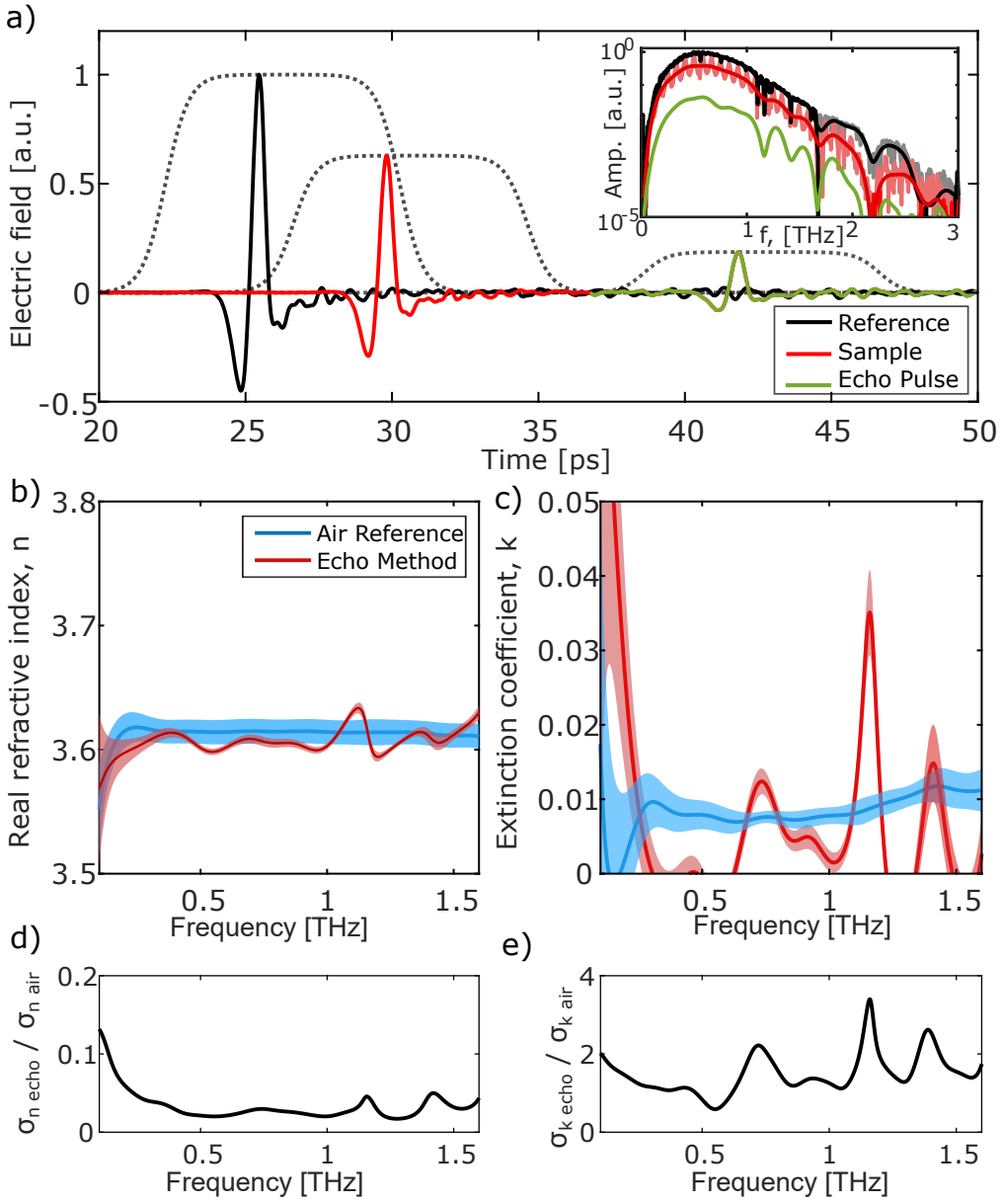

FIGURE 2.9: Echo method with a sample of a 500 μm thick, semi-insulating GaAs. Multiple scans of the sample are taken and both the first transmitted pulse and the first echo pulse due to internal reflections are windowed in the time-domain(a), along with a reference air scan for comparing extraction methods. The inset shows the amplitude of the FFTs performed on these measurements. The real (b) and imaginary (c) refractive indices are extracted and averaged using the air-reference approach (blue) and echo method (red), with the error shown as a shaded region. The relative standard deviation of the two methods is shown for the real (d) and imaginary (e) components.

2.5.3 Small k approximation with the echo method

As previously discussed in the chapter, it is possible to analytically solve an approximation to the zero reflection transfer function if the real refractive index is greater than

the imaginary part. Therefore, only the real refractive index in the propagation coefficients is taken to be responsible of the phase of the transfer function and so it follows that

$$n_b(\omega) = -\phi_H(\omega) \cdot \frac{c}{2\omega L}, \tag{2.26}$$

where ϕ_H is the phase of the echo transfer function. By approximating that the reflection coefficients are real, the imaginary part of the refractive index effects the amplitude of the transfer function, $|H_{echo}(\omega)|$, from the attenuation from the propagation, it is possible to evaluate it as

$$k_b(\omega) = \frac{c}{2\omega L}\left(|H_{echo}(\omega)| - 2\ln\left[\frac{n_a(\omega) - n_b(\omega)}{n_a(\omega) + n_b(\omega)}\right]\right). \tag{2.27}$$

2.5.4 Results of the echo method with the small k approximation: lithium niobate

Lithium niobate ($LiNbO_3$) has a high refractive index (n > 6 and a low imaginary component (k < 0.1) which makes it an ideal candidate for the use of the approximation echo method.

The measurements were performed almost identically to that of the SI-GaAs sample but with longer scans in this case, as each of the 25 scans was given 2 minutes of averaging improving the SNR. Both the reference and sample were placed in a 10 mm aperture, with no focusing lenses making alignment easier and avoiding any potential issues with the sample not being at focus, and ensuring the radiation can be approximated to a plane-wave.

Figure 2.10 shows the results for the small approximation fitting for the echo method, with literature values included for comparison. Terahertz spectra of lithium niobate varies a significant amount in the literature, as it is sensitive to growing conditions, as well as phase unwrapping issues due to the resonance present at low terahertz frequencies, as the THz-TDS signal starts to diminish. It can be seen, as compared to the air referencing method, that the variation between measurements is significantly less with the echo method, up until a frequency of 1 THz. The Fourier transform of the pulses and reference reveal that the available bandwidth of the echo pulse is less than the first pulse, which in turn is significantly less than that of the reference. These measurements are an improvement on the previous, $GaAs$ measurements, benefiting from higher signal to noise thanks to the longer acquisition times. Still, it follows that this method, as a function of both these shorter bandwidth signals would therefore have a more limited applicability with regards to frequency. The extinction coefficient shows little

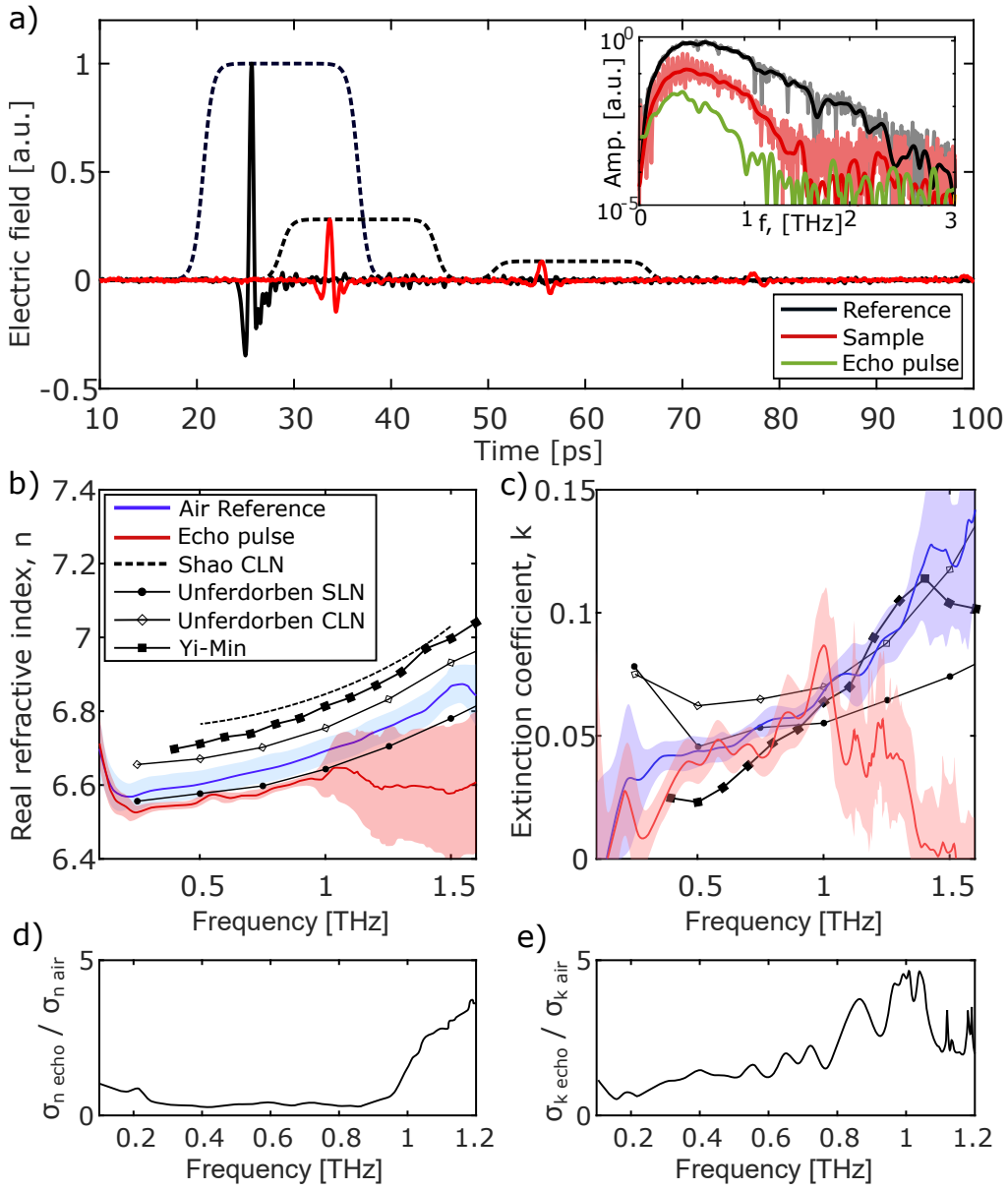

FIGURE 2.10: Echo method with a sample of a 500 μm thick lithium niobate. The time-domain signals and the applied window functions (a) with the corresponding frequency spectrum shown in the inset, including that of the isolated echo pulse. The real (b) and imaginary (c) refractive indices are extracted and averaged using the air-reference approach (blue) and echo method (red), with the error shown as a shaded region. The relative standard deviation of the two methods is shown for the real (d) and imaginary (e) components.

improvement with the echo method at any frequency, but remains consistent with the air reference measurement until 1 THz. Figure 2.10(d) and (e) more clearly show that the relative error of the echo method is superior for the real refractive index but not for

the extinction coefficient. It is likely that the improved stability in the extracted extinction coefficient as compared to the previous $GaAs$ is due to the slightly high absorption of the $LiNbO_3$ sample and the longer acquisition time.

In both the cases of the $LiNbO_3$ and SI-GaAs measurements, the small k approximation echo method, and fitting by Newton-Raphson was performed and showed negligible difference in the mean and standard deviation of values. They are not shown here but are discussed in the next chapter, with a comparison to machine learning methods. Regardless of which type of echo fitting method is used, the results are the same, and the limits appear to be due to the loss of signal.

2.6 Conclusions

THz-TDS benefits from easy phase retrieval and therefore straightforward fitting for the extraction of refractive index. However, there are a wide range of extraction techniques with different benefits, and there is still much interest in finding more reliable techniques in particular. In this chapter the methods in measuring a sample with a THz-TDS are introduced, with the fundamental aspects of parameter extraction with Fresnel modelling presented. One aspect, the poor fitting that comes with complicated transfer functions is highlighted, and a new method is presented, with one if its advantages being the stable fitting whilst benefiting from the multiple reflections observed in samples of sufficient thickness and refractive index. This new method, the "echo method" is presented that determines the complex refractive index from THz-TDS data in the absence of an air reference measurement. The resulting transfer functions are shown with measurements of semi-insulating gallium arsinide and lithium niobate used to demonstrate that for frequencies where the SNR is sufficient, the echo method extracted parameters achieve lower standard deviation in real refractive index values as compared to the air reference method, but at the cost of a lower usable frequency bandwidth than that of the air reference method. However, other advantages of the echo method include reduced effects of laser power drift, less need for purging of the spectrometer chamber with inert gas, and all the data required for the measurement can be obtained while keeping the sample in place. Future work will include combining the echo method with the air reference method to produce a transfer function that requires no thickness measurement, and even output the thickness without knowing the refractive index, with a frequency dependent uncertainty. In the case for aqueous protein samples, where absorption is too high to have measurable internal reflections, the concepts of this method can be applied with the use of a highly reflective substrate preceding the protein solution.

Chapter 3

Machine learning for THz-TDS parameter extraction

By virtue of measuring the terahertz pulse in the time-domain, phase information is retained that would otherwise be lost in typical absorption spectroscopy. This makes frequency resolved parameter extraction a far simpler process, mathematically speaking. However, there are still practical challenges involved, and steps taken to address these are not consistent between research groups [98]. In this chapter, machine learning is presented as an excellent candidate to automate the process of parameter extraction, where the implementation can be easier to implement, runs faster, more accurate and has the potential to extend beyond conventional data-analysis. The work discussed in this chapter has been been published in [2].

3.1 Introduction

The key principle in using THz-TDS to obtain material parameters is to measure the transmitted or reflected electric field of a reference and a sample, whereby obtaining the sample's complex refractive indices is achieved without the need to use Kramers-Kronig (K-K) relations. The amplitude of electric field (not merely the intensity) is mapped, in time, and therefore the phases information of the signal is easily extracted when performing a Fourier transform. Despite this advantage however, it has yet to completely break through into the industrial setting. A key issue is that obtaining materials parameters can still be a challenging process, as multiple analysis steps are required, each of which can be prone to introducing errors into the calculations, such as phase unwrapping and transfer function convergence. This aspect contributes to the wide variability in results between researchers around the world which is an issue important enough to warrant its own research [98, 99]. Additionally, there is a time penalty for both the implementation and computational run-time for the iterative,

root-finding algorithms that provide the accuracy required for sensitive spectroscopic measurements. To this end, it is advantageous to look towards automated methods of parameter extraction. The previous chapter contains a full discussion of the intricacies of these methods.

Machine learning is a candidate for such an alternative approach. By using an appropriate training algorithm with training data, an artificial neural network (ANN) is able to map the relationship between the input training data and its associated output. This is achieved by adjusting parameters inside the network (the neuron connections containing weights and biases) to improve the prediction autonomously, through a training algorithm. A neural network of just one hidden layer of neurons has enough flexibility to be able to approximate any function [100] (with an arbitrary activation function [57]). Therefore, it is expected that such a network could analyse THz-TDS data and output the associated material's refractive index, with the accuracy being determined by the size of the data set and number of neurons in the hidden layer.

The application of ANNs to THz-TDS has primarily been concerned with classification problems [101, 102, 103, 104], such as organic compounds [51], food [105, 106], and biological samples [52, 53]. The classification algorithms can be used to analyse terahertz images [107] for medical [108, 109, 110], and security applications [111]. Beyond classification problems, machine learning can be utilised for regression fitting, such as to find concentrations of specific molecules [54, 55, 112, 113, 114] or for finding the thickness of thin samples of known refractive index [115, 116].

These neural fitting methods demonstrate the powerful opportunities that machine learning has to offer in THz-TDS analysis, however they rely on the prior knowledge of the sample's refractive index, which can be a problematic assumption as the refractive index of a known sample is often altered by the manufacturing history or growth of the material. In the following chapter, machine learning is demonstrated as a method which has the potential to overcome the issues raised where it is shown that an ANN can estimate the complex, frequency dependent refractive index. With a training set comprised of simulated data, the fitting of the experimental data to a theoretical model can be replaced with a trained ANN, whereby minimal coding is required, benefiting in speed of implementation (and run-time) over an iterative method. Further, the potential for the neural network to replace data processing steps in the extraction procedure by employing deep neural network (DNN) and convolutional neural network (CNN) architectures could be possible and therefore improve accuracy. In this case the extracted phase and amplitude of the experimental transfer function from the measurement can be directly fed, showing the potential for a more generalised model. The training data is simulated to encompass any THz-TDS transmission setup, as neural networks are trained over a wide range of material parameter values and frequencies to ensure wide applicability of the neural networks in real world applications.

3.2 The target for machine learning: THz-TDS Extraction

As discussed in detail in the previous chapter, for analysis of THz-TDS data to obtain the complex refractive index three common methods are generally used.

1. In the simplest case the time domain of a spectrum can be analysed to measure the time difference in the peak amplitude of the pulse travelling through air as compared to travelling through the sample, Δt, and calculate an average refractive index using the equation $n = 1 + c\Delta t/L$. This method is extremely simple, and provides no spectroscopic information, however it does give an estimate of the material parameters, especially in the case of a non-dispersive medium.

2. The second widely used method, which shall be continued to be referred to here as the 'analytical method', involves performing a Fourier transform on the time domain data to obtain the intensity and phase of the pulse normalised to the air reference: the experimentally derived transfer function, H_{exp}. From this stage the complex refractive index can be extracted at each frequency by equating the experimental transfer function to a theoretical transfer function that is a function of n and k. The theoretical transfer function is obtained by modelling the light-matter interaction of the sample and the terahertz pulse with the Fresnel equations for transmission through a boundary and a complex exponential attenuation coefficient, however they rely on the assumption that n » k to render the Fresnel coefficients completely real. This assumption allows one to separate the real and imaginary components of the theoretical transfer function. This permits to analytically obtain a value for n and k for every frequency,

$$n(\omega) = -\left[\frac{\phi_H(\omega)}{L} \cdot \frac{c}{\omega} + 1\right] \tag{3.1}$$

$$k(\omega) = \ln\left[\frac{4n(\omega)}{|H(\omega)| \cdot (n(\omega)+1)^2}\right] \cdot \frac{c}{\omega L} \tag{3.2}$$

   where $\phi(\omega)$ is the phase, L is the sample thickness, c is the speed of light and ω is the angular frequency. The amplitude of the transfer function, $|H(\omega)|$, is calculated by division of the sample pulse with the reference pulse (air). The result decreases in accuracy as the loss of the material increases, as well as with reducing frequency, limiting the usefulness of this method.

3. To avoid the approximation issues with the analytical method one can find solutions to the transfer function equations numerically. The experimental transfer function is equated to the theoretical transfer function but without the assumption employed in the previous case. In order for a solution for n and k to be found an iterative root-finding method, such as the 'Newton-Raphson' method,

is commonly used for THz-TDS analysis, where one fits the experimentally derived transfer function to a theoretical model. This method can obtain values of complex refractive index with excellent accuracy, however, like all root finding and minimisation algorithms, it is possible for the method to converge on a local, not global, solution. Furthermore, many iterations are required which can take a long time to process.

Here, machine learning techniques are introduced for analysing THz-TDS data to extract spectroscopic parameters. The ultimate goal is to replace the entire extraction process, but to begin with this approach, the scope is limited to solving the theoretical transfer function,

$$\widetilde{H}_{theo}(\omega) = \frac{4\tilde{n}(\omega)}{(\tilde{n}(\omega)+1)^2} e^{-i(\tilde{n}(\omega)-1)\omega L/c} \tag{3.3}$$

for values of $\tilde{n}(\omega)$, the complex refractive index, which is a function of angular frequency ω; L, which is the known material thickness and c, which is the speed of light. The real part of the complex refractive index is the real refractive index, $n(\omega)$ and the imaginary part is given by the extinction coefficient, $k(\omega)$. The function $H_{theo}(\tilde{n})$ is the mapping of a pulse passing through a material of thickness L, with no additional reflection terms. It is non-bijective (there are multiple values of phase that produce the same output) by virtue of the complex exponential term, meaning that it can only be analytically solved in one direction: inputting n and k to obtain amplitude and phase. By creating a large data set of amplitude and phase values for various values of n and k, an ANN can be trained to predict from the phase and amplitude the target values of n and k.

As the application is terahertz spectroscopy n, k and ω are restricted to well defined bounds. THz-TDS also has a well defined bandwidth of usable signal and measuring materials of limited refractive indices and thicknesses is targeted therefore it is only necessary to train a model to solve the equation for a limited parameter space. Figure 3.1(a) shows how for many materials, their parameters at 500 GHz typically lay at $1 < n < 10$ and $0 < k < 1$ [3, 117, 118, 119, 120, 121]. Training a model with values within these limits reduces the scope of the problem significantly without making any mathematical approximations. Of course, the neural network approach is in itself a form of approximation, and its accuracy will be explored in this work.

3.3 Replacing traditional fitting with a neural network

Figure 3.1(b) shows an outline of the processing steps, with the ultimate goal of replacing as many of these as possible to reduce human error and loss of data for example

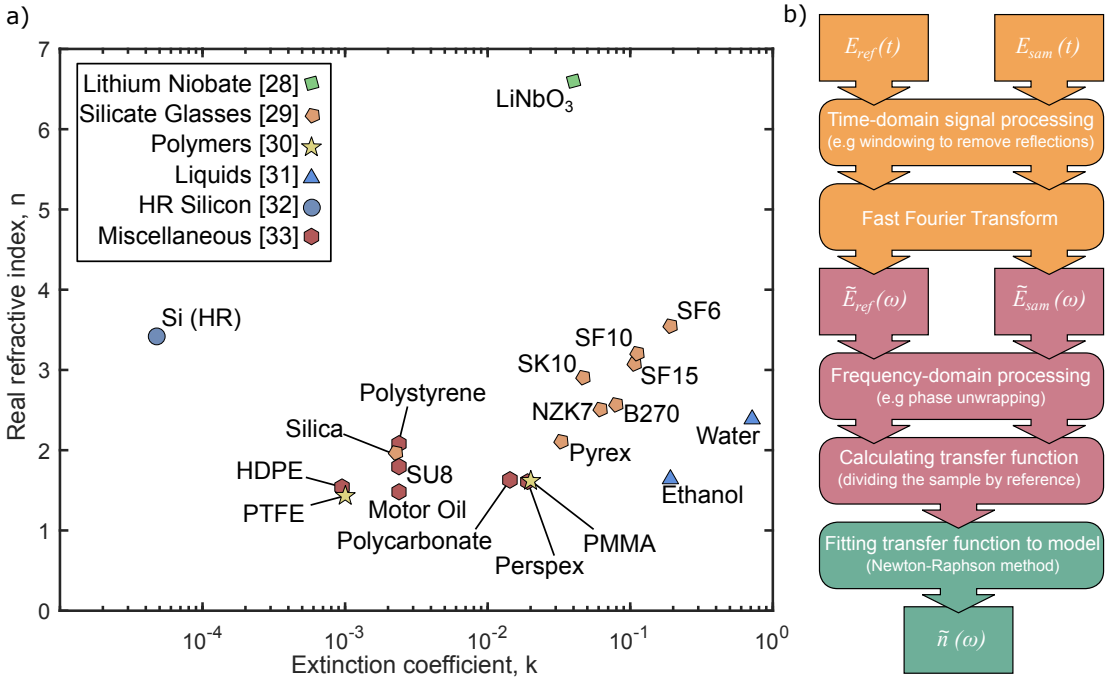

FIGURE 3.1: (a) Examples of materials characterised with terahertz spectroscopy, showing the real and imaginary refractive index at 500 GHz. (b) Parameter extraction flow chart outlining the steps to extract the complex refractive index from time-domain signals, including the time-domain processing steps (orange), frequency-domain processing (pink) and fitting the theoretical model to obtain the complex refractive index (green). Each processing step can be a point of error if care is not taken with analysis.

removing reflections by windowing. Initially, ANNs were explored as candidates to replace the final stage in the parameter extraction process for THz-TDS data, the fitting function.

3.3.1 Creation of Training Sets

To generate simulated training sets for the machine learning algorithm firstly an input pulse is required. The input pulses are created by defining a function within Matlab, which creates a shape similar to that of an experimentally measured pulse, characteristic of the PCA emitter used in our lab. It is an approximation of the physics of a photo-conductive antennae (PCA) emitters. The generated photo-current an incident optical, femtosecond pulse is given by [122]

$$J(t) \propto \left\{ \exp\left(\frac{\tilde{\tau}_{las}^2}{4\tau_{em}^2} - \frac{A\cdot t}{\tau_{em}}\right) \operatorname{erfc}\left(\frac{\tilde{\tau}_{las}}{2\tau_{em}} - \frac{A\cdot t}{\tilde{\tau}_{las}}\right) - \exp\left(\frac{\tilde{\tau}_{las}^2}{4\tilde{\tau}_{em}^2} - \frac{A\cdot t}{\tilde{\tau}_{em}}\right) \operatorname{erfc}\left(\frac{\tilde{\tau}_{las}}{2\tilde{\tau}_{em}} - \frac{A\cdot t}{\tilde{\tau}_{las}}\right) \right\}, \quad (3.4)$$

where $\tilde{\tau}_{las} = \tau_{las}/(2\sqrt{\ln 2})$ and $\tilde{\tau}_{em}$. τ_{las} is the laser pulse duration, τ_{em} is the recombination time, and t is time. The additional term A that multiplies the time axis is used to

simply stretch or compress the resulting pulse to adjust bandwidth while maintaining the approximate shape in time and frequency. The amplitude of the resulting electric field in the time-domain is proportional to the derivative of the generated photocurrent,

$$E_{THz}(t) \propto \frac{dJ(t)}{dt} \tag{3.5}$$

The function produces the characteristic curve of an experimentally measured pulse in the frequency domain, and can be seen in figure 3.2. Important features to note here is the sharp rise in signal at low frequencies and a much slower decline in signal after 500 GHz. These features will become important when developing a more sophisticated model, where noise is also simulated, but for this implementation any arbitrary pulse shape can be used. It is preferential to use a simulated pulse opposed to a real, measured pulse as in the case of linear materials, the resultant transfer function will be independent of the pulse shape.

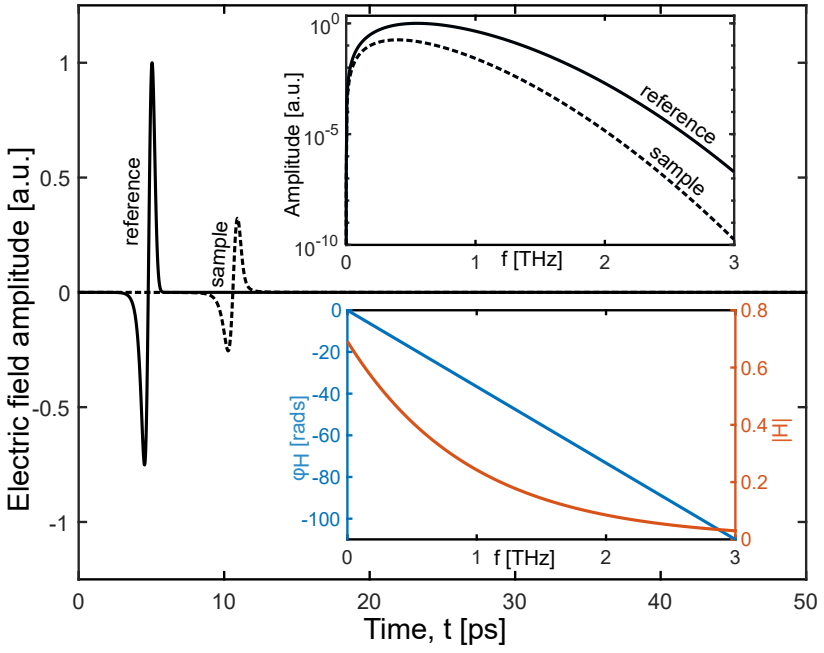

FIGURE 3.2: Simulated pulse in the time domain propagated through air and a sample of n = 3.5, k = 0.1, L = 0.5 mm. The top inset shows the frequency spectra of the reference and sample, and the bottom shows the (unwrapped) phase and magnitude of the transfer function when the two signals are divided.

Once the pulse has been generated it is propagated through the material for given material parameters of refractive index n, k, and thickness L, where only the transmission coefficients in and out the sample and the propagation coefficient are modelled (no reflections). Additionally the pulses are propagated through a column of air of equal thickness to the sample. The data set values of n are varied from 2 to 8, values of k

are varied from 0 to 0.3, and thickness varied from 0.1 to 1 mm, and frequency varied from 0 to 2 THz. Tightening the limits of n and k to these values decreases the training time while also still encompassing a large variety of real material parameters. To create a simulated data set in this way in Matlab is a quick process. 10,000 pulses with 1024 point resolution in frequency required around 80 seconds on a modest laptop computer, demonstrating the ability to generate large training sets relatively quickly. Figure 3.2 shows the time domain signal, along with the frequency spectra of the pulses propagated through the reference (air) and sample in the upper inset, with the magnitude and phase of the transfer function calculated by division shown in the lower inset. The phase of the transfer function requires unwrapping of the phases of the Fourier Transforms of the spectra before the difference is calculated.

Neural networks accept an input array of floating point numbers and produce an output array. The training method employed will be supervised, meaning that the network accepts an input array and is expected to predict a known output array.

As the ANN is to analyse data with the relationship of equation 3.3, each data entry only needs a single value frequency, the corresponding value of amplitude and phase of the transfer function, and the thickness. Each of these values is termed a 'feature'. The target for an entry consists of the values of n and k for the sample material and the corresponding frequency, known as 'labels'. Therefore, each entry corresponds to a random material at a single, random frequency. The data is then normalised, assisting the training of the model. Each label is scaled such that the maximum possible value is 1, and the minimum possible value is 0. For example, a refractive index, n = 2 is given the value 0, and n = 5 becomes 0.5 as our data set only contains values of n between 2 and 8. When the training algorithm calculates its accuracy, it is a function of error in estimated n and k. The k values are much smaller and therefore, if not normalised, errors in its prediction will be weighted less than those in n.

The machine learning is performed in the Matlab Neural Net Fitting Toolbox, employing a standard Levenberg-Marquard (LM) back-propagation function and a mean-squared error loss function. This method takes the weights and biases of a network, and calculates the gradient of the parameter space with respect to the loss function. The weights and biases of the network are then adjusted in such a way to converge to the minimum. For every training session, the simulated data were randomly divided into training (70 %), testing (15 %) and validation (15 %) sets to avoid over-fitting. No noise was inserted into the data model and so the over-fitting was avoided by halting the training once the trained network's performance decreased on the validation data set after six consecutive training iterations. Various configurations of the neural net were trialled, but a network architecture of one hidden layer of 200 fully connected neurons provided sufficient accuracy and training times. This neuron count gave the model a good compromise between flexibility to enable accurate predictions without over fitting - too many degrees of freedom in a model can result in a model that maps

the inputs and outputs of the training data specifically, rather than developing any generality in the relationship between the input and output. For the mapping of four input values to two output values, and with data sets of 10^6 entries and six values each (four features, two labels), the LM method was used as it resulted in excellent fitting, and computational times and memory usage manageable on a desktop PC.

3.3.2 Results and discussion

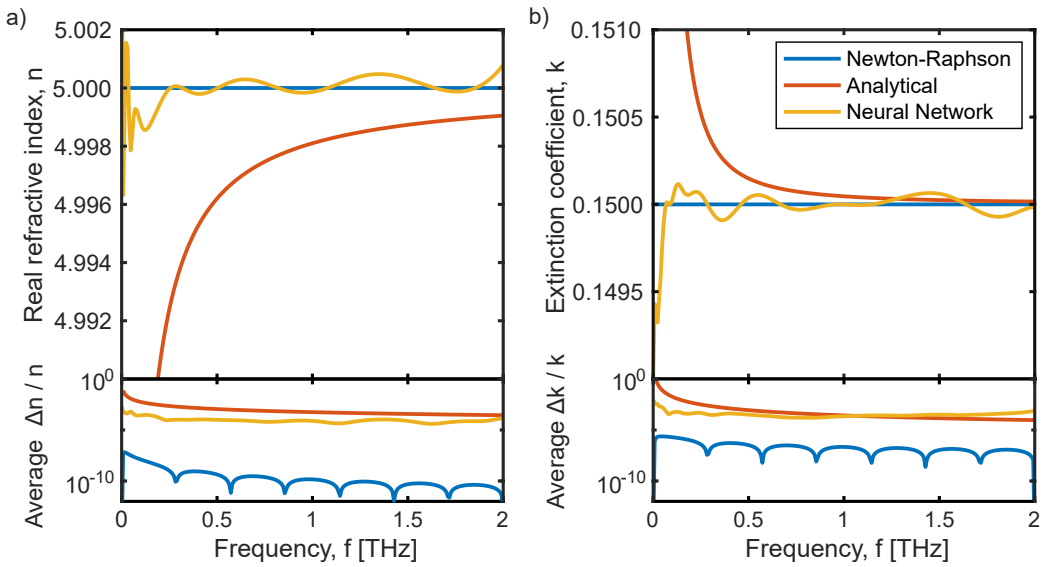

FIGURE 3.3: Examples showing the performance of different extraction methods for the complex refractive index of a sample of n=5, and k=0.15. Below each plot is the average fractional difference across the entire parameter space of training data plotted against frequency.

The networks performance was evaluated in two ways: accuracy in prediction of simulated data generated by the same scheme, and accuracy in predicting real-world, experimental data. For a quantitative evaluation of the training performance, the trained network is used to predict simulated data the model has not seen before, by simply creating another randomised data set. The root-mean-squared-error (RMSE) was then calculated from the differences between the model's prediction and the target ground truth. Additionally, the Newton-Raphson method was used to fit the theoretical transfer function (equation 3.3) to the simulated data for comparison with a typically used iterative method, along with the analytical extraction with k«n approximation. Figure 3.3 compares the results of the three methods. It can be seen that the approximation method is not as accurate as our trained model in predicting the the ground truth values, with the accuracy being surpassed by the iterative method. For this noiseless, simulated data, this is not surprising as the Newton-Raphson method to the transfer function corresponding to the propagation through a material with no reflections has generally stable convergence, the accuracy of which can be set by the user. In doing so this comes at a cost of computation time. The iterative method must loop until a

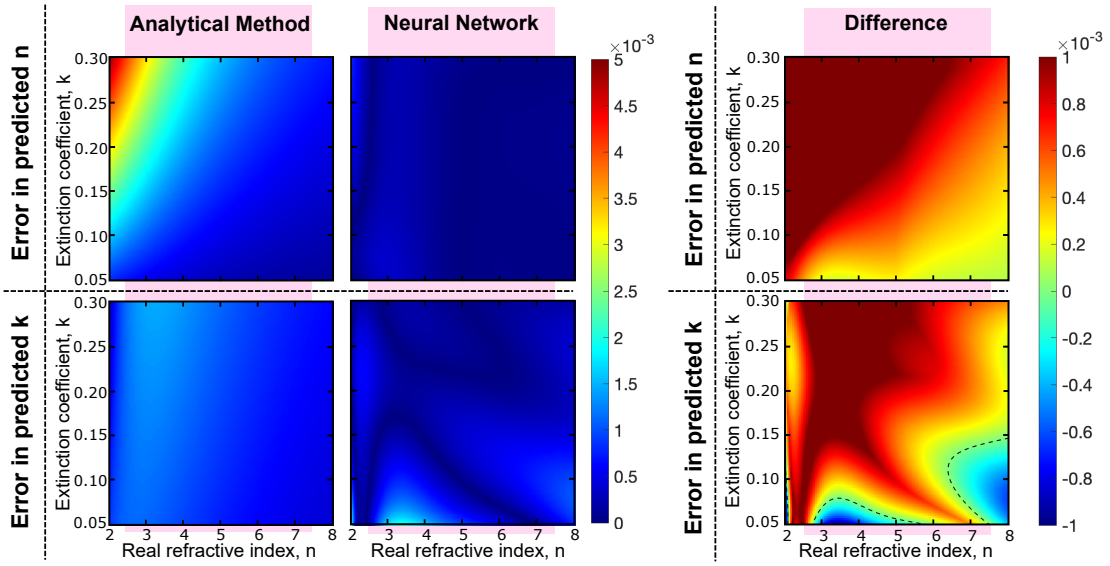

FIGURE 3.4: Surface plots showing the fractional error between the prediction in and n and k from the analytical approximation and neural network implementation methods and their true values for a range of values of n and k (at 500 GHz). The fractional error of the neural network prediction is subtracted from the analytical approximation and the are shown in b), where generally, the neural network improves over the analytical method, as can be seen by values greater than zero.

chosen accuracy is satisfied, or is timed out, which will always vary depending on the quality of the data, the fitting model chosen etc. A fully trained network results in a fitting that will generally be faster, and with a much less variable run-time. In the case of fitting to real data, the average run-time over 1000 iterations for a standard Newton-Raphson was 118 $\pm$ 9 ms, whereas the trained ANN took 60.4 $\pm$ 7.6 ms to complete. Unsurprisingly, the approximated analytical method took much less time to run: 0.296 ms $\pm$ 0.046 ms.

Figure 3.4 expands on figure 3.3 by visualising the accuracy of the neural network and the analytical method for a fixed frequency (f = 500 GHz) and fixed thickness (d = 500 μm) for the complete range of values of n and k used for training. Figure 3.4(a) shows the errors of both methods independently. As expected, the analytical method's accuracy in evaluating the real part of the refractive index drops with decreasing n and increasing k. The error in the evaluation of k is less pronounced, approaching 0.2 % at lower values of n. The neural network's prediction shows a significant improvement when evaluating the real part of the refractive index, and is improved across the entire parameter space. The prediction of k is improved over the analytical method, but not everywhere. Features are present in this surface plot, but they are random due to the nature of the training algorithm, and the impossibility of the LM minimisation finding a global minimum in the loss function. To better compare the two methods, the error of the neural network prediction was subtracted from the evaluation by the analytical method and shown in figure 3.4b). Green in these surface plots show equivalence of the two methods, with increasing to red indicating the neural networks improvement,

and blue its deficit, when compared to that of the analytical method. The landscape of the difference in error of the real refractive index shows that the expected breaking-down of the analytical method towards high k and low n values superseding any error from the neural net prediction. For the error in k plot, there are 'islands' where the neural network is marginally worse than that of the analytical method, but they are small and in theory could be improved with more adjustment to training, as their location is random, and a symptom of the stochastic nature of the machine learning technique.

To further evaluate the model, the refractive indices are extracted from experimental data. Figure 3.5 shows the extracted complex refractive index of lithium niobate measured with a time-domain spectrometer, described by our previous work using the analytical, Newton-Raphson method and the neural network [3]. For this case, it can be seen that in the real and imaginary parts predicted by the neural network are extremely close to the alternative methods.

3.4 Developing an artificial neural network for direct analyses of spectral data

To advance the complexity of our networks to account for more of the stages in the parameter extraction process, input and output data sets must accommodate full frequency-domain information of a measurement rather than a single point in frequency. The neural net trained in the previous section creates a prediction for value of n and k that solves a mathematical function for a given frequency, where there is no context on what the values of n and k should be when looking at the spectrum as a whole. Any errors introduced before the phase and transfer function are calculated will be carried through to the output, namely, the pre-processing of the phase of the transfer function. It is expected that as frequency tends to zero, the phase value also approaches zero when observing the FFT of a broadband signal. However, as the SNR diminishes at low frequencies, the extracted value is consequentially erroneous. This means that when subtracting the phases of the two signals, the phase difference at every frequency will be shifted by the difference of the two incorrect zero-frequency values for each spectrum. Therefore, particularly at low frequencies, the extracted refractive index will be incorrect. For samples with a flat, non-dispersive refractive index, this effect is countered by fitting a straight line, weighted by the SNR, to the phase of each spectrum, and then applying an offset where this line intersects the y-axis. However, as can be seen in figure 3.5(c), lithium niobate is dispersive and has potential features at low frequencies ($< 200\,\mathrm{GHz}$) [3], where the SNR of the spectrometer is low and therefore a linear fit on the phase would give a wrong zero intercept. One has to know the approximate shape of the phase as a function of frequency to correctly counter the phase offset issue. As such, the data sets containing many values for the whole frequency range are

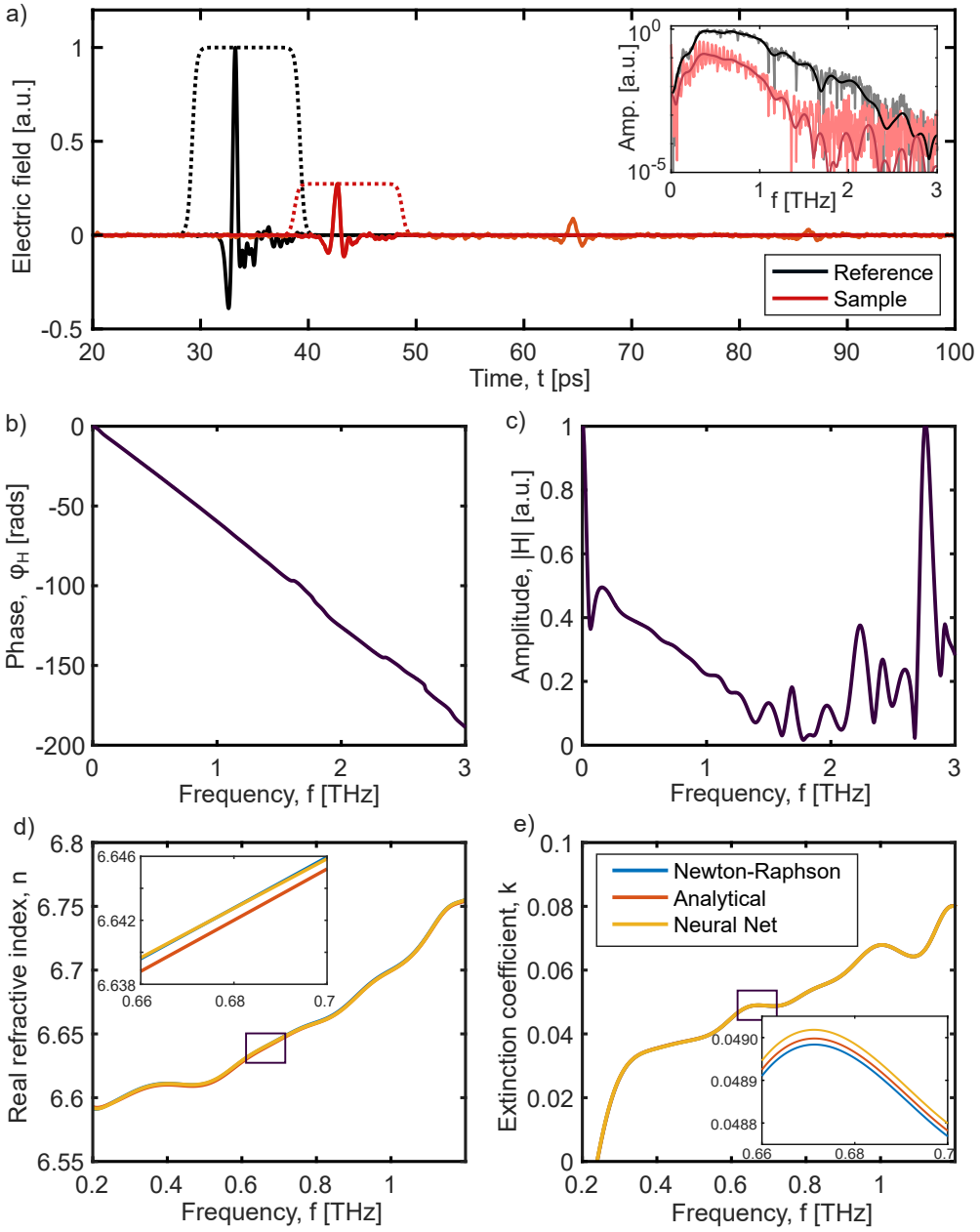

FIGURE 3.5: Performance of the fitting methods on real data, of 500 micron thick lithium niobate. The time domain traces, (a), of the reference and lithium niobate sample, with the window function in place and the resulting spectra shown in the inset. The windowing simplifies the phase (b) and amplitude (c) of the resulting transfer functions to simplify fitting, at the cost of resolution. In the case of this samples, the extracted real (d) and imaginary (e) refractive index are almost identical between methods.

much larger requiring more computer processing power and memory usage. Using the LM method is no longer tractable, scaled conjugate gradient backpropagation (SCG) is instead employed. While one single network layer has the flexibility to approximate any function, more complex network architectures are required for efficiently training models to recognise patterns in data sets with more dimensions, such as pixel data in an images. Spectroscopic data is analogous to image data, but only has two, one dimensional arrays (phase and magnitude). Use of a multiple hidden layer network is attempted, often categorised as a deep neural net (DNN) to analyse data sets with a complete frequency axis.

3.4.1 Creating dispersive training sets with noise

In order to motivate the network to ultimately replace manual steps from the data analysis, more sophisticated training data is required. Pulses propagating through samples with dispersion, modelled simply by generating values of n and k with sums of sinusoids with random frequencies and amplitudes are simulated. It is worth noting, that the simulated samples were not with any particular features corresponding to resonances, but with arbitrary shapes that provide enough variation that the neural network can correlated the changes in phase and amplitude with respect to frequency. Furthermore, noise is added to the propagated pulses, with a simulated 70 dB SNR between the signal at its peak and the noise floor. This introduces a small, but significant offset error to the unwrapped phase, as previously described. For this early experiment, the number of points on the frequency axis to were minimised to 32 points between 0 and 1 THz, creating each entry of the training data sets to be arrays of 64 in length (for both magnitude and phase of the transfer function). Similarly, the target data sets have 32 points each for the real and imaginary parts of the refractive index. Additionally, a thickness value is negated and only 500 μm thick samples are simulated, to further decrease computation requirements. Using the Matlab Deep Learning Toolbox, a model with two hidden layers of 128 neurons and 512 neurons was created. The network was trained with the same parameters as previous, but with training algorithm switched to *scaled conjugate gradient backpropagation*, (SCG). The algorithm is analogous to the LM algorithm in that it uses the gradient of the weights and biases versus the loss function but works on small subsets of the parameters (called batches) rather than the entire network. A *convolutional neural network*, (CNN) [123], was additionally trialed, as this architecture is much more efficient for image processing, where the training data are higher dimensional arrays. The training function here was Matlab's implementation of *adaptive moment estimation*, (Adam) [124] and training a network of one convolution layer and one hidden layer of 200 neurons.

3.4.2 Results and discussion

By training neural nets with data containing simulated noise, and increasing the number of neurons on the input layer to accept a linearly spaced range of points in frequency, a network that still extracts the refractive index is produced. Figure 3.6 shows the current state of the model. It accepts data without looping through each frequency. It has the advantage of being more stable at low SNR as can be seen in the low frequency data points. The Newton-Raphson implementation has unrealistic values here, where real refractive index sharply drops off and the imaginary refractive index oscillates wildly. A trained CNN (one convolutional layer, and one hidden layer of 200 neurons) is also included, to show that with dramatically shorter training timescales, something close is produced, and indicates the potential future of our ANNs.

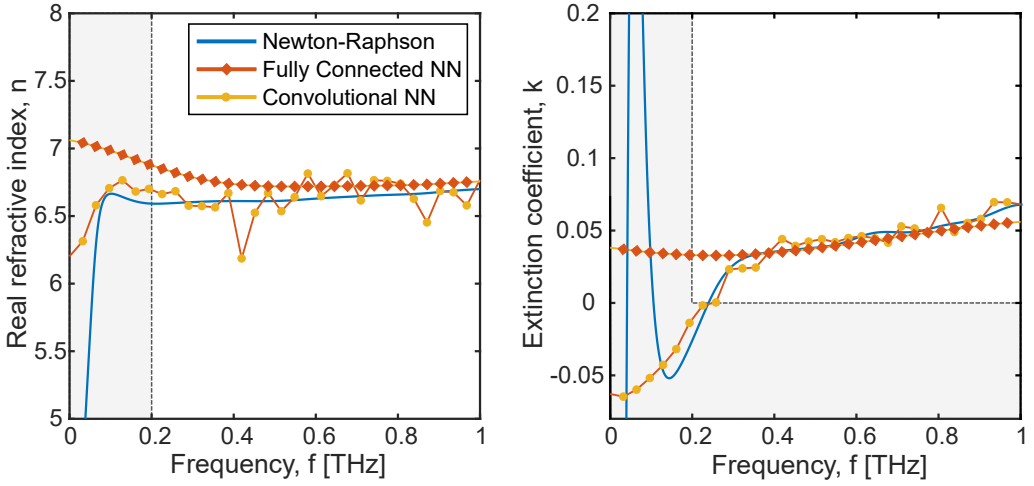

FIGURE 3.6: DNN model predictions (blue), Newton-Raphson fitting (black) and CNN model predictions of the refractive index of lithium niobate. The real (a) and imaginary (b) components. 100,000 training sets split 70/15/15 for training, testing and validation respectively. Hidden layer [128,512]. Convolution neural network has one 2D convolutional layer and one 200 fully connected hidden layer. The shaded regions of both plots show what is typically excluded from measurements due to low SNR or unphysical results.

The reported performance of the networks during training is much higher than what can be seen from the example fitting in figure 3.6 with real-world data not included in the training data sets, indicating a bias towards features that are in the training set itself. Continuing to train on fully connected networks for such problems is possible, but inefficient. Using CNNs allows for increasing the size of the input layers without dramatically increasing the computational overhead, thanks to the convolution layer. Both results show promise to provide more stable results at lower SNR than direct fitting methods. The CNN produces more stable and physically possible results at lower frequencies, at the cost of some deviation from the Newton-Raphson result. Further development of these networks will require careful selection of training data sets to produce trustworthy results.

In this chapter, machine learning has been shown to replace the fitting function, and further replacing steps in the extraction process as shown in figure 3.1(b) by accepting spectral information. With a flexible enough network architecture and carefully simulated training sets, neural networks can be used to replace the entire extraction process and accept time-domain data. This would provide a universal approach that avoids pre-processing that can introduce artefacts and reduces consistency and reliability - essential for real-world applications. Furthermore, it could be possible, by retaining the full time-domain data to have a neural network make refractive index predictions without requiring thickness information, windowing or Fourier transforms, and therefore could work well for thin samples where multiple internal reflections are problematic.

In order to expand the scope of the neural networks to accept such a level of unprocessed data, two approaches could be employed. One is to train networks for each step independently and either combine the separately trained models into one master function where each is applied sequentially to the data, or to use this collection of models to train a single network, a process known as *knowledge distillation*[125]. An approach such as this has the advantage of maintaining a limited parameter space for the training sets whilst also being able to monitor the accuracy of the networks at each step. An alternative method is to simulate an even larger array of data, and attempting to train one unified network. Theoretically, this should be possible, but the run-time and number of data sets required may be too large, and relegates all the steps of parameter extraction to effectively one black box, where more advanced inspection tools will be required to understand the how the network is developing.

3.5 Conclusion

Terahertz time-domain data retains phase information in a way that simplifies parameter extraction as compared to other techniques. However, care is still required when a parameter extraction algorithm is implemented. Artificial neural networks have been shown to be used in place of traditional fitting functions with excellent accuracy, fast run-time and straightforward implementation. Furthermore, deeper, more generalised networks can replace more steps of the extraction process, automating data analysis and increasing consistency. Generalised networks can surpass traditional methods in the low frequency range, especially in the presence of materials exhibiting low frequency resonances. Advancing the sophistication of the networks will lead to more generalised models that can potentially improve on current methods so they can extract thickness, refractive index and optimise the windowing process, all while retaining rapid run-times. Further, it is easy to envisage in the future that neural networks can be developed to operate directly on time-domain data and thereby create a new paradigm in THz-TDS parameter extraction. Future work will include training CNNs

to not only extract frequency dependent parameters, but also length parameter approximations by avoiding windowing in the time-domain. Additionally, due to the flexibility of training networks to achieve any desired function, it is possible to use a different model to simulate data to develop models that extract the DC conductivity of graphene from imaging data, mitigating the issues of 'jitter' [90] typically caused by deviations in substrate thickness across large area samples.

Chapter 4

A surface tension confined PTFE microfluidic flow cell

In this chapter, the development of a novel polytetrafluoroethylene (PTFE) microfluidic flow cell is documented. PTFE surfaces are modified to both increase and decrease its wettability confining aqueous samples by surface tension forces. The device is terahertz compatible, robust, non-fouling, expandable, quick to manufacture and cost effective. The work was published in [1], and is expanded on in this chapter. The work was performed by the thesis author, with contact angle measurements made by B. Bowden.

4.1 Introduction

To study the terahertz response of proteins in their native, fully hydrated environment, short beam path-lengths are required in order to mitigate attenuation of the transmitted pulse. Microfluidic devices manufactured from materials that are terahertz compatible (low absorption and refractive index) provide an excellent solution to maximise signal integrity when performing real-time spectroscopy. Furthermore, such platforms provide additional sample manipulation and expanded functionality to progress towards an all-in-one platform, often termed: *Lab-on-chip* [126, 127].

Terahertz compatible devices are often comprised of silicon or quartz [70, 71], which have good spectroscopic properties at terahertz frequencies whilst benefiting from mechanical rigidity. However, specialised manufacturing processes are required to fabricate such devices, and prototyping can be slow. A faster and cheaper pathway is to use polymer based devices. Polydimethylsiloxane (PDMS) is a commonly used material in microfluidic devices, however its terahertz absorption characteristics are significantly dependent on its preparation methods [72].

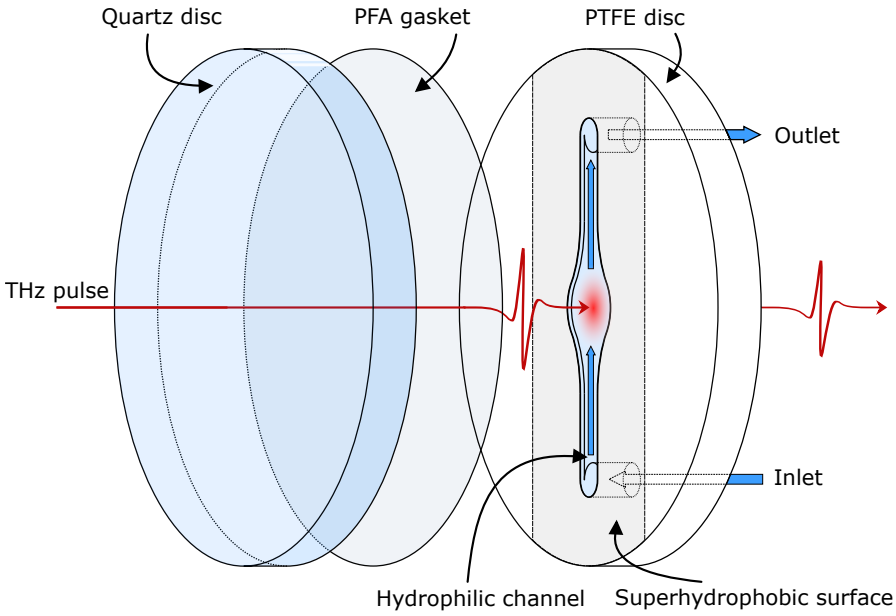

FIGURE 4.1: The geometry and core components of the microfluidic flow-cell. A microfluidic channel with a 6 mm wide zone for the transmission of the terahertz beam is machined into a 5 mm thick, 50 mm diameter PTFE disc by a PCB milling machine. Assembly pressure provided by a 3D printed manifold (not shown) lightly contacting a 5 mm thick, 50 mm diameter quartz disc and 50 µm thick PFA gasket on top of the device. The 1.6 mm inlet holes were introduced using the same PCB milling machine.

The chosen scheme in this work was to develop a device based on machined polytetrafluoroethylene (PTFE), a schematic of which can be seen in figure 4.1). The material has desirable properties for both use in a terahertz spectrometer and as a basis for a microfluidic device. Primarily, these are its low, non-dispersive refractive index between 300 GHz and 3 THz, negligible absorption over a broad terahertz spectrum, and its well known chemical resistance and bio-compatibility. Whilst these attributes are appealing, two key problems emerge when trying to use PTFE for microfluidics: its lack of rigidity and the incompatibility with adhesives, complicating encapsulation methods. PTFE has the advantage of being hydrophobic, which can aid sealing in devices that rely on compression forces, where any possible open pathways reject water due to capillary action [73]. The limits of such a device are discussed later in the chapter.

It is possible to exploit contrasts in surface wetting characteristics to make microfluidic devices that rely on capillary forces alone and no physical boundary or walls to control flow paths [77, 78, 79, 80]. Such a scheme negates the need for problematic encapsulation by relaxing the requirement for alignment and bonding. Additionally, having an easily accessible liquid-gas interface allows the application of gases that may react with the liquid and/or analytes on-chip, whilst serving to help to then remove introduced air bubbles from the system.

This boundary could serve to assist in the investigation of biological processes. For example, while terahertz spectra of single concentrations of oxy- and deoxy-hemoglobin do not show any fine spectral features in absorption in the terahertz band [128], changes in structure of globular proteins are detectable with terahertz spectroscopy by studying the effect of the macromolecule on the surrounding hydrogen bond network [44]. Differences in hydration dynamics between the two states of hemoglobin have been observed with microwave radiation [129], however, by using terahertz radiation instead, and by flowing oxygen and nitrogen gases adjacent to the protein solution (oxygenating and de-oxygenating the protein respectively [130, 131]), the investigation of longer range influences can be made with protein modification occurring *in situ*. Hemoglobin's function to flexibly bind and relinquish oxygen is coupled with the surrounding water network, and further investigation into its mechanism may provide greater insight into a wide range of hemoglobin disorders [132]. Furthermore, with the addition of a gas-liquid interface, such microfluidic devices can be used for the detection of airborne analytes such as drugs [133] or hazardous contaminants (ammonia) [134] by diffusion into the controlled liquid medium.

In this chapter, the development of a PTFE based microfluidic device for terahertz spectroscopy is shown, whereby a simple, closed cell design is enhanced using a novel combination of RIE, machining and plasma ashing to produce a surface tension confined device. The robust solution to water confinement in PTFE microfluidics presented allows the material to be used in a sandwich scheme but manufacturing tolerances can be relaxed. Key theoretical concepts of microfluidic flows and surface tension effects are introduced, followed by the challenges associated with PTFE based microfluidics, highlighting the difficulty in preventing leaks from compression sealing a chemically inert, flexible material. Then, the results of the surface modification of PTFE are shown, where the hydrophobicity of the material is both increased and decreased to enable dramatic contrast in wetting characteristics. The working device is then used to take spectroscopic measurements of water and bovine serum albumin (BSA) for different concentrations in the fabricated microfluidic prototype flow-cell with a Terahertz Time-Domain Spectrometer (THz-TDS).

4.2 Microfluidics theory

4.2.1 Fluid flow in a microfluidic channel

When coupled with the right materials, geometries and appropriate flow rates, water channels manufactured on the micron scale exhibit properties that are advantageous for the study of multiple aqueous samples with terahertz radiation. The two main properties to be exploited for this work are shallow geometries to reduce the signal attenuation of water, and the potential for automation for measuring multiple samples

consecutively with minimal interruption. The automation aspect of a channel design comes from the flow needing to be **laminar**, where there are no turbulences such that when introducing another fluid, its mixing will be limited by diffusion. This occurs when the fluid that's flowed at a given rate is confined to a sufficiently narrow channel. The consequence of such a flow is that a second introduced liquid can reliably replace the first.

To determine whether the flow of a fluid in a channel is going to be laminar or turbulent, the system's Reynolds number, R_e, can be calculated [81]. It has been found that if R_e<2300, the flow is laminar, where as if R_e>2900, then the flow is turbulent. In between these values, the flow will be intermittent, where the type of flow will oscillate between laminar and turbulent. Equation 4.1 shows the how the Reynolds number is calculated, by taking into account the fluid characteristics, flow rate and the size of the channel. The contribution from the channel's dimensions are given by the **hydraulic diameter**, a length value that depends on the shape of the cross section of the channel. In the case of a wide but shallow channel (a duct), this value is given by equation 4.2.

$$R_e = \frac{QD_H}{\nu A} \tag{4.1}$$

where the Reynolds number, Re, is given in terms of volumetric flow rate, Q, kinematic velocity, ν, hydraulic diameter, D_H, and cross-sectional area, A.

$$D_H = \frac{2H}{1 + \frac{H}{W}} \tag{4.2}$$

where the hydraulic diameter for a wide duct, D_H, is given in terms of the height, H, and the width, W. In the case of pure water, flowed through a channel at $50\mu l/min$, where this channel is a wide duct of 100 μm by 1.6 mm, it gives a Reynolds number of 10^{-7}. It is clear that in this case, laminar flow will occur.

For a wide channel ($w >> h$) the maximum pressure difference required for fluid flow, ΔP_{flow}, for a flow-rate Q, is given by [135]:

$$\Delta P_{flow} = \frac{12\eta LQ}{(h+d)^3 w[1 - 0.63\frac{h+d}{w}]} \tag{4.3}$$

where the channel length L and fluid of viscosity η contribute proportionally.

4.2.2 Contact angles and capillary effects

The choice of PTFE as a material removes the need for expensive fabrication techniques but at the sacrifice of having a soft material that cannot be reliably sealed by compression, especially flow-cell dimensions needed for terahertz compatible devices, motivating an alternative approach to confinement. It has been shown that selectively modifying the material surface in order to control the wetting properties of parallel coupled, etched capillaries allows for microfluidic function with a stable liquid-gas interface that relies on the Laplace pressure for confining the liquid in the channel [136]. It is feasible for this concept to be applied to the entirety of the channel, such that the principal form of confinement is from the liquid's surface tension forces. In doing so, the tolerances on device manufacturing can be relaxed, and the primary method of sealing will no longer be attributed to the elasticity of a gasket layer or by the requirement of bonding. To selectively modify the PTFE, a three stage approach is employed, shown in figure 4.8. In the first instance, the PTFE is plasma treated with RIE, creating a superhydrophobic surface across the entire substrate. Next, the channel is milled exposing native PTFE by completely machining the desired geometry, penetrating through the microstructured layer. Finally, the PTFE is treated by another plasma treatment, making the now exposed PTFE hydrophilic. Crucially, the last stage does not alter the initially created microstructures, thereby maintaining superhydrophobicity. This provides a contrast in wetting characteristics for the confinement of the sample within the channel beneath an upper hydrophobic substrate contacting (but not compressed against) the modified PTFE substrate.

4.3 Compression-only sealing a PTFE microfluidic device

The initial design of the PTFE based device relies on the malleability of a gasket layer to occupy defects, with the hydrophobicity of the substrate and the gasket materials resisting leaking through negative capillary action. The material of the gasket layer was chosen to be PFA, due to its similar spectroscopic and chemical properties as PTFE, but with a smaller Young's modulus. The initial design is shown in figure 4.2. One PTFE disc has a milled channel and inlet holes, with a second blank PTFE (or quartz for optical inspection) disc surrounding a PFA gasket layer. Each disc has a diameter of 50 mm and a thickness of 5 mm, with the channel milled using a PCB milling machine to the desired depth. A 3D printed manifold, and twelve bolts are used to provide even compression across the assembly.

The goal is to achieve a water tight seal around the channel from direct contact between the PTFE substrate and the PFA gasket layer. In addition, if any subtle deformations or defects exist in the substrate, the hydrophobicity of both layers were thought to prevent leaks from spreading across the substrate due to negative-capillary action.

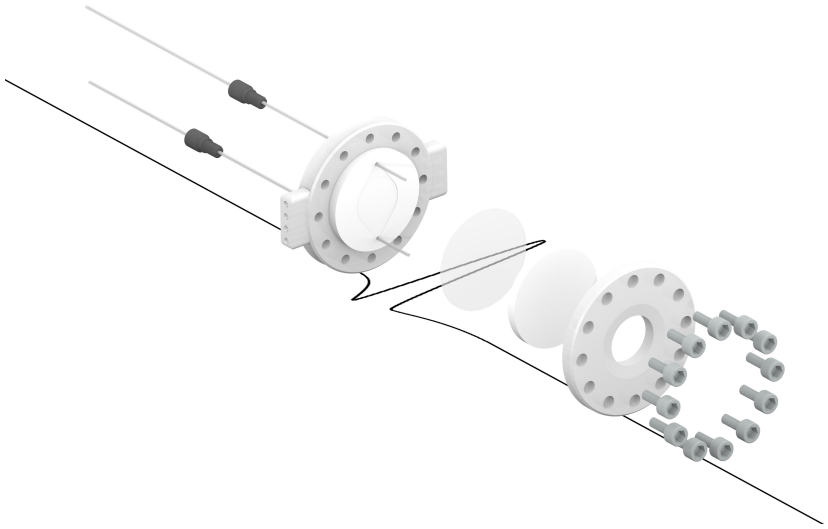

FIGURE 4.2: Initial prototype of the PTFE microfluidic flow-cell, with a 3D printed manifold to apply compression force across the 'sandwich' of PTFE discs with a PFA gasket layer in-between. Sealing was suspected to be achieved thanks to the PFA's hydrophobicity and malleability to occupy machining defects in the PTFE.

In the case of pure, de-ionised (DI) water, and low concentrations of salt solutions, this scheme is sufficient for proper confinement. However, care is required when tightening the device, as leaks would occur if either the screws were over-tightened, or tightened in such a away to produce uneven pressure on the cell. Evidence of cell warping from over-tightening can be seen in figure 4.3(a). As the cell is tightened, flex is visible due to an introduction of oscillation into the transfer function of the empty, loosely assembled cell, to the flexed cell where the cavity inside is affected. The oscillation is caused by the beating of two different etalon frequencies as the cell expands due to warping.

Further, if a flow rate of 100 µL/min of water was passed through the cell, the pressure inside would be too great and cause further leaking. The main concern with a flow-cell that leaks, beyond the practicalities of contamination and mess, is the disruption to ideal flow characteristics. In particular, if a new solution of a different viscosity is introduced, such as a high concentration of protein, proper removal of the previous sample will not occur.

Attempts were made to enhance the sealing by using a machined, metal manifold as opposed to the 3D printed one shown in figure 4.3(b). The aim was to apply more pressure to the configuration as well as reducing the possibility of the cell warping due to flexing of the manifold. Leaking still occurred, whether it was a loose even fit, or a tightened fit. To provide a better seal, PDMS was trialed as a gasket layer. Although not so hydrophobic, the higher malleability was thought to help provide a physical seal. However, leaks still occurred and were observed to be more severe than that with

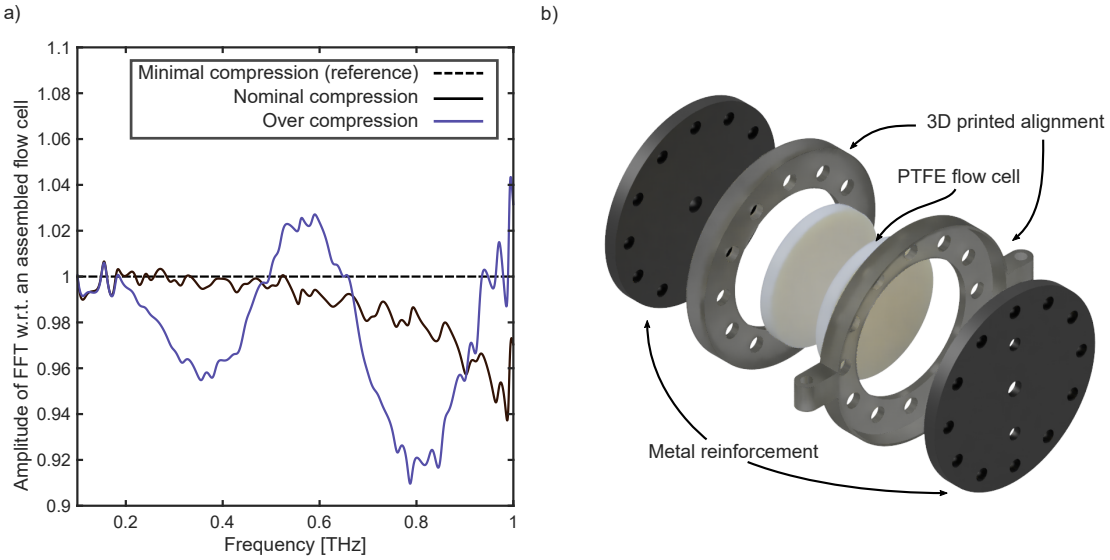

FIGURE 4.3: Magnitude of the transfer function of the empty flow cell of various compression forces with reference to itself at the loosest possible fitting (a). As the cell is tightened, oscillations appear in the transfer function due to the beating of two etalons as the cell warps from the reference cavity length. *Nominal compression* (black line) was the typical sealing pressure achieved with finger tightening, with *over compression* (blue) being the result of using a screwdriver to compress the cell. (b) Greater reinforcement with a metal manifold attempted still resulted in leaking, most likely due to cell warping.

the PFA gasket, indicating that the hydrophobicity of the gasket layer played a more crucial role then initially hypothesised.

4.4 Surface tension microfluidics for THz spectroscopy

4.4.1 Criteria for surface tension confinement in a retangular channel

The following explains how such a device is able to confine water from surface tensions forces alone. The requirement to have a balance of pressures on the sample, such that forces acting to confine the fluid are not exceeded by those acting to expel it.

In general for a surface of a fluid/liquid with surface tension, γ, where the curvature is defined by two radii R_1 and R_2, the difference in pressure between either side of the boundary can be expressed as [135]

$$\Delta P_{surf} = \gamma(\frac{1}{R_1} + \frac{1}{R_2}) \tag{4.4}$$

The concept described above is demonstrated in figure 4.5(b), in which the geometry formed from water due to its surface tension and contact angles with three types of surface are visualised.In this case the radius of curvature in the direction of the channel

$R_2 \to \infty$ is infinite and with the geometric description shown, the radius of curvature can be given in terms of the gap between the two surfaces, d, which results from etching the top surface, and the contact angle at the boundary where the water meets surface a and surface b as θ_a and θ_b respectively. When the system is in equilibrium, this contact angle is the same as that which can be simply measured on a flat surface with a sufficiently small droplet. It then follows that equation 4.4 can then be expressed as

$$\Delta P_{surf_{ab}} = -\frac{\gamma}{d}(\cos\theta_a + \cos\theta_b) \tag{4.5}$$

As indicated in Fig4.5(b), contact angles a and b are greater than 90 °, leading to a net pressure acting on the liquid resisting the filling of the space between the two materials, d. Similarly, this applies for the filling of the channel itself, where the pressure difference due to the upper hydrophobic substrate and the channel base, $\Delta P_{surf_{ac}}$, can be calculated by

$$\Delta P_{surf_{ac}} = -\frac{\gamma}{h+d}(\cos\theta_a + \cos\theta_c) \tag{4.6}$$

where the total space between the two surfaces now include the channel depth, h. Therefore, it can be seen that to reduce back-pressure inside the channel, reducing the contact angle of the channel surface will be beneficial. Indeed, if the contact angle of just the channel surface is sufficiently reduced, capillary action will cause the channel to be self-wetting. It is evident that being able to localise surface modification to increase or decrease contact angles would be beneficial to aid confinement on filling or reintroduction of new liquids.

For confinement to occur, the pressures responsible for retaining the liquid in the channel must be greater than those responsible for expelling it. Therefore, the condition for confinement can be expressed as

$$\Delta P_{surf_{ab}} - \Delta P_{surf_{ac}} - \Delta P_{flow} - \Delta P_{hydro} > 0 \tag{4.7}$$

The additional term, ΔP_{hydro}, that is included in this expression is the hydrostatic pressure caused by height difference in the medium and the density of the fluid. In microfluidics this is often ignored, especially in horizontal configurations or when the device is completely, physically sealed. However, such a term in the case of a surface tension confined, vertically orientated device, of which is a desirable configuration for aligning horizontal beam paths of the spectrometer. Here, it is noted that $\Delta P_{flow} \propto \frac{1}{h^3}$ and also $\Delta P_{flow} \propto Q$, as the requirement is to minimise h for optimum terahertz wave transmission, meaning that in doing so, the confinement pressures required increase. While the flow rate can be simply reduced to mitigate leaking, it was found that there

are practical lower limits for Q as too little flow increases the likelihood of air bubble trapping and poor displacement of previous liquid samples, compromising one of the benefits of microfluidic devices: reliable automation.

4.4.2 Design of a surface tension confined microfluidic device

For a terahertz compatible device, the height of the channel greatly affects the resulting available bandwidth and sensitivity of detected radiation in transmission. Figure 4.4(a) shows a typical spectrum of the terahertz spectrometer through an empty PTFE flow-cell, with 100 µm and 400 µm thicknesses of water measured in earlier prototype devices.

It is clear that the bandwidth and peak signal-to-noise ratio of a single terahertz pulse frequency spectrum is significantly reduced with increasing path length. Of course, an increase in path length will increase sensitivity to a change in the absorbing medium. By using the measured absorption coefficients for a BSA solution of 33 mg/ml [41] and that of water [61], a calculation of transmission with varying thicknesses can be made. Sensitivity was chosen to be defined by subtracting the two theoretical transmissions for different thickness to estimate at which path length the maximum difference, indicated in figure 4.4(b), where a peak delta in transmission occurs just below 100 µm of water, indicating that for both retention of signal strength and for sensitivity, a channel depth of 100 µm should be targeted.

The implications of reducing the channel height of a 1.6 mm wide channel are explored in figure 4.4(c). Equation 4.6 was used to calculate the Laplace pressures that arise when filling both a treated ($\theta_c = 90\ ^\circ$) and untreated ($\theta_c = 120\ ^\circ$) PTFE channel, with increasing channel height. Additionally, equation 4.3 was used to calculate the pressure difference required for a constant flow rate of 50 µL/min, again with varying channel height. The sum of the Laplace and fixed flow rate pressure drops shown in this figure act against the confinement of the device. The contribution from the fixed flow rate becomes the dominant term below channel heights of 100 µm, and sharply increases as it is reduced. By implementing a simple Newton-Raphson method, equation 4.7 can be solved for the maximum allowed gap for confinement, d_{max}, for varying channel contact angle and repelling surface contact angle, the results of which are shown in figure 4.4(d). Aqueous protein solution typically has a reduced surface tension and increased viscosity. To indicate such effects on confinement criteria, parameters were chosen that roughly coincide with 100 mg/mL BSA solution, a very high concentration, but not unheard of in terahertz spectroscopy analysis. A solution of this concentration has a viscosity, $\eta_{BSA} \approx 2\ mPa \cdot s$ [137] (as opposed to water, $\eta_{water} = 1\ mPa \cdot s$), and taking an approximate surface tension, $\gamma_{BSA} \approx 50$ mN/m (opposed to water, $\gamma_{water} = 72.8$ mN/m) [138]. It can be seen that while the tolerance for leaking due to gap sizes for high concentrations of protein decrease, the effect can be entirely accounted for with

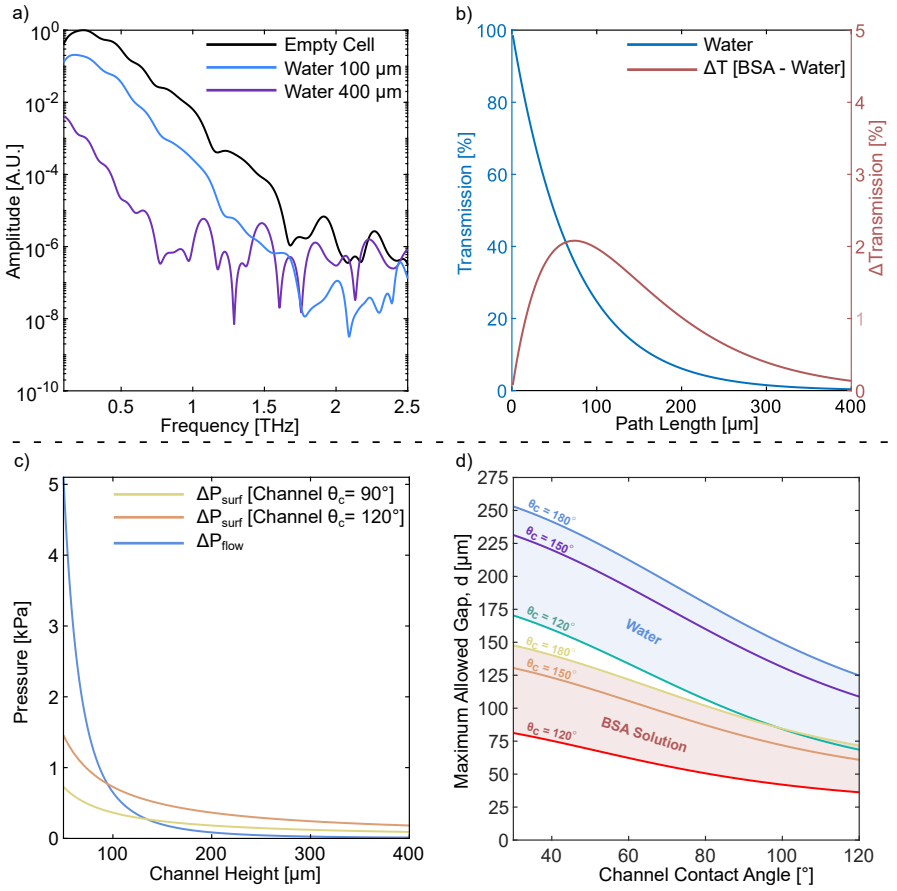

FIGURE 4.4: Theoretical calculations for optical and confinement parameters in a surface tension confined flow-cell. a) THz transmission frequency spectra of an empty PTFE flow-cell, 100 µm and 400 µm of water. b) Transmission of 500 GHz radiation through water of increasing depth up to 400 µm (blue) and delta transmission between water and 33 mg/ml of BSA solution (red). c) Back pressures caused by filling the channel against channel height, for a channel contact angle of 90 ° (yellow), 120 ° (orange) both with a cap of contact angle 120 °. Calculated pressure difference required for steady flow in a 1.6 mm wide channel (blue). d) Theoretical maximum gap for surface tension confinement of water (blue) and a 100 mg/ml BSA solution (red), where the gap required is plotted against the contact angle of the surface in a 100 µm deep, 1.6 mm wide channel, for a range of contact angles of the surrounding substrate, with example values plotted: 120 °, 150 ° and 180 °. For BSA solution, approximations for viscosity η was taken to be twice that of water and surface tension, γ, to be two thirds of that of water. Values obtained by numerically solving equation 4.7 for gap, d.

surface modifications. Confinement is maintained with an achievable 90 ° channel contact angle and a 150 ° repelling surface contact angle, with much greater enhancement as the channel and repelling surfaces are made hydrophilic and superhydrophobic respectively. Such an increase in tolerance to defect induced leaking allows PTFE to be used in a for terahertz compatible device (100 µm thickness) as the contact angle contrast will confine liquids of higher viscosity than water at the same flow rate.

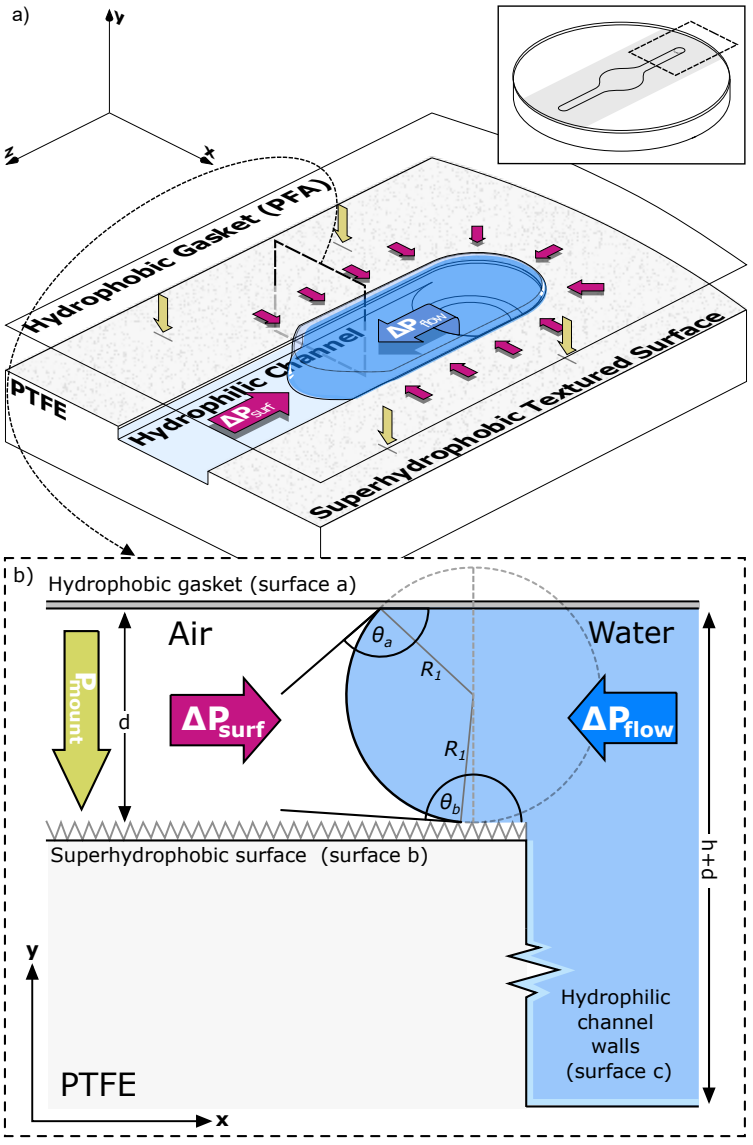

FIGURE 4.5: The fundamentals of surface tension based confinement in a microchannel. (a) displays a section of straight, rectangular channel, typical of what milling the PTFE substrate will produce, with the inset illustrating the location of the cross section. The surfaces are labelled with their wettability. (b) a cross section of one of the boundaries responsible for sealing the device showing the liquids meniscus with respect to the contact angles produced at the two surfaces interfaces, the geometry of which can be used to calculate the capillary pressure produced by the surface tension.

4.5 Fabrication of the surface tension confined device

4.5.1 PTFE surface modification

The hydrophobicity of native PTFE ($\theta_c \approx 120$ °) can be both increased and decreased through plasma treatments of various powers and ionic species. High power, directional argon and oxygen plasmas can modify PTFE's surface to be superhydrophobic (contact angle, $\theta_c > 150$ °) whereas a low power, ambient plasma exposure can increase

the hydrophilicity of the surface [74, 75], with reported contact angles as low as 40 degrees [76]. X-ray photoelectron scattering measurements have revealed that in the case of a low power exposure, it is the surface chemistry that is modified [139]. In contrast, exposure to higher energy plasmas, more specifically the accelerated particles found in reactive ion etching (RIE), changes the topology of the PTFE surface by creating significant roughness at the micro and nano-scales such that its roughness supersedes chemical effects to define a superhydrophobic surface [140].

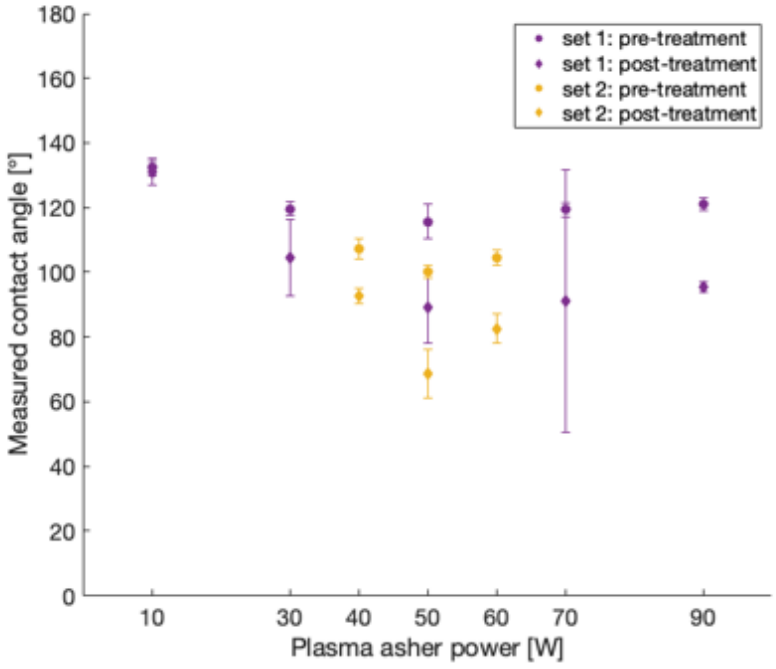

FIGURE 4.6: Contact angle measurements of 20 µL water droplets on PTFE substrates, with and without treatment (six total per disc), for a range of plasma asher powers. A second set of measurements were made after a finer machining process for finishing the PTFE substrates was established, closer to the perceived optimum for increase in wetting. The target is to have a the measured contact angle below 90 °.

To modify the surface chemically, such that it becomes hydrophilic, PTFE discs parted from an extruded rod were treated with oxygen plasma in a plasma asher for three minutes at a range of power settings with maximum flow rate. Figure 4.6 shows the difference in hydrophobicity between native and treated PTFE with the contact angle measurements of 20 µL water droplets. The objective is to reach a treated surface with a measured contact angle of less than 90 °, to remove the back-pressure from filling a channel. The figure also illustrates the importance of machining consistency, as the second set of data was also performed on PTFE discs that were machined more carefully, resulting in more consistency in measured contact angle. Both sets of data indicate that an approximate optimum power setting exists near 50 W, and this value met the criteria for a hydrophobic surface.

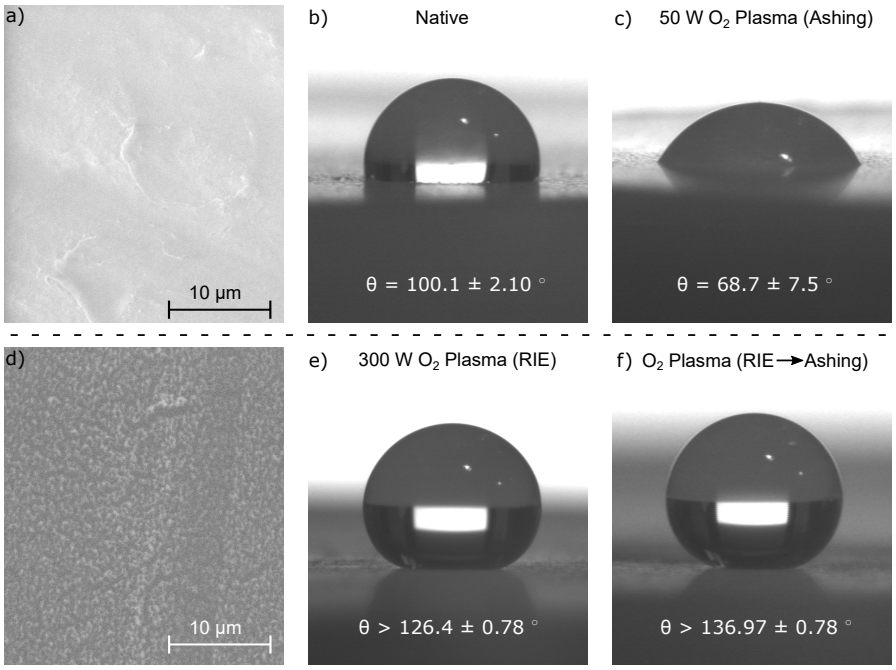

FIGURE 4.7: Examples of images used for contact angle measurements proceeded by SEM images of plasma modified PTFE. All droplets have a volume of 20 µL, as this was the minimum volume size required for the water to leave the needle. a) SEM image of untreated PTFE, showing typical native surface, (b) image of water on an untreated PTFE substrate, (c) a substrate after O_2 plasma ashing, clearly showing hydrophilic behaviour, (d) SEM image of PTFE after the RIE treatment, (e) attempt at contact angle measurement of PTFE after 30 mins of RIE treatment. Droplet migrated to a stable point where captured contact angle is actually lower than that typical of a smooth surface. (f) Is an attempt to measure the contact angle of PTFE after RIE and then the ashing process, but the etched surface roughness causes the surface to retain superhydrophobicity. Similarly to (e), the drop still moved to a defect on the disc and measured angles from this image will be an underestimate. (f) A sharp increase in surface roughness indicates the cause for the superhydrophobic behaviour.

The 50 W treatment reduced the contact angle of the surface from a measured 100.1 $\pm$ 2.1 $^\circ$ to 68.7 $\pm$ 7.5 $^\circ$ illustrated by figure 4.7(b) and (c) respectively, with (a) showing a scanning electron microscope (SEM) image (taken in partial vacuum) of an untreated, PTFE surface. Physio-chemical surface modification is expected to change over time [141] however ageing with the PTFE exposed to water is far slower than in air, and the hydrophilicity can possibly be rejuvenated. It is predicted that with regular use in a microfluidic device the reduced contact angle of these surfaces will remain stable, however, with the subsequently described treatment, the device will still be able to confine most aqueous samples if the channel's wettability fails over said long period.

A superhydrophobic surface was created by etching micro-structures with RIE with oxygen species. If the surface tension is high enough such that the liquid bridges over the microscopic valleys of the structures, the contact angle is enhanced to become an effective contact angle. This is a function of the Young's contact angle, θ, of the unmodified material and that of air present in the valleys of the structures. When this occurs, the system is said to be in the Cassie-Baxter state [82]

$$\cos\theta_{CB} = f_1 \cos\theta - f_2 \tag{4.8}$$

where f_1 and f_2 are fractions of solid/liquid and air/liquid of the effective surface respectively. It can be seen from this equation that whilst the measured Cassie-Baxter contact angle θ_{CB} does indeed depend on the material's surface Young's contact angle, θ_{CB} will still show an enhancement even in the case of a hydrophilic substrate. To achieve a sufficiently rough topology, the PTFE was treated for 30 minutes with RIE (300 W, 5 sccm of O_2, 50 mTorr, 600 V DC bias produced). Contact angle measurements show that the surface was superhydrophobic. However, it was not possible to obtain a reliable measurement, as the water droplet resisted leaving the deposition needle, requiring a larger droplet size to be effectively deposited. The water then accelerated away from the target and came to rest at a defect, with the image of the droplet in figure 4.7(d). The measurements made at this point will be an underestimate, so only the contact angle enhancement of $\theta_{CB} > 126.4 \pm 0.78\ ^\circ$ can reported. Observation under an SEM shows the topological result of the treatment in figure 4.7(f) where a high degree of roughness is visible. Additionally, the roughness can be seen with the naked eye as a matting of the otherwise slightly glossy PTFE surface.

Equation 4.8 suggests that if the fraction of air to the material surface is high enough, then the contact angle will increase even in the case of lower Young's contact angles. By exposing an already etched substrate of PTFE to the O_2 plasma from the plasma asher, the surface roughness contribution to the effective contact angle supersedes the effect of the chemical change and the surface remains superhydrophobic, as evident in figure 4.7(e). Indeed this was the case as it was similarly difficult to quantify what the contact angle was as with just the RIE exposure alone: $\theta_{CB} > 136.97 \pm 0.78^\circ$. With these methods, a high contrast in wettability can be created without the need for a mask, by first treating a PTFE substrate with RIE, milling the desired geometry (exposing unaltered, native PTFE), and then treating the whole substrate in a plasma asher. After doing so the milled channel will be relatively hydrophilic where as the upper, micro-structured surfaces will continue to be superhydrophobic.

4.5.2 Frabricating a device with localaised wetting

A 50 mm diameter, 5 mm thick PTFE disc was treated with the RIE process previously described. During the RIE process two glass slides were used to mask the edges of the material, producing a step down to the superhydrophobic micro-structured area and protects this area from being damaged by contact when the device is assembled. The channel geometry and inlet holes were milled and drilled respectively on a PCB milling machine (LPKF ProtoMat s64) with a 380 µm diameter mill tool and a 1.6 mm diameter drill tool. The inlet channel width was chosen to match the 1.6 mm inlet holes both

to minimise required pressures for steady flow while also allowing for opportunity for smooth transitions between milled and drilled areas on the substrate. The substrate was sonicated in isopropyl alcohol and de-ionised water each for ten minutes to remove any machining fragments and residue from handling. The whole substrate was exposed to the O_2 plasma treatment outlined previously, reducing the contact angle of the newly exposed PTFE in the channel. For enclosing the device, a z-cut quartz disc of 50 mm diameter and 5 mm thickness was chosen for its terahertz and optical transparency to visually inspect the flow in the device. A 50 µm thick tetrafluoroethylene-perfluoro(alkoxy vinyl ether) (PFA) gasket layer with a hydrophobic surface (and similar terahertz spectroscopic characteristics to PTFE) was used to enclose the channel.

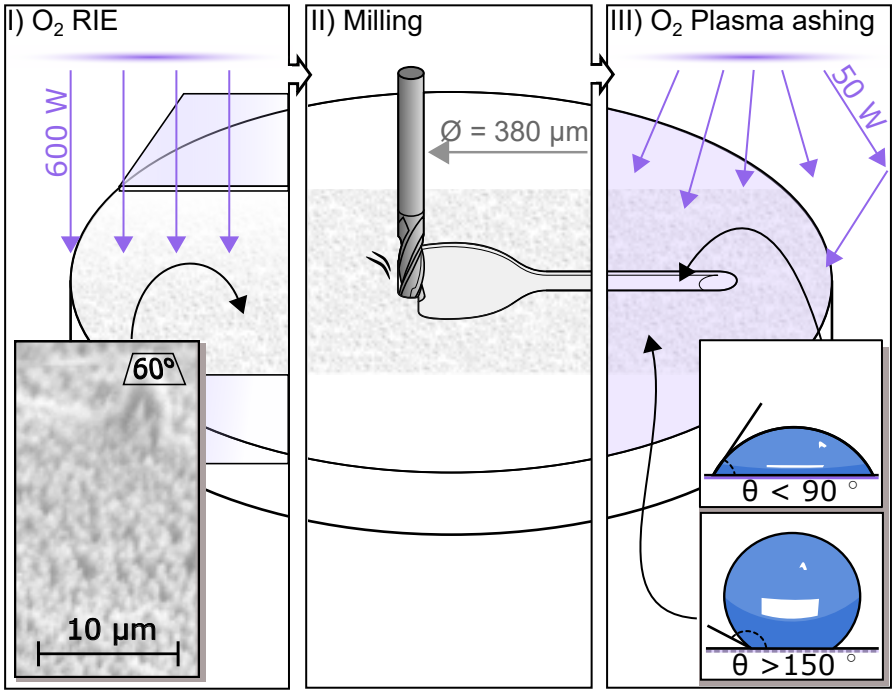

FIGURE 4.8: Illustration of the microfluidic channel fabrication process, showing the stages of plasma treatment of a PTFE disc. The whole disc is first treated with reactive ion etching (RIE), creating a superhydrophobic surface. The desired channel geometry is then milled with a PCB milling machine, exposing native PTFE. The whole device is then exposed to the non-directional, O_2 plasma in a plasma asher where a hydrophilic channel is produced surrounded by a superhydrophobic upper surface. The leftmost inset is an image of the etched surface captured with a scanning electron microscope (SEM), with the sample tilted by 60 ° from the normal.

It is possible for another PTFE disc to be used instead of quartz and PFA if visual inspection is not needed. Substituting PTFE would reduce reflection losses and simplify the device further. The assembly does not require a large compression force to provide seal but instead relies on surface tension forces only. As such, a simple 3D printed manifold was used enabling faster prototyping than the large, metallic manifold that has previously been required for a similar scheme (without the surface modification) for microwave resonance spectroscopy [73]. Avoiding metallic elements has the addition

benefit of making the device lighter and easier to align in the time-domain spectrometer.

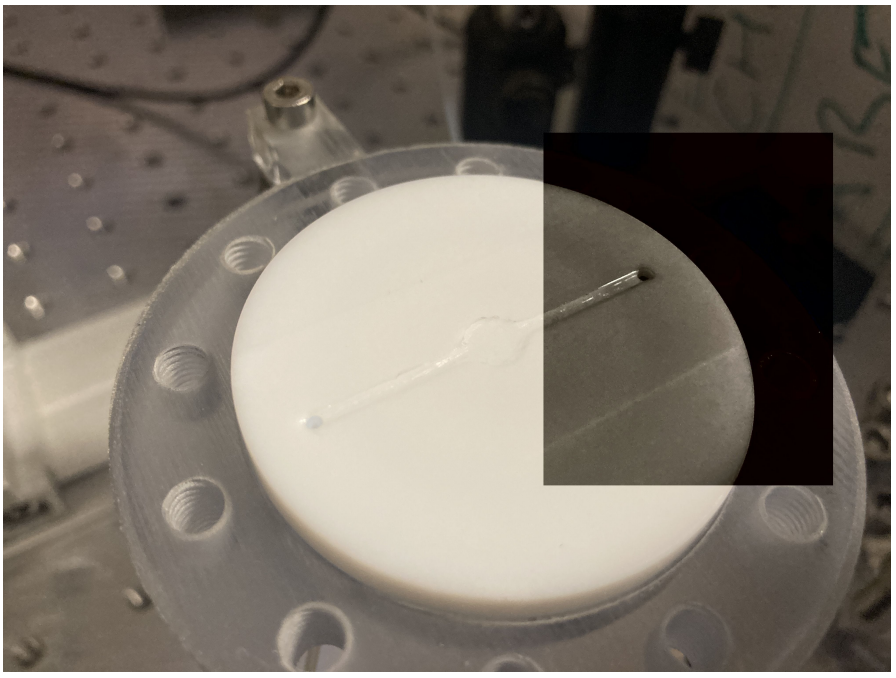

FIGURE 4.9: Fabricated flow-cell channel, with the cropped window contrast adjusted to make the different surfaces more visible. The matted section is due to microstructures being present.

A stylus profiler was used to measure the surface roughness of the channel and the relative step sizes. The RIE treatment produced an etch of 21.63 $\pm$ 0.99 µm deep, and the milling created a channel of a further 89.90 $\pm$ 1.64 µm below the etch height. The surface roughness of the milling, calculated as the line length ratio, r, to be almost imperceptible larger than one, which is comparable to the native roughness of the substrate, r = 1.01 $\pm$ 0.0043. Evidence of swarf at channel walls extending up to 80 µm above the channel, meaning, depending on assembly pressure, the total effective channel height when assembled will be < 170 µm. The length parameter can be calibrated by fitting to previously measured water spectra, but in the future, there will have a more refined machining process. Figure 4.9 is a photo of the machined and treated PTFE surface. The microstructured section is visible as a matted surface on the otherwise more glossy material.

The assembled device accepts water and more viscous protein solutions at the same flow rate, with the added benefit of removing any air bubbles introduced by syringe exchange. As well as tolerating air introduction, this also demonstrates the sealing of the device being due to surface tension forces, as a gas-liquid interface must be present.

4.6 Performance of the device in a time domain spectrometer

The THz-TDS system used to measure protein solutions has been outlined in chapter 2, with figure 4.10 showing the additional microfluidic components. The emitted terahertz broadband pulse is collimated and focused through the centre of the flow-cell, then re-collimated and subsequently focused onto the detector. The flow-cell is placed in the metallic box with the terahertz components and is purged with nitrogen gas to remove water vapour absorption lines. The syringe pump is placed above the box and the protein waste jar can be placed on top (as this device is stable even with the additional hydrostatic pressure of the outlet tube rising approximate 40 cm above the cell) or under the cell inside the box.

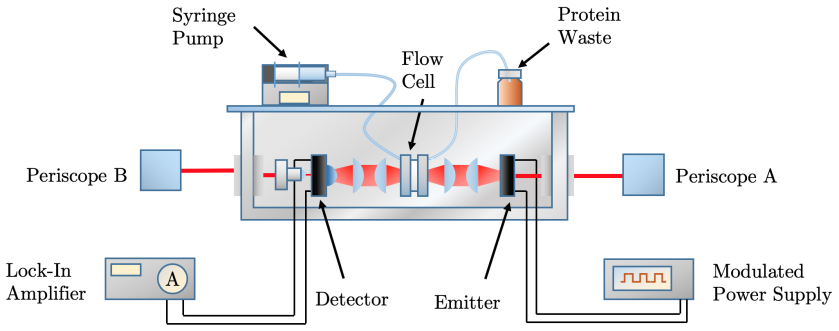

FIGURE 4.10: Flow cell in the nitrogen box with microfluidic apparatus. The syringe pump and protein waste jar sit outside of the box allowing sample exchange without disturbing any of the spectrometer's core components and surrounding atmosphere.

Lyophilised BSA powder (Sigma A7906) 98% purity was weighed and mixed with deionised water, then sonicated for 10 minutes. Each concentration was introduced to the cell via a syringe pump and a PTFE inlet tube at a flow of 50 µL/min. Initially, a time-domain transmission spectrum of the empty cell was measured and then starting with water, each new solution was passed through and two minute long scans were recorded to get a terahertz spectrum. The spectrum of each solution was divided by the reference spectrum (the empty cell) to obtain an experimental transfer function; the complex refractive index was extracted by fitting a theoretical transfer function to the data using the Newton-Raphson method. Due to water's severe attenuation, no echo reflections were detectable and therefore the straightforward *zero reflection* transfer function (equation 2.11) can be used, with the refractive index of the reference taken as that of PTFE, n = 1.4. A further transfer function was trialed, whereby the internal reflections in the empty cavity were modelled, but due to the low refractive index of PTFE, the difference in extracted parameters was negligible.

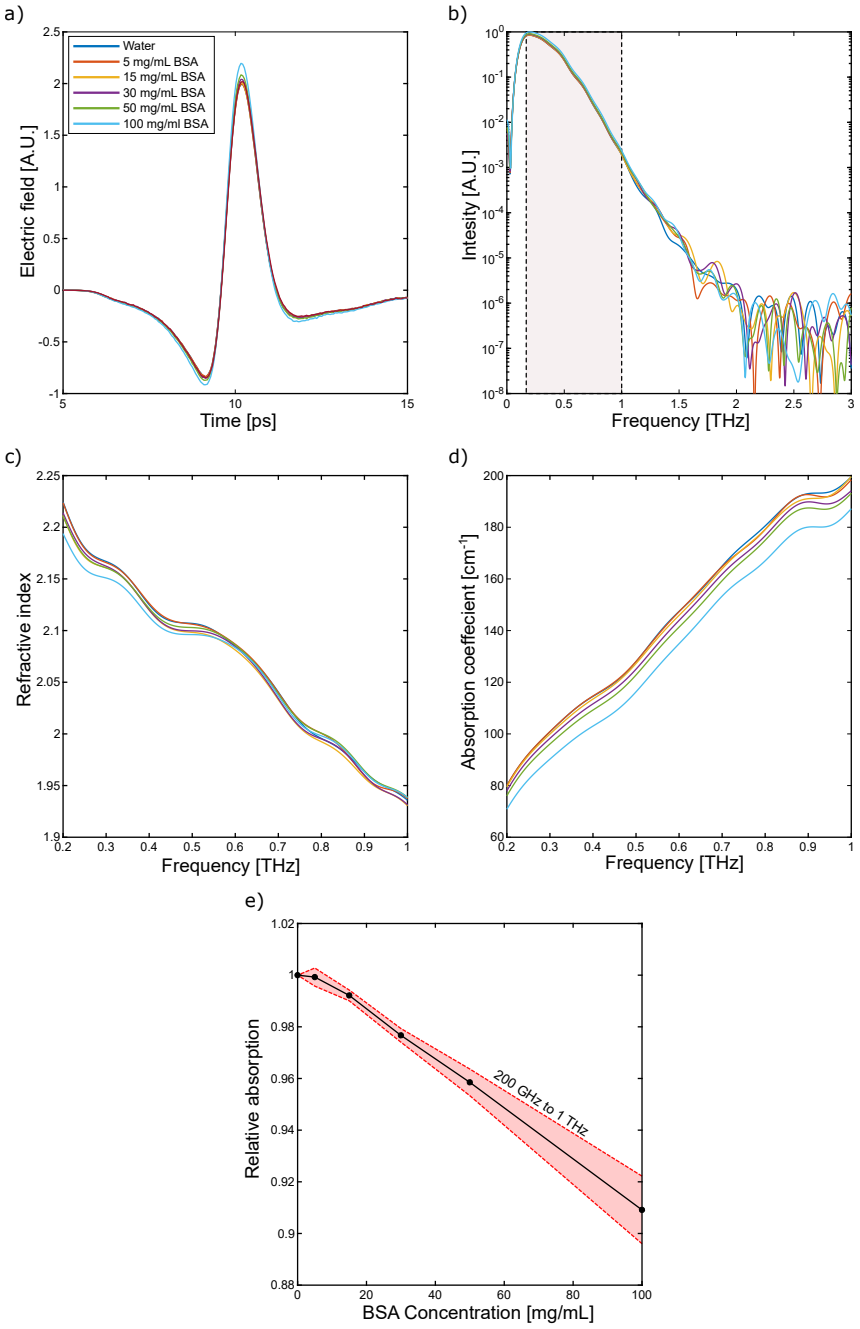

FIGURE 4.11: THz-TDS scans and extracted parameters for water and aqueous BSA concentrations up to 100 mg/mL. The time domain pulses (a) and the associated frequency spectra (b). Focusing on a range of 200 GHz to 1 THz, the extracted refractive index (c) and absorption coefficient (d) are shown, with small variations due to protein concentration. By calculating the relative absorption by division of the measured water absorption for each frequency, the average relative absorption for each concentration (black circles) is plotted with the standard deviation (dotted red) to show the spread across frequencies.

Figure 4.11 shows the terahertz scans of the water and BSA solutions. By looking purely at the time-domain and frequency-domain spectra (figure 4.11(a) and (b) respectively).

The small path length of the microfluidic channel allows for a wide bandwidth of terahertz spectra to remain above the noise floor. The small differences in frequency dependent complex refractive indices between the samples are not obvious, as attenuation of signal can be from both reflections dominated by the real refractive index as well as absorption directly from the extinction coefficient. Extracting the real refractive index and the absorption coefficient (figure 4.11(c) and (d) respectively) confirms that the most significant contribution on signal change due to presents of BSA protein is the absorption coefficient, where only comparably small changes in the real refractive index are visible. The relative absorption was calculated by dividing each frequency dependent absorption coefficient for BSA concentrations by the measured water absorption. The average of this was made over the 200 GHz and 1 terahertz frequency range which is shown, along with the standard deviation, is shown in figure 4.11. As mentioned in the introduction, a key area of interest is the terahertz response of protein hydration and dynamic water bond networks. A non-linear trend in the concentration dependent absorption coefficient indicates the presence of a third component (in addition to protein and bulk water volume): hydration shell water. The trend is not what would be expected in accordance with the simple exchange of water with protein volume. In these measurements, it is possible that such a non-linear trend exists, with a small inflection point in the region of 5-10 mg/mL.

It is clear, however, that the apparent deviation from a linear trend is very subtle, and there is not enough data to confidently claim the detection of a hydration shell. To improve on this experiment, more data points are required, specifically at the lower concentrations. Additionally, longer scan times can be taken and more repeats of the experiment to provide a tighter confidence interval.

4.7 Integrating artificial neural networks with protein hydration terahertz measurements

The previous chapter introduced a new method of parameter extraction through the use of artifical neural networks (ANNs). The same, frequency point wise trained network was trialed on the protein solution measurement data. Figure 4.12(a) shows the extracted real refractive index of water and 100 mg/mL of BSA obtained by the Newton Raphson fitting and the ANN prediction. At lower frequencies, there is a discrepency between the methods, and this is most likely due to this ANN being trained on a data set that contains samples of up to k = 0.3, which is significantly below that of water at k $\approx$ 0.5 at 1 THz. It follows then that the extracted absorption coefficient (figure 4.12(b)) has greater variation between the techniques, where the network's prediction undershoots the numerically fitted values.

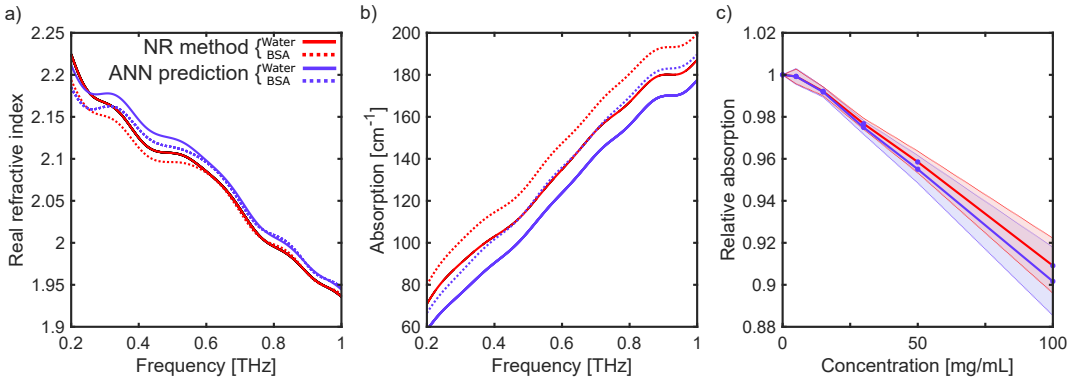

FIGURE 4.12: Extracting the refractive index of water and BSA with an artificial neural network and the standard, Newton-Raphson fitting to a theoretical model. The real refractive index (a) and absoprtion coefficient (b) of water and 100 mg/mL of BSA are predicted and compared with the numerical fitting. The concentration dependence (c) of average absorption between 200 GHz and 1 THz is produced with both methods.

A more promising demonstration is in figure 4.12(c), where the concentrentration dependence shares similar a similar trend to that shown in the Newton-Raphson extraction. If a training set that encompased a wider range of values were to be used, it is likely that the ANN prediction would produce very similar results.

4.8 Conclusion

A PTFE flow-cell for terahertz spectroscopy of biological aqueous samples has been demonstrated in this chapter. PTFE offers low absorption at terahertz frequencies and the low surface free energy characteristic of PTFE produces a non-fouling and bio-compatible surface, desirable for many microfluidic applications. However, this characteristic hampers efforts for bonding and encapsulating the microfluidic channels. New encapsulation approaches are therefore required for the effective implementation of PTFE in microfluidics, and here a cost effective and robust solution is demonstrated. A novel combination of plasma treatments of PTFE to provide a strongly localised contrast in wetting characteristics on a single substrate has been demonstrated. This leads to realisation of a flow-cell with optimised path-length of aqueous media that exploits surface tension forces to achieve easy confinement of the liquid medium, enabling terahertz spectroscopic measurements of aqueous solutions of biological macromolecules. The cell is demonstrated for terahertz spectroscopy of aqueous solutions of the common model protein bovine serum albumin determining the complex refractive index and hence absorption coefficient as a function of both frequency and BSA concentration. The straightforward fabrication process will allow for rapid prototyping and customisation for further development of device functionality, and the gas-liquid interface inherent to the confinement method provides an excellent platform for future

study of the effects of gas interactions on protein hydration, such as hemoglobin oxygenation, potentially allowing for the detection of small changes in conformation of gas-sensitive proteins with the reaction being performed 'on-chip'. Furthermore, the integration with an artifical nerual network for the complex refractive index extraction was trialed. The water and BSA solutions measured exhibit extinction coefficents that were outside of the ANN's training parameters but still showed similar absoprtion with respect to concentration as the numerical fitting method. The next step is to widen the simulation paramters for the neural network's training set, and integrate more complex neural network architectures to further the automated and integrated nature of the developed platform.

Chapter 5

Conclusions and Future Work

5.1 THz-TDS parameter extraction methods

By virtue of the way in which THz-TDS retains the phase information of the spectroscopic data through probing the amplitude of the electric field, instead of merely the intensity, it is possible to extract the complex refractive index of a sample without the need for overly-complicated fitting procedures. However, care still has to be taken when trying to fit a theoretical model to the data, especially with regards to fitting terahertz scans containing multiple reflections. Typically, algorithms such as the Newton-Raphson method is used to converge onto the correct parameters for TDS information. However, in this thesis it is shown that complications can arise due to instability in the algorithm. To counter this it is common to apply a windowing function in the time domain to remove the reflections, simplifying the transfer function and therefore the fitting procedure, at the cost of reducing information and spectral resolution. In this chapter a new method was presented where the reflections are still used for fitting, and removing the need for an air reference measurement. The so called 'echo method' is self referencing, and has shown to produce a more consistent refractive index extraction between separate scans, while being straightforward to fit and avoiding the convergence issues presented with conventional multiple reflection fitting. It has also been proposed in this chapter that further transfer functions can be constructed evolving from these principles where the thickness can be ignored and possibly even calculated simultaneously, without the prior knowledge of the sample's refractive index. This will be the topic of future work, along with further characterisation of the conventional fitting methods outlined in the chapter to further understand its limits.

5.2 Machine learning for terahertz parameter extraction

Machine learning methods, or more specifically, artificial neural networks (ANN), have seen a huge wave of adoption in mainstream computing in the last ten years. In this chapter, ANNs were used to replace the Newton-Raphson and analytical approximation fitting steps in the extraction of refractive indices of THz-TDS data with excellent accuracy and fast run-time with its simple implementation. Unlike more conventional training of neural networks where huge, real world data sets are required, the data used for training was simulated using the Fresnel coefficients outlined in the previous chapter. This means that the simulations required to create large data sets for training are straightforward and can be rapidly produced.

More advanced network architectures were introduced. Multi-layered networks, or deep neural networks (DNNs) and convolutional neural networks (CNNs) showed that fitting was still possible when feeding in spectral data rather than independent, frequency point wise data, with simulated noise added. The future of this work is to progress the trained models to remove as many of the parameter extraction steps as possible, eliminating the problems associated with the variation due to data pre-processing and the poor convergence of numerical methods on fitting to un-optimised data.

Future work includes the use of large data sets obtained experimentally from the imaging of large area graphene layers with a THz-TDS. The data sets will be incorporated into the training to produce a CNN to extract DC conductivity from the samples, and aims to be a more reliable approach to the minimisation methods currently employed. This will also be the first demonstration of direct time-domain analysis, potentially providing a solution that requires no pre-processing at all, thanks to the simpler model of the equations that govern conductivity in sub-wavelength samples.

5.3 PTFE microfluidics for terahertz spectroscopy

One of the greatest appeals of terahertz spectroscopy is that many physical phenomena have a response in the terahertz frequency domain. One such example is the collective motions of proteins and the interaction with their surround hydrogen bond network when in aqueous solution. A technical challenge presents itself with probing these processes while overcoming the strong absorption of the surrounding solvent. Hydration shell dynamics is a growing field of study as it illuminates another degree of structure in protein dynamics. With respect to terahertz spectroscopy, it also in fact compliments what was seen as a strong disadvantage in the field. An ideal way to probe these dynamics is through the use of microfluidics to measure the response of multiple concentrations of protein solution. This chapter showed a novel device comprised

of surface modified PTFE, whereby the confinement of the water is achieved through surface tension effects. This has many benefits, not limited to water confinement but also provides a boundary where solutions can be interacted with gases. The device was demonstrated to work with these boundaries present, and so the future study with this device is to measure the oxygenation effects on hydration dynamics on heme-group proteins with the gas exchange occurring on the device. Furthermore, it is possible to rapidly prototype other channel geometries and with the surface chemistry altered, deposit metallic layers onto the PTFE enhancing terahertz signals or providing an even more dramatic contrast in wetting allowing unrepresented sample control for such a terahertz compatible device. The techniques established here provide a flexible platform with the options of rapid prototyping to host a huge range of experiments with proteins and terahertz spectroscopy, all in a single microfluidic device.

5.4 Integrating artificial neural networks with the microflduic platform

Initial results of using ANNs with the data obtained with the flow cell show promise in the further automisation of using THz-TDS to probe biological material and processes. To fully marry the two areas of this work, expanding the training set for the neural networks are required to encompass a wider range of parameters. Further, implementation of CNNs could provide a hands-off method to reliably extract the refractive index of water by mitigating the effects of phase retrieval error and varience in the extraction of parameters of such a dispersive sample.

The combination of which will provide a reliable, robust and flexible solution, that has the potential to increase consistancy and ease of use, whilst providing a foundation for studying more complex processes with terahertz time-domain spectroscopy. This work presents the basis for a improving on the accessability of biological processess with terahertz radiation, overcoming the attenuation of accompanying water matrix of biological material and the potential to minimise the variability seen across the wider THz comunity with the implimentation of machine learning. The next steps are to increase accuracy and generality of the neural networks, to more appropriately extract spectral information from aqueous solutions presented to the spectrometer, while also increasing the functionality of the platform itself, unlocking more probable biological mechanisims with terahertz radiation.

Appendix A

Appendix A: Publications produced during the work undertaken for this thesis

A.1 Journal publications

N. T. Klokkou, D. J. Rowe, B. M. Bowden, N. P. Sessions, J. J. West, J. S. Wilkinson, and V. Apostolopoulos, “Structured surface wetting of a PTFE flow-cell for terahertz spectroscopy of proteins,” *Sensors and Actuators B: Chemical*, vol. 352,p. 131003, (2022).

N. Klokkou, J. Gorecki, J. S. Wilkinson, and V. Apostolopoulos, "Artificial neural networks for material parameter extraction in terahertz time-domain spectroscopy," *Opt. Express*, 30, 15583-15595, (2022)

J. Gorecki, N. Klokkou, L. Piper, S. Mailis, N. Papasimakis, and V. Apostolopoulos, “High-precision thz-tds via self-referenced transmission echo method,” *Appl. Opt.*, vol. 59, pp. 6744–6750, (2020).

N. Phanchat, W. Talataisong, N. Klokkou, R. Chitaree, V. Apostolopoulos, M. Beresna, and G. Brambilla, "Extruded TOPAS hollow-core anti-resonant fiber optimized for THz guidance at 0.9THz," *Opt. Express 30*, 13059-13069, (2022)

A.2 Conference proceedings

N. T. Klokkou, D. J. Rowe, B. M. Bowden, N. P. Sessions, J. J. West, J. S. Wilkinson, V. Apolostolopoulos, "THz Spectroscopy of BSA in a Surface-Tension Confined Flow-Cell," *2021 46th International Conference on Infrared, Millimeter and Terahertz Waves (IRMMW-THz)*, 2021, pp. 1-2.

N. Klokkou, J. Gorecki and V. Apostolopoulos, "THz-TDS Parameter Extraction via Machine Learning," *2021 46th International Conference on Infrared, Millimeter and Terahertz Waves (IRMMW-THz)*, 2021, pp. 1-2.

Bibliography

[1] N. T. Klokkou, D. J. Rowe, B. M. Bowden, N. P. Sessions, J. J. West, J. S. Wilkinson, and V. Apostolopoulos, "Structured surface wetting of a ptfe flow-cell for terahertz spectroscopy of proteins," *Sensors and Actuators B: Chemical*, vol. 352, p. 131003, 2022.

[2] N. Klokkou, J. Gorecki, J. S. Wilkinson, and V. Apostolopoulos, "Artificial neural networks for material parameter extraction in terahertz time-domain spectroscopy," *Opt. Express*, vol. 30, pp. 15583–15595, Apr 2022.

[3] J. Gorecki, N. Klokkou, L. Piper, S. Mailis, N. Papasimakis, and V. Apostolopoulos, "High-precision thz-tds via self-referenced transmission echo method," *Appl. Opt.*, vol. 59, pp. 6744–6750, Aug 2020.

[4] N. T. Klokkou, D. J. Rowe, B. M. Bowden, N. P. Sessions, J. J. West, J. S. Wilkinson, and V. Apostolopoulos, "Thz spectroscopy of bsa in a surface-tension confined flow-cell," in *2021 46th International Conference on Infrared, Millimeter and Terahertz Waves (IRMMW-THz)*, pp. 1–2, 2021.

[5] N. Klokkou, J. Gorecki, and V. Apostolopoulos, "Thz-tds parameter extraction via machine learning," in *2021 46th International Conference on Infrared, Millimeter and Terahertz Waves (IRMMW-THz)*, pp. 1–2, 2021.

[6] M. J. E. Golay, "A pneumatic infra-red detector," *Review of Scientific Instruments*, vol. 18, no. 5, pp. 357–362, 1947.

[7] G. N. Harding, M. F. Kimmitt, J. H. Ludlow, P. Porteous, A. C. Prior, and V. Roberts, "Emission of sub-millimetre electromagnetic radiation from hot plasma in zeta," *Proc. Phys. Soc. (London)*.

[8] S. Hargreaves and R. A. Lewis, "Terahertz imaging: materials and methods," *Journal of Materials Science: Materials in Electronics*, vol. 18, no. 1, p. 299, 2007.

[9] P. L. Richards, "High-resolution fourier transform spectroscopy in the far-infrared," *J. Opt. Soc. Am.*, vol. 54, pp. 1474–1484, Dec 1964.

[10] K. M. Evenson, D. A. Jennings, and F. R. Petersen, "Tunable far-infrared spectroscopy," *Applied Physics Letters*, vol. 44, no. 6, pp. 576–578, 1984.

[11] T. Y. Chang and T. J. Bridges, "Laser action at 452, 496, and 541 m in optically pumped ch3f," *Optics Communications*, vol. 1, no. 9, pp. 423–426, 1970.

[12] J. FAIST, F. CAPASSO, D. SIVCO, C. SIRTORI, A. HUTCHINSON, and A. CHO, "QUANTUM CASCADE LASER," *SCIENCE*, vol. 264, pp. 553–556, APR 22 1994.

[13] E. Bründermann, D. R. Chamberlin, and E. E. Haller, "High duty cycle and continuous terahertz emission from germanium," *Applied Physics Letters*, vol. 76, no. 21, pp. 2991–2993, 2000.

[14] P. R. Smith, D. H. Auston, and M. C. Nuss, "Subpicosecond Photoconducting Dipole Antennas," *IEEE Journal of Quantum Electronics*, vol. 24, no. 2, pp. 255–260, 1988.

[15] P. Y. Han, M. Tani, M. Usami, S. Kono, R. Kersting, and X.-C. Zhang, "A direct comparison between terahertz time-domain spectroscopy and far-infrared fourier transform spectroscopy," *Journal of Applied Physics*, vol. 89, no. 4, pp. 2357–2359, 2001.

[16] N. M. Burford and M. O. El-Shenawee, "Review of terahertz photoconductive antenna technology," *Optical Engineering*, vol. 56, no. 1, pp. 1 – 20, 2017.

[17] V. Apostolopoulos and M. E. Barnes, "THz emitters based on the photo-dember effect," *Journal of Physics D: Applied Physics*, vol. 47, p. 374002, aug 2014.

[18] D. H. Auston, K. P. Cheung, J. A. Valdmanis, and D. A. Kleinman, "Cherenkov radiation from femtosecond optical pulses in electro-optic media," *Phys. Rev. Lett.*, vol. 53, pp. 1555–1558, Oct 1984.

[19] A. Schneider, M. Neis, M. Stillhart, B. Ruiz, R. U. A. Khan, and P. Günter, "Generation of terahertz pulses through optical rectification in organic dast crystals: theory and experiment," *J. Opt. Soc. Am. B*, vol. 23, pp. 1822–1835, Sep 2006.

[20] T. Seifert, S. Jaiswal, U. Martens, J. Hannegan, L. Braun, P. Maldonado, F. Freimuth, A. Kronenberg, J. Henrizi, I. Radu, E. Beaurepaire, Y. Mokrousov, P. M. Oppeneer, M. Jourdan, G. Jakob, D. Turchinovich, L. M. Hayden, M. Wolf, M. Münzenberg, M. Kläui, and T. Kampfrath, "Efficient metallic spintronic emitters of ultrabroadband terahertz radiation," *Nature Photonics*, vol. 10, no. 7, pp. 483–488, 2016.

[21] E. P. J. Parrott, Y. Sun, and E. Pickwell-Macpherson, "Terahertz spectroscopy: Its future role in medical diagnoses," *Journal of Molecular Structure*, vol. 1006, no. 1-3, pp. 66–76, 2011.

[22] P. U. Jepsen, D. G. Cooke, and M. Koch, "Terahertz spectroscopy and imaging - Modern techniques and applications," *Laser and Photonics Reviews*, vol. 5, no. 1, pp. 124–166, 2011.

[23] S. Fan, B. S. Y. Ung, E. P. J. Parrott, V. P. Wallace, and E. Pickwell-Macpherson, "In vivo terahertz reflection imaging of human scars during and after the healing process," *Journal of Biophotonics*, vol. 9, pp. 1–9, 2016.

[24] T. Safonova, A. Angeluts, I. Ozheredov, A. Shkurinov, P. Solyankin, M. Prokopchuk, M. Mischenko, A. Balakin, and A. Larichev, " In vivo THz sensing of the cornea of the eye ," *Laser Physics Letters*, vol. 15, no. 5, p. 055601, 2018.

[25] C. Kang, D.-C. Shin, J.-H. Shin, J.-M. Park, S. Kim, J. W. Lee, G. B. Jung, Y.-S. Lee, and C.-S. Kee, "Terahertz spectroscopy of human sclera," *Current Applied Physics*, vol. 15, no. 10, pp. 1156–1159, 2015.

[26] G. Png, R. Flook, B.-H. Ng, and D. Abbott, "Terahertz spectroscopy of snap-frozen human brain tissue: an initial study," *Electronics Letters*, vol. 45, no. 7, p. 343, 2009.

[27] W.-G. Yeo, N. Srinivasan, O. Gurel, N. L. Lehman, S. Park, N. K. Nahar, P. D. King, and K. Sertel, "Terahertz Imaging and Electromagnetic Model of Axon Demyelination in Alzheimer's Disease," *IEEE Transactions on Terahertz Science and Technology*, vol. 7, no. 6, pp. 711–721, 2017.

[28] F. Palombo, E. Hendry, P. Hobson, R. S. Edginton, R. I. Stantchev, and J. C. Mansfield, "Subwavelength hyperspectral THz studies of articular cartilage," *Scientific Reports*, vol. 8, no. 1, pp. 1–8, 2018.

[29] B. Carl-Ivar and J. Tooze, *Introduction to protein structure*. Garland Pub., 2009.

[30] A. G. Markelz, A. Roitberg, and E. J. Heilweil, "Pulsed terahertz spectroscopy of DNA, bovine serum albumin and collagen between 0.1 and 2.0 THz," *Chemical Physics Letters*, vol. 320, no. 1-2, pp. 42–48, 2000.

[31] C. Zhang, E. Tarhan, A. K. Ramdas, A. M. Weiner, and S. M. Durbin, "Broadened far-infrared absorption spectra for hydrated and dehydrated myoglobin," *Journal of Physical Chemistry B*, vol. 108, no. 28, pp. 10077–10082, 2004.

[32] C. Ro/nne, L. Thrane, P.-O. Åstrand, A. Wallqvist, K. V. Mikkelsen, and S. R. Keiding, "Investigation of the temperature dependence of dielectric relaxation in liquid water by THz reflection spectroscopy and molecular dynamics simulation," *The Journal of Chemical Physics*, vol. 107, no. 14, pp. 5319–5331, 1997.

[33] W. Doster, S. Cusack, and W. Petry, "Dynamical transition of myoglobin revealed by inelastic neutron scattering," 1989.

[34] A. G. Markelz, J. R. Knab, J. Y. Chen, and Y. He, "Protein dynamical transition in terahertz dielectric response," *Chemical Physics Letters*, vol. 442, no. 4-6, pp. 413–417, 2007.

[35] Y. He, P. I. Ku, J. R. Knab, J. Y. Chen, and A. G. Markelz, "Protein dynamical transition does not require protein structure," *Physical Review Letters*, vol. 101, no. 17, pp. 2–5, 2008.

[36] S. Ebbinghaus, J. K. Seung, M. Heyden, X. Yu, U. Heugen, M. Gruebele, D. M. Leitner, and M. Havenith, "An extended dynamical hydration shell around proteins," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 52, pp. 20749–20752, 2007.

[37] B. Born, S. J. Kim, S. Ebbinghaus, M. Gruebelebc, and M. Havenith, "The terahertz dance of water with the proteins: the effect of protein flexibility on the dynamical hydration shell of ubiquitin," *Faraday Discussions*, vol. 141, p. 161, 2008.

[38] S. Ebbinghaus, J. K. Seung, M. Heyden, X. Yu, M. Gruebele, D. M. Leitner, and M. Havenith, "Protein sequence- and pH-dependent hydration probed by terahertz spectroscopy," *Journal of the American Chemical Society*, vol. 130, no. 8, pp. 2374–2375, 2008.

[39] B. Born, H. Weingärtner, E. Bründermann, and M. Havenith, "Solvation dynamics of model peptides probed by terahertz spectroscopy. observation of the onset of collective network motions," *Journal of the American Chemical Society*, vol. 131, no. 10, pp. 3752–3755, 2009.

[40] K. Meister, S. Ebbinghaus, Y. Xu, J. G. Duman, A. Devries, M. Gruebele, D. M. Leitner, and M. Havenith, "Long-range protein-water dynamics in hyperactive insect antifreeze proteins," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, no. 5, pp. 1617–1622, 2013.

[41] J. W. Bye, S. Meliga, D. Ferachou, G. Cinque, J. A. Zeitler, and R. J. Falconer, "Analysis of the hydration water around bovine serum albumin using terahertz coherent synchrotron radiation," *Journal of Physical Chemistry A*, vol. 118, no. 1, pp. 83–88, 2014.

[42] O. Sushko, R. Dubrovka, and R. S. Donnan, "Sub-terahertz spectroscopy reveals that proteins influence the properties of water at greater distances than previously detected," *Journal of Chemical Physics*, vol. 142, no. 5, 2015.

[43] Y. Sun, Y. Zhang, and E. Pickwell-MacPherson, "Investigating antibody interactions with a polar liquid using terahertz pulsed spectroscopy," *Biophysical Journal*, vol. 100, no. 1, pp. 225–231, 2011.

[44] Y. Sun, J. Zhong, C. Zhang, J. Zuo, and E. Pickwell-MacPherson, "Label-free detection and characterization of the binding of hemagglutinin protein and broadly neutralizing monoclonal antibodies using terahertz spectroscopy," *Journal of Biomedical Optics*, vol. 20, no. 3, p. 037006, 2015.

[45] M. Heyden, E. Bründermann, U. Heugen, G. Niehues, D. M. Leitner, and M. Havenith, "Long-range influence of carbohydrates on the solvation dynamics of water - Answers from terahertz absorption measurements and molecular modeling simulations," *Journal of the American Chemical Society*, vol. 130, no. 17, pp. 5773–5779, 2008.

[46] D. M. Leitner, M. Gruebele, and M. Havenith, "Solvation dynamics of biomolecules: Modeling and terahertz experiments," *HFSP Journal*, vol. 2, no. 6, pp. 314–323, 2008.

[47] A. S. Miller, B. Blott, and T. Hames, "Review of neural network applications in medical imaging and signal-processing," *Medical & Biological Engineering & Computing*, vol. 30, pp. 449–464, SEP 1992.

[48] D. E. Rumelhart and J. L. McClelland, *A General Framework for Parallel Distributed Processing*, pp. 45–76. 1987.

[49] L. A. Garcia and A. Shigidi, "Using neural networks for parameter estimation in ground water," *Journal of Hydrology*, vol. 318, no. 1, pp. 215–231, 2006.

[50] S. C. Raisinghani, A. K. Ghosh, and P. K. Kalra, "Two new techniques for aircraft parameter estimation using neural networks," *The Aeronautical Journal (1968)*, vol. 102, no. 1011, p. 25–30, 1998.

[51] M. R. Nowak, R. Zdunek, E. Pliński, P. Świątek, M. Strzelecka, W. Malinka, and S. Plińska, "Recognition of pharmacological bi-heterocyclic compounds by using terahertz time domain spectroscopy and chemometrics," *Sensors*, vol. 19, p. 3349, July 2019.

[52] C. Cao, Z. Zhang, X. Zhao, and T. Zhang, "Terahertz spectroscopy and machine learning algorithm for non-destructive evaluation of protein conformation," *Optical and Quantum Electronics*, vol. 52, Apr. 2020.

[53] A. I. Knyazkova, A. V. Borisov, L. V. Spirina, and Y. V. Kistenev, "Paraffin-embedded prostate cancer tissue grading using terahertz spectroscopy and machine learning," *Journal of Infrared, Millimeter, and Terahertz Waves*, Feb. 2020.

[54] Y. Sun, P. Du, X. Lu, P. Xie, Z. Qian, S. Fan, and Z. Zhu, "Quantitative characterization of bovine serum albumin thin-films using terahertz spectroscopy and machine learning methods," *Biomedical Optics Express*, vol. 9, p. 2917, June 2018.

[55] X. Sun, J. Liu, K. Zhu, J. Hu, X. Jiang, and Y. Liu, "Generalized regression neural network association with terahertz spectroscopy for quantitative analysis of benzoic acid additive in wheat flour," *Royal Society Open Science*, vol. 6, p. 190485, July 2019.

[56] I. Basheer and M. Hajmeer, "Artificial neural networks: fundamentals, computing, design, and application," *Journal of Microbiological Methods*, vol. 43, no. 1, pp. 3–31, 2000. Neural Computting in Micrbiology.

[57] Stinchcombe and White, "Universal approximation using feedforward networks with non-sigmoid hidden layer activation functions," in *International 1989 Joint Conference on Neural Networks*, pp. 613–617 vol.1, 1989.

[58] M. Hagan and M. Menhaj, "Training feedforward networks with the marquardt algorithm," *IEEE Transactions on Neural Networks*, vol. 5, no. 6, pp. 989–993, 1994.

[59] T. L. Paine, H. Jin, J. Yang, Z. L. Lin, and T. S. Huang, "Gpu asynchronous stochastic gradient descent to speed up neural network training," *CoRR*, vol. abs/1312.6186, 2014.

[60] R. Liu, M. He, R. Su, Y. Yu, W. Qi, and Z. He, "Insulin amyloid fibrillation studied by terahertz spectroscopy and other biophysical methods," *Biochemical and Biophysical Research Communications*, vol. 391, no. 1, pp. 862–867, 2010.

[61] J. Xu, K. W. Plaxco, and S. J. Allen, "Absorption spectra of liquid water and aqueous buffers between 0.3 and 3.72THz," *The Journal of Chemical Physics*, vol. 124, no. 3, p. 036101, 2006.

[62] E. Castro-Camus and M. B. Johnston, "Conformational changes of photoactive yellow protein monitored by terahertz spectroscopy," *Chemical Physics Letters*, vol. 455, no. 4-6, pp. 289–292, 2008.

[63] J. Xu, K. W. Plaxco, and S. J. Allen, "Collective dynamics of lysozyme in water: Terahertz absorption spectroscopy and comparison with theory," *Journal of Physical Chemistry B*, vol. 110, no. 47, pp. 24255–24259, 2006.

[64] S. Cheng and Z. Wu, "Microfluidic electronics," *Lab Chip*, vol. 12, pp. 2782–2791, 2012.

[65] Y. Qi, D. J. Rowe, V. Mittal, M. Banakar, Y. Wu, M. Nedeljkovic, J. S. Wilkinson, and G. Z. Mashanovich, "Integration of mid-infrared SOI photonics with microfluidics," in *Silicon Photonics XIV* (G. T. Reed and A. P. Knights, eds.), vol. 10923, pp. 47 – 55, International Society for Optics and Photonics, SPIE, 2019.

[66] V. Miralles, A. Huerre, F. Malloggi, and M.-C. Jullien, "A review of heating and temperature control in microfluidic systems: Techniques and applications," *Diagnostics*, vol. 3, no. 1, pp. 33–67, 2013.

[67] C.-Y. Lee, C.-L. Chang, Y.-N. Wang, and L.-M. Fu, "Microfluidic mixing: A review," *International Journal of Molecular Sciences*, vol. 12, no. 5, pp. 3263–3287, 2011.

[68] H. Yin and D. Marshall, "Microfluidics for single cell analysis," *Current opinion in biotechnology*, vol. 23, no. 1, pp. 110–119, 2012.

[69] R. N. Zare and S. Kim, "Microfluidic platforms for single-cell analysis," *Annual review of biomedical engineering*, vol. 12, pp. 187–201, 2010.

[70] X. Hu, G. Xu, L. Wen, H. Wang, Y. Zhao, Y. Zhang, D. R. Cumming, and Q. Chen, "Metamaterial absorber integrated microfluidic terahertz sensors," *Laser and Photonics Reviews*, vol. 10, no. 6, pp. 962–969, 2016.

[71] X. Zhao, M. Zhang, D. Wei, Y. Wang, S. Yan, M. Liu, X. Yang, K. Yang, H.-L. Cui, and W. Fu, "Label-free sensing of the binding state of MUC1 peptide and anti-MUC1 aptamer solution in fluidic chip by terahertz spectroscopy," *Biomedical Optics Express*, vol. 8, no. 10, p. 4427, 2017.

[72] S. Alfihed, M. H. Bergen, J. F. Holzman, and I. G. Foulds, "A detailed investigation on the terahertz absorption characteristics of polydimethylsiloxane (pdms)," *Polymer*, vol. 153, pp. 325–330, 2018.

[73] D. J. Rowe, A. Porch, D. A. Barrow, and C. J. Allender, "Microfluidic device for compositional analysis of solvent systems at microwave frequencies," *Sensors and Actuators, B: Chemical*, vol. 169, pp. 213–221, 2012.

[74] H. S. Salapare, F. Guittard, X. Noblin, E. Taffin de Givenchy, F. Celestini, and H. J. Ramos, "Stability of the hydrophilic and superhydrophobic properties of oxygen plasma-treated poly(tetrafluoroethylene) surfaces," *Journal of Colloid and Interface Science*, vol. 396, pp. 287–292, 2013.

[75] S. Zanini, R. Barni, R. D. Pergola, and C. Riccardi, "Modification of the PTFE wettability by oxygen plasma treatments: Influence of the operating parameters and investigation of the ageing behaviour," *Journal of Physics D: Applied Physics*, vol. 47, no. 32, 2014.

[76] J. Nakamatsu, L. F. Delgado-Aparicio, R. D. Silva, and F. Soberon, "Ageing of plasma-treated poly(tetrafluoroethylene) surfaces," *Journal of Adhesion Science and Technology*, vol. 13, no. 7, pp. 753–761, 1999.

[77] P. Lam, K. J. Wynne, and G. E. Wnek, "Surface-tension-confined microfluidics," *Langmuir*, vol. 18, no. 3, pp. 948–951, 2002.

[78] J. West, A. Michels, S. Kittel, P. Jacob, and J. Franzke, "Microplasma writing for surface-directed millifluidics," *Lab Chip*, vol. 7, pp. 981–983, 2007.

[79] S. Bouaidat, O. Hansen, H. Bruus, C. Berendsen, N. K. Bau-Madsen, P. Thomsen, A. Wolff, and J. Jonsmann, "Surface-directed capillary system; theory, experiments and applications," *Lab on a Chip*, vol. 5, no. 8, pp. 827–836, 2005.

[80] B. Zhao, J. S. Moore, and D. J. Beebe, "Surface-directed liquid flow inside microchannels," *Science*, vol. 291, no. 5506, pp. 1023–1026, 2001.

[81] N. Nguyen, S. Wereley, and S. Wereley, *Fundamentals and Applications of Microfluidics*. Artech House MEMS library, Artech House, 2002.

[82] A. B. D. Cassie and S. Baxter, "Wettability of porous surfaces," *Trans. Faraday Soc.*, vol. 40, pp. 546–551, 1944.

[83] R. N. Wenzel, "Resistance of solid surfaces to wetting by water," *Industrial and Engineering Chemistry*, vol. 28, no. 8, pp. 988–994, 1936.

[84] A. Marmur, "Wetting on hydrophobic rough surfaces: To be heterogeneous or not to be?," *Langmuir*, vol. 19, no. 20, pp. 8343–8348, 2003.

[85] W. Lai, H. Cao, J. Yang, G. Deng, Z. Yin, Q. Zhang, B. Pelaz, and P. del Pino, "Antireflection self-reference method based on ultrathin metallic nanofilms for improving terahertz reflection spectroscopy," *Opt. Express*, vol. 26, pp. 19470–19478, Jul 2018.

[86] S. Huang, P. C. Ashworth, K. W. C. Kan, Y. Chen, V. P. Wallace, Y. ting Zhang, and E. Pickwell-MacPherson, "Improved sample characterization in terahertz reflection imaging and spectroscopy," *Opt. Express*, vol. 17, pp. 3848–3854, Mar 2009.

[87] F. Vandrevala and E. Einarsson, "Decoupling substrate thickness and refractive index measurement in thz time-domain spectroscopy," *Opt. Express*, vol. 26, pp. 1697–1702, Jan 2018.

[88] A. Redo-Sanchez and X.-C. Zhang, "Self-referenced method for terahertz wave time-domain spectroscopy," *Opt. Lett.*, vol. 36, pp. 3308–3310, Sep 2011.

[89] L. Duvillaret, F. Garet, and J.-L. Coutaz, "Highly precise determination of optical constants and sample thickness in terahertz time-domain spectroscopy," *Appl. Opt.*, vol. 38, pp. 409–415, Jan 1999.

[90] P. R. Whelan, K. Iwaszczuk, R. Wang, S. Hofmann, P. Bøggild, and P. U. Jepsen, "Robust mapping of electrical properties of graphene from terahertz time-domain spectroscopy with timing jitter correction," *Opt. Express*, vol. 25, pp. 2725–2732, Feb 2017.

[91] J. Choi, W. S. Kwon, K.-S. Kim, and S. Kim, "Nondestructive material characterization in the terahertz band by selective extraction of sample-induced echo signals," *Journal of Nondestructive Evaluation*, vol. 34, no. 1, p. 269, 2014.

[92] F. Sanjuan, A. Bockelt, and B. Vidal, "Determination of refractive index and thickness of a multilayer structure with a single terahertz time domain spectroscopy measurement," *Appl. Opt.*, vol. 53, pp. 4910–4913, Aug 2014.

[93] N. Greenall, L. Li, E. Linfield, A. Davies, J. Cunningham, and A. Burnett, "Multilayer extraction of complex refractive index in broadband transmission terahertz time-domain spectroscopy," 07 2016.

[94] U. Tayvah, J. A. Spies, J. Neu, and C. A. Schmuttenmaer, "Nelly: A user-friendly and open-source implementation of tree-based complex refractive index analysis for terahertz spectroscopy," *Analytical Chemistry*, vol. 93, no. 32, pp. 11243–11250, 2021. PMID: 34355556.

[95] L. Duvillaret, F. Garet, and J.-L. Coutaz, "A reliable method for extraction of material parameters in terahertz time-domain spectroscopy," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 2, no. 3, pp. 739–746, 1996.

[96] M. C. Beard, G. M. Turner, and C. A. Schmuttenmaer, "Subpicosecond carrier dynamics in low-temperature grown GaAs as measured by time-resolved terahertz spectroscopy," *Journal of Applied Physics*, vol. 90, no. 12, pp. 5915–5923, 2001.

[97] D. Grischkowsky, S. Keiding, M. van Exter, and C. Fattinger, "Far-infrared time-domain spectroscopy with terahertz beams of dielectrics and semiconductors," *J. Opt. Soc. Am. B*, vol. 7, pp. 2006–2015, Oct 1990.

[98] M. Naftaly, "An international intercomparison of THz time-domain spectrometers," *International Conference on Infrared, Millimeter, and Terahertz Waves, IRMMW-THz*, vol. 2016-November, pp. 0–1, 2016.

[99] P. U. Jepsen, "Phase Retrieval in Terahertz Time-Domain Measurements: a "how to" Tutorial," *Journal of Infrared, Millimeter, and Terahertz Waves*, no. January, pp. 395–411, 2019.

[100] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks*, vol. 2, no. 5, pp. 359–366, 1989.

[101] R. Mitsuhashi, K. Murate, S. Niijima, T. Horiuchi, and K. Kawase, "Terahertz tag identifiable through shielding materials using machine learning," *Optics Express*, vol. 28, p. 3517, Jan. 2020.

[102] Siuly, X. Yin, S. Hadjiloucas, and Y. Zhang, "Classification of THz pulse signals using two-dimensional cross-correlation feature extraction and non-linear classifiers," *Computer Methods and Programs in Biomedicine*, vol. 127, pp. 64–82, Apr. 2016.

[103] X.-X. Yin, S. Hadjiloucas, and Y. Zhang, "Complex extreme learning machine applications in terahertz pulsed signals feature sets," *Computer Methods and Programs in Biomedicine*, vol. 117, pp. 387–403, Nov. 2014.

[104] X.-X. Yin, S. Hadjiloucas, J. He, Y. Zhang, Y. Wang, and D. Zhang, "Application of complex extreme learning machine to multiclass classification problems with high dimensionality: A THz spectra classification problem," *Digital Signal Processing*, vol. 40, pp. 40–52, May 2015.

[105] H. Luo, J. Zhu, W. Xu, and M. Cui, "Identification of soybean varieties by terahertz spectroscopy and integrated learning method," *Optik*, vol. 184, pp. 177–184, May 2019.

[106] A. Ren, A. Zahid, A. Zoha, S. A. Shah, M. A. Imran, A. Alomainy, and Q. H. Abbasi, "Machine learning driven approach towards the quality assessment of fresh fruits using non-invasive sensing," *IEEE Sensors Journal*, vol. 20, pp. 2075–2083, Feb. 2020.

[107] M. W. Ayech and D. Ziou, "Terahertz image segmentation using k -means clustering based on weighted feature learning and random pixel sampling," *Neurocomputing*, vol. 175, pp. 243–264, Jan. 2016.

[108] X.-X. Yin, Y. Zhang, J. Cao, J.-L. Wu, and S. Hadjiloucas, "Exploring the complementarity of THz pulse imaging and DCE-MRIs: Toward a unified multi-channel classification and a deep learning framework," *Computer Methods and Programs in Biomedicine*, vol. 137, pp. 87–114, Dec. 2016.

[109] W. Liu, R. Zhang, Y. Ling, H. Tang, R. She, G. Wei, X. Gong, and Y. Lu, "Automatic recognition of breast invasive ductal carcinoma based on terahertz spectroscopy with wavelet packet transform and machine learning," *Biomedical Optics Express*, vol. 11, p. 971, Jan. 2020.

[110] J. Shi, Y. Wang, T. Chen, D. Xu, H. Zhao, L. Chen, C. Yan, L. Tang, Y. He, H. Feng, and J. Yao, "Automatic evaluation of traumatic brain injury based on terahertz imaging with machine learning," *Optics Express*, vol. 26, p. 6371, Mar. 2018.

[111] X. Yang, H. Guo, N. Wang, B. Song, and X. Gao, "A novel symmetry driven siamese network for THz concealed object verification," *IEEE Transactions on Image Processing*, vol. 29, pp. 5447–5456, 2020.

[112] X. Sun, K. Zhu, J. Liu, J. Hu, X. Jiang, Y. Liu, and Z. Gong, "Terahertz spectroscopy determination of benzoic acid additive in wheat flour by machine learning," *Journal of Infrared, Millimeter, and Terahertz Waves*, vol. 40, pp. 466–475, Mar. 2019.

[113] A. Zahid, H. T. Abbas, A. Ren, A. Zoha, H. Heidari, S. A. Shah, M. A. Imran, A. Alomainy, and Q. H. Abbasi, "Machine learning driven non-invasive approach of water content estimation in living plant leaves using terahertz waves," *Plant Methods*, vol. 15, Nov. 2019.

[114] Y. Jiang, H. Ge, and Y. Zhang, "Quantitative analysis of wheat maltose by combined terahertz spectroscopy and imaging based on boosting ensemble learning," *Food Chemistry*, vol. 307, p. 125533, Mar. 2020.

[115] M. Luo, S. Zhong, L. Yao, W. Tu, W. Nsengiyumva, and W. Chen, "Thin thermally grown oxide thickness detection in thermal barrier coatings based on SWT-BP neural network algorithm and terahertz technology," *Applied Optics*, vol. 59, p. 4097, Apr. 2020.

[116] W. Tu, S. Zhong, Y. Shen, A. Incecik, and X. Fu, "Neural network-based hybrid signal processing approach for resolving thin marine protective coating by terahertz pulsed imaging," *Ocean Engineering*, vol. 173, pp. 58–67, Feb. 2019.

[117] M. Naftaly and R. E. Miles, "Terahertz time-domain spectroscopy of silicate glasses and the relationship to material properties," *Journal of Applied Physics*, vol. 102, no. 4, 2007.

[118] S. Wietzke, C. Jansen, M. Reuter, T. Jung, D. Kraft, S. Chatterjee, B. M. Fischer, and M. Koch, "Terahertz spectroscopy on polymers: A review of morphological studies," *Journal of Molecular Structure*, vol. 1006, no. 1-3, pp. 41–51, 2011.

[119] G. J. Wilmink, B. L. Ibey, T. Tongue, B. Schulkin, N. Laman, X. G. Peralta, C. C. Roth, C. Z. Cerna, B. D. Rivest, J. E. Grundt, and W. P. Roach, "Development of a compact terahertz time-domain spectrometer for the measurement of the optical properties of biological tissues," *Journal of Biomedical Optics*, vol. 16, no. 4, p. 047006, 2011.

[120] J. Dai, J. Zhang, W. Zhang, and D. Grischkowsky, "Terahertz time-domain spectroscopy characterization of the far-infrared absorption and index of refraction of high-resistivity, float-zone silicon," *Journal of the Optical Society of America B*, vol. 21, no. 7, p. 1379, 2004.

[121] M. Naftaly and R. E. Miles, "Terahertz time-domain spectroscopy for material characterization," *Proceedings of the IEEE*, vol. 95, no. 8, pp. 1658–1665, 2007.

[122] L. Duvillaret, F. Garet, J. F. Roux, and J. L. Coutaz, "Analytical modeling and optimization of terahertz time-domain spectroscopy experiments using photoswitches as antennas," *IEEE Journal on Selected Topics in Quantum Electronics*, vol. 7, no. 4, pp. 615–623, 2001.

[123] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[124] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, pp. 1–15, 2015.

[125] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," 2015.

[126] P. A. George, W. Hui, F. Rana, B. G. Hawkins, A. E. Smith, and B. J. Kirby, "Microfluidic devices for terahertz spectroscopy of biomolecules," *Optics Express*, vol. 16, no. 3, p. 1577, 2008.

[127] L. Liu, R. Pathak, L. J. Cheng, and T. Wang, "Real-time frequency-domain terahertz sensing and imaging of isopropyl alcohol-water mixtures on a microfluidic chip," *Sensors and Actuators, B: Chemical*, vol. 184, pp. 228–234, 2013.

[128] S. Nagatomo, K. Saito, K. Yamamoto, T. Ogura, T. Kitagawa, and M. Nagai, "Heterogeneity between Two α Subunits of $\alpha 2\beta 2$ Human Hemoglobin and O2 Binding Properties: Raman, 1H Nuclear Magnetic Resonance, and Terahertz Spectra," *Biochemistry*, vol. 56, no. 46, pp. 6126–6136, 2017.

[129] L. Latypova, G. Barshtein, A. Puzenko, Y. Poluektov, A. Anashkina, I. Petrushanko, S. Fenk, A. Bogdanova, and Y. Feldman, "Oxygenation state of hemoglobin defines dynamics of water molecules in its vicinity," *Journal of Chemical Physics*, vol. 153, no. 13, 2020.

[130] S. Longeville and L. R. Stingaciu, "Hemoglobin diffusion and the dynamics of oxygen capture by red blood cells," *Scientific Reports*, vol. 7, no. 1, pp. 1–10, 2017.

[131] K. Z. Chng, Y. C. Ng, B. Namgung, J. K. S. Tan, S. Park, S. L. Tien, H. L. Leo, and S. Kim, "Assessment of transient changes in oxygen diffusion of single red blood cells using a microfluidic analytical platform," *Communications Biology*, vol. 4, no. 1, pp. 1–10, 2021.

[132] B. G. Forget and H. Franklin Bunn, "Classification of the disorders of hemoglobin," *Cold Spring Harbor Perspectives in Medicine*, vol. 3, no. 2, pp. 1–12, 2013.

[133] T. Frisk, D. Rönnholm, W. Van Der Wijngaart, and G. Stemme, "A micromachined interface for airborne sample-to-liquid transfer and its application in a biosensor system," *Lab on a Chip*, vol. 6, no. 12, pp. 1504–1509, 2006.

[134] J. D. Greenwood, Y. Liu, D. E. Busacker, D. Cheng, and H. Jiang, "Collection of gaseous and aerosolized samples using microfluidic devices with Gas-liquid interfaces," *IEEE Sensors Journal*, vol. 10, no. 5, pp. 952–959, 2010.

[135] H. Bruus, *Theoretical Microfluidics*, vol. 18. Oxford university press Oxford, 01 2008.

[136] A. Hibara, S. Iwayama, S. Matsuoka, M. Ueno, Y. Kikutani, M. Tokeshi, and T. Kitamori, "Surface modification method of microchannels for gas-liquid two-phase flow in microchips," *Analytical Chemistry*, vol. 77, no. 3, pp. 943–947, 2005.

[137] S. Yadav, S. Shire, and D. Kalonia, "Viscosity analysis of high concentration bovine serum albumin aqueous solutions," *Pharmaceutical research*, vol. 28, pp. 1973–83, 04 2011.

[138] S. McClellan and E. Franses, "Effect of concentration and denaturation on adsorption and surface tension of bovine serum albumin," *Colloids and Surfaces B: Biointerfaces*, vol. 28, pp. 63–75, 04 2003.

[139] D. J. Wilson, R. L. Williams, and R. C. Pond, "Plasma modification of PTFE surfaces - Part I: Surfaces immediately following plasma treatment," *Surface and Interface Analysis*, vol. 31, no. 5, pp. 385–396, 2001.

[140] J. Ryu, K. Kim, J. Park, B. G. Hwang, Y. Ko, H. Kim, J. Han, E. Seo, Y. Park, and S. J. Lee, "Nearly Perfect Durable Superhydrophobic Surfaces Fabricated by a Simple One-Step Plasma Treatment," *Scientific Reports*, vol. 7, no. 1, p. 1981, 2017.

[141] D. J. Wilson, R. L. Williams, and R. C. Pond, "Plasma modification of PTFE surfaces - Part II: Plasma-treated surfaces following storage in air or PBS," *Surface and Interface Analysis*, vol. 31, no. 5, pp. 397–408, 2001.