Extracting the Refractive Index from THz measurements

Patric Boardman

September 2022

Abstract

This paper outlines the method that I developed to extract the value of the complex refractive index from a time domain based terrahertz measurement of a thin sample of material. The code was designed to batch import data taken throughout an experimental session, and, for each measurement, extract the complex value and thereafter produce a graph of how the refractive index changed during the course of the experiment. This code formed part of an overall project in which the specific aim was to quantify the amount of water within a tissue sample by measuring the index. However it can also be applied more generally to measure the refractive index of any piece of material on the order of microns.

1 Introduction

Terrahertz (THz) waves are electromagnetic waves that are located in the far infrared region of the electromagnetic spectrum, with wavelengths typically ranging from $\lambda \sim 0.1$ - 1mm. They are produced by an ultrashort laser pulse via a physical process known as non-linear optical rectification. [1] In order to measure the properties of a material, an ultrafast detector measures the amplitude of the pulse throughout its duration, which typically lasts on the order of 10^{-12} seconds. The amplitude of the detected pulse will vary depending on its path from the emitter to the detector; for example if the pulse is unimpeded (i.e travelling through the air), a larger peak will be detected than if it has to pass through a dielectric material. This means that we can place a sample of a certain thickness in the path of the pulse, and by comparing it to the reference pulse, the refractive index of the sample can be extracted.

There are a couple of further complexities. Firstly, the refractive index is dependent on the frequency of the probing radiation (in this case, the THz pulse, which covers a range of frequencies), which means that the code has to take this into account and produce the correct value of the index at a chosen frequency. Secondly, the refractive index has to be a complex number. [2] This is because as it has to account for absorption of the radiation, and the complex part of the index, referred to as the *extinction coefficient*, physically represents the amount of absorption occurring within the sample. In this document I outline the methodology to extract the refractive index for batch THz measurements from the raw data taken periodically over the course an experiment, to a plot of the frequency profile of the index for each measurement of the experiment.

2 Theory and Methodology

The fundamental approach of this analysis relies on comparing the (processed) data to an equation, which will be derived in Section 2.1. This equation has a variety of parameters, one of which includes the refractive index. The code will iterate through a range of refractive index values, and find the point where the equation and data have the least difference, which will most likely correspond to a solution. Since the refractive index has a real part and an imaginary part, we will therefore be minimising the data to two parameters of the equation. The code will then repeat this exact process across the entire frequency range, in order to produce the frequency dependence of the refractive index.

2.1 Derivation of the Equation

The equation in which the data will be minimized is, in essence, a ratio between the measured electric field strength E of the measurements; one measurement with the sample, and one without. This ratio represents the *transmission* of the THz pulse through the sample, and can be any value between 0 (complete absorption or reflection), or 1 (perfect transmission). For a general pulse, the transmitted electric field E_{Tr} is the original incident electric field E_{In} multiplied by the transmission through the sample, t: [3]

$$E_{Tr} = t(\tilde{n}, \omega, L)E_{In}. \tag{1}$$

This transmission has three dependencies: the complex refractive index $\tilde{n} = n + i\kappa$, the frequency of the radiation (given here as angular frequency) $\omega = 2\pi f$, and the thickness of the sample L. These are related by

$$t(n, \kappa, \omega, L) = \frac{t_{12}t_{23}e^{i\frac{(n+i\kappa)\omega L}{c}}}{1 + r_{12}r_{23}e^{2i\frac{(n+i\kappa)\omega L}{c}}}. \tag{2}$$

Here, $t_{ij} = 2n_i/(n_i + n_j)$ and $r_{ij} = (n_i - n_j)/(n_i + n_j)$ are the Fresnel transmission and reflection coefficients through each boundary. A diagram of this can be shown in Figure 1

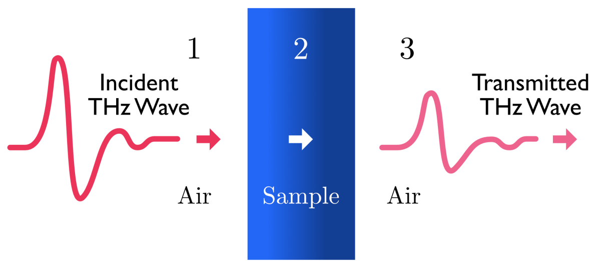


Figure 1: THz transmission from air (1), through a sample (2) to air on the other side (3).

Note that for this analysis, we take air to be completely transparent, hence $\tilde{n}_1 = \tilde{n}_3 = 1 + 0i$. For the reference THz pulse, all three media are air. However, the pulse will still accumulate phase propagating through the same space as the sample. To account for this phase shift, we multiply Equation (2) by a factor of $e^{-i\omega L/C}$, which results in subtracting the term in the exponential in the numerator by 1.

We use this, along with simplifying the Fresnel coefficients to rewrite Equation 2 as

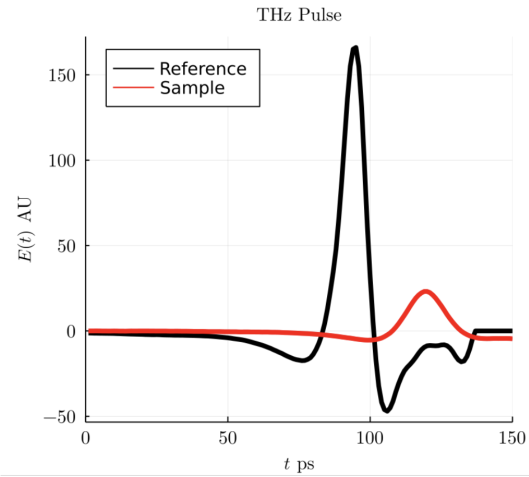
$$t(n, \kappa, \omega, L) = \frac{4(n + i\kappa)e^{i\frac{(n+i\kappa-1)\omega L}{c}}}{(1 + (n + i\kappa))^2 - (1 - (n + i\kappa))^2 e^{2i\frac{(n+i\kappa)\omega L}{c}}}. \tag{3}$$

We note that in order to extract values for n and κ, the frequency $\omega = 2\pi f$ and the thickness L must be treated as constants such that a solution can be found.

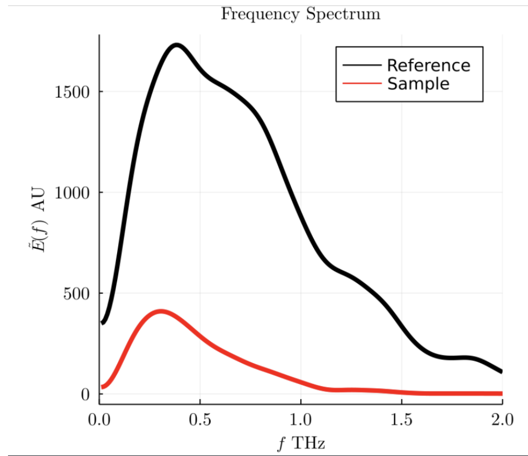
2.2 Processing the Raw Data

The code must be able to repeat the minimisation procedure to Equation 3 for a range of frequencies, and for multiple THz measurements at once. This meant that it had to be engineered to iterate on large scales, therefore performance became a key factor for consideration. I wrote the code in `Julia`, a scientific programming language which had comparatively good performance benchmarks [4]. This document will show plots all originating from the same experiment, carried out on the 29th April. There were two scripts used to carry out this analysis: `DataSorter.jl` and `BulkExtractor.jl`. The majority of the data processing and extraction process was done in `BulkExtractor.jl`. However the script `DataSorter.jl` was used to collate the data such that it was ready for batch processing.

The first stage of this analysis was to collate all of the data into a single `.csv` file. This was accomplished by the script `DataSorter.jl`, which took a working directory with ordered data files, and imported the data. The data was collated into a single DataFrame, with each column being a measurement. This was done for all reference measurements, and all sample measurements, such that there were two files for export onto a `.csv` file. `BulkExtractor.jl` initially imported the raw THz data for both the reference measurements and the sample measurements, whilst also defining several experimental constants. Since the sample was drying out over time, measurements were made at regular intervals. In order to get the data in the frequency domain for analysis, the code then performed a Fast Fourier Transform (FFT) on each column. To improve the resolution of the FFT, the code added a certain number of zeros on the end of each column, a process known as zero padding. An initial plot of one particular measurement in both the time and frequency domain is shown in Figure 2.



(a) Plot of the electric field measurement of the THz pulses in the time domain. Black: Reference, Red: Sample.



(b) Modulus of the FFT of the electric field measurement of the THz pulses in the frequency domain. Black: Reference, Red: Sample.

Figure 2: Results of initial data extraction for a particular measurement taken on the 29th April. Note that this plot isn't necessary for the refractive index extraction process, it is for the purpose of illustrating the process.

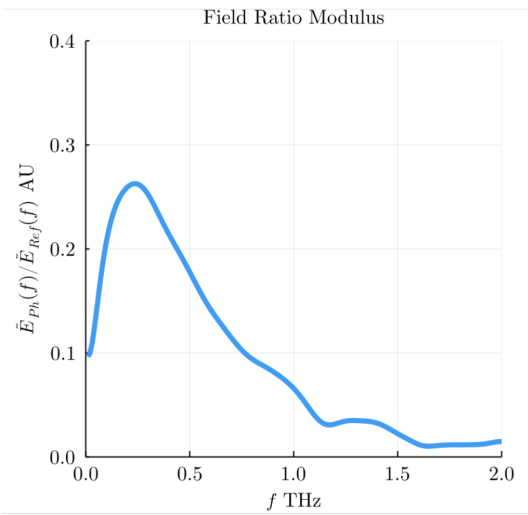
We then took the ratio between the sample and the reference for each frequency step. Note that when performing the FFT, a complex result is obtained with both a real and imaginary part. For our purposes, we are most interested in the magnitude and the phase of each of the complex data points in the set. By convention, we ensure that the phase of subsequent data points is increasing rather than decreasing, which `BulkExtractor.jl` does by reversing the sign of the complex angle. The absolute value of this ratio physically represents the transmission through the sample as a function of frequency. A plot of both the absolute value of the ratio along with the phase is given in Figure 3.

It is the data shown in Figure 3*a* which we will attempt to fit to Equation (3), for extraction of the values of n and κ.

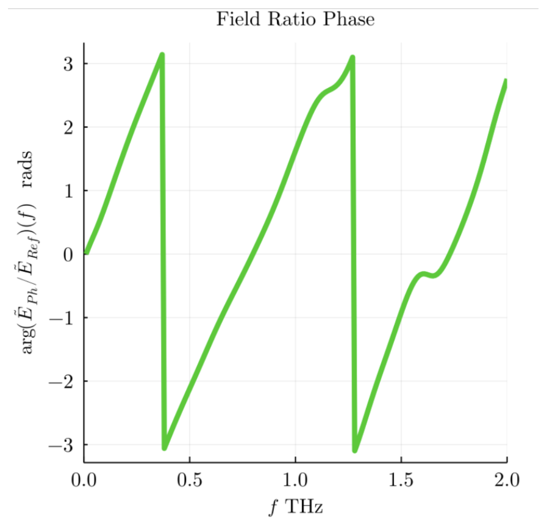
2.3 Solution Extraction

In order to extract a single solution, we must first select a single frequency value within the range and input the thickness of the sample. For a given measurement, we then choose that data point that corresponds to that particular frequency. Using the known value of the thickness and the chosen frequency, we evaluate Equation (3) for a range of n and κ, and store the results in a 2D complex array. The boundaries of these ranges are upper and lower guesses and must be manually chosen. For our experiments, we were unlikely to be analysing samples with a real part of the refractive index being above 5, and an imaginary part being above 4. Since n and κ cannot physically be lower than 1 and 0 respectively, this gave us a suitable guess range. We then subtract each value in the array, from the complex value of the data point at that frequency. A 2D plot of this is given in Figure 4.

The region of this plot where the value is zero corresponds to the solution. However, it is obvious from Figure 4 that there are two distinct regions in which this difference approaches zero, implying that there are multiple refractive index solutions. Of course, physically, this is not the case as any material can only have a single value for the index. The reason for the oscillatory behaviour in Figure 4 is due to the fact that there can be multiple values which all correspond to the same phase angle, and since n is directly linked to the phase, the plot will show all values of n which have the same phase solution.



(a) Plot of the absolute value of the Fourier transformed ratio between the sample and the reference pulses. This physically represents the transmission sepctrum through the sample.



(b) Plot of the phase of the Fourier transformed ratio between the sample and reference pulses. Due to convention, we define the phase to be positively accumulating.

Figure 3: Plots of the Fourier transformed field ratio between the sample and the reference pulses for the same data point taken on 29th April.

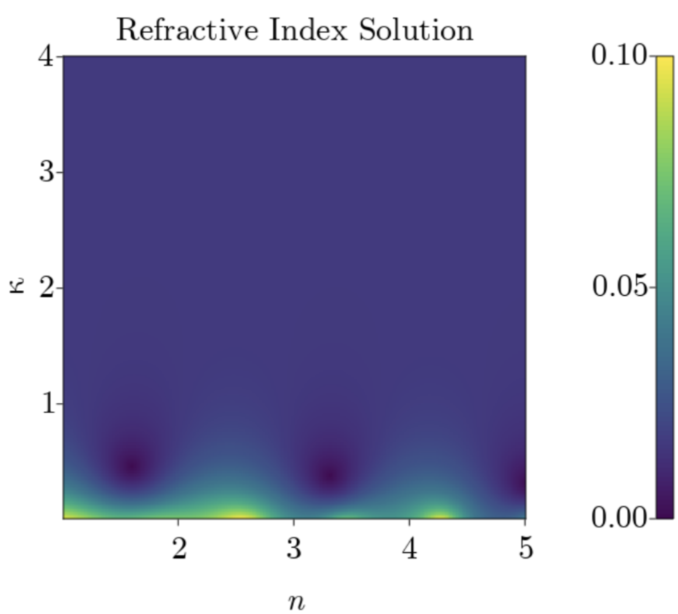
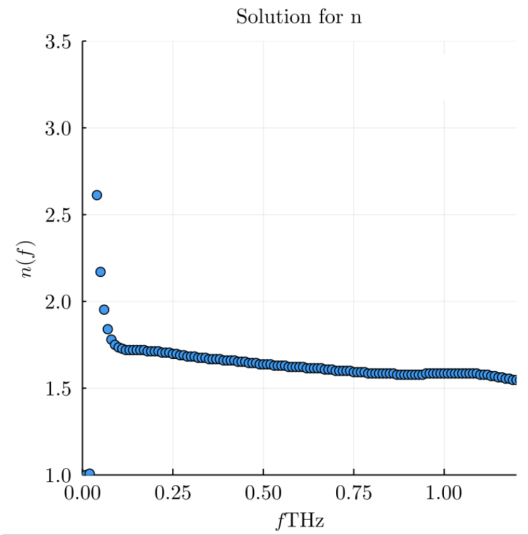
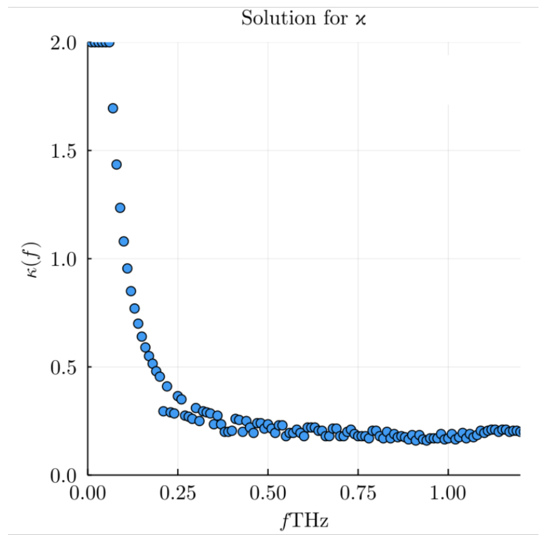


Figure 4: Plot of the difference between the 500GHz datapoint and the function, across a range of n and κ. The minima of this plot correspond to the solutions of the refractive index at this frequency, from this measurement.

In order to obtain the correct solution, a guess of the solution at a certain frequency must be made. This must be manually input into `BulkExtractor.jl`. With this initial guess known, the code was made to select the solution that was the closest to the initial guess. This meant, for each frequency, all solutions must be located, and the solution that was closest to the previous one was chosen. This approach ensured that the dependence of n on f was continuous, since it wouldn't make physical sense for there to be discontinuities. With this implemented, I was able to extract the frequency dependence for a particular measurement. A plot of the solutions for n and κ for the same measurement is shown in Figure 5.



(a) Frequency profile of the real part for the 29th April measurement.



(b) Frequency profile of the imaginary part for the 29th April measurement.

Figure 5: Frequency profiles for both components of the refractive index, plotted within the optimal frequency range of the THz pulse.

With the frequency profile obtained for each measurement, we continued to batch process and extract the profile for each measurement, which was later implemented in `BulkExtractor.jl`. The next stage of the project was to fit this data to various physical models for the frequency dependence of soft materials. This process is carried out within the scripts `EffectiveMediumFitter.jl` and `ExtractedModelParameterPlots.jl`, and from these scripts we can estimate the water content within these samples. For the way in which these scripts work, the documentation can be found here. However `BulkExtractor.jl` could more generally be applied to any thin sample to find its refractive index.

3 Performance and Code Limitations

3.1 Performance Time

With a typical experiment containing up to 15 measurements, execution speed became more important to consider. The primary factor which had the most influence on performance was the number of steps in the guess range for n and κ. This was because the code had to calculate the difference between each datapoint at a certain frequency, and the evaluation of Equation (3) throughout the entire n-κ two-space. Therefore, doubling the number of iterations would lead to a four-fold increase in performance time.

3.2 Thin Film Limit

The general transmission equation given by Equation (2) is only valid under the condition that the thickness of the film is small in compare to the wavelength of the radiation. With frequencies of the probing THz pulse typically ranging from 0.05-1.5THz (shown explicitly in Figure 2b), the corresponding wavelengths of this radiation are 0.2-6 mm. Therefore, if the sample thickness is too thick (on the order of mm), then solutions to the refractive index will become too many and therefore difficult to distinguish. This analysis works best if there are only one or two solutions within the estimation domain for n and κ, else it is highly likely that the code will automatically latch onto a non physical solution.

3.3 Thickness Measurement

Due to a particular measurement being directly proportional to the product of thickness, frequency and refractive index (from Equation (3)), an accurate measurement of the thickness of the sample is imperative, such that the correct index value can be extracted. For a solid sample, this can be achieved to a high degree of accuracy using standard laboratory equipment. However, the experiments that I was dealing with were using soft materials which, due to their water content, dried out over time, affecting the thickness. This created some experimental difficulty with the particular project. However this would largely not need to be considered if this code were to be applied to any solid samples.

4 Acknowledgements

I would like to acknowledge my supervisor, Professor Euan Hendry from the University of Exeter physics department, for the exceptional amount of assistance given to me on this project and the co-development of this code. I am very grateful for the assistance I was given whilst working on this project. The experience I gained during that year I have no doubt will provide invaluable for future endeavors. I would also like to acknowledge Dr. Sonal Saxenna, a post-doctorate who was working closely alongside me on this project. Many of the measurements in this experiment were conducted by her, without which this analysis would not be possible.

References

[1] X.-C. Zhang and J. Xu, "Generation and detection of thz waves," in *Introduction to THz wave photonics*, pp. 27–48, Springer, 2010.

[2] D. J. Segelstein, *The complex refractive index of water*. PhD thesis, University of Missouri–Kansas City, 1981.

[3] R. Ulbricht, E. Hendry, J. Shan, T. F. Heinz, and M. Bonn, "Carrier dynamics in semiconductors studied with time-resolved terahertz spectroscopy," *Reviews of Modern Physics*, vol. 83, no. 2, p. 543, 2011.

[4] "Julia micro-benchmarks." `https://julialang.org/benchmarks`. Accessed: 2022-05-08.